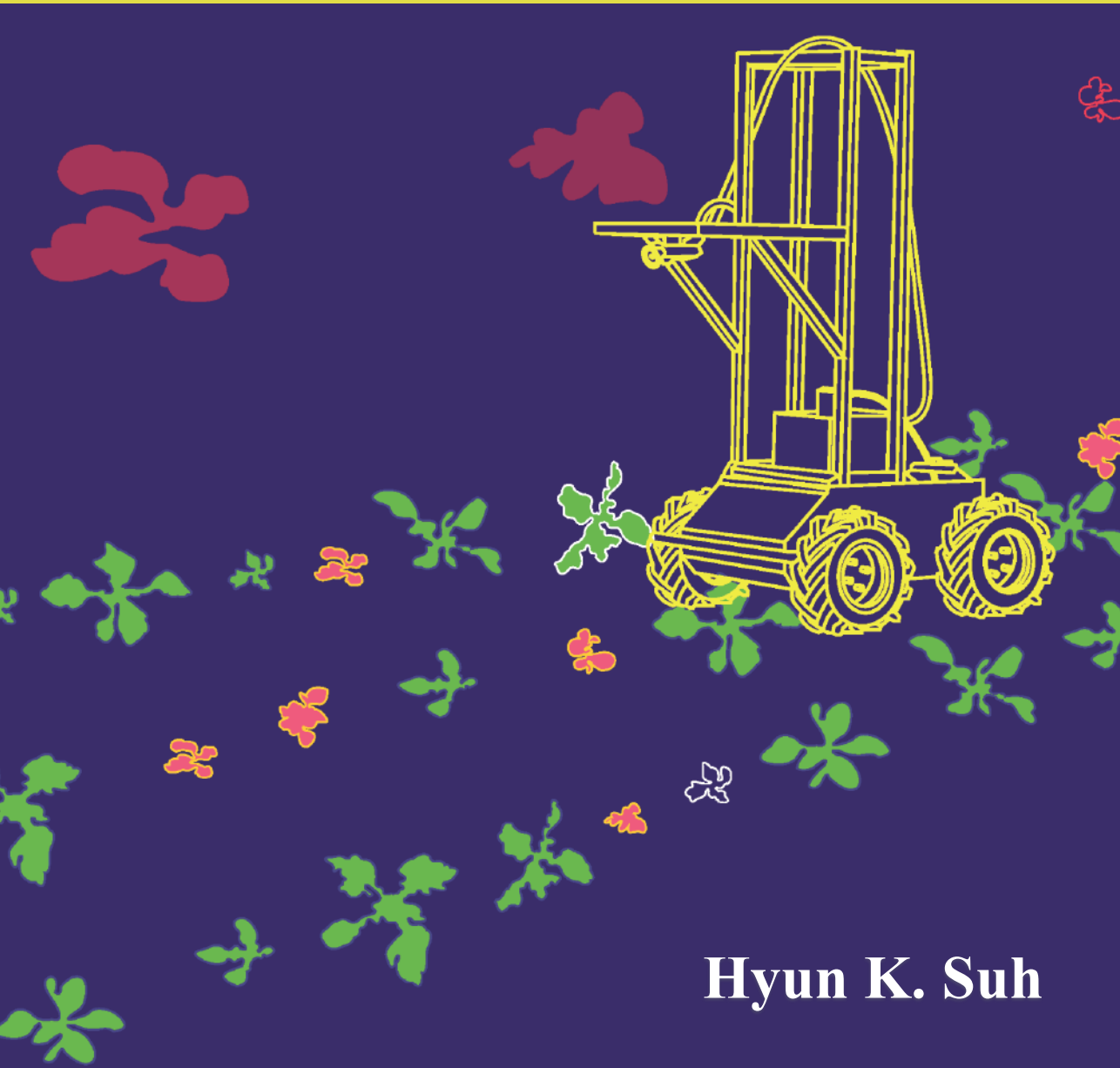
Advanced classification of volunteer potato in a sugar beet field



Hyun K. Suh

Advanced classification of volunteer potato in a sugar beet field

Hyun K. Suh

Thesis committee

Promotor
Prof. Dr E.J. van Henten
Professor of Farm Technology
Wageningen University & Research

Co-promotors
Dr J.W. Hofstee
Lecturer, Farm Technology Group
Wageningen University & Research

Dr J.M.M. IJsselmuiden
CEO of Track32

Other members
Prof. Dr D. de Ridder, Wageningen University & Research
Prof. Dr T. Gevers, University of Amsterdam
Prof. Dr W. Saeys, KU Leuven, Belgium
Dr F.K. van Evert, Wageningen University & Research

This research was conducted under the auspices of the Graduate School of Production Ecology & Resource Conservation (PE&RC).

Advanced classification of volunteer potato in a sugar beet field

Hyun K. Suh

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 19 September 2018
at 4 p.m. in the Aula.

Hyun K. Suh
Advanced classification of volunteer potato in a sugar beet field,
192 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summary in English

ISBN: 978-94-6343-791-2
DOI: https://doi.org/10.18174/452437

Propositions

1. Deep learning is such a powerful technology that it will be the basis for all future weed control systems.
   (this thesis)

2. For weed control to be really successful, weed classification should be accompanied by confidence values of the estimates.
   (this thesis)

3. Nature is the source of answers to all the questions in science.

4. For automotive applications, the hydrogen fuel cell offers a far more sustainable solution than the electric energy stored in Li-ion batteries.

5. Bram Peper's report has had most influence on the Dutch agricultural sector in the past one and a half decade.
   ("Duurzame Kennis, Duurzame Landbouw - Een advies aan de Minister van Landbouw, Natuurbeheer en Visserij over de kennisinfrastructuur van de landbouw in 2010")

6. Computer programming should be a compulsory course starting at the first classes in primary school.

7. In legislation, smoking nearby children should be treated as an act of violence.

Propositions belonging to the thesis, entitled:
"Advanced classification of volunteer potato in a sugar beet field"

Hyun K. Suh
Wageningen, 19-September-2018

Table of Contents

CHAPTER 1

General Introduction

1.1 A brief history of weed control

When biblical Adam and Eve were banished from the Garden of Eden, they were promised to have thorns and thistles (Genesis 3:18). Perhaps from then on, weeds (thorns and thistles) have troubled humanity and made mankind fight against them in crop productions (Young & Pierce, 2014). The earliest known weed control method was, according to the drawings made in ancient Egypt and Mesopotamia (6000 BC), hand-weeding which merely implied pulling weeds by hand or cutting plants out with a knife and with hoes (Timmons, 2005). Since then, hand-weeding has long been a primary means of "technology" for weed control (Bell, 2015). In due course of time, however, the weeding technology was gradually improved, thanks to the advancement of the weeding tools as well as the use of animal power, by utilising T-shaped wooden implements, V-shaped tools tipped with bronze and A-shaped logs with pegs which were often pulled by animals such as cattle or horses (Smith & Frederiksen, 2000). Only in the mid-19th century, animal-power used for ploughing was replaced by steam-engine which was ultimately substituted by combustion engines in the 1930s.

Chemicals have also been used for weed control for quite a long time. The very first known chemical treatment for weed control was to use a rock salt as was done by the Romans during the 1st century BC (Zimdahl, 2013). Since then, salt had

long been used as a "weed-killer" because salt was known to destroy plant life quite well (Dreiling, 2017). Only in the beginning of the 20th century, the modern ways of chemical treatment for weed control have begun with the introduction of effective synthetic herbicides. Chemical weed control has gradually become a viable alternative for laborious mechanical weed removal (Kelton & Price, 2011). In addition, the application of herbicides using a tractor-pulled spraying system significantly reduced the labour burdens associated with the weeding. Nowadays with the advent of automation and precision agriculture technologies, site-specific weed management is becoming feasible. In site-specific weed management, i.e. differentiated application of herbicides on an individual plant level provides an efficient way to minimise herbicide costs and environmental impact.

1.2 A specific case: volunteer potato control in sugar beet

Potato and sugar beet are major crops grown in the Netherlands (IRS, 2005). Potato is frequently rotated followed by sugar beet because a proper sequence of crops is beneficial for crop production. However, some of the potato tubers that remain in the field after harvest may survive a mild winter and will emerge in the next crop (sugar beet) during the following spring and summer or even the year after. These emerged potatoes are known as volunteer potatoes and are considered to be a weed.

Volunteer potato is a major problem in sugar beet production in the Netherlands. Not only because volunteer potato may overgrow the sugar beet (Rahman, 1980) but also volunteer potato competes with sugar beet for water, nutrients and space in the field, which in most cases lead to a loss of crop yield (Boydston & Seymour, 2002; Nieuwenhuizen et al., 2010). In a sugar beet field, for example, five volunteer potato plants per square meter can lead to a loss of sugar beet yield of up to 16.5 t/ha (MacEwan et al., 2017). Moreover, volunteer potato provides a hideout for harmful diseases such as nematodes and pests, and is also known as the point source for the spread of one of the most notorious potato diseases, called late blight, caused by *Phytophthora infestans*. This is a major threat to potato production in north-western Europe and is one of the most devastating plant pathogens in agriculture (Cooke et al., 2011; Moushib et al., 2013).

These are unwanted effects, and therefore adequate control of volunteer potato is critical. This is stressed by a statutory obligation in the Netherlands under which farmers have to remove volunteer potato plants from their fields before the 1st of July in the growing season, to a maximum level of two remaining plants per square meter (Kienhuis & Berge, 2003).

For the control of volunteer potato, Dutch farmers generally apply glyphosate to the weeds, and this chemical application is mainly done manually and selectively. Such manual application is not the desired solution for weed control in the field as it is labour-intensive and time-consuming which typically comes with high labour cost. On top of the manual application, mechanized methods are also carried out in practice but are less effective, like dipping the volunteer potato plants with glyphosate using a pass of a roll or a stick by their higher height as compared to the crop (Boonman, 2013). For a proper control in practice, each stem of the volunteer potato plant has to be handled with glyphosate to destroy it completely.

As manual control was becoming too expensive and yielded incomplete control, an automated system for detection and control of volunteer potato was developed by Nieuwenhuizen et al. (2010) (Figure 1.1a). This tractor-pulled system was equipped with machine vision device inside the hood and with a weeding actuator on the back. Once volunteer potato plants were detected and identified by machine vision, a micro sprayer deposited a droplet of 3.2 µL for the selective control of volunteer potato. This system demonstrated the potential of automated and plant-specific weed control with a vision-based approach for an agricultural field application. However, with a success rate of 83% the system did not reach the required 95% control of volunteer potato plants. The system yielded an unsatisfactory result in field conditions because the classification between sugar beet and volunteer potato was mainly based on colour features. Besides, due to the use of colour as a primary discriminative feature, artificial lighting (five xenon lamps) was needed to attain a constant and sufficient level of illumination under the hood which required a large amount of energy (Figure 1.1b). The whole system was designed to be mounted behind a tractor, which was a limiting factor as well. The use of a (human-driven) tractor for weed control may bring about other issues such as soil compaction, environmental pollution, additional fuel consumption and driver fatigue in a case for human-driven tractor (Stemp, 2005).

These issues can be minimised or avoided with a small-sized and lightweight autonomous system because such a system may have less environmental impact with

lower usage of energy as well as less soil compaction. Additionally, such a system may have less critical safety issues than heavy and large agricultural machinery (Pedersen et al., 2006; Sørensen et al., 2010; Van Henten et al., 2009).

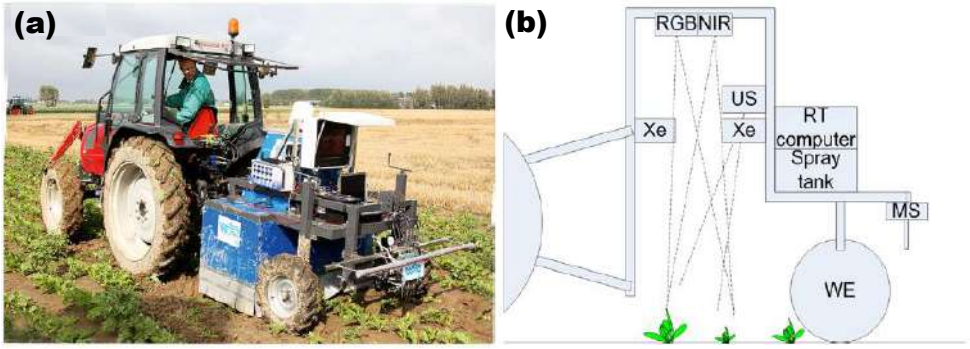


Figure 1.1: (a) Automated system for detection and removal of volunteer potato developed by Nieuwenhuizen et al. (2010). (b) Schematic drawing of measurement setup inside of the system. Two cameras and five xenon work lamps (XE) are located under the blue cover with grey plastic flaps. Two cameras (RGB, NIR), ultrasonic sensors (US), the micro sprayer (MS) and wheel encoder (WE) are connected on the system.

1.3 The EU SmartBot project: a small-sized robot for volunteer potato control

In 2011, the EU SmartBot project[1], a cross-border collaboration project which involved 24 different partners from Germany and the Netherlands, was initiated to develop a robotic system for several applications including for agricultural use. In AgroBot, part of the SmartBot project, a small-sized and vision-based autonomous weed control system was to be developed for effective control of volunteer potato plants in a sugar beet field. As a robotic platform, the Clearpath Husky A200 UGV (Unmanned Ground Vehicle) was to be used in this project (Figure 1.2). The robot was to be equipped with a camera on the front and a weeding actuator on the back.

[1]Interreg IVa, European Fund for the Regional Development of the European Union and Product Board for Arable Farming.

Due to the reduced carrying capacity of the robotic platform (Husky), additional infrastructure like a hood was not a viable option. Moreover, artificial lighting was not considered feasible either because the mobile platform was battery operated. Thus, the system should be able to perform robustly in scenes that are fully exposed to ambient lighting conditions.

Requirements

Within the context of the SmartBot weeding application, the following requirements were set, similar to those of the previous study of Nieuwenhuizen (2009). The resulting automatic weeding system should:

- effectively control more than 95% of the volunteer potato;
- ensure less than 5% of undesired control of sugar beet plants;
- ensure a classification time of less than 1 second per field image for real-time operation in the field.

The overall control accuracy depends both on the accuracy of the classification and the accuracy of the weed control (actuation) device. In a real-life situation, it is questionable that the actuation device would perform with 100% success in weed control. Therefore, to achieve in the end the required volunteer control of 95% or more, the classification accuracy should be considerably higher than 95%.

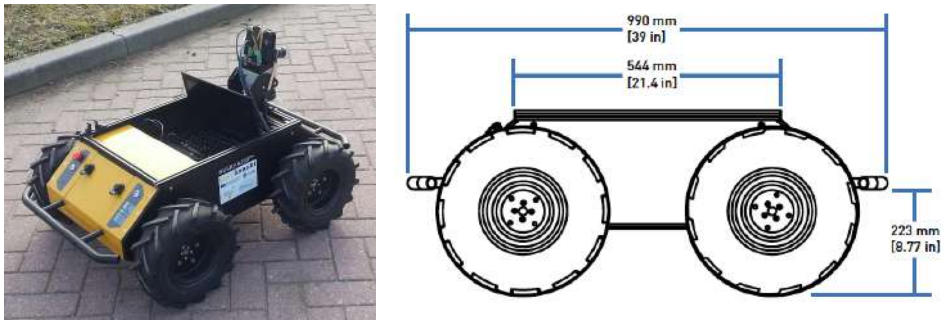


Figure 1.2: The robotic platform, Clearpath Husky A200 UGV (Unmanned Ground Vehicle) (left), and side view of the Husky robot with its dimension (right) (Clearpath, 2014).

1

1.4 A fundamental pipeline for weed control and scope of this thesis

For autonomous weed control in an agricultural field, following three core functionalities are typically required (Slaughter et al., 2008): I) (autonomous) vehicle navigation in the field, II) weed identification and classification, and III) actuation of a weed removal device. A fundamental pipeline for weed control is presented in Figure 1.3.

Among these three core functionalities, the identification and classification of weeds from cash crop using machine vision under agricultural field conditions still remains the greatest challenge (Liu et al., 2014; Slaughter et al., 2008). Consequently, this thesis focuses on II) weed identification and classification (in Figure 1.3), specifically for volunteer potato detection within SmartBot project. For this, three essential steps are needed as follows:

(a) Any plant materials are first segmented in an acquired field image. During this vegetation segmentation, background pixels (soil-related pixels) are removed, and foreground pixels (plant-related pixels) are left (Figure 1.3a).
(b) Individual objects (plants) are identified (Figure 1.3b).
(c) Each individual object is classified either as a sugar beet plant or a volunteer potato plant (Figure 1.3c).

Of the three steps above, this thesis covers two processes, the first (a) and the third (c) steps, to identify and classify volunteer potato in a sugar beet field: 1) vegetation segmentation (Figure 1.3a), and 2) classification of vegetation into volunteer potato plants (weeds) and sugar beet plants (crop) (Figure 1.3c). Dealing with variability in colour, shape and size of the plants and varying light conditions are amongst the research challenges. The details are explained in the following sections.

The identification of individual plants, the second step (b), was manually carried out mainly using blob detection, and this was not covered in this thesis. Overlapping plants of similar and different species were not considered in this study.

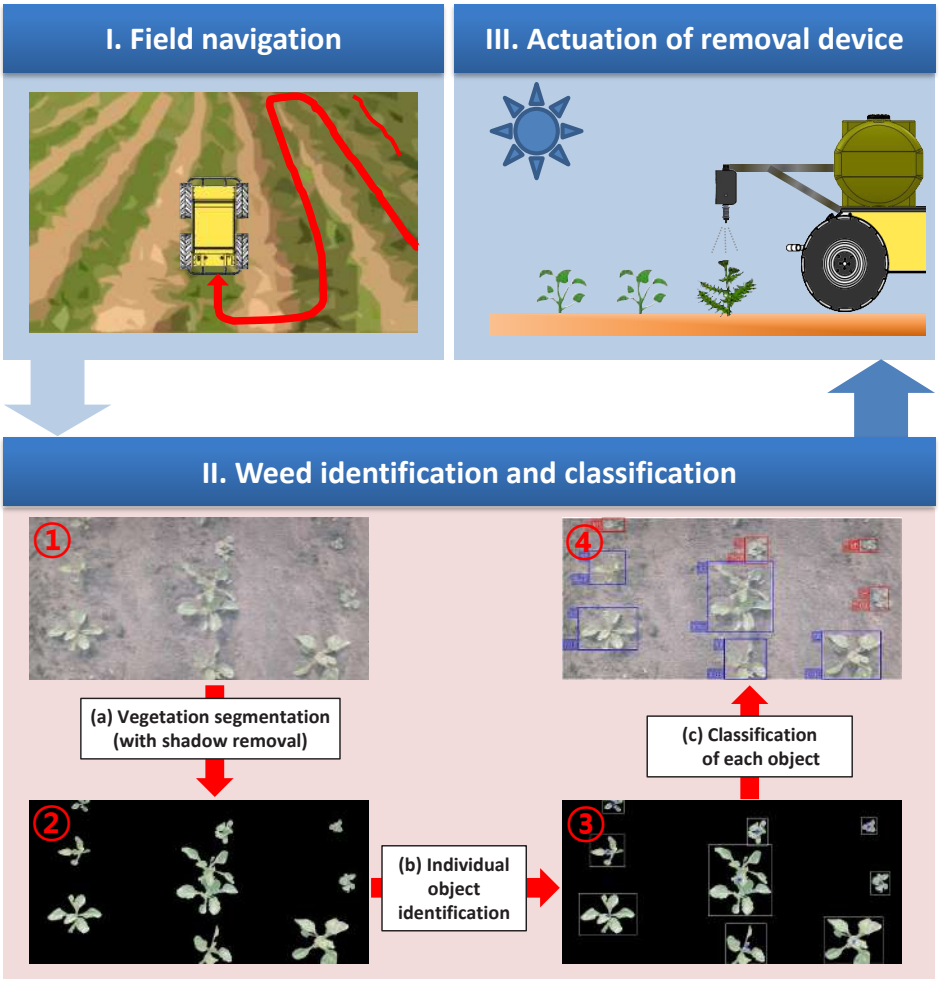


Figure 1.3: A fundamental pipeline for autonomous weed control. An autonomous weed control system requires three core functionalities: I) (autonomous) vehicle navigation in the field, II) weed identification and classification, and III) actuation of a weed removal device. This thesis focuses on II) volunteer potato (weed) identification and classification. The required three steps are listed: (a) vegetation segmentation removes soil background and segments plant materials in a field image, (b) the individual objects (plants) are identified, and then (c) each identified object is classified either as a sugar beet plant (blue square) or a volunteer potato plant (red square).

1

1.5 Problem description

Concerning the application of a vision-based system in an agricultural field environment, researchers have commonly addressed the issues of varying illumination and shadows by using a hood covering both the scene and the vision acquisition device (Figure 1.4). By doing so, any ambient light was blocked (Ahmed et al., 2012; Haug et al., 2014; Lottes et al., 2017), and constant illumination under the cover was then obtained using artificial lighting (Nieuwenhuizen et al., 2010; Polder et al., 2014).

However, such an approach using a hood/cover was not feasible within the context of the SmartBot project because a small-sized mobile robotic platform (Husky) was to be used (Figure 1.2). Additional infrastructure such as a hood and lighting equipment to overcome the challenges of ambient lighting conditions in the field, as for example were used by Nieuwenhuizen et al. (2010) and Haug et al. (2014) in Figure 1.4, was not considered viable. The resulting system in SmartBot should perform robustly in scenes that are fully exposed to ambient lighting conditions in the field.

There are two challenging issues for vision-based applications that are fully exposed to ambient lighting conditions in an agricultural field: 1) strongly varying natural illumination (Jeon et al., 2011; Wang et al., 2012); 2) shadows under direct sunlight conditions (Guo et al., 2013; Zhang et al., 2010). In a field environment, illumination conditions constantly change depending on the sky and weather conditions, and this change affects colour pixel values of acquired field images and leads to an inconsistent

Figure 1.4: Example applications of vision-based systems in an agricultural field. Despite different appearances, these systems have one common design approach: the vision system was placed under a cover to block any ambient light, and constant illumination under the cover was then obtained using artificial lighting. These applications were developed by Polder et al. (2014), Nieuwenhuizen et al. (2010), and Haug et al. (2014) (from left to right).

colour representation of plants (Sojodishijani et al., 2010; Teixidó et al., 2012). The dynamic range of the scene in such field environments is much larger than a traditional machine-vision camera covers (Dworak et al., 2013), and thus a traditional RGB camera may not be able to capture stable and reliable images. In addition, shadows often create extreme illumination contrasts, causing substantial luminance differences within a single image scene under direct sunlight conditions when the scene in the field is not covered by a hood or similar structure. Shadows influence colour values of the object in an image scene, and in many cases in vegetation segmentation, shadows tend to be classified as part of the foreground, i.e., as vegetation regions. These issues make vegatation segmentation a very challenging task.

When it comes to the classification of weeds amongst cash crops, the use of conventional features (intuitive features), such as colour, shape (biological morphology) and texture leads to relatively poor classification result. These conventional features on an individual basis or as a combination of multiple of them have been commonly used for the classification of weed and crop (Ahmed et al., 2012; Åstrand & Baerveldt, 2002; Gebhardt & Kühbauch, 2007; Pérez et al., 2000; Persson & Åstrand, 2008; Slaughter et al., 2008; Swain et al., 2011; Zhang et al., 2010). These features are intuitive and easy-to-implement, but may have limited discriminative power under widely varying natural light conditions in an agricultural field. The use of colour features, for example, may not yield robust classification in a system that has to work under ambient light conditions (Lee et al., 2010). For a case of volunteer potato and sugar beet, it is sometimes hard if not impossible to differentiate between them using colour features. Usually, volunteer potato has a darker green colour than sugar beet (Figure 1.5a) which results in a separable pixel distribution in the EG-RB colour plane (Figure 1.5c). However, as is shown in Figure 1.5b, volunteer potato occasionally has the same or very similar colour as sugar beet which then yields an inseparable distribution in the EG-RB colour plane (Figure 1.5d). Besides, the colour of plants may change depending on the growth stage and nutritional status, and the green plant leaves sometimes even turn yellow in the summer time (Muñoz-Huerta et al., 2013). Shape and texture may also not be sufficiently discriminating features for successful classification of sugar beet and volunteer potato in the field. Therefore, a substantial effort has to be made in the classification of weeds amongst cash crops using novel discriminative features; however, it is still unknown which features work best for the classification of sugar beet and volunteer potato in agricultural field conditions.

1

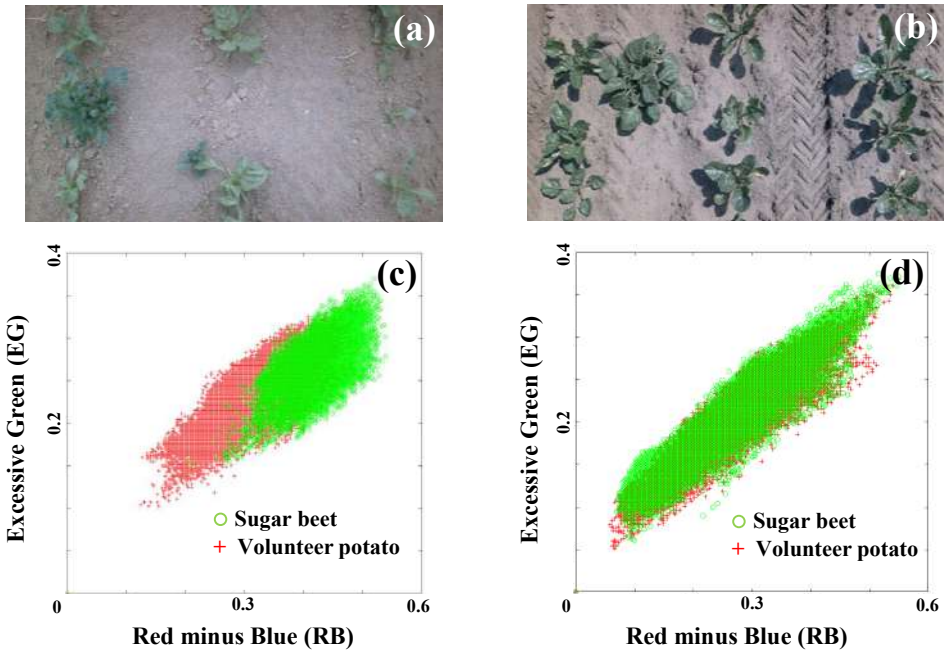


Figure 1.5: (a) In general, volunteer potato has a darker green colour than sugar beet. (c) In such a case, sugar beet and volunteer potato are separable (based on the colour) in the EG-RB plane. (b) An example case of volunteer potato having the same colour distribution as sugar beet. (d) Sugar beet and volunteer potato are then visually inseparable in the EG-RB plane. The EGRBI transformation was used to compare the colour difference between sugar beet and volunteer potato (Nieuwenhuizen et al., 2007).

1.6 Objectives, research questions and thesis outline

Within the SmartBot project the objective of this research was:

to develop a computer vision procedure
that detects volunteer potato plants
under ambient light conditions in a sugar beet field

Based on this main objective, three sub-objectives were formulated addressing two critical processes to discriminate between volunteer potato and sugar beet: vegetation

segmentation (sub-objectives 1 and 2) and classification of vegetation into volunteer potato and sugar beet (sub-objective 3). In line with these three sub-objectives, research questions were derived. The research questions are dealt with in the four research chapters in this thesis (Chapter 2 to 5) and are written in bold in below:

Sub-objective 1: *To develop an algorithm for robust vegetation segmentation that can cope with the influence of shadows and natural illumination conditions in a sugar beet field, using a camera that has a wide dynamic range.*

a. **Does a ground shadow detection and removal method enhance the performance of vegetation segmentation under natural illumination conditions in the field, using a High Dynamic Range (HDR) camera?**

   Chapter 2 proposes an algorithm for ground shadow detection and removal based on colour space conversion and a multilevel threshold. The advantage of using the proposed algorithm was assessed for vegetation segmentation with field images that were acquired by an HDR camera under natural illumination. An HDR camera was used because it enables the capture of stable and reliable images even under intense and direct solar radiation or under faint starlight (Reinhard et al., 2010).

Sub-objective 2: *To evaluate which combination of colour indices and threshold techniques performs best under field conditions given varying illumination conditions, presence of shadows and differences in plant size.*

a. **Do different combinations of colour index and threshold technique result in different segmentation performance when evaluated on field images? Do certain combinations stand out positively in performance compared to others and, the other way around, do certain combinations stand out negatively when compared to others?**

b. **If differences in segmentation performance do exist, which combination works the best given the field conditions like illumination intensity, shadow presence and plant size?**

c. **Given the varying conditions in the field, is it better to use one combination (at all times) or should the combination be adapted to the conditions at hand for best segmentation performance?**

1

d. **Do results obtained from a-c hold true when validated on a different independent image dataset?**

Chapter 3 evaluates the performance of 40 combinations of eight colour and five thresholding techniques to identify which combination performs best under field conditions given varying illumination conditions, presence of shadows and differences in plant size. The performance of one combination at all times was compared with the combinations that were adapted to the conditions. In this way, it was assessed if it is better to adapt the combinations based on the conditions for best segmentation performance.

Sub-objective 3: *To develop algorithms for the classification of volunteer potato (weed) among sugar beet (cash crop) using discriminative features that are not dependent on illumination, colour, and shadow.*

a. **Does an algorithm using a Bag-of-Visual-Words (BoVW) model and SIFT or SURF descriptors meet the requirements set for the classification of volunteer potato and sugar beet under natural and varying daylight conditions?**

b. **b. If the BoVW model does not meet the requirements, does a deep learning approach, particularly transfer learning based on Convolutional Neural Network (ConvNet, or CNN) provide an effective and better performance to meet the requirements with limited amount of dataset?**

d. **Are the processing times (or calculation times) fast enough for real-time application?**

Chapter 4 discusses an algorithm using a Bag-of-Visual-Words (BoVW) model and SIFT or SURF descriptors as well as crop row information in the form of the Out-of-Row Regional Index (ORRI) was proposed for the classification of sugar beet and volunteer potato under natural and varying daylight conditions. The BoVW approach has demonstrated good performance in many computer vision applications such as object and scene classification (Law et al., 2014; Tsai, 2012; Zhou et al., 2013). The SIFT descriptor has been used for weed classification and recognition in several recent studies (Kazmi et al., 2015a; Wilf et al., 2016).

The performance difference between SIFT and SURF was verified by assessing classification accuracy and computation time. Crop row information in the form of ORRI was added to the feature set to assess any performance improvement. Three different classifiers (SVM, random forest, and neural network) were compared to get more insight into performance differences amongst classifiers. A posterior probability of the output of the SVM was calculated using a method proposed by Platt (1999).

Chapter 5 evaluates a transfer learning procedure with three different implementations of AlexNet (Part I) and then assesses the performance difference amongst the six network architectures (Part II). In Part I, AlexNet was used as a pre-trained ConvNet. Based on two available options in transfer learning (use of ConvNet as a feature extractor and use of ConvNet as a classifier), three scenarios for transfer learning were formulated: 1) pre-trained AlexNet as a fixed feature extractor followed with a classifier, 2) modified and fine-tuned AlexNet as a binary classifier, and 3) modified and fine-tuned AlexNet as a fixed feature extractor followed with a classifier. In Part II, following six pre-trained deep networks were evaluated to assess the classification performance amongst different ConvNet architectures: AlexNet, VGG-19, GoogLeNet, ResNet-50, ResNet-101 and Inception-v3. These networks are available as pre-trained ConvNets which have been trained on ImageNet Dataset, and are used as pre-trained networks. These nets were used to classify sugar beet and volunteer potato images taken under ambient varying light conditions in agricultural environments. The classification performance in both Part I and II was analysed regarding classification accuracy as well as training and classification time.

Lastly, in Chapter 6, the main results are evaluated with a general discussion in relation to the objectives of the research. The major contribution of the thesis is reviewed in a broader perspective and further discussed in connection with existing literature. Furthermore, the limitations and implications of the results are discussed for practical application of a weeding robot, and recommendations are proposed for future research.

CHAPTER 2

Improved vegetation segmentation with ground shadow removal using an HDR camera

Hyun K. Suh
Jan Willem Hofstee
Eldert J. van Henten

The contents of this chapter have been published in *Precision Agriculture (2018), 19(2), 218-237* as a paper entitled: Improved vegetation segmentation with ground shadow removal using an HDR camera.

Abstract

A vision-based weed control robot for agricultural field application requires robust vegetation segmentation. The output of vegetation segmentation is the fundamental element in the subsequent process of weed and crop discrimination as well as weed control. There are two challenging issues for robust vegetation segmentation under agricultural field conditions: (1) to overcome strongly varying natural illumination; (2) to avoid the influence of shadows under direct sunlight conditions. A way to resolve the issue of varying natural illumination is to use High Dynamic Range (HDR) camera technology. HDR cameras, however, do not resolve the shadow issue. In many cases, shadows tend to be classified during the segmentation as part of the foreground, i.e. vegetation regions. This study proposes an algorithm for ground shadow detection and removal, which is based on color space conversion and a multilevel threshold, and assesses the advantage of using this algorithm in vegetation segmentation under natural illumination conditions in an agricultural field. Applying shadow removal improved the performance of vegetation segmentation with an average improvement of 20%, 4.4%, and 13.5% in precision, specificity and modified accuracy, respectively. The average processing time for vegetation segmentation with shadow removal was 0.46 s, which is acceptable for real-time application (< 1 s Required). The proposed ground shadow detection and removal method enhances the performance of vegetation segmentation under natural illumination conditions in the field and is feasible for real-time field applications.

2.1 Introduction

This work was part of the EU-funded project SmartBot, a project with the research goal to develop a small-sized vision-based robot for control of volunteer potato (weed) in a sugar beet field. Such a vision-based weed control robot for agricultural field application requires robust vegetation segmentation, i.e. a vegetation segmentation that has good performance under a wide range of circumstances. The output of vegetation segmentation is the fundamental element in the subsequent process of weed and crop discrimination as well as weed control (Meyer & Camargo Neto, 2008; Steward et al., 2004). There are two challenging issues for robust vegetation segmentation in agricultural field conditions: 1) to overcome the strongly varying natural illumination (Jeon et al., 2011; Wang et al., 2012); 2) to avoid the influence of shadows under direct sunlight conditions (Guo et al., 2013; Zheng et al., 2009).

Illumination conditions constantly change in an agricultural field environment depending on the sky and weather conditions. These illumination variations greatly affect (RGB) pixel values of acquired field images and lead to the inconsistent color representation of plants (Sojodishijani et al., 2010; Teixidó et al., 2012). In addition, shadows often create extreme illumination contrast, causing substantial intensity/luminance differences within a single image scene. These extreme intensity differences make vegetation segmentation a very challenging task.

Researchers addressed the above two problems by using a hood covering both the scene and the vision acquisition device. By doing so, any ambient visible light was blocked (Ahmed et al., 2012; Åstrand & Baerveldt, 2002; Haug et al., 2014; Lee et al., 1999). Constant illumination under the cover was then obtained using artificial lighting (Nieuwenhuizen et al., 2010; Polder et al., 2014).

Such a solution was not feasible within the framework of the Smartbot project because a small-sized mobile robotic platform was to be used. An extra structure for the cover was not viable due to the reduced carrying capacity of the platform. Moreover, using additional energy for artificial lighting would be another critical issue, considering the mobile platform was battery operated. Therefore, a solution was needed that uses the ambient light while overcoming the drawbacks mentioned earlier.

A way to resolve the issue of varying natural illumination and substantial intensity differences within a single image scene is to use High Dynamic Range (HDR) camera technology as has been indicated by a number of studies (Graham, 2011; Hrabar et al.,

2009; Irie et al., 2012; Lapray et al., 2012; Mann et al., 2012; Slaughter et al., 2008). Under direct sunlight conditions, the dynamic range of the scene is much larger than a traditional non-HDR camera covers, especially if an image scene contains shadows (Dworak et al., 2013). Having a larger dynamic range, an HDR camera enables the capture of stable and reliable images even under strong and direct solar radiation or under faint starlight (Reinhard et al., 2010).

HDR cameras, however, do not resolve the shadow issue. When the scene in the field is not covered by a hood or similar structure, shadows are inevitable. In many cases in vegetation segmentation, shadows tend to be classified as part of the foreground, i.e. vegetation regions (Figure 2.1). Therefore, shadows need to be detected and preferably removed for better segmentation performance. However, shadow detection is extremely challenging especially in an agricultural field environment because shadows change dramatically throughout the day depending on position and intensity of the sun. Besides, shadows have no regular shape, size, or texture, and can even be distorted on an uneven ground surface. In recent years, many shadow detection and removal algorithms were proposed in computer vision research area using a feature-based or a brightness/contrast compensation (Sanin et al., 2012). However, these shadow detection and removal algorithms are difficult to implement and require a significant amount of computation time, which is an important issue for real-time field applications. Moreover, these algorithms provide poor shadow removal output for outdoor scenes (Al-Najdawi et al., 2012). Therefore, a simple and effective shadow detection and removal algorithm is needed for real-time weed detection and control application in an agricultural field environment.

This paper proposes an algorithm for ground shadow detection and removal, and

Figure 2.1: Example of shadow images (top), and vegetation segmentation output with excess green (ExG) segmentation (bottom). Shadows are partially segmented as vegetation.

assesses the effectiveness of using this algorithm in vegetation segmentation under natural illumination conditions in an agricultural field. The proper quantitative measure to evaluate the performance of vegetation segmentation are discussed.

2.2 Materials and Methods

2.2.1 High Dynamic Range (HDR) camera

A common definition of the dynamic range of an image is the ratio of maximum and minimum illuminance in a given scene. More precisely, Bloch (2007) defines dynamic range as the logarithmic ratio between the largest and the smallest readable signal (an image is treated as a signal from the camera hardware aspect):

$$Dynamic\ Range\ (dB) = 20 \times \log_{10}(\frac{MaxSignal}{MinSignal}) \tag{2.1}$$

The illumination difference in a real-life image scene can easily exceed a dynamic range of 80 dB. In outdoor field conditions, the dynamic range can exceed 120 dB (Radonjić et al., 2011). Human eyes have a dynamic range of around 200 dB, while a conventional imaging device such as a non-HDR CCD digital camera typically has a dynamic range of around 60 dB (Bandoh et al., 2010; Ohta, 2007). Under direct sunlight conditions, the dynamic range of the scene can be much higher than a traditional non-HDR camera can cover, especially when the image scene contains sharp dark shadows. Thus, a conventional imaging device is not feasible for machine vision applications in a natural agricultural environment, because strong direct solar radiation and shadows frequently cause extreme lighting intensity changes. Piron et al. (2010) used an exposure fusion method to generate a high dynamic range scene of plant images and reported that high dynamic range acquisition supported obtaining a quality image of the scene with a strong signal to noise ratio. In the past few years, HDR cameras have become commercially available at an affordable price.

In this study, a HDR camera (NSC1005c, New Imaging Technologies, Paris, France) having a dynamic range of 140 dB and a bit depth of 36 bits per pixel was used (Figure 2.2). This camera has two identical CMOS sensors providing the stereo images (left and right), but only the left sensor image was used in this study.

Example images of a similar scene made with the HDR and traditional non-HDR

CCD cameras are shown in Figure 2.3. The HDR camera captures the objects even in the dark shadow region (Figure 2.3a) whereas a traditional non-HDR CCD camera (Sony NEX-5R) captures no objects but produces black pixels (Figure 2.3c). The histogram of the HDR image is well balanced across the darkest and lightest margins (Figure 2.3b) while the histogram of a traditional non-HDR CCD camera image is imbalanced with peaks both on the left and right edges due to clipping (Figure 2.3d).

An example field image that was acquired with an HDR camera under very bright sunny conditions is shown in Figure 2.4. Some pixels in the green leaves were bright due to specular reflection; while some pixels in the shadow region were very dark. The extreme lighting intensity difference with a high dynamic range is often found in a field image scene. In such a condition in the field, a conventional non-HDR imaging device would not be able to adequately capture the objects in both the bright regions as well as in the dark shadow regions but an HDR camera does adequately capture these objects under these lighting conditions.

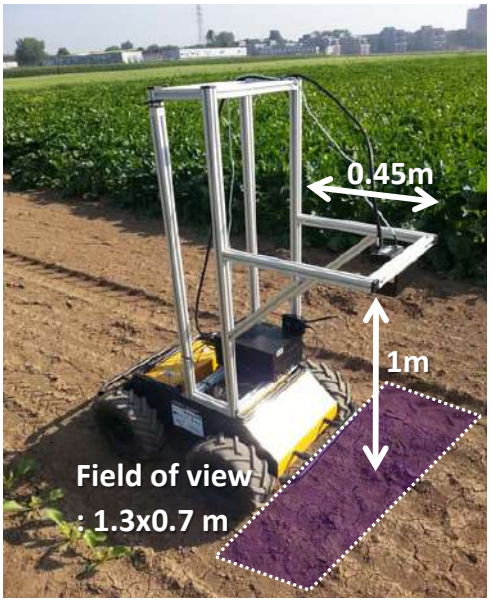


Figure 2.2: Field images were acquired with an HDR camera (left) which was mounted at a height of 1 m *viewing perpendicular to the ground surface, resulting in a field of view:* 1.3 m × 0.7 m *(right).*

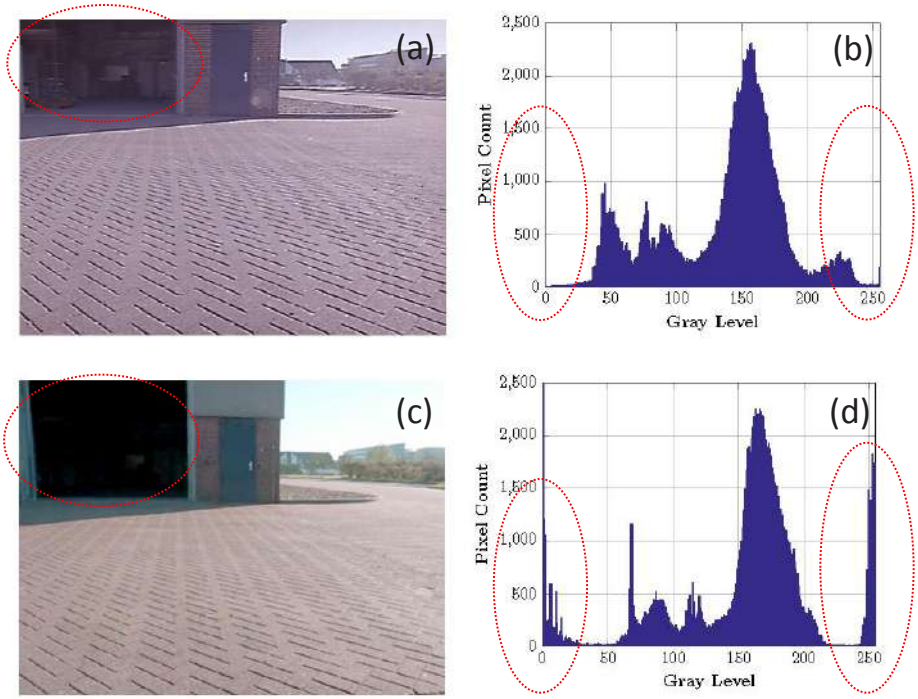


Figure 2.3: An example outdoor image scene on a sunny day: (a) HDR camera image with (b) image histogram, (c) traditional non-HDR CCD camera (Sony NEX-5R, ISO 100, 1/80, f/11, dynamic range optimizer activated) image with (d) image histogram. The red ellipses indicate that (b) the histogram of the HDR image is well balanced across the darkest and lightest margins, but (d) the histogram of a traditional non-HDR CCD camera image is imbalanced with peaks both on the left and right edges due to clipping.

2.2.2 Algorithm - Ground shadow detection and removal

As was shown in Figure 2.1, shadows in agricultural field images are often classified as part of vegetation when applying a commonly used vegetation segmentation method based on the excessive green index (2g-r-b), ExG (Woebbecke et al., 1995). To further process the shadows, a ground shadow detection algorithm was developed using color space conversion. Color pixel values in RGB space can be highly influenced by the illumination conditions because illumination and color parts are not separated in this color representation (Florczyk, 2005). Using a different color space (or conversion of

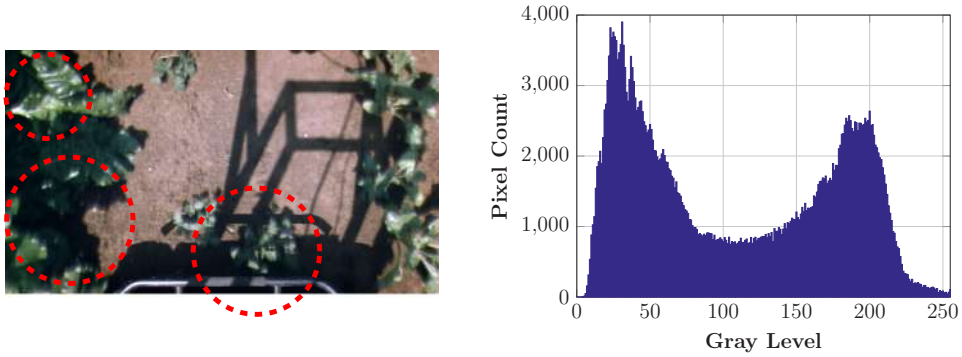


Figure 2.4: HDR camera image in a sugar beet field (left) and its image histogram (right). The red circles indicate that some pixels in the green leaves are bright due to specular reflection; while some pixels in the shadow region are very dark.

the RGB image to another color space) that uses a color representation separating color and illumination, pixel values are less influenced by the illumination conditions with the shadow detection procedure. Many studies have shown that a color space conversion approach is simple to implement and computationally inexpensive; thus, is very useful for real-time field applications (Sanin et al., 2012).

In this study, the XYZ color space was chosen because the normalized form of this color space separates luminance from color (or rather from chromaticity). Also this color space is based on how a human would perceive light (Pascale, 2003). The XYZ system provides a standard way to describe colors and contains all real colors (Corke, 2011). Besides, this particular color space has been shown to be robust under illumination variations (Lati et al., 2013a).

The procedure used for ground shadow detection and removal is shown in Figure 2.5. Two main processes are shown: 1) ExG with Otsu (1979) threshold (Figure 2.5 steps a to c), and 2) ground shadow detection and removal (Figure 2.5 steps d to h).

The left column in Figure 2.5, steps (a) to (c), shows the conventional vegetation segmentation procedure. ExG, one of the most commonly used methods, was used in this study to compare the performance of vegetation segmentation before and after shadow removal because ExG showed good performance in most cases in our preliminary studies. The Otsu threshold was used because the Otsu method showed good performance in a preliminary study.

The right column in Figure 2.5, step (d) to (h), shows the ground shadow detection

and removal procedure. The three individual steps (d) to (f) are referred to as a ground shadow detection, and pixel-by-pixel subtraction in step (g) is referred to as ground shadow removal. The detected ground shadow region was subtracted from ExG with Otsu threshold (ExG+Otsu) which resulted from step (c). Then, the shadow-removed image (Figure 2.5h) was compared with ExG+Otsu (Figure 2.5c) to evaluate the performance improvement when using vegetation segmentation after shadow detection and removal. The details of the algorithms are described below.

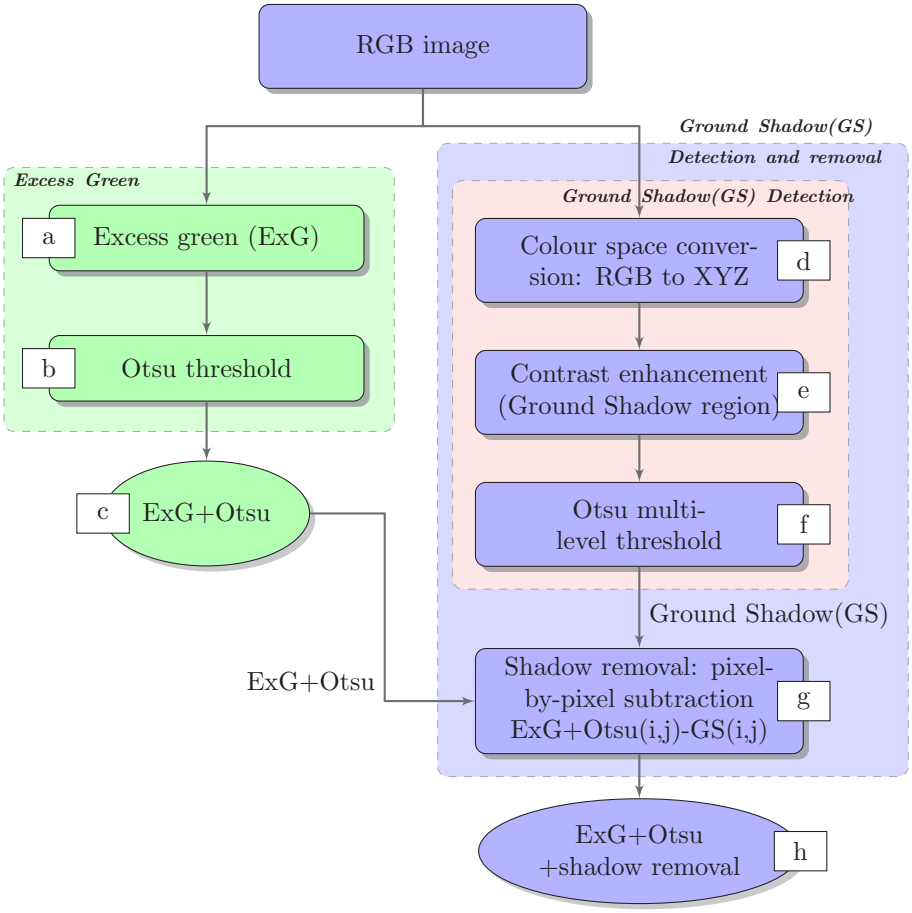


Figure 2.5: Flow diagram of ground shadow detection and removal algorithm.

Step a: Excess Green (ExG)

The excess green index ($ExG = 2g - r - b$) was applied to the RGB image (Woebbecke et al., 1995). The normalized spectral r, g and b components, in the range [0,1], were obtained (Gée et al., 2008).

$$r = \frac{R_n}{R_n + G_n + B_n}, \quad g = \frac{G_n}{R_n + G_n + B_n}, \quad b = \frac{B_n}{R_n + G_n + B_n} \tag{2.2}$$

where R_n, G_n, and B_n are the normalized RGB coordinates ranging from 0 to 1. They were obtained as follows:

$$R_n = \frac{R}{R_{max}}, \quad G_n = \frac{G}{G_{max}}, \quad B_n = \frac{B}{B_{max}} \tag{2.3}$$

where $R_{max} = G_{max} = B_{max} = 255$

Step b: Otsu threshold

The Otsu threshold method was applied to obtain an optimum threshold value. The pixels of the image were divided into the two classes: C_0 for $[0, \cdots, t]$ and C_1 for $[t+1, \cdots, L]$, where t was the threshold value ($0 \leq t < L$), and L was the number of distinct intensity levels. An optimum threshold value t^* was chosen by maximizing the between-class variances, σ_B^2 (Otsu, 1979):

$$t^* = \underset{0 \leq t < L}{\operatorname{argmax}} \{\sigma_B^2(t)\} \tag{2.4}$$

Step c: ExG+Otsu

With an optimum threshold value t^* (Eq. 2.4), vegetation pixels were classified.

$$\begin{cases} Background\,region & \text{if} \quad ExG(i,j) < t^* \\ Vegetation\,region & \text{if} \quad ExG(i,j) \geq t^* \end{cases}$$

where $ExG(i,j)$ was the ExG value of the pixel (i,j).

Step d: Color space conversion

The first step involved color space conversion. The RGB values were converted to the 1931 International Commission on Illumination (CIE) XYZ space using the following

matrix (Lati et al., 2013a):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.5}$$

where R, G, B were pixel values in RGB color space in the range [0, 255]. X, Y, Z were pixel values in XYZ color space. Finally, XYZ values were then normalized using the following equation:

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z} \tag{2.6}$$

Step e: Contrast enhancement of the ground shadow region

The contrast of the ground shadow region was enhanced from the rest of the image. This contrast enhancement was achieved by dividing the product of the chromaticity values x and y by z:

$$Contrasted\ Ground\ Shadow(CGS(i,j)) = \frac{x(i,j) \cdot y(i,j)}{z(i,j)} \tag{2.7}$$

where $CGS(i,j)$ was the contrasted pixel (i,j) of the ground shadow, and $x(i,j)$, $y(i,j)$, $z(i,j)$ were normalized values of the pixel (i,j) in the XYZ color space.

Step f: Otsu multi-level threshold

The Otsu multi-level threshold method was applied to the image based on the observation that the shadow image contained three components - ground shadow, plant material and soil background. The previous steps (d) and (e) made the ground shadow region more distinct from other components. Thus, the Otsu multi-level threshold enabled to separate the ground shadow region, which had the lowest intensity level, from plant material and soil background. The lowest intensity level was selected as the ground shadow region, but plant material and soil background regions were not separated because they were not clearly distinct from each other.

All pixels of the image obtained in the previous step (e) were divided into the following three classes: C_0 for $[0, \cdots, t_1]$, C_1 for $[t_1{+}1, \cdots, t_2]$, and C_2 for $[t_2{+}1, \cdots, L]$, where t_1 and t_2 were threshold values ($0 \leq t_1 < t_2 < L$), and L was the number of

distinct intensity levels. An optimal set of threshold values t_1^* and t_2^* was chosen by maximizing the between-class variances, σ_B^2 (Otsu, 1979):

$$\{t_1^*, t_2^*\} = \underset{0 \leq t_1 < t_2 < L}{\operatorname{argmax}} \{\sigma_B^2(t_1, t_2)\} \tag{2.8}$$

For the ground shadow detection, the optimal threshold value t_1^* was used. The threshold value t_2^* was ignored since it did not have any added value in this ground shadow detection process. Consequently, the ground shadow pixels were classified in two classes as follows

For the ground shadow detection, threshold value t_1 was used, and ground shadow pixels were selected as follows:

$$\begin{cases} \textit{Ground shadow } (GS) & \text{if} \quad CGS(i,j) < t_1^* \\ \textit{Non-ground shadow } (NGS) & \text{if} \quad CGS(i,j) \geq t_1^* \end{cases}$$

Step g: Ground shadow removal by subtraction

Once the ground shadow region was identified, the shadow-removed image was generated by a pixel-by-pixel subtraction from the ExG+Otsu. The shadow-removed pixel values were simply the values of ExG minus the corresponding pixel values from the ground shadow region image (Eq. 2.9).

$$F(i,j) = ExG(i,j) - GS(i,j) \tag{2.9}$$

where $F(i,j)$ was the shadow-removed pixel (i,j), and $GS(i,j)$ was the detected ground shadow pixel (i,j).

2.2.3 Field image collection

For crop image acquisition, the HDR camera was mounted at a height of 1 m viewing perpendicular to the ground surface on a custom-made frame carried by a mobile platform (Husky A200, Clearpath, Canada), as was shown in Figure 2.2. The camera was equipped with two identical Kowa 5 mm lenses (LM5JC10M, Kowa, Japan) with a fixed aperture. The camera was set to operate in automatic acquisition mode with automatic point and shoot, having an image resolution of 1280×580 pixels per image of left and right sensors. The ground-covered area was $1.3\,\text{m} \times 0.7\,\text{m}$, corresponding to

three sugar beet crop rows. The acquisition program was implemented in LabVIEW (National Instruments, Austin, USA) to acquire five images per second. Field images were taken while the mobile platform was manually controlled with a joystick and driven along crop rows using a controlled traveling speed of 0.5 m/s.

Sugar beet was seeded three times (Spring, Summer, and Fall) in 2013 and 2014 in two different soil types (sandy and clay soil) on the Unifarm experimental sites in Wageningen, The Netherlands. Crop images were acquired under various illumination and weather conditions on several days in June, August and October of 2013 as well as in May, July and September of 2014.

2.2.4 Image dataset

The following image datasets were chosen for this study: 1) Set 1: only containing images with shadow to purely test and evaluate the performance of the shadow detection algorithm against human generated ground truth, and 2) Set 2: containing a mix of images with and without shadows to assess the effectiveness of shadow removal on segmentation.

Set 1 consisted of 30 field images that all contained shadows ranging from shallow to dark with various shadow shapes (Figure 2.6). The images in this set were acquired on several days under various weather conditions at different growth stages of the crop. Ground truth images for shadow regions was manually generated.

For Set 2, a total of 110 field images was selected from all acquired field images. During the selection of this set, a wide range of natural conditions was considered, including different stages of plant growth, various illumination conditions from a cloudy dark to sunny bright day conditions and extreme illumination scenes caused by strong

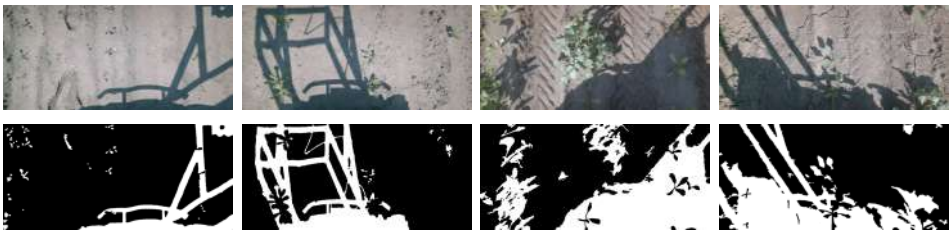


Figure 2.6: Flow diagram of ground shadow detection and removal algorithm.

direct solar radiation and shadows. Half of the images in this set contained shadows (from shallow to dark shadows) while the other half contained no shadows. Vegetation regions were manually labeled for ground truth. All images were processed with the Image Processing Toolbox™ in Matlab (The MathWorks Inc., Natick, USA).

2.2.5 Quantitative performance measures of vegetation segmentation

The vegetation segmentation results were compared and evaluated at pixel level with human-labelled ground truth images. The ground truth images were generated by two people. In this study, a set of quantitative measures based on the confusion matrix (Table 2.1) was used to assess the performance of the vegetation segmentation. Positive prediction value (precision), true-positive rate (recall or sensitivity), true-negative rate (specificity) and modified accuracy (MA) were used. Each of these has a different goal to measure, thus assessing above measures altogether helps to evaluate the performance of vegetation segmentation in a balanced way. The details of the measures are described below (Metz, 1978; Prati et al., 2003):

$$Precision\ (Positive\ Predict\ Value) = \frac{TP}{TP + FP} \tag{2.10}$$

$$Recall\ (True\ positive\ rate\ or\ Sensitivity) = \frac{TP}{TP + FN} \tag{2.11}$$

$$Specificity\ (True\ negative\ rate) = \frac{TN}{TN + FP} \tag{2.12}$$

where TP is true-positive; FP is false-positive; TN is true-negative, and FN is false-negative.

Precision indicates how many of the positively segmented pixels are relevant, and it refers to the ability to minimize the number of false-positives. Recall indicates how well a segmentation performs in detecting the vegetation and thus relates to the ability to correctly detect vegetation pixels that belong to the vegetation region (true-positive). Specificity, on the other hand, specifies how well the segmentation algorithm performs in avoiding false-positive error, which also indicates the ability to correctly detect non-vegetation pixels that belong to non-vegetation regions (true-negative). A single

Table 2.1: Confusion matrix
(TR:true-positive, TN:true-negative, FP:false-positive, and FN:false-negative)

		Algorithm	
		Vegetation	Background
Ground truth	Vegetation	TP	FN
	Background	FP	TN

measure above does not fully reflect the performance of vegetation segmentation because each can have a biased value under certain conditions. For example, if a segmentation produces vegetation pixels only when there are strong green components, precision will have a higher value (close to 1) in a poorly segmented image. Moreover, if a segmentation always identifies all the pixels as vegetation, recall will attain large values.

Accuracy is commonly used as a single representative performance indicator in the literature. However, this measure has a drawback if there is a significant imbalance between vegetation and background (Bac et al., 2013; Rosin & Ioannidis, 2003). An alternative way to measure the performance would be balanced accuracy, i.e. the average of sensitivity and specificity. However, this measure can also provide a biased value if segmentation output has a large number of false-positives in case an image contains only a small amount of vegetation. Therefore, the amount of vegetation (foreground area) needs to be considered to reflect better the performance of vegetation segmentation. Sezgin & Sankur (2004) used the relative foreground area error (RAE) and combined this indicator with accuracy (misclassification error). Many studies have used this combined approach since then (Guan & Yan, 2013; Nacereddine et al., 2005; Navarro et al., 2010; Shaikh et al., 2011). Inspired by these studies, the modified accuracy (MA) was defined in this study. This performance indicator uses a harmonic mean of relative vegetation area error (RVAE) and balanced accuracy (BA). Both measures have values between 0 and 1, where 0 represents very poor segmentation, and 1 represents perfect segmentation. The harmonic mean indicates if there is a large imbalance between these two measures, thus providing a better indication of the performance. The equations are described below:

$$Balanced\ accuracy\ (BA) = \frac{Recall + Specificity}{2} \tag{2.13}$$

$$Relative\ Vegetation\ Area\ Error\ (RVAE) = \begin{cases} 1 - \frac{A_{GT} - A_{SEG}}{A_{GT}} & \text{if } A_{SEG} < A_{GT} \\ 1 - \frac{A_{SEG} - A_{GT}}{A_{SEG}} & \text{if } A_{GT} \leq A_{SEG} \end{cases} \tag{2.14}$$

$$Modified\ accuracy\ (MA) = \frac{2 \cdot BA \cdot RVAE}{BA + RVAE} \tag{2.15}$$

where A_{GT} is the vegetation area in ground truth (TP+FN); A_{SEG} is the vegetation area in segmented image (TP+FP).

In addition, receiver operating characteristic (ROC) and precision-recall curves were used. These curves have been used in several studies on image processing helping to visually assess the segmentation performance (Bai et al., 2014; Bulanon et al., 2009). One performance indicator that is often used in these curves is the Area Under Curve (AUC), a measure represented with a single scalar value ranging from 0 to 1. AUC indicates how reliably the segmentation can be performed. An AUC value of 1 indicates a perfect segmentation (Mery & Pedreschi, 2005).

Finally, processing time was measured to indicate how fast the algorithm performed. The processing time was measured on a PC equipped with an Intel® Core™ i7-377T 2.5 GHz processor and 8 GB RAM running 64-bit Windows 7.

2.3 Results

2.3.1 Ground shadow detection

The performance measures of the ground shadow detection in Set 1 is shown in Table 2.2. Ground shadow detection was generally successful under natural lighting conditions ($modified\ accuracy \geq 0.9$). An average processing time (0.33 s) is satisfactory for real-time application as well. Example images in Set 1 and their ground shadow detection output are shown in Figure 2.7. From the original field images (Figure 2.7a), enhanced contrast of the ground shadow region is shown in the second column (Figure 2.7b). The third column shows the detected ground shadow region (Figure 2.7c), and its ground truth images are displayed in the fourth column (Figure 2.7d). The last column (Figure 2.7e) displays the difference image between detected shadow and ground truth.

Table 2.2: Quantitative performance measure of ground shadow detection in Set 1 (30 images). The mean, median, and standard deviation (SD) of the performance measures are indicated.
(PPV:Positive Predictive Value, TNR:True Negative Rate)

	Precision (PPV)	Recall (Sensitivity)	Specificity (TNR)	Modified Accuracy	Processing time (sec)
Mean	0.94	0.87	0.99	0.92	0.33
Median	0.96	0.87	0.99	0.92	0.33
SD	0.07	0.04	0.01	0.03	0.01

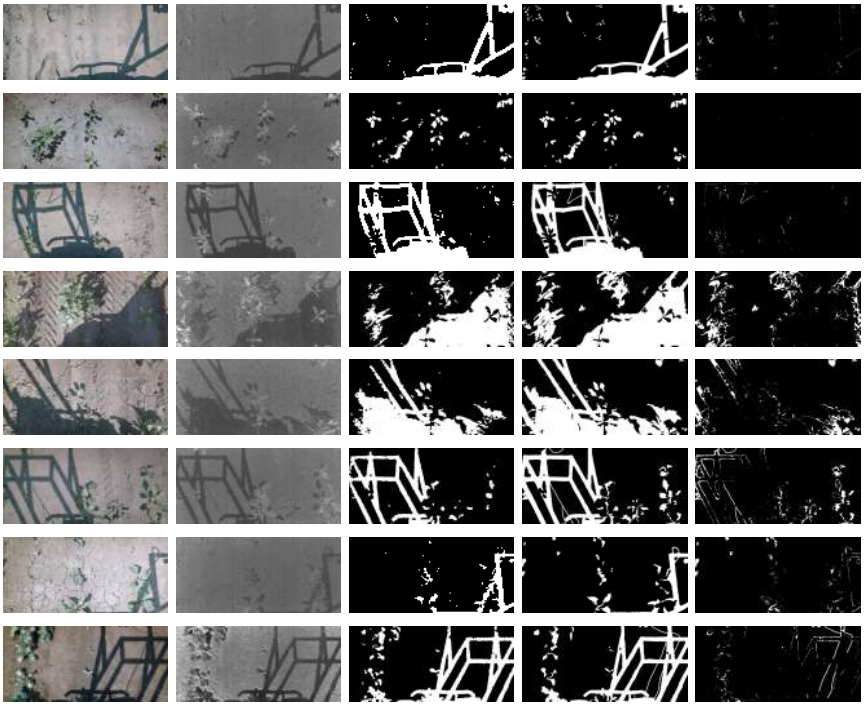


Figure 2.7: Example images in Set 1 using the ground shadow detection process: (a) original image, (b) contrasted ground shadow region, (c) ground shadow detected, (d) ground truth, and (e) difference between (c) and (d).

2.3.2 Ground shadow removal and vegetation segmentation performance

The images with shadow removal (ExG+Otsu+shadow removal, Figure 2.5h) were compared with those without shadow removal (ExG+Otsu, Figure 2.5c) to assess the performance improvement in vegetation segmentation when using ground shadow removal. The quantitative performance comparison is shown in Figure 2.8. When it comes to precision, sensitivity and modified accuracy, the figure indicates that the vegetation segmentation with shadow removal has a higher performance than the segmentation without shadow removal. The average values of precision, specificity and modified accuracy for vegetation segmentation with shadow removal were 0.67, 0.96 and 0.71, respectively, indicating 20%, 4.4% and 13.5% improvement over indicator values achieved without shadow removal. A T-test revealed that these improvements are significantly different ($P < 0.001$). Only recall indicated that there were some losses of true-positive pixels (vegetation pixels) due to the shadow subtraction process.

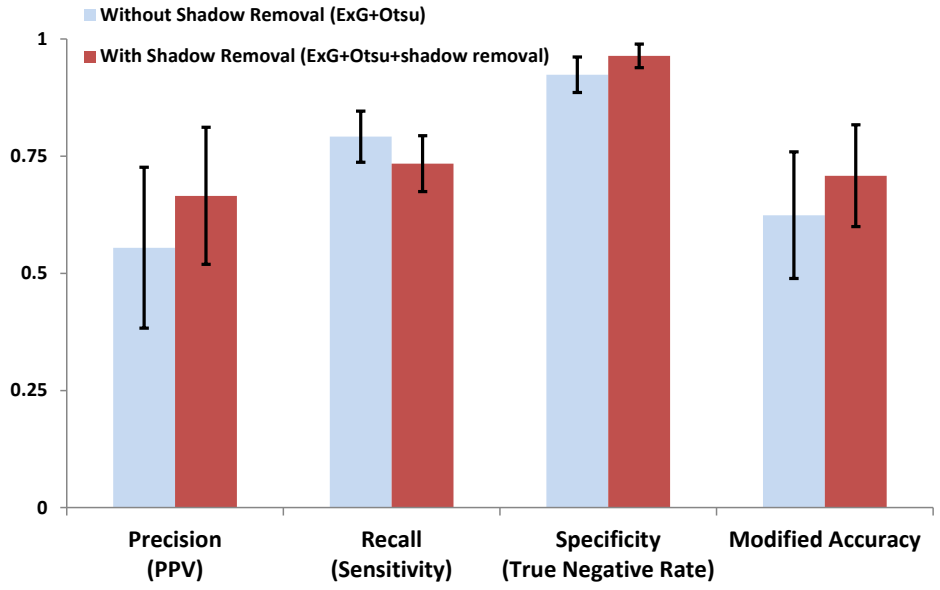


Figure 2.8: The performance of vegetation segmentation: without shadow removal (ExG+Otsu) vs. with shadow removal (ExG+Otsu+shadow removal).

This minor loss was mainly observed when the shadow removal was applied to images without shadows (see the last column in Figure 2.9). The combined measure, modified accuracy, indicated considerable improvement (13.5%) in vegetation segmentation. The average processing time for vegetation segmentation without shadow removal was 0.12 s, and was 0.46 s for segmentation with shadow removal. This is acceptable for a real-time application (< 1 s *Required*).

Figure 2.9 shows the results of the segmentation process with ground shadow removal applied to Set 2, including images without and with shadows. When there was a shadow in the image scene, the ground shadow detection algorithm was, in general, successful in detecting the ground shadow region; but when there was no shadow in the image scene, almost the entire soil background was classified as a ground shadow region (Figure 2.9d). In both cases, however, green-related pixels (plant materials) were not included in the ground shadow region, leading to no significant loss of vegetation pixels in the shadow removal process (Figure 2.9e). The last column in Figure 2.9 contains some examples of vegetation pixel loss with shadow removal indicated by circles.

The ROC and precision-recall curves with a shadow image before and after the ground shadow removal are shown in Figure 2.10. The AUC before and after shadow removal in ROC analysis were 0.944 and 0.987 respectively, and those in the precision-recall analysis were 0.729 and 0.908 respectively. Both curves showed that after ground shadow removal the performance improved and the vegetation segmentation (with a shadow image) succeeded ($AUC \geq 0.9$).

The ROC and precision-recall curves are shown in Figure 2.11 for segmentation with ground shadow removal applied to an image which contained no shadows. Then, the AUC values before and after shadow removal in ROC analysis were 0.981 and 0.980 respectively, and those in the precision-recall analysis were 0.957 and 0.951 respectively. Both curves showed that the performance was not considerably different before and after ground shadow removal. The vegetation segmentation in an image without shadows was successful even after ground shadow removal. The ground shadow removal led to better performance of vegetation segmentation when applied to an image containing a shadow and did not negatively affect the result when applied to an image without shadows (Figure 2.10 and 2.11).

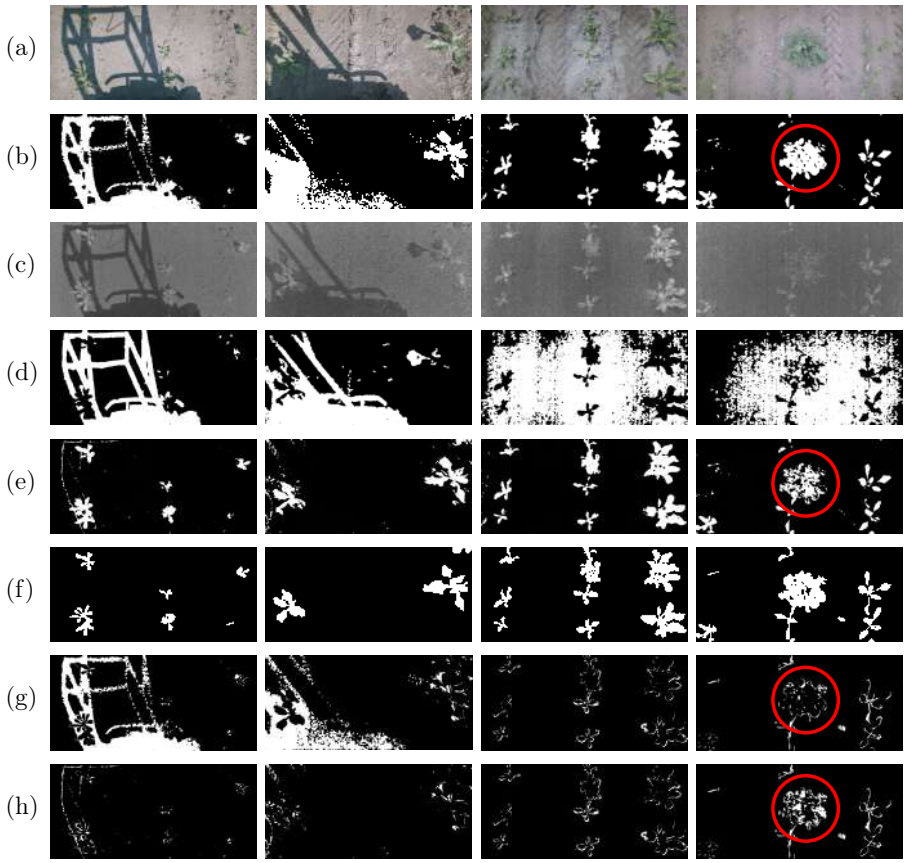
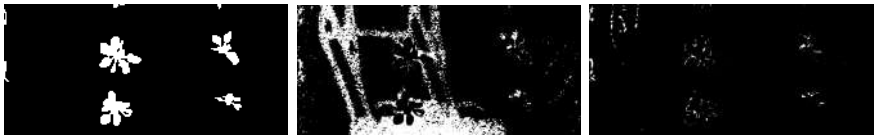
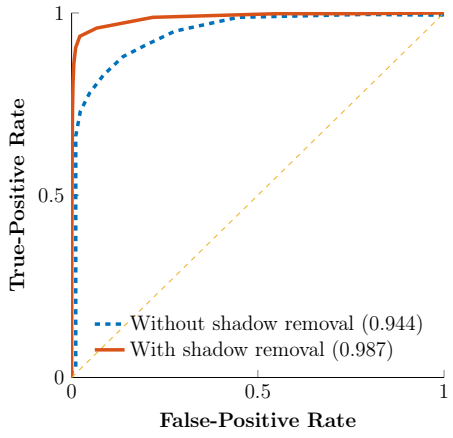


Figure 2.9: Example of images with vegetation segmentation and ground shadow removal process: (a) original image, (b) vegetation segmentation without shadow removal (ExG+Otsu), (c) contrasted ground shadow region, (d) ground shadow detected, (e) vegetation segmentation with shadow removal (ExG+Otsu+shadow removal), (f) vegetation ground truth, (g) difference between (b) and (f), and (h) difference between (e) and (f). Indicated with circles in the last column are some vegetation pixels lost during shadow removal.

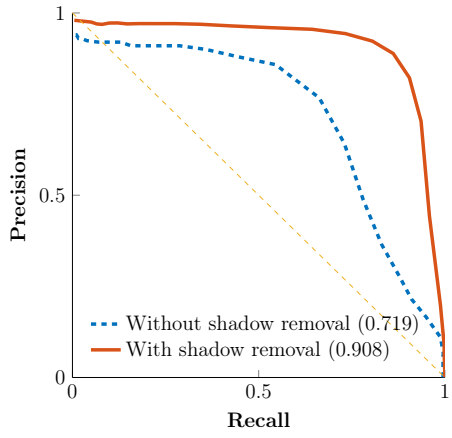
(a) Original image (b) Without shadow removal (ExG+Otsu) (c) With shadow removal (ExG+Otsu+SR)



(d) Ground truth (GT) (e) Difference between (b) and (d) (f) Difference between (c) and (d)

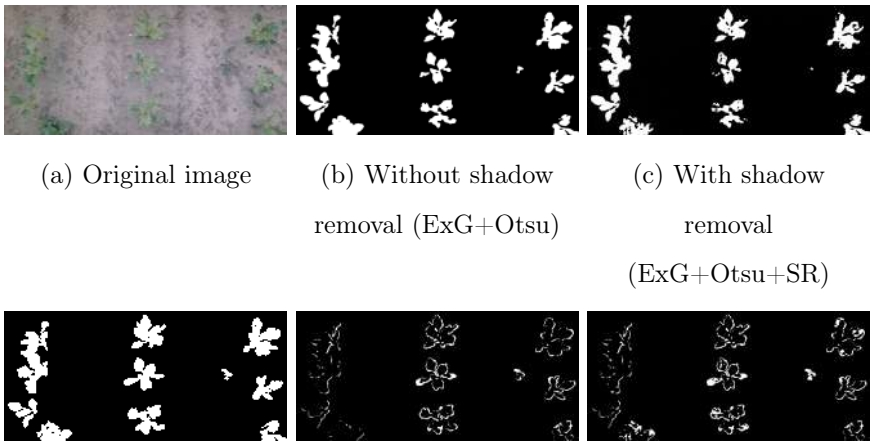


(g) ROC curve



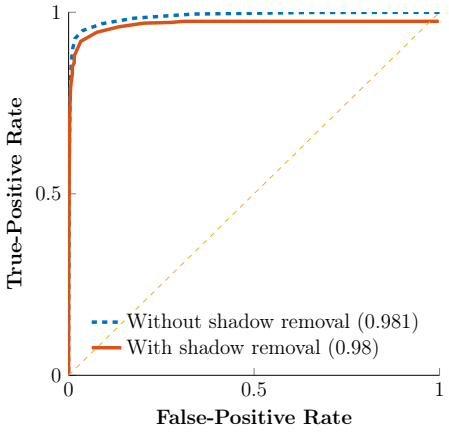
(h) Precision-Recall curve

Figure 2.10: Segmentation of a shadow image with ground shadow removal and its performance analysis: (a) original image, (b) vegetation segmentation without shadow removal (ExG+Otsu), (c) vegetation segmentation with shadow removal (ExG+Otsu+shadow removal), (d) ground truth, (e) difference between (b) and (d), (f) difference between (c) and (d), (g) and (h) ROC and precision-recall curves for vegetation segmentation with and without shadow removal, the area under curve (AUC) in parenthesis.

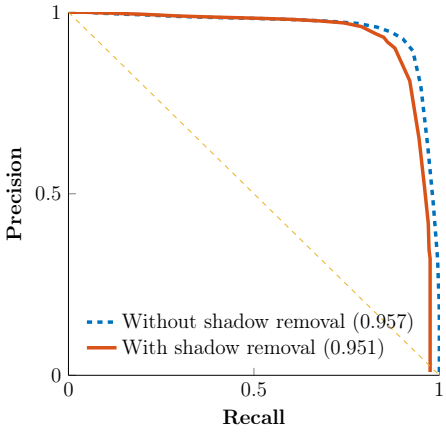


(a) Original image (b) Without shadow removal (ExG+Otsu) (c) With shadow removal (ExG+Otsu+SR)

(d) Ground truth (GT) (e) Difference between (b) and (d) (f) Difference between (c) and (d)



(g) ROC curve



(h) Precision-Recall curve

Figure 2.11: Segmentation of a non-shadow image with and without ground shadow removal and its performance analysis: (a) original image, (b) vegetation segmentation without shadow removal (ExG+Otsu), (c) vegetation segmentation with shadow removal (ExG+Otsu+shadow removal), (d) ground truth, (e) difference between (b) and (d), (f) difference between (c) and (d), (g) and (h) ROC and precision-recall curves for vegetation segmentation with and without shadow removal, the area under curve (AUC) in parenthesis.

2.4 Discussion

2.4.1 High Dynamic Range (HDR) camera

The HDR camera enabled capture of quality images in the high dynamic range scene of the field. During the field image acquisition, no image saturation caused by strong direct solar radiation was observed, and plants under sharp dark shadows were still clearly noticeable. Two other studies (Dworak et al., 2013; Piron et al., 2010) also reported that high dynamic range acquisition enabled a strong signal to noise ratio for all pixels of the image as well as a better Normalized Difference Vegetation Index (NDVI). In this study, however, an HDR and a conventional non-HDR cameras were not simultaneously used in parallel in the field. Thus, a quantitative comparison between these two cameras under agricultural field conditions could not be made. However, the added value of using a HDR camera is expected in the agricultural field under natural light conditions.

2.4.2 Shadow detection and removal

Although the proposed algorithm effectively detects and removes ground shadows, and thus improves the performance of vegetation segmentation, the algorithm itself alone does not extract any green material. The algorithm has to be combined with vegetation extraction methods (vegetation index), such as ExG, NDVI and CIVE. However, the shadow detection and removal algorithm is not limited to any specific vegetation extraction method because the algorithm is a separate procedure that can work as an add-on process.

The algorithm is based on color space conversion and chromaticity difference. This approach is simple, easy-to-implement and computationally inexpensive. Sanin et al. (2012) reported that the color space conversion approach needed the least computation time among the reviewed methods. However, the color space conversion approach requires the selection of an optimal threshold value that relies on the assumption that the image scene consists of a fixed number of components. In this study, a hypothesis was made that the field image scene can be divided into three classes: vegetation (green plants), background (soil) and ground shadow. Although hardly any other materials than these three were found in field images, a crop image scene may contain, according to Yang et al. (2015), various kinds of straw, straw ash or non-green plants. If an

image scene contains a significant amount of the above mentioned or other materials, the algorithm may have limited performance.

There were few losses of true-positive pixels (vegetation pixels) during the shadow removal process. Although this loss was not critical, there are two ways to improve this procedure: post-image processing, and selective application of shadow removal. Post-image processing such as a hole filling or erode/dilate operation can recover some true-positive pixels that were lost during the removal process. Alternatively, applying shadow removal only when an image contains a shadow can improve the performance. In this study, shadow removal was applied to all images (images with and without shadows). However, technically there is no need to apply shadow removal when the image contains no shadows. This selective approach, however, would require a procedure that detects the presence of shadows in a given image scene. An alternative might be to use an illuminant-invariant image based on physical models of illumination and colors (Álvarez & Lopez, 2011; Finlayson et al., 2006).

The processing time for vegetation segmentation with shadow removal was 0.46 s, and this should be acceptable in a real-time application (< 1 s *Required*). There is a way to further reduce the processing time. If the processing time is highly critical for certain applications, a faster processor with multiple/parallel processing implementations might be an alternative approach to reduce the processing time.

2.5 Conclusions

In this study, a ground shadow detection and removal method based on color space conversion and multi-level threshold was proposed. This method is to be used in a real-time automated weed detection and control system that has to operate under natural light conditions. Then vegetation segmentation is challenging due to shadows.

Applying shadow removal improved the performance of vegetation segmentation with an average improvement of 20%, 4.4% and 13.5% in precision, specificity and modified accuracy, respectively, compared with no shadow removal. The average processing time for vegetation segmentation with shadow removal was 0.46 s, which is acceptable for the real-time application (< 1 s *Required*).

The proposed method for ground shadow detection and removal enhances the performance of vegetation segmentation under natural illumination conditions in the field, and is feasible for real-time field applications and does not reduce segmentation

performance when shadows are not present.

2.6 Acknowledgements

The work presented in this study was part of the Agrobot part of the Smartbot project and funded by Interreg IVa, European Fund for the Regional Development of the European Union and Product Board for Arable Farming. We thank Gerard Derks at experimental farm Unifarm of Wageningen University & Research for arranging and managing the experimental fields. We also thank Wim-Peter Dirks for his contribution to creating the ground truth.

CHAPTER 3

Investigation on combinations of colour indices and threshold techniques in vegetation segmentation for volunteer potato control in sugar beet

Hyun K. Suh
Jan Willem Hofstee
Eldert J. van Henten

The contents of this chapter have been submitted to *Computers and Electronics in Agriculture* as a paper entitled: Investigation on combinations of colour indices and threshold techniques in vegetation segmentation for volunteer potato control in sugar beet.

Abstract

Robust vegetation segmentation is required for a vision-based weed control robot in an agricultural field operation. The output of vegetation segmentation is a fundamental element in the subsequent process of weed/crop discrimination as well as weed control actuation. Given the abundance of colour indices and thresholding techniques, it is still far from clear how to choose a proper threshold technique in combination with a colour index for vegetation segmentation under agricultural field conditions. In this research, the performance of 40 combinations of eight colour indices and five thresholding techniques was assessed to identify which combination works the best given varying field conditions in terms of illumination intensity, shadow presence and plant size. It was also assessed whether it was better to use one specific combination at all times or whether the combination should be adapted to the field conditions at hand. A clear difference in performance, represented in terms of MA (Modified Accuracy), was observed among various combinations under the given conditions. On the image dataset that was used in this study, CIVE+Kapur showed the best performance while VEG+Kapur showed the worst. When adapting the combination to the given conditions yielded a slightly higher performance than when using a single combination for all (in this case CIVE+Kapur). Consistent results were obtained when validated on a different independent image dataset. The expected advantage of adapting the combination to the field condition is not large because it seems that for practical use, the slight improvement when adapting the combination to the field conditions does not outweigh the investment in sensor technology and software needed to accurately determine the different conditions in the field.

3.1 Introduction

Within the EU-funded project SmartBot, a small-sized robot was developed to be used for vision-based precise control of volunteer potato (weed) in a sugar beet field. Due to its small size and the necessary battery operation, the platform design had to refrain from additional infrastructure and should be able to robustly detect weeds in a scene that is fully exposed to ambient lighting conditions. Additional infrastructure such as a hood and lighting equipment to overcome the challenges of ambient lighting conditions in the field, like the one used by for instance Nieuwenhuizen et al. (2010) and Haug et al. (2014), was considered not viable. Additionally, the system had to deal with different sizes of cash crop and weeds as well as shadows which are unavoidable in a system like this.

Such a vision-based weed control robot requires robust vegetation segmentation, i.e. a vegetation segmentation that has good performance under a wide range of circumstances. The output of vegetation segmentation is the fundamental element in the subsequent process of weed and crop discrimination as well as weed control (Suh et al., 2018b). It is challenging to come up with a proper segmentation of vegetation from soil under field conditions with varying natural illumination (Hernández-Hernández et al., 2016).

The segmentation of vegetation can be done in three ways (Guijarro et al., 2015): 1) using colour-based indices, 2) learning-based methods, and 3) discrete wavelet transform. The colour-based indices are easier to comprehend as well as simple to implement and they are the most commonly used approaches in agricultural applications; while learning-based methods and the discrete wavelet transform require extensive domain knowledge (Guijarro et al., 2015; Guo et al., 2013; Romeo et al., 2013). While the latter will have to be explored further for potential benefits, colour-based indices remain popular and common in many applications nowadays because the colour is a crucial feature for plant recognition (Hernández-Hernández et al., 2016). Still, in colour-based segmentation, some questions remained unanswered so far, and this paper aims to address some of these questions. They will be defined in more detail hereafter.

The segmentation of vegetation using colour-based indices principally contains two steps (Guerrero et al., 2013; Tellaeche et al., 2008): 1) transformation of the RGB image into a near-binary intensity image (monochrome), and 2) application of a

threshold to convert the near-binary image to a full-binary image

Several methods have been used for the transformation of the RGB image into a near-binary image: ExG (Excess Green), CIVE (Colour Index of Vegetation Extraction), NDI (Normalized Difference Index), ExGR (Excess Green minus Excess Red), VEG (Vegetative Index), COM (Combination of Green), GA (Greenness Accentuation), and HIT (Hue-Invariant Transformation) (Guerrero et al., 2013; Hague et al., 2006; Hamuda et al., 2016; Kataoka et al., 2003; Lati et al., 2013b; Meyer & Camargo Neto, 2008). Each of these indices uses different mathematical formulae for near-binary transformation, but all of them have essentially been proposed to enhance the differentiation between the pixels associated with the vegetation (green pixels) and the pixels related to the background (soil pixels). Hamuda et al. (2016) assessed the performance of different colour indices; however, the same test data were not used in all cases which made a direct comparison more difficult. Besides, for segmentation, a colour index always requires a threshold to yield a segmentation result.

For thresholding, several options exist too. A fixed threshold value, typically determined by an empirical analysis, has been widely used. The biggest disadvantage of this type of threshold is that it produces a poor output when the image scene is exposed to varying natural light conditions. Consequently, the threshold needs to be reset depending on the illumination conditions (Burgos-Artizzu et al., 2010). The Otsu method (Otsu, 1979), which uses variance to separate foreground and background classes, is considered a good option to automatically calculate a threshold value for a given image, and it was used in several studies (Guo et al., 2013; Romeo et al., 2013; Shrestha & Steward, 2005). At the same time, other more sophisticated thresholding methods are also reported in the literature, for instance the iterative threshold (Ridler & Calvard, 1978), the max-entropy threshold (Kapur et al., 1985), the minimum-error threshold (Kittler & Illingworth, 1986), and the unimodal threshold (Rosin, 2001). Some studies used abovementioned thresholds for vegetation segmentation. However, hardly any studies have investigated the details of these thresholds and compared the performance of vegetation segmentation in agricultural field conditions.

After all, when it comes to the choice of a combination of colour index and threshold technique for vegetation segmentation, some studies used ExG with Otsu threshold (ExG+Otsu), while some others used ExG with Kapur threshold (ExG+Kapur) or NDI with Otsu threshold (NDI+Otsu) (Meyer & Camargo Neto, 2008; Montalvo et al., 2013; Tellaeche et al., 2008). It is still far from clear how to choose a proper

threshold technique in combination with a colour index for vegetation segmentation under agricultural field conditions. To the best of our knowledge, comparative studies on this matter seem to be lacking so far, and it is still unclear which combination performs best under field conditions given varying illumination conditions, the presence of shadows and differences in plant size.

Based on the above, in this paper three questions will be addressed:

1) Do different combinations of colour index and threshold technique result in different segmentation performance when evaluated on field images? Do certain combinations stand out positively in performance compared to others and, the other way around, do certain combinations stand out negatively when compared to others?

2) If differences in segmentation performance do exist, which combination works the best given the field conditions like illumination intensity, shadow presence and plant size?

3) Given the varying conditions in the field, is it better to use one combination (at all times) or should the combination be adapted to the conditions at hand for best segmentation performance?

3) Do results obtained from 1-3 hold true when validated on a different independent image dataset?

Section 3.2 describes the collection of colour indices as well as threshold techniques used in this research. Section 3.3 describes the experimental setup including field image dataset collection and implemented procedure as well as the performance measures used for evaluation of the segmentation techniques. Then, in Section 3.4, experimental results are then presented followed by the discussion. Lastly, the conclusions are drawn.

3.2 Materials and Methods

In this study, 40 different combinations of colour indices and threshold techniques were evaluated. Details of the eight colour indices and five threshold techniques are described in sections 3.2.1 and 3.2.2, respectively. Section 3.3.1 presents the image

acquisition as well as the selection and categorisation of the images for the image sets used to assess the segmentation performance. Section 3.3.2 contains the performance criteria used to quantify the segmentation performance.

3.2.1 Colour based indices

The RGB images were first normalised using following equations (Guerrero et al., 2012):

$$r = \frac{R_n}{R_n + G_n + B_n}, \quad g = \frac{G_n}{R_n + G_n + B_n}, \quad b = \frac{B_n}{R_n + G_n + B_n} \tag{3.1}$$

where R_n, G_n, and B_n are the normalized RGB coordinates ranging from 0 to 1 and are obtained as follows:

$$R_n = \frac{R}{R_{max}}, \quad G_n = \frac{G}{G_{max}}, \quad B_n = \frac{B}{B_{max}} \tag{3.2}$$

where $R_{max} = G_{max} = B_{max} = 255$ for 24-bit colour images.

ExG (Excess Green)

Woebbecke et al. (1995) introduced the excess green index described as follows:

$$ExG = 2 \cdot g - r - b \tag{3.3}$$

NDI (Normalized Difference Index)

Pérez et al. (2000) used the normalised difference index (NDI), equal to the ratio of the difference and the sum of the green and red colour channels:

$$NDI = \frac{G - R}{G + R} \tag{3.4}$$

CIVE (Colour Index of Vegetation Extraction)

Kataoka et al. (2003) introduced the colour index of vegetation extraction (CIVE) to enhance green information in the image. This index was derived based on the analysis

of the principal components of acquired crop images. The CIVE is defined as follows:

$$CIVE = 0.441 \cdot r - 0.811 \cdot g + 0.385 \cdot b + 18.78745 \tag{3.5}$$

ExGR (Excess Green minus Excess Red)

Meyer et al. (2004) proposed the difference of excess green index (ExG) and excess red index (ExR), based on the observation that red pixel values were found in certain soils and crop residues.

$$ExGR = ExG - ExR = (2 \cdot g - r - b) - (1.4 \cdot r - g) \tag{3.6}$$

VEG (Vegetative Index)

Hague et al. (2006) used a vegetative index (VEG) that is insensitive to illumination changes. They reported that VEG provided good contrast between plant and soil. This index was based on Marchant & Onyango (2000)'s blackbody approximation.

$$VEG = \frac{G}{R^a \cdot B^{(1-a)}} \tag{3.7}$$

where a is a constant value equal to 0.667.

COM (Combination of Green)

Guerrero et al. (2012) introduced a combination of three colour indices: ExG, CIVE, and VEG. They showed that this combination approach provided better performance than the individual application of each index. Each index was weighted based on its relative importance having an overall sum of 1. The combination is defined as follows:

$$COM = 0.36 \cdot ExG + 0.47 \cdot CIVE + 0.17 \cdot VEG \tag{3.8}$$

GA (Greenness Accentuation)

Guerrero et al. (2013) proposed a greenness accentuation based on the reasoning that the pixels associated with plants should have dominant green colour. They obtained a

greenness accentuation by multiplying the COM (Eq. 3.8) by g (Eq. 3.1). The formula for GA is as follows:

$$GA = COM \cdot g \tag{3.9}$$

HIT (Hue-Invariant Transformation)

Lati et al. (2013b) used a hue-invariant transformation based on the xyY colour space. The illumination invariant image was obtained by converting an RGB image into the xyY model followed by log transformation. The formula for HIT is as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1804 \\ 0.2126 & 0.7151 & 0.0721 \\ 0.0193 & 0.1191 & 0.9503 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.10}$$

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z} \tag{3.11}$$

$$HIT = a^{\frac{x}{y}} \tag{3.12}$$

where a is the hue calibration parameter which depends on the camera manufacturer (camera sensor). In this study, the value of 0.45 was used.

An example of an original image and the resulting near-binary intensity images after applying each of the transformations presented before is shown in Figure 3.1.

3.2.2 Thresholding techniques

Thresholding is one of the most common and straightforward techniques for image segmentation. The goal of this process is to convert the near-binary image into a full-binary image. In this binary image, the vegetation pixels are generally represented as white while the background soil pixels are represented as black. From a near-binary image $I(x, y)$, a full-binary image $B(x, y)$ is obtained by applying some threshold T (gray level) as in Eq. 3.13

$$B(x,y) = \begin{cases} 1, \text{if } I(x,y) \geq T \\ 0, \text{if } I(x,y) < T \end{cases} \tag{3.13}$$

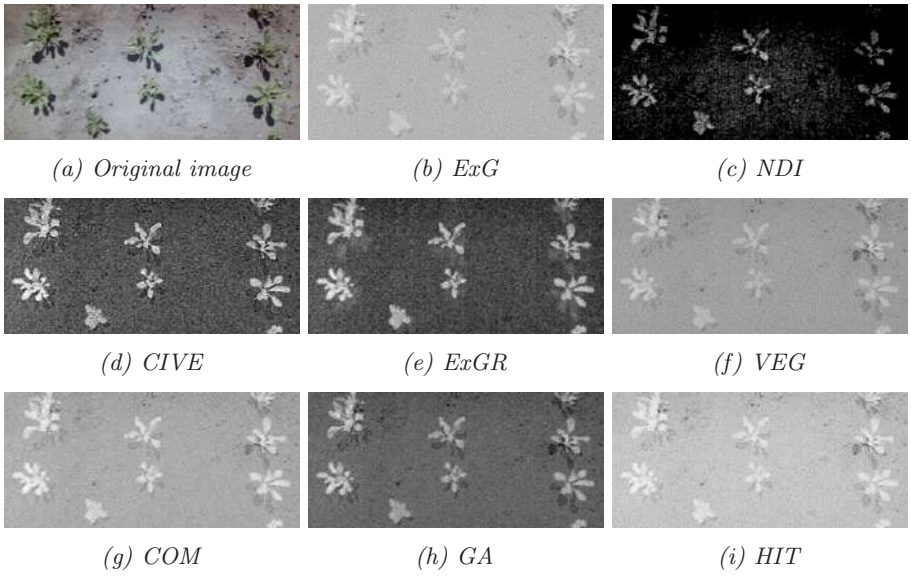
The threshold techniques used in this research are shortly described hereafter.

Otsu (Variance-based threshold)

Otsu's method finds an optimal threshold that maximises the between-class variances, σ_B^2, of the foreground and background classes. Otsu's method is known to be very robust, and provides good results when the intensity distribution of the pixels in an image is bimodal (Chaki et al., 2014). The optimal threshold value, T_{opt}, is calculated as follows (Otsu, 1979):

$$T_{opt} = \underset{0 \leq t < L}{\operatorname{argmax}} \{\sigma_B^2(t)\} \tag{3.14}$$

where T is the threshold value ($0 \leq T < L$), and L is the number of distinct intensity levels.



(a) Original image *(b) ExG* *(c) NDI*

(d) CIVE *(e) ExGR* *(f) VEG*

(g) COM *(h) GA* *(i) HIT*

Figure 3.1: An example of near-binary intensity images after applying the colour transformations to the original image shown in (a).

Ridler (Iterative threshold)

This threshold technique was proposed by Ridler & Calvard (1978). An optimal threshold value is obtained with an intensity value using an iterative procedure. From an initially estimated threshold (e.g. mean image intensity), the optimal threshold value is reestimated in an iterative process using two class means. The iteration continues until the threshold value does not change or the change becomes small. The iterative threshold process is as follows Kaur (2013):

1) An initial threshold value, T, is estimated using a mean image intensity (average intensity of image is assumed as a good initial threhold value).

2) Pixels above and below the threshold are assigned to the crop and soil background, respectively.

3) Of both classes, crop and soil background, mean values, u_1 and u_2, are calculated of the grey values of each class.

4) A new threshold valuee, T, is computed such that value of T is the average of two means u_1 and u_2:

$$T = \frac{u_1 + u_2}{2} \tag{3.15}$$

5) Iterate steps 2-4 until the change in threshold value is less than a very small number.

Kapur (Max-Entropy threshold)

Kapur et al. (1985) proposed entropy-based thresholding. This technique uses the entropy of the foreground and background regions within an image. The optimal threshold is obtained by maximising the sum of the entropy values which can be explained as measures of class compactness and separability (Drobchenko et al., 2011). Sezgin & Sankur (2004) reported that the best thresholding performance was achieved with this max-entropy method. Tellaeche et al. (2008) used this max-entropy thresholding for vegetation segmentation. Following is the formula:

$$T_{opt} = \underset{T}{\operatorname{argmax}}\{H_f(T) + H_b(T)\} \tag{3.16}$$

where the entropy of the foreground, $H_f(T)$ and the entropy of the background, $H_b(T)$, as a function of threshold T, are defined as follows:

$$H_f(T) = -\sum_{g=0}^{T} \frac{h(g)}{P_f(T)} \log \frac{h(g)}{P_f(T)}, \quad H_b(T) = -\sum_{g=T+1}^{G} \frac{h(g)}{P_b(T)} \log \frac{h(g)}{P_b(T)} \tag{3.17}$$

Kittler (Min-Error threshold)

Kittler & Illingworth (1986) considered an error measure in calculating the optimal threshold and proposed minimum error thresholding that is computationally efficient. This method uses a cost function, which is based on the Bayesian classification rule, with the underlying assumption that foreground and background grayscale values are normally distributed (Chaki et al., 2014). The optimal threshold value, T_{opt}, is calculated as follows (Kittler & Illingworth, 1986):

$$\begin{aligned} T_{opt} = \underset{T}{\operatorname{argmax}}\{&[P_f(T)\log\sigma_f(T) + P_b(T)\log\sigma_b(T)] \\ &- [P_f(T)\log P_f(T) + P_b(T)\log P_b(T)]\} \end{aligned} \tag{3.18}$$

where $\sigma_f(T)$ and $\sigma_b(T)$ are the standard deviation of the foreground and background, respectively.

Rosin (Unimodal threshold)

Abovementioned threshold techniques assume that the intensity histogram of the grayscale image is non-unimodal. However, unimodal histogram distribution in an image can also be observed depending on the conditions in an agricultural field. For example, if either one of the classes (foreground or background) dominates the histogram, causing an unbalance between the classes, the intensity histogram becomes unimodal. In such circumstances, many of the standard threshold selection algorithms will fail. Rosin's threshold method (Rosin, 2001) was designed to deal specifically with unimodal histograms. The unimodal threshold process is as follows (Figure 3.2):

1) A straight line is drawn from the peak of the histogram to the last non-zero element of the histogram

2) The optimal threshold is selected at the point of the histogram furthest from the straight line

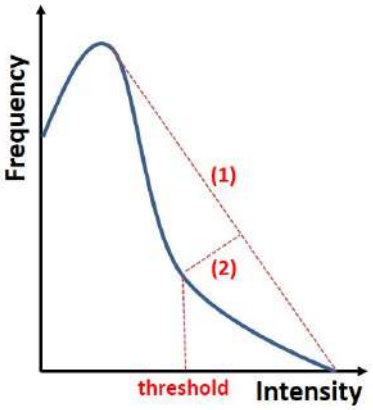


Figure 3.2: The unimodal threshold process of Rosin is described. (1) A straight line is drawn from the peak of the histogram to the last non-zero element of the histogram. (2) The optimal threshold is selected at the point of the histogram furthest from the straight line.

3.3 Experimental setup

3.3.1 Image dataset and experiment

Image acquisition is described in detail in Suh et al. (2018b). Sugar beet was sown three times (Spring, Summer, and Fall) each year in 2013, 2014 and 2015 in sandy and clay soil at Unifarm experimental sites in Wageningen, The Netherlands. One week after sowing the sugar beet, the potato was planted in random locations throughout the fields. The plant images were acquired under a wide range of illumination and weather conditions during several days in June, August and October 2013, in May, June, July and September 2014 and in May, June, July and October 2015.

From all acquired images, a total of 200 images was selected for further analysis. The 200 images were categorized according to plant size (estimated crown diameter of the plant $0-50mm$ indicated with 'S' for category Small; $51-150mm$: indicated with 'M' for category Medium; $>150mm$: indicated with 'L' for category Large), illumination

Table 3.1: The image dataset categories based on plant size (estimated crown diameter of the plant $0-50mm$ indicated with 'S' for category Small; $51-150mm$: indicated with 'M' for category Medium; $> 150mm$: indicated with 'L' for category Large), illumination condition ('S' indicating sunny conditions or 'C' indicating cloudy conditions), presence of shadows ('Y' indicating 'Yes' or 'N' indicating 'No'). The number of images in each group is shown in the last column. A total of 200 images was divided over two Image Subsets, Subset 1 and Subset 2, containing each 100 images and with an almost equal distribution over the 9 image categories as indicated in the Table.

Category	**Plant size**	**Illumination**	**Shadows**	**Number of instances (Image Subset 1 and 2)**
SSY	Small	Sunny	Yes	25 (training set:13, validation set:12)
SSN	Small	Sunny	No	19 (training set:9, validation set:10)
SCN	Small	Cloudy	No	22 (training set:11, validation set:11)
MSY	Medium	Sunny	Yes	26 (training set:13, validation set:13)
MSN	Medium	Sunny	No	23 (training set:11, validation set:12)
MCN	Medium	Cloudy	No	27 (training set:14, validation set:13)
LSY	Large	Sunny	Yes	18 (training set:9, validation set:9)
LSN	Large	Sunny	No	20 (training set:10, validation set:10)
LCN	Large	Cloudy	No	20 (training set:10, validation set:10)
Total				200 (Subset 1: 100, Subset 2: 100)

condition ('S' indicating sunny conditions or 'C' indicating cloudy conditions), presence of shadows ('Y' indicating 'Yes' or 'N' indicating 'No'). This yielded nine categories of field conditions in total, i.e. SSY, SSN, SCN, MSY, MSN, MCN, LSY, LSN and LCN.

While assuring that all nine categories were equally distributed, the 200 images were randomly divided into two subsets of 100 images each, which will be referred to as Image Subset 1 and Image Subset 2. See Table 3.1 for details. Example images of each group are shown in Figure 3.3.

Then the following procedure was implemented to answer the four research questions mentioned in the introduction:

1) To answer research question 1, of all 40 combinations of vegetation index and

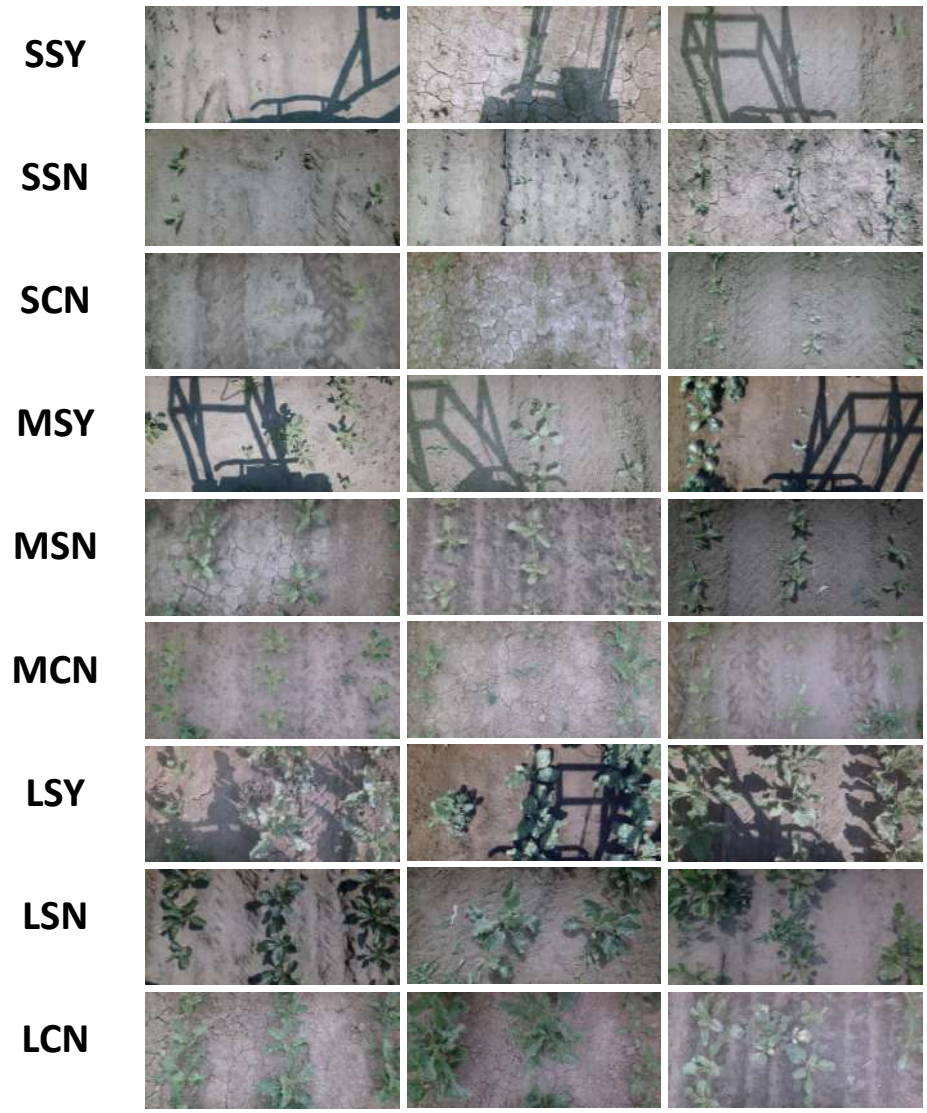


Figure 3.3: Example images in each group in the image dataset (Table 3.1). During the selection of the image dataset, a wide range of natural conditions was considered, including different stages of plant growth, illumination conditions from a cloudy to a sunny day, and extreme illumination scenes caused by strong direct sunlight and resulting in shadows.

threshold technique the segmentation performance was assessed on Image Subset 1 with the aim to identify combinations that stand out positively in performance when compared to others and to identify combinations that stand out negatively when compared to others.

2) To answer research question 2, the segmentation performance on Image Subset 1 was analysed to identify which combination works the best for each of the nine categories of field conditions: SSY, SSN, SCN, MSY, MSN, MCN, LSY, LSN and LCN.

3) To answer research question 3, segmentation performance on the whole Image Subset 1 was assessed using 1) a single fixed combination of colour index and threshold technique for the whole dataset (the best combination obtained in procedure 1 was used), or 2) using the best combination of colour index and threshold technique for each of the nine categories, as obtained in procedure 2. In this way, the potential advantage of adapting the colour index and threshold combination to the field conditions was evaluated.

4) The results obtained under 1 to 3 were validated on Image Subset 2 to assess the results obtained from the procedure 1 to 3 on Image Subset 1 hold true on a different independent image dataset.

Vegetation regions were manually labelled for ground truth. All images were processed with Image Processing Toolbox™ in Matlab 2015b (The MathWorks Inc., Natick, USA) on a PC equipped with an Intel® Core™ i7-377T 2.5 GHz processor and 8 GB RAM running 64-bit Windows 7.

3.3.2 Criteria for evaluating segmentation performance

The segmentation results were compared and evaluated pixel-to-pixel with human-labelled ground truth images, using the modified accuracy (MA) as a representative overall performance measure (Eq. 3.21). The modified accuracy (MA) is the harmonic mean of relative vegetation area error (RVAE) and balanced accuracy (BA) and is valued between 0 and 1, where 0 represents the poorest segmentation, and 1 the best segmentation. The harmonic mean indicates if there is a significant imbalance between

RVAE and BA, and thus provides a better description of the performance. The details of the performance measures are described in Suh et al. (2018b).

$$Balanced\ accuracy\ (BA) = \frac{Recall + Specificity}{2} \tag{3.19}$$

$$Relative\ Vegetation\ Area\ Error\ (RVAE) = \begin{cases} 1 - \frac{A_{GT} - A_{SEG}}{A_{GT}} & \text{if } A_{SEG} < A_{GT} \\ 1 - \frac{A_{SEG} - A_{GT}}{A_{SEG}} & \text{if } A_{GT} \leq A_{SEG} \end{cases} \tag{3.20}$$

$$Modified\ accuracy\ (MA) = \frac{2 \cdot BA \cdot RVAE}{BA + RVAE} \tag{3.21}$$

where: A_{GT} is the vegetation area in ground truth (TP+FN); A_{SEG} is the vegetation area in segmented image (TP+FP).

3.4 Results

3.4.1 Vegetation segmentation performance of all combinations of colour indices and threshold techniques on Image Subset 1

Among 40 combinations of colour indices and threshold techniques, the top five highest-performing combinations, as well as the bottom five lowest-performing combinations on Image Subset 1, are shown in Table 3.2. The top five high performing combinations were found to be CIVE+Kapur, CIVE+Rosin, CIVE+Kittler, ExGR+Kapur, and GA+Rosin; while the five poorest performing combinations were NDI+Kittler, ExG+Kapur, NDI+Rosin, NDI+Kapur, and VEG+Kapur.

Figure 3.4 shows the MA of all combinations of colour indices and threshold techniques on Image Subset 1 as box-and-whisker plots. MA values widely vary over the whole range from 0 to 1 for all combinations except for CIVE+Kapur, a combination that showed less variations in MA.

The results for CIVE+Kapur, CIVE+Rosin, CIVE+Kittler, and GA+Rosin that were listed in the top five high performing combinations (Table 3.2) show that in about 75% of the images they produced a MA of 0.6 and higher in vegetation segmentation.

Table 3.2: Segmentation performance expressed as MA (Modified Accuracy) of the top five highest-performing combinations as well as the bottom five lowest-performing combinations, assessed on Image Subset 1.

	Rank	Combination	MA (Modified Accuracy)
The five highest-performing combinations	1	CIVE+Kapur	0.87
	2	CIVE+Rosin	0.81
	3	CIVE+Kittler	0.79
	4	ExGR+Kapur	0.73
	5	GA+Rosin	0.73
The five lowest-performing combinations	36	NDI+Kittler	0.56
	37	ExG+Kapur	0.53
	38	NDI+Rosin	0.47
	39	NDI+Kapur	0.44
	40	VEG+Kapur	0.43

Among these combinations, CIVE+Kapur showed the highest performance as about 75% of the images produced a MA value of 0.82 and higher. The minimum MA value obtained with CIVE+Kapur was 0.46, which was considerably higher than any other combination's minimum value. Figure 3.5 and 3.6 show example cases where CIVE+Kapur showed better performance in vegetation segmentation than other combinations under sunny and cloudy conditions. However, in some cases, CIVE+Kapur produced a poor vegetation segmentation result while the other methods such as CIVE+Ridler, CIVE+Kittler, CIVE+Otsu, and CIVE+Rosin yielded good results (Figure 3.7). In this case, the threshold value of Kapur was considerably higher than others. It is worth noting that for that case the histogram has a multimodal distribution as shown in Figure 3.7i.

In general, various thresholds performed better in combination with CIVE than in combination with other colour indices. CIVE in combination with Kapur, Kittler, and Rosin showed an average MA of 0.79 and higher. However, CIVE together with Otsu and Ridler showed large performance variations.

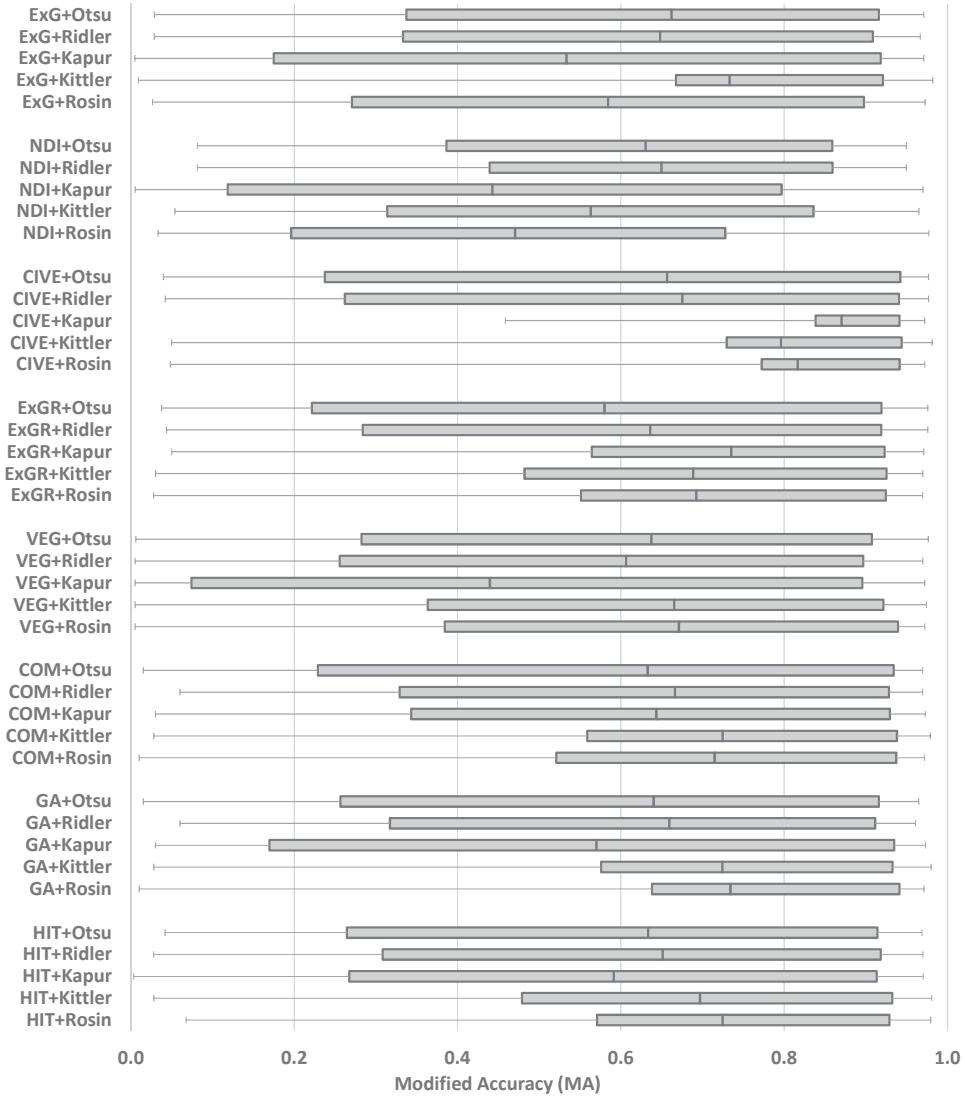
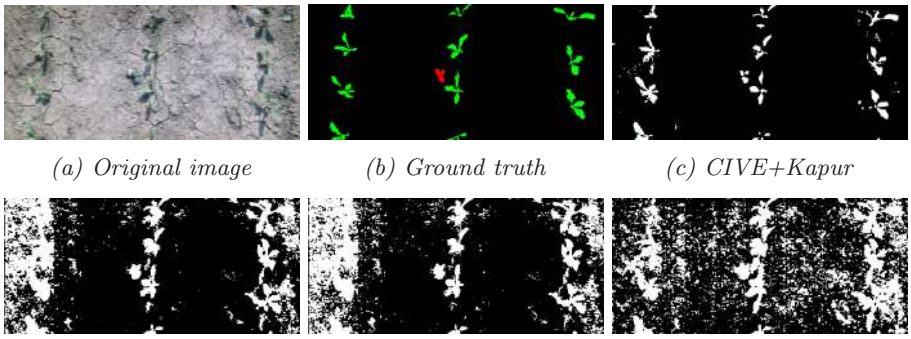


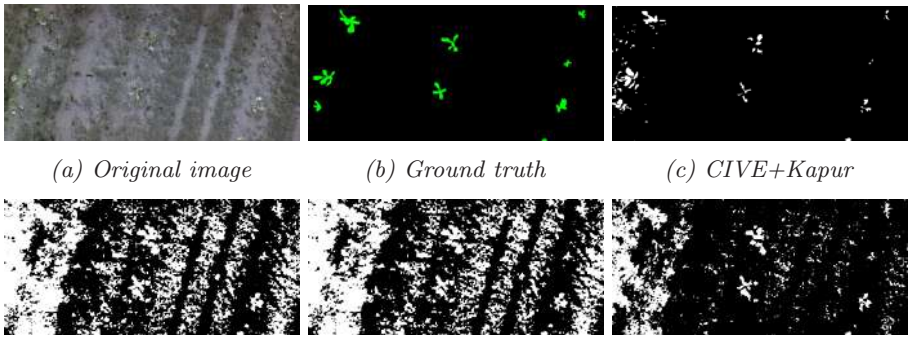
Figure 3.4: Box-and-whiskers plot of the MA (Modified Accuracies) for all the combinations of colour indices and threshold techniques evaluated on Image Subset 1. The left and right bars indicate the min to max values for each combination, whereas the box and centre line indicate the Q1 (1st quartile), mean, and Q3 (3rd quartile).



(a) Original image *(b) Ground truth* *(c) CIVE+Kapur*

(d) CIVE+Otsu *(e) ExG+Rosin* *(f) NDI+Ridler*

Figure 3.5: An example case for which CIVE+Kapur showed better performance in vegetation segmentation than any other combinations under sunny conditions.



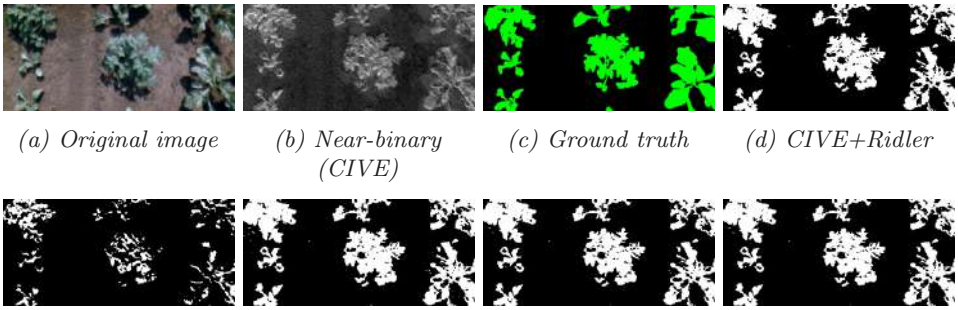
(a) Original image *(b) Ground truth* *(c) CIVE+Kapur*

(d) COM+Otsu *(e) GA+Rosin* *(f) VEG+Kittler*

Figure 3.6: An example case for which CIVE+Kapur showed better performance in vegetation segmentation than any other combinations under cloudy condition.

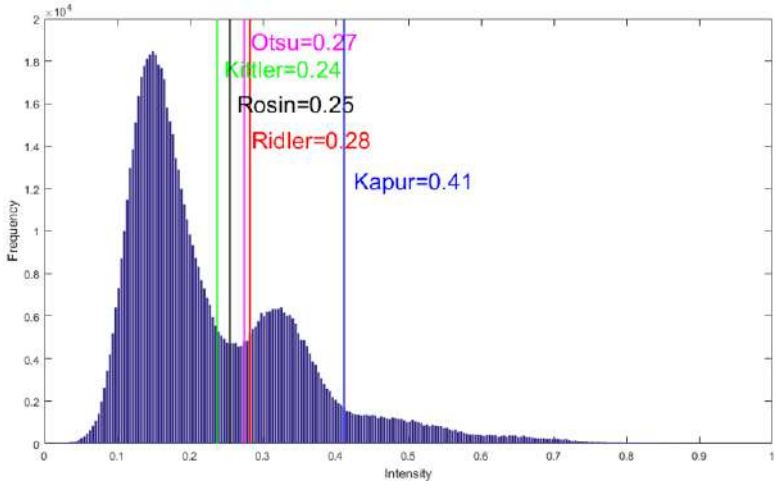
3.4.2 Assessing the best combination for each of the nine image categories in Image Subset 1

Table 3.3 shows which combination performs best for each of the nine image categories in Image Subset 1. In total, six different combinations appeared to be the best combination for a particular image category; CIVE+Kapur in case of four different image categories and CIVE+Kittler, CIVE+Otsu, COM+Kapur, GA+Rosin, and HIT+Kittler each in case of one of the categories



(a) Original image (b) Near-binary (CIVE) (c) Ground truth (d) CIVE+Ridler

(e) CIVE+Kapur (f) CIVE+Kittler (g) CIVE+Otsu (h) CIVE+Rosin



(i) Histogram with threshold values

Figure 3.7: An example case for which CIVE+Kapur produced poor vegetation segmentation while the other methods CIVE+Ridler, CIVE+Kittler, CIVE+Otsu, and CIVE+Rosin yielded good results. (a) The original image, (b) near-binary intensity image transformed from RGB image using CIVE, (c) ground truth for vegetation segmentation, (d)-(h) resulting images of the combinations of CIVE with threshold techniques, and (i) histogram of near-binary intensity image with threshold techniques and threshold values (drawn in vertical lines). In this example, the threshold value of Kapur was considerably higher than others. It is worth noting the histogram has a multimodal character.

Table 3.3: The best combination of colour index and threshold technique is shown for each category on Image Subset 1. In total, six different combinations appeared to be the best combination; CIVE+Kapur in case of four different image categories and CIVE+Kittler, CIVE+Otsu, COM+Kapur, GA+Rosin, and HIT+Kittler each in case of one of the categories.

Category	Best combination	Mean MA	Std. Deviation MA
SSY	CIVE+Kapur	0.76	0.11
SSN	CIVE+Kittler	0.81	0.14
SCN	COM+Kapur	0.91	0.04
MSY	CIVE+Kapur	0.87	0.09
MSN	CIVE+Kapur	0.90	0.06
MCN	GA+Rosin	0.95	0.02
LSY	CIVE+Kapur	0.88	0.11
LSN	CIVE+Otsu	0.92	0.03
LCN	HIT+Kittler	0.94	0.04

3.4.3 Assessing the advantages of adapting the combination to the field conditions for Image Subset 1

Table 3.4 shows for Image Subset 1 the overall MA when using one combination for all image categories (in this case CIVE+Kapur) and when using for each image category the combination which performed best, as found in section 3.4.2. When adapting the combination to the image category yielded a slightly higher MA (0.88) than when using a single combination for the whole image set (MA of 0.86).

3.4.4 Validation of results on Image Subset 2

Finally, the results obtained in steps 1 to 3 (sections 3.4.1 to 3.4.3) were validated on Image Subset 2 to assess whether the results obtained from the procedure 1 to 3 on Image Subset 1 hold true on a different independent image dataset, Image Subset 2.

To start with, for Image Subset 2, the performance of all 40 combinations were assessed. The five highest-performing and five lowest-performing combinations are

Table 3.4: Overall MA of two approaches is shown on Image Subset 1: using one combination for all categories (CIVE+Kapur) vs. using the best performing category specific combination.

Category	Using one combination obtained for all categories in Image Subset 1	Using the best performing category specific combination for Image Subset 2
SSY	CIVE+Kapur	CIVE+Kapur
SSN	CIVE+Kapur	CIVE+Kittler
SCN	CIVE+Kapur	COM+Kapur
MSY	CIVE+Kapur	CIVE+Kapur
MSN	CIVE+Kapur	CIVE+Kapur
MCN	CIVE+Kapur	GA+Rosin
LSY	CIVE+Kapur	CIVE+Kapur
LSN	CIVE+Kapur	CIVE+Otsu
LCN	CIVE+Kapur	HIT+Kittler
Overall MA	**0.86**	**0.88**

shown in Table 3.5. For this image dataset, the top best performers were CIVE+Kapur, CIVE+Rosin and CIVE+Kittler; whereas the poor performers were NDI+Rosin, NDI+Kapur and VEG+Kapur.

The three highest-performing combinations were consistent with the results obtained on Image Subset 1 as shown in section 3.4.1. Consistent results were also found for the four lowest-performing combinations. In section 3.4.1, ExGR+Kapur and GA+Rosin were the fourth and fifth highest-performing combinations respectively on Image Subset 1, but on Image Subset 2 they were ranked in sixth and seventh place with a MA of 0.74 and 0.73, respectively. These performances, however, did not significantly differ from the performances of the two combinations in the fourth and fifth highest-performing combinations (MA of 0.75 and 0.74) for Image Subset 2. Similarly, in section 3.4.1 for Image Subset 1, NDI+Kittler was ranked in 36th

Table 3.5: Segmentation performance expressed as Modified Accuracy of the top five of high performing combinations as well as the five lowest-performing combinations, assessed on Image Subset 2.

	Rank	Combination	MA (Modified Accuracy)
Top five highest-performing combinations	1	CIVE+Kapur	0.85
	2	CIVE+Rosin	0.82
	3	CIVE+Kittler	0.78
	4	ExG+Kittler	0.75
	5	GA+Kittler	0.74
Bottom five lowest-performing combinations	36	GA+Kapur	0.56
	37	ExG+Kapur	0.53
	38	NDI+Rosin	0.48
	39	NDI+Kapur	0.40
	40	VEG+Kapur	0.38

place with a MA of 0.56, but this combination was ranked in 34th place with MA of 0.58 for Image Subset 2. However, the performance was not considerably different between NDI+Kittler and the one that was ranked in 36th place on Image Subset 2 (GA+Kapur).

In Table 3.6, overall MA of three different approaches validated on Image Subset 2 is shown:

1) Using one combination obtained in Image Subset 1 for all categories in Image Subset 2 (in this case CIVE+Kapur);

2) Using the category specific combination obtained from Image Subset 1 (best performers for each of the nine categories on Image Subset 1 in section 3.4.2) on Image Subset 2;

3) Using the best performing category specific combination for Image Subset 2.

Again, using an adapted combination for each image category yielded a slightly better MA (0.86 and 0.87 in approach 2 and 3, respectively) than using one combination

Table 3.6: Overall MA of different approaches validated on Image Subset 2 is shown: 1) using one combination obtained in Image Subset 1 for all categories in Image Subset 2, 2) using the category specific combination obtained from Image Subset 1 on Image Subset 2, and 3) using the best performing category specific combination for Image Subset 2.

Category	Using one combination obtained in Image Subset 1 for all categories in Image Subset 2	Using the category specific combination obtained from Image Subset 1 on Image Subset 2	Using the best performing category specific combination for Image Subset 2
SSY	CIVE+Kapur	CIVE+Kapur	CIVE+Kapur
SSN	CIVE+Kapur	CIVE+Kittler	CIVE+Rosin
SCN	CIVE+Kapur	COM+Kapur	HIT+Kapur
MSY	CIVE+Kapur	CIVE+Kapur	CIVE+Kapur
MSN	CIVE+Kapur	CIVE+Kapur	CIVE+Kapur
MCN	CIVE+Kapur	GA+Rosin	CIVE+Rosin
LSY	CIVE+Kapur	CIVE+Kapur	CIVE+Kapur
LSN	CIVE+Kapur	CIVE+Otsu	CIVE+Otsu
LCN	CIVE+Kapur	HIT+Kittler	HIT+Kittler
Overall MA	**0.85**	**0.86**	**0.87**

for all image categories (MA of 0.85). Although the differences are minute, the adaptive approach 3 performed the best as was expected. Adaptive approach 2, based on Image Subset 2, performed only very slightly less. And in line with the results represented in Table 3.4, using one combination yielded again a slightly worse performance.

3.5 Discussion

Under the given conditions of this research, some combinations of colour index and threshold technique performed consistently well while some other combinations performed consistently poor. For example, the combination of CIVE and Kapur performed best on both Image Subset 1 and 2. Interestingly enough, however, to the best of the knowledge, CIVE+Kapur has not been proposed or used for vegetation segmentation

under natural field conditions before. Supporting our findings, the colour index CIVE was used in recent studies. For instance, Ye et al. (2015) reported that CIVE showed better and stable performance over ExG, NDI and VEG. Hamuda et al. (2016) suggested using CIVE under cloudy and overcast conditions because CIVE showed a better performance under these conditions than other colour-based indices. Their findings are in line with the results obtained in this study.

The combination of ExG and Kapur, for example, performed consistently poor on both Image Subset 1 and 2. Yet Tellaeche et al. (2008) used ExG+Kapur for vegetation segmentation to identify weeds in corn crops. They indicated that the best performance was achieved with Kapur threshold, but the paper did not provide sufficient detail to get proper insight into these results. These contradicting findings might indicate that selection of a combination of colour index and threshold technique is sensitive to plant type and thus this needs further validation in a future study.

Regarding the Kapur threshold, several studies indeed reported Kapur generally performed better than Otsu, Ridler, or Rosin thresholds (Bhandari et al., 2015; Oliva et al., 2014; Rosin & Ioannidis, 2003; Sezgin & Sankur, 2004). Again, in this study, the highest performance was shown in the combination of CIVE with Kapur on both Image Subset 1 and 2. Under noisy conditions, however, the Kapur threshold was reported to tend to perform poorly leading to a multimodal intensity histogram (Prasad et al., 2011; Su & Amer, 2006). This was observed in this study as well, as was shown in Figure 3.7, when the images produced a multimodal histogram due to noise or irregular illuminations, Kapur performed poorer than other thresholds. In a recent study, Zheng et al. (2017) proposed an entropic thresholding based on Kapur that takes into account the spatial correlation between pixels to improve the performance of Kapur threshold. They reported their proposed approach performed better than a traditional Kapur threshold when using some of the images from the Berkeley image segmentation dataset (Martin et al., 2001). A future study topic might be to evaluate their approach for vegetation segmentation.

The results of this research indicate that for volunteer potato in sugar beet, using different combinations for vegetation segmentation depending on the field conditions like light intensity, shadow presence and plant size might be beneficial although the performance difference compared to using a single combination for all images, in this case CIVE+Kapur, was not very large. To the best of our knowledge, there is no evidence in the scientific literature supporting or contradicting these findings. It seems

worth a while to investigate this aspect further and to expand the set of conditions to include for instance different crop types. Given the current findings, it seems that for practical use, the slight improvement when adapting the combination to the field conditions does not outweigh the investment in sensor technology and software needed to accurately determine the different conditions in the field.

3.6 Conclusions

In this paper, the performance of 40 combinations of eight colour and five thresholding techniques was evaluated under natural field conditions. A clear difference in performance, represented in terms of MA (Modified Accuracy), was observed among various combinations under the given conditions. CIVE+Kapur showed the best performance on Image Subset 1, while VEG+Kapur showed the worst performance on the dataset.

In a total of nine image categories related to light intensity, shadow presence and plant size in Image Subset 1, six different combinations appeared to be the best combination; CIVE+Kapur in case of four different image categories and CIVE+Kittler, CIVE+Otsu, COM+Kapur, GA+Rosin, and HIT+Kittler each in case of one of the categories. When adapting the combination to the image category yielded a slightly higher MA (0.88) than when using a single combination for the whole image set (MA of 0.86) on Image Subset 1. The expected advantage of adapting the combination to the field condition was not large.

The results obtained from the procedure 1 to 3 on Image Subset 1 were consistent when validated on a different independent image dataset, Image Subset 2.

Using different combinations for vegetation segmentation depending on the field conditions like light intensity, shadow presence and plant size might be beneficial although the performance difference compared to using a single combination for all images, in this case CIVE+Kapur, was not very large. Given the current findings, it seems that for practical use, the slight improvement when adapting the combination to the field conditions does not outweigh the investment in sensor technology and software needed to accurately determine the different conditions in the field.

3.7 Acknowledgements

The work presented in this study was part of the Agrobot part of the Smartbot project and funded by Interreg IVa, European Fund for the Regional Development of the European Union and Product Board for Arable Farming. We thank Gerard Derks at experimental farm Unifarm of Wageningen University & Research for arranging and managing the experimental fields. We also thank Wim-Peter Dirks for his contribution to creating the ground truth.

CHAPTER 4

Sugar beet and volunteer potato classification using Bag-of-Visual-Words model, Scale-Invariant Feature Transform, or Speeded Up Robust Feature descriptors and crop row information

Hyun K. Suh
Jan Willem Hofstee
Joris IJsselmuiden
Eldert J. van Henten

The contents of this chapter have been published in *Biosystems Engineering (2018), 166, 210-226* as a paper entitled: Sugar beet and volunteer potato classification using Bag-of-Visual-Words model, Scale-Invariant Feature Transform, or Speeded Up Robust Feature descriptors and crop row information.

Abstract

One of the most important steps in vision-based weed detection systems is the classification of weeds growing amongst crops. In EU SmartBot project it was required to effectively control more than 95% of volunteer potatoes and ensure less than 5% of damage of sugar beet. Classification features such as colour, shape and texture have been used individually or in combination for classification studies but they have proved unable to reach the required classification accuracy under natural and varying daylight conditions. A classification algorithm was developed using a Bag-of-Visual-Words (BoVW) model based on Scale-Invariant Feature Transformation (SIFT) or Speeded Up Robust Feature (SURF) features with crop row information in the form of the Out-of-Row Regional Index (ORRI). The highest classification accuracy (96.5% with zero false-negatives) was obtained using SIFT and ORRI with Support Vector Machine (SVM) which is considerably better than previously reported research although its 7% false-positives deviated from the requirements. The average classification time of $0.10 - 0.11$ s met the real-time requirements. The SIFT descriptor showed better classification accuracy than the SURF, but classification time did not vary significantly. Adding location information (ORRI) significantly improved overall classification accuracy. SVM showed better classification performance than random forest and neural network. The proposed approach proved its potential under varying natural light conditions, but implementing a practical system, including vegetation segmentation and weed removal may potentially reduce the overall performance and more research is needed.

keywords

Weed classification; Bag-of-Visual-Words; SIFT; SURF; posterior probability

4.1 Introduction

Within the EU-funded project SmartBot (SmartBot), a small-sized robot was developed for vision based precision control of volunteer potatoes (weed) in a sugar beet field (Figure 4.1). Due to the small size of the robot and its battery operation, the platform design had to refrain from using additional infrastructure and should be able to robustly detect weeds in scenes that are fully exposed to ambient lighting conditions (Suh et al., 2018b). Additional infrastructure such as a hoods and lighting, as for example were used by Nieuwenhuizen et al. (2010) and Haug et al. (2014), was not considered viable

One of the most important steps in vision-based weed detection is the classification of weeds among crops. The output of this classification is a fundamental element in the subsequent process of weed control either by chemical spraying or mechanical actuation (Behmann et al., 2015). In a system for weed detection, vegetation segmentation is followed by classification of the segmented vegetation into weeds and crop. This classification step traditionally involves two aspects: 1) selection of the discriminative features and 2) selection of the classification technique (classifier) to differentiate between weeds and crop.

Figure 4.1: The robotic platform for volunteer potato control in a sugar beet field.

Regarding the features used for discrimination, many studies have used colour, shape (biological morphology) and texture on an individual basis or in combination (Ahmed et al., 2012; Åstrand & Baerveldt, 2002; Gebhardt & Kühbauch, 2007; Pérez et al., 2000; Persson & Åstrand, 2008; Slaughter et al., 2008; Swain et al., 2011; Zhang et al., 2010). These features are intuitive and easy-to-implement but may have limited discrimination ability under ambient lighting conditions.

In a system that has to work under ambient light conditions, the use of colour features may not yield robust classification (Lee et al., 2010). In the field, illumination constantly changes because of the varying sunlight and weather conditions. These variations in illumination greatly affect the Red-Green-Blue (RGB) pixel values of the acquired field images and lead to an inconsistent colour representation of plants (Sojodishijani et al., 2010; Teixidó et al., 2012). Additionally, irrespective of the illumination, it is sometimes hard, if not impossible, to differentiate between volunteer potato and sugar beet using colour features. Usually, volunteer potato has a darker green colour than sugar beet (Figure 4.2a) which results in a separable pixel distribution in the EG-RB colour plane (Figure 4.2c). However, as is shown in Figure 4.2b, volunteer potato occasionally has the same colour as sugar beet which makes them inseparable in the EG-RB colour plane (Figure 4.2d). Also, the colour of plants may change depending on their growth stage and nutritional status with plant leaves sometimes even turning yellow in the summer (Muñoz-Huerta et al., 2013) (Figure 4.3).

Shape and texture may also not be sufficiently discriminating features for successful classification of sugar beet and volunteer potato in the field. Camargo Neto et al. (2006), Swain et al. (2011), and Rumpf et al. (2012) showed that leaf edge information, plant orientation, and shape could serve as discriminative features. However, results obtained under laboratory conditions in a highly structured environment do not easily translate to real field conditions. Wind, shadow, and specular reflection of sunlight make it difficult for clear recognition of the shape of the plants in the field (Kazmi et al., 2015a). Some studies have shown that texture has the potential to discriminate between broad- and narrow-leaf plants as both have clearly different textural properties (Gebhardt & Kühbauch, 2007; Ishak et al., 2009; Van Evert et al., 2009). However, sugar beet and volunteer potato have similar textural properties that cannot easily be discriminated (Vollebregt, 2013). Therefore, a solution was needed to classify sugar beet and volunteer potato that would not depend on colour, shape, and textural features.

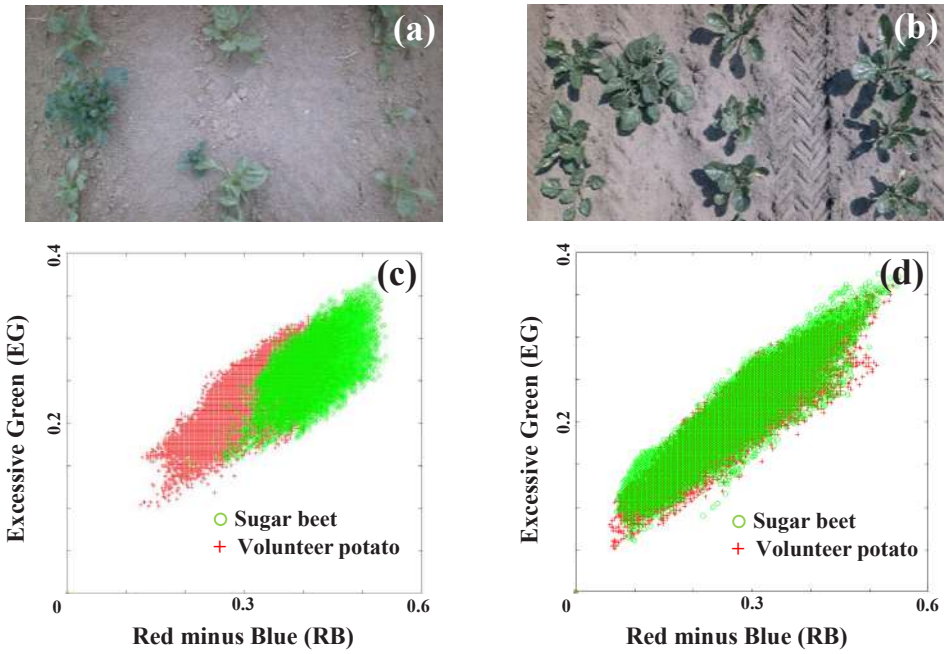


Figure 4.2: (a) In general, volunteer potato has a darker green colour than sugar beet. (c) In such a case, sugar beet and volunteer potato are separable (based on the colour) in the EG-RB plane. (b) An example case of volunteer potato having the same colour distribution as sugar beet. (d) Sugar beet and volunteer potato are then visually inseparable in the EG-RB plane. The EGRBI transformation was used to compare the colour difference between sugar beet and volunteer potato (Nieuwenhuizen et al., 2007).

Figure 4.3: Example plant images in the field. The plant leaves often turn yellow in the summer as indicated in squares.

A potential method to resolve the afore-mentioned issues and meet the performance requirements is to use counter-intuitive features (i.e. local descriptors) extracted by Scale-Invariant Feature Transform (SIFT) (Lowe, 2004) or Speeded Up Robust Features (SURF) (Bay et al., 2008). Both SIFT and SURF are invariant to illumination and colour while providing strong performance against noise. The SIFT descriptor has been used for weed classification and recognition in several recent studies (Kazmi et al., 2015a; Kounalakis et al., 2016; Wilf et al., 2016). Using the SIFT descriptor, Wilf et al. (2016) proposed a leaf identification procedure based on a machine learning approach. Although they acquired images under controlled environmental conditions with the manual arrangement of the leaves, their study showed the potential of the SIFT descriptor for leaf classification. Kazmi et al. (2015a) used both SIFT or SURF descriptors to classify sugar beet and creeping thistle under field conditions. Their study showed the potential of using local descriptor features for thistle detection. They combined these local descriptors with the features of surface colour and edge shapes. Using k-Nearest Neighbours (kNN) and SVM classifiers a very promising classification performance was achieved. However, their study was limited to detecting creeping thistle in a sugar beet crop, two species having clearly different textural features. Also, the field images were mostly acquired using a cover preventing direct access of sunlight to the scene, quite a distinct difference with the daylight conditions the SmartBot robot is confronted with.

A common way for classifying images using SIFT or SURF descriptors is to use a Bag-of-Visual-Words (BoVW) approach. The BoVW approach has demonstrated good performance in many computer vision applications such as object and scene classification (Law et al., 2014; Tsai, 2012; Zhou et al., 2013). The BoVW evolved from the original Bag-of-Words methodology which was first proposed in the field of text analysis and information retrieval (Bosch et al., 2007). In text analysis and information retrieval, each appearance of a word is recognised as a feature and is represented in the form of a bag of words, an orderless document representation of vocabulary (Salton & McGill, 1983). Once the Bag-of-Words model learns a vocabulary from all the documents, then each document can be classified by the number of times each word appears (occurrence). The same methodology and concept are applied in image classification in BoVW. The extracted features from an image are treated as a visual word, and the BoVW model is formed based on the occurrence of each visual word. Once the BoVW approach has learned each visual word from all the images, then each

image can be classified by the number of times each visual word appears (occurrence).

This paper presents a classification algorithm using a Bag-of-Visual-Words model, SIFT or SURF descriptors. SIFT is known to provide better classification performance than SURF, but it is said to be several times slower than SURF (Csurka et al., 2004; Khan et al., 2011; Wu et al., 2013). This research aimed to verify the difference in performance between SIFT and SURF by assessing classification accuracy and computation time on similar datasets (images) obtained in the field in 2015. Since neither SIFT nor SURF uses location related features, crop row information was used as an additional feature and added to the feature set to assess whether that would improve the classification accuracy.

SURF, SIFT and crop row information provide the features but require further processing for classification. Due to the challenging nature of the agricultural environment, and complexity of plant materials, it is hard to select a-priori one particular classifier which performs best in the classification task at hand (Suh et al., 2018b). To provide more insight into the performance differences found amongst different classifiers, the Support Vector Machine (SVM), random forest, and neural network classifiers were compared. These classifiers have been used in many agricultural applications (Ahmed et al., 2012; Cho et al., 2002; Jeon et al., 2011; Lottes et al., 2017).

To estimate the amount of certainty of the classification output, a posterior probability of the output of the SVM was calculated using a method proposed by Platt (1999). The posterior probability might provide useful information for weed control in practice since the action of removing volunteer potato should only be applied to those potato plants that are classified with a high confidence, while the control action should be skipped for those potato plants that are classified with a low confidence to prevent undesired destruction of the sugar beet.

Within the context of the SmartBot weeding application, following requirements were set by the previous study of Nieuwenhuizen (2009): the resulting automatic weeding system should be able to effectively control more than 95% of the volunteer potatoes as well as ensuring less than 5% of damage of the sugar beet plants. Therefore, classification accuracy should be considerably higher than 95% with a misclassification level of both sugar beet (false-negative) and volunteer potato (false-positive) of less than 5%. In addition, a classification time of less than 1 s per field image is required for feasible real-time field application. In this paper the classification process is evaluated in view of these requirements.

The first section of this paper describes the processing method of the BoVW model construction using the SIFT or SURF features. The following section describes the acquisition and selection of the image dataset, quantitative performance measure, and estimation of the posterior probability of SVM outputs. The experimental results are shown with the corresponding discussions. Lastly, conclusions are drawn.

4.2 The classification process

The classification process consists of the following procedures: 1) feature extraction using SIFT or SURF descriptors as well as crop row information, 2) feature clustering for visual vocabulary generation, 3) feature quantisation, 4) classification with SVM, random forest or neural network classifiers. The image classification process is shown in Figure 4.4, and each component will be described in more detail in the following section.

4.2.1 Feature extraction with SIFT or SURF descriptors and Out of Row Regional Index (ORRI)

The first step involved the extraction of local features from the training images (Figure 4.4a, Figure 4.5a). For the selection of the feature extraction point (keypoint) within an image, a regular grid-point based sampling was used as several studies reported that it provided robust performance (Fei-Fei & Perona, 2005; Law et al., 2014; Tsai, 2012). Grid size refers to the density of the feature extraction within a given image. In a preliminary study (Table 4.4) it was found that a grid size of 3×3 pixels proved to perform best for SIFT and 6×6 pixels for SURF.

During the generation of the visual vocabulary, the spatial location of the feature within an image was ignored. However, the spatial location may contain some valuable information especially for weed and crop discrimination in the field. Uijlings et al. (2009) reported that the classification performance of BoVW using SVM was considerably improved when they included spatial information (contextual information) into the algorithm.

In a classification problem with one single object in an image scene, the location of the object within an image may not carry any additional and useful information. However, with weed detection in the field, the location of each plant can play a

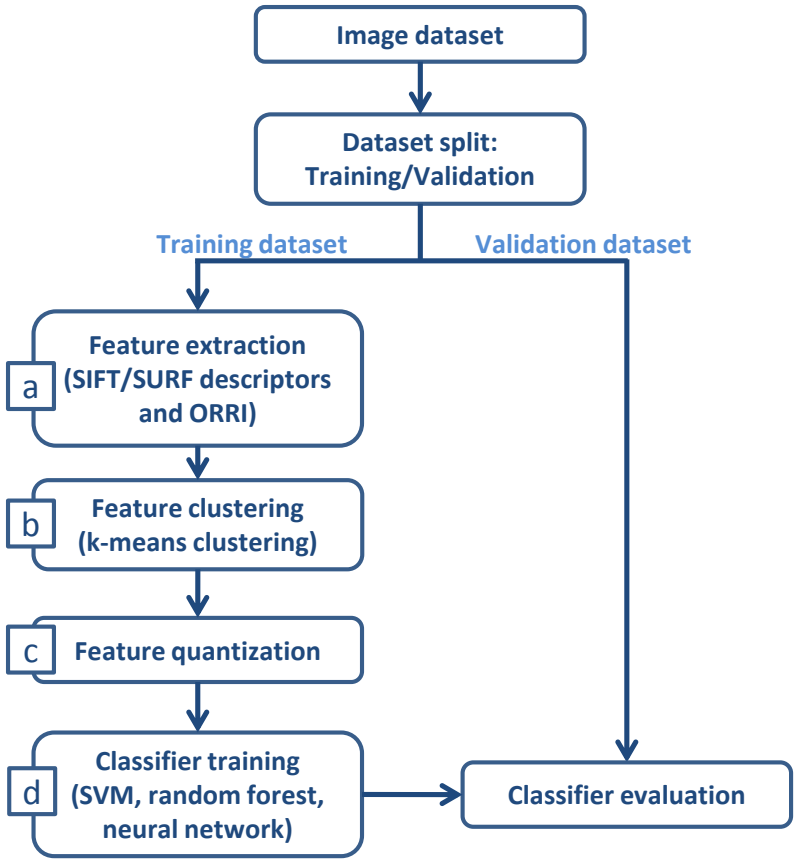


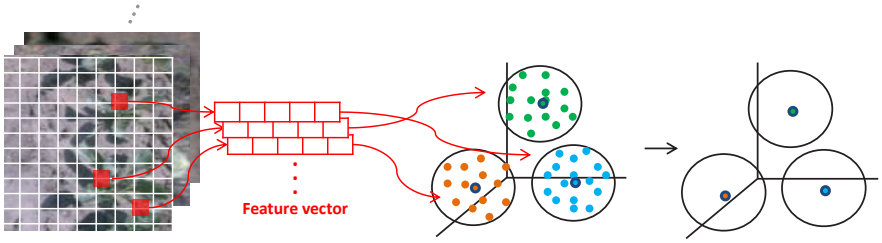
Figure 4.4: Flowchart of image classification using Bag-of-Visual-Words.

significant role in the plant recognition. For example, sugar beet plants are cultivated in rows (Åstrand & Baerveldt, 2002). Due to precision seeding, the crop row width and plant spacing within a row are fixed. For this reason, most of the sugar beet are found inside crop rows whilst weeds can be found randomly distributed across the field. Any green plant that is located far away from the crop rows is unlikely to be a crop but very likely to be a weed.

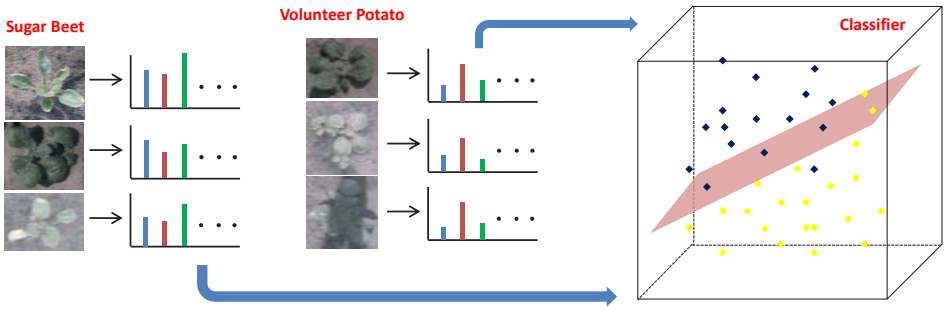
Inspired by the details mentioned above, an out-of-row regional index (ORRI) was generated for each plant on the basis of the out-of-row distance (Figure 4.6), a distance

between the centre of the plant to the nearest crop row. The ORRI was added to the BoVW feature set. Identifying weeds as weeds when located outside the crop row may sound trivial, which it is. However, it was hypothesised that adding ORRI information during the learning process might add an extra discriminatory dimension, and thus might enhance the discriminative power in the classification. The details of the ORRI generation are described below.

First, the location of three crop rows was manually estimated. Second, a distance between the centers of each plant to the nearest crop row, the out-of-row distance, was estimated. Third, each plant received a value for the ORRI from the set $[0.3, 0.6, 0.9]$



(a) Feature extraction (SIFT, SURF descriptors) (b) Feature clustering using k-means clustering



(c) Feature quantization to construct a histogram of visual words (d) Classifier training

Figure 4.5: Overview of BoVW model generation. (a) SIFT or SURF features (local descriptors) were extracted from the training images. (b) The extracted features were then clustered for visual vocabulary generation using k-means clustering. (c) A histogram of visual words was constructed from each training image, (d) which was used for classifier training.

based on the following rules:

$$ORRI = \begin{cases} 0.3, & \text{if out-of-row distance} < 80 \text{ pixels} \\ 0.6, & \text{if } 80 \leq \text{out-of-row distance} < 160 \text{ pixels} \\ 0.9, & \text{otherwise} \end{cases} \tag{4.1}$$

where the out-of-row distance is represented as a pixel value (one pixel corresponds to approximately 1 mm in the field).

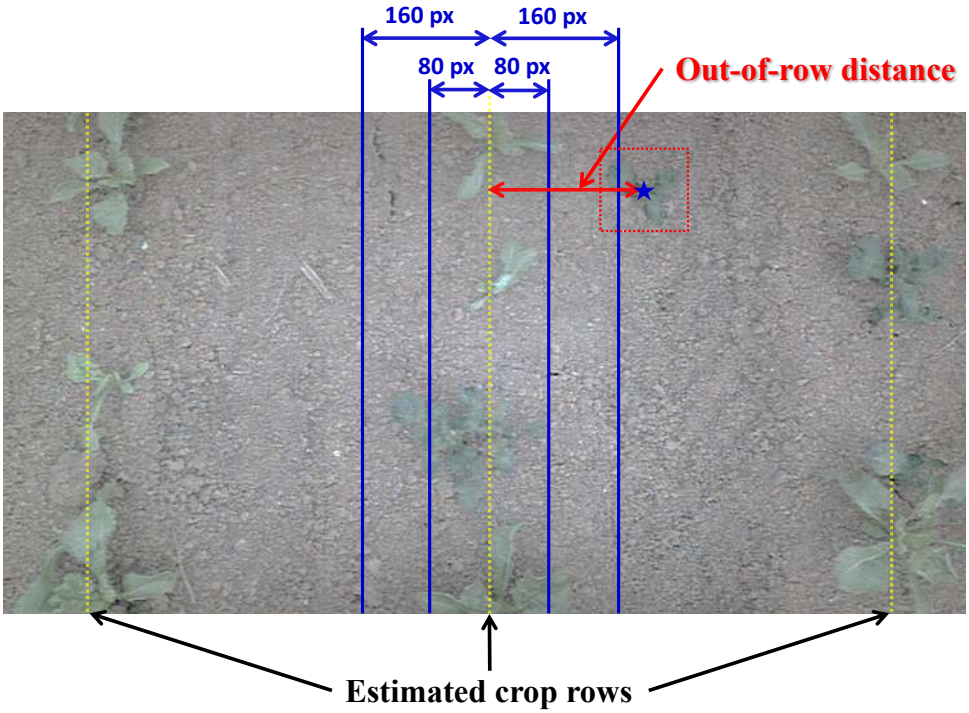


Figure 4.6: The location of the three crop rows in the field of view was manually estimated (yellow dotted lines). An individual plant was extracted, then the distance between the centre position of a plant (marked as a star) to the nearest crop row, the out-of-row distance, was estimated. Two distances from the central crop row (80 and 160 pixels) are shown (blue lines).

For the regional index discrete values of 0.3, 0.6 and 0.9 were used instead of continuous values because it was expected that the estimation of the crop rows and centre point of the plant would be likely to introduce noise.

4.2.2 Feature clustering for visual vocabulary generation

In this step, extracted features were clustered using k-means clustering, a common method for visual vocabulary generation (Figure 4.4b, Figure 4.5b). Each cluster centroid determined by k-means clustering was considered as a visual word. Based on a preliminary study, the number of clusters and thus the vocabulary size was set to 500 (Table 4.4).

If the vocabulary size (number of clusters) is too small, the set of visual words may be too limited to represent all the important features of images, and thus may lead to poor classification performance (Yang et al., 2007). On the other hand, if the vocabulary size is too large, there is a higher chance of overfitting the training dataset. In addition, a large size of the vocabulary also requires more processing power.

4.2.3 Feature quantisation

Once the visual vocabulary was generated, the features (descriptors) extracted from each image were assigned to each visual word to construct a histogram of visual word occurrences (Figure 4.4c, Figure 4.5c). Using the Euclidean distance, each extracted feature was allocated to its nearest visual word (nearest neighbour). A histogram of visual words was then generated by counting the number of features that were assigned to each visual word. The length of the histogram was equal to the number of cluster centres generated by k-means clustering, where the n^{th} value in the histogram was the occurrence of the n^{th} visual word. This process is commonly called feature quantisation (Kato & Harada, 2014). A histogram of visual word occurrence generated from images of sugar beet and volunteer potato is shown in Figure 4.7.

4.2.4 Classification based on supervised learning

Supervised learning was used to train the classifiers for differentiation between sugar beet and potatoes (Figure 4.4d, Figure 4.5d). Three classifiers were used in this study: SVM, random forest and a neural network. In the SVM, three different polynomial

kernels (linear, quadratic and cubic) were assessed. For the evaluation of the classifiers, 10-fold cross-validation was used. Some details of random forest and neural network are described below.

Support Vector Machine (SVM)

The SVM is a supervised learning model based on the theory of statistical learning (Vapnik, 1995). SVM is one of the most widely used classification models in machine learning applications and often reaches high performance in high-dimensional problems with small sample problems (Csurka et al., 2004; Li, 2011). The basic principle of SVM is to find the optimal hyperplane which separates classes with minimum error.

Random forest (Ensemble Classifier)

A random forest classifier, an ensemble method that consists of multiple decision trees, was used for this study. Random forest, as the name says, is constructed from decision trees, more precisely it is a collection of tree-structured classifiers. Each

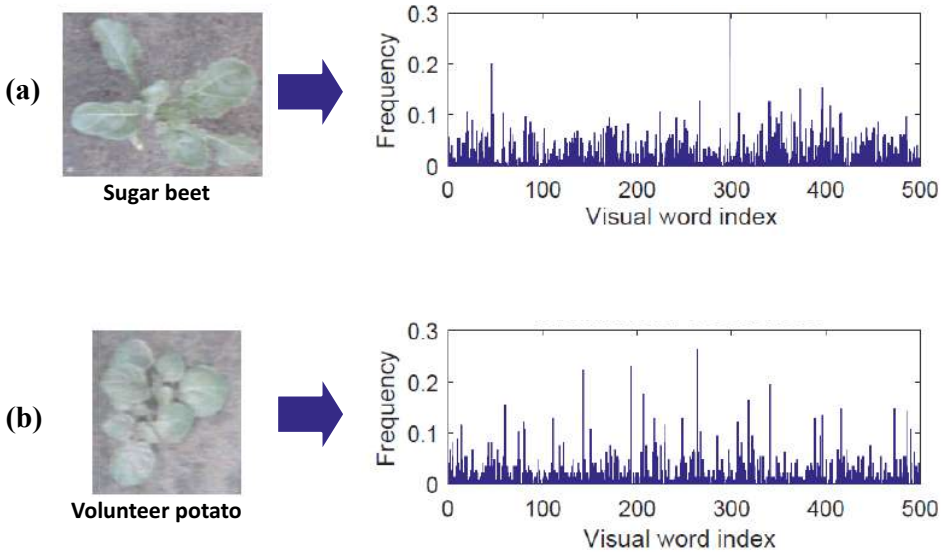


Figure 4.7: Images of (a) sugar beet and (b) volunteer potato on the left, with the associated histograms of visual word occurrences on the right.

decision tree provides a classification "vote," and the majority vote is selected for the final classification (Chan & Paelinckx, 2008; Liaw & Wiener, 2002; Polikar, 2006). Breiman (2001) reported that the performance of a random forest was superior to other learning algorithms. Rodriguez-Galiano et al. (2012) indicated that the random forest is relatively robust to outliers and noise as well as computationally less expensive than other tree ensemble methods.

Neural network

The artificial neural network consists of multiple nodes and neurons that are connected in the layers. Compared to other classifiers, according to Behmann et al. (2015), a neural network requires less prior information and is robust to noise thus particularly suitable for the modeling of optical sensor data. In this study, a feed-forward back propagation neural network was used. The neural network used in this research consists of one hidden layer with 150 neurons besides an input and an output layer. In the input layer, histograms of visual words were utilized, and in the output layer, sugar beet was represented by $[1, 0]$ while volunteer potato was represented by $[0, 1]$.

4.3 Experiment setup

4.3.1 Field image collection and image dataset

To acquire crop images, a camera was mounted at the height of 1 m perpendicular to the ground on a custom-made frame carried by a mobile platform (Husky A200, Clearpath, Canada) (Figure 4.8). A stereo camera (NSC1005c, NIT, France) was equipped with two Kowa 5 mm lenses (LM5JC10M, Kowa, Japan) with a fixed aperture. The camera was set to operate in an automatic acquisition mode with default settings. The camera images from left and right sensors were acquired each having an image resolution of 1280×580 pixels. The ground-covered area was $1.3\,\text{m} \times 0.7\,\text{m}$ per image (pair), corresponding to three crop rows of sugar beet. The acquisition program was implemented in LabVIEW (National Instruments, Austin, TX, USA) and acquired five images per second. Raw format images (TIFF) were initially acquired in the field, and debayer was processed offline to convert the raw format image into RGB colour. Field images were taken while the mobile platform was manually controlled with a joystick and driven along crop rows using a controlled traveling speed of 0.5 m/s.

Sugar beet were sown in April 2015 in sandy and clay soil at Unifarm experimental sites in Wageningen, The Netherlands. One week after sowing the sugar beet, potatoes were planted in random locations throughout the fields. Crop images were acquired for two days in the morning and afternoon on 1-June and 5-June, 2015.

For the labelled image dataset used in this study, a total of 400 individual plant images was manually extracted from selected field images: 200 sugar beet plants and 200 volunteer potato plants. During the selection of this image dataset, images with different illuminations levels were considered as well as images containing shadows. The size of each plant image in the dataset varied from the smallest size of 65×65 pixels to the largest of 305×315 pixels. Example images in the dataset are shown in Figure 4.9.

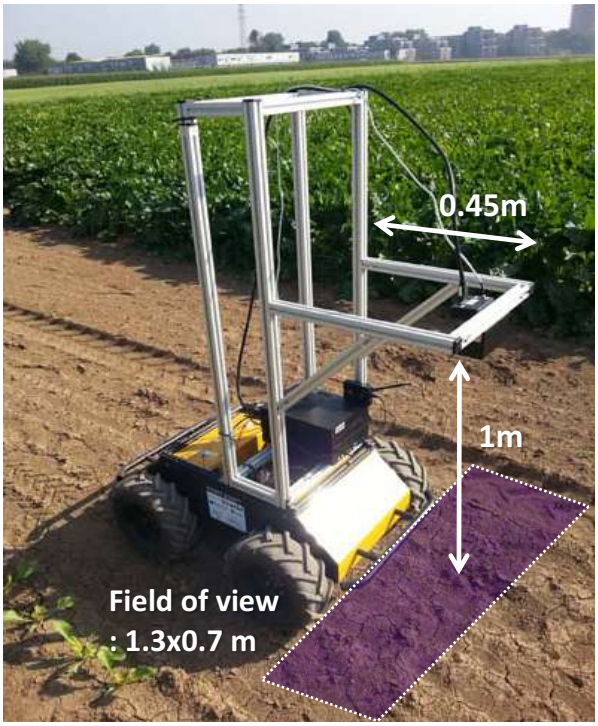


Figure 4.8: Field images were acquired with a stereo camera mounted at the height of 1 m *viewing perpendicular to the ground surface resulting in a field of view of* $1.3\,\mathrm{m} \times 0.7\,\mathrm{m}$. *A mobile platform, Clearpath Husky, was manually controlled with a joystick and driven along crop rows using a controlled travelling speed of* $0.5\,\mathrm{m/s}$.

In the image dataset, all sugar beet were found inside crop rows (out-of-row distance < 80 pixels), having an ORRI of 0.3. On the other hand, volunteer potatoes were found inside and outside crop rows. The number of volunteer potatoes found inside the crop row (out-of-row distance < 80 pixels), i.e. ORRI $= 0.3$, was 55; while the number of volunteer potatoes found outside the crop row (out-of-row distance ≥ 80 pixels), i.e. ORRI > 0.3, was 145.

4.3.2 Performance measure and system platform

In this study, a binary classification was carried out; i.e. sugar beet or volunteer potato. The classification performance measures in this study are described below.

A confusion matrix (Table 4.1) was used to assess and compare the classification performances. The classification accuracy was calculated along with training and classification time since this approach should, in the end, yield a real-time field application. Each classifier was validated using 10-fold cross-validation. The classification accuracy and training time were averaged over ten trials with a random split of the dataset. The training time included times for classifier training as well as extracting

Figure 4.9: Example images from the field image dataset containing a total of 400 plant images with 200 sugar beet (top) and 200 volunteer potatoes (bottom). During the generation of this dataset, images with different illumination levels were selected as well as images containing shadows.

Table 4.1: Confusion matrix
(TR:true-positive, TN:true-negative, FP:false-positive, and FN:false-negative)

		Predicted Class	
		Sugar Beet (SB)	Volunteer Potato (VP)
Actual Class	Sugar Beet (SB)	TP	FN
	Volunteer Potato (VP)	FP	TN

features and building a visual vocabulary. The classification time was measured for the prediction of one plant image. All images were processed in Matlab® 2015b (The MathWorks Inc, Natick, MA, USA) using the Computer Vision System Toolbox™, Neural Network Toolbox™, and VLFeat library for Matlab (Vedaldi & Fulkerson, 2008). Processing time was measured on an Intel® Core™ i7-377T 2.5 GHz processor with 8 GB memory running 64-bit Windows 7.

$$ClassificationAccuracy = \frac{TP + TN}{TP + FN + FP + FN} \tag{4.2}$$

where TP is true-positives; FP is false-positives; TN is true-negative, and FN is false-negative.

4.3.3 Estimated posterior probability of SVM outputs

Platt (1999) proposed a method using a sigmoid function to calculate and estimate the posterior probability for SVM classifier. Since then, this method has been used in many applications as it is a useful measure to provide the degree of certainty (belief) of the classification output (Lin et al., 2007). In this study, a posterior probability was estimated for the SVM using a linear kernel and employing the ORRI in the feature set.

4.4 Results

The classification performances of BoVW using SIFT or SURF descriptors are summarized with true-positive (TP), false-negative (FN), false-positive (FP), true-negative (TN), classification accuracy, training time and classification time in Table 4.2 and Table 4.3. In these tables, it is also indicated whether the ORRI was used.

Table 4.2: The classification performance using SIFT features is shown. The classifiers were trained and validated with a total of 400 images (200 of sugar beet and 200 of volunteer potato) using 10-fold cross-validation. The final classification performance was averaged over ten repetitions. The training time includes the time for training of the classifier as well as for extracting SIFT features and building a visual vocabulary. The classification time includes the time required to classify the class of a single plant image using the trained classifier.
(TR:true-positive, TN:true-negative, FP:false-positive, FN:false-negative, and ORRI:Out-of-Row Regional Index)

Classifier models			**TP**	**FN**	**FP**	**TN**	**Classification accuracy (%)**	**Training time (s)**	**Classification time (s/image)**
			(% of total)						
SVM	**Linear**	**without ORRI**	183	17	20	180	90.8	218.6	0.107
			(91.5)	(8.5)	(10)	(90)			
		with ORRI	200	0	20	180	95.0	221.4	0.108
			(100)	(0)	(10)	(90)			
	Quadratic	**without ORRI**	186	14	17	183	92.3	216.6	0.106
			(93)	(7)	(8.5)	(91.5)			
		with ORRI	200	0	14	186	96.5	218.8	0.107
			(100)	(0)	(7)	(93)			
	Cubic	**without ORRI**	188	12	18	182	92.5	219.3	0.106
			(94)	(6)	(9)	(91)			
		with ORRI	196	4	17	183	94.8	222.6	0.106
			(98)	(2)	(8.5)	(91.5)			
Random forest		**without ORRI**	172	28	38	162	83.5	228.9	0.109
			(86)	(14)	(19)	(81)			
		with ORRI	183	17	21	179	90.5	238.9	0.108
			(91.5)	(8.5)	(10.5)	(89.5)			
Neural network		**without ORRI**	187	12	23	177	91.2	245.4	0.125
			(93.5)	(6)	(11.5)	(88.5)			
		with ORRI	195	5	12	188	95.8	260.5	0.130
			(97.5)	(2.5)	(6)	(94)			

Table 4.3: The classification performance using SURF features is shown. The classifiers were trained and validated with a total of 400 images (200 of sugar beet and 200 of volunteer potato) using 10-fold cross-validation. The final classification performance was averaged over ten repetitions. The training time includes the time for training of the classifier as well as for extracting SURF features and building a visual vocabulary. The classification time includes the time required to classify the class of a single plant image using the trained classifier.
(TR:true-positive, TN:true-negative, FP:false-positive, FN:false-negative, and ORRI:Out-of-Row Regional Index)

Classifier models			TP	FN	FP	TN	Classification accuracy (%)	Training time (s)	Classification time (s/image)
			(% of total)						
SVM	**Linear**	**without ORRI**	175	25	42	158	83.3	175.8	0.099
			(87.5)	(12.5)	(21)	(79)			
		with ORRI	200	0	22	178	94.5	182.9	0.099
			(100)	(0)	(11)	(89)			
	Quadratic	**without ORRI**	179	21	35	165	86.0	175.7	0.099
			(89.5)	(10.5)	(17.5)	(82.5)			
		with ORRI	196	4	18	182	94.5	182.9	0.105
			(98)	(2)	(9)	(91)			
	Cubic	**without ORRI**	176	24	29	171	86.8	175.7	0.099
			(88)	(12)	(14.5)	(85.5)			
		with ORRI	195	5	20	180	93.8	183.1	0.101
			(97.5)	(2.5)	(10)	(90)			
Random forest		**without ORRI**	170	30	55	145	78.8	178.9	0.106
			(85)	(15)	(27.5)	(72.5)			
		with ORRI	179	21	42	159	84.5	186.2	0.104
			(89.5)	(10.5)	(21)	(79.5)			
Neural network		**without ORRI**	165	35	27	173	84.5	195.1	0.115
			(92.5)	(17.5)	(13.5)	(86.5)			
		with ORRI	190	10	21	179	92.3	190.1	0.119
			(95)	(5)	(10.5)	(89.5)			

4.4.1 Classification accuracy

In Table 4.2, using SIFT features and ORRI, the highest classification accuracy obtained was 96.5%; while the lowest classification accuracy obtained was 83.5%. Three classifier models (SVM linear, SVM quadratic, and neural network) showed the classification accuracies $\geq 95\%$, thus meeting the requirements. Likewise, in Table 4.3, using SURF features and ORRI, the highest classification accuracy obtained was 94.5%; while the lowest classification accuracy obtained was 84.5%. None of the classifier models showed a classification accuracy of $\geq 95\%$, and thus using SURF features and ORRI did not meet the requirements set at the beginning of this research.

4.4.2 Misclassification rate (false-positive and false-negative)

The false-negative values obtained for the cases with the highest classification accuracies using SIFT features with ORRI and using SURF features with ORRI were both zero (Table 4.2 and Table 4.3). Meeting the requirements, in these cases all the sugar beet plants were correctly classified as a sugar beet, and thus no crop would be eliminated by a weed control operation (0% of undesired control of sugar beet plants). However, in these cases the false-positive values obtained with the highest classification accuracies using SIFT with ORRI, and using SURF with ORRI were 14 (7%) and 22 (11%), respectively. So, 7% and 11% of volunteer potato were classified as sugar beet, respectively, and thus would not be destroyed. These false-positive values do not meet the requirements (misclassification: less than 5%).

4.4.3 Training and classification time

Training time in this work includes the time needed for training of the classifiers as well as for extracting SIFT or SURF features and building the visual vocabulary. SVMs required $218-222$ s and $175-183$ s of training time using SIFT with ORRI and SURF with ORRI, respectively; while the neural network required 260 s and 190 s of training time using SIFT with ORRI and SURF with ORRI, respectively. The training times needed by all classifiers were reasonable, considering the training can be done offline and may not have to be repeated very often.

The classification time indicates the time required to classify the class of a single plant image using a trained classifier. For all classifiers, an average time of $0.10-0.11$ s

was needed for classification, which is a reasonable value when real-time application in the field is considered.

4.4.4 SIFT compared to SURF

SIFT is known to provide better classification performance than SURF, however, at the expense of more computation time. In view of classification accuracy, this observation was confirmed in this research. Overall, in line with findings reported in the literature, using SIFT features resulted in better classification accuracy than using SURF features. Without ORRI, the accuracy improved on average 6.2% when using SIFT features instead of using SURF features. With ORRI this difference reduced, and on average, the accuracy improved by 2.6% when using SIFT features instead of SURF features. SIFT features required more training time than SURF features. On average 46 s more training time was required when using SIFT instead of SURF. Classification time did not differ much for SIFT and SURF; however, this result does not match with observations reported in the literature. On average 0.11 s and 0.10 s was needed when using SIFT and SURF, respectively.

4.4.5 Out-of-Row Regional Index (ORRI)

For all classifiers classification accuracy improved with ORRI. It was earlier hypothesized that adding spatial information (ORRI) during the learning process adds an extra discriminatory dimension which enhances the discriminative power of the classification of sugar beet and volunteer potato. This hypothesis was confirmed by the results, showing that the classification accuracy considerably improved when using the ORRI. Averaged over all classifiers, the improvement in classification accuracy using the ORRI was 4.5% and 8% when using the SIFT and SURF features, respectively.

For comparison, it is worth noting that using the ORRI as the only feature, a classification accuracy of 86.3% was obtained in all classifiers with TP, FN, FP and TN of 200, 0, 55, 145, respectively. This is a relevant result because, as mentioned earlier, in the dataset a total of 255 plants (200 sugar beet and 55 volunteer potatoes) were found inside crop rows (out-of-row distance < 80 pixels, having an ORRI of 0.3). In Table 4.3, it can be seen that adding ORRI to SURF and classifying with a SVM and a linear kernel results in a change of classification for 45 plants (FN:from 25 to 0, FP:from 42 to 22). Further analysis of the individual images revealed that 29 of

these 45 images had an ORRI 0.3, so were inside crop rows: 25 of them were sugar beet plants, and four of them were volunteer potato plants. Interestingly enough, these 25 sugar beet, though being inside the crop rows, were not properly classified by SURF only (without ORRI). This is no surprise because SURF does not employ any locational feature. More interesting is to note that four of the images were volunteer potato plants. So, by adding a location feature in training improved the classification for volunteer potato inside crop rows, which is a real challenge in weed classification.

When training time is considered with ORRI, SIFT required on average 6.7 s more training time when training without ORRI. Likewise, training with ORRI using SURF required on average 7.2 s more time than training without ORRI. When it comes to classification, however, the use of ORRI did not lead to a considerable increase in calculation time.

4.4.6 Comparison of SVM, Random forest and Neural network classifiers

SVM classifiers with a linear and quadratic showed better classification accuracies than random forest and neural network, though the SVM and neural network did not differ much. In Table 4.2, using SIFT features and ORRI, the highest classification accuracy of 96.5% was obtained with a SVM and a quadratic kernel; while the lowest classification accuracy of 90.5% was obtained with the random forest. In Table 4.3, using SURF features and ORRI, the highest classification accuracy of 94.5% was obtained with a SVM and both a linear and a quadratic kernel; while the lowest classification accuracy of 84.5% was obtained with the random forest.

4.4.7 Grid size and vocabulary size

Classification accuracy with different sizes of grid and vocabulary are compared in Table 4.4. Using small grid sizes tended to produce better result than large grid sizes. However, vocabulary size did not seem to produce any regular pattern of performance. In fact, grid and vocabulary size are not formally related, but a certain combination (in this case, a grid size of 6×6 and vocabulary size of 500) showed a better performance than others in this study. Therefore, a grid size of 6×6 pixels and vocabulary size of 500 were used as an optimal combination when employing the SURF descriptor because the highest classification accuracy (94.5%) was achieved with these settings.

Table 4.4: Comparison of classification accuracy (%) with different grid and vocabulary sizes. Using SURF descriptor, the classification accuracy of SVM linear with ORRI is shown.

		Vocabulary size					
		100	**200**	**300**	**400**	**500**	**600**
Grid size (pixels)	**4x4**	92.5	93.2	91.2	93.7	93.8	92.7
	6x6	91.7	93.0	93.5	92.5	**94.5**	92.7
	8x8	90.5	91.0	93.7	92.0	90.7	92.2
	10x10	91.2	92.7	93.6	93.2	92.7	92.7
	12x12	91.2	91.5	91.5	91.5	91.0	91.5

For the SIFT descriptor, a grid size of 3×3 pixels with a vocabulary size of 500 was used as the highest classification accuracy was achieved with these settings.

4.4.8 Estimated posterior probability

The posterior probabilities of the SVM with linear kernel using SIFT features and ORRI were calculated and visualized in the form of a box-and-whiskers plot in Figure 4.10. All sugar beet images were correctly classified as sugar beet (true-positive), and on average the posterior probability was 0.96 with a standard deviation of 0.09. A total of 180 volunteer potato images (out of 200) was correctly classified as volunteer potatoes (true-negative), and for these images the average posterior probability was 0.98 with a standard deviation of 0.02. However, 20 volunteer potato images were incorrectly classified as sugar beet (false-positive). With an average value of 0.49 and standard deviation of 0.27, in these cases, the average posterior probability was lower than in the true-positive and true-negative cases. These results indicate that the classifier was more confident in case of correct classification than when making a false classification.

The above results indicate that the posterior probability might provide useful information for weed control in practice. Using the posterior probability, the action to remove volunteer potato should only be applied to those plants that are classified with a high confidence. Figure 4.11a, for example, shows the classification results with

the posterior probability with a field image. Plants 1 to 6 are sugar beet whereas plants 7 to 9 are volunteer potatoes (Figure 4.11b). In Figure 4.11c, plants 2 to 6 are correctly classified as sugar beet with a posterior probability of 0.86 and higher; and plants 7 to 9 are correctly classified as volunteer potatoes with a posterior probability of 1.0. However, plant 1 (sugar beet) is incorrectly classified as a volunteer potato (false-negative). In this case, the posterior probability is 0.54 and considerably lower than the others. In such a case, based on the lower posterior probability, it might be beneficial to skip the weed control action because since it would lead to the destruction of the crop.

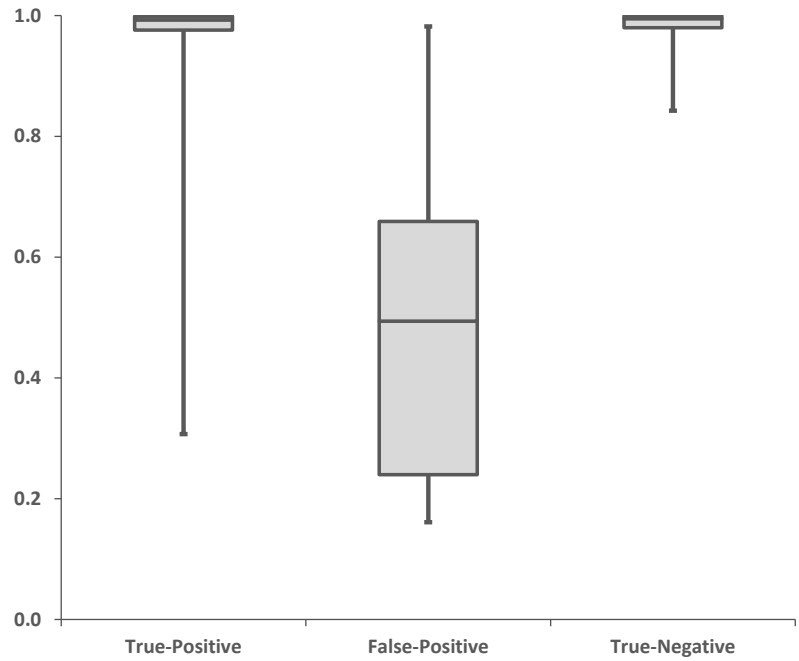


Figure 4.10: A box-and-whisker plot of the estimated posterior probabilities of true-positive, false-positive and true-negative classifications using the SVM with linear kernel on SIFT features and ORRI. All sugar beet images were correctly classified as sugar beet (true-positive) with an average posterior probability of 0.96. A total of 180 volunteer potato images (out of 200) was correctly classified as volunteer potatoes (true-negative) with an average posterior probability of 0.98. However, 20 volunteer potato images were incorrectly classified as sugar beet (false-positive) with a Q1 (1st quartile), median and Q3 (3rd quartile) of 0.24, 0.49 and 0.66, respectively.

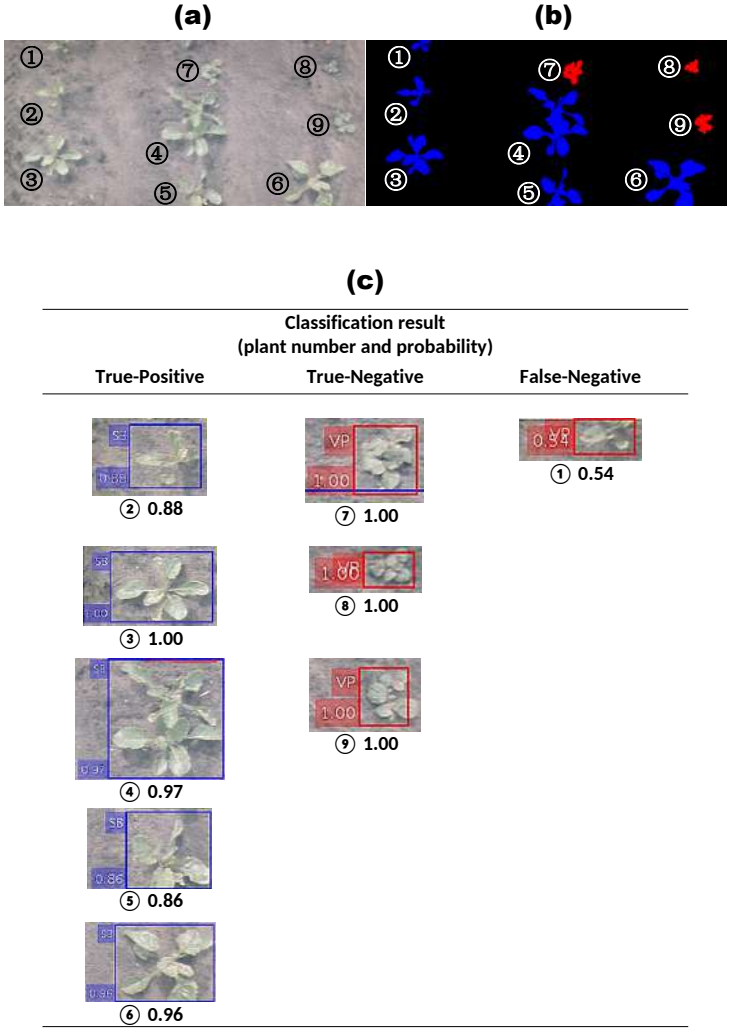


Figure 4.11: An example of the classification results with posterior probability with a field image. (a) A field image with plant number. Each plant was manually extracted, and then put into the classification algorithm proposed in this study. (b) The ground truth of the given image. Plants 1 to 6 are sugar beet, and plants 7 to 9 are volunteer potatoes. (c) Classification results with posterior probability. Plants 2 to 6 are correctly classified as sugar beet (true-positive) with a posterior probability of 0.86 and higher, and plants 7 to 9 are correctly classified as volunteer potatoes (true-negative) with a posterior probability of 1.0. However, plant 1 is incorrectly classified as a volunteer potato (false-negative) and results in a posterior probability of 0.54.

4.5 Discussion

4.5.1 Classification accuracy

The classification accuracy obtained using BoVW approach with ORRI exceeded previously reported accuracies; e.g. Nieuwenhuizen et al. (2010) and Persson & Åstrand (2008). Considering the different illuminations levels and shadows in the image dataset, the highest classification accuracy (96.5%) obtained in this study is considerably better than any other approaches with colour, shape, and texture features in the literature for weed classification. However, the overall performance of weed control also depends on the performance of vegetation segmentation as well as the actuation performance of the weeding device. If the individual performance of either one of these two operations would be ¡ 100%; thus the classification accuracy should be considerably higher than 95% in order for the automatic weeding system to effectively control more than 95% of the volunteer potatoes in the field. In this regard, the highest classification accuracy achieved in this study (96.5%) may not be enough to satisfy the overall performance of volunteer potato control since it is not significantly higher than 95%.

The obtained results were based on manually extracted plant images. Thus, the proposed approach itself does not lead to the precise detection of volunteer potato in field images. To make a complete system for the use of weed control in the field, vegetation segmentation and weed removal operation needs to be integrated. During integration, overlapping plant cases need to be considered as well.

4.5.2 Misclassification rate (false-positive and false-negative)

For weed control in practice, it is critical to have a large as possible number of true-positives as well as a large as possible number of true-negatives. Not only that, but it is also important to consider both the number of false-negatives (the number of sugar beet plants that are classified as volunteer potatoes) and the number of false-positives (the number of volunteer potato plants that are classified as sugar beet). The false-negatives lead to the removal of the cash crop caused by the misclassification, thus keeping the number of false-negatives as small as possible is critical (Lottes et al., 2016). However, it is desirable to keep the number of false-positives as small as possible. If there are many left over volunteer potato plants caused by misclassification,

then a weed control robot may need to drive repetitively acrodd the field to meet the statutory regulation in the Netherlands (Nieuwenhuizen, 2009). The economic consequences of false-negative and false-positive detections require further research.

4.5.3 Calculation time

From general observations of the field images, there is an average of 6-8 plants in one image. Based on these number of plants in an image, the whole plant classification of one field image may take up to 0.8 seconds (including all other steps in the image processing) using SVM classifiers, which is acceptable for our real-time application (< 1 s for one field image). The classification time, of course, depends on the size of each plant found in an image, and can be further improved with a parallel-processing approach. In addition, the size of the grid and vocabulary also influences the classification and processing time. If the processing time is highly critical for certain applications, grid and vocabulary size can be changed to reduce the processing time at the expense of classification accuracy.

4.5.4 SIFT and SURF

Several studies have indicated that SURF is rapid for computation and matching (Khan et al., 2011; Panchal et al., 2013; Wu et al., 2013; Zagoris et al., 2014). In this research SIFT required more training time than SURF. However, the classification times required for SIFT and SURF were not considerably different in this study. This result is not accord with the literature. The different grid sizes used for SIFT and SURF in this study may have caused classification times to be similar.

There is room for improvement in terms of the classification accuracy. During the extraction of SIFT and SURF descriptors, dataset images were converted to greyscale ignoring all the colour information (RGB) because SIFT and SURF operate on intensity information only. However, colour may carry some discriminative information for the classification of sugar beet and volunteer potato. To overcome the abovementioned weakness of SIFT and SURF descriptors, several variations of SIFT and SURF have been proposed in the literature using colour features such as rgSIFT, Transformed colour SIFT, and Color-SURF (Fan et al., 2009; Van De Sande et al., 2008) to improve classification accuracy. Similarly, Rassem & Khoo (2011) proposed not to convert RGB image to greyscale but to apply the feature extraction on each RGB channel. The

extracted features from the individual colour channels may add extra discriminative power for classification, and validating this hypothesis is, therefore, a topic of a future study. As indicated in Figure 4.2, the added value of using also colour might be limited in cases where, as here, crop and weed plants have similar colour values.

4.5.5 Out-of-Row Regional Index (ORRI)

Combining ORRI considerably improved the classification accuracy enhancing the discriminative power of the classification. However, spatial information of each plant (ORRI) including crop rows and out-of-row distance was manually estimated in this study. For an automated field application using a mobile robot, the estimation of crop rows and out-of-row distance should be automated as well. Algorithms for crop row detection have been presented in several studies (Guerrero et al., 2013; Hiremath et al., 2014; Kise et al., 2005; Leemans & Destain, 2006; Romeo et al., 2012; Søgaard & Olsen, 2003), but these algorithms are likely to introduce noise. Thus, in the current approach, regional index (0.3, 0.6 and 0.9) was used instead of a precise number for the out-of-row distance to compensate any potential noise.

4.5.6 Classifiers

Based on the results obtained in this study SVM classifiers would be an easy and plain choice for field applications, not only because SVM classifiers showed better classification performance in most cases than random forest and neural network, but also because SVMs are easier to implement than other classifiers. However, the neural network also performed quite well, showing similar classification performance as SVMs, although a simple network structure (1 hidden layer) was used in this study. Kanellopoulos & Wilkinson (1997) indicated that multi-layer network architecture might be potentially more powerful than a simple network. This has been confirmed over the past few decades in various applications (LeCun et al., 2015). Thus, adding more layers is likely lead to better classification performance.

4.5.7 Posterior probability

The posterior probability estimated by Platt's method offers additional information during the weed control action, which can be useful in practice. Using this posterior

probability, the action to remove volunteer potato should only be applied to those volunteer potato plants that are classified with a high confidence. Volunteer potato plants that are classified with lower confidence might be better skipped because it might lead to the undesired destruction of the sugar beet. However, the characteristics and applicability of this approach need further study.

Two studies have indicated that probability estimation using Platt's method could be ineffective in some cases especially for large datasets (Niculescu-Mizil & Caruana, 2005; Pérez-Cruz et al., 2007). To compensate for the weakness of Platt's method, Lin et al. (2007) proposed an improved algorithm which theoretically avoids numerical difficulties. When large datasets are concerned, their proposed method for probability estimation might be a better choice.

In this study, the posterior probability was estimated only for SVM classifier. However, the posterior probability for other classifiers, such as random forest and neural network, can also be estimated using a method proposed by Niculescu-Mizil & Caruana (2005). They reported that random forest and the neural network classifiers provided well-calibrated probabilities having no bias compared to SVM. Investigating the posterior probability for other classifiers would be a future study topic.

4.5.8 Reflection on contribution to weed control

In this study, binary classification (between sugar beet and volunteer potato) was proposed based on the assumption that in most cases plants found in sugar beet fields are either sugar beet or volunteer potato. However, in an agricultural field, a variety of different weed species is likely to be found. A future study topic might include a multiclass classification of weed species within a crop. Classification of other crop species may also benefit from the proposed approach.

4.6 Conclusions

In this study, an algorithm using a Bag-of-Visual-Words model and SIFT or SURF descriptors as well as crop row information in the form of the ORRI (Out-of-Row Regional Index) was proposed for the classification of sugar beet and volunteer potato under natural and varying daylight conditions. In EU SmartBot project it was required to effectively control $> 95\%$ of volunteer potatoes (weed) and to ensure $< 5\%$

of undesired control of the sugar beet crop. Considering the different illuminations levels and shadows in the image dataset, the highest classification accuracy of 96.5% with false-negative of 0% which was obtained using SIFT features and ORRI with SVM classifier is considerably better than any other approaches found in the literature that used colour, shape and textural features. Therefore, the proposed approach proved its potential under ambient light conditions although the false-positive rate of 7% deviates from the requirements (misclassification: $< 5\%$). An average time of $0.10 - 0.11$ s was needed for classification, which is a reasonable value when the real-time application in the field is considered and is well within the required 1 s. However, implementing a full pipeline including vegetation segmentation and weed removal operation by actuator may potentially further reduce the overall performance. The SIFT descriptor showed better classification accuracy than using the SURF descriptor. Using SIFT required more training time than SURF, but the classification time required for SIFT and SURF was not considerably different.

Adding crop row information as an additional feature (ORRI) significantly improved the overall classification accuracy. However, for an automated field application using a weed control robot, the estimation of crop rows and out-of-row distance should be automated and might potentially introduce noise.

In this application, SVM classifiers showed better classification performance than random forest and neural network. However, a neural network with multi-layer architecture would potentially improve the performance.

The posterior probability estimation can be useful in practice which provides an another decision moment for weed control action, but characteristics and applicability of it need further study.

This study has shown the potential benefit of using counter-intuitive features such as SIFT and SURF instead of colour, shape and texture for weed classification under natural daylight conditions.

4.7 Acknowledgements

The work presented in this paper was part of the Agrobot part of the Smartbot project and funded by Interreg IVa, European Fund for the Regional Development of the European Union and Product Board for Arable Farming. We thank Gerard Derks at experimental farm Unifarm of Wageningen University & Research for arranging and

managing the experimental fields.

CHAPTER 5

Transfer learning for the classification of sugar beet and volunteer potato under field conditions

Hyun K. Suh
Joris IJsselmuiden
Jan Willem Hofstee
Eldert J. van Henten

The contents of this chapter have been published in *Biosystems Engineering (2018), 174, 50-65* as a paper entitled: Transfer learning for the classification of sugar beet and volunteer potato under field conditions.

Abstract

Classification of weeds amongst cash crops is a core procedure in automated weed control. Addressing volunteer potato control in sugar beets, in the EU Smartbot project the aim was to control more than 95% of volunteer potatoes and ensure less than 5% of undesired control of sugar beet plants. A promising way to meet these requirements is deep learning. Training an entire network from scratch, however, requires a large dataset and a substantial amount of time. Then, transfer learning can be a promising solution. This study evaluates, in Part I, a transfer learning procedure with three different implementations of AlexNet and then, in Part II, assesses the performance difference amongst the six network architectures: AlexNet, VGG-19, GoogLeNet, ResNet-50, ResNet-101 and Inception-v3. All nets had been pre-trained on the ImageNet Dataset. These nets were used to classify sugar beet and volunteer potato images taken under ambient varying light conditions in agricultural environments. In Part I, the highest classification accuracy of 98.0% was obtained with an AlexNet architecture modified to generate binary output. In Part II, the highest classification accuracy of 98.7% was obtained with VGG-19 modified to generate binary output. Transfer learning proved to be effective and showed robust performance with plant images acquired in different periods of the various years on two types of soils. All scenarios and pre-trained networks were feasible for real-time applications ($classification\ time < 0.1\,\mathrm{s}$). Classification is only one step in weed detection, and a complete pipeline for weed detection may potentially reduce the overall performance.

5.1 Introduction

Volunteer potato is a source of potato blight (*Phytophthora infestans*) and viral diseases. Volunteer potato in a sugar beet field can reduce the crop yield by 30% (O'Keeffe, 1980). Under statutory obligation, sugar beet farmers in the Netherlands are required to control volunteer potato plants up to two remaining plants per m^2 by 1st of July (Nieuwenhuizen, 2009). For the automated control of volunteer potato in a sugar beet field, a vision-based and small-sized robot was developed within the EU-funded project SmartBot. Due to the small size of the robot and the required battery operation, the platform design had to refrain from additional infrastructure and should be able to robustly detect weeds in a scene that is fully exposed to ambient lighting conditions. Additional infrastructure such as a hood and lighting equipment, as used for instance by Nieuwenhuizen et al. (2010) and Lottes et al. (2016), was not considered viable. The robotic platform is shown in Figure 5.1.

Figure 5.1: The robotic platform for volunteer potato control in a sugar beet field.

The classification of weeds amongst cash crops, i.e. weed/crop discrimination, is the core procedure for automated weed detection. In a pipeline for weed detection, vegetation segmentation is followed by classification of the segmented vegetation into weeds and crop. This classification step traditionally involves two aspects: selection of the discriminative features as well as selection of the classification techniques (Suh et al., 2016).

Regarding the features used for discrimination, many studies have used color, shape (*biological morphology*) and texture on an individual basis or as a combination of multiple features (Ahmed et al., 2012; Åstrand & Baerveldt, 2002; Gebhardt & Kühbauch, 2007; Pérez et al., 2000; Persson & Åstrand, 2008; Slaughter et al., 2008; Swain et al., 2011; Zhang et al., 2008). However, these features have shown poor performance under widely varying natural light conditions (Suh et al., 2018a). Other features such as Scale Invariant Feature Transform (SIFT) (Lowe, 2004) and Speeded Up Robust Features (SURF) (Bay et al., 2008), have shown their potential in recent studies in the classification of plant species (Kazmi et al., 2015a; Suh et al., 2016; Wilf et al., 2016). However, the highest classification accuracy using SIFT and SURF obtained in Suh et al. (2016) is still not satisfactory for the requirements set by the previous study of Nieuwenhuizen (2009): the resulting automatic weeding system should effectively control more than 95 % of the volunteer potatoes as well as ensure less than 5 % of undesired control of the sugar beet plants. Therefore, within the framework of the EU Smartbot Project, a solution was needed that achieves a classification accuracy of 95 % or more as well as a misclassification of both sugar beet (false-negative) and volunteer potato (false-positive) of less than 5 %. In addition, a classification time of less than 0.1 s per image was also needed because these algorithms should be used in a real-time field application.

A promising way to meet above mentioned requirements is to use a deep learning approach. In recent studies, the deep neural network has shown its potential in an agricultural context for plant identification and classification. Grinblat et al. (2016) used a convolutional neural network (ConvNet, or CNN), a specific type of deep network, for plant identification from leaf vein patterns. Although the binary images of vein patterns were used, the study showed the potential of ConvNet for plant identification. Sun et al. (2017) used a residual network (ResNet), one of the most common ConvNet architectures used for classification tasks, for plant species identification with images acquired by mobile phones. A 91.78% of classification accuracy was obtained, but they

needed 10 000 images to train the network. Dyrmann et al. (2016) classified 22 plants species using a ConvNet and obtained 86.2% of classification accuracy. In their study, images were acquired under controlled conditions, a quite distinct difference with the conditions that SmartBot is confronted with, and the number of images needed to train the network from scratch was even more than 10 000. Obtaining such a large number of images, however, is a challenging task in agricultural fields (Xie et al., 2016). Besides, training an entire ConvNet from scratch requires a substantial amount of time (Jean et al., 2016; Yosinski et al., 2014) and is an expensive task that may be hard to realise in practice. Then, transfer learning can be a promising solution.

The objective and novelty of this paper are to deal with crop/weed classification under uncontrolled agricultural environments as well as to reduce the amount of data and time using transfer learning.

Transfer learning has gained its success in real-world applications (Jean et al., 2016; Shin et al., 2016; Sun et al., 2014; Xie et al., 2016). Transfer learning, according to Goodfellow et al. (2016), refers to exploiting what has been learned from one setting into another different setting. In transfer learning, the base network is trained on a base dataset and task, and then the (pre-)trained network is reused for another task (Yosinski et al., 2014). Interestingly enough, though the ConvNet is trained with a specific dataset to perform a specific task, the generic features extracted from ConvNet seem to be powerful and perform very well on other classification tasks as well (Donahue et al., 2014; Razavian et al., 2014). Transfer learning has recently been applied in several agricultural applications such as disease detection (Fuentes et al., 2017); however, the transfer learning procedure has not yet been investigated in detail in plant classification.

In this study, in Part I, three different transfer learning scenarios were evaluated using AlexNet (Krizhevsky et al., 2012). In Part II, the performance of following six pre-trained networks was compared: AlexNet, VGG-19 (Simonyan & Zisserman, 2015), GoogLeNet (Szegedy et al., 2015), ResNet-50 and -101 (He et al., 2016a) and Inception-v3 (Szegedy et al., 2016). The classification performance in both Part I and II was analysed regarding classification accuracy as well as training and classification time, given the fact that this approach should be used in a real-time field application.

The first section of this paper describes ConvNets and their popular architectures. The following section, Part I, describes three different scenarios in transfer learning. In Part II, the performance assessment amongst six pre-trained networks is described.

The experimental setup including field image dataset collection and the performance measures to be used are described. Then, the experimental results are shown with the corresponding discussions. In the end, conclusions are drawn.

5.2 Convolutional neural networks and popular architectures

Convolutional neural networks (ConvNets, or CNNs) are a specialised type of neural networks that are designed to process multi-dimensional data such as signals (1D), images (2D), and videos (3D) (LeCun et al., 1998, 2015). ConvNets have gained huge success in many applications since AlexNet has won the ImageNet competition in 2012 with a breakthrough performance (Sainath et al., 2013; Schwing & Urtasun, 2015; Sermanet et al., 2014; Zeiler & Fergus, 2014).

Motivated by the success of AlexNet, further deep ConvNets were proposed in the recent literature such as VGG-19, GoogLeNet, ResNet and Inception-v3. These ConvNets contain from several to hundred layers of convolutions with non-linear activation functions, such as Sigmoid, Tanh, and ReLU (Rectified Linear Units), applied to the results. A different set of convolution filters is applied over each layer, and then the output of the convolutions are combined to maintain the local connectivity between neurons of adjacent layers. This local connectivity enables each neuron to be connected only to a small local subset of the given image which helps to reduce the number of parameters in the whole network as well as to make the computation more efficient (Chen et al., 2014). Such a deep layered ConvNet structure enables the network to learn the best features during the training process automatically and will in most cases outperform hand-crafted feature extractors which generally require an extensive engineering skill and knowledge (Hu et al., 2015; LeCun et al., 2015).

AlexNet, one of the first ConvNets, contains seven layers besides input and output layers (Figure 5.2). The first five layers are convolutional layers (Conv layers) each followed by ReLU and max-pooling, which are non-linear activation and down-sampling functions to enhance the training time efficiency. The last three layers are fully-connected layers (FC layers) composed of two FC layers each with a 4096 dimensional activation vector followed by one FC layer (Softmax layer) with 1000 activation neurons, thus producing the classification score in terms of 1000 different categories.

In VGG-19, only 3×3 filters are used in all convolutional layers to reduce the number of parameters in the network. Furthermore, the usage of max pooling in between convolutional layers largely reduces the network volume (Simonyan & Zisserman, 2015). Like AlexNet, the last layers are two fully-connected layers, each with a 4096-dimensional activation vector, followed by a softmax layer.

In GoogLeNet, the Inception Module was introduced to process the required operations in parallel. The Inception Module acts as an efficient multi-level feature extractor and makes the network considerably smaller and faster. Szegedy et al. (2015) reported that GoogLeNet was smaller and faster than VGG-19 even though GoogLeNet contained more layers (22 layers) than VGG-19 (19 layers).

ResNet (Residual Network) consists of several basic residual blocks which provide a shortcut connection between layers. This shortcut connection makes it possible to train hundreds or more layers while achieving enhanced performance. ResNet is primarily designed for large-scale data analysis and is developed with many different numbers of layers including 50 and 101 (Alom et al., 2018). ResNet-50 and ResNet-101 contain, respectively, 50 and 101 convolutional layers including one fully-connected layer at the end of the network (He et al., 2016b).

Inception-v3 extends the original GoogLeNet implementation and enhances the Inception Module to improve the accuracy by factorisation of convolutions and improved

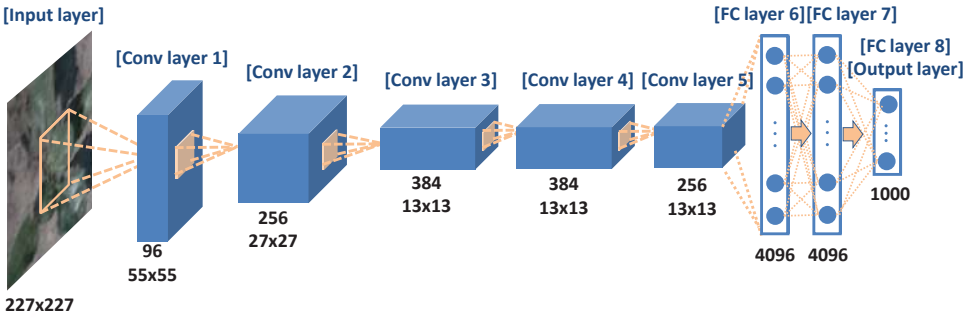


Figure 5.2: The overall structure of AlexNet. The network is composed of five convolutional layers (Conv layer 1-5) and three fully-connected layers (FC layers 6-8). FC layer 6 and 7 produce a 4096 *dimensional activation vector. The last layer, FC layer 8, is the output layer which produces classification scores on* 1000 *categories as the AlexNet was originally designed to classify* 1000 *different classes. The output size of each layer changes as a convolution process is being applied (Conv layer 1-5).*

normalisation (Szegedy et al., 2016, 2017). V3 simply indicates that this network is the 3rd version, updated and released by Google.

All these networks are available as pre-trained ConvNets which have been trained on ImageNet Dataset[1] to classify 1000 object categories such as desk, chair, keyboard, animals, etc. These ConvNets are used as pre-trained networks in this paper.

5.3 Part I: three scenarios for transfer learning

Transfer learning aims to overcome the shortage of training data and time by transferring information or features that are extracted from the pre-trained ConvNets (Oquab et al., 2014; Weiss et al., 2016). In Part I, AlexNet was used as a pre-trained ConvNet. Two options are available in transfer learning: use of ConvNet as a feature extractor and use of ConvNet as a classifier. Based on these available options, three scenarios for transfer learning were formulated based on the following hypotheses:

1) Scenario 1: In this scenario the hypothesis was tested whether or not without retraining AlexNet, a classification accuracy of 95 % or more could be achieved using the features extracted from FC6 and FC7 and using conventional classifiers.

2) Scenario 2: AlexNet is modified to produce binary classification output (i.e. sugar beet or volunteer potato). Once AlexNet was modified, it is fine-tuned with training images of sugar beet and volunteer potato. In this case, using more training data leads to a better classification accuracy than the one obtained in scenario 1.

3) Scenario 3: Once AlexNet was modified and fine-tuned as in scenario 2, the hypothesis was that an improved classification accuracy might be achieved using the features extracted from FC6 and FC7 and using a conventional classification scheme as used in scenario 1.

A total of 1100 labelled plant images was used. Each plant image was resized to 227×227 pixels (RGB) using a default image resizing function in Matlab as AlexNet has a predefined 227×227 pixel input size. No data augmentation was applied in Part I. In all scenarios, the classification performance was averaged over ten repetitions.

[1] The ImageNet Dataset contains 1.2 million labelled training images and 50 000 test images, with each image labeled with one of 1000 classes (Yosinski et al., 2014).

The classifiers in scenario 1 and 3 were validated by 10-fold random cross-validation over ten repetitions on a separate set of random images.

To get more insight into performance differences amongst different classifiers for scenario 1 and 3, the Support Vector Machine (SVM), random forest and linear discriminant analysis (LDA) were used for classification. These classifiers have been used in many agricultural applications (Ahmed et al., 2012; Longchamps et al., 2009; Lottes et al., 2016; Zhang et al., 2008), but it was not known a priori which classifier performs best on the classification task at hand. In the SVM, three different polynomial kernels (linear, quadratic and cubic) were evaluated. The classification performance was analysed regarding classification accuracy as well as training and classification time, given the fact that this approach should be used in a real-time field application.

5.3.1 Scenario 1 - AlexNet as a fixed feature extractor

One of the transfer learning approaches is to use a pre-trained ConvNet as a feature extractor (Jean et al., 2016). Without retraining the whole network, the features extracted from the last layers in ConvNets can be used as a feature vector which has generic properties applicable to other tasks using a conventional classification scheme (Donahue et al., 2014; Gong et al., 2013). In this scenario, the 4096 dimensional feature vector, was extracted from each of the last two fully-connected layers FC layer 6 (FC6) and FC layer 7 (FC7) as AlexNet yields vectors with 4096 feature values in these last layers. To investigate the difference in classification performance between the two layers of FC6 and FC7 in AlexNet, the extracted features in FC6 and FC7 were used individually to train and validate the following classifiers: SVM (with three different kernels), random forest, and LDA (Figure 5.3). The flowchart of scenario 1 is shown in Figure 5.4.

5.3.2 Scenario 2 - Modified and fine-tuned AlexNet as a binary classifier

Inspired by Papadomanolaki et al. (2016) who modified a pre-trained ConvNet to classify satellite data, in this scenario, AlexNet was modified to generate a binary classification output: sugar beet or volunteer potato. The original AlexNet was designed to classify 1000 objects, having 1000 activation neurons in the last layer of the network (Krizhevsky et al., 2012). This last layer in the original AlexNet was

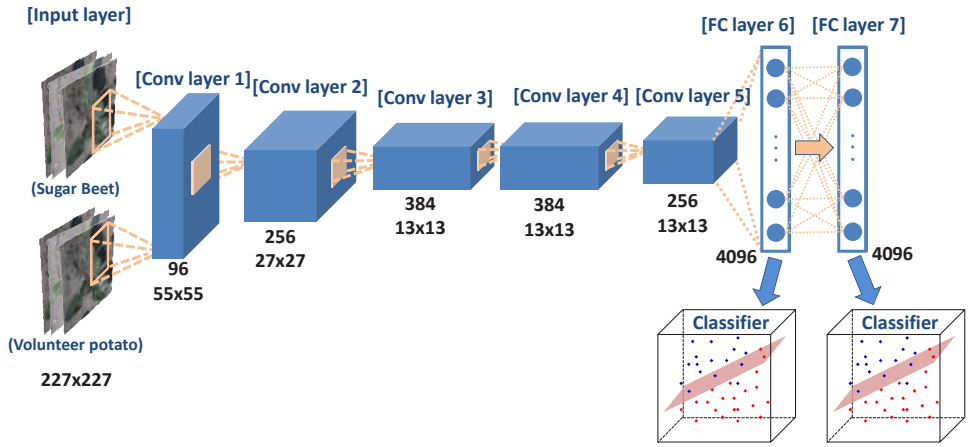


Figure 5.3: Scenario 1 - AlexNet as a feature extractor. Two fully-connected layers, FC layer 6 and 7, were each used individually as a feature extractor. For the classification of sugar beet and volunteer potato, three classifiers were evaluated: SVM (with linear, quadratic, and cubic kernels), random forest, and LDA.

5

removed, and two new fully-connected layers (FC layer 8' and 9') were added. The modification details are as follows (Figure 5.5):

1) The last fully-connected layer in AlexNet was removed (FC layer 8 in Figure 5.2).

2) A new fully-connected layer with 64 neurons (a square root of 4096) was added to the end, followed by a ReLU (Rectified Linear Unit), as ReLU was applied to the output of every layer in the original AlexNet.

3) A new fully-connected layer with two neurons was added with a 2-way softmax to produce the binary classification output.

Between FC layer 7 (size of 4096 neurons) and FC layer 9' (size of two neurons for binary output), FC layer 8' with 64 neurons was added to help smooth the dimensional reduction from 4096 to two (Figure 5.5). Preliminary experimentation showed that this addition produced slightly better performance compared to having no layer in between.

The modified AlexNet was then fine-tuned on our image dataset, which had been acquired during three different periods of three different years on two different soil

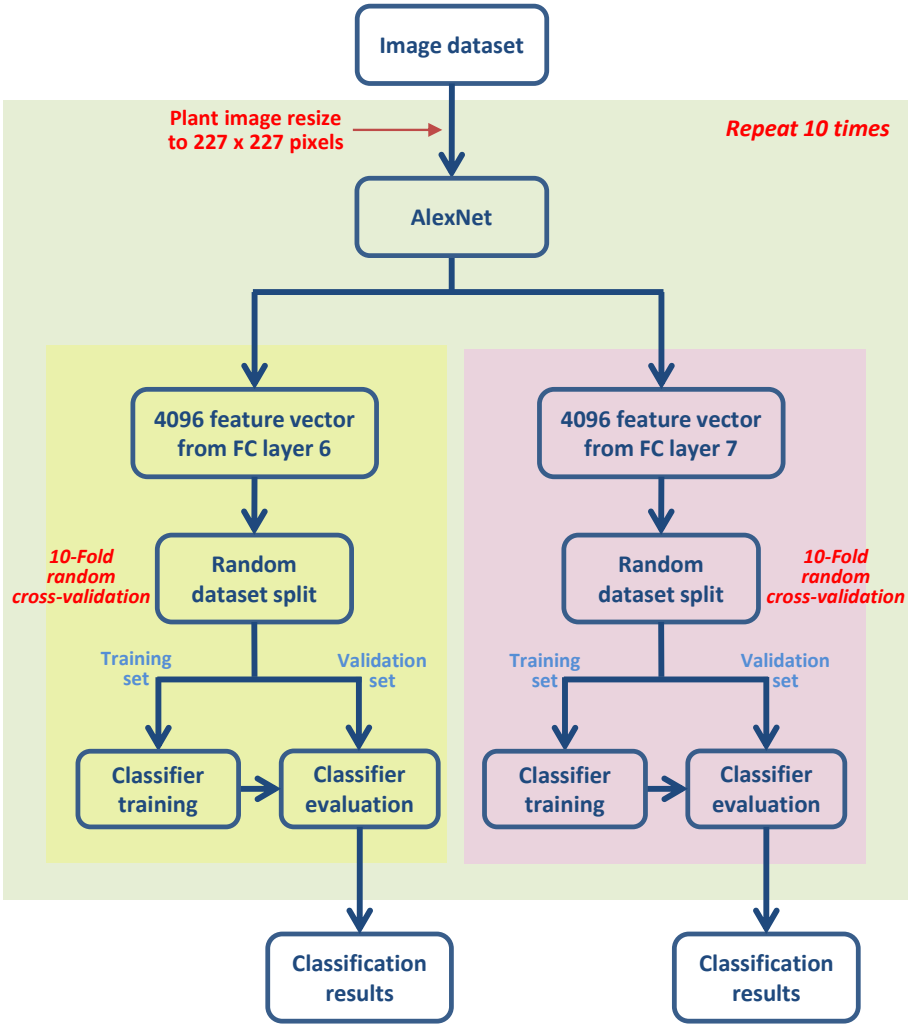


Figure 5.4: Flowchart of scenario 1. A total of 1100 *labelled plant images was used. Each plant image was resized to* 227×227 *pixels (RGB). The classifier results were validated by 10-fold random cross-validation over ten repetitions.*

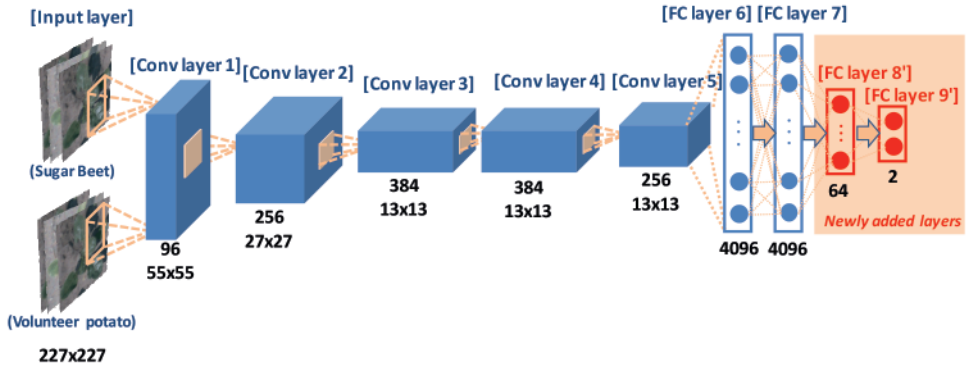


Figure 5.5: Scenario 2 - AlexNet was modified to generate a binary classification output: sugar beet or volunteer potato. The original FC layer 8 was removed, and new FC layers 8' and 9' were added in the end for AlexNet to generate a binary classification into sugar beet and volunteer potato plants.

5

types. The modified AlexNet was trained using a stochastic gradient descent (SGD) method with a batch size of 128 examples and a momentum of 0.9 as these parameters had also been used for training the original AlexNet (Krizhevsky et al., 2012). In order not to change the parameters of original convolutional layers too much, the learning rate was fixed to 0.001 and the learning was stopped after 20 epochs.

The classification performance of the fine-tuned network was expected to depend on the number of training images used. To find the optimal number of training images needed for fine-tuning the AlexNet, the classification performance was evaluated while varying the number of training images from 200 to 900 with an increment of 100. Training images were randomly selected out of the 1100 available images in the dataset, and from the remaining images, 200 images were randomly selected for validation.

5.3.3 Scenario 3 - Modified and fine-tuned AlexNet as a fixed feature extractor

This scenario is a combination of scenario 1 and 2. The original AlexNet was first modified to generate a binary classification output (sugar beet or volunteer potato) as described in scenario 2. The modified AlexNet was fine-tuned with two different numbers of images, 300 and 800, which were randomly selected out of the 1100 plant

images in the dataset. Then, new classifiers including SVM, random forest, and LDA were trained and validated using features extracted from the fully-connected layers 6 and 7 (FC6 and FC7) separately as described in scenario 1. For classifier training and validation, 300 images were randomly selected from the remaining images.

During the fine-tuning process, the modified AlexNet was expected to adjust its node weights based on the given plant images. This change would alter the value of the activation vectors in (all) layers in the network, which was likely to improve the classification performance compared to scenario 1.

5.4 Part II: classification performance amongst different ConvNet architectures

Following six pre-trained deep networks were evaluated to assess the classification performance amongst different ConvNet architectures: AlexNet, VGG-19, GoogLeNet, ResNet-50, ResNet-101 and Inception-v3. Each network was modified to produce binary classification output of sugar beet and volunteer potato, as was done in scenario 2 with AlexNet (section 5.3.2), by removing the last original layer and adding two new fully-connected layers. Then, each modified network was fine-tuned with 500 randomly selected images of sugar beet and volunteer potato; while the remaining 600 images were used for validation. The number of 500 was chosen for fine-tuning based on our preliminary studies as well as based on the fact that in scenario 2 with AlexNet, the accuracy improvement started to flatten after 500 (Figure 5.8). Plant images were resized to correspond to the input size of each network. Unlike in Part I, data augmentation was applied in Part II based on image transformations such as translation, rotation and flipping. Data augmentation includes a wide range of techniques used to generate new training images from the original ones by applying above-mentioned random image transformations. Data augmentation is to increase the generalizability of the model, and in most cases leads to an improvement in classification accuracy. It can then be assessed if applying data augmentation in Part II may yield a better performance compared to no data augmentation in scenario 2 in Part I (in the case of AlexNet).

The training parameters in Part II were the same as those described in Part I, but two different epochs, 20 and 30, were used for training to gain insight into the

performance difference between the two number of training epochs. No layers were frozen during training as it yielded slightly better performance than freezing some portion of the layers in our preliminary examination. The classification performance of each network in Part II was averaged over five repetitions.

5.5 Experimental setup

5.5.1 Field image collection and image dataset

For crop image acquisition, a camera was mounted at the height of 1 m perpendicular to the ground on a custom-made frame carried by a mobile platform (Husky A200, Clearpath, Canada) (Figure 5.6). A camera (NSC1005c, NIT, France) was equipped with two Kowa 5 mm lenses (LM5JC10M, Kowa, Japan) with a fixed aperture. The camera was set to operate in automatic acquisition mode with default settings. The camera had two identical CMOS sensors providing left and right images. Though this camera is intended to be used for stereovision, this feature was not used in this research. Left and right images were individually treated and separately used having each image resolution of 1280×580 pixels.

Sugar beet was sown three times (Spring, Summer, and Autumn) each year in 2013, 2014 and 2015 in sandy and clay soil at Unifarm experimental sites in Wageningen, The Netherlands. One week after sowing the sugar beet, potatoes were planted in random locations throughout the fields. The plant images were acquired under a wide range of illumination and weather conditions for several days in June, August and October of 2013, in May, June, July and September of 2014 and in May, June, July and October of 2015.

For the labelled image dataset used in this study, a total of 1100 individual plant images was manually extracted from selected field images: 550 sugar beet plants and 550 volunteer potato plants. During the selection of this dataset, images with different ambient light conditions were included as well as images containing various stages of plant growth and shadows which were caused by neighbouring plants and/or robotic platform. The size of each plant image in the dataset varied from the smallest size of 73×60 pixels to the largest of 310×315 pixels. Example images from this dataset are shown in Figure 5.7

5.5.2 Software and hardware platform

All procedures were implemented in Matlab® (The MathWorks Inc, Natick, MA, USA) using the Neural Network Toolbox™, Statistics and Machine Learning Toolbox™ and MatConvNet toolbox (Vedaldi & Lenc, 2015). As computing hardware platforms, two cloud servers were used from Amazon Elastic Compute Cloud (EC2) and Paperspace GPU Cloud. Amazon EC2 was used in Part I, and Paperspace GPU Cloud was used in Part II. These cloud servers provided a simple and easy setup of a high-performance computing platform with reduced cost of maintenance. Amazon EC2 was equipped with an Intel® Xeon® CPU E5-2670 2.5 GHz processor, 15 GB memory and Nvidia Grid™ K520 GPU running 64-bit Windows Server 2012. Paperspace GPU cloud was equipped with an Intel® Xeon® CPU E5-2623 2.6 GHz processor, 30 GB memory and Nvidia Quadro™ P5000 16GB GPU running 64-bit Ubuntu 16.04 LTS.

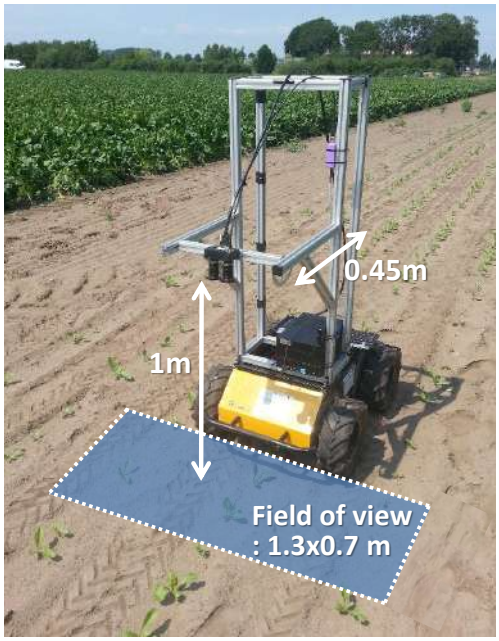


Figure 5.6: Field images were acquired with a camera mounted at the height of 1 m *viewing perpendicular to the ground surface resulting in a field of view of* 1.3 m × 0.7 m. *A mobile platform, Clearpath Husky, was manually controlled with a joystick and driven along crop rows using a controlled travelling speed of* 0.5 m/s.

5.5.3 Performance measures

A binary classification was performed in this study: sugar beet or volunteer potato. The classification performance measures for this study are described below.

A confusion matrix (Table 5.1) was used to assess and compare the classification performance. The classification accuracy was calculated along with training and classification time considering this work is for real-time field application. The classification accuracy and training time were averaged over ten and five trials in Part I and II, respectively. The classification time was measured for the time required to classify a single plant image on the cloud servers.

$$ClassificationAccuracy = \frac{TP + TN}{TP + FN + FP + FN} \tag{5.1}$$

where TP is true-positives; FP is false-positives; TN is true-negative, and FN is false-negative.

Figure 5.7: Example images from the field image dataset containing a total of 1100 *plant images with* 550 *sugar beet (top) and* 550 *volunteer potato plants (bottom). During the selection of this dataset, images with different ambient light conditions were included taken on both in sandy and clay soils as well as images containing shadows and various stages of plant growth.*

Table 5.1: Confusion matrix
(TR:true-positive, TN:true-negative, FP:false-positive, and FN:false-negative)

		Predicted Class	
		Sugar Beet (SB)	Volunteer Potato (VP)
Actual Class	Sugar Beet (SB)	TP	FN
	Volunteer Potato (VP)	FP	TN

5.6 Results

5.6.1 Part I: three scenarios for transfer learning

Scenario 1 - AlexNet as a fixed feature extractor

Three classifiers were trained, using supervised learning, based on the 4096 feature values that were extracted from each of AlexNet's two fully-connected layers FC6 and FC7 separately. The classification performance was evaluated with TP, FN, FP, TN, classification accuracy, training time and classification time as shown in Table 5.2.

Using the features from FC6, the highest classification accuracy of 97.0 % was obtained with a SVM with a quadratic kernel; while the lowest classification accuracy of 90.8 % was obtained with LDA. Likewise, using the features extracted from FC7, a highest classification accuracy of 95.8 % was obtained with a SVM and a quadratic kernel; while the lowest classification accuracy of 91.9 % was obtained with LDA. Using the features extracted from FC6 provided a better classification accuracy with the SVMs and the random forest than using the features from FC7; while with LDA, using the features extracted from FC7 provided a better classification accuracy than using the features from FC6.

The smallest false-negative and false-positive values were 21 and 12, respectively, which were obtained using the features extracted from FC6 and the SVM with a quadratic kernel. The false-negative number of 21 indicates that in total 3.8 % of sugar beet was classified as volunteer potato, and thus would be eliminated by the weed control robot. The false-positive of 12 indicates that 2.2 % of volunteer potato was classified as sugar beet, and thus would not be killed.

Table 5.2: Scenario 1 - The classification performance is shown using features that were extracted from each of AlexNet's two fully-connected layers in FC6 and FC7 separately. The classifiers were trained and validated with a total of 1100 images (550 of sugar beet and 550 of volunteer potato) using 10-fold random cross-validation. The final classification performance was averaged over ten repetitions. The training time includes times for feature extraction as well as training of the classifier. The classification time was measured for the time required to classify the class of a single plant image using a trained classifier.
(TR:true-positive, TN:true-negative, FP:false-positive, and FN:false-negative)

Input layer and classifier models			TP	FN	FP	TN	Classification accuracy (%)	Training time (s)	Classification time (s/image)
			(% of total)						
FC6	**SVM**	**Linear**	526	24	18	532	96.2	13.3	0.0143
			(95.6)	(4.4)	(3.3)	(96.7)			
		Quadratic	529	21	12	538	97.0	13.3	0.0143
			(96.2)	(3.8)	(2.2)	(97.8)			
		Cubic	527	23	16	534	96.5	13.3	0.0142
			(95.8)	(4.2)	(2.9)	(97.1)			
	Random forest		513	37	45	505	92.5	15.5	0.0154
			(93.3)	(6.7)	(8.2)	(91.8)			
	LDA		490	60	41	509	90.8	13.9	0.0217
			(89.1)	(10.9)	(7.5)	(92.5)			
FC7	**SVM**	**Linear**	515	35	24	526	94.6	14.6	0.0160
			(93.6)	(6.4)	(4.4)	(95.6)			
		Quadratic	523	27	19	531	95.8	14.6	0.0161
			(95.1)	(4.9)	(3.5)	(96.5)			
		Cubic	524	26	21	529	95.7	14.6	0.0161
			(95.3)	(4.7)	(3.8)	(96.2)			
	Random forest		512	38	47	503	92.3	16.7	0.0170
			(93.1)	(6.9)	(8.5)	(91.5)			
	LDA		499	51	38	512	91.9	15.2	0.0229
			(90.7)	(9.3)	(6.9)	(93.1)			

The training time includes the time needed for feature extraction as well as training of the classifier itself. Using the features extracted from FC6, the SVMs and LDA required 13 –14 s of training time while the random forest required 16 s of training time. Similarly, using the features extracted from FC7, the SVMs and LDA required 15 s of training time while the random forest required 17 s of training time. The average training time for one plant image was 0.014 s. The training times needed by all classifiers are reasonable, considering the training can be done offline and may not have to be repeated very often.

The classification time indicates the time required to classify (or predict) the class of a single plant image using a trained classifier. For all classifiers, an average of 0.016 s was needed using the features extracted from FC6, and an average of 0.018 s was needed using the features extracted from FC7. This classification time is fast enough for real-time application in the field ($classification\ time < 0.1$ s)

Scenario 2 - Modified and fine-tuned AlexNet as a binary classifier

The classification performance of the modified and fine-tuned AlexNet is shown in Figure 5.8. As expected, when the number of training images increased from 200 to 900, the classification accuracy increased from 89.1 % to 98.0 %. However, the classification accuracy did not linearly increase with the number of training images. The largest improvement in classification accuracy (4.4 %) was found when the number of training images changed from 200 to 300; while the smallest improvement in classification accuracy (0.3 %) was found when the number of training images was changed from 800 to 900.

The highest classification accuracy obtained in scenario 1 was 97.0 % as shown in Table 5.2. However, in scenario 2, only when more than 700 training images were used, a classification accuracy higher than 97.0 % was obtained.

The training time required for fine-tuning of the AlexNet linearly increased with the number of training images. For fine-tuning with 200 and 900 images, a training time of 94.9 s and 656.4 s was needed, respectively. The average training time for one plant image was 0.6 s. Comparing this training time with the training time in scenario 1 (0.014 s), training the deep network is found to be computationally more expensive than training the conventional classifiers.

In all cases, the classification time required to classify (or predict) the class of a single plant was 0.012 s, showing the fastest classification time among all scenarios.

Scenario 3 - Modified and fine-tuned AlexNet as a fixed feature extractor

The classification performance when the modified AlexNet was fine-tuned with 300 plant images is shown in Table 5.3. Using the features extracted from FC6, the highest classification accuracy of 96.7 % was obtained with a SVM and linear kernel; while the lowest classification accuracy of 93.0% was obtained with the random forest and LDA. A similar trend in the results was found when using the features from FC7. The highest classification accuracy of 96.3% was obtained with SVM and linear kernel; while the lowest classification accuracy of 91.0% was obtained with LDA.

In Table 5.3, the smallest false-negative and false-positive were 6 and 4, respectively, which were obtained using the features extracted from FC6 and the SVM with a linear kernel. The false-negative value of 6 indicates that in total 4.0% of sugar beet was classified as volunteer potato, and thus would be killed by the weed control robot. The false-positive value of 4 indicates that 2.7% of volunteer potato was classified as sugar beet, and thus would be left untreated by the weed control robot.

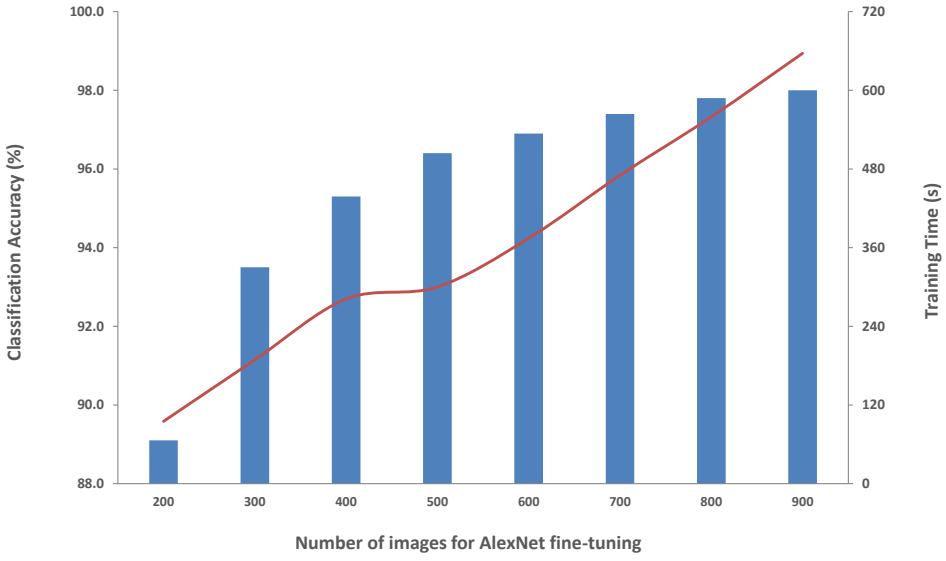


Figure 5.8: Scenario 2 - AlexNet was modified to produce binary output for the classification of sugar beet and volunteer potato. The modified AlexNet was fine-tuned with a varying number of training images. The bars are classification accuracy and the line is training time as a function of the number of images for fine-tuning of AlexNet.

In Table 5.4, the classification performance is shown when the modified AlexNet was fine-tuned with 800 plant images. Using the features extracted from FC6, a highest classification accuracy of 97.3% was obtained with the SVM and linear kernel; while the lowest classification accuracy of 95.3% was obtained with LDA. Using the features from FC7, the highest classification accuracy of 97.3% was obtained with random forest; while the lowest classification accuracy of 96.0% was obtained with the LDA.

In Table 5.4, the smallest false-negative and false-positive were 4 and 4, respectively, which were obtained using the features extracted from FC7 and using the random forest classifier. This misclassification indicates that 2.7% of sugar beet would be treated by the weed control device; while 2.7% of volunteer potato would not be treated. It should be noted that the same classification accuracy of 97.3% was achieved using FC6 with SVM linear and using FC7 with random forest. However, different false-negative and false-positive were obtained. Using FC6 with SVM linear, false-negative and false-positive values were 5 and 3, representing misclassification of 3.3% of sugar beet and 2.0% of volunteer potato, respectively. Furthermore, using FC7 with random forest, false-negative and false-positive values were 4 and 4, representing misclassification of 2.7% of sugar beet and 2.7% of volunteer potato, respectively.

Fine-tuning with 800 images produced better classification accuracy in all classifiers when compared to using 300 images for fine-tuning: a 0.6% and 1.3% increase in the highest classification accuracies using the features from FC6 and FC7, respectively; and a 2.4% and 4.7% increase in the lowest classification accuracies obtained when using the features from FC6 and FC7, respectively.

The training time includes the time needed for fine-tuning of AlexNet as well as training of the classifier. Using 300 images for fine-tuning, 195 to 197 s of training time was needed for all classifiers; while using 800 images for fine-tuning, 583 to 586 s of training time was needed for all classifiers. The average time for training the classifiers was 4.3 s showing that the major part of the required training time was used for fine-tuning of AlexNet. Again, training the deep network is a computationally more expensive task than training conventional classifiers.

The classification time required to classify a single plant image was 0.013 to 0.022 s for all classifiers. SVMs showed the fastest classification time; while LDA showed the slowest classification time among all classifiers. This classification time is fast enough for real-time application in the field ($classification\ time < 0.1$ s).

Table 5.3: Scenario 3 - After fine-tuning of the AlexNet with 300 images, the classifiers were trained with features extracted each from FC6 and FC7 separately. A total of 300 *training images for fine-tuning was randomly selected from* 1100 *plant images in the dataset. From the remaining images, a total of 300 images was randomly selected for classifier training and validation. The training time includes times for fine-tuning of AlexNet and training of the classifier. The classification time was measured for the time required to classify the class of a single plant image using a trained classifier. (TR:true-positive, TN:true-negative, FP:false-positive, and FN:false-negative)*

Input layer and classifier models			TP	FN	FP	TN	Classification accuracy (%)	Training time (s)	Classification time (s/image)
			(% of total)						
FC6	**SVM**	**Linear**	144	6	4	146	96.7	195.8	0.0130
			(96.0)	(4.0)	(2.7)	(97.3)			
		Quadratic	142	8	5	145	95.7	196.3	0.0131
			(94.7)	(5.3)	(3.3)	(96.5)			
		Cubic	142	8	6	144	95.3	195.2	0.0130
			(94.7)	(5.3)	(4.0)	(96.0)			
	Random forest		140	10	11	139	93.0	198.5	0.0134
			(93.3)	(6.7)	(7.3)	(92.7)			
	LDA		138	12	9	141	93.0	197.9	0.0180
			(92.0)	(8.0)	(6.0)	(94.0)			
FC7	**SVM**	**Linear**	143	7	4	146	96.3	196.3	0.0144
			(95.3)	(4.7)	(2.7)	(97.3)			
		Quadratic	143	7	5	145	96.0	197.1	0.0143
			(95.3)	(4.7)	(3.3)	(96.7)			
		Cubic	142	8	6	144	95.3	196.4	0.0143
			(94.7)	(5.3)	(4.0)	(96.0)			
	Random forest		143	7	9	141	94.7	197.9	0.0146
			(95.3)	(4.7)	(6.0)	(94.0)			
	LDA		135	15	12	138	91.0	196.4	0.0183
			(90.0)	(10.0)	(8.0)	(92.0)			

Table 5.4: Scenario 3 - After fine-tuning of the AlexNet with 800 *images, the classifiers were trained with features extracted each from FC6 and FC7 separately. A total of* 800 *training images for fine-tuning was randomly selected from* 1100 *plant images in the dataset. The remaining images,* 300 *images, were used for classifier training and validation. The training time includes times for fine-tuning of AlexNet and training of the classifier. The classification time was measured for the time required to classify the class of a single plant image using a trained classifier.*
(TR:true-positive, TN:true-negative, FP:false-positive, and FN:false-negative)

Input layer and classifier models			TP	FN	FP	TN	Classification accuracy (%)	Training time (s)	Classification time (s/image)
			(% of total)						
FC6	**SVM**	**Linear**	145	5	3	147	97.3	581.4	0.0135
			(96.7)	(3.3)	(2.0)	(98.0)			
		Quadratic	146	4	5	145	97.0	584.8	0.0135
			(97.3)	(2.7)	(3.3)	(96.7)			
		Cubic	145	5	5	145	96.7	583.2	0.0136
			(96.7)	(3.3)	(3.3)	(96.7)			
	Random forest		145	5	6	144	96.3	586.2	0.0140
			(96.7)	(3.3)	(4.0)	(96.0)			
	LDA		142	8	6	144	95.3	584.9	0.0204
			(94.7)	(5.3)	(4.0)	(96.0)			
FC7	**SVM**	**Linear**	145	5	4	146	97.0	583.9	0.0148
			(96.7)	(3.3)	(2.7)	(97.3)			
		Quadratic	146	4	5	145	97.0	584.5	0.0148
			(97.3)	(2.7)	(3.3)	(96.7)			
		Cubic	146	4	5	145	97.0	585.4	0.0149
			(97.3)	(2.7)	(3.3)	(96.7)			
	Random forest		146	4	4	146	97.3	586.9	0.0159
			(97.3)	(2.7)	(2.7)	(97.3)			
	LDA		143	7	5	145	96.0	585.8	0.0221
			(95.3)	(4.7)	(3.3)	(96.7)			

All scenarios: summary

The classification accuracies are summarised in Figure 5.9. Among all scenarios, both the highest and lowest classification accuracies were obtained in scenario 2. The highest classification accuracy achieved in scenario 2 was 98.0% while highest accuracy in scenarios 1 and 3 were 97.0% and 97.3%, respectively. In scenario 2, a higher classification accuracy than 97.3% was obtained when the number of images used for fine-tuning was more than 700. On the other hand, the lowest classification accuracy achieved in scenario 2 was 89.1% while those in scenario 1 and 3 were 90.8% and 91.0%, respectively. In case only a small number of training dataset was used, training conventional classifier yielded a better performance than fine-tuning the AlexNet. However, a large number of images for fine-tuning resulted in a better classification accuracy compared to the conventional classifier training.

Using the conventional classifiers in scenario 1 and 3, SVMs showed better classi-

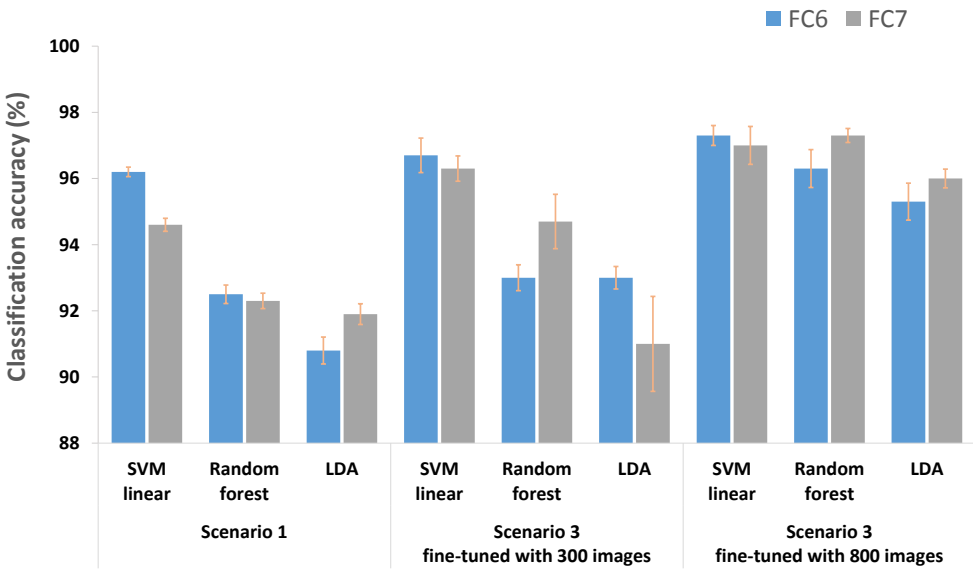


Figure 5.9: The obtained classification accuracies using SVM with a linear kernel, random forest and LDF in scenario 1 and 3 are summarised. The classifiers were trained using the features extracted each from FC6 (fully-connected layer 6) and FC7 (fully-connected layer 7) separately. In scenario 3, the AlexNet was fine-tuned with 300 and 800 images separately.

fication accuracy than the other classifiers. However, the difference in classification accuracy among the classifiers tended to decrease as AlexNet was fine-tuned with more training images.

5.6.2 Part II: classification performance amongst different pre-trained networks

The classification performance of the modified and fine-tuned deep networks (based on scenario 2) is shown in Table 5.5.

When the training was stopped after 30 epochs, again the highest classification accuracy of 98.7% was obtained with VGG-19; while the lowest classification accuracy of 94.8% was obtained with Inception-v3. However, this accuracy obtained with Inception-v3 was largely improved compared to when the training was stopped after 20 epochs. Yet AlexNet, VGG-19 and GoogLeNet did not yield such improvements.

The values in Table 5.5 indicate an average over five repetitions. Together with the stochastic nature of the training, this will result in some variation in the accuracy. This probably explains the decrease in accuracy with AlexNet when the training was

Table 5.5: The classification performance among six pre-trained deep networks was evaluated with two training epochs (20 and 30). Based on scenario 2, each network was modified and fine-tuned to classify sugar beet and volunteer potato. Randomly selected 500 images were used for training, while the remaining 600 images were used for validation. The classification performance was averaged over five repetitions and validated with classification accuracy, training time and classification time.

	Training 20 epoch			**Training 30 epoch**		
	Accuracy (%)	**Training time (m)**	**Classification time (s/image)**	**Accuracy (%)**	**Training time (m)**	**Classification time (s/image)**
AlexNet	97.9	9.0	0.0038	97.7	15.6	0.0040
VGG-19	98.4	37.4	0.0130	98.7	71.4	0.0124
GoogLeNet	97.0	23.8	0.0033	97.3	36.9	0.0035
ResNet-50	96.2	40.3	0.0072	97.2	69.8	0.0075
ResNet-101	97.5	106.6	0.0118	98.5	162.7	0.0111
Inception-v3	90.8	88.7	0.0088	94.8	133.3	0.0086

stopped after 20 epochs (97.9%) and 30 epochs (97.7%).

In Figure 5.10, the loss and accuracy for each epoch of the training with Inception-v3 and VGG-19 are shown. The accuracy for Inception-v3 still gradually improved even after 20 epochs; while the accuracy for VGG-19 more or less stabilised after only a small number of epochs. Likewise, the loss for Inception-v3 slowly reduced even after 20 epochs; while the loss for VGG-19 rapidly reduced from the first to five epochs and then reasonably stabilised although values were fluctuating a bit in between zero and 0.15. This result indicates that Inception-v3 requires more epochs to reach the highest accuracy and lowest loss than VGG-19. A similar trend in the results was also found with ResNet-50 and ResNet-101. Again, GoogLeNet required less training time than VGG-19. Also, in classification time, GoogLeNet still required the shortest classification time, even less than AlexNet, while VGG-19 required the longest classification time among all networks. The classification time in all networks was found to be fast enough for real-time application in the field.

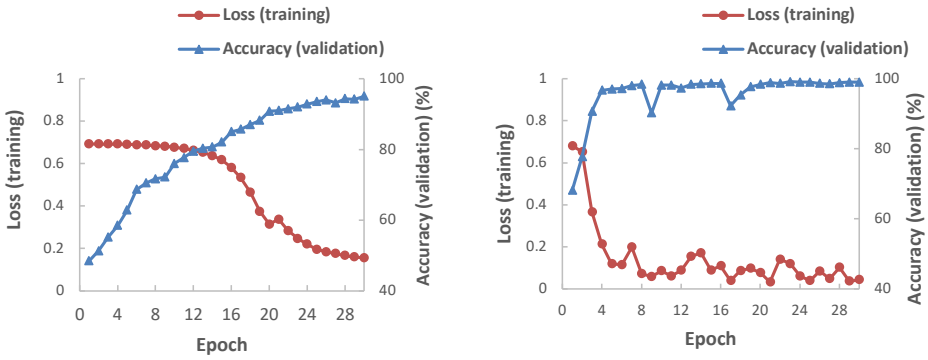


Figure 5.10: The loss and accuracy for each epoch of the training with Inception-v3 (left) and VGG-19 (right).

5.7 Discussion

The classification performance obtained using transfer learning in this study exceeds previously reported accuracies for instance by Persson & Åstrand (2008), Nieuwenhuizen et al. (2010), and Suh et al. (2016). Given the widely varying circumstances in

natural fields, the highest classification accuracy (98.7%) obtained in this study is considerably better, to the best of our knowledge, than any other approaches mentioned in the literature for crop and weed classification. To further substantiate the claim of considerable progress being made by using transfer learning, it would be beneficial to compare different algorithms, including ones previously used, on the same dataset. However, associating previously mentioned algorithms with this study might not yield a proper comparison because most, if not almost all, algorithms developed so far were based on image acquisition hardware including a hood covering the scene from ambient light and by illuminating the scene with artificial light. Algorithms were tuned for that purpose and for those conditions. However, in the current research approach no hood or artificial lighting were used. To compare algorithms on the current dataset would require retraining of the previously used algorithms which would not result in a fair comparison of results. Though a proper comparison is lacking, it seems fair to claim that with transfer learning of a ConvNet progress can be made in this field.

The proposed approach in this study was a partial implementation of a full pipeline for weed detection. The obtained results were based on manually extracted plant images, and vegetation segmentation procedure was not integrated. Implementing a full pipeline may potentially reduce the performance. In other words, the proposed approach does not lead to the precise detection of volunteer potato in field images. The individual plant detection procedure needs to be integrated as well. Sa et al. (2016) proposed a fruit detection system using ConvNet that detects each fruit in the large image even under occlusion. A similar approach can be used to detect each individual plant in crop fields.

Part I: proper scenario selection in transfer learning

When using transfer learning with ConvNets for weed classification, the most proper scenario needs to be selected based on the number of available training images. It was in scenario 2 that the highest classification accuracy was obtained, as the results in Part I indicated, only if a large number of images were used for fine-tuning. If a large number of images is not available, scenarios 1 and 3 would provide a better classification accuracy than scenario 2. Determining the required number of training images for the selection of scenarios may not be trivial, and perhaps may need more study as well. At least in Part I, 700 training images were needed in scenario 2 to obtain better classification accuracy than in scenarios 1 and 3. However, by applying

data augmentation in Part II, even 500 training images were enough to obtain higher classification accuracy than in scenarios 1 and 3. Nevertheless, some studies did not apply data augmentation as the performance improvement was not considered to be significant (Wan et al., 2013; Wang et al., 2017; Yosinski et al., 2014).

In addition, the number of classes (plant species in this case) may need to be considered for a proper scenario selection as well. Hasan et al. (2016) reported that when the number of classes was large, the performance of ConvNet was worse than SVM classifier in their classification task. It is unclear how a different number of classes would influence the performance in crop and weed classification. Only binary classification was performed in this study based on the assumption that in most cases plants found in sugar beet fields, especially in The Netherlands, are either sugar beet or volunteer potato. However, in some agricultural fields, several different weed species are often found together. A future study topic might be the multiclass classification for crop and several weed species as well as the classification performance assessment among a different number of plant species.

Using a SVM in scenarios 1 and 3, the extracted features from FC6 provided better classification accuracy than using features from FC7. Hu et al. (2015) also reported that the extracted features from the first FC layer consistently provided better performance compared to the second FC layer. However, when using other classifiers such as random forest and LDA, the extracted features from FC7 provided better classification accuracy than the ones from FC6. The behaviour or function of each layer in the deep network is not yet fully understood, and the deep network is still seen as a "black-box." Research has revealed that the first layer in many deep neural networks, when trained on images, tends to learn general features similar to the Gabor filter and colour blobs (Yosinski et al., 2014). More understanding of the function of each layer is, therefore, a topic for the future study.

Part II: different ConvNets architectures for weed classification

AlexNet showed a classification accuracy of 97.9% in Part II. Considering the fact that AlexNet contains far less layers than the other networks used in the study, the classification accuracy obtained with AlexNet was surprisingly good even compared to the top performers such as VGG-19 (98.7%) and ResNet-101 (98.5%). Moreover, the training time required by AlexNet was considerably less than others. This training time can even be further reduced, without sacrificing the performance, if training stops

after only 5-10 epochs since the accuracy during training was shown to be more or less stabilised after 5 epochs. Regarding the classification time, AlexNet was one of the fastest networks for classification, followed by GoogLeNet but only by a very small margin, which very well suits real-time applications. Given these results, it seems fair to use AlexNet in our application for the classification of sugar beet and volunteer potato, although AlexNet is already considered to be an "old-fashioned" network.

The number of epochs for training largely influences the classification performance depending on the network architecture. To reach the highest desired accuracy, some shallow networks such as AlexNet and VGG-19 require only a small number of epochs; while some deeper networks such as ResNet-101 and Inception-v3 require a relatively large number of epochs. It is unclear how to choose the optimum number of epochs for the training of the deep networks since selecting the optimum number has been mainly based on empirical experience (Jozefowicz & Com, 2015; Schmidhuber, 2015). For this reason, monitoring the training process with the loss and accuracy is particularly important to determine when to stop the training. Also, training of a deep network depends on other parameter settings such as learning rate, momentum and batch size, which in many cases also relies on empirical knowledge (LeCun et al., 2015). All these parameter settings are likely to influence the classification performance which may also influence the required number of training images and epochs needed to obtain high classification accuracy. It is worth investigating and thus worth to better understand the influence of various parameters used in the deep learning on the performance of the deep neural network.

VGG-19 is known to be an expensive and complex architecture regarding computational cost and number of parameters, which makes the network less suitable for real-time applications (Canziani et al., 2016). He et al. (2016a) discussed that VGG-19 has higher complexity and requires more computations than ResNet-101, even though VGG-19 has considerably less layers than ResNet. This was confirmed in Part II (Table 5.5) as VGG-19 needed the longest classification time than any other networks.

According to Yosinski et al. (2014), the effectiveness of transfer learning is expected to decline if there is less similarity between the network's original task and the new task at hand. All networks in our study were originally (pre-)trained with ImageNet Dataset which contained object images commonly found in ordinary life such as desk, computer, car, etc. ImageNet Dataset is quite distinctly different, so to speak, from sugar beet and volunteer potato images; yet, the performance obtained using transfer

learning in this study is still very impressive. If the networks were (pre-)trained with crop/weed field image dataset, further promising performance may likely be achieved as similarity will be higher between the network's original task and the new task at hand.

Practical considerations for weed control

For weed control in practice, it is critical to have a large as possible number of true-positives as well as a large as possible number of true-negatives. Not only that, it is also important to consider both the number of false-negatives (the number of sugar beet plants that are classified as volunteer potatoes) and the number of false-positives (the number of volunteer potato plants that are classified as sugar beet). The false-negatives lead to the removal of the cash crop caused by the misclassification, thus keeping the number of false-negatives as small as possible is critical (Lottes et al., 2016). At the same time, however, keeping the number of false-positives as small as possible is also desired. If there are many left over volunteer potato plants caused by misclassification, the weed control robot may have to drive repetitively through the field for Dutch farmers to meet the statutory regulation in the Netherlands (Nieuwenhuizen, 2009). The economic consequences of the different numbers of false-negatives and false-positives deserve further research.

Training and application of deep neural network require sophisticated hardware; high-performance GPUs. This requirement has been a limiting factor in many applications (Sa et al., 2016). However, cloud services (e.g. Amazon Elastic Compute Cloud and Paperspace GPU Cloud), as used in this study, provided a simple and easy way of using high-performance computing hardware without having to acquire and maintain the hardware on site.

Although the calculation time was measured on an EC2 server in this study, it is reasonable to think that a similar calculation speed (or even better calculation speed) can be achieved during on-field application because high-performance PCs (e.g. gaming laptops with high-performance GPUs) compared to the EC2 server used in this study are already available in the market.

5.8 Conclusions

This study evaluated a transfer learning procedure and assessed the performance amongst different ConvNet architectures for the classification of sugar beet and volunteer potato under ambient varying light conditions. Three different implementation scenarios were assessed using AlexNet in Part I, and the performance of following six pre-trained networks was compared in Part II: AlexNet, VGG-19, GoogLeNet, ResNet-50, ResNet-101 and Inception-v3.

Transfer learning provided very promising performance for the classification of sugar beet and volunteer potato images under ambient varying light conditions. In Part I, the highest classification accuracy (98.0%) was obtained with AlexNet in Scenario 2. In scenario 1 and 3, the highest classification accuracy of 97.0% and 97.3% were obtained, respectively.

All three scenarios were feasible for real-time field applications, but training the deep network was a computationally more expensive task than training the conventional classifiers.

The highest classification accuracy (98.7%) obtained in Part II, to the best of our knowledge, was considerably better than any other approaches mentioned in the literature for crop and weed classification. Data augmentation may improve the classification accuracy. VGG-19 yielded the highest classification accuracy but needed the longest classification time. AlexNet required the shortest training time; while ResNet-101 required the longest training time. With Inception-v3 using 30 epochs instead of 20 epochs for training yielded a significant improvement in performance. Such improvements were not observed when using more training epochs with AlexNet, VGG-19 and GoogLeNet.

Three different scenarios as well as six different ConvNet architectures for transfer learning showed robust performance with the plant images acquired in different periods of the various years with two types of soils. However, implementing a full pipeline for weed detection may potentially reduce the overall performance.

5.9 Acknowledgements

The work presented in this paper was part of the Agrobot part of the Smartbot project and funded by Interreg IVa, European Fund for the Regional Development of the

European Union and Product Board for Arable Farming. We thank Gerard Derks at experimental farm Unifarm of Wageningen University & Research for arranging and managing the experimental fields.

CHAPTER 6

Conclusions, General Discussion and Recommendations

As a final chapter of the thesis, this chapter lists the conclusions of the research and the sub-objectives each corresponding to Chapter 2 to 5 in Section 6.1. Then, section 6.2 will put the results obtained into perspective by reflecting on the required functionality of a fully autonomous weeding robot and by doing so identify directions for future research.

6.1 Conclusion

As was described in Chapter 1, the primary objective of this research was "*to develop a computer vision procedure that detects volunteer potato plants under ambient light conditions in a sugar beet field.*" The developed procedure was to be used for a weed control robot in the framework of the EU SmartBot project.

For a complete weed control pipeline, including weed detection and weed removal, the following requirements were set. The resulting automatic weeding system should:

- effectively control more than 95% of the volunteer potato;
- ensure less than 5% of undesired control of sugar beet plants;
- ensure a classification time of less than 1 second per field image for real-time operation in the field.

It was indicated that due to the potential non-perfect performance of actual weed removal, classification accuracy should be considerably higher than 95%.

To realize the main objective, various sub-procedures for vegetation segmentation (Chapter 2 and 3) and sugar beet/volunteer potato classification (Chapter 4 and 5) were developed.

Chapter 2 addressed the research question: *"Does a ground shadow detection and removal enhance the performance of vegetation segmentation under natural illumination conditions in the field?"*

In Chapter 2, an algorithm was described and evaluated for ground shadow detection and removal based on colour space conversion and a multilevel threshold. The advantage of using the proposed algorithm was assessed for vegetation segmentation with field images that were acquired by a High Dynamic Range (HDR) camera under natural illumination. Compared with no shadow removal, applying shadow removal enhanced the performance of vegetation segmentation under natural illumination conditions in the field with an average of 20%, 4.4% and 13.5% in precision, specificity and modified accuracy, respectively, and did not reduce segmentation performance when shadows were not present. The average processing time was 0.46 s, which is feasible when real-time application in the field is considered.

Chapter 3 addressed the research question: *"Do different combinations of colour index and threshold technique result in different segmentation performance when evaluated on field images? Given the varying conditions in the field, is it better to use one specific combination at all times or the combination should be adapted to the field conditions at hand for best segmentation performance?"*

In Chapter 3, the performance of 40 combinations of eight colour indices and five threshold techniques for vegetation segmentation were evaluated. A clear difference in performance, represented in terms of MA (Modified Accuracy), was observed among various combinations under the given conditions of this research. CIVE+Kapur showed the best performance, while VEG+Kapur showed the worst on the dataset. When adapting the combination to the given conditions yielded

a slightly higher performance than when using a single combination for all (in this case CIVE+Kapur). Consistent results were obtained when validated on a different independent image dataset. The expected advantage of adapting the combination to the field condition is not large because it seems that for practical use, the slight improvement when adapting the combination to the field conditions does not outweigh the investment in sensor technology and software needed to accurately determine the different conditions in the field.

Chapter 4 and 5 focussed on classification and addressed the following research questions: *"Does an algorithm using a Bag-of-Visual-Words (BoVW) model and SIFT or SURF descriptors meet the requirements set for the classification of volunteer potato and sugar beet under natural and varying daylight conditions? If the BoVW model does not meet the requirements, does a deep learning approach, particularly transfer learning based on Convolutional Neural Network (ConvNet, or CNN) provide an effective and better performance to meet the requirements with limited amount of dataset? Are the processing times (or calculation times) fast enough for real-time application?"*

For the classification of sugar beet and volunteer potato under ambient varying daylight conditions, Chapter 4 proposed a classification algorithm using a Bag-of-Visual-Words (BoVW) model based on SIFT or SURF features as well as crop row information in the form of the Out-of-Row Regional Index (ORRI). The highest classification accuracy of 96.5% with false-negative of 0% obtained using SIFT and ORRI with SVM is considerably better than previously reported approaches for weed classification; however, the false-positive rate of 7% deviates from the requirements since misclassification should be less than 5%. The average classification time of 0.10 - 0.11 s met the real-time requirements. Adding location information (ORRI) improved overall classification accuracy significantly. The proposed approach proved its potential under varying natural light conditions.

Since the required classification accuracy was not obtained in Chapter 4, further research was carried out for the classification of sugar beet and volunteer potato under ambient varying daylight conditions. Chapter 5 evaluated a transfer learning procedure with three different implementations of AlexNet (Part I), and then assessed the performance amongst different ConvNet architectures (Part II): AlexNet, VGG-19, GoogLeNet, ResNet-50, ResNet-101 and Inception-v3. In Part I, the highest classification accuracy (98.0%) was obtained with AlexNet in

Scenario 2. In scenario 1 and 3, the highest classification accuracy of 97.0% and 97.3% were obtained, respectively. In Part II, the highest classification accuracy of 98.7% was obtained. This result, to the best of our knowledge, was considerably better than any other approaches mentioned in the literature for crop and weed classification. Transfer learning provided very promising performance for the classification of sugar beet and volunteer potato images under ambient varying light conditions. A deep learning approach based on ConvNet provided better performance than the one in Chapter 4, and satisfied the requirements. All procedures were feasible for real-time field applications (the classification time < 0.1 s).

As was indicated in the introduction in Chapter 1, the full pipeline for volunteer potato detection consists of three steps: 1) vegetation segmentation, i.e. separating pixels in an image into plant pixels and non-plant pixels, 2) individual object (plant) identification, i.e. identification of individual plants in the set of plant pixels obtained after segmentation, and 3) classification of the plants into two classes, sugar beet (crop) and volunteer potato (weed).

In this thesis, steps 1 and 3, i.e. image segmentation and classification of sugar beet/volunteer potato were successfully addressed. Step 2, the identification of individual plants in the images was not addressed. Despite this limitation, it can be concluded that significant progress has been made in this area of study, given the fact that reported algorithms were developed using images captured in full daylight with significant variations in light colour and intensity; a distinct challenge that so far has been circumvented by using hoods and artificial lighting. Yet, the question remains unanswered whether a full pipeline, including all three steps, would be able to meet the requirements identified at the onset of the research.

With current hardware and suitable implementation of software, it seems that the requirement of 1 s per image for real-time operation of a weed control system can be attained. The highest classification accuracy of 98.7% obtained in Chapter 5 is supportive in meeting the required 95% control of volunteer potatoes, but when the ConvNet classification would be implemented in a full pipeline also containing vegetation segmentation and individual plant identification, a degraded performance can be expected. Preliminary results of Li (2017), in which a full pipeline for weed detection was implemented, indicate that a bottleneck in individual object identifica-

tion, step 2 in the pipeline, results in an overall classification performance of 87.5%, which does not meet or exceed the required 95% correct classification. Non-perfect weed control based on such a weed detection algorithm will therefore show an even lower performance of the overall weed control system. Yet, the results of Li (2017) show the potential of this approach as they were obtained under varying daylight conditions, and compare favourably with classification results in the range of 85-90% that were obtained in various previous researches using hoods and artificial lighting (e.g. Nieuwenhuizen et al. (2010) and Haug et al. (2014)). Therefore, it is safe to say that this research has laid the foundation for a small-sized robotic platform to come into action for weed control in the field.

6.2 Reflections and future directions

The introduction in Chapter 1 outlined the functionalities of a fully autonomous weed control robot: 1) autonomous navigation of a mobile platform carrying a sensor based weed removal device, 2) classification and identification of weed and crop, 3) removal of weeds. Each of these functionalities will be (shortly) addressed and reflected upon to identify potential future directions for research.

6.2.1 An autonomous small-sized mobile robotic platform

An increasing desire and concern in sustainable and environment-friendly production in arable farming has raised an issue with heavy machinery usage. Heavy machinery is often associated with soil compaction and irreparable damage to the soil as well as with adverse impacts of fossil fuel consumption. For instance, some of the most massive agricultural machines weigh 60 tonnes which leave a trail of soil compaction that could last for years (King, 2017). These issues can be minimised or avoided with a small-sized and autonomous robotic platform because such a lightweight system would not crush the soil. Moreover, a small robotic platform may have less critical safety issues than large machines as well as less environmental impact with lower usage of energy or battery operation.

A single small-sized robotic platform, however, may not be practical to cover a vast area of the crop field due to the limited operating hours. For example, a robotic platform used in this study, Husky A200 (Figure 1.2), can drive only up to two hours

due to the limited battery capacity. This working time will likely be further reduced when spraying actuation in the field is concerned. A potential way to increase the operating hours would be to mount a solar panel on the platform which can generate extra electricity during field operation, such as RIPPA (Robot for Intelligent Perception and Precision Application), the University of Sydney's solar-powered prototype robot. RIPPA has an array of solar cells mounted on the top, and it can run more than 20 hours converting solar radiation continuously into electric energy in the field (Sukkarieh, 2016). Another option to extend the operation time would be to use a supplementary power generator mounted to the rear or top side of the mobile robotic platform (Furgale & Barfoot, 2010). A compact size and lightweight power generator may add weight and mass to the overall system, but will provide extended operating time in the field. Along with these, a collection of small-sized robotic platforms collaborating as a fleet of robots, a swarm robotics approach, may have the potential to cover a large area for weed control (Emmi & Gonzalez-de Santos, 2017). The practicability of a swarm robotics approach in agricultural field applications is still an open research topic.

Autonomous navigation is one of the critical functionalities required for an automated weed control system using a small-size robotic platform. Autonomous navigation of such systems would require not only following the crop row without damaging any cash crops, but also moving to the next crop row when the system gets to the headland, in other words, the headland turn (Backman et al., 2015; English et al., 2014). Within the SmartBot project, crop row following functionality was implemented using a Husky platform during BSc and MSc thesis works (Janssen, 2015; Jol, 2015). In this work, a particle filter was used based on the method reported by (Hiremath et al., 2014). Satisfactory performance was achieved during the short field test in 2015 (Janssen, 2015). However, for complete autonomous navigation in agricultural fields, headland turn functionality needs to be developed and integrated into the robotic platform.

Additionally, if the robotic platform uses the battery, the battery needs to be continuously checked to avoid any potential stop due to battery depletion in the middle of the crop field. This issue needs to be considered and taken into account in autonomous navigation as well. Moreover, the system should have obstacle awareness as the system may encounter puddles or (dead) animals during navigation, although this is an unlikely situation in crop fields in the Netherlands.

6.2.2 Volunteer potato classification and identification

To bring vision-based weed classification and identification further to practice various issues require attention. Here are the issues for further consideration.

Camera technology

In this study, a HDR camera (NSC1005c, New Imaging Technologies, Paris, France) having a dynamic range of 140 dB was used. This camera was one of the cameras on the market in 2013 having a very high dynamic range. As was stated in Chapter 2, a HDR camera was known to provide a way to resolve the issue of varying natural illumination and substantial intensity differences within a single image scene in agricultural field conditions. Thus, using a HDR camera in an agricultural field under natural light conditions was expected to have added value. However, the difference in performance between a HDR camera and a traditional non-HDR camera was not quantitatively evaluated, as this was out of the scope of this research. It is suggested to investigate the performance of different dynamic range cameras under agricultural field conditions in a future study.

While the HDR camera is expected to bring an added value in agricultural applications, the use of deep learning seems to reduce the importance of camera performance compared to when traditional computer vision algorithms are used. In recent studies, even with a simple and cheap camera, deep learning has shown very promising performance under challenging agricultural environments where traditional computer vision algorithms most likely fail to achieve any successful results (Fuentes et al., 2017; Mohanty et al., 2016; Tibbetts, 2018). Therefore, it is worth investigating the benefit of the HDR camera in relation to the processing pipeline in the future study.

Limitation of colour- and threshold-based approach for vegetation segmentation

For vegetation segmentation in Chapter 2 and 3, all the methods were mainly based on RGB pixel values, which means that vegetation segmentation in this research depended on colour. Colour is indeed one of the most discriminative features for discriminating vegetation and soil background; however, using a colour-based approach may not yield sufficient vegetation segmentation performance in a system that has to work under

ambient light conditions (Yu et al., 2015). Unlike the controlled indoor environment, the illumination in an agricultural field changes dramatically with time and weather conditions, which often causes serious misclassification. These illumination variations significantly affect colour RGB pixel values of acquired field images and lead to the inconsistent colour representation of plants and soil background, which makes a colour-only approach very challenging to resolve. Moreover, the colour-only approach can be sensitive to noise disturbance especially in uncontrolled lighting conditions (Mythili & Kavita, 2011). To overcome these limitations, several recent studies have proposed hybrid approaches for plant segmentation. A hybrid approach proposed by Chopin et al. (2016) utilizes basic apriori information about the plant shape and local image orientations. Mancini et al. (2017)'s proposed algorithm takes into account muti-spectral as well as synthetic features that were derived from their developed models. Pande-Chhetri et al. (2017) utilized an object-based analysis using spectral, textural and geometrical object features computed from the individual pixels within each object. These hybrid approaches seem to offer the potential for future research.

On the other hand, thresholding techniques, in general, have their limitations because they only consider the intensity of the given images, not any spatial coherence between the pixels nor any consideration of object structure in the image. Thus, the pixels or objects identified by the threshold are not contiguous. Solomon & Breckon (2011) reported that using only the intensity of histogram does not guarantee for an optimum threshold value. Besides, the thresholds in Chapter 3 were based on the assumption that an image contains only plants and soil background (two classes), in which the histogram of the near-binary image needs to be partitioned into two classes. Although hardly any other objects than soil and plants were found in field images, a crop image scene may contain various kinds of straw, straw ash, and rocks (Yang et al., 2015). If an image scene contains a significant amount of the above mentioned or other materials, the thresholds would have limited performance. To overcome these limitations, Liu et al. (2012) proposed a new threshold method that utilises spatial information by combining image gradient with class uncertainty, and thus requires no predefined number of classes. Alternatively, several other methods for an improved threshold are also mentioned in the literature such as multiband thresholding, thresholding from a texture in the combination of regional boundaries, multiple thresholding criteria, and thresholding based on conditional histograms (Russ & Neal, 2015). Moreover, other threshold techniques than these are also found in the

literature: Sauvola (Sauvola & Pietikäinen, 2000), improved Sauvola (Shafait et al., 2008), Wang threshold using Parzen window technique (Wang et al., 2008), Ramesh threshold using functional approximation of the given histogram (Ramesh, 1995), etc. These threshold techniques are not reported to be used in agricultural applications and would be interesting research subjects in the future.

Individual object (plant) identification and overlapping plants

In a full pipeline for weed detection (Chapter 1), individual object (plant) identification is followed by vegetation segmentation, and each identified plant is the fundamental element in the subsequent process of weed and crop classification. Blob-based detection and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) are commonly used algorithms for the identification of individual objects in computer vision applications. In agricultural applications, these algorithms were also used to identify individual plant and fruit in recent studies (Li, 2017; Li et al., 2016; Yamamoto et al., 2014). They can be potential options for a full pipeline in weed detection.

Blob-based detection and DBSCAN, however, do not resolve the overlapping issue (Kurtulmuş & Kavdir, 2014). In an agricultural field, weeds are often found to be overlapped by crops. When crops and weeds grow close together and thus overlap each other, they tend to be identified as a single plant (Persson & Åstrand, 2008; Xia et al., 2013). This misidentification causes a substantial error in crop/weed classification, and therefore a solution is needed to identify an individual plant when they are overlapped.

A potential way to solve this issue is perhaps to use 3D imaging since depth information over 2D images provided satisfactory results to separate individual plant even overlapping conditions in a recent study of Li & Tang (in press). Young & Pierce (2014) also discussed the potentials of 3D imaging for an individual plant detection in overlapping conditions. However, for a practical application in the field, computation time needs to be reduced since 3D imaging requires a significant amount of processing time (Kapach et al., 2012). Kazmi et al. (2015b) further discussed that 3D sensing may have its own set of challenges especially when it comes to outdoor field applications. Alternatively, a recent study of Wang et al. (2018) proposed an algorithm based on Chan-Vese model and Sobel operator to segment overlapping plant regions. Although potential results were obtained using the images acquired in a partially controlled environment, they discussed that their proposed algorithm was not robust against

illumination changes and direct sunlight.

Other potential ways to detect the overlapping plants are to use Selective Searches or Region-based ConvNet (R-CNN) (Girshick et al., 2014; Uijlings et al., 2013). R-CNN has shown its potential to detect overlapping objects in many applications (Schmidhuber, 2015). Sa et al. (2016) showed the promising performance of R-CNN for the detection of overlapping fruits in an orchard.

Interestingly enough, Faster R-CNN is known to provide an "end-to-end" solution, producing detection and classification results simultaneously by merging all the required steps such as region proposal extraction, feature extraction and object classification, and bounding box regression (which can translate in this study vegetation segmentation and weed/crop classification) into the CNN (Lu et al., 2016; Tychsen-Smith & Petersson, 2017). Multi-task training and significant weight sharing within the CNN have not only enabled higher detection speeds, but also ensured higher detection quality than a non-end-to-end approach (Akselrod-Ballin et al., 2016). Such a solution offers a convenient way for detection because managing a pipeline of sequentially-trained tasks is no longer needed. In literature, Faster R-CNN has shown promising performance for real-time detection applications in challenging conditions in agricultural fields (Bargoti & Underwood, 2017; Fuentes et al., 2017; Sa et al., 2016). However, such solution requires a large amount of training data, which is one of the drawbacks of using deep neural network training in practice.

Field and crop conditions

Field and crop conditions need to be considered for the actual application of the weed control system in an agricultural field. Nieuwenhuizen (2009) reported that soil type might have a significant influence on the performance of vision-based weed detection, which in consequence will affect the control performance. In this research, both clay and sandy soil fields were considered during image acquisition. The proposed methods in this study seemed to work quite well in vegetation segmentation and weed/crop classification on both types of soils. However, the algorithms that are tuned to a particular soil type may likely perform better on that soil, which leaves a potential topic for future study. Also, the other types of soils need to be studied for further verification.

In addition, depending on irrigation management, tillage intensity and weather

situation, the soil field conditions can vary and influence the system performance. Under dry soil conditions, for example, when the weed control system rides on the field, a cloud of dust may continuously arise from the dry ground which negatively affects the performance of the camera system.

Similarly, different crop conditions shown by shape, colour, and growth pattern can be observed depending on cultivar, culture type, nutritional status, etc., and this may also influence the performance of weed detection and classification. Deep learning appears to be able to offer a promising solution in plant species classification even in irregular shape and pattern (Dyrmann et al., 2016; Ghazi et al., 2017); however, the performance details need to be further investigated.

Image dataset and ground truth

The image datasets in Chapter 2 and 3 contained a broad range of natural illumination encountered in the field, including extreme situations, and was meant to be representative of the conditions that can appear in the field (from different days and different seasons). Although a representative selection was made with at least some extreme situations, a substantial number of images in the dataset might have brought more insights. However, this would require ground truth assessment of many more images which was not considered feasible within the current project. The proper identification of different conditions was not considered feasible either.

In Chapter 3, the image dataset was divided into nine groups based on the plant size, illumination, and presence of shadow. Illumination and the presence of shadows were categorised only by two conditions: either sunny or cloudy (illumination), and either yes or no (the presence of shadow). In an agricultural field environment, however, illumination and shadow conditions are more complicated than categorising into merely two different conditions. More diverse categories could be generated based on some quantified criteria for better grouping of the environmental conditions.

The ground truth images in Chapter 2 and 3 were labelled by only two persons. Although the images were binary-labelled, either plant materials or background soil, the ground truth labelling in this study may bring a concern due to the lack of validation of additional annotators. Bac (2015) indicated that labelling of objects in images could differ among annotators, and thus one should employ several annotators to improve the reliability of ground truth. However, it is uncertain how many annotators might

be needed for reliable ground truth labelling.

The use of the term "ground truth" might be debatable because strictly speaking there might be no "truth" to the values measured in situ versus values measured proximally or remotely by a digital device. There is always a chance that measurement errors may occur in the process of making in situ measurements on the ground using digital devices. The use of an alternative term such as "ground observation" or "in-situ observation" might be an option although these alternative terms may likely create more confusion and ambiguity. Still, the term "ground truth" is one of the most commonly used terms in computer vision and image processing applications.

6.2.3 Actuation system for weed control

Even though there has been an environmental concern regarding the use of chemicals to control volunteer potatoes in a sugar beet field, the chemical application is considered one of the most cost-effective and practical control methods (Kunz et al., in press; Pedersen et al., 2006). Once the precise location of a volunteer potato plant is detected, the minimum amount of chemical deposition to the exact target location is required not only to minimise the environmental impact but also to prevent sugar beet plants being damaged by chemical drift.

For precision chemical application, some studies have attempted to control and regulate micro sprayer nozzles. Midtiby et al. (2011) developed a micro-spraying system based on inkjet printer nozzles. They showed the potential of a micro-spraying system for real-time weed control although they discussed the system lacked sufficient timing precision to target and control small plants. Nieuwenhuizen et al. (2010) developed micro-sprayers in an automated weed control system, showing selective weed control with minimum and precise use of chemicals in the field. For weeding in SmartBot, the micro-sprayer developed by Nieuwenhuizen et al. (2010) was to be reused with suitable modification to fit into the small-sized robotic platform. Precise timing and positioning of chemical droplets may need to be further enhanced for a small-sized robotic platform.

Other options than chemical control are also available in literature including a mechanical device using cutting tools, thermal flaming, and laser (Fennimore et al., 2016). Such weeding devices are advantageous for organic farmers as they are organic-compliant, and thus they can be a promising alternative for integrated weed man-

agement. However, further improvements seem to be needed for a practical usage in the field using a small-sized robotic platform. Frasconi et al. (2017) developed a weeding system based on mechanical and thermal flaming to remove weeds in maize fields. Although a promising performance was achieved, the total mass of the machine of more than 900 kg makes it difficult to be used with a small-sized robotic platform. Xiong et al. (2017) developed a prototype laser weeding system and showed the potential of laser weeding on a small-sized robotic platform. However in their research, the performance test was conducted under laboratory conditions, and thus further evaluation in an agricultural field condition is needed.

References

Ahmed, F., Al-Mamun, H. A., Bari, A. H., Hossain, E., & Kwan, P. (2012). Classification of crops and weeds from digital images: A support vector machine approach. *Crop Protection*, *40*, 98–104.

Akselrod-Ballin, A., Karlinsky, L., Alpert, S., Hasoul, S., Ben-Ari, R., & Barkan, E. (2016). A Region Based Convolutional Network for Tumor Detection and Classification in Breast Mammography. In *Carneiro G. et al. (eds) Deep Learning and Data Labeling for Medical Applications* (pp. 197–205). LABELS 2016, DLMIA 2016. Lecture Notes in Computer Science, vol 10008. Springer, Cham.

Al-Najdawi, N., Bez, H. E., Singhai, J., & Edirisinghe, E. A. (2012). A survey of cast shadow detection algorithms. *Pattern Recognition Letters*, *33*, 752–764.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Hasan, M., Van Esesn, B. C., Awwal, A. A. S., & Asari, V. K. (2018). The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. *arXiv*, .

Álvarez, J. M., & Lopez, A. M. (2011). Road detection based on illuminant invariance. *IEEE Transactions on Intelligent Transportation Systems*, *12*, 184–193.

Åstrand, B., & Baerveldt, A. J. (2002). An agricultural mobile robot with vision-based perception for mechanical weed control. *Autonomous Robots*, *13*, 21–35.

Bac, C., Hemming, J., & Van Henten, E. (2013). Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper. *Computers and Electronics in Agriculture*, *96*, 148–162.

Bac, C. W. (2015). Improving obstacle awareness for robotic harvesting of sweet-pepper, PhD thesis, Wageningen University.

Backman, J., Piirainen, P., & Oksanen, T. (2015). Smooth turning path generation for agricultural vehicles in headlands. *Biosystems Engineering*, *139*, 76–86.

Bai, X., Cao, Z., Wang, Y., Yu, Z., Hu, Z., Zhang, X., & Li, C. (2014). Vegetation segmentation robust to illumination variations based on clustering and morphology modelling. *Biosystems Engineering*, *125*, 80–97.

Bandoh, Y., Qiu, G., Okuda, M., Daly, S., Aach, T., & Au, O. C. (2010). Recent advances in high dynamic range imaging technology. In *17th IEEE International Conference on Image Processing (ICIP 2010)* (pp. 3125–3128). Hong Kong, China: IEEE.

Bargoti, S., & Underwood, J. (2017). Deep fruit detection in orchards. In *IEEE International Conference on Robotics and Automation (ICRA 2017)* (pp. 3626–3633). Marina Bay Sands, Singapore: IEEE.

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

Behmann, J., Mahlein, A.-K., Rumpf, T., Römer, C., & Plümer, L. (2015). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture*, *16*, 239–260.

Bell, C. (2015). A Historical View of Weed Control Technology, University of California Weed Science.

Bhandari, A., Kumar, A., & Singh, G. (2015). Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur's, Otsu and Tsallis functions. *Expert Systems with Applications*, *42*, 1573–1601.

Bloch, C. (2007). The HDRI Handbook: High Dynamic Range imaging for photographers and CG artists, Rocky Nook.

Boonman, J. C. P. (2013). Improving vision based classification between sugar beet and volunteer potato plants using regular plant spacing features, MSc thesis, Biosystems Engineering, Wageningen University.

Bosch, A., Muñoz, X., & Martí, R. (2007). Which is the best way to organize/classify images by content? *Image and Vision Computing*, *25*, 778–791.

Boydston, R. A., & Seymour, M. D. (2002). Volunteer Potato (Solanum tuberosum) Control with Herbicides and Cultivation in Onion (Allium cepa). *Weed technology*, *16*, 620–626.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

Bulanon, D., Burks, T., & Alchanatis, V. (2009). Image fusion of visible and thermal images for fruit detection. *Biosystems Engineering*, *103*, 12–22.

Burgos-Artizzu, X. P., Ribeiro, A., Tellaeche, A., Pajares, G., & Fernández-Quintanilla, C. (2010). Analysis of natural images processing for the extraction of agricultural elements. *Image and Vision Computing*, *28*, 138–149.

Camargo Neto, J., Meyer, G. E., Jones, D. D., & Samal, A. K. (2006). Plant species identification using Elliptic Fourier leaf shape analysis. *Computers and Electronics in Agriculture*, *50*, 121–134.

Canziani, A., Paszke, A., & Culurciello, E. (2016). An Analysis of Deep Neural Network Models for Practical Applications. *arXiv*, .

Chaki, N., Shaikh, S. H., & Saeed, K. (2014). A Comprehensive Survey on Image Binarization Techniques. *Exploring Image Binarization Techniques, Studies in Computational Intelligence*, *560*, 5–16.

Chan, J. C. W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, *112*, 2999–3011.

Chen, L., Wu, C., Fan, W., Sun, J., & Naoi, S. (2014). Adaptive Local Receptive Field Convolutional Neural Networks for Handwritten Chinese Character Recognition. *Pattern recognition communications in computer and information science*, *484*, 455–463.

Cho, S. I., Lee, D. S., & Jeong, J. Y. (2002). Weed-plant Discrimination by Machine Vision and Artificial Neural Network. *Biosystems Engineering*, *83*, 275–280.

Chopin, J., Laga, H., & Miklavcic, S. J. (2016). A Hybrid Approach for Improving Image Segmentation: Application to Phenotyping of Wheat Leaves. *PLOS (Public Library of Science) One*, *11*.

Clearpath (2014). Husky UGV - Outdoor Field Research Robot by Clearpath.

Cooke, L. R., Schepers, H., & Hermansen, A. (2011). Epidemiology and integrated control of potato late blight in Europe. *Potato Research*, *54*, 183–222.

Corke, P. (2011). Light and Color. In *Robotics, Vision and Control* (pp. 223–250). Springer Berlin Heidelberg.

Csurka, G., Dance, C. R., Fan, L., Willamowski, J., & Bray, C. (2004). Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV 2004* (pp. 1–22). Prague, Czech Republic: Springer-Verlag LNCS.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In *31st International Conference on Machine Learning (ICML 2014)* (pp. 647–655). Beijing, China: Springer.

Dreiling, L. (2017). Agronomist offers weed control history, High Plains/Midwest Ag Journal.

Drobchenko, A., Kamarainen, J.-K., Lensu, L., Vartiainen, J., Kälviäinen, H., & Eerola, T. (2011). Thresholding-based detection of fine and sparse details. *Frontiers of Electrical and Electronic Engineering in China*, *6*, 328–338.

Dworak, V., Selbeck, J., Dammer, K., 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*, *13*, 1523–38.

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering*, *151*, 72–80.

Emmi, L., & Gonzalez-de Santos, P. (2017). Mobile robotics in arable lands : current state and future trends. In *European Conference on Mobile Robots (ECMR 2017)* (pp. 256–261). Paris, France: IEEE.

English, A., Ross, P., Ball, D., & Corke, P. (2014). Vision based guidance for robot navigation in agriculture. In *IEEE International Conference on Robotics and Automation (ICRA 2014)* (pp. 1693–1698). Hong Kong, China: IEEE.

Fan, P., Men, A., Chen, M., & Yang, B. (2009). Color-SURF: A surf descriptor with local kernel color histograms. In *IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC 2009)* (pp. 726–730). Beijing, China: IEEE.

Fei-Fei, L., & Perona, P. (2005). A Bayesian Hierarchical Model for Learning Natural Scene Categories. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)* (pp. 524–531). San Diego, USA: IEEE.

Fennimore, S. A., Slaughter, D. C., Siemens, M. C., Leon, R. G., & Saber, M. N. (2016). Technology for Automation of Weed Control in Specialty Crops. *Weed Technology*, *30*, 823–837.

Finlayson, G. D., Hordley, S. D., Lu, C., & Drew, M. S. (2006). On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 59–68.

Florczyk, S. (2005). Robot Vision: Video-based indoor exploration with autonomous and mobile robots, Wiley-VCH.

Frasconi, C., Raffaelli, M., Emmi, L., Fontanelli, M., Martelloni, L., & Peruzzi, A. (2017). An Automatic Machine Able to Perform Variable Rate Application of Flame Weeding: Design and Assembly. *Chemical Engineering Transactions*, *58*, 301–306.

Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, *17*, 2022.

Furgale, P., & Barfoot, T. (2010). Visual path following on a manifold in unstructured three-dimensional terrain. In *IEEE International Conference on Robotics and Automation (ICRA 2010)* (pp. 534–539). Anchorage, USA: IEEE.

Gebhardt, S., & Kühbauch, W. (2007). A new algorithm for automatic Rumex obtusifolius detection in digital images using colour and texture features and the influence of image resolution. *Precision Agriculture*, *8*, 1–13.

Gée, C., Bossu, J., Jones, G., & Truchetet, F. (2008). Crop/weed discrimination in perspective agronomic images. *Computers and Electronics in Agriculture*, *60*, 49–59.

Ghazi, M. M., Yanikoglu, B., & Aptoula, E. (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*, *235*, 228–235.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)* (pp. 580–587). Columbus, USA: IEEE.

Gong, Y., Jia, Y., Leung, T., Toshev, A., & Ioffe, S. (2013). Deep Convolutional Ranking for Multilabel Image Annotation. *CoRR*, (pp. 1–9).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.

Graham, D. J. (2011). Visual perception: Lightness in a high-dynamic-range world. *Current Biology*, *21*, R914–R916.

Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, *127*, 418–424.

Guan, P. P., & Yan, H. (2013). A hierarchical multilevel image thresholding method based on the maximum fuzzy entropy principle. In *Image Processing: Concepts, Methodologies, Tools, and Applications* (pp. 274–302). Hershey: IGI Global.

Guerrero, J., Guijarro, M., Montalvo, M., Romeo, J., Emmi, L., Ribeiro, A., & Pajares, G. (2013). Automatic expert system based on images for accuracy crop row detection in maize fields. *Expert Systems with Applications*, *40*, 656–664.

Guerrero, J., Pajares, G., Montalvo, M., Romeo, J., & Guijarro, M. (2012). Support Vector Machines for crop/weeds identification in maize fields. *Expert Systems with Applications*, *39*, 11149–11155.

Guijarro, M., Riomoros, I., Pajares, G., & Zitinski, P. (2015). Discrete wavelets transform for improving greenness image segmentation in agricultural images. *Computers and Electronics in Agriculture*, *118*, 396–407.

Guo, W., Rage, U. K., & Ninomiya, S. (2013). Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model. *Computers and Electronics in Agriculture*, *96*, 58–66.

Hague, T., Tillett, N. D., & Wheeler, H. (2006). Automated Crop and Weed Monitoring in Widely Spaced Cereals. *Precision Agriculture*, *7*, 21–32.

Hamuda, E., Glavin, M., & Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, *125*, 184–199.

Hasan, M., Kotov, A., Idalski Carcone, A., Dong, M., Naar, S., & Brogan Hartlieb, K. (2016). A study of the effectiveness of machine learning methods for classification of clinical interview fragments into a large number of categories. *Journal of Biomedical Informatics*, *62*, 21–31.

Haug, S., Michaels, A., Biber, P., & Ostermann, J. (2014). Plant classification system for crop/weed discrimination without segmentation. In *IEEE Winter Conference on Applications of Computer Vision (WACV 2014)* (pp. 1142–1149). Steamboat Springs, Colorado, USA: IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016a). Deep residual learning for image recognition. In *29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)* (pp. 770–778). Las Vegas, Nevada, USA: IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016b). Identity Mappings in Deep Residual Networks. In *European Conference on Computer Vision (ECCV 2016)* (pp. 630–645). Amsterdam, The Netherlands: Springer International Publishing.

Hernández-Hernández, J. L., García-Mateos, G., González-Esquiva, J. M., Escarabajal-Henarejos, D., Ruiz-Canales, A., & Molina-Martínez, J. M. (2016). Optimal color space selection method for plant/soil segmentation in agriculture. *Computers and Electronics in Agriculture*, *122*, 124–132.

Hiremath, S., Van Evert, F. K., Braak, C. T., Stein, A., & Van der Heijden, G. (2014). Image-based particle filtering for navigation in a semi-structured agricultural environment. *Biosystems Engineering*, *121*, 85–95.

Hrabar, S., Corke, P., & Bosse, M. (2009). High dynamic range stereo vision for outdoor mobile robotics. In *IEEE International Conference on Robotics and Automation (ICRA 2009)* (pp. 430–435). Kobe, Japan: IEEE.

Hu, F., Xia, G.-S., Hu, J., & Zhang, L. (2015). Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sensing*, *7*, 14680–14707.

Irie, K., Yoshida, T., & Tomono, M. (2012). Outdoor localization using stereo vision under various illumination conditions. *Advanced Robotics*, *26*, 327–348.

IRS (2005). *IRS - Sugar beet growing in the Netherlands*. Technical Report Institute of sugar beet research.

Ishak, A. J., Hussain, A., & Mustafa, M. M. (2009). Weed image classification using Gabor wavelet and gradient field distribution. *Computers and Electronics in Agriculture*, *66*, 53–61.

Janssen, E. (2015). Navigation of a Husky robot with a particle filter, BSc thesis, Biosystems Engineering, Wageningen University.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*.

Jeon, H., Tian, L. F., & Zhu, H. (2011). Robust crop and weed segmentation under uncontrolled outdoor illumination. *Sensors*, *11*, 6270–6283.

Jol, A. (2015). Particle filter based navigation of a mobile robot, MSc thesis, Biosystems Engineering, Wageningen University.

Jozefowicz, R., & Com, I. G. (2015). An Empirical Exploration of Recurrent Network Architectures. *the 32nd International Conference on Machine Learning (ICML 2015)*, *37*, 2342–2350.

Kanellopoulos, I., & Wilkinson, G. G. (1997). Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, *18*, 711–725.

Kapach, K., Barnea, E., Mairon, R., Edan, Y., & Ben-Shahar, O. (2012). Computer Vision for Fruit Harvesting Robots - State of the Art and Challenges Ahead. *International Journal of Computational Vision and Robotics*, *3*, 4–34.

Kapur, J., Sahoo, P., & Wong, A. (1985). A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram. *Computer Vision Graphics and Image Processing*, *29*, 273–285.

Kataoka, T., Kaneko, T., Okamoto, H., & Hata, S. (2003). Crop growth estimation system using machine vision. In *2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)* (pp. 1079–1083). Kobe, Japan: IEEE.

Kato, H., & Harada, T. (2014). Image reconstruction from bag-of-visual-words. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2014)* (pp. 955–962). Columbus, Ohio, USA: IEEE.

Kaur, P. (2013). Performance Evaluation of Various Thresholding Methods using Canny Edge Detector. *International Journal of Computer Applications*, *71*, 26–32.

Kazmi, W., Garcia-Ruiz, F., Nielsen, J., Rasmussen, J., & Andersen, H. J. (2015a). Exploiting affine invariant regions and leaf edge shapes for weed detection. *Computers and Electronics in Agriculture*, *118*, 290–299.

Kazmi, W., Garcia-Ruiz, F. J., Nielsen, J., Rasmussen, J., & Andersen, H. J. (2015b). Detecting Creeping thistle in Sugar beet fields using vegetation indices. *Computers and electronics in agriculture*, *112*, 10–19.

Kelton, J. A., & Price, A. J. (2011). Weed Science and Management. *Soils, Plant Growth and Crop Production*, *3*, 1–10.

Khan, N. Y., McCane, B., & Wyvill, G. (2011). SIFT and SURF performance evaluation against various image deformations on benchmark dataset. In *International Conference on Digital Image Computing: Techniques and Applications (DICTA 2011)* (pp. 501–506). St Lucia, QLD, Australia: IEEE.

Kienhuis, J., & Berge, R. (2003). Verordening hpa bestrijding phytophthora infestans bij aardappelen (Legislation main board of arable products on the control of phytophthora infestans in potato).

King, A. (2017). Technology: The Future of Agriculture. *Nature*, *544*, S21–S23.

Kise, M., Zhang, Q., Rovira Más, F., & Mas, F. R. (2005). A stereovision-based crop row detection method for tractor-automated guidance. *Biosystems Engineering*, *90*, 357–367.

Kittler, J., & Illingworth, J. (1986). Minimum error thresholding. *Pattern recognition*, *19*, 41–47.

Kounalakis, T., Triantafyllidis, G. A., & Nalpantidis, L. (2016). Weed recognition framework for robotic precision farming. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST 2016)* (pp. 466–471). Crete Island, Greece: IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *25th International Conference on Neural Information Processing Systems (NIPS 2012)* (pp. 1097–1105). Lake Tahoe, Nevada, USA: NIPS.

Kunz, C., Weber, J. F., Peteinatos, G. G., Sökefeld, M., & Gerhards, R. (in press). Camera steered mechanical weed control in sugar beet, maize and soybean. *Precision Agriculture*, *XX*.

Kurtulmuş, F., & Kavdir, I. (2014). Detecting corn tassels using computer vision and support vector machines. *Expert Systems with Applications*, *41*, 7390–7397.

Lapray, P., Heyrman, B., Rosse, M., & Ginhac, D. (2012). High Dynamic Range real-time vision system for robotic applications. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)* (pp. 18–22). Vilamoura-Algarve, Portugal: IEEE.

Lati, R. N., Filin, S., & Eizenberg, H. (2013a). Estimating plant growth parameters using an energy minimization-based stereovision model. *Computers and Electronics in Agriculture*, *98*, 260–271.

Lati, R. N., Filin, S., & Eizenberg, H. (2013b). Plant growth parameter estimation from sparse 3D reconstruction based on highly-textured feature points. *Precision Agriculture*, *14*, 586–605.

Law, M. T., Thome, N., & Cord, M. (2014). Fusion in Bag-of-words image representation: Key ideas and further insight. In *Fusion in Comuter Vision: Understanding complex visual content* chapter Ch2. (pp. 29–52). Springer International Publishing.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, *86*, 2278–2324.

Lee, W., Alchanatis, V., Yang, C., Hirafuji, M., Moshou, D., & Li, C. (2010). Sensing technologies for precision specialty crop production. *Computers and Electronics in Agriculture*, *74*, 2–33.

Lee, W. S., Slaughter, D. C., & Giles, D. K. (1999). Robotic weed control system for tomatoes. *Precision Agriculture*, *1*, 95–113.

Leemans, V., & Destain, M.-F. (2006). Line cluster detection using a variant of the Hough transform for culture row localisation. *Image and Vision Computing*, *24*, 541–550.

Li, G. Z. (2011). Machine learning for clinical data processing (Chapter 4.9). In *Machine Learning: Concepts, Methodologies, Tools and Applications* (pp. 876–878). IGI Global.

Li, J., & Tang, L. (in press). Crop recognition under weedy conditions based on 3D imaging for robotic weed control. *Journal of Field Robotics*, *XX*.

Li, S. (2017). Construction and evaluation of a plant detection pipeline for weed control in sugar beets field, MSc thesis, Biosystems Engineering, Wageningen University.

Li, Y., Cao, Z., Lu, H., Xiao, Y., Zhu, Y., & Cremers, A. B. (2016). In-field cotton detection via region-based semantic image segmentation. *Computers and Electronics in Agriculture*, *127*, 475–486.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R news*, *2*, 18–22.

Lin, H. T., Lin, C. J., & Weng, R. C. (2007). A note on Platt's probabilistic outputs for support vector machines. *Machine Learning*, *68*, 267–276.

Liu, H., Lee, S. H., & Saunders, C. (2014). Development of a machine vision system for weed detection during both of off-season and in-season in broadacre no-tillage cropping lands. *American Journal of Agricultural and Biological Science*, *9*, 174–193.

Liu, Y., Liang, G., & Saha, P. K. (2012). A new multi-object image thresholding method based on correlation between object class uncertainty and intensity gradient. *Medical physics*, *39*, 514–32.

Longchamps, L., Panneton, B., Samson, G., Leroux, G. D., & Thériault, R. (2009). Discrimination of corn, grasses and dicot weeds by their UV-induced fluorescence spectral signature. *Precision Agriculture*, *11*, 181–197.

Lottes, P., Hoeferlin, M., Sander, S., Müter, M., Schulze Lammers, P., & Stachniss, C. (2016). An Effective Classification System for Separating Sugar Beets and Weeds for Precision Farming Applications. In *IEEE International Conference on Robotics and Automation (ICRA 2016)* (pp. 5157–5163). Stockholm, Sweden: IEEE.

Lottes, P., Hörferlin, M., Sander, S., & Stachniss, C. (2017). Effective Vision-based Classification for Separating Sugar Beets and Weeds for Precision Farming. *Journal of Field Robotics*, *34*, 1160–1178.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, *60*, 91–110.

Lu, Y. H., Kadin, A. M., Berg, A. C., Conte, T. M., Debenedictis, E. P., Garg, R., Gingade, G., Hoang, B., Huang, Y., Li, B., Liu, J., Liu, W., Mao, H., Peng, J., Tang, T., Track, E. K., Wang, J., Wang, T., Wang, Y., & Yao, J. (2016). Rebooting Computing and Low-Power Image Recognition Challenge. In *2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD 2015)* (pp. 927–932). Austin, USA: IEEE.

MacEwan, C., Stevens, M., Bowen, S., & Broom, C. (2017). *Sugar beet reference book by British Beet Research Organization*. Technical Report British Beet Research Organization Norwich, UK.

Mancini, A., Dyson, J., Frontoni, E., & Zingaretti, P. (2017). Soil/crop segmentation from remotely sensed data acquired by Unmanned Aerial System. In *2017 International Conference on Unmanned Aircraft Systems (ICUAS 2017)* (pp. 1410–1417). Miami, USA: IEEE.

Mann, S., Lo, R. C. H., Ovtcharov, K., Gu, S., Dai, D., Ngan, C., & Ai, T. (2012). Realtime HDR (High Dynamic Range) video for eyetap wearable computers, FPGA-based seeing aids, and glasseyes (EyeTaps). In *25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2012)*. Montreal, Canada: IEEE.

Marchant, J. a., & Onyango, C. M. (2000). Shadow-invariant classification for scenes illuminated by daylight. *Journal of the Optical Society of America*, *17*, 1952–1961.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluation segmentation algorithms and measuring ecological statistics. In *8th International Conference on Computer Vision (ICCV 2001)* (pp. 416–423). Vancouver, Canada: IEEE.

Mery, D., & Pedreschi, F. (2005). Segmentation of colour food images using a robust algorithm. *Journal of Food Engineering*, *66*, 353–360.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, *8*, 283–298.

Meyer, G. E., & Camargo Neto, J. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, *63*, 282–293.

Meyer, G. E., Camargo Neto, J., Jones, D. D., & Hindman, T. W. (2004). Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. *Computers and Electronics in Agriculture*, *42*, 161–180.

Midtiby, H. S., Mathiassen, S. K., Andersson, K. J., & Jørgensen, R. N. (2011). Performance evaluation of a crop/weed discriminating microsprayer. *Computers and Electronics in Agriculture*, *77*, 35–40.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, *7*.

Montalvo, M., Guerrero, J. M., Romeo, J., Emmi, L., Guijarro, M., & Pajares, G. (2013). Automatic expert system for weeds/crops identification in images from maize fields. *Expert Systems with Applications*, *40*, 75–82.

Moushib, L. I., Witzell, J., Lenman, M., Liljeroth, E., & Andreasson, E. (2013). Sugar beet extract induces defence against Phytophthora infestans in potato plants. *European Journal of Plant Pathology*, *136*, 261–271.

Muñoz-Huerta, R. F., Guevara-Gonzalez, R. G., Contreras-Medina, L. M., Torres-Pacheco, I., Prado-Olivarez, J., & Ocampo-Velazquez, R. V. (2013). A review of methods for sensing the nitrogen status in plants: advantages, disadvantages and recent advances. *Sensors*, *13*, 10823–10843.

Mythili, M. C., & Kavita, D. V. (2011). Efficient Technique for Color Image Noise Reduction. *The Research Bulletin of Jordan ACM*, *2*, 41–44.

Nacereddine, N., Hamami, L., Tridi, M., & Oucief, N. (2005). Non-parametric histogram-based thresholding methods for weld defect detection in radiography. *World Academy of Science, Engineering and Technology*, *1*, 1237–1241.

Navarro, P., Iborra, A., Fernández, C., Sánchez, P., & Suardíaz, J. (2010). A sensor system for detection of hull surface defects. *Sensors*, *10*, 7067–7081.

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *22nd international Conference on Machine learning (ICML 2005)* 1999 (pp. 625–632). Bonn, Germany: ACM.

Nieuwenhuizen, A. T. (2009). Automated detection and control of volunteer potato plants, PhD thesis, Wageningen University.

Nieuwenhuizen, A. T., Hofstee, J. W., & Van Henten, E. J. (2010). Performance evaluation of an automated detection and control system for volunteer potatoes in sugar beet fields. *Biosystems Engineering*, *107*, 46–53.

Nieuwenhuizen, A. T., Tang, L., Hofstee, J. W., Müller, J., & Van Henten, E. J. (2007). Colour based detection of volunteer potatoes as weeds in sugar beet fields using machine vision. *Precision Agriculture*, *8*, 267–278.

Ohta, J. (2007). *Smart CMOS Image Sensors and Applications*. CRC Press, Taylor & Francis.

O'Keeffe, M. G. (1980). The control of Agropyron repens and broad-leaved weeds pre-harvest of wheat and barley with the isopropylamine salt of glyphosate. In *Proceedings 1980 British Crop Protection Conference - Weeds* (pp. 53–60). Brussels, Belgium: British Crop Protection Council.

Oliva, D., Cuevas, E., Pajares, G., Zaldivar, D., & Osuna, V. (2014). A Multilevel thresholding algorithm using electromagnetism optimization. *Neurocomputing*, *139*, 357–381.

Oquab, M., Bottou, L. B., Laptev, I. L., & Sivic, J. (2014). Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks. In *IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR 2014) (pp. 1717–1724). Columbus, OH, USA: IEEE.

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

Panchal, P. M., Panchal, S. R., & Shah, S. K. (2013). A Comparison of SIFT and SURF. *International Journal of Innovative Research in Computer and Communication Engineering*, *1*, 323–327.

Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., & Wilhelm, V. L. (2017). Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. *European Journal of Remote Sensing*, *50*, 564–576.

Papadomanolaki, Vakalopoulou, M., Zagoruyko, S., & Karantzalos, K. (2016). Benchmarking deep learning frameworks for the classification of high resolution satellite multispectral data. In *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* (pp. 83–88). Prague, Czech Republic: ISPRS Congress.

Pascale, D. (2003). *A Review of RGB Color Spaces, from xyY to R'G'B'*. Technical Report The BabelColor Company Technical report, The BabelColor Company, Montreal, Canada.

Pedersen, S. M., Fountas, S., Have, H., & Blackmore, B. S. (2006). Agricultural robots-system analysis and economic feasibility. *Precision Agriculture*, *7*, 295–308.

Pérez, A. J., López, F., Benlloch, J. V., & Christensen, S. (2000). Colour and shape analysis techniques for weed detection in cereal fields. *Computers and Electronics in Agriculture*, *25*, 197–212.

Pérez-Cruz, F., Martinez-Olmos, P., & Murillo-Fuentes, J. J. (2007). Accurate posterior probability estimates for channel equalization using gaussian processes for classification. In *IEEE 8th Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2007)* (pp. 1–5). Helsinki, Finland: IEEE.

Persson, M., & Åstrand, B. (2008). Classification of crops and weeds extracted by active shape models. *Biosystems Engineering*, *100*, 484–497.

Piron, A., Heijden, F., & Destain, M. F. (2010). Weed detection in 3D images. *Precision Agriculture*, *12*, 607–622.

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, *10*, 61–74.

Polder, G., Van der Heijden, G. W. A. M., Van Doorn, J., & Baltissen, T. A. (2014). Automatic detection of tulip breaking virus (TBV) in tulip fields using machine vision. *Biosystems Engineering*, *117*, 35–42.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, *6*, 21–45.

Prasad, M. S., Krishna, V. R., & Reddy, L. S. S. (2011). Investigations on entropy based threshold methods. *Asian Journal of Computer Science and Information Technology*, *5*, 132 – 137.

Prati, A., Mikic, I., Trivedi, M. M., & Cucchiara, R. (2003). Detecting moving shadows: algorithms and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*, 918–923.

Radonjić, A., Allred, S. R., Gilchrist, A. L., & Brainard, D. H. (2011). The dynamic range of human lightness perception. *Current Biology*, *21*, 1931–1936.

Rahman, A. (1980). Biology and control of volunteer potatoes - a review. *New Zealand Journal of Experimental Agriculture*, *8*, 313–319.

Ramesh, N. (1995). Thresholding based on histogram approximation. *IEE Proceedings - Vision, Image, and Signal Processing*, *142*, 271–279.

Rassem, T. H., & Khoo, B. E. (2011). Object class recognition using combination of color SIFT descriptors. In *2011 IEEE International Conference on Imaging Systems and Techniques (IST 2011)* (pp. 290–295). Batu Ferringhi, Malaysia: IEEE.

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. In *Workshop on IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)* (pp. 512–519). Columbus, OH, USA: IEEE.

Reinhard, E., Ward, G., Pattanaik, S., Debevec, P., Heidrich, W., & Myszkowski, K. (2010). *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann.

Ridler, T. W., & Calvard, S. (1978). Picture Thresholding Using an Iterative Selection Method. *IEEE Transactions on Systems, Man, and Cybernetics*, *8*, 630–632.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93–104.

Romeo, J., Pajares, G., Montalvo, M., Guerrero, J., Guijarro, M., & de la Cruz, J. (2013). A new Expert System for greenness identification in agricultural images. *Expert Systems with Applications*, *40*, 2275–2286.

Romeo, J., Pajares, G., Montalvo, M., Guerrero, J. M., Guijarro, M., & Ribeiro, A. (2012). Crop row detection in maize fields inspired on the human visual perception. *The Scientific World Journal*, *2012*, 1–10.

Rosin, P. L. (2001). Unimodal thresholding. *Pattern Recognition*, *34*, 2083–2096.

Rosin, P. L., & Ioannidis, E. (2003). Evaluation of global image thresholding for change detection. *Pattern Recognition Letters*, *24*, 2345–2356.

Rumpf, T., Römer, C., Weis, M., Sökefeld, M., Gerhards, R., & Plümer, L. (2012). Sequential support vector machine classification for small-grain weed species discrimination with special regard to Cirsium arvense and Galium aparine. *Computers and Electronics in Agriculture*, *80*, 89–96.

Russ, J. C., & Neal, F. B. (2015). Chapter 7. Segmentation and thresholding. In *The Image Processing Handbook* chapter 7. (pp. 381–434). Boca Raton, FL: CRC Press, Taylor & Francis Group. (7th ed.).

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors*, *16*, 1222.

Sainath, T. N., Mohamed, A. R., Kingsbury, B., & Ramabhadran, B. (2013). Deep convolutional neural networks for LVCSR. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)* (pp. 8614–8618). Vancouver, Canada: IEEE.

Salton, G., & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. New York, USA: McGraw-Hill, Inc.

Sanin, A., Sanderson, C., & Lovell, B. C. (2012). Shadow detection: A survey and comparative evaluation of recent methods. *Pattern Recognition*, *45*, 1684–1695.

Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, *33*, 225–236.

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Schwing, A. G., & Urtasun, R. (2015). Fully Connected Deep Structured Networks. *arXiv preprint*, (pp. 1–10).

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2014). OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. In *International Conference on Learning Representations (ICLR 2014)*. Banff, Canada: ICLR.

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, *13*, 146–165.

Shafait, F., Keysers, D., & Breuel, T. (2008). Efficient implementation of local adaptive thresholding techniques using integral images. In *SPIE 6815, Document Recognition and Retrieval XV*. San Jose, California, USA: SPIE.

Shaikh, S. H., Maiti, A., & Chaki, N. (2011). Image binarization using iterative partitioning: A global thresholding approach. In *International Conference on Recent Trends in Information Technology (ICRTIT 2011)* (pp. 281–286). Chennai, India: IEEE.

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE transactions on medical imaging*, *35*, 1285–1298.

Shrestha, D. S., & Steward, B. L. (2005). Shape and size analysis of corn plant canopies for plant population and spacing sensing. *Applied Engineering in Agriculture*, *21*, 295–303.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations (ICRL)* (pp. 1–14).

Slaughter, D. C., Giles, D. K., & Downey, D. (2008). Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture*, *61*, 63–78.

Smith, C. W., & Frederiksen, R. A. (2000). *Sorghum : origin, history, technology, and production*. Wiley Series in Crop Science.

Søgaard, H., & Olsen, H. (2003). Determination of crop rows by image analysis without segmentation. *Computers and Electronics in Agriculture*, *38*, 141–158.

Sojodishijani, O., Ramli, A. R. R., Rostami, V., Samsudin, K., & Saripan, M. I. I. (2010). Just-in-time outdoor color discrimination using adaptive similarity-based classifier. *IEICE Electronics Express*, *7*, 339–345.

Solomon, C., & Breckon, T. (2011). *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Chichester, UK: John Wiley & Sons, Ltd.

Sørensen, C. G., Jørgensen, R. N., Maagaard, J., Bertelsen, K. K., Dalgaard, L., & Nørremark, M. (2010). Conceptual and user-centric design guidelines for a plant nursing robot. *Biosystems Engineering*, *105*, 119–129.

Stemp, G. (2005). Agriculture: Green Farming Equipment. *Environmental Health Perspectives*, *113*, A590.

Steward, B. L., Tian, L. F., Nettleton, D. S., & Tang, L. (2004). Reduced-dimension clustering for vegetation segmentation. *Transactions of the ASAE*, *47*, 609–616.

Su, C., & Amer, A. (2006). A real-time adaptive thresholding for video change detection. In *IEEE International Conference on Image Processing (ICIP 2006)* (pp. 157–160). Atlanta, USA: IEEE.

Suh, H. K., Hofstee, J. W., IJselmuiden, J., & Van Henten, E. J. (2016). Discrimination between Volunteer Potato and Sugar Beet with a Bag-of-Visual-Words Model. In *International Conference on Agricultural Engineering (CIGR-AgEng)*. Aarhus, Denmark: EurAgEng.

Suh, H. K., Hofstee, J. W., IJsselmuiden, J., & Van Henten, E. J. (2018a). Sugar beet and volunteer potato classification using Bag-of-Visual-Words model, Scale-Invariant Feature Transform, or Speeded Up Robust Feature descriptors and crop row information. *Biosystems Engineering*, *166*, 210–226.

Suh, H. K., Hofstee, J. W., & Van Henten, E. J. (2018b). Improved vegetation segmentation with ground shadow removal using an HDR camera. *Precision agriculture*, *19*, 218–237.

Sukkarieh, S. (2016). We are bipolar about robotics but growers are going for it. *FOCUS August - Australian Academy of Technology and Engineering*, (pp. 14–16).

Sun, Y., Liu, Y., Wang, G., & Zhang, H. (2017). Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience*, *2017*, 1–6.

Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)* (pp. 1891–1898). Columbus, Ohio, USA: IEEE.

Swain, K. C., Nørremark, M., Jørgensen, R. N., Midtiby, H. S., & Green, O. (2011). Weed identification using an automated active shape matching (AASM) technique. *Biosystems Engineering*, *110*, 450–457.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In *the 31st AAAI Conference on Artificial Intelligence (AAAI-17)* (pp. 4278–4284). San Francisco, USA: AAAI-17.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*. Boston, USA: IEEE.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)* (pp. 2818–2826). IEEE.

Teixidó, M., Font, D., Pallejà, T., Tresanchez, M., Nogués, M., & Palacín, J. (2012). Definition of linear color models in the RGB vector color space to detect red peaches in orchard images taken under natural illumination. *Sensors*, *12*, 7701–7718.

Tellaeche, A., Burgos-Artizzu, X. P., Pajares, G., & Ribeiro, A. (2008). A vision-based method for weeds identification through the Bayesian decision theory. *Pattern Recognition*, *41*, 521–530.

Tibbetts, J. H. (2018). The Frontiers of Artificial Intelligence: Deep learning brings speed, accuracy to the life sciences. *BioScience*, *68*, 5–10.

Timmons, F. L. (2005). A History of Weed Control in the United States and Canada. *Weed Science*, *53*, 748–761.

Tsai, C. F. (2012). Bag-of-Words Representation in Image Annotation: A Review. *ISRN Artificial Intelligence*, *2012*, 1–19.

Tychsen-Smith, L., & Petersson, L. (2017). DeNet: Scalable Real-time Object Detection with Directed Sparse Sampling. In *IEEE International Conference on Computer Vision (ICCV 2017)* (pp. 428–436). Venice, Italy: IEEE.

Uijlings, J., Smeulders, A., & Scha, R. (2009). What is the Spatial Extent of an Object? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)* (pp. 770–777). Miami, USA: IEEE.

Uijlings, J. R. R., Van De Sande, K. E. A., Gevers, T., & Smeulders, A. W. M. (2013). Selective Search for Object Recognition. *International Journal of Computer Vision*, *104*, 154–171.

Van De Sande, K. E. A., Gevers, T., & Snoek, C. G. M. (2008). A comparison of color features for visual concept classification. In *2008 International Conference on Content-based Image and Video Retrieval (CIVR 2008)* (pp. 141–150). Niagara Falls, Canada: ACM.

Van Evert, F., Polder, G., Van Der Heijden, G., Kempenaar, C., & Lotz, L. (2009). Real-time vision-based detection of Rumex obtusifolius in grassland. *Weed Research*, *49*, 164–174.

Van Henten, E. J., Van Asselt, C. J., Bakker, T., Blaauw, S. K., Govers, M. H. A. M., Hofstee, J. W., Jansen, R. M. C., Nieuwenhuizen, A. T., Speetjens, S. L., Stigter, J. D., Van Straten, G., & Van Willigenburg, L. G. (2009). WURking: a small sized autonomous robot for the Farm of the Future. In V. E. J. Henten, D. Goense, & C. Lokhorst (Eds.), *Precision Agriculture 2009 - the 7th European conference on precision agriculture* (pp. 833–840). Wageningen, The Netherlands: Wageningen Academic.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory* volume 8. Springer-Verlag New York, Inc.

Vedaldi, A., & Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms. In *18th ACM International Conference on Multimedia* (pp. 1469–1472). Firenze, Italy: ACM.

Vedaldi, A., & Lenc, K. (2015). MatConvNet: Convolutional neural networks for MATLAB. In *23rd ACM International Conference on Multimedia* (pp. 689–692). Brisbane, Australia: ACM.

Vollebregt, M. (2013). Texture based discrimination between sugar beet and volunteer potato, MSc thesis, Biosystems Engineering, Wageningen University.

Wan, L., Zeiler, M., Zhang, S., LeCun, Y., & Fergus, R. (2013). Regularization of neural networks using dropconnect. In *30th International Conference on Machine Learning (ICML 2013)* (pp. 1058–1066). Atlanta, USA: PMLR.

Wang, J., Luo, C., Huang, H., Zhao, H., & Wang, S. (2017). Transferring Pre-Trained Deep CNNs for Remote Scene Classification with General Features Learned from Linear PCA Network. *Remote Sensing*, *9*, 225.

Wang, Q., Wang, H., Xie, L., & Zhang, Q. (2012). Outdoor color rating of sweet cherries using computer vision. *Computers and Electronics in Agriculture*, *87*, 113–120.

Wang, S., Chung, F.-l., & Xiong, F. (2008). A novel image thresholding method based on Parzen window estimate. *Pattern Recognition*, *41*, 117–129.

Wang, Z., Wang, K., Yang, F., Pan, S., & Han, Y. (2018). Image segmentation of overlapping leaves based on Chan-Vese model and Sobel operator. *Information Processing in Agriculture*, *5*, 1–10.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). *A survey of transfer learning* volume 3. Springer International Publishing.

Wilf, P., Zhang, S., Chikkerur, S., Little, S. A., Wing, S. L., & Serre, T. (2016). Computer vision cracks the leaf code. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 3305–3310.

Woebbecke, D., Meyer, G., Von Bargen, K., & Mortensen, D. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, *38*, 259–269.

Wu, J., Cui, Z., Sheng, V. S., Zhao, P., Su, D., & Gong, S. (2013). A Comparative Study of SIFT and its Variants. *Measurement Science Review*, *13*, 122–131.

Xia, C., Lee, J.-M., Li, Y., Song, Y.-H., Chung, B.-K., & Chon, T.-S. (2013). Plant leaf detection using modified active shape models. *Biosystems Engineering*, *116*, 23–35.

Xie, M., Jean, N., Burke, M., Lobell, D., & Ermon, S. (2016). Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. In *30th AAAI Conference on Artificial Intelligence (AAAI 2016)* (pp. 3929–3935). Phoenix, USA: AAAI Press.

Xiong, Y., Ge, Y., Liang, Y., & Blackmore, S. (2017). Development of a prototype robot and fast path-planning algorithm for static laser weeding. *Computers and Electronics in Agriculture*, *142*, 494–503.

Yamamoto, K., Guo, W., Yoshioka, Y., & Ninomiya, S. (2014). On Plant Detection of Intact Tomato Fruits Using Image Analysis and Machine Learning Methods. *Sensors*, *14*, 12191–12206.

Yang, J., Jiang, Y. G., Hauptmann, A. G., & Ngo, C. W. (2007). Evaluating bag-of-visual-words representations in scene classification. In *International Workshop on Multimedia Information Retrieval (MIR 2007)* (pp. 197–206). Augsburg, Germany: ACM.

Yang, W., Wang, S., Zhao, X., Zhang, J., & Feng, J. (2015). Greenness identification based on HSV decision tree. *Information Processing in Agriculture*, *2*, 149–160.

Ye, M., Cao, Z., Yu, Z., & Bai, X. (2015). Crop feature extraction from images with probabilistic superpixel Markov random field. *Computers and Electronics in Agriculture*, *114*, 247–260.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems 27 (Proceedings of NIPS)*, *27*, 1–9.

Young, S. L., & Pierce, F. J. (2014). *Automation: The future of weed control in cropping systems*. Dordrecht: Springer Netherlands.

Yu, Z., Cao, Z., Wu, X., Bai, X., Qin, Y., Zhuo, W., Xiao, Y., Zhang, X., & Xue, H. (2015). Automatic image-based detection technology for two critical growth stages of maize: Emergence and three-leaf stage. *Agricultural and Forest Meteorology*, *174-175*, 65–84.

Zagoris, K., Pratikakis, I., Antonacopoulos, A., Gatos, B., & Papamarkos, N. (2014). Distinction between handwritten and machine-printed text based on the bag of visual words model. *Pattern Recognition*, *47*, 1051–1062.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 818–833). volume 8689 LNCS.

Zhang, Z., Kodagoda, S., Ruiz, D., Katupitiya, J., & Dissanayake, G. (2008). Classification of Bidens in wheat farms. In *15th International Conference on Mechatronics and Machine Vision in Practice (M2VIP 2008)* (pp. 505–510). Auckland, New Zealand: IEEE.

Zhang, Z., Kodagoda, S., Ruiz, D., Katupitiya, J., & Dissanayake, G. (2010). Classification of Bidens in wheat farms. *International Journal of Computer Applications in Technology*, *39*, 123–129.

Zheng, L., Zhang, J., & Wang, Q. (2009). Mean-shift-based color segmentation of images containing green vegetation. *Computers and Electronics in Agriculture*, *65*, 93–98.

Zheng, X., Ye, H., & Tang, Y. (2017). Image Bi-Level Thresholding Based on Gray Level-Local Variance Histogram. *Entropy*, *19*, 191.

Zhou, L., Zhou, Z., & Hu, D. (2013). Scene classification using a multi-resolution bag-of-features model. *Pattern Recognition*, *46*, 424–433.

Zimdahl, R. L. (2013). *Fundamentals of weed science*. Elsevier Academic Press.

Summary

Volunteer potato is a major problem in sugar beet production in the Netherlands, and adequate control of volunteer potato is critical. This is stressed by a statutory obligation in the Netherlands under which farmers have to remove volunteer potato plants from their fields before the 1st of July in the growing season every year, to a maximum level of two remaining plants per square meter.

In 2011, the EU SmartBot project, a cross-border collaboration project which involved 24 different partners from Germany and the Netherlands, was initiated to develop a robotic system for several applications including for agricultural use. In AgroBot, part of the SmartBot project, a small-sized and vision-based autonomous weed control system was to be developed for effective control of volunteer potato plants in a sugar beet field. As a robotic platform, the Clearpath Husky A200 UGV (Unmanned Ground Vehicle) was to be used in this project. Due to the reduced carrying capacity of the robotic platform (Husky), additional infrastructure like a hood was not a viable option. Moreover, artificial lighting was not considered feasible either because the mobile platform was battery operated. Thus, the system should be able to perform robustly in scenes that are fully exposed to ambient lighting conditions.

Within the EU SmartBot project, the primary objective of this research was identified as:

to develop a computer vision procedure
that detects volunteer potato plants
under ambient light conditions in a sugar beet field

For a complete weed control pipeline, including weed detection and weed removal, the following requirements were set. The automatic weeding system should:

- effectively control more than 95% of the volunteer potato;

- ensure less than 5% of undesired control of sugar beet plants;
- ensure a classification time of less than 1 second per field image for real-time operation in the field.

It was indicated that due to the potential non-perfect performance of actual weed removal, classification accuracy should be considerably higher than 95%.

The steps required to fulfil the above-mentioned objective form the main line of this thesis including vegetation segmentation (Chapter 2 and 3) and sugar beet/volunteer potato classification (Chapter 4 and 5).

Chapter 2 addressed the research question: *"Does a ground shadow detection and removal enhance the performance of vegetation segmentation under natural illumination conditions in the field?"*

In Chapter 2, an algorithm was described and evaluated for ground shadow detection and removal based on colour space conversion and a multilevel threshold. The advantage of using the proposed algorithm was assessed for vegetation segmentation with field images that were acquired by a High Dynamic Range (HDR) camera under natural illumination. Compared with no shadow removal, applying shadow removal enhanced the performance of vegetation segmentation under natural illumination conditions in the field with an average of 20%, 4.4% and 13.5% in precision, specificity and modified accuracy, respectively, and did not reduce segmentation performance when shadows were not present. The average processing time was 0.46 s, which is feasible when real-time application in the field is considered.

Chapter 3 addressed the research question: *"Do different combinations of colour index and threshold technique result in different segmentation performance when evaluated on field images? Given the varying conditions in the field, is it better to use one specific combination at all times or the combination should be adapted to the field conditions at hand for best segmentation performance?"*

In Chapter 3, the performance of 40 combinations of eight colour indices and five threshold techniques for vegetation segmentation were evaluated. A clear difference in performance, represented in terms of MA (Modified Accuracy), was observed among various combinations under the given conditions of this research. CIVE+Kapur showed the best performance, while VEG+Kapur showed the worst on the dataset.

When adapting the combination to the given conditions yielded a slightly higher performance than when using a single combination for all (in this case CIVE+Kapur). Consistent results were obtained when validated on a different independent image dataset. The expected advantage of adapting the combination to the field condition is not large because it seems that for practical use, the slight improvement when adapting the combination to the field conditions does not outweigh the investment in sensor technology and software needed to accurately determine the different conditions in the field.

Chapter 4 and **5** focussed on classification and addressed the following research questions: *"Does an algorithm using a Bag-of-Visual-Words (BoVW) model and SIFT or SURF descriptors meet the requirements set for the classification of volunteer potato and sugar beet under natural and varying daylight conditions? If the BoVW model does not meet the requirements, does a deep learning approach, particularly transfer learning based on Convolutional Neural Network (ConvNet, or CNN) provide an effective and better performance to meet the requirements with limited amount of dataset? Are the processing times (or calculation times) fast enough for real-time application?"*

For the classification of sugar beet and volunteer potato under ambient varying daylight conditions, Chapter 4 proposed a classification algorithm using a Bag-of-Visual-Words (BoVW) model based on SIFT or SURF features as well as crop row information in the form of the Out-of-Row Regional Index (ORRI). The highest classification accuracy of 96.5% with false-negative of 0% obtained using SIFT and ORRI with SVM is considerably better than previously reported approaches for weed classification; however, the false-positive rate of 7% deviates from the requirements since misclassification should be less than 5%. The average classification time of 0.10 - 0.11 s met the real-time requirements. Adding location information (ORRI) improved overall classification accuracy significantly. The proposed approach proved its potential under varying natural light conditions.

Since the required classification accuracy was not obtained in Chapter 4, further research was carried out for the classification of sugar beet and volunteer potato under ambient varying daylight conditions. Chapter 5 evaluated a transfer learning procedure with three different implementations of AlexNet (Part I), and then assessed the performance amongst different ConvNet architectures (Part II): AlexNet, VGG-19, GoogLeNet, ResNet-50, ResNet-101 and Inception-v3. In Part I, the highest classifica-

tion accuracy (98.0%) was obtained with AlexNet in Scenario 2. In scenario 1 and 3, the highest classification accuracy of 97.0% and 97.3% were obtained, respectively. In Part II, the highest classification accuracy of 98.7% was obtained. This result, to the best of our knowledge, was considerably better than any other approaches mentioned in the literature for crop and weed classification. Transfer learning provided very promising performance for the classification of sugar beet and volunteer potato images under ambient varying light conditions. A deep learning approach based on ConvNet provided better performance than the one in Chapter 4, and satisfied the requirements. All procedures were feasible for real-time field applications (the classification time $< 0.1\,\mathrm{s}$).

The full pipeline for weed detection consists of three steps: 1) vegetation segmentation, i.e. separating pixels in an image into plant pixels and non-plant pixels, 2) individual object identification, i.e. identification of individual plants (objects) in the set of plant pixels obtained after segmentation, and 3) classification of the plants into two classes, sugar beet (crop) and volunteer potato (weed).

In this thesis, steps 1 and 3, i.e. image segmentation and classification of sugar beet/volunteer potato were successfully addressed. Step 2, the identification of individual plants in the images was not addressed. Despite this limitation, it can be concluded that significant progress has been made in this area of study, given the fact that reported algorithms were developed using images captured in full daylight with significant variations in light colour and intensity; a distinct challenge that so far has been circumvented by using hoods and artificial lighting. Yet, the question remains unanswered whether a full pipeline, including all three steps, would be able to meet the requirements identified at the onset of the research.

With current hardware and suitable implementation of software, it seems that the requirement of 1 s per image for real-time operation of a weed control system can be attained. The highest classification accuracy of 98.7% obtained in Chapter 5 is supportive in meeting the required 95% control of volunteer potatoes, but when the ConvNet classification would be implemented in a full pipeline also containing vegetation segmentation and individual plant identification, a degraded performance can be expected. Given the fact that the images in this research were obtained under varying daylight conditions, the results showed potential of the proposed approach and compared favourably with classification results in the range of 85-90% that were

obtained in various previous researches using hoods and artificial lighting. Therefore, it is safe to say that this research has laid the foundation for a small-sized robotic platform to come into action for weed control in the field.

Acknowledgements

I consider all the moments in my PhD are valuable since I have gained more knowledge in this field of the study as well as since I have realised what I can do more to move this field forward. However, the most significant values I have gained during my PhD are the following two things: firstly, to know about genuine humility as I realise how small and vulnerable I am before Nature and secondly, learned how to be happy no matter what the circumstances are.

I look back on my past at this particular moment, the moment where I put the period in my thesis and the moment I often consider myself did all the works. However, soon I become to realise how foolish I am to believe that "I did it all." I admit that none of this work would have been remotely possible without all the advice from the people around. The insights from the books and papers, even from the media, have shaped my way in one way or the other.

A very special gratitude goes out to all down to EU Interreg IVa SmartBot Project, European Fund for the Regional Development of the European Union and Product Board for Arable Farming, for providing the funding for the work.

I am thankful for the supervision I received in FTE. Jan Willem and Eldert are the biggest contributors, as supervisors, for this work. Thanks to you two, now I know how to utilise my weapons in this battle of the research field to stand up as an independent researcher. Thanks to you two, I know how to position myself among my comrades in this battlefield of the research. Those who want to sharpen his or her arms for the battle should first seek Jan Willem and Eldert.

All staff in FTE are the great supporters: Peter, Bert, Joris, Sam, Arni, Simon, Frank, and Ehud. My special thanks go to Sam who managed all the technical issues with a professional attitude. FTE former and current PhDs deserve my gratitude

and thanks as well: Wouter, Bastiaan, Dennis, Peter, Liansun, Ellen, Inge, Monique, Francisco, Song, Dejan, David, and Dorji.

Special gratitude goes out to all my colleagues in Glastuinbouw and AFR (Agro Food Robotics) Team. With a special mention to Jochen, Silke, Gerrit, Jos, Toon, Pieter, Koen, Ruud, Angelo, and Bart. I am privileged to have the opportunity to work with you all.

I would like to thank the Wageningen Korean Community and the members of the Eindhoven Korean Church. It was you all who encouraged me the most during the moment of distress and trouble. Your support during the first year in NL was also invaluable to my family.

Last, but not least, I am grateful to my family for the patience, love, and support of this work as well as for the tolerance of my absence.

About the author

Hyun Kwon Suh was born on 16 October 1979 in Gumi, Republic of Korea (South Korea). After graduating from high school in 1998 (from Geochang High School), he started a bachelor study in Computer Science and Electronics Engineering at Handong Global University in Pohang. During his sophomore in this university, he served in the Korean Army for two years as a secretary to the Battalion Commander. Soon after discharge from the Army, he started to work at Wisebooktopia Co. in Seoul as an E-book production publisher. In 2004, he was selected to an international exchange program, which was funded by Korean Ministry of Information and Communication, and thus went to the United States as an exchange student. At Iowa State University (ISU) in the U.S., he was fascinated by agriculture and its prospects for automation, he changed his major to Agricultural Systems Technology in the Department of Agricultural Engineering. In 2007, he participated in 2007 ASABE (American Society of Agricultural and Biological Engineers) Robotics Competition at the ASABE Annual International Meeting in Minneapolis, where his team (ISU CyBot Team) won the second place in competition challenge & demonstration. After that, he went to Yanji University, China, where he worked as a network administrator as well as a part-time lecturer. During this one-year stay and work in China, he learned to speak Chinese as well as to have better understanding of the dynamics between countries around northeast China. In 2008, he started his graduate study in Biosystems Engineering at Seoul National University (SNU). During his graduate study, he involved several research projects particularly related to the agricultural automation and machine vision. After completing the

combined MSc & PhD courses, he worked as a part-time lecturer in Biosystems Engineering at SNU. In 2012, he continued his PhD research, working in Wageningen University, the Netherlands, focusing on an automated weed detection and control which was funded by EU SmartBot project. In Winter 2016, he finished his PhD work which produced this thesis. Since Spring 2017, he works in a post-doctoral position in Greenhouse Horticulture (business unit) in Wageningen University & Research (WUR).

List of publications

Peer-reviewed publications in journals

Suh, H. K., J.W. Hofstee, & E.J. van Henten (2018). Improved vegetation segmentation with ground shadow removal using an HDR camera. *Precision Agriculture*, 19(2), 218-237.

Suh, H. K., J.W. Hofstee, Joris IJsselmuiden, & E.J. van Henten (2018). Sugar beet and volunteer potato classification using Bag-of-Visual-Words model, Scale-Invariant Feature Transformation, or Speeded Up Robust Feature descriptors and crop row information. *Biosystems Engineering*, 166, 210-226.

Suh, H. K., J.W. Hofstee, Joris IJsselmuiden, & E.J. van Henten (2018). Transfer learning for the classification of sugar beet and volunteer potato under field conditions. *Biosystems Engineering*, 174, 50-65.

Blok, P.M., **H.K. Suh**, K. van Boheemen, H.J. Kim, & G.H. Kim (2018). Autonomous in-row navigation of an orchard robot with a 2D LIDAR scanner and particle filter with a laser-beam model. *Journal of Institute of Control, Robotics and Systems*, 24(8), 726-735.

Cho, S. I., Y.R. Kim, J.W. Lee, D.S. So, **H.K. Suh**, & Y.J. Cho (2010). Review on the application of nanotechnology in food processing and packaging. *Food Engineering Progress*, 14(4), 283-291.

Manuscripts submitted for publication or in preparation

Suh, H. K., J.W. Hofstee, & E.J. van Henten (under review). Investigation on combinations of colour indices and threshold techniques in vegetation segmentation for volunteer potato control in sugar beet. Manuscript submitted to *Computers and Electronics in Agriculture*.

Suh, H. K., J.W. Hofstee, & E.J. van Henten. Sugar beet and volunteer potato detection under uncontrolled natural environment using deep networks (Manuscript in preparation).

Conference proceedings

Suh, H. K. (2017). Smart agriculture in the Netherlands and Northwest Europe. Europe-Korea Conference on Science and Technology (EKC), July 26-29, 2017, Stockholm, Sweden.

Suh, H. K., J.W. Hofstee, Joris IJsselmuiden, & E.J. van Henten (2016). Discrimination between volunteer potato and sugar beet with Bag-of-Visual-Words model. International Conference of Agricultural Engineering (CIGR-AgEng), June 26-29, 2016, Aarhus, Denmark.

Suh, H. K., J.W. Hofstee, & E.J. van Henten (2014). Shadow-resistant segmentation based on illumination invariant image transformation. International Conference of Agricultural Engineering (AgEng), July 6-10, 2014, Zurich, Switzerland.

Suh, H. K., J.W. Choi, S.M. Park, S. Chung, & S.I. Cho (2012). Developing romaine harvesting manipulator for plant factory with RoboticsLab simulation. The 6th International Symposium on Machinery and Mechatronics for Agriculture and Biosystems Engineering (ISMAB), June 18-20, 2012, Jeonju, Korea.

Choi, J. W., **H.K. Suh**, S.M. Park, & S.I. Cho (2011). Vision-guided detection of harvest cutting position algorithm for romaine lettuce in plant factory. Proceedings

of the 2011 Summer Conference, Korean Society for Agricultural Machinery, 16(2), 279-283 (In Korean).

Taylor, T. A., S. S-F. Smith, & **H-K. Suh**. (2007). A Virtual Harp with Physical String Vibrations in an Augmented Reality Environment. 2007 ASME International Design Engineering Technical Conference, Las Vegas, NV, September 07 (DETC2007-35408).

Other publications

Hemming, J., & **H.K. Suh** (2018). Monitoring system for spider mite damage and yellow sticky traps: PeMaTo-EuroPep Project.

Hoste, R., **H.K. Suh**, & H. Kortstee (2017). Smart farming in pig production and greenhouse horticulture: an inventory in the Netherlands. Wageningen University & Research, Report 2017-097. ISBN: 978-94-6343-218-4.

Min, S.K., & **H.K. Suh** (2017). Big data, the future of agriculture. RDA Interrobang, Report No.199. ISSN: 2233-5056, Reg No.: 11-1390000-002866-03.

Hofstee, J.W., **H.K. Suh**, & E.J. van Henten (2014). Modulaire unit voor detectie van aardappelopslag onder verschillende lichtomstandigheden. Leerstoelgroep Agrarische bedrijfstechnologie (In Dutch).

Cho, S. I., Y.R. Kim, D.S. So, **H.K. Suh**, & J.W. Lee (2009). Study of nano research trend on novel(nano) food materials and safety management research. Korean Food & Drug Administration, Report 2009-214 (In Korean).

PE&RC PhD Training Certificate

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities).

- **Review of literature (4.5 ECTS)**
  - Automated weed control systems (2013)
- **Writing of project proposal (4.5 ECTS)**
  - Mobile-based autonomous system for real-time detection and removal of volunteer potatoes in sugar beet field (2012)
- **Post-graduate courses (4.1 ECTS)**
  - Bayesian statistics; PE&RC (2013)
  - Multivariate analysis; PE&RC (2014)
  - Machine learning; University of Amsterdam (2015)
  - Machine learning; Stanford University (2015, 2016)
- **Deficiency, refresh, brush-up courses (3 ECTS)**
  - Scientific skills training; WIAS (2012)

- **Competence strengthening / skills courses (7.3 ECTS)**
  - Project and time management; WGS (2013)
  - Competence assessment; WGS (2013)
  - Scientific writing; WGS (2014)
  - Teaching and supervising thesis student; WGS (2014)
  - Efficient writing strategies; WGS (2014)
- **PE&RC Annual meetings, seminars and the PE&RC weekend (1.8 ECTS)**
  - PE&RC Weekend (2012)
  - PE&RC Day (2012 - 2014)
- **Discussion groups / local seminars / other scientific meetings (6.6 ECTS)**
  - Modelling and Statistics Network (MSN) (2013-2015)
  - KOSEA Seminars & meetings; Korean Scientists and Engineers Association in Netherlands (2013-2016)
  - GreenVision meetings (2014-2016)
  - IEEE-AgRA Meetings (2014-2016)
- **International symposia, workshops and conferences (4 ECTS)**
  - International Conference of Agricultural Engineering; Zurich, Switzerland (2014)
  - International Conference of Agricultural Engineering; Aarhus, Denmark (2016)
- **Supervision of BSc & MSc students**
  - Particle filter based navigation of a mobile robot
  - Colour constancy of plant material with a HDR camera
  - Navigation of a Husky robot with a particle filter
  - Machine vision based navigation in a sugar beet crop field of a Husky robot
  - Navigation of the Husky in a sugar beet field

Notes

Colophon

Funding: The research described in this thesis was funded by Interreg IVa (project SmartBot), European Fund for the Regional Development of the European Union and Product Board for Arable Farming.
Cover design: Hyun K. Suh, Yoon Choe
Printing: GVO drukkers & vormgevers B.V., Ede, the Netherlands