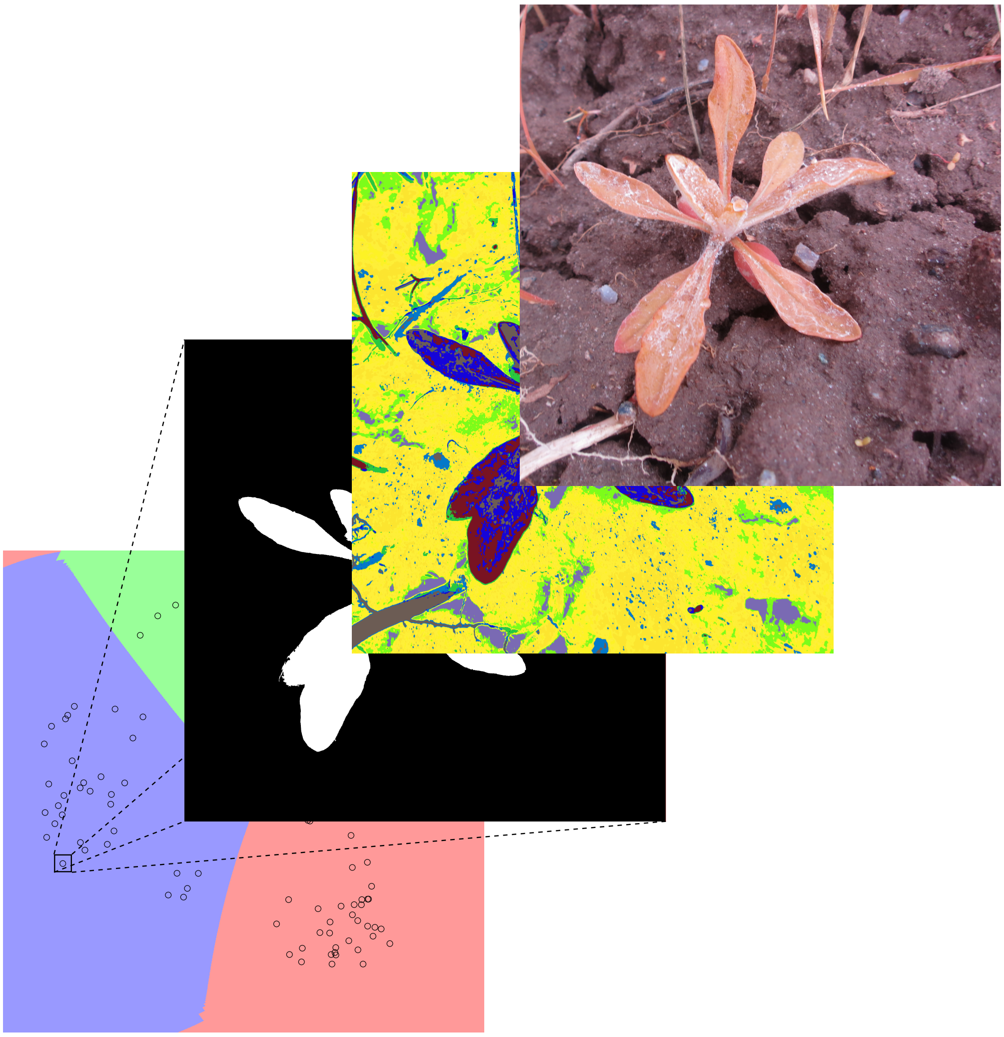
CHALMERS
UNIVERSITY OF TECHNOLOGY



Weed classification and measurement

From computer vision to machine learning

Master's thesis in Complex Adaptive System

Victor Nilsson

Department of Mathematical Siences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2017

Master's thesis 2017:NN

Weed classification and measurement

From computer vision to machine learning

VICTOR NILSSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology
Gothenburg, Sweden 2017

Weed classification and measurement
From computer vision to machine learning
VICTOR NILSSON

Supervisor: Mats Rudemo, Department of Mathematical Sciences
Examiner: Mats Rudemo, Department of Mathematical Sciences

Supervisor @ farmdrones: Therese von Hackwitz

,

Master's Thesis 2017:NN
Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: A schematic view of the computer vision to machine learning process.

Typeset in LaTeX
Printed by [Name of printing company]
Gothenburg, Sweden 2016

Weed classification and measurement
From computer vision to machine learning
VICTOR NILSSON
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Keywords: lorem, ipsum, dolor, sit, amet, consectetur, adipisicing, elit, sed, do.

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Victor Nilsson, Gothenburg, June 2017

Contents

List of Figures

List of Tables

1
Introduction

"Boot up"

Since the dawn of computers, humans have had access to computational power previously far beyond reach. In the beginning this was used to numerically solve Ordinary Differential Equations, (ODE), that had no exact explicit solution. With mathematical computational, the computer was able to retrieve solutions to problems previously far beyond human reach. Although, with the immense computational power that the computer possessed it still lacked something that made man superior still. Intuition, creativity and being able to reason about results given prior knowledge for similar problems. As the computer had to be programmed and do exactly as programmed, it was as good as the code giving it instructions. This made the computer static, the program was never able to learn from previous problems.

This limitation is something that is progressively being erased today. Machine learning is a field of science, in which one studies the constructions of algorithms that learn from data to make statistical predictions on new data. In this thesis, difference machine learning algorithms are discussed and compared on an application in image recognition.

1.1 Background

Over-fertilization of crops and heavy use of herbicides in weed control introduces chemicals into the ecosystem. Reducing the amount of these substances is therefore of a problem requiring attention. In order to make best use of these products, local measurements of the required chemicals can be used for optimal distribution. For this to work, information about the system, as well as good analytic tools is needed to interpret the current state. Using cameras to acquire images over the fields will enable a database to be used in machine learning algorithms that will enable extraction of information over the system.

1.2 Aim

The primary goal of this thesis is to be able to extract and classify different kinds of weeds that are located in a farming field. Given an image of a section of a field, information about the field is in form of location and density of the weeds. This information will be crucial when determining the state of the field and what is the proper course of action in order to both maximize the yeild and growth of the intended plants.

1.3 Limitations senare

There exists different kinds of limitations within the framework of this project. The first limitations is due to small amount of data. In order to give machine learning algorithms a chance to do its work, a large amount of data is required to make proper assumptions on the dataset. For the initial stages of the project, a dataset of 8 different kinds of weeds are provided with 27 images each. On this dataset, the different algorithms are tested and evaluated.

1.4 Temporär information

Name	Command
Chapter	`\chapter{`*Chapter name*`}`
Section	`\section{`*Section name*`}`
Subsection	`\subsection{`*Subsection name*`}`
Subsubsection	`\subsubsection{`*Subsubsection name*`}`
Paragraph	`\paragraph{`*Paragraph name*`}`
Subparagraph	`\paragraph{`*Subparagraph name*`}`

2

Computer Vision

"A picture is worth a thousand words"

2.1 Image

2.2 Computer Vision

2.3 Images

Using images is a way to represent data in a visual manner. An image is a 2-dimensional object that is defined by its individual pixels. The information contained in each of these pixel is expressed using a function, $I(x, y)$, where x and y are the so called *spacial coordinates* of the pixel often represented by a pixel on a screen. The value of this function $I(x, y)$ is called the *intensity* in this pixel. More often then not, $I(x, y)$ contains more information than is given by only one value, e.g. the intensity of different colors such as red, green and blue in a *RGB* valued image. These different intensities can be represented using different images stacked on each other, i.e. $I_R(x, y)$, $I_G(x, y)$, and $I_B(x, y)$ for the RGB images. This can be encoded into only one concatenated image, I, but not expressed in spatial coordinates but rather using different *channels*. The final form of the image function is $I(x, y; c)$ where c is the channel and $c \in \{0, 1, \ldots, d\}$ where d is the number of channels the image.

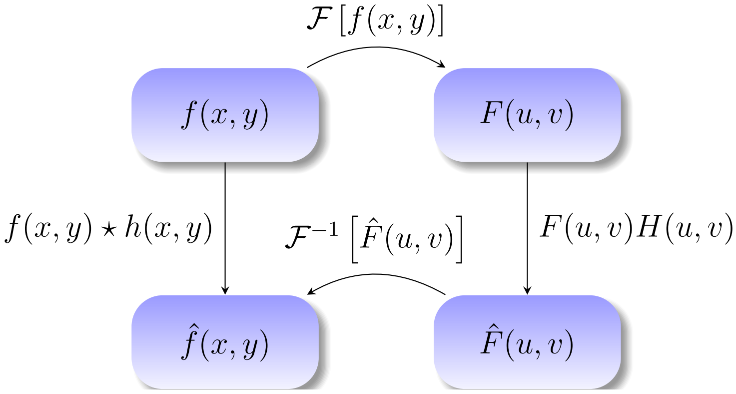


Figure 2.1: How to procedure in the fourier filtering domatin workds

2.3.1 Information in images

An image $I(x, y; c)$ usually takes discrete values based on the number of bits represented by each pixel, which is called the pixel depth. The three values of R (red), G (green), and B (blue) are usually represented using 8 bits each, so the pixel depth is 24 bits. This means that each pixel is built upon 3 digits in the range of $\{0, \ldots, 255 = 2^8 - 1\}$ which gives rise to $(2^8)^3 = 16777216$ different combinations of colors. The set of spatial coordinates in an Image are discrete are defined by a width and height (M, N), that is $\{(x, y) \in \mathbb{R}^2 | 0 \leq x \leq M - 1, 0 \leq y \leq N - 1\}$. Combining both the different pixel values with the spatial domain the whole image space of a given size contains a total of $(2^8)^3 * M \times N$ different elements. Considering that M and N are usually of order $\approx 10^2 - 10^4$ in applications, the amount of different images are immense. Surely there must be a way to classify the message of an image but the use of local information. In figure 2.2 we can see two binary images (each pixel only takes either 0 or 1 in one channel) that both represents the digit five. As seen in the middle picture, their overlap are of little importance for understanding the separate images. This shows that a given message embedded into a specific image can be just one in a subset of images conveying the same information. This subset of images defines a specific class, a class that gives meaning to all the images in this subset, such as the digit five. In order to determine whether a given image is of this subset or not, a certain number of criteria, or features, are required to be fulfilled. In the images seen below, this classification might require a 'S'-like shape with flatter line on the top. In the rest of this section the prerequisites for digitally extract meaningful features are presented.

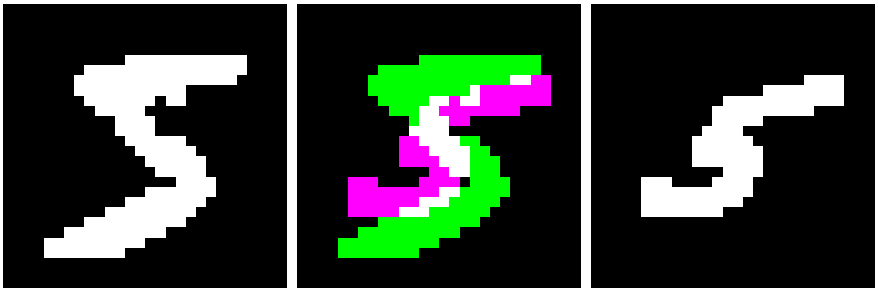


Figure 2.2: Above we see two binary images of size 28×28 that represents the digit 5. Individually there is no ambiguity of their illustrations, but studying their overlap in the middle picture they might as well come from different classes as the conjoined sections are far smaller than the secluded parts.

2.3.1.1 Color arithmetics

As previously discussed, an image contains information in different channels and normally in the form of an RGB image. The RGB color space is convenient for several reasons, first it is close to how humans perceive colors from the photoreceptor in our eyes and that the colors red, green and blue can be combined in an additive manner to obtain all other colors. Additive colors works in an intuitive way, imagine a completely dark room with three dimmer switches, each controlling the intensity of three different lamps. Each lamp emits either red, green, or blue light. One would find that different intensity settings would yield different colors on the walls and if all three would be maximized, white light would be shown. The opposite is true for subtractive color, here one would start with a completely white room, and the three dimmers would determine how much of each colors should be removed. The subtractive color combination is used for example when printing. One starts with the white color and then the printer receives how 'little' of each of the colors (cmy) cyan, magenta, and yellow should be present on the printed paper. The differences of the additive and the subtractive color combinations can be seen in Fig. B.2.

(a) **(b)**

Figure 2.3: Combining colors can either be done additively (*(a)*: The normal RGB color space combines red and blue to magenta,blue and green to cyan, green and red to yellow, and all three to white) or subtractively (*(b)*: the cmy color space combines colors subtractively by removing the selected colors from a pure white color).

2.3.1.2 Color spaces

In image processing there is often a need to separate colors into different groups, e.g. when trying to separate a green plant from a brown soil. A problem might be that the different colors are best for representing the different groups, the groups could share many pixels with similar intensity in several channels. To make the colors more distinct, the use of another color space could help give more useful information for the different groups.

2.3.1.2.1 RGB A color space is the coordinated system used to represent the different kinds of available colors. The most common one, the RGB color space can be seen as a orthogonal coordinate system with each of the axis representing one of the colors and the distance from the origin represent the intensity.

- RGB
- HSI
- NDVI
- MSAVI
- MSAVI2
- SAVI

(a) **(b)**

Figure 2.4: Color coordinate systems

2.3.1.3 Layers of information

If we yet again consider the images in figure 2.2, the classification problem does not use the local information in the pixels to make the classification, nor does it use small regions of information but rather uses the whole image space to fully extract the important features. This means that we could change the gray level values of some pixels without changing the output, which for example could originate from a noisy signal. The result would not be altered even when some regions of the image is changed, for example, the same classification should be done even if we somehow made the five thinner or wider. From this we can conclude that there are different sizes where information is present. Generally we can divide such sizes into three main groups namely,

- Local, where each pixel is considered separately
- Regional, A neighbourhood around the pixel is used to determine some spatial information locally
- Global, The whole image as a whole is considered

We shall later see how we can use information from each of these groups as inputs to different classification algorithms. The most intuitive ones uses the highest level of information, namely the from the global group. Some features from this group, used in the classification of digits problem, could be,

- Number of white pixels, Area
- Its curliness, Convexity
- Number of holes, Euler number

2.4 Image segmentation

In order to extract useful information from an image, we need to segment the image into the parts we want to extract information for, e.g. if we want to study a plant in a field we need to remove the background dirt.

2.4.1 Histogram-based thresholding

If the part of the image is fundamentally different from the rest of the image, e.g. a white cloud on a blue sky, we can use a technique of thresholding. In thresholding, we represent the image using only one intensity, i.e. only one channel, and creates a threshold value, T_I. Using this threshold we create a binary image where 1 represents the intense part of the image and 0 the

(motsats intense)

part of the image using,

$$I_S(x, y) = \begin{cases} 1 & I(x, y) \geq T_I \\ 0 & I(x, y) < 0 \end{cases} \tag{2.1}$$

sätt in en bild som förklarar.

Choosing this threshold might be trivial for some images, e.g. images which have a bright and a dim? region, but sometimes the intensities are very close or even overlapping.

3

Machine Learning

"Practice makes perfect"

qinterGeneral machine learning, supervised learning, unsipervised learning
Börja med linear regression,
Sedan Klassifiseringsmetoder som Fisher's discriminant for multiple classes.
Prata om SVM?
Börja sedan med preceptron algorithmen för linjärt separerbara 2 klasser.
Sedan in på multilayer erceptron, till neural nets.
Sedan till convoluted neural nets.
Sedan gå över till image theory flr information extraction.

3.1 Machine Learning

Solve problem without explicitly program the procedure.
Learn and make prediction based on data.
Algorithm operates by a built model from example inputs to make predictions on data previously not presented.
Big challenges are optimization of these models.
These algorithms make predictions based on a training session, and these training sessions can be divided into several categories:

- Unsipervised Learning
  - Unknown Problem/Properties
  - Unknown Features
- Reinforced Learning
  - Not Known result but get reward based on the whole population
  - Input -> Ouput not presented
- Supervised Learning
  - Know which input should yield a specific output in training set and adjust based on the current output

3.1.1 Supervised learning

Learn to predict an output vector when given a input vector
algorithms e.g. (neural net, rnn) feed forward neural net using different learning algorithms, e.g. backpropagation
There are generally two groups of supervised learning

- Regression - Output is range of real numbers

- Classification - output is a target class label

Model class $y = f(x;\omega)$, where f uses numerical parameters ω to map input vector x to predicted output y. During the learning, reduce the discrepency of target t to y.
for regression, minimize a energy function $\frac{1}{2}(y-t)^2$

3.1.2 Reinforced learning

Learn to select an action to maximize payoff.
algorithms e.g. feed forward neural net but learning algorithm e.g. using evolutionary algorithms

3.1.3 Unsipervised learning

Find good internal representation of the input, could be used before the supervised learning to adapt the inputdata into more graspable data for the supervised learning. This method is often used to reduce training time of the supervised learning as you can wire the network more efficiently and thus less weights needs to be altered to match the output.
This stage could be omitted if you can already extract useful features from the input.

3.1.4 Quadratic Discriminant Analysis

Our first multiclass classification will be the so called Quadratic Discriminant. For this classifier we will model our classes with a class conditional distribution, $P(X|K = k)$. This distribution tells us what is the probability of seeing the studied data X given that we are in the class k. We will also assume that the features from plants in a class vary as a multvariate Gaussian distribution, i.e.

$$P(X|K=k) = \frac{1}{(2\pi)^D|\Sigma_k|^{1/2}} exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right), \tag{3.1}$$

where $\boldsymbol{x} = x_1, \cdots, x_D$ are the features of the studied plant X, $\boldsymbol{\mu}_k = \{\mu_{k_1}, \cdots, \mu_{k_D}\}$ and

$$\Sigma_k = \begin{bmatrix} \sigma_{k_{11}} & \sigma_{k_{12}} & \sigma_{k_{13}} & \cdots & \sigma_{k_{1D}} \\ \sigma_{k_{21}} & \sigma_{k_{22}} & \sigma_{k_{23}} & \cdots & \sigma_{k_{2D}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{k_{D1}} & \sigma_{k_{D2}} & \sigma_{k_{D3}} & \cdots & \sigma_{k_{DD}} \end{bmatrix}$$

,
are the mean and covariance of the feature distribution of plant K and D is the number of features that describes the plant.

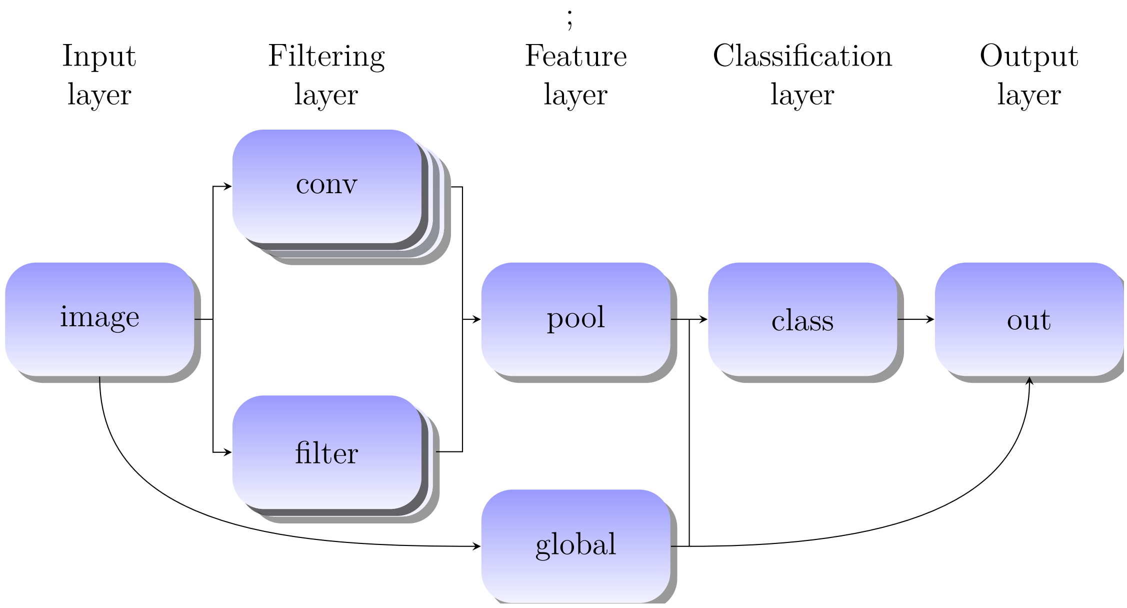


Figure 3.1: this is the struct of the network

4

Information Extraction

"Ball is in your court"

4.1 Descriptive Object Features

4.1.1 Size dependent features

4.1.1.1 Area

4.1.1.2 Perimeter

4.1.1.3 ConvexHUll

4.1.1.4 ConvexHullPerimeter

4.1.1.5 Thickness

4.1.2 Size independent features

4.1.2.1 Mean Color

4.1.2.2 Standard devisation color

4.1.2.3 Area moments

4.1.2.4 area perimterer

4.1.2.5 formfactor

4.1.2.6 Elognatedness

4.1.2.7 convexities

4.1.2.8 solidities

5

Methods

"Actions speak louder than words"

5.1 Method

First part of the project is data acquisition, and currently there are two different desirable approaches, both with different advantages and disadvantages.

One of the approaches would be to grow the crops and weeds oneself in a controlled environment, alongside with a "wild" grown field of unspecified weeds directly taken from a real field. This approach would give me complete control over the data, when it should be extracted but would yield a small database as I don't have neither the time nor the place nor knowledge go grow an entire field.

The other would be to get in contact with a real farmer, taking continuous pictures from a real field. This could give direct feedback of the state of the field, as expert knowledge is in the vicinity. Although, having a "real life" dataset might seems desirable, I might have too little control of the environment, e.g. the farmer might be busy, the fields might be under some treatment which will alter the plants, and there might be some other uncontrollable factors.

Regardless of the approach, the database will be obtained from the fields using a Canon S110 camera, which can give pictures in $4000 \times 3000 pixels$. To get as good pictures as possible, the camera will be mounted on a custom made tripod, which will give pictures from a desirable height.

When the dataset has been specified, a sequence of images over the same area for different times will be overlapped so a time-lapse of the plants will be available. This is a non trivial task if done autonomously as is desirable since the project should be able to scale. The method that will be used for this process is image warping and matching. Possible problems that might occur is that the images are taken to far from each other in time that the images are too different, and some preprocessing might be used if this is the case.

The next part is image segmentation in order to extract the parts of the pictures which represents the different plants. The first part of this process is to take into account of the color channels in the image, since the plants usually have a distinct green color against the background soil which is usually in a brownish color which consist of a heavy redness. Then different edge detection methods will be used, such as applying a laplacian filter, in order to differentiate plants that resides close to each other.

When the position of each plant is determined, it is time to perform two classification problems on the dataset. The first is a binary classification, is the plant a crop or a weed, and the second is only performed on the weeds given from the first classification. During this part the kind of weed should be determined. Given the distinct features in each of these classes different methods will be considered. If the features is easily distinguished between the classes, e.g. shape and color, then classification methods such as linear or quadratic discrimination will be considered as well as support vector machines. If the classification part seems to be more complex, a machine learning approach be more appropriate, such as a convolutional neural network. This approach is actually preferable as it is easier to scale to include more classes in the future, but I might be limited by the amount of data accessible, since this approach requires a large amount of data.

5.2

Material beskriver vilka komponenter som ingår i de laborationer som har utförts samt övrigt material som har använts. Utelämna inget men texten bör vara koncis snarare än uttömmande. Försök därför lyfta ut och beskriva centrala delar ur exempelvis laborationsinstruktioner (exempelvis instrument, instrumentinställningar, kemikaliekoncentrationer), och kravspecifikationer i texten. Lista inte allt utan lägg hellre långa instruktioner och protokoll som bilagor och hänvisa till dessa. Syftet med Material och Experiment/Metod är att en laboration ska kunna återskapas eller ett projekt ska kunna utvärderas mot bakgrund av metodval och genomförande. Material och Metod kan vara två separata avsnitt/kapitel.

6

Results

"Cut to the chase"

6.1 Method

6.2

Återger resultat från observationer och datorstödda analyser/mätningar. Vanligtvis återges resultat i form av figurer, tabeller, grafer, eller fotografier men det är avgörande att även förse dessa med tillräckliga kommentarer för att belysa resultatets värde och eventuella avvikelser eller oväntade resultat. Givetvis ska resultatavsnittet också bestå av text som binder samman illustrationerna. Beroende på experiment eller metod bör skribenter ange och berättiga beräkningar och formler som ligger till grund för tolkningen av resultat och datorstödd analys. I vissa fall måste även primärdata anges för att resultat ska bli meningsfulla och kunna diskuteras. Med primärdata avses obehandlade mätdata. Notera att detta inte är tillämpligt i alla enskilda fall. Fråga din ämneshandledare.

6.3

Återger resultat från observationer och datorstödda analyser/mätningar. Vanligtvis återges resultat i form av figurer, tabeller, grafer, eller fotografier men det är avgörande att även förse dessa med tillräckliga kommentarer för att belysa resultatets värde och eventuella avvikelser eller oväntade resultat. Givetvis ska resultatavsnittet också bestå av text som binder samman illustrationerna. Beroende på experiment eller metod bör skribenter ange och berättiga beräkningar och formler som ligger till grund för tolkningen av resultat och datorstödd analys. I vissa fall måste även primärdata anges för att resultat ska bli meningsfulla och kunna diskuteras. Med primärdata avses obehandlade mätdata. Notera att detta inte är tillämpligt i alla enskilda fall. Fråga din ämneshandledare.

6.4 Segmentation

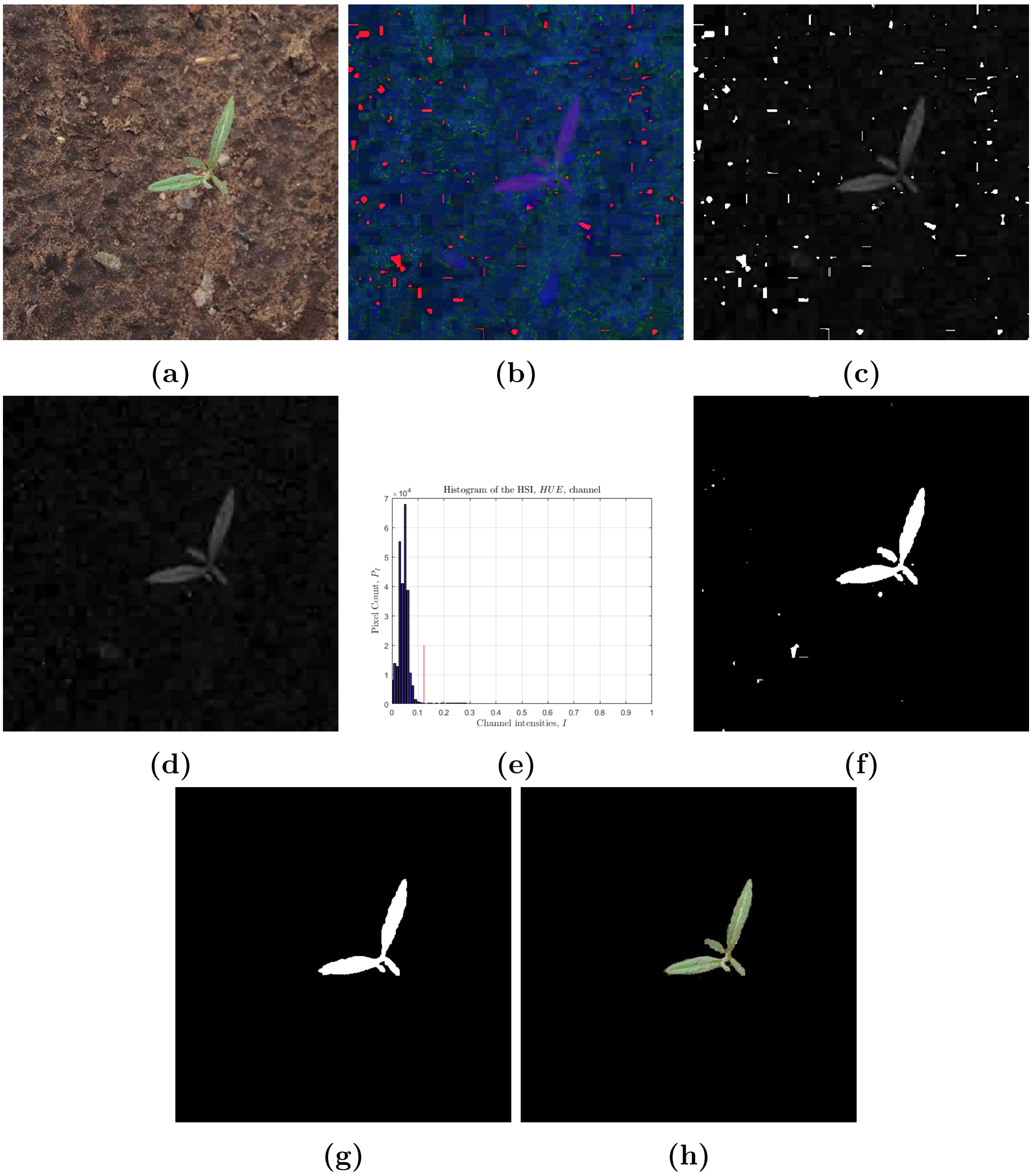


Figure 6.1: The segmentation process as follows: *(a):* The original image is loaded and prepared for processing. *(b):* The image converted to HSI but still represented in RGB colours. *(c):* For the segmentation to work properly the image needs to be represented in one channel that represents the two segments well, and this is what the Hue channel in HSI representation does. *(d):* The Hue channel clean up. *(e):* The optimal threshold between the two segments is represented by the red line *(f):* Binarization of the HSI image using clipping. *(g):* Clean up the binary image by removing small objects. *(h):* Final segmented image.

(a)

(b)

Figure 6.2: Directly applying the same segmentation algorithm on non-perfect noise data does not really give satisfying results.

Using the same algorithm on the data received on a field without alterations, and also quite late in the development does not give good result as can be seen in figure 6.2. This is due several reason, first we can see that there are more than one interesting plant in the picture, and from the results we see that only a part of a plant is given. This is where the convolutional neural networks come in for the first time, using a convolutional neural network one can train the networks to segment the image into several important parts. Using this network to classify segments of an image could be done as prework and then the normal classification using a normal feed forward network or just the quadratic classifier, but why stop here when we have already integrated a CNN, why not use it in series with another CNN to only train one big network instead of using several different parts of the classification.

6.5 Results from quadratic discrimination

Table 6.1: Feature names along with their corresponding feature number during feature selection phase.

Feature Number	1	2	3	4	5
Feature Name	From	Elog	Conv	Solid	mean Red
Feature Number	6	7	8	9	10
Feature Name	Mean green	Mea blue	std Red	Std green	Std blue
Feature Number	11	12	13	14	
Feature Name	area moment 1	area moment 2	area moment 3	perimeter moment	

Figure 6.3

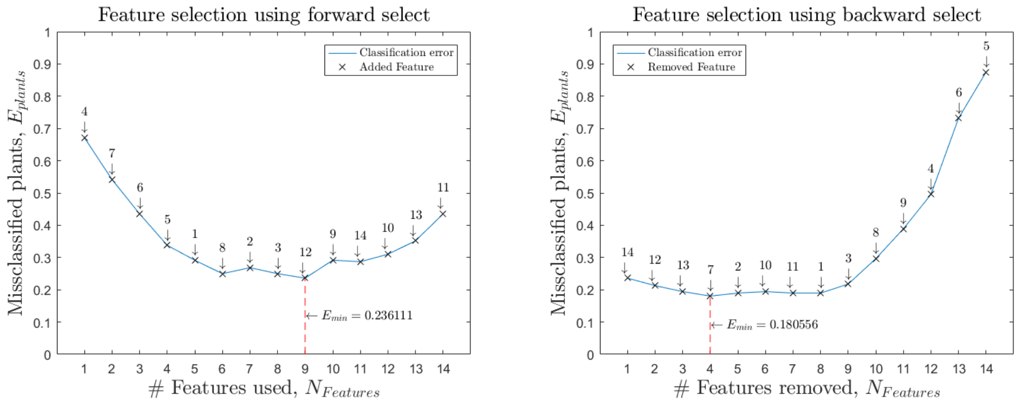


Figure 6.4: Using either forward or backward selection for the best feature combination yields different results. E.g. Feature number 7 (mean blue) is the second feature to be added in the forward selection algorithm and is the last to be removed in the backward selection. The corresponding features can be found in Table 6.1.

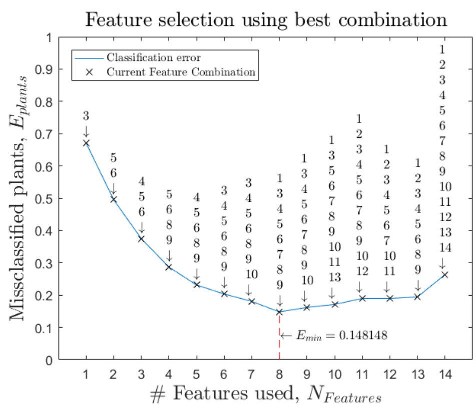
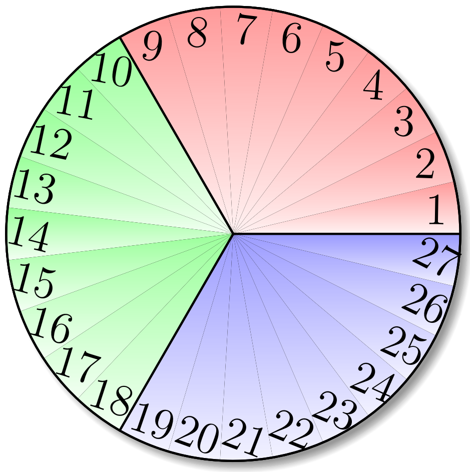
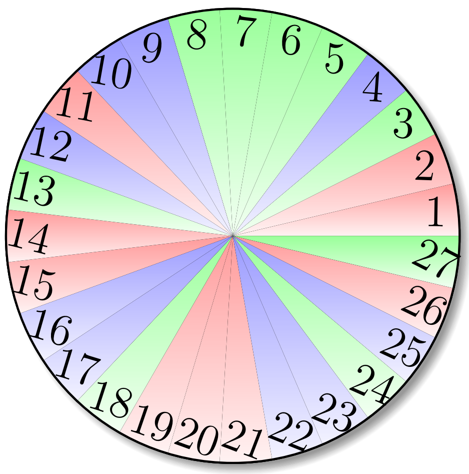
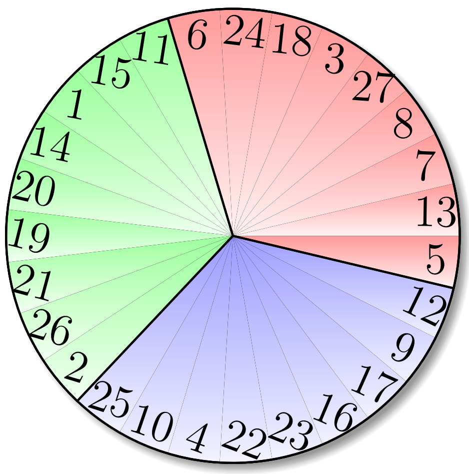


Figure 6.5: With immense computer power the best n combinations can be brute forced. This ensures to get the best available combination using n features as all combinations are compared to each other.

As we can see in Figure 6.4, many of the defined features are present when selecting the best combinations of features using both the backward and forward selection algorithm. In both cases, the class of features which are neglected the most are the moment invariant features. During these algorithms $K-$crossvalidation was used with $K = 3$, i.e. the different plants was randomly divided into 9 different group with 3 plants in each. This grouping is different in each run and might thus yield different results for different runs. To make this more consistent, the grouping could be done using increasing the dart wheel as in Figure **??** indexing instead of the random distribution in Figure **??**.
With much patients the absolute best combination was found by combining the best n features together. This can be seen in Figure 6.5. The result from using







this algorithm will only be used in a best case scenario when comparing to other methods later as this algorithm is very slow compared to the other. The feature selection process usually takes place before the classification is done, thought, if one would find a new feature that might give more information this whole process would need to be made again and thus a fast algorithm is preferred.

6.6 Results from using Neural Networks

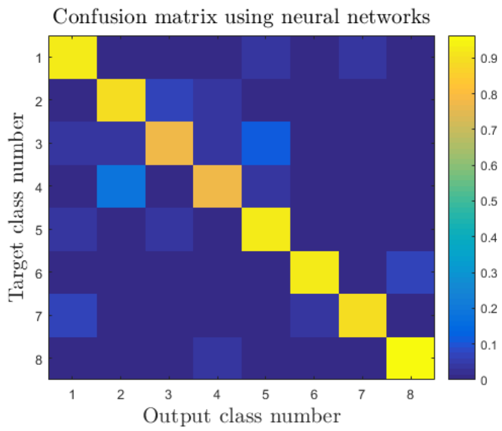


Figure 6.6

7
Discussion

"From a to z"

7.1

Här behandlas resultaten i förhållande till teori och metod och utvärderas med utgångspunkt från syftesformuleringen i inledningen. Diskussionen omfattar även det förväntade resultat som referenslitteraturen (teori, metod, databas) indikerar samt förklarar enskilda specifika resultat och mätdata. Även avvikande resultat och data måste identifieras och förklaras genom synpunkter angående material, teori och utförande. Diskussionen behöver dessutom också anknyta till eller hänvisa till inledningen och problemställningen.

7.1.1

I vissa situationer kan de två funktionerna resultat och diskussion kombineras till ett avsnitt i en rapport. Fråga din ämneshandledare.

8

Conclusion

"To make a long story short"

8.1

Här redovisas experimentets eller projektets slutsatser mot bakgrund av diskussion och resultatredovisning och ofta genom att också anknyta till inledningen av rapporten. För många undersökande projekt är det även vanligt att slutsatser leder fram till rekommendationer. Notera att vissa texter kombinerar avsnitten diskussion och slutsatser. Fråga din ämneshandledare

Bibliography

[1] Rafael C. Gonzalez . Richar... *Digital Image Processing: International Edition*. Pearson, 2009. ISBN: 0132345633 (cit. on p. III).

[2] John Weier and David Herring. *Measuring Vegetation*. 2010. URL: `https://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_2.php` (visited on 07/01/2017) (cit. on p. IX).

[3] *Soil-adjusted Vegetation Index*. 2013. URL: `http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:soil-adjusted_vegetation_index` (visited on 07/01/2017) (cit. on p. IX).

[4] *Modified Soil-adjusted Vegetation Index*. 2012. URL: `http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:modified_soil-adjusted_vegetation_index` (visited on 07/01/2017) (cit. on p. IX).

A

Convolutions

Some dummy text

A.1 adnd also a dummy section

A.1.1 with a dummy subsection

A.1.1.1 with yet another dummy layer

B

Color representation

This part extracts infromation from the course book Digital Image Processing, Gonzales [1] unless other sources stated.

B.1 Primary Colors

The use of images is a great way of providing information visually. An image is represented by a rectangular grid where each cell, or hereafter pixel, emits light of a certain wavelength. When the color of each pixel is determined and is ready to be displayed, there is a problem of how to convey the color information from the source to the recipients eyes. On way would be to many different light sources for each pixel, where each source is responsible for a small region of wavelengths. This is problematic due to our wide range of visible colors and in order to reduce the number of light sources we need, these wavelength regions must be chosen with care. This problem is slightly simplified by the fact that our eyes consists of three different sensors that react to different wavelengths, and these are roughly equal to blue, green and red colors and their absorption can be seen in Fig. B.1.

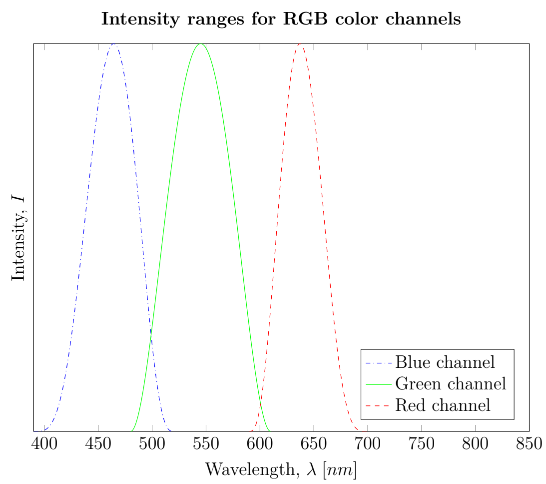
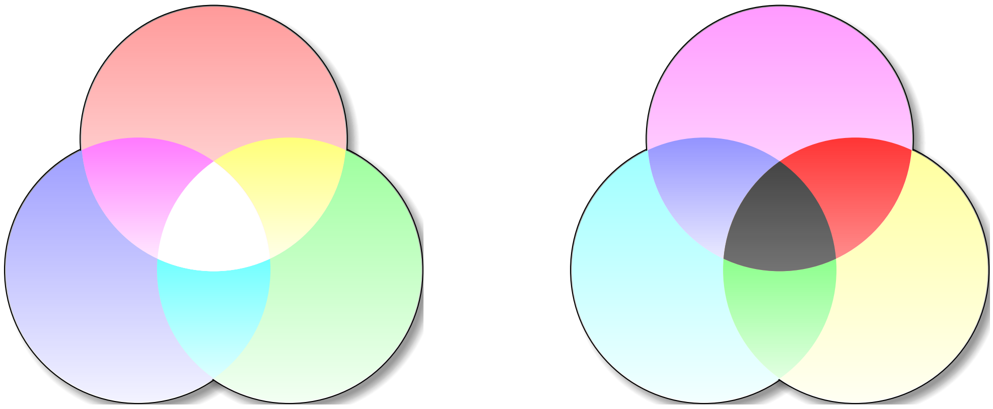


Figure B.1: When incoming photons with different wavelengths λ hits each receptor, they get excited depending on their ranges. The figure shows, not accurate ranges nor intensities, but illustrates a schematically the different ranges for the different colors.

These three colors, red, green, and blue are therefore called the *primary colors* and will be the basis for image representations. There are two different ways of using these colors to provide the targeted color, in a additive or a subtractive manner. When adding the primaries we start with a completely black canvas and increase the intensity for each of the colors until we reach a white color. For the subtractive nature we start with a white canvas and determine how much of each color we want to subtract until a black color is reached. The difference between these can be thought of as absorption and emittance, and the distinction can be seen in Fig. B.2.



(a) **(b)**

Figure B.2: Combining colors can either be done additively (*(a)*: The normal **RGB** color space combines red and blue to magenta,blue and green to cyan, green and red to yellow, and all three to white) or subtractively (*(b)*: the **CMY** color space combines colors subtractively by removing the selected colors from a pure white color).

B.2 Color Spaces

B.2.1 RGB and CMY

In the introduction of this appendix we talked about the primary colors and how they are used to represent a wide range of colors. This representation of colors is the **RGB** (red, green, blue) color space for the additive colors, and **CMY** (cyan, magenta, yellow) for the subtractive. Many other color spaces exists and are used in different situations. Ultimately they represents the same color, only through other variables than red, green, and blue. Often the change of color spaces is done by changing the basis for the color representations. E.g. transforming from **RGB** to **CMY** color space is done by the linear transformation,

$$\begin{pmatrix} c \\ m \\ y \end{pmatrix} = 1 - \begin{pmatrix} r \\ g \\ b \end{pmatrix}, \tag{B.1}$$

which makes it clear why the other is called subtractive when the first is additive. Although we can change the color space in which we represent an image, the fact that our screens use red, green, and blue colors to mix colors is still present. So to view a new color space we simply use these light sources and pretend that our new values still lies in a **RGB** color space. The result of this representation can be seen in Fig. B.3.

(a) (b)

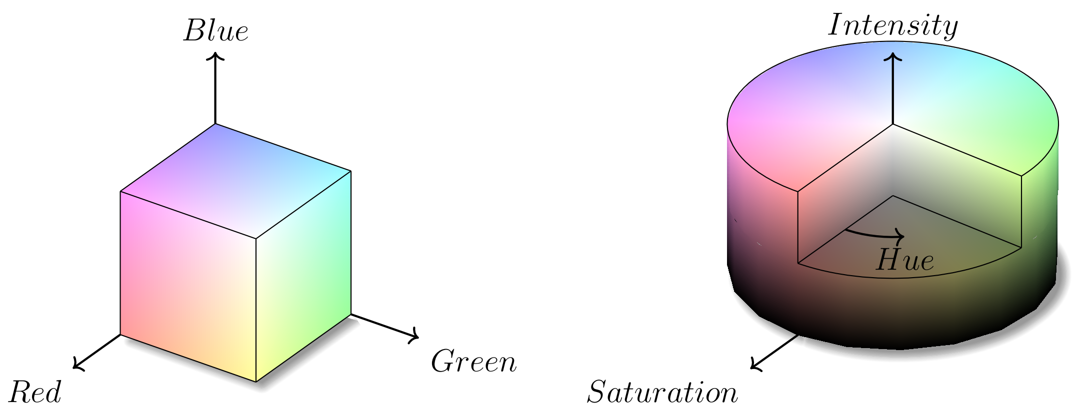
Figure B.3: Presenting images using other color spaces than the ordinary **RGB** color space can yield interesting result. *(a)*: Image show in the ordinary **RGB** color space. *(b)*: The same image as in *(a)* but show in **RGB** representation using **CMY** color values.

Sometimes the information that is interesting in the image in not needed to be expressed using three color channels and a color conversion might only produce a single index for each pixel. These will be called color indices, and basically any color space is build by three different color indices, due to the three different kinds of receptors. The difference between three random color indices and a color space is that the color space spans the entire "color room", meaning that a color space can represent any color and conversion between spaces exists, which is not always true for color indices. Several color spaces/indices will be presented as well as some reason why using them.

B.2.2 HSI

In many computer vision applications there is a need to separate segments of images based on the color of these. This could be tedious if done using the **RGB** color space. E.g. suppose you would want to select a yellowish color, but you care not for an exact color and allow some deviation. A color value is chosen as base, and small deviations in the red color channel will yield about the same color, but also small deviations in the green and blue channels will produce almost the same result. So to select the appropriate range of colors that are representative is dependent on three different deviations, and the target lies in a three dimensional shape. So a color space in which the information of the color is embedded in one of the channels could prove useful. We can imagine that all available colors are represented on a disk where the colors changes on the viewing angle from the origin, and the prevalence of this color increases with distance. So to choose a specific color we would simply choose an angle from the origin which points at the targeted color. Then we would go in that direction until the amount of color is reached and then we increase/decrease the intensity of the selected color. This is a new color space which we call **HSI**,

Hue, **S**aturation, and **I**ntensity. The color coordinates for this color space is here represented by an angle, a radius and a height compared to three distances for the **RGB** color space, the differences is show in Fig. B.4



(a) **(b)**

Figure B.4: Different color spaces uses different coordinates to represent the color room. In *(a)* we see the **RGB** color space which is spanned by orthonormal basis, whereas in *(b)* which shows the **HSI** color space uses an angle, a radius and a height instead.

B.2.2.1 RGB to HSI

Converting between **RGB** and **HSI** is a non-linear transfrom and can thusly not be written in matrix form. In the following calculations the values in **RGB** space is written as $R, G, \text{and} B$ and **HSI** $H, S, \text{and} I$. Also a **RGB** color is assumed to be in $C_{RGB} \in [0,1]^3$ and **HSI** $C_{HSI} \in [0°, 360°] \times [0,1]^2$. The Hue component is an angle and trigonometric calculations are therefore necessary,

$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360° - \theta & \text{if } B > G \end{cases}, \tag{B.2}$$

where,

$$\theta = \cos^{-1}\left\{\frac{\frac{1}{2}\left[(R-G)+(R+B)\right]}{\left[(R-G)^2+(R-B)(G-B)\right]^{1/2}}\right\}, \tag{B.3}$$

whereas the Saturation and Intensity is easier to calculate,

$$S = 1 - \frac{3}{R+G+B}\left[min(R,G,B)\right], \tag{B.4}$$

and

$$I = \frac{1}{3}(R+G+B). \tag{B.5}$$

B.2.2.2 HSI to RGB

Converting back to RGB colors is also a tedious task, mostly due to that we need three different algorithms depending on the value of H.

$0° \leq H < 120°$

$$\begin{aligned} R &= I\left[1 + \frac{S\cos H}{\cos(60° - H)}\right] \\ G &= 3I - (R + B) \\ B &= I(1 - S) \end{aligned} \tag{B.6}$$

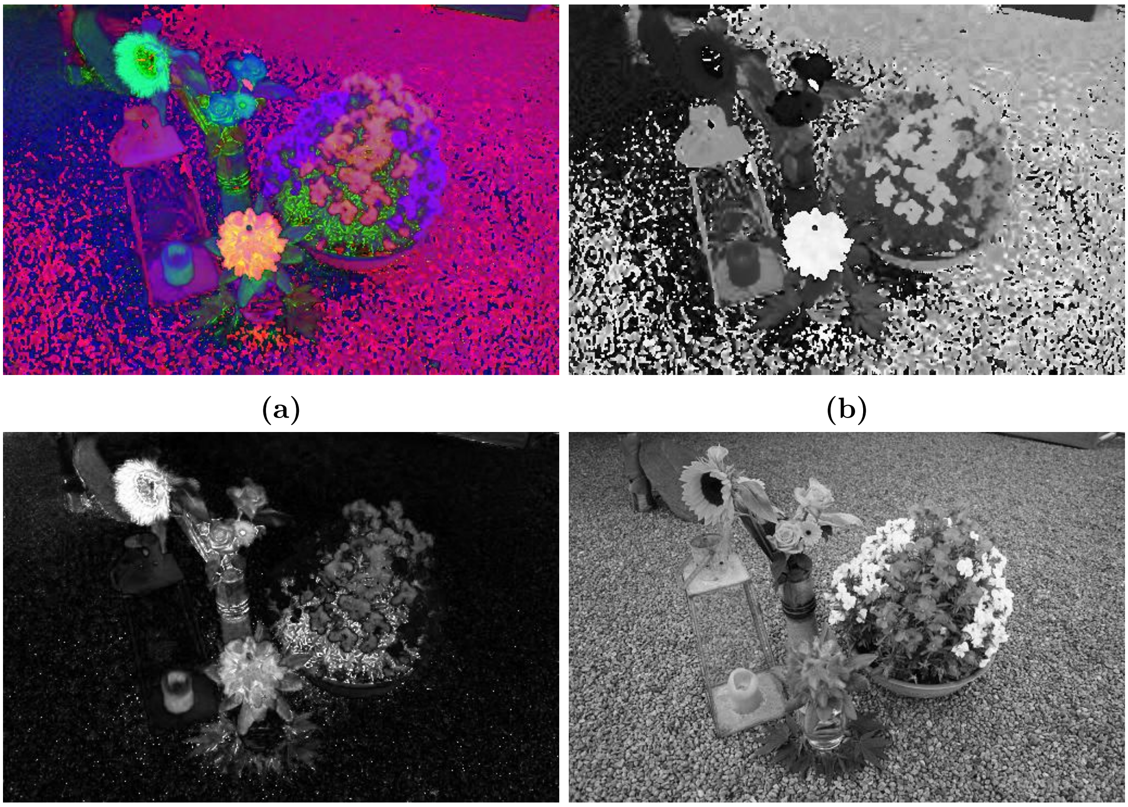
$120° \leq H < 240°$

$$\begin{aligned} H' &= H - 120° \\ R &= I(1 - S) \\ G &= I\left[1 + \frac{S\cos H'}{\cos(60° - H')}\right] \\ B &= 3I - (G + B) \end{aligned} \tag{B.7}$$

$240° \leq H \leq 240°$

$$\begin{aligned} H' &= H - 240° \\ R &= 3I - (G + B) \\ G &= I(1 - S) \\ B &= I\left[1 + \frac{S\cos H'}{\cos(60° - H')}\right] \end{aligned} \tag{B.8}$$

The applicability of the **HSI** color space is hopefully evidently shown in Fig. B.5.



(a) **(b)** **(c)** **(d)**

Figure B.5: It is not evident from the **RGB** representation of the **HSI** color image shown in *(a)* that the colors can be grouped based on one color channel, but is clearer from the Hue channel in *(b)*, which shows large groupings for different parts of the image. The Saturation channels shows occasional grouping in *(c)* but is not as uniform as the Hue, and the Intensity in *(d)* shows large variations.

B.3 Color Indices using Near Infrared wavelengths

One great thing with using electromagnetic radiation (light) as a source of information is that we do not need to use wavelengths comprehensible to our brains. The implications of this gives us the ability to use light information from wavelengths outside the visible spectrum. This can be of immense help when gathering information from biological living things which usually function in certain ways due to physical or chemical consequences. To further explain this we realise that light affect biological cells by exciting its building components to a higher energy state, and if the energy in a photon does not lie within the range of the possible excitation energies, there is no reason for a cell to absorb that light. So the reason why we cannot see certain wavelengths is due to almost full reflectance of those energies in our eyes receptors, and the same is true for most biological molecules as they are of the same order of magnitude in size and should be affected by roughly the same wavelengths, albeit in different ways. If we could somehow capture the near infra red light ($\lambda \approx 750 - 1400\,\mathrm{nm}$) we could "see" the healthiness of a plant as a sick and degenerate cells does not behave as a living, healthy one as the molecules starts to crumble and behave differently.

B.3.1 NDVI

NDVI or **N**ormalized **D**ifference **V**egetation **I**ndex is used to differentiate healthy vegetation from either unhealthy or sparse vegetation sections[2]. What enables the usage of this index is that active chlorophyll in plants strongly absorbs visible light and a living cell strongly reflects near infrared light as it can not use this energy internally, and absorbing it would cause overheating. On the other hand, inactive or dead chlorophyll does not absorb the visible light to the same degree whereas the cell reflects roughly the same amount of near infrared light even if the chlorophyll is degenerate. The index is calculated by the formula,

$$NDVI = \frac{nR - VIS}{nR + VIS}, \tag{B.9}$$

where nR is the near infrared index and VIS is the representation of the visible light. This index ranges from -1 to +1 where a number close to +1 represents a healthy plant. Besides determining the health of plants, this index can also be used to determine what part of an image is a plant or not as a low index corresponds to sparse or no vegetation.

B.3.2 SAVI

The previously discussed index, **NDVI**, is great when determining the health of vegetation that is dense. If, on the other hand, the vegetations is sparse, then the reflectance of the soil could come to matter. Thus a **S**oil-**A**djusted **V**egetation **I**ndex [3], **SAVI**, could help with this problem which is a modification of **NDVI** using a *Soil brightness correction factor*, L is introduced to and extends Eq. (B.9) to,

$$SAVI = \frac{nR - VIS}{nR + VIS + L}(1 + L), \tag{B.10}$$

where $SAVI(L = 0) = NDVI$. The factor L is adjusted by the amount of green vegetation, where $L = 1$ means a low amount of greenness and $L = 0$ a large amount.

B.3.3 MSAVI

The biggest problem with the **SAVI** index is the parameter L which is determined for each problem by trial-and-error. The **M**odified **S**oil-**A**djusted **V**egetation **I**ndex [4], **MSAVI** is calculated the same way as **SAVI** Eq. (B.10),

$$MSAVI = \frac{nR - VIS}{nR + VIS + L}(1 + L), \tag{B.11}$$

only that in this case, L is not a constant but is calculated using,

$$L = 1 - \frac{2s \cdot (nR - VIS)(nR - s \cdot VIS)}{nR + VIS}, \tag{B.12}$$

where s is yet another parameter, but is a equipment dependent parameter rather than application dependent one.

B.3.4 MSAVI2

The final Color index we will consider using the near infrared channel is completely parameter free, but the expression looks nastier. This new index is derived from **MSAVI** using recursion (the details will not be given here) and the final expression is calculated by,

$$MSAVI2 = \frac{\left(2nR + 1 - \sqrt{(2nR+1)^2 - 8\left(nR - VIS\right)}\right)}{2}. \tag{B.13}$$

Use the indices for healthy and unhealthy plants, make new figure showing the comparisons

These four color indices can be seen in Figures B.6c to B.6f.

B.4 nRGB to RGB

Above it is mentioned that near infrared light can come to help in various image analysis problems. However, this does not come for free, unless having access to really special equipment, most image light sensors only provide three different channels for capturing. Thus, the inclusion of a new channel, near infrared, implies the exclusion of another. The simplest would be to change it with the red channel as it lies closest in wavelength. If viewed using the **RGB** format, the resulting image will most likely be very reddish in color as most materials reflect the near infrared light, the comparison can be seen in Figures B.6a and B.6b

Use this image to show rgb and nrgb channels

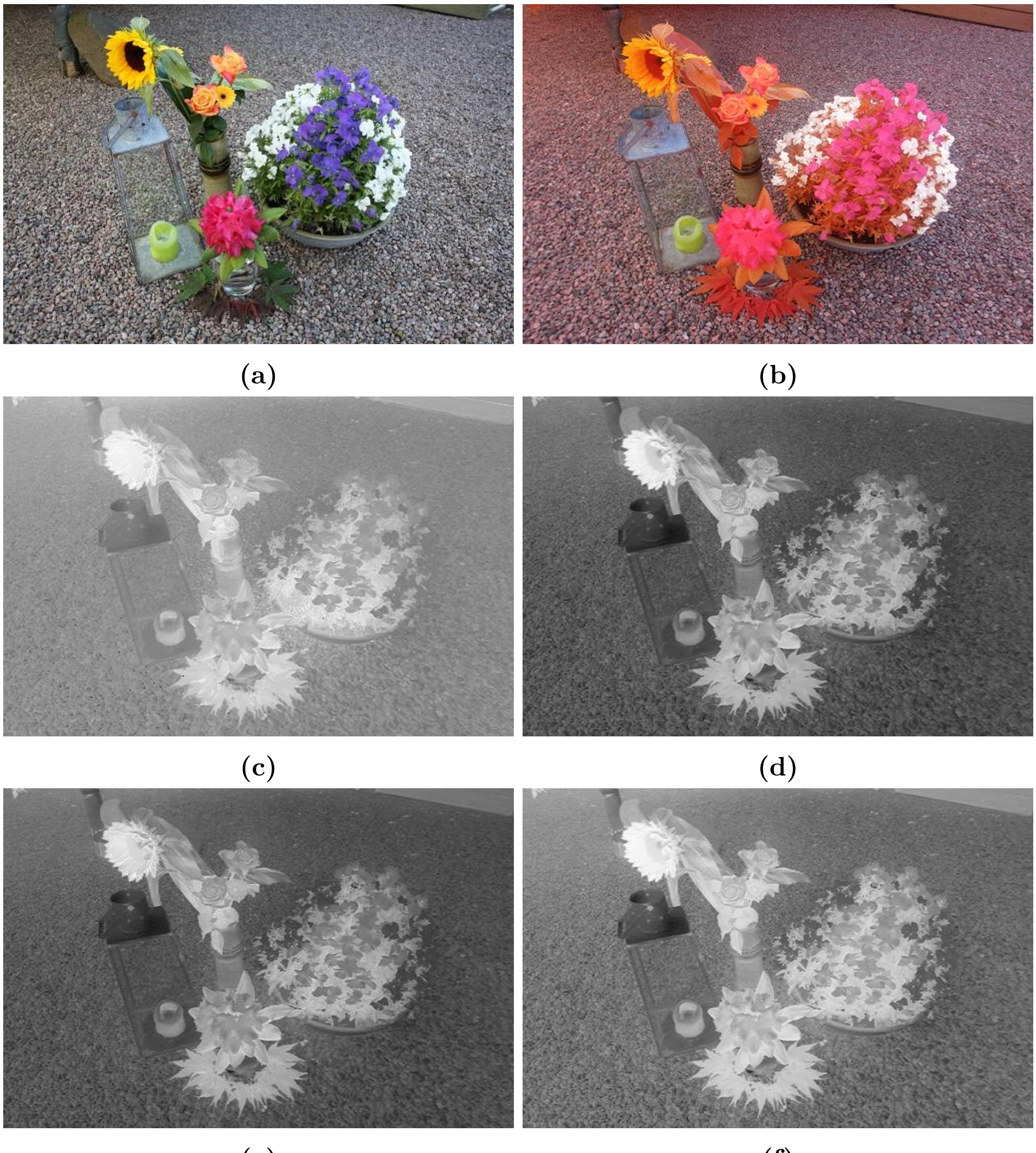


Figure B.6

In order to show the **nRGB** image we want to "simulate" it as it was an ordinary **RGB** image. So we want to move the wavelengths of the new channel to a targeted one. Mathematically this is done using a convolution, described in Chapter. , with convolving function, $\delta(x)$ where x is the targeted channels mean. This modification, replacing the red channel with the near infrared channel can be seen in Figure. **??**.

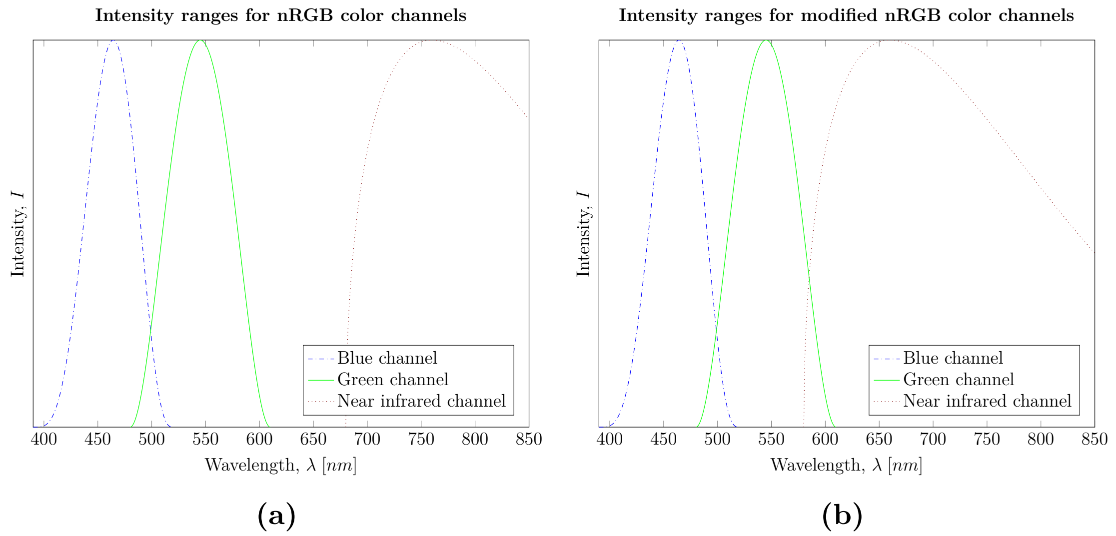


Figure B.7: In *(a)* we can see the capturing of light in the **nRGB** color space. In order to display it using **RGB** color channels, we need to modify the wavelengths to the ordinary of **RGB** colors. This modification of the near infrared channel is showns in *(b)*.