

Project title:	Early detection of stress in strawberry plants using hyperspectral image analysis
Project number:	SF 144
Project leader:	Andrew French University of Nottingham, Jubilee campus, Wollaton Road, Nottingham, NG8 1BB
Report:	Annual report, March 2015
Previous report:	None
Key staff:	Andrew French (Joint Supervisor) Nicola Harrison (Joint Supervisor) Amy Lowe
Location of project:	University of Nottingham East Malling Research
Industry Representative:	Marion Regan Hugh Lowe Farms, Barons Place Mereworth, Maidstone, ME18 5NF
Date project commenced:	1 May 2014
Date project completed (or expected completion date):	31 April 2017

DISCLAIMER

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

© Agriculture and Horticulture Development Board 2015. No part of this publication may be reproduced in any material form (including by photocopy or storage in any medium by electronic mean) or any copy or adaptation stored, published or distributed (by physical, electronic or other means) without prior permission in writing of the Agriculture and Horticulture Development Board, other than by reproduction in an unmodified form for the sole purpose of use as an information resource when the Agriculture and Horticulture Development Board or AHDB Horticulture is clearly acknowledged as the source, or in accordance with the provisions of the Copyright, Designs and Patents Act 1988. All rights reserved.

[The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.]

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

[Name] Dr Andrew French

[Position] Project Leader

[Organisation] University of Nottingham

Signature..... A. French Date.08/04/2015.....

[Name] Dr Nicola Harrison

[Position] Project Leader

[Organisation] East Malling Research

Signature..... N. Harrison Date.08/04/2015.....

[Name] Amy Lowe

[Position] Ph.D Student

[Organisation] University of Nottingham & East Malling Research

Signature A. Lowe Date 08/04/2015.....

Report authorised by:

[Name]

[Position]

[Organisation]

Signature Date

[Name]

[Position]

[Organisation]

Signature Date

CONTENTS

Grower Summary	1
Headline.....	1
Background and expected deliverables	1
Summary of the project and main conclusions	1
Financial benefits.....	2
Action points for growers	2
 Science Section	 3
Introduction	3
Materials and methods	7
Results.....	12
Discussion	20
Conclusions	24
Knowledge and Technology Transfer	25
Glossary.....	25
References	26
Appendices	27

GROWER SUMMARY

Headline

- This project is developing a system to detect indicators of stress in strawberry plant physiology using hyperspectral image analysis, a non-invasive technique.

Background and expected deliverables

Regular crop inspection by experienced growers and trained agronomists is an essential activity necessary to maintain the health and productivity of any crop. When signs of biotic or abiotic stress are detected within a crop, there is a high probability that the causative stimulus has already had an impact on the productivity of the affected plants and may potentially spread to the whole crop depending on what the cause is.

Finding ways to automatically detect adverse stress would be beneficial to the horticultural industry and could lead to a reduction in the volume of fertiliser and pesticides used through targeted applications, and early intervention would also reduce the damage caused to the crop. The physical parameters of a plant can be measured using automated phenotyping, which is an automatic high-throughput process that scans and analyses the physical features of a plant. This process decreases the time taken to measure the quantitative parameters of a plant, for example: leaf area or chlorophyll pigment concentration. Hyperspectral imaging is a recent tool in phenotyping and includes extra colour information that cannot be observed by the human eye or a digital camera.

The aim of this project is to use hyperspectral imaging to analyse and detect the onset of stress in strawberry plants in relation to certain diseases, pests or environmental conditions. Strawberry plants will be subjected to diseases, pests or drought in order to capture a time series of how the plants are responding to these biotic and abiotic stresses. The images will be collected using hyperspectral cameras and will include both spatial information, (the location of the pixels in the image), and spectral information, (the narrow bands of contiguous wavelengths from visible light to near infra-red light). The plants will be imaged at East Malling Research (EMR) and the University of Nottingham.

Once the images have been captured, the strawberry plants need to be identified in the images using a technique known as 'segmentation'. This means labelling objects within the image by finding similar properties such as colour, shape or texture. Once the plants have been located in the images, the hyperspectral information can then be analysed over time. Hyperspectral data contains important information related to physical processes, such as the chlorophyll content and the cell structure. For example, during drought the amount of

chlorophyll reduces which means a relative increase in light—specifically red wavelengths—being reflected by the plant instead of being absorbed.

This work is being done as part of the AHDB Horticulture Studentship scheme, designed to train new scientists to work in the UK horticultural industry. The scientist working on this project is Amy Lowe who is jointly supervised by Andrew French (University of Nottingham) and Nicola Harrison (East Malling Research).

Summary of the project and main conclusions

To date, eight strawberry varieties, with three replicates of each, have been grown at EMR then transferred to glasshouses at the University of Nottingham for imaging using the hyperspectral camera system. This initial dataset is a time series of images from well-watered strawberry plants through to just-visible onset of drought. The strawberry plants were imaged once a day over four consecutive days.

The images have been segmented using Non-negative matrix factorisation (NMF). This uses the hyperspectral information to label the image areas based on the different materials in the image. NMF can also discriminate between the leaves, flowers, stamen and overturned leaves, for example the matrix factorisation is by parts. When this uses spectral information it will identify the signatures within the input matrix. The matrix could also focus on features to separate the parts within the image such as the leaf veins.

Further image analysis techniques will be implemented in 2015 along with analysing the data to find stress markers. To build-on and enhance the pilot dataset collected in 2014, we will generate new datasets using the hyperspectral imaging system at EMR, with the aim of investigating strawberry responses to powdery mildew (*Podosphaera aphanis*), two-spotted spider mite (*Tetranychus urticae*) and drought.

In addition, further computational analyses will be undertaken with focus on image recognition of individual leaves on the plant and determining the orientation of the leaf to build a geometric model of the plant. This will allow analysis on the flat facing leaves and also to analyse the same leaf over time.

Financial benefits

For this annual report it is not appropriate to undertake a cost/benefit analysis.

Action points for growers

- There are no action points for growers at this stage of the project.

SCIENCE SECTION

Introduction

Current methods for detecting the health and stress of plants

One way of detecting stress is manually checking the plants for visible signs of stress. This involves growers and agronomists regularly checking their plants for symptoms of disease or adverse conditions such as drought or nutrient deficiencies. This method is time-consuming and costly.

As an alternative, the health of a crop could be automatically measured to increase efficiency and reduce costs to the grower. Automated phenotyping is a high throughput screening of the plants, measuring the observable properties such as the leaf area and leaf pigment content. A plant's phenotype can be described as observable complex traits influenced by the genotype and the environmental conditions (Dhondt, 2013).

Automated phenotyping can be expensive due to the advanced technology and technical equipment needed, although research is aimed at reducing the cost by using low cost cameras and phenotyping systems (Tsaftaris and Noutso, 2009). Using a conveyor system is not viable when the plants are difficult or impossible to move; instead multiple cameras can be used to capture plants as reported on peppers by van der Heijden et al, (2012). Hyperspectral imaging is a recent addition in phenotyping technology, which can measure extra colour information that cannot be observed by eye or with a standard digital camera.

Techniques for measuring the health of a crop also include vegetation indices. There are indices for detecting specific characteristics which combines the reflectance properties of two or more wavelengths; for example normalised difference vegetation index (NDVI) is used to detect stress caused by a cereal pest in wheat (Genc et al, 2008). There are over 150 vegetation indices so far but only a small number have been systematically tested (Verrelst et al, 2006).

Examples of stress signs

Common examples of causative agents of stress and their associated symptoms in the strawberry plants are described in Swift (2011) and include:

- Powdery mildew caused by a fungus (*Podosphaera aphanis*) shows some or all of the visible indicators: rolling of infected leaves, purplish or reddish blotches on leaves, the infected flowers and fruit will have a fine white covering of powder;

- Leaf spot is also caused by a fungus (*Mycosphaerella fragariae*) that creates spores and can stunt the growth of the plant. Typical signs are: purple or brown spots on the leaves and other areas of the plant;
- Two spotted spider mite (*Tetranychus urticae*) is a very small arthropod pest that infest a range of host plants. The infestation symptoms include bronze coloured leaves and eventually leaf death because the mites pierce cells in the leaves and extract the cell contents. If the infestation is severe it can reduce fruit production by up to 80% (Livinali *et al*/2014);
- A few visible signs of stress due to drought include overturning, curling and lighter coloured leaves.

There are different signs depending on the biotic or abiotic stress involved and the aim of this studentship is to determine whether we can detect indicators of stress before these signs are visible. This project will use hyperspectral imaging to analyse the spectrum range from visible to near infrared wavelengths. Further computational analysis will characterise the plants' 'healthy' spectral signatures to then enable the identification of 'unhealthy' spectral signatures that may indicate a particular stress. This could lead to a new, specific index being developed for strawberry crops.

Hyperspectral imaging

A hyperspectral camera captures the electromagnetic spectrum in contiguous narrow bands of wavelengths. A spectrum will show the intensity per wavelength from short to long wavelengths (UV to infrared respectively).

A hyperspectral image contains spatial and spectral information, which is represented in three dimensions. The spatial information is the pixels location along the x- and y-axis and the spectrum is the z-axis. This can be visually displayed in a hypercube (Figure 1) where pink is high reflectance and black is low reflectance. Each band of the spectrum can be displayed as a grey scale image that displays the intensity of the light. Figure 2 shows a select number of grey scale images at every 50th band from 1 to 400 along the spectrum range of 400nm to 1000nm.

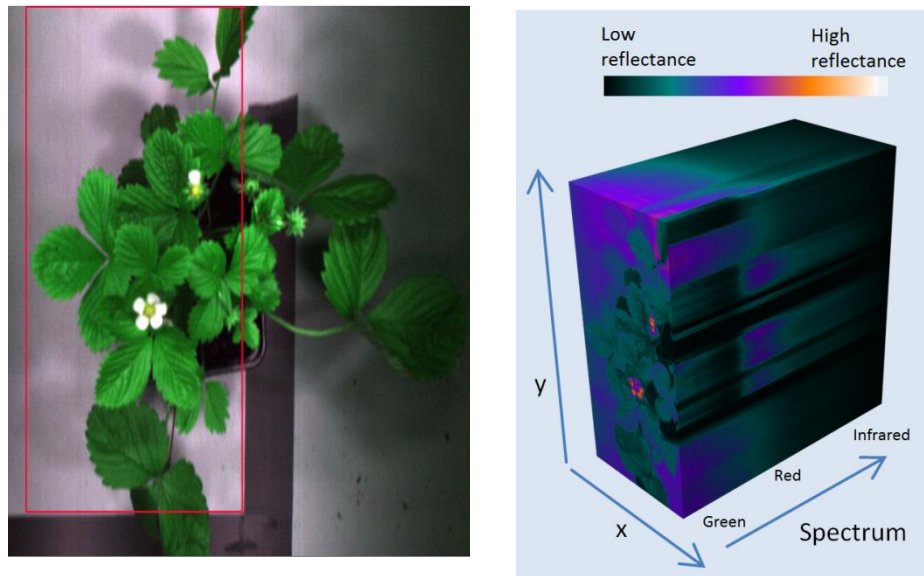


Figure 1 –Hypercube of the *Fragaria* sp. plant cv. Hawaii-4. Left: the RGB image of the cv. Hawaii-4 plant where the red box indicates the section that has been imaged. Right: the hypercube of the selected area. The x and y axis are the pixels in the spatial location and the spectrum is the z axis from green to near infra-red wavelengths.

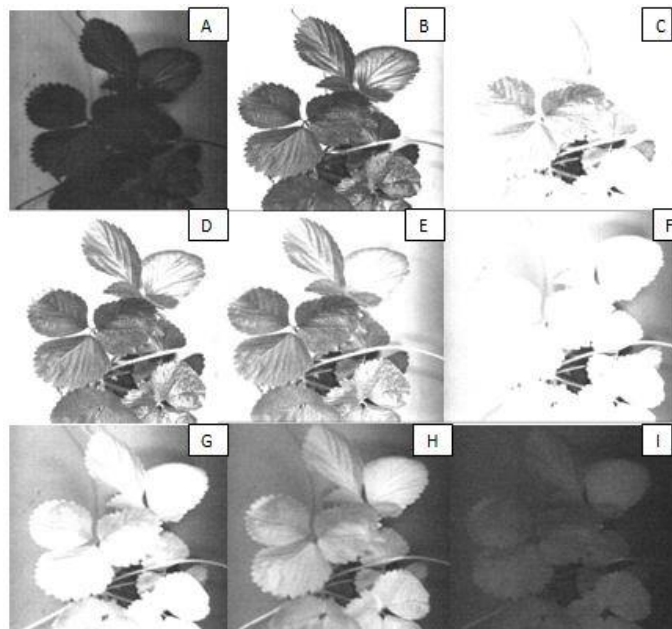


Figure 2 –Select greyscale images over the spectrum of cv. Redgauntlet. Redgauntlet replicate 1's hyperspectral images displayed in grayscale images at 50th band intervals over the spectrum where the grayscale images are the intensity of the light at a specific band: A is 399nm, B is 472nm, C is 546nm, D is 620nm, E is 694nm, F is 768nm, G is 842nm, H is 916nm and I is 990nm.

Healthy leaves and vegetation look green because the leaf pigments absorb most of the blue and red wavelengths (~450nm and ~650nm) while most of the green wavelengths (~550nm) are reflected back. The red and blue wavelengths are used for photosynthesis and creating chlorophyll. When the leaves become unhealthy there is less chlorophyll content therefore some red wavelengths are also reflected instead of being absorbed, which is why the leaves start to look yellow. In the infrared range, the wavelengths are longer which allows them to travel through the leaf to the cell structure, specifically the spongy mesophyll (Liew, 2008). More light is reflected back in the Infrared range than in the visible range of the spectrum for vegetation; however the human eye cannot see this. Figure 3 shows the spectral signature of a healthy strawberry cultivar Redgauntlet plant.

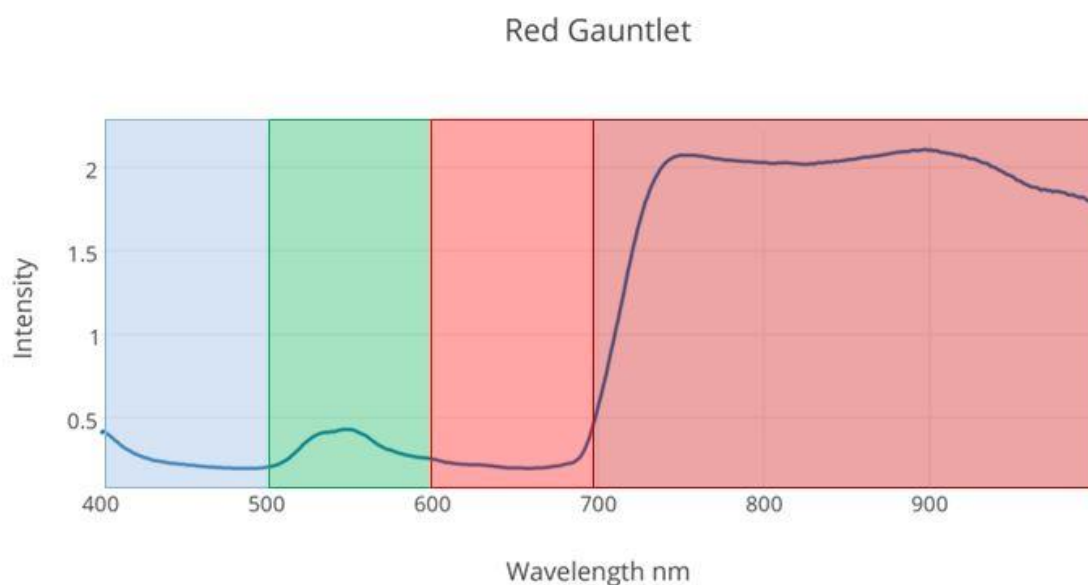


Figure 3 – The spectral signature of cv. Redgauntlet. The blue, green and red sections refer to the perceived colouring in the visible range, and the dark red box (700-1000nm) represents the near infrared range. The plot shows the reflectance of the light from the plant over the spectrum, known as a spectral signature. This data covers the range from 400nm to 1000nm and each band width is 1.48nm.

Previous literature on hyperspectral imaging of strawberry plants is limited. ElMasry *et al* (2006) used hyperspectral imaging and analysis to determine the quality of the internal attributes such as dry matter content and sugar content. Romer *et al* (2012) investigated early drought stress in cereal using hyperspectral image analysis. The method used is simplex volume maximisation (SiVM) which is a matrix factorisation technique that calculates the plants spectra in comparison to typical spectra to determine how stressed the plant is. This method indicated drought stress four days before established indices and five days before visual detection (manual detection).

Materials and methods

A selection of eight strawberry cultivars and related *Fragaria* species (six replicates of each) were grown at EMR and transferred to Nottingham in June 2014 where they were imaged with a hyperspectral camera in the laboratory at the University of Nottingham (Figure 4). The *Fragaria x ananassa* cultivars selected were: Elsanta, Elegance, Redgauntlet, Malling Centenary, Vibrant and Cambridge Favourite. The species selected were: *Fragaria vesca* Hawaii-4, *Fragaria chiloensis* BSP14 and *Fragaria virginiana* RH30. The initial experiment involved a time series of the strawberry plants, imaged consecutively over four days from 'well watered' through to just-visible onset of drought.

The camera at the University of Nottingham captures visible and near infrared (VNIR) wavelengths, which has the range 400-1000nm and is comprised of contiguous bands where each band is 0.74nm wide with a total of 812 bands. For the drought experiment, 406 bands were used (where each band covered 1.481nm) to reduce data size and decrease capture time.

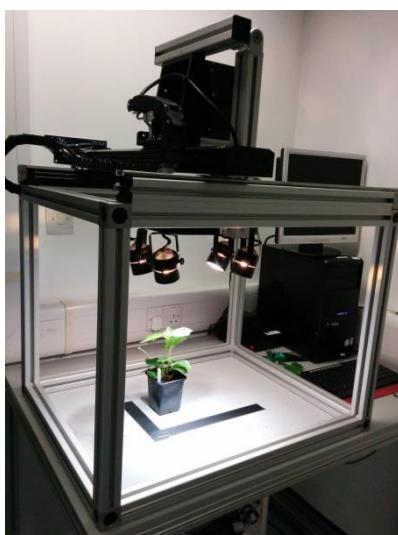


Figure 4 – Hyperspectral camera imaging a strawberry plant. The square frame holds the camera and the cables so the camera can move and capture the plant. The system has the capability to move the camera in the x-and y- axis (where x is left-right, and y is forward-back) but for imaging the strawberry plants the camera only needs to move along the -y axis.

Calibration

When using cameras and lighting they need to be calibrated to make sure that the images produced are not interfered with by the lighting or electrical noises. First, covering the camera lens or keeping the camera shutter closed records the dark images. Then the dark images are analysed to determine the internal electrical noises of the sensor in the absence of light. (Robles-Kelly and Huynh 2012)

A white board (white balance) is also imaged that is designed to reflect 99% of the light over the spectrum. This detects the non-uniformity of the illumination. The images can be normalised using the white balance to correct for the difference in the illumination. Without normalising the data with the white balance, the strawberry plants' spectral signature would appear low in the near-infrared range (Figure 5). This is known to be wrong, because healthy leaves reflect higher amounts of light in the near infrared range of the spectrum. The wavelengths are longer and the leaf pigment cannot absorb these wavelengths so they reach the mesophyll cells and reflect back, only a small portion is absorbed, (Liew, 2008), hence, image correction is a necessary step for subsequent analyses.

To normalise the data, the plant's spectral signature (Figure 5) is divided by the white balance (Figure 6), this produces the final spectral signature for analysis (Figure 7). Calibration corrects the spectra's signature of the strawberry plant, as can be seen in Figure 7, where there is now high reflectance in the near infra-red section of the spectrum as expected. The white balance corrects the reflectance problem in the near infrared because there is 99% reflectance through the spectrum.

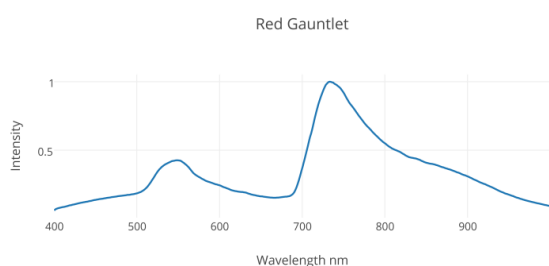


Figure 5 – Cv. Redgauntlet's uncalibrated signature over the spectrum where each band is the light reflected from the plant.

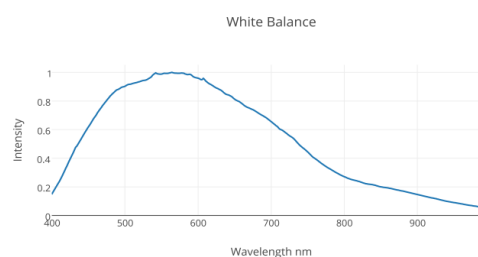


Figure 6 – White calibration target reflects 99% light back per band over the spectrum reflection (so the spectrum of the light bulbs is shown here).

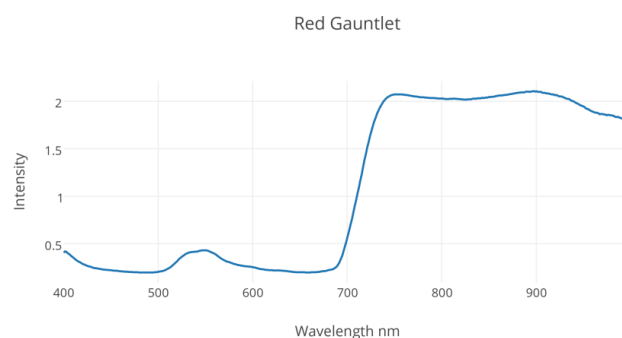


Figure 7– Cv. Redgauntlet’s spectral signature after calibration. The data in Figure 5 is divided by that in Figure 6 to calibrate the data.

Segmentation - Identifying the plants in the image

The plants need to be identified within the image. This is called image segmentation, which partitions the image into distinct regions where the pixels in the same region have similar features and if the neighbouring pixels are in a different region then they have different features such as: colour, intensity or texture. The regions represent the different labelled objects within the image.

Another feature that can be used is the hyperspectral information. The plant and the background have different spectral signatures because they reflect and absorb light differently along the spectrum. One approach to segmenting the hyperspectral images collected is known as ‘spectral unmixing’, which utilises this effect.

Spectral unmixing

This is the decomposition of a mixed pixel into distinct spectra or ‘endmembers,’ and the proportion of each endmember within the pixel. An endmember refers to a material such as plant, soil or background. If a pixel is identified as pure it means that only one endmember is contained in that pixel. Keshava and Mustard (2002) list the three steps for unmixing the hyperspectral data:

1. Dimension reduction (optional);
2. Endmember selection;
3. Inversion (proportion of each endmember within the pixel – fractional abundance).

Steps 2 and 3 can happen together, depending on the algorithm.

Dimension reduction takes the data and reduces the size by keeping the most important information. Hyperspectral data is usually large, so this process would reduce the size of

the data and increase the computation speed. For example each plant that has been imaged for the drought experiment has the dimensions (552x1004x406) which correspond to (y, x, z) respectively. There are over 200 million pixels that need to be processed and the amount of pixels would be doubled if there were 812 bands instead of 406 bands.

The endmember selection step finds all of the endmembers within the input data matrix. Then the inversion finds the fractional abundance matrix. Keshava (2003) lists the endmember and inversion techniques available

Spectral unmixing will either use the linear mixture model or non-linear mixture model depending on if the data is linear or non-linear. The data is determined to be linear if there is a linear combination of the endmembers weighted by the fractional area of each endmember in the pixel, i.e. the light interacts with only one component for each distinct endmember. The data is non-linear if the endmembers are mixed spatially and smaller than the path length of photons in the mixture. The light interacts with different components, as it is scattered onto multiple endmembers (Keshava and Mustard, 2002).

If the wrong mixture model is selected, the proportions of the endmembers within the pixel will be significantly wrong. Keshava and Mustard (2002) demonstrates the results of using a linear mixture model with non-linear data.

Linear mixture model

When the light reflects once off the surface of the endmember in the pixel then the endmember will have a clear signature. This means that the fractional abundance per endmember within the pixel will have a linear relationship and the endmembers abundance will sum to 1. A pixel in the observed matrix can be expressed as:

$$Y_{ij} = \sum_{a=1}^r W_{ia} H_{aj} + N_{ij} \quad (1)$$

Where $N \in R^{m \times n}$ accounts for additive noise. Then Equation 1 can be written as:

$$Y = W H + N \quad (2)$$

Where $Y \in R^{m \times n}$ is the hyperspectral input data matrix, m is the number of spectral bands and n is the pixels, and $W \in R^{m \times r}$ is the endmember's signatures where r is the number of endmember's and $H \in R^{r \times n}$ is the abundance matrix containing a fraction for each endmember per pixel. Also each column of H sums to 1, Tang W. (2012).

Non-negative matrix factorisation (NMF)

The input matrix is decomposed into two lower rank non-negative matrices containing the endmember signatures and the abundance fractions of each endmember per pixel. The input data matrix is a two dimensional matrix where the pixels spatial location is dismissed and the pixels are converted into a matrix to vector alignment.

$$Y \approx W H \quad (3)$$

Equation 3 uses Equation 2 and follows the linear mixture model. Matrices W and H are either initialised using random non-negative values or initialised using Vertex component analysis for more robust results (Rajabi and Ghassemian 2014). Also r (number of endmembers) is either already known, found using Virtual Dimensionality (Chang and Du 2004) or estimated/ trial and error. NMF uses a multiplicative update rule to keep the matrix values above zero and to minimise the matrices. There are a few update rules that can be used. Two examples are Lee-Seung (LS), shown in Equations 4 and 5, and Alternating Least Squares (ALS) (see Appendix A). Two minor problems with the update rules are that LS slowly converge and ALS can reach local minima and diverge (Soltuz et al 2009).

$$H = H((W^T Y) ./ (W^T W H)) \quad (4)$$

$$W = W((Y H^T) ./ (W H H^T)) \quad (5)$$

$$\sum_{r=1}^r H_{rn} = 1 \quad (6)$$

Equations 4, 5 and 6 are the 'LS' update rule used in Berne (2013) and Tang. (2012). After the first iteration the estimated matrices are tested for convergence using equation 5. This update continues until the matrices converge or the input iteration number is exceeded. The input data matrix (Y) and the estimated data matrices (WH) are compared to check the difference per pixel and if the convergence is below a certain value (i.e. <0.0001) it will stop. Frobenious norm, Equation 7, is a popular minimisation equation however other equations can also be used.

$$\frac{1}{n} ||Y - WH||_F^2 \quad (7)$$

However there may not be a unique solution since $WH = (WD)(D^{-1}H)$ for any nonnegative invertible matrix D which is explained in Jia S. (2009). This General NMF can be improved using sparseness and smoothness constraints.

Results

Cvs. Redgauntlet (replicate one, day two) and Hawaii-4 (replicate three, day one) were chosen to illustrate the successful application of NMF for finding the endmembers and abundance matrices.

Before testing how well the NMF works, the spectral signatures need to be found in the original images in order to compare the spectra (Figures 9 and 10).

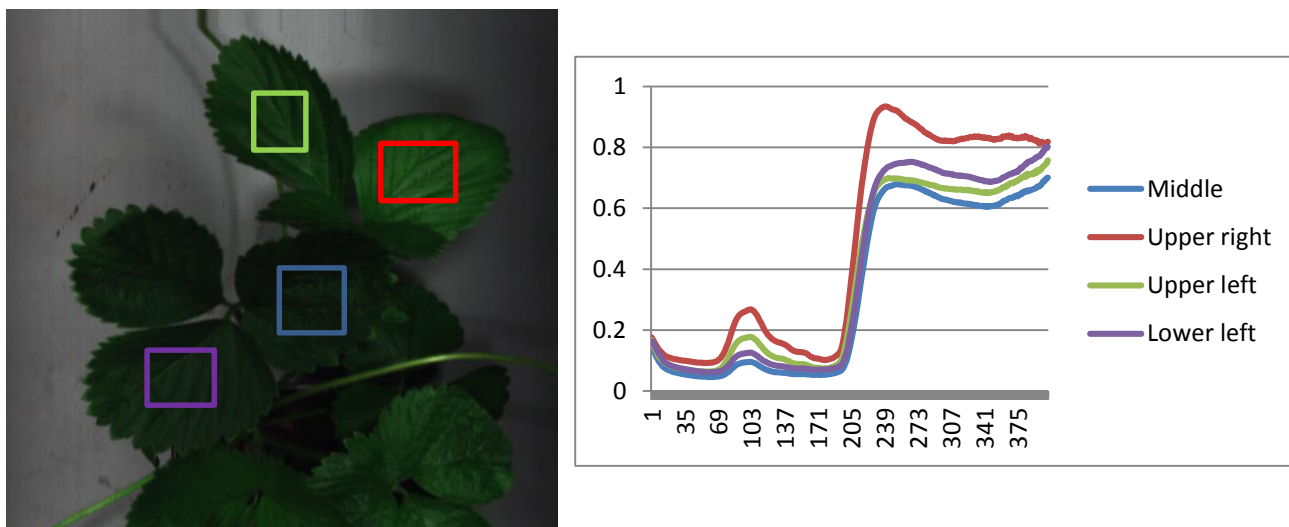


Figure 9 –Spectral signatures of strawberry leaves. The average reflectance value per square (left) is collected using FIJI, which is ImageJ (Schindelin et al, 2012), and then plotted in MS Excel (right). This displays the spectral signatures of the leaves to display the difference. The difference can be due to shadows and lighting since the lights did not reach the same intensity near the sides.

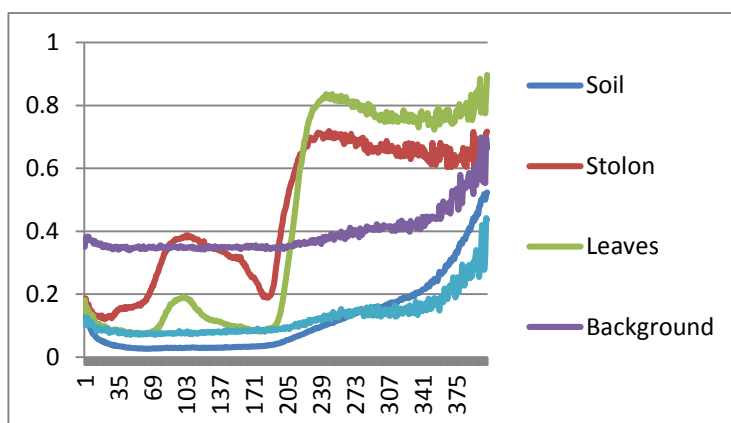


Figure 10 – The spectral signatures of the endmembers from the original hyperspectral images. The signatures are taken from the original data (after calibration) to compare with the NMF's estimated signatures. They were collected in FIJI and plotted in MS Excel.

Two repetitions of NMF using cv. Redgauntlet

With NMF, the endmember number either needs to be known or estimated. Applying NMF to the whole image stack will classify differences in the background as well as the plant. The matrix factorisation will separate endmember signatures depending on how different they are. NMF will be repeated twice, the first NMF is with two endmembers to find the plant and background. Then using just the plants information, NMF will be applied again with three endmembers to find the leaves, soil and stolon.

The reason the background is being removed first is to not blend in with other parts of the plant, such as the stolon. This technique only uses the spectral information and disregards the spatial information. By removing the background, the analysis is concentrated on the important sections of the image.

NMF 1 – Two endmembers

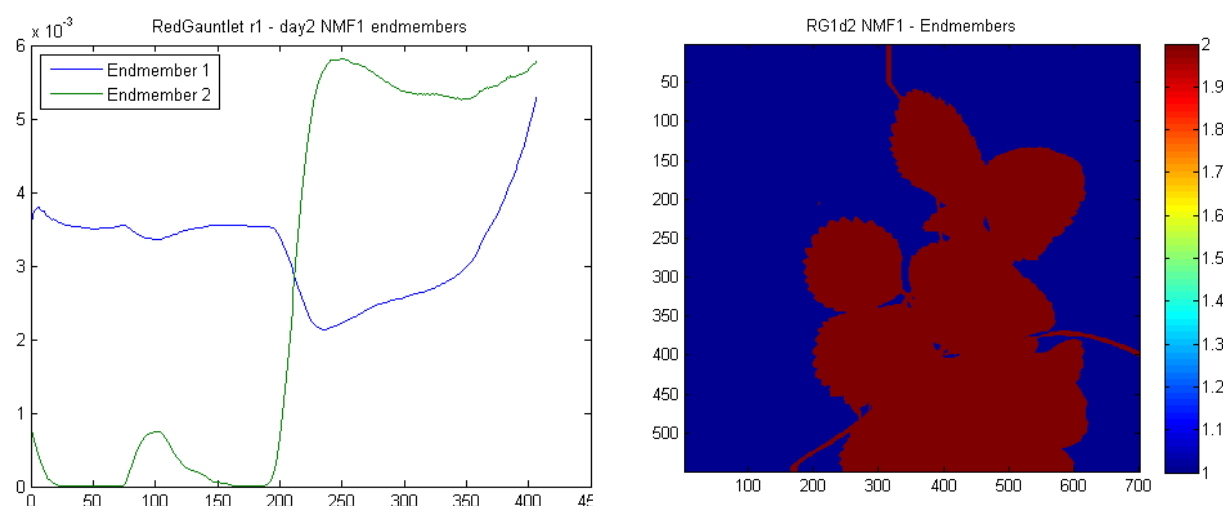


Figure 11 –NMF 1 using two endmembers produces the endmember signatures plotted (left) and the abundance matrix (right). Left is the spectral signatures produced by the NMF when two endmembers are the input parameter. Right is an image displaying the highest value endmember per pixel.

NMF produces two matrices W and H where W is the endmembers' spectral signatures (Figure 11 - left) and H is the abundance matrix (Figure 11 – right) however it is slightly modified to display the endmember in the pixel with the largest value, in a colour map. For the actual H matrix, the images are displayed in Figure 12.

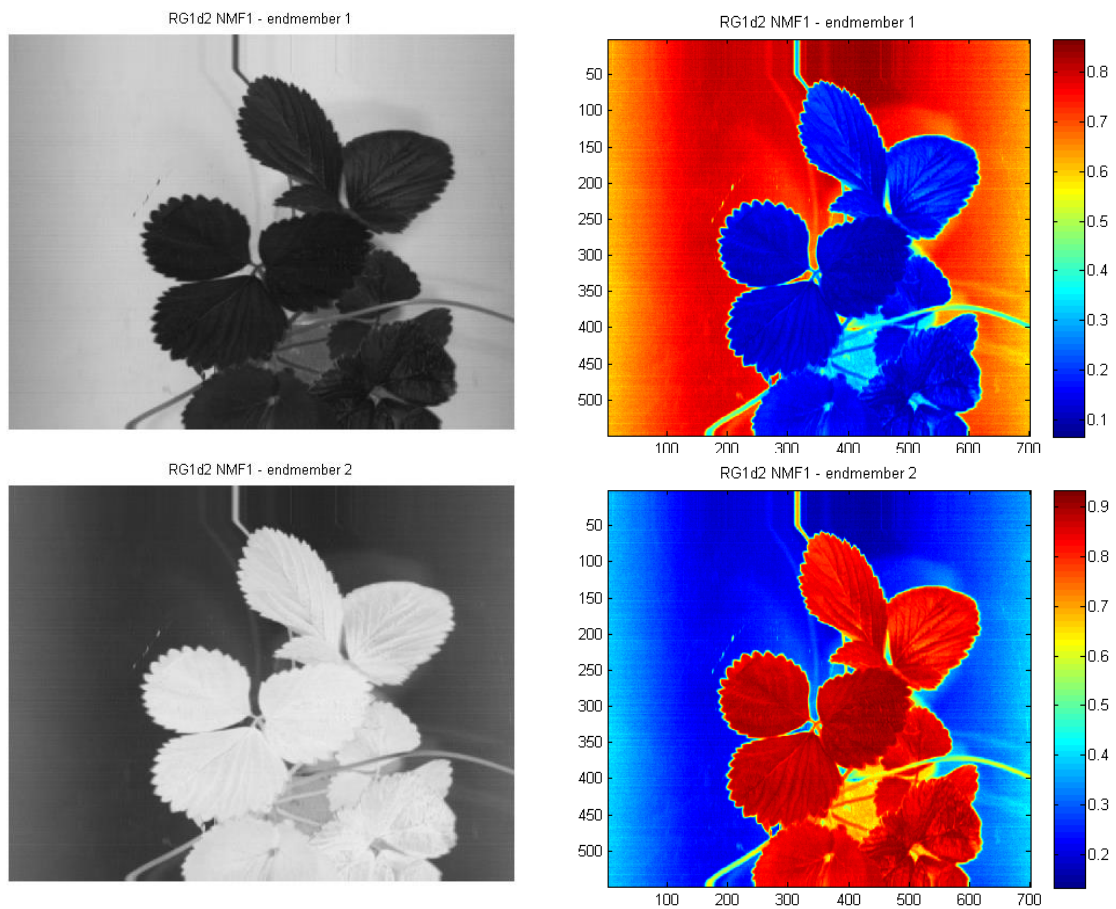


Figure 12 – Endmembers' intensity images and colour maps of the abundance matrix. Top left is Endmember 1 intensity image and top right is the colour map. Bottom left is Endmember 2 intensity image and bottom right is the colour map. If the image pixels are white/light grey then it is a part of that endmember, likewise if the pixels are black/dark grey then it is not a part of the endmember. The colour map displays a scale.

Mask for NMF2

The mask is created from the plant endmember 1 from the abundance matrix in NMF1. By setting the threshold value of endmember 2 (Figure 12, bottom right) to 0.5 this eliminates the background. This is because NMF1 only has two endmembers and there is a distinct difference. The plant endmember is used as the input matrix for NMF2. The mask is a logical matrix where the red is true and the blue is false (Figure 13). This means that once the new matrices are created from NMF2, this mask will be used to recreate the image.

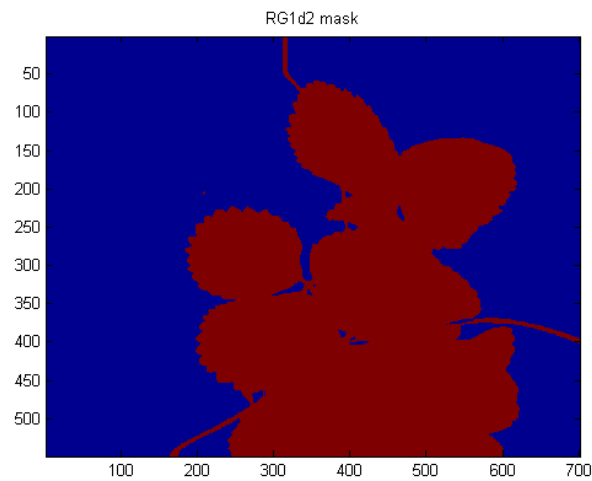


Figure 13 – A mask to keep the plant and disregard the background. This is a logical matrix with red pixels holding true values and blue pixels holding false values.

NMF2- Three endmembers

After removing the background NMF can now be applied to the plant endmember (and soil).

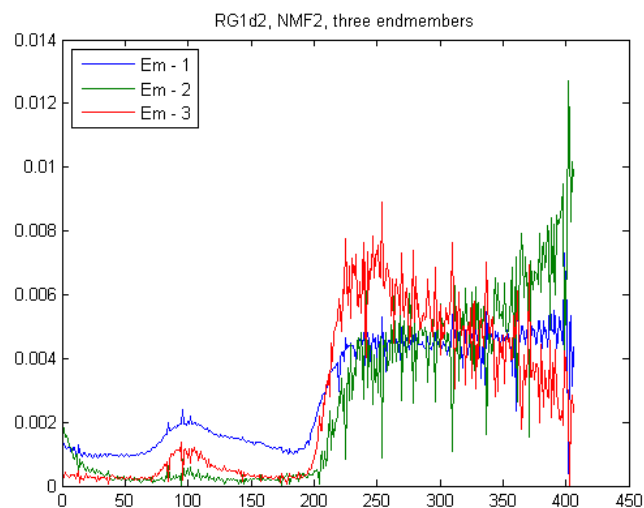


Figure 14 – Plot of endmembers (W) using three endmembers for the second NMF. The endmembers' spectral signatures are plotted in a graph to show the three endmembers.

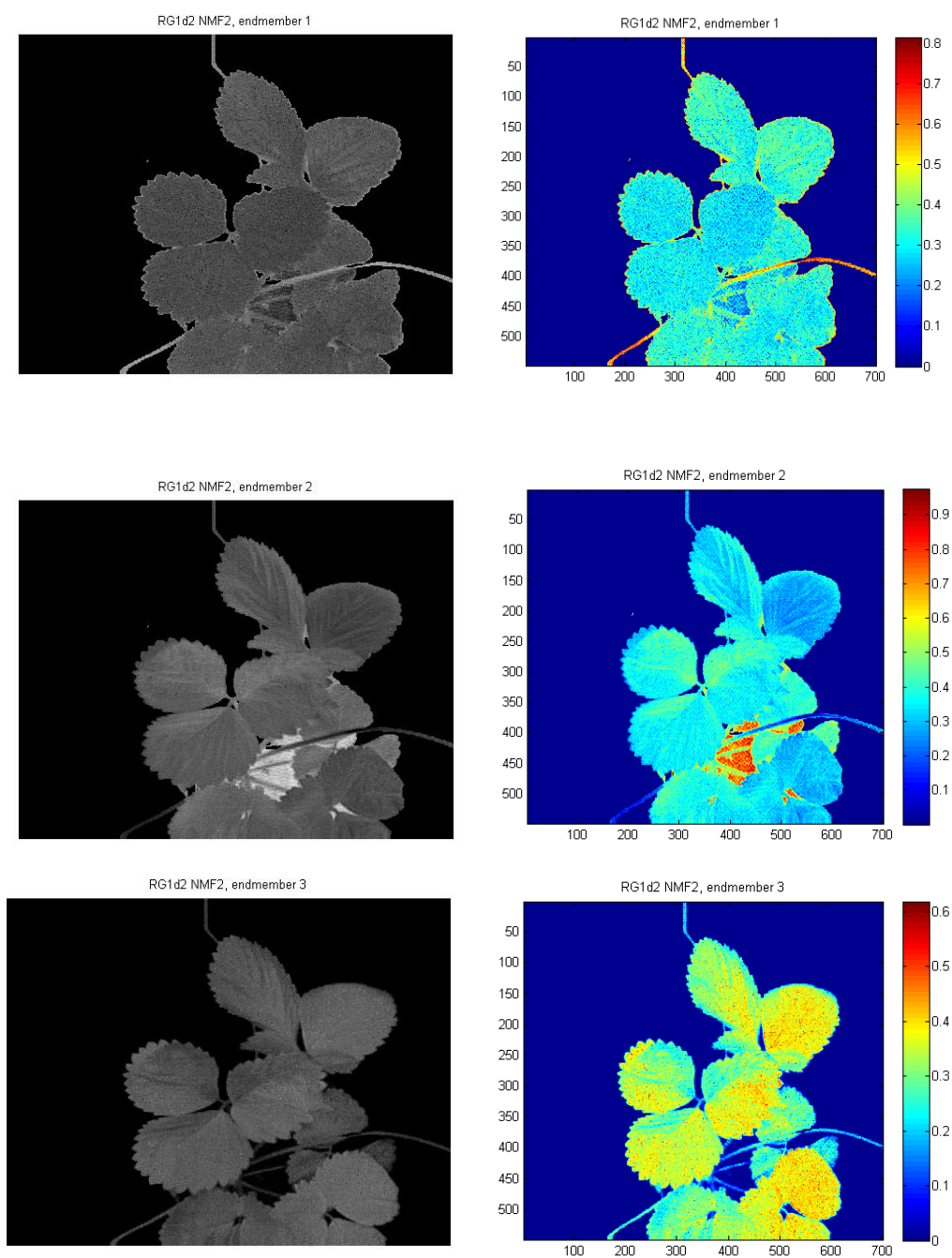


Figure 15 –The abundance matrix H represented in images. Top left is the intensity image for endmember 1 and top right is the colour map, middle left is the intensity image for endmember 2 and middle right is the colour map, bottom left is the intensity image for endmember 3 and bottom right is the colour map.

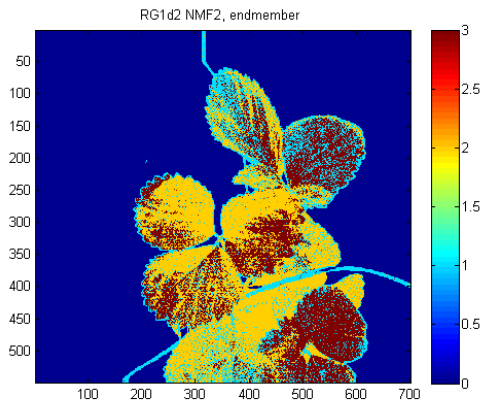


Figure 16 – A colour map of the endmembers. The endmember in the pixel with the largest value becomes the pixel and this is represented in a colour map where light blue, yellow and red are endmembers 1, 2 and 3 respectively.

Figure 16 is the abundance matrix but displayed as one endmember per pixel where the endmember with the highest value in the pixel is displayed as that endmember. However the pixel could be 0.49 endmember 1 and 0.51 endmember 2, this is just to visualise the endmembers together in one image. A good comparison of this would be to apply NFINDR and then compare the two images. NFINDR finds the pure pixels and not mixed pixels where NMF covers pure and mixed pixels but usually if the endmember is pure the pixel will probably still have a small amount of the other endmembers by the way the matrices are multiplied.

Hawaii-4, replicate-3, day 1

This analysis with cv. Hawaii-4 (replicate 3) will include the background because there are flowers in the image and applying a two endmember NMF would remove the background, partial leaves and also the flowers.

First Hawaii is going to be analysed with NMF using two endmembers, then three endmembers and finally six endmembers. This is a trial and error analysis to see what happens and also because the number of endmembers is currently estimated. From visually looking at the image there are definitely two different endmembers: 'background' and 'plant'. Also referring to Figure 17 the flower visually has a different signature and the soil as well. The amount of endmembers can be calculated using Virtual Dimension and other endmember detector algorithms which are currently being tested.

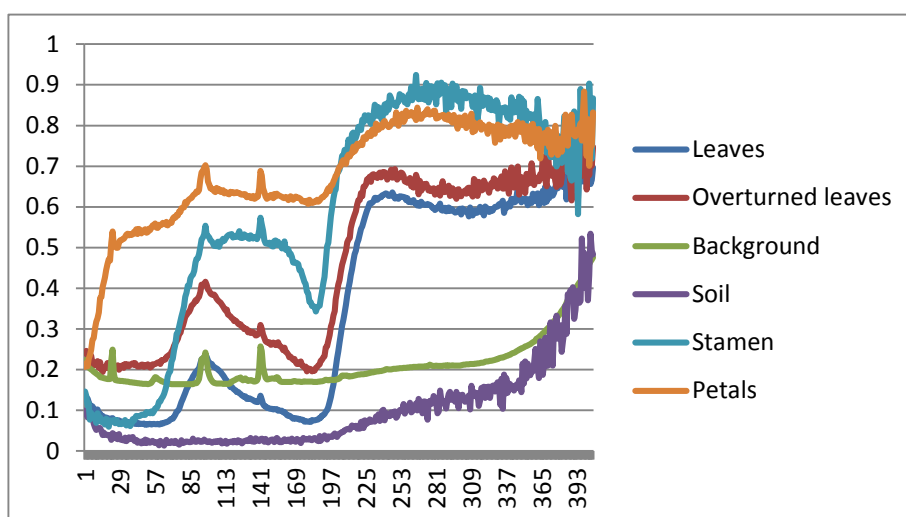


Figure 17 – The True endmembers' signatures collected from the data using FIJI. The plot shows the endmembers' signatures using FIJI to collect the average intensities and MS Excel to normalise and plot the data.



Figure 18 – A Greyscale and RGB image of cv. Hawaii-4 replicate 3, day 1 of the drought experiment. The left image is a greyscale image and the right image is an RGB image.

Two endmembers

This is NMF using two endmembers which will produce two endmember signatures and an abundance matrix containing two columns (corresponding endmember abundance).

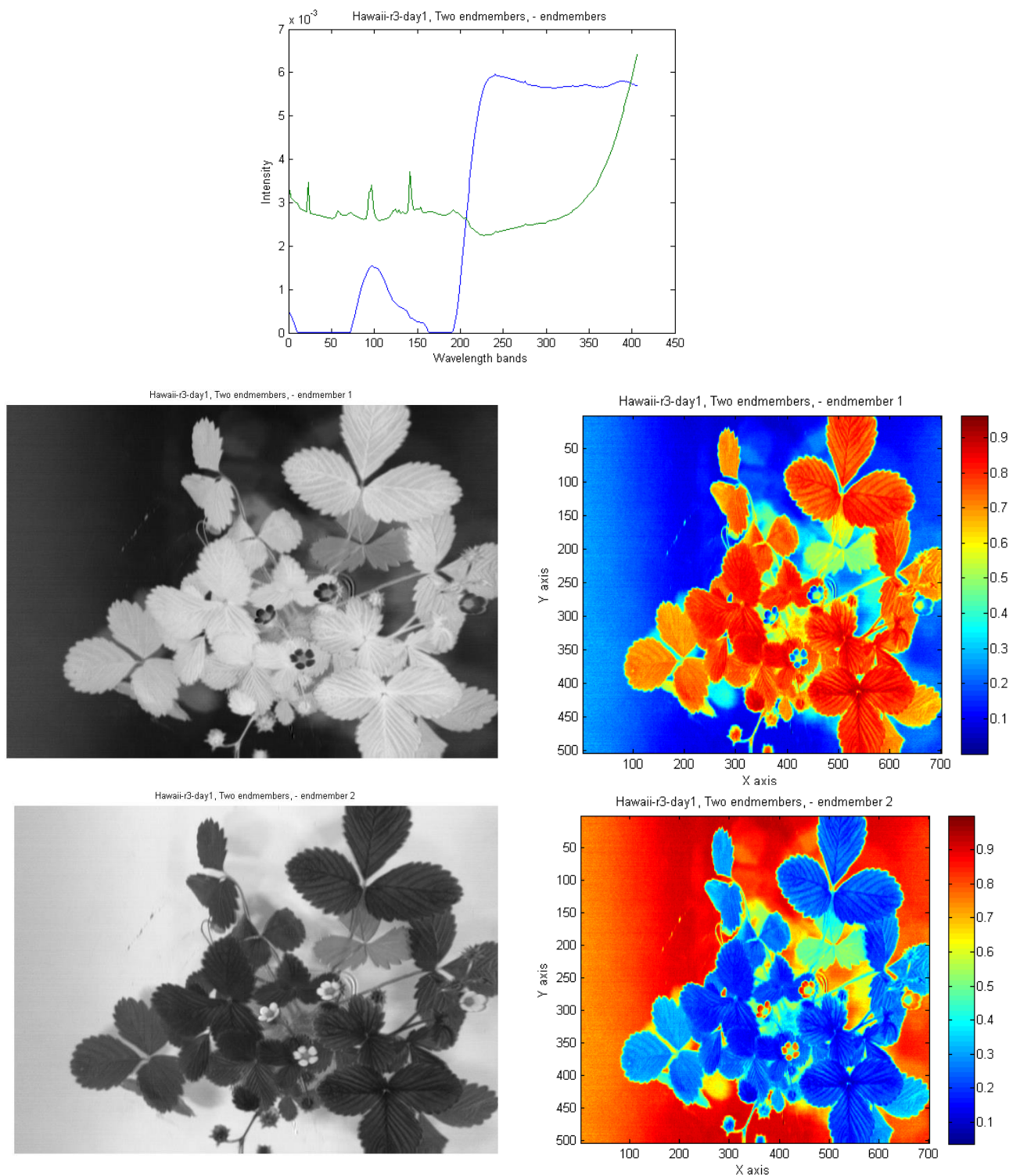


Figure 19 – The endmember signature matrix (W) and abundance matrix (H) with two endmembers. The top graph is the endmember signatures for two endmembers, middle left image is the intensity values in the endmember 1 row of H matrix and middle right image is the corresponding colour map. Bottom left image is the intensity values in endmember 2 and bottom right is the corresponding colour map.

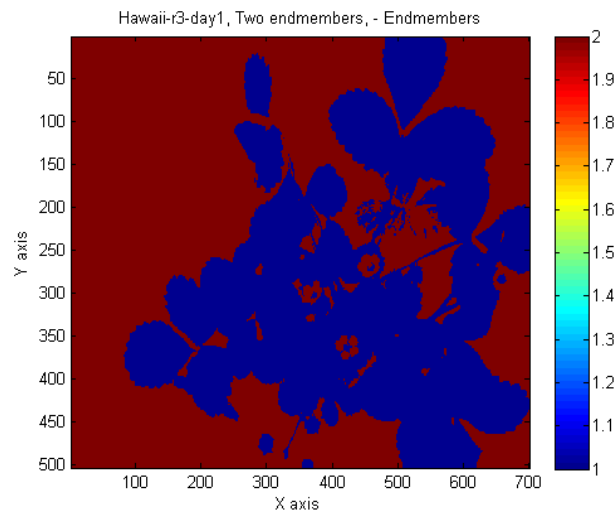


Figure 20 – A colour map displaying the highest value endmember per pixel. The endmember with the highest value in the pixel is displayed in the pixel. The colour map displays the division of the two endmembers. Part of the overturned leaf and the petals are being classified as the background.

Three endmembers selected for NMF

This time three endmembers will be selected to see what happens. If the endmember number selected is below the minimum amount of endmembers then the estimated signatures produced from the NMF will combine together. This is partly why they will not look like Figure 17's true signatures.

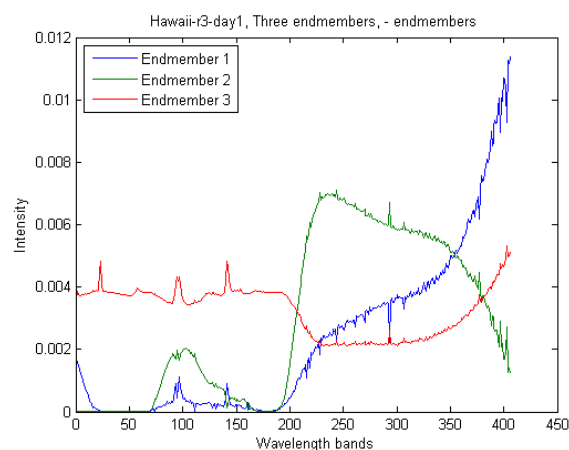


Figure 21 - Endmember signatures of matrix W using three endmembers. The endmembers spectral signatures are plotted in a graph. Endmember 1 represents the leaves, Endmember 2 represents the background and Endmember 3 represents the soil/underneath the plant.

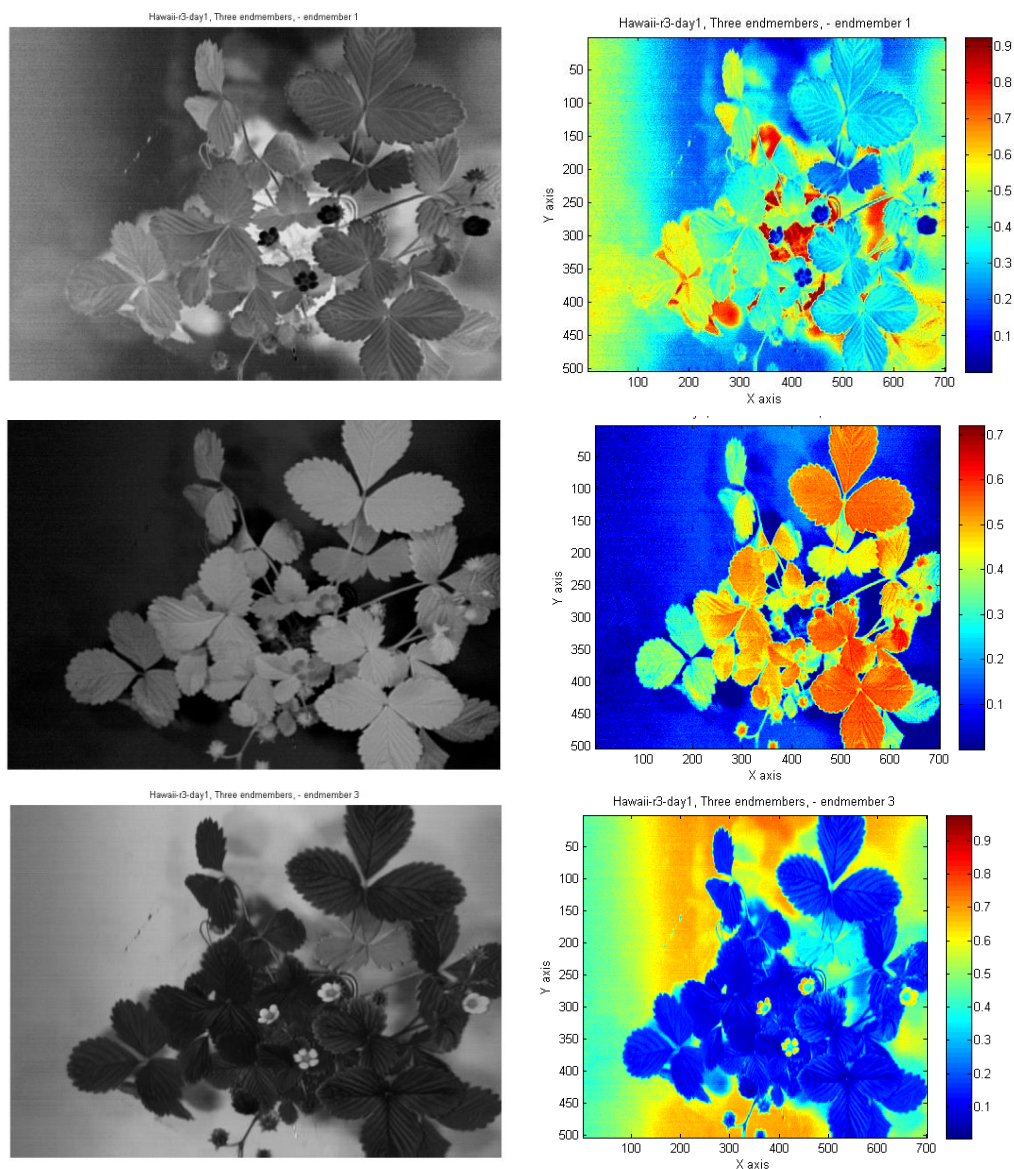


Figure 22 - Abundance matrix H with three endmembers can be visualised as three greyscale images where each image is focused on a different endmember. Top left is the intensity value image for endmember 1 and top right is the corresponding colour map. Middle left is the intensity value image for endmember 2 and middle right is the colour map. Bottom left is the intensity value image for endmember 3 and bottom right is the colour map.

From looking at Figure 21 the three endmembers appear to be: endmember 2 is the leaves; endmember 3 is the background and partially the flowers; and endmember 1 is the soil/underneath the plant. There appears to be some crossover as the leaves are blending in with endmember 1 and 2.

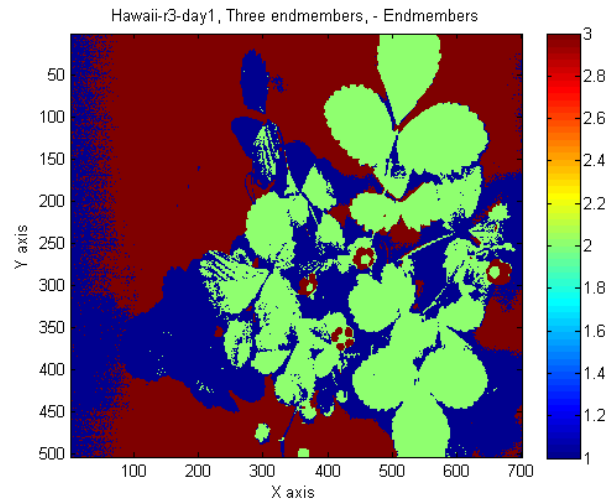


Figure 23 – The endmember with the highest value in the pixel is displayed in the pixel. The colour map displays the division of the three endmembers.

The largest endmember value in each pixel can be represented in Figure 23, however the difference could be so small that another NMF would display that pixel as a different endmember. Figure 22 are images from the actual abundance matrix produced from the NMF.

Overturned leaves

Although overturned leaves are not displaying as an endmember in the abundance matrix with the highest value pixel becoming that endmember (i.e. Figure 23) there is visually a difference of 0.1-0.2 on the colour map in Figure 24. No analysis has been performed yet to test if the overturned leaves can be found using this technique.

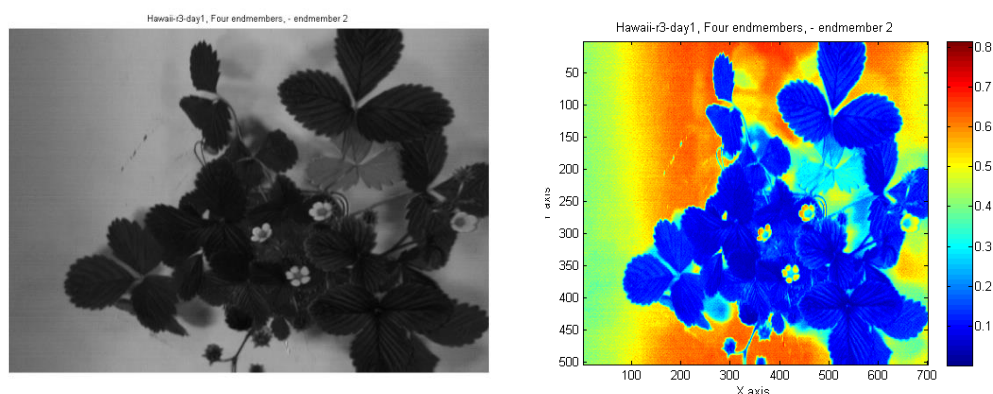


Figure 24 – Endmember 2 greyscale image showing the intensity of the light. The colour map is the right image showing the scale from 0 to 0.8. No analysis has been performed yet to detect overturned leaves.

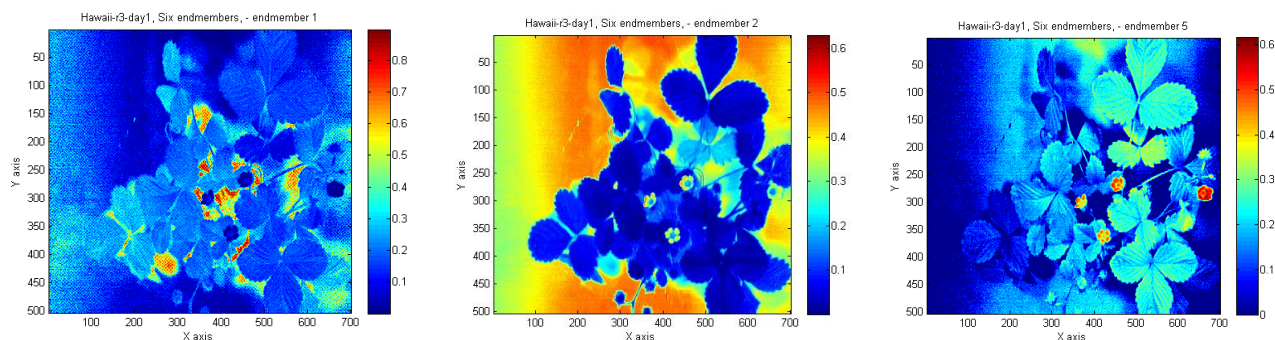


Figure 25 – Three colour maps produced from six endmember NMF to show the intensities. Left image displays the soil and shadows endmember. Middle image displays the background prominently with the overturned leaves in the middle of the range bar. Right image displays the flowers.

Six endmembers

NMF with six endmembers resulted in three noisy endmembers and three endmembers displaying the background, soil/shadows and the flowers. There are too many images and graphs produced as more endmembers are selected. Therefore Figure 25 only displays three colour maps which are for three of the endmembers.

The amount of endmembers needed was estimated for this report; however the actual number of endmembers will be used in the future using Virtual Dimensionality. By estimating the amount of endmembers, it showed how the non-negative matrix works in finding the endmembers, and when a smaller number is chosen the other endmembers combine. For example, in the cv. Hawaii-4 section in Figure 19 the plant and soil are combined into one endmember, however when three endmembers are selected the soil is a separate endmember (Figure 22 – endmember 3). This happens with the flowers; in the sections with two and three endmembers the flowers are classified with another endmember however in Figure 25 when six endmembers are selected the flowers are a separate endmember.

Discussion

The image analysis using NMF produces two matrices, where W contains the endmember signatures and H contains the abundance proportions of the endmembers per pixel. This separates the data into meaningful information that can be analysed once NMF is improved.

Firstly, by looking at the endmember signatures for cvs. Redgauntlet and Hawaii'-4, the results produced by NMF's W matrices are similar to the spectral signatures taken from the original images using FIJI. There are small differences between the manually calculated

spectral signatures and NMF's estimated spectral signatures for example cv. Redgauntlet displays a dip in the background's signature which is not in the original data, this is also present with cv. Hawaii-4 background signature. The estimated signatures can be tested against the true signatures for accuracy.

However NMF is not that effective at segmenting the images. When the image contains just a simple background and the plant then it is easy to distinguish them apart, as demonstrated for cv. Redgauntlet, however in the segmentation for cv. Hawaii-4, the presence of flower petals led to the misclassification of the petals as background. This is because there is no spatial information being used. Finding the plant first and then applying spectral unmixing will be more useful because the analysis can be applied to the leaves, petioles or flowers rather than on the pixels disregarding the location.

Segmentation using shape-based techniques such as active contours are currently being developed and implemented. Also segmentation techniques using features such as the veins are being developed.

The objectives for year 2:

- New data collection at EMR with different diseases and pests (powdery mildew, two spotted spider mite and drought) using varieties that have both similar and dissimilar phenotypes;
- Continue implementing different shape-based segmentation techniques and focus on the features;
- Identify the leaves and orientation to build a geometric model of the plant;
- Analyse the plants over time using the hyperspectral information to detect changes. Use Spectral unmixing techniques by improving NMF and also using a simplex based algorithm.

Conclusions

NMF separates the original images into meaningful data that includes the spectral signatures and the corresponding abundance fractions, which can be improved by initialising W and H using Vertex component analysis (VCA) and also by using extra constraints on the algorithm such as smoothness and sparseness.

However, NMF ignores the spatial location of the pixels and classifies part of the plant with the background especially if the background is noisy or includes other plants. Therefore shape-based segmentation is needed before applying NMF or other unmixing techniques.

Knowledge and Technology Transfer - 2014

- Study visit to EMR June-July 2014 to collect data and transfer strawberry plants for the drought experiment.
- Visit to Hugh Lowe Farms, June 2014, with Marion Regan, industry representative for this studentship.
- AHDB Horticulture conference 16-17 September 2014, poster session and tour. Awarded 2nd Prize in the poster competition.
- Poster presented at the International Workshop on Image Analysis Methods for the Plant Sciences (IAMPS, Aberystwyth, September 2014)
- EMRA/AHDB Horticulture Soft Fruit Day 26 November 2014 at EMR. Dissemination of progress in this studentship through a poster exhibition to both industry and academia.

Glossary

Endmember – An endmember refers to a material such as plant, background, soil etc.

Endmember signature – the material's reflected light over a selected range of the spectrum, for this project the range is from visible light to near-infrared light.

Hypercube – All of the hyperspectral bands collected together to represent a 3D image where the first two dimensions are the pixel location and the third dimension is the hyperspectral information (wavelength bands over the spectrum).

Hyperspectral – narrow contiguous bands of light (wavelengths) over a spectrum

Projection (mentioned in VCA's definition) – This is a mathematical term which transforms the points and lines onto a plane, which is then treated as a two dimensional surface.

Non-negative Matrix Factorisation (NMF) – This refers to an algorithm that takes the input data (image) with non-negative values and produces two matrices. The first matrix contains the endmembers signatures and the second matrix contains the proportion of each endmember per pixel.

Vertex Component analysis (VCA) – This algorithm finds the minimum number of endmembers in the data by projecting all of the pixels in a random direction and the largest projection is considered as the endmember. This is repeated to find all of the endmembers

and each new endmember is compared with the previous endmembers to make sure they have not already been found.

References

- Berné O**, Boulais A, Deville Y, Joblin C, Pety J, Teyssier D, Gerin M, (2013) "NMF-based spectral unmixing of Herschel observations of the horsehead nebula", Proceedings of the 5th Workshop on Hyperspectral Image and Signal Processing (WHISPERS 2013), Gainesville, Florida, USA, 25-28 June 2013.
- Chang, C. I.**, and Du, Q. (2004). Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(3), 608-619.
- Dhondt, S.**, Wuyts, N., and Inzé, D. (2013). Cell to whole-plant phenotyping: the best is yet to come. *Trends in plant science*, 18(8), 428-439.
- ElMasry, G.**, Wang, N., ElSayed, A., and Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, 81(1), 98-107.
- Genc, H. et al (2008). Vegetation indices as indicators of damage by the sunn pest (Hemiptera: Scutelleridae) to field grown wheat. *African Journal of Biotechnology*, 7(2).
- Jia S** and Qian Y (2009) "Constrained Nonnegative Matrix Factorisation for Hyperspectral Unmixing" *IEEE transactions of geosciences and remote sensing*, Vol 47, No 1, pg 161-173, January 2009.
- Keshava, N.** and Mustard, J. F. (2002). Spectral unmixing. *Signal Processing Magazine, IEEE*, 19(1), 44-57.
- Keshava, N.** (2003). A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, 14(1), 55-78.
- Liew, O.W**, Chong, P.C.J.; Li, G.; Asundi, A.K (2008) Signature optical clues: emerging technologies for monitoring plant health. *Sensors* **2008**, 8, 3205–3239
- Livinali, E.**, Sperotto, R. A., Ferla, N. J., and Souza, C. F. V. D. (2014). Physicochemical and nutritional alterations induced by two-spotted spider mite infestation on strawberry plants. *Electronic Journal of Biotechnology*, 17(5), 193-198.
- Rajabi, R.**, and Ghassemian, H. (2014). Spectral Unmixing of Hyperspectral Imagery Using Multilayer NMF.
- Robles-Kelly, A.**, and Huynh, C. P. (2012). *Imaging spectroscopy for scene analysis*. Springer Science & Business Media, pg14

Römer, C., Wahabzada, M., Ballvora, A., Pinto, F., Rossini, M., Panigada, C. and Plümer, L. (2012). Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis. *Functional Plant Biology*, 39(11), 878-890.

Schindelin, J.; Arganda-Carreras, I. and Frise, E. et al. (2012), "Fiji: an open-source platform for biological-image analysis", *Nature methods* **9(7)**: 676-682.

Soltuz, S. M., Wang, W., and Jackson, P. J. (2009). A hybrid iterative algorithm for nonnegative matrix factorization. In *Statistical Signal Processing, 2009.SSP'09.IEEE/SP 15th Workshop on* (pp. 409-412). IEEE.

Swift C.E (2011) *Strawberry Diseases*, Colorado State University.<http://www.ext.colostate.edu/pubs/garden/02931.html> [accessed: 16.02.15]

Tang, W., Shi, Z., and An, Z. (2012). Nonnegative matrix factorization for hyperspectral unmixing using prior knowledge of spectral signatures. *Optical Engineering*, 51(8), 087001-1.

Tsaftaris, S.A. and Noutsos, C. (2009) Plant phenotyping with low cost digital cameras and image analytics. In *Proceedings of the 4th International Symposium on Information Technologies in Environmental Engineering*. Berlin: Springer, pp. 238–251.

van der

Heijden G, Song Y, Horgan G, Polder G, Dieleman A, Bink M, Palloix A, van Eeuwijk F, Glasbey C (2012) "SPICY :towards automated phenotyping of large pepper plants in the greenhouse." *Functional Plant Biology* 39, 870–877.

Verrelst, J., Koetz, B., Kneubühler, M., and Schaepman, M. (2006). Directional sensitivity analysis of vegetation indices from multi-angular Chris/PROBA data. In *ISPRS Commission VII Mid-term symposium* (pp. 677-683).

Appendices

Appendix - A

Alternating Least Squares (ALS) equations:

$$H_{n+1} = (W_n^T W_n)^{-1} (W_n^T Y) \quad (A)$$

$$W_{n+1} = Y H_{n+1}^T (H_{n+1}^T H_{n+1})^{-1} \quad (B)$$

The matrices may not be non-negative because of the inverse. After equation A and B if W or H (pixel values) are less than zero make that pixel zero otherwise keep the current value. (Soltuz, 2009)