Technical University of Crete
Electronic and Computer Engineering Department

Hyperspectral Imaging and Spectral Classification Algorithms in Plant Pathology

Anthony Sgouros

Thesis Committee

Costas Balas, Associate Professor (supervisor)
Michael Zerbakis, Professor
Kostas Kalaitzakis, Professor

Chania, December 2008

Abstract

Hyperspectal imaging acquires a plurality of narrow-band images across the visible and the non-visible part of the spectrum. In the obtained dataset the intensity re-emitted from the object under analysis is recorded as a function of both wavelength and location. In other words a full-spectrum can be resolved and calculated per image pixel. We have employed a hyperspectral camera to study variations of chlorophyll and carotene content under various plant stress conditions. The acquired spectra are classified using spectral Angle Mapper and Nearest Neighbour. The software used to perform the analysis is based on Python for GUI, data reference and engine API and a combination of Pyrex and C for the computation of the main routine. Key features of the software are its fast computation in comparison to optimized Matlab code (first implementation), threads to take advantage of multicore CPUs, fully customizable User Interface, multi-platform. The spectral classification algorithms return quantitative parameter expressing the similarity degree between spectra. On the other hand the same parameter can be used as a threshold to cluster and to map areas with similar spectra characteristics. The developed algorithms have been validated in plants under different stress conditions (including nutritional and virus stress). It has been shown that the combination of hyperspectral image acquisition with spectra classification algorithms detect and map efficiently alterations in plant chromophores (chlorophyll, carotene). This finding highlights the potential of the developed technology in plant pathology and remote sensing.

Acknowledgements

First of all i would like to thank my family for their invaluable support through the years of my studies.

I would like to thank my supervisor professor Costas Balas for his support and guidance through my thesis. He entrusted me an unknown field and the challenge broadened my mind and scientific background.

I would like to thank Antonis Makris of Mediteranean Agronomical Instistute of Chania, for the support on the experimental procedure. My colleagues, Phd Candidate Athanasios Tsapras for his support and suggestions and George Epitropou for sharing his knowledge on classification techniques.

Special thanks goes to all my friends and people I've met through the years of my studies in Chania. I was fortunate that the journey to becoming a better person, was not a lonely one.

Contents

1 Preface **1**
1.1 Problem Definition 3

2 Hyperspectal Imaging and Classification **5**
2.1 Physics 5
2.2 Hyperspectal Imaging 10
2.3 Hyperspectral Imaging Classification 12

3 Light and Plants **19**
3.1 Photosynthesis and Chromophores 19
3.2 Plant Tissue and Interaction with Light 27
3.3 Remote Sensing in Plants 30
3.3.1 Quality assessment of food 31
3.3.2 Vegetation Identification 33
3.3.3 Plant Pathology 36
3.4 Hyperspectral Imaging and Virus 38

4 Materials and Methods **42**
4.1 Hyperspectal Imaging System 42
4.2 Experiment Procedure - Protocol 43
4.3 Classification Algorithms, Chlorophyll-Carotene Indices 47
4.3.1 Classification Algorithms 47
4.3.2 Vegetation Indices 51
4.4 Defining the Reference Vectors 53

5 Results and Assumptions **55**
5.1 Results 57
5.1.1 Tobacco plants 58
5.1.2 Mulberry Leafs 64
5.1.3 Band Reduction Experiment 67
5.1.4 Noise Removal 70

5.1.5 Veins Classification 72
5.2 Assumptions 72

6 Appendix A **76**

List of Figures

2.1 Electromagnetic spectrum 6
2.2 Light interactions with matter 7
2.3 Absorption-Emission 7
2.4 Jablonski diagram 8
2.5 Second Jablonski diagram 8
2.6 Absorption-Emission in atomic system 9
2.7 Spectral cube and signatures 11
2.8 2nd example of spectral cube 12
2.9 Spectral Signatures of various materials 13
2.10 Summary of the differences between various types of Spectal Imaging 13
2.11 Spectal data analysis steps 14
2.12 Steps in supervised and unsupervised classification methods . 17

3.1 The procedure of Photosynthesis. 20
3.2 Absorbance spectra of plant's chromophores 21
3.3 Color of leaves 21
3.4 Chlorophyll α and β absorption spectra. 22
3.5 Molecular structures of Chlorophyll α and β 22
3.6 Leaf structure. 25
3.7 Leaf structure and kinetics during photosynthesis. 25
3.8 Interaction of light and plants 28
3.9 Leaf spectra response 29
3.10 Reflectance of light by leaf in three phases. 29
3.11 Satellite Spectral Imaging 31
3.12 French fries spectral response example 33
3.13 Food quality evaluation based on CCD cameras. 34
3.14 Various vegetation indices tested by Sims and Gamon 35
3.15 Spectra response of a plant in different damage levels. 37
3.16 Chaerle et al. results 38
3.17 Chaerle et. al thermal imaging results 39
3.18 Chaerle et al. measurement setup. 40

3.19 Scopoletin and Salicylic Acid 41

4.1 Spectral angle example 49
4.2 Block diagram of implemented algorithms 52

5.1 Explanation of a Classifier result figure. The colorbar is explained, as value 1 is a healthy class, value 6 is a damaged class and intermediate scale constitute the progress classes. . . 56
5.2 Explanation of a mND result. The colorbar is explained, as we proceed to higher values, higher concentration 56
5.3 First Mulberry leaf, RGB and 4 bands 57
5.4 Reference Vectors (Tobacco Plants) 58
5.5 Reference Noise and Veins vectors (Tobacco Plants) 58
5.6 Difference between vein and leaf 59
5.7 All reference matrix, tobacco 59
5.8 1st Leaf Tobacco results, SPAM 60
5.9 2nd Leaf Tobacco results, SPAM 61
5.10 3rd Leaf Tobacco results, SPAM 62
5.11 2nd Leaf Tobacco, SPAM-NEAR comparison 63
5.12 Reference Vectors (Mulberry leaves) 64
5.13 1st Mulberry Leaf results, SPAM 65
5.14 3rd Mulberry Leaf results, SPAM 66
5.15 Band reduction 1st mulberry leaf, SPAM 68
5.16 Band reduction 1st mulberry leaf, Nearest Neighbour 69
5.17 Noise removal fyllo1 NEAR 70
5.18 Noise removal fyllo1 SPAM 70
5.19 Noise removal fyllo2 NEAR 71
5.20 Noise removal fyllo2 SPAM 71
5.21 Veins Fyllo1 72
5.22 Veins Fyllo2 72

List of Tables

3.1 Hyperspectral Imaging application fields in plants. Differences and similarities. 30

4.1 Measurements summary, stage one 45
4.2 Measurements summary, stage two 45
4.3 Measurements summary, stage three 46
4.4 Proposed method for extracting reference vectors from a given dataset, related to leaf senescence 54

Chapter 1

Preface

Our world is full of light and colors.

The existence of colors is the simplest form of interaction between light and matter. We understand light and color through vision, which is our most basic sense. Color differences is the simplest and most basic perceptual criteria when categorizing things in a basket. Light travels through space and When it hits the surface of an object, it interacts with it. That interaction is the base of many physical sciences that try to describe all forms of responses. Understanding light and its properties was crucial, and many theories evolved in an attempt to give answers for all the phenomena that took place in the world and in scientific experiments. So light can be seen both as an electromagnetic wave and as a particle.

Color of objects is formed through the absorption and reflection of parts of the visible spectrum and is something that our eye can catch, so a question emerges as whether there are other forms of interactions between light and matter that our eye can't see. What if our eye can only catch a portion of the light's spectrum?

To move one step forward, what if we could extend the visible part of the spectrum by technical means so that our eye can *see* more things. Many systems exist and various science evolved to provide the means for extending our vision beyond the visible part of the spectrum. Extending our vision also implies researching on other aspects of light-matter interaction, depending on which part of the spectrum is chosen as a *diagnostic tool*. The absorption and transmittance of light is the base of many known imaging systems that all share key principles, although they utilize different bands of the spectrum.

One such system and thus evolving science (through its specific applications) is Spectral Imaging. We could claim that Spectral Imaging is the kind of sensor system, where the spectra of an object is the excitation of the system, and thus provides information for the object. One key factor that distinguishes Hyperspectral imaging from other Imaging technologies [1], is the part of the spectrum that is utilized in applications. We refer to wavelengths of 300nm to 1000, 1200 or 2500nm. Waves of such wavelengths have small amounts of energy and we can easily simulate a similar excitation to physical light with technical means, using common lighting gear and equipment. This is important for medical and other applications where targets should not be radiated in contrast to X-ray (radiation), or MRI (highly sophisticated equipment).

Hyperspectral Imaging is a field of great importance due to its discrimination ability, it can be applied with day light or technical source of light with small amounts of energy, so it does not affect the tissues or specimens under investigation. It provides a vast source of information and besides the challenges found in evolving the technology itself, there are many fields where Hyperspectral Imaging can be applied such as mineralogy, medicine, agriculture, sea sciences.

Using a Hyperspectral Imaging System on plants, we aim at first on clarifying our perceptual knowledge regarding plants health status. We know that a plant is healthy through the color of their leaves. It is common to think that the greener the plant is, the healthier it is. On the same time, color acts as an indicator when something is wrong, and informs the gardener, farmer or agriculturalist. Then, they only realise the damage providing the color difference between health and damaged leaves, is notable enough. Experience and constant monitoring is crucial to be able to identify in time the danger through the visible symptoms and apply the necessary treatment. The color of the leaves if formed due to the absorption of certain wavelenghts of light, because of the plant's chromophores. Chlorophyll and carotenes are the most basic chromophores and their role in the process of photosynthesis is to be clarified later on. Since color of the leaves is linked to chlorophyll and carotenes and these chromophores are crucial to the process of photosynthesis, we could assume that there is a link between health status and chromophores concentration. Also we might be able to identify various species and states as youth, maturity and senescence through the optical properties of the chromophores.

[1] By imaging technologies we refer to MRI, X-ray etc.

Finally the ability to take measurements remotely and in situ, highlights the prospects of applying Hyperspectral Imaging in plants.

Our study suggest that Hyperspectral imaging is successful on estimating the chlorophyll content while depicting on the same time the distribution of the plant's vital compounds on a leaf. Despite estimating the progress of total concentration of chlorophyll given a damaging factor, we can also depict the dynamics of the stress on leaf level. Providing characteristic reference spectra we could identify the stress factor. This capability is available by implementing classification techniques in the acquired spectra. Further less we examine the minimum bands required for a Hyperspectral Imaging system to provide rigid information regarding the plants physiology.

1.1 Problem Definition

We briefly list the problem definition, although some aspects and terms are to be clarified later on the text.

We conduct an experiment described in sec.4.2, to acquire a dataset from plants suffering various types of stress, using the Hyperspectral Imaging System described in sec.4.1. We proceed in analyzing the dataset collected, implementing two classification algorithms and two vegetation indices both described in sec.4.3. Based on the proposed methodology of sec.4.4 for acquiring and defining reference vectors and the software developed, we experimented on the dataset collected. Our research focused on:

- Investigating the diagnostic capabilities of classification techniques on a plant's spectral cube.
- Investigating the best classification algorithm [2] in terms of diagnostic accuracy and consistency.
- Application of vegetation indices to monitor stress effect.
- Comparing the classification techniques to vegetation indices.
- Investigating the minimum number of bands required for classification techniques compared to vegetation indices.

What follows:

[2]Comparison between the two algorithms implemented

Chapter 2. We describe the physical phenomena on which Hyperspectal Imaging is based, the technology itself and Classification methodology to provide our mathematical background.

Chapter 3. We describe the fields of Plant Research in combination with imaging technologies and other methods to model, and define chlorophyll content, stress factors and mechanisms.

Chapter 4. We dive into detailed information regarding our HyIS, the Chlorophyll content indexes used, the protocol of our experiment and the classification techniques implemented to analyze our spectra.

Chapter 5. Results of our experiments and analysis and the concluded assumptions.

Appendix A. Technical Information on the software used.

Chapter 2

Hyperspectal Imaging and Classification

This chapter covers the physics background in light-matter interaction and all the basic principles of spectral imaging. Further investigation regarding different kinds of spectral imaging is also covered. We also introduce the field of classification as an analytical technique to extract vital information from the collected spectral data.

2.1 Physics

Light is a form of electromagnetic wave. As such it can be also seen both as a particle and as wave as Einstein indicated. The key features of electromagnetic wave are the frequency and the wavelength. The energy content of an electromagnetic wave of given frequency and wavelength can be estimated with the formula:

$$E = h \cdot f \tag{2.1}$$

$$\text{and since } f = \frac{c}{\lambda} \tag{2.2}$$

$$E = h \cdot \frac{c}{\lambda} \tag{2.3}$$

where h is Planks constant, f is the frequency of the wave, λ is the wavelenght of the wave and c is the speed of light.

All possible different wavelenghts consist the whole electromagnetic spectrum, as seen in fig. 2.1. We then cluster the electromagnetic spectrum to various bands, by means of energy content, application specifics and the interactions implied by the wave attributes (frequency, energy content, etc.).

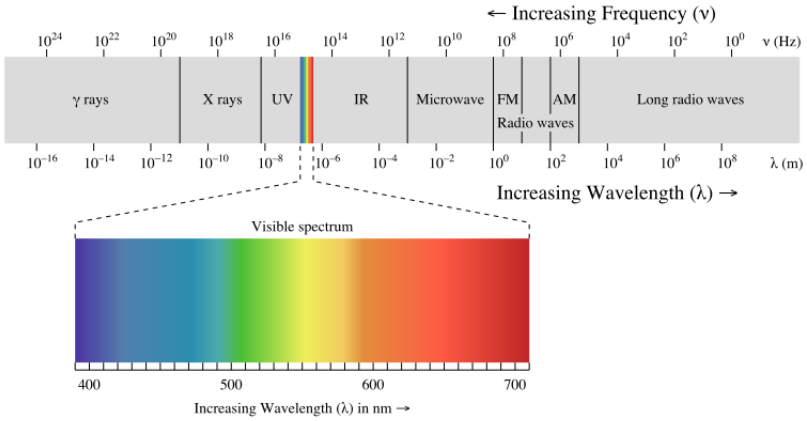


Figure 2.1: Electromagnetic spectrum and divisions to different bands, based on frequency of the wave and possible uses.

When light travels through space without any obstacles it has no loss, and can be formulated with given equations 2.1 to 2.3. When travels through different matters or objects, four basic interactions occur. We have absorption, reflection, emission and scattering of light as depicted in fig. 2.2. Each physical phenomena, depends on the frequency (or wavelength), the properties of the medium (molecule structure etc), and duration in time each interaction holds.

Phasmatoscopy is the scientific field that examines the interaction of electromagnetic radiation with matter. One of the first experiments that indicated the interaction between light and matter, was the discovery of the photoelectric effect by Hertz.

Absorption - Emission

To describe absorption we regard an atomic system that is in ground state. We name that state as $E0$. Given enough excitation by means of radiation, light, or heat, an electron from $E0$ can jump to the next energy level, providing that the energy absorbed is equal or greater than the energy gap between $E0$ and $E1$. That phenomena is called absorption. When in ground state the atom's tendency is to remain in low energy levels. So after a while, when an amount of energy is absorbed and there is a transition from ground state to an excited one, the electron returns back to the ground state while emitting a photon of energy similar to the energy gap between the two states.

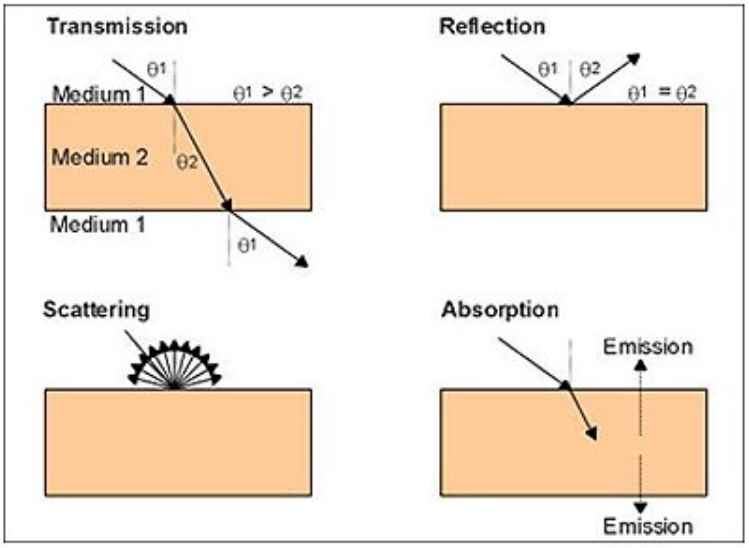


Figure 2.2: When light hits the surface of another medium, the depicted phenomena takes place.

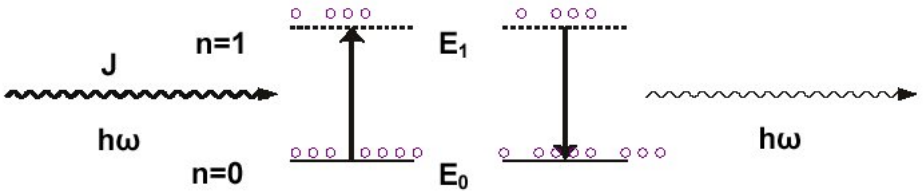


Figure 2.3: Absorption and emission of a photon from ground state to excited state. A photon of energy $J = h \cdot \omega$ (ω is frequency of the photon) is absorbed and the electron jumps from $E0$ to $E1$. It then returns to ground state. (n stands for basic quantum number)

The transition from an excited state to ground state, is not straight forward. It may be accompanied with emission of a photon but it can also return to lower energy states, without any radiation. Taking into account these facts and the duration of the emitting radiation, we give different definition for each phenomena. So by **luminescence** we refer in general to the transition of an electron to lower states while emitting radiation. There are two discrete types of luminescence, named as **fluorescence** and **phosphorescence**. The main difference between the two, is the duration of the

emmiting photon. Also there is a difference in the total number of steps required to return to ground state. The greater the gap, the larger the energy and thus the duration of the emitted photon. Nonetheless, many times an electron returns to ground states, following some intermediate states. These phenomena are depicted using the Jablonski diagrams.

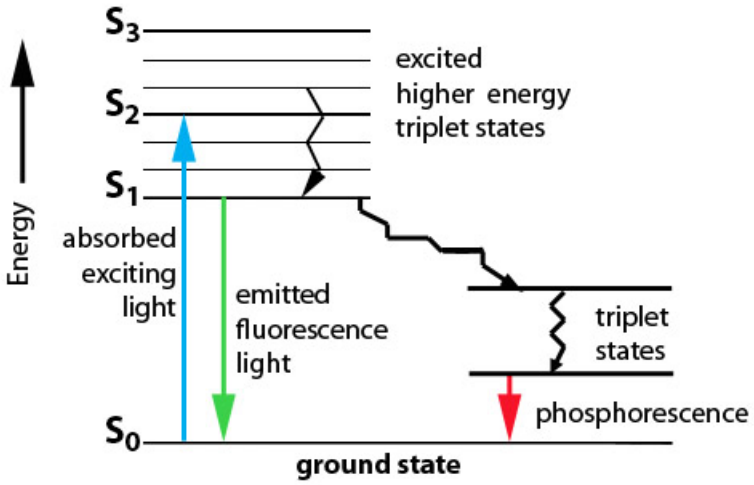


Figure 2.4: This figure known as Jablonski diagram, depicts the absorption and emission

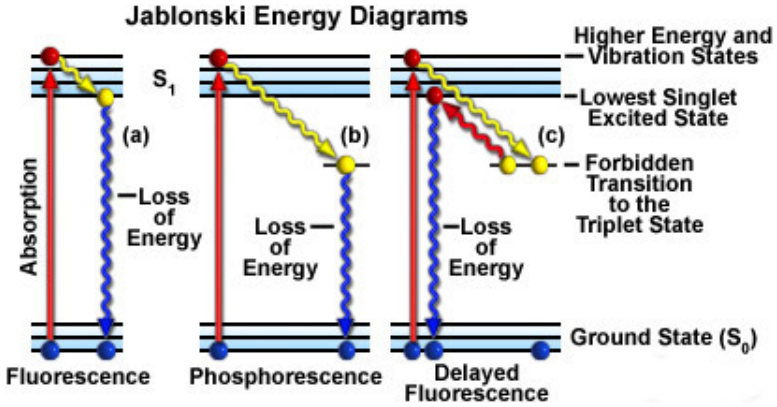


Figure 2.5: Second Jablonski diagram

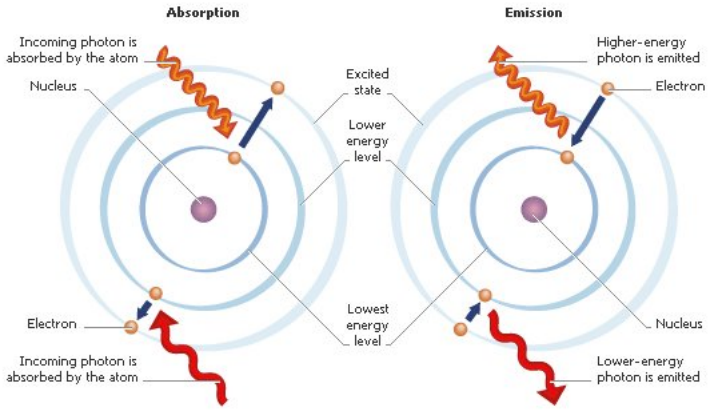


Figure 2.6: This figure demonstrates the absorption and emission in an atomic system.

Scattering

Scattering of light is the phenomena where one system absorbs the energy of a photon and then re-emits the photon with different attributes. These attributes consist of the wavelength, phase, direction and polarization and may differ from the radiation entering the system. Scattering depends mainly on the nature of the system, thus its material, its size in relation to the wavelenghts of the entering radiation, and its relative orientation in space.

Due to the altering factors of scattering, there is a categorization depending on the dominant factor, and so we have the following types of scattering:

- Rayleigh scattering
- Raman scattering
- Mie
- Stokes and Anti-Stokes
- Brillouin

2.2 Hyperspectal Imaging

We can talk about spectral image analysis as an extension of spectroscopy. Spectroscopy is the study of light as a function of wavelength that has been transmitted, emitted, reflected or scattered from an object. Chemical bonds within the object under investigation absorb light energy at specific wavelengths. The variety of absorption processes and their wavelenght dependency allows us to derive information about the chemistry of the object.

Hyperspectral Imaging was a breakthrough in remote sensing applications. It is widely used in agriculture, mineralogy, physics and sea sciences. A Hyperspectral Imaging system collects and processes information from across the electromagnetic spectrum.

A Hyperspectral Imaging system consists of a camera or CCD sensor, band-pass filters to allow particular segments of the spectrum to pass in, lighting gear and a computer to gather the data and perform the analysis on the spectral data acquired. The breakthrough of the Hyperspectal Imaging System is the evolution of hyperspectral sensors being able to combine the spatial resolution of imaging sensors and the spectral resolution of spectrometers.

Hyperspectral Imaging, consists of repeated foto stills of a target stacked together to form what is known as spectra cube. Each slice of the cube is a band-limited light response of a target to a certain wavelength. So when a pixel is chosen, in the two-dimensional space, we also get the spectra of that pixel as the z-dimension. In other words, we could say: the acquisition of images in hundreds or registered, continuous spectral bands such that for each picture element of an image it is possible to derive a complete reflectance spectrum. The nature of imagery data is typically multidimensional, spanning three spatial, on spectral and one temporal dimensions. Each point in this multidimensional space is described by the intensity of the radiance which is emitted, reflected, or a combination of both (depending on the phenomenology under investigation)

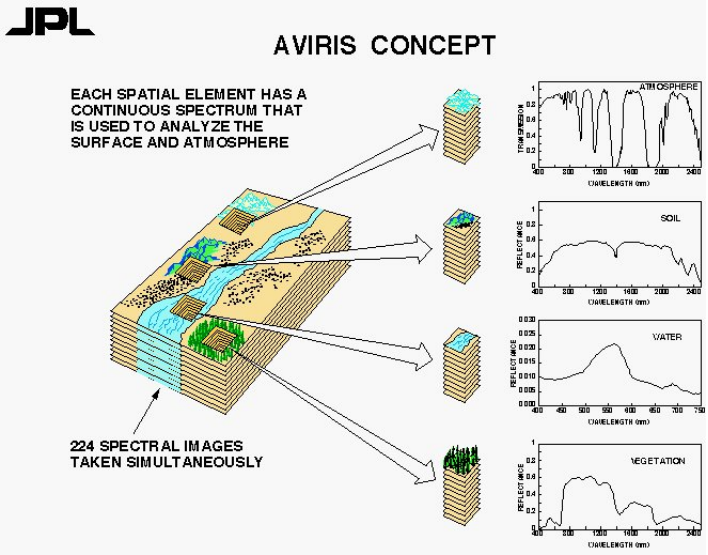


Figure 2.7: Example of a spectral cube and spectral signatures of different materials.

Spectral image cubes are analogous to a stack of pictures of an object, a sample or a scene, where each image is acquired at a narrow spectral band. Each pixel in the image cubes, therefore, represents the spectrum of the scene of that point, as shown in fig.2.7 and fig.2.8.

Different materials have different molecular structure, and so their interaction with light results in different spectra for each material. So it might

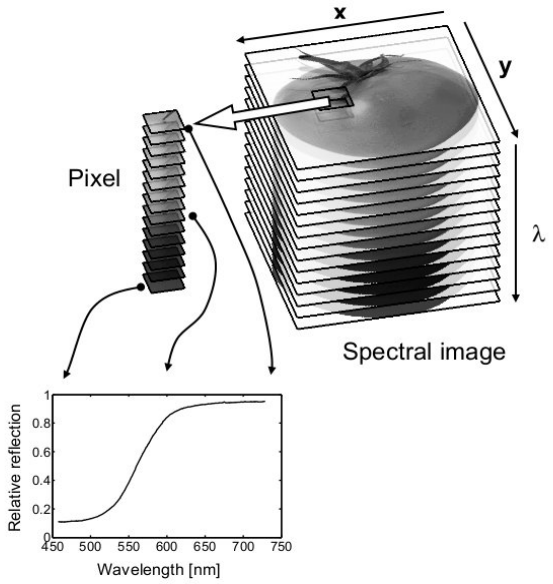


Figure 2.8: A spectral image is a three dimensional data cube with two spatial dimensions (x and y) and one spectral dimension (λ) resembling the reflection spectrum at every pixel.

be possible (providing the appropriate equipment) to have unique spectra responses for each kind of materials. That spectra constitutes a spectral signature for known materials such as soil, water, vegetation, etc, as shown in fig.2.7 and fig.2.9.

Hyperspectral Imaging systems need high resolution in space and spectra. We can also discriminate between the number of bands that consist the cube. Less bands lead to multi-spectral systems, that have band-limited filters with wider bandwidth compared to hyperspectral imaging, causing poorer spectral resolution. Hyperspectral imaging utilizes narrow-bandwidth providing high spectral resolution. In fig.2.10 we summarize the basic differences between different types of Spectral Imaging.

2.3 Hyperspectral Imaging Classification

Hyperspectral Imaging and all related technology provides a vast amount of information. The process from acquiring the spectral cube to extracting specific rigid information about an object or target is not straightforward, as

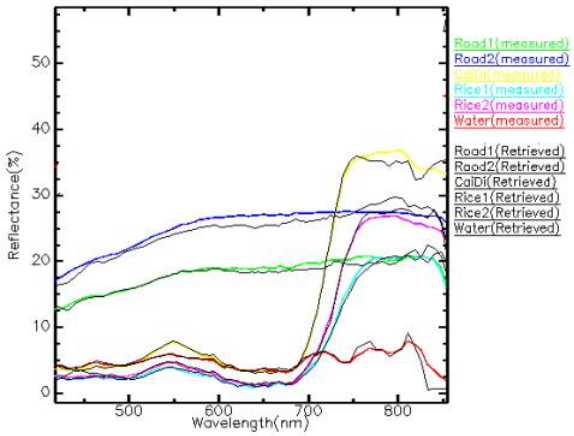


Figure 2.9: Spectral Signatures of various materials

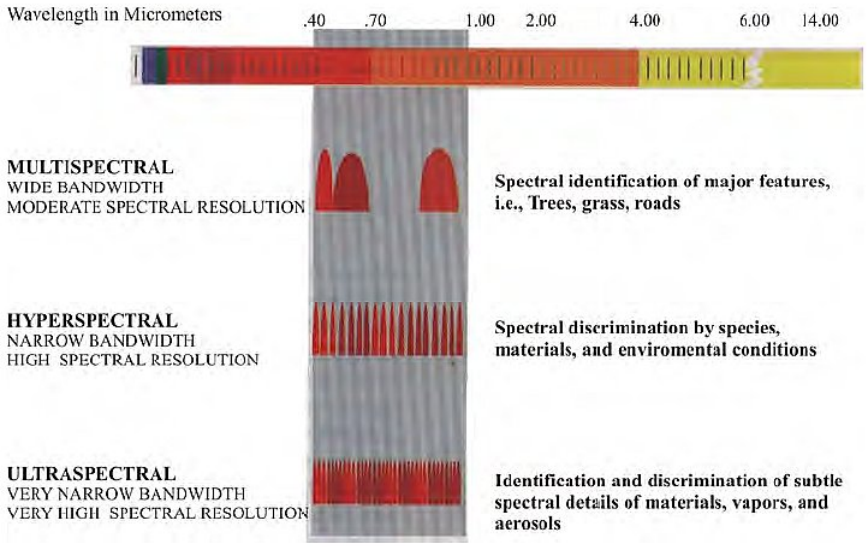


Figure 2.10: Summary of the differences between various types of Spectal Imaging

valuable information is often mixed with noise, different materials etc. Remote sensing in general involves many different kind of noises, such as lighting conditions, relative motion between instrument and target[1], scattering and

[1]e.g. instrument apparatus mounted on a plane

interpolation of light, atmospheric conditions [2] etc.

Also, a Hyperspectal Imaging system is obliged to provide efficient discretion between objects with similar spectral signatures. We should be able to distinguish and uniquely identify different factors without any errors, strict specifications stem from applications in medicine, agriculture.

Basically there are four steps when we want to get information using a Hyperspectral Imaging System.

- Acquiring the Spectral cube
- Pre-processing
- Feature Extraction
- Classification

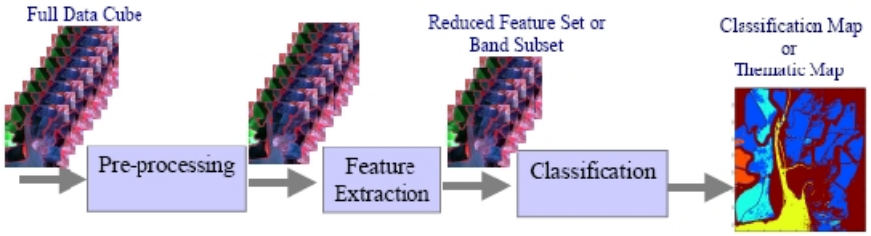


Figure 2.11: Steps required from acquiring the spectral cube to depicting vital information

Although acquiring the spectral cube is self-explanatory, the rest steps need further investigation.

Pre-Processing

The process of acquiring an image in a remote sensing environment, involves the participation of various factors that imply noise, a spectral cube that is not aligned, etc. Before we can use the spectral cube in our analysis environment, we should limit, remove or enhance some aspects of the cubes,

[2] e.g. remote sensing from plane, or satellite

to remove the imperfections imposed during measurement. This involves pre-processing the spectral cube. To name some methods applied, Atmospheric Correction, Radiometric Calibration, Image-Registration.

Atmospheric Correction

In remote sensing from planes or satellites, radiation is absorbed and reflected in air particles, ozone layer, and thus the spectral cube acquired also consists of the spectra of atmosphere. These environmental parameters should be taken into account during measurement and thus remove their contribution during processing.

Radiometric Calibration

The spherical form of the earth, the relative motion of the measuring apparatus (mounted on plane or satellite), the various alterations of the speed of tha plane, implies noise and various methods are used to calibrate the image, taking some marked points on the ground as references.

Image-Registration

Despite relative motion between the measuring apparatus and the target, optical properties of the optical gear result in misplacement of images in the stack of the spectral cube. By registering the images of the spectral cube we align the pixels, so that selecting a pixel results in the same pixel for each slice of the spectral cube.

Feature Extraction

Feature Extraction is the method of identifying key features or attributes in the collected data. When acquiring a spectral cube, all pixels represent different materials. The spectral signatures of all materials in a spectral cube consist the feature space. With feature extraction we try to find the key features of objects in the data, and through a process of mathematical analysis, we extract information based on amplification or selection of portions of data that uniquely characterize the objects. This process is optional and used to reduce computational times, as it helps the classification routines. As an example we can refer to fig.2.7, where the basic spectra of three materials are identified. In this picture, three features, namely vegetation, soil and water, have very discrete spectral signatures.

Classification

Classification is the method of assigning each pixel of the spectral cube to a certain category, by means of the statistical properties of the function of the light intensity represented by the spectra of that pixel. Various criteria and algorithms exist to account for various statistical properties or other attributes available, that are best in terms of classification efficiency. Classification efficiency[3] is determined in practice and we may use a training data set, to provide some base intuition. Classification is a part of a wider field, of pattern recognition.

Through classification we aim in creating an automated system to classify our data, leading to a pseudo-color map where each color represents pixels with similar spectra characteristics. The similarity degree is defined by the classification algorithm and the decision function used[4].

Starting from the feature space, we declare each object with a vector $x = [x(1), x(2), ..., x(N)]$ where $x(i), (i = 1 \text{ to } N)$ represents one feature from a total of N.

In the case of spectral cubes our feature space is a 2-dimensional collection of vectors. Each pixel of the spectral cube is to be classified. Thus every pixel-vector $x(i, j)$ represents the intensity values of the cube in position (i, j). The end result of the process is the classification of the pixel-vector $x(i, j)$ to a class ω_j where $j = 1$ to K, where K is the total number of classes. Each pixel-vector of our feature space is compared to a reference vector (or the objects $x(i)$).

$$refvec = [ref(1), ref(2), ..., ref(K)] \tag{2.4}$$

where K is the number of classes, using a decision function $f(x)$ such that:

$$f(x) = \omega_j, j = 1 \text{ to } K \tag{2.5}$$

Depending on the classification or classification process, there are Supervised and Unsupervised methods. In Supervised methods we take advantage on previous knowledge regarding the spectral characteristics of some objects in the spectral cube. Then we train our system using that information as a training set. On the other hand, unsupervised methods rely on automated clustering of the spectral cube, based on specific statistical attributes.

[3]Because of the vast number of methods and procedures, classification efficiency is the successful **decision** of the correct class for a given spectra in a number of experiments. It highly depends on the decision function and the reference spectra used.

[4]Similarity or distance from some reference spectra.

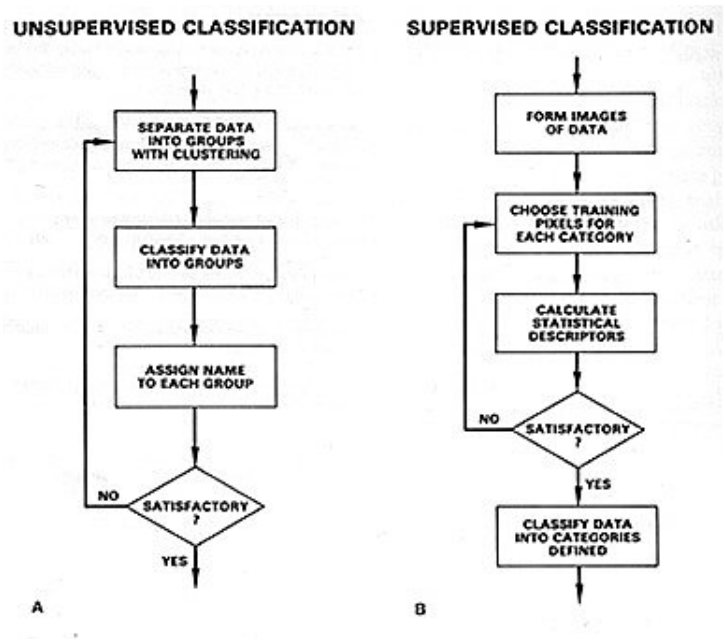


Figure 2.12: Steps in supervised and unsupervised classification methods

We conclude this chapter by mentioning various classification methods.

Nearest Neighbour

The decision function (2.5) in the *Nearest Neighbour* classifier is a distance critiria. Each pixel-vector is classified to the nearest cluster, and there are three approaches as far as the distance critiria is concerned:

1. Classify each sample to the closest reference cluster. The reference clusters are estimated in the training process.

2. Classify each sample to the closest sample, forming new clusters based on a threshold distance to form new clusters.

3. Given a sample we seek the k closest reference samples taken during training. k is a predifined parameter, and then the sample is classified by majority terms.

The most common and widely implemented Nearest Neighbour classifier, is based on the first distance critiria. It is widely used and also found in man

Euclidean Distance

The *Euclidean Distance* classifier takes C initial samples in the spectral cube as reference, and C represents the number of different materials that exist in our dataset. Then the distance between each sample and the reference is calculated:

$$Dist = \sqrt{(BV_{ijk} - \mu_{ck})^2 + (BV_{ijl} - \mu_{cl})^2} \tag{2.6}$$

where BV_{ijk} and BV_{ijl} are the intensity values of pixel ij, in bands k and l respectively, and μ_{ck} and μ_{cl} are the mean values of all pixels of c class, in the bands k and l respectively.

The algorithm uses regression and terminates when an extra iteration makes no changes in the pixel's class. The steps are:

```
* Get C initial means
* For each iteration
        * Calculates pixel distances and assign membership
        * Calculates new means based on the class members
* Returns a classification vector.
```

Fisher's Linear Discriminant - Mahalanobis Distance

Euclidean Distance classifier takes the same distances for all classes. *Fisher's Linear Discriminant* takes into account the covariance matrix between the classes to include information regarding the similarity between different bands.

In the case of *Mahalanobis Distance* every class has its own covariance matrix.

Maximum Likelihood - Bayes classifier

The *Maximum Likelihood* classifier is based on statistical analysis. It regards the classes as Independent Variables with Gaussian distribution. The decision function is of the form:

$$\text{if } p(x|\omega_i) \cdot p(\omega_i) > p(x|\omega_j) \cdot p(\omega_j)\ \forall j, i \neq j \text{ then } x \exists \omega_i \tag{2.7}$$

where x is the current vector and ω_i is a class.

Bayes classifier takes also into account the cost of a right or wrong classification decision.

Chapter 3

Light and Plants

Photosynthesis one of the most vital procedures in our world as every form of life depends on it, directly or indirectly through the food chain. This basic procedure found in plants, algae and some bacteria consumes CO_2 and produces O_2 which is vital for life. We examine the process of photosynthesis in plants and then the interaction between light and plant matter. We briefly examine remote sensing in plants and in more depth in specific applications.

3.1 Photosynthesis and Chromophores

Photosynthesis is a metabolic pathway that converts light energy into chemical energy. Its initial substrates are carbon dioxide and water; the energy source is sunlight (electromagnetic radiation); and the end-products are oxygen and (energy-containing) carbohydrates, such as sucrose, glucose or starch. This process is one of the most important biochemical pathways, since nearly all life on Earth either directly or indirectly depends on it as a source of energy. It is a complex process occurring in plants, algae, as well as bacteria such as cyanobacteria.

Only a portion of sunlight is used during photosynthesis. It is known as Photosynthetically Active Radiation, often abbreviated PAR, and designates the spectral range of solar light from 400 to 700 nanometers that is useful to terrestrial plants in the process of photosynthesis.

Chlorophyll, the most abundant plant pigment, is most efficient in capturing red and blue light. Horticulturists say that blue light is the most important for leaf growth and that red light encourages flowering. Accessory pigments such as carotenes and xanthophylls harvest some green light and pass it on to the photosynthetic process, but enough of the green is reflected

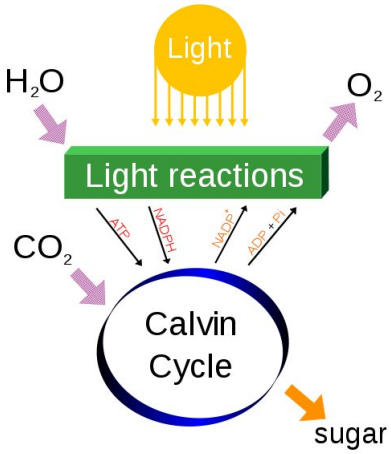


Figure 3.1: The procedure of Photosynthesis.

to give leaves their characteristic color. An exception to the predominance of chlorophyll is autumn, when chlorophyll decays earlier than the accessory pigments that remain to color the leaves red, yellow and orange. In fig.3.2 we can see the wavelengths where chlorophyll and carotene absorb radiation, and in fig.3.3 we can see the relation between the health status of the leaf, its color and wavelengths of the radiation absorbed.

Chlorophyll is vital for photosynthesis, which allows plants to obtain energy from light. Chlorophyll molecules are specifically arranged in and around pigment protein complexes called photosystems which are embedded in the thylakoid membranes of chloroplasts. In these complexes, chlorophyll serves two primary functions. The function of the vast majority of chlorophyll (up to several hundred molecules per photosystem) is:

- to absorb light
- to transfer that light energy by resonance energy transfer to a specific chlorophyll pair in the reaction center of the photosystems.

Because of chlorophyll's selectivity regarding the wavelength of light it absorbs, areas of a leaf containing the molecule will appear green. There are currently two accepted photosystem units, Photosystem II and Photosystem I, which have their own distinct reaction center chlorophyll, named P680 and P700, respectively. These pigments are named after the wavelength (in nanometers) of their red-peak absorption maximum, as depicted in fig.3.4.

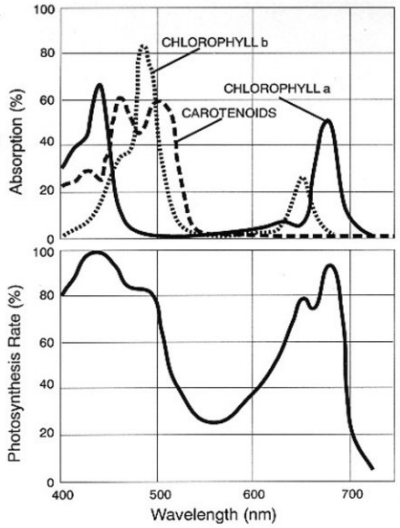


Figure 3.2: We can see the absorbance spectra of chlorophyll-α, chlorophyll-β and carotenes. The efficiency of photosynthesis on the same wavelenghts is depicted in the second diagram.

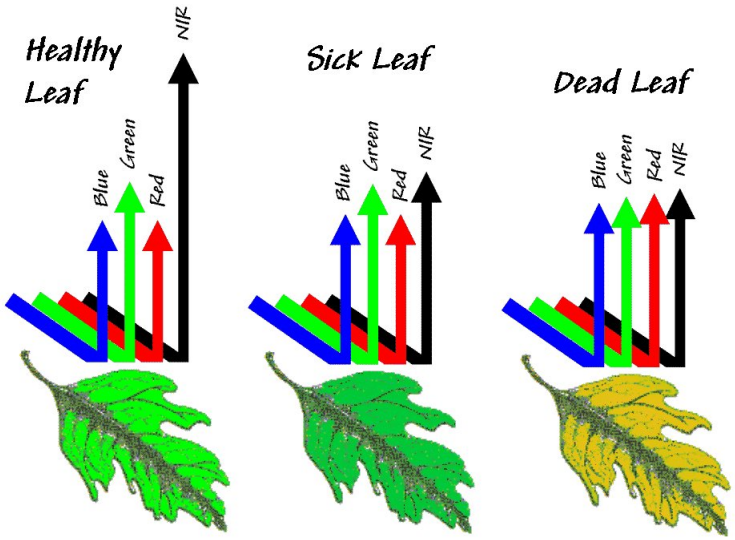


Figure 3.3: The color of the leaves in relation to the wavelength of the light absorbed in different states.

The identity, function and spectral properties of the types of chlorophyll in each photosystem are distinct and determined by each other and the protein structure surrounding them.

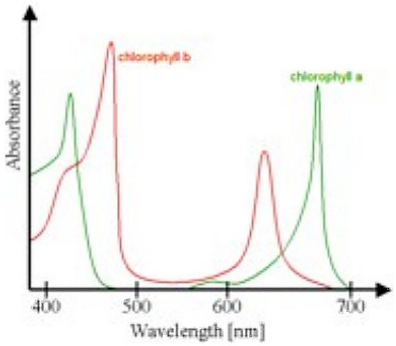
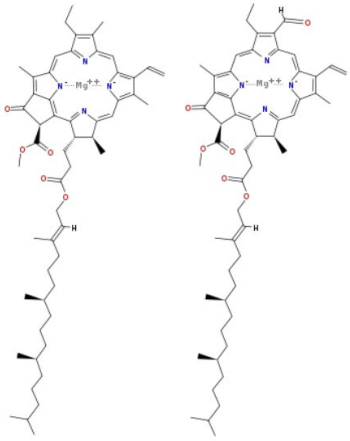


Figure 3.4: Chlorophyll α and β absorption spectra.



(a) Chlorophyll a (b) Chlorophyll b

Figure 3.5: Molecular structures of Chlorophyll α and β

The function of the reaction center chlorophyll is to use the energy absorbed by and transferred to it from the other chlorophyll pigments in the photosystems to undergo a charge separation, a specific redox reaction in which the chlorophyll donates an electron into a series of molecular intermediates called an electron transport chain. The charged reaction center

chlorophyll (P680+) is then reduced back to its ground state by accepting an electron. In Photosystem II, the electron which reduces P680+ ultimately comes from the oxidation of water into O2 and H+ through several intermediates. This reaction is how photosynthetic organisms like plants produce O2 gas. Photosystem I typically works in series with Photosystem II, thus the P700+ of Photosystem I is usually reduced, via many intermediates in the thylakoid membrane, by electrons ultimately from Photosystem II. Electron transfer reactions in the thylakoid membranes are complex, however, and the source of electrons used to reduce P700+ can vary. The electron flow produced by the reaction center chlorophyll pigments is used to shuttle H+ ions across the thylakoid membrane, setting up a chemiosmotic potential mainly used to produce ATP chemical energy, and those electrons ultimately reduce NADP+ to NADPH a universal redundant used to reduce CO2 into sugars as well as for other biosynthetic reductions. Reaction center chlorophyll-protein complexes are capable of directly absorbing light and performing charge separation events without other chlorophyll pigments, but the absorption cross section (the likelihood of absorbing a photon under a given light intensity) is small. Thus, the remaining chlorophylls in the photosystem and antenna pigment protein complexes associated with the photosystems all cooperatively absorb and funnel light energy to the reaction center.

Photosynthesis uses light energy and carbon dioxide to make triose phosphates (G3P). G3P is generally considered the first end-product of photosynthesis.[citation needed] It can be used as a source of metabolic energy, or combined and rearranged to form monosaccharide or disaccharide sugars, such as glucose or sucrose, respectively, which can be transported to other cells, stored as insoluble polysaccharides such as starch, or converted to structural carbohydrates, such as cellulose or glucans.

A commonly used slightly simplified equation for photosynthesis is:

$$6CO_{2(g)} + 12H_2O_{(l)} + \text{photons} \rightarrow C_6H_{12}O_{6(aq)} + 6O_{2(g)} + 6H_2O_{(l)} \quad (3.1)$$

carbon dioxide + water + light energy → glucose + oxygen + water

Photosynthesis occurs in two stages, as depicted in fig.3.1. In the first stage, light-dependent reactions or photosynthetic reactions (also called the Light Reactions) capture the energy of light and use it to make high-energy molecules. During the second stage, the light-independent reactions (also called the Calvin-Benson Cycle, and formerly known as the Dark Reactions)

use the high-energy molecules to capture and chemically reduce carbon dioxide (CO2) (also called carbon fixation) to make the precursors of carbohydrates.

The energy for photosynthesis ultimately comes from absorbed photons and involves a reducing agent,which is water in the case of plants, releasing oxygen as product. The light energy is converted to chemical energy (known as light-dependent reactions), in the form of ATP and NADPH, which are used for synthetic reactions in photoautotrophs. The overall equation for the light-dependent reactions under the conditions of non-cyclic electron flow in green plants is:

$$2H_2O + 2NADP^+ + 2ADP + 2P_i + light \rightarrow 2NADPH + 2H^+ + 2ATP + O_2 \quad (3.2)$$

Most notably, plants use the chemical energy to fix carbon dioxide into carbohydrates and other organic compounds through light-independent reactions. The overall equation for carbon fixation (sometimes referred to as carbon reduction) in green plants is:

$$3CO_2 + 9ATP + 6NADPH + 6H^+ \rightarrow C_3H_6O_{3phosphate} + 9ADP + 8P_i + 6NADP^+ + 3H_2O \quad (3.3)$$

To be more specific, carbon fixation produces an intermediate product, which is then converted to the final carbohydrate products. The carbon skeletons produced by photosynthesis are then variously used to form other organic compounds, such as the building material cellulose, as precursors for lipid and amino acid biosynthesis, or as a fuel in cellular respiration.

Plants absorb light primarily using the pigment chlorophyll while often supported by other accessory pigments such as carotenes and xanthophylls. Both chlorophyll and accessory pigments are contained in organelles (compartments within the cell) called chloroplasts. Although all cells in the green parts of a plant have chloroplasts, most of the energy is captured in the leaves. The cells in the interior tissues of a leaf, called the mesophyll, can contain between 450,000 and 800,000 chloroplasts for every square millimeter of leaf. The surface of the leaf is uniformly coated with a water-resistant waxy cuticle that protects the leaf from excessive evaporation of water and decreases the absorption of ultraviolet or blue light to reduce heating. The transparent epidermis layer allows light to pass through to the palisade mesophyll cells where most of the photosynthesis takes place.

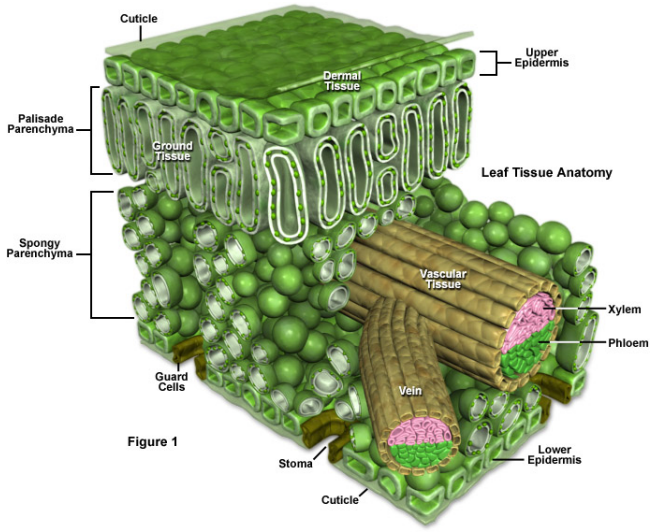


Figure 3.6: Leaf structure.

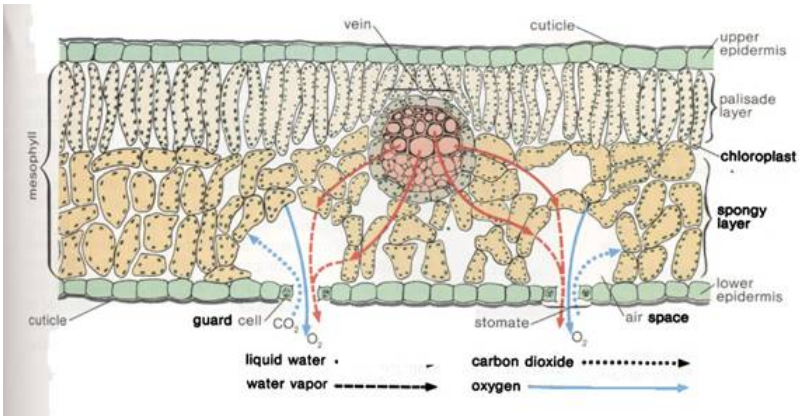


Figure 3.7: Leaf structure and kinetics during photosynthesis.

Light-harvesting Complexes

Green plants contain chloroplasts which house the photosynthetic apparatus. Chloroplasts contain the thylakoid membrane, which holds the pigment protein complexes, known as LHC, light-harvesting complexes, I and II. The tightly spaced regions in the thylakoid membrane, known as grana, hold predominantly LHCII, which is the most abundant pigment-protein complex in green plants.

It is important to note that both photosystems are almost simultaneously

excited; thus, both photosystems begin functioning at almost the same time.

Steps

1. The excited electron is passed along until it reaches P680 chlorophyll.
2. The excited electron is passed to the primary electron acceptor. Photolysis in the thylakoid takes the electrons from water and replaces the P680 electrons that were passed to the primary electron acceptor. (O2 is released into the air as a waste product)
3. The electrons are passed to photosystem I via the electron transport chain (ETC) and in the process used to pump protons across the thylakoid membrane into the lumen.
4. The stored energy in the proton gradient is used to produce ATP which is used later in the Calvin-Benson Cycle.
5. P700 chlorophyll then uses light to excite the electron to its second primary acceptor.
6. The electron is sent down another ETC and used to reduce NADP+ to NADPH.
7. The NADPH is then used later in the Calvin-Benson Cycle to remove PGA that is produced from RuBisCO reaction and releases enzyme for continuation of steady state reaction.

Role of carotenoids

Chlorophylls and carotenoids are important in light-harvesting complexes present in plants. Chlorophyll b is almost identical to chlorophyll a except it has a formyl group in place of a methyl group. This small difference makes chlorophyll b absorb light with wavelengths between 400 and 500 nm more efficiently. Carotenoids are long linear organic molecules which have alternating single and double bonds along their length. Such molecules are called polyenes. Two examples of carotenoids are lycopene and These molecules also absorb light most efficiently in the 400 500 nm range. Due to their absorption region, carotenoids appear red and yellow and provide most of the red and yellow colours present in fruits and flowers. The carotenoid molecules also serve a safeguarding function. Carotenoid molecules suppress damaging photochemical reactions, particularly those including oxygen, which exposure to sunlight can cause. Plants that lack carotenoid molecules quickly die upon exposure to oxygen and light. The chlorophylls and carotenoids present in

the light-harvesting complexes are referred to as accessory pigments. These accessory pigments are held inside the light harvesting proteins in a highly uniform fashion.

Not all wavelengths of light can support photosynthesis. The photosynthetic action spectrum depends on the type of accessory pigments present. For example, in green plants, the action spectrum resembles the absorption spectrum for chlorophylls and carotenoids with peaks for violet-blue and red light. In red algae, the action spectrum overlaps with the absorption spectrum of phycobilins for blue-green light, which allows these algae to grow in deeper waters that filter out the longer wavelengths used by green plants. The non-absorbed part of the light spectrum is what gives photosynthetic organisms their color (e.g., green plants, red algae, purple bacteria) and is the least effective for photosynthesis in the respective organisms.

3.2 Plant Tissue and Interaction with Light

In section 2.1 we discussed the interaction of light with matter, in fig.3.8 we can see the phenomena that take place when light hits plant matter.

In the previous section we discussed the role of chlorophyll and carotene. We considered the visible portion of the spectrum that is named PAR, due to its significance in the process of photosynthesis. The interaction between light and plant matter is extended more than the visible spectrum, thus visible, near- and shortwave infrared wavelengths can be divided into three categories when evaluating the spectral properties of leaves. These categories are related to the different components of a leaf. These categories are:

- 400nm to 750nm: plant pigments, especially chlorophyll α and β, carotenes and xanthophylls.
- 750nm to 1350nm: internal leaf structure.
- 1350nm to 2500nm: water concentration in plant tissue.

During early stages of plant development, chlorophyll concentration is relatively low compared to carotenoids and anthocyanins and the visual appearance of a plant leaf is pale green. Blue and red light is absorbed and used as an energy source for photosynthesis and photochemical reactions. As the plant matures, the chlorophyll concentration increases relative to other plant pigments, and the plant appears green to dark green. Upon senescence, chlorophyll synthesis has ceased, resulting in less light absorption in the blue

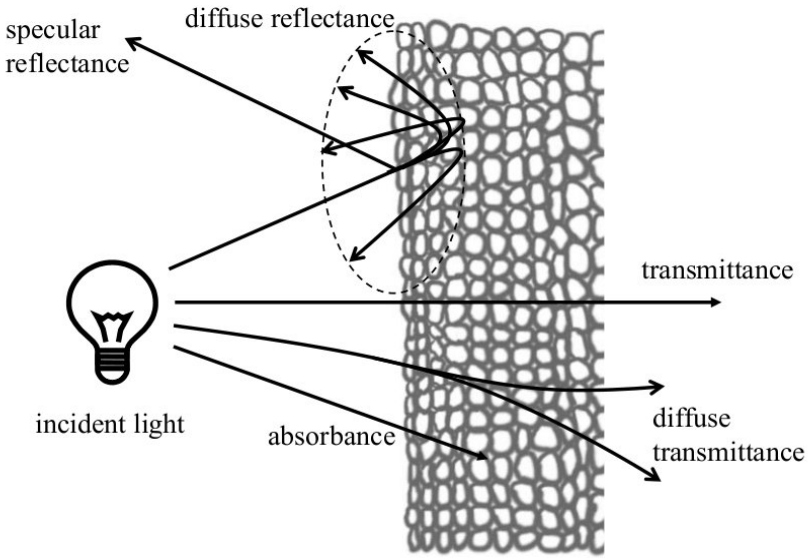


Figure 3.8: Incident light on the tissue cells of food products results in specular reflectance, diffuse reflectance (diffuse) transmittance and absorbance. These strongly depend on the object material and wavelength.

and red wavelenghts. The senescent plant is dominated by other pigments (other than chlorophyll), resulting in a yellow to brown visual appearance.

Reflectance in the next category, from 750nm to 1350nm, is affected by the internal leaf structure. Relatively higher reflectance occurs in this waveband when compared to visible light, with over 50% of the incoming radiation reflected by mature leaves. During early stages of leaf development, mesophyll air spaces appear as cell walls pull apart. Incident near-infrared light passes through the upper layers of the leaf into the spongy mesophyll, reflecting off the cell walls in the mesophyll air spaces. As the air spaces increase, ans subsequently an increase in wall surface area, the reflectance in the near-infrared increases. Reflectance is at its highest at full maturity and decreases as the cell walls breakdown during senescence, as seen in fig.3.10.

In section 3.1 we described the process of photosynthesis and the role of chlorophyll and carotenes. In fig.3.2 the absorption spectra of the chromophores is depicted. In vivo these spectral signatures may have differences because of plant matter and molecular structure. We now combine the spec-

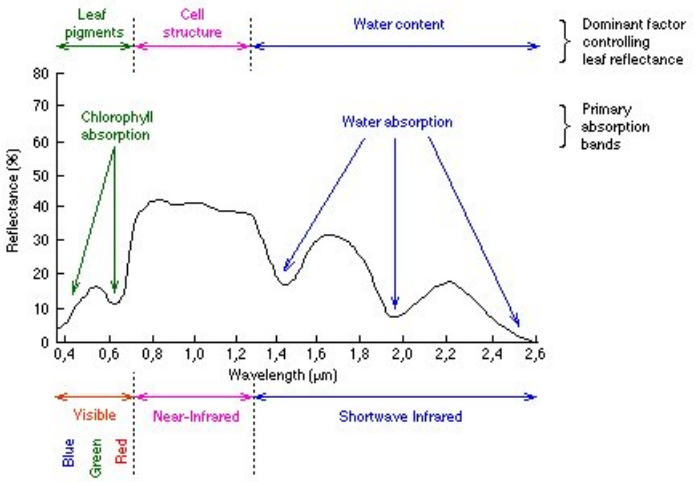


Figure 3.9: Leaf Spectra response and categorization depending on leaf components responsible.

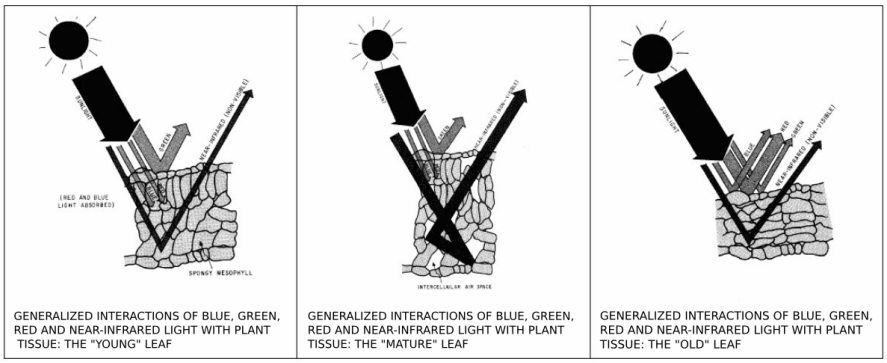


Figure 3.10: Reflectance of light by leaf in three phases.

tral signatures of the chromophores with total spectral response of a plant, as in fig.3.9. The peaks and deeps of absorbance or reflectance spectra can be used as a diagnostic tool. For this reason many indices have been defined to calculate total chlorophyll content, carotene content, chlorophyll to carotene ratios etc. In section 3.4.2 we refer to vegetation indices implemented for use in remote sensing applications.

	Mounting apparatus	**Space Resolution**	**Spectral Resolution**	**Processing Requirements**
Plant Pathology	Lab equipment	Low	Very High	High
Vegetation Identification	Plane, Satelite or Portable device	Low to very High	Limited	Limited
Quality control fruits	Portable device, Platform	Low	Limited	Very High

Table 3.1: Hyperspectral Imaging application fields in plants. Differences and similarities.

3.3 Remote Sensing in Plants

Hyperspectral Imaging is a method in remote sensing. By remote sensing we refer to the method of taking measurements of an object from a distance. So the measuring apparatus[1] could be mounted on a satellite, plane or in a kind of platform in field situations (or laboratory experiments). As far as research and experiments on plants implementing remote sensing is concerned, there are three major fields of interest. The principles of remote sensing are the same, although differences do exist, regarding methodology, specifications, field of interest and various limitations implied in the field.

Current research fall for three major categories.

- Quality assessment of fruits
- Vegetation Identification
- Plant Pathology

To give a brief description of the differences between the categories mentioned, we should focus on experimental procedure, lighting, already available data and references, different kind of distortions implied in the field. The following table, describes the basic similarities and differences.

[1] A Spectral Imaging system

We should also point out that all these fields of application have different impact on economy and in the process of agriculture and food production. These technologies in combination:

- can reduce the amount of fertilizers used,
- can automate production and packaging,
- given a damaging factor one can apply a more sophisticated and targeted treatment,
- provide efficient means for water management,
- can enhance quality of food production.

Satellites characteristics

	Spectral bands	Nadir IFOV	Swath Width	Revisit at equator	Bits per pixels
NOAA/AVHRR	Red-IR-TIR	1 Km	2700 Km	2 passes/day	16 MS, 8 NDVI
MODIS Terra-Aqua	blue-red-IR-TIR	250 m (bands 1-2) 500 m (bands 3-7) 1000 m (bands 8-36)	2300 Km	1 to 2 days	16
Landsat TM	VIS-IR-TIR	30m MS, 120m TIR	185 Km	16 days	8
Landsat ETM+	VIS-IR-TIR PAN	30m MS, 120m TIR,15m PAN	185 Km	16 days	8
SPOT	VIS-IR-PAN	20m MS, 10m PAN	60 Km	26 days	8
ERS1/2	C-band	30 m	100 Km	35 days	
Ikonos	VIS-IR-PAN	4m MS, 1m PAN	11 Km		11
QuickBird	VIS-IR-PAN	2,44 m MS, 0,61 m PAN	16.5 Km		11

MS = Multispectral; PAN = Panchromatic; VIS = Visible; IR = Reflected Infrared; TIR = Therrmal Infrared.

Figure 3.11: Characteristics of the Spectral Imaging Systems of various satellites.

In the remaining chapter we discuss various research approaches, results in these fields, while we dive into more details in plant pathology

3.3.1 Quality assessment of food

Quality control of foods, is a flourishing field because of its significance in society. There is a growing interest in providing better food and following

high quality standards that people can rely on. Also automated production becomes more sophisticated and mandatory and technology is used to assure that the standards are met.

In food production in general, it is crucial to distinguish between good and bad fruits, meat etc. Right know in a food production facility the quality control is a human task. It is a tiresome and boring task, no mistakes are allowed and also a certain pace should be maintained throughout a shift. There are lines of people that select the fruits, chickens etc. That is a task that could be automated while providing the versatility of a human's ability to make decisions. To cut a long story short many research groups focus on identifying damaged objects in an automated machine that sorts "good" and "bad" objects.

Another reason for implementing spectral imaging in food quality control is the ability to distinguish various types of damage. So there is a discretion between good and bad food, but also the kind of damage is recorded and can be offline processed to deduce valuable information regarding the source and kind of problem. Fig.3.12 illustrates defects and diseases in case of french fries. The idea is based in the unique spectral characteristics of a damaged area for example. To extend this idea, defects and diseases in food has analogous spectral response, compared to the expected healthy spectral response. Analogous examples may be found in case of tomatoes, apples, chickens [25].

First attempts used simple RGB cameras and computers in a simple machine vision system that tried to investigate the quality of fruits. The system was promising although it still lacked the accuracy required. Many researchers [18],[9],[19], have found that Hyperspectral Imaging approaches outperform RGB images since many defects are not visible in the color images. It provides more information so that we could distinguish between different states of maturity, kind of damage, disease or defect. Labeling and classification accuracy is highly improved.

Applying Hyperspectral Imaging in food production has been proven although, there are some major disadvantages. The system should be able to provide results in real-time. It should process a vast amount of information

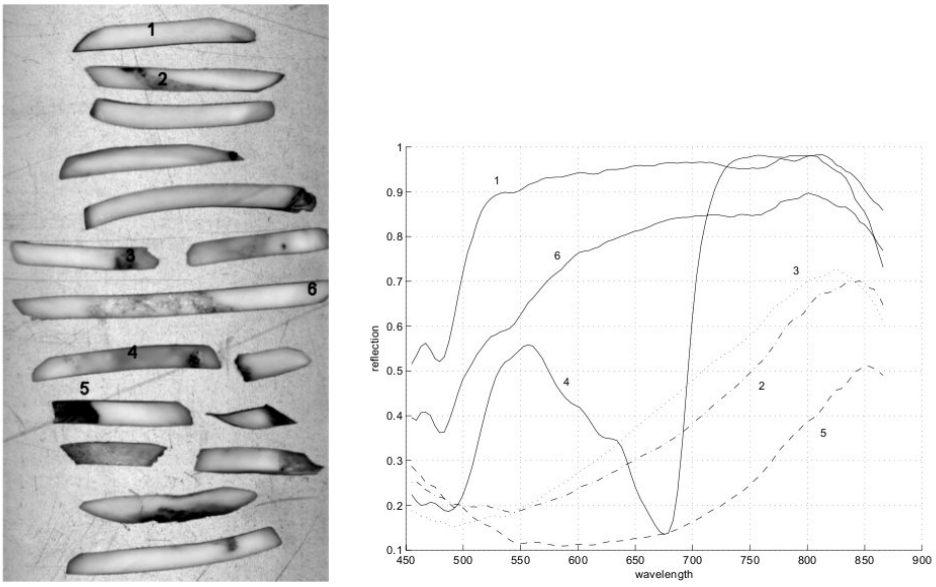


Figure 3.12: Several typical French fries defects and diseases with their corresponding spectra. 1=potato flesh, 2=peel, 3=damage, 4=greening, 5=external rot, 6=browning

to maintain an acceptable rate in the distinguishing machine, successfully and efficiently identifying stress factors and defects.

Researchers now focus on providing even more accuracy, taking advantage of various spectra signatures, such as ripping tomatoes, certain stress impacts, etc, providing real-time results. The following image (fig.3.13) illustrates various attempts in food quality evaluation, taken from [9].

3.3.2 Vegetation Identification

Vegetation Identification if the field of remote sensing of plant matter, where the aim is to extract information of the quality of plant matter. In previous sections we examined the reflectance and absorbance spectra of chlorophyll and carotenes. Leaf chlorophyll content provides valuable information about physiological status of plants. Reflectance and absorbance measurements by means of remote sensing, makes it possible to quickly and non-destructively assess, in situ, the chlorophyll content of leaves. Many large scale project

Table 1. Summary of CCD camera applications for food quality evaluation

Category	Products	Applications	References
Fishery	Bivalve	Study of larval growth	Pontual *et al.* (1998)
	Crassostrea	Detection of hinge lines	Jung and Fred (2002)
	Fish	Sorting fish	Zion *et al.* (1999)
Fruit	Apple	Defect segmentation	Leemans *et al.* (1999)
	Cherry	Analysing fruit shape	Beyer *et al.* (2002)
	Orange	Location and characterization of the stem-calyx area	Ruiz *et al.* (1996)
	Pistachio nuts	Detection of early split	Pearson and Slaughter (1996)
Grain	Rice	Quality classification	Wan *et al.* (2002)
	Wheat	Classification	Utku and Koksel (1998)
Meat	Beef	Using image texture features as indicators of tenderness	Li *et al.* (1999)
	Pork	Color evaluation	Lu *et al.* (2000)
	Poultry carcasses	Classification	Park *et al.* (2002)
Vegetable	Asparagus	Defect inspection	Rigney *et al.* (1992)
	Chicory	Study of visual preference	Coppenolle *et al.* (2002)
Others	Cheese	Evaluation of the functional properties	Wang and Sun (2001)
	Noodle	Influence of sprout damage on appearance	Hatcher and Symons (2000)
	Pizza	Quality evaluation	Sun and Brosnan (2003a, 2003b)
	Sausage	Estimation of sensory properties	Ioannou *et al.* (2002)

Figure 3.13: Food quality evaluation based on CCD cameras.

such as AVIRIS[2] and CHRIS[3] scan large areas to identify vegetation.

Proven the ability to examine and identifying vegetation remotely from planes, while on the same time discriminating different kinds of vegetation, there is an effort to expand the capabilities and quality of information extracted. Vegetation Identification makes use of several spectral bands, where the correlation between spectra and chlorophyll is high. These spectra bands are the wavelengths where chlorophyll, carotenes and other plant pigments give their maxima and minima in terms of absorption or reflectance of light. These maxima and minima are strongly related to the concentration of chlorophyll, carotenes and so on. Several indices have been defined that make use of these bands, in an attempt to define chlorophyll (and other pigments) concentration directly from spectral information.

Many of these indices have proved their ability, in terms of successfully estimating chlorophyll and other pigments concentration. Still though there

[2]Airborne Visible/Infrared Imaging Spectrometer
[3]Compact High Resolution Imaging Spectrometer

is a great debate, as which bands provide better results. The debate stems from the fact that although chlorophyll and carotene spectral signatures are known to have maximum absorption in known bands as seen in fig.3.2, fig.3.4 and fig.3.9, these responses differ in different species and also measurements in situ, are also affected by factors such as leaf structure, physiology etc. In fig.3.14 we can see a summary of many indices that have been defined and tested by [6].

Table 1
Summary of the spectral reflectance indices used in this study

Pigment estimated	SR indices	Normalized difference indices	Others
Chlorophyll	$\mathrm{SR}_{680} = \frac{R_{800}}{R_{680}}$	$\mathrm{ND}_{680} = \frac{(R_{800} - R_{680})}{(R_{800} + R_{680})}$	$\lambda_{re} = \lambda_{\max \frac{dR}{d\lambda}}$
	$\mathrm{SR}_{705} = \frac{R_{750}}{R_{705}}$	$\mathrm{ND}_{705} = \frac{(R_{750} - R_{705})}{(R_{750} + R_{705})}$	$\mathrm{mSR}_{705} = \frac{R_{750} - R_{445}}{R_{705} - R_{445}}$
			$\mathrm{mND}_{705} = \frac{(R_{750} - R_{705})}{(R_{750} + R_{705} - 2R_{445})}$
Carotenoid/chlorophyll		$\mathrm{PRI} = \frac{(R_{531} - R_{570})}{(R_{531} + R_{570})}$	$\mathrm{SIPI} = \frac{(R_{800} - R_{445})}{(R_{800} - R_{680})}$
			$\mathrm{PSRI} = \frac{(R_{680} - R_{500})}{R_{750}}$
Anthocyanin	$\frac{\mathrm{Red}}{\mathrm{Green}} = \frac{\sum_{i=600}^{699} R_i}{\sum_{i=500}^{599} R_i}$		

R_{xxx} refers to leaf reflectance at wavelength xxx in nanometers. See text for full names and sources for the indices.

Figure 3.14: Various vegetation indices tested by Sims and Gamon

It is crucial to define indices and means of estimating plant's pigments concentration that are generic and provide immunity in between species. Some indices such as Red-Edge[4], and NDVI[5] have been validated through different species.

The main advantages of airborne remote sensing, is of course covering great areas and districts. Remote sensing via satellites is of proven efficiency although only capable of distinguishing between discrete spectral signatures, such as between soil and vegetation due to the limitation of low field resolution. Airborne remote sensing is better than satellites, although limitations still exists in analogous way, that are more application specific. We should

[4]Red-Edge is the minima of chlorophyll reflectance in 680nm
[5]Normalized Difference Vegetation Index

point out, that the main trade-off in remote sensing is covering large areas in favour of resolution leading to lack of accuracy and discrimination efficiency.

3.3.3 Plant Pathology

The field of Plant Pathology is the offspring of applying remote sensing techniques in plants. When referring to Plant Pathology, we mean all kinds of stresses and diseases that occur in a plant. Plant Pathology is the science where the effects and the mechanisms of plants homeostasis is investigated. Those stresses can also be categorized to physical ones, such as lack of water, nutritional abnormalities of deficiencies, and physical damage. On the other hand we examine the effect of viruses, and diseases in general that rely on third party involvement.

Thus, plant pathology is of great importance due to the key role diseases and stresses play in vegetation in general and in agriculture. Till now scientists only notice the impact of a virus or stress through their visual effects on leaves, fruits, or plant matter. Another way of investigation in field, is a destructing one, taking samples from the crop or the specimen of investigation, and through a destructing bio-chemical analysis conclude the chemical concentrations and factors that produce the problem. Those methods suffer from lack of efficiency, due to the sampling limitations in a field, and the huge manpower needed in the sampling process and the experiment procedure. Also many times, when the effects of a virus become visible, the crop already is infected or the damage is significant. It is of vital importance to diagnose the problem in time, so that the appropriate measures can be applied. Then we maximize the efficiency of the treatment and also we limit it's application to the suffering area and not to the whole field.

The growing interest in dealing with the various stresses and viruses on plants, is depicted in various research efforts such as [4, 5, 13, 8, 1, 15, 12, 17]. These efforts have many differences in the methods and experiment procedure. These differences also stem from the focus of the efforts. They focus on various species or damaging factors.

The work of the group [4, 5, 13], differentiate as for the methods and procedures and their target that is applying Hyperspectral Imaging in the leaf level.

Pathology Spectra

When a plant is infected or damaged by a stress factor, we only realize the problem, when visible symptoms appear. Literature related to early stages of an infection is very limited. There is plenty of information regarding the spectral response of the evolution of chlorophyll through a leaf's circle of life, although measurements under stress conditions are limited, as biochemical measurements are destructive.

When something goes wrong in a plant, and its homeostasis is disturbed, defence mechanisms try to compensate for the impact of the problem. So if we regard a healthy spectra prior to the emergence of a damaging factor, as a starting observation point, the spectral response of the damaged plant is as in fig.3.15, as indicated by [17].

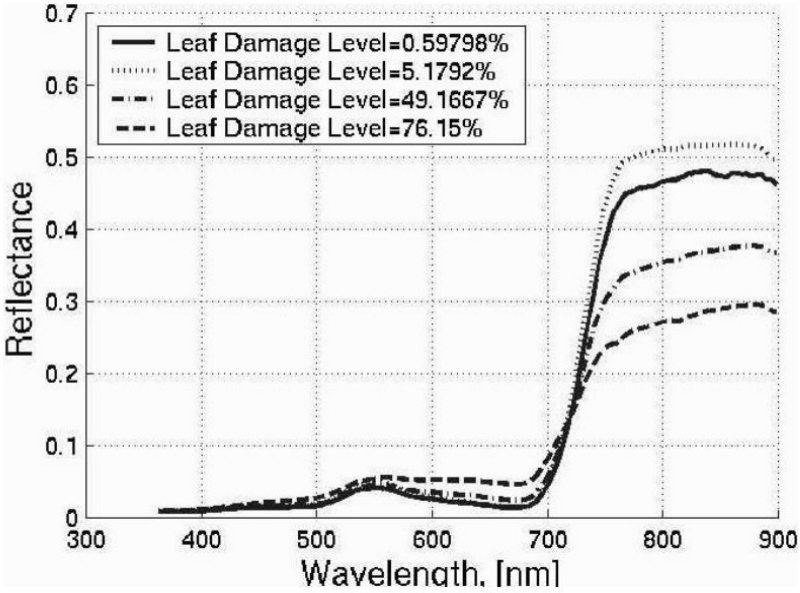


Figure 3.15: Spectra response of a plant in different damage levels.

So for some time, after the initiation of the defence mechanisms, the spectral response of a damaged plant, tends to be "healthier". When the damage progresses the spectra falls to what is expected. We should note that fig.3.15 demonstrates spectra in terms of damage percentage, and not versus time.

3.4 Hyperspectral Imaging and Virus

In the previous sections we described the basics of hyperspectral imaging in plants and we gave brief information for specific applications. We already described the importance of virus detection. What about hyperspectral imaging? In this section we try to answer the question and describe the mechanics of the effect TMV (Tobacco Mosaic Virus) has on Tobacco leaves. We focus on the mechanism and the results of the work of [4],[5],[13]

By implementing a multi-spectral system the mentioned research group has managed to depict the impact of the virus many hours before the effects appear to the human eye (fig.3.16). TMV has a circle of 48 hours, and through their imaging systems, they managed to see the effect in 24 hours.

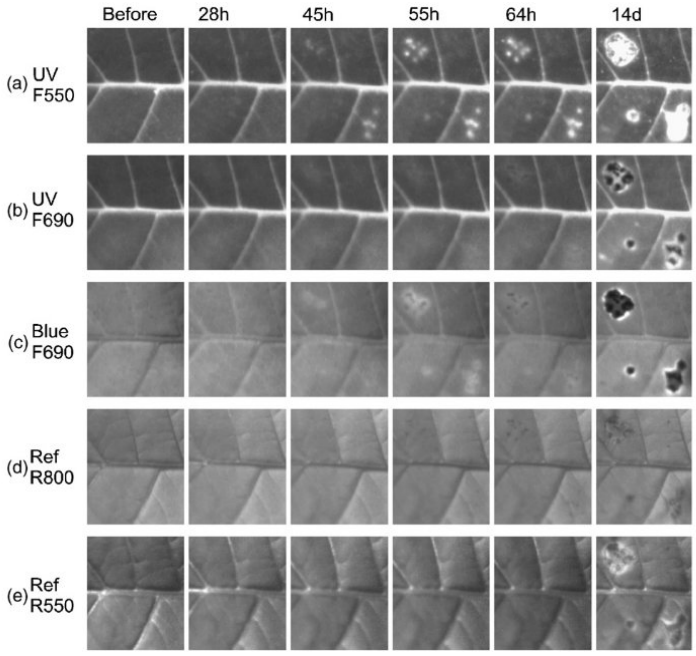


Figure 3.16: Chaerle et al. results

They were able to identify the location of the leaf where the damage progress, many hours before the visual symptoms. In a following experiment they also used a thermal camera, so that they could monitor the minor changes of heat in the surface of the contaminated leaf (fig.3.17). Afterwards

in the most recent publication they examine the prospects of hyperspectral imaging in the leaf level and the possible applications.

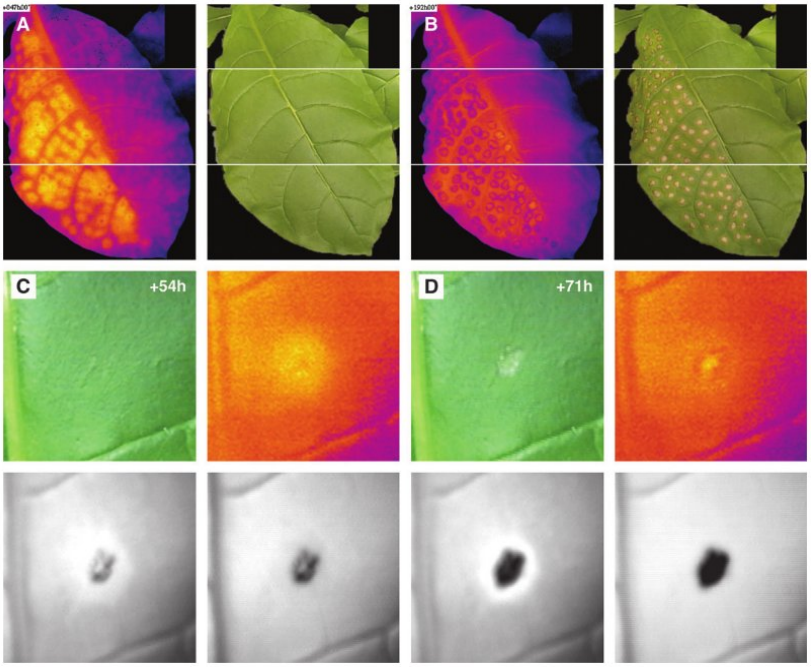


Figure 3.17: Chaerle et. al thermal imaging results. We can see the pinpoints where the damage originates, these are the points where high temperature was identified.

The system used (fig.3.18) was:

- **fluorescence excitation:** Xenon flash lamp, filter wheel in front of the Xenon lamp
- **camera:** gated intensified video camera with a 50mm NIKON lens. Filter-wheel behind the lens and achieved wavelength ranges are: blue 440nm, red 690nm, far red 740nm, all with half bandwidth of 10nmm, green 550nm and near infrared 800nm with half-bandwidth of 40nm. Image resolution 565x754 8bit. Measured leaf area was 2x8mm

So the description of the infecting process is as follows. Upon stress, plants may synthesize specific compounds, depending on the causal agent. Such compounds may alter the absorption of the light impinging on plant

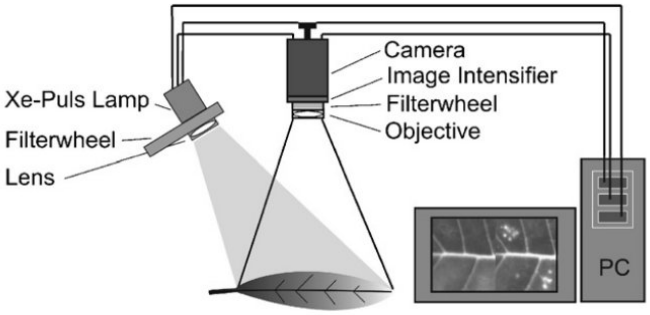


Figure 3.18: Chaerle et al. measurement setup.

leaves, hence the spectrum of reflected, re-emitted, and transmitted light changes. UV- excited fluorescence imaging specifically allows visualization of the accumulation of phenolic compounds, e.g. those associated with the hypersensitive response to pathogens.

Given a damaging factor, the plant tries to compose various chemicals compounds to compensate for the damage. The defence of the plant composes chemicals that relate to the damaging factor. There is a rise of the concentration of these compounds around the area where the initial infection or damage is located. Those compounds may alter the spectral response of the leaf. So we get different reflectance or transmittance in that area due to those compounds. It is claimed that through UV-excited fluorescence imaging we can identify the concentration of phenolic compounds.

Pathogens, in general, also affect the photosynthetic electron transport or the downstream metabolic reactions, resulting in an increase of chlorophyll fluorescence at the early stages of infection. For both the tobacco - TMV and beet - Cercospora plant- pathogen interactions, spots of higher chlorophyll fluorescence intensity co-located with the thermal symptoms (and with the subsequently spreading visual damage). The thermal symptoms is caused by stomatal closure resulting from accumulation of salicylic acid.

Their limited system permitted to discern the key components salicylic acid and scopoletin, fig.3.19

Based on the above-mentioned multi-sensor approaches, it is suggested that is should be possible to establish a robust 'stress-catalogue' that might be used to diagnose and quantify different stress responses on the basis of

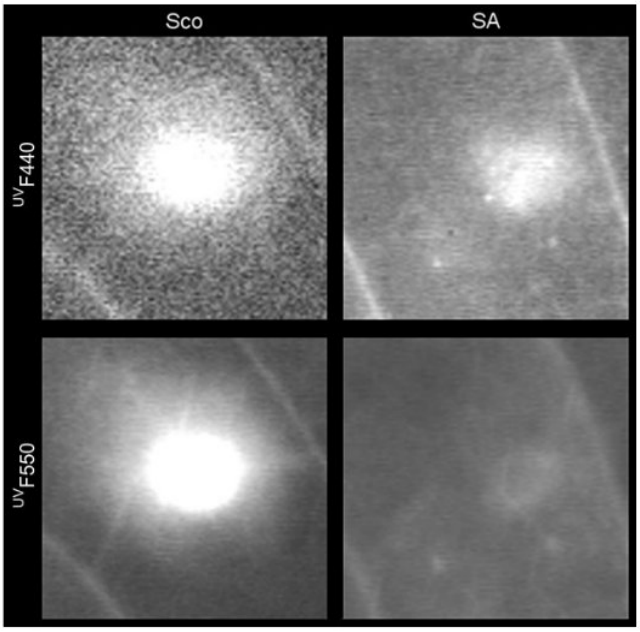


Figure 3.19: Bright areas indicate the high concentration of scopoletin (Sco, left side) and Salicylic acid (SA, right side)

their characteristic stress-specific signatures. This could evolve towards an early warning expert system.

Chapter 4

Materials and Methods

In the following chapter we provide information regarding our experimental procedure, the instruments used and the software developed.

4.1 Hyperspectal Imaging System

The Hyperspectral Imaging System used in our experiment is based on a patent pending, all-optical imaging monochromator. Displacement of the optical elements of the latter results in the tuning of the imaging wavelength, which is performed with the aid of electromechanical manipulators, is controlled from the PC via microcontroller. The system is capable of acquiring spectral images of 5 nm full width half maximum (FWHM), with 3 nm tuning step, in the spectral range 380–1000 nm. The minimum transmittance is 40% across its operational spectral range, which determines the high throughput of the developed monochromator. The tuning spectral range matches the responsivity spectral range of the charge coupled device (CCD) image sensor but it can be extended for longer wavelengths, up to the mid-infrared range. The output feedback signal of the monochromator carries information for the state of the tunable filter, thus enabling its synchronization with the image capturing procedure. The monochromator is interfaced with a black and white, megapixel CCD camera, based on the IEEE-1394 data transferring protocol, capable of acquiring images at a rate of 15 frames/s at full resolution and of more than 30 frames/s at VGA resolution.

Specially developed software, which is compiled under turbo C++, is employed for the control of the camera and monochromator as well as for the spectral image analysis. The system operates in two modes: the spectroscopy mode and the spectrometry mode. The former enables the random selection

and real time visualization of desired spectral images, while the spectrometry mode performs synchronized spectral scanning and image capturing and, finally, calculation of one full spectrum per image pixel. In both cases, a special calibration procedure [2] is executed before these imaging procedures, in order to compensate for the wavelength dependence of the response of the electrooptical parts of the system, such as CCD, illuminators, etc. A Ba2SO4 white plate with unity reflectance across the 380–1000 nm spectral range is used as a calibration specimen. The specimen is placed in the field-of-view of the lens and the gray value of the central area of the image is real time displayed. Then the monochromator scans the total spectral range and in each tuning step the camera shutter and gain is automatically regulated so that the displayed gray value approaches the value of 255. This ensures that the dynamic range of the CCD is fully exploited. The shutter and gain values, used to obtain 255 gray level, are stored in each wavelength-tuning step, together with the image of the white specimen, constituting the calibration data set of the system. These settings determine the sensitivity level of the camera, which increases as the imaging wavelength is tuned to shorter or to longer wavelengths than the wavelength range at which the maximum light throughput and efficiency of the system is obtained. This makes the system's response almost independent from the wavelength, thus ensuring a "device-independent" spectral imaging and spectrometry. The stored spectral images of the white specimen are used in order to correct for the uneven of image brightness due to the non-uniform transfer function of the optics (flat field correction). By running the spectrometry code section, the system performs synchronized tuning of the imaging wavelength and image capturing and storing of the area under analysis. In each step, the sensitivity of the camera is automatically regulated according to the stored, during calibration, shutter and gain values. From the stored stack of spectral images, a spectrum can be calculated from the gray values of the corresponding pixel spectral column and displayed for any spatial point of the image. The spatial resolution of the detector determines the number of the spectra that can be collected in one experiment run. With the described configuration, 1 000 000 spectra can be collected in approximately 2 min scanning time. The system also embodies a fast saving procedure at a lesser (VGA) resolution, which reduces the scanning time of the system to 10 s.

4.2 Experiment Procedure - Protocol

Our experimental procedure consisted of three stages. The first two stages leaves from tobacco plants (Nicotiana Rustica) in vivo where under investi-

gation, whereas in the last stage we examined three leaves of mulberry ex vivo. In order to get the appropriate consistency in the collected data, we created a certain strict protocol. We aimed in limiting the mistakes and random factors of an experimental procedure, since we were dealing with in vivo examination. Also there is a great amount of data collected so that each measurement should not be mistaken.

First and Second Stage

The experiments of these stages took place in MAICH[1]. These stages involve plants of tobacco that where put under investigation while attempting to get the plants infected with virus and monitor the process in vivo with our Hyperspectral Imaging System. The procedure was as follows:

First stage

Three plants were chosen as the plants to be infected, and one plant was chosen as control plant and reference. The plants where marked as fyto1, fyto2, fyto3 and fyto4 both in their pots and in the experiment log. Then certain leaves of the plants where chosen. Our criteria was to investigate three to four leaves of each plant. Each leaf that was chosen was marked using colored fibre (cotton), in the base of the petiole.

Each measurement results in a spectral cube and so the naming convention used to store the data is of vital importance for clarity and latter use. So each cube was named after the plant under investigation, the date of the measurement, the leaf under investigation and the color mark. The naming convention is as follows:

plant - date (dd/mm/year) - leaf - color mark
e.g. fyto1 - 140507 - fyllo1 - prasino

The following table summarizes the number of measurements that were held and the number of spectral cubes acquired:

[1]MAICH stands for Mediterranean Agriculture Institute of Chania

# Plant	# Leaf	# Cubes Collected
1	1	11
1	2	11
2	1	9
2	2	9
3	1	10
3	2	10
4	1	7
4	2	9
4	3	2

Table 4.1: Measurements summary, stage one

Plant	Leaf	Cubes Collected
1b	1	10
1b	2	10
1b	3	10
1b	4	3
2b	full plant	10

Table 4.2: Measurements summary, stage two

Second stage

In the second stage of the experiment, two tobacco plants where chosen and were put under virus stress. The plants where marked as fyto1b and fyto2b, where the letter ***b*** used to represent the second stage while preserving the same naming convention with the first stage of the experiment. The following table summarizes the number of measurements that were held and the number of spectral cubes acquired:

The protocol of the experiment procedure, after selecting the plants is common for both the first and second stage. We describe the protocol here:

```
Measurement procedure:
1. Calibrating the lighting system.
2. Calibrating the HyIS system (procedure explained in previous section).
3. We pick the first leaf of the first plant.
4. Each leaf is cleared of dust and water.
```

Leaf	Mark	Cubes Collected
1	km	15
2	mm	15
3	pm	14
4	-	12

Table 4.3: Measurements summary, stage three

```
5. We orientate the leaf in the same direction from day1 to day2
   (reference is the camera)
6. We start the measurement process.
7. We investigate the cube for any mistakes or faults.
8. If no mistakes are found we proceed to the next step, otherwise
   we repeat steps 5 to 7.
8. We store the spectral cube with the naming convention mentioned.
9. We repeat steps 3 till 8, until our set is over.
```

Third Stage

The third stage was held in the Electronics Laboratory of TUC.

In the third stage we collected 4 leaves from the surroundings of TUC. As far as their species is concerned, all the leaves were taken from a mulberry tree. Our aim was to depict the chlorophyll decay due to nutritional and water stress. That experiment is ex vivo. All the leaves were left in the atmospheric conditions of the Electronics Laboratory, with the exception of the first leaf of mulberry, that was held inside a glass of water [2].

We marked the leaves of the mulberry, by the length of their stalk, with the exception of the last leaf that was very different from all others, and so the naming convention of the third state was:

leaf - kotsani - date dd/mm/year
e.g. fyllo1 - km - 260608

The table 4.3 summarizes the number of measurements that were held and the number of spectral cubes acquired.

[2] with its petiole outside the water

The protocol of the third state is the same as the one already mentioned, with minor changes because we dealt with four discrete leaves.

4.3 Classification Algorithms, Chlorophyll-Carotene Indices

The acquired spectra dataset were classified using Spectral Angle Mapper and Nearest Neighbour. We also implemented two widely used indices to detect chlorophyll concentration and the chlorophyll to carotene ratio.

4.3.1 Classification Algorithms

Current literature on remote sensing in the field of vegetation, is limited as classification techniques are involved. The reasons for this fact are found in the specific applications and thus the equipment used. We already referred to the trade-off in space and spectral resolution. Classification techniques are most valuable in multidimensional data, because of their efficiency that improves[3] with dimensionality of data.

Most attempts in remote sensing in plants, that involves classification, resend in the field of satellite investigation, where projects as AVIRIS and CHRIS provide vast amount of information. In the following paragraph we mention many attempts and we focus on their methodology and field of application. Many efforts use Genetic Algorithms, so [24] applied GA to spectrometer spectra records of 16 tropical mangrove species, [23] applied ROC[4] and statistical properties to data taken with spectroradiometer, Partial Spectal Mixture Unmixing [3] on data acquired by aerial means, various features selection methods in AVIRIS data [22], linear discriminant analysis [12], Self-Organizing Map (SOM) neural network [8], CHRIS data and Look-up tables [10], Nearest Neighbour classifier - spectroradiometer system [17],kernel-driven RossThick-LiSparse model [11].

Our goal is to provide classification analysis in the field of Plant Pathology. From the referred work, it is clear that there are many different methodologies but none of which refer to our goal. Comparisons are difficult, although it is clear that despite the differences in the application specifics or the data

[3] in general, although some techniques improve their performance with few bands
[4] Region Of Convergence

acquired, mathematical formality is crucial and classification techniques improve the accuracy and efficiency of every method. The previous observation is a common place for experimenting and researching on multi-dimensional data. Literature regarding spectral classification is vast and if we exclude the field of Plant Pathology, we should be able to make use of this source of information. In an attempt to deduce our requirements as far the classification methods to apply in our case, we combine knowledge from the general field of spectral classification to specific information involving Plant Pathology.

Requirements

Our Hyperspectral Imaging System[5] acquires spectral cubes with great resolution. We get more space resolution in comparison to spectrometers and 30 spectral bands. We are able to take measurements on leaf level, while preserving high spectral resolution. Taking measurements on leaf level implies that our system should be able to distinguish subtle changes in spectral response and thus classify each sample to the appropriate class. This procedure should be as generic as possible, and immunized against different plant species and environments.

Chlorophyll and carotene spectra, are available (fig.3.2,fig.3.4) and also an example spectra of a damaged plant (fig.3.15). The classifier should be able to take into account the spectra characteristics of those pigments while maintaining a high degree of independence on lighting issues. Our classifiers should be widely implemented to enable comparisons and provide a solid base for debate as for performance and efficiency in Plant Pathology is concerned.

Finally, we choose to implement supervised classifiers. The problem of determining or defining reference vectors, is crucial to every supervised method. The reference vectors play a very important role in the classifiers performance. In sec.4.4 we propose a method to define reference vectors as a key to enhance our classifier's performance.

Spectral Angle Mapper

The Spectral Angle Mapper (SpAM) is a supervised classification algorithm. We have already seen that each pixel in the spectral cube, has $\boldsymbol{n}$ different values, where $\boldsymbol{n}$ is the number of bands of the spectral cube. Each pixel then, is regarded as a vector in a n-dimensional space. As a vector it has two attributes, namely length and direction. In this representation the length of the vector represents the luminance of the pixel, and the direction its spectral characteristics.

[5]described in sec.4.1

The spectral signature is the angle between the origin of the n-dimensional space and the vector. The classification process, consists of comparing each pixel-vector with some reference vectors, whereas their number defines the number of classes. Then each pixel is classified to the class represented by the reference vector, that leads to the smallest angle. The reference vectors are chosen with various criteria, such as:

- randomly in the spectral cube
- chosen from spectral libraries
- training dataset

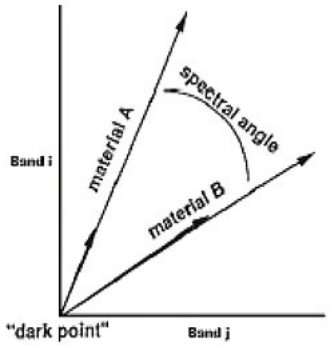


Figure 4.1: Spectral angle example

The spectral angle is computed using the following equations.

$$a = cos^{-1}\left(\frac{\vec{t}\cdot\vec{r}}{\|\vec{t}\|\cdot\|\vec{r}\|}\right) \tag{4.1}$$

or

$$a = cos^{-1}\left(\frac{\sum_{i=1}^{nb} t_i \cdot r_i}{\left(\sum_{i=1}^{nb} t_i^2\right)^{\frac{1}{2}} \cdot \left(\sum_{i=1}^{nb} r_i^2\right)^{\frac{1}{2}}}\right) \tag{4.2}$$

where $\vec{t}$ is the current pixel vector to be classified, $\vec{r}$ is the current reference vector and nb is the number of bands where the Spectral Angle a is calculated. In fig.4.2 we can see the block diagram of the algorithm implemented.

There are some main advantages towards choosing SpAM as our first classification algorithm. First of all, it is widely implemented [14, 20, 16, 21] and on the same time it has the ability to evaluate the spectra similarity while repressing the influence of the shading to accentuate the target reflectance characteristics [7]. In other words lighting issues, or shadows that occur during a the measurement process, have none or small effect on the computation and classification procedure.

This is very important is the case of plants, as various shades occur because of the leaf's physiology. So a wrinkle of the leaf, or a shade that could shift the chlorophyll absorption spectra, should not lead to different classification results. Another important factor of SPAM, is that many researchers suggest that the higher the number of classes, the better the performance of classification. This fact is also very important to our case, as utilizing a large number of classes, enables us to model different states of spectral responses. These states are modeled by specific reference vectors.

To conclude, non-uniform lighting of the specimen under investigation should not affect the classification process, and SPAM classification algorithm is supposed to succeed in fulfilling the requirements stated.

Nearest Neighbour

The Nearest Neighbour algorithm is also a supervised classification algorithm. We referred in the previous section in the n-dimensional space, where n represents the number of spectral bands. In the Nearest Neighbour the classification process consists of comparing each pixel-vector with a number of reference vectors. The criteria is based on the minimum distance between the current pixel-vector and the reference vectors. The distance is computed for each pixel-vector and each reference vector, and the pixel is assigned the class where the minimum distance is achieved.

The distance is determined using the following equation:

$$distance = \|\vec{t} - \vec{r}\| \tag{4.3}$$

where $\vec{t}$ is the current pixel vector to be classified and $\vec{r}$ is the current reference vector that is used in current comparison. In fig.4.2 we can see the block diagram of the implemented algorithm.

Nearest Neighbour is the most used classifier. It is used in many different applications and is also found in commercial packages. We choose to implement Nearest Neighbour Classifier as a reference to SPAM and other classification techniques mentioned. On the same time, we test the classifier's response to different number of classes, as this number plays an important role in the classifiers performance.

4.3.2 Vegetation Indices

The base of remote sensing in plants is estimating the leaf area index. An area is scanned by means of remote sensing and the data is processed to extract information regarding vegetation. The various indices defined (such as fig.3.14), do not make use of all the bands available with Hyperspectral Imaging Systems. Many research efforts try to examine the performance of the empirical indices, to overcome their poor performance across different species, lighting conditions, and phenomena under investigation. Right now, only the concentration and analogies of the major plant's pigments can be estimated.

In order to examine the prospect of classification in Plant Pathology, we also implemented two widely accepted and implemented indices to compute chlorophyll concentration and chlorophyll to carotenes ratio, that is ***mND*** and ***SIPI***. The two indices make use of the red-edge in 680-700nm that is proven as the wavelength where the chlorophyll absorbance is maximized, and 800nm where reflectance is maximized. Both indices take the 440nm band as a reference band that is related to the average spectral response due to the surface of the leaf etc. The spectral response of 440nm is a normalizing factor, and included in the equations to provide some immunity to various leaf structures, the phenomena that takes place in the leaf surface, such as scattering, existence of dust and water. Previous work and research, states that these indices also show immunity to different species.

The equations of the indices are (R stands for reflectance):

$$mND = \frac{R_{800} - R_{700}}{R_{800} + R_{700} - 2 \cdot R_{440}} \tag{4.4}$$

$$SIPI = \frac{R_{800} - R_{440}}{R_{800} + R_{680}} \tag{4.5}$$

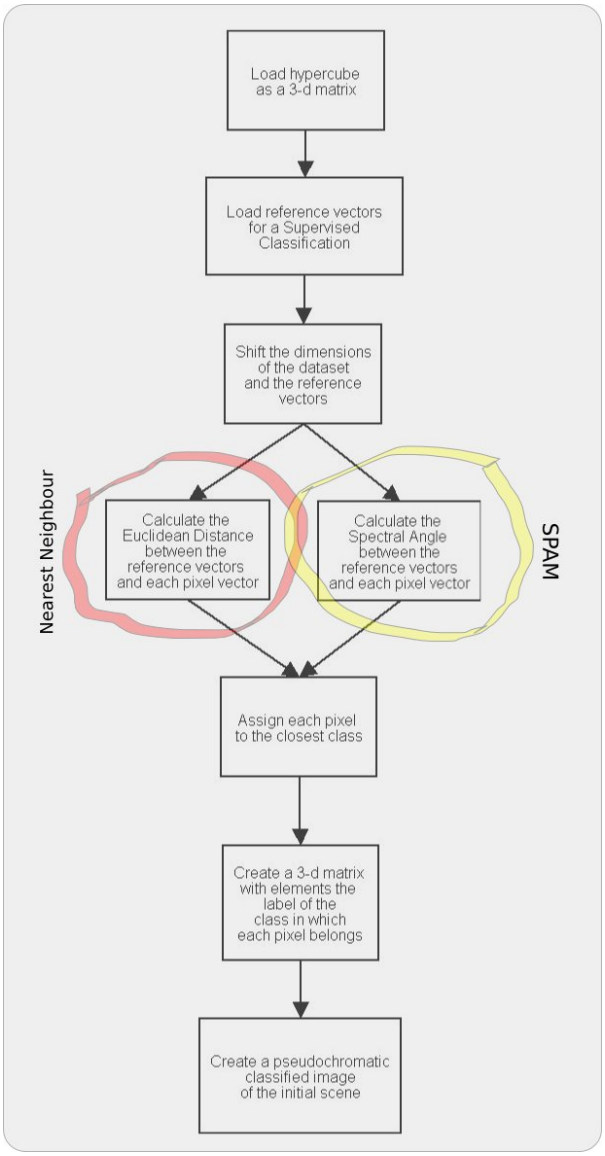


Figure 4.2: Block diagram of SpAM and Nearest Neighbour classification algorithms implemented.

4.4 Defining the Reference Vectors

In our study we also propose a method to extract reference vectors from our dataset. A common problem in comparing different classification algorithms and procedures in Hyperspectral Imaging, is declaring the reference vectors according to which the distance criteria is calculated. There are no known spectral libraries as far as plants in leaf level are concerned. Also no modelling of the chlorophyll decay is found. We do know from literature and many works, the spectra of a leaf in youth, when it is mature and in senescence, but due to instrument and measurement limitations, no information to the concentration or distribution of the chlorophyll in leaf level is found. Given the vast number of spectra given by a Hyperspectral Imaging System, it is vital to have a robust procedure when declaring the reference vectors.

In our attempt our aim is to declare reference vectors that represent various states of the leaf. We took advantage of the provided information regarding the spectra of the chlorophyll as far as aging is concerned. In our experimental procedure we tried to define various states of a damaged leaf. The decay of the chlorophyll is not uniform in the leaf area, as generally it starts from periphery and proceeds to the center of the leaf. When a stress factor or defect appear, the decay starts from the damaged area and proceeds in an outward motion.

Our method for estimating proper reference vectors, is described in table 4.4

The proposed procedure has a high cost in computational power. We try to correlate our knowledge based on theoretical research, to a plant that is a living being. The dynamic character of the experiment, implies that great effort is made when the vectors are taken. This is only possible by providing fast and robust computational tools. Limiting the computational time, was of vital importance, because of the enormous amount of iterations and experimentation within the dataset required.

The last step leads to the reference matrix for the leaf under investigation, to be used in the classification procedure. The reference matrix is different for each species. Given the species of the leaf, the reference matrix used is the same for various kinds of stress. In the last part of our study we discuss the results of the proposed method.

#	**Actions**
1	Find the first vector based on health criteria
2	Find the worst vector
3	Apply the SPAM and Nearest Neighbour classification algorithms with random vectors
4	Collect a number of vectors of a specific area of the leaf for each day
5	Sort the collected vectors, taking known first and last vectors as reference
6	We limit the number of the sorted vectors from step 5, based on the number of classes using a threshold to the distance criteria
7	The vectors are checked to be consistent with the literature
8	A reference matrix is consisted from the chosen vectors.

Table 4.4: Proposed method for extracting reference vectors from a given dataset, related to leaf senescence

Noise Vectors

Each spectral cube was also manually sampled to create reference vectors for noise implicated by background, shadows and soil. We created a large matrix consisting of various waveforms of noise found in our dataset, in an attempt to remove it, by classification terms. So our reference matrix also included noise vectors that were manually sampled.

Chapter 5

Results and Assumptions

We briefly gave our problem definition in sec.1.1. In chapter 4, we provided our materials and methods, where we explained our choices regarding the classification algorithms and indices used. In the following chapter, we provide our results and accordingly the assumptions. First of all, we provide two examples regarding the observation of the figures. In fig.5.1 we explain the figure of a SPAM of Nearest Neighbour result, and what the colorbar means. The value *0* represents noise, and values from *1* to *6*[1] represent the leaf's spectral progress from health to senescence[2].

As far as vegetation indices are concerned, in the case of mND, higher values in the colorbar represent higher concentration values of Chlorophyll, whereas in the case of SIPI, higher values represent higher Chlorophyll to Carotenes ratio, as seen in fig.5.2 .

Note 1 In this section we use two terms, *Full Bands* and *Less Bands*. *Full Bands* stand for all the spectral bands available in our cube, from 420nm to 1000nm. *Less Bands* stand for all the spectral bands from 420nm to 800nm. We examine the Classifiers performance using the reference vectors presented before the result figures. We discriminate between *Full Bands* and *Less Bands* to take into account the spectral information as listed in fig.3.9

Note 2 In the *Results* section, in each case at first we present the reference vectors used and we proceed in the results figures.

[1] or maximum class number, in case of greater number of classes

[2] exact vector selection procedure explained in sec.4.4

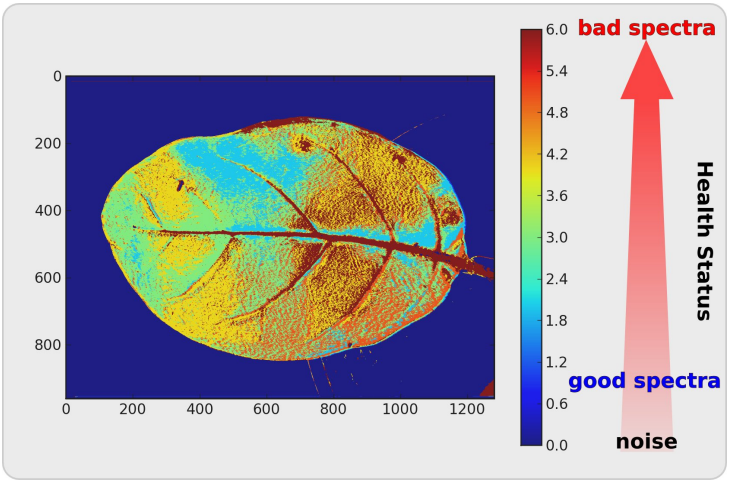


Figure 5.1: Explanation of a Classifier result figure. The colorbar is explained, as value 1 is a healthy class, value 6 is a damaged class and intermediate scale constitute the progress classes.

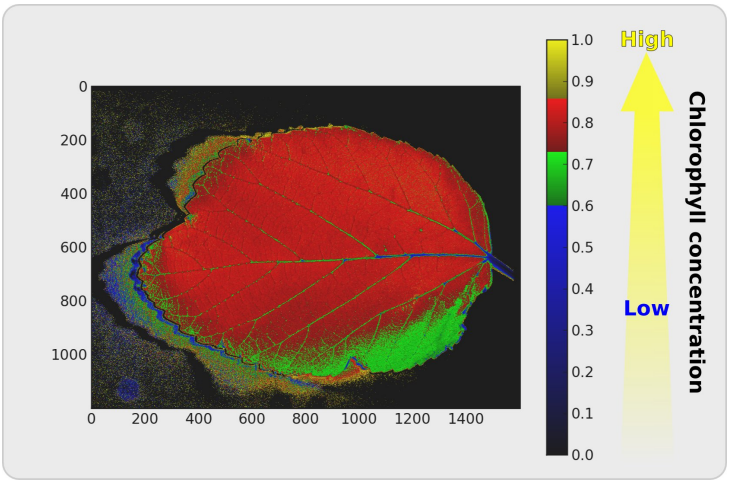


Figure 5.2: Explanation of a mND result. The colorbar is explained, as we proceed to higher values, higher concentration

5.1 Results

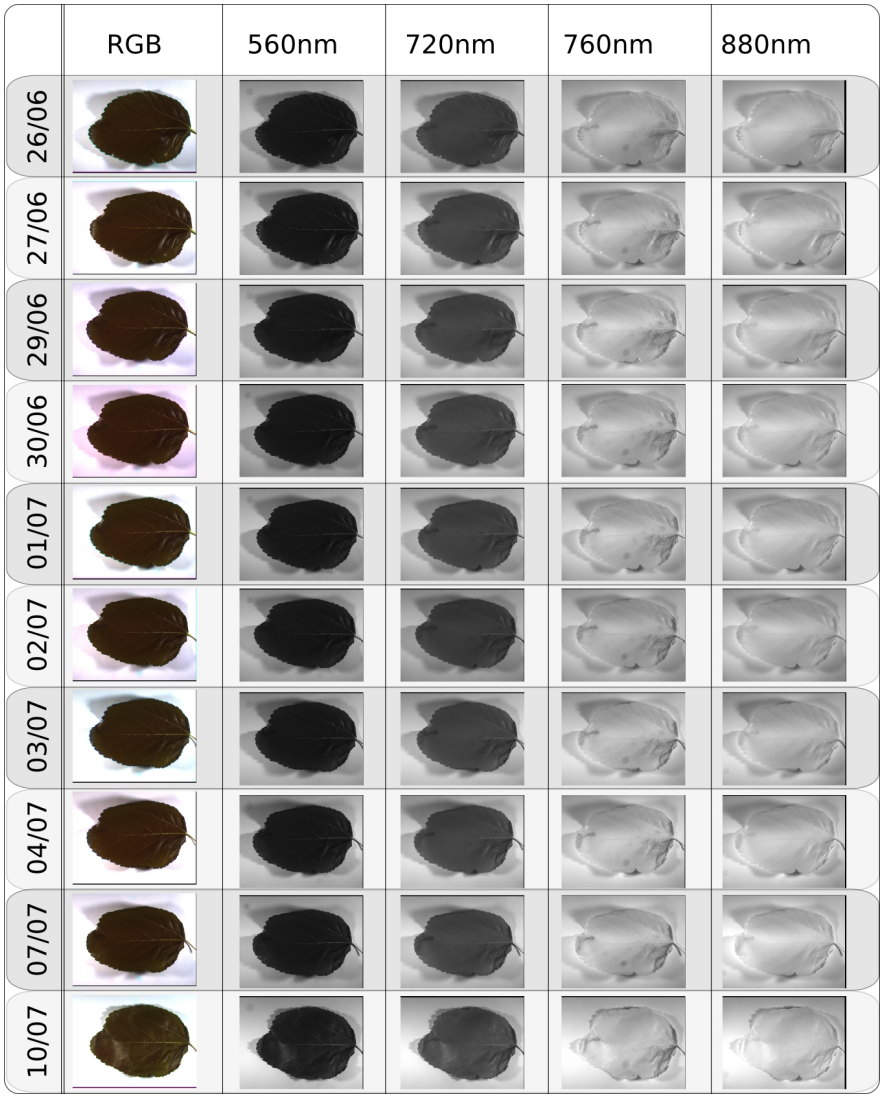
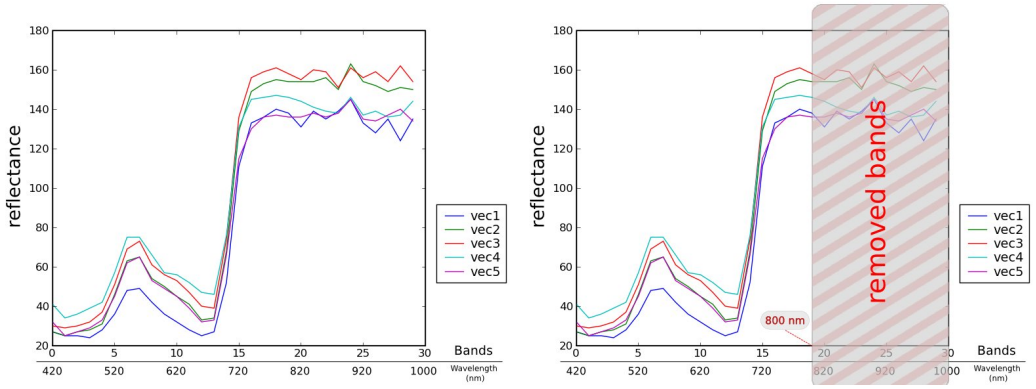


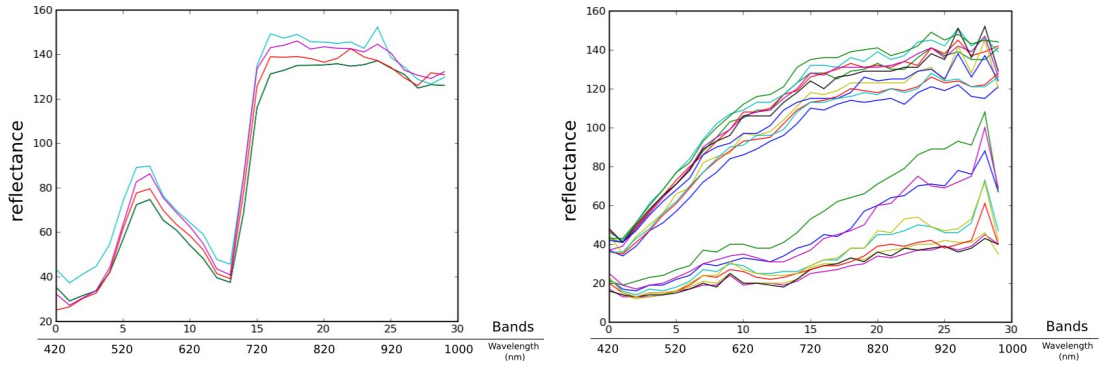
Figure 5.3: First mulberry leaf. We can see the RGB stills and stills of 4 discrete bands. A mark on the far-right edge is visible from day 26/06 that is not visible in the RGB stills.

5.1.1 Tobacco plants



(a) Full Bands (b) Less Bands

Figure 5.4: Reference Vectors (Tobacco Plants). In (a) are the reference vectors used in *full bands*, whereas in (b) the same vectors are band-selected to 800nm, so wavelenghts over 800nm are cut-off yielding *less bands*



(a) Vein Vectors (b) Noise Vectors

Figure 5.5: In (a) the vein reference vectors are plotted, and in (b) various noise vectors. We can see in (a) that the vein vectors have great more reflectance on 520-560nm compared to the reference vectors of tobacco in fig.5.4a

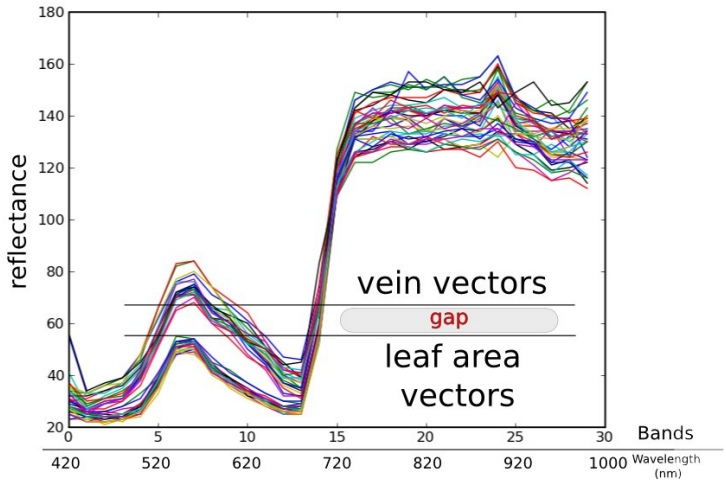


Figure 5.6: In this graph, various vectors from veins and leaf's green matter are plotted. We note the distinct gap in 520-560nm.

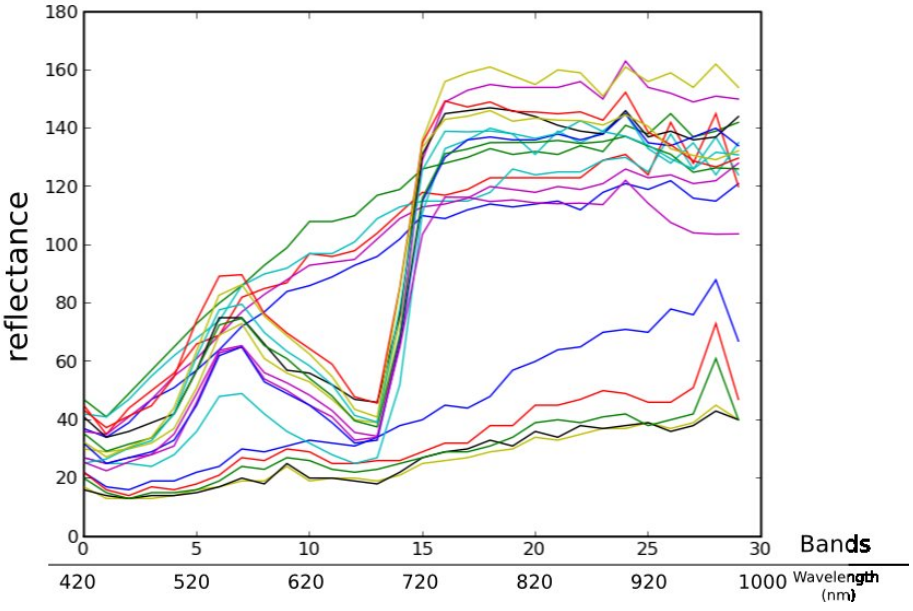


Figure 5.7: All Reference Vectors, total of 20 classes (Tobacco plants), as input during classification process in each iteration and classifier.

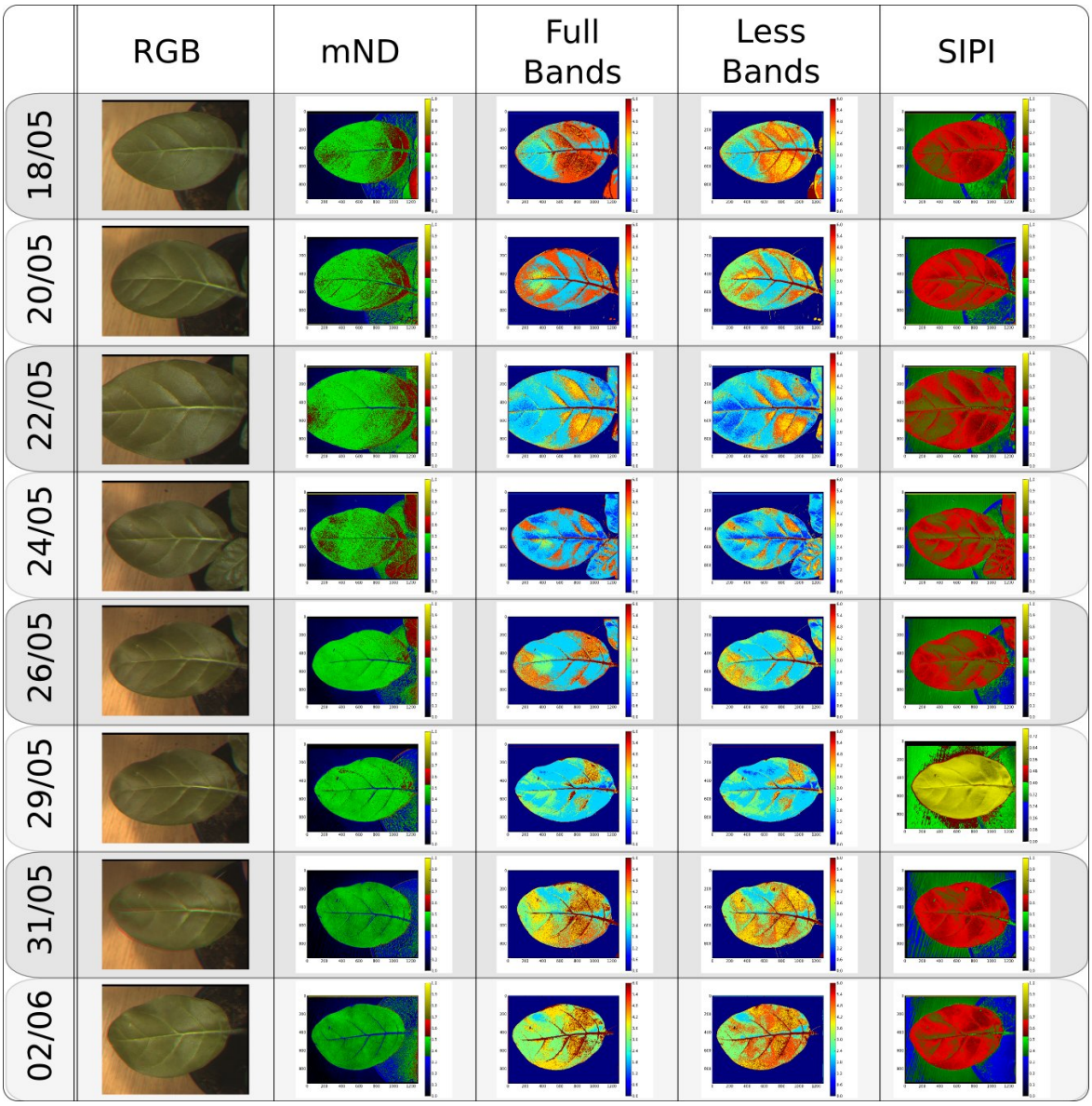


Figure 5.8: SPAM classifier 20 classes.*Fyllo1* of tobacco, healthy leaf. Noise and veins classes summarized to one class.

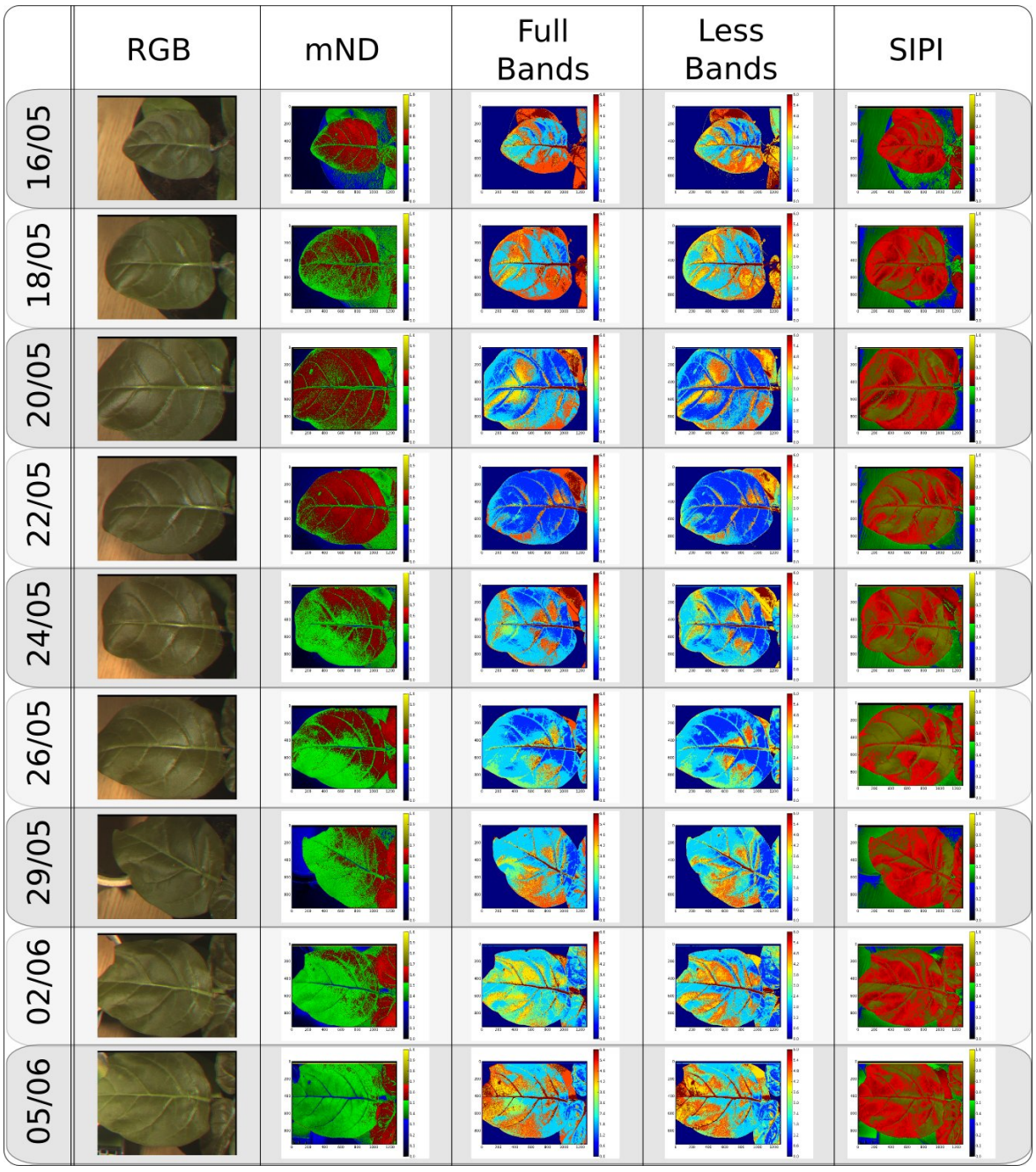


Figure 5.9: SPAM classifier 20 classes.*Fyllo2* of tobacco, damaged leaf in the center vein, near the petiole.Noise and veins classes summarized to one class.

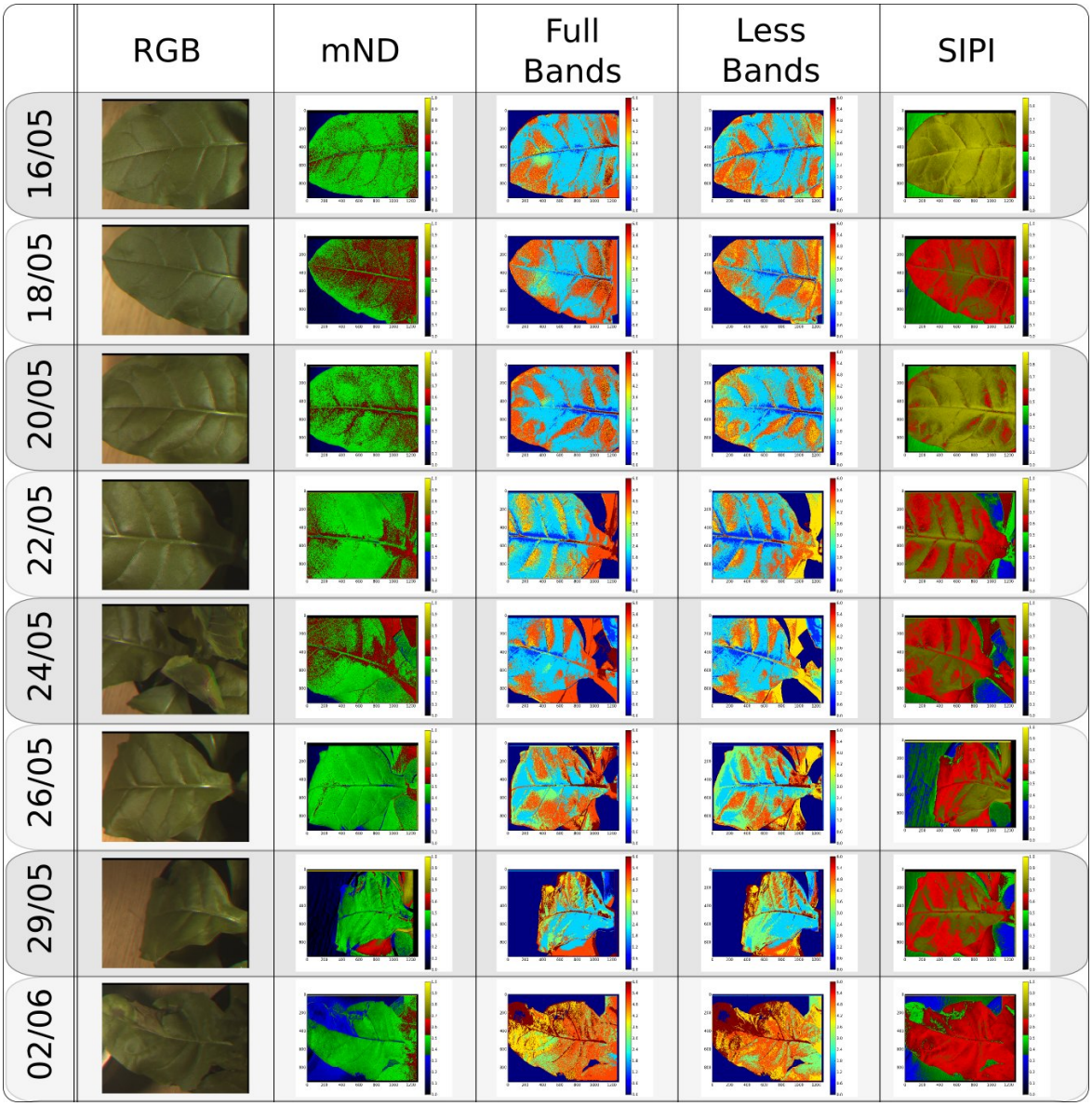


Figure 5.10: SPAM classifier 20 classes.*Fyllo3* of tobacco, damaged leaf in the far-right edge.Noise and veins classes summarized to one class.

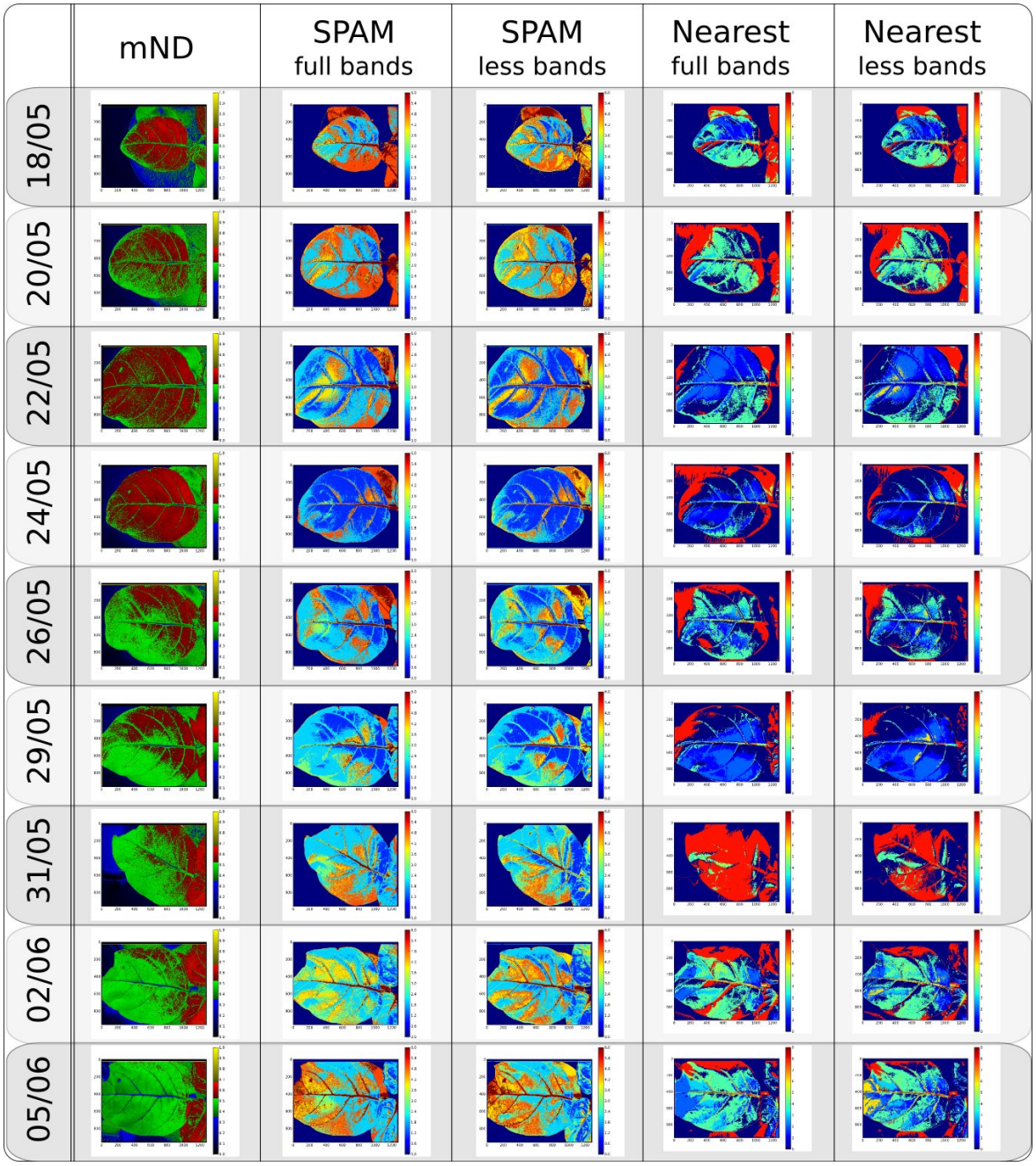
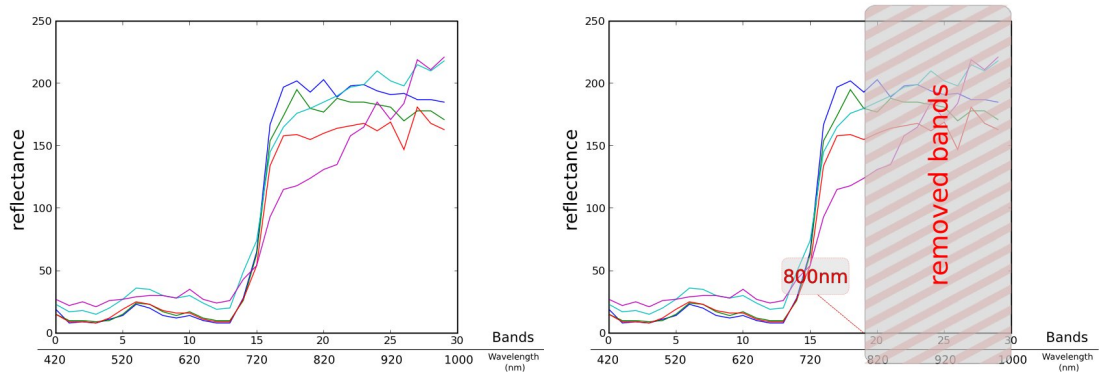


Figure 5.11: *Fyllo2* of tobacco, results from SPAM and Nearest Neighbour classifiers.

5.1.2 Mulberry Leafs



(a) Full Bands (b) Less Bands

Figure 5.12: Reference Vectors (Mulberry Leaves). In (a) are the reference vectors used in *full bands*, whereas in (b) the same vectors are band-selected to 800nm, so wavelenghts over 800nm are cut-off, yielding *less bands*

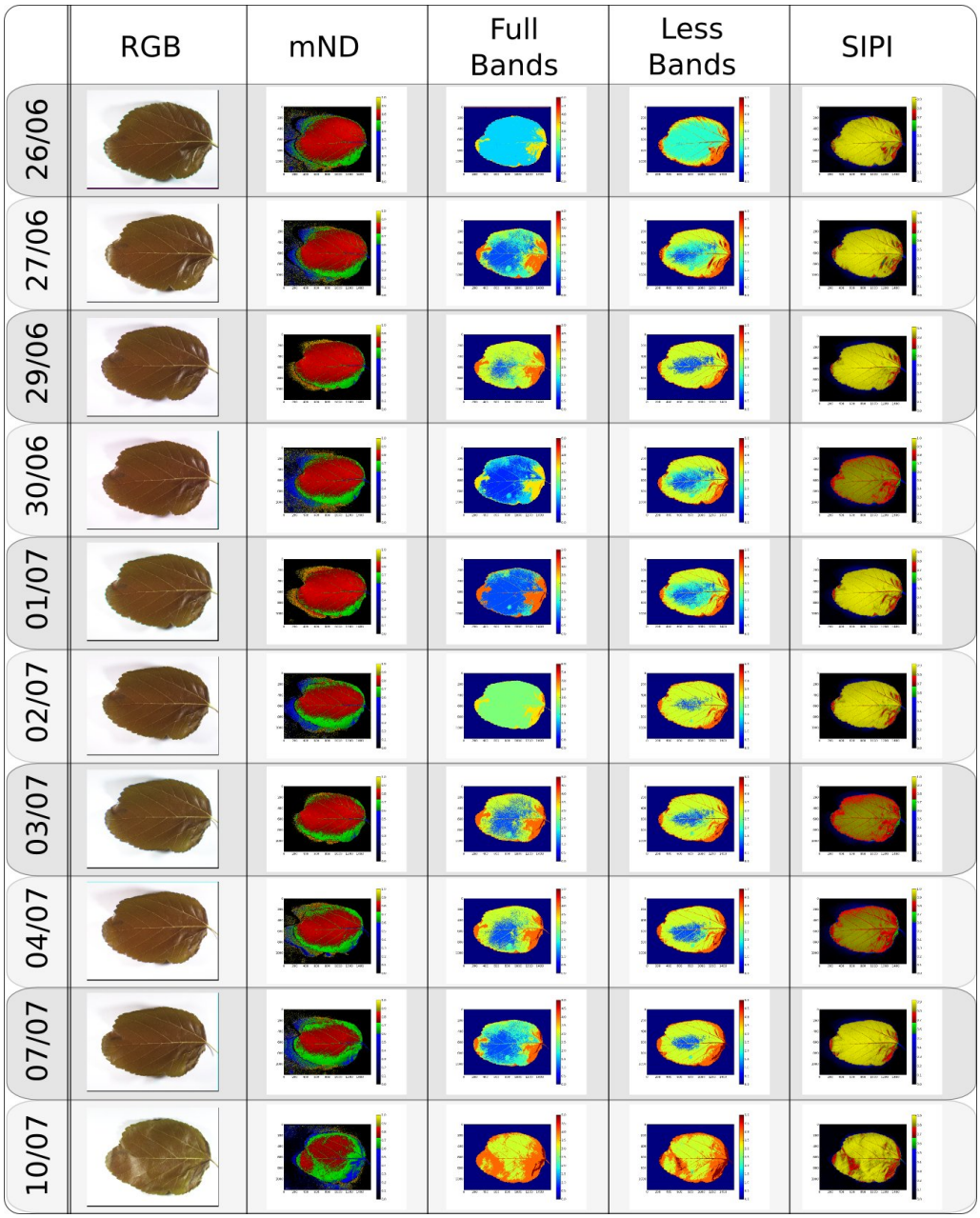


Figure 5.13: SPAM classifier.*Fyllo1* of mulberry. It was cut off from the tree and preserved in water, with its petiole kept off the water.

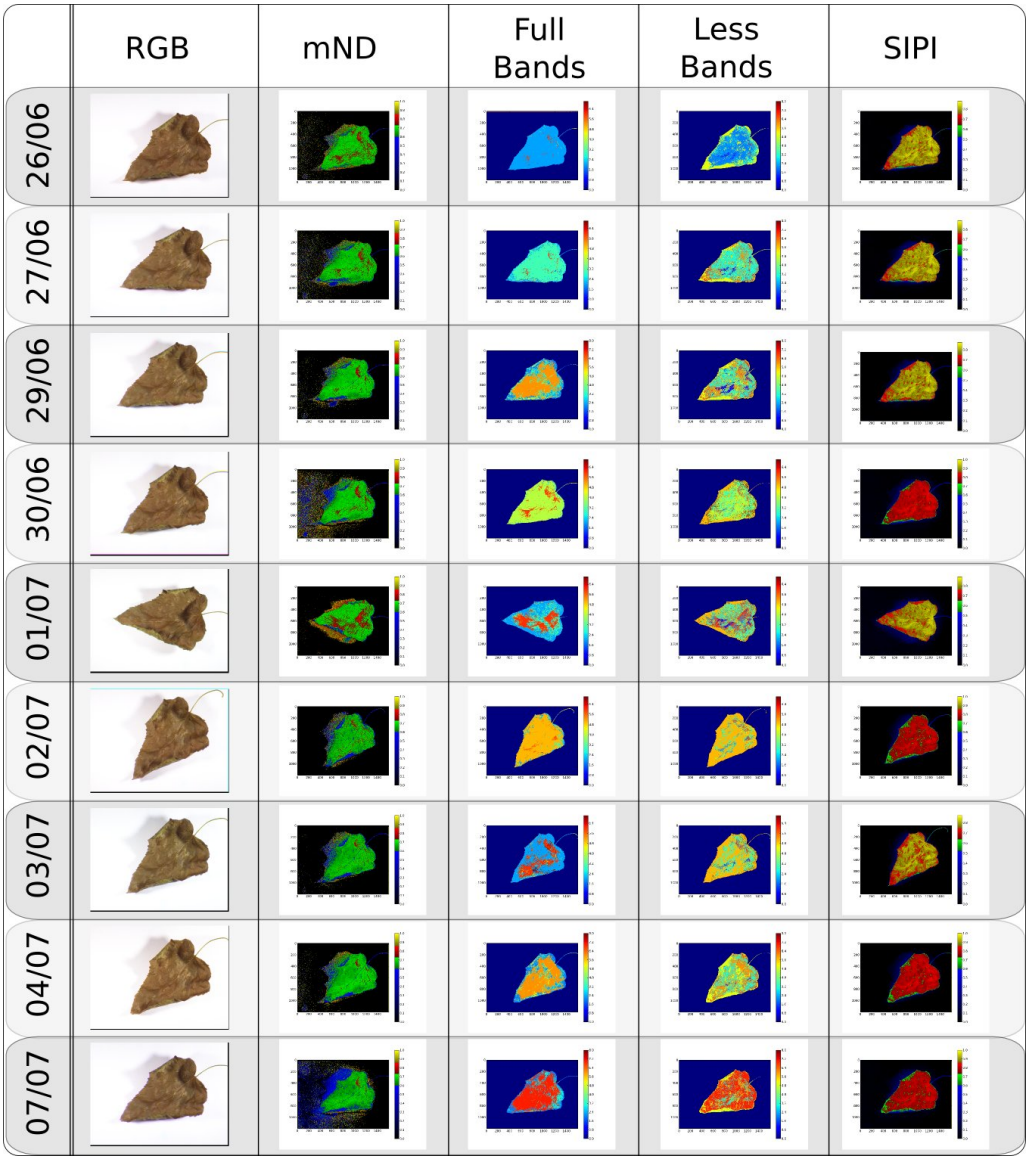


Figure 5.14: SPAM classifier.*Fyllo3* of mulberry. It was cut off from the tree and left exposed to environmental conditions of our lab.

5.1.3 Band Reduction Experiment

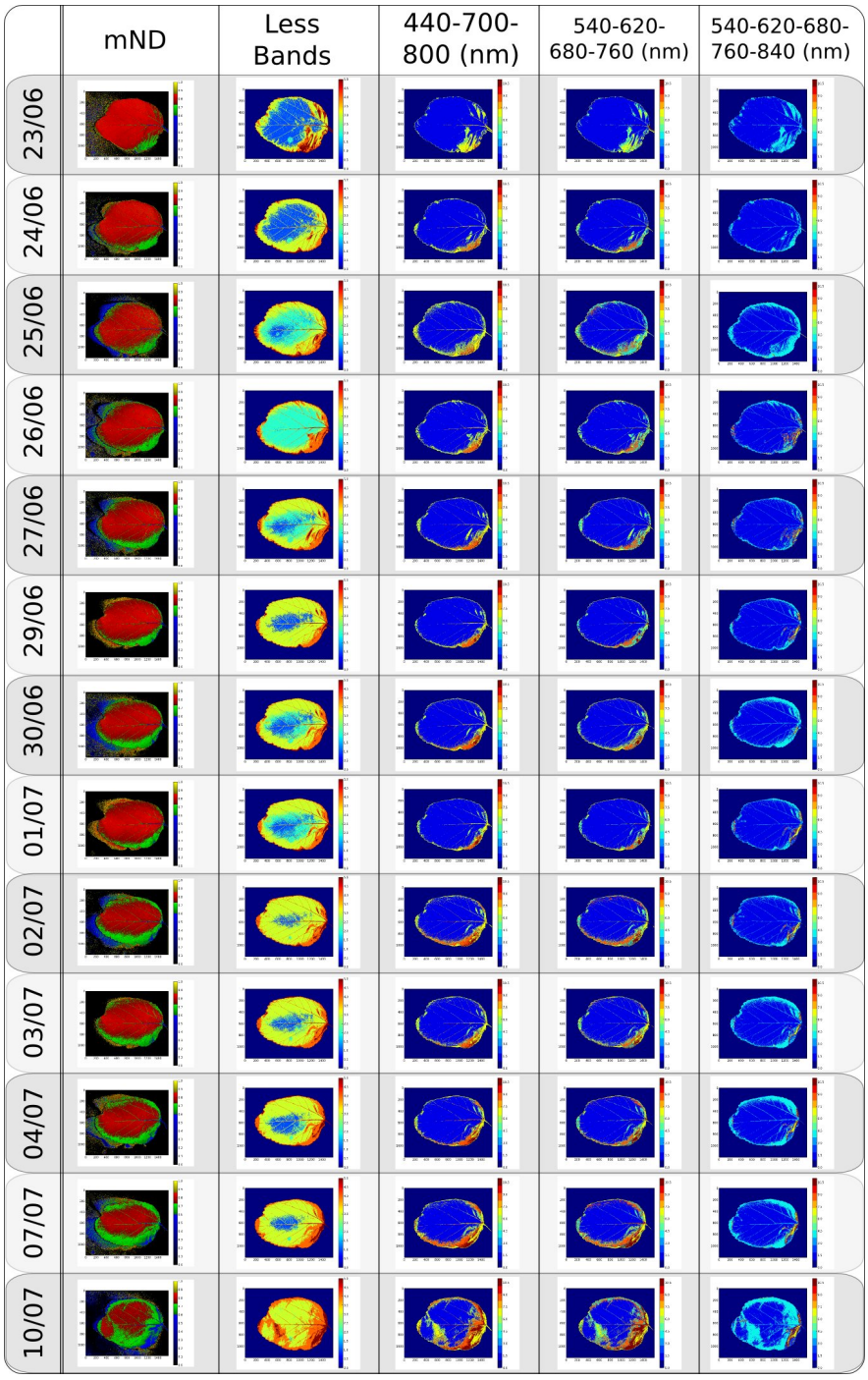


Figure 5.15: SPAM classifier.*Fyllo1* of mulberry. We compare the mND results to SPAM results taking into account the bands labeled in each case. We examine the similarity between the indice and the classification results, and the minimum number of bands required for the classifier to achieve the same results.(reference vectors listed in fig.5.12a). We see the high degree of resemblance between mND and 440-700-800(nm)

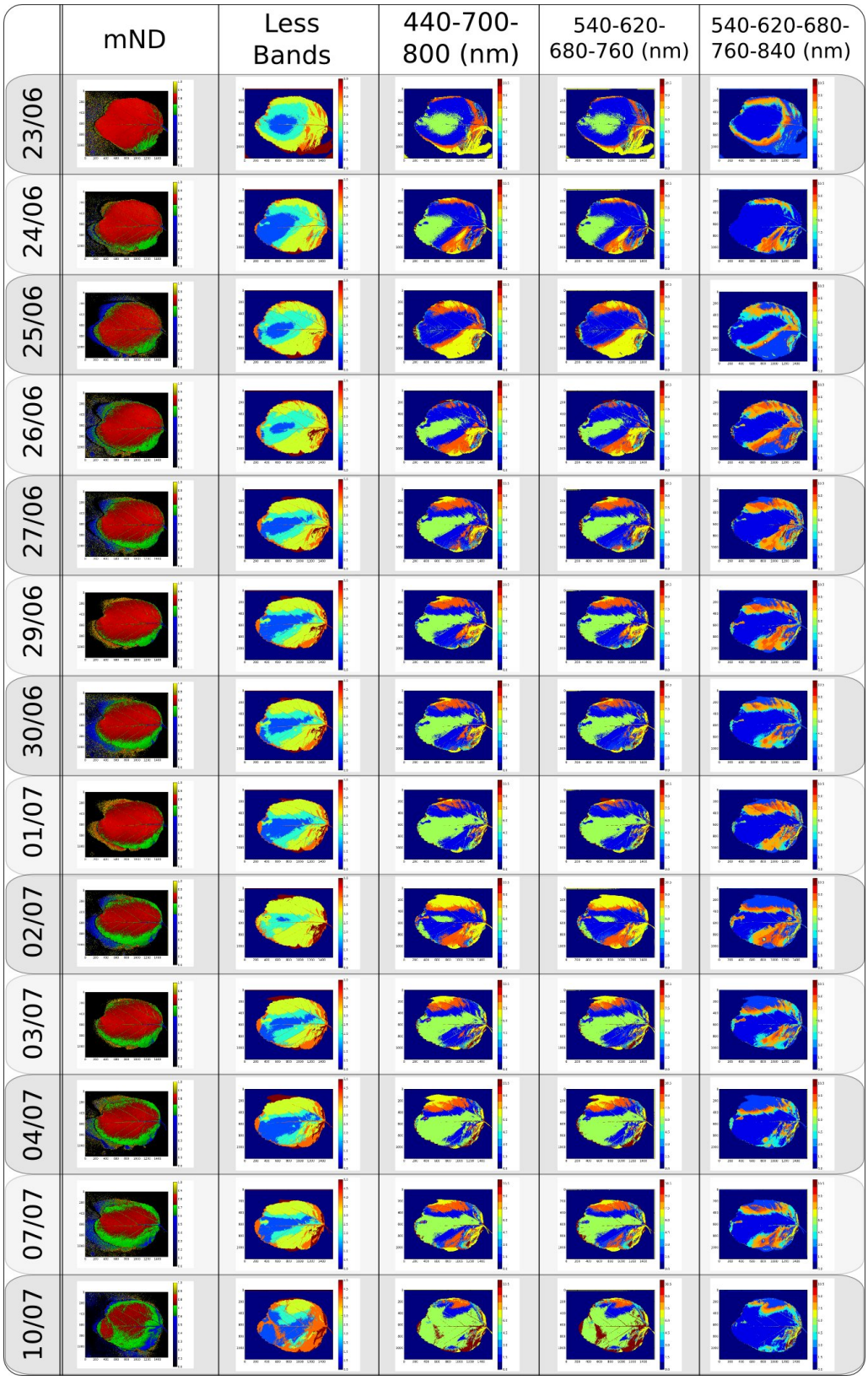
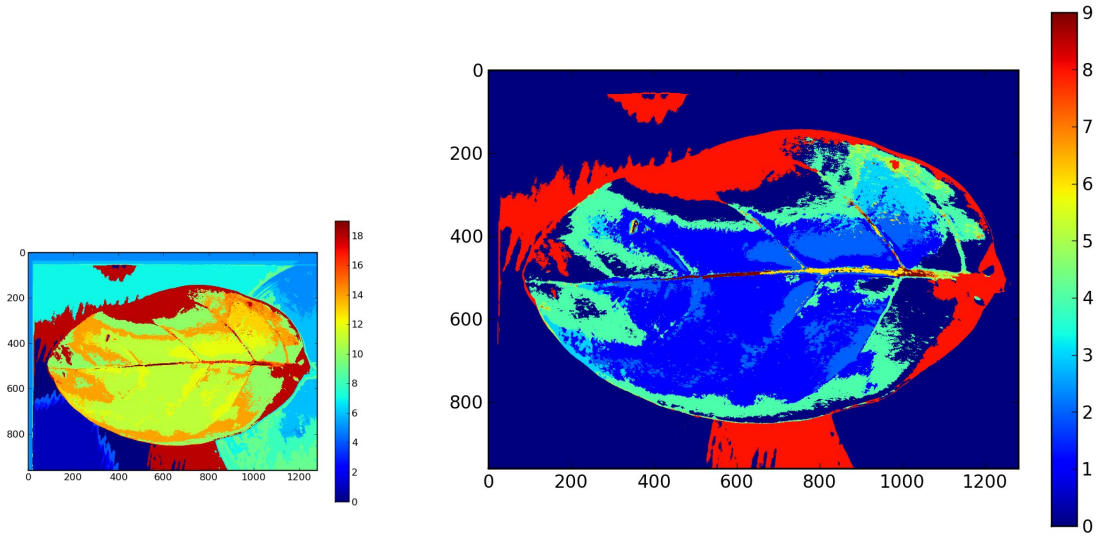


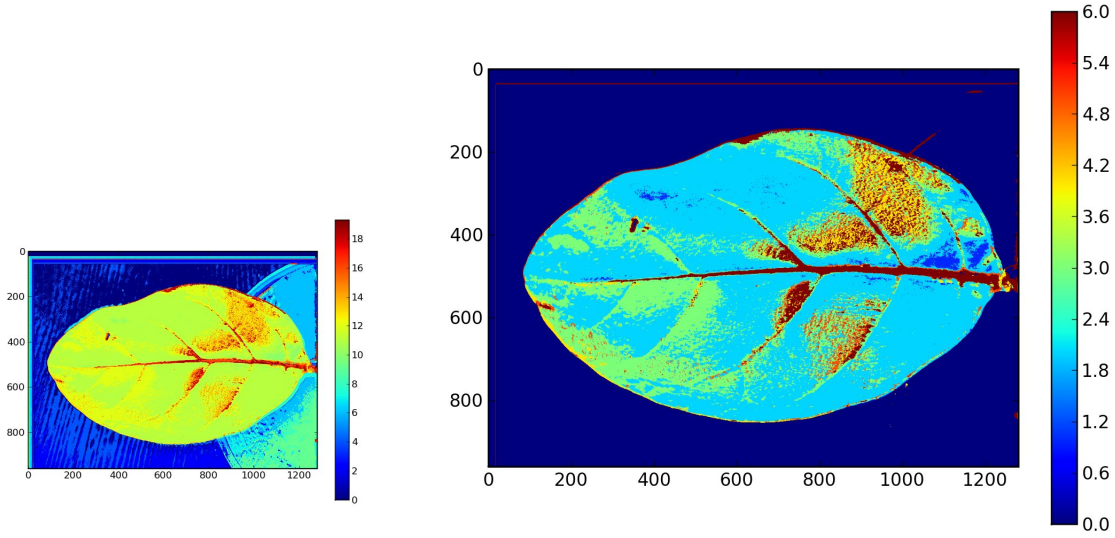
Figure 5.16: Nearest Neighbour classifier.*Fyllo1* of mulberry. Comparison of mND to Nearest Neighbour classifier. (The same as in fig.5.15 apply)

5.1.4 Noise Removal



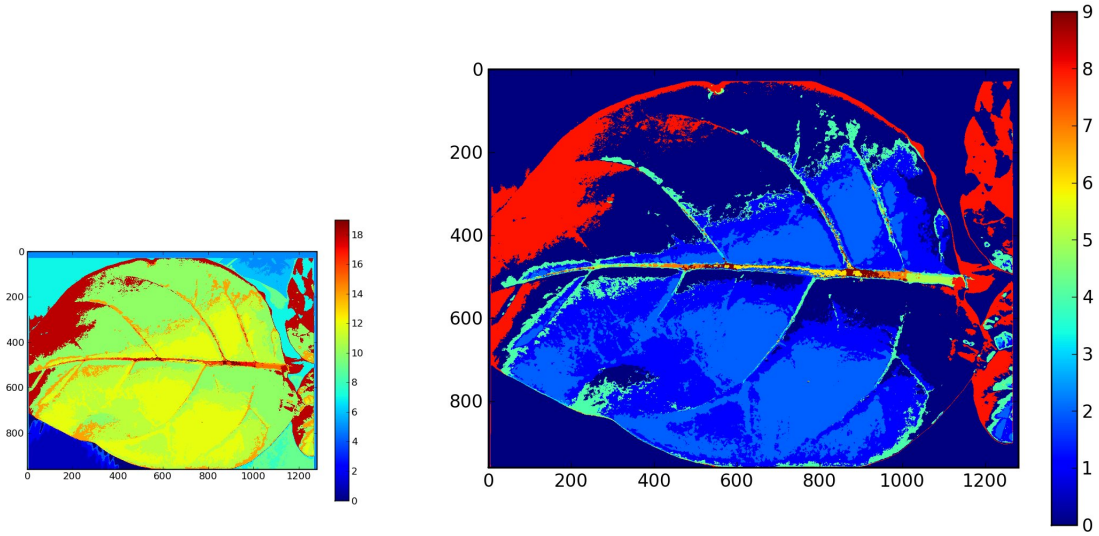
(a) NEAR Plus Noise (b) NEAR Without Noise

Figure 5.17: Fyllo1. Noise identification and removal, Nearest Neighbour classifier.



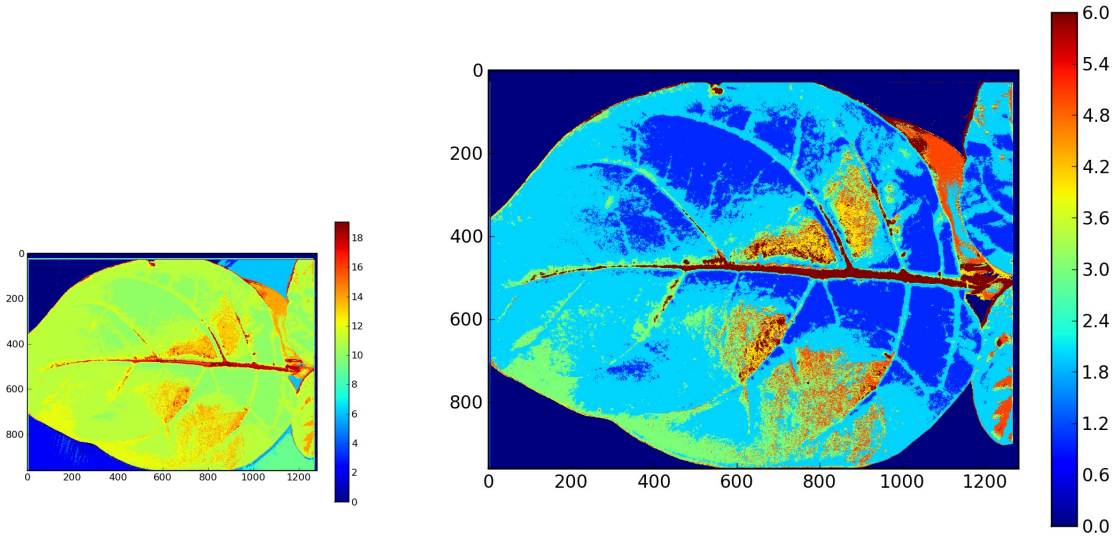
(a) SPAM Plus Noise (b) SPAM Without Noise

Figure 5.18: Fyllo1. Noise identification and removal, SPAM classifier.



(a) NEAR Plus Noise (b) NEAR Without Noise

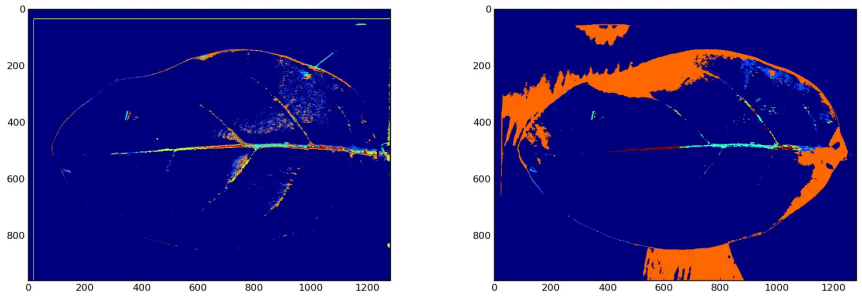
Figure 5.19: Fyllo2. Noise identification and removal, Nearest Neighbour classifier.



(a) SPAM Plus Noise (b) SPAM Without Noise

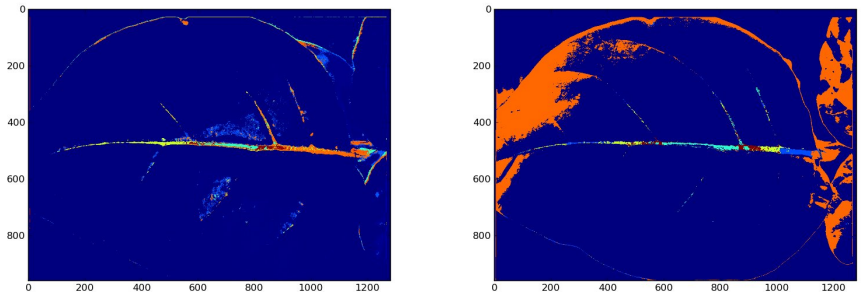
Figure 5.20: Fyllo2. Noise identification and removal, SPAM classifier.

5.1.5 Veins Classification



(a) SPAM (b) NEAR

Figure 5.21: Fyllo1. Veins Identification and comparison between the two classifiers (vein reference vectors plotted in fig.5.5a).



(a) SPAM (b) NEAR

Figure 5.22: Fyllo2. Veins Identification and comparison between the two classifiers (vein reference vectors plotted in fig.5.5a).

5.2 Assumptions

Our first question was about the ability to extend or enhance the available information of a leaf by means of Hyperspectral Imaging. Fig.5.3 shows that various defects are not visible to the human eye till the defect progresses, although through Hyperspectral Imaging we can clearly identify a dark spot

in the far-right corner of the leaf, from day 27/06 while it is not visible in RGB fotos, till day 01/07.

We can clearly see through all the result figures, that in the progress of measurements, there is a strong correlation between RGB fotos, and the results of both Classifiers and Indices. The chlorophyll concentration indice mND, depicts successfully what is visible through the progress of the experiments, for both species.

Comparing mND results to SPAM classifier (fig.5.8, fig.5.9,fig.5.10,fig.5.13,fig.5.14), we make two observations:

- We list the results of *Full Bands* and *Less Bands*, where it is clear that SPAM and *Less Bands* provide great discrimination ability, results coincide with mND, though more accurately.

- *Full Bands* and *Less Bands* differ in the case of Tobacco leaves, although the difference is very significant in the case of mulberry leaves.

This difference follows the difference observed in the reference vectors of mulberry leaves. We know that plant's pigments respond to light from 400-800nm, and the response from 800-1300nm is related to leaf's cell structure. Thus, the tobacco leaves were measured in situ and the response from 800nm to 1000nm is almost straight line. In the case of mulberry the response between 800-1000nm is not a straight line (fig.5.12), implying changes in cell structure that is consistent to the mulberry leaf's type of stress (nutritional, lack of water). We can also assume the cell structure is affected by looking the eq.3.1 and taking into account the photosynthesis procedure. If Photosynthesis is ceased, the protein complexes are decomposed.

Comparing Spectral Angle Mapper and Nearest Neighbour classifier (fig.5.11), we conclude that SPAM performs better. SPAM successfully overcomes lighting problems, and it performed better in determining and assigning pixels to correct classes. Nearest Neighbour classifier, is influenced by shadows and noise factors. Also it fails to provide the same discrimination efficiency as SPAM.

Taking into account the discrimination efficiency as far as noise and veins classification are concerned (fig.5.17,fig.5.18, fig.5.19, fig.5.20, fig.5.21, fig.5.22), it is clear that SPAM performs better than Nearest Neighbour.

Both vegetation indices have lighting issues. mND performs better and succeeds to provide valid information between species. SIPI on the other hand, fails to provide consistent results, as it is highly susceptible to light intensity alterations.

In fig.5.15 we can see that it is possible to get similar to mND results, by utilizing SPAM classifier in three bands 440-700-800(nm). Those bands were choosen based on the bands used on the indices and the peaks and deeps of leaf's spectra. More bands lead to more discrimination efficiency. On the other hand, information available regarding other factors of leaf, is lost. The trade-off is clear in fig.5.13, where the "inconsistency" between *Full Bands* and *Less Bands* is evident.

The selection of reference vector was successful. We can also see from fig.5.8, that the progress of a healthy leaf can also be monitored, providing the sufficient reference vectors.

Conclusions

The analysis of Hyperspectral Imaging cubes by classification means, lead to successful estimation of chlorophyll content, and identification of subtle differences is the spectra, as those between vein-vector and green-matter vector. Our work suggest that Spectral Angle Mapper classifier provides greate efficiency through different species and stress factors. Vegetation Indices such as mND, successfully estimate the chlorophyll content, although they suffer from light conditions and cannot take the most of the available spectral information.

Our work suggest that:

- Classification algorithms succesfully expand the capabilities of Hyperspectral Imaging in the field of Plant Pathology.
- SPAM classifier performs better than Nearest Neighbour and mND.
- It is possible to provide rigid results based on SPAM and a limited number of bands, although more tests are needed.
- We successfully classified and removed background noise.
- We are able to monitor the growth of a plant.
- Subtle changes in spectra are identified.

In terms of application field, our results indicate that a Hyperspectral Imaging System combined with classification algorithms, can be deployed in an automated plant monitoring environment, as in greenhouses, or mounted in a movable-platform in situ. As a tool for agriculturalist or plant pathologist, a spectral library could be developed. Experiments on different species and stress factors, could lead to extraction of unique spectral signatures that characterize the pathogen. Besides the pathogen's spectral signature, information regarding the growth and expansion of the damage can be collected and thus provide the base of modelling and simulating various diseases.

The impact of Hyperspectral Imaging on Plant Pathology is great due to in situ monitoring, the small cost of the method and finally in time warning of a stress or defect in a plant. Classification techniques highly improve the quality of information extracted. The efficiency of classification is related to the methodology of defining the reference vectors and the processing power required. Sophisticated software is crucial for classification as processing times may be very large and cost in time limits the experiment on analyzing spectral data. Taking advantage of recent developments in computers, could lead to real-time performance.

Chapter 6

Appendix A

The main drawback of classification techniques is their need for processing power. The dimensionality and thus the size of the spectral cubes and the complexity of the algorithms, make computation hard on time. In our case, data cubes loaded and ready for computation are of 120MB. Due to the plurality of cubes acquired during the three stages of our experiment, and the fact that no spectral reference existed, implied high experimentation on the collected dataset. Execution time, should be as small as possible, and reducing it was a crucial key to get results.

Optimized Matlab code for Spectral Angle Mapper and Nearest Neighbour classifiers, provided poor performance. Both algorithms initially took 25min to complete, when after optimization we reduced the time to 4min for both algorithms. Also a Graphical User Interface (GUI) was needed and we choose to change platform and use a software written in Python.

Software is released under GPL contrarily to the proprietary MATLAB. It is based in Python for the software engine, interface to the main routines, data structures and GUI. The main computation routine is written in a combination of Pyrex and C. The software is available under GPL.

The software used, succeeded in reducing dramatically the computational times and it also took advantage of multi-core CPUs. The end result is 25sec for both algorithms. It is a great improvement compared to MATLAB code, and enabled us to experiment vastly on the collected dataset. Our method for selecting the reference vectors, highly depends on computational time, as it is a costly procedure to exhaustingly compare all collected vectors to each other and sort them according to SPAM and Nearest Neighbourhood classifier.

Bibliography

[1] A.Krumov, A.Nikolova, V.Vassilev, and N.Vassilev. Assessment of plant vitality detection through fluorescence and reflectance imagery. *Advances in Space Research*, 41:1870–1875, 2008.

[2] C. Balas. A novel optical imaging method for the early detection, quantitative grading, and mapping of cancerous and precancerous lesions of cervix. *Biomedical Engineering, IEEE Transactions on*, 48(1):96–104, Jan. 2001.

[3] Harm Bartholomeus, Gerrit Epema, and Michael Schaepman. Determining iron content in mediterranean soils in partly vegetated areas, using spectral reflectance and imaging spectroscopy. *Applied Earth Observation and Geoinformation*, 9:194–203, 2007.

[4] Laury Chaerle, Sandor Lenk, Dik Hagenbeek, Claus Buschmann, and Dominique Van Der Straeten. Multicolor fluorescence imaging for early detection of the hypersensitive reaction to tobacco mosaic virus. *Journal of Plant Physiology*, 164:253–262, 2006.

[5] Laury Charle, IIkka Leinonen, Hamlyn G.Jones, and Dominique Van Der Straeten. Monitoring and screening plant populations with combined thermal and chlorophyll fluorescence imaging. *Journal of Experimental Botany*, 58:773–784, 2007.

[6] John A. Gamon Daniel A. Sims. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81:337–354, 2002.

[7] O. de Carvalho Jr., R. Guimaraes, R. Gomes, A. de Carvalho, N. da Silva, and E. Martins. Spectral multiple correlation mapper. *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*, pages 2773–2776, 31 2006-Aug. 4 2006.

[8] D.Moshou, C.Bravo, R.Oberti, J.West, L.Bodria, A.McCartney, and H.Ramon. Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using kohonen maps. *Real-Time Imaging*, 11:75–83, 2005.

[9] Cheng-Jin Du and Da-Wen Sun. Recent developments in the applications of image processing techniques for food quality evaluation. *Food Science & Technology*, 2004.

[10] G.D'Urso, L.Dini, F.Vuolo, L.Alonso, and L.Guanter. Retrieval of leaf area index by inverting hyper-spectral, multi-angular chris/proba data from sparc 2003. In *2nd CHRIS/Proba Workshop*, 2004.

[11] Baoxin Hu, James Freemantle, John Miller, and Anne Smith. Vegetation classification using hyperspectral and multi-angular remote sensing data. *IEEE*, 2007.

[12] P.J. Komi, M.R. Jackson, and R.M. Parkin. Plant classification combining colour and spectral cameras for weed control purposes. *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on*, pages 2039–2042, June 2007.

[13] Sandor Lenk, Laury Chaerle, Erhard E. Pfundel, Gabriele Langsdorf, Dik Hagenbeek, Hartmut K. Lichtenthaler, Dominique Van Der Straeten, and Claus Buschmann. Multispectral fluorescence and reflectance imaging at the leaf level and its possible applications. *Journal of Experimental Botany*, 58:807–814, 2007.

[14] I. McCubbin, R. Green, H. Lang, and D. Roberts. Mineral mapping using partial unmixing at ray mine, az. *7th Annual Airborne Earth Science Workshop Pasadena, CA, USA*, 1998.

[15] J.E. McMurtrey, E.M. Middleton, L.A. Corp, P.K.E. Campbell, L.M. Butcher, and C.S.T. Daughtry. Optical reflectance and fluorescence for detecting nitrogen needs in zea mays l. *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, 7:4602–4604 vol.7, July 2003.

[16] Bogliolo M.P, Teggi S., Buongiorno M.F, and Pugnaghi S. Retrieving ground reflectance from mivis data: A case study on vulcano island (italy). *1st EARSeL Workshop on Imaging Spectroscopy, Remote Sensing Laboratories, University of Zurich, Switzerland*, pages 403–416, 1998.

[17] Hamed Hamid Muhammed. Using hyperspectral reflectance data for discrimination between healthy and diseased plants, and determination of damage-level in diseased plants. In *31st Applied Imagery Pattern Recognition Workshop*, volume 0-7695-1863-X/02, 2002.

[18] Jan Corstiaan Noordam. *Chemometrics in multispectral imaging for quality inspection of postharvest products.* PhD thesis, Radboud Universiteit Nijmegen, 2005.

[19] Gerrit Polder. *Spectral imaging for measuring biochemicals in plant material.* PhD thesis, Technical University of Delft, 2004.

[20] Kaufmann Riaza, A Zock, and A Müller. Mineral mapping in makteshramon (israel) using dais 7915. *1st EARSeL Workshop on Imaging Spectroscopy, Remote Sensing Laboratories*, pages 365–373, 1998.

[21] L. L. Richardson and Kruse F. A. Identification and classification of mixed phytoplankton assemblages using aviris image-derived spectra. *Summaries of the 8th JPL Airborne Earth Science Workshop.*, 1998.

[22] Sebastiano B. Serpico and Gabriele Moser. Extraction of spectral channels from hyperspectral images for classification purposes. *IEEE Transactions on Geoscience and Remote Sensing*, 45, 2007.

[23] H. Tamhankar, L. Mann Bruce, B. Henry, and D. Shaw. Detection of moisture stress effects on plants using hyperspectral data. *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, 3:1529–1531 vol.3, June 2002.

[24] Chaichoke Vaiphasa, Andrew K. Skidmore, Willem F. de Boer, and Tanasak Vaiphasa. A hyperspectral band selector for plant species discrimination. *Photogrammetry & Remote Sensing*, 62:225–235, 2007.

[25] Chun-Chieh Yang, Kuanglin Chao, and Yud-Ren Chen. Development of multispectral image processing algorithms for identification of wholesome, septicemic, and inflammatory process chickens. *Journal of Food Engineering*, 69:225–234, 2005.