Technical University of Crete
School of Electrical and Computer Engineering
Electronics Laboratory

Smart Spectral Imaging for Material Identification

by

Vastaroucha Stergiani

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF POLYTECHNICAL DIPLOMA IN ELECTRICAL AND COMPUTER ENGINEERING

Thesis Commitee

Professor Balas Constantinos, *Thesis Supervisor*
Professor Garofalakis Minos
Associate Professor Samoladas Vasilios

Chania, October 2018

Abstract

Hyperspectral Imaging is a powerful analytical tool that enables the acquisition of a series of images in narrow spectral bands. This technique makes it possible to extract both spatial and spectral information about the scene under investigation. Therefore, it is widely used for non-destructive and non-invasive analysis in a variety of fields, ranging from food quality assessment to biomedical applications. Material Identification is the key to all these applications, which is achieved by using a library of spectral signatures of materials of interest and Spectral Similarity Measurements. Finding the right Spectral Similarity Measure is an important step and many studies have been conducted for their evaluation in terms of accuracy. However, in these studies the evaluation is made using members of the libraries (labeled data) and time performance is never assessed. This study proposes a series of steps that should be followed in Spectral Similarity Measures evaluation, including both labeled and unlabeled data for comparison. More specifically, reflectance measurements of various materials were gathered from online public spectral libraries to construct a database of spectral signatures. Furthermore, a series of Spectral Similarity algorithms were implemented and tested for their speed and accuracy. The accuracy of the algorithms has been assessed on the basis of their ability to produce a right match when an unknown spectrum is compared against the reference spectra of the database. The Spectral Similarity algorithms tested in this study are: a) SAM, b) ED, c) SID, d) SCA, e) SGA, f) SID-SAM, g) SID-SCA, h) AWN, i) SSS, j) NS3, k) SSD and l) SPM. Finally, hyperspectral measurements of skin lesions have been used as a test case for the final evaluation of the implemented comparison methods. The combination of a spectral database of references with the right Spectral Similarity algorithms can provide a valuable tool for material identification, with applications in a variety of scientific and industrial fields.

Acknowledgements

I would like to thank professor Costas Balas for his support and guidance through this thesis. I would also like to thank professor Minos Garfalakis and associate professor Vasilis Samoladas for their participation in the committee and the evaluation of this thesis.

I would also like to thank the members of the Electronics Laboratory of the Technical University of Crete for their guidance and suggestions and especially Giannis Gkouzionis, Thanasis Papathanasiou, Thanasis Tsapras, and Manos Vourdoulakis.

Special thanks go to my family for their support through all the years of my studies and to all the valuable friends I made during my study years at Chania.

Contents

List of Figures

List of Tables

To my family and friends...

Chapter 1

Preface

1.1 Thesis Outline

Chapter 2 provides a theoretical background of the of the principals behind spectroscopy and spectral imaging.

Chapter 3 gives an outline of the Classification methods used on Spectral Imaging data, as well as a detailed analysis of the Spectral Similarity Algorithms implemented for the purposes of this thesis.

In Chapter 4 the implementation methods used for the Spectral Library and Spectral Matching are analyzed.

In Chapter 5 a test case is used to present the results about the performance of the methods discussed on Chapter 4. The test case consists of Hyperspectral data acquired from human nevi.

In Chapter 6 the conclusions of this thesis are provided and the possible future research directions on the problem.

Chapter 2

Spectral Imaging

2.1 Electromagnetic Radiation

Electromagnetic radiation is the energy carried by electromagnetic waves. An electromagnetic wave consists of an oscillating electric field E and an oscillating magnetic field M. The two fields are perpendicular to each other as well as to the propagation direction of the wave. The key features describing an electromagnetic wave are its wavelength λ and its frequency f. The wavelength corresponds to the horizontal distance of a full oscillation and the frequency to the number of oscillations that occur per second. The two measures are inversely proportional, as the shorter the wavelength gets, the higher the frequency becomes.

This relationship of wavelength and frequency is described by Maxwells equation:

$$c = \lambda f, \tag{2.1}$$

where C is the speed of light, which equals to 299.792.458 m/s in a vacuum. With this equation, Maxwell proved that light is an electromagnetic wave.

Einstein, on the other hand, proved that electromagnetic radiation and therefore light can be treated as a continuous flow of wave energy packets, called photons. The energy content of each photon is given by:

$$E = hf = \frac{hc}{\lambda} \tag{2.2}$$

where h is Planck's constant ($6.6261 \times 10^{-34} Js$), c is the speed of light, v is the frequency and is the wavelength of the radiation.

The above two principals led to todays perception of light being both a wave and a flow of particles. When light travels it behaves as an electromagnetic wave and whenever it interacts with matter it behaves as a particle.

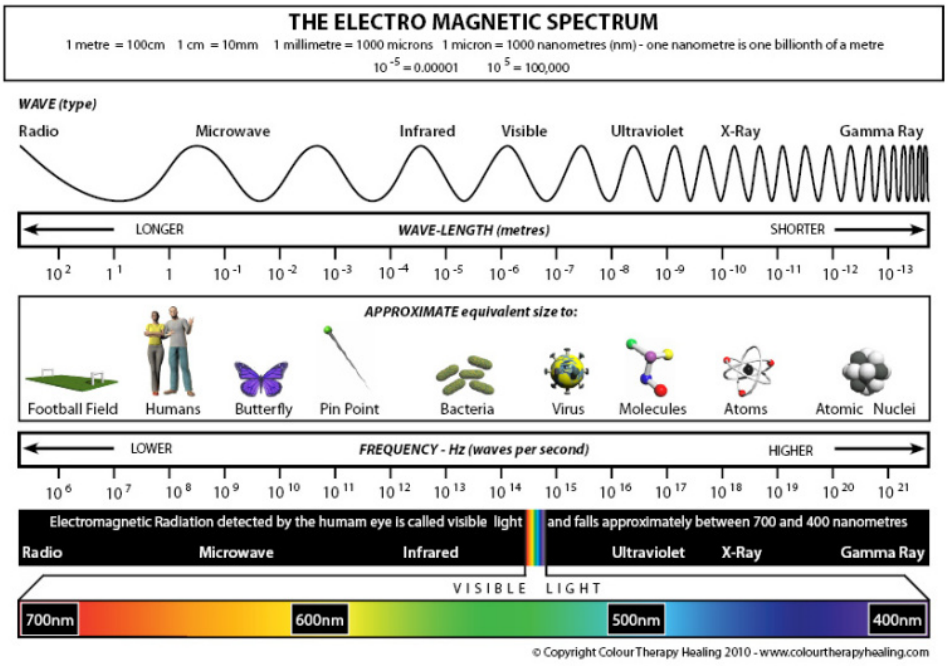


Figure 2.1: *The electromagnetic spectrum.*

The continuous range of frequencies of electromagnetic radiation is known as the electromagnetic spectrum. This range is divided into different regions, called bands, mostly on the basis of how electromagnetic waves of each region interact with matter. The electromagnetic spectrum ranges from the longest radio waves to the very short gamma rays. The transition from one band to another is gradient and not instant and therefore there are more categories in between bands. Visible light, as part of the electromagnetic spectrum, occupies only a narrow band of it, from 400nm to 700nm approximately.

2.2 Spectroscopy-Spectrometry

Spectroscopy refers to the study of how electromagnetic energy interacts with matter. This interaction can result in energy being absorbed, reflected, transmitted or scattered by matter.

Matter is comprised of atoms and empty space. When electromagnetic energy passes through a material it may interact with its atoms. The probability of photons interacting with the atoms of a material depends on numerous factors, such as the photon's energy or the atomic composition. When a photon does interact with a particle it transfers energy to it. The amount of energy transferred depends on the electromagnetic wave's frequency as described

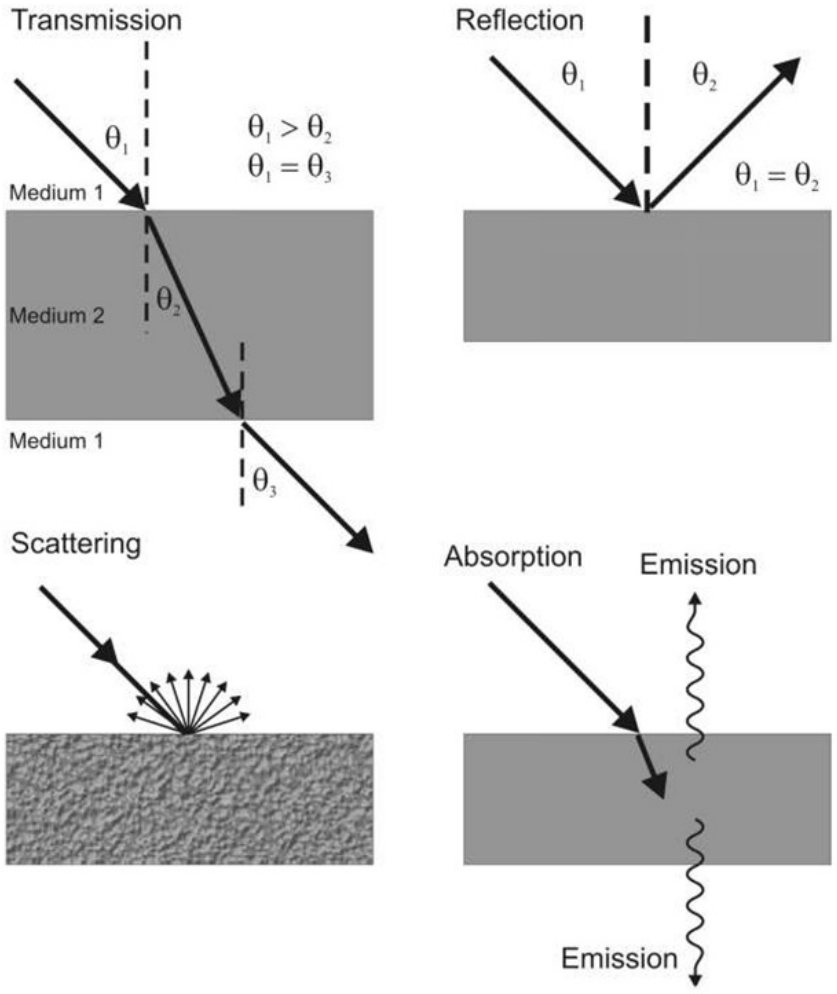


Figure 2.2: *Ways in which electromagnetic radiation interacts with matter.*

in Equation 2.2. Therefore, low energy photons belonging to the infrared side of the spectrum will only cause a vibration to the particles they interact with and as a result increase the heat of the material. On the other hand, photons that are part of the visible spectrum, have enough energy to cause outer atom electrons to get elevated to higher energy levels. Lastly, when photons of the x-ray and γ-ray bands of the spectrum interact with matter, they can cause the excitation of core atomic electrons.

Spectrometry is the technique used to measure this radiation to material interactions and produce quantifiable results. Essentially, spectrometry is the application of spectroscopy. Spectrometers measure the intensity of the light emerging from the sample as a function of the wavelength. Spectrometry is used in chemistry for the identification of substances, by analyzing the absorbed or reflected spectrum of these substances.

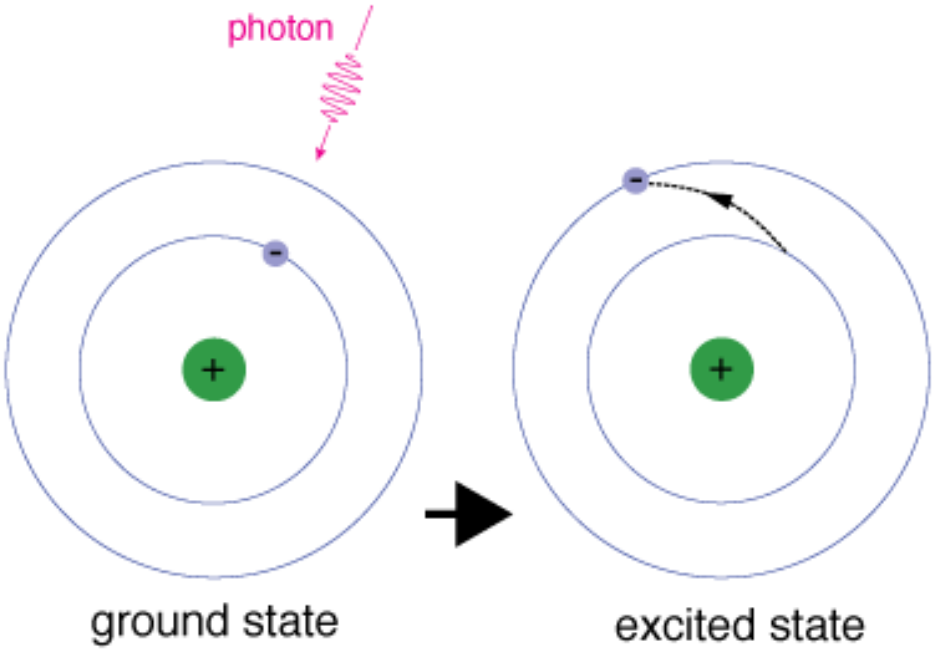


Figure 2.3: *Atomic excitation.*

2.2.1 Reflectance Spectroscopy

As was previously mentioned, reflection is one of the interactions that may occur between electromagnetic radiation and matter. When a beam of radiation hits the boundary that separates two mediums, a fraction of it bounces back into the initial medium. This phenomenon is called reflection. When the reflected beam has the same angle as the initial beam, then the reflection is called specular. In the opposite case, the reflection is called diffused.

Reflectance spectroscopy is widely used in scientific measurements as it can offer information about the materials comprising the measured sample. Since, the fraction of light that is not reflected, gets absorbed by the material, it can give an intuition about its chemical composition.

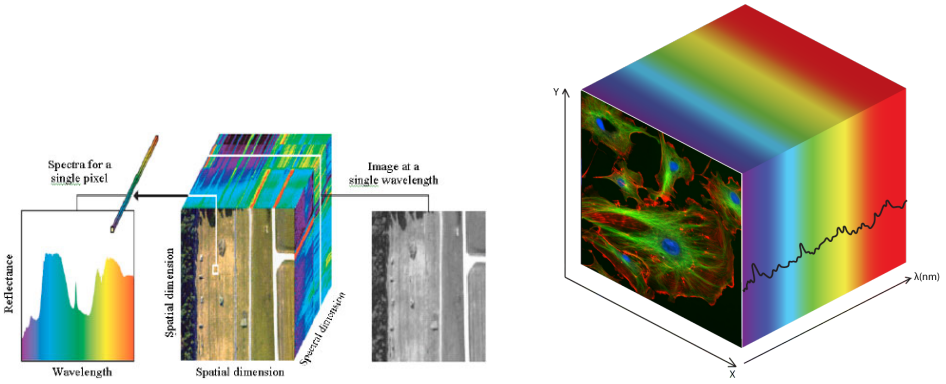
2.3 Spectral Imaging (SI)

Spectral imaging combines spectroscopy with photography in a way that enables spectral information to be collected at every location in an image plane. Spectral Imaging systems record light intensity as a function of both wavelength and location. Over the years, different techniques have been applied, creating categories of Spectral Imager. In regards to spectral range, two distinctions are made: Multispectral Imaging and Hyperspectral Imaging. Their main difference lies in the number of bands each system can acquire and how narrow these bands

can be. A Multispectral imager typically acquires fewer and wider bands than a Hyperspectral Imager.

2.3.1 Hyperspectral Imaging

HyperSpectral Imaging systems can acquire data in a wide range of continuous bands. The hyperspectral imager takes a full image at each individual wavelength. When these images are put together they form a three-dimensional data set of spectral and spatial information, known as a spectral or hyperspectral cube. In other words, a spectral cube can be viewed as a stack of images of the same scene, each of them representing a different wavelength. This way, a fully resolved spectrum can be recorded at each pixel, providing millions of individual spectra per scene.



(a) Spectal curve extracted from a pixel *(b) The coordinates of a spectral cube*

Figure 2.4: *Hyper-spectral Cubes*

It is this large amount of information contained in a measurement that makes Hyperspectral Imagery a powerful analytical tool, which has been widely used in a variety of applications like satellite or airborne remote sensing, industrial quality control, astronomy, military target detection, internal medicine, chemometrics, molecular biology and so many more.

2.3.2 Spectral Signatures

Each material has its own chemical composition that differentiates it from others. Different materials reflect differently and in different sections of the electromagnetic spectrum. This property makes it possible to uniquely identify through the spectrum in which it reflects. The radiation reflected from an object as a function of wavelength, that uniquely identifies it is called Spectral Signature.

In Hyperspectral Imaging spatial resolution is an important factor to take into account, while processing data from spectral cubes. When the spatial resolution is low, a single pixel may

contain many spectral signatures mixed together. In this case, special methods of Spectral Unmixing must be performed so as to reveal all hidden signatures. In most cases, though, and when spatial resolution is higher, a collection of pixels may represent the same spectral signature. This process is called Classification and it will be further discussed in Chapter 3.

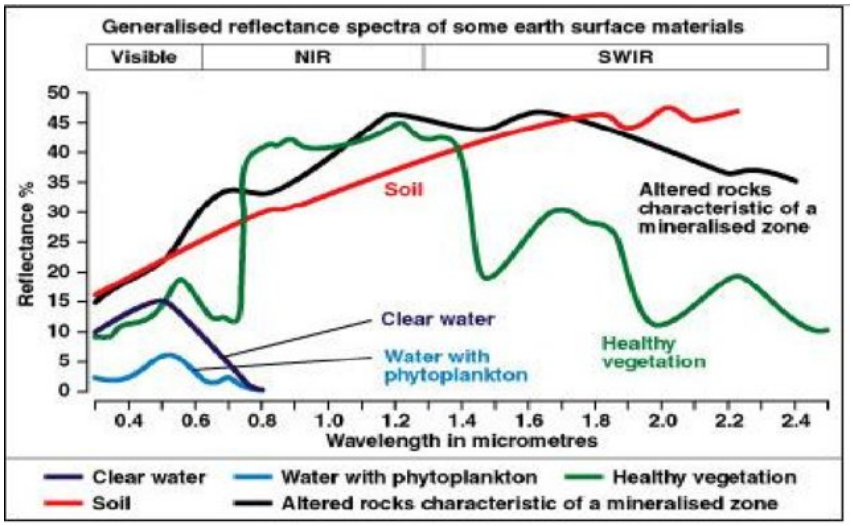


Figure 2.5: *Spectral signatures of different surface materials*

2.3.3 Target Detection and Material Identification

Hyperspectral Image Analysis is used for Anomaly Detection, Target Detection, and Material Identification. Anomaly detection is used to identify pixels that are different from the background. Target detection is used to identify pixels in the image that contain spectra of a known type, called target spectra. In that essence, known spectra from a spectral library are compared with the acquired spectra and pixels containing the target are separated from the rest. Essentially, in both anomaly and target detection pixels are either labeled as anomalies/targets or background.

Target detection is actually a binary classification technique. In material identification, on the other, a spectral library of known spectra is used to label pixels regardless of whether they belong to the background or not. Material identification can be viewed as target detection used multiple times over the entire scene. Spectral similarity measures are the most common technique used for material identification. In order to speed up the process of labeling, pixels are usually first grouped together based on their spatial and spectral information through classification (Multiclass Classification). In the following chapter the major principles of Spectral Classification are discussed.

Chapter 3

Hyperspectral Image Classification

3.1 Introduction

Classification comprises a basic step in Hyperspectral Image Analysis, as it enables the analyst to extract important information about the scene under investigation. As a process, classification aims to partition the pixels of an image into groups (called classes), such as those pixels of the same group have similar characteristics. Spectral classification techniques use the spectral information of every pixel in order to assign it to a predefined class. In general, the number of classes contained in an image should be exhaustive so as to reflect the complexity of the scene. Choosing the correct number of classes will guarantee that all pixels are correctly classified. Correct separability means that pixels within a class have similar spectral signatures, which are dissimilar to spectral signatures of neighboring classes.

The Classification process has two main stages. In the first stage, which is usually called Prototyping or Training, the number and nature of the classes are determined. In the second stage - usually called Identification or Labelling - every unknown element is assigned to one of the predefined classes, according to its level of similarity to the basic pattern. The result of image classification is a thematic map that shows the segmented regions, each of which corresponds to a class.

Generally, the classification techniques are traditionally divided into two categories: Unsupervised Classification and Supervised Classification. The division is done on the basis of the analyst' s involvement in the classification process. In Supervised Classification, an expert analyst with prior knowledge of the scene guides the classifier during prototyping and training of the system. On Unsupervised Classification, on the other hand, the classifier is trained using statistical information about the data. Unsupervised classification is preferred, when there exists no prior knowledge about the dataset under analysis. Other methods have been proposed in literature through the years, which are actually exploiting or combining features of

the two traditional Classification techniques. These methods are called Partially Supervised or Semi-Supervised Classifications and are preferred when a priori knowledge of the scene exists for some data, but not for all them. Recently, semi-supervised methods have gained attention, in an attempt to produce semi-automated classifiers. Before applying any technique, though, a series of preprocessing steps should be performed on the spectral cube. In the following section, some common preprocessing techniques are described.

3.2 Preprocessing

Hyperspectral measurement systems can cause unwanted effects on data during acquisition. In order to achieve a meaningful interpretation of data, there are several issues that should be handled before the analysis of a sample. Hyperspectral images acquired for different applications may require different preprocessing methods to be applied. Airborne hyperspectral data, for example, require atmospheric and topographic corrections, while in medical applications a common preprocessing step is the removal of artifacts that are not useful for diagnosis. The following methods are common preprocessing steps applied in the majority of applications.

3.2.1 Noise Reduction

Hyperspectral sensors, no matter how advanced they are, will add noise to the acquired data. Instrumental noise includes thermal, quantization and shot noise which causes corruption in the spectral bands by varying degrees. Noise will cause distortion to the original data and lead to vague results during the analysis. There are three main methods of noise reduction for hyperspectral data:

1. Smoothing Filtering
2. Image Transformation
3. Wavelet Transformation

3.2.2 Image Registration

A sudden movement of the measuring system or the hyperspectral sensor will cause the images of the spectral cube to be misaligned in regards to one another. The use of image registration algorithms helps to align all images of the cube according to a reference or target image.

3.2.3 Feature Extraction

As was mentioned in Chapter 2, a hyperspectral cube can contain millions of spectra. Processing all that data can be computationally intensive and time-consuming. Feature extraction is a method used to find key features that best describe a data set. A feature is a single element of a pattern. Transforming the image set to the feature set is often essential since it reduces the dimensionality of the data and simplifies the calculations performed by classifiers.

One of the most commonly used methods for dimensionality reduction is that of the Principal Component Analysis (PCA). This method computes an orthonormal basis derived from the eigenvectors of the covariance matrix, which in turn correspond to the largest eigenvalues (also known as Principal Components). Generally, the first several principal components contain most of the necessary information and the rest can be discarded with no great loss of information.

3.3 Unsupervised Classification

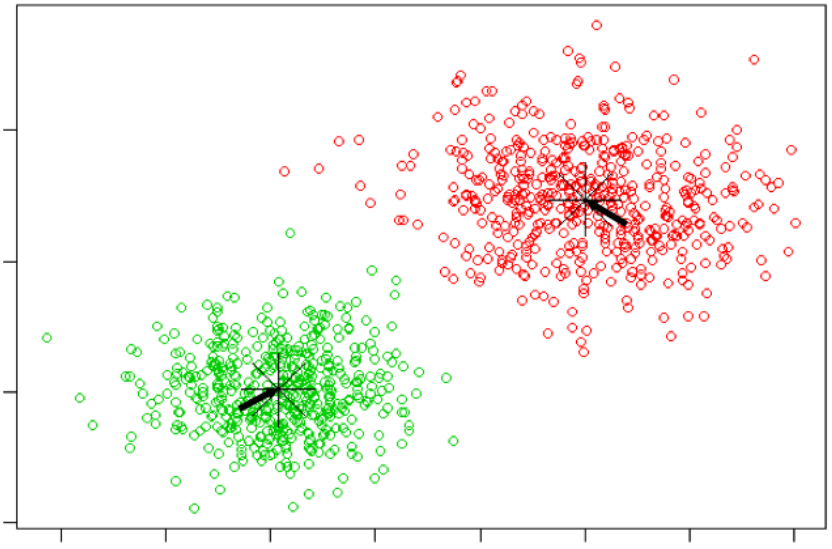


Figure 3.1: *Data points formed into classes using Unsupervised classification.*

The Unsupervised Classification method is used when the analyst has no a priori knowledge of the acquired scene and the classes that it might actually contain. In cases like these, where there is insufficient reference information available the unsupervised method helps to reveal the hidden structures of the hyperspectral cube. The user has to specify the number of classes and maybe some statistical measures, depending upon the algorithm used. The algorithms, then, use a predefined Spectral Similarity (Section 3.6) measure to create clusters of pixels that have similar features. When the clustering has completed the user has to spend some time to label

the generated classes based on his/her knowledge of the scene. However, because the clustering procedure does not require the analyst' s involvement, Unsupervised Classifiers are considered automated.

Clustering has been used for several decades in various fields for grouping data. There are numerous clustering algorithms that can be used to determine the classes present in the data set, each having its own characteristics. Some of the most popular clustering algorithms used are k-means, fuzzy C-means, ISODATA, statistical clustering methods, and the SOM (self-organizing feature maps), an unsupervised neural classification method.

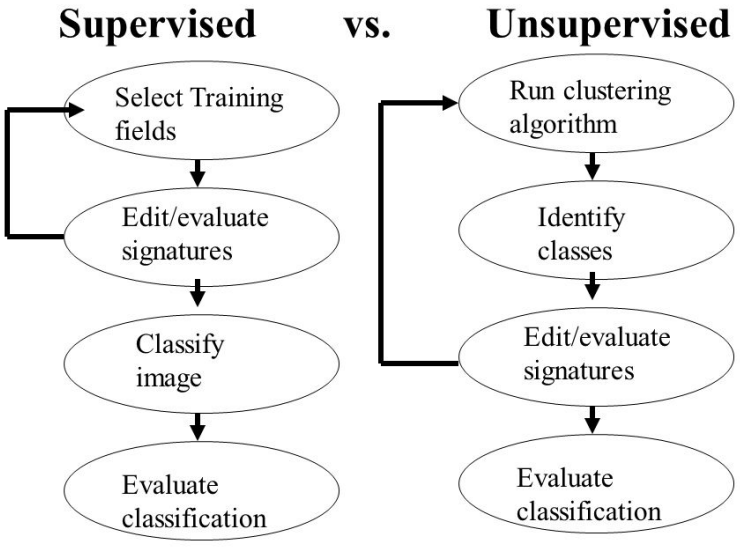


Figure 3.2: *Differences of Supervised and Unsupervised Classification.*

3.4 Supervised Classification

Supervised classification methods are based on the knowledge of the area to be classified. It may be defined as the process of identifying unknown objects by using the spectral information derived from training data provided by the analyst. In this method, the analyst selects and specifies representative samples on the image of a known cover type called Training Sites/Areas. The training samples should be selected from homogenous regions and cover the variability within the image. The extracted data are used to find the properties of each individual class. A computer software, then, uses the training set and classifies the whole image. The result of the classification is the assignment of unknown pixels to pre-defined groups. Ideally, the classification result should provide classes that do not overlap. The selected references that comprise the Training Areas play a critical role in Supervised Classification. If they are not accurate or

representative of the complexity of the image, then the classification will give inaccurate results. On the other hand, Supervised Classification performed with a good training set can give better and more accurate results than Unsupervised Classification.

Supervised classification is performed in three stages. In the first stage, called the training stage, the analyst defines the regions that will be used to extract training data, from which statistical estimates of the data properties are computed. At the classification stage, which is the second stage, every unknown pixel in the test image is labelled in terms of its spectral similarity to specified land cover features. As a result, a thematic map is produced, showing every pixel with a class label. Finally, in the third stage, the results of the classification are validated in terms of accuracy and performance. The accuracy of supervised classification is determined partly by the quality of the ground truth data and partialy by how well the set of ground truth pixels are representative of the full image. In order to measure the accuracy, it is common practice to use only part of the ground truth data for training the classifier and to use the remaining pixels for testing, that is to see if the classifier output corresponds to reality.

The size of the training data set is very important in supervised classification if statistical estimates are to be reliable. The analyst that selects the training sites must have very good knowledge of the variability in the image. The sample size is mainly related to the number of features whose statistical properties are to be estimated. Supervised classification methods require more user interaction, especially in the collection of training data. The majority of work is done before the actual classification of a hyperspectral image. In traditional Supervised Classification, the training set extracted from one hyperspectral cube cannot be used on an another. This is a timely and costly process that made researchers look for other ways to have the classification accuracy of Supervised methods in a more autonomous way.

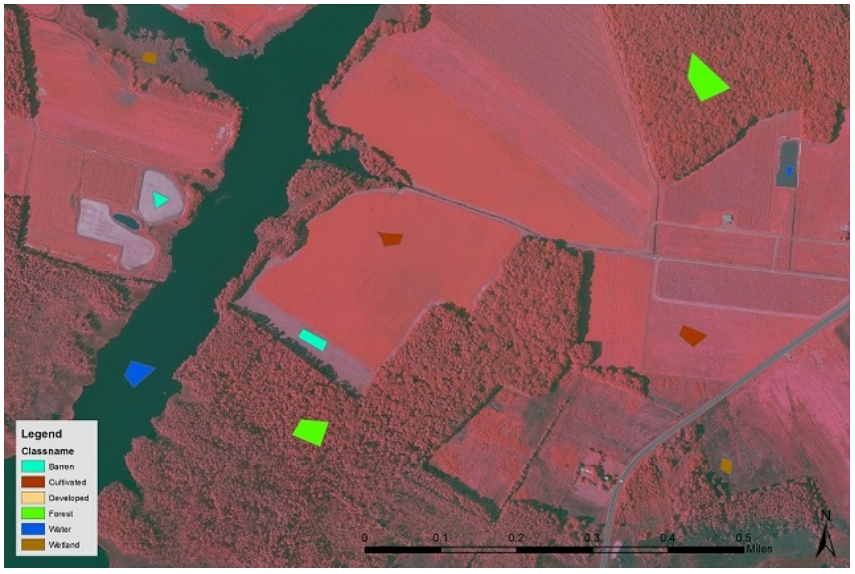


Figure 3.3: *Training Areas selection to perform Supervised Classification.*

3.5 Semi - Supervised Classification

The main difference between Supervised and Unsupervised Classification is that Supervised Classification uses a training set of spectral data which are previously labelled by an analyst. Unsupervised Classification algorithms, on the other hand, are used on unlabelled data and must determine the classes in an image on their own, based on statistics and patterns of the data. In general, supervised classification is more accurate when the training dataset is good, but it is time-consuming and costly. Semi-Supervised Classification exploits the two traditional methods, by using both labeled and unlabeled data for classification. Including unlabeled data on the classification process has proved to provide better classification results and improve the classification accuracy, while at the same time reduces time and cost. This way provides a semi-automatic classification procedure. Several approaches to semi-supervised classification have been proposed, which can be categorized as co-training, self-training or generative models. A common method used in hyperspectral imaging is the "Knowledge Transfer" model, which belongs to the generative category of Semi-Supervised models. In this model, the analyst uses reference data extracted from an already labeled image. The labeled data are used on the target image to determine if there are any spectral data present in this image and what that may be. Then an unsupervised classifier is trained on the unlabeled data in order to define the boundaries of the classes or even identify new ones. This way the analyst does not have to identify training areas from each and every image to be classified. However, this method requires that the labeled data used for the classification of the target image come from a nearby and preferably overlapping area.

Another semi-supervised approach is the "Cluster-then-Label" technique [34], which also belongs to the generative group of models. This method uses an Unsupervised algorithm first in order to cluster the image and produce class mean values. Then, the mean of each class is labeled using a Supervised algorithm and a set of already labeled data. Since the supervised method is used for a much smaller set of data compared to the millions that may be on a hyperspectral image, the whole process of classification becomes much faster. A spectral library can be used to assist to the labeling of the classes' mean spectrum or centroid.

Spectral libraries [34-35],[37] have been widely used by analysts to assist in the interpretation of spectral data. The benefit of using spectral libraries is that they allow reusability of already labeled data and of the spectral knowledge gained from images acquired at different periods of time. Furthermore, apart from labeled hyperspectral data they can also contain laboratory spectral measurements acquired with other equipment like a spectrometer. Spectral library search is emerging as a promising approach in semi-supervised classification models and especially in the field of material identification and mapping. Material identification by spectral library search requires:

1. a spectral library of reflectance spectra that contains the application specific materials,
2. appropriate Spectral Similarity metrics to perform the search in the library and find the right matches, and
3. performance evaluation criteria.

The quality of the spectra in the library is of key importance in the classification process. This is why a processing sequence to normalize the spectral data in the library should be followed so as to yield better results during the labeling process.

The following section describes the Spectral Similarity Metrics that are commonly used in Classification procedures and especially in spectral library searching.

3.6 Spectral Similarity Metrics for Material Identification

Spectral Similarity Metrics are used in all Classification approaches in order to quantify the variance between the spectral data at hand. Such algorithms compare two spectra with each other and produce a score of similarity. In the majority of cases, as this score tends to zero the similarity between the spectra increases, while a score equal to zero indicates a perfect match. Reflectance intensities are represented as n-dimensional vectors, where n is the number of spectral bands used during the acquisition. The magnitude of the vector corresponds to brightness, while direction corresponds to the spectral shape. Several spectral similarity measures have been proposed in the literature that can be categorized as either stochastic or deterministic. In this study, the following twelve algorithms have been used in order to evaluate their performance in a spectral library search.

3.6.1 Spectral Angle Mapper (SAM)

SAM [42] algorithm is one of the most widely used Spectral Similarity algorithms. This algorithm uses the dot product between two spectra and finds the angle difference between them. SAM algorithm is defined as:

$$SAM(\mathbf{x}, \mathbf{y}) = \arccos\left(\frac{< \mathbf{x}, \mathbf{y} >}{||\mathbf{x}||_2||\mathbf{y}||_2}\right) \tag{3.1}$$

where $\mathbf{x}, \mathbf{y}$ are two n-dimensional spectra, and n is the number of spectral bands. The resulted angle is expressed in radians and smaller angles represent a closer match. SAM is invariant to the intensity changes and relies heavily upon the spectral features to produce a result.

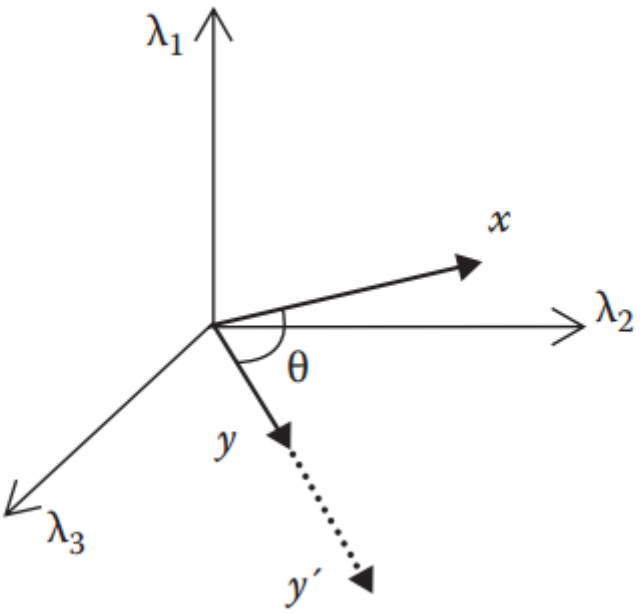


Figure 3.4: *Vector representations of two reflectance vectors* $\mathbf{x}, \mathbf{y}$ *in 3D orthonormal space created by three different wavelengthts. SAM measure calculates the angle* (θ) *between the two vectors.*

3.6.2 Euclidean Distance (ED)

The Euclidean Distance algorithm calculates the distance between two n-dimensional vectors of spectral curves using the following equation:

$$ED(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|| = \sqrt{\sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{y}_i)^2} \tag{3.2}$$

To remove the dependency on the number of the spectral bands, a normalized Euclidean Distance can be calculated by:

$$ED'(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{y}_i)^2} \tag{3.3}$$

Naturally, an increase of the distance between the two spectral curves means higher dissimilarity. ED is a measure that takes into account the intensity difference between the two vectors, which makes it more suitable in cases where spectra differ mainly in intensity characteristics.

3.6.3 Spectral Information Divergence (SID)

SID [43] algorithm is based on the concept of divergence of information theory. This algorithm considers each spectral curve used in the comparison process as a random variable and then measures the discrepancy of probabilistic behaviors between them. The definition of SID algorithm is:

$$SID(\mathbf{x}, \mathbf{y}) = D(\mathbf{x}||\mathbf{y}) + D(\mathbf{y}||\mathbf{x}) \tag{3.4}$$

where,

$$D(\mathbf{x}||\mathbf{y}) = \sum_{i=1}^{n} p_i \log_(\frac{p_i}{q_i}) \tag{3.5}$$

and

$$D(\mathbf{y}||\mathbf{x}) = \sum_{i=1}^{n} q_i \log_(\frac{q_i}{p_i}) \tag{3.6}$$

with p_i and q_i being,

$$p_i = \frac{x_i}{\sum_{j=1}^{n} x_j} \tag{3.7}$$

and

$$q_i = \frac{y_i}{\sum_{j=1}^{n} y_j} \tag{3.8}$$

The term $D(\mathbf{x}||\mathbf{y})$ is known as the KullbackLeibler information function and represents the relative entropy of $\mathbf{y}$ with respect to $\mathbf{x}$.

3.6.4 Spectral Correlation Mapper (SCM) and Angle (SCA)

Spectral Correlation Mapper [41] is a measure that effectively calculates a statistical measure of independence between two spectral vectors, which is known as Pearson' s correlation coefficient. In probability theory and statistics, correlation refers to the strength and direction of a linear relationship between two random variables. SCM metric is calculated by the following equation:

$$SCM(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n}(x_i-\overline{\mathbf{x}})(y_i-\overline{\mathbf{y}})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{\mathbf{x}})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{\mathbf{y}})^2}} \tag{3.9}$$

where $\mathbf{x}$, $\mathbf{y}$ are two n-dimensional spectral vectors, with n being the number of bands used during acquisition and $\overline{\mathbf{x}}$, $\overline{\mathbf{y}}$ being their respective means. The correlation coefficient produces values in the range [-1, +1], with -1 indicating a negative correlation, while +1 a positive one. SCM algorithm has been created as an improvement on the SAM algorithm, as it has the ability to centralize itself in the mean of $\mathbf{x}$ and $\mathbf{y}$. This means that SCM can distinguish between negative and positive correlations between the two vectors, while SAM takes into account only the absolute value. Using the Pearson' s correlation coefficient of Equation 3.9, the correlation angle is calculated by:

$$SCA(\mathbf{x},\mathbf{y}) = \arccos\left(\frac{SCM(\mathbf{x},\mathbf{y})+1}{2}\right) \tag{3.10}$$

where SCA $\in [0, \frac{\pi}{2}]$. As SCA angle gets closer to zero $\mathbf{x}$, $\mathbf{y}$ are becoming more similar.

3.6.5 Spectral Gradient Angle (SGA)

The Spectral Gradient Angle [71] is an algorithm that also uses the angle to produce similarity results. The difference lies in the fact that SGA takes into consideration the slope changes within the vectors used in the comparison. Given an n-dimensional vector $\mathbf{x}$, the spectral gradient is calculated as:

$$SG(\mathbf{x}) = (x_2 - x_1, x_3 - x_2, ..., x_n - x_{n-1}) \tag{3.11}$$

In order to calculate the gradient angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ the following equation is used:

$$SGA(\mathbf{x},\mathbf{y}) = SAM(abs(SG(\mathbf{x}), abs(SG(\mathbf{y})))) \tag{3.12}$$

where SAM is the Spectral Angle Mapper algorithm as that was defined in eq. 3.1. SGA is also invariant to illumination distance, just like SAM is and can discriminate materials with distinct reflectance features.

3.6.6 SID - SAM

SID - SAM [45] algorithm is a combination of the two separate algorithms in one. SID - SAM has two expressions:

$$SID - SAM_{tan}(\mathbf{x}, \mathbf{y}) = SID(\mathbf{x}, \mathbf{y}) \tan SAM(\mathbf{x}, \mathbf{y}) \tag{3.13}$$

or

$$SID - SAM_{sin}(\mathbf{x}, \mathbf{y}) = SID(\mathbf{x}, \mathbf{y}) \sin SAM(\mathbf{x}, \mathbf{y}) \tag{3.14}$$

3.6.7 SID - SCA

Another combinational algorithm is that of SID - SCA [46] and it too can be expressed in two different ways:

$$SID - SCA_{tan}(\mathbf{x}, \mathbf{y}) = SID(\mathbf{x}, \mathbf{y}) \tan SCA(\mathbf{x}, \mathbf{y}) \tag{3.15}$$

or

$$SID - SCA_{sin}(\mathbf{x}, \mathbf{y}) = SID(\mathbf{x}, \mathbf{y}) \sin SCA(\mathbf{x}, \mathbf{y}) \tag{3.16}$$

3.6.8 Adaptive Wiener Normalization (AWN)

Wiener estimation [36] has been widely used as a reflectance reconstruction technique. A modified Wiener method was proposed in 2007, that requires no prior knowledge of the spectral characteristics of the sample in order to reconstruct its reflectance. In this method, the spectral similarity measure of Equation 3.17 was proposed in order to select the training samples that are more similar to the estimated reflectances and which will be subsequently used to calculate the correlation matrix of reflectances.

$$AWN(\mathbf{x}, \mathbf{y}) = a \cdot mean|\mathbf{x}' - \mathbf{y}'| + (1 - a) \cdot max|\mathbf{x}' - \mathbf{y}'| \tag{3.17}$$

where $\mathbf{x}'$ and $\mathbf{y}'$ are n-dimensional vectors of reflectances normalized to the sum as follows:

$$\mathbf{x}' = \frac{x_i}{\sum_{i=1}^{n} x_i} \tag{3.18}$$

$$\mathbf{y}' = \frac{y_i}{\sum_{i=1}^{n} y_i} \tag{3.19}$$

so as $\sum_{i=1}^{n} x_i$ and $\sum_{i=1}^{n} y_i$ equal to 1 and a is a scaling factor equal to 0.5. This normalization has the advantage of keeping the statistical information regarding the shape and not the magnitude of the spectral curve. In AWN as the resulted score gets closer to 0, so does the similarity between the two vectors increase.

3.6.9 Spectral Similarity Scale (SSS)

The SSS [47] algorithm uses the ED and SCM algorithms to create the new similarity measure, such as:

$$SSS(\mathbf{x}, \mathbf{y}) = \sqrt{ED'(\mathbf{x}, \mathbf{y})^2 + (1 - SCM'(\mathbf{x}, \mathbf{y}))^2} \tag{3.20}$$

where ED' is the described in Equation 3.3 and SCM' is normalized to remove dependence of the number of bands:

$$SCM'(\mathbf{x}, \mathbf{y}) = \left(\frac{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{\mathbf{x}})(y_i - \overline{\mathbf{y}})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{\mathbf{x}})^2} \sqrt{\sum_{i=1}^{N} (y_i - \overline{\mathbf{y}})^2}} \right)^2 \tag{3.21}$$

SSS algorithm exploits the benefits of both algorithms to produce a new similarity measurement. ED algorithm measures the brightness difference between the two spectra and is insensitive to shape differences, while SCM compares the features and is insensitive to brightness. As a result, SSS takes into consideration both the brightness differences as well as the shape differences between the compared spectra. Finally, as the score gets closer to zero, the similarity of the two spectra increases.

3.6.10 NS3

NS3 [50] is another combinational measurement that uses the SAM (Equation 3.1) and the normalized ED' (Equation 3.3) algorithms to create a new similarity algorithm, which is described as:

$$NS3(\mathbf{x},\mathbf{y}) = \sqrt{A'(\mathbf{x},\mathbf{y})^2 + (1-\cos SAM(\mathbf{x},\mathbf{y}))^2} \tag{3.22}$$

where:

$$A'(\mathbf{x},\mathbf{y}) = \frac{ED'(\mathbf{x},\mathbf{y}) - ED'_{min}(\mathbf{x},\mathbf{y})}{ED'_{max}(\mathbf{x},\mathbf{y}) - ED'_{min}(\mathbf{x},\mathbf{y})} \tag{3.23}$$

is the normalization of ED' algorithm in the range of [0,1] so as to allow the comparison of the spectral vectors in a common baseline. From a geometrical point of view, NS3 is using the Pythagorean theorem to calculate the hypotenuse connecting the edge of the perpendicular measures A' and SAM. In this algorithm, the similarity becomes greater as the score gets closer to zero, too.

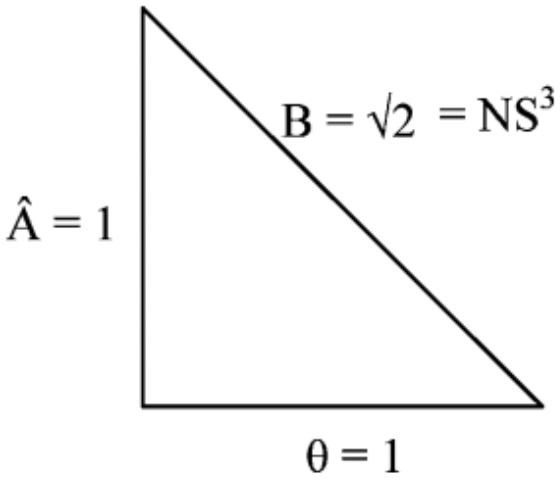


Figure 3.5: *Geometrical representation of the NS3 measure.*

3.6.11 Spatio-Spectral Decomposition (SSD)

This algorithm uses the Pearson' s correlation coefficient (Equation 3.9) to create a similarity measure:

$$SSD(\mathbf{x},\mathbf{y}) = \left(\frac{1-SCM(\mathbf{x},\mathbf{y})}{2}\right)^2 \tag{3.24}$$

since SCM score result in values in the range [-1,+1], SSD is actually a normalization of SCM correlations in the range [0,1]. SSD [48] algorithm takes into consideration the features of the spectral curves to produce a similarity score.

3.6.12 Spectral Pan-Similarity Measure (SPM)

SPM [48] combines ED, SSD and SID algorithms as follows:

$$SPM(\mathbf{x},\mathbf{y}) = SID(\mathbf{x},\mathbf{y}) \tan\left(\sqrt{ED'(\mathbf{x},\mathbf{y})^2 + SSD(\mathbf{x},\mathbf{y})^2}\right) \tag{3.25}$$

SPM algorithm is yet another effort to measure the similarity between two spectra using both the amplitude and the shape of the curves.

3.7 Accuracy Assesment of Spectral Similarity Measurements

A classification process is not complete until its accuracy is assessed. The purpose of accuracy assessment is to produce quantitative results about the classifier' s capacity to produce classes that describe the complexity of the scene exhaustively. When a spectral library search is performed for the identification of target spectra, the Spectral Similarity measurements are assessed based on their capacity to find the right match from within the spectral library. Given a spectral library, Δ and a set of spectral signatures $\{\mathbf{s}_i\}_{i=1}^{M} \in \Delta$, the Spectral Discriminatory Probability [50], and Spectral Discriminatory Entropy[50] are used to assess the accuracy of a Spectral Similarity algorithm.

3.7.1 Spectral Discriminatory Probability

Let $m(.,.)$ be any spectral similarity measure and t be a target spectrum to be identified using Δ. The Spectral Discriminatory Probability (SDP) is defined as:

$$SDP(\mathbf{t},\mathbf{s}_i) = \frac{m(\mathbf{t},\mathbf{s}_i)}{\sum_{j=1}^{J} m(\mathbf{t},s_j)} \tag{3.26}$$

SDP calculates the likelihood that the target spectrum will be identified as a member of the spectral library. A small SDP value indicates a difficulty for the measurement $m(.,.)$ to distinguish between the target spectrum and a library reference. In other words, a small SDP means that the measure is likely to find the best match in the library Δ.

3.7.2 Spectral Discriminatory Entropy

Using SDP, the Spectral Discriminatory Entropy (SDE) of a measurement $m(.,.)$ can be defined as:

$$SDE(\mathbf{t}, \mathbf{s}_i) = -\sum_{j=1}^{J} SDP(\mathbf{t}, s_j) \log SDP(\mathbf{t}, s_j) \tag{3.27}$$

SDE provides a measurement for the uncertainty of the identification of target t using Δ. As the entropy value gets lower, the possibility to find a match of t in the spectral library increases. Thus, t can be easily determined.

3.8 The Problem at Hand

Many studies have been conducted in an effort to evaluate Spectral Similarity metrics and find the most accurate one. These studies create a library of spectral signatures acquired from online sources or from cover types selected from a hyperspectral image as reference data for Supervised Classification. These libraries are generally, relatively small in their number of entries (usually about 400 signatures are used). In the first case, in which online acquired spectral signatures are used, a member of the library is selected and compared against all the other members in the library using different Spectral Similarity algorithms. In the second case, where reference data from a hyperspectral image are used, a Supervised Classification is performed using different Spectral Similarity algorithms. In this essence, the data used for evalution are already labelled, which cannot provide an intution on the metrics capability to identify unknown data or its robustness.

In this study, a spectral library is used to evaluate the twelve spectral similarity metrics mentioned above. The evaluation procedure is comprised from three steps. In the first step, spectral signatures are chosen from the spectral library and concidered as target spectra to identify. Then, the chosen signatures are compared to the library using the spectral similarity metrics one by one. As the spectrum treated as unknown is aready a member of the library this step is used to verify the correct implementation of the metrics, since every one of them must be able to find the match in the library. In the second step, reference data are collected from a hyperspectral image and stored in the library. Pixels are then selected from the cube and using the similarity metrics and the library the get identified. So in this step, the image is a mixture of labelled and unlabelled data, which are both used during the evaluation process. In the last step, the algorithms are tested on a hyperspectal cube containing unlabelled data, meaning that none of the spectral signatures contained on the cube are members of the library. The scene should be known to the analyst for this step to be effectively realized and give an accurate evaluation of the metrics.

Chapter 4

Implementation of a Spectral Library and Spectral Similarity Algorithms

4.1 Introduction

Spectral Libraries have been widely used by spectroscopists in their analysis of spectral data. Lately, there is an effort to incorporate spectral libraries for material identification on Hyperspectral Data. By exploiting the spectral signatures already stored in the library the analyst does not need to manually select reference data from the scene when performing Supervised Classification. This way, the Supervised Classification process becomes more automatic.

In this study, a Library containing spectral signatures from various materials was implemented on SQLite. Then, the algorithms described in Section 3.6 were used for the library search. Library search was performed in three steps. First, a signature was selected from the library and compared with the rest of the signatures in it. Since the selected signature was already a member of the library, the first best match for each algorithm was the signature itself. In the second step, reference data were collected from a hyperspectral image and stored in the library. Then, pixels from the same image were selected at random and compared against the library signatures. The discriminatory probabilities and entropies, as well as time performance, were measured during this step. Since some data from the image are already stored in the library they are labelled, while the rest are unlabelled. In the last step, a hyperspectral image was used for the final evaluation of the algorithms. This image was not labelled, which means that none of the spectral data it contains were part of the library. The areas of the image and its classes were already known to the analyst. This final step can give conclusive results on the Spectral Similarity metrics accuracy and performance. Moreover, it shows the possibility that

using a Spectral Library instead of manually selecting reference data in the Supervised method can have accurate classification results.

4.1.1 Data Collection

The first step of this study was to find publicly available online spectral libraries that contained the spectral signatures of interest. In this study, the reflectances of various materials were collected from the following spectral libraries:

1. United States Geological Survey (USGS) Spectral Library: USGS library contains spectral reflectances of thousands of materials measured from 200nm-2000nm. The spectral library was assembled to facilitate the identification and mapping of manmade materials, vegetation, and minerals.
2. Aster Spectral Library: ASTER spectral library was created by NASA and contains spectral signatures of manmade materials, soils, rocks, minerals and so much more.
3. ECOSTRESS Spectral Library: ECOSTRESS is an extension of the ASTER spectral library in which 1100 new reflectance measurements from vegetation were added.
4. EcoSIS: EcoSIS is a spectral library containing thousands of spectral signatures of vegetation.
5. National Institute of Standards and Technology: NIST offers a publicly available dataset of a 100 reference reflectance spectra of human skin tissue measured in 250nm-2500nm.

The total number of spectral signatures selected is 5000. Each of them belongs in one of the following three categories of interest:

1. Man-made Materials
2. Soils and Soil Mixtures
3. Vegetation
4. NIST Skin
5. Macbeth Colorchecker
6. Munsell Colors
7. Color Panels of Man-made Pigments
8. Nevi Samples

The aforementioned categories were very general and did not help in the accurate categorization of the collected spectra. Thus, subcategories were used to better organize the spectral signatures. The subcategories have been created on the basis of the relative information accompanying each sample from the source it was acquired. The final categorization creates the tree-like structure of Figure 4.1

4.1.2 Dataset Preparation

The spectra collected from the open libraries were not homogeneous in their form. Different instruments have been used for the measurements in each case, while the bands selected for each measurement differ. In addition, some spectral signatures have been stored in their raw form without being processed, and as a result, they were noisy. Finally, the reflectances in some cases were provided on a scale of 0-100, while in others were normalized on the 0-1 scale. These data inconsistencies were not preferable so it was necessary to process the spectral curves before they were stored in the database.

The processing of the spectra was done in MATLAB's environment. Initially, the spectral curves were displayed using MATLAB's plot() function to see if they contained noise. In cases that noise was found, a moving median smoothing filter was applied to eliminate the noise. Then, the bands of interest were selected. Different spectra have been measured in different bands with a different step. In this study, the reflectance data were used in the Ultraviolet (UV), Visible (VIS) and Near Infrared (NIR) range, and more specifically from 370 nm to 1000 nm, and the step selected was 10 nm. So, the acquired data had to be upsampled or downsampled to be represented in that form.

4.2 Relational Database Implementation

After preprocessing all the acquired spectral signatures and bringing them into the chosen form, a database system was chosen to store and handle all the data. There are many choices out there for data storage, but a Relational Database Management System (RDBMS) was chosen as the most appropriate for the dataset of this study. RDBMSs are a great choice for structured data and provide high availability and consistency. Furthermore, data search, comparisons, and analysis are faster and easier with RDBMSs. For this study, an RDBMS that can be stored both locally and also be used in a client-server implementation was wanted.

SQLite is an RDBMS that is great for device-local storage, but it can also handle medium traffic in a client-server application. SQLite stores the database at a single disk file and each process that wants to access the database reads and writes directly on the disk file. This method

improves performance and readability as it can be faster than a normal filesystem, with reduced cost and complexity. On the downside, most systems have an upper limit on the size of files in the disk, thus the database is limited to about 140 TB in size. In this study, the sqlite3 version of SQLite was used. The following subsection describes the schema used to store the acquired spectral signatures.

4.2.1 The Database Schema

As it can be seen in Figure 4.2 the database schema consists of three tables namely "Category", "Sample" and "Metadata". The "Category" table was used to store information about the various categories and subcategories in which the spectral data belong to. An integer number called, "categoryID" was used to assign a unique number to each category or subcategory.

Since each category may have multiple subcategories, an Adjacency List Model was used to store the hierarchy as that is depicted in Figure 4.1. Adjacency lists are computationally a cheap model when it comes to inserting or deleting nodes in the hierarchy, but finding a node's level or ancestry is quite expensive as it requires to perform multiple JOINs. For these reasons, Adjacency Lists are preferable when the hierarchical tree does not grow very deep. In this study case, the height of the hierarchy does not go above the fourth level, so the Adjacency List Model seemed the right choice.

The "parent_categoryID" attribute was used to show in which category each subcategory belongs to. In the case of root categories, this attribute is assigned to Null. Additionally, in order to make things easier in regards to finding the level and the ancestry of a node, two more attributes were added namely "level" and "lineage". The later holds the path from the root to that node for each record.

As soon as the hierarchical structure of categories was established, a separate table, called "Sample", was created to hold the individual spectral signatures. Each spectral signature is uniquely identified by an integer called "sampleID". The "categoryID" attribute is a reference to the "Category" table and was used to enable the assignment of each spectral curve to a category or subcategory. The relationship between the "Category" table and the "Sample table is that of one to many (depicted as 1:N in the schema), which means that each category may have multiple samples assigned to it, but each "Sample" record can belong to only one category. The "name" attribute holds the name that characterizes each sample for a more user-friendly discrimination between the spectral signatures. Then, the "wavelength" and "reflectance" attributes are used to hold the actual spectral data for each spectral signature. The vectors of wavelength and reflectance measurements are stored as comma separated strings for each spectral signature.

Finally, the "Metadata" table is used to store some additional information concerning each sample. Because this table was created in order to further describe the original spectral data and does not hold information that can be uniquely identified and stand on their one, it is called a weak entity. The attribute "sample_ID" is used to refer to the "sampleID" attribute of the "Sample" table. The "measurement" attribute stores information about the equipment used to perform the reflectance measurement, while the "description" holds a few information about the sample used or the conditions during which the measurement was conducted. Lastly, the "source" attribute holds the name of the online Spectral Library from which the specified reflectance was acquired.

4.2.2 Database Constraints

A database system is useful as long as the information stored in it is accurate. Constraints are rules forced onto a database schema so as to exclude invalid records from being stored in the database. This way data integrity is guaranteed throughout the schema.

The attributes "categoryID", "sampleID" and "sample_ID" are primary keys of the tables "Category", "Sample" and "Metadata" respectively. A primary key uniquely identifies the records in all possible valid cases of relationship instances. On the other hand, foreign keys are used to establish relationships between tables on the schema. A foreign key of one table is connected with the primary key on another table in order to form the relationship. To ensure data integrity, foreign keys may only refer to existing primary keys on the reference table. In this study's schema the attribute "parent_categoryID" is a foreign key referring to the primary key "categoryID" on the same table. The attribute "categoryID" of the "Sample" is also a foreign key to the "Category" table. Finally, "sample_ID" attribute of table "Metadata", apart from being the primary key of that table, is also a foreign key pointing to the "sampleID" of the "Sample" table. Referential integrity is ensured throughout the schema by cascading to the foreign keys any updates or deletes that may happen on the primary key with which they are associated in each case.

Apart from the key constraints, a NOT NULL constraint was used on the "name" columns of tables "Category" and "Sample". The NOT NULL constraint was also forced on attributes "wavelength" and "reflectance". NOT NULL constraints ensure that this attributes will always have values upon a new insert on the table. Furthermore, a CHECK constraint on both "wavelength" and "reflectance" columns verify that the inserted string for each record contains only numerical characters and commas using regular expressions. A UNIQUE constraint on "reflectance" ensures that there will be no duplicate reflectance measurement in the "Sample" table.

4.3 Implementation of Spectral Similarity Algorithms

Since the database schema has been created and the database has been populated with the acquired spectral signatures from the online Spectral Libraries, the next step was to test various Spectral Similarity Algorithms, so as to verify how the use of a Spectral Database can assist in semi-automatic classification. This section describes how the algorithms presented in Section 3.6 were implemented and the early tests that were performed on them.

4.3.1 Database Interface

A database interface is software program build to connect and interact with the database. It provides the user with all the necessary tools to interact with the database without directly using the SQL language. For the purposes of this study, a simple interface was created using MATLAB and the ODBC library. The interface first establishes a connection with the database and then retrieves all the stored reflectances and their respective names from it using the SQL SELECT statement. Subsequently, the retrieved spectral signatures are converted from the comma separated strings to numerical vectors.

Since a lot of the algorithms defined in Section 3.6 make use of other similarity measures in order to produce a similarity score, each one of them was implemented as a separate MATLAB function. To validate that the measurements were implemented correctly, a spectral signature was chosen from the library and it was compared against all the spectral signatures contained in it. If the algorithm was correct, then it would match the selected spectral signature with a score equal to zero. Apart from the validation of the algorithms' correctness, this stage was used to find proper thresholds for each one.

Comparing a spectral signature from the database with the spectral library helped to see the trends that the similarity algorithms have, but was not enough to assess their performance. In order to validate the spectral similarity measurements for their performance in hyperspectral data labelling, the interface described in this section was incorporated in a GUI built for Hyperspectral Data Analysis.

4.3.2 Spectral Suite

The Interface and the algorithmic functions were incorporated in a Graphical User Interface (GUI), called Spectral Suite, which was created in Electronics Laboratory of Technical University of Crete (Figure 4.3). This Suite was designed to assist in Hyperspectral Imaging Analysis after the acquisition of a spectral cube. It consists of two windows namely "Cube Viewer" and "Spectrum Viewer".

The "Cube Viewer" window is the main window where all functionalities lie. It consists of a menu bar, a toolbar, and two axes. The user can upload a hyperspectral cube as a stack of images or as Matlab file (.mat) using the "Import" tab on the menu bar. During the import process, the user is prompted to choose the spatial resolution of the cube among a list of options and the case-specific category from the database. The resolutions provided are:

1. The original spatial dimensions of the cube.
2. 1920×1080
3. 1280×720
4. 800×600
5. 640×480
6. 320×200
7. 150×100

As soon as the import of the cube is completed, it is displayed on the right axis. The user can view the images in the stack one by one using the slider below the axis. As the slider moves through the images, a text label displays the wavelength in which the current image was acquired.

The Spectral Suite software also provides three preprocessing filters that can be applied on the cube. The filters can be selected from the "Process Image" tab on the menu bar. The offered filters are Median, Gaussian, and Wiener. The selected filter is applied to every image in the cube and the user can either save the changes or reset the cube to its original form. Furthermore, the user can reduce the spectral resolution through the "Dimension Handling" menu tab. When this tab is chosen, a list of the wavelengths in which the imported cube was acquired is shown. The user can then choose manually the wavelengths which he/she would like to keep for subsequent analysis.

Apart from the preprocessing steps, the Suite offers Spectral Classification methods that can be performed on the cube. This is the core of the Suite. Spectral Classification is performed through the "Clustering" tab on the menu bar. The "Clustering" tab enlists a series of Unsupervised Classification Algorithms namely: K-means, K-medoids, FCM, DBSCAN, Spectral-Clustering, Gaussian Mixture Model and Fast K-means. The latter method is an improvement of the classical K-means algorithm that was implemented on the Electronics Laboratory of the Technical University of Crete. The difference between the classical K-means algorithm and the Fast K-means lies in the fact that the latter does not require the number of classes to be inserted from the user and is able to find the optimum number of clusters in the hyperspectral dataset.

Dimensionality reduction is performed automatically before each Unsupervised algorithm using the Principal Component Analysis (PCA) method. After evaluation, it was proved that the first five principal components contain most of the information and they could be used in unsupervised classification without great loss. Each clustering algorithm produces a thematic map of clusters as a result. A random color is assigned in each class and the final thematic map is displayed on the left axis on the "Cube Viewer" window.

Below the menu bar, a toolbar gives access to "Zoom In" and "Zoom Out" operations on the currently displayed image of the cube or the pseudocolor map. Additionally, the toolbar provides a "Data Cursor", which offers the ability of pixel selection. By clicking on it, a cursor appears on the screen that allows the user to select a pixel from either of the two axes. When a pixel is selected the respective reflectance intensity is extracted from it. The intensity is, then, transformed in the range [0,1] to match the stored spectral signatures in the spectral library and it is displayed on the "Spectrum Viewer" window. Figure 4.4 shows an example of a pixel selection from the cube and its respective reflectance being displayed on the axis of "Spectrum Viewer". The x-axis on "Spectrum Viewer" takes the values of wavelength in which the imported cube was acquired (wavelength unit used is nanometers), while the y-axis represents the reflectance intensity values of a chosen pixel in the range from 0 to 1.

The analyst has the option to save the chosen pixel reflectances in a text file through the "Store Spectrum" button. When this button is clicked a window opens that allows the user to place a name for the .txt file and choose the directory in which it will be saved. The user can save multiple pixels in the defined text file. Each time a pixel reflectance is stored, its curve is also stored on the "Spectrum Viewer" and a marker on the image indicates the place of the selected pixel. The first utility offers the ability to compare at once the chosen reflectances, while the latter prevents the analyst from choosing the same pixel twice. The stored reflectances can be deleted by clicking on the "Clear Spectra" button.

4.3.2.1 The Labelling Menu Tab

The "Labelling" menu tab contains the twelve Spectral Similarity Algorithms implemented for the purposes of this thesis. The analyst can select any of the algorithms on the list in order to label the previously stored pixel reflectances. When a Similarity algorithm is chosen, the entries on the .txt file are read one by one and compared against the entries of the Spectral Library. The algorithms provide the best match found in the Library for each stored spectrum and plot them both on the same axes.

4.3.3 Cluster-then-Label Implementation

Apart from their use in the "Spectral Suite" for pixel labeling, the implemented Spectral Similarity metrics were assessed for their ability to label cluster centroids[1]. An Unsupervised Clustering is first performed on a hyperspectral cube. The resulted centroids are then compared against each entry on the database using each of the twelve algorithms. Subsequently, the spectral similarity measurements are assessed for their ability to produce a right match for each of the unlabelled centroids.

In Chapter 4 the final results of the Spectral Similarity algorithms are presented for both pixel labelling and centroid labelling. The algorithms are tested on hyperspectral images taken from skin lesions in-vitro.

[1] It is important to note that this utility is not part of the "Spectral Suite" at the moment.

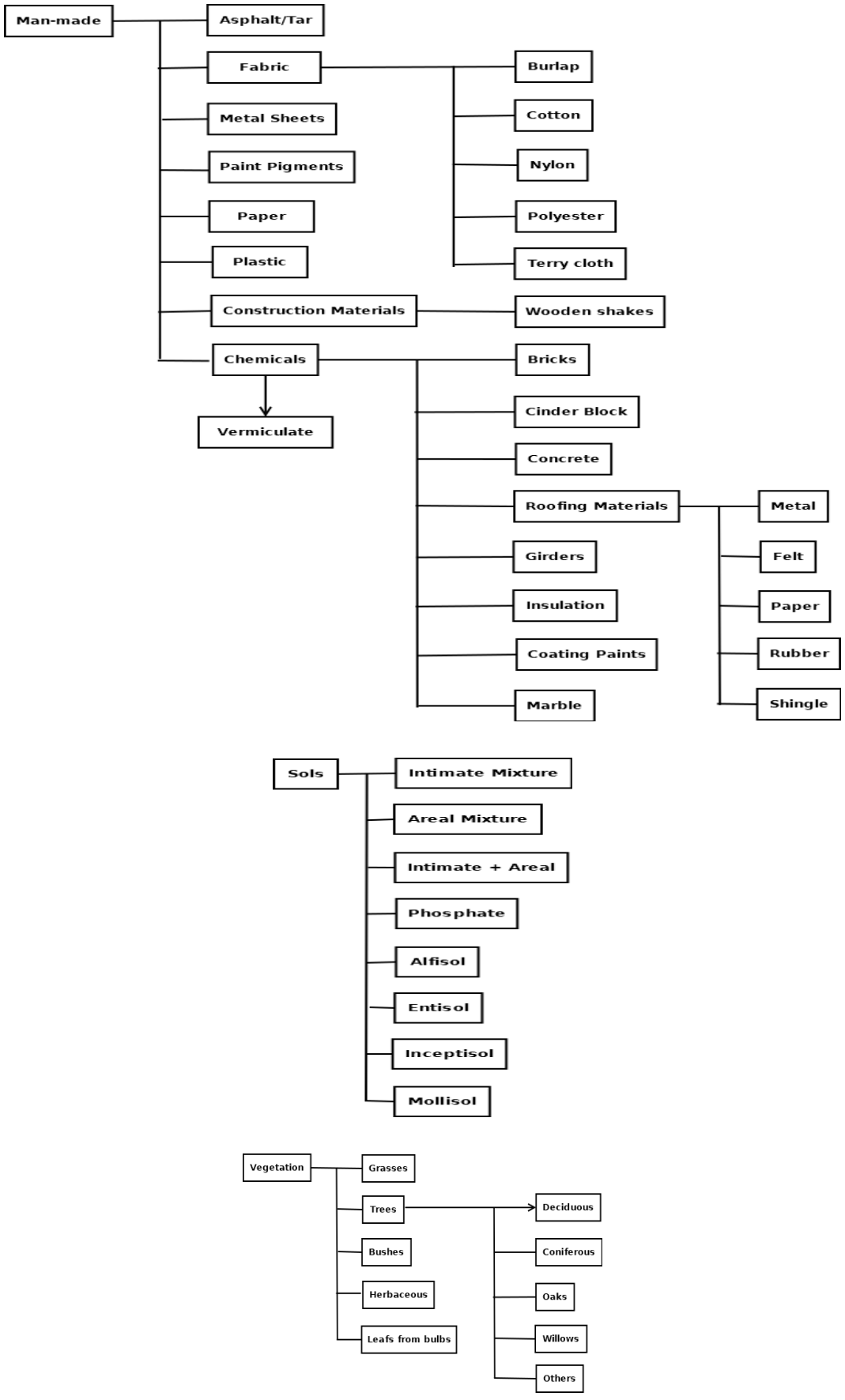


Figure 4.1: *Categories and subcategories of the collected spectral signatures. Macbeth Colorchecker, Color Panels, Munsell and NIST Skin categories have no subcategories so they are not included in this depiction.*

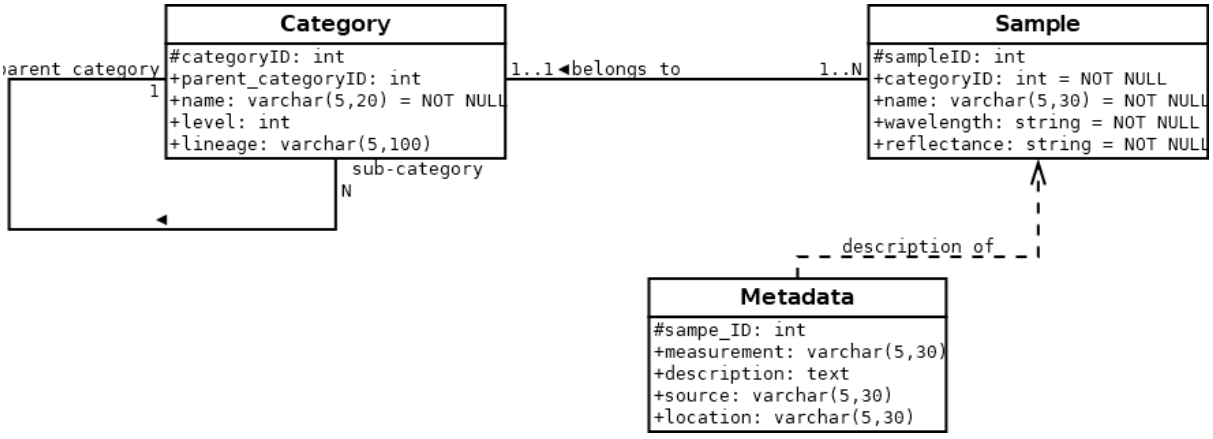
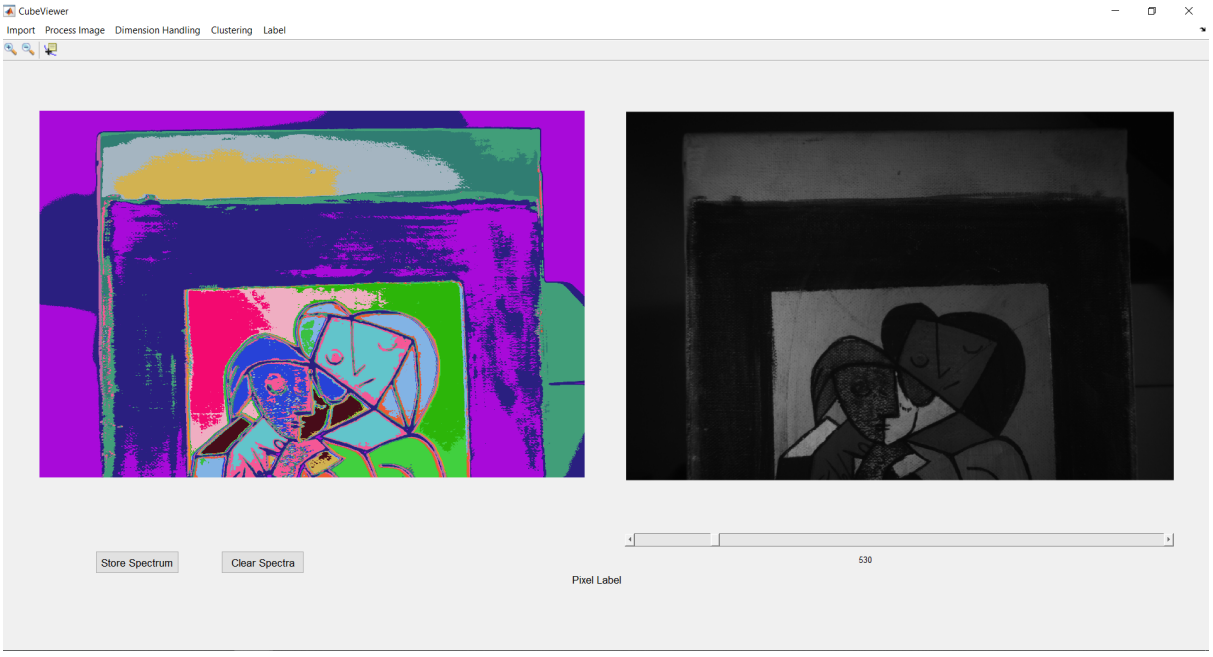
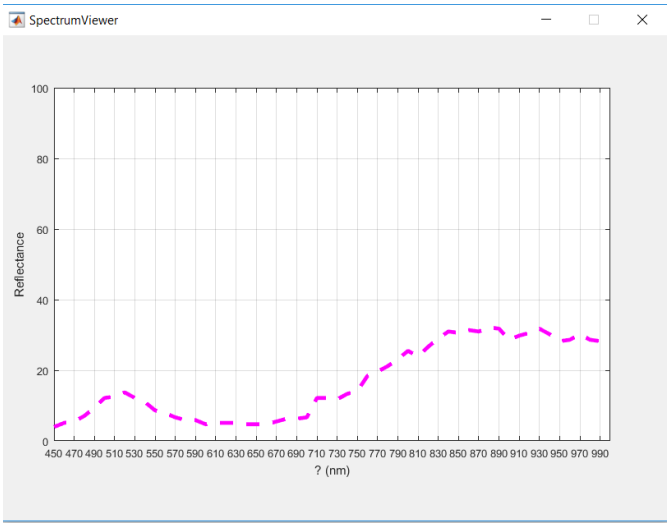


Figure 4.2: *The schema followed for the RDBMS implementation.*



(a) *The "Cube Viewer" window*



(b) *The "Spectrum Viewer" window*

Figure 4.3: *Outlook of the Spectral Suite*

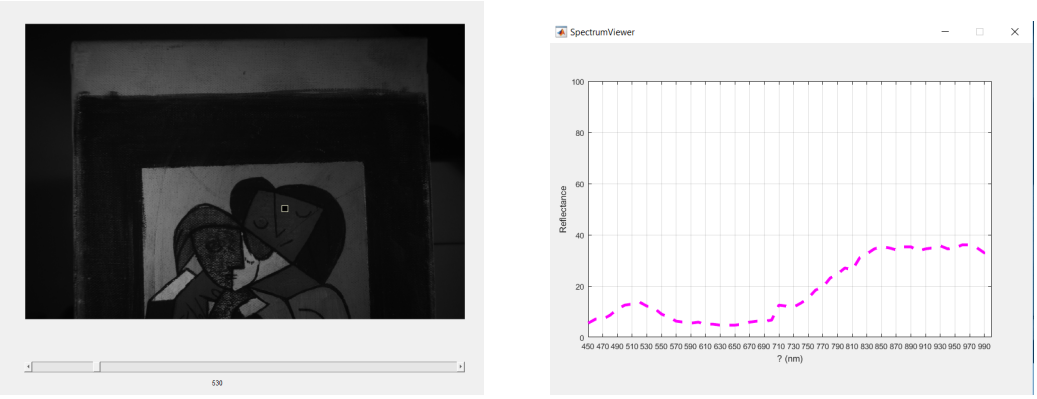
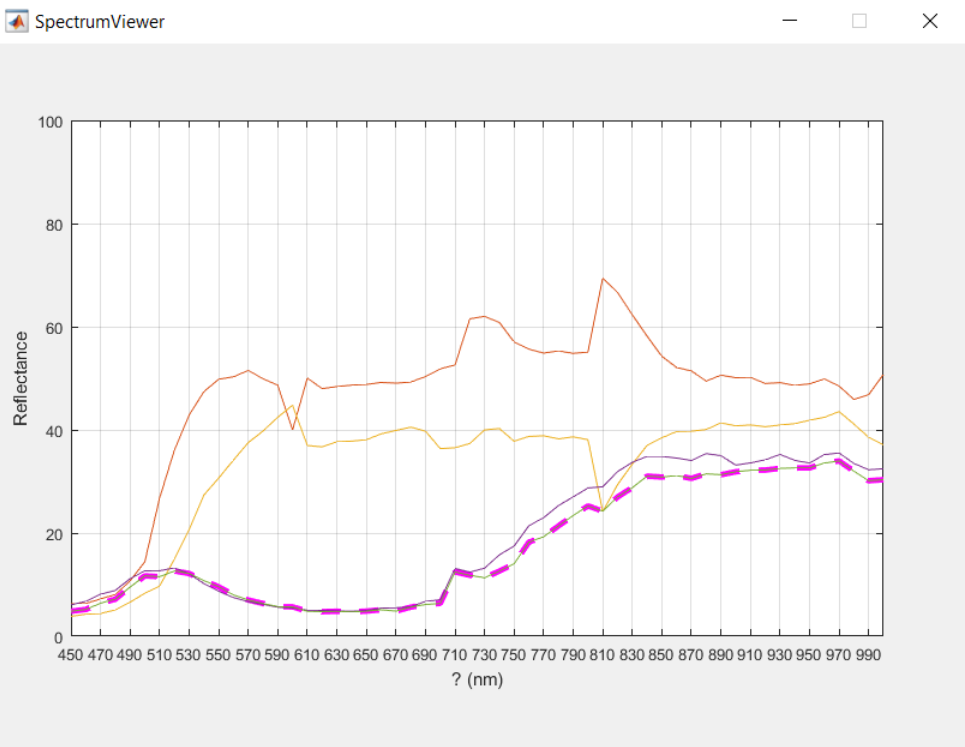


Figure 4.4: *An example of pixel selection from the hyperspectral cube and its respective reflectance intensity plotted on "Spectrum Viewer"*



(a) *Spectral curves from multiple stored spectra*

(b) *Markers are placed on the selected pixels*

Figure 4.5: *Each time a pixel is stored "Spectrum Viewer" holds its respective spectral curve and markers apear on the thematic map."*

Chapter 5

Methods and Results (Case Study: Identification of Skin Lesions)

5.1 Introduction

Skin cancer is the most common form of cancer, with about a million new cases in the U.S. each year [72]. Skin cancer is mainly divided into three categories:

1. Basal-cell skin cancer (BCC)
2. Squamous-cell skin cancer (SCC)
3. Melanoma

the first two are benign forms, while melanoma is a malignant form of cancer. Often, skin cancer is difficult to diagnose non-invasively, as malignant skin lesions can closely resemble the benign ones. Different lesion types can have similar characteristics, making it difficult to discriminate among them. In this Chapter, a dataset of skin nevi samples was assembled in order to test if identification by spectral library searching can help in skin cancer diagnosis.

5.2 Spectral Imaging on Nevi Classification

Early detection and treatment of skin cancer can significantly improve patient outcomes. In clinical practice, visual examination determines whether a skin lesion is cancerous based on the ABCDE rule (asymmetry, border, color, diameter, and evolution) and the change in the

appearance of a mole or pigmented area over a period of time. However, clinical diagnostic sensitivity and specificity vary greatly, depending on the expertise and visual skills of the clinician. Consequently, histopathologic examination of the excised suspicious element still remains the gold standard. However, a biopsy is an invasive procedure and leaves a scar at the biopsy site, which otherwise would be unnecessary in the case of benign lesions.

The absorption of light can provide information on the biochemical composition of the skin. The light scattering properties of skin can provide information regarding its micro-architecture. Recently, great attention has been given to skin cancer diagnosis with non-invasive spectroscopic and Spectral Imaging methods [5-27]. However, these studies are based on heuristic techniques to classify a nevus as one type or another. More specifically, instead of using all the features of the spectral curves, they only choose two wavelengths where the peak of melanin and hemoglobin absorptions occur. Since each lesion has different spectral features these methods are not optimal and leave out information that could be of great importance in the lesions accurate type classification.

In this study, a Hyperspectral Imaging system has been used to acquire reflectance data from multiple types of pigmented skin lesions, in the wavelength range from 420-1000nm. The resulted cubes are imported on the "Spectral Suite" Software and reference data are selected to be stored in the database. Then pixels are selected from the cubes and compared to the library of spectral signatures using the whole spectral curve. The pixels are compared first with the case-specific spectra acquired as references from the lesions cubes and then with all the spectral signatures there are in the library. These two steps comprise the last steps of the Spectral Similarity measures evaluation.

5.2.1 Acquasition System

The MuSES-9 HS Hyperspectral Imaging System was used in this experiment to acquire spectral cubes from nevi samples in-vitro. The system is capable of acquiring spectral images in a range of 370-1000 nm (UV-NIR) with a spectral resolution of 5nm to 15nm using a 1/1.8" square pixel CMOS sensor. The camera can acquire 30 frames/sec with spatial resolution up to 3096x2080pixels in 32s -15s integration time. Thus, up to 150 discrete spectral bands can be acquired in real time.

A specially developed software is employed for the control of the camera and the CMOS sensor as well as for the spectral image analysis. The system operates in two modes: the spectral imaging and spectroscopy in both reflectance and fluorescence. 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 is executed before these imaging

procedures. A white surface with unity reflectance across the 370-1000 nm spectral range is used for calibration. The shutter and gain values are automatically generated as the sensor moves through the specified spectral range. These settings determine the sensitivity level of the camera at each wavelength to ensure the efficiency of the system. During the acquisition of a sample, the sensitivity of the camera is automatically regulated according to the stored calibration. The resulted cube is stored and 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 less than 1 min scanning time.

For this study, hyperspectral cubes of skin lesions were acquired in the range from 420nm to 1000nm with a 20nm step.

5.3 Data Selection

Table 5.1 shows the number of samples used for this study per skin lesion type. Each of the samples was surgically removed and sent for histopathological analysis. After the biopsy was conducted a hyperspectral cube was acquired in-vitro from each sample using the aforementioned system (Subsection 5.2.1). Figure 5.1 shows examples of the four nevi lesions in selected wavebands. As it can be observed healthy nevi regions get transparent as the measurement proceeds into the NIR. On the other hand, some patches or spots on the nevi remain visible even after 800nm. Especially, in the case of the melanoma, the cancerous region of the nevus remains dark until 1000nm. This is due to the high melanin concentration in this regions on the nevi. Melanin is the main skin absorber in the infrared region of the spectrum. As the malignancy of a nevus increases, so does the melanin concentration in the skin. So, dark spots that remain after 750nm can be indicators of a problematic nevus.

Table 5.1: *Number of nevi samples for each lesion type.*

Dysplastic	Compound	Junctional	Melanoma
7	8	1	1

In this study, reference spectra were selected from samples that were most representative of the lesions. The biopsy results were used as a map during this stage of reference data selection. In the case of the junctional and melanoma nevi, only one sample is available in each type, so reference data were collected from them. In the case of dysplastic and compound skin lesions, however, where multiple samples were available only one sample was used for the acquirement of references. Figure 5.1 shows the compound and dysplastic samples used as a reference in

selective bands[1]. After the selection of reference data was completed, a new category was created in the database to store the newly selected spectral signatures.

Figure 5.2 shows indicative spectral signatures collected from a nevus. As it can be seen, the spectral signatures of melanoma and those of normal skin have distinctive features and can be separated from each other. On the other hand, it is very difficult to discriminate between the compound and dysplastic spectra as their signature are overlapping. Furthermore, the junctional and the melanoma signatures overlap at a certain degree, however, their features are not similar. This area between melanoma signatures and normal skin signatures is of tremendous importance for early skin cancer detection.

5.4 Pixel Identification Results

Since the dataset of reference was inserted into the database, a series of experiments were conducted to assess the accuracy of the spectral similarity algorithms. The hyperspectral cubes acquired from each sample were imported in the "Spectral Suite" software. Various pixels were selected from each sample and their respective reflectance intensities were compared to the database entries using each similarity measure separately. In this section, two examples of a pixel selection and its comparison to the database are presented. At first, the melanoma sample is used for spectral similarity evaluation. Instances of the melanocytic nevus are shown on Figure 5.1 d). A pixel was selected from the melanoma region that is visible beyond 760nm. The extracted pixel was preprocessed to remove noise using a moving median smoothing filter and then normalized in the [0-1] region.

In the second example, a dysplastic nevus is used for evaluation. The nevus used in the second example is depicted in Figure 5.3. Figure 5.4 depicts the image of the nevus at 700nm and a pixel selected from the cube represented with a square on the image. The extracted pixel reflectance was also normalized in the region from 0 to 1 by dividing the reflectance values by the maximum intensity of 255. A moving median smoothing filter was used for noise reduction.

5.4.1 When the a specified category is searched

At first, the reflectances of selected pixels were compared with only the relevant reference data collected from the nevi samples. Figure 5.5 shows the first match that each algorithm produced for the melanoma sample. Table 5.2 shows the resulted entropies and time performances for each algorithmic implementation. The resulted signatures of the dysplastic sample and the respective entropies and time performances are shown on Figure 5.6 and Table 5.3 respectively.

[1]Pay attention to the transitioning of melanin reflectance in the bands 760nm and 880nm.

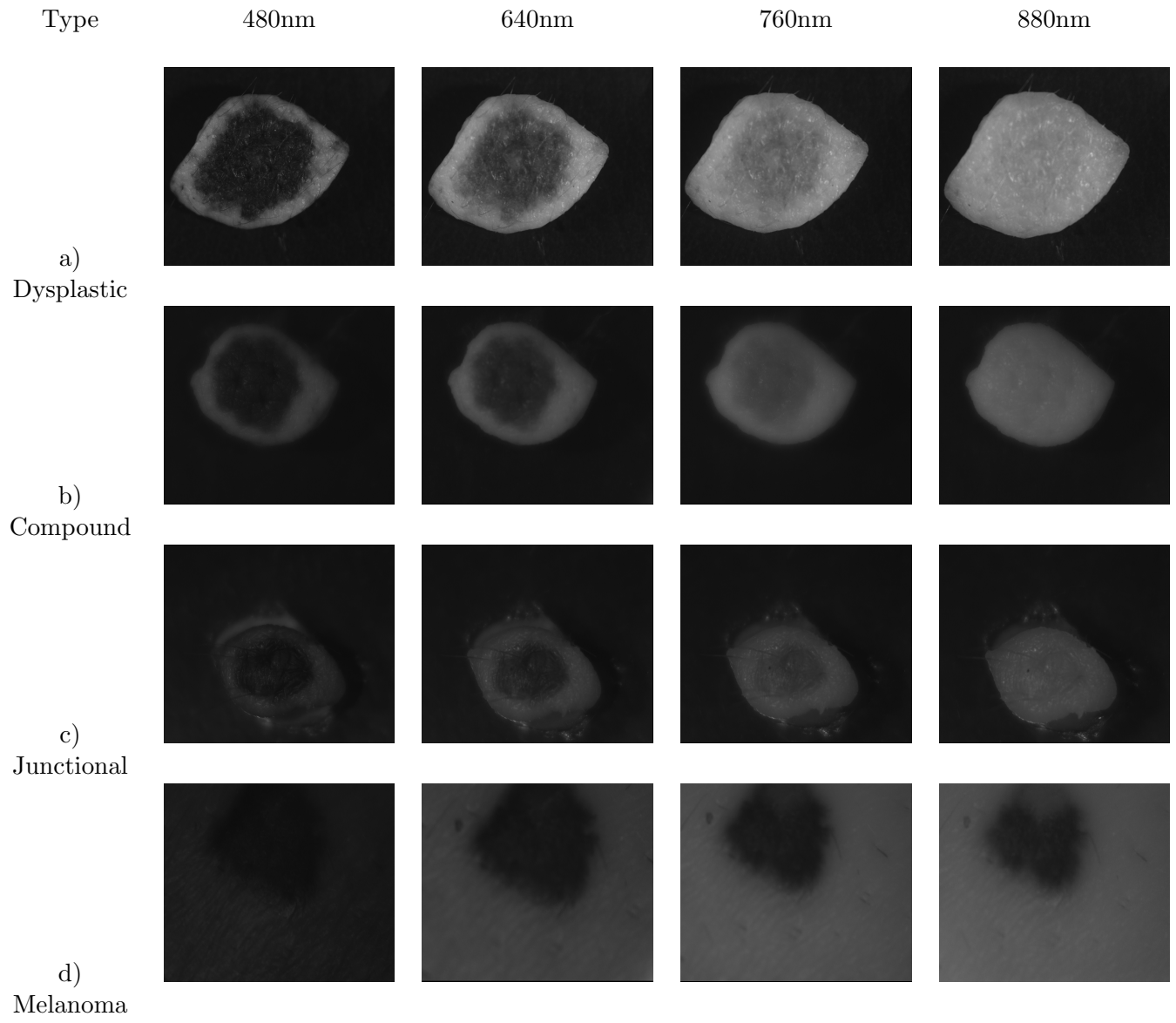


Figure 5.1: *The nevi used to comprise the reference dataset for the four different lesions types. The lesions are depicted in four different wavebands (480nm, 640nm, 760nm, and 880nm) to show the differences in melanin absorption. On the first row is the dysplastic nevus, on the second the compound nevus. These two samples were selected because of the homogeneity they exhibit in regards to melanin absorption. On the third row is the junctional nevus and finally on the last row the melanoma.*

As it can be seen by the retrieved spectral signatures of Figures 5.5 and 5.6 AWN, SAM, SCA, SGA, and SSD matched the unknown spectrum according to its spectral feature and did not take into consideration the intensity level of reflectance, which is in compliance with the characteristics of these algorithms. From the remaining algorithms, SID-SAM and SID-SCA results are very close to the unknown spectrum. The respective entropies of SID-SAM and SID-SCA have the lowest value according to Tables 5.2, 5.3 and their entropies are numerically close compared to the rest algorithms. On the other hand, their time performance falls somewhere in the middle. Finally, on Table 5.4 similarity entropies are ordered according to their respective

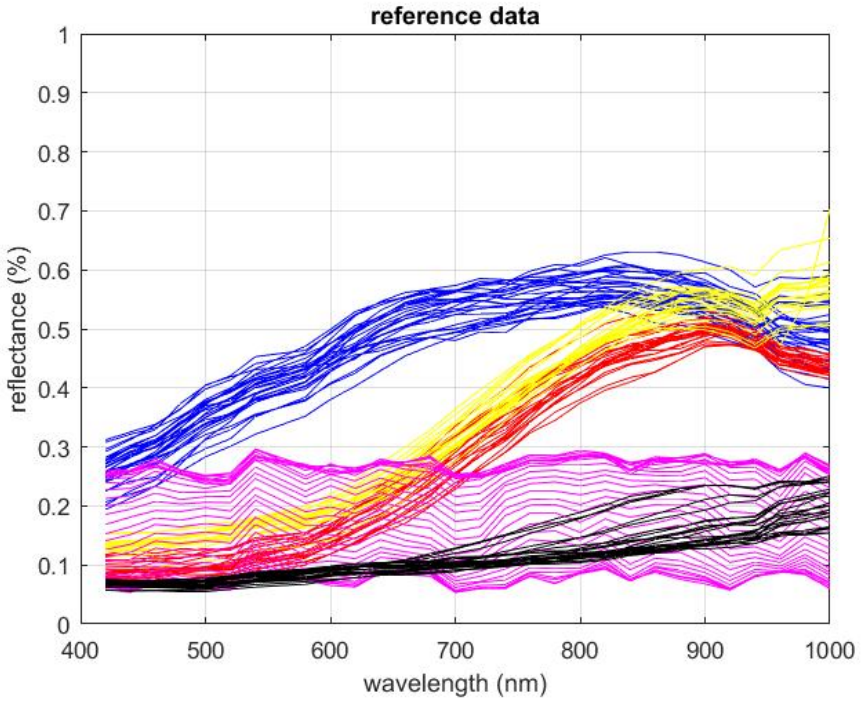


Figure 5.2: *Spectral signatures of the acquired reference data. In blue are normal skin spectral signatures, in yellow the compound spectral signatures, in red the dysplastic spectral signatures, in magenta the junctional spectral signatures and in black are the spectral signatures of melanoma.*

Table 5.2: *Melanoma Sample: Results on entropy and time performance of each Spectral Similarity Measurement, when a pixel is compared with nevi reference data in the Spectral Library*

Similarity measurement	Entropy	Time (sec)
AWN	7.6214	0.0378
ED	0.19491	0.030
NS3	0.1969	0.0341
SAM	8.1915	0.0296
SCA	8.2159	0.0333
SGA	8.323	0.0435
SID	0.09302	0.028
SID-SAM	0.0038	0.031
SID-SCA	0.022	0.0317
SPM	0.91447	0.0369
SSD	6.5468	0.0358
SSS	0.299	0.0514

entropies and time performances from lowest to highest.

Figure 5.3: *The dysplastic nevus used as an example in this chapter*

Table 5.3: *Dysplastic Sample: Results on entropy and time performance of each Spectral Similarity Measurement, when a pixel is compared with nevi reference data in the Spectral Library*

Similarity measurement	Entropy	Time (sec)
AWN	7.426	0.026
ED	0.18461	0.0224
NS3	0.1865	0.0264
SAM	0.1865	0.047
SCA	8.0594	0.027
SGA	8.2052	0.0418
SID	0.062	0.0242
SID-SAM	0.0015	0.0284
SID-SCA	0.00937	0.029
SPM	2.0275	0.0312
SSD	5.4898	0.0302
SSS	0.286	0.0302

5.4.2 When the whole database is searched

The same pixel reflectances were used for comparison, but this time they were tested against all the entries in the database. This was done in order to check the similarity measurements robustness and their ability to correctly identify a material from a dataset of diverse spectral signatures. Figures 5.7, 5.8 depict the best match that each algorithm retrieved from the

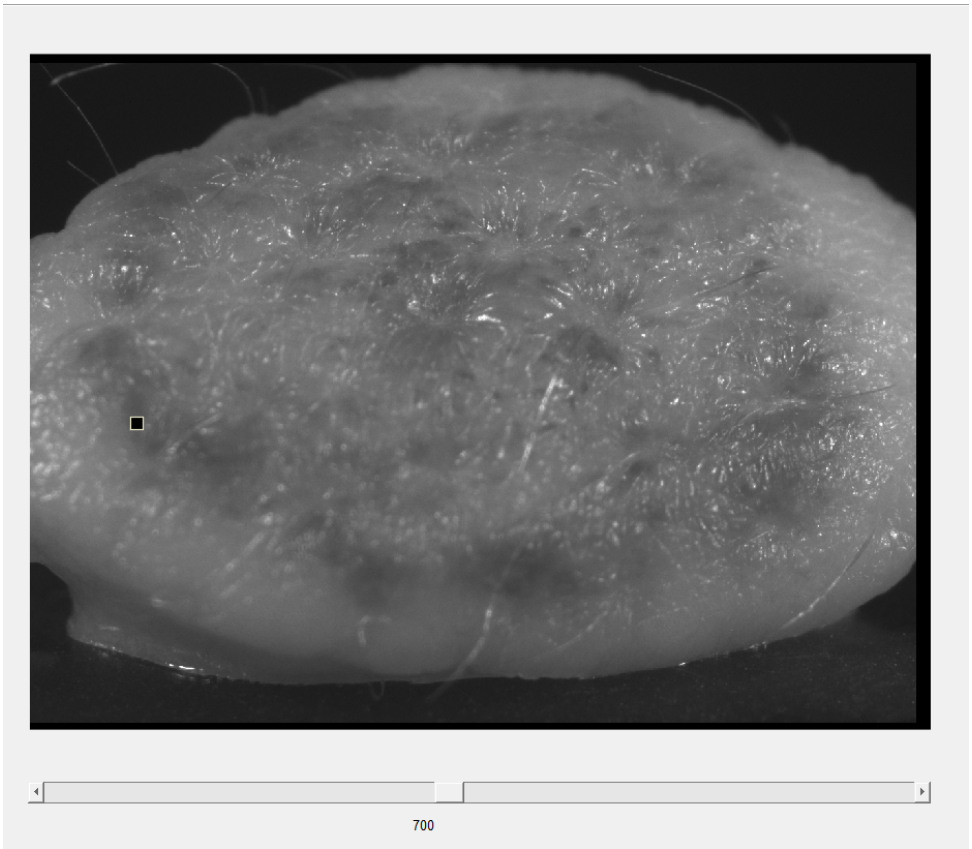


Figure 5.4: *A pixel (X=167, Y=540) is selected from the nevi cube to be compared with the database entries. The selected pixel is represented with a square on the image. The image shown is at 700nm and the pixel was selected from a dark spot on the nevus.*

Table 5.4: *Algorithms order from lowest to highest in respect to their mean of time performance and resulted entropies, when a pixel spectrum is compared with a specific category in the database.*

Time	Entropy
SID	SID-SAM
ED	SID-SCA
SID-SAM	SID
SCA	ED
NS3	NS3
SID-SCA	SSS
AWN	SPM
SSD	SAM
SPM	SSD
SAM	AWN
SSS	SCA
SGA	SGA

database, in the case of melanoma and dysplastic nevus respectively, while Tables 5.5,5.6 present the relative entropies and time performances for each algorithm.

Table 5.5: *Melanoma Sample: Results on entropy and time performance of each Spectral Similarity Measurement, when a pixel is compared with all entries in the Spectral Library*

Similarity measurement	Entropy	Time (sec)
AWN	11.932	0.3899
ED	3.0965	0.3289
NS3	3.7681	0.4081
SAM	12.037	0.4337
SCA	12.013	0.4238
SGA	12.206	0.719694
SID	1.8149	0.3662
SID-SAM	0.1329	0.4154
SID-SCA	1.6249	0.4375
SPM	7.4487	0.4761
SSD	10.143	0.4094
SSS	3.952	0.4114

Table 5.6: *Dysplastic Sample: Results on entropy and time performance of each Spectral Similarity Measurement, when a pixel is compared with all entries in the Spectral Library*

Similarity measurement	Entropy	Time (sec)
AWN	12.041	0.366
ED	2.6745	0.378
NS3	3.187	0.422
SAM	12.113	0.3732
SCA	11.914	0.399
SGA	12.211	0.7317
SID	1.2358	0.3496
SID-SAM	0.665	0.4209
SID-SCA	0.78076	0.4277
SPM	7.6017	0.4519
SSD	9.9456	0.4108
SSS	3.6112	0.4114

Figures 5.7,5.8 show clearly that AWN, SAM, SCA, SGA, SID-SAM, SID-SCA, and SSD are not affected by changes in the intensity levels and can produce a correct match even if the reference spectra show great diversity. Yet again, SID-SAM and SID-SCA manage to follow both the spectral features and the reflectance intensity of the unknown spectral curve. Finally, on Table 5.7 similarity entropies are ordered according to their respective entropies and time performances from lowest to highest. Although algorithms such as AWN and SAM produced correct matches for the unknown spectral curve, their respective entropies are very high compared to the rest. Entropy can only evaluate the ability of a method to discriminate among a number of spectral signatures, based on the numerical value distribution the similarity score produces. The entropy cannot be used to evaluate the actual spectral identification accuracy of

a method, but only provides inside on the distribution of the measurement' s results.

Table 5.7: *Algorithms order from lowest to highest in respect to their mean of time performance and resulted entropies, when an pixel spectrum is compared against all entries in the database*

Time	Entropy
ED	SID-SAM
SID	SID-SCA
AWN	SID
SAM	ED
SSD	NS3
SCA	SSS
SSS	SPM
NS3	SSD
SID-SAM	SCA
SID-SCA	AWN
SPM	SAM
SGA	SGA

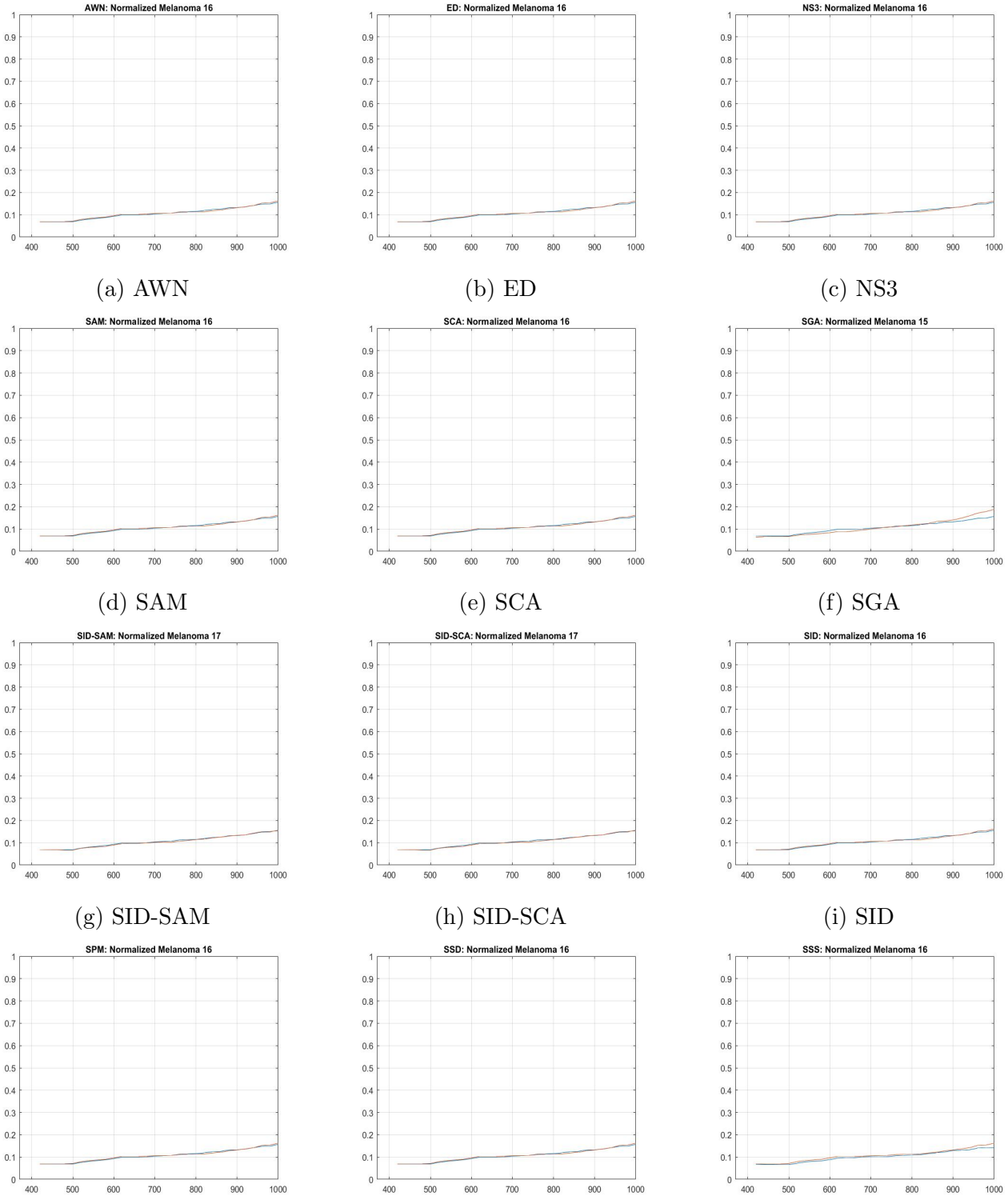
5.5 Cluster Centroids Identification Results

Another purpose of this study was to assess the Spectral Similarity measurements capacity to correctly label cluster means or centroids. Two examples are presented using the sample characterized as melanoma by the histopathological examination, and another one characterized as a compound. Each sample' s spectral cube was imported in the "Spectral Suite" Software and clustering was performed on them using the provided Unsupervised algorithms on the original cube spatial resolution. Here, the clustering results of the K-means Fast method will be used as an example to present the respective cluster labeling results. The colors used on the thematic maps are assigned randomly, but a matrix is used to hold each cluster' s RGB value. Figures 5.9, 5.10 show the image at 680nm and the thematic map produced by clustering for the compound and the melanoma samples respectively.

Once the clustering was completed, the resulted centroids were labelled using each one of the twelve similarity measures. Each centroid was preprocessed before comparison and normalized in the range [0-1]. Figure 5.11 shows the resulted labels for the melanoma case when the comparison is performed only on the "Reference Data of Nevi Data" category.

As it can be seen from the resulted spectral signatures, the melanoma class (pink color) is identified by all algorithms. Subsequently, the remaining nevus region (green) is labeled as junctional from the majority of methods, except SAM, SCA, and SGA that identified it like normal skin. However, the centroids spectral features, are more similar to those of dysplastic or compound lesions. Finally, the purple area of normal skin tissue is correctly labeled as such.

Figure 5.12 presents the resulted labels for the compound case when the comparison is performed on the case-specific category on the database. As it can be seen the lesion area is given different labels for each similarity measurement. However, AWN, SAM, SCA, SGA, and SID-SAM labels manage to better follow the features of the centroid curve. The normal skin region is correctly labelled from all similarity measurements. Finally, the black background is labeled as melanoma by all algorithms. This is due to the fact that the search is performed only on the reference data and so the closest resemblance to the black background is the low-intensity features of the melanocytic region.



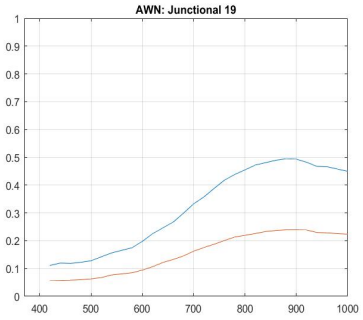
(a) AWN (b) ED (c) NS3

(d) SAM (e) SCA (f) SGA

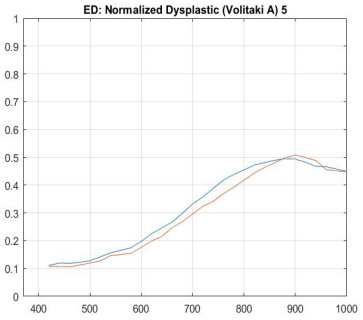
(g) SID-SAM (h) SID-SCA (i) SID

(j) SPM (k) SSD (l) SSS

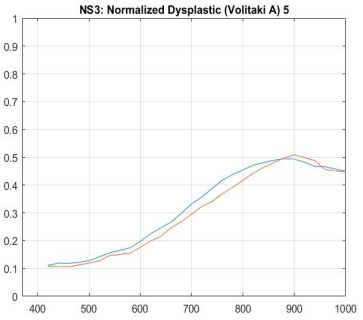
Figure 5.5: *Melanoma Sampe: The spectral signatures returned from each algorithm as a best match, when the unknown pixel is compared against the "Skin Lesions" category.*



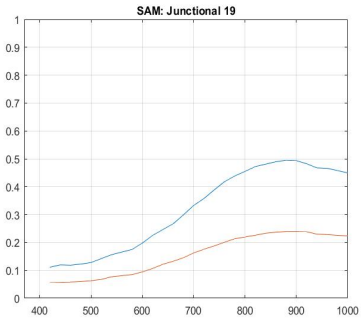
(a) AWN



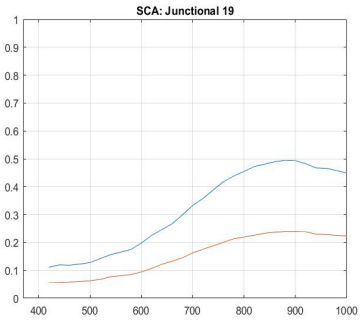
(b) ED



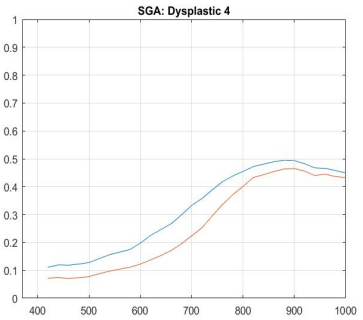
(c) NS3



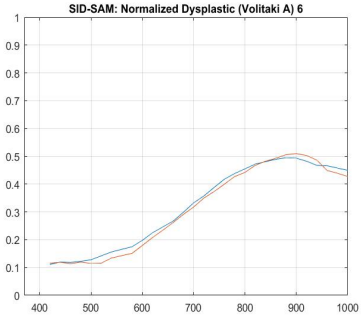
(d) SAM



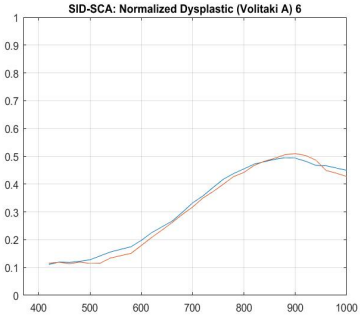
(e) SCA



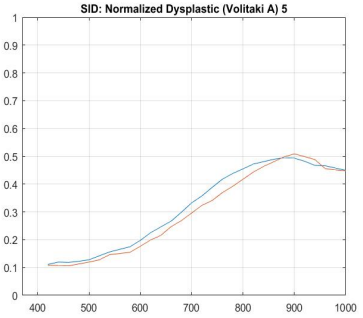
(f) SGA



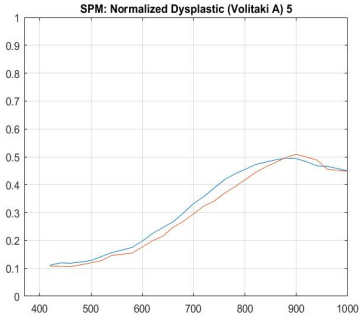
(g) SID-SAM



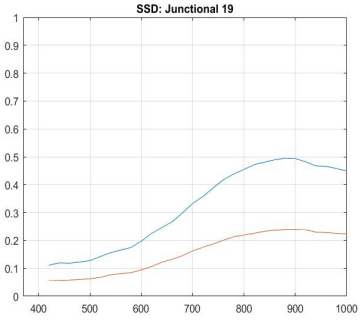
(h) SID-SCA



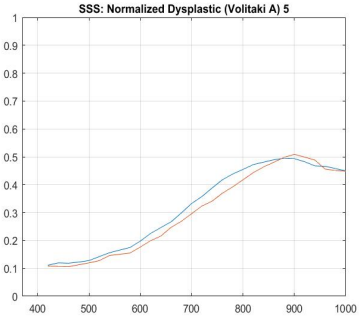
(i) SID



(j) SPM

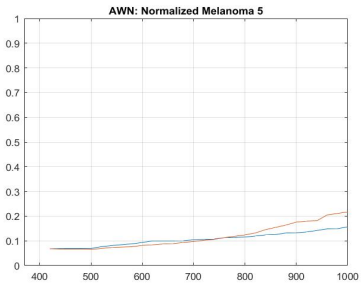


(k) SSD

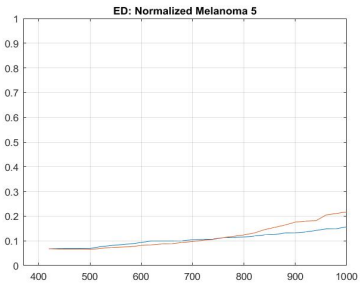


(l) SSS

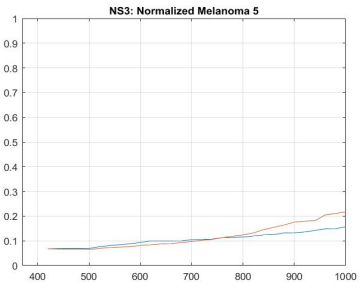
Figure 5.6: *Dysplastic Sampe: The spectral signatures returned from each algorithm as a best match, when the unknown pixel is compared against the "Skin Lesions" category.*



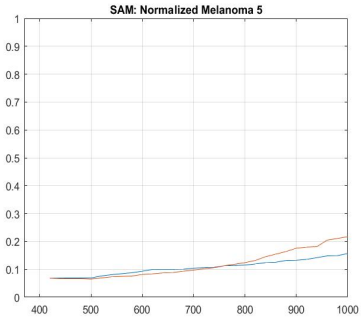
(a) AWN



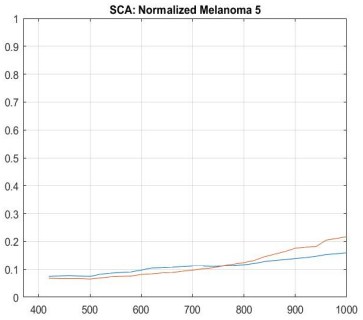
(b) ED



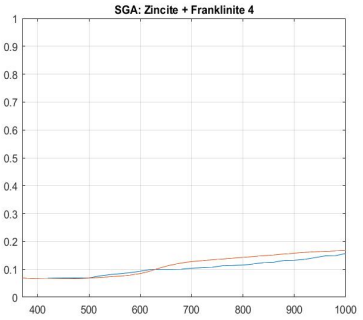
(c) NS3



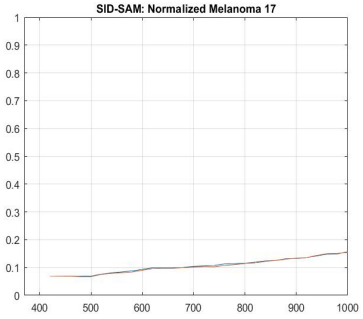
(d) SAM



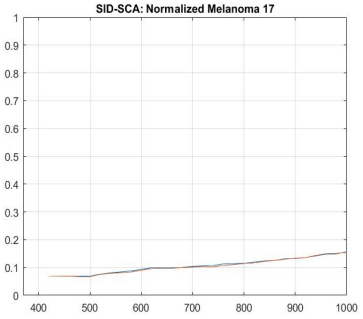
(e) SCA



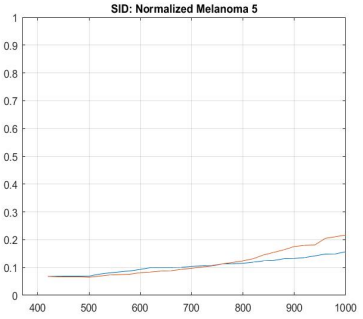
(f) SGA



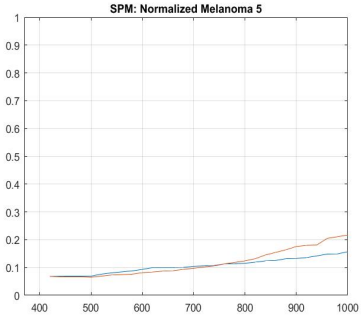
(g) SID-SAM



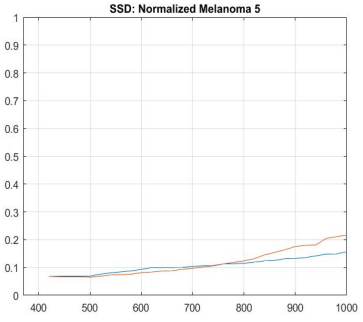
(h) SID-SCA



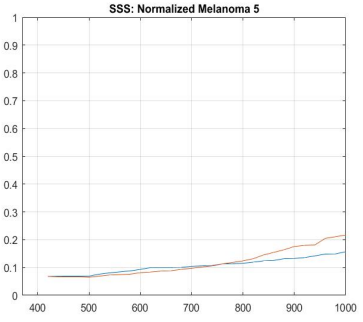
(i) SID



(j) SPM

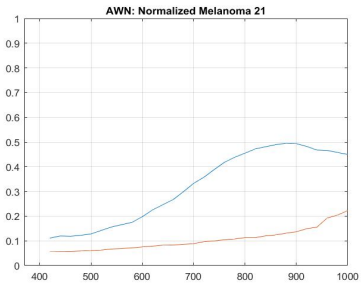


(k) SSD

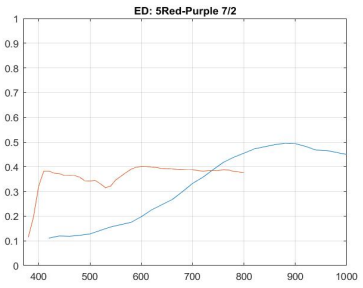


(l) SSS

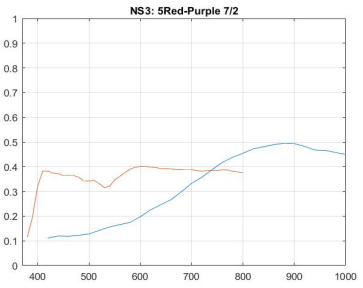
Figure 5.7: *Melanoma Sample: The spectral signatures returned from each each algorithm as a best match, when the unknown pixel is compared against the whole database.*



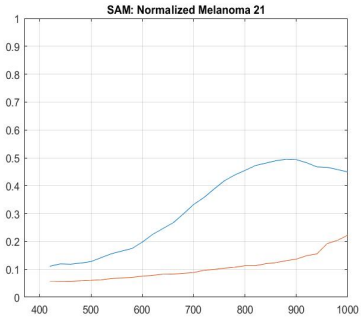
(a) AWN



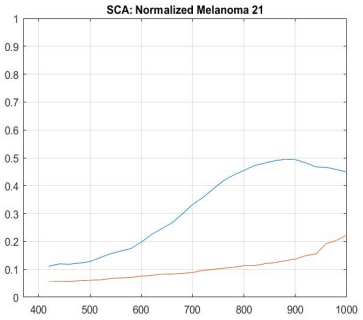
(b) ED



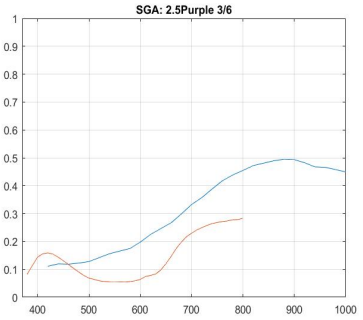
(c) NS3



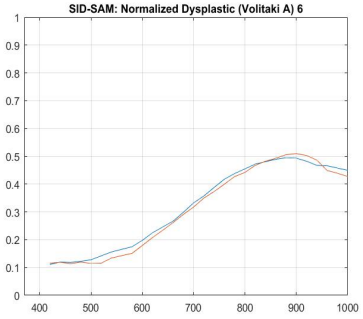
(d) SAM



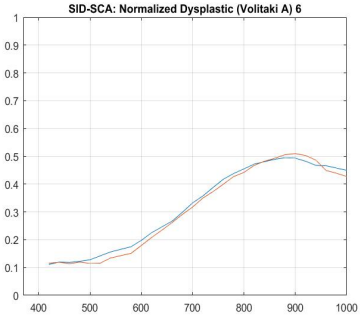
(e) SCA



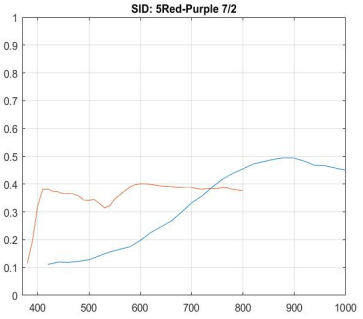
(f) SGA



(g) SID-SAM

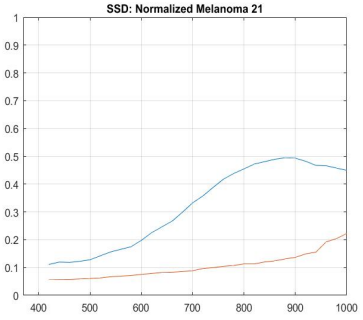


(h) SID-SCA



(i) SID

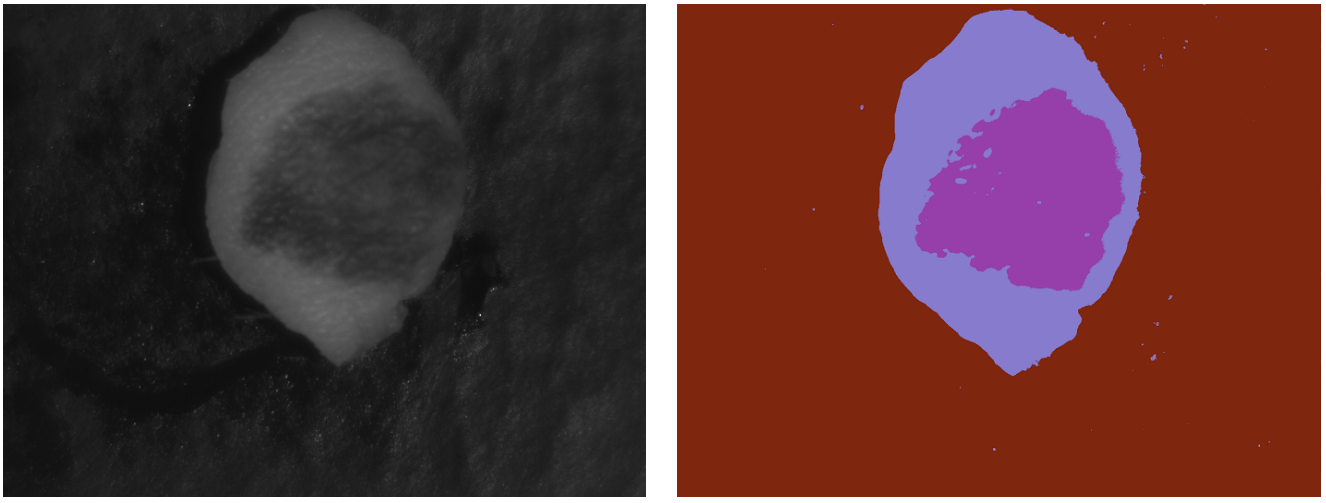
(j) SPM



(k) SSD

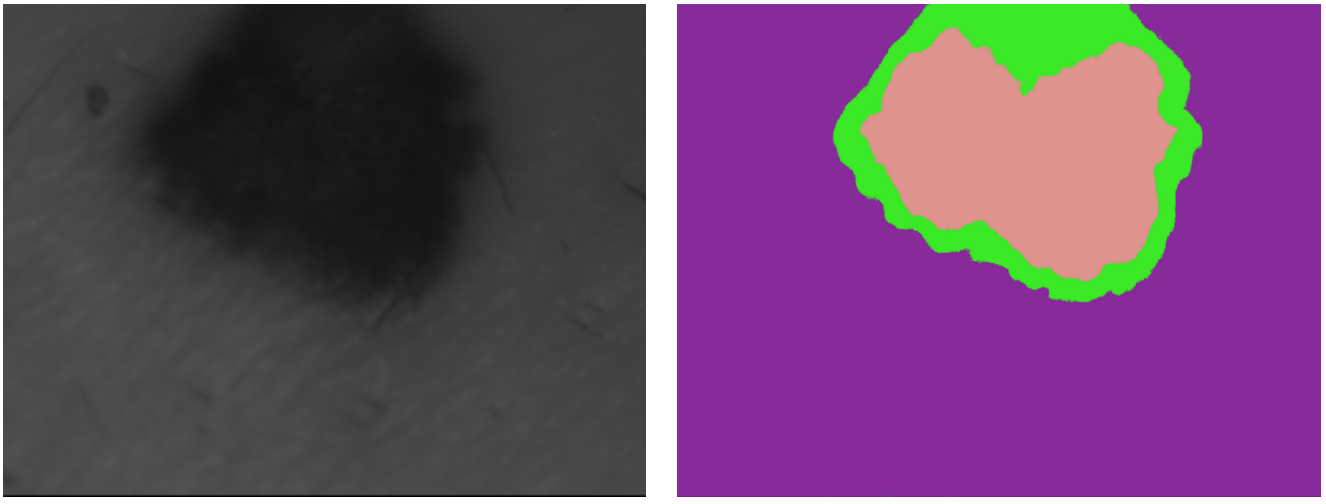
(l) SSS

Figure 5.8: *Dysplastic Sample: The spectral signatures returned from each each algorithm as a best match, when the unknown pixel is compared against the whole database.*



(a) Saple at 680nm (b) Thematic map

Figure 5.9: *Compound sample used for semi-supervised classification: a) shows the lesion at 680nm and b) the thematic map resulted after performing K-means Fast method for clustering*



(a) Saple at 680nm (b) Thematic map

Figure 5.10: *Melanoma sample used for semi-supervised classification: a) shows the lesion at 680nm and b) the thematic map resulted after performing K-means Fast method for clustering*

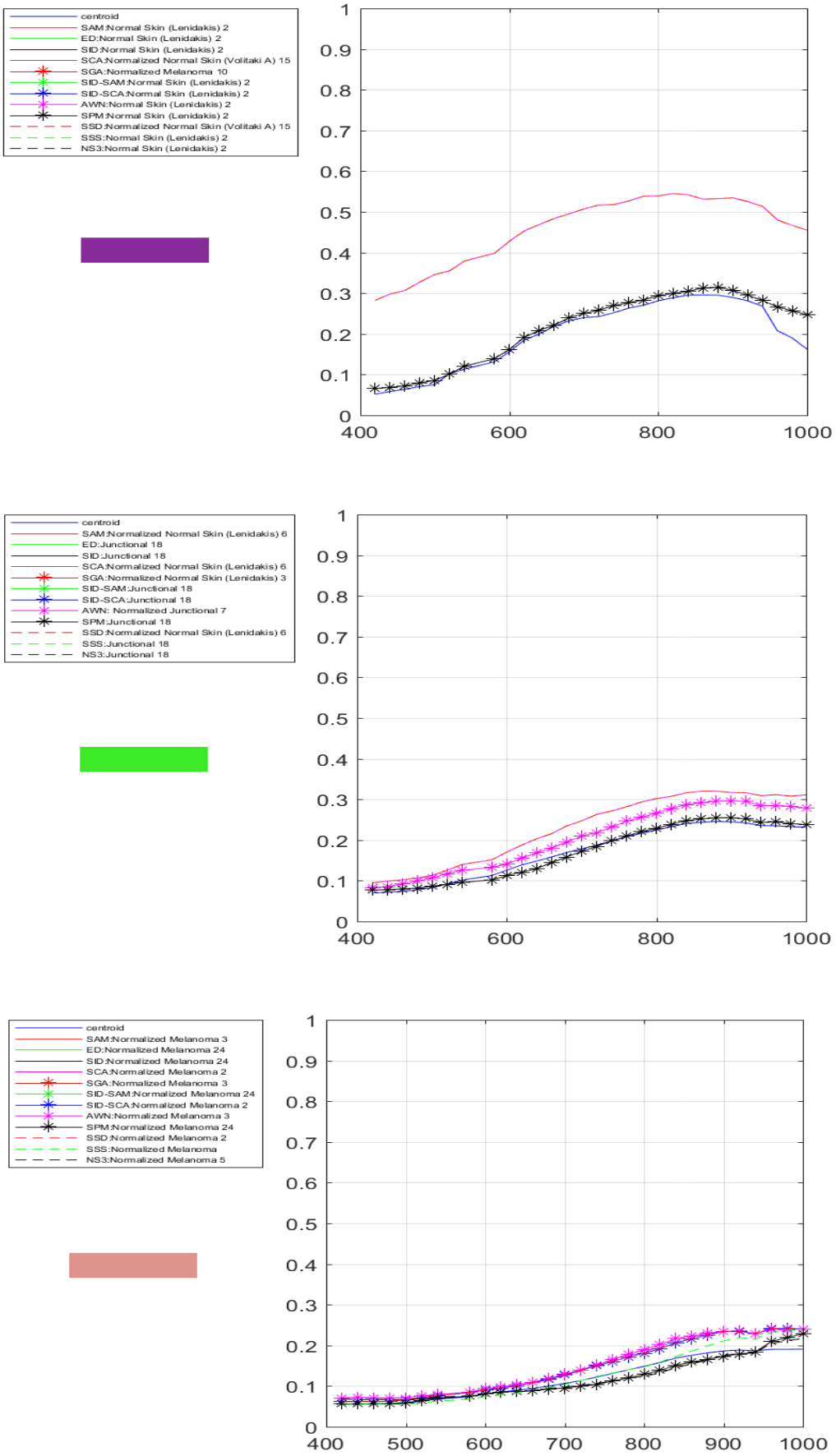


Figure 5.11: *Labelling results for the melanoma case.*

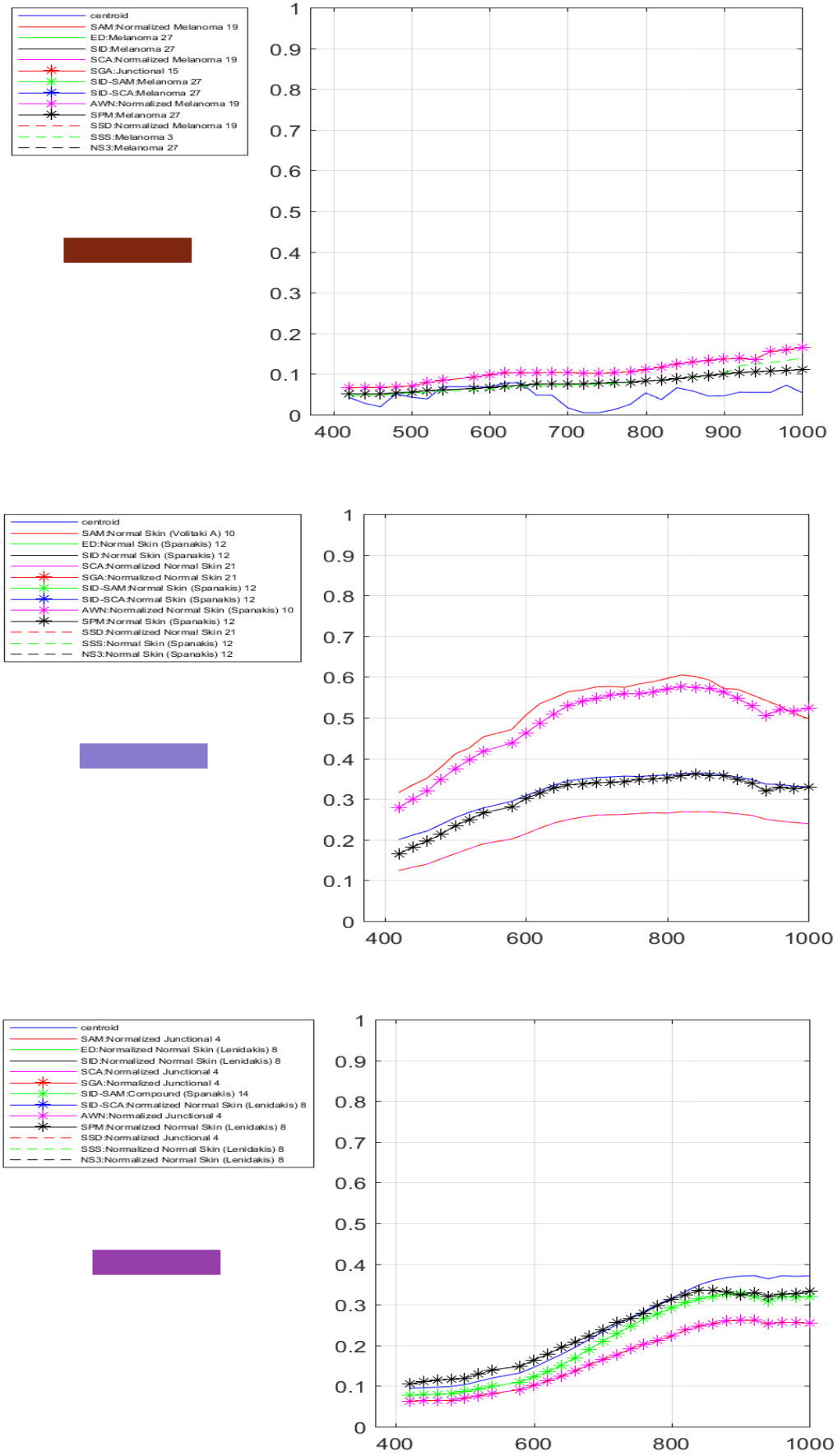


Figure 5.12: *Labelling results for the melanoma case.*

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This study was conducted in order to evaluate the use of a Spectral Database in material identification applications. A database was implemented and populated with spectral signatures of various materials. The spectral library was designed in such a way to allow for multiple categories to be stored in one library. This way it can be used in multiple applications. To perform spectral searching in the library, twelve Spectral Similarity algorithms were implemented and evaluated for their capacity to correctly label an unknown spectral signature. The similarity measures were then incorporated in the Spectral Suite software. This software offers the necessary tools for Hyperspectral Imaging analysis. The similarity measures were tested for pixel labeling and cluster labeling. In the first case, hyperspectral cubes are imported into the software and multiple pixels are selected from it. Subsequently, the extracted reflectances from the selected pixels are compared with the database records. Then, algorithms were assessed for their accuracy and time performance. In the second case, an Unsupervised classification was first performed prior to labelling. Then, the mean of each resulted class was labelled through a spectral library search.

In both pixel selection and cluster labelling, the similarity measurements were evaluated for their accuracy and time performance. The results showed that SID-SAM and SID-SCA methods are the most effective in retrieving a correct spectral signature from the database. This is due to the fact that by combining an algorithm sensitive to intensity changes like SID with an algorithm sensitive to feature changes (like SCA or SAM), the resulted method takes into account both characteristics to find the most similar spectral curve. On the other hand, AWN, SAM, SCA, and SGA also produced very good results, but their insensitivity in intensity changes can be a drawback for applications that intensity levels are of key importance. In regards to time performance, algorithms have small variations from one another, with SGA being the slowest

in all cases. Furthermore, the entropy can evaluate better measurements based on distances. Generally, since it is based on the numerical scores of the similarity measurement it cannot evaluate the accuracy of the produce spectral signature.

Additionally, the proposed implementation was tested on skin lesions. In-vitro samples of skin lesions were used for this purpose and spectral cubes were acquired from them using a Hyperspectral Imaging system. The database was populated with a dataset of references selected using the biopsy analysis for each lesion type. The reference data showed that spectral signatures from different lesion types tend to overlap with one another. The reference set was then used to label separate pixels and centroids of unsupervised classification. The results showed that SID-SAM and SID-SCA manage to retrieve spectral signatures that are very similar to the unknown pixel signature. On the clustering labelling, on the other hand, the results were ambiguous. The classes of normal skin were correctly labeled for each sample, but the identification of lesions was not always accurate.

The overlapping of spectral features of skin lesions types has gained considerate attention from researchers. An increasing number of studies[5-27] is conducted in an effort to find ways to discriminate the lesion types that lie between the normal skin and the melanoma spectral signatures. These studies use only two features in the spectral curves to classify the nevi in one of the skin cancer types. This study has shown that exploiting all the features in a spectral curve can assist in the accurate classification of a nevus. The proposed methods and algorithms have shown that the spectral signature of a lesion can be accurately identified even in a collection of completely diverse materials signatures.

An accurate classification of a skin lesion can play an important role in early skin cancer diagnosis thus saving thousands of lives. Moreover, it can help in avoiding unnecessary surgical operations in cases where a nevus is considered suspicious.

6.2 Future Work

The problem of the hard discrimination between skin lesion types is widely exploratory and interdisciplinary. For that reason, there is a number of expansions that could be made upon this. Some suggestions are:

1. Better feature selection and lesion discrimination: Some better feature selection and better class discrimination (by using trained and experienced doctor) can help the proposed method achieve higher levels of prediction.

2. Directly connect the spectral library and the Spectral Suite with a hyperspectral imaging system in order to provide live labeling: The final connection of the HSI system with

implemented software can provide dermatologists a vital tool for the non-invasive analysis of skin lesions.

3. Test the spectral similarity measurements for in-vivo applications: Although the tests for this thesis, have been performed on biopsies, the results can give quite an optimistic approach for in-vivo applications, thus using Hyper Spectral Imaging most useful tool: non-destructive analysis

Bibliography

[1] D. A. Boas, C. Pitris, and N. Ramanujam. *Handbook of Biomedical Optics*, chapter 7, pages 131–164. Taylor & Francis, 2011.

[2] T. Vo-Dinh. *Biomedical Photonics Handbook*, Taylor & Francis, 2010.

[3] R. Ramakrishnan, J. Gehrke *Database Management Systems*, McGraw-Hill, 2012.

[4] A. Tsapras, E. Terzakis, A. Makris, E. Papadakis, G. Papoutsoglou, E. Papagiannakis, C. Tsatsanis, E. Stathopoulos, and C. Balas. *"Hyper-Spectral Imaging for Skin Cancer Diagnosis in Mice"*. 6^{th} European Symposium on Biomedical Engineering, June 2008.

[5] A. Garcia-Uribe, J. Zou, M. Duvic, J. H. Cho-Vega, V. G. Prieto and L. V. Wang. *"In vivo diagnosis of melanoma and non-melanoma skin cancer using oblique incidence diffuse reflectance spectrometry "*. 2012. Available here.

[6] L. M. McIntosh, R. Summers, M. Jackson, H. H. Mantsch, J. R. Mansfield, M. Howlett, A. N. Crowson and J. W. P. Toole. *"Towards Non-Invasive Screening of Skin Lesions by Near-Infrared Spectroscopy"*. 2001. Available here.

[7] M. Kyriakidou, J. Anastassopoulou, A. Tsakiris, M. Koui and T. Theophanides. *"FT-IR Spectroscopy Study in Early Diagnosis of Skin Cancer "*. Available here.

[8] H. Lui, J. Zhao, D. McLean and H. Zeng. *"Real-time Raman Spectroscopy for In Vivo Skin Cancer Diagnosis "*. 2012. Available here.

[9] L. Rey-Barroso, F. J. Burgos-Fernndez, X. Delpueyo, M. Ares, S. Royo, J. Malvehy, S. Puig and M. Vilaseca. *"Visible and Extended Near-Infrared Multispectral Imaging for Skin Cancer Diagnosis "*. 2018. Available here.

[10] Murphy, B. W. Webster, R. J. Turlach, B. A. Quirk, C. J. Clay, C. D. Heenan. *"Toward the discrimination of early melanoma from common and dysplastic nevus using fiber optic diffuse reflectance spectroscopy "*. 2005. Available here.

[11] Zonios G., Dimou A., Bassukas I., Galaris D., Tsolakidis A., Kaxiras E. *"Melanin absorption spectroscopy: new method for noninvasive skin investigation and melanoma detection"*. 2008. Available here.

[12] A. Bjorgan, M. Milanic, L. Lyngsnes Randeberg. *"Estimation of skin optical parameters for real-time hyperspectral imaging applications "*. 2014. Available here.

[13] E. Borisova, Ts. Genova-Hristova, P. Troyanova, I. Terziev, E.A. Genina, A.N. Bashkatov, O. Semyachkina-Glushkovskaya, V. Tuchin, L. Avramov. *"TowardsOptical UV-VIS-NIR spectroscopy of benign, dysplastic and malignant cutaneous lesions ex vivo "*. 2018. Available here.

[14] I. Diebele, I. Kuzmina, J. Kapostinsh, A. Derjabo and J. Spigulis. *"Melanoma-nevus differentiation by multispectral imaging "*. 2011. Available here.

[15] I. Diebele, I. Kuzmina, J. Kapostinsh, A. Derjabo and J. Spigulis. *"Melanoma-nevus differentiation by multispectral imaging "*. 2011. Available here.

[16] Zonios G, Dimou A, Carrara M, Marchesini R. *"In vivo optical properties of melanocytic skin lesions: common nevi, dysplastic nevi and malignant melanoma "*. 2010. Available here.

[17] Gaudi Sudeep, Meyer Rebecca , Ranka Jayshree, Granahan James C., Israel Steven A., Yachik Theodore R., Jukic Drazen M. *"Hyperspectral Imaging of Melanocytic Lesions"*. 2014. Available here.

[18] Asad Safi, Victor Castaneda, Tobias Lasser, Nassir Navab. *"Skin Lesions Classification with Optical Spectroscopy"*. 2010. Available here.

[19] L. A. Zherdeva, I. A. Bratchenko, M. V. Alonova, O. O. Myakinin, D. N. Artemyev, A. A. Moryatov, S. V. Kozlov, V. P. Zakharov. *"Hyperspectral imaging of skin and lung cancers"*. 2016. Available here.

[20] R. Jolivot, Y. Benezeth, F. Marzani. *"kin Parameter Map Retrieval from a Dedicated Multispectral Imaging System Applied to Dermatology/Cosmetology"*. 2013. Available here.

[21] W. Verkruysse, R. Zhang, B. Choi, G.Lucassen, L. O. Svaasand, J. S. Nelson. *"A library based fitting method for visual reflectance spectroscopy of human skin"*. 2004. Available here.

[22] I. A. Bratchenko, V. P. Sherendak, O. O. Myakinin, D. N. Artemyev, A. A. Moryatov, E. Borisova, L. Avramov, L. A. Zherdeva, A. E. Orlov, S. V. Kozlov, V. P. Zakharov. *"In vivo hyperspectral imaging of skin malignant and benign tumors in visible spectrum"*. 2017. Available here.

[23] I. Diebele, A. Bekina, A. Derjabo, J. Kapostinsh, I. Kuzmina, J. Spigulis. *"Analysis of skin basalioma and melanoma by multispectral imaging"*. 2012. Available here.

[24] V. Zheludev, I. Plnen, N. Neittaanki-Perttu, A. Averbuch, P. Neittaanmki, M. Grnroos, H. Saari. *"Delineation of Malignant Skin Tumors by Hyperspectral Imaging Using Diffusion Maps Dimensionality Reduction"*. 2015. Available here.

[25] R. R. Winkelmann, D. S. Rigel, L. Ferris, A. Sober, N. Tucker, C. J. Cockerell. *"Correlation Between the Evaluation of Pigmented Lesions by a Multi-spectral Digital Skin Lesion Analysis Device and the Clinical and Histological Features of Melanoma"*. 2016. Available here.

[26] E. Borisova, Ts. Genova-Hristova, P. Troyanova, I. Terziev, E.A. Genina, A.N. Bashkatov, O. Semyachkina-Glushkovskaya, V. Tuchin, L. Avramov. *"Optical UV-VIS-NIR spectroscopy of benign, dysplastic and malignant cutaneous lesions ex vivo"*. 2018. Available here.

[27] D.J. Robinson, T.A. Middelburg, C.L. Hoy, E.R.M. de Haas, T.E.C. Nijsten, A. Amelink. *"Quantitative optical spectroscopy in the skin"*. 2013. Available here.

[28] J. Manuel Amigo, H. Babamoradi, S. Elcoroaristizabal. *"Hyperspectral image analysis. A tutorial"*. 2015.

[29] Hoong-Ta Lim, Vadakke Matham Murukeshan. *"Hyperspectral imaging of polymer banknotes for building and analysis of spectral library"*. 2017.

[30] C. Balas, K. Rapantzikos. *"Hyperspectral imaging: potential in non-destructive analysis of palimpsests"*. *IEEE-International Conference on Image Processing (ICIP)*, page 11–14, September 2005. Available here.

[31] D. Manolakis, D. Marden, and G. Shaw. *"Hyperspectral Image Processing for Automatic Target Detection Applications"*. *Lincoln Laboratory Journal*, 14(1), 2009. Available here.

[32] V. V. Tuchin. *"Handbook of Coherent-Domain Optical Methods: Biomedical Diagnostics, Environmental Monitoring, and Materials Science"*. Springer-Verlag GmbH, 2013.

[33] C. I. Chang. *"Hyper-spectral Data Processing: Algorithm Design and Analysis"*. Wiley, 2013.

[34] B. D. Bue, E. Mernyi, B. Csath. *"Automated Labeling of Materials in Hyperspectral Imagery"*. 2009.

[35] R. R. Nidamanuri, R. Anandakumar. *"Spectral identification of materials by reflectance spectral library search"*. 2015.

[36] H. Shen, P. Cai, S. Shao, and J. H. Xin. *"Reflectance reconstruction for multispectral imaging by adaptive Wiener estimation"*. 2007.

[37] S. Robila. *"An Investigation Of Spectral Metrics in Hyperspectral Image Preprocessing for Classification"*. 2014.

[38] J. Burger, A. Gowen. *"Data handling in hyperspectral image analysis"*. 2011.

[39] G. Healey, D. Slater. *"Models and Methods for Automated Material Identification in Hyperspectral Imagery Acquired Under Unknown Illumination and Atmospheric Conditions"*.

[40] S. Homayouni, M. Roux. *"Hyperspectral Image Analysis for Material Mapping Using Spectral Matching"*.

[41] Osmar Abilio de Carvalho Jnior, M. Roux. *"Spectral Correlation Mapper (SCM): An Improvement on the Spectral Angle Mapper (SAM)"*. 2017.

[42] Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., and Goetz, A. F. H. *"The spectral image processing system (SIPS)interactive visualization and analysis of imaging spectrometer data"*. 1993.

[43] Chang, C. *"An information theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis"*. 2000.

[44] van der Meer, F., and Bakker, W. *"Cross correlogram spectral matching: application to surface mineralogical mapping by using AVIRIS data from Cuprite, Nevada"*. 1997.

[45] Du, Y., Chang, C. I., Ren, H., Chang, C. C., Jensen, J. O., and DAmico, F. M.. *"New hyperspectral discrimination measure for spectral characterization"*. 2004.

[46] Naresh Kumar, M., Seshasai, M. V. R., Vara Prasad, K. S., Kamala, V., Ramana, K. V., Dwivedi, R. S., and Roy, P. S.. *"A new hybrid spectral similarity measure for discrimination among Vigna species"*. 2011.

[47] Granahan, J. C., and Sweet, J. N.. *"An evaluation of atmospheric correction techniques using the spectral similarity scale"*. 2001.

[48] Kong, X., Shu, N., Tao, J., and Gong, Y.. *"A new spectral similarity measure based on multiple features integration"*. 2011.

[49] Nidamanuri, R. R., and Zbell, B.. *"Normalized spectral similarity score (NS3) as an efficient spectral library searching method for hyperspectral image classification"*. 2011.

[50] Chang, C. I. *"Hyperspectral imaging: techniques for spectral detection and classification"*. 2003.

[51] C. A. Glasbey, G. W. Horgan. *"Image Analysis for the Biological Sciences"*. Wiley, July 1995.

[52] T. C. Poon, P. P. Banerjee. *"Contemporary Optical Image Processing with MATLAB"*. Elsevier Science, 2001.

[53] J. R. Parker. *"Algorithms for Image Processing and Computer Vision"*. Wiley, 2010.

[54] M. Petrou, C. Petrou. *"Image Processing: The Fundamentals"*. Wiley, 2010.

[55] Warren J. Smith. *"Modern Optical Engineering"*. McGraw Hill, 2000.

[56] Dimitris G. Manolakis. *"Hyperspectral Imaging Remote Sensing: Physics, Sensors and Algorithms"*. Cambridge University Press, 2016.

[57] Y. Garini, I. Young, and G. McNamara. *"Hyperspectral Spectral Imaging: Principles and Applications"*. 2006 International Society of Analytical Oncology, 2006. Available here.

[58] Di Wu, Da-Wen Sun. *"Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review – Part I: Fundamentals"*. Elsevier Science, 2013.

[59] R. Vadivambal, Digvir S. Jayas. *"Bio-Imaging: Principles, Techniques, and Applications"*. CRC Press, 2016.

[60] L. V. Wang and W. Hsin-L. *"Biomedical Optics Principles and Imaging"*. Wiley, 2007.

[61] Michael R. Hamblin, Pinar Avci, Gaurav K. Gupta. *"Imaging in Dermatology"*. Elsevier Science, 2016.

[62] R. R. Anderson, J. A. Parrish. *"The optics of human skin"*. J Invest Dermatol 1981;77(1):13-9.

[63] Petya Pavlova, Ekaterina Borisova, Lachezar Avramov, Elmira Petkova and Petranka Troyanova (2011). *"Investigation of Relations Between Skin Cancer Lesions' Images and Their Reflectance and Fluorescent Spectra, Melanoma in the Clinic - Diagnosis, Management and Complications of Malignancy"*, Prof. Mandi Murph (Ed.), ISBN: 978-953-307-571-6, InTech, Available here.

[64] Klaus Wolff, Richard A. Johnson, Arturo P. Saavedra. *"Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology"*. McGraw-Hill Professional, 2016.

[65] I. Diabele, I. Kuzmina, A. Lihachev, J. Kapostinsh, A. Derjabo, L. Valeine, J. Spigulis. *"Clinical evaluation of melanomas and common nevi by spectral imaging"*. Biomedical Optics Express 467, Vol. 3, No. 3, 2012.

[66] B. Farina, C. Bartoli, A. Bono, A. Colombo, M. Lualdi, G. Tragni, R. Marchesini. *"Multispectral imaging approach in the diagnosis of cutaneous melanoma: potentiality and limits"*. Phys. Med. Biol. 45(5), pp. 1243-1254, 2000.

[67] Estee L. Psaty, Allan C. Halpern. *"Current and emerging technologies in melanoma diagnosis: the state of the art"*. Clinics in Dermatology, Volume 27, Issue 1, pp. 35-45, 2009.

[68] M. Herlyn. *"Lessons from melanocyte development for understanding the biological events in naevus and melanoma formation"*. Melanoma Research, 2000. 10(4): pp. 303-12.

[69] F. Meier. *"Molecular events in melanoma development and progression"*. Frontiers in Bioscience: a Journal and Virtual Library, 1998. 3: pp. D1005-10.

[70] R. Gonzalez, R. Woods. *"Digital image processing"*. Reading, Massachusetts, Addison-Wesley Publishing Company. 1993. pp. 148-56.

[71] Angelopoulou. E., S. W. Lee, and R. Bajcsy . *"Spectral gradients: A material descriptor invariant to geometry and incident illumination "*. 1999.

[72] J. G. Ding, X. B. Li and L.Q. Huang. *"A Novel Method for Spectral Similarity Measure by Fusing Shape and Amplitude Features "*. 2015.

[73] American Cancer Society website Melanoma Skin Cancer. here.