

TECHNICAL UNIVERSITY OF CRETE

DIPLOMA THESIS

---

**Development of Supervised and  
Unsupervised Methods for Spectral  
Estimation in Hyper-Spectral Imaging**

---

*Author:*  
Fragkoulis LOGOTHETIS

*Supervisor:*  
Dr. Costas BALAS

*Comittee:*  
George Karystinos  
Vassilis Digalakis

*A thesis submitted in fulfillment of the requirements  
for the degree of Electrical and Computer Engineering  
in the*

Department of Electrical and Computer Engineering

October 3, 2018

*“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”*

Dave Barry

TECHNICAL UNIVERSITY OF CRETE

## *Abstract*

Technical University of Crete  
Department of Electrical and Computer Engineering

Electrical and Computer Engineering

### **Development of Supervised and Unsupervised Methods for Spectral Estimation in Hyper-Spectral Imaging**

by Fragkoulis LOGOTHETIS

Hyperspectral imaging (HSI) is an emerging technology that integrates conventional imaging and spectroscopy to attain both spatial and spectral information from an object. The spectral images collected in the spectral cube are tens of hundreds and the information we receive from them is crucial for many applications, such as bio-medical technology, remote sensing, microscopy etc. Nevertheless, the current state of the art includes HSI systems which need long acquisition time, something that prevents them from observing any dynamically developing phenomena. Also, are expensive and sizeable, which makes them inaccessible for many important applications. To address these limitations, a real-time snapshot spectral imaging system (SNI) that captures only six spectral bands, and by using dimensionality expansion techniques provides real time HIS was developed by us. To achieve that goal, a plenty of spectral estimation methods were leveraged and compared in terms of minimizing the estimation error of the uncaptured spectral images. Wiener, PCA, Random Forest Regression, Linear Regression, Neural Networks, and Support vector Regression are some of the methods that were applied to expand the dimensionality of the spectral domain. To do this more accurately, we propose a new Adaptive Hybrid method that overperforms the above spectral estimation models and achieves the lowest estimation error. The novelty of the proposed method stems mainly both from the combination of Wiener and PCA models and from the manner of the training sample selection. Looking at the same problem from another perspective, in order to increase the performance of the models, training procedure should be carried out using the most informative and distinctive bands as features. On that account, a great amount of band selection techniques, which are based either on geometric approaches or on similarity-based formulas were analyzed and compared. The main shortcoming is that there is no band extraction method specific for spectral estimation approaches. For that reason, a new band selection method specific for the spectral dimensionality expansion was devised by us. Experimental results show that the proposed method locate the most featured bands, which decrease dramatically the root mean square error between the reference and the estimated spectrum, compared with the well-known OSP method. These findings were set the basis for the development of a powerful SNI.

## *Acknowledgements*

After five years of truly hard yet exciting work and before presenting my thesis, I cannot but express my gratitude towards the people who were by my side and made everything I achieved possible.

First and foremost, I would like to thank my advisor, *Prof. Costas Balas*, for exposing me to research and pointing me towards exciting topics. I am also grateful towards the other two members of the committee. I had numerous intriguing discussions with *Prof. George Karystinos* and *Prof. Vassilis Digalakis* about technical issues. I am also very grateful to Papanathanasiou Athanasios, M.Sc Candidate, for his guidance and suggestions to improve my diploma thesis. I owe a lot to many, among my other Professors of TUC for their wise advice when I had to make critical decisions.

Nothing would have been possible without my family. I wish to thank my family and my close friends for their advising and support all these years of my studies.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction to Hyper Spectral Imaging - Estimation . . . . .	1
1.2 Light as electromagnetic wave and also as particle - Photoelectric Effect	2
1.3 Spectroscopy/ spectrometry . . . . .	3
1.4 Imaging spectroscopy . . . . .	4
1.5 Hyperspectral Imaging . . . . .	5
1.6 Single Exposure or Instantaneous Spectral Imagers . . . . .	6
1.7 Color vs. Spectral Imaging . . . . .	7
1.8 Spectral Imaging Applications . . . . .	7
1.9 What is Spectral Estimation . . . . .	8
1.9.1 Spectral Acquisition Chain of a Snapshot Spectral Imager . .	10
<b>2 Spectral Estimation - Problem Statement and Methods</b>	<b>11</b>
2.1 Measures of Spectral Similarities . . . . .	11
2.1.1 Goodness of Fit Coefficient (GFC) . . . . .	11
2.1.2 Spectral Angle Mapper (SAM) . . . . .	11
2.1.3 Euclidean Distance (ED) . . . . .	11
2.1.4 Root Mean Square Error (RMSE) . . . . .	12
2.2 Problem Statement . . . . .	12
2.3 Spectral Estimation Algorithms . . . . .	14
2.3.1 Estimation of Spectral Reflectances Using Wiener Method . . .	14
2.3.2 Estimation of Spectral Reflectances Using Projection on PCA	14
2.3.3 Estimation of Spectral Reflectances Using Fourier Analysis . .	15
2.3.4 Estimation of Spectral Reflectances Using Wavelets Analysis .	15
2.3.5 Estimation of Spectral Reflectances Using Hybrid Method . .	16
2.3.6 Estimation of Spectral Reflectances Using Adaptive-Wiener Method . . . . .	17
2.3.7 Estimation of Spectral Reflectances Using Adaptive Hybrid Method . . . . .	19
2.3.8 Estimation of Spectral Reflectances Using Multivariate Linear Regression Method . . . . .	19
2.3.9 Estimation of Spectral Reflectances Using Random Forest Re- gression . . . . .	21
2.3.10 Estimation of Spectral Reflectances Using Artificial Neural Networks . . . . .	23
2.3.11 Estimation of Spectral Reflectances Using Support Vector Re- gression . . . . .	25

<b>3</b>	<b>Spectral Estimation - Experiments</b>	<b>29</b>
3.1	Training Sets and Target Surfaces	29
3.1.1	Macbeth Color Checker	29
3.1.2	Munsell Mate Color Spectra Data	30
3.2	Hardware of the experiments	30
3.3	Evaluation of each Spectral Estimation Method	31
3.3.1	Evaluation Thresholds of Spectral Estimation Algorithms	31
3.3.2	Scoring the algorithms	32
3.3.3	Information about the experiments	32
3.4	First Series of the Experiments	36
3.4.1	Wiener Method	37
3.4.2	PCA Method	39
3.4.3	Hybrid Method	41
3.4.4	Adaptive Wiener Method	43
3.4.5	Adaptive Hybrid Method	45
3.4.6	Linear Regression Method	47
3.4.7	Random Forest Method	49
3.4.8	Neural Network Method	51
3.4.9	SVR method	53
3.4.10	First Comparison of the Methods	55
3.5	Second Series of Experiments	57
3.5.1	Wiener Method	58
3.5.2	PCA Method	60
3.5.3	Hybrid Method	62
3.5.4	Adaptive Wiener Method	64
3.5.5	Adaptive Hybrid Method	66
3.5.6	Linear Regression Method	68
3.5.7	Random Forest Method	70
3.5.8	Neural Network Method	72
3.5.9	SVR method	74
3.5.10	Second Comparison of the Methods	76
3.6	Total Scoring	78
3.7	Execution Times of the Spectral Estimation Algorithms	79
<b>4</b>	<b>Band Selection - Feature Extraction</b>	<b>80</b>
4.1	Introduction to Band Selection	80
4.2	Data Preprocessing	81
4.3	Properties of Similarity-Based Band Selection	81
4.4	Band Selection Algorithms	82
4.4.1	Information Entropy	82
4.4.2	Principal Component Analysis Ranking (PCAr)	82
4.4.3	Linear prediction (LP)	83
4.4.4	Orthogonal Subspace Projection (OSP)	83
	Band Used as the Initial for Band Selection for OSP and LP	84
4.4.5	Correlation Measure (CorM)	84
4.4.6	Sparse Principal Component Analysis of a constant-rank matrix	84
4.4.7	BS (Band Selection Algorithm)	85
4.4.8	Band selection using Linear Interpolation (LIBS)	86
4.5	Practical Considerations	88
4.6	Experiments for evaluation the band selection methods	88
4.6.1	Information about the experiments	88

4.6.2	Conclusions of the Band Selection Experiments . . . . .	91
<b>5</b>	<b>Conclusions and Future Work</b>	<b>92</b>
	<b>Bibliography</b>	<b>94</b>

# List of Figures

1.1	A Spectral Cube (source: Quantative BioImaging Laboratory,(QBIL) [1].)	1
1.2	Light as an electromagnetic wave (source: Physicstuff [2].)	2
1.3	Spectroscopy using a simple prism (source: Astronomy, Cosmos [3]).	3
1.4	Principle of (remote sensing) imaging spectroscopy (source: Markelowitz [4]).	4
1.5	Electromagnetic spectrum (source: Wikiversity [5]).	5
1.6	Typical (hyper)spectral imaging approaches. (A) Point scan. (B) Line scan (i.e. “pushbroom”). (C) Wavelength scan. (D) snapshot (source: Wang, Yu and P. Reder [6]).	6
1.7	RBG vs Spectral Cube (source: [7]).	7
1.8	Spectral and spatial domain	8
1.9	Spectral Acquisition Chain of a Snapshot Spectral Imager	10
2.1	Synopsis of the spectral model of the acquisition process in a multi-spectral system (source: Alamin Mansouri [9]).	12
2.2	Hybrid model with different values of w.	17
2.3	Two typical examples of the training sets for the given candidate samples. Channel number C=6-Adaptive Wiener (source: Hui-Liang Shen [14]).	18
2.4	Flow Chart of Decision Tree Regression Algorithm.	20
2.5	Example of partitions of decision trees.	21
2.6	Example of random forest regression (source: Assignmentpoint.com) [16]).	21
2.7	Model of a biological neuron (source: Wikipedia).	23
2.8	The Neural Network for the spectral estimation problem.	23
2.9	Model of a biological neuron (source: Wikipedia).	24
2.10	The soft margin loss setting for a linear SVM (source: Wei-Feng Zhang [19]).	27
2.11	Architecture of a regression machine constructed by the SV algorithm (source: Wei-Feng Zhang[19]).	28
3.1	Macbeth Color Checker (source: phononet.com [23]).	29
3.2	Munsell Mate Color Spectra Data.	30
3.3	MUSES-9 HS (source: spectricon.com [25]).	31
3.4	Macbeth Color Checker (96 isolated patches).	32
3.5	Software for experiments using Matlab Tools.	33
3.6	Software for experiments using Qt cross-platform SDK.	34
3.7	PCA score images of Macbeth.	34
3.8	Example of a Color-Map.	35
3.9	First-Series of Experiments. Training/Validation set contains only the patches into the green grid.	36
3.10	First Experiments - Wiener Color Map.	37
3.11	First Experiments - Wiener Performance.	37

3.12	First Experiments - Wiener, Patch 2M-Patch 8M.	38
3.13	First Experiments - PCA Color Map.	39
3.14	First Experiments - PCA Performance.	39
3.15	First Experiments - PCA, Patch 2M-Patch 8M.	40
3.16	First Experiments - Hybrid Color Map.	41
3.17	First Experiments - Hybrid Performance.	41
3.18	First Experiments - Hybrid, Patch 2M-Patch 8M.	42
3.19	First Experiments - Adaptive Wiener Color Map.	43
3.20	First Experiments - Adaptive Wiener Performance.	43
3.21	First Experiments - Adaptive Wiener, Patch 2M-Patch 8M.	44
3.22	First Experiments - Adaptive Hybrid Color Map.	45
3.23	First Experiments - Adaptive Hybrid Performance.	45
3.24	First Experiments - Adaptive Hybrid, Patch 2M-Patch 8M.	46
3.25	First Experiments - Linear Regression Color Map.	47
3.26	First Experiments - Linear Regression Performance.	47
3.27	First Experiments - Linear Regression, Patch 2M-Patch 8M.	48
3.28	First Experiments - Random Forest Color Map.	49
3.29	First Experiments - Random Forest Performance.	49
3.30	First Experiments - Random Forest, Patch 2M-Patch 8M.	50
3.31	First Experiments - Neural Network Color Map.	51
3.32	First Experiments - Neural Network Performance.	51
3.33	First Experiments - Neural Network, Patch 2M-Patch 8M.	52
3.34	First Experiments - SVR Color Map.	53
3.35	First Experiments - SVR Performance.	53
3.36	First Experiments - SVR, Patch 2M-Patch 8M.	54
3.37	First Comparison of the Methods.	55
3.38	First ranking of the Methods.	55
3.39	Isolated patches of Munsell in the greed grid for experiments.	57
3.40	Second Experiments - Wiener Color Map.	58
3.41	Second Experiments - Wiener Performance.	58
3.42	Second Experiments - Wiener, Patch 2M-Patch 8M.	59
3.43	Second Experiments - PCA Color Map.	60
3.44	Second Experiments - PCA Performance.	60
3.45	Second Experiments - PCA, Patch 2M-Patch 8M.	61
3.46	Second Experiments - Hybrid Color Map.	62
3.47	Second Experiments - Hybrid Performance.	62
3.48	Second Experiments - Hybrid, Patch 2M-Patch 8M.	63
3.49	Second Experiments - Adaptive Wiener Color Map.	64
3.50	Second Experiments - Adaptive Wiener Performance.	64
3.51	Second Experiments - Adaptive Wiener, Patch 2M-Patch 8M.	65
3.52	Second Experiments - Adaptive Hybrid Color Map.	66
3.53	Second Experiments - Adaptive Hybrid Performance.	66
3.54	Second Experiments - Adaptive Hybrid, Patch 2M-Patch 8M.	67
3.55	Second Experiments - Linear Regression Color Map.	68
3.56	Second Experiments - Linear Regression Performance.	68
3.57	Second Experiments - Linear Regression, Patch 2M-Patch 8M.	69
3.58	Second Experiments - Random Forest Color Map.	70
3.59	Second Experiments - Random Forest Performance.	70
3.60	Second Experiments - Random Forest, Patch 2M-Patch 8M.	71
3.61	Second Experiments - Neural Network Color Map.	72
3.62	Second Experiments - Neural Network Performance.	72

3.63	Second Experiments - Neural Network, Patch 2M-Patch 8M. . . . .	73
3.64	Second Experiments - SVR Color Map. . . . .	74
3.65	Second Experiments - SVR Performance. . . . .	74
3.66	Second Experiments - SVR Performance. . . . .	75
3.67	Second Comparison of the Methods. . . . .	76
3.68	Second ranking of the Methods. . . . .	76
3.69	Total scoring of the Methods. . . . .	78
3.70	Execution Times of the Spectral Estimation Algorithms. . . . .	79
4.1	LIBS algorithm. . . . .	87
4.2	Procedure of finding feature points. . . . .	87
4.3	Comparison of Band Selection Methods. . . . .	90
4.4	Ranking of Band Selection Methods. . . . .	90

# List of Tables

3.1	Evaluation Thresholds of Spectral Estimation Algorithms. . . . .	31
3.2	Scoring Algorithms. . . . .	32
3.3	Color-Map and the quality of estimation. . . . .	36
3.4	Metrics of Error for the First Series of Experiments. . . . .	56
3.5	Metrics of Error of the Second Series of Experiments. . . . .	77
4.1	Selected bands from the band selection algorithms. . . . .	89
4.2	Metrics of Error for the Band Selection Experiments. . . . .	89

# Physical Constants

$$\begin{array}{ll} \text{Speed of Light} & c = 2.997\,924\,58 \times 10^8 \text{ m s}^{-1} \\ \text{Planck constant} & h = 6.626\,070\,04 \times 10^{-12} \text{ m}^2 \text{ kg/s} \end{array}$$

*Dedicated to my family...*

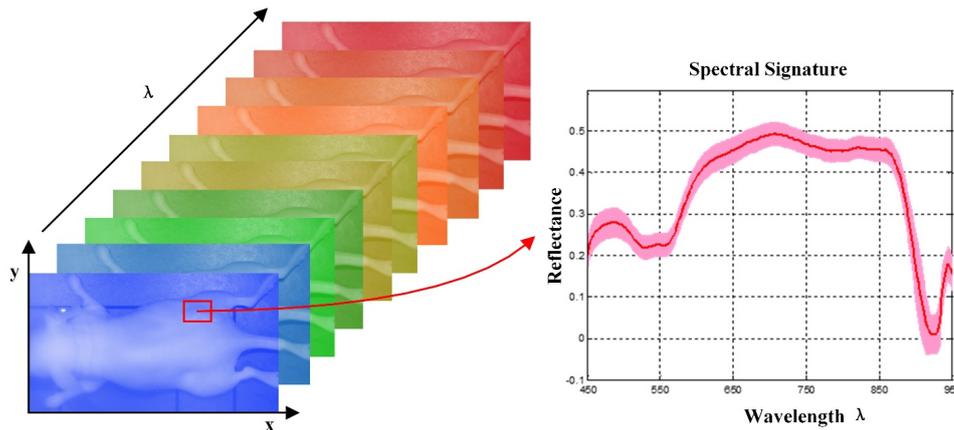


FIGURE 1.1: A Spectral Cube (source: Quantative BioImaging Laboratory,(QBIL) [1].)

## Chapter 1

# Introduction

### 1.1 Introduction to Hyper Spectral Imaging - Estimation

Over the last two decades, the field of optical imaging has developed very rapidly providing color imaging (CI) systems with very high resolution that nowadays allow 3D imaging as well as video capture with a very high frame rate. These systems try to emulate human vision in order to reproduce a result (image) that resembles the actual scene as it was perceived by a human eye. Usually these systems produce three-dimensional data (RGB is more commonly used) where each of the three dimensions represents the intensity and chrominance of light.

Although CI systems can provide an accurate representation of the scene at hand, there is a great deal of information that cannot be perceived with these systems; information that may be hidden in the visible spectrum or beyond it, i.e. UV or Infrared regions of the spectrum. In the direction to obtain the hidden information, spectral imaging systems are used.

Spectral Imaging (SI) is the application of reflectance spectroscopy to every pixel in a spatial image. Every spatial image captured represents a different wavelength within the electromagnetic spectrum and each pixel represents the spectral power distribution of the scene at that point. The stack of images created from this system is called *Spectral Cube* (Fig. (1.1)), and the data are represented in multidimensional, spanning spatial and spectral dimensions ( $x, y, \lambda$ ).

Spectroscopy can be used to detect individual absorption features due to specific chemical bonds in a solid, liquid, or gas. Solids can be either crystalline (i.e. minerals) or amorphous (such as glass). Every material is formed by chemical bonds and

## Electromagnetic Wave

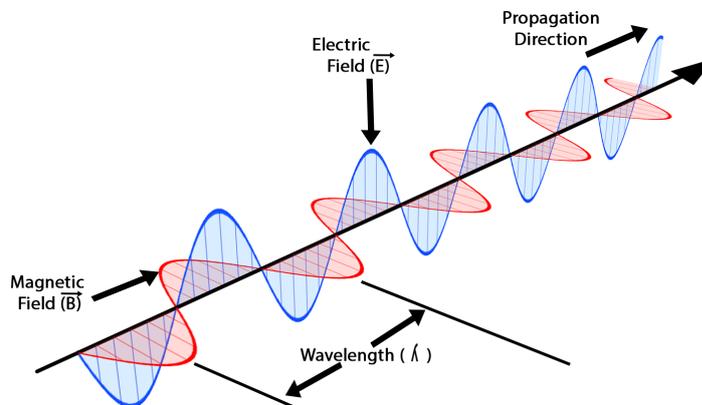


FIGURE 1.2: Light as an electromagnetic wave (source: Physicstuff [2].)

has the potential for detection with spectroscopy. Actual detection is dependent on the spectral coverage, spectral and spatial resolution, and signal-to-noise of the SI system, the abundance of the material, and the strength of absorption features for that material in the wavelength region measured.

SI systems are widely used nowadays in numerous fields such as medicine, astronomy, industry, military etc. and due to the continuous need to improve the preexisting techniques and methods, many innovations and advances were developed.

The information provided by these systems, which is usually not discernible to the human eye, allows useful facts and phenomena to be revealed. Unfortunately, SI systems are expensive and sizable, which makes them inaccessible for many applications. The acquisition and computational time that is needed is very high, which prevents the system from observing any dynamically developing phenomena.

To address these limitations, real-time snapshot spectral imaging systems need to be developed that allow simultaneous Hyperspectral imaging (HSI) and has lower cost and size. We have developed and proposed a novel real-time HSI system that can simultaneously acquire six spectral bands and provide spectral information in any desired wavelength within the visible and infrared spectrum.

## 1.2 Light as electromagnetic wave and also as particle - Photoelectric Effect

Starting from the beginning of that science, it is important to be clarified that light is one of the fundamental phenomena of electromagnetism that behaves as wave and also as photon particle. It is characterized as wave when is emitted and propagated and as particle when it is absorbed. As an electromagnetic wave, it has both electric and magnetic field components, which oscillate in a fixed relationship to one another, perpendicular to each other and perpendicular to the direction of the energy and wave propagation as it is illustrated in Fig. (1.2). Each electromagnetic wave is characterized by the frequency or wavelength of its wave. The electromagnetic spectrum, in order of increasing frequency and decreasing wavelength, consists of radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays and gamma rays. The eyes of various organisms sense a somewhat variable but relatively small

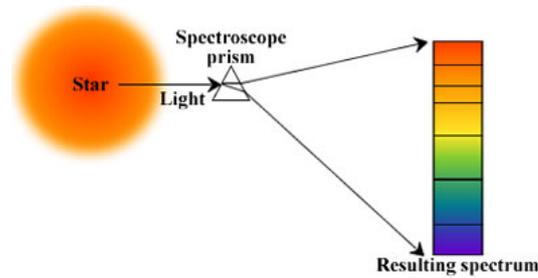


FIGURE 1.3: Spectroscopy using a simple prism (source: Astronomy, Cosmos [3]).

range of frequencies called the visible spectrum of the light. Higher frequencies correspond to proportionately more energy carried by each photon; for instance, a single gamma ray photon carries far more energy than a single photon of visible light. This phenomenon was described first time by Einstein, who realized that the energy( $E$ ) of any photon is given by the equation

$$E = h \frac{c}{\lambda}$$

where  $h$  is the Planck constant,  $c$  is the speed of the light in vacuum and  $\lambda$  is the photon's wavelength.

In classical physics, electromagnetic radiation is considered to be produced when charged particles are accelerated by forces acting on them. Electrons are responsible for emission of most electromagnetic radiation because they have low mass, and therefore are easily accelerated by a variety of mechanisms. Rapidly moving electrons are most sharply accelerated when they encounter a region of force, so they are responsible for producing much of the highest frequency electromagnetic radiation observed in nature. Quantum processes can also produce electromagnetic radiation, such as when atomic nuclei undergo gamma decay, and processes such as neutral pion decay.

As it mentioned before light behaves as particle when is absorbed. In other words, that is exactly the reason of the photoelectric effect, in which “packets” of energy, called photons increase the maximum kinetic energy of the photoelectrons emitted from a given metal. Einstein got the Nobel Price in 1921 for his paper publication advancing the hypothesis that light energy is carried in discrete quantized packets to explain experimental data and the photoelectric effect. Photoelectric effect lead the scientific community to implement image sensors, which could read those “packets” of energy and produce a real digital image characterized by high color fidelity.

In this work, focusing in the duality of light is needed. The science of spectroscopy could bring out interesting information about the interaction between radiation and the chemical structure of the objects as a function of the wavelength  $\lambda$ .

### 1.3 Spectroscopy/ spectrometry

Spectroscopy was originally the study of the interaction between radiation and the structure of matter as a function of the wavelength  $\lambda$ . In fact, historically, spectroscopy referred to the use of visible light dispersed according to its wavelength, e.g. by a prism (Fig. (1.3)). Later, the concept was expanded greatly to comprise any measurement of a quantity as function of either wavelength or frequency. Thus, it

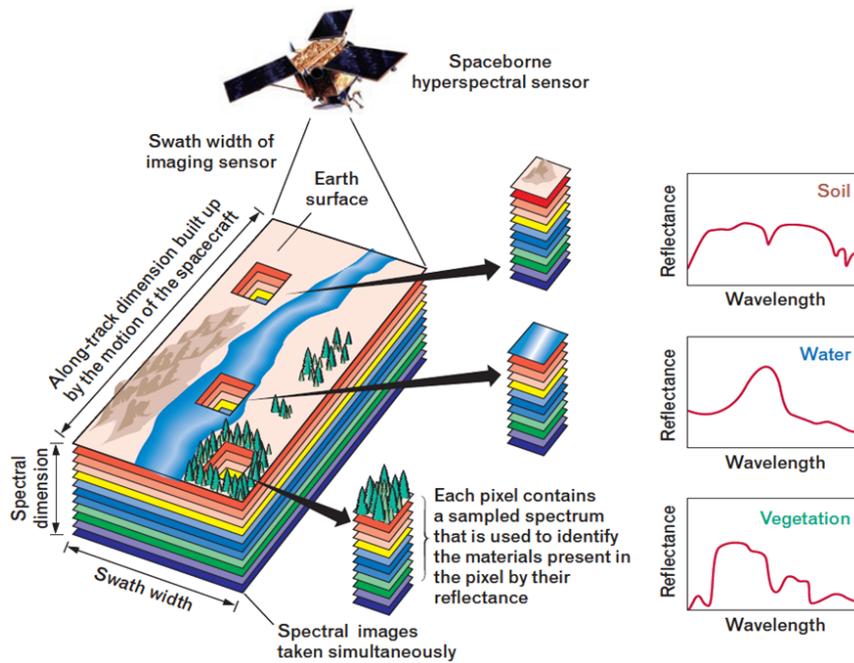


FIGURE 1.4: Principle of (remote sensing) imaging spectroscopy (source: Markelowitz [4]).

can also refer to interactions with particle radiation or to a response to an alternating field or varying frequency  $\nu$ . A further extension of the scope of the definition added energy ( $E$ ) as a variable, once the very close relationship

$$E = h\nu$$

for photons was realized as it mentioned in Section 1.2. Spectrometry is the spectroscopic technique used to assess the concentration or amount of a given species. In those cases, the instrument that performs such measurements is a spectrometer or spectrograph.

Spectroscopy/spectrometry is often used in physical and analytical chemistry for the identification of substances through the spectrum emitted from or absorbed by them. Spectroscopy/spectrometry is also heavily used in astronomy and remote sensing. Most large telescopes have spectrometers, which are used either to measure the chemical composition and physical properties of astronomical objects or to measure their velocities from the Doppler shift of their spectral lines.

## 1.4 Imaging spectroscopy

Imaging spectroscopy (also spectral imaging or chemical imaging) is the application of reflectance spectroscopy to every pixel in a spatial image. In remote sensing situations (Fig. (1.4)), the surface materials mapped must be exposed in the optical surface (e.g., to map surface mineralogy it must not be covered with vegetation), and the diagnostic absorption features must be in regions of the spectrum that are reasonably transparent to the atmosphere. The optical surface is the same as what the geologist sees in the field with his or her eyes. Spectroscopy can be used in laboratories on hand samples, in the field with portable field spectrometers (spatial resolution in the millimeter to several meter range), from aircraft, and in the future from satellites. The aircraft systems can image large areas in short time (2 sq. km per second),

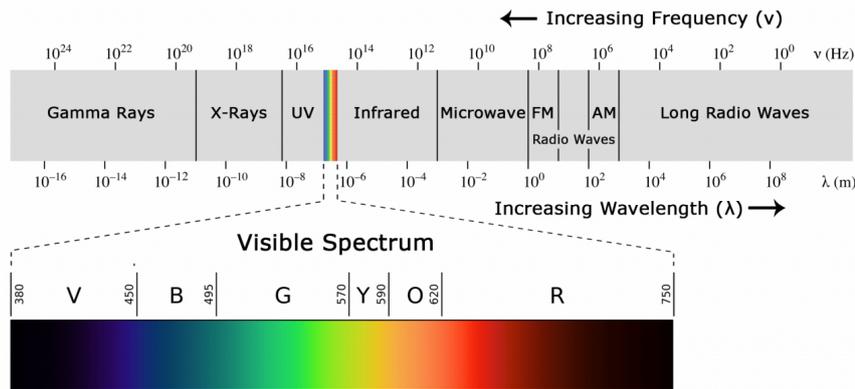


FIGURE 1.5: Electromagnetic spectrum (source: Wikiversity [5]).

producing spectra for each pixel that can be analyzed for specific absorption bands and thus specific materials. These measurements can then be used for the unambiguous direct and indirect identification of surface materials and atmospheric trace gases, the measurement of their relative concentrations, subsequently the assignment of the proportional contribution of mixed pixel signals (e.g., the spectral unmixing problem), the derivation of their spatial distribution (mapping problem), and finally their study over time (multi-temporal analysis).

Imaging spectroscopy can be considered as the equivalent of color photography, but each pixel needs to acquire many bands of light intensity data from the spectrum, instead of just the three bands of the RGB color model. More precisely, it is the simultaneous acquisition of spatially co-registered images in many spectrally contiguous bands. In order to improve the spectroscopy imaging the scientific community turned into the HSI.

## 1.5 Hyperspectral Imaging

Hyperspectral Imaging collects and processes information from across the electromagnetic spectrum. Unlike the human eye, which just sees visible light, HSI is more like the eyes of the “mantis shrimp”, which can see visible light as well as from the ultraviolet to infrared. Hyperspectral capabilities enable the “mantis shrimp” to recognize different types of coral, prey, or predators, all which may appear as the same color to the human eye.

Humans build sensors and processing systems to provide the same type of capability for application in agriculture, mineralogy, physics, and surveillance and other fields of science. Hyperspectral sensors collect information as a set of “images”. Each image represents a range of the electromagnetic spectrum (Fig. (1.5)), and is also known as a spectral band. Hyperspectral sensors look at objects using a vast portion of the electromagnetic spectrum. Certain objects leave unique “fingerprints” across the electromagnetic spectrum. These “fingerprints” are known as *spectral signatures* and enable identification of the materials. For example, having the spectral signature for oil helps mineralogists find new oil fields.

The precision of these sensors is typically measured in spectral resolution, which is the width of each band of the spectrum that is captured. If the scanner picks up a large number of fairly small wavelengths, it is possible to identify objects even if said objects are only captured in a handful of pixels. However, spatial resolution is a factor in addition to spectral resolution. If the pixels are too large, then multiple

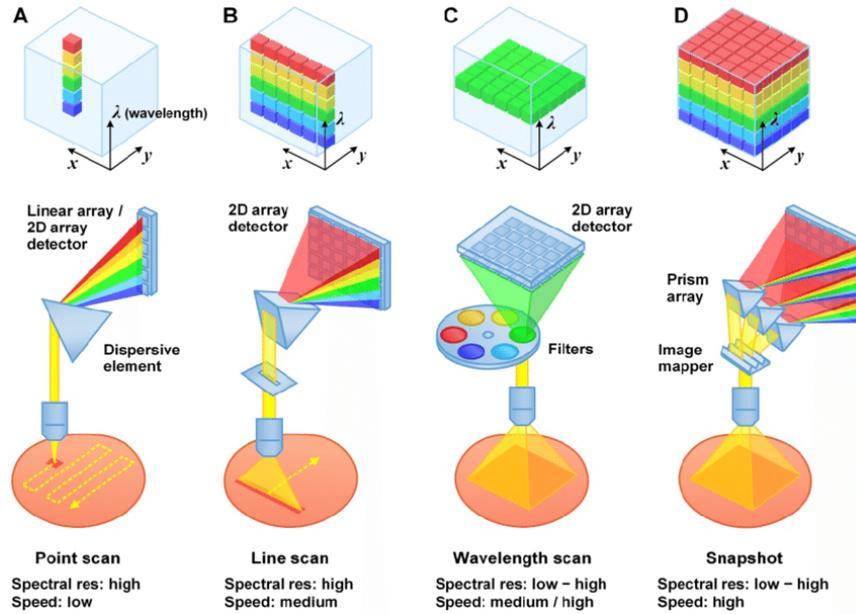


FIGURE 1.6: Typical (hyper)spectral imaging approaches. (A) Point scan. (B) Line scan (i.e. “pushbroom”). (C) Wavelength scan. (D) snapshot (source: Wang, Yu and P. Reeder [6]).

objects are captured in the same pixel and become difficult to identify. If the pixels are too small, then the energy captured by each sensor-cell is low, and the decreased signal-to-noise ratio reduces the reliability of measured features.

## 1.6 Single Exposure or Instantaneous Spectral Imagers

In order to record spectral images, we use tunable filtered-based spectral imaging systems. Those systems record spectral images in a time-sequentially manner and obtain the spectra from post hoc assembly of the time-sequential data. This reveals a major disadvantage of these systems. When we have phenomena that are changing on a time scale that is shorter than the duration required for recording the spectral cube, the SI systems described cannot perform accurately. Furthermore, the scene recorded from SI systems must be static. Otherwise problems will be created in the co-registration of the spectral images, which is needed in order to provide accurate spectra. There are numerous of applications (i.e biomedical and etc) that require spectral imaging and analysis of transient moving scenes. This is why single shot or instantaneous spectral imagers (Fig. (1.6)) need to be developed. Specifically, there are four techniques to develop a spectral scanner. The most simple structure of a HS imager is the point scan (Fig. (1.6.A)), where a line array detector acquires all the spectral information. However, point scan technique is characterized by really low speed. Furthermore, line scan and wavelength scan are often used, but it is more profitable to use the snapshot idea in order to handle quick changing effects. Snapshot imager systems have many advantages over SI systems such as fast acquisition of accurately registered images, high device robustness and reliability, low cost etc. But, there are many trade-offs in order to achieve all that. Due to current technological limitations there is a trade-off between spatial and spectral resolution. Specifically, a snapshot imager can capture either a small number of spectral images with high resolution or a great number of spectral images with really low spatial resolution. Therefore, in the case of stationary and invariant scenes the spectral imaging systems,

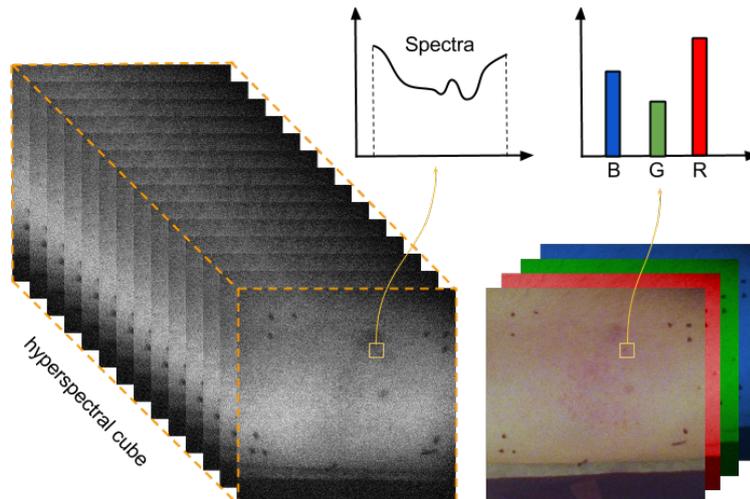


FIGURE 1.7: RGB vs Spectral Cube (source: [7]).

described before, are utilized since the spatial and spectral resolutions are superior to SNI systems. However, whether a spectral snapshot imager captures a small number of predetermined bands, SE systems are preferred for all the advantages that described above.

## 1.7 Color vs. Spectral Imaging

Color cameras emulate the human vision for color reproduction and are real-time devices since they record three spectral bands simultaneously. The color imaging devices that emulate human vision have three parameters. The Red, Green, and Blue (RGB) parameters. These parameters can be interpreted easily since the model familiar color perception processes. But, unfortunately the human vision has some limitations that are shared by these devices. They allocate the incoming light into the three RGB coordinates, thus missing the important spectral information. So, objects with the same RGB values might have different spectral components. This phenomenon is known as *metamerism*. So, due to this effect, the RGB imaging systems don't have the ability to distinguish material with the same color appearance but different chemical composition. In the SI systems for each pixel a series of images corresponding to many wavelengths are appeared. With these images a spectral cube can be created, which contains the spectral information of the object. The nature of the spectral cubes data is  $(x, y, \lambda)$  (spatial coordinates and wavelength, Fig. (1.7)).

## 1.8 Spectral Imaging Applications

As we mentioned before, there are numerous of application that the HSI can be used due to the importance of the spectral information.

HS Imaging is nowadays commonly used in biomedicine. For instance, DYSIS[8] is a new company, which found out a new colposcope with an advanced cervical scan. The DYSIS colposcope is a high resolution digital tool with an adjunctive map. The DYSIS spectral map is generated by a proprietary technology that measures the aceto-whitening reaction and summarises it in the form of an intuitive map. The DYSIS map is overlaid on the live image of the cervix to help with the identification of the most relevant biopsy sites.

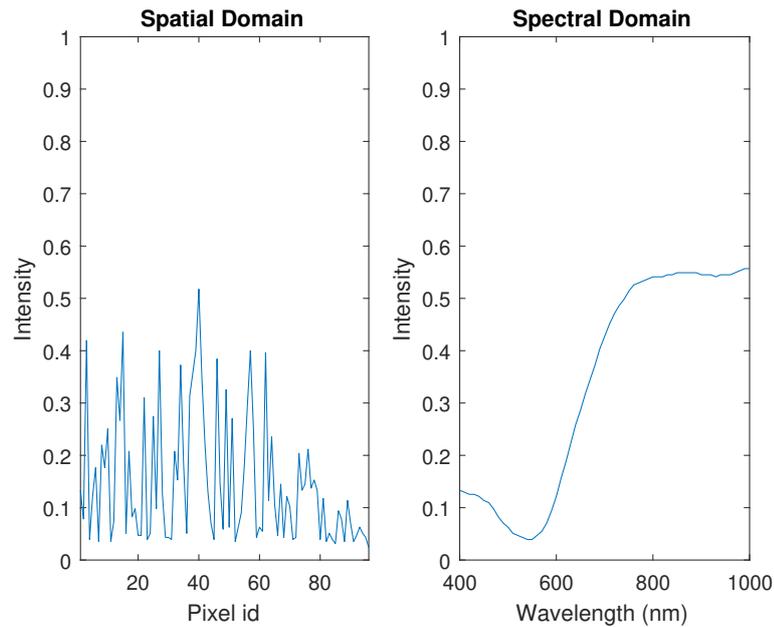


FIGURE 1.8: Spectral and spatial domain

HS Imaging is also useful to detect the chemical composition of plants, which can be used to detect the nutrient and water status of wheat in irrigated systems. On a smaller scale, NIR HSI can be used to rapidly monitor the application of pesticides to individual seeds for quality control of the optimum dose and homogeneous coverage.

In the food processing industry, HS imaging, combined with intelligent software, enables digital sorters (also called optical sorters) to identify and remove defects and foreign material (FM) that are invisible to traditional camera and laser sorters. By improving the accuracy of defect and FM removal, the food processor's objective is to enhance product quality and increase yields.

In astronomy, HS imaging is used to determine a spatially-resolved spectral image. Since a spectrum is an important diagnostic, having a spectrum for each pixel allows more science cases to be addressed. In astronomy, this technique is commonly referred to as integral field spectroscopy, and examples of this technique include FLAMES<sup>1</sup> and SINFONI<sup>2</sup> on the Very Large Telescope, but also the Advanced CCD Imaging Spectrometer on Chandra X-ray Observatory uses this technique.

From all the above examples, the importance of HSI is undisputed.

## 1.9 What is Spectral Estimation

As we mentioned before the RGB imaging devices miss important spectral information. Additionally, objects with different spectral information may have same RGB values, which means that RGB devices cannot detect the chemical difference (metamerism). Furthermore, as we studied in Section 1.6, snapshot imagers is a new trend of technology, combined with a variety of advantages (e.g Speed of acquisition). Looking at the same problem from another perspective, it is more cheaper and quicker to acquire a static number of images in specific wavelengths and estimate all the other bands, using a spectral estimation algorithm. An important question is

<sup>1</sup>“FLAMES – Fibre Large Array Multi Element Spectrograph”.

<sup>2</sup>“SINFONI – Spectrograph for integral Field Observations in the Near Infrared”.

why is better to estimate signals in spectral domain and not in spatial domain. The spatial domain is characterized from features with high spatial frequency (i.e. edges) that are those that change greatly in intensity over short image distances. These big, non-periodic intensity variation is very hard to be estimated if we also consider the fact that a natural scene is in most cases unique. There is no actual pattern or collection of images that can describe these variations accurately because of the randomness of natural scene characteristics (from objects to lighting conditions). On the other hand, spectral domain signals in the UV-VIS-NIR (400nm to 1000nm) are broad with a few peaks and valleys, as it is illustrated in figure 1.8. The absorption in this spectral range, comes from the electronic transitions (according to the selection rules for each atom) that are happening in the complex molecular structures of matter and the combination of them give these smooth characteristics to the signal. Most approaches for building a real time HSI device have the already mentioned trade-off between spectral and spatial resolution that stems by optoelectronic limitations. As the spectral band number increases the spatial resolution decrease and vice versa. The estimation of missing information (spectral or spatial) is a strategic decision for designing these systems. Taking into consideration the above we choose to preserve the spatial resolution and estimate the missing spectral information as it is quite obvious that the estimation procedure is less demanding. Our goal is find most appropriate algorithms in terms of minimizing the estimation error.

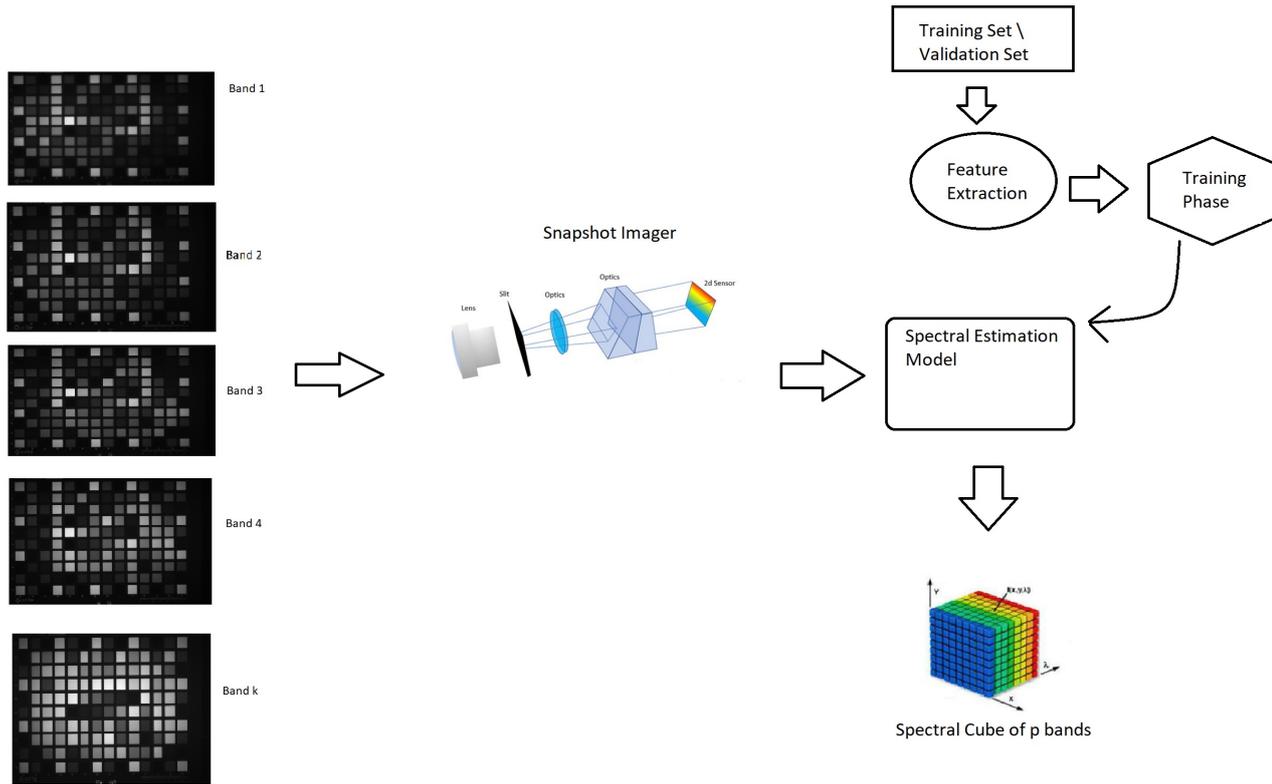


FIGURE 1.9: Spectral Acquisition Chain of a Snapshot Spectral Imager

### 1.9.1 Spectral Acquisition Chain of a Snapshot Spectral Imager

The figure 1.9 shows the procedure that take place in order to build a powerful spectral snapshot imager.

Firstly, the snapshot imager acquires a low number of spectral images, let this number be equal to  $k$ . In this case, the spectral bands that are captured from the snapshot system are  $\text{Band}_1, \text{Band}_2, \dots, \text{Band}_k$ . The identity and the number of these bands are investigated in this thesis.

Secondly, a pre-trained spectral estimation model which takes as input the  $k$  bands and returns as output the bands  $\text{Band}_1, \text{Band}_2, \dots, \text{Band}_p$  will be leveraged. In order to achieve that goal, we should utilize the most informative training set and we should extract the  $k$  more featured bands that force the model to minimize the estimation error. All the above are described analytically at this thesis. In Chapter 2 the most well-known spectral estimation methods are reported and in Chapter 3 are compared. Conclusions about the effectiveness of each are extracted. Also in Chapter 3 the training set, validation set, and detailed information about the experiments are depicted. In Chapter 4 a great amount of feature selection techniques are represented and analyzed in terms of capturing the most informative and distinctive bands.

To make it clear, at this thesis a complete spectral snapshot imager will be developed.

## Chapter 2

# Spectral Estimation - Problem Statement and Methods

### 2.1 Measures of Spectral Similarities

Before the analyzation of the spectral estimation methods, all mathematical models that measure the similarity of the original reflectance and the estimated reflectance should be taken into consideration. There are plenty of spectral similarity measures but we are going to focus on the most well-known. In the following metrics,  $\mathbf{r}_o$  is a  $p \times 1$  vector and will be the original spectrum. Furthermore,  $\mathbf{r}$  is the estimated spectrum, which has the same size as  $\mathbf{r}_o$ . Note that  $p$  is the number of the reflectance bands, which were estimated.

#### 2.1.1 Goodness of Fit Coefficient (GFC)

GFC is defined as the cosine of the angle between the recovered and original signal, thus

$$\text{GFC}(\mathbf{r}_o, \mathbf{r}) = \frac{\mathbf{r}_o^T \mathbf{r}}{\|\mathbf{r}_o\| \|\mathbf{r}\|}. \quad (2.1)$$

#### 2.1.2 Spectral Angle Mapper (SAM)

The spectral angle mapper (SAM) determines the spectral similarity between the measured and the reference spectra. The spectra are treated as vectors in a space with dimensionality equal to the number of bands, and the angle that is formed between these vectors is used as a metric of the spectral similarity. Smaller angles represent closer matches to the reference spectrum. SAM has also been used as a feature selection method for selecting an optimal subset of spectral bands. The angle ( $\theta$ ) between pixel vectors as a discrimination measure is given by the following formula:

$$\theta = \text{SAM}(\mathbf{r}_o, \mathbf{r}) = \arccos(\text{GFC}(\mathbf{r}_o, \mathbf{r})) = \arccos\left(\frac{\mathbf{r}_o^T \mathbf{r}}{\|\mathbf{r}_o\| \|\mathbf{r}\|}\right) \quad \text{where } 0 \leq \theta \leq \frac{\pi}{2}. \quad (2.2)$$

SAM is independent of lighting. Also, it is independent to the multiplications of a vector with a natural number since it only increases its length and doesn't change the angle. Hence, SAM is a non-prosthetic distance function.

#### 2.1.3 Euclidean Distance (ED)

In mathematics, the Euclidean distance or Euclidean metric is the distance between two points that one would measure with a ruler and is given by the Pythagorean

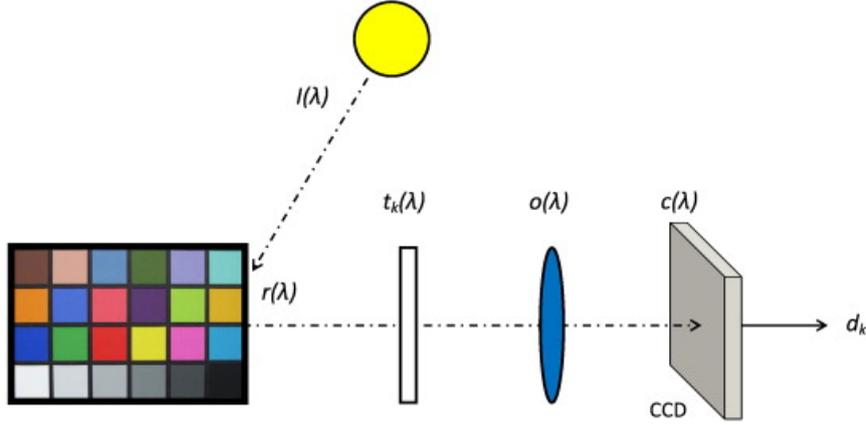


FIGURE 2.1: Synopsis of the spectral model of the acquisition process in a multispectral system (source: Alamin Mansouri [9]).

formula. By using this formula as distance, Euclidean space (or even any inner product space) becomes a metric space. The associated norm is called the Euclidean norm. Older literature refers to the metric as Pythagorean metric.

$$\text{ED}(\mathbf{r}_o, \mathbf{r}) = \|\mathbf{r}_o - \mathbf{r}\| = \sqrt{\sum_{i=1}^p (r_{o_i} - r_i)^2}. \quad (2.3)$$

#### 2.1.4 Root Mean Square Error (RMSE)

In statistics, the mean square error or MSE of an estimator is one of many ways to quantify the difference between an estimator and the true value of the quantity being estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. MSE measures the average of the square of the error. The error is the amount by which the estimator differs from the quantity that is estimated. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimation.

$$\text{RMSE}(\mathbf{r}_o, \mathbf{r}) = \frac{1}{\sqrt{p}} \text{ED}(\mathbf{r}_o, \mathbf{r}) = \sqrt{\frac{\sum_{i=1}^p (r_{o_i} - r_i)^2}{p}}. \quad (2.4)$$

## 2.2 Problem Statement

The generally used spectral model of the acquisition chain in a multispectral system is illustrated in Fig. (2.1), where  $I(\lambda)$  is the spectral radiance of the illuminant,  $r(\lambda)$  is the spectral reflectance of the surface,  $o(\lambda)$  is the spectral transmittance of the optical system,  $t_k(\lambda)$  is the spectral transmittance related to the  $k^{\text{th}}$  filter,  $c(\lambda)$  is the spectral sensitivity of the camera, and  $n_k$  represents the spectral noise for the  $k^{\text{th}}$  channel,  $k = [1, 2, \dots, K]$ . The camera output  $d_k$ , related to the channel  $k$  for a single pixel of the image is given by

$$d_k = \int_{\lambda_{\min}}^{\lambda_{\max}} I(\lambda)r(\lambda)o(\lambda)c(\lambda)t_k(\lambda) d\lambda + n_k. \quad (2.5)$$

If the noise is assumed to be removed by preprocessing, and assuming a linear opto-

electronic transfer function, we can replace  $I(\lambda)$ ,  $c(\lambda)$ ,  $o(\lambda)$ , and  $t_k(\lambda)$  by the spectral sensitivity  $s_k(\lambda)$  of the  $k$ th channel. Then, Eq. (2.5) becomes

$$d_k = \int_{\lambda_{min}}^{\lambda_{max}} s_k(\lambda)r(\lambda) d\lambda \approx d_k = \mathbf{s}_k^T \mathbf{r} \quad (2.6)$$

where  $\mathbf{s}_k = [s_k(\lambda_1), s_k(\lambda_2), \dots, s_k(\lambda_p)]^T$  is a column vector containing the spectral sensitivity of the acquisition system related to the  $k$ th channel for a single pixel,  $\mathbf{r} = [r(\lambda_1), r(\lambda_2), \dots, r(\lambda_p)]$  is a column vector of the spectral reflectances of the scene for a single pixel, and  $T$  is the transpose vector operator. Considering the system with all channels, Eq. (2.6) can be written as

$$\mathbf{d} = \mathbf{S}^T \mathbf{r} \quad (2.7)$$

where  $\mathbf{d}$  is a column vector containing all  $k$  camera outputs  $[d_1, d_2, \dots, d_k]$  for a single pixel, and  $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k]$  is the matrix containing the channels spectral sensitivities. The final goal is to recover  $\mathbf{r}$  from the camera output according to Eq. (2.7). This is obtained by finding an operator  $\mathbf{Q}$  that solves for the following equation:

$$\mathbf{r} = \mathbf{Q}\mathbf{d}. \quad (2.8)$$

Depending on how the matrix  $\mathbf{S}$  is determined, three paradigms to assess the  $\mathbf{Q}$  matrix exist.

- If  $\mathbf{S}$  is obtained by a direct physical system characterization (combination of all spectral responses of the acquisition system components, that provided by the manufacturer), then the operator  $\mathbf{Q}$  is the inverse of  $\mathbf{S}$ . However, because  $\mathbf{S}$  is not usually a square matrix, its inverse does not exist. A choice could be the pseudo-inverse  $\mathbf{Q} = \mathbf{S}^-$ , where  $\mathbf{S}^-$  is the pseudo-inverse matrix of  $\mathbf{S}$ , which is equal to  $(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T$ .

- If  $\mathbf{S}$  is obtained indirectly by matching a set of  $J$  color patches for which the theoretical reflectances are already known, we only have to capture an image of these patches with the multispectral camera. Then we produce a set of corresponding pairs  $(\mathbf{d}_j, \mathbf{r}_{oj})$  for  $j = 1, 2, \dots, J$ , where  $\mathbf{d}_j$  is a vector of dimension  $k$  containing the camera responses and  $\mathbf{r}_{oj}$  is a vector of dimension  $p$  representing the spectral reflectance of the  $j$ th patch. The reflectances  $\mathbf{r}_{oj}$  are gathered in the matrix  $\mathbf{R} = [\mathbf{r}_{o1}, \mathbf{r}_{o2}, \dots, \mathbf{r}_{oJ}]$  and the camera outputs for the  $J$  patches are gathered in the matrix  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_J]$ , in order that  $\mathbf{D} = \mathbf{S}^T \mathbf{R}$ . The operator  $\mathbf{Q}$  is straightforwardly assessed by any optimization method (such methods will be described at the next sections).

- A third paradigm for spectral reflectance estimation consists of direct interpolation of the camera outputs  $d_k$ . In this case,  $\mathbf{S}$  is a selection matrix and specifically a submatrix of the identity matrix. Nevertheless, rigorous conditions about the spectral shape of the filters as well as well-calibrated and normalized data are required for this kind of reconstruction. The reconstruction can be performed by any interpolation operator (linear, spline, etc.).

Ending up, the main subject of this problem is to find the optimal  $\mathbf{Q}$  matrix that both minimizes the error of estimation and expands the spectral dimension. There is no simple way to calculate that matrix and we should be really careful during choosing the most appropriate method. In the following section, a plenty of methods, that are based on second paradigm will be analyzed and evaluated.

## 2.3 Spectral Estimation Algorithms

### 2.3.1 Estimation of Spectral Reflectances Using Wiener Method

The purpose of the Wiener estimation (Stigell et al. [10]) is to make estimations from low-dimensional data into high-dimensional data, for example, from three-filter camera responses into reflectance spectra. The Wiener estimation is one of the conventional estimation methods which provides accurate estimations. The purpose of the Wiener model is to assess the matrix  $\mathbf{Q}$  in terms to minimize the square error (MSE) between reference  $\mathbf{r}_o$  reflection signal and the estimated  $\mathbf{r}$ . Hence, the above problem can be formulated as

$$\mathbf{Q}^{\text{Wiener}} = \arg \min_{\mathbf{Q}} \mathbb{E}[\|\mathbf{r}_o - \mathbf{r}\|^2]. \quad (2.9)$$

Let us assume that the data are centered. Then, the solution of the Eq. (2.9) is denoted by

$$\mathbf{Q}^{\text{Wiener}} = R_{\mathbf{r}_o \mathbf{d}} R_{\mathbf{d} \mathbf{d}}^{-1} = \mathbb{E}[\mathbf{r}_o \mathbf{d}^T] (\mathbb{E}[\mathbf{d} \mathbf{d}^T])^{-1} \approx \frac{1}{J^2} \mathbf{R} \mathbf{D}^T (\mathbf{D} \mathbf{D}^T)^{-1} \quad (2.10)$$

where  $\mathbb{E}$  is the expected value. In Eq. (2.10),  $R_{\mathbf{r}_o \mathbf{d}}$  is the cross-correlation matrix of vectors  $\mathbf{r}_o$  and  $\mathbf{d}$ . Matrix  $R_{\mathbf{d} \mathbf{d}}$  is the auto-correlation matrix of the vector  $\mathbf{d}$ .

The complexity of Wiener algorithm is  $O(pk)$ . The training phase costs a constant number of multiplications due to the fact that takes place only one time.

### 2.3.2 Estimation of Spectral Reflectances Using Projection on PCA

Principal component analysis (PCA) was invented in 1901 by Karl Pearson et al. [11] and is a technique extensively used for dimensionality reduction in a data set. It consists of finding an orthogonal basis composed of vectors called principal components. Each component is associated with an energy that indicates the statistical relevance of the vector in the data. Technically speaking, PCA is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data lies on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. In the field of multispectral imaging, PCA has been largely used for data compression but also in spectral reflectance reconstruction. The main difference from Wiener estimation is that we approaching  $\mathbf{r}$  by using the basis matrix  $\mathbf{B}$  and the weight vector  $\mathbf{a}$  such that each reflectance  $\mathbf{r}$  could be written as

$$\mathbf{r} = \mathbf{B} \mathbf{a}. \quad (2.11)$$

Let us assume that data are centered. Then the  $p \times p$  covariance matrix  $\mathbf{C}$  of  $\mathbf{r}_o$  is given by

$$\mathbf{C} = \mathbb{E}[\mathbf{r}_o \mathbf{r}_o^T] \approx \frac{1}{J} \sum_{i=1}^J \mathbf{r}_{o_i} \mathbf{r}_{o_i}^T = \frac{1}{J} \mathbf{R} \mathbf{R}^T. \quad (2.12)$$

It is a symmetric matrix and, therefore, it can be diagonalized. Hence,  $\mathbf{C}$  of  $\mathbf{r}_o$  can be rewritten as

$$\mathbf{C} = \mathbf{V} \mathbf{L} \mathbf{V}^T \quad (2.13)$$

where  $\mathbf{V}$  is a matrix of eigenvectors (each column is an eigenvector) and  $\mathbf{L}$  is a diagonal matrix with eigenvalues  $\lambda_i$  in the decreasing order on the diagonal. The eigenvectors are called principal axes or principal directions of the data. Projections of the data on

the principal axes are called principal components. Note that the eigenvectors could be also calculated using singular value decomposition. We define the  $p \times m$  matrix  $\mathbf{B}$  as

$$\mathbf{B} = m \text{ first column vectors of matrix } \mathbf{V}. \quad (2.14)$$

The basis functions are themselves functions of wavelength but free of constraints such as being positive or constrains to be limited to the range  $[0 - 1]$ . Eq. (2.7) can be written as

$$\mathbf{d} = \mathbf{S}^T \mathbf{B} \mathbf{a} \quad (2.15)$$

where  $\mathbf{a}$  is a  $m \times 1$  vector that holds the weights that define the particular spectrum which we are trying to reconstruct. Furthermore,  $\mathbf{S}^T$  is a  $k \times p$  matrix which it is assessed using one of the paradigms that are described in Section 2.2. An approximate solution of Eq. (2.15) is

$$\mathbf{a} = (\mathbf{S}^T \mathbf{B})^{-1} \mathbf{d}. \quad (2.16)$$

Totally,

$$\mathbf{Q}^{\text{PCA}} = \mathbf{B}(\mathbf{S}^T \mathbf{B})^{-1}. \quad (2.17)$$

The complexity of PCA spectral estimation algorithm is  $O(pk)$ . The training phase costs a constant number of multiplications due to the fact takes place only one time.

Notice that a mean function is subtracted from the training set before deriving the basis and this mean function should be added when using the basis for the reconstruction. Mansouri et al. [12] studied the performance of PCA spectral estimation, alongside with Fourier (Section 2.3.3), and Wavelets analysis (Section 2.3.4).

### 2.3.3 Estimation of Spectral Reflectances Using Fourier Analysis

To perform a Fourier analysis, we use the Fourier transform [12]. This is based on the assumption that it is possible to take any periodic function of time (wavelength) and write it as equivalent to an infinite summation of cosine and sine waves with frequencies being integer multiples of a base frequency  $f$ ,  $f = \frac{1}{T}$  (where  $T$  is the period of the function). We use Fourier transform implemented by the FFT (Fast Fourier Transform), which decomposes a signal into its phase and frequency components. We consider a given reflectance function, and decompose this function into a sum of basis function written as:

$$B(\lambda) = \cos(2\pi f\lambda + \varphi) \quad (2.18)$$

where the function  $\mathbf{B}$  is entirely determined by its frequency  $f$  and its phase  $\varphi$ . In this way, the Fourier transform is a linear operator that maps functions to other functions. In a certain way, the Fourier transform decomposes a function into a continuous spectrum of its frequency components and the inverse transform synthesizes a function from its spectrum of frequency components. Then, using the function of Eq. (2.17), the spectral estimation goal is achieved.

### 2.3.4 Estimation of Spectral Reflectances Using Wavelets Analysis

To extract information not readily available from the time (wavelength) domain representation, it is useful to project the function onto a set of basis functions [12]. The basis functions are building blocks and the transform determines how those blocks are combined to build the function. The Fourier transform provides tools for studying

global function properties that are constant throughout the function, that is, properties that are stationary. To study local or transient function characteristics, Gabor et al. [13] formulated a windowed Fourier transform which correlates a function with time-frequency atoms. But wavelets transforms often give a better signal representation using multiresolution analysis. Wavelet analysis refers to the representation of a function in terms of scaled and translated copies (known as daughter wavelets) of a finite length or fast decaying oscillating waveform (known as the mother wavelet). So, a given reflectance spectra  $\mathbf{r}$  could be decomposed into a wavelet basis  $\Psi(\frac{\lambda-\alpha}{b})$  and  $\Phi(\frac{\lambda-\alpha}{b})$ , where  $\Psi$  is the wavelets function at the scale 0 and  $\phi$  is the scaled associated function.  $b$  and  $a$  are respectively scale and translation factors. Then, using the function of Eq. (2.17), the spectral estimation goal is achieved.

### 2.3.5 Estimation of Spectral Reflectances Using Hybrid Method

Hybrid is an innovate finite-dimensional algorithm conducted by us in order to achieve better estimations with the minimum error. This algorithm is based on the combination of Wiener (Stigell et al. [10]) and PCA (Mansouri et al. [12]). Specifically, we have already known from Eq. (2.10) that Wiener Estimation is given by

$$\mathbf{r} = R_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1}\mathbf{d} \quad (2.19)$$

additionally, from Eq. (2.11)

$$\mathbf{r} = \mathbf{B}\mathbf{a}. \quad (2.20)$$

As a result, a least squares approximation of  $\mathbf{a}$  is

$$R_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1}\mathbf{d} = \mathbf{B}\mathbf{a}' \implies \mathbf{a}' = \mathbf{B}^{-}R_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1}\mathbf{d}. \quad (2.21)$$

Note that  $\mathbf{B}$  is a orthogonal matrix, therefore,  $\mathbf{B}^T\mathbf{B} = \mathbb{I}$  where  $\mathbb{I}$  is the identity matrix and  $\mathbf{B}^{-} = \mathbf{B}^T$ . Ending up, the Eq. (2.20) can be rewritten as

$$\mathbf{r} = \mathbf{B}\mathbf{B}^TR_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1}\mathbf{d}. \quad (2.22)$$

In order to achieve the contribution of the two above methods, we linearly merge Eq. (2.19) and Eq. (2.22), therefore, we can express the Hybrid model as

$$\mathbf{r}'' = wR_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1}\mathbf{d} + (1-w)\mathbf{B}\mathbf{B}^TR_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1}\mathbf{d} \quad (2.23)$$

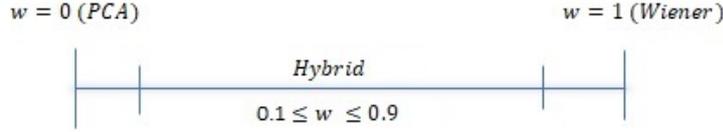
where  $w \in [0.1 \ 0.9]$  is scaling factors between Wiener and PCA model that were presented in Eq. (2.22). Hence,  $\mathbf{Q}^{\text{Hybrid}}$  it can be written as

$$\mathbf{Q}^{\text{Hybrid}} = wR_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1} + (1-w)\mathbf{B}\mathbf{B}^TR_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1} = [w\mathbb{I} + (1-w)\mathbf{B}\mathbf{B}^T]R_{\mathbf{r},\mathbf{d}}R_{\mathbf{d}\mathbf{d}}^{-1}. \quad (2.24)$$

Moreover, the Gradient Multipliers technique will be utilized to find the value of  $w$ . Hence, finding  $w$  could be described as a linear programming problem.

$$\begin{aligned} \min_w \quad & \mathbb{E} \|\mathbf{r}_o - \mathbf{r}''\|^2 \\ \text{subject to} \quad & 0.1 \leq w \leq 0.9. \end{aligned} \quad (2.25)$$

When  $w = 0$  the Hybrid model becomes PCA model and when  $w = 1$  becomes Wiener model as it is illustrated in Fig. (2.2). The first constraint ensures that the weightings of the two techniques summarize to 1, while the second constraint forces positive contribution ( $\geq 0.1$ ) of each technique. The objective functions together with the constraints can be solved using linear or quadratic programming. Note

FIGURE 2.2: Hybrid model with different values of  $w$ .

that the weightings that are assessed from the objective functions are suboptimal due to the fact that linear or quadratic programming methods may not return the optimal solution of the mathematical problem. Nevertheless, experimental results will show that this strategy can lead to spectral estimation improvement. Hybrid is supported to be more efficient and robustness due to the fact that reduces the noise levels included either in the training set or in the captures image. It keeps only the  $m$  more informative non-noisy features (Eq. (2.22)) in contrast with Wiener that needs apriori knowledge about the white noise of the camera in order to eliminate it. Thus, our spectral estimation is seems to be weighted up both by the most informative data of the training set using the PCA technique, and by second order statistics using the Wiener model. Furthermore, the number of the  $m$  basis function are critical for the preciseness of the Hybrid model. Often, a wondering about the number of the most informative principal components (i.e four, five or six) exists. This question does not have a specific answer as it depends on the spectral application. On that account, a powerful spectral estimation model should automatically adjust the difference between the possible inclusion of the important  $i^{th}$  eigenvector and the ommitance of the important  $i^{th}$  eigenvector. Hence, the weights ( $w_1, w_2$ ) are included between Wiener and new PCA model in order to calibrate the PCA contribution on the estimation result. The complexity of Hybrid algorithm is  $O(pk)$ . The training phase costs a constant number of multiplications due to the fact that takes place only one time.

### 2.3.6 Estimation of Spectral Reflectances Using Adaptive-Wiener Method

Shen et al. [14] proposed a method to reconstruct spectral reflectance by using a modified Wiener estimation method, without a prior knowledge of the spectral characteristics of the samples being imaged. Shen investigated whether the accuracy of reflectance estimation can be improved if the training samples for calculating the reflectance characteristics are adaptively selected and appropriately weighted. The novelty of the proposed method is mainly in the manner of training sample selection and autocorrelation matrix construction. Suppose that the response of the spectral camera is  $\mathbf{d}$ , and then its corresponding reflectance  $\mathbf{r}$  can be calculated according to the traditional Wiener estimation using Eq. (2.10). The training samples with reflectance  $\mathbf{r}_{oi}$  can then be selected according to their spectral similarity to  $\mathbf{r}$ . In the calculation of spectral similarity, the reflectance is normalized so that its summation is equal to 1. The reason for normalization is that statistical information of the reflectance is mainly decided by its shape, not magnitude. The spectral similarity consists of two terms, i.e., mean spectral distance and maximum spectral distance

$$\text{dist}_i = \alpha \frac{1}{p} \left\| \frac{\mathbf{r}_{oi}}{\|\mathbf{r}_{oi}\|_\infty} - \frac{\mathbf{r}}{\|\mathbf{r}\|_\infty} \right\|_1 + (1 - \alpha) \left\| \frac{\mathbf{r}_{oi}}{\|\mathbf{r}_{oi}\|_\infty} - \frac{\mathbf{r}}{\|\mathbf{r}\|_\infty} \right\|_\infty \quad (2.26)$$

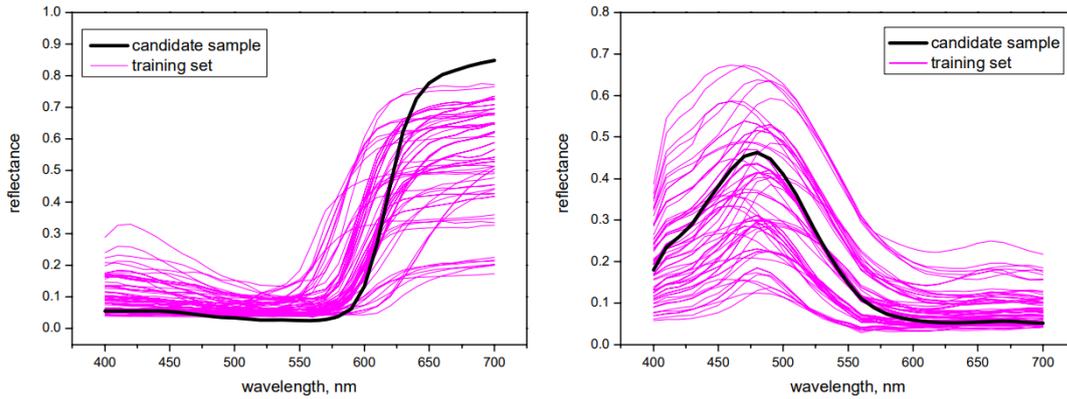


FIGURE 2.3: Two typical examples of the training sets for the given candidate samples. Channel number  $C=6$ -Adaptive Wiener (source: Hui-Liang Shen [14]).

where  $\alpha$  is a scaling factor,  $\frac{1}{p} \|x\|_1$  is the mean value of  $x$ , and the  $\|x\|_\infty$  is the maximum value of  $x$ . For two reflectances with similar shapes both of the mean and maximum distances should not be large. When  $\mathbf{r}$  is very similar to  $\mathbf{r}_{oi}$ ,  $\text{dist}_i$  is close to 0. We select  $L$  training samples according to their spectral similarities and sort the similarity measurements in ascending order, or more specifically,  $\text{dist}_1 \leq \text{dist}_2 \leq \dots \leq \text{dist}_L$ . Among the selected  $L$  training samples it is reasonable to assume that the spectral characteristic of the more similar  $\mathbf{r}_{oi}$  ( $1 \leq i \leq L$ ) should be more close to that of the candidate sample. Then the recalculation of the Wiener model is needed using the new adapted training set. It is noted that Wiener need to be calculated per pixel in a multispectral image and thus the proposed method is *computationally expensive* than the traditional Wiener estimation. This is actually the common shortcoming for almost all adaptive methods. Nevertheless, it is worthwhile to adopt the adaptive method when the spectral estimation accuracy is very important.

Figure 2.3 shows the selected training sets for the given candidates. As expected, the shapes of the training samples are similar to those of the candidate samples. The performance of the proposed adaptive Wiener estimation is investigated for the different channel numbers. Experimental results of Shen, indicate that the proposed method is significantly better than the Wiener estimation when the channel number is not large (for example, 6 or 7), while is slightly better than or close to the traditional Wiener estimation when channel number 11 or more.

As it mentioned before, adaptive methods are really computationally expensive. The complexity of Adaptive-Wiener algorithm is  $\Theta(\text{pixels} \times (2pk + J \log J))$ , where  $pk$  is a constant number describes the number of multiplications need to be done to calculate two times the vector  $\mathbf{r}$  (one for similarity check and one for spectral estimation),  $\text{pixels}$  is the number of pixels of the image, and  $J \log J$  is the complexity of sorting a  $J$ -element similarity matrix. Furthermore, in order to implement adaptive methods we need a really large training set to guarantee the adaptation of different surfaces and colors. In other words, adaptive methods can product decent spectral estimation given that there are similar reflectances in the database. Therefore, somebody must be really careful while using an adaptive method.

### 2.3.7 Estimation of Spectral Reflectances Using Adaptive Hybrid Method

Working on the same way as adaptive Wiener we can adopt the adaptation method for the Hybrid algorithm. The only difference is that instead of using Wiener estimation, we should use the Hybrid algorithm. Generally, adaptation method could be inherited by any spectral estimation method.

### 2.3.8 Estimation of Spectral Reflectances Using Multivariate Linear Regression Method

In spectral estimation, Linear Regression(LR) called also pseudo-inverse method [15]. In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis. Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the “lack of fit” in some other norm (as with least absolute deviations regression), or by minimizing a penalized version of the least squares cost function as in ridge regression (L2-norm penalty) and lasso (L1-norm penalty). Conversely, the least squares approach can be used to fit models that are not linear models. Thus, although the terms “least squares” and “linear model” are closely linked, they are not synonymous. In spectral estimation approach our goal is to find the best linear estimation (regression). So,  $\mathbf{r}$  is given by

$$\mathbf{r} = \mathbf{Q}\mathbf{d} + \mathbf{b} \implies \mathbf{r} = [\mathbf{Q} \ \mathbf{b}] \begin{bmatrix} \mathbf{d} \\ 1 \end{bmatrix} \quad (2.27)$$

where  $\mathbf{Q}$  is a  $p \times k$  matrix and  $\mathbf{b}$  is a  $p \times 1$  vector. Note that Shen et al. [15] expresses the same equation but without interception part ( $b$ ). For that reason, the experiments in [15] showed that the pseudo-inverse method has approximately the same performance as Wiener. So the criterion is to minimize the sum of squared error. To achieve that we set

$$\mathbf{Q}' = [\mathbf{Q} \ \mathbf{b}] \quad (2.28)$$

where  $\mathbf{Q}'$  is a  $p \times (k + 1)$  matrix. Additionally, we define  $\mathbf{d}^* = [\mathbf{d} \ 1]^T$  which is a  $(k + 1) \times 1$  vector. Due to this fact, the training set pairs will be  $(\mathbf{r}_{oi}, \mathbf{d}_i^*)$ . A choice

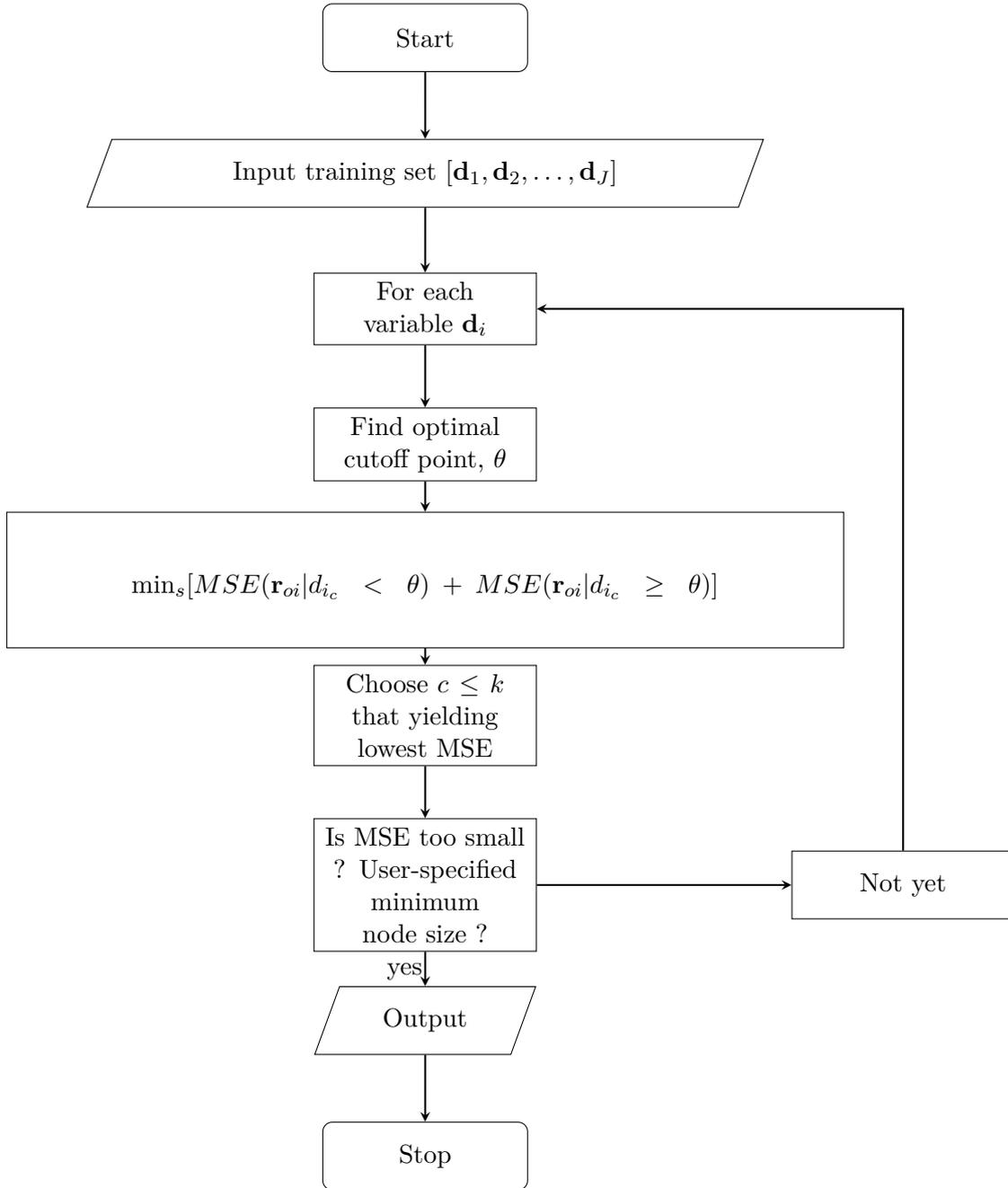


FIGURE 2.4: Flow Chart of Decision Tree Regression Algorithm.

to estimate the  $\mathbf{Q}'$  matrix is to minimize the sum of square errors. Therefore,

$$J(\mathbf{Q}') = \sum_{i=1}^J (\mathbf{r}_{oi} - \mathbf{Q}' \mathbf{d}_i^*)^2. \quad (2.29)$$

The formulation of the problem is the same as Wiener, hence the only difference is that  $\mathbf{d}$  should be replaced by  $\mathbf{d}^*$ . Then, we use Eq. (2.27) to calculate  $\mathbf{r}$ . Conclusively, LR seems to be an efficient tool to estimate reflectances due to the fact that spectral estimation is a linear problem according to Eq. (2.7).

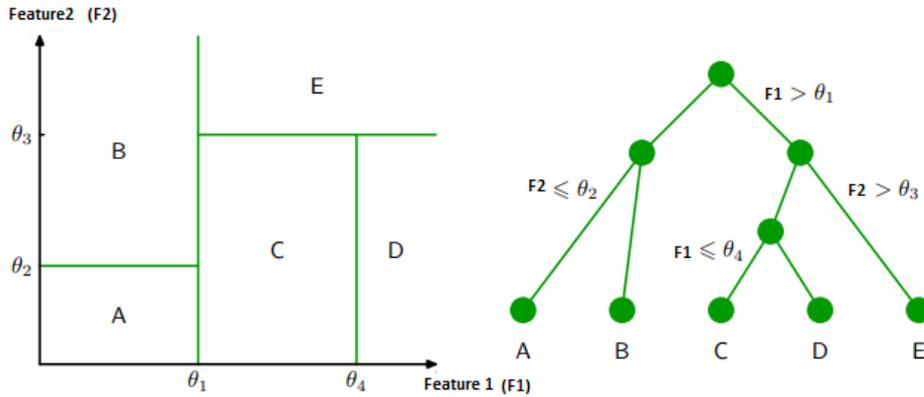


FIGURE 2.5: Example of partitions of decision trees.

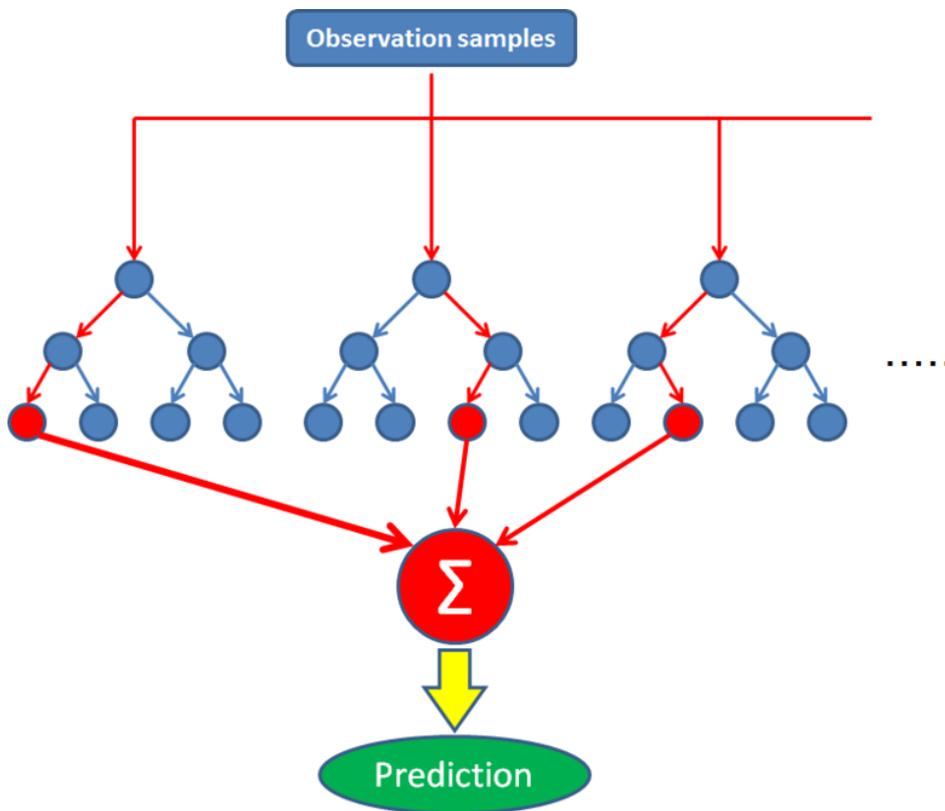


FIGURE 2.6: Example of random forest regression (source: Assignmentpoint.com) [16]).

### 2.3.9 Estimation of Spectral Reflectances Using Random Forest Regression

All regression techniques contain a single output (response) variable and one or more input (predictor) variables. The output variable is numerical. The general regression tree building methodology allows input variables to be a mixture of continuous and categorical variables. A decision tree is generated when each decision node in the tree contains a test on some input variables value. The terminal nodes of the tree contain the predicted output variable values.

A regression tree is built through a process known as binary recursive partitioning, which is an iterative process that splits the data into partitions or branches, and then

continues splitting each partition into smaller groups as the method moves up each branch (Fig. (2.5)).

Initially, all records in the training set are grouped into the same partition. The algorithm then begins allocating the data into the first two partitions or branches using every possible binary split on every field. The algorithm selects the split that minimizes the sum of the squared deviations from the mean in the two separate partitions  $(\theta_1, \theta_2)$ . This splitting rule is then applied to each of the new branches. This process continues until each node reaches a user-specified minimum node size and becomes a terminal node. If the sum of squared deviations from the mean in a node is zero, then that node is considered a terminal node even if it has not reached the minimum size. The above algorithm is described in the following flowchart (Fig. (2.4)).

Random forest is a really flexible algorithm that produces even without hyperparameter tuning a great result most of the times. It is also one of the most used algorithms, because it can be used for both classification and regression tasks. Specifically, random forest builds multiple decision-regression trees and merges them together to get a more accurate and stable prediction. We can see it from its name which is to create a forest by some way and make it random. There is a direct relationship between the number of trees in the forest and the results it can get; the larger the number of trees, the more accurate the result. There are two stages in random forest algorithm, one is random forest creation, the other is to make a prediction from the random forest trees created in the first stage. The algorithm selects randomly  $t$  features from total  $k$  features, where  $t < k$ . Among the  $t$  features, it calculates the node  $x$  using the best split point. Then, it splits the node into daughter nodes using a split criterion. Then, by repeating the steps for  $n$  times,  $n$  number of trees are created. Simplify the above, random forest build  $n$  decision-regression trees and it use them to make predictions. The final prediction in the case of random forest prediction it comes from averaging the regression values of the  $n$  trees (Fig. (2.6)).

Jason Deglint et al. [17] leverage the random forest modeling to construct the function  $\mathbf{Q}$  using a comprehensive set consisting of 10.000 reflectance spectra and their corresponding sensor measurements based on the numerical forward model for the characterized image sensor. Random forest regression models the complex relationships between reflectance spectra and sensor measurements without imposing strong assumptions about the nature of the relationship.

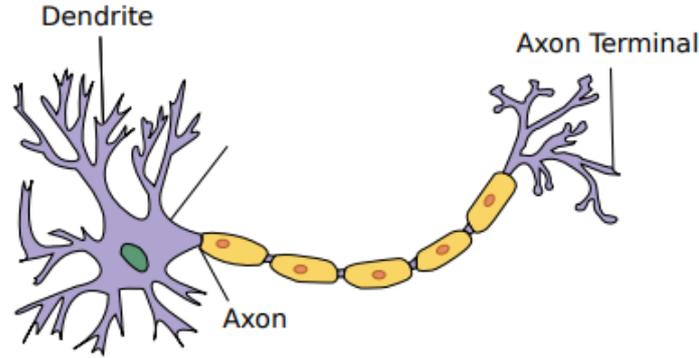


FIGURE 2.7: Model of a biological neuron (source: Wikipedia).

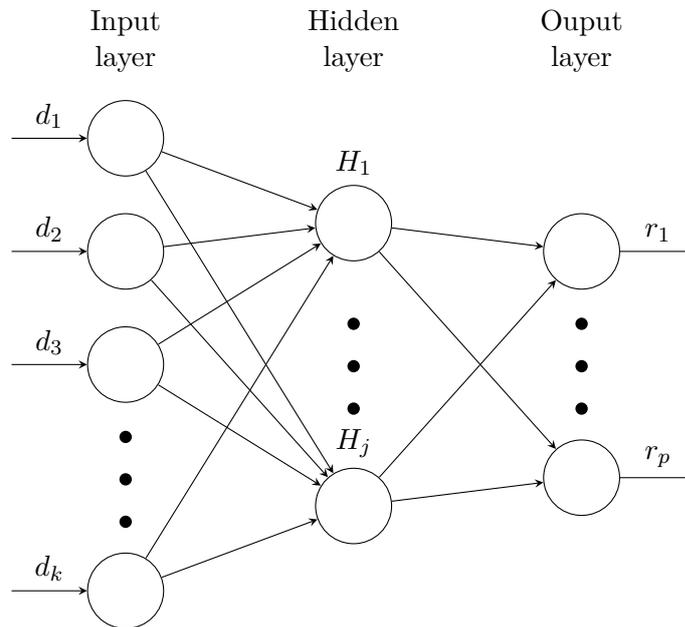


FIGURE 2.8: The Neural Network for the spectral estimation problem.

### 2.3.10 Estimation of Spectral Reflectances Using Artificial Neural Networks

A. Mansouri et al. [18] proposed Artificial Neural Networks as a way for spectral estimation. Inspired by biological nervous systems, artificial neural networks (ANNs) aim at reaching their versatility through learning. ANNs are commonly employed in artificial intelligence, machine learning and pattern recognition. There has been substantial research into how the human brains structure achieves such a high level of versatility. This research has provided some important insights, however the conclusions are far from completely explaining the complex functioning of the brain. Even though we have not been able to replicate the brain so far, the field of artificial intelligence offers very effective solutions to many problems by simulating the observations of biological research of various nervous systems. It is estimated, that the average human brain contains 86 billion neurons. Together they form a huge network. Even if we knew the detailed inner structure of the human brain, we would still not be able to simulate it with current technology because of its robustness. Our efforts are therefore rather different. We want to build a neural network with a good ratio between its size and its effectiveness.

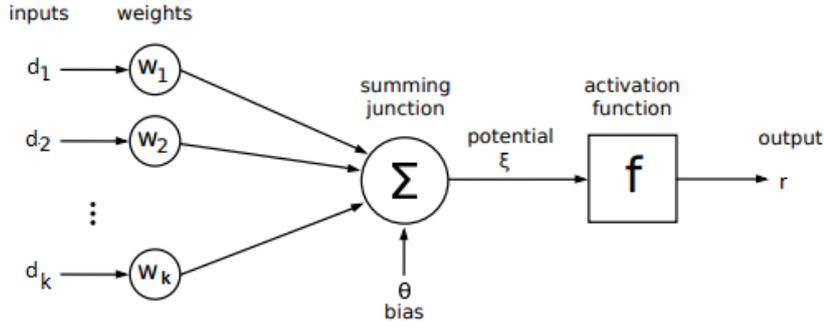


FIGURE 2.9: Model of a biological neuron (source: Wikipedia).

Generally, ANNs consist of a set of artificial neurons. Formally, an artificial neuron has  $k$  inputs represented as a vector  $\mathbf{d}$ . Inputs in an artificial neuron correspond to the dendrites in a biological neuron, while a single output of an artificial neuron corresponds to the axon in a biological neuron, which is depicted in figure 2.7. Each input  $i$ ,  $1 \leq i \leq k$ , has an assigned weight  $w_1, \dots, w_k$ . Weighted input values are combined and run through an activation function producing some output  $\mathbf{r}$ , as shown in figure 2.9. The network is formed by connecting the neuron output with the input of a different neuron. ANN is therefore effectively described as an oriented graph as shown in figure 2.8, where vertices represent the neurons, and oriented edges represent the output-input connections between them. A set of input neurons consists of the neurons which are the first ones in any complete path in the graph. All input neurons have exactly one input, and all inputs together represent an instance of the problem to be solved by the ANN. A set of output neurons consists of the neurons which are the last ones in any complete path in the graph. All output neurons have exactly one output, and all outputs together represent a possible solution to the problem to be solved by the ANN. A set of hidden neurons consists of the neurons which are not input, nor output neurons. Their number and organization into layers may vary even for the same problem, but is a key feature of the network vastly influencing its performance. An ANN works by feeding the data into the input neurons. The data flows in the direction of oriented edges and ends when the output neurons are hit. The result is interpreted from the values obtained in the output neurons. Formally, an ANN is a 6-tuple  $M = (N, C, k, p, w, t)$ , where

- $N$  is a finite non-empty set of neurons,
- $C \subseteq N \times N$  is a non-empty set of oriented edges between the neurons.
- $k \subset N$  is a non-empty set of neurons in the input layer.
- $p \subset N$  is a non-empty set of neurons in the output layer.
- $w : C \mapsto \mathbb{R}$  is a weighting function.
- $t : N \mapsto \mathbb{R}$  is a function for network bias

Let us consider neuron  $j$  with its input  $\mathbf{d}^{(j)} = (d_1^{(j)}, \dots, d_k^{(j)})$ , weights  $w_1^{(j)}, \dots, w_k^{(j)}$  and bias  $\theta^{(j)}$ . Then the potential of the neuron is computed

$$\xi^{(j)} = \sum_{i=1}^k w_i^{(j)} d_i^{(j)} + \theta^{(j)}. \quad (2.30)$$

Consider the following activation function:

$$f(\xi) = \frac{1}{1 + \exp(-\xi^{(j)})}. \quad (2.31)$$

Then the output  $r_i$  of the neuron  $j$  is computed:

$$r^{(j)} = f(\xi^{(j)}) = f\left(\sum_{i=1}^k w_i^{(j)} d_i^{(j)} + \theta^{(j)}\right). \quad (2.32)$$

Considering ANN containing  $p$  such neurons in the output layer, we obtain the output of the network as  $(r_1, \dots, r_p)$ . During the learning process, ANNs may change their weights, bias, or in some networks even the number of neurons and their setup. In contrast, our definition does not allow such modifications and  $M$  does not change in the process of learning. The Multilayer perceptron (MLP) is a feed-forward neural network consisting of multiple mutually interconnected layers of neurons. The layers are stacked one onto each other. Every neuron in one layer is connected to every neuron in the following layer. The motivation behind designing multilayer networks is to be able to solve more complex tasks. The unit step function is usually not a suitable activation function to be used with the perceptrons. Because of continuity and greater flexibility, the sigmoid function is most commonly used instead. When choosing the most suitable activation function, we want it to be differentiable at every point of its domain, and to be non-linear. Non-linearity is important, because in general we want the output to be non-linearly dependent on the given input. Perceptrons are arranged into  $k \geq 2$  layers. Let us consider a network  $M$  with  $z$  layers. The set of neurons  $C$  is split into mutually distinct subsets called layers  $L_1, \dots, L_z$ . More formally it holds  $\forall i, j : 1 \leq i, j \leq z, (L_i \neq \emptyset \wedge L_i \cap L_j \neq \emptyset) \implies i = j$ . The network layers are stacked one onto each other,  $L_1$  being the input layer,  $L_2, \dots, L_{z-1}$  being the hidden layers and  $L_z$  being the output layer. As shown in figure 2.8, the edges are all oriented in the direction from the input layer  $L_1$  towards the output layer  $L_z$ . Each neuron in layer  $L_i$  is connected to every neuron in layer  $L_{i+1}$ . In other words all neighboring layers form complete bipartite graphs.

The output of the network is computed sequentially, layer by layer. The neural network is trained using the Backpropagation algorithm. The purpose of the back-propagation algorithm is to adjust the synaptic weights of neurons, so that the network produces the desired output. The algorithm describes the process of training (also called learning). The result of this algorithm is a neural network configured to minimize the error when solving given problem. Training must be performed on labeled data ( $\mathbf{d}_i \mapsto \mathbf{r}_{oi}$ ) and therefore is supervised.

Considering the spectral estimation problem, the ANN should not be deep (Fig. (2.8)) due to the fact that the different spectra in nature are countable and their features are not many. Ending up, given that our training sets are small, we should be more than careful, during tuning the hyper-parameters of the ANN model (i.e. learning rate).

### 2.3.11 Estimation of Spectral Reflectances Using Support Vector Regression

Wei-Feng Zhang et al. [19] proposes a novel estimating approach based on the support vector regression method. Due the fact that in most of the cases SVR has a single output, the training set becomes  $(\mathbf{x}_i = [d_1, d_2, \dots, d_k] \mapsto y_i = r_{oi}, \forall i = [1, 2, \dots, J], j = [1, 2, \dots, p])$ . As a result of this transformation, the training set increases by a factor of  $p$ . For instance, if we want to recover 31 bands and have the 24 patches of the Macbeth ColorChecker chart, the composite training set contains 744 training samples. Therefore, that SVR can estimate only a band, due to this fact we need to train  $p - 1$  more SVR's. A large set increases the complexity of the

computation, but also offers an opportunity for a sparse representation in terms of SVs. In  $\epsilon$ -SV regression [20], our goal is to find a function  $f(\mathbf{x})$  that has at most  $\epsilon$  deviation from the actually obtained targets  $y_i$  for all the training data, and at the same time is as flat as possible. In other words, we do not care about errors as long as they are less than  $\epsilon$ , but will not accept any deviation larger than this. This may be important if you want to be sure not to lose more than  $\epsilon$  money when dealing with exchange rates, for instance. Firstly, we describing the case of linear functions  $f$ , taking the form

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \quad (2.33)$$

where  $\langle \cdot, \cdot \rangle$  denotes the dot product in  $x$ . Flatness in the case of (2.33) means that one seeks a small  $\mathbf{w}$ . One way to ensure this is to minimize the norm, ( i.e.  $\|\mathbf{w}\|^2$ ). We can write this problem as a convex optimization problem so that

$$\begin{aligned} \min_{\mathbf{w}} \quad & 0.5 \|\mathbf{w}\|^2 \\ \text{subject to} \quad & y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \epsilon \\ & \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \leq \epsilon. \end{aligned} \quad (2.34)$$

The tacit assumption in (2.34) was that such a function  $f$  actually exists that approximates all pairs  $(\mathbf{x}_i, y_i)$  with  $\epsilon$  precision, or in other words, that the convex optimization problem is feasible. Sometimes, however, this may not be the case, or we also may want to allow for some errors. Analogously to the “soft margin” loss function (Bennett and Mangasarian et al. [21]) which was adapted to SV machines by Cortes and Vapnik [20], one can introduce slack variables  $\xi_i, \xi_i^*$  to cope with otherwise infeasible constraints of the optimization problem (2.34). Hence, we arrive at the formulation stated in [20]

$$\begin{aligned} \min_{\mathbf{w}} \quad & 0.5 \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i + \xi_i^* \\ \text{subject to} \quad & y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \epsilon + \xi_i \\ & \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0. \end{aligned} \quad (2.35)$$

The constant  $C > 0$  determines the trade-off between the flatness of  $f$  and the amount up to which deviations larger than  $\epsilon$  are tolerated. This corresponds to dealing with a so called  $\epsilon$ -insensitive loss function  $|\xi|_\epsilon$  described by

$$|\xi|_\epsilon := \begin{cases} 0, & \text{if } |\xi| \leq \epsilon. \\ |\xi| - \epsilon, & \text{otherwise.} \end{cases} \quad (2.36)$$

Fig. (2.10) depicts the situation graphically. Only the points outside the shaded region contribute to the cost insofar, as the deviations are penalized in a linear fashion. It turns out that in most cases the optimization problem can be solved more easily in its dual formulation. Moreover, as we will see, the dual formulation provides the key for extending SV machine to nonlinear functions. Hence we will use a standard dualization method utilizing Lagrange multipliers, as described in e.g.[22]. The key idea is to construct a Lagrange function from the objective function (it will be called the primal objective function in the rest of this Section) and the corresponding constraints, by introducing a dual set of variables. It can be shown that this function has a saddle point with respect to the primal and dual variables at the solution. We proceed as follows:

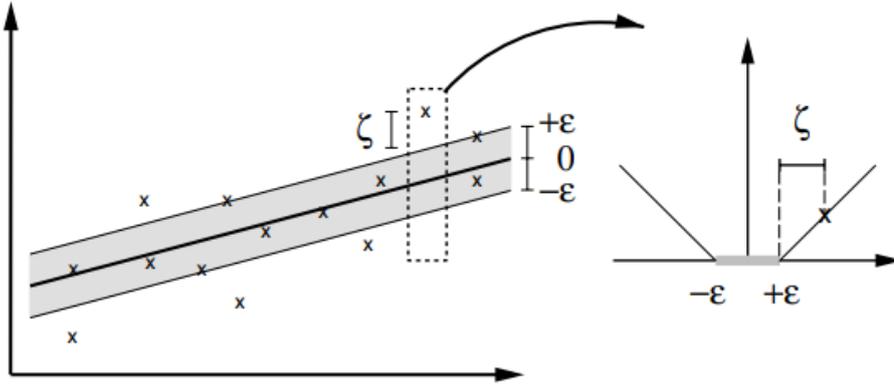


FIGURE 2.10: The soft margin loss setting for a linear SVM (source: Wei-Feng Zhang [19]).

$$\begin{aligned}
 L := & 0.5 \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \\
 & - \sum_{i=1}^l \alpha_i (\epsilon + \xi_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) - \sum_{i=1}^l \alpha_i^* (\epsilon + \xi_i^* - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b). \quad (2.37)
 \end{aligned}$$

Here  $L$  is the Lagrangian and  $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$  are Lagrange multipliers. Hence, the dual variables in (2.37) have to satisfy positivity constraints, i.e.

$$\alpha_i^{(*)}, \eta_i^{(*)} \geq 0. \quad (2.38)$$

Note that by  $\alpha_i^{(*)}$ , we refer to  $\alpha_i$  and  $\alpha_i^*$ . It follows from the saddle point condition that the partial derivatives of  $L$  with respect to the primal variables  $(\mathbf{w}, b, \xi_i, \xi_i^*)$  have to vanish for optimality

$$d_b L = \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \quad (2.39)$$

$$d_{\mathbf{w}} L = \mathbf{w} - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \mathbf{x}_i = 0 \quad (2.40)$$

$$d_{\xi_i^*} C - \alpha_i^{(*)} - \eta_i^{(*)} = 0. \quad (2.41)$$

Substituting (2.37), (2.40), and (2.41) into (2.39) yields the dual optimization problem.

$$\begin{aligned}
 & \text{maximize} \quad -0.5 \sum_{i=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad \text{and} \quad -\epsilon \sum_{i=1}^l (\alpha_i - \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \\
 & \text{subject to} \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) \mathbf{x}_i \quad \text{and} \\
 & \quad \quad \quad a_i - a_i^* \in [0, C]. \quad (2.42)
 \end{aligned}$$

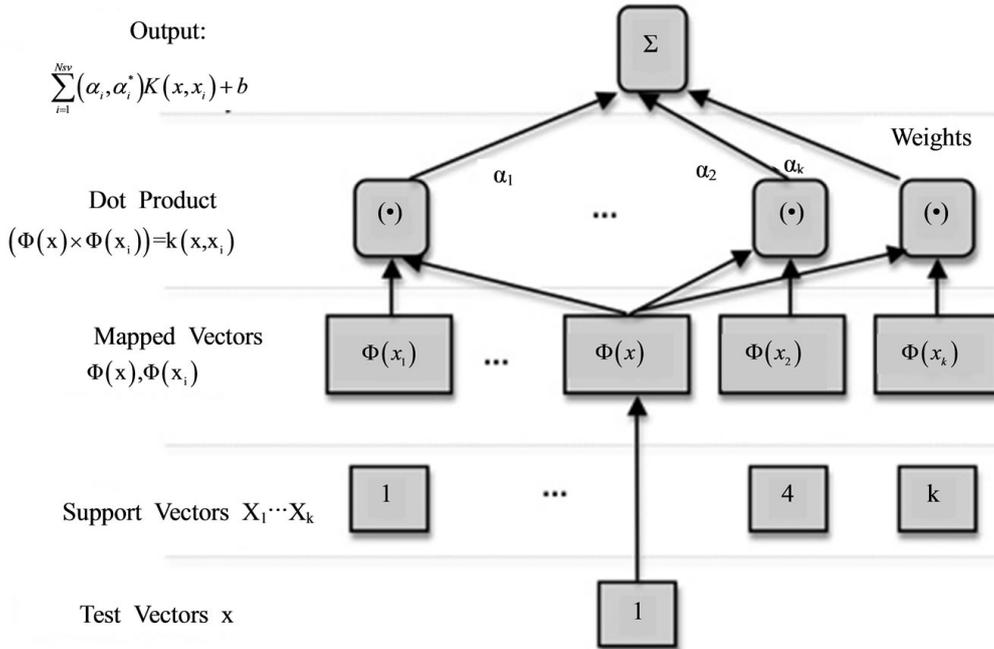


FIGURE 2.11: Architecture of a regression machine constructed by the SV algorithm (source: Wei-Feng Zhang[19]).

In deriving (2.42) we already eliminated the dual variables  $\eta_i, \eta_i^*$ . Eq. (2.40) can be rewritten as follows

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \mathbf{x}_i \text{ thus } f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{x} \rangle + b. \quad (2.43)$$

Clearly this approach is not feasible and we have to find a computationally cheaper way. *Kernel tricks* is a more efficient way to compute  $f(\mathbf{x})$ , as the latter depends on dot products between patterns  $\mathbf{x}_i$ . That allow us to restate the SV optimization.

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \phi(\mathbf{x}_i) \text{ thus } f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) k \langle \mathbf{x}_i, \mathbf{x} \rangle + b \quad (2.44)$$

where  $k \langle \mathbf{x}_i, \mathbf{x} \rangle := \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle$ .

The difference to the linear case is that  $\mathbf{w}$  is no longer given explicitly. Also note that in the nonlinear setting, the optimization problem corresponds to finding the flattest function in feature space, not in input space. The entire algorithm is given in Fig. (2.11).

## Chapter 3

# Spectral Estimation - Experiments

### 3.1 Training Sets and Target Surfaces



FIGURE 3.1: Macbeth Color Checker (source: phononet.com [23]).

#### 3.1.1 Macbeth Color Checker

Specifically designed to meet the needs of digital photographers, the Digital Macbeth ColorChecker SG target includes the highest quality color reference standards available. Each of the 140 patches was chosen for its location in color space to expand the color gamut, so we can create profiles that capture the full capabilities of any camera and spectral scanner.

The ColorChecker Digital SG target includes the colors from the standard ColorChecker target, many of which represent natural objects, such as human skin tone, foliage and blue sky. Additional skin-tone reference colors deliver greater accuracy and consistency over a wide variety of skin tones, and gray scale steps provide accurate control of camera balance to maintain a neutral aspect, regardless of light source. Macbeth ColorChecker SG can also be used to create a white balance with your digital camera to guarantee precise, uniform, neutral white under any lighting condition.

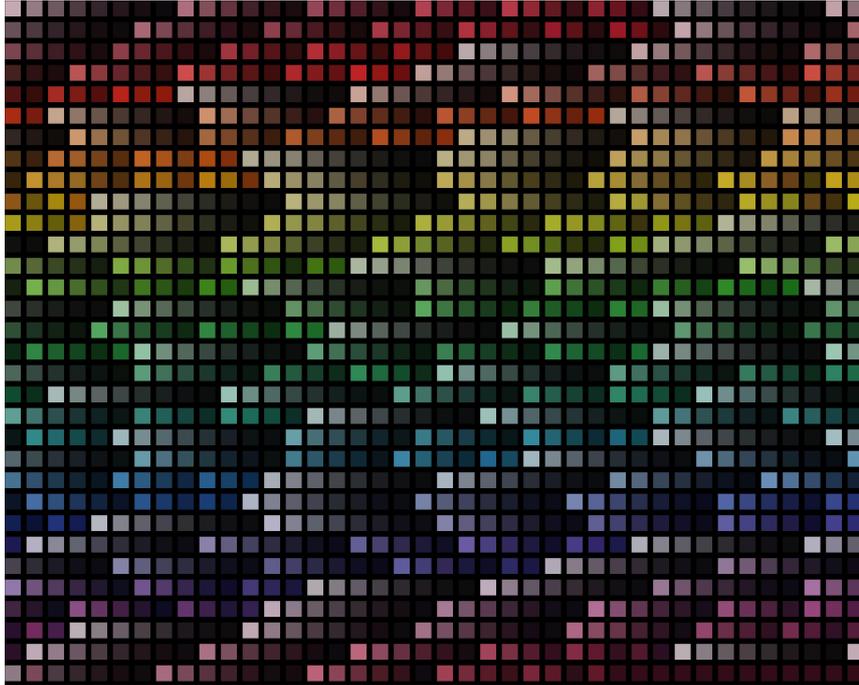


FIGURE 3.2: Munsell Mate Color Spectra Data.

### 3.1.2 Munsell Mate Color Spectra Data

This data set was obtained from the University of Joensuu Color Group [24]. This database contains in ASCII format the reflectance values of 1269 spectra in one file. Every single entry includes spectral information from 421 channels, between 380 nm and 800 nm with one nanometer step. In order to visualize the color which each spectra represents, we designed an artificial pallet of colors like Macbeth ColorChecker. Figure 3.2 depicts the new artificial Munsell ColorChecker and will be used as training and validation set in our experiments.

## 3.2 Hardware of the experiments

MuSES-9 HS is a powerful Hyperspectral imager for Spectroscopy and Spectrometry in a wide spectral range from 360 nm (UV) to 1000 nm (Near Infrared). Spectral resolution ranges from 7 nm (for visible or NIR) to 18 nm (for both visible and NIR). More than 110 spectral bands can be acquired in less than 1 min. From the collected images, millions of spectra can be obtained and per-point spectral and color analysis can be performed through a graphical user interface. MuSES-9 HS (Fig. (3.3)), technical features are:

- Wavelength tuning range: 360-1000nm (extendable)
- Bandwidth (FWHM): 7-18nm
- Sensor-type: CMOS
- Resolution: 6.4 Mega Pixels  $3096 \times 2080$
- Exposure Range:  $32\mu s - 1000s$
- Shutter speed (integration time):  $1/100.000 s - 2.0 s$
- Sensor Size  $7.4mm \times 5mm$
- Sensor Size  $7.4mm \times 5mm$
- Imaging Spectroscopy: Automatic spectral scanning, sequential spectral image storing (spectral cube) and calculation of a full spectrum per image pixel

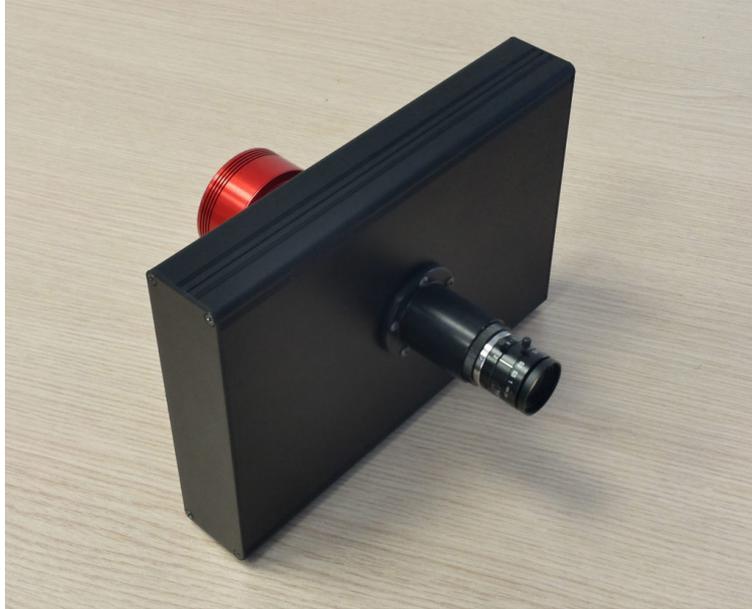


FIGURE 3.3: MUSES-9 HS (source: spectricon.com [25]).

- Imaging modes: Spectral imaging and spectroscopy in both reflectance and fluorescence modes

The main difference we have between the MuSES-9 HS and the simple Spectrometer is that MuSES-9 has an acquisition integral of 10nm in contrast with Spectrometer, which has lower than 1nm. This affects our data in the following way: if we want to use the data from MuSES-9 we must first make a smoothing process. Also when using Spectrometer we need to fit the data in order to remove all kinds of noise the Spectrometer might have. Finally, one of the main reasons we wanted to use the MuSES-9 HS camera was that we would have all the devices needed in one camera. Therefore, we didn't have to make different calibration for each device or have a fault in aligning the second device in the same spot as the first. So, to keep it sort, MuSES-9 HS makes our work easier.

### 3.3 Evaluation of each Spectral Estimation Method

#### 3.3.1 Evaluation Thresholds of Spectral Estimation Algorithms

In this Section we will discuss the accuracy of each method by taking into advantage the error measurements. Since, we are interested purely in spectra, thresholds that can indicate the quality of the estimation should be determined. Thus, the following thresholds according to the RMSE between the reference and the estimated spectra were used.

RMSE value	Quality of Estimation
$RMSE \leq 1.32\%$	Perfect Estimation
$RMSE \leq 2.10\%$	Very Good Estimation
$RMSE \leq 2.81\%$	Good Estimation
$RMSE > 2.81\%$	Bad Estimation

TABLE 3.1: Evaluation Thresholds of Spectral Estimation Algorithms.



FIGURE 3.4: Macbeth Color Checker (96 isolated patches).

We utilized RMSE, as it focuses in the magnitude difference between the estimated and the reference spectra and not in the angle difference (GFC, SAM). There is no doubt that those thresholds are really strict by cause of the perfect estimation is given only whether the root mean square difference is 1.32%, something that is undeniably “a perfect fit”.

### 3.3.2 Scoring the algorithms

In order to turn out which is the most appropriate method for spectral estimation, all the methods are compared and scored according to their performance. Each algorithm take points whether it gives more perfect estimations than good or bad estimations. On that account, we marking the methods using the criteria that are depicted in the following table.

Percentage on Score.	Number of Patches that achieved:
40%	Perfect Estimation (PE)
35%	Very Good Estimation (VGE)
24%	Good Estimation (GE)
1%	Bad Estimation (BE)

TABLE 3.2: Scoring Algorithms.

$$\text{Score}_i = 0.4 \text{ PE} + 0.35 \text{ VGE} + 0.24 \text{ GE} + 0.01 \text{ BE}. \quad (3.1)$$

Using the 3.1 formula we calculate the score for each  $i^{\text{th}}$  spectral estimation method and we normalize to the highest score. Finally, all the methods have Total score  $\in [0 \ 1]$ .

### 3.3.3 Information about the experiments

Firstly, we acquired the spectral cube of Macbeth using MuSES-HS in wavelengths between 400nm and 730nm (Visible spectrum). Then, we divided our experiments into two parts. In both parts, we validated each algorithm according to its estimation

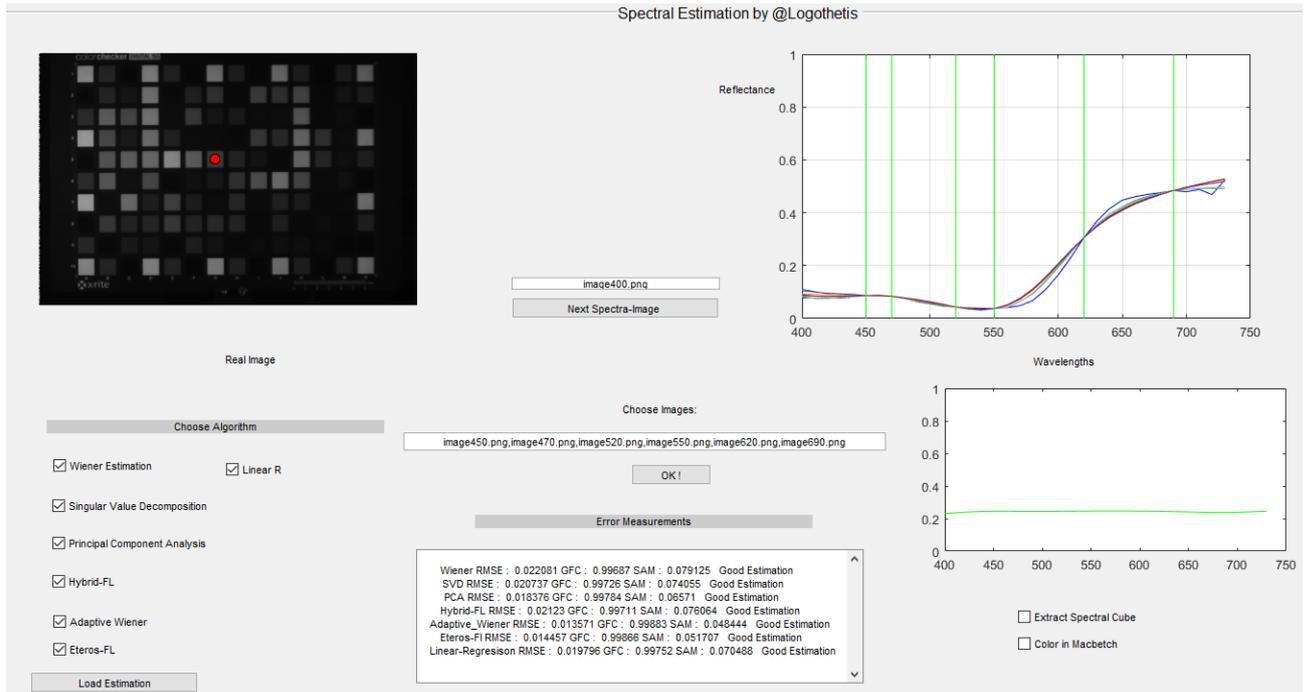


FIGURE 3.5: Software for experiments using Matlab Tools.

accuracy. In the first part, Munsell’s patches were the training set and Macbeth Color Checker the validation set. In the second part, we swapped the training and the validation set and we analyzed again the estimation preciseness. It is important to mention that we didn’t use all the Macbeth’s patches for training, but we isolated only 96; the most informative and the non-dark ones that are included into the red grid (Fig. (3.4)). These 96 patches were contained in the training set in the first part of experiments and were the target surface in the second part of experiments.

It is crucial for our research not only to evaluate the efficiency of the dimensionality expansion methods, but to extend our work in a real new spectral snapshot imager. On that account, we decided to keep the spatial resolution in high levels and capture only six spectral bands. Due to the fact that each spectrum has a few features, peaks and valleys, six spectral bands are almost enough to estimate the missing spectrum. After a market research, we found that the bands [450 nm, 460 nm, 470 nm, 480 nm, 520 nm, 530nm, 550nm, 580 nm, 610 nm, 620 nm, 690nm] are already available in the market and being sold by optical manufacturers. In our experiments we used [460 nm, 480 nm, 530nm, 580 nm, 610 nm, 690nm] as sampled spectral bands.

To compare the preciseness of each spectral estimation algorithm, we used both the scoring metric in 3.1 and a appropriate software specific for our experiments. On that account, two machines were implemented by us; the first using Matlab tools (Fig. (3.5)) and the second using Qt cross-platform SDK (Fig. (3.6)). Both software tools are available online<sup>1</sup> and provide:

- Load and read a spectral cube
- Extract a spectral cube using a specific estimation method
- Choosing dynamically the training set and the surface target
- Real time visualization of spectral estimation
- Real time visualization of error metrics

<sup>1</sup>[www.github.com/flogothesis](http://www.github.com/flogothesis)

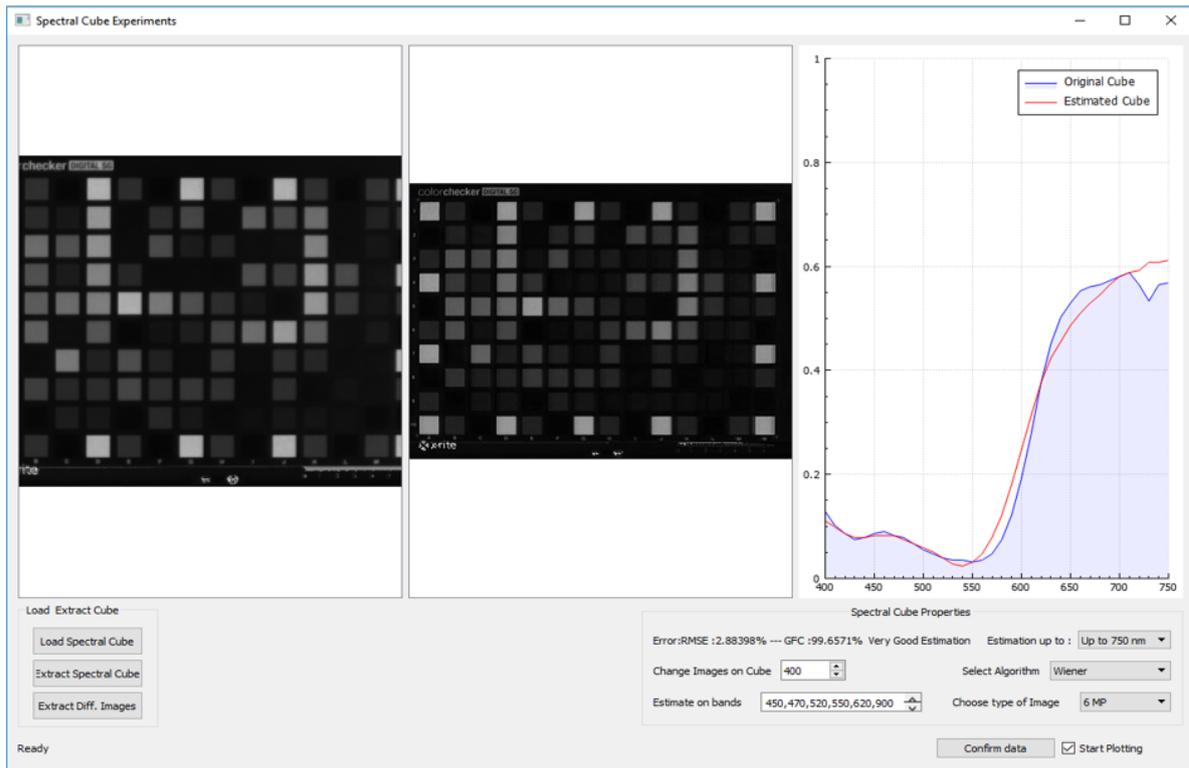


FIGURE 3.6: Software for experiments using Qt cross-platform SDK.

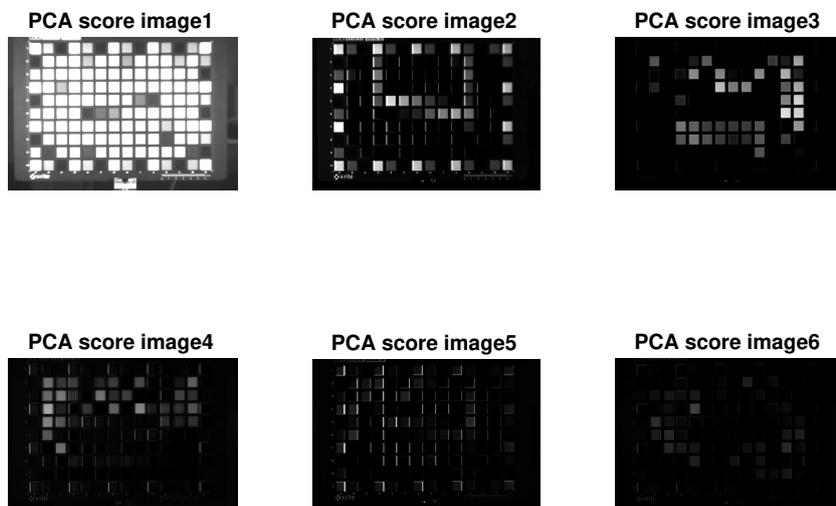


FIGURE 3.7: PCA score images of Macbeth.

- Real time spectral estimation on a specific pixel by hovering the mouse
- Dynamic selecting of sampled bands
- Dynamic selection of spectral estimation algorithm
- Zoom In-Zoom Out over the spectral cube to select a region of interest
- Compare the estimation accuracy of different spectral estimation methods.

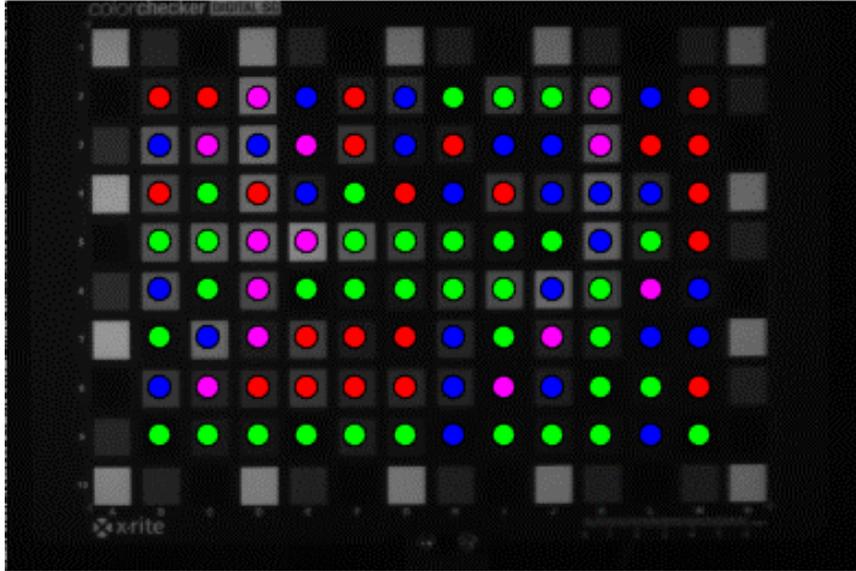


FIGURE 3.8: Example of a Color-Map.

Moreover, machine learning algorithms (Neural Nets, SVR, Random Forest) were implemented using the Tensor-Flow Tool. Additionally, the hyper-parameters of the models were investigated and are presented in this paragraph. Wiener method has no hyper parameters that need to be tuned. On the other hand, PCA's accuracy based on the number of the  $m$  eigenvectors. To find the most informative principal components, we extracted all the PCA scores as images of Macbeth, which are illustrated in Fig. (3.7). The first five score images are the most informative. The  $six^{th}$  pca score image is almost black due to the fact that there is no useful information in the  $six^{th}$  eigenvector. Hence,  $m$  was set equal to 5. Furthermore,  $\mathbf{Q}^{PCA}$  matrix was obtained using the training set pairs, thus  $\mathbf{Q} = \mathbf{R}\mathbf{D}^{-}$ . The above considerations are also extended to Hybrid method, therefore,  $m$  is equal to 5 eigenvectors and  $w_1, w_2$  were calculated equal to 0.5 respectively. Adaptive Wiener and Adaptive Hybrid achieve better estimation by setting  $a$  equal to 0.5 and  $L$  equal to 31 for small training sets (i.e Macbeth), but 100 for larger training sets (i.e Munsell). Random Forest Regression needs on average 8000 different decision trees and maximum depth equal to 90-100. Last but not least, the Neural Network that we built it is not deep. It is turned out that deep networks are not the optimal solution for the spectral estimation problem, due to the fact that spectrum has a low amount of features. On the contrary, a small neural network with one hidden layer that contains 10-15 neurons is a really efficient predictor. Furthermore, we used the Momentum optimizer and the Elu as activation function. SVR model was fitted with the following parameters; penalty  $C$  equal to 1.0,  $epsilon$  equal to 0.1. Radial basis function is the kernel type that was selected, which is based on the squared Euclidean distance between the two feature vectors  $k(x, x') = \exp(-\gamma \|x - x'\|^2)$ , where  $gamma$  is equal to  $\frac{1}{\text{number of features}}$ . In this study, the number of features are equal to the number of captured bands. Ending up, all the data had been regularized before the training phase.



FIGURE 3.9: First-Series of Experiments. Training/Validation set contains only the patches into the green grid.

### 3.4 First Series of the Experiments

As it mentioned before, the experiments at this section were carried out using the Munsell database as training set and Macbeth SG Color Checker as validation set. The bands that we used for sampling are [460 nm, 480 nm, 530nm, 580 nm, 610 nm, 690nm] and we reconstructed the entire spectrum from 400nm - 730nm, totally 36 bands. For each method a table will be demonstrated, which includes the error metrics between estimated and reference spectrum. Moreover, a color-map that includes round points is depicted, in order to visualize the quality of estimation on each patch. For instance, a sample color map is depicted in Fig. (3.8). The relation between the color of the points and the quality of estimation is given in Table (3.3).

Color	Quality of Estimation
<i>Green</i>	Perfect Estimation
<i>Blue</i>	Very Good Estimation
<i>Magenta</i>	Good Estimation
<i>Red</i>	Bad Estimation

TABLE 3.3: Color-Map and the quality of estimation.

Furthermore, a bar chart will be depicted, in order to count the number of patches that were estimated perfectly, goodly, and badly etch. More than that, we illustrate the estimated spectra of the patches 2M-9M that are in the green grid (Fig. (3.9)).

### 3.4.1 Wiener Method

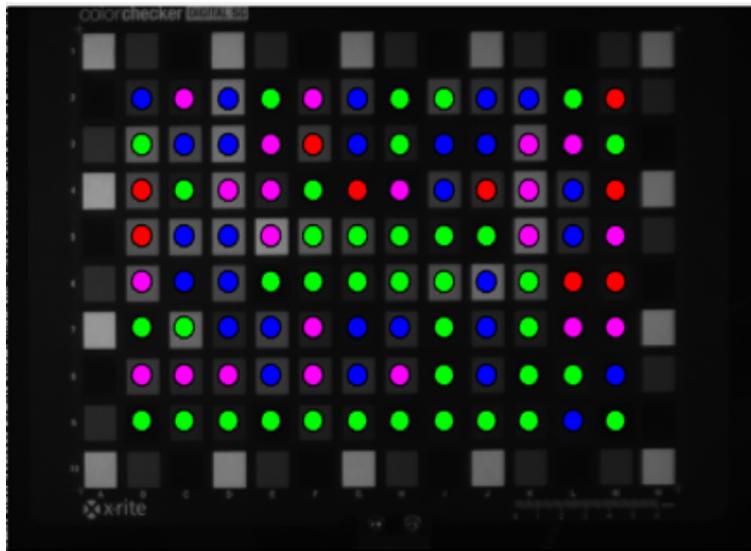


FIGURE 3.10: First Experiments - Wiener Color Map.

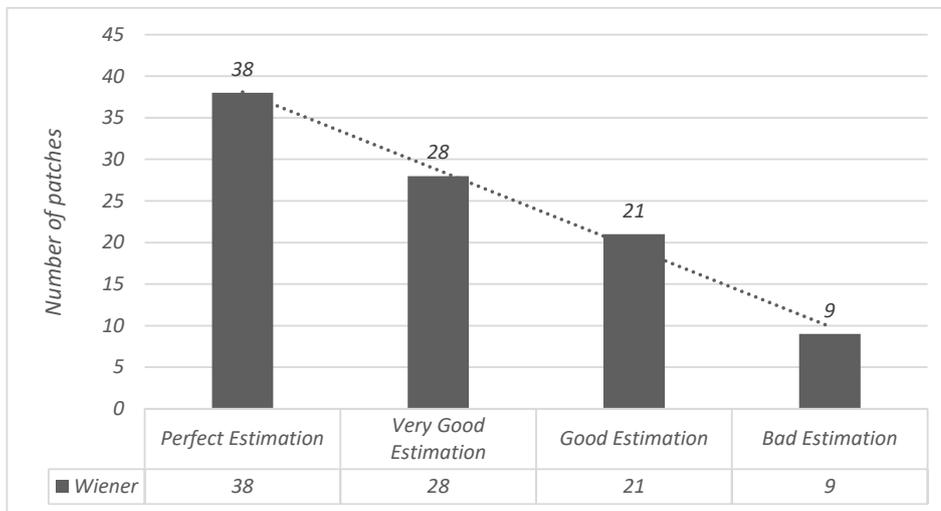


FIGURE 3.11: First Experiments - Wiener Performance.

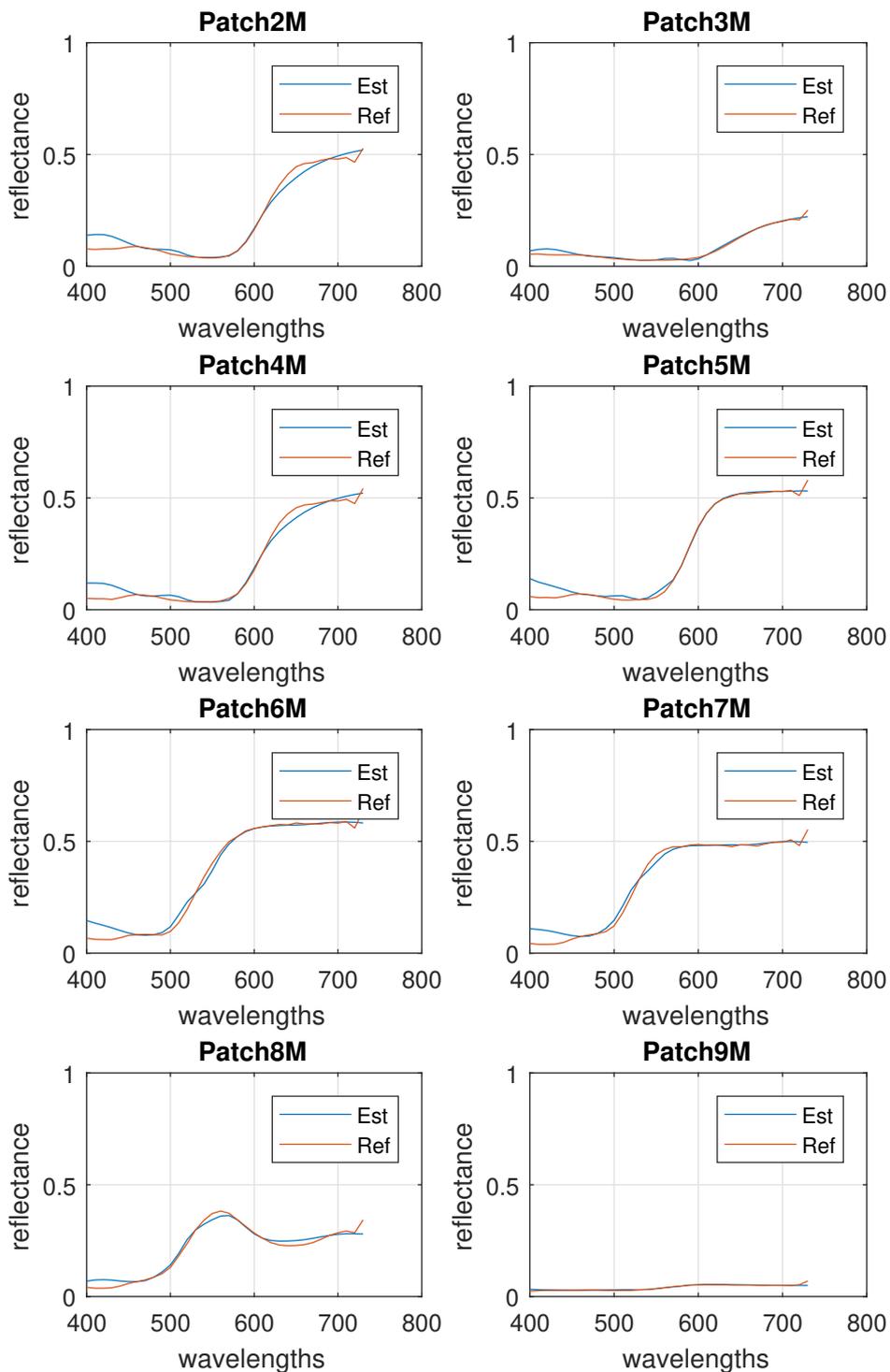


FIGURE 3.12: First Experiments - Wiener, Patch 2M-Patch 8M.

### 3.4.2 PCA Method

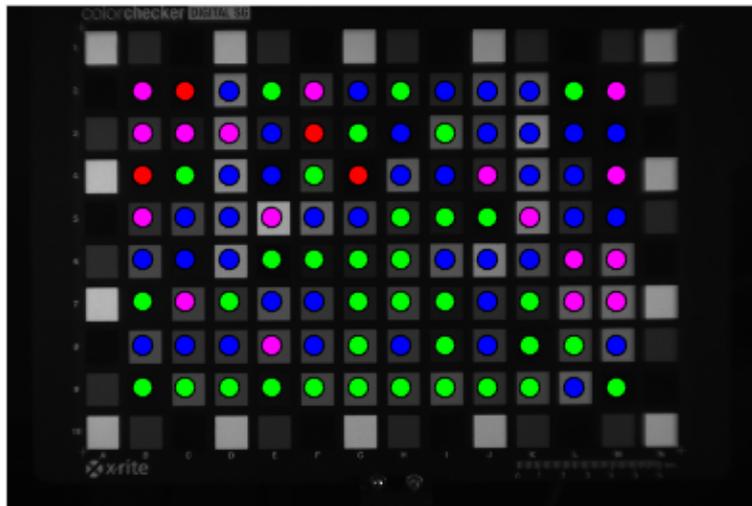


FIGURE 3.13: First Experiments - PCA Color Map.

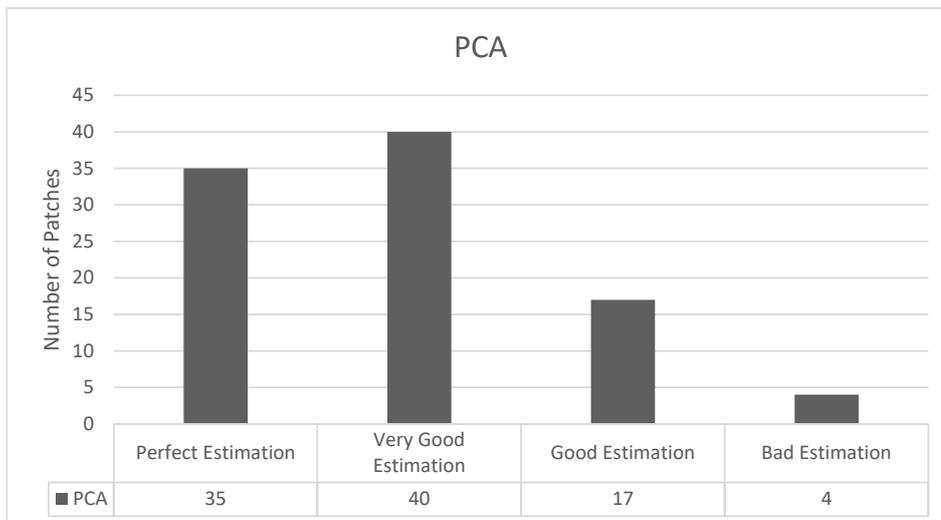


FIGURE 3.14: First Experiments - PCA Performance.

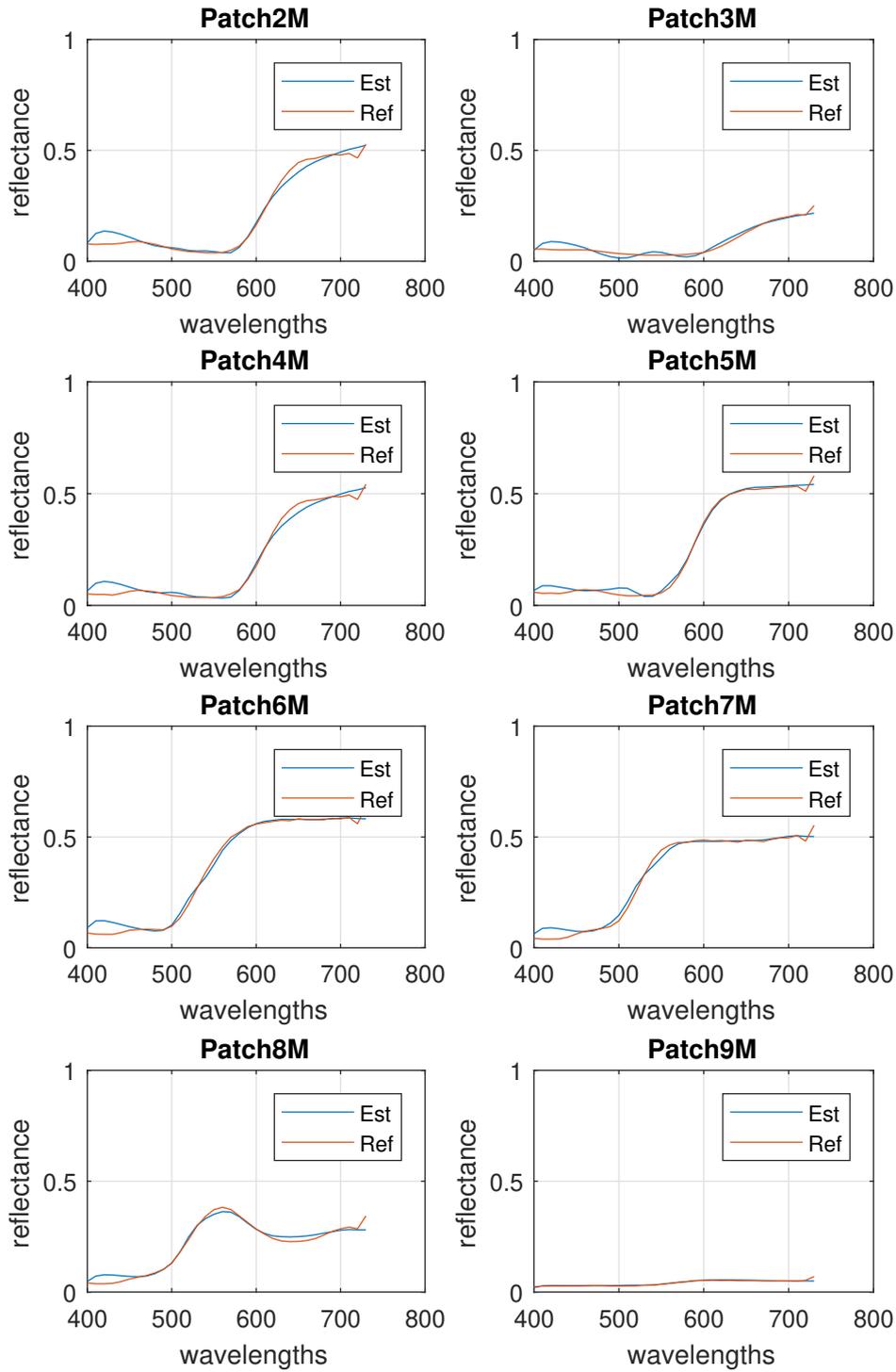


FIGURE 3.15: First Experiments - PCA, Patch 2M-Patch 8M.

### 3.4.3 Hybrid Method

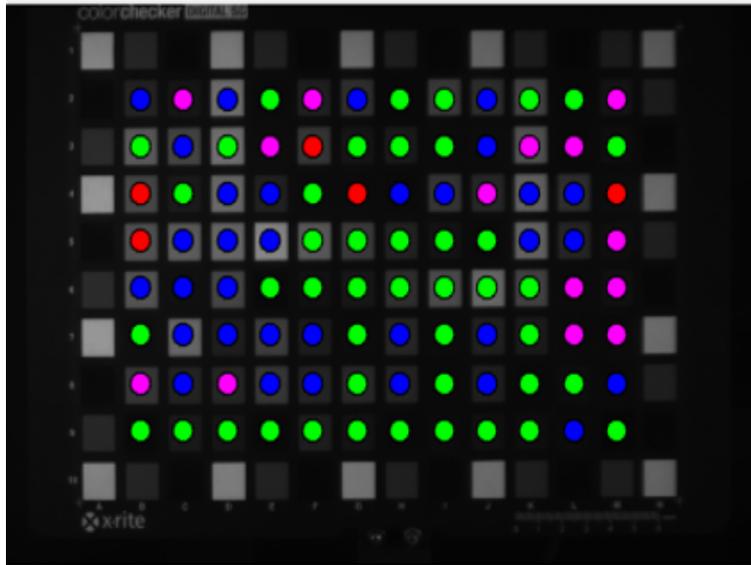


FIGURE 3.16: First Experiments - Hybrid Color Map.

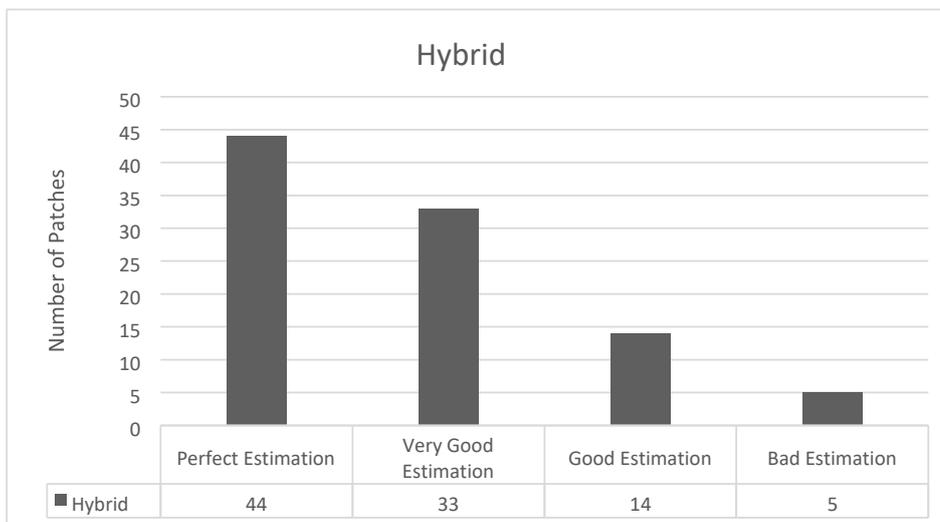


FIGURE 3.17: First Experiments - Hybrid Performance.

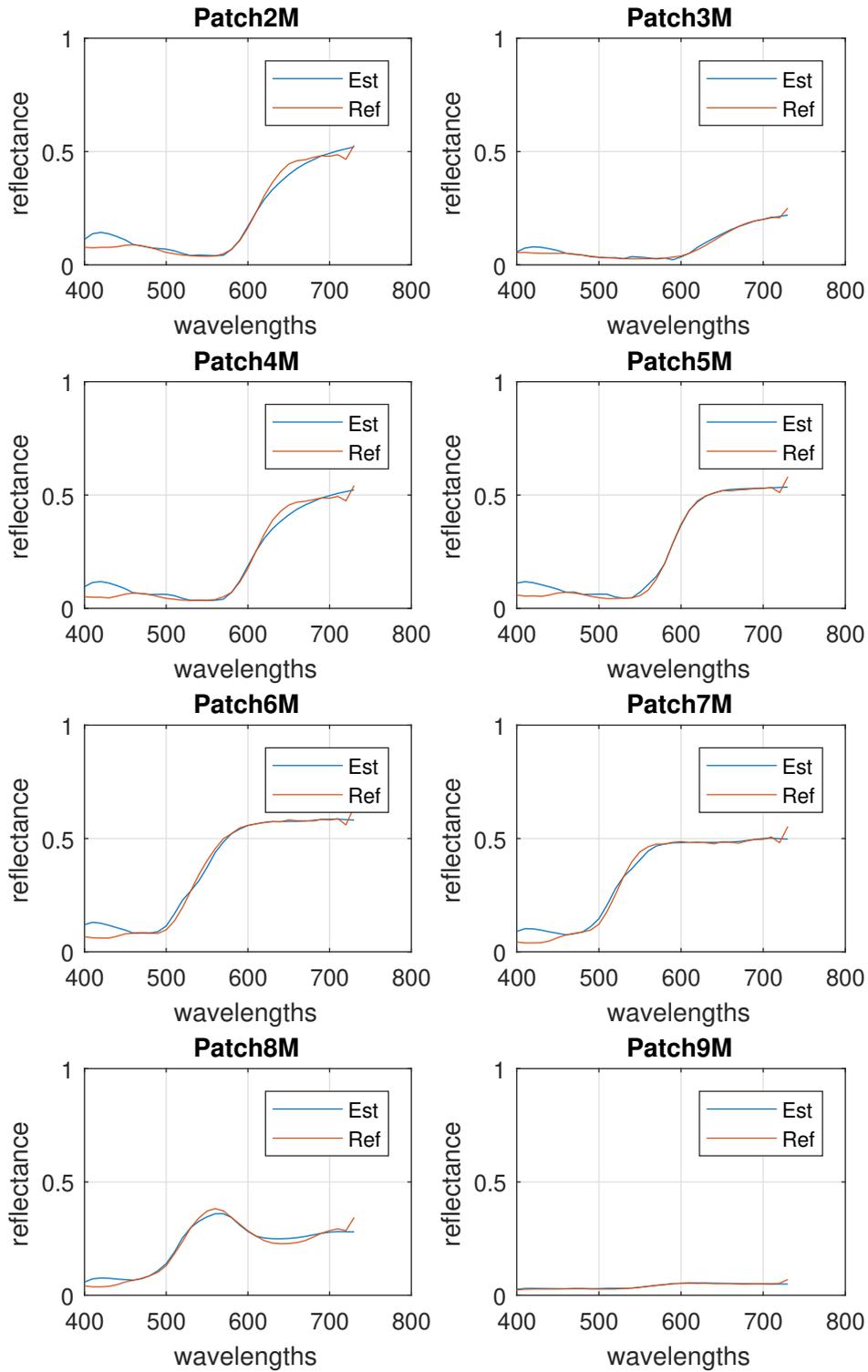


FIGURE 3.18: First Experiments - Hybrid, Patch 2M-Patch 8M.

### 3.4.4 Adaptive Wiener Method

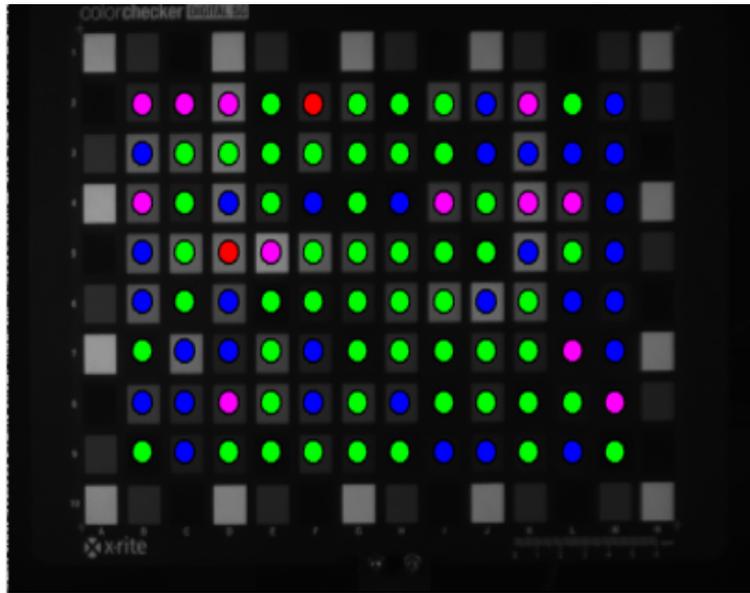


FIGURE 3.19: First Experiments - Adaptive Wiener Color Map.

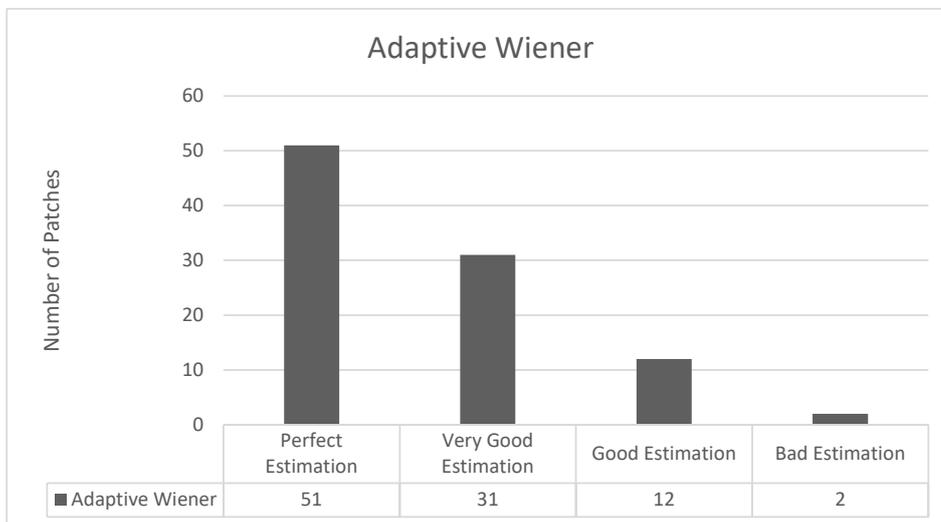


FIGURE 3.20: First Experiments - Adaptive Wiener Performance.

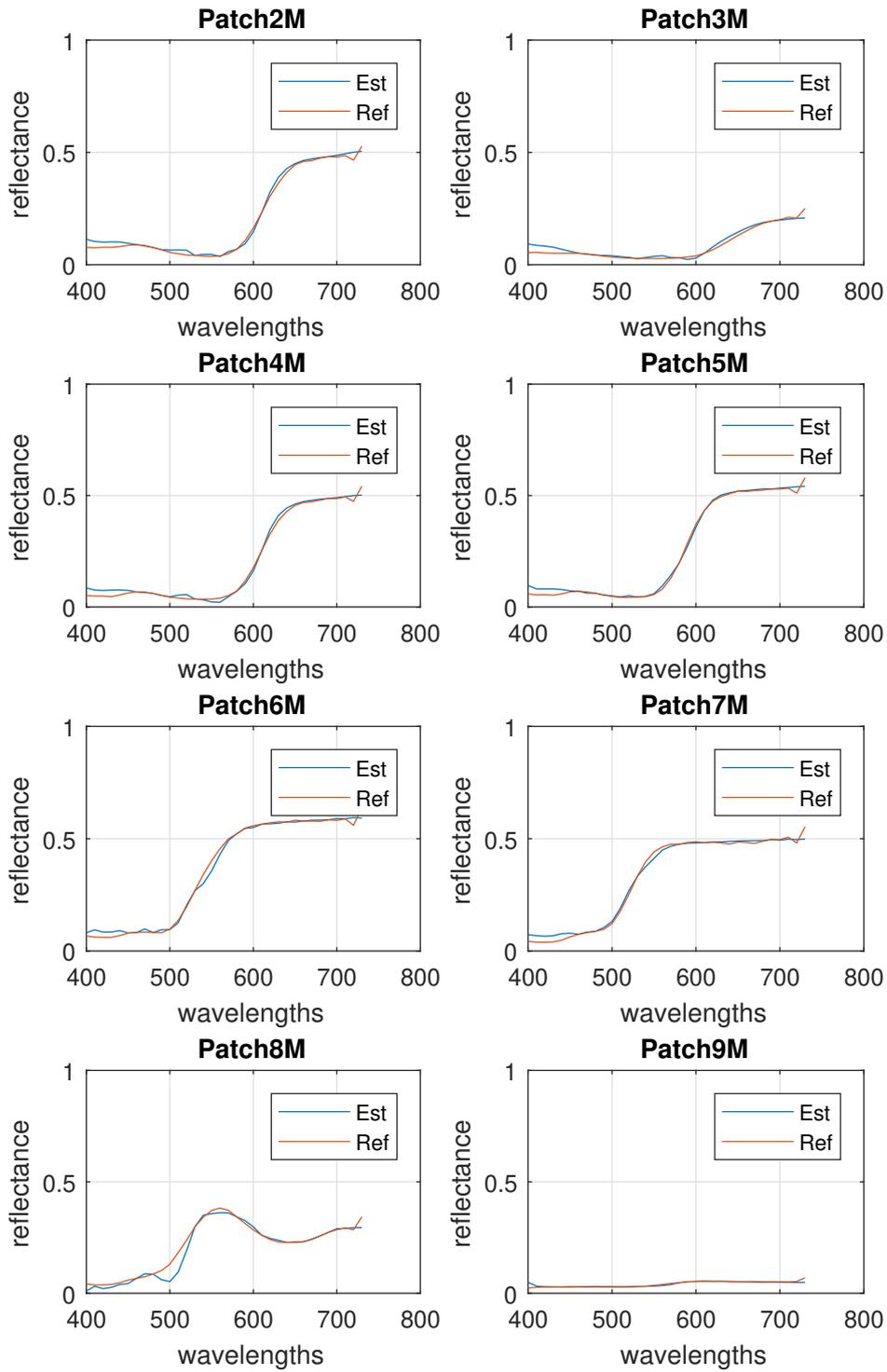


FIGURE 3.21: First Experiments - Adaptive Wiener, Patch 2M-Patch 8M.

### 3.4.5 Adaptive Hybrid Method

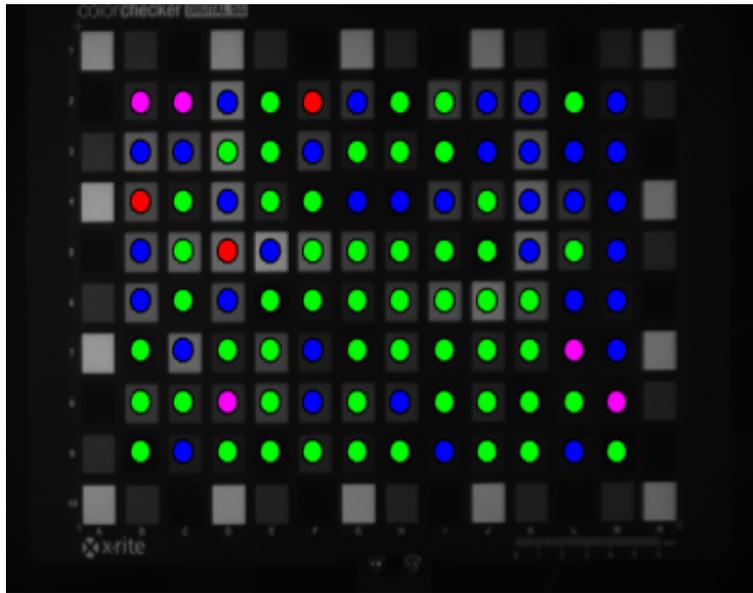


FIGURE 3.22: First Experiments - Adaptive Hybrid Color Map.

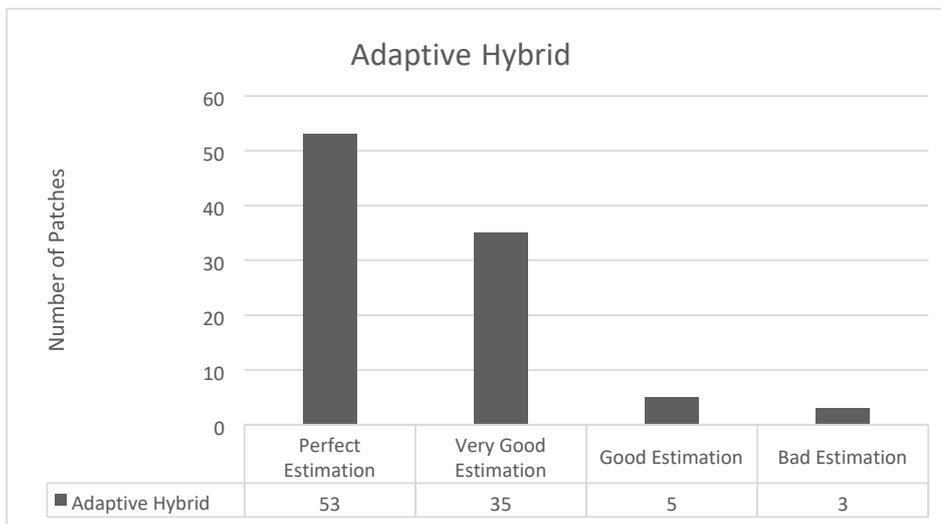


FIGURE 3.23: First Experiments - Adaptive Hybrid Performance.

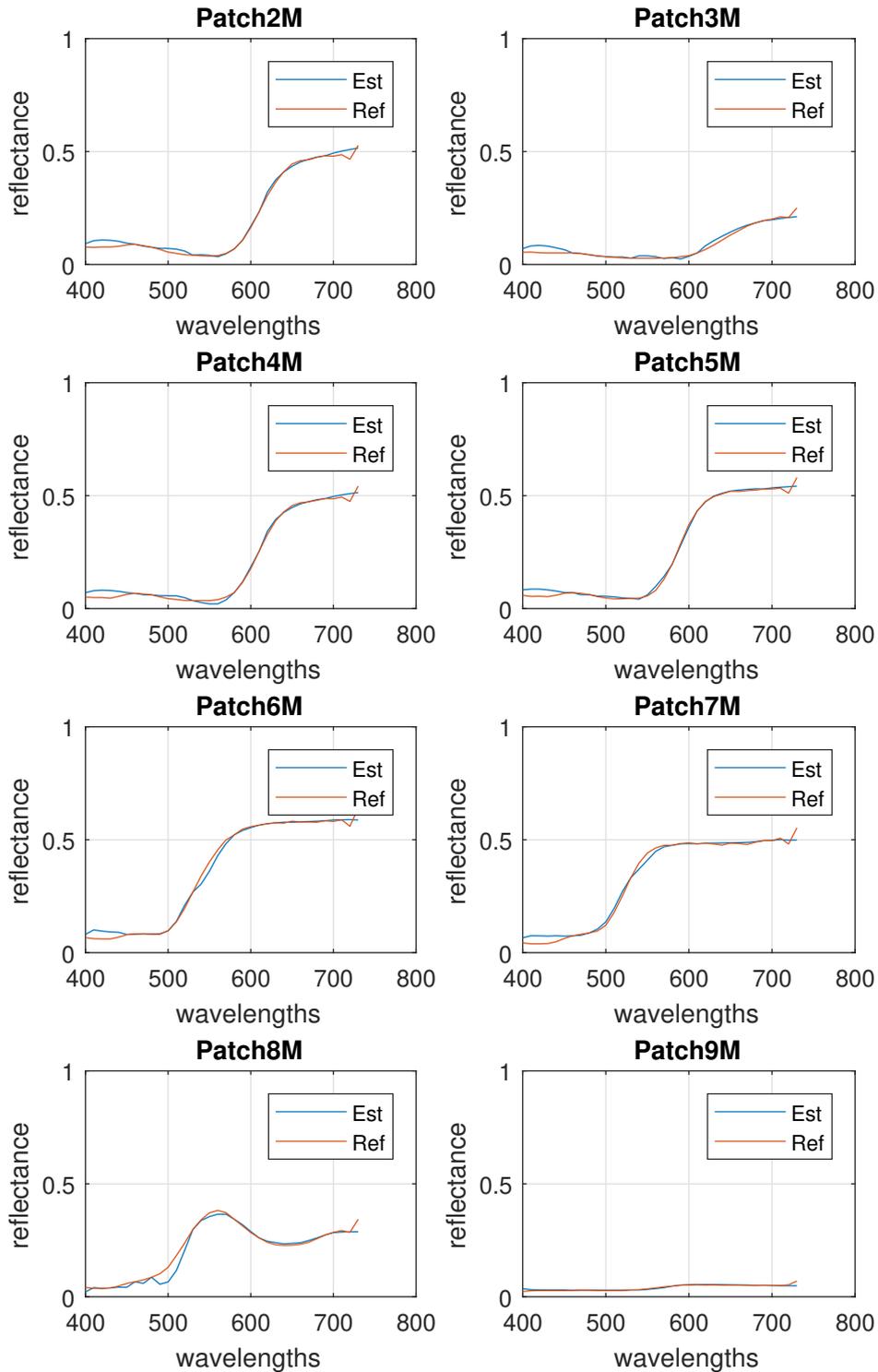


FIGURE 3.24: First Experiments - Adaptive Hybrid, Patch 2M-Patch 8M.

### 3.4.6 Linear Regression Method

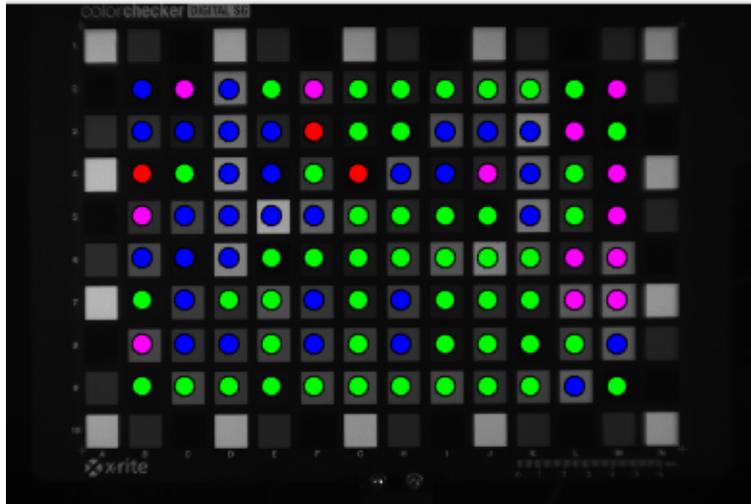


FIGURE 3.25: First Experiments - Linear Regression Color Map.



FIGURE 3.26: First Experiments - Linear Regression Performance.

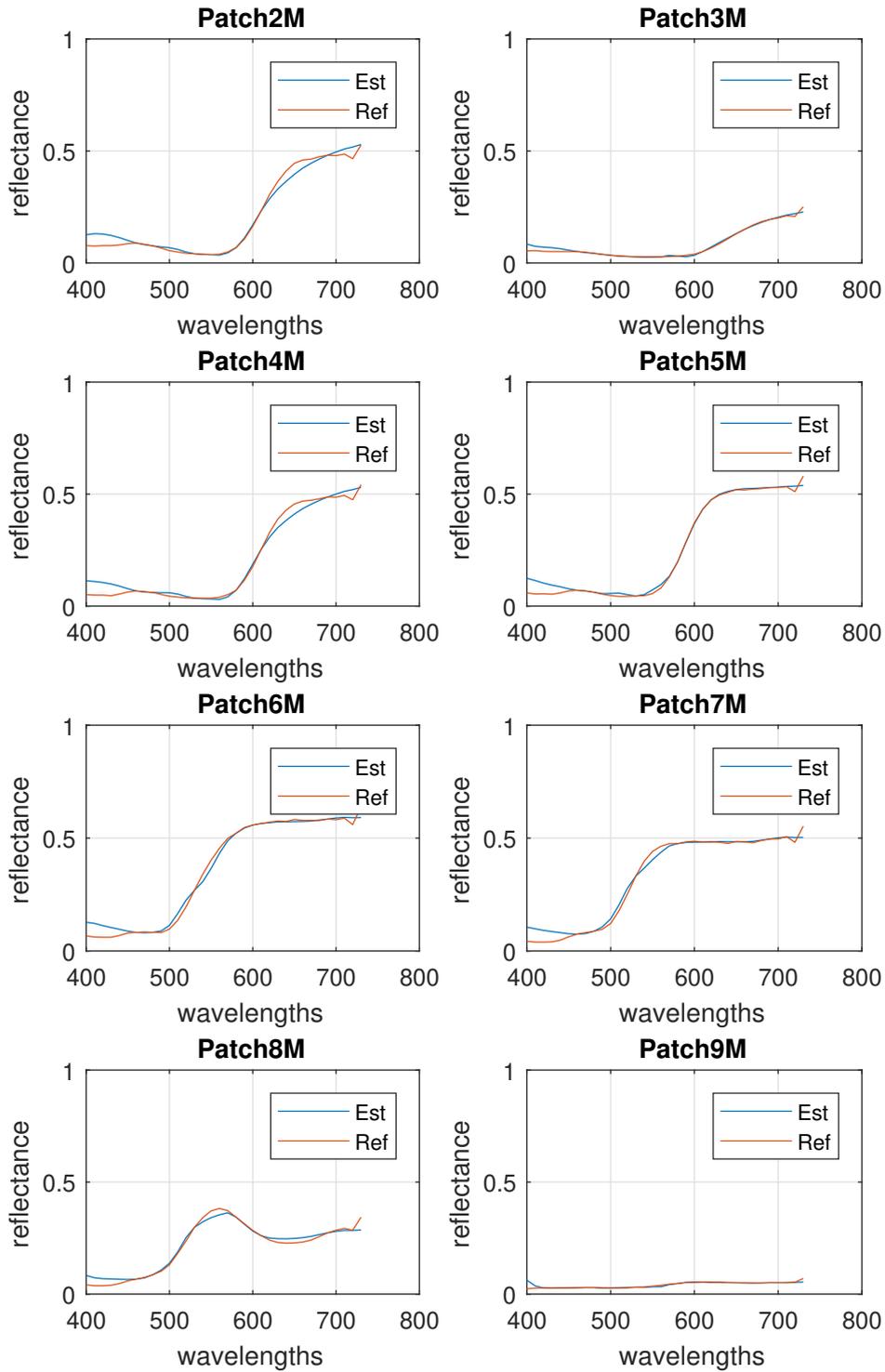


FIGURE 3.27: First Experiments - Linear Regression, Patch 2M-Patch 8M.

### 3.4.7 Random Forest Method

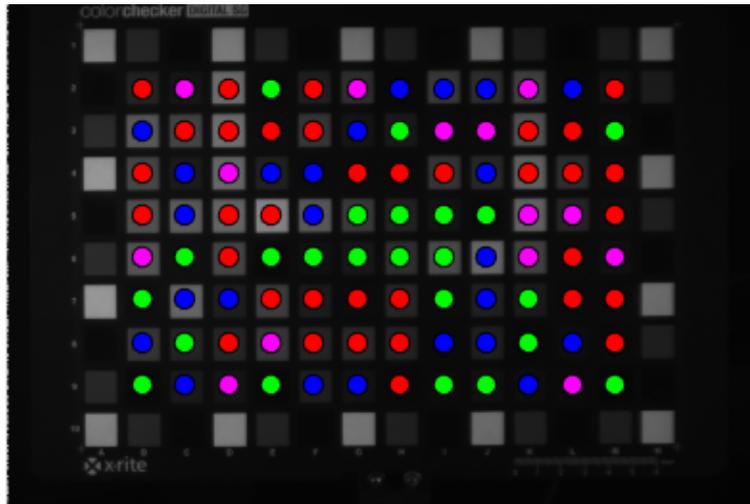


FIGURE 3.28: First Experiments - Random Forest Color Map.

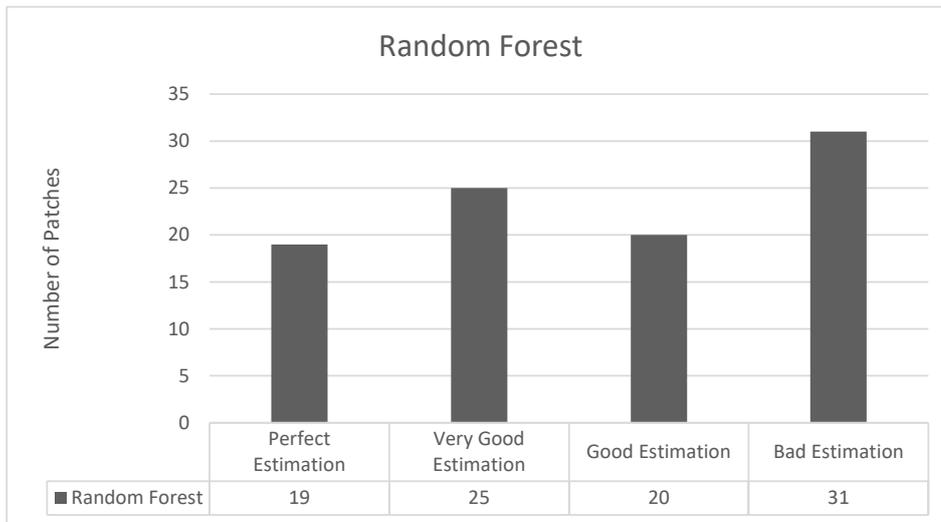


FIGURE 3.29: First Experiments - Random Forest Performance.

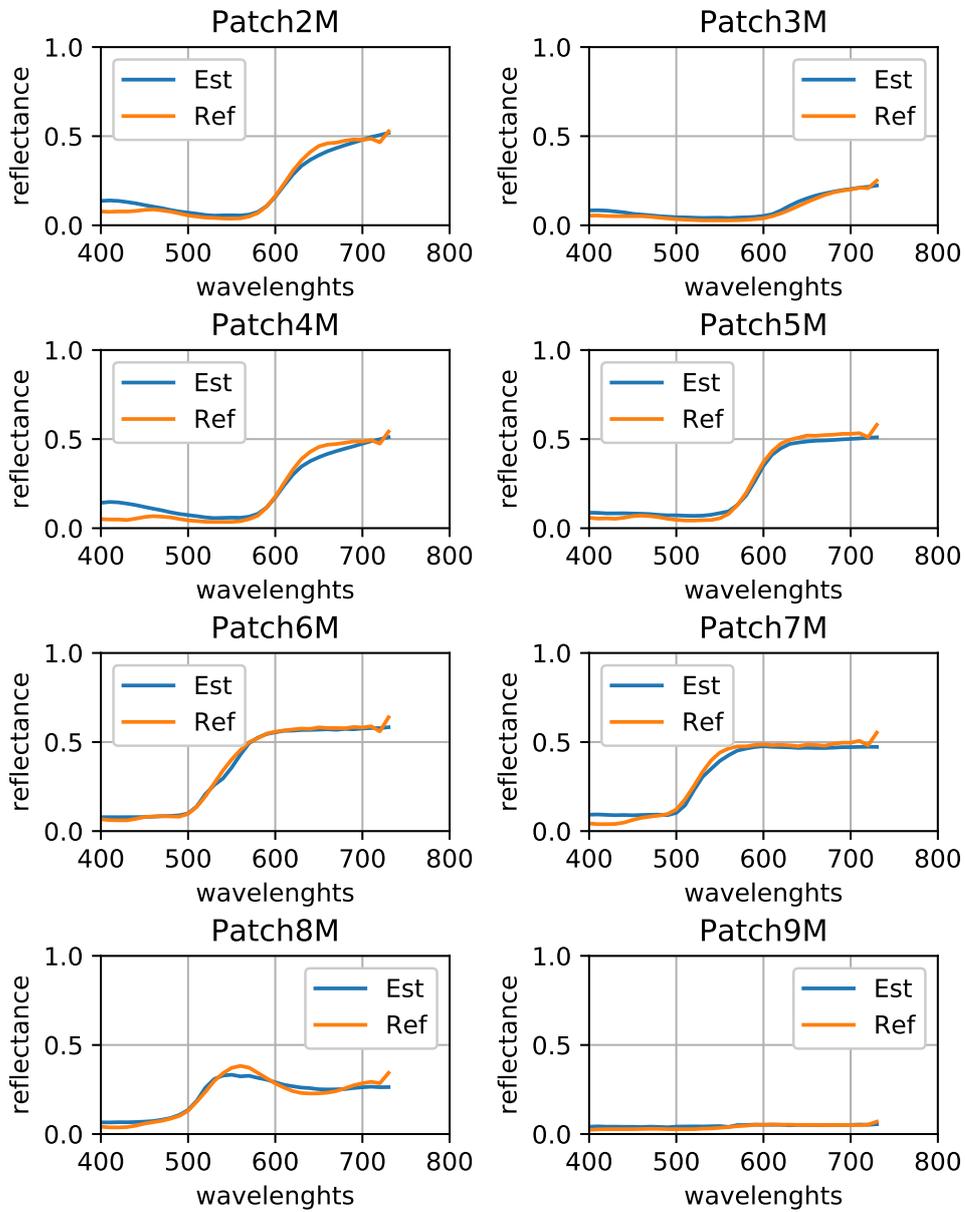


FIGURE 3.30: First Experiments - Random Forest, Patch 2M-Patch 8M.

### 3.4.8 Neural Network Method

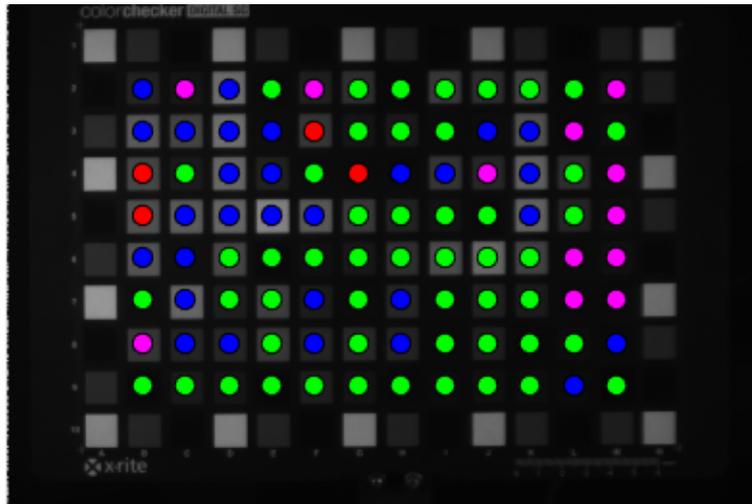


FIGURE 3.31: First Experiments - Neural Network Color Map.

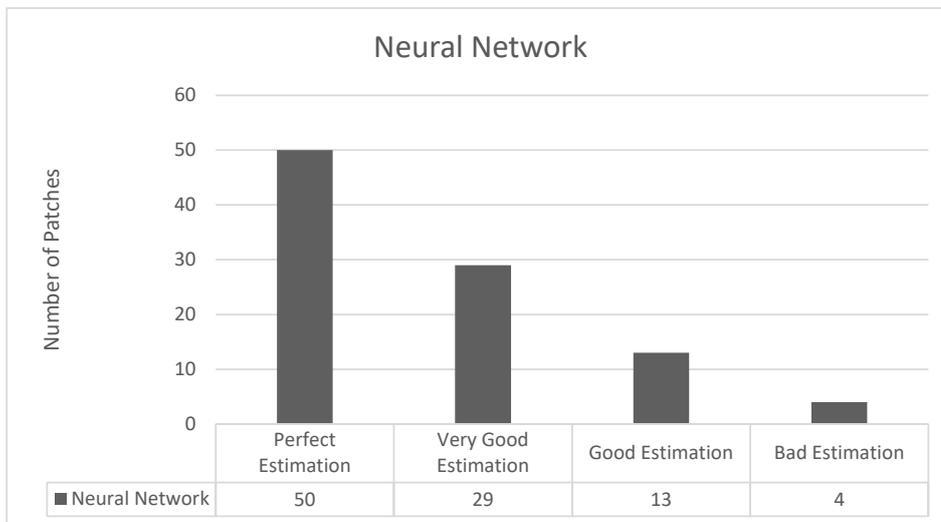


FIGURE 3.32: First Experiments - Neural Network Performance.

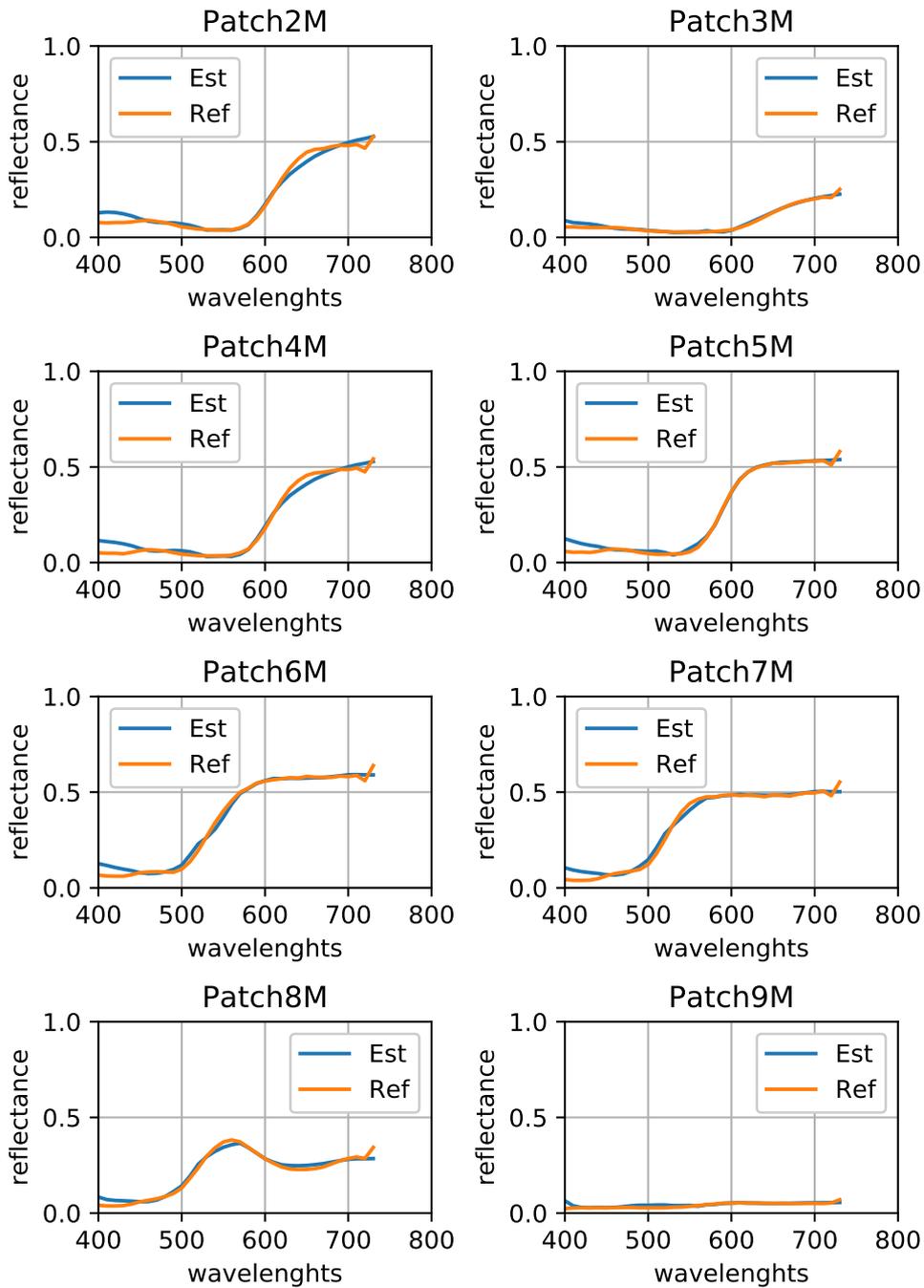


FIGURE 3.33: First Experiments - Neural Network, Patch 2M-Patch 8M.

### 3.4.9 SVR method

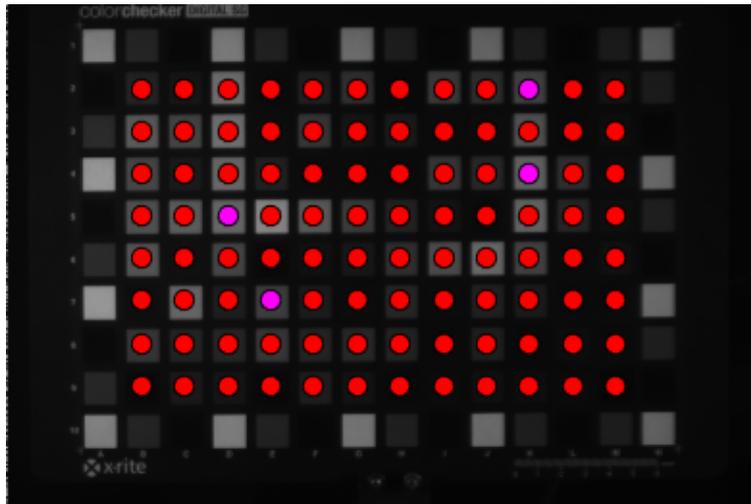


FIGURE 3.34: First Experiments - SVR Color Map.

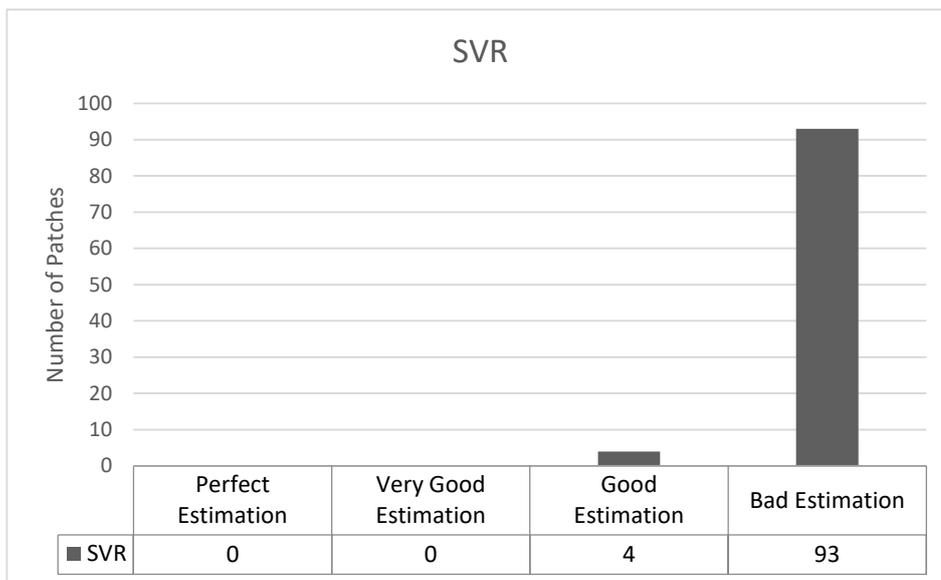


FIGURE 3.35: First Experiments - SVR Performance.

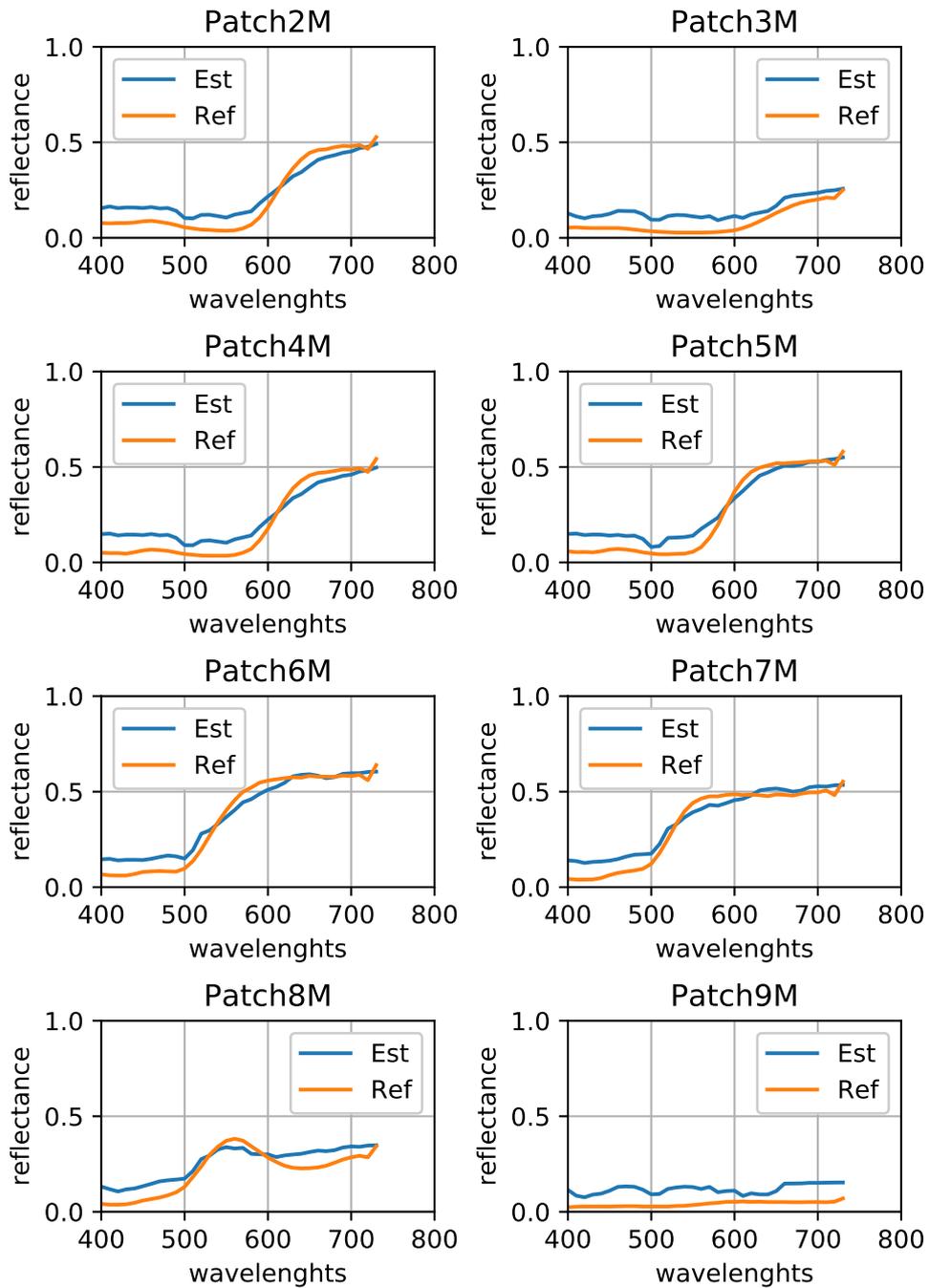


FIGURE 3.36: First Experiments - SVR, Patch 2M-Patch 8M.

### 3.4.10 First Comparison of the Methods

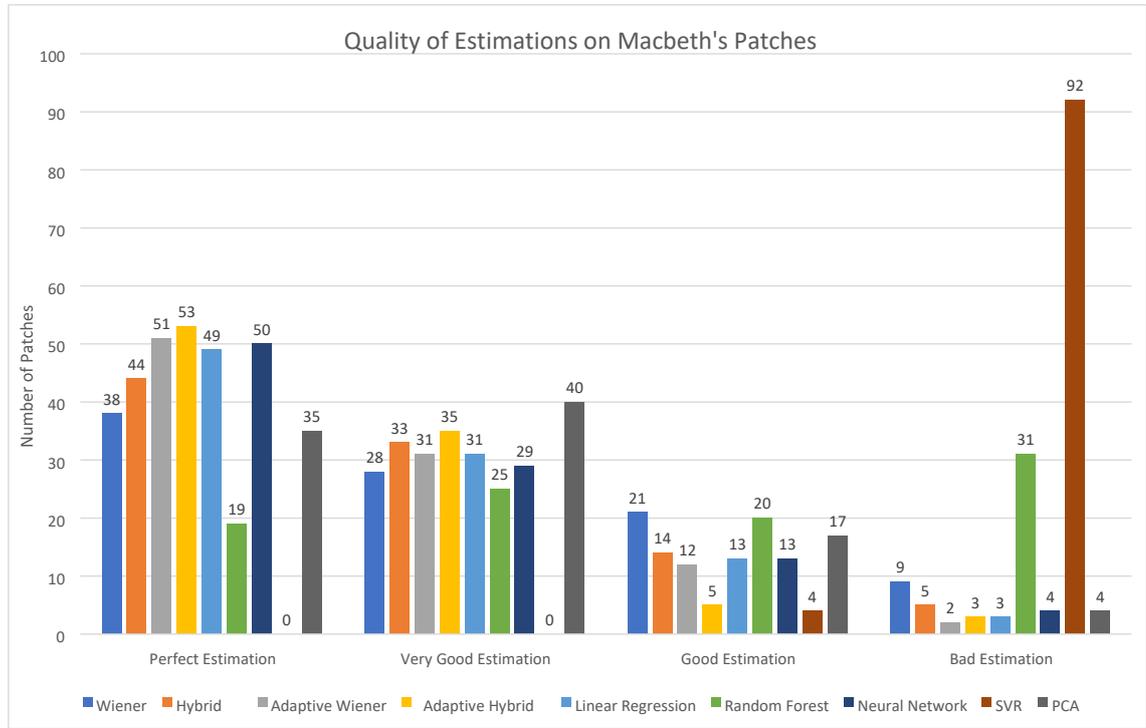


FIGURE 3.37: First Comparison of the Methods.

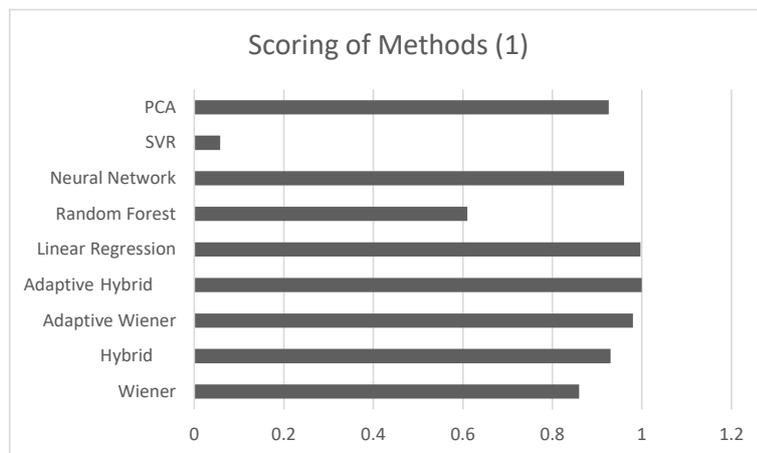


FIGURE 3.38: First ranking of the Methods.

Method	MEAN RMSE(%)	MAX RMSE(%)	MIN RMSE(%)	MEAN GFC
Wiener	1.6224	3.4977	0.2919	0.9964
PCA	1.6224	3.4977	0.2919	0.9964
Hybrid	1.5447	4.0273	0.2623	0.9965
Adaptive Wiener	1.3504	3.4608	0.3399	0.9970
Adaptive Hybrid	1.2611	3.0715	0.2770	0.9974
Linear Regression	1.4156	3.0680	0.3974	0.9963
Neural Network	1.4344	3.1105	0.4054	0.9963
Random Forest	2.4	5.669	0.2904	0.9863
SVR	5.517	8.1642	2.5079	0.9652

TABLE 3.4: Metrics of Error for the First Series of Experiments.

The conduction of the above comparisons led to important conclusions about the spectral estimation methods. Figure 3.37 illustrates the performance of the algorithms and specifically the number of patches that were estimated perfectly, very goodly, goodly, and badly. Alongside, Table 3.4 indicates analytically the mean, max, min RMSE, and the mean GFC of each method according to the corresponding experiment.

Overall, it is clear from Fig. (3.38) and Table 3.4 that Linear Regression and Adaptive Hybrid algorithms are the two main methods that achieve the most considerable estimations. In the opposite direction, SVR and Random Forest are the algorithms that did not estimate accurately the spectrum from 400nm to 730nm.

It can be seen in the first bar chart (Fig. (3.37)), that Wiener method comes with only 38 perfectly estimated patches in contrast with Hybrid, which comes with 44 perfectly estimated patches. On that account, Hybrid seems to perform better than the classic Wiener, as it gives less bad estimations and statistical less mean RMSE. The performance of the Adaptive Wiener is too close to Linear Regression as the Adaptive Wiener gives 51 perfect patches and 2 bad and Linear Regression gives 49 and 3 respectively. That observation is more than useful and very encouraging, for the reason that a HS system could achieve approximately the same estimation accuracy with constant time complexity, compare to the Adaptive Wiener model which due to the adaptation of the training set it's time expensive. The estimation accuracy is risen by using the Adaptive Hybrid, which accounted for the largest amount of patches that were estimated perfectly with mean GFC equal to 0.9970. Neural network seems to be close to both Linear Regression and Adaptive Wiener according to its score. Conversely, SVR and Random Forest Regression got the two lowest scores. Random Forest is considered to be a robust tool, which avoid over-fitting and a famous classifier but technically speaking, in that problem cannot compose an accurate estimation because it averages the regression decision of the decision trees, and alongside, it averages the values that are laid on the same decision bounds. This double averaging declines the performance of the final estimation. Ending up, SVR gives the worst estimations and it seems that cannot be a way to achieve the spectral estimation goal, but it is one of the most practical ways for classification.

The execution of the above experiments give us undeniably useful information about the effectiveness of the compared methods. But, an important question is what will be happened, if the target surface is more complex in spectral domain compare with the training set. Munsell database includes a great amount of spectra that by observing Fig. (3.1) and Fig. (3.2), it seems that the majority of Macbeth patches are already included in the Munshell database. Therefore, a way to answer

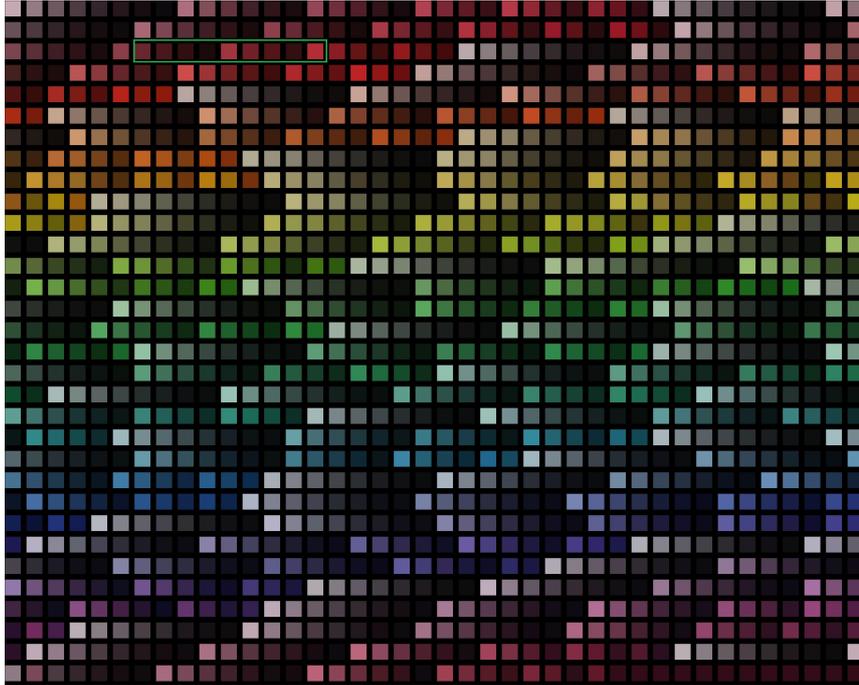


FIGURE 3.39: Isolated patches of Munsell in the greed grid for experiments.

this question is to swap the experiments by setting Macbeth as training set and Munsell as validation set. In the following section we execute one more time the same experiment using the same algorithms but swapped training and validation set.

### 3.5 Second Series of Experiments

The experiments at this section were carried out using the Macbeth SG Color Checker as training set and Munsell database as validation set. The bands that we used for sampling are [460 nm, 480 nm, 530nm, 580 nm, 610 nm, 690nm] and we strive to reconstruct the entire spectrum from 400nm - 730nm, totally 36 bands. All the other terms are the same as the first part of experiments but the estimation figures will be presented for the patches into the green grid in Fig. (3.39)

### 3.5.1 Wiener Method

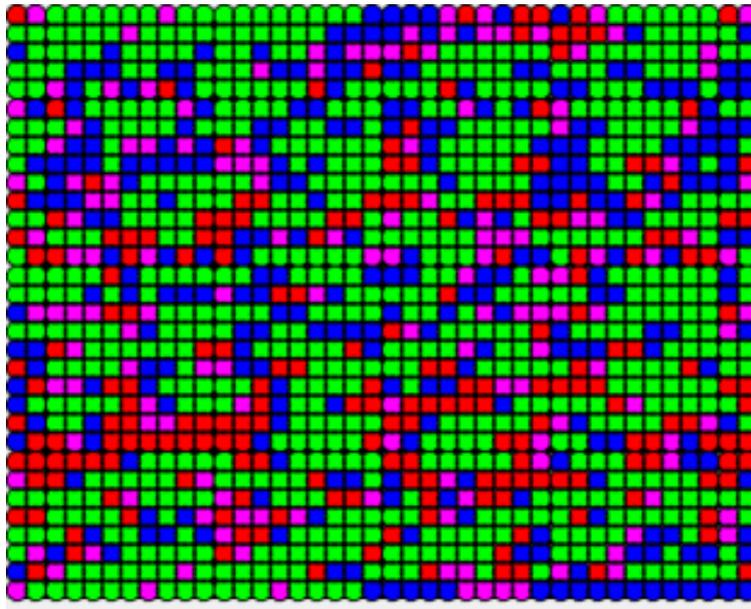


FIGURE 3.40: Second Experiments - Wiener Color Map.

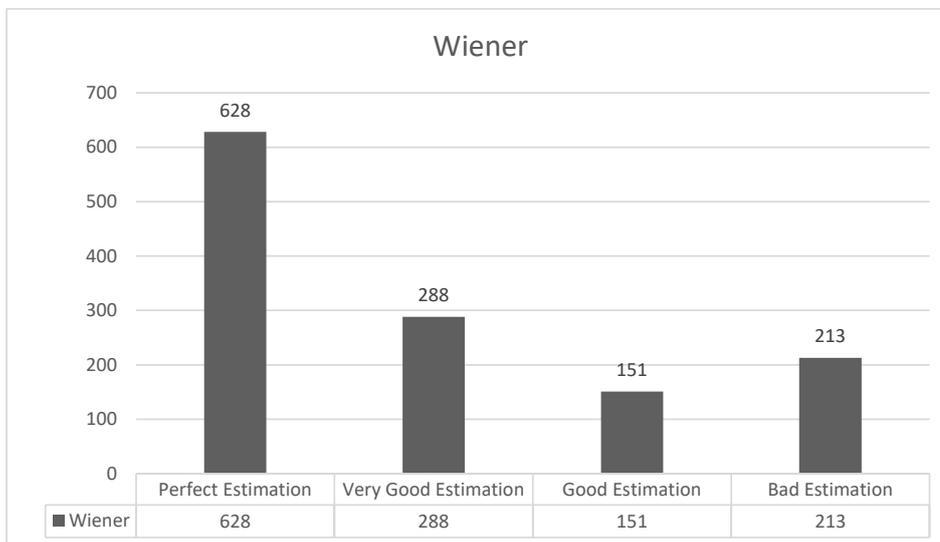


FIGURE 3.41: Second Experiments - Wiener Performance.

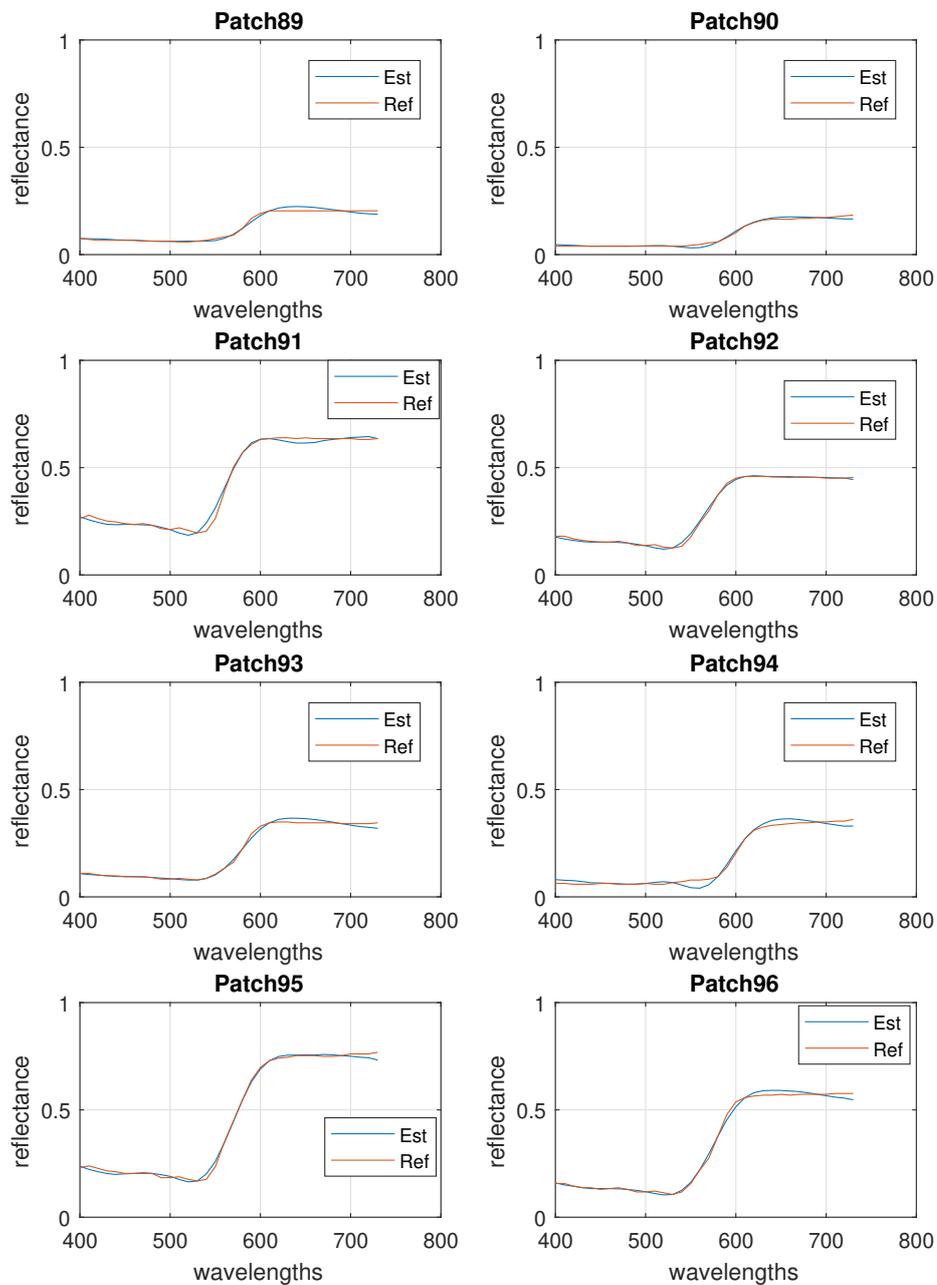


FIGURE 3.42: Second Experiments - Wiener, Patch 2M-Patch 8M.

### 3.5.2 PCA Method

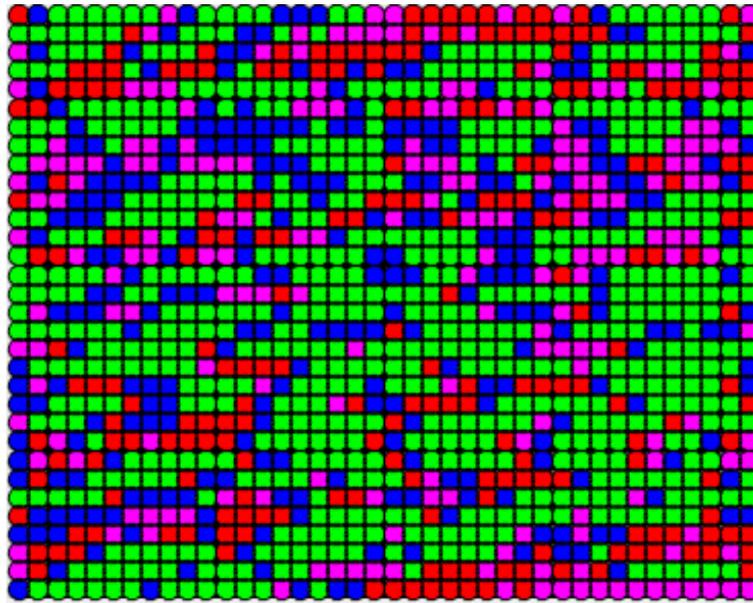


FIGURE 3.43: Second Experiments - PCA Color Map.

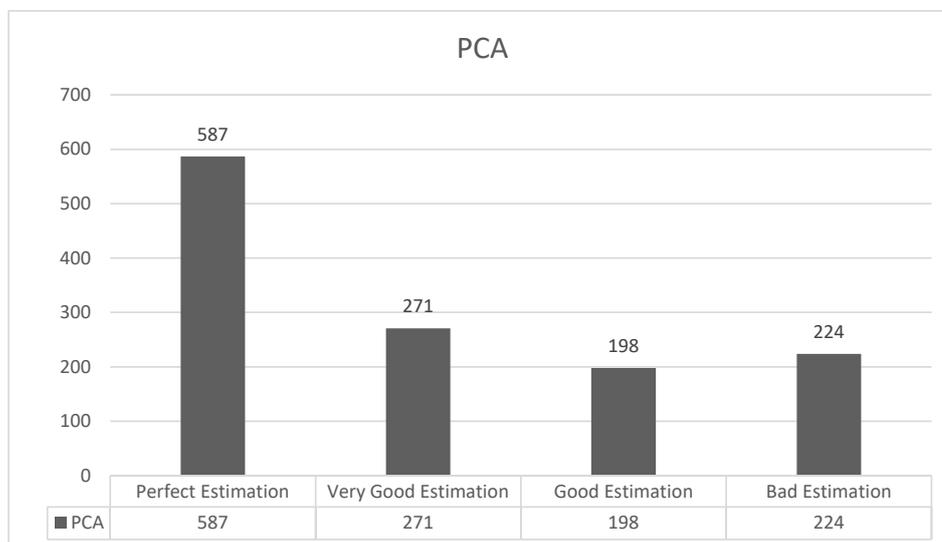


FIGURE 3.44: Second Experiments - PCA Performance.

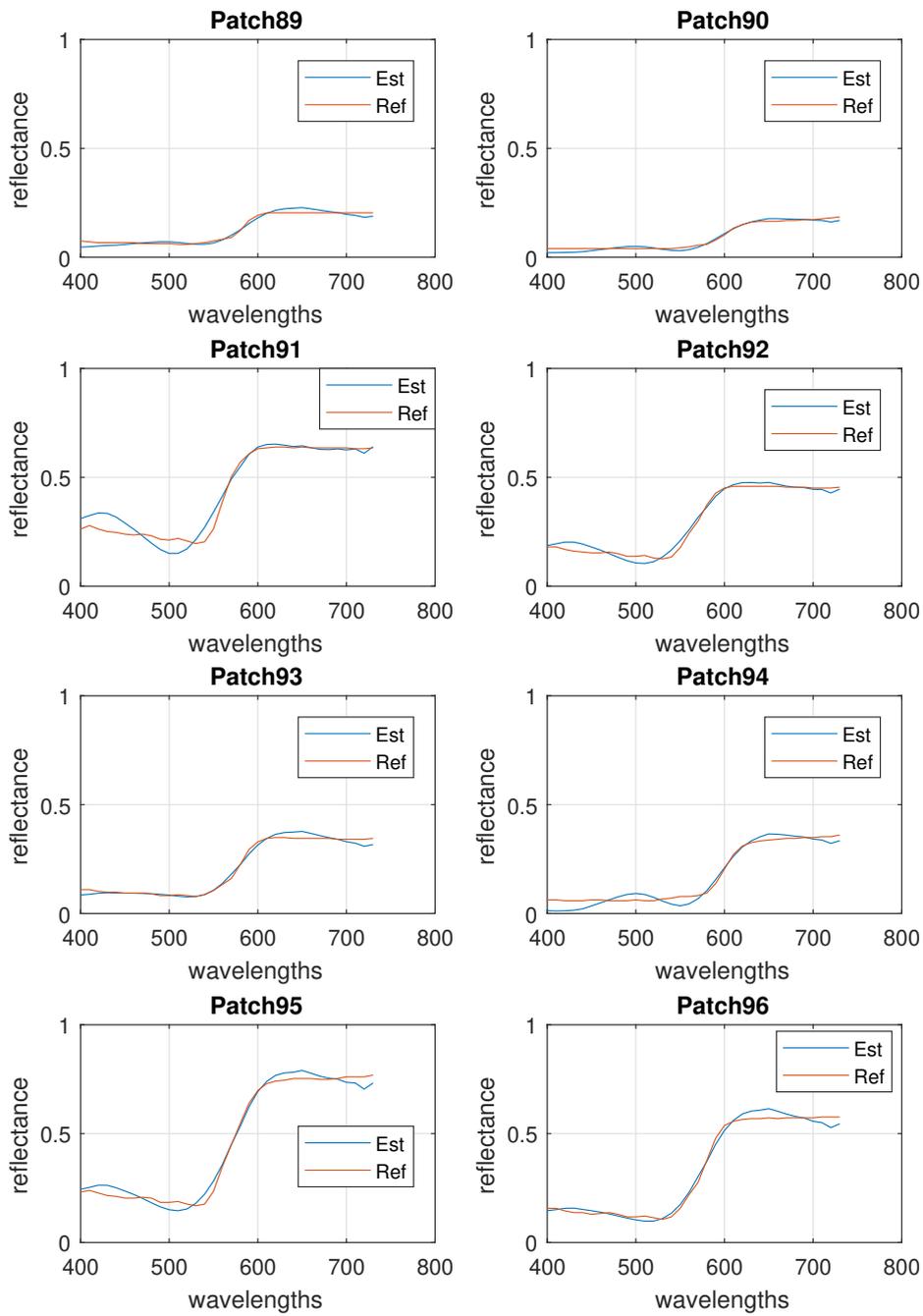


FIGURE 3.45: Second Experiments - PCA, Patch 2M-Patch 8M.

### 3.5.3 Hybrid Method

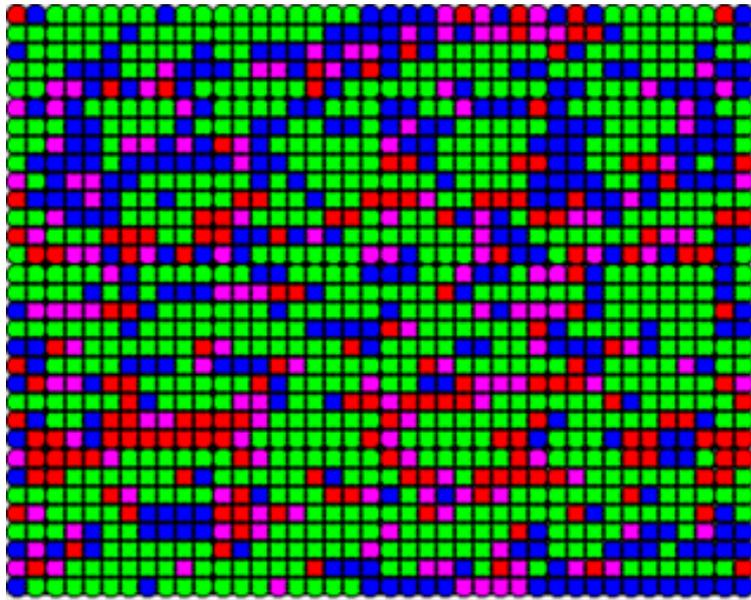


FIGURE 3.46: Second Experiments - Hybrid Color Map.

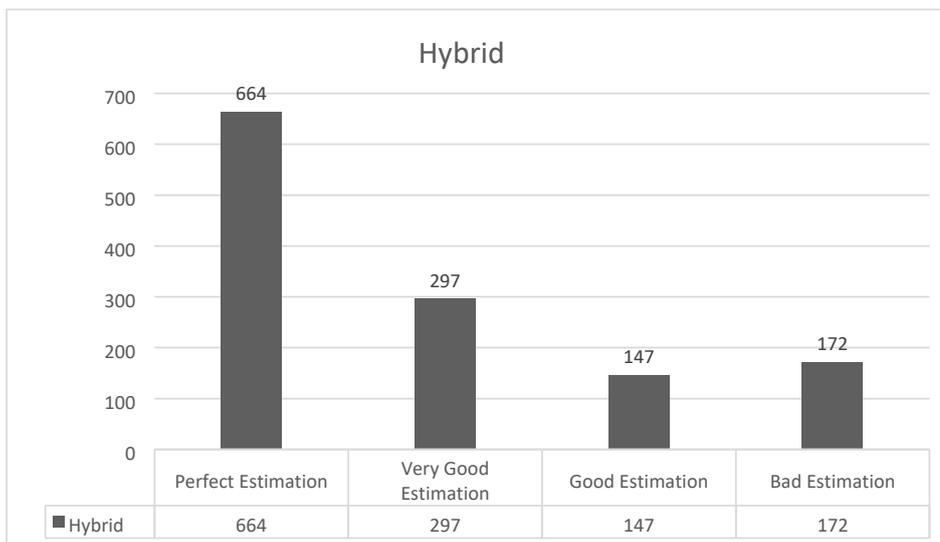


FIGURE 3.47: Second Experiments - Hybrid Performance.

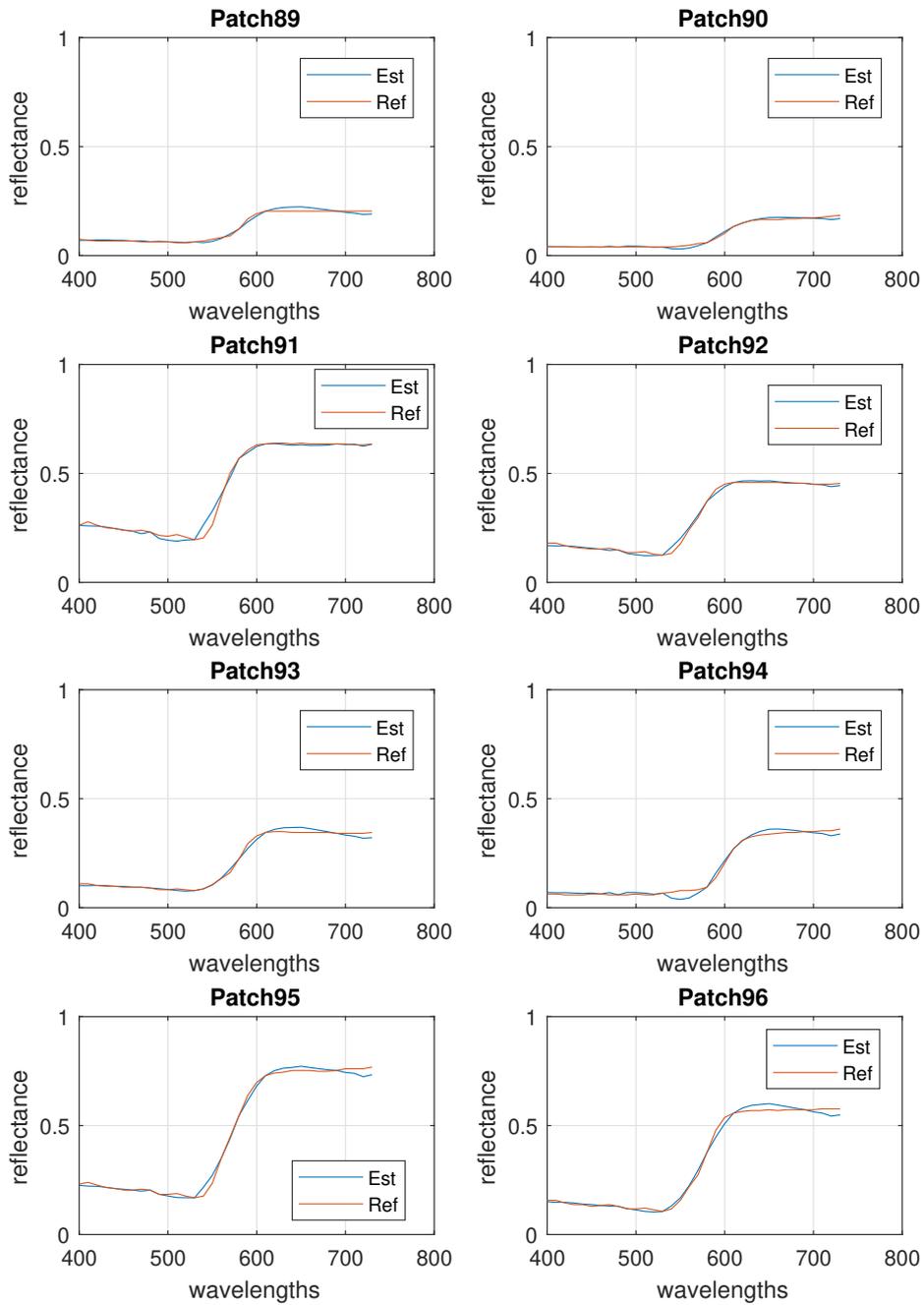


FIGURE 3.48: Second Experiments - Hybrid, Patch 2M-Patch 8M.

### 3.5.4 Adaptive Wiener Method

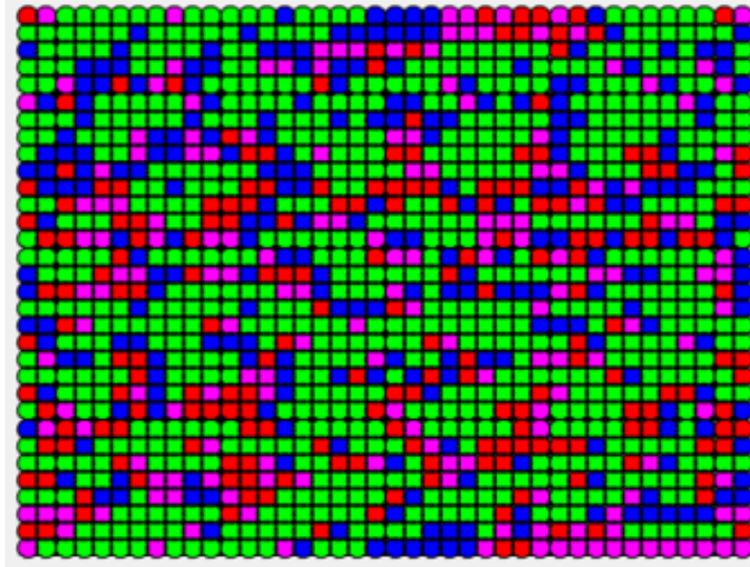


FIGURE 3.49: Second Experiments - Adaptive Wiener Color Map.

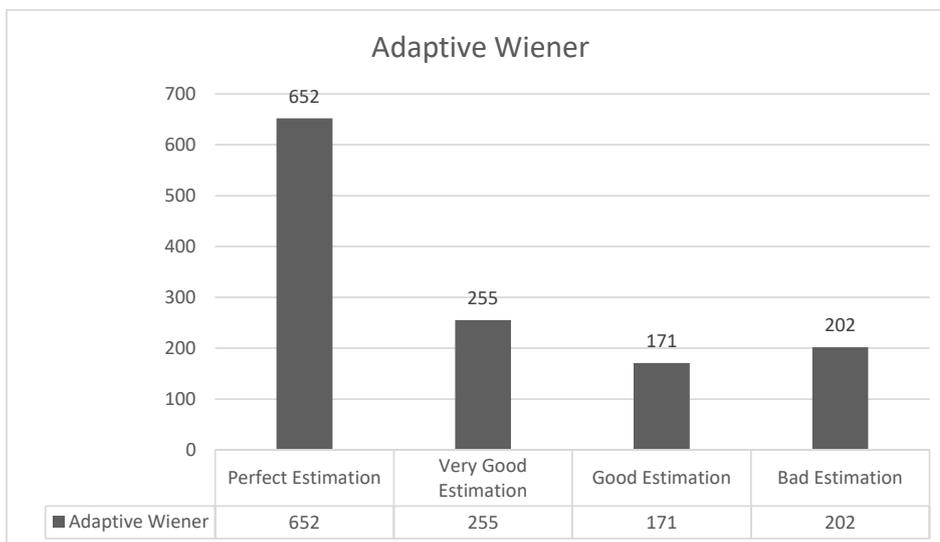


FIGURE 3.50: Second Experiments - Adaptive Wiener Performance.

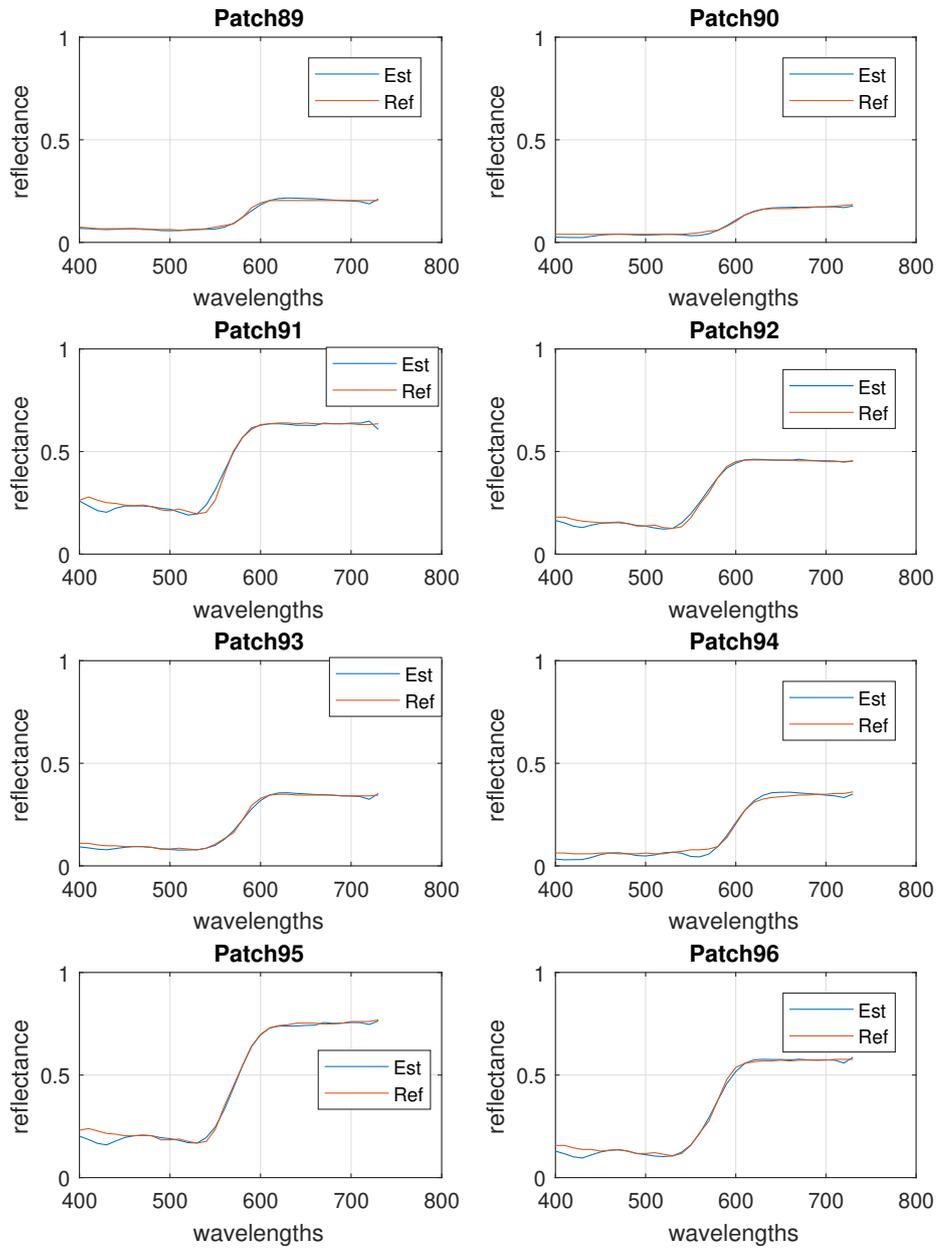


FIGURE 3.51: Second Experiments - Adaptive Wiener, Patch 2M-Patch 8M.

### 3.5.5 Adaptive Hybrid Method

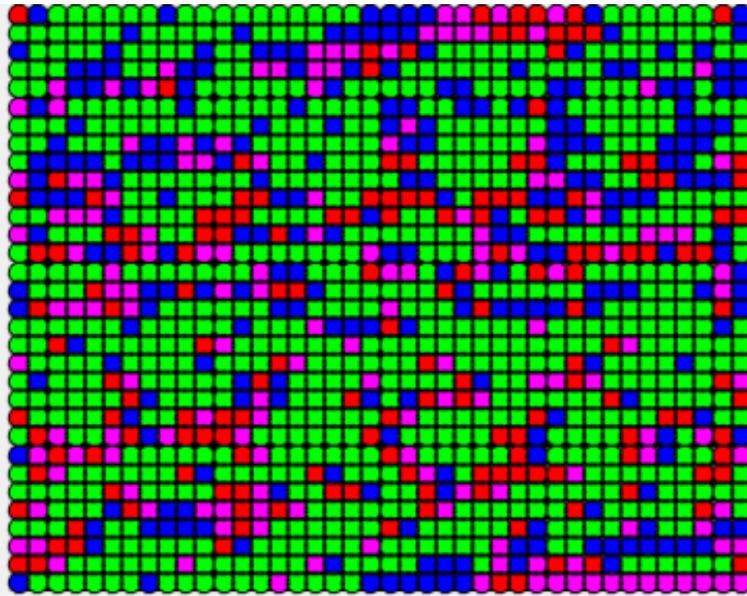


FIGURE 3.52: Second Experiments - Adaptive Hybrid Color Map.

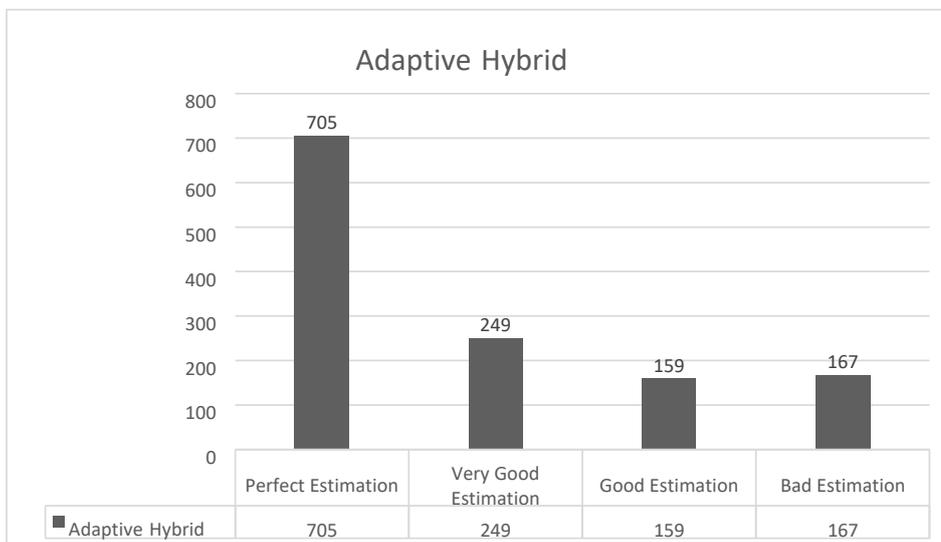


FIGURE 3.53: Second Experiments - Adaptive Hybrid Performance.

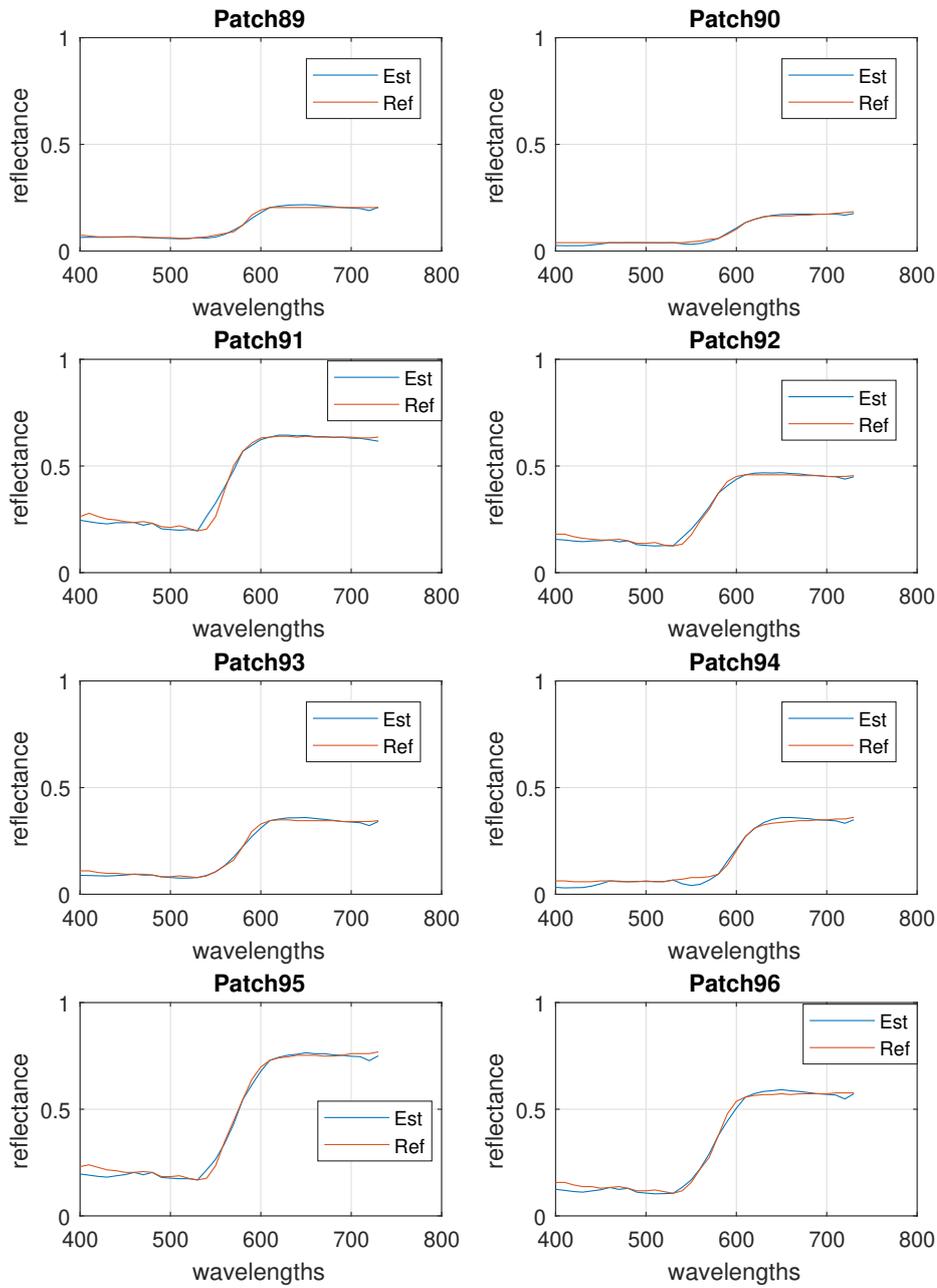


FIGURE 3.54: Second Experiments - Adaptive Hybrid, Patch 2M-Patch 8M.

### 3.5.6 Linear Regression Method

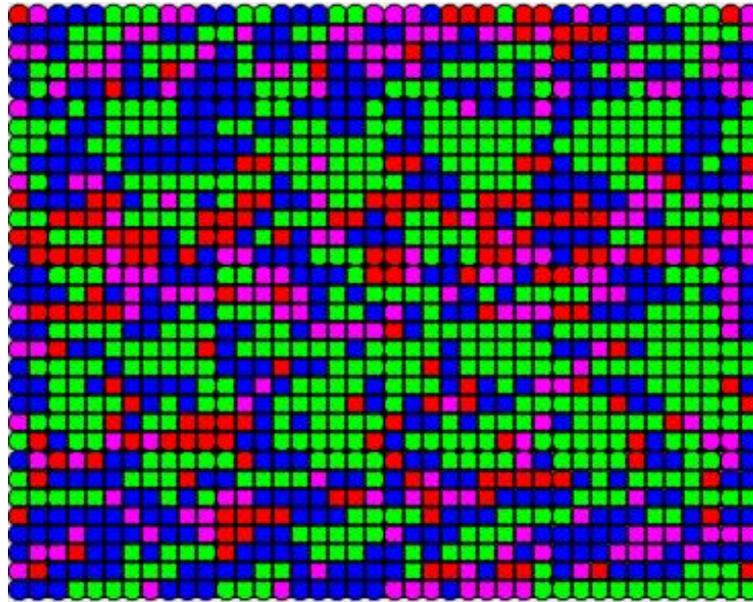


FIGURE 3.55: Second Experiments - Linear Regression Color Map.



FIGURE 3.56: Second Experiments - Linear Regression Performance.

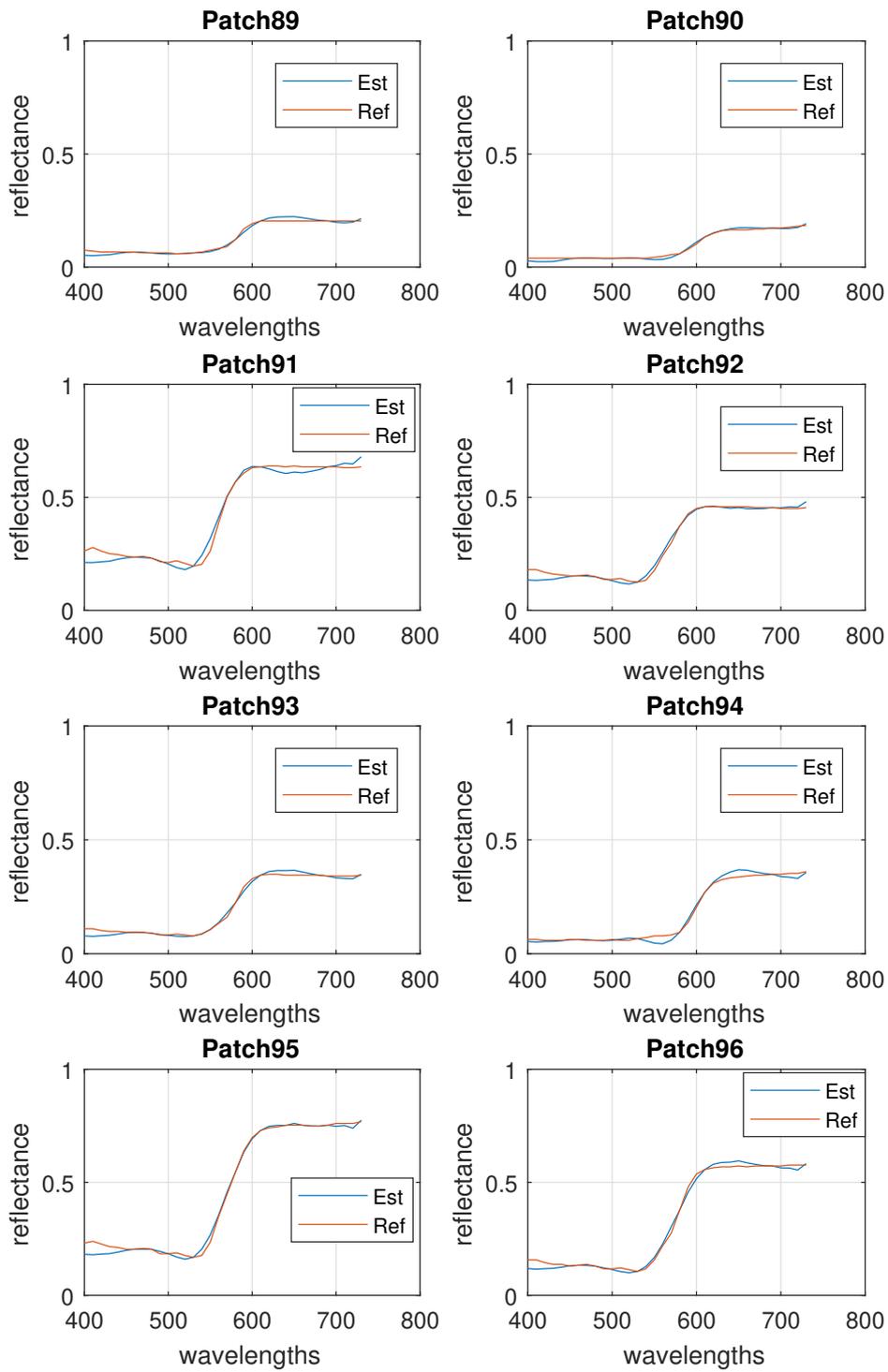


FIGURE 3.57: Second Experiments - Linear Regression, Patch 2M-Patch 8M.

### 3.5.7 Random Forest Method

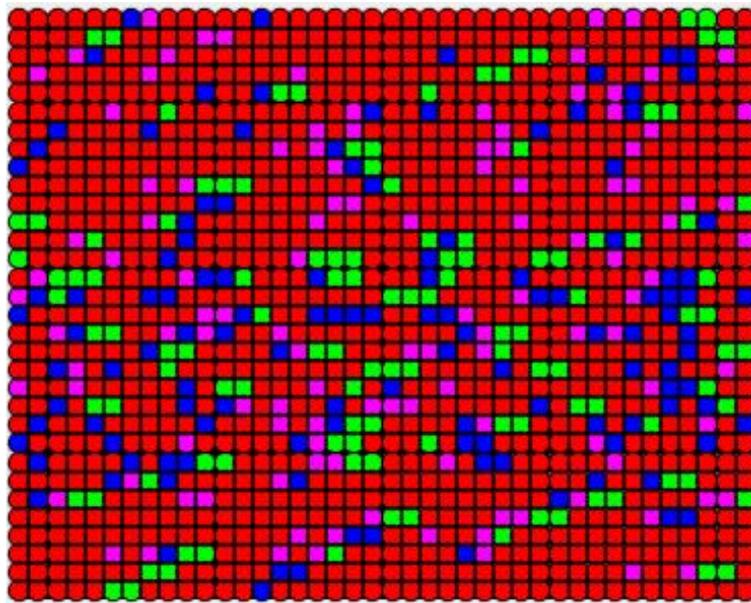


FIGURE 3.58: Second Experiments - Random Forest Color Map.

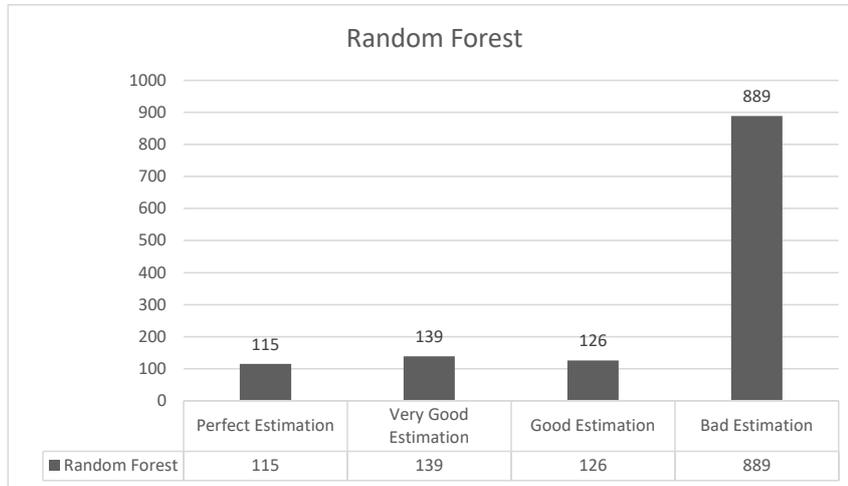


FIGURE 3.59: Second Experiments - Random Forest Performance.

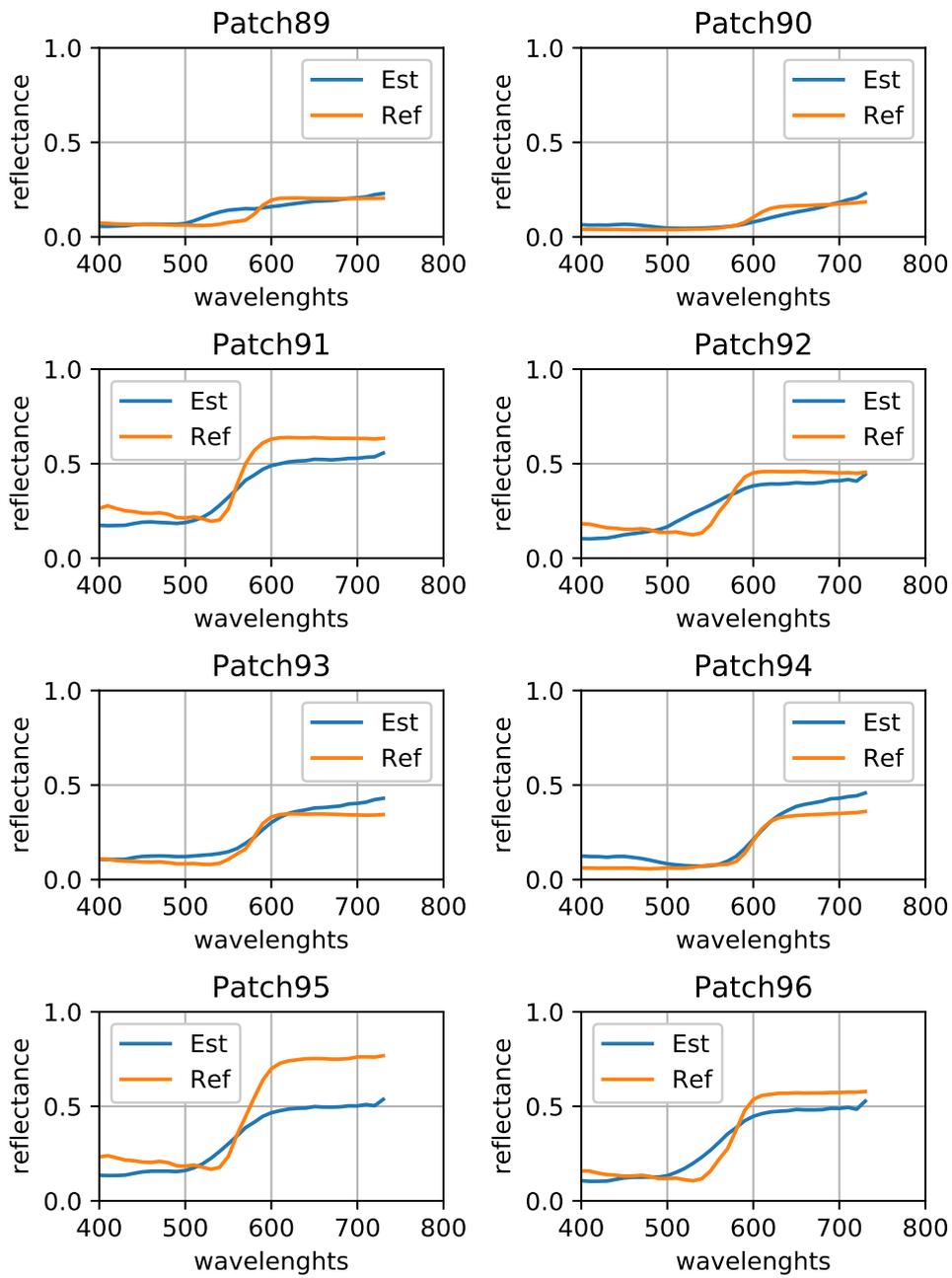


FIGURE 3.60: Second Experiments - Random Forest, Patch 2M-Patch 8M.

### 3.5.8 Neural Network Method

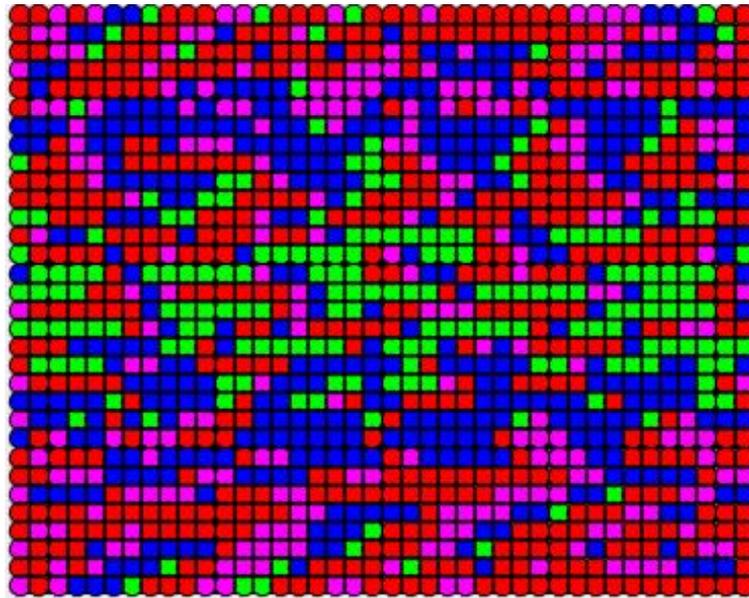


FIGURE 3.61: Second Experiments - Neural Network Color Map.

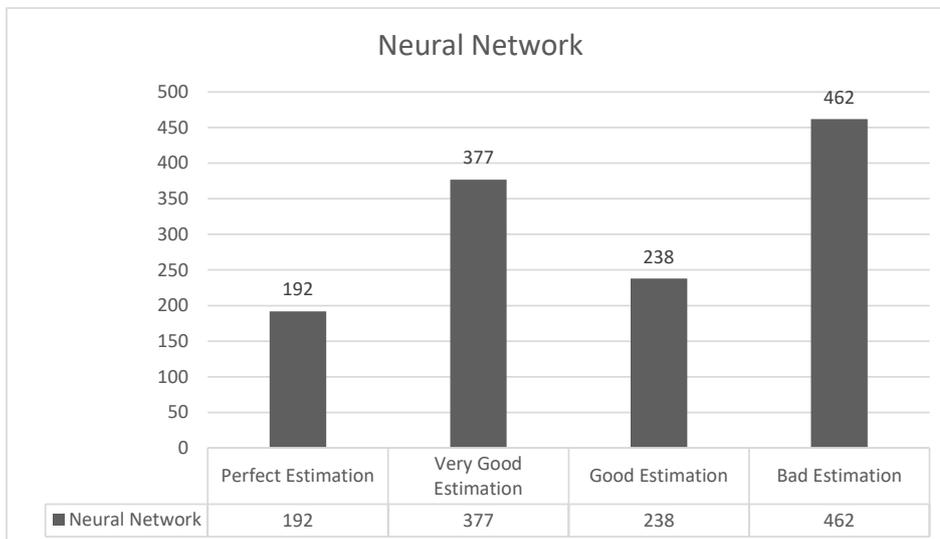


FIGURE 3.62: Second Experiments - Neural Network Performance.

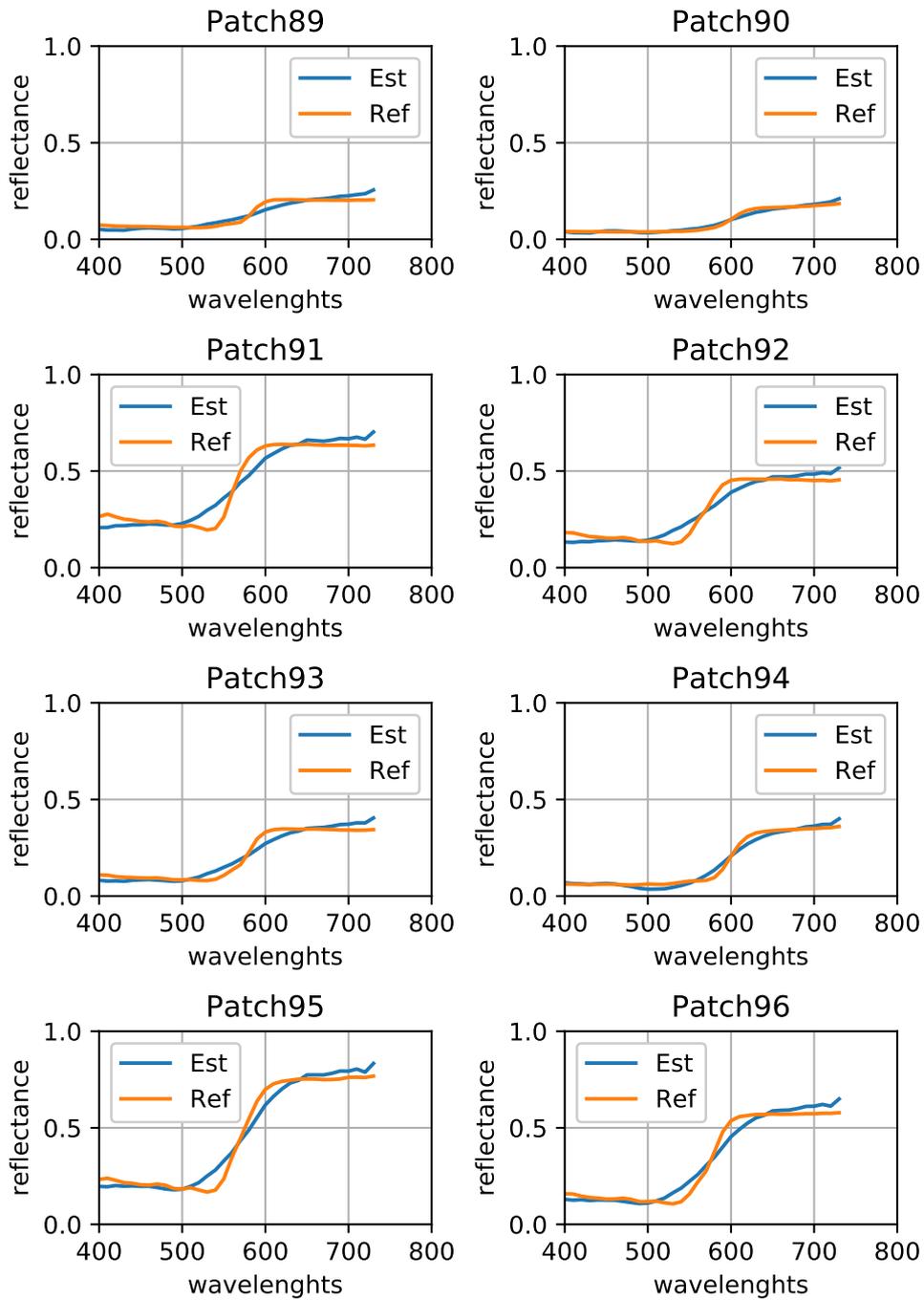


FIGURE 3.63: Second Experiments - Neural Network, Patch 2M-Patch 8M.

### 3.5.9 SVR method

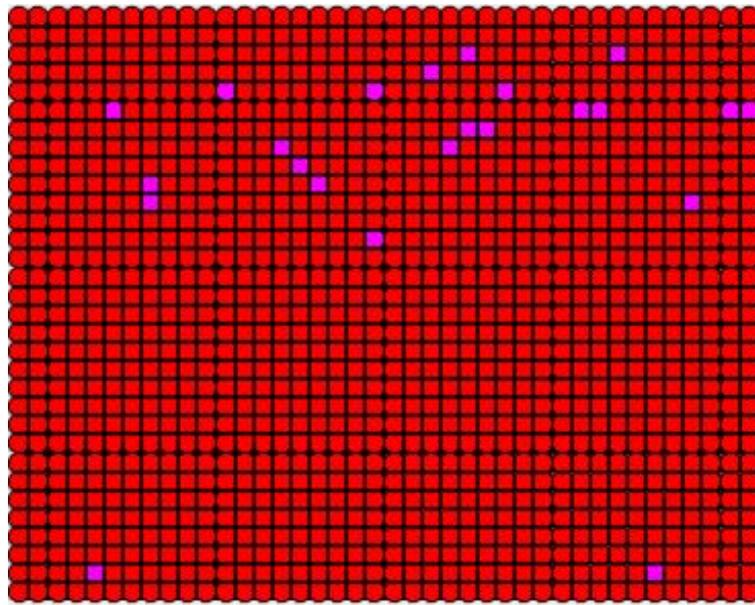


FIGURE 3.64: Second Experiments - SVR Color Map.

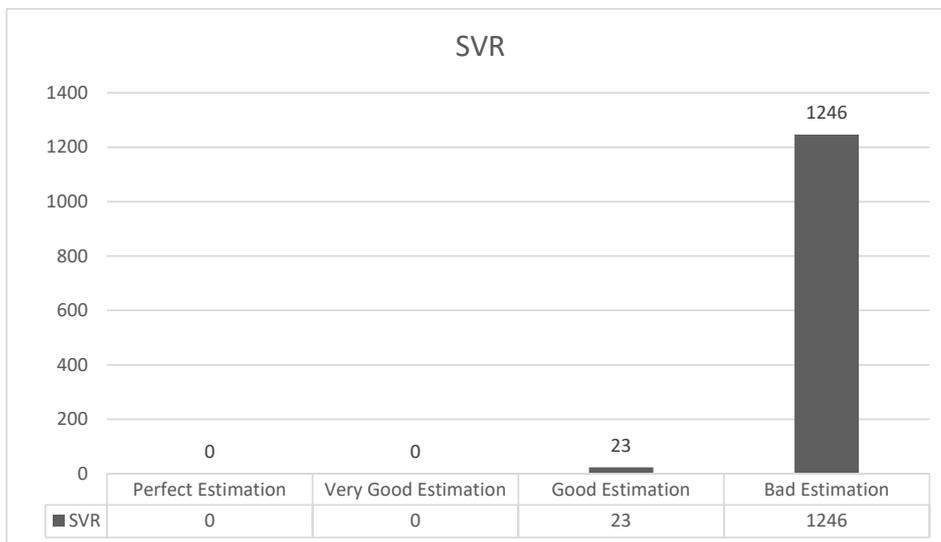


FIGURE 3.65: Second Experiments - SVR Performance.

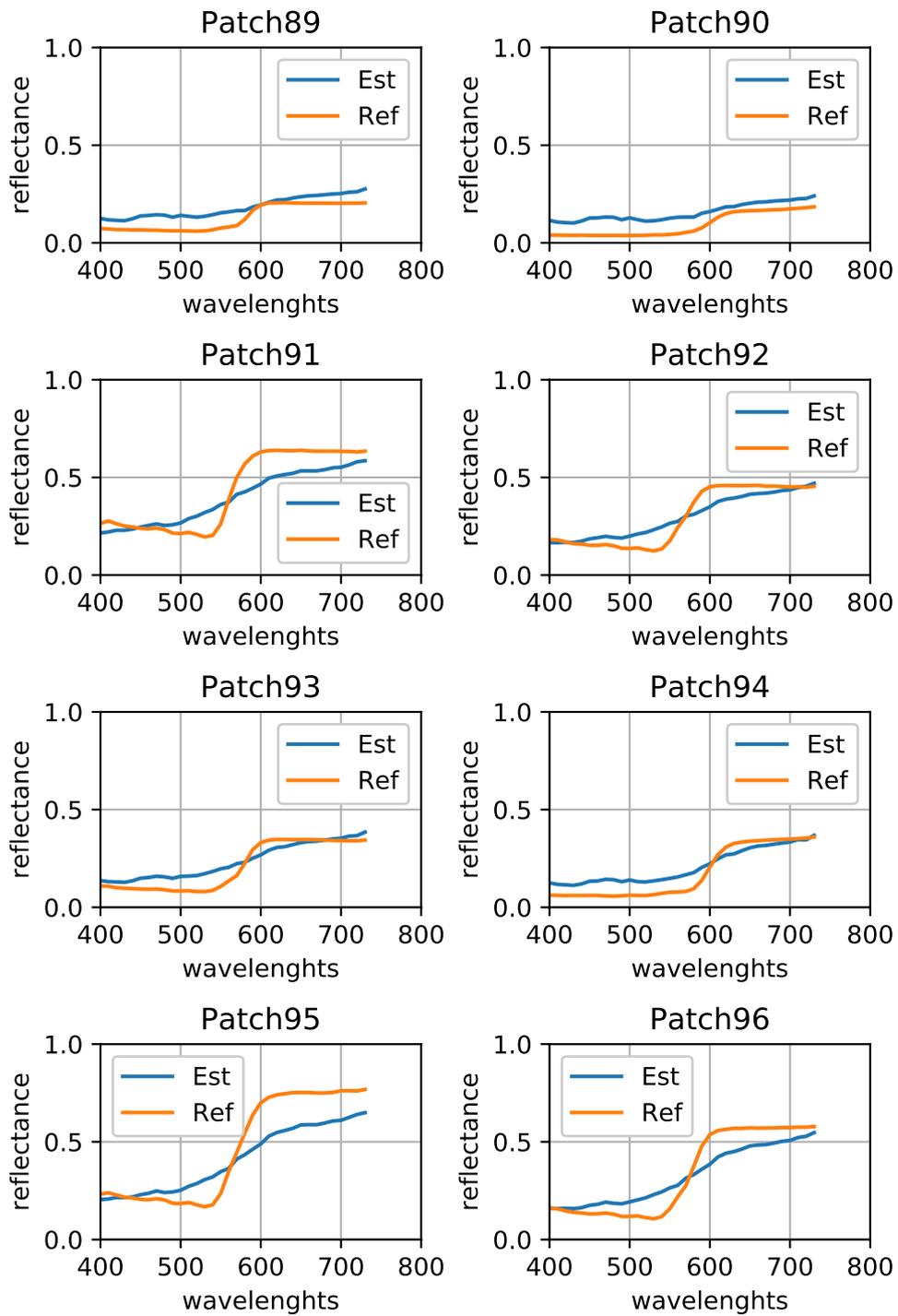


FIGURE 3.66: Second Experiments - SVR Performance.

### 3.5.10 Second Comparison of the Methods

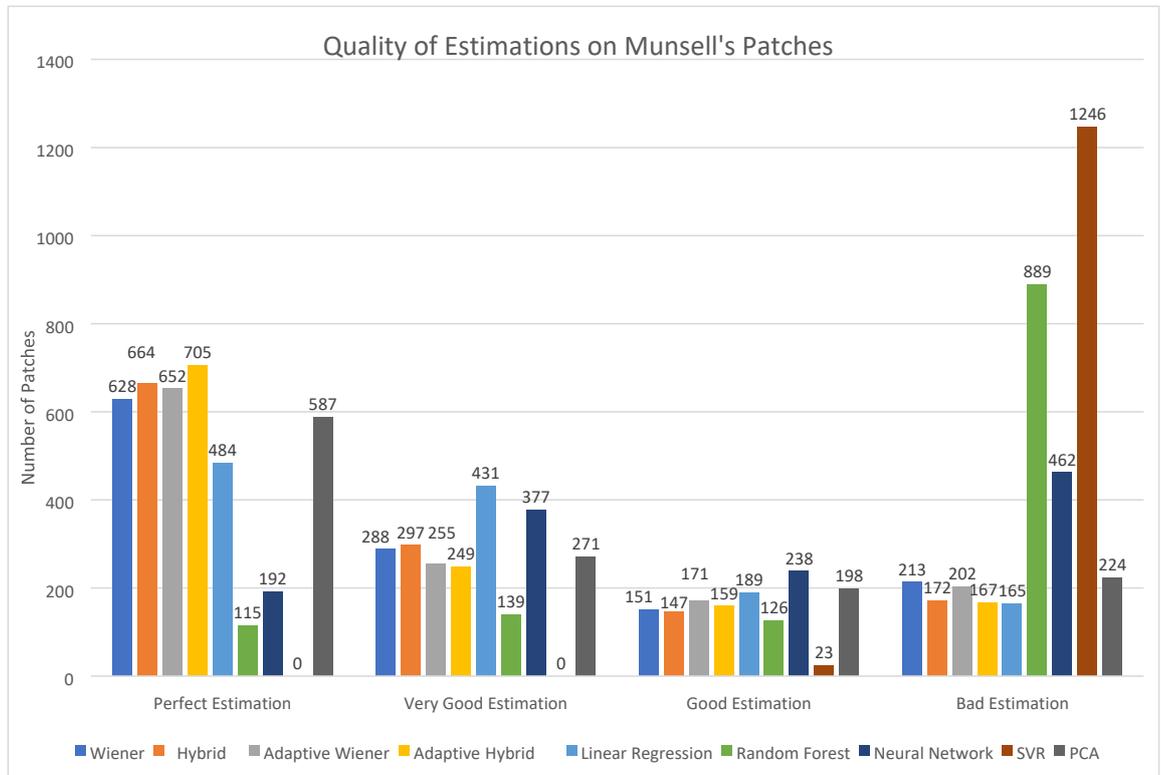


FIGURE 3.67: Second Comparison of the Methods.

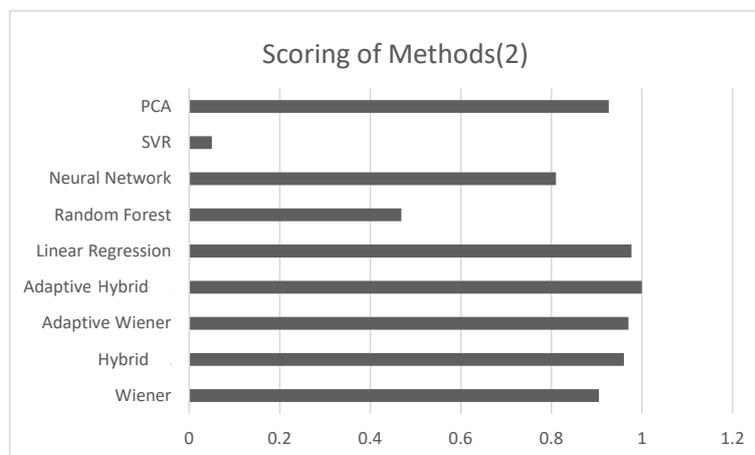


FIGURE 3.68: Second ranking of the Methods.

Method	MEAN RMSE(%)	MAX RMSE(%)	MIN RMSE(%)	MEAN GFC
Wiener	1.6902	7.2230	0.1443	0.9978
PCA	1.7213	6.3235	0.2085	0.9974
Hybrid	1.5450	6.2968	0.2623	0.1574
Adaptive Wiener	1.6488	6.4958	0.1464	0.9978
Adaptive Hybrid	1.5065	5.8194	0.1719	0.9981
Linear Regression	1.7619	4.8335	0.4079	0.9963
Neural Network	2.023	8.456	0.4054	0.9960
Random Forest	5.562	24.73	0.2769	0.9955
SVR	7.2715	16.955	2.3936	0.9951

TABLE 3.5: Metrics of Error of the Second Series of Experiments.

The reconduction of the above comparisons led to important conclusions about the spectral estimation methods using a smaller training set. Figure 3.67 illustrates the performance of the algorithms and specifically the number of patches that were estimated perfectly, very goodly, goodly, and badly. Error measurements are depicted in Table 3.5.

Overall, it is clear from Fig. (3.67) and Table.(3.5) that Hybrid and Adaptive Hybrid algorithms are the two main methods that achieve the most considerable estimations. In the opposite direction, SVR and Random Forest are the algorithms that did not estimate accurately the spectrum from 400nm-730nm as they accounted for the biggest mean RMSE.

On the second series of experiments, Hybrid still gives more perfectly and goodly estimated patches than traditional Wiener. Wiener, Adaptive Wiener, and Linear Regression perform in the same way as got the same score. Specifically, Linear Regression seems to be more powerful as it achieves the lowest amount of badly estimated patches and the lowest max RMSE. Therefore, Linear Regression outperforms Wiener and Adaptive Wiener. Adaptive Hybrid and simple Hybrid methods accounted for the highest ranking in the score bar chart. Adaptive Hybrid gives 705 perfectly estimated patches, 249 and 159 very good and good estimations respectively. Nevertheless, Hybrid and Linear Regression are characterized by stability and effectiveness comparing the experiments with the larger training set and with the smaller. SVR and Random Forest are the most inefficient methods due to the reasons that described in Section 3.4.10.

Conclusively, that experiment answered the question that had been posed in Section 3.4.10. A smaller training set can train a model in way that the model can estimate a great amount of unknown spectra. In other words, all the models were trained with only 96 spectra and, for instance, Adaptive Hybrid estimated badly only 167 patches out of 1269 patches. That is an optimistic result for the spectral estimation approach due to the fact that a HS imager could be trained with the database of Munsell and with high probability achieves accurate estimations in an unknown environment.

### 3.6 Total Scoring

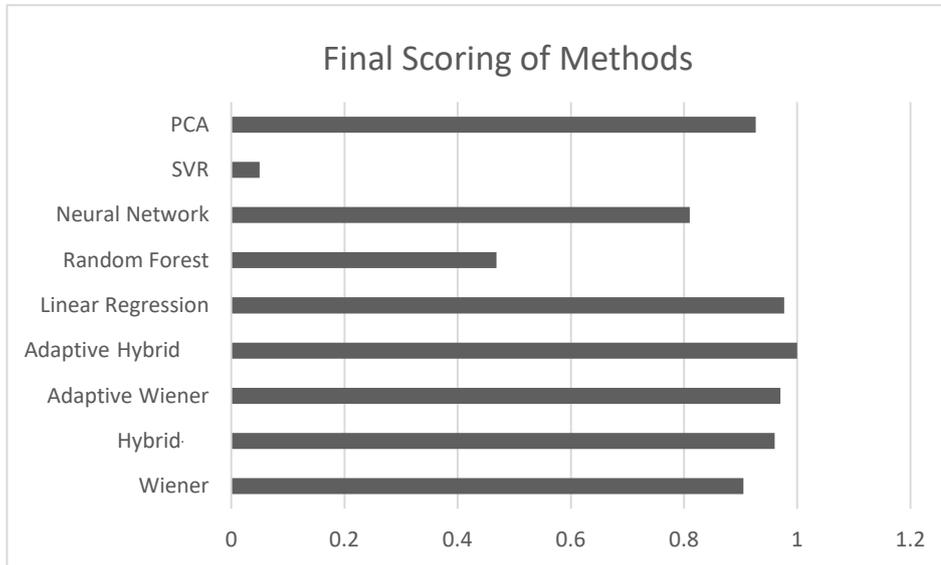


FIGURE 3.69: Total scoring of the Methods.

At this section, a final scoring of the algorithms took place. Specifically, both the scores from the first series of experiments and from the second were summarized and normalized to the highest score. This calculation produced the Figure 3.69, which illustrates the performance of the algorithms as a function of the above two series of experiments.

Overall, it is clear from Fig. (3.69) that Linear Regression and Adaptive Hybrid algorithms are the two main methods that achieve the most considerable estimations.

It comes without doubt that Hybrid and PCA outperform the traditional Wiener method, but at the same time Hybrid raise the number of the perfectly estimated patches. Moreover, Hybrid model reached that highest accuracy, when the available training set is small (i.e Mars Missions). The Adaptive Wiener seems to achieve the same score as the Linear Regression. According to that result, Linear Regression is a better candidate method than the Adaptive Wiener, which has time complexity  $O(\text{pixels} \times (2pk + J \log J))$ . Adaptive Hybrid keeps the first place in both experiments, therefore, whether the high accuracy of estimation is demanded Adaptive Hybrid is the best algorithm.

Ending up, Linear Regression and Hybrid-FL are the most powerful whether the real time execution and a good estimation are demanded, and Adaptive Hybrid whether the high spectral accuracy is needed regardless the time complexity.

### 3.7 Execution Times of the Spectral Estimation Algorithms

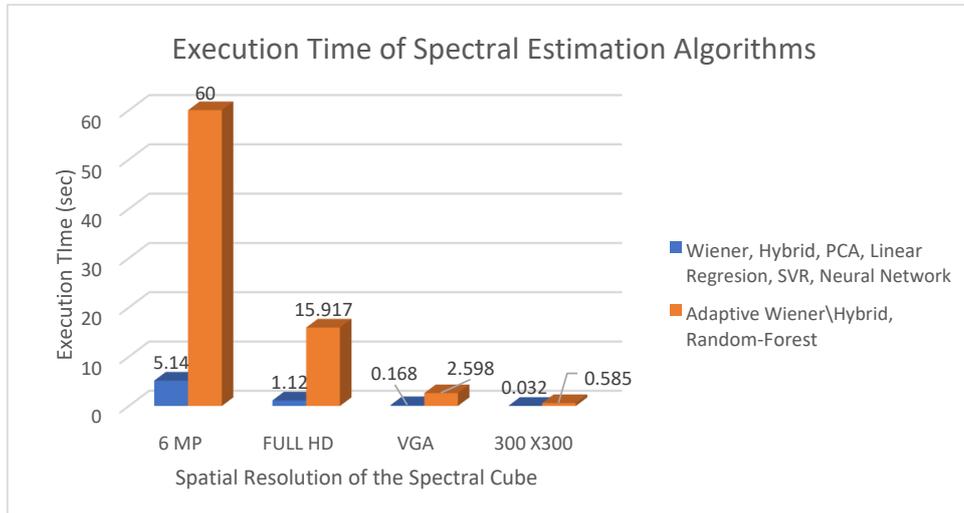


FIGURE 3.70: Execution Times of the Spectral Estimation Algorithms.

Figure 3.70 illustrates the execution times of each spectral estimation algorithm. The execution times are presented in four different forms of the possible spatial resolutions. Specifically, we have acquired a 6MP spectral cube by using the MuSES-9 HS camera and then, we utilized our software (Fig. (3.6)) both to scale the resolution of the 6MP spectral cube into FULL HD, VGA, 300 × 300, and to measure the execution times as a function of the spatial resolution.

Considering the Figure 3.70 and the time complexity of each algorithm, it was predictable that due to the fact that Wiener, Hybrid, PCA, Linear Regression, SVR, and the Neural network methods have  $O(pk)$  complexity their execution time will be approximately the same and equal to the minimum among all methods. Specifically, Hybrid method produces a really high spatial resolution cube by locking only for approximately 5 seconds the Intel I7 processor. As the resolution is declined the execution time is dropped off.

On the other side, adaptive methods have  $O(\text{pixels} \times (2pk + J \log J))$  time complexity, therefore, this is the reason that the execution time of estimating a 6MP spectral cube is approximately 1 minute. All the adaptive methods were implemented using multi-threading approaches, hence, the procession time could be fell down by raising up the number of the processor cores.

Ending up, the two main algorithms Linear Regression and Adaptive Hybrid could be executed in real time (less than a minute) and especially Linear Regression could also be implemented in a smart phone with a miniaturized HS camera for hands-on applications

## Chapter 4

# Band Selection - Feature Extraction

### 4.1 Introduction to Band Selection

A hyperspectral image cube contains hundreds of spectral bands with very fine spectral resolution. In particular, the very high data dimensionality presents a challenge to many traditional image analysis algorithms. One approach of reducing the data dimensionality is to transform the data onto a low dimensional space by using certain criteria. For instance, the objective of principal component analysis (PCA) is to maximize the variance of the transformed data (or minimize the reconstruction error). However, these methods usually change the physical meaning of the original data because the channels in the low-dimensional space do not correspond to individual original bands but their linear combinations. Another dimensionality reduction approach is the band selection approach in order to select a subset of the original bands without losing their physical meaning. Many band selection methods have been proposed. In terms of object information availability, band selection techniques can be divided into two categories: supervised and unsupervised. Supervised methods are to preserve the desired object information, which is known a priori, whereas unsupervised methods do not assume any object information. For example, canonical analysis was employed for band selection in [26]; the Jeffries–Matusita distance, divergence, and Bhattacharya distance between classes were used as selection criteria in [27]–[28], respectively. Although these supervised techniques clearly aim at selecting bands that include important object information and the selected bands can provide better detection or classification than those from unsupervised techniques, the required prior knowledge may be unavailable in practice. For instance, the bands that are more informative and they contribute in order to achieve more considerable spectral estimation are unknown and there is no a priori knowledge. Therefore, it is a need to develop reliable unsupervised band selection methods that can generally offer good performance regardless of the types of surfaces they estimate (or classify).

Because the basic idea of unsupervised band selection methods is to find the most distinctive and informative bands, a plenty of algorithms have been proposed over the last years. The major difference is that the algorithms are applied in the spatial domain, instead of being applied in the spectral domain for endmember extraction. There are quite few endmember extraction algorithms and a comparative study can be found in [29]. In general, endmember extraction algorithms can be divided into the following two categories: one extracting distinctive pixels based on similarity measurement and the other using the geometry concept, such as simplex. In this Chapter, we propose the application of the similarity-based endmember extraction algorithms in the spectral domain. Practical considerations such as algorithm initialization, the number of pixels to be involved in the band selection process are investigated.

## 4.2 Data Preprocessing

To select the most distinctive but informative bands, water absorption and low SNR bands need to be preremoved. This is because they can be very distinctive but not informative. Instead of manual selection, we compute the spectral correlation coefficients between original bands; those bands that have very low correlation coefficients with adjacent bands are considered as bad bands and will be preremoved[30]. The noise component in different bands is varied. If the noise component is larger, a band may look more different from others, although it may not be informatively distinct. Thus, noise whitening is needed, which requires noise estimation. It is known that noise estimation is a difficult task. In [31], it was demonstrated that the net effect of noise whitening and data whitening is similar. Therefore, we apply data whitening to the original bands (after bad band removal), which can be easily achieved by the eigendecomposition of the data covariance matrix. Then, the whitened bands actually participate in the following band selection process. Note that the selected bands are the original ones, not the whitened ones.

## 4.3 Properties of Similarity-Based Band Selection

To select the most distinctive and dissimilar bands, a similarity metric needs to be designated (widely used metrics include distance, correlation, etc). The measurement is taken on each pair of bands. The proposed band selection algorithms are using the same concept in endmember extraction except for Entropy and PCA approaches which just rank each band using a certain criterion. In addition, due to the large number of original bands, the exhaustive search for optimal band combinations is computationally prohibitive. The sequential forward search can save significant computation time [27]. It begins with sub-optimal two bands combination, and then, this band combination is subsequently augmented to three, four, and so on until the desired number of bands is selected. The proposed band selection algorithms using the endmember extraction concept adopt this sequential forward search strategy. The basic steps of that method can be described as follows:

1. Initialize the algorithm by choosing a pair of bands  $\mathbf{b}_1$  and  $\mathbf{b}_2$ . Then, the resulting selected band subset is  $\Phi = \{\mathbf{b}_1 \mathbf{b}_2\}$ .
2. Find a third band  $\mathbf{b}_3$  that is the most dissimilar to all the bands in the current  $\Phi$  by using a certain criterion. Then, the selected band subset is updated as  $\Phi = \Phi \cup \{\mathbf{b}_3\}$ .
3. Continue on Step 2) until the number of bands in  $\Phi$  is large enough.

Another advantage is that, it is less dependent on the number of bands to be selected because those bands that are already being selected do not change with this value; increasing this value simply means to continue the algorithm execution with the bands being selected, whereas decreasing this number simply means to keep enough bands from the selected band subset (starting with the first selected band) as the final result. Moreover, in this section a new similarity based algorithm (BS), specific for spectral estimation will be introduced and compared with other well-known band selection methods. BS algorithm does not need initialization of bands, something that is really important comparing with other band extraction methods, which need initialization of the first two bands bands. Furthermore, we propose a second band selection algorithm that is based on linear interpolation technique (LIBS).

## 4.4 Band Selection Algorithms

Eight band selection algorithms are evaluated here. Entropy, Correlation, Principal component analysis ranking-based algorithm, Linear prediction and Orthogonal subspace projection, Sparse PCA, LIBS, and BS algorithm will be discussed and analyzed in this Section.

### 4.4.1 Information Entropy

This method is based on evaluating each band separately using the information entropy measure (Russ et al. [32]) defined as follows:

$$H(\lambda) = - \sum_{i=1}^l p_i \ln p_i \quad (4.1)$$

where  $H$  is the entropy measure,  $p$  is the probability density function of reflectance values, and  $l$  is the number of distinct reflectance values. The probabilities are estimated by computing a histogram of reflectance values. Generally, if the entropy value  $H$  is high, then the amount of information in the data is large. Thus, the bands are ranked in the ascending order from the band with the highest entropy value (large amount of information) to the band with the smallest entropy value (small amount of information).

### 4.4.2 Principal Component Analysis Ranking (PCAr)

Principal component analysis has been used very frequently for band selection in the past (Campbell et al. [33]). The method transforms a multidimensional space to one of an equivalent number of dimensions where the first dimension contains the most variability in the data, the second most, and so on. The process of creating this space gives two sets of outputs. The first is a set of values that indicate the amount of variability each of the new dimensions in the new space represents, which are also known as eigenvalues. The second is a set of vectors of coefficients, one vector for each new dimension, that define the mapping function from the original coordinates to the coordinate value of a particular new dimension. The mapping function is the sum of the original coordinate values of a data point weighted by these coefficients. As a result, the eigenvalue indicates the amount of information in a new dimension and the coefficients indicate the influence of the original dimensions on the new dimension. The PCA-based ranking system (PCAr) makes use of these two facts by scoring the bands (the “original” dimensions in the above discussion) as follows:

$$PCAr(\lambda_i) = \sum_{i=1}^j |\epsilon_i c_{ij}| \quad (4.2)$$

where  $\lambda_i$  is the central wavelength,  $\epsilon_j$  is the eigenvalue for the  $j^{th}$  principal component, and  $c_{ij}$  is the mapping coefficient of the  $i^{th}$  central wavelength in the  $j^{th}$  principal component. Because the procedure for computing the eigenvalues and coefficients is both complex and available in most data analysis texts (Duda et al. [34]), it is omitted.

### 4.4.3 Linear prediction (LP)

Qian et al. [35] proposed the LP band selection method. The concept in the LP-based band selection was originally used in the UFCLSLU for endmember pixel selection in [36], which means that a pixel with the maximum reconstruction error, using the linear combination of existing endmember pixels, is the most distinctive pixel. The difference here is that, for band selection, there is no constraint imposed on the coefficients of linear combination. Assume that all the pixels in a spectral cube for each band are gathered in  $\mathbf{B}$ , where  $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p]$  is a  $N \times p$  matrix and there are two bands  $\mathbf{b}_1$  and  $\mathbf{b}_2$  in  $\Phi$  with  $N$  pixels each. To find a band that is the most dissimilar to  $\mathbf{b}_1$  and  $\mathbf{b}_2$ ,  $\mathbf{b}_1$  and  $\mathbf{b}_2$  are used to estimate a third band  $\mathbf{b}$ , i.e.,

$$a_0 + \mathbf{b}_1 a_1 + \mathbf{b}_2 a_2 = \mathbf{b}' \quad (4.3)$$

where  $\mathbf{b}'$  is the estimate or linear prediction of band  $\mathbf{b}$  using  $\mathbf{b}_1$  and  $\mathbf{b}_2$ , and  $a_0, a_1$ , and  $a_2$  are the parameters that can minimize the linear prediction error. Let the parameter vector be  $\mathbf{a} = (a_0 \ a_1 \ a_2)^T$ . It can be determined using a least squares solution

$$\mathbf{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (4.4)$$

where  $\mathbf{X}$  is an  $N \times 3$  matrix whose first column is one, second column includes all the  $N$  pixels in  $\mathbf{b}_1$ , and third column includes all the pixels in  $\mathbf{b}_2$ .  $\mathbf{y}$  is an  $N \times 1$  vector with all the pixels in  $\mathbf{b}$ . The band that yields the maximum error  $e_{min}$  (using the optimal parameters) is considered as the most dissimilar band to  $\mathbf{b}_1$  and  $\mathbf{b}_2$  and will be selected as  $\mathbf{b}_3$  for  $\Phi$ . Obviously, the similar procedure can be easily conducted when the number of bands in  $\Phi$  is larger than two.

### 4.4.4 Orthogonal Subspace Projection (OSP)

Qian et al. [35] proposed the OSP algorithm for band selection approaches. Assume that there are two bands  $\mathbf{b}_1$  and  $\mathbf{b}_2$  in  $\Phi$ . To find a band that is the most dissimilar to  $\mathbf{b}_1$  and  $\mathbf{b}_2$ , an orthogonal subspace of  $\mathbf{b}_1$  and  $\mathbf{b}_2$  is constructed as

$$\mathbf{P} = \mathbb{I} - \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \quad (4.5)$$

where  $\mathbb{I}$  is an  $N \times N$  identity matrix, and  $\mathbf{Z}$  is an  $N \times 2$  matrix whose first column includes all the pixels in  $\mathbf{b}_1$  and second column includes all the pixels in  $\mathbf{b}_2$ . Then, the projection  $\mathbf{y}_o = \mathbf{P}^T \mathbf{y}$  is computed, where  $\mathbf{y}$  includes all the pixels in  $\mathbf{b}$  and  $\mathbf{y}_o$  is the component of  $\mathbf{b}$  in the orthogonal subspace of  $\mathbf{b}_1$  and  $\mathbf{b}_2$ . The band that yields the maximum orthogonal component  $\|\mathbf{y}_o\|$  is considered as the most dissimilar band to  $\mathbf{b}_1$  and  $\mathbf{b}_2$  and will be selected as  $\mathbf{b}_3$  for  $\Phi$ . The similar procedure can be easily conducted when the number of bands in  $\Phi$  is larger than two. It can be proved that the least squares solution to the unconstrained linear unmixing problem is mathematically equivalent to the OSP solution. Therefore, the LP-based approach (which can be considered as the unconstrained linear unmixing problem) actually yields the identical selected bands from the OSP based approach. Nevertheless, we will present the results from both LP and OSP algorithms. Compared to the OSP solution, the LP-based approach is computationally more efficient because it involves matrices with relatively smaller size. The remaining problem is to find the initial bands for OSP, LP, and CorM algorithms.

### Band Used as the Initial for Band Selection for OSP and LP

The initial band pair is critical to the performance of the proposed algorithms. Intuitively, we should use the two bands whose dissimilarity is the largest. Instead of the exhaustive search, the following algorithm can be applied to an original  $p$ -band data set.

1. Randomly select a band  $\mathbf{A}_1$ , and project all the other  $p - 1$  bands to its orthogonal subspace  $\langle \mathbf{A}_1 \rangle^\perp$ .
2. Find the band  $\mathbf{A}_2$  with the maximum projection in  $\langle \mathbf{A}_1 \rangle^\perp$ , which is considered as the most dissimilar to  $\mathbf{A}_1$ .
3. Project all the other  $p - 1$  bands to the orthogonal subspace  $\langle \mathbf{A}_2 \rangle^\perp$ , and find the band  $\mathbf{A}_3$  with the maximum projection.
4. If  $\mathbf{A}_3 = \mathbf{A}_1$ ,  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are confirmed to be the pair with the most significant dissimilarity, and the algorithm is terminated; if  $\mathbf{A}_3 \neq \mathbf{A}_1$ , go to the next step.
5. Continue the algorithm until  $\mathbf{A}_{i+1} = \mathbf{A}_{i-1}$ , then either  $\mathbf{A}_{i-1}$  or  $\mathbf{A}_i$  can be used as the band selection initial  $\mathbf{b}_1$  (or  $\mathbf{A}_{i-1}$  and  $\mathbf{A}_i$  are used as the initial band pair).

Qian et al. [35] found out that this algorithm can always extract the two most distinctive bands, regardless of its initial  $\mathbf{A}_1$ , although it will result in a suboptimal set of bands. In the following experiments we exhaustively searched the two initial-optimal bands, due to the fact that a suboptimal initialization could easily decline the accuracy of the spectral estimation algorithms.

#### 4.4.5 Correlation Measure (CorM)

One of the standard measures of band similarity is normalized correlation (Duda et al. [34]). The normalized correlation metric is a statistical measure that performs well if a signal-to-noise ratio is large enough. This measure is also less sensitive to local mismatches because it is based on a global statistical match. The correlation based band ordering computes the normalized correlation measure for all adjacent pairs of bands similar to the spatial autocorrelation method (Warner et al. [37]). After selecting the first least correlated band based on all adjacent bands, the subsequent bands are chosen as the least correlated bands with the previously selected bands. This type of ranking is based on mathematical analysis of Jia and Richards et al. [38] where spectrally adjacent blocks of correlated bands are represented in a selected subset.

#### 4.4.6 Sparse Principal Component Analysis of a constant-rank matrix

The principal components (PCs) of a set of observations on  $N$  variables capture orthogonal directions of maximum variance and offer a distance-optimal, low-dimensional representation that for many purposes conveys sufficient amount of information. Without additional constraints, the PCs of a data set can be computed in polynomial time using the eigenvalue decomposition. A disadvantage of conventional PCA is that, in general, the extracted components are expected to have nonzero elements in all their entries, therefore, PCA does not provide information about the source of the

principal components. On that account, George Karystinos et al. [39] and Megasthenis Asteris proposed a non NP-hard sparse PCA algorithm that computes the most  $k$ -sparse principal components alongside with the their support vectors in real time. Authors proved that the sparse principal components of a  $p \times p$  correlation matrix  $\mathbf{C}$  can be obtained in polynomial time under a new sufficient condition: when  $\mathbf{C}$  can be written as a sum of a scaled identity matrix and a positive semidefinite update, i.e.,  $\mathbf{C} = \sigma \mathbf{I}_p + \mathbf{A}$ , and the rank  $D$  of the update  $\mathbf{A}$  is not a function of the problem size. Under this condition, they show that sparse PCA is solvable with time complexity  $O(p^{D+1})$ . The  $\mathbf{C}$  matrix is defined as the auto-correlation matrix of the  $\mathbf{B}$  matrix, which is a  $N \times p$  matrix, where  $N$  is the number of pixels of the spectral cube,  $p$  is the number of observations-wavelengths, and  $k$  is the number of bands we want to isolate as the most informative and distinctive. Because the procedure of computing the sparse PCA is complex and available in paper [39], it is omitted. Nevertheless, intuitively, the original problem of the  $k$ -sparse PCA is formed as

$$X_{opt} = \underset{\|x\|=1, \|x\|_0 \leq k}{\operatorname{argmax}} \{x^T \mathbf{C} x\}. \quad (4.6)$$

Solving the original problem, it is needed to find the  $k$ -row sub-matrix of  $\mathbf{V}$ , where  $\mathbf{V}$  is given by the decomposition of  $\mathbf{A}$  ( $\mathbf{A} = \mathbf{V}\mathbf{V}^T$ ) whose principal singular value ( $\sigma$ ) is the maximum one among all sub-matrices. The indices that are contained in the optimal support  $I_{opt}$ , where  $I_{opt}$  is given by

$$I_{opt} \triangleq \underset{I \subseteq [N], |I|=k}{\operatorname{argmax}} \{\sigma_{max}(\mathbf{V}_{I,:})\} \quad (4.7)$$

corresponding to the nonzero loadings of the solution  $X_{opt}$ . Note that  $\mathbf{V}_{I,:}$  is the left singular vector.

#### 4.4.7 BS (Band Selection Algorithm)

It can be easily observed that none of the above algorithms are specific for spectral estimation approaches. The described methods search for the best subset of bands that are informative and distinctive, but this is not enough for the spectral estimation approach. The spectral estimation procedure is based on estimation algorithms and on their performance. Each spectral estimation method is constructed using different criteria and mathematical formulas, so different methods perform better using different subset of bands.

Furthermore, all the above band selection methods and especially the most well-known (OSP, LP) obtain the initial two bands either using geometry projections or using exhaustive methods like experiments in [35]. According to [35], geometry projection methods return sub-optimal initial bands, something that effects negatively the performance of the spectral estimation algorithms.

Moreover, PCA and Correlation methods locate the bands using statistical methods under low SNR conditions. However, these methods usually change the physical meaning of the original data, due to the fact that the channels in the low-dimensional space do not correspond to individual original bands but their linear combinations.

Hence, we need an algorithm individual of PCA and Correlation methods, specific for spectral estimation aim without need of initialization.

All the above requirements are satisfied by BS. BS is a band selection algorithm specific for spectral estimation. Furthermore, the most important achievement is that it works without need of searching for initial bands. The BS algorithm works from high dimensional space to a lower. Specifically, the algorithm uses all the bands that

a spectral cube contains as initial bands. Then it locate the band  $j$ , which can be subtracted from the entire set of bands and will not affect negatively the reconstruction of the spectral cube, using a spectral estimation algorithm. In other words, in each step a band that it does not raise significantly the RMSE of reconstruction of the spectral cube is subtracted from the initial set. The algorithm returns a set with the  $k$  most important bands, capable for spectral estimation. The following pseudo-code describes the BS method analytically. BS takes as input parameter a  $N \times p$  matrix that contain  $N$  pixel of the spectral cube and their reflectance values in  $p$  wavelengths. The time complexity of this algorithm is in the worst case  $O(p^2)$ , where  $p$  is the number of wavelengths that are contained in the spectral cube.

---

**Algorithm 1** BS Algorithm
 

---

```

1: procedure BS( $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p] \in \mathbb{R}^{N \times p}$ , number of bands to keep  $k$ )
2:    $\mathbf{B}^{tmp} = \{\mathbf{B}\}$ 
3:   RMSE = 0
4:   index = 0
5:   for  $i$  in range  $1 : p - k$  do
6:     maxRMSE =  $\infty$ 
7:     for  $j$  in range  $1 : (p + 1) - i$  do
8:        $\mathbf{v} = \mathbf{B}^{tmp} - \{\mathbf{B}_j^{tmp}\}$ 
9:
10:      EstimatorMatrix = SpectralEstimationAlgorithm( $\mathbf{v}, \mathbf{B}$ )
11:      FinalEstimation =  $\mathbf{v} * \mathbf{EstimatorMatrix}$ 
12:      RMSE = CalculateRMSE(FinalEstimation,  $\mathbf{B}$ )
13:
14:      if RMSE < maxRMSE then
15:        maxRMSE = RMSE;
16:        index =  $j$ 
17:
18:      end if
19:    end for
20:     $\mathbf{B}^{tmp} = \mathbf{B}^{tmp} - \{\mathbf{B}_{index}^{tmp}\}$ 
21:  end for
22:  return  $\mathbf{B}^{tmp} \in \mathbb{R}^{N \times k}$ 
23: end procedure

```

---

This algorithm returns a sub-optimal subset of bands, due to the fact that the subtraction of a band for the initial set take place without prediction of the next-subtracted band. On this way, algorithm may exclude the  $i^{th}$  band from the initial set, but the spectral estimation may be more considerable whether after the exclusion of the  $j^{th}$  next-subtracted band, the remain subset of bands included the  $i^{th}$  band gives better spectral estimation than the returned subset of the BS. Nevertheless, the heuristic BS algorithm is adapted for spectral estimation purposes and the experiments shew that the method returns more capable bands to augment the spectral estimation accuracy compare with other methods.

#### 4.4.8 Band selection using Linear Interpolation (LIBS)

Another idea to fulfill the aim of the band selection is to consider as features, the bands of the spectrum that can be united using linear interpolation technique, and that union yields the reconstruction of the spectrum with the minimum error. In

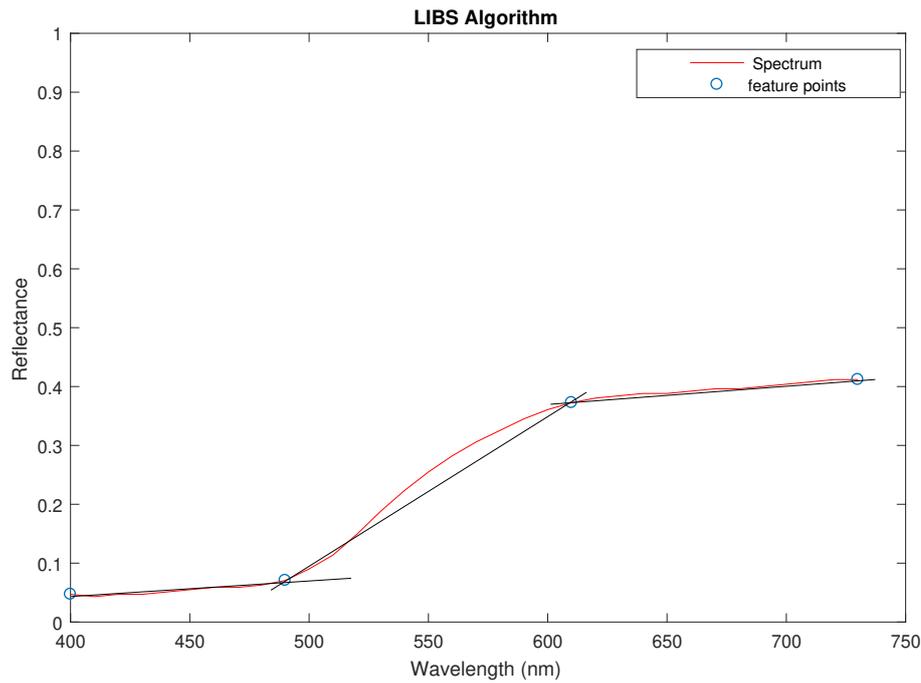


FIGURE 4.1: LIBS algorithm.

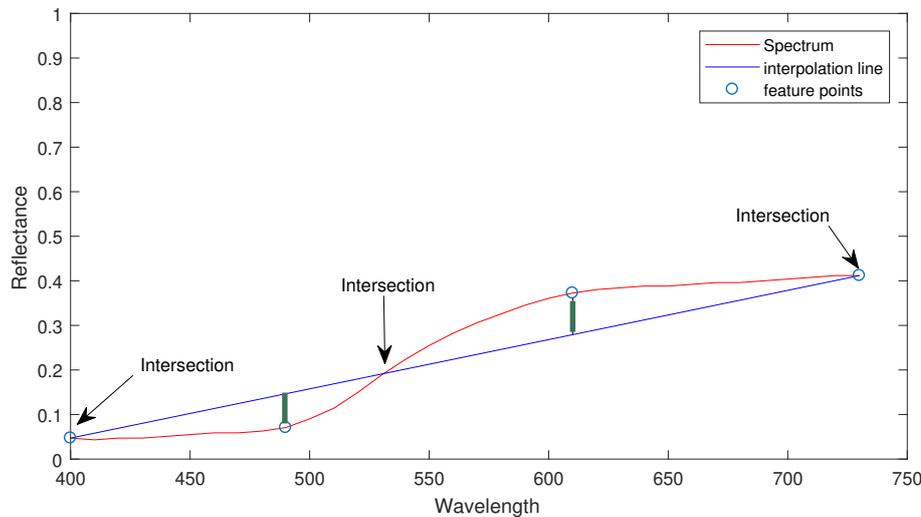


FIGURE 4.2: Procedure of finding feature points.

Chapter 2 we shew that a way to reconstruct the spectrum can be performed using any interpolation operator among -captured bands. For instance, such bands are those that are characterized in Figure 4.1 as points. The red line illustrates the reference spectrum and the black line the reconstruction of the spectrum using linear interpolation between these points.

Considering that, a new approach for band selection was implemented. Each band is voted positive whether the algorithm find a point that lies on this band. Then, the method counts all the votes and elects as more bands the  $k$  bands that were voted by the majority pixels of the spectral cube. That procedure is known as *majority voting*. The way that the algorithm follows to detect all the point is illustrated in Figure 4.2.

Firstly, the method calculates the line that passes through the first and the last band using the following model:

$$y - I(b_1) = \frac{I(b_p) - I(b_1)}{b_p - b_1}(x - b_1) \quad (4.8)$$

where  $I$  is the intensity value,  $b_1$  is the first band of spectrum and  $b_p$  is the last band of the spectrum. We have observed that we reconstruct more accurately the spectrum, whether the first and the last bands of the spectrum are parts of the selected subset of bands. Then, we detect the intersection points of the line with the reference spectrum. Between each pair of intersection points, the band that has the maximum projection on the line (green vertical line in Fig. (4.2)) is considered as the most feature point. Observing Fig. (4.1), linear interpolation among these bands can reconstruct accurately the spectrum. This procedure is repeated for each pixel of the spectral cube and the algorithm returns the  $k$  most bounds of the spectrum using the majority voting technique. The steps of algorithm are described below:

1. For each spectrum of each pixel of the spectral cube, find the line that passes through the first and the last band.
2. Detect the intersection points of the line with the reference spectrum
3. Between each pair of intersection points, find the band that has the maximum projection on the line
4. Adding +1 vote for each band, including the first and the last band.
5. At the end, count the votes of each band and elect the  $k$  most high voted sets, as the most .

## 4.5 Practical Considerations

1. **Number of Pixels Involved in the Band Selection Process:** The size of a spectral cube is very large due to the large number of pixels in an image. Fortunately, authors of [35] found out that using a small subset of pixels in the band selection process will not change the results in most cases. This is because each band image is spatially highly correlated. Therefore, in the experiments, we randomly choose 10% or 1% of  $N$  pixels for band similarity assessment.
2. **Number of Bands to be Selected:** In practice, it is difficult to know how many bands should be selected. It is more than sure that by increasing the number of the selected bands, the estimation accuracy will be augmented. The number of selected bands is depend on the optical limitations of the hardware and on the application that is studied.

## 4.6 Experiments for evaluation the band selection methods

### 4.6.1 Information about the experiments

To evaluate the performance of the band selection algorithms, a real band selection procedure took place. The effectiveness of each method was evaluated according to the preciseness of estimation using the  $k$  bands as features. In this experiment,

the Macbeth SG ColorChecker was used to extract the six most informative and distinctive bands and the Munsell's patches as target of the estimation. The scoring of each band selection method achieved using Root Mean Square Error (RMSE) and Goodness of Fit (GFC) metrics. Specifically, max, mean, and min RMSE and mean GFC are presented for each band selection method in Table 4.2. In Table 4.1 the selected bands from each method were depicted. Fig. (4.3) shows the success percentage of the different band selection methods and Fig. (4.4) illustrates the total ranking.

Method	Selected Bands
Entropy	[600nm, 620nm, 630nm, 650nm, 660nm, 720nm]
CorM	[440nm, 490nm, 550nm, 700nm, 710nm, 720nm]
PCAr	[560nm, 570nm, 580nm, 590nm, 670nm, 690nm]
OSP	[430nm, 500nm, 560nm, 610nm, 650nm, 710nm]
BS	[430nm, 480nm, 560nm, 600nm, 650nm, 700nm]
LP	[430nm, 500nm, 550nm, 600nm, 660nm, 730nm]
LIBS	[400nm, 500nm, 530nm, 560nm, 600nm, 730nm]
Sparse PCA	[440nm, 550nm, 560nm, 570nm, 630nm, 700nm]

TABLE 4.1: Selected bands from the band selection algorithms.

Method	MEAN RMSE(%)	MAX RMSE(%)	MIN RMSE(%)	MEAN GFC
Entropy	8.053	32.74	0.5160	0.9620
CorM	2.1343	9.0090	0.3816	0.9950
PCAr	6.1260	28.7439	0.6102	0.9631
OSP	1.4578	4.7359	0.2279	0.9982
BS	1.4896	4.594	0.3143	0.980
LP	1.5369	5.4226	0.3448	0.9978
LIBS	2.44	8.9687	0.1749	0.9963
Sparse PCA	2.6187	10.5467	0.5187	0.9932

TABLE 4.2: Metrics of Error for the Band Selection Experiments.

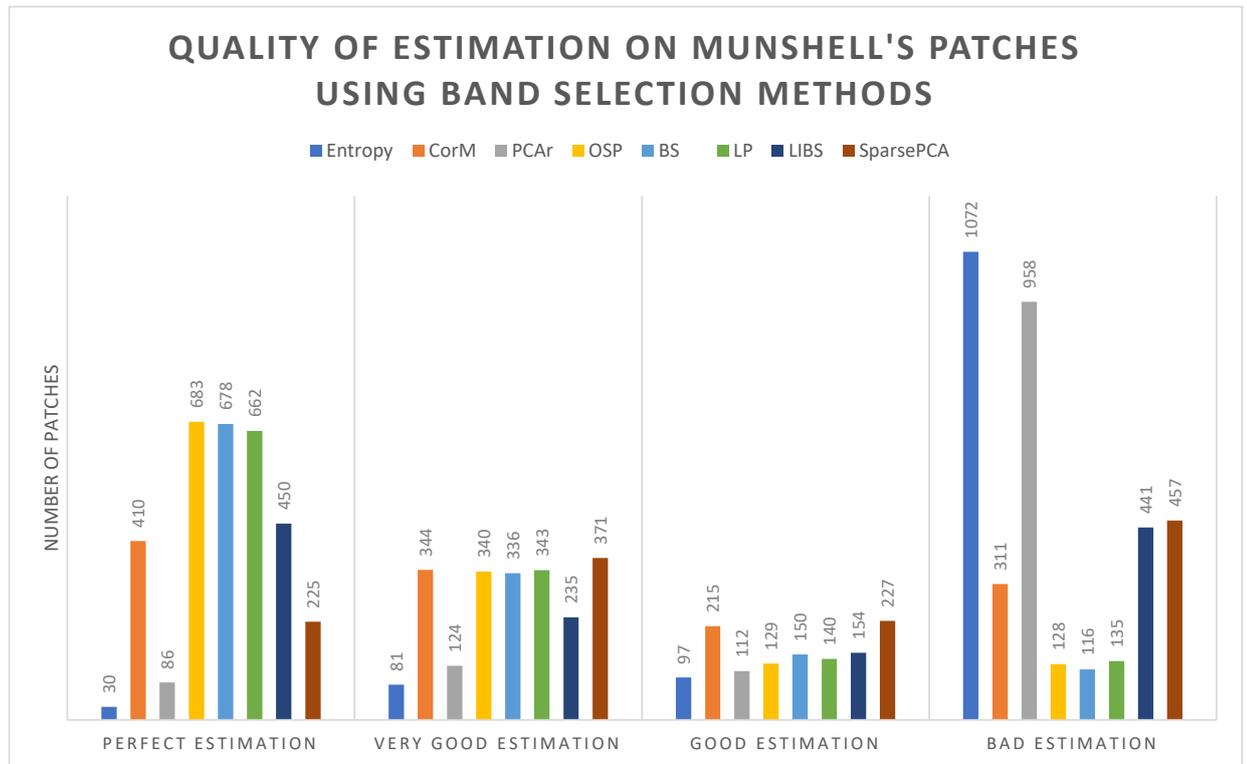


FIGURE 4.3: Comparison of Band Selection Methods.

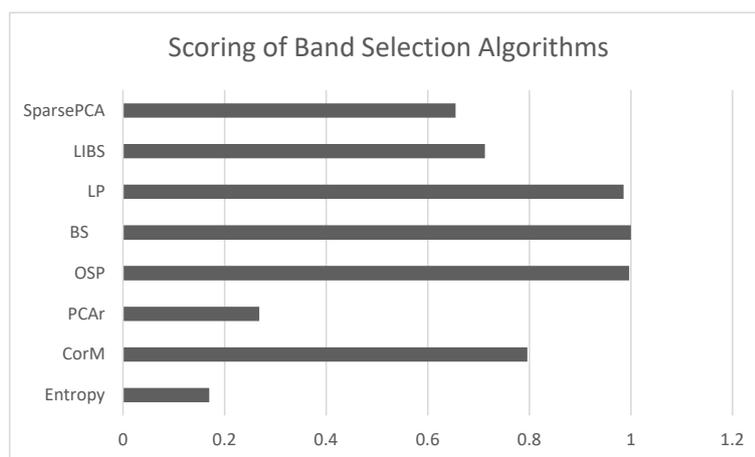


FIGURE 4.4: Ranking of Band Selection Methods.

### 4.6.2 Conclusions of the Band Selection Experiments

The above tables and charts provide an analysis of the performance among different band selection algorithms using Linear Regression as spectral estimation model.

Overall, it can be seen in Fig. (4.3) that BS, OSP, and LP are the best scored methods that produce the most considerable spectral estimations. On the other hand, Entropy and PCAr get the lowest rank in scoring table, therefore, they did not raise up the estimation preciseness.

It is clear from Table 4.1 that Entropy selects the bands around 600nm and 720nm, something that is expected due to the fact that Entropy finds the most informative parts of data. As a result, spectrum between 600nm and 720nm is the most informative part of the visible spectrum but not the most distinctive. PCAr method gives more informative and distinctive bands since the method gives a rank for each band according to both the eigenvectors and eigenvalues. CorM method raises significantly the perfectly estimated patches from 86 of PCAr to 410, as it is illustrated in Fig. (4.3), and declines the mean RMSE error from 6.12% of PCAr to 2.13%. Moreover, the selected bands of Sparse PCA and LIBS are more capable for spectral estimation purposes than those of PCAr, but they did not move up the mean GFC compared with CorM. Nevertheless, it comes without doubt that PCAr, CorM, and Sparse PCA change the physical meaning of the original data, something that effects negatively the spectral estimation accuracy. In the opposite direction, OSP, LP, and BS improve the spectral estimation. OSP gives 683 perfectly estimated patches and 128 badly estimated compared with LP that produces 135 badly estimated and bigger mean RMSE. The projection of OSP on a orthogonal subspace is the key of its success, due to the fact that combines similarity based metrics with geometric approaches. Slightly up, with the highest score is the BS algorithm. The novelty and the success of the proposed method stems from the manner of the algorithm. Specifically, BS selects bands that can reconstruct accurately a spectral cube and for that reason it achieved the lowest estimation error among the other methods. Using the bands of the BS algorithm we can estimate the half of the Munsell database perfectly and only 116 out of 1269 patches badly. Note that BS has bigger mean RMSE than OSP but the max RMSE is the lowest compare to the other methods.

Ending up, BS method is the most suitable band selection method for the spectral estimation aim.

## Chapter 5

# Conclusions and Future Work

Throughout this thesis we studied and compared methods of spectral estimation alongside with band selection methods.

In Chapter 1, a plenty information about spectroscopy and hyper-spectral imaging were provided. Furthermore, we shew that simple exposure imagers reveal a major disadvantage. When we have phenomena that are changing on a time scale that is sorter than the duration required for recording the spectral cube, the HS system cannot perform accurately. On that account, a need for a snapshot spectral imager that acquires in appropriate time the spectral data was revealed. Then, we utilized spectral estimation methods due to the fact that snapshot imagers cannot obtain a huge number of wavelengths without rapid loss of the spatial resolution.

In Chapter 2, all the well-known spectral estimation methods were compared according to the quality of their estimation using RMSE and GFC metrics. A new algorithm called (Adaptive) Hybrid proposed by us, in order to overcome shortcomings of the other methods.

In Chapter 3, detailed experiments were carried out using Munsell and Macbeth ColorChecker as training and validation set and vise versa. Then, a scoring of the algorithms took place to evaluate the spectral accuracy. It turned out that Linear Regression is the best candidate algorithm for spectral estimation approaches whether the execution time of the estimation should be minimized and high estimation performance is demanded. When the training set is small, Hybrid method achieves more considerable estimation than the Linear Regression. Adaptive Hybrid seems to gives the lowest spectral reconstruction error but its time complexity is undeniable higher compare with Linear Regression. (Adaptive) Hybrid algorithm is more efficient than the Wiener model, something that was indicated in experiments.

In Chapter 4, we strived to optimize the performance of the spectral estimation methods by leveraging band selection techniques in order to select as features the most informative and distinctive bands. A great amount of band selection algorithms were analyzed, which based either on geometric approaches or on similarity formulas. Additionally, a new promising band selection algorithm called BS proposed by us and it turned out that BS performs more accurately than all the other methods. BS algorithm is based on finding the most informative and distinctive bands, but alongside, bands that will provide errorless spectral estimation using any spectral estimation method.

Since Linear Regression and (Adaptive) Hybrid need small time period for processing, we can create a system that will have six band pass filters and then have online

processing to create a spectral cube in real time. That can also be used for mobile applications (since spectral imaging nowadays is mostly for static applications). This is extremely hopeful since in the present moment there is no miniaturized hyperspectral imaging device that can provide real time results.

# Bibliography

- [1] “Quantative bioimaging laboratory,(qbil)”, [Online]. Available: [http://feilab.org/Research/Research\\_HSI.htm](http://feilab.org/Research/Research_HSI.htm).
- [2] “Physicstuff”, [Online]. Available: <http://physicstuff.com/tag/3d/>.
- [3] “Astronomy,cosmos”, [Online]. Available: <http://astronomy.swin.edu.au/cosmos/S/Spectroscopy>.
- [4] “Hyperspectral imaging”, [Online]. Available: <http://www.markelowitz.com/Hyperspectral.html>.
- [5] “Wikiversity”, [Online]. Available: [https://en.wikiversity.org/wiki/Electromagnetic\\_wave](https://en.wikiversity.org/wiki/Electromagnetic_wave).
- [6] Y. Wang, N. P. Reder, S. Kang, A. Glaser, and J. Liu, “Multiplexed optical imaging of tumor-directed nanoparticles: A review of imaging systems and approaches”, vol. 1, pp. 369–388, Aug. 2017.
- [7] G. Lua and B. Feia, “Medical hyperspectral imaging : A review”, 2014.
- [8] “Dysis company”, [Online]. Available: <https://dysismedical.com/products/colposcope/>.
- [9] C. S. Chane, A. Mansouri, F. S. Marzani, and F. Boochs, “Integration of 3d and multispectral data for cultural heritage applications: Survey and perspectives”, *Image and Vision Computing*, vol. 31, no. 1, pp. 91–102, 2013, ISSN: 0262-8856. DOI: <https://doi.org/10.1016/j.imavis.2012.10.006>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885612001862>.
- [10] P. Stigell, K. Miyata, and M. Hauta-Kasari, “Wiener estimation method in estimating of spectral reflectance from rgb images”, *Pattern Recognition and Image Analysis*, vol. 17, pp. 233–242, 2007.
- [11] K. P. F.R.S., “Liii. on lines and planes of closest fit to systems of points in space”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901. DOI: [10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720). eprint: <https://doi.org/10.1080/14786440109462720>. [Online]. Available: <https://doi.org/10.1080/14786440109462720>.
- [12] A. Mansouri, T. Sliwa, J. Y. Hardeberg, and Y. Voisin, “Representation and estimation of spectral reflectances using projection on pca and wavelet bases”, *Color Research & Application*, vol. 33, no. 6, pp. 485–493, DOI: [10.1002/col.20442](https://doi.org/10.1002/col.20442). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/col.20442>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/col.20442>.
- [13] G. D. T. of communication. *J IEE* 1946;93:429–457,

- [14] H.-L. Shen, P.-Q. Cai, S.-J. Shao, and J. H. Xin, "Reflectance reconstruction for multispectral imaging by adaptive wiener estimation", *Opt. Express*, vol. 15, no. 23, pp. 15 545–15 554, Nov. 2007. DOI: [10.1364/OE.15.015545](https://doi.org/10.1364/OE.15.015545). [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-23-15545>.
- [15] H.-L. Shen, J. H. Xin, and S.-J. Shao, "Improved reflectance reconstruction for multispectral imaging by combining different techniques", *Opt. Express*, vol. 15, no. 9, pp. 5531–5536, Apr. 2007. DOI: [10.1364/OE.15.005531](https://doi.org/10.1364/OE.15.005531). [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-9-5531>.
- [16] "Random forest", [Online]. Available: <http://www.assignmentpoint.com/science/computer/random-forest.html>.
- [17] "Numerical demultiplexing of color image sensor measurements via non-linear random forest modeling", vol. 27, Jun. 2016.
- [18] A. Mansouri, F. Marzani, and P. Gouton, *Neural networks in two cascade algorithms for spectral reflectance reconstruction*, Oct. 2005.
- [19] W.-F. Zhang and D.-Q. Dai, "Spectral reflectance estimation from camera responses by support vector regression and a composite model", *J. Opt. Soc. Am. A*, vol. 25, no. 9, pp. 2286–2296, Sep. 2008. DOI: [10.1364/JOSAA.25.002286](https://doi.org/10.1364/JOSAA.25.002286). [Online]. Available: <http://josaa.osa.org/abstract.cfm?URI=josaa-25-9-2286>.
- [20] C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, ISSN: 1573-0565. DOI: [10.1023/A:1022627411411](https://doi.org/10.1023/A:1022627411411). [Online]. Available: <https://doi.org/10.1023/A:1022627411411>.
- [21] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets", *Optimization Methods and Software*, vol. 1, no. 1, pp. 23–34, 1992. DOI: [10.1080/10556789208805504](https://doi.org/10.1080/10556789208805504). eprint: <https://doi.org/10.1080/10556789208805504>. [Online]. Available: <https://doi.org/10.1080/10556789208805504>.
- [22] R. Fletcher, *Practical Methods of Optimization; (2Nd Ed.)* New York, NY, USA: Wiley-Interscience, 1987, ISBN: 0-471-91547-5.
- [23] "Macbeth color checker", [Online]. Available: <http://photonet.com>.
- [24] "Munsell database", [Online]. Available: <https://www.uef.fi/web/spectral/munsell-colors-matt-spectrofotometer-measured>.
- [25] "Muses 9hs", [Online]. Available: <http://www.spectricon.com/products/muses-9-hs/>.
- [26] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification", *IEEE Trans. Geoscience and Remote Sensing*, vol. 37, pp. 2631–2641, 1999.
- [27] A. Ifarraguerri and M. W. Prairie, "Visual method for spectral band selection", *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 2, pp. 101–106, Apr. 2004, ISSN: 1545-598X. DOI: [10.1109/LGRS.2003.822879](https://doi.org/10.1109/LGRS.2003.822879).
- [28] S. D. Backer, P. Kempeneers, W. Debruyne, and P. Scheunders, "A band selection technique for spectral classification", *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 3, pp. 319–323, Jul. 2005, ISSN: 1545-598X. DOI: [10.1109/LGRS.2005.848511](https://doi.org/10.1109/LGRS.2005.848511).

- [29] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection", *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 7, pp. 793–802, 2004, ISSN: 0099-1112. DOI: [doi:10.14358/PERS.70.7.793](https://doi.org/10.14358/PERS.70.7.793). [Online]. Available: <https://www.ingentaconnect.com/content/asprs/pers/2004/00000070/00000007/art00001>.
- [30] S. Cai, Q. Du, and R. J. Moorhead, "Hyperspectral imagery visualization using double layers", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3028–3036, Oct. 2007, ISSN: 0196-2892. DOI: [10.1109/TGRS.2007.894922](https://doi.org/10.1109/TGRS.2007.894922).
- [31] Q. Du, H. Ren, and C.-I. Chang, "A comparative study for orthogonal subspace projection and constrained energy minimization", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1525–1529, Jun. 2003, ISSN: 0196-2892. DOI: [10.1109/TGRS.2003.813704](https://doi.org/10.1109/TGRS.2003.813704).
- [32] J. C. Russ and F. B. Neal, *The Image Processing Handbook*, 7th. Boca Raton, FL, USA: CRC Press, Inc., 2015, ISBN: 149874026X, 9781498740265.
- [33] B. Campbell, *Introduction to Remote Sensing, Second Edition*. The Guilford Press, New York, 1996.
- [34] D. G. S. Richard O. Duda Peter E. Hart, *Pattern Classification, 2nd Edition*.
- [35] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis", *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 564–568, Oct. 2008, ISSN: 1545-598X. DOI: [10.1109/LGRS.2008.2000619](https://doi.org/10.1109/LGRS.2008.2000619).
- [36] D. C. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, Mar. 2001, ISSN: 0196-2892. DOI: [10.1109/36.911111](https://doi.org/10.1109/36.911111).
- [37] T. A. Warner, K. Steinmaus, and H. Foote, "An evaluation of spatial autocorrelation feature selection", *International Journal of Remote Sensing*, vol. 20, no. 8, pp. 1601–1616, 1999. DOI: [10.1080/014311699212632](https://doi.org/10.1080/014311699212632). eprint: <https://doi.org/10.1080/014311699212632>. [Online]. Available: <https://doi.org/10.1080/014311699212632>.
- [38] X. Jia and J. A. Richards, "Efficient maximum likelihood classification for imaging spectrometer data sets", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 2, pp. 274–281, Mar. 1994, ISSN: 0196-2892. DOI: [10.1109/36.295042](https://doi.org/10.1109/36.295042).
- [39] M. Asteris, D. S. Papailiopoulos, and G. N. Karystinos, "The sparse principal component of a constant-rank matrix", *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2281–2290, Apr. 2014, ISSN: 0018-9448. DOI: [10.1109/TIT.2014.2303975](https://doi.org/10.1109/TIT.2014.2303975).