Doctoral Dissertation
Doctoral Program in ELECTRICAL, ELECTRONICS AND COMMUNICATIONS ENGINEERING (33th Cycle)

Landmine detection using hyper-spectral imaging

Mahdi Khodor

* * * * * *

Supervisors
Prof. Tiziano.Bianchi.,Supervisor
Prof. Massimo.Zucchetti, Co-Supervisor

Doctoral Examination Committee:
Prof. Farid. Melgani., Reviewer, University of Trento
Prof. Paul. Honeine., Reviewer, University of Rouen
Prof. Mohamad. Khalil., Examiner, Lebanese University
Prof. Enrico. Magli., Examiner, Politecnico di Torino University
Prof. Fabrizio. Argenti., Examiner, Degli Studi Di Firenze University
Prof. Rafic.Younes., Supervisor, Lebanese University
Dr. Jihan.Khoder., Co-Supervisor, Lebanese University

Lebanese University
March 04, 2021

Acknowledgment

I would like to express my deep gratitude to my supervisors Prof. Rafic Younes, Dr. Jihan Khoder, Prof Tiziano Bianchi and Prof. Massimo Zucchetti for their continuous support, kindness, and availability during the three years of work on the thesis. I would like to thank the external reviewers Prof. Farid Melgani and Prof. Paul Honeine for accepting the participation in the final dissertation and for their valuable comments that helped me in improving the quality of the report. I want to thank also the examiners Prof. Mohamad Khalil and Prof. Fabrizio Argenti and Prof. Enrico Magli for their positive contribution in the jury.

Also, I would like to thank the Lebanese army for providing real samples of landmines extracted from minefields in Lebanon. These samples were used in the collection of spectra of landmines in different conditions. Also, I would like to thank Dr. Ihab Makki for his help in the acquisition of the landmines' spectra using the spectroradiometer.

A special thanks goes to AZM center and Erasmus Committee for the financial support in Lebanon and in Italy.

Finally, the constant support of my family, friends and colleagues allowed me to successfully conclude this PhD.

Thank you

Abstract

This PhD thesis explores the possibility to detect landmines utilizing hyperspectral imaging. By using this innovation, we can get at every pixel of the image spectral information in many wavelengths. In this way, at every pixel we acquire a reflectance range that is utilized as a unique signature to recognize the materials of that pixel, and essentially in our project help us to identify the presence of landmines.

The proposed procedure functions as follows: a preconfigured drone will convey the hyperspectral camera. This customized drone is responsible of flying over the zone to be cleared of landmines to acquire image from a safe distance. Different image processing tools will be used to detect the landmine from the image. When the presence of a mine or explosives is suspected, an alert sign is sent to the base station giving data about the landmine type, its location and the clear path that could be taken by the mine removal team for demining process.

There are several reasons why hyperspectral imaging advantages over other demining methods has: first, this method is safer than the other methods due to the absence of persons in the demining area. Second, using this method we can scan a large area in a single day. Finally, this technique is able to identify each type of landmine even when the difference between landmines is small.

First, a brief presentation that focus on the problem of landmine referring to some statistics taken from the UN organization is addressed. Also, different types of used landmines are presented. In addition, dimensionality reduction in hyperspectral imaging is presented. A summary of all detection techniques used in hyperspectral imaging is presented with the advantages and disadvantages of each method. How to improve the performance of existing detection methods in terms of complexity and detection rate is then discussed.

In this work, we study the importance of dimensionality reduction in hyperspectral imaging, by developing a new bands selection method. Also, we give a summary about detection techniques in hyperspectral imaging, identifying what are the most performing methods that give high detection probability with low false alarm rate. These methods are tested in different scenarios with different statistical and linear unmixing based methods. In addition, we introduce the use of pixel intensity

methods in order to detect rare pixels at subpixel level: this will reduce the number of pixels in the images, meanwhile reducing the complexity and improving the performance of different existing target detection methods.

An investigation of the impact of dimensionality reduction using multicriteria and net analyte signal methods are also provided. A comparative study was made based on computational time and detection performance after image reduction.

In a field experiment, we study how the spectral signature of landmines is affected depending on different environmental factors affecting the area where the landmine is planted (soil, grass, etc.). We take the signature of different landmine types in different conditions:

- In the lab
- Landmine covered by grass
- Landmine covered by soil

Also, a hyperspectral image is acquired from a real area containing mines buried in different conditions in either soil or grass. Dimensionality reduction and target detection are applied to this image to study the possibility to detect landmine in a realistic scenario. Finally, a brief introduction about thermal infrared hyperspectral imaging, we discuss the advantages of this technique to detect fully buried objects, especially landmines in different scenarios.

Contents

List of figures

List of tables

Chapter 1

Introduction to landmine problems

1.1 Introduction

Landmines and cluster munitions are a major barrier to society growth and return to normal life after a ceasefire has been reached. With the end of the war, fear of death and destruction will not stop, but will continue with the existence of cluster munitions, landmines, unexploded ordnance and improvised explosive devices. This type of weapons does not know when the war ends and stay active for years or even decades threatening innocent people in their daily lives.

Recent statistics [1] show that 80% of deaths caused by landmines involve kids who have nothing to do with the war or its causes. Therefore, the use of this sort of blind guns must be prohibited. Efforts have begun to ban landmines and we have a global campaign to ban landmines and cluster ammunition with numerous signing countries [1].

During a conflict, however, it is very difficult to apply any regulation and landmines have been used in latest clashes (e.g. in 2016 in both Libya and Syria) [1]. Therefore, it is necessary to discover quick and reliable detection methods to remedy this situation.

To detect landmines, various methods have been discussed. Each technique has its benefits and disadvantages. The metal detector is one of the oldest and most used techniques.

This sort of detectors can detect items that contain metal under the soil due to electrical induction events. Although this method is inexpensive, it has several drawbacks: it detects all metals, either landmines or inert metals, so it has very high false alarm rates; in addition, recent landmines contain fewer metals, making them more difficult to detect. Other landmine detection methods will be discussed in the next sections.

1.2 Landmines issues and current solutions

1.2.1 Landmines problems

In this section we present an overview of the different problems caused by landmines, including their legacy on affected regions.

The social and economic growth of the regions experiencing the presence of landmines has been severely affected by this issue: not only innocent people are killed by landmines, but also the areas being cultivated are reduced [1, 2] [3]. As is known, landmines are responsible for hundreds of thousands deaths or limb amputations. In Cambodia, for example, landmine explosions have caused more than 35,000 amputees [2, 3] and many of the wounded were killed due to lack of transportation to hospitals [3]. Moreover, mines may reduce the area being exploited in a country, discourage tourism, and contribute to economic regression.

There are several barriers preventing the removal of deployed landmines, such as the loss or lack of maps and information regarding the types of landmines used or where these were originally found, the shift of landmines due to climate and physical factors, the broad variety of landmines used and the high cost of detecting and removing landmines. Landmines are also very harmful to individuals because they can explode even after a long time in adverse environmental conditions.

Mines in an enemy battle can still kill or harm civilians decades later. Landmines, ammunition clusters and other explosive remnants of war (ERW) continue to kill or damage at least 4,300 individuals annually [1]. According to landmine monitor report in 2018, compared to the military and safety personnel, civilians accounted for the vast majority of victims, 87% were civil victims in 2017, where the status was recognized.

Table 1 States with the most recorded Mine/ERW causalities

Country	Casualities
Afghanistan	2,300
Syria	1,906
Ukraine	429
Iraq	304
Pakistan	291
Nigeria	235
Myanmar	202
Libya	184
Yemen	160

As demonstrated in Table 1 [1], In countries facing civil war and large-scale violence, particularly in Afghanistan and Syria, mines/ERW caused a very high number of casualties in 2017.

Casualties by type of mine/ERW in 2017 are shown in Figure 1

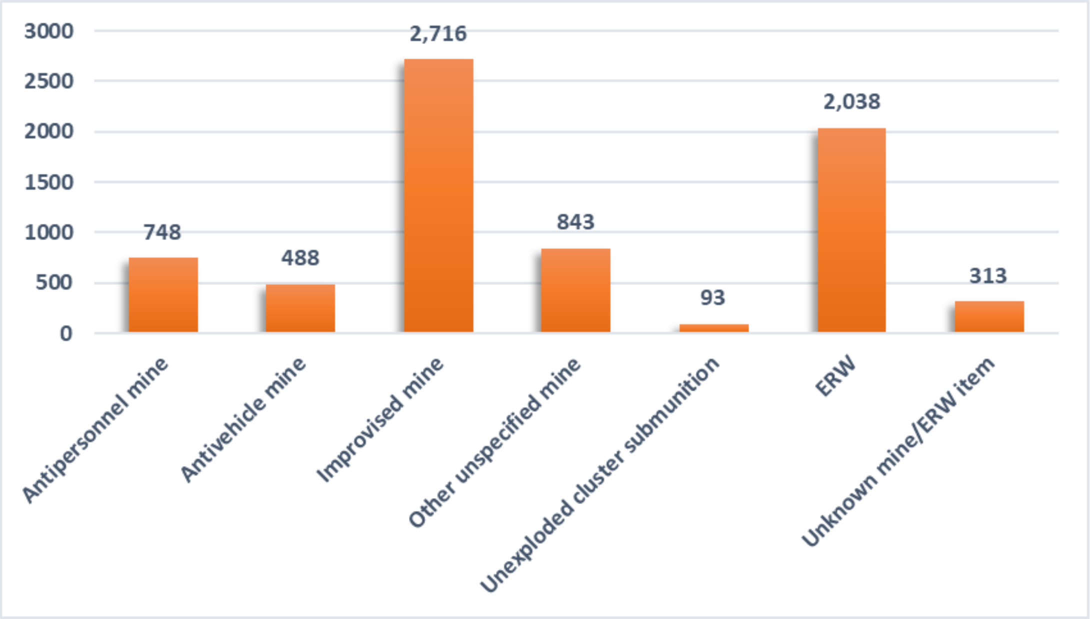


Figure 1 Casualties by type of mine/ERW in 2017

In 2017, 18 countries are listed. During 2017, Most improvised mine casualties occurred in Afghanistan (1093 losses) and Syria (887 losses) suffered the largest number of explosive mine casualties.

862 deaths were registered in the classification of non-specified mine types in 16 countries and regions. About three-quarters (1,208, corresponding to 74 %) of those victims were registered in Syria.

The most reported casualties of anti-vehicle mines were Pakistan (132 losses) and Ukraine (103 losses) in 2017, with at least 488 casualties occurring in twenty states and regions.

When a mine explodes, the effect of the explosion decreases as the distance from the mine increases, since the blast wave has maximum power at the point of the explosion and is dissipated when moving in the atmosphere. For example, compared to a device with a significantly greater hazardous charge, a mine with small quantities of explosives (e.g. a mine under foot) can have high killing power while producing a relatively small damage within a few meters. As a result, the number of wounds from landmines and Unexploded ordnance (UXO) injuries is enormous. Two basic results are mine / UXO wounds. First, they influence the lives of the wounded and their families; secondly, they influence the health base of the affected country. The main financial consequences on the survivors are the deprivation of the ability to earn income to support themselves and their families. This situation may also cause mental harm in addition to the apparent physical injuries. The condition of women can be especially powerless, since the widespread physical harm can severely limit their marriage likelihood. Treatment of mine injuries subtracts resources to the health care system of developing nations, as such injuries usually require several operations to be properly treated. Moreover, such injuries often require a prothesis or a wheelchair, and people also needs intense physiotherapy before they can benefit from a prothesis.

1.2.2 landmines types

Mines can be classified as anti-personnel or anti-tank. Anti-personnel (AP) mines are aimed at individuals, while anti-tank (AT) mines target thrash tanks or other protected vehicles [4].

Anti-tank mines contain a large quantity of explosive (more than 5 kilograms) are wide (usually larger than a man's sole) and are designed to be activated by heavy vehicles. These mines contain enough explosives to kill people in or near the vehicle and destroy the vehicle that runs across them. Anti-Tank mines are usually placed where enemy vehicles are expected: on highways, bridges and roads.

Anti-personnel mines are much easier to activate and are designed to cause injuries to people walking on them. They have less explosive and are much smaller and lighter than anti-tank mines. A typical AP mine could only weigh around 50 grams and be as thin as a packet of cigarettes. Anti-personnel mines are made from various materials and come in all shapes and colors.

While the AP mines can kill a person, they are designed primarily for serious injury, an injured person must be rescued, which takes more time and money for the enemy. Anti-personnel mines can be mounted anywhere, set up in a range of ways stepping on them, pulling the wire or just shaking them. Anti-personnel mines can also explode if an element is destroyed [5].

Two types of AP mines occur in general: explosive mines and fragmentation mines.

Blast mine

Blast mines are buried near the soil surface and are usually activated by pressure. When someone steps on and triggers a blast mine, the mine detonates and causes an explosion shock wave of hot gases moving at an extremely high speed.

Fragmentation mine

These landmines release pieces in all directions or can be positioned in one direction for the transmission of fragments. Such mines can injure people up to 200 meters from the surface. The fragments used are either metal or glass. [6]

Since anti-tank mines are meant for the immobilization or destruction of vehicles, this type of mines are blast mines. A fragmentation anti-tank mine is not necessary.

Some countries and military forces use landmines to secure key installations and the main bases against the enemy. When a military base is constructed in an open environment, it is vulnerable to enemy attacks from all sides. In these situations, landmines are used to restrict open areas and concentrate defense forces on one side.

While landmines may have readily identifiable military uses, they do necessarily lead to civilian casualties due to their purpose, design, and deployment.

Mining neutralization requires training and is still a slow and dangerous operation. Mining is often designed and implemented to make it difficult to detect.

1.2.3 Techniques for landmine detection

The main techniques used for detecting mines are shown in this section. This section is aimed at showing already used techniques and discuss their advantages and disadvantages, comparing their performance to the hyperspectral image technique.

The most common detection method used to detect mines is based on the same techniques developed during World War II and involve people directly. Today the standard deminer's tool kit is much like that of 50 years ago: a metal detector and a prodding instrument.

Various techniques for demining were designed and developed. Each technique may be more or less appropriate depending on the type of mine case, explosive material, and soil.

We will describe the detection techniques currently used while studying the potential of hyperspectral imaging for detecting landmines. This section is not intended to examine the complicated physics principles of how the sensors work, but rather to provide a brief overview of these techniques, to show the strength and limits of hyperspectral imagery over other methods.

In general, most landmines are detected by 3 main units, a landmine signature detector, a signal, or imaging system to organize the information collected and a decision-making unit to determine whether a landmine exists.

Electromagnetic, acoustic, atomic, biological, chemical, or mechanical detectors can be used.

1.2.3.1 Electromagnetic Methods

The deminer stays near the ground and sweeps it slowly through the field being investigated with an electro-magnetic induction detector (EMI) handle (Fig. 2). A current flowing through the first plate, the 'transmission coil,' generates a magnetic field in the ground that changes over time. In addition, this magnetic primary field causes electrical (eddy) flows in metal objects. A weaker secondary magnetic field is produced by the currents from the buried objects. The second spiral, the "receiving spiral," is used to detect voltage changes caused by the secondary magnetic field as indicated. These changes in electrical potential are then converted into an audible signal by the detector. [6]

Figure 2 Landmine detection using metal detector

1.2.3.2 Acoustic/Seismic method

Such methods are unique among methods of detection that detect the mine case, since they are based on mechanical properties and not on electromagnetic characteristics. The A / S technique is used for the detection of landmines by vibrating them with acoustic or seismic waves, respectively, generated and obtained by non-contact transductors. The transmission system may consist of electronic shakers or acoustic loudspeakers. When the recipient senses a reflected energy, an object may be buried.[Fig.3] [7].

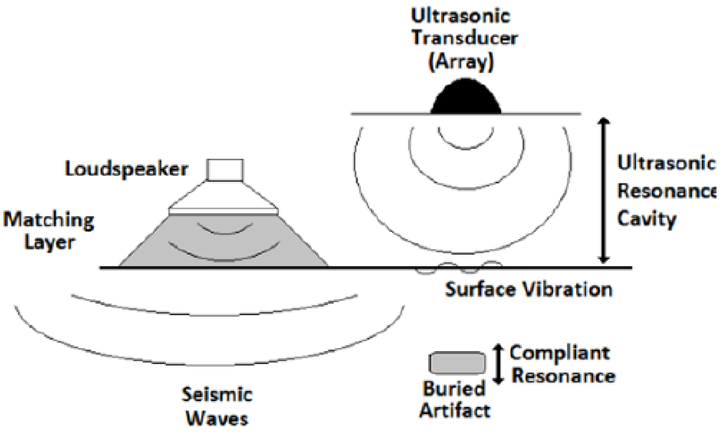


Figure 3 Acoustic/Seismic method for landmine detection

1.2.3.3 Ground Penetrating Radar (GPR)

This method is being established by difficulties in detecting small quantities of metal in a plastic land mine using a metal detector [8]. GPR assesses buried objects by emitting Radio waves (ranging from approximately 10 MHz until several GHz) into the ground and then analyzing wave reflections in the return signals caused by a different index of soil-landmine or soil-broad rock refraction on any subsurface discontinuity. The GPR system (Fig.4) uses an antenna pair (small fixed distance transmitter and receiver) to transmit short bursts of electromagnetic energy into the ground and then record the returned signals. A computerized signal processing system uses an audio image to determine the form or location of the object to interpret the return signal.

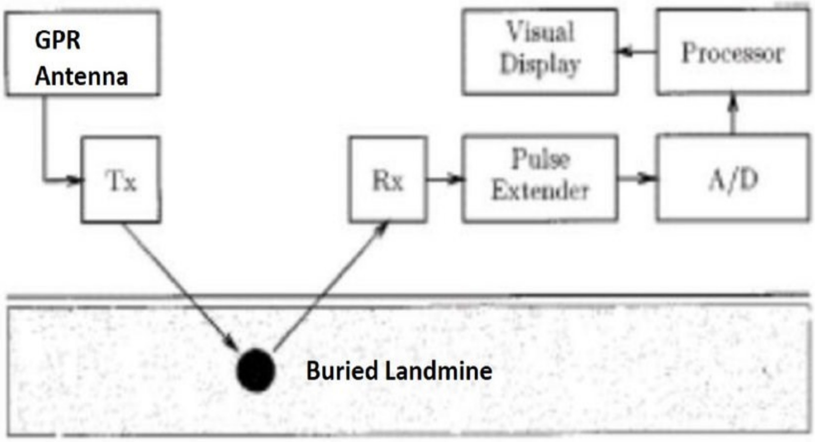


Figure 4 GPR method principle

1.2.3.4 Infrared/Hyperspectral Systems

Infrared radiation in microwave regions consists of a wavelength of 0.7µm to 1 mm. Infrared / Hyperspectral techniques can detect unusual variations in electromagnetic radiation that are reflected or released immediately above buried mines by surface mines or soil and vegetation. Two modes of action include active and passive irradiation using an extended range of electromagnetic wavelengths: A passive IR system detects natural object radiation, and an active system provides a heat source and detects heated object radiation. The methods of thermal detection exploit diurnal temperature fluctuations in areas in the vicinity of mines in contrast. The physical activity of the mining process influences the distribution of natural soil particles by bringing small particles into the air, which affects the way the soil disperses light. There may well also be systematic changes in the moisture level of vegetation immediately over buried mines [9].

1.2.3.5 Vapor sensors

A small proportion of the explosive comes out of the mine, by wind, through cracks and mine structures. The goal is to detect the presence of explosives vapor. In this subject, there are two research lines: biology and chemistry.

The use of mammals, insects, microorganisms, or plants are different biological methods that can be used to detect the presence of explosive substances [7]. Numerous non-biological mechanisms have been studied in recent years to detect low concentrations of explosives in air or in soil samples and to establish highly sensitive odor detection devices. When an explosive specimen moves between the sensor sheets, some of the explosive attaches to the polymer and reduces the amount of fluorescent light one slide releases. A small photovoltaic multiplier indicates the existence of explosive material by detecting this reduction in the intensity of radiation received [7].

1.2.3.6 Nuclear Quadruple Resonance (NQR)

This technique is based on radiofrequency to detect certain chemical compounds such as explosives. This consists of a transmitter that transmits a radiation with the frequency of explosive material resonance. It excites the nucleus of an element and releases further radiation, which induces electric potential at the receptor spindle, when restored to its stable state. By detecting the presence of explosives, the presence of a Landmine is noticed [10].

1.2.3.7 Mechanical methods

In some situations, it is possible to use big blinded vehicles to clean up the minefields when terrain and soil conditions are suitable. During a war, the military prefers this approach because the mines cannot be detected, identified and removed in time. This requires large and expensive vehicles to be used. The threat is reduced since the demining personnel are either in a well-protected area or track the vehicles remotely. But the environment is virtually destroyed by this mechanical technique. Furthermore, a mine can be buried deeper or partly damaged, and remain unexploded. The report was not released [7].

1.2.3.8 Dog detection method

Dogs are capable, in challenging post-war conditions, of checking contaminated soil more fast than manual demining (e.g. human) teams, of providing relatively low-tech and versatile alternative, of detecting minerals with low or no metal contents, and of working where heavy metal waste prevents using standard metal detectors

The search is performed with the dog on a long leash .The dog can detect the scent of the explosives up to 10 cm from the target. Once the position is marked the deminers inspect the target up to 25 cm deep using metal detectors [11].

Figure 5 Dog detection method

1.2.4 Our contribution

In this thesis, we address the problem of landmine detection using Hyperspectral Imaging or Imaging Spectroscopy, which is a novel technique in this scenario. This method provides the capacity to measure the part of light reflected in hundreds of wavelengths at each picture unit (pixel). A hypercube consisting of two spatial dimensions and a third dimension containing spectral data will thus be obtained. This method is widely used in several remote sensing applications such as mapping, agriculture, astronomy, food tracking, surveillance and others.

When light hits an object, it is either absorbed or reflected. The reflected part of light depends on the size of the object's molecules, specifically on the relationship between the intermolecular distances and the radiation wavelength. Each material composed of different components could reflect light of various wavelengths in a different manner. Therefore, by using this technology we have the ability to identify the materials remotely. In our case, we will use the hyperspectral images, spectral and spatial information to detect landmines and cluster ammunition without the presence of field deminers.

Due to the huge amount of data captured, dimensionality reduction is the main step before processing hyperspectral images. Different methods are used for this purpose. Some of these methods have high performance but also high complexity. And the other methods have low performance with low complexity. Therefore, we developed a new method for dimensionality reduction, taking into consideration the performance and the complexity. We tested our method in different scenarios. Also, a comparative study is made between the developed method and the main other methods usually used in the research.

There are several methods for target detection using hyperspectral imagery: some are supervised where the information spectrum is previously recognized; others are not supervised, based on the search for targets that are spectrally distinct from their environment. The latter sort of algorithms do not require the knowledge of the target spectrum in advance. However, as we will see in the next sections, this sort of detectors is characterized by an elevated false alarm rate because rare occurrences in the image will be labeled as targets, different from their context.

Some of the existing algorithms are characterized by a high false alarm rate during target detection and some of them have high computational complexity [12]. To solve these problems, in this thesis we propose a technique based on anomaly detection. We detect anomalies using a Pixel Intensity similarity measure at the pixel level in order to reduce the number of pixels, before applying the best performing detection algorithms.

The rest of this thesis is organized as follows.

Chapter 2 presents an introduction to hyperspectral imagery with dimensionality reduction in this field. Chapter 3 presents a new developed method for dimensionality reduction in hyperspectral imaging. Chapter 4 reviews current approaches for the detection of landmines using hyperspectral imagery and introduces a novel technique based on anomaly detection and the dimensionality reduction technique presented in the previous chapter. In Chapter 5, a discussion about thermal hyperspectral imaging used to detect buried landmine is presented to be a perspective of our future work. Chapter 6 illustrate the conclusions.

Chapter 2

Introduction to hyperspectral imaging with dimensionality reduction

2.1 Hyperspectral Imaging

This chapter presents an introduction concerning hyperspectral imaging. In addition, different existing reduction methods for dimensionality reduction in hyperspectral imaging are discussed. Finally, we propose a new reduction method aimed at obtaining a better trade-off between performance and complexity with respect to the existing methods..

In the field of remote sensing, hyperspectral imaging is a trending technique. It is based on the acquisition of images in the visible and infrared domain in quasi-continuous bands. Through this, we get a reflection spectrum at each pixel that lets us recognize the constituents of the object materials. In many fields, this technology is now fully operational, including remote sensing, astronomy, medicine, food safety, and forensics.

Hyperspectral imaging is the result of several advances in electro-optic technology. The progress in the manufacturing of sensors has made it possible for an object to be acquired in hundreds of wavelengths (fig 6). Such imaging techniques are usually referred to as multispectral imaging and hyperspectral imaging. The main difference between the two imaging techniques is the number of bands: more precisely, multispectral images may contain from 3 to 10 bands, whereas hyperspectral images usually have hundreds or thousands of bands.

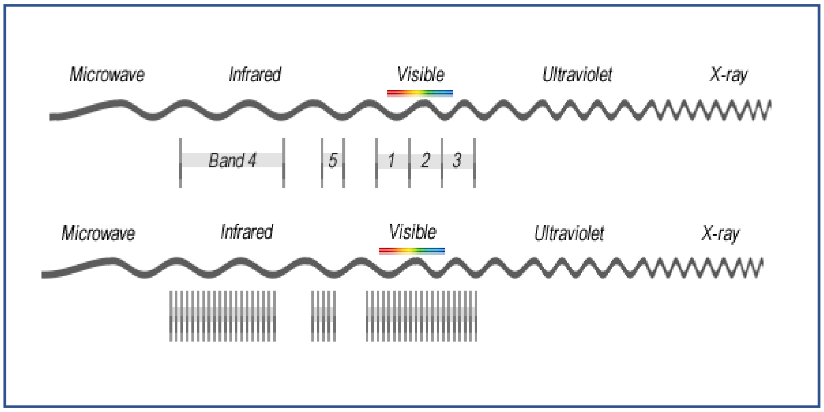


Figure 6 Difference between broadband, multispectral, hyperspectral and Ultraspectral Imaging

2.1.1 Hyperspectral image acquisition

For hyperspectral data acquisition, there are mainly two types of hyperspectral sensors: whiskbroom and pushbroom. To sample the received signal by wavelength, both use a dispersive element: sometimes a prism, but more often a grating. The difference is at the line acquisition level.

The whiskbroom sensor has a CCD array, where each element acquires a different wavelength. The acquisition of a complete line is done by a moving mirror system to image a different portion of the swath (image width) at different times (see figure7). The incoming energy is divided into multiple independently sensed spectral components [12].

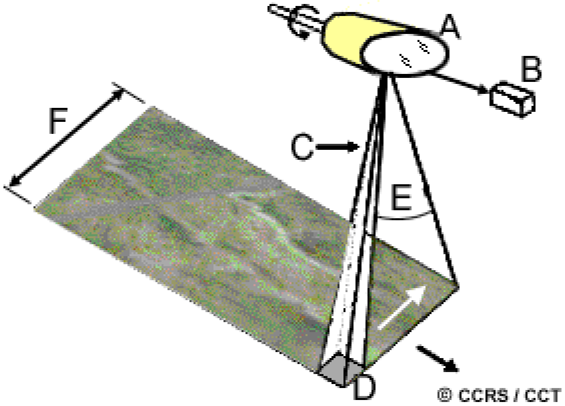


Figure 7Wiskbroom scanning principle

The pushbroom sensor has a CCD matrix where one of the dimensions corresponds to the different wavelengths and the other one to the swath of the instrument (Figure 8). The acquisition of the different lines is done in both cases by moving the sensor.

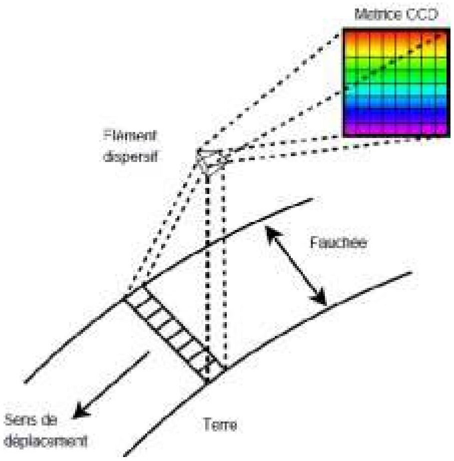


Figure 8 Pushbroom scanning principle

The disadvantage of the whiskbroom is that there is a moving part (the mirror), so this system is less reliable in a space system context. The pushbroom is more robust, but generally produces scratched images because the individual columns of the CCD matrix do not have all the same sensitivity. This defect can be partially corrected by calibration. The current trend for space-based sensors is clearly oriented towards pushbroom sensors.

2.1.2 Hyperspectral sensors used for remote sensing

Several hyperspectral imagers have been developed over the last few decades in various unrelated domains such as chemistry, mapping, military, food quality control, agriculture and so on. In this section, we review the most used ones. The list is not exhaustive, as the number of manufacturers is continually increasing as the application of this technology extends to new areas.

2.1.2.1 AVIRIS

The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) is an airborne instrument operated by NASA starting from 1987. The instrument measurements broadcast, reflected and distributed solar energy through wide areas at high spatial and spectral resolutions throughout the Earth's surface and atmosphere on 224 channels with wavelengths from 400 to 2500 nanometers [13]. On the other hand, the images are readily available, and many places are covered. In addition, the technical documentation available is very complete, and all specifications are available on the site.

2.1.2.2 CASI, CASI-2

Compact Spectrographic Imager (CASI), produced by Itres Research of Canada in 1988, is a pushbroom imaging spectrogram on two-dimensional CCD array .this sensor has a spectral range between 405 and 950nm with 288 spectral channels [14]. CASI is a widespread and widely used commercial hyperspectral instrument. It is the leading commercial device for hyperspectral image acquisition in the domain of the visible.

2.1.2.3 HYMAP

HyMap is an airborne, hyperspectral imaging sensor produced by Integrated Spectronics in Australia. This sensor provides a full spectrum from 450 to 2480 nm except in the water absorption bands around 1400 and 1900 nm. There are about 5 copies in use at present. Some differences exist between the versions (TIR bands on some versions). The instrument can be adapted according to the needs of the users. HyVista, an operator, owns one of these instruments, and is in charge of selling the images [15] [16].

2.1.2.4 Hyperion

Hyperion is a high resolution hyperspectral imaging instrument produced by NASA in November 2000. The Hyperion has 220 contiguous spectral bands, with a spectral range from 400 nm to 2.5 µm, with spatial resolution of 30 m. It has been manufactured to be the first sensor to transmit hyperspectral data from space, Hyperion was sent to demonstrate the feasibility of hyperspectral space applications. The SNR is not very good compared with airborne instruments (e.g. AVIRIS) [13, 17].

2.1.2.5 CHRIS-PROBA

This satellite sensor on board Proba, launched in October 2001 and operational since April 2002. CHRIS provides 19 spectral bands with a spectral range (400 - 1050 nm) at a GSD (Ground Sampling Distance) of 17 m..it has had different data acquisition modes, favoring either the spatial resolution, or spectral resolution. This compromise is necessary because of the limited transmission capacity. The objective of this instrument is to collect BRDFs (Bidirectional Reflectance Distribution Function) in order to better understand the spectral reflectance. At the technological level, CHRIS is also being used to explore the capabilities of Spectro imagers on small, handy satellites [18].

2.1.2.6 MERIS

MERIS (Medium Resolution Imaging Spectrometer) is a European space sensor of the ESA. It's not really a hyperspectral sensor, because it only transmits 15 bands [19]. therefore, we no longer have spectral continuity. In addition, the spatial resolution is quite low (200 to 300 m). An exhaustive study will therefore not be carried out for this sensor.

2.1.2.7 HYDICE

The Hydice developed by Hughes Danbury Optical Systems and operational since 1994, it has has been specified to demonstrate the usefulness of hyperspectral techniques in the fields of intelligence, military, but also civilian [20]. This sensor has a spectral range from 400 to 2500nm with 210 spectral bans. Its configuration is quite similar to that of CASI, but it has a higher spatial resolution (1-4 m), but for a swath lower (about 10°). Due to its military nature, images are not available.

2.2 Dimensionality reduction for hyperspectral imaging

Generally, data size can be measured along two dimensions: the number of variables and the number of bits to represent each variable. These two dimensions can take very high values, which can pose a problem when exploring and analyzing these data. For this purpose, it is fundamental to set up tools for the processing of data that allow a better understanding of the value of knowledge available in these data. The reduction in dimensions is one of the oldest approaches to provide a solution to this problem. Its objective is to select or extract an optimal subset of relevant characteristics according to a previously established criterion. This selection/extraction thus makes it possible to reduce the dimension of the signal space of the examples and to make the data set more representative of the problem. This reduction has a two-fold objective: the first one is to reduce the redundancy and the second one is to facilitate later treatments (extraction, classification, and pattern recognition, reduce the storage space required, reduce learning and use time, identify relevant factors,...) and thus interpretation of the data.

Techniques for the reduction of dimensionality in hyperspectral images can be classified according to two approaches: those based on the projection (linear or non-linear) of large data over well chosen subspaces and those based on the selection of spectral bands using complementarity-redundancy criteria (Entropy, Contrast, Correlation, Mutual Information) which consists in choosing the most representative spectra in the measurement space. The methods used in each of these two classes for the extraction of useful characteristics are very varied. We will briefly recall the principles of the methods using linear/non-linear projections, and then we will describe the methods for selecting spectral bands that have been proposed in the literature over the last years.

2.2.1 LINEAR PROJECTION APPROACHES

Linear data projection in a subspace of smaller dimension is a preliminary step in dimensional reduction. It reduces the dimension of the original space and make the data set more representative of the problem [21]. The new data representation is a linear combination of the original data, i.e.

Equation 1

$yki = w_{k1} \,.\, x_{1i} + w_{k2} \,.\, x_{2i} + \cdots + w_{km} \,.\, x_{mi} \, \forall \, K \in \{1,2,..,r\} \, , \forall \, i \in \{1,2,..,n\}$

Or in matrix notation: $Y = \Omega X$ with

Equation 2

$X = \{ xij \, , 1 \leq j \leq m \, , 1 \leq i \leq n \}, Y = \{ yki \, , 1 \leq k \leq r, 1 \leq i \leq n \}$ and

$\Omega = \{w_{kj}, 1 \leq k \leq r, 1 \leq j \leq m\}$ the linear transformation matrix. Several methods of linear projection exist in the literature and are introduced in the paragraphs below.

2.2.1.1 Principal Component Analysis (PCA)

Principal component analysis [22] is an unsupervised projective method that maximizes the variance of the projected data. It is assumed that that the source data lie in a hyperplane and can be expressed by means of the vectors that define this hyperplane. If the hypothesis proves to be true, we find the intrinsic dimensionality of the data. The data vectors $X \subset R^m$ are projected to the sub-dimensional space $Y \subset R^r$, $r < m$ such that $Y = W X$, where $W \subset R^{m\times r}$.

Equation 3

$$E[XW] = E[X]W = \mu W$$

Equation 4

$$\mathrm{Var}(XW) = E[(XW - \mu W)^2] = E[W'(X-W)'(X-\mu W)W] = W'\Sigma W$$

The main component of the projection is W_1. Since the direction of this vector is the most relevant information, it is considered that $\|W_1\| = 1$

$$Y_1 = X W_1 \text{ and } \mathrm{Var}(Y_1) = W'\Sigma W_1$$

We are looking for W_1 such that $\mathrm{Var}(Y_1)$ is maximized by respecting the constraint $W_1{}' W_1 = 1$. The Lagrangian is: $L(\alpha) = W_1{}'\Sigma W_1 - \alpha (W_1{}' W_1 - 1)$. Thus, deriving from W_1 one finds $\Sigma W_1 - \alpha W_1 = 0$.

The relation $\Sigma W_1 - \alpha W_1 = 0$ is valid if W_1 is an eigenvector of Σ and α is the corresponding eigenvalue. We then have: $\alpha W_1{}' W_1 = \alpha$.

The eigenvector with the largest eigenvalue is chosen so that the variance is maximized. The second main component must also maximize the variance, have a unitary norm and, in addition, be orthonormal to W_1.

Equation 5

$$\text{So, Max } L(\alpha) = W_2{}' \Sigma W_2 - \alpha (W_2{}' W_2 - 1) - \beta (W_1{}' W_1 - 1)$$

Deriving from W_1 and equalling to 0:

Equation 6

$$2\,\Sigma\, W_2 - \alpha W_2 - \beta W_1 = 0$$

Multiplying equation (6) by W_1 we see that $\beta = 0$. We have: $\Sigma\, W_2 = \alpha\, W_2$. So, W_2 is also an eigenvector of Σ corresponding to the 2nd largest eigenvalue. The rth component can be derived in the same way. Σ is symmetrical, singular, its rank (the intrinsic dimensionality of the data) is r with $r < m$. Usually, the initial data are centered before computing the projection, i.e.,: $Y = (X - \mu)/W$.

After this linear transformation, the projected data lie in an r-dimensional space spanned by the r eigenvectors of Σ corresponding to the r largest eigenvalues and the variance of projected data for each of these dimensions correspond to their associated eigenvalues. The vectors W_i , $i=1,\ldots r$ define an orthonormal base for the projection space. Looking at the spectrum of eigenvalues, we get an indication of the intrinsic dimensionality of the data. The directions that correspond to the smallest eigenvalues are normally interpreted as noise. PCA is very effective for removing redundancies from the data but may not preserve some information that is still relevant for classification tasks, in particular rare events.

2.2.1.2 Locality Preserving Projections (LPP):

Another linear algorithm in dimensionality reduction is LPP [23, 24]. It constructs a neighborhood graph of the data set using the notion of the Laplacian, then compute a transformation matrix which maps the data points to a subspace. This linear transformation optimally preserves local neighborhood information in a certain sense. It is relatively insensitive to outliers and noise.

The LPP algorithm is formally described as follows:

1. **Construct the adjacent graph**: let G is the graph with k nodes. If x_i and x_j are close, Then an edge is putted between i and j. there are two variations:
    a) ϵ-neighborhoods: where $\epsilon \in \mathbb{R}$. So, Nodes i and j are connected by an edge if $||xi - xj||^2 < \epsilon$ where the norm is the usual Euclidean norm in R^n.

    b) k nearest neighbors. [parameter $k \in N$] Nodes i and j are connected by an edge if i is among k nearest neighbors of j or j is among k nearest neighbors of i.

2. **Choosing the weights**: Here, as well, we have two variations for weighting the edges. W is a sparse symmetric $m \times m$ matrix with W_{ij} having the weight of the edge joining vertices i and j, and 0 if there is no such edge.

    a. Heat kernel [parameter $t \in R$]. If nodes i and j are connected

Equation 7

$$Wij= exp^{\frac{||xi-xj||^2}{t}}$$

    b. Heat core with parameter t = 1. If vertices i and j are connected, $W_{ij}= 1$.

3. Eigenmaps: Compute the eigenvectors and eigenvalues for the generalized eigenvector problem:

Equation 8

$$XL X^T a = \lambda XD X^T a$$

Where D is a diagonal matrix of weights whose elements are the sum of each line

of W, $D_{ii}=\Sigma_j W_{ij}$, L is the Laplacian matrix L = D - W. L is a symmetrical semi defined positive, with $a_0, a_1, ..., a_{k-1}$ are the solutions of equation (8) which corresponds to values $\lambda_0 < \lambda_1 < ... < \lambda_{k-1}$, $x_i \rightarrow y_i = A^T x_i$, $A= (a_0, a_1,..., a_{m-1})$, y_i is a vector to m dimensions in the confined space.

2.2.1.3 Multi-Dimensional Scaling (MDS)

The Multidimensional Scaling Method (MDS) [25] takes a dissimilarity matrix D where D_{ij} represents the dissimilarity between points i and j and produces a mapping on a lower dimension, preserving the dissimilarities as closely as possible. The dissimilarity matrix could be observed or calculated from the given dataset [26]. The classic example is to get the map of a country based on the knowledge of the distances between each pair of cities. On the other hand, this method is very sensitive to the variation of parameters of data and to noise degradation, as verified in [26] based on tests on a natural database and an artificial database. Moreover, MDS does not preserve rare events [25].

Algorithm

MDS calculates a dissimilarity approximation in lower dimension r < m Suppose a distance matrix D approximates the inter-point distances of a configuration of points X in a space of low dimensionality p (usually p = 1, 2, or 3). That is, the elements of D, denoted d_{ij} , may be calculated from X using the following formula:

Équation 9

$$d_{ij}=\sqrt{\sum_{1}^{p}(xik-xjk)^{2}}$$

the algorithm steps are as follow:

i. Calculate $A =-\frac{1}{2}\, d_{ij}{}^{2}$

ii. Calculate $B = \{a_{ij} - a_{i.} - a_{.j} + a_{..}\}$, where $a_{i.}$ is the average of all a_{ij} across j.

iii. Find the p largest eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_p$ of B and corresponding eigenvectors

$L = (L_{(1)}, L_{(2)},\ldots L_{(p)})$ which are normalized so that $L_i' L_i = \lambda_i$. (We are assuming that p is selected so that the eigenvalues are all relatively large and positive.)

iv. the rows of L are the coordinate of the objects.

2.2.2 NONLINEAR PROJECTION APPROACHes

In many applications, the Euclidean distance does not make much sense; it assumes in particular that all the variables are comparable with each other (therefore, they must have been properly standardized). Hilbert's space theory allows to define other scalar products, based on kernel functions K(x, y). K is then a measure of similarity between the items in the set to be addressed. The K kernel implicitly defines an application from the origin space to a "feature space" H. The dimension of H-space is eventually infinite. Many statistical methods can express themselves using only scalar products between the points to be addressed and the learning examples. If we replace the usual scalar product by a K kernel, we render the method non-linear; this is the kernel trick, which has been the subject of numerous researches since its introduction by Vapnik in [27] within the framework of support vector machines (SVM). Sometimes the source data belong to non-linear subspaces.

The general principle of these methods is to consider all of the initial data like samples of a variety of Riemann surface. Therefore, to each value of x there correspond a finite number of values of y. From there, a solution is built approximating the metric of the variety that is used to eliminate small dimensions significant variability of the set. The goal is to identify low dimensional structures in high dimensional spaces.

2.2.2.1 ISOMAP (Isometric Characteristics Mapping)

The Isomap (Isometric Characteristics mapping) dimensionality reduction method has been introduced by Tenenbaum [28] and is based on the Multidimensional Scaling Method (MDS). for the knowledge of computing a matrix of pairwise dissimilarities substituting the Euclidean distance with a different metric . The goal this time is to find a (non-linear) variety containing the data. The Isomap method estimates the geodetic distance [29, 30] as follows: In a first pass, the neighborhood of each of the points is calculated. Once the neighborhoods are known, a graph is constructed by connecting all neighboring points(Each point is connected to other if it is a K nearest neighbor.). Each point of the graph is then weighted by the Euclidean distance [29] between the two ends of the edge. Finally, the geodetic distance between two points is estimated by the sum of the lengths of the edges along the shortest path between these points. In practice, the shortest path between two vertices of the graph is calculated using the Dijkstra algorithm.

Algorithm

From an algorithmic point of view, the Isomap projects a surface in an m-dimensional space,

represented by n points, as follows:

i. Search for vectors of the set X = $\{xi ,..., x_n\} \subset R^m$ neighbors. According to the Euclidean distances d_{ij} a neighbourhood is determined for each point x_i, either. with the k-nearest neighbour criterion, or by considering all the points inside a sphere of radius α centered on x_i. Euclidean distances are approaching the geodesics when the vectors are at small distance.

ii. The geodetic distances d(G_{ij}) are estimated between all x_i points. Isomap constructs a graph whose vertices are the points and edges are the distances between them.

iii. The Dijkstra algorithm is applied to obtain the matrix of distances D, which is contains the geodetic distances between each point and its neighbours. The Dijkstra algorithm computes the shortest path between two vertices of the graph. A vertex is adjacent to another only if they have been defined as neighbours (Step 1). The geodetic distance is estimated between each data pair by the length of the shortest path on the graph (Floyd's or Djikstra's algorithm).

iv. Finally, the MDS method is applied to the matrix of distances (dG_{ij}), i, j =1,..., n to obtain a new Euclidean coordinate system $Y \subset R^r$, (r < m) which preserves the intrinsic geometry of the variety.

2.2.2.2 Locally linear embedding(LLE)

Roweis and Saul [31] have proposed a method of dimensionality reduction called "Locally Linear Embedding" which this time is no longer based on neural methods, but algebraic. This method allows to project a variety of vectors belonging to a space of dimension m in a space of smaller dimension r by considering that the data are locally linear.

Local Linear Embedded (LLE) was presented at the same time as Isomap and discusses the same problem by a different route. Each point is here characterized by its reconstruction from its closest neighbors. LLE builds a projection to a linear space of small dimension preserving the neighborhood.

The LLE method considers that the data, which are globally non-linear, are linear in small neighborhoods. The idea is to determine for each vector $x_i \in E$ its k neighbors and to consider that in this neighborhood the data are linear. Once the neighbors of x_i are determined, the w_{ij} reconstruction weights associated with each pair (x_i, x_j) are calculated, where x_j belongs to neighborhood V(x_i). These reconstruction weights, which represent the neighborhood of each vector x_i, will allow to reconstruct the topology of the neighborhood in the space of projection. The LLE method consists of three main steps with the following functions:

i. Search for k neighbors,

ii. Calculation of reconstruction weights w_{ij}

iii. Vector projection by minimizing the cost function E(W)

Search for neighbors: This step determines the neighborhood of each vector in which we will make the hypothesis that the data are linear. We are simply looking for the vectors closest to the xi vector whose neighborhood we want to determine, and to do so we simply use the Euclidean distance.

Calculation of the reconstruction weights: The neighborhoods of each of the vectors of the set of data have been determined, the cost function that will be required for the new data must now be defined and minimized to determine the matrix w reconstruction weight. The cost function E(W) is defined by:

Equation 9

$$E\ (\mathrm{w}) = \sum_{i=1}^{n} |\mathrm{xi} - \Sigma \mathrm{j} \in \mathrm{V(xi)wijxi}|^2$$

The minimization of E(w) is done under two constraints:

i. each vector x_i is reconstructed only thanks to its neighbors, i.e. $w_{ij} = 0$ if x_j is not close to x_i

ii. $\forall$ i $\Sigma_{j \in V(\mathrm{xi})}(xiwij\ ) = 1$

With the above constraints, the reconstruction weights can be calculated as follows. Let x_i be a particular vector of the surface to be unfolded in E space; where $(\xi_1 \ldots, \xi_k)$ are the nearest neighbors of x_i and w_{ij} the associated reconstruction weights. The principle is thus to reconstruct linearly the neighborhood of x_i based on its neighbors in the initial space. For each vector x_i

Equation 10

$$E_i(\mathrm{w}) = |\sum_{j=1}^{k} \mathrm{wij(xi - xj)}^2 = \sum_{j=1}^{k} \sum_{n=1}^{k} \mathrm{wijwin}\emptyset\mathrm{jn}\ |$$

We have the following cost function: where ϕ^i is a K x K matrix defined by $\phi_{jn}{}^{i} = (x_i - x_j)^t(x_i - x_n)$ we poses $R_i = (\phi^i)^{-1}$. Solve the problem $\sum_{j=1}^{k} \mathrm{wij} = 1$ which is a problem of least squares, with constraint, leads to:

Equation 11

$$W_{ij} = \frac{\sum_{n=1}^{k} R_{jn}^{i}}{\sum_{p=1}^{k} \sum_{q=1}^{k} R_{pq}^{i}}$$

Vector projection

This last step aims at reconstructing the local topology, represented by the matrix of weights W, in Euclidean space F. Therefore, it allows to unfold the initial area represented by the vectors (x_i) in a space of smaller dimension. This unfolding is obtained by minimizing the following cost function:

Equation 12

$$\phi(y)=\Sigma_i |y_i - \Sigma_i w_{ij} y_j|^2$$

Where y are all the vectors projected on F.

ϕ (y) is written in matrix form: ϕ (y) = Y^TM Y, this minimization must be carried out under the following two constraints:

- Any translation of the yj vectors must not modify ϕ . To remove this degree of freedom, we will impose to the vectors to be centered in the origin $\Sigma_i y_j = 0$

- To avoid degenerate solutions, the y_j vectors are constrained to have a same covariance

$\frac{1}{N}\Sigma_i y_i \otimes y_i = I$

The last step is to align the global coordinates. We calculate the (d+1) plus small eigenvectors of B correspond to (d+1) eigenvalues.

2.2.2.3 Local Linear Coordination (LLC)

Local Linear Coordination LLC calculates a number of linear models and then performs a global alignment [32]. This process consists of two steps:

i. Mixing of local linear models on the data using the expectation-maximization (EM) algorithm.

ii. Aligning local linear models to obtain a representation of the data in the reduced space using an Locally linear embedding (LLE) variant.

LLC constructs m components analysis mixture using the EM algorithm [33-35]. The local linear models in the mixture are used to construct *m* data representations zij and their corresponding responsibilities r_{ij} (where j ∈ {1, . . . , m}) for every datapoint x_i , r_{ij} describes the extent to which the data points in data x_i correspond to the model j by satisfying:

$$\sum_{j=1}^{m} rij = 1$$

Using the local models and the corresponding responsibilities, responsibility-weighted data representations $u_{ij} = r_{ij}z_{ij}$ are computed. The responsibility-weighted data representations u_{ij} are stored in a $n \times mD$ block matrix U. The alignment of the local models is performed based on U and on a matrix M that is given by

$M = (I_n - W)^T(I_n - W)$. Moreover, the matrix W contains the reconstruction weights computed by LLE, and I_n denotes the n × n identity matrix. LLC aligns the local models by solving the

generalized eigenproblem.

$$Av = \lambda Bv$$

For the d smallest nonzero eigenvalues. In the equation, A is the inner product of M^T U and B is the inner product of U. The d eigenvectors vi form a matrix L that defines a linear mapping from the responsibility-weighted data representation U to the underlying low-dimensional data representation Y. The lowdimensional data representation is thus obtained by computing Y = UL. The main weakness of LLC is that the fitting of the mixture of factor analyzers is susceptible to

presence of local maxima in the log-likelihood function. LLC has been successfully applied to face images of a single person with variable pose and expression, and to handwritten digits.

2.2.2.4 Sammon Mapping

Sammon mapping is a non-linear algorithm [26], which is related to MDS. The purpose of the algorithm to project data from a space at r dimension in m dimensions such that r<<m. During the projection, the Sammon mapping tends to preserve the distances between all points in the dataset taken pairwise distance in order to preserve the data structure, so that the distances after the mapping are close to the distances within the original space. Each object is represented by a point in a space at r dimensions. dij, ∀ i, j =1...n is the distance between two points in the original r-dimensional space. δij, ∀ i, j =1...n is the distance between two points in the projected m-dimensional space.

In order to identify the n points within the projected space whose pairwise distances best approximate those in the original space, a cost criterion is used. The latter is generally referred to as a "stress function" denoted as S. It measures the difference between the configurations in both spaces.

Equation 13

$$S = \frac{1}{\sum_{i=1}^{n}\sum_{i=j+1}^{n} \delta ij} \sum_{i=1}^{n}\sum_{i=j+1}^{n} \frac{(\delta ij - dij)^{2}}{\delta ij}$$

the key drawback of this technique is that focuses on the preservation of the large pairwise distances, and not on preserving the small pairwise distances, which are much more important to the geometry of the data. On the other hand, it has the power to minimize the error function, which is often referred to as Sammon stress and retain the local structure of data (weight, distance).

2.2.2.5 Diffusion Maps

Diffusion Maps (DM) is a method based on hybridization between a heat diffusion model and a stochastic process (Markov chain) [36].

$X = \{x^{(1)}, x^{(2)},\dots, x^{(k)}\} \in R^{m}$, $x^{(i)}=\{ x_{1}^{(i)}, x_{2}^{(i)},\dots, x_{m}^{(i)}\}$ i=1,...,k

$Y= \{y^{(1)}, y^{(2)},\dots, y^{(k)}\} \in R^{r}$, $y^{(i)}=\{ y_{1}^{(i)}, y_{2}^{(i)},\dots, y_{r}^{(i)}\}$, j =1,...,k

Where X are the data points and Y are the lower-dimensional set of X.

With $r \ll m$, $\{x^{(1)}, x^{(2)},\dots, x^{(k)}\} \in M$, or M a variety in R^{m},. the goal is to find a lower-dimensional set of feature vectors, $Y_{1\dots k}$, for which the distance matrix, DY [i, j] =d(y_i, y_j), minimises a cost function, $\rho(D_X, D_Y)$. The procedure is described below in three steps.

Algorithm

i. Construct the adjacency matrix W of the graph whose weights Wij connecting the nodes

i and j are given by the kernel: $W_{ij}= exp^{\frac{-\|xi-xj\|2}{\varepsilon}}$, $\varepsilon = 2\,\sigma^{2}$

ii. Construct the diagonal normalized k x k matrix and the Laplacian matrix respectively as $Dii = \sum_{j=1}^{k} Wij$ and L= D-W. Note also that L is symmetrical and semi definite positive.

iii. Calculate the eigenvalues and eigenvectors of the generalized problem:

Equation 14

$$L\,y = \lambda\,D\,y$$

The eigenvectors $y_0, y_1..., yk_{-1}$ corresponding to eigenvalues $\lambda_0, \lambda_1..., \lambda_{k-1}$ are solutions of equation (11).

2.2.2.6 Laplacian Eigenmaps

The Laplacian Eigenmaps algorithm calculates a small Y-dimensional representation of set X, preserving the proximity relations of the data in the original space [37]. The algorithm is similar to the LLE method; however it has different theoretical justification: the neighborhood information is retrieved using a graph, but the small coordinates are obtained from the Laplacian matrix of the graph [37].

Algorithm

Determination of the neighbors for each point (with the techniques defined for Isomap and LLE). In the construction of the graph, the vertices are the points and the edges are non-void only if xi and xj are close. In the Edge Weighting, there are two possibilities:

- Core with a parameter t ∈ R. If vertices i and j are connected

Equation 15

$$W_{ij} = exp^{\frac{-\|xi-xj\|2}{t}}$$

- Core with parameter t = 1. $W_{ij} = 1$ if vertices i and j are connected by an edge and Wij = 0 if vertices i and j are not connected by an edge), the eigenvalues and vectors of the system:

Equation 16

$$L\,f = \lambda Df$$

Where D is a diagonal matrix of weights whose elements are the sum of each line of W, $D_{ii} = \Sigma_j\, w_{ij}$. L is the Laplacian matrix L = D - W. L is a symmetrical matrix positive semidefinite which can be considered as the Laplacian operator of the functions defined in graph G.

where D is diagonal weight matrix, and its entries are column (or row, since W is symmetric) sums of W, $D_{ii} = \sum_j W_{ji}$. L = D − W is the Laplacian matrix. Laplacian is a symmetric, positive semidefinite matrix that can be thought of as an operator on functions defined on vertices of G.

Let $f_0,\dots,f_{k-1}$ be the solutions of equation 13, ordered according to their eigenvalues:

$Lf_0 = \lambda_0 Df_0$

$Lf_1 = \lambda_1 Df_1$

…

$Lf_{k-1} = \lambda_{k-1} Df_{k-1}$

$0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{k-1}.$

We leave out the eigenvector f_0 corresponding to eigenvalue 0 and use the next m eigenvectors for embedding in m-dimensional Euclidean

space:

$x_i \rightarrow (f_1(i), \dots, f_m(i)).$

2.3 Conclusion

In this chapter, we gave a brief introduction of hyperspectral imaging, presenting the two general nonparametric and unsupervised approaches for the reduction of the dimensionality of the images and their principles. It is worth mentioning that the aim of this thesis is that of learning non supervised, non-parametric representations by preserving rare events and that in this context, we are interested in automatic dimensionality reduction methods for large images. Unsupervised techniques based on projections are either limited by their linear nature (ACP, MDS), or difficult to use because of their algorithmic complexity (Isomap, LLE...) when working on large databases and the majority of them are sensitive to the different variations they undergo (parameters, noise degradation) and to the loss of information. That is why they do not fully meet the thesis objective. Taking into account the local/global evaluation criteria presented in the following chapter, the authors have shown, the relevance of some algorithms of dimension reduction by projection (LTSA, H.LLE, Isomap, MVU) compared with other techniques (PCA, K.PCA, DM, LLE, LEM, Sammon, Autoencoder, LLC, MC, LP, LE) to preserve the local/global geometric data structure.

Chapter 3: Multicriteria dimensionality reduction method

3.1 Introduction

This chapter mostly covers material presented in a research article published in the journal of applied remote sensing [38].

We develop a multicriteria dimensionality reduction approach that is adapted to hyperspectral applications in the sense that it preserves the data's "explainability." The method is based on the fact that images from neighboring (adjacent) bands of hyperspectral images are highly correlated and often convey almost the same information about the scene. By exploiting this characteristic, the set of images, which constitute the hyperspectral image, are examined and classified based on multiple criteria. The images carrying similar information are put in the same class, and one image from each class is chosen to represent its class. The set of these representative images form a subset of the original set of images. This subset represents the reduced hyperspectral image; thereby, the dimension will be reduced to the number of representative images chosen (or number of classes).

The developed algorithm classifies the set of images corresponding to the different spectral bands in a hyperspectral image based on multiple criteria. These criteria evaluate the similarity of two images taking into account different aspects. The images which are classified within the same class are the ones which satisfy certain thresholds of similarity, i.e. convey almost the same information. A single image from each class is chosen to represent its class in the final set of data. Hence, the final set of data is a subset of the original one, whereby a certain number $(< n_\lambda)$ of spectral bands from the original set are chosen to estimate the original hyperspectral image. This resulting subset of images represents the reduced hyperspectral image.

3.1.1 Classification Principle

The images are examined from the lowest spectral band to the highest one. The values of three classification criteria are calculated between two images. These criteria are the **mutual information**, **fidelity**, and the **cross-correlation coefficient**. These three calculated values form a vector $\vec{V}$, this vector $\vec{V}$ is compared to a threshold vector $\vec{T}$. $\vec{T}$ contains threshold values for each of the three criteria previously mentioned. If the vector $\vec{V}$ calculated for two images is above the threshold vector $\vec{T}$, then these two images are considered to convey almost the same information, and thereby are classified in the same class. Noting that in this discussion $\vec{V} > \vec{T}$ means each of the three values in $\vec{V}$ is above its corresponding threshold value in $\vec{T}$. In case the value of any criterion among the three is not above its corresponding threshold value, then $\vec{V}\ ! > \vec{T}$ and these two images are not considered to convey almost the same information and therefore are not classified in the same class.

As previously mentioned, and since images belonging to the same class are the ones which convey almost the same information, only one image from each class is chosen to represent its class. The selection of the class representative image is based on the one which has maximal **entropy** within its class. This choice is based on the fact that the image which has maximal entropy is the one which conveys maximal information. This particular idea will be further elaborated in the next sections. By the end of the previously described procedure, a subset of the original set of images will be obtained. Each spectral band in the original hyperspectral image was represented initially by one image. However, in the reduced hyperspectral image, one or more adjacent spectral bands are represented by one image. Therefore, the dimensionality will be reduced to be equal to the number of classes, which is also equivalent to the number of representative images in the final set of data. The idea is illustrated by an example in figure 9.

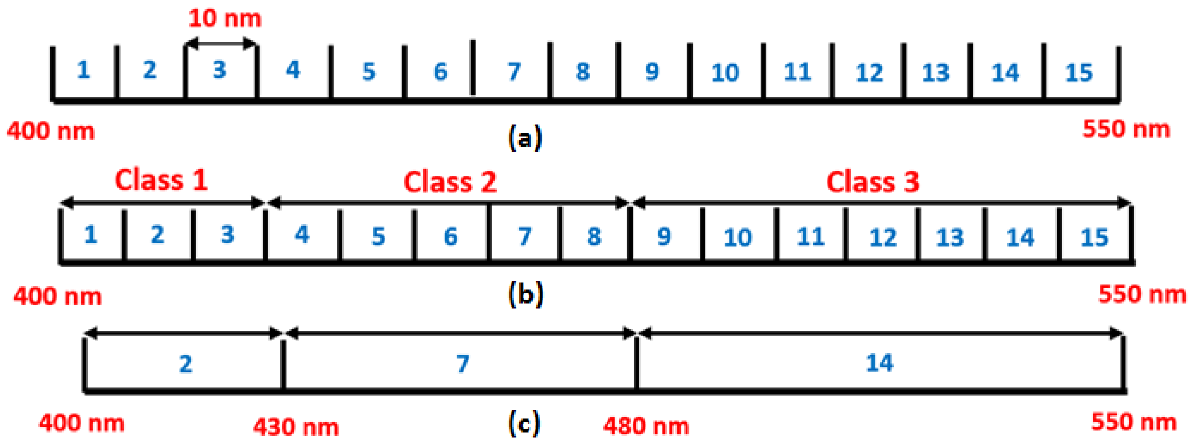


Figure 9 Classification principle (a) Illustration of a Hyperspectral Image (n_λ=15), (b) Classification Results , (c) Illustration of The Reduced Hyperspectral Image (〖n'〗_λ=3)

Figure 9(a) is a schematic illustration of the first 15 images $(n_\lambda = 15)$ of a hyperspectral image. Each blue numbered value corresponds to a spectral band of 10 nm width, and also corresponds to an image in the set of images of the hyperspectral image. Supposing that the previously discussed multi-criteria reduction method was applied on this part of the hyperspectral image, the results of the classification are those shown in figure 9 (b).

As previously mentioned, the image which has maximal entropy within its class is the one chosen to represent its class. Supposing image 2, 7, and 14 are the ones having maximal entropy within class 1, 2, and 3 respectively. Then the reduced version of the hyperspectral image is the one represented in figure 9 (c). The reduced hyperspectral image which is illustrated in figure 10(c) contains three images from the set of images of the original hyperspectral image. Therefore, the set of images of the reduced hyperspectral image is a subset of the original set of images. Where each image in the reduced hyperspectral image represents more than one 10 nm spectral band. The dimensionality in this example is reduced from 15 to 3 by applying the multi-criteria reduction method on the hyperspectral image presented in figure 9 (a).

3.1.1.1 **Rare Event Constraint**

Hyperspectral sensors can now uncover many unknown signal sources which cannot be identified by visual inspection. These signal sources may include unknown interfering sources, which cannot be identified by prior knowledge. So, when reducing the redundancy, it is important to preserve also those weaker signals. There are several signals processing algorithm developed to remove signal redundancies while preserving rare signals (vector) detected from a weak source; Oleg Kuybeda et al [39], proposed an algorithm for redundancy reduction of high-dimensional noisy signals, named Maximum Orthogonal Complement Analysis (MOCA), which is designed for applications where a representation of both abundant and rare vectors is essential; José M et al [40] present the signal subspace method (Hsym method) which estimates the signal and the noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the minimum mean square error sense. Also Nicola Acito [41, 42] et al present two signal subspace algorithm: the (MOCA) algorithm presented above and the Robust Signal Subspace Estimator (RSSE) algorithm to address both the abundant and the rare signal components. Therefore, rare signals may contain information about the rare objects present in the hyperspectral scene. After the identification of rare signals due to weak sources, this rare vector should be preserved in our method after the signal processing tools. Therefore, a new criterion was added in our proposed method which is based on the mean squared error (MSE), as explained in the following.

The classification, previously discussed, is based on the values of the three criteria which are the mutual information, fidelity, and the cross-correlation coefficient. The values of these criteria are determined by global processing of the set of images of the hyperspectral image. However, global processing alone proves to be inefficient in cases where rare events exist in small spatial zones. Rare events are defined as certain objects which appear in a scene and are different from their surroundings. An example of a rare event is the tulip which appears in figure 10. Such events are details in an image, which appear in small areas of the image. If the spatial dimensions of such events are very small compared to the total spatial dimensions of the image, their presence is likely to be ineffective on the result of global processing. For such reasons, local processing is also needed alongside the global one in order to process rare events.

In the context of the multi-criteria reduction method developed in this thesis, a constraint should be added to ensure the preservation of rare events. In fact, when classifying the set of images of the original hyperspectral image, it is important to ensure the presence of images which contain rare events in the final reduced hyperspectral image. The three global criteria may not be enough to ensure the preservation of rare events for the reasons previously discussed .

Figure 10 Image Containing a Rare Event

To further elaborate the problem, let us consider the case where an image containing a relatively small rare event is classified with other images that do not contain this rare event. This is possible when only *global* processing is present, since this small rare event may not affect the final result of processing. Now, since the chosen class representative image will be the only one present in the final reduced hyperspectral image, if the image having maximal entropy within this class is not the one containing the rare event, then the rare event will be absent in the final result of reduction. This would be a big flaw in the reduction process whereby important and unique details will be lost.

To solve this problem, a fourth criterion was added to the reduction method. This fourth criterion is the MSE (Mean Squared Error) and is used for local processing of pixel zones where rare events are detected. In addition to the three criteria previously mentioned, this criterion presents a constraint in the reduction process. Consider the case when during the reduction process and when comparing two images, the three original criteria are satisfied globally. In this case additional steps should be taken before classifying these two images in the same class.

The pixels of the two compared image where rare events exist should be processed locally by calculating their respective mean squared errors. If the MSE is above a certain threshold at the level of any rare event, that rare event is considered to be unpreserved, and the images are not classified in the same class even if the three other *global* criteria are satisfied. However, if the MSE is below a certain threshold, then the two images are classified in the same class. This rare event constraint therefore presents the fourth criterion in the multi-criteria reduction method we developed.

To detect the presence of rare events in the pixels of an image, we used the method proposed in [43]. The method states that in order to determine if a pixel a_i is considered as a rare event, we calculate its relevance based on :

Equation 17

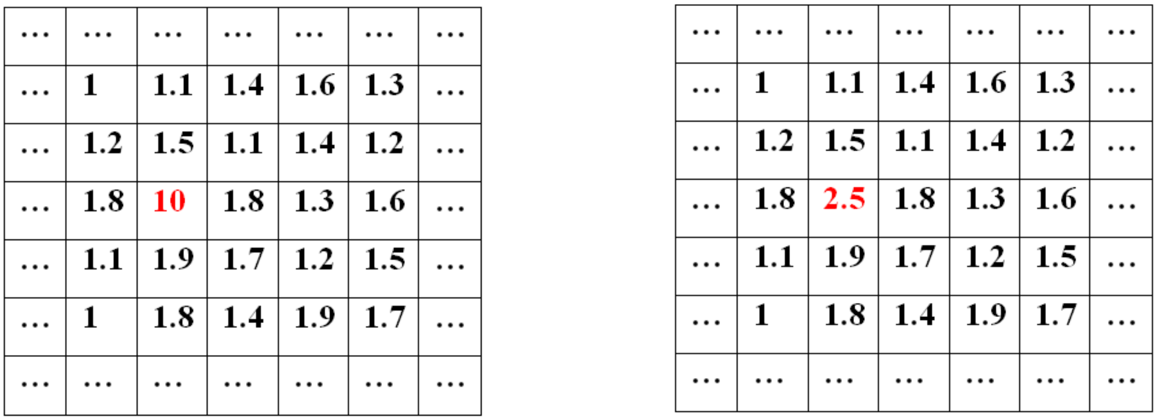
$$Relevance(a_i) = R(a_i) = \frac{\|a_i - \bar{\mu}\|^2}{\mu^2 + \sigma^2}$$

μ and σ^2 are the mean value and the variance of the set of pixels containing a_i and its neighborhood. $\bar{\mu}$ is the mean value of only the pixels which are in the neighborhood of a_i (excluding a_i). Based on Equation 18 it is decided whether a_i is a rare event or not.

Equation 18

$$\begin{cases} a_i\ is\ not\ rare \quad if\ \ 0 < R(a_i) < 0.5 \\ \qquad a_i\ is\ rare\ \ if\ R(a_i) \geq 0.5 \end{cases}$$

Examples are presented in figures 11 (a) and (b). In figure 11 (a) the pixel highlighted in red has a value of 10 and is different from its surroundings. The relevance of this pixel is calculated to be 12.094 using Equation 17 and using Equation 18 we can verify that this pixel is indeed a rare event. In figure 11 (b) the pixel highlighted in red has a value of 2.5 which is different but close to its surroundings. The relevance of this pixel is calculated to be 0.49 , using Equation 18we can say that this pixel is on the limit of being considered rare.



…	…	…	…	…	…	…
…	1	1.1	1.4	1.6	1.3	…
…	1.2	1.5	1.1	1.4	1.2	…
…	1.8	10	1.8	1.3	1.6	…
…	1.1	1.9	1.7	1.2	1.5	…
…	1	1.8	1.4	1.9	1.7	…
…	…	…	…	…	…	…

(a)

…	…	…	…	…	…	…
…	1	1.1	1.4	1.6	1.3	…
…	1.2	1.5	1.1	1.4	1.2	…
…	1.8	2.5	1.8	1.3	1.6	…
…	1.1	1.9	1.7	1.2	1.5	…
…	1	1.8	1.4	1.9	1.7	…
…	…	…	…	…	…	…

(b)

Figure 11 (a) Image Representations 1, (b) Image Representation 2

$$R(2.5) = \frac{(2.5 - 1.4375)^2}{2.3036} = 0.49 < 0.5 \qquad R(10) = \frac{(10 - 1.4375)^2}{6.062} = 12.094 > 0.5$$

3.1.1.2 **Multi-Criteria Analysis**

As previously explained, the classification-reduction method we developed is based on the similarity between images from consecutive (adjacent) spectral bands. The similarity of images is evaluated using multiple criteria (image similarity measures) which are the **mutual information**, **cross-correlation coefficient, fidelity, and MSE.** These measures provide a quantitative measure of the degree of match between two images by taking into account the spatial alignment.

Image similarity can be evaluated with or without spatial alignment. Similarity measures without spatial alignment work by comparing the probability distributions of two images without taking into account the spatial positioning. On the other hand, similarity measures with spatial alignment return values which consider the spatial positioning [44]. In our reduction method we are interested in image similarity measures with spatial alignment such as the four criteria which were chosen. The reason for choosing such measures lies in the nature of hyperspectral imaging, where the scene is observed on multiple spectral bands while keeping the same spatial alignment during acquisition. Therefore, taking into account spatial alignment while comparing images is essential in order to respect the nature of the original data and achieve efficient reduction. In the following, the criteria used will be analyzed and their mathematical expressions will be presented. The two images to be compared are denoted by A and B; a_k and b_k are, respectively, the gray-levels of the k^{th} pixel in A and B. The total number of pixels is K.

- **Mean Squared Error (MSE)**

The mean squared error is probably the simplest measure among the similarity measures. The MSE should be used when the input images have been captured with the same sensor under similar conditions. This measure is sensitive to outliers and illumination changes [45]. However, this problem is not serious in local processing of relatively small areas in an image. The simplicity of the MSE makes it a popular image similarity measure especially in local processing tasks. The mathematical expression of MSE is defined in:

Equation 19

$$MSE(A,B) = \frac{1}{K}\sum_{k}(a_k - b_k)^2$$

In the case where the image intensities of A and B are scaled from 0 to 1, the resulting MSE is a value also between 0 and 1. Image similarity, based on the MSE value, is evaluated as follows: A value close to 1 indicates that the two images are not similar, while a value close to 0 indicates that the two images are similar.

In the reduction method we developed, the MSE was used as the single local processing criterion which ensures the preservation of rare events as explained in the previous section.

- **Image Fidelity**

Image fidelity is expressed in terms of NMSE (Normalized Mean Squared Error), so these two have similar properties. The mathematical expression of the image fidelity IF is defined in:

Equation 20

$$IF(A,B) = 1 - \frac{\sum_k (a_k - b_k)^2}{\sum_k (a_k)^2}$$

In (17) image A is chosen as the reference (normalization) image. Image similarity, based on the fidelity value, is evaluated as follows: A value close to 1 indicates that the two images are similar, while a value close to 0 indicates that the two images are not similar. In the reduction method we developed, the fidelity was used as a global processing criterion which is one of the three global measures which determine the similarity of two images, i.e. whether the two images are classified in the same class or not.

Cross-Correlation Coefficient

The cross-correlation coefficient is another similarity measure used in image processing. This coefficient evaluates the similarity of two images from the monotonic/linear relationship (dependence) point of view. The cross-correlation coefficient is more robust to changes of illumination (especially linear ones) than the MSE [46]. The mathematical expression of the cross-correlation coefficient ρ of images A and B is defined in Equation 22; where $\bar{A}$ and $\bar{B}$ are, respectively, the mean gray-levels of A and B.

Equation 21

$$\rho(A,B) = \frac{\sum_k (a_k - \bar{A})(b_k - \bar{B})}{\sqrt{\sum_k (a_k - \bar{A})^2 \sum_k (b_k - \bar{B})^2}}$$

Image similarity, based on the value of the cross-correlation coefficient, is evaluated as follows: A value close to 1 indicates that the two images are similar (correlated), while a value close to 0 indicates that the two images are not similar.

In the reduction method we developed, the cross-correlation coefficient was used as a second global processing criterion (alongside fidelity), which also evaluates image similarity considering the linear dependence aspect.

- **Mutual Information**

The mutual information (MI) is also a measure widely used in evaluating the similarity of two images. This measure relies on the probabilistic relation and the distribution of intensities in the images used. MI is highly robust to changes in illumination; even to non-linear ones [47]. The mathematical expression of the MI between A and B, of size m x n, is defined in:

Equation 23

$$MI(A,B) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{AB}(a_i, b_j)\log_2 \frac{p_{AB}(a_i, b_j)}{p_A(a_i)p_B(b_j)}$$

Where $p_A(a_i)$ is the probability that a pixel in A has a gray-level a_i, $p_B(b_j)$ is the probability that a pixel in B has a gray-level b_j, and $p_{AB}(a_i, b_j)$ is the probability that a pixel in A has a gray-level a_i and the same pixel in B has a gray-level b_j.

The probability distributions are determined using the normalized histograms of the images. In the histogram of an image, each possible gray-scale value has its corresponding number of occurrences in the image. Normalizing the histogram by the total number of pixels determines the probability distribution of the gray-scale pixel values of an image.

The higher the value of the MI, the higher the similarity between two images. The MI describes the stochastic association between the two images and evaluates image similarity from the probabilistic aspect.

In the reduction method we developed, the mutual information was used as a third global processing criterion (alongside fidelity and cross-correlation coefficient), which is also an image similarity measure with spatial alignment that evaluates similarity from the probabilistic point of view.

By analyzing these criteria, we can see that they evaluate image similarity considering different aspects. In our reduction method, the three *global* criteria should be satisfied at the same time in order to classify two images in the same class. Only one image from each class will remain in the final set of data, which is the one having maximal entropy (class representative). Getting rid of images may seem a risky operation, however the used criteria evaluate image similarity considering different aspects and having all three of them satisfied at the same time highly decreases the probability of losing any critical information. Therefore, the class representative image explains highly the vast majority of information within its class. The rare event constraint provided by the fourth criterion (MSE), which is implemented *locally,* also is an additional factor which prevents the loss of critical information. The use of these criteria also makes the reduction method highly robust against illumination changes which could occur during the acquisition of the hyperspectral image. The previous ideas are summarized in table 2.

Table 2 Criteria Analysis

Criterion	**Processing**	**Illumination Changes**	**Similarity Aspect**	**Used For**
MSE	Local	Very sensitive	Structural	Rare Event Preservation
Image Fidelity	Global	Sensitive	Structural	Similarity Measure
Cross-Correlation Coefficient	Global	Robust to Linear Changes	Linear Dependence	Similarity Measure
Mutual Information	Global	Very Robust	Probabilistic	Similarity Measure

Another important point to note is based on the analysis of these criteria from the hyperspectral point of view. These criteria are related to the quality criteria provided in [48], and relative to the different categories of quality criteria which were introduced in [43]:

- MSE and MI ↔ Criteria Relative to Information Loss
- Correlation ↔ Criteria Relative to Stability
- Fidelity ↔ Criteria Relative to Spectral Similarity

In conclusion, the criteria used cover various aspects of image similarity and hyperspectral image quality. Respecting these various aspects, provides a strong basis for reduction and promises efficient results.

- **The Entropy**

In information theory, entropy is the expected value (average) of the information contained in each message received. A source which has high entropy is a highly informative source. Entropy is expressed in bits, the value of entropy in bits also represents the minimum number of bits needed to code the data of a source without any loss. The mathematical expression of the entropy was provided in [49].

The entropy is maximum when all messages have equal probability. In this equiprobable case, the probability of each message is $1/n$ where n is the total number of messages. The mathematical expression in the case of maximal entropy, of a source X, is given in:

Equation 22

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) = -\sum_{i=1}^{n} \frac{1}{n}\log_2 \frac{1}{n} = \log_2(n)$$

The minimum value of the entropy is zero. Entropy is minimum when one of the outcomes is certain, i.e. has a probability equal to one, while all the other outcomes have a probability equal to zero. Events that always occur do not communicate information.

In our work we are interested in the entropy of images whose probability distributions are obtained from their normalized histograms. In our classification-reduction method, we used the entropy as a property of an image which determines which image is chosen to be the representative of its class. As mentioned before, the selection of the class representative was based on the image having maximal entropy.

Choosing the entropy as the property which determines the class representative has two main reasons:

- The images classified in the same class are similar and convey almost the same information. Out of these images, the one which has maximal entropy is the one which conveys the most information. Therefore, the image having maximal entropy is the most qualified image to be the class representative since it is the most informative one.

Image A

120	120	120	120
120	120	120	120
120	120	120	120
120	120	120	120

Image B

118	120	120	120
120	120	120	120
120	120	120	120
120	120	120	120

Figure 12 Gray-scale Pixel Values of Images A and B

A simple example is demonstrated in figure 12, two (4x4) highly similar images A and B. In image A all the gray-scale pixel values are 120; in image B all the gray-scale pixel values are 120 except for one detail represented by a value of 118. The histogram-based probability distributions of A and B is shown in figure 13. By calculating the entropies of images A and B based on Equation 24 we obtain $H(A) = 0\ bits$ (120 is the only possible outcome) and $H(B) = 1.72\ bits$.

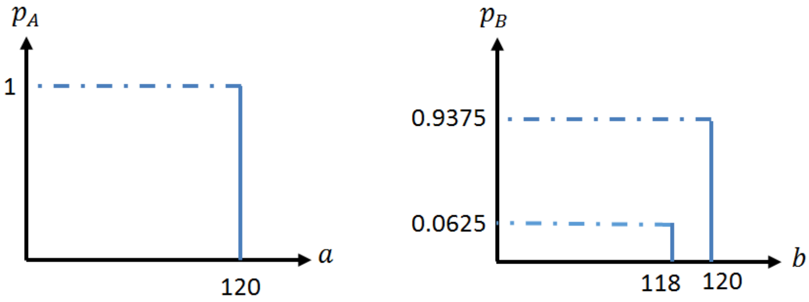


Figure 13 Probability Distributions of Images A and B

Based on the entropies calculated, image B is more informative than image A. The two images are very similar, however image B contains a small detail which is absent in image A. This is why in classification applications where images A and B are classified in the same class, choosing image B to represent its class is a better choice than choosing image A, in order to decrease the risk of losing any detail which may turn out to be critical.

- In our application, other statistical properties such as the variance or standard deviation may be also qualified to determine the representative image of a class since they also are indicators of the presence of details. These two measures indicate how far a set of data are spread from the average (mean), and the presence of details in an image increases their respective values. However, since our work is adapted to hyperspectral imaging, it is important to consider the possibility of outliers or illumination changes in acquisitions. As indicated in [40], properties which are based on the probability distributions are the most robust to outliers and illumination changes. Therefore, the entropy, which is a probabilistic property, is the most reliable property in our desired hyperspectral application.

3.1.1.3 **Detailed Algorithm and Flowchart**

For each criterion in our algorithm, there is a corresponding threshold in the vector $\vec{T}$. The values of these thresholds are very critical in the comparison and classification process. Therefore, the determination of threshold depends on the number of bands that we need after reduction and also depends on the application (some application need a high reduction rate and some need low reduction rate). In our reduction method there are two ways to determine the threshold:

1. The thresholds may be predefined by the user. The higher of threshold values are set, the higher of the final dimensionality and the lower is the resulting reduction rate, whereas relaxing the thresholds will result in a lower dimensionality in the reduced hyperspectral image and thus a higher reduction rate . However, experienced users in HIS may prioritize setting reduction standards in their applications .[48] Also, selecting the threshold depends on the user experiences in HIS.
2. The threshold is determined based on the desired reduction rates. In other words, in certain applications the final dimensionality is the ultimate specification and the goal is reaching a certain reduction rate. In the latter case, the user predefines only the final desired dimensionality without predefining thresholds. Therefore, the choice of

reduction rate is absolutely depending on the application. Some application needs low reduction rate (military, mines detection...), while some applications are less sensitive (agriculture, algae detection), therefore we choose a high reduction rate.

3.1.1.4 Detailed algorithm with predefined threshold

Consider the case of reduction of the original hyperspectral image $I(x, y, \lambda)$ having n_λ spectral bands. In the context of our algorithm and flowchart we consider the following definitions:

- r : is the representative image of the class.
- a_i : refers to the image of I having spectral index i where$1 \leq i \leq n_\lambda$.
- $rare\_preservation\_flag$: is a flag which determines whether rare events are preserved (1 ⇒ preserved ; 0 ⟹ unpreserved).
- MSE_{thresh} : is the threshold value of the MSE for local processing of rare events.
- b : represents the reduced hyperspectral image.
- $\vec{V}$ & $\vec{T}$ are the vectors which were defined in previous section.
- $IF, \rho , MI,$ and H refer to the fidelity, cross-correlation coefficient, mutual information, and entropy respectively.
- C_j : represents class number j. Note: this variable is unnecessary from the programming aspect, but we present it here for intuition purposes.
- i and j are counters. After reduction is complete, the final value of j represents n'_λ which is the new dimensionality of the reduced hyperspectral image.
- $verify\_rare\_events(r , a_i , MSE_{thresh})$ *:* is a function which verifies if the rare events (pixels which have a relevance > 0.5) are preserved between r and a_i for a value of MSE_{thresh}. The function returns 1 if they are preserved and 0 if not.

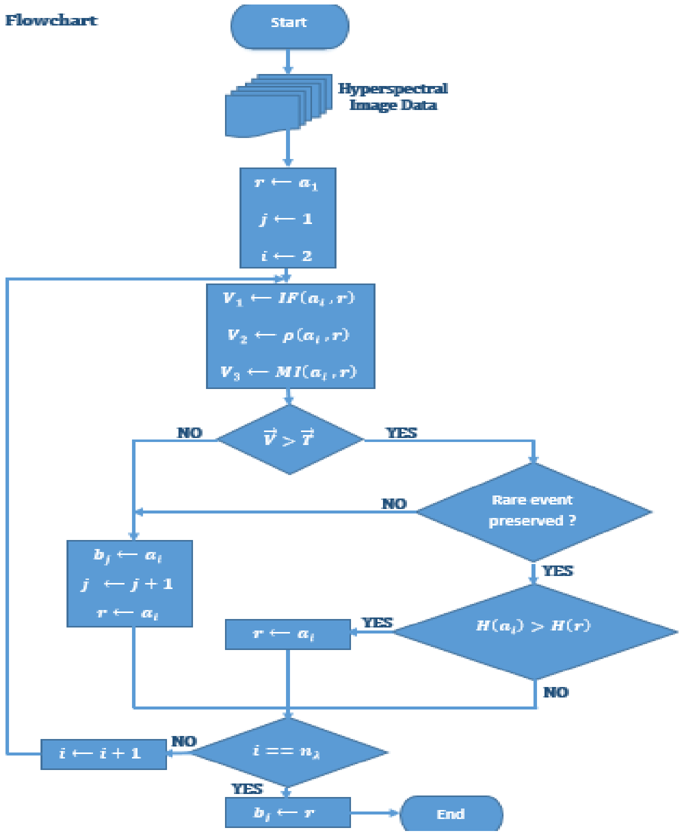
ALGORITHM

```
$r = a_1$ ;
$j = 1$ ;
```

```
$for\ i = 2 : n_\lambda$
    $V_1 = IF(a_i\ ,r);\qquad V_2 = \rho(a_i\ ,r);\qquad V_3 = MI(a_i\ ,r);$
    $rare\_preservation\_flag = 0\ ;$
    $if\left(\vec{V} > \vec{T}\right)$
        $rare\_preservation\_flag = verify\_rare\_events(r\ ,a_i\ ,MSE_{thresh})$
    $end$

    $if(rare\_preservation\_flag)$
        $x = [\ r\ ,a_i\ ]\ ;$
        $r = \underset{x}{\operatorname{argmax}}\, H(x)\ \ ;$
    $Else$
          $b_j = a_i\ ;$
          $j = j + 1\ ;$
          $r = a_i\ ;$
    $end$
    $\%classify\ a_i\ in\ C_j\ ;$
$end$
$b_j = r\ ;$
```

FLOWCHART



3.1.1.5 **Detailed algorithm with Iterative Determination of the Thresholds**

We developed an iterative algorithm which takes the desired dimensionality and the original hyperspectral image as input, and calculates threshold values which allow reaching the desired dimensionality.

We started this iterative algorithm by setting the thresholds to their higher values. Then in the first iteration, we decrease these thresholds based on the ratio between the desired dimensionality and the original one. This ratio is what determines the decreasing rate of the thresholds, the smaller the ratio the faster the thresholds are decreased and vice versa. At each iteration, we use the up to date thresholds to apply the reduction algorithm and determine the resulting dimensionality. If the dimensionality is higher than the desired one, we decrease the thresholds in the next iteration, whereas if the dimensionality is lower than the desired one, we increase the thresholds in the next iteration. The process is repeated until convergence, i.e. obtaining certain thresholds which yield the desired dimensionality.

As mentioned, the thresholds are updated based on the ratio between the desired and the initial dimensionality. The reason this ratio was used was in order to adapt the convergence to the input of the iterative algorithm and decrease the total number of iterations. Consider the following two examples where: n_λ and n'_λ are the initial and the desired dimensionalities respectively ($n'_\lambda < n_\lambda$) , x denotes the previously discussed ratio where $= n'_\lambda / n_\lambda$.

- Example 1 : $n_\lambda = 40 ; n'_\lambda = 10 \Longrightarrow x_1 = 0.25.$
- Example 2 : $n_\lambda = 40 ; n'_\lambda = 30 \Longrightarrow x_2 = 0.75.$

In the first example the desired dimensionality is far from the initial one, while in the second example it is near. Therefore, in the first iteration of the first example we need to decrease the thresholds faster than the second example in order to reach convergence faster and decrease the total number of iterations. So, our objective is to adapt the decreasing/increasing rates according to x and also update the thresholds at every iteration based on the current and desired dimensionalities. For this reason, we propose using a function f which takes x as an argument and operates based on Equation 25. Let $dim = n'_\lambda$ be the desired dimensionality to be reached, d_i the dimensionality in iteration i, and t_i the threshold values at iteration i.

Equation 25

$$t_{i+1} = t_i \times f(x) \text{ where } x \in]0\,;1[\text{ such that } \begin{cases} f(x) < 1 \;\; if\; d_i > dim \\ f(x) > 1 \;\; if\; d_i < dim \end{cases}$$

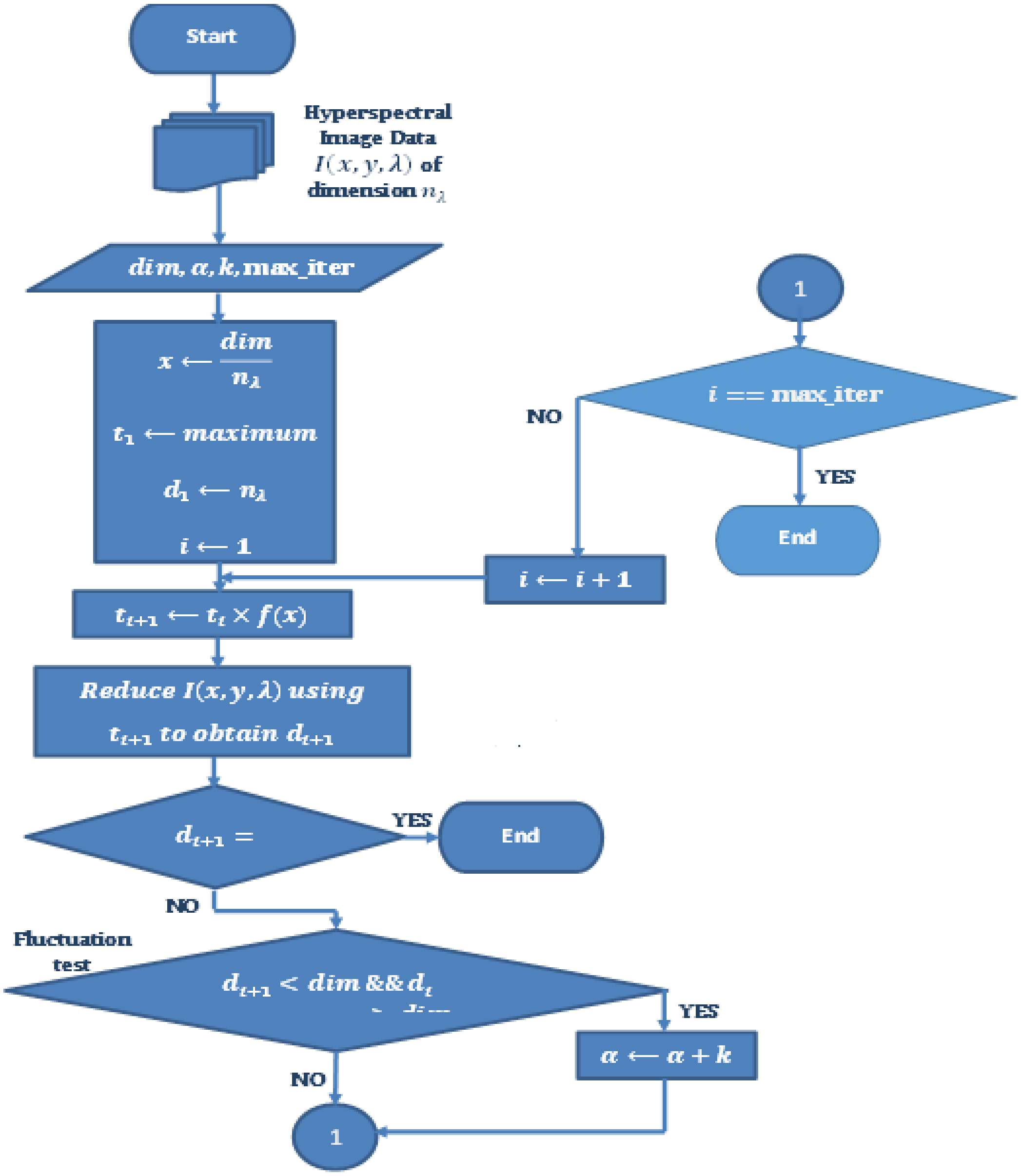
The two objectives of the definition in Equation 25 are: First, to increase the thresholds at iteration $i + 1$ if the dimensionality reached at iteration i is less than the desired one ($d_i < dim$), and to decrease the thresholds at iteration $i + 1$ if the dimensionality reached at iteration i was greater than the desired one ($d_i > dim$). Second, to adapt the decreasing/increasing rate to the value of x.

Using a suitable linear function is a possibility: $f(x) = \alpha x$ where $\alpha \in \mathbb{R}$. However, practice shows that non-linear models prove to be more efficient and adaptable. In our practice the exponential function provided better results than the linear, so f was chosen as defined in:

Equation 26

$$\boldsymbol{f(x) = \begin{cases} 1 - e^{-\alpha x}, \;\; d_i > dim \\ 1 + e^{-\alpha x}, \;\; d_i < dim \end{cases}}$$

In such iterative algorithms α is defined as the learning or convergence rate. In (22) smaller values of α will increase the decreasing/increasing (modification) rates of the thresholds. This may lead to achieving convergence faster, however it will also the increase the risk of instability. The original value of α is set by the user, in our algorithm every time d_i fluctuates about dim we increase the value of α by k in order to decrease the threshold modification rates (smaller steps). This will allow reaching the exact value of dim without fluctuating continuously about it which will result in the instability of the algorithm. We also set a maximum number of iterations, if the algorithm didn't converge within the maximum number of iterations then the final value of d_i reached will be the resulting dimensionality. The flow chart of the algorithm is provided below:



3.2 EXPERIMENTS

3.2.1.1 **Simulation Results**

The reduction algorithm we developed alongside the iterative determination of thresholds was tested and some results will be shown in this section. The Indian Pines, Group Of Fruits, and

Figure 14 Hyperspectral Scene (a) Indian Pines Scene (b) Group of Fruits Scene (c) Female Face Scene

Female face three hyperspectral data samples provided in http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes ,

http://www.imageval.com/scene-database-4-faces-1-meter/ and /sites.google.com/site/hyperspectralcolorimaging of the scenes presented in figure 14 (a), (b) & (c) respectively.

The results obtained in table 3, 4 and 5 have important indications. Our reduction algorithm was successful in reducing these images to various predefined dimensions. The iterative algorithm converged on all occasions with low numbers of iterations. The dimensions of the three hyperspectral images are:

- Indian Pines : (145x145x220)$\Longrightarrow n_\lambda = 220$. Spectral ranges between 400 and 2500nm with spectral resolution of 10nm.
- Group of Fruits: (1312x1924x49)$\Longrightarrow n_\lambda = 49$. Spectral ranges between 400 and 950nm with spectral resolution of 10nm
- Female Face: (1377x1008 x148)$\Longrightarrow n_\lambda = 148$. Spectral ranges between 400 and 1900nm with spectral resolution of 10nm.

The images above have a spectral range between 400 and 2500nm with a spectral resolution of 10nm.

Below are results obtained by implementing our dimensionality reduction algorithm on these three data samples by predefining the values of n'_λ. IF, ρ and MI are the image fidelity, cross correlation and mutual information thresholds respectively.

$\alpha = 4,\ k = 2, \text{max\_iter} = 100$, and fixed $MSE = 3\%$

Table 3 Indian Pines Simulation Results

n'_λ	IF	ρ	MI	**Number of iterations**	**Convergence**
165	**0.97**	**0.97**	**3.88**	**5**	**Yes**
110	**0.78**	**0.78**	**3.11**	**7**	**Yes**
90	**0.7**	**0.7**	**2.8**	**13**	**Yes**
55	**0.52**	**0.52**	**2.1**	**5**	**Yes**

Table 4 Group of Fruits Simulation Results

n'_λ	IF	ρ	MI	**Number of iterations**	**Convergence**
36	**0.999**	**0.999**	**3.97**	**18**	**Yes**
24	**0.99**	**0.99**	**3.96**	**15**	**Yes**
20	**0.985**	**0.985**	**3.94**	**13**	**Yes**
12	**0.808**	**0.808**	**3**	**6**	**Yes**

Table 5 Female Face Simulation Results

n'_λ	IF	ρ	MI	**Number of iterations**	**Convergence**
111	**0.988**	**0.988**	**3.76**	**8**	**Yes**
74	**0.84**	**0.84**	**3.72**	**7**	**Yes**
60	**0.78**	**0.78**	**3.5**	**15**	**Yes**
37	**0.6**	**0.6**	**3.12**	**3**	**Yes**

The number of iterations was low for the reduction of the Indian Pines hyperspectral image. For $n'_\lambda = 110$, which is equivalent to a reduction rate of 50%, relatively high levels of fidelity (78%), cross-correlation coefficient (78%), mutual information (3.88) were maintained. As for the Group of Fruits hyperspectral image, the results indicate that their sets of images are less spectrally redundant than that of the Indian Pines which due to the low number of bands ($n_\lambda = 49$). This is due to the fact that very minor modifications in the thresholds caused huge reduction rates such as the case where $n'_\lambda = 24\ (50\%\ \text{reduction rate})$. This particularity of the Group of Fruits hyperspectral image led to an increase in the number of iterations for the lower reduction rates (higher n'_λ). As for Female Face hyperspectral image with $n'_\lambda = 74(50\%\ \text{reduction rate})$, the results show high level of fidelity (84%), cross-correlation coefficient (84%), mutual information (3,72) which means that the sets of images are more spectrally redundant than Indian Pines and Group of Fruits hyperspectral images.

In order to carry out the experimental analysis aimed to assess the effectiveness of the proposed approach; we compared the performance of our algorithm with other existing dimensionality reduction algorithms.

3.2.1.2 Comparison Properties and Methodology

In order to compare dimensionality reduction techniques with respect to each other, some performance metrics (comparison properties) are needed. The values of these properties must be calculated for the reduced hyperspectral image ($n'_\lambda < n_\lambda$), and compared with the reference image of dimension n_λ. We propose using the three following performance metrics for this purpose.

1. **The Total Spectral Standard Deviation**

Each pixel in a hyperspectral image has a corresponding spectral vector. The values of the pixels vary with respect to the wavelengths present in the hyperspectral image. The standard deviation of a spectral vector, which corresponds to a certain pixel, indicates how far the brightness values of this pixel are spread from the mean value. Throughout the reduction process, the standard deviations of the spectral vectors in the hyperspectral image may vary. This variation has important indications about the reduction process. Consider the case where in the reduced hyperspectral image, the standard deviations of the spectral vectors increase or maintain their values compared to the original reference image. In such a case the reduction is considered efficient, since an increase in the standard deviations indicates that values close to the mean (redundant) were the ones removed. Maintaining the original standard deviation after reduction also is considered as a satisfying result for similar reasons. While in the case where the standard deviations decreases, the opposite reasoning is true, and thus the lower the standard deviations the worse the result of the reduction technique is. The total spectral standard deviation is given by Equation 26 and is based on [49].

Equation 23

$$\sigma(I) = \sum_{x,y} \sqrt{Var\big(I(x,y,:)\big)}$$

Where $I(x,y,:)$ is the spectral vector of the pixel having spatial coordinates (x,y) in the hyperspectral image I.

2. The Total Spectral Entropy

The concept of the total spectral entropy is similar to that of the total spectral standard deviation. However, this time the entropies of the spectral vectors are calculated instead of the standard deviations. The general concept of the entropy was explained in the previous section: in this application the higher the total entropy is, the more the set of spectral vectors are informative. Therefore, it is important to monitor the total entropy throughout the reduction process in order to evaluate the performance of the various dimensionality reduction algorithms. The total spectral entropy is given by:

Equation 27

$$\mathrm{H(I)} = \sum_{\mathrm{x,y}} \mathrm{H}\big(\mathrm{I(x,y,:)}\big)$$

Where $I(x, y, :)$ is the spectral vector of the pixel having spatial coordinates (x, y) in the hyperspectral image I.

3. The Average of the First Order Spectral Image Derivatives

This concept is the one provided in [46] in the proposed approach in previous section. the average of the first order spectral derivatives relative to spectrally adjacent images in I is calculated. The obtained value $D(I)$ evaluates the redundancy of the hyperspectral image data. The lower its value, the higher is the level of redundancy. Therefore, a reduction technique which maintains a high value of $D(I)$ is better at exploiting the redundancy present in images from adjacent spectral bands.

4. Comparison Methodology

The reduction techniques are compared based on the values of the three-performance metrics described above. Suppose using a certain reduction method a hyperspectral image $I(n_\lambda)$ was reduced to another hyperspectral image $\tilde{I}(n'_\lambda)$ where $n'_\lambda \leq n_\lambda$. The properties of $\tilde{I}(n'_\lambda)$ are compared to that of $\tilde{I}(n_\lambda)$ by calculating the respective ratios of the performance metrics. $\tilde{I}(n_\lambda)$

Represents the reference image for comparison, our goal is to monitor the properties of $\tilde{I}(n'_\lambda)$ with respect to $\tilde{I}(n_\lambda)$ as n'_λ varies. $\tilde{I}(n_\lambda)$ is obtained by applying the reduction algorithm on $I(n_\lambda)$ while keeping the same dimensionality n_λ. In the context of our developed reduction technique, $\tilde{I}(n_\lambda)$ is equivalent to $I(n_\lambda)$, however in other techniques which are based on subspace projection, these two images are different. The ratios to be calculate are the ones in:

Equation 24

$$r_\sigma = \frac{\sigma\left(\tilde{I}(n'_\lambda)\right)}{\sigma\left(\tilde{I}(n_\lambda)\right)} \quad ; \quad r_H = \frac{H\left(\tilde{I}(n'_\lambda)\right)}{H\left(\tilde{I}(n_\lambda)\right)} \quad ; \quad r_D = \frac{D\left(\tilde{I}(n'_\lambda)\right)}{D\left(\tilde{I}(n_\lambda)\right)}$$

Note that n_λ is fixed and represents the original dimensionality of the hyperspectral image, while n'_λ varies by taking the value of the desired dimensionality. As n'_λ varies the ratios in (Equation 28) vary, the various reduction techniques are compared by analyzing the ratios defined in (Equation 28) for certain values of n'_λ. For example, consider the case of comparing two dimensionality reduction methods A and B by reducing a certain hyperspectral image ($n_\lambda = 100$) to $n'_\lambda = 20$ (80% reduction rate). Suppose the results obtained for $n'_\lambda = 20$ are the ones shown in table 6.

Table 6 Reduction Results for 〖n'〗_λ=20

$n'_\lambda = 20$	r_σ	r_H	r_D
Method A	**0.98**	**1**	**0.8**
Method B	**0.7**	**1**	**0.8**

The results of table 6 indicate that for this particular dimension($n'_\lambda = 20$) the performances of methods A and B are similar in terms of the total spectral entropy and the average of the first order spectral image derivatives. However, the values of r_σ indicate that for a reduction rate of 80% method A was capable of maintaining 98% of the original total spectral standard deviation, while method B only maintained 70%. So, the performance of method A is superior to that of method B in terms of the total spectral standard deviation. The previous example was provided to illustrate the comparison methodology which was implemented.

An additional step is still to be noted in the comparison methodology. Before calculating the ratios in (Equation 29), the hyperspectral data is standardized (zero mean and unit variance) so that the comparison between the techniques becomes valid and unbiased.

Equation 25

$$\frac{\mathrm{X} - \mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}}$$

The mean of the set of data (μ_X) is subtracted from every instance of the data, and the difference is divided by the standard deviation of the data (σ_X). Hence, $\tilde{I}(n'_\lambda)$ and $\tilde{I}(n_\lambda)$ are standardized before the calculation of the ratios in (Equation 28). The standardization is done image by image, whereby every image in the set of images of the hyperspectral data is standardized. This means that all the images in the data to be compared will have a zero mean and a variance/standard deviation equal to one.

We also compared the performance of our algorithm with other existing dimensionality reduction algorithms in terms of the preservation of known rare event present in the scene. A quantitative measure based on the relevance of $\boldsymbol{X_i}$ will be used to assess the rate of preservation of this rare event $\boldsymbol{X_i}$. This measure, denoted as T ($\boldsymbol{X_i}$), is calculated as the ratio of relevance before the reduction R($\boldsymbol{X_i}$) $_{\boldsymbol{Before}}$ and the relevance after the reduction R($\boldsymbol{X_i}$) $_{\boldsymbol{After}}$ .This rate of preservation is defined by:

Equation 30

$$T(X_i) = \frac{\mathrm{R}(X_i)_{Before}}{\mathrm{R}(X_i)_{After}} \quad \forall\, \mathrm{i} = \mathrm{l}, \dots, \mathrm{n}$$

$$\boldsymbol{T(X_i)} = \begin{cases} \boldsymbol{T(X_i) = 1\ \ The\ rare\ event\ \ X_i\ is\ fully\ preserved} \\ \boldsymbol{T(X_i) > 1\ \ The\ rare\ event\ X_i\ is\ not\ \ preserved} \end{cases}$$

3.2.1.3 Results: comparative study with the existing methods

The dimensionality reduction techniques, provided in [50], which we used in our comparative study are:

PCA (Principal Component Analysis)
MDS (Classical Multidimensional Scaling)
SPE (Stochastic Proximity Embedding)
Diffusion Maps
EMPCA (Probabilistic PCA)
SNE(Stochastic Neighbor Embedding)
Fa(Factor Analysis)
Cfa(coordinate factor analysis)
NPE (Neighborhood Preserving Embedding)
LPP (Locality Preserving Projection)
Kernel PCA
Sammon Mapping
Sym SNE (Symmetric SNE)
Isomap
LAPLACIAN_EIGEN
GPLVM(Gaussian Process Latent Variable Model)

The comparative study was performed on three hyperspectral images, which are: Indian Pines, Group of Fruits and high definition Female Face. The results of the ratios in (26) were simulated for 75%, 50%, and 25% reduction rates for the various reduction techniques.

- Indian Pines, Group of Fruits and hight definition Female Face hyperspectral images have an original dimensionality $n_\lambda = 220, n_\lambda = 49, n_\lambda = 148$, respectively.So the reduction rates of 25%, 50%, and 75% correspond to values of :
- n'_λ Equal to 165, 110, and 55 for Indian Pines.
- n'_λ Equal to 36, 24, and 12 for Group of Fruits.
- n'_λ Equal to 111, 74, and 37 for high definition Female Face.

The results are shown in the bar graphs of figure 15; the horizontal axis represents the various techniques simulated, and the vertical axis represents the ratios.

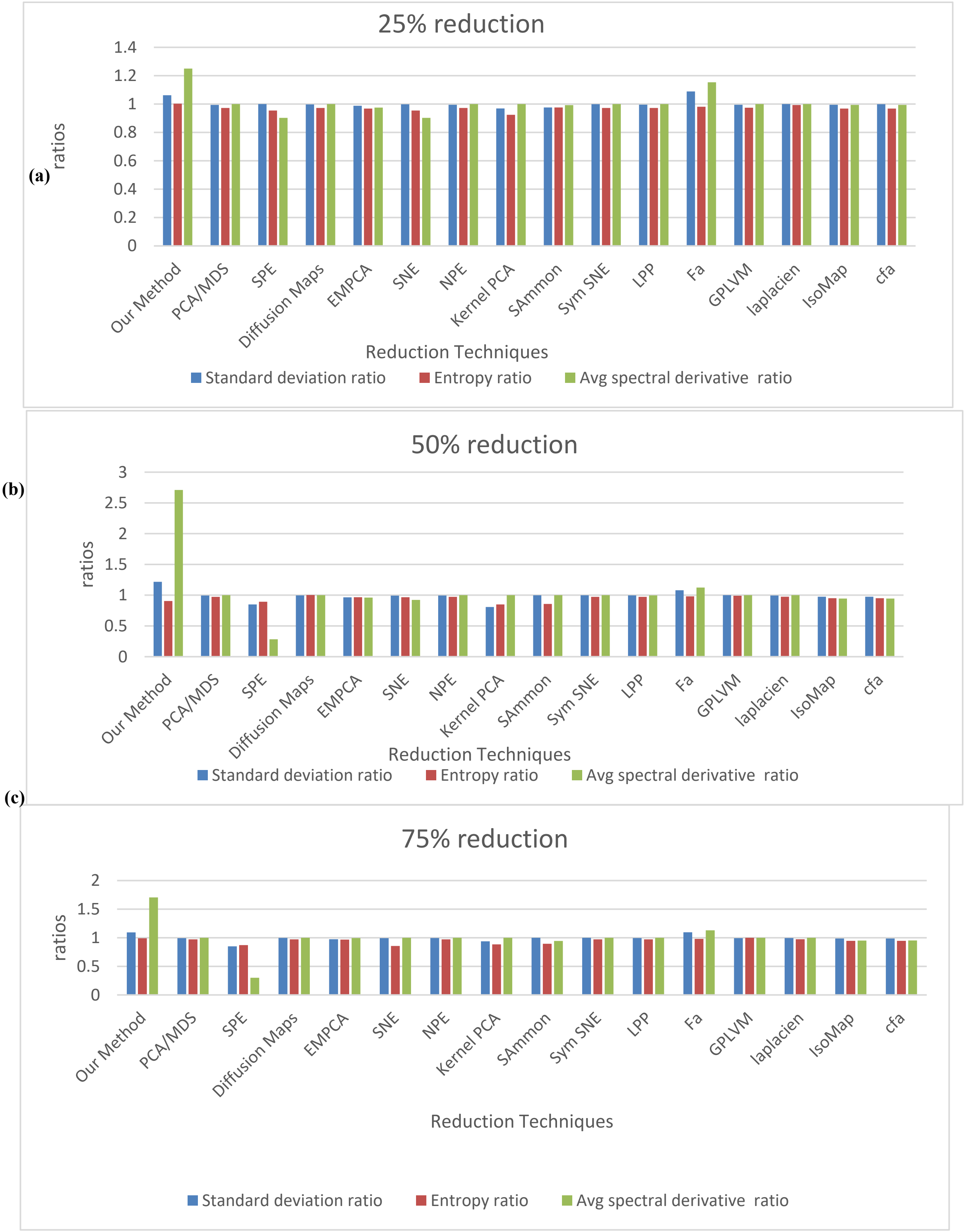


Figure 15 Results for Indian Pines Scene, (a) 25% reduction , (b) 50% reduction, (c) 75% reduction

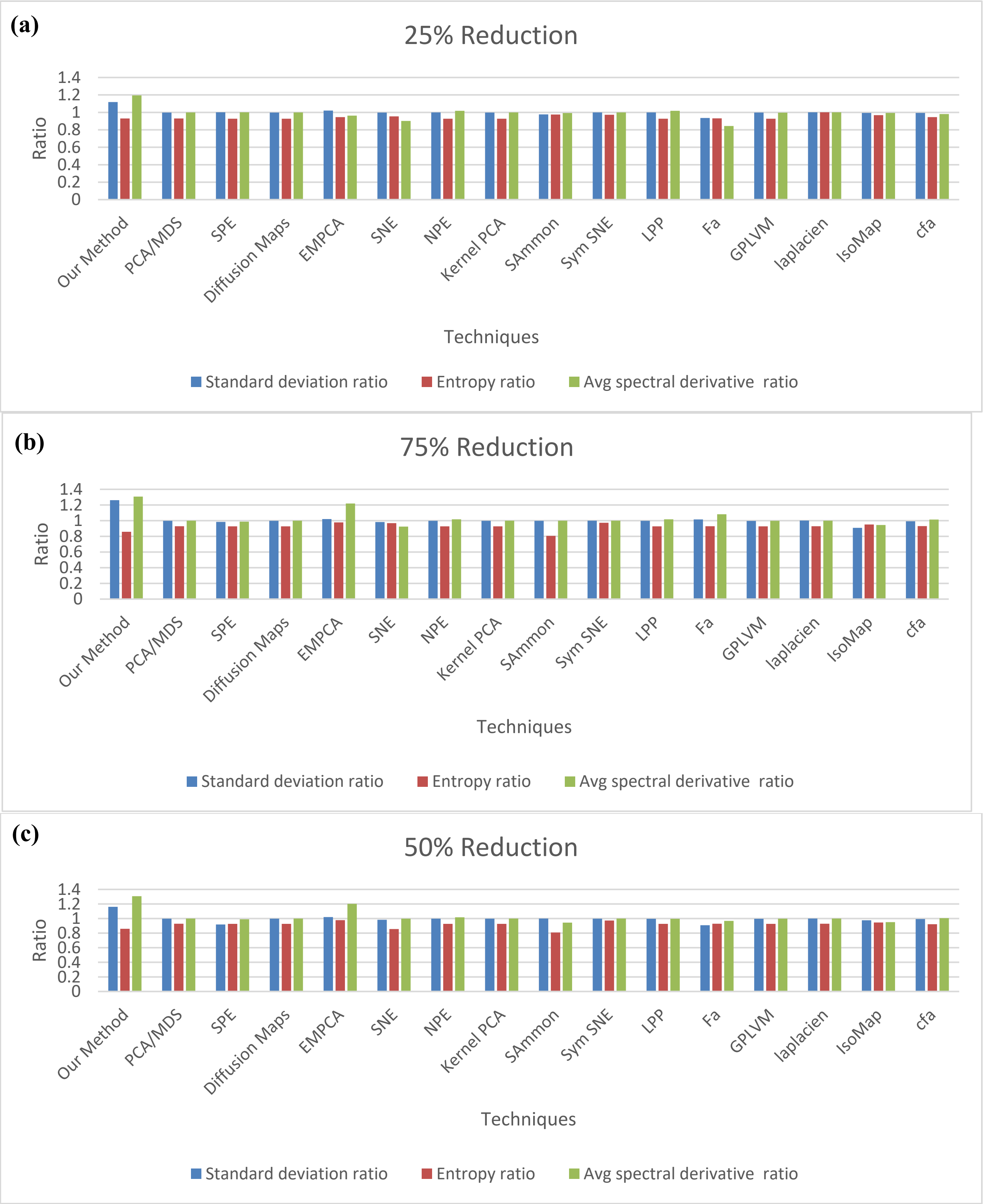


Figure 16 Results for Group of Fruits Scene (a) 25% reduction, (b) 50% reduction, (d) 75% reduction

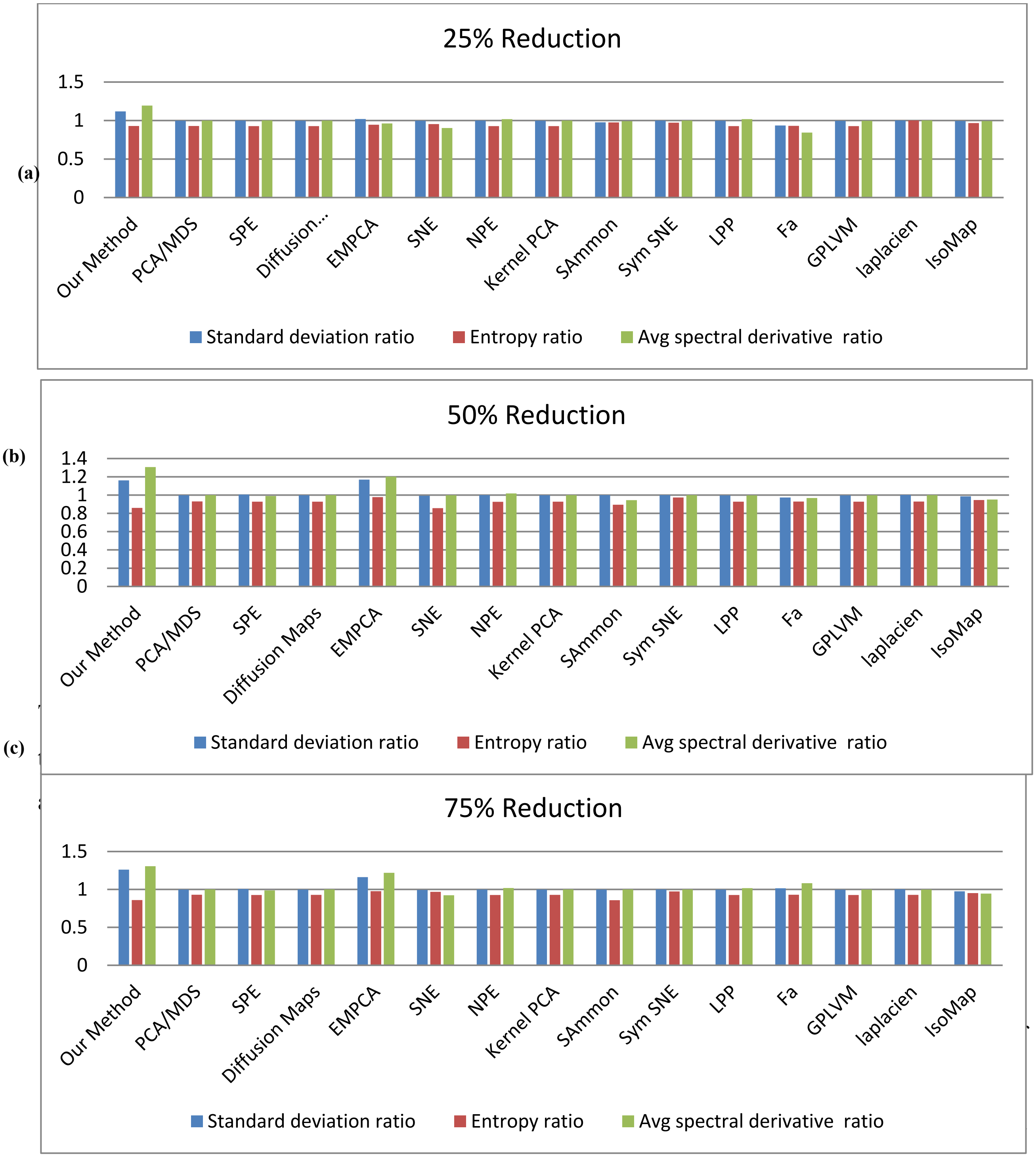


Figure 17 Female Face Scene (a) 25% reduction, (b) 50% reduction, (c) 75% reduction

The results of Figs. 15–17 present the values of rσ, rH, and rD for various reduction techniques with three different images, including our multicriteria reduction algorithm. By analyzing the results of each ratio for the three reduction rates, we can conclude the following:

- Standard deviation ratio rσ: for a reduction rate of 25%, the performances of the techniquesin the three hyperspectral images are almost alike except the Fa reduction method in the female face image, which has a slight advantage for our method. A reduction rate of 25% is considered a low rate of reduction, so the techniques did not lose their potential of maintaining high percentages of the original standard deviation. For a reduction rate of 50%, some techniques start losing their performance: SPE, kernel PCA, and Cfa for the Indian pines; SPE, Fa, and Isomap for group of fruits; and Sym Sne, Fa and Cfa for female face image. As for a reduction rate of 75%, which is considered a high reduction rate, the performances of the SPE, kernel PCA, Cfa, Fa, Isomap, and Sym Sne remained the poorest in the three images. The Sammon and SNE were successful in maintaining their potential

  regardless of the high reduction rate. Our method proved to have the best performance since its ratio value was the highest; the ratio also increased for this reduction rate, which means that our method was successful in eliminating the data that were close to the mean (redundant data), which led to an increase in the total spectral standard deviation.
- Entropy Ratio r_H : for a reduction rate of 25%, the performances of the techniques are almost the same in the three types of images. For a reduction rate of 50%, the SPE and Kernal PCA have the poorest performance for Indian Pines, while the Sammon and Fa have the poorest for Group of Fruits and Female Face images respectively. While the GPLVM, EMPCA and multi-criteria methods have the best one for Indian Pines, Group of Fruits and Female Face images respectively. The performances of the other methods are almost the same. As for a reduction rate of 75%, for Indian Pines image, the ratio maintains the same as the case of 50% reduction for all the methods except the SNE method which begin the poorest one. With group of fruits image, all the methods maintain the same ratio and the Sammon method has the worst performance. For the Female Face image, the ratio of Fa is the worst one and the other methods have the same ratios as 75% . The best performance in terms of entropy remains that of the GPLVM, SNE and Isomap

for Indian Pines, Group of Fruits and Female Face images respectively. Our method remains competitive as it ranks third in the three types of images.

- Average fist order spectral image derivative ratio r_D : for a reduction rate of 25%, the performances of the techniques are almost the same in the three images, except the Fa method in the Female Face image, with an advantage for our method. This advantage is clearer in the case of 50% reduction, where the ratio increases to reach a value of 1.6 for Indian Pines and 1.3 for Group of Fruits and Female Face images. The ratio for the SPE decreases strictly for the Indian Pines which signifies a high level of redundancy within the images of the reduced data compared to the original one. Also, a small decreasing in ratio in Sammon technique with Group of Fruits image and a small value of ratio for the Fa in the Female Face image can be observed. As for a reduction rate of 75%, the ratio for our method further increases to reach a value of 2.65 for Indian Pines and 1.3 for Group of Fruits and 2.2 for the Female Face image. The change in the ratios for other techniques is negligible. Our method proves to have the best performance in terms of spectral image derivative, the reason behind this is that our algorithm is image processing oriented, whereby one image was taken for each class of similar images.

3.2.1.4 Rare event Preservation Rate results

The comparative study was performed on two natural hyperspectral images presented in figure 18 provided in https://**sites.google.com/site/hyperspectralcolorimaing/dataset/general-scene** and http://**personalpages.manchester.ac.uk/staff/david.foster/**.

The first image (Figure 18(a)) has a size of **1312*1903*49** and it contains a colors bar rare event were the second image (Figure 18(b)) has a size of **1017*1338*33** and it contains a flower rare object. It is important to mention that rare objects in this two image are larger than a single pixel. This is not a problem in our method, because we evaluate locally each pixel to obtain if this pixel is rare or not.

Figure 18 Scene with rare event (a) Colors bar rare object in natural scene, (b) flower rare object in natural scene

To evaluate the performance of our reduction method in terms of preservation of rare event we proceed as follows: the two rare objects present in figures 18(a, b) are known before reduction, so we can determine if the selected rare objects are preserved after reduction. The results of the preservation rate were simulated for different reduction rate from 5% to 95% for the various reduction techniques. The results are shown in the bar graphs of figure 20, the horizontal axis represents the various techniques simulated at different reduction rate, and the vertical axis represents the preservation rate.

The results of figure 19 present the preservation rate of rare event in the scene for various reduction techniques with two different images, including our multi-criteria reduction algorithm. By analyzing the results of each method, it is clear that the multicriteria reduction method has the same performance at different reduction rate, were the other reduction methods lose their performance when the reduction rate is larger than 25%. So, our method guaranteed the preservation of rare event with high reduction rate.

So, our method guaranteed the preservation of rare event with high reduction rate.

To further test the performance of our method in case of preservation of rare events we use a new geological hyperspectral image taken from: http://rsipg.dii.unipi.it with a size of 1024*320*512. 7 rare target are deployed in figure 20. We use three of the 7 deployed targets (white panel, V1 and P1)

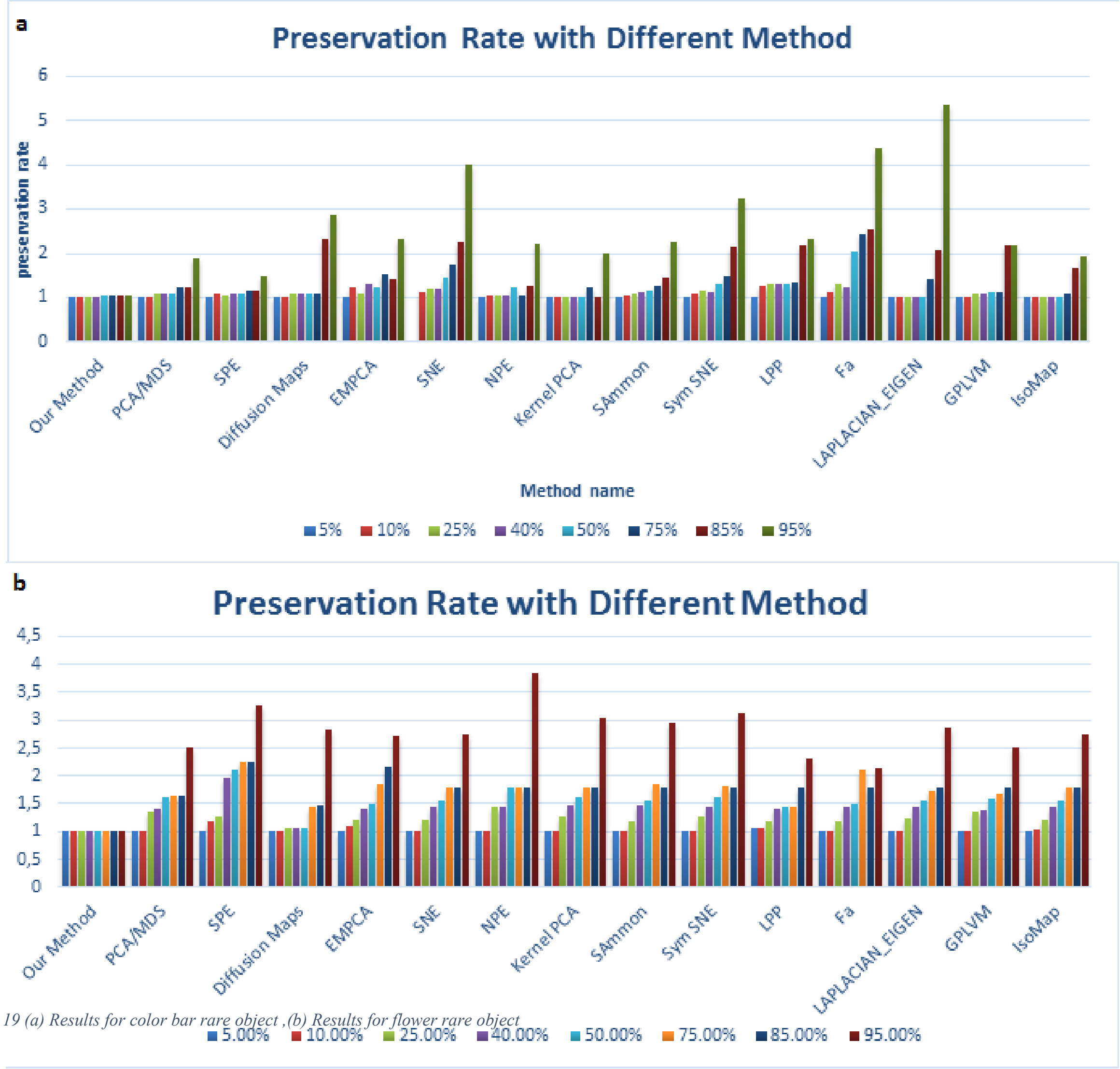


Figure 19 (a) Results for color bar rare object ,(b) Results for flower rare object

Figure 20 Geological image with three rare event

There are three different rare events present in Figure 20 represented by three areas:

1. Area represented by the blue frame(white panel)
2. Area represented by the orange frame(p1)
3. Area represented by the green frame(v1)

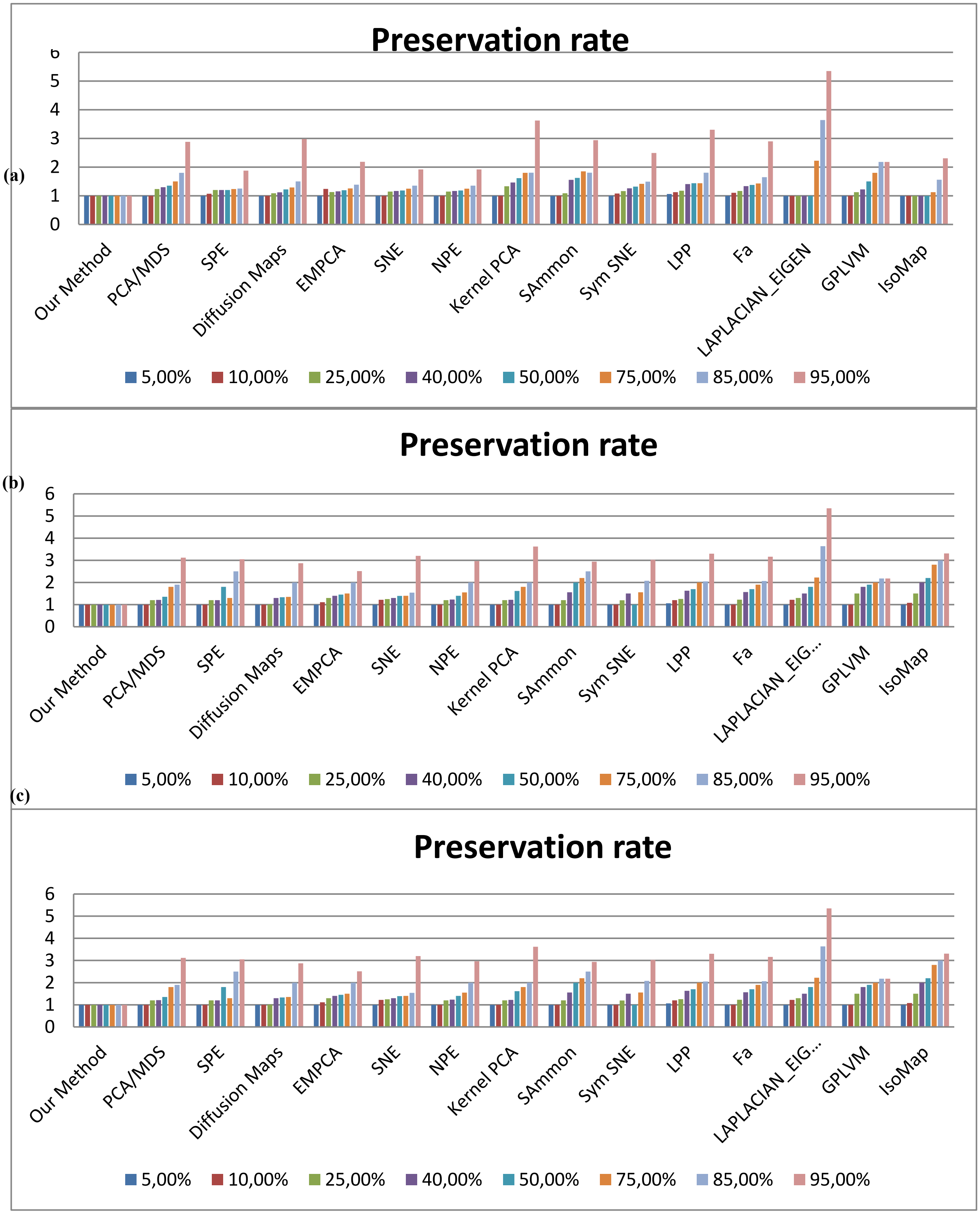


Figure 21 (a) Results for the area located at the blue frame , (b) Results for the area located at the orange frame, (c) Results for the area located at the green frame

As we expected, in Figures 21 our developed method can preserve more than one rare event present in the same hyperspectral image, where all the other method start losing rare event at 25% of reduction.

➢ Elapsed Time

Another comparative study was performed to evaluate the effectiveness of our method based on the elapsed time taken by each method during reduction. The elapsed times were simulated for the different reduction techniques when applied to the Group of Fruits image . The results are shown in the bar graphs of figure 22 were the horizontal axis represents the various techniques, and the vertical axis represents the elapsed time taken by each method to provide a three reduction rate (25%,50% and 75%) of each image.

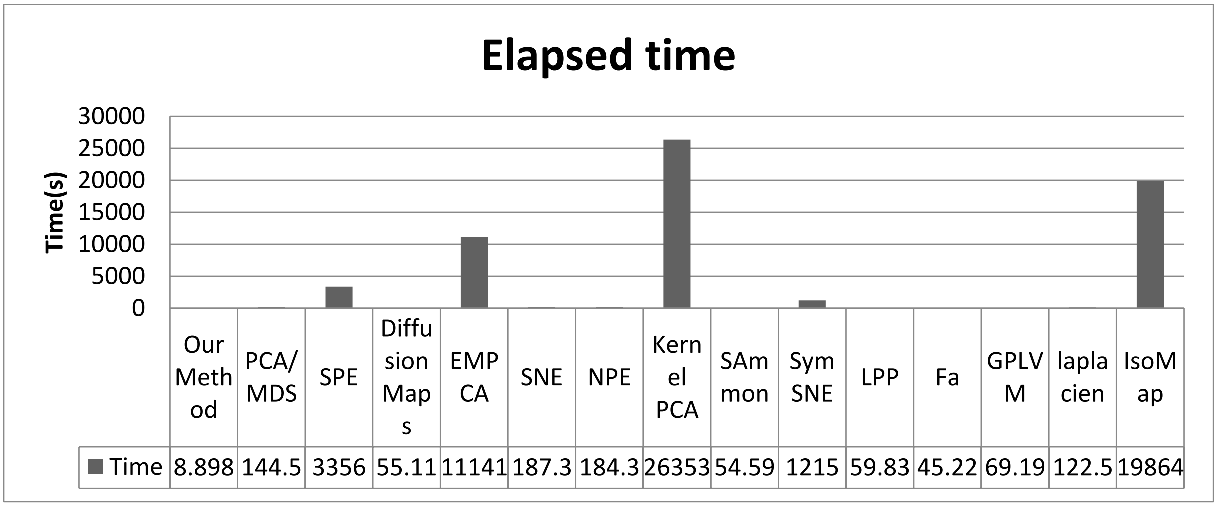


Figure 22 Elapsed Time

By analyzing the results from figure 22, it is clear that EMPCA, Kernal PCA and IsoMap take more time to finish the three reduction rate, were our method has the minimum elapsed time (8.89s) to obtain the three reduced images.

➢ multicriteria method performance evaluation using K-means clustering

To better appreciate the performance of the proposed method, a new performance evaluation was added in terms of the K-means classification result before and after reduction.

Two different images are used and taken from https://sites.google.com/site/hyperspectralcolorimaging/dataset/:

- Figure 23 (a): dimension : 1312*2374***49**
- Figure 23(b): dimension: 1312*2498***49**

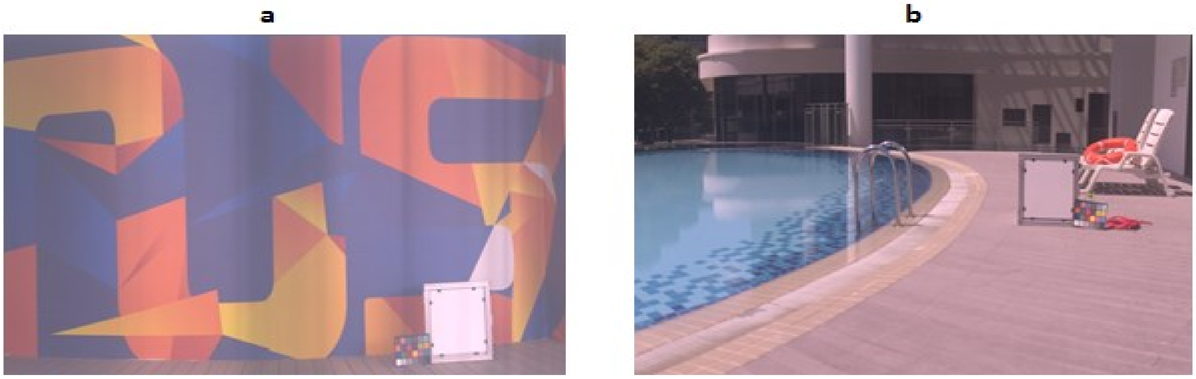


Figure 23 Hyperspectral image (a) scene07, (b) scene08

For reduction rate estimation, the Laplacian eigenmaps intrisinc dimentionality was used and taken from DR toolbox at *drtoolbox* [22] and used before apply our reduction method. 67% of reduction is estimated for scene07 and 65% is estimated for scene08. A K-means clustering method was applied to the represented image before and after 67% of reduction, in order to evaluate the classification accuracy at the pixel level [51] [52]. The classification accuracy allows us to assess how the image elements are changed after the reduction process.

Four classes were used for scene07 clustering, and 5 classes for scene08 clustering. The results of classification are shown in figures 23(a, b) and 24(a, b).

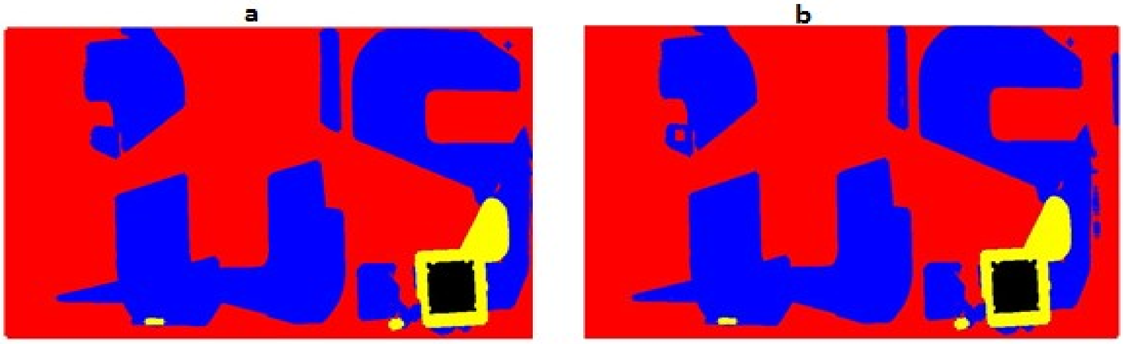


Figure 24 K-means clustering of scene07, (a) before reduction, (b) after 67% of reduction

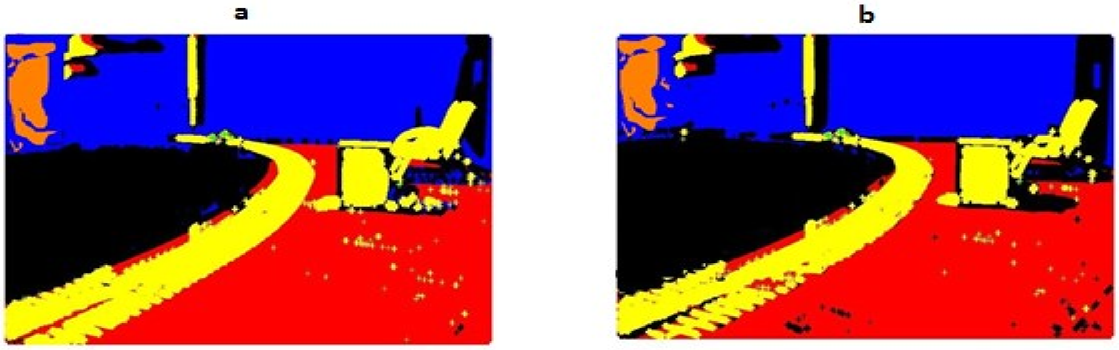


Figure 25 K-means clustering of scene08, (a) before reduction, (b) after 65% of reduction

The zones black , yellow and red are appear at the place of window in the scene07 , because of the reflection into the glace of this window, whereas figure 25(a) shown, it is clear that the water, land and soil are appear reflected into the glace.

The classification accuracy is used to evaluate if the classification process is the same after reduction. And the results are shown in the table 7 below:

Table 7 Classification accuracy of two represented scene

Hyperspectral image	**Classification accuracy (CA %)**
Scene07 with 67% reduction rate	98.88%

Scene08 with 65% reduction rate	97.97%

(CA %) is evaluated based on the difference between each pixel, before and after reduction. So, with high accuracy we have a high similarity between pixels before and after reduction, therefore we have a reduction without losing information.

Table 7 has an important indication. Our developed method can achieve a high reduction rate (67 and 65%) maintaining a high classification accuracy. So, the developed method can produce high reduction rate without losing information about the structure and the elements in the image. Also, from figure 24and 25, we can note that the rare event (colorbar) is preserved and correctly classified after reduction.

After clustering and determining classification accuracy, the RGB bands are selected (figure26) to show how the image appears after the reduction and to verify if we have lost information after reduction.

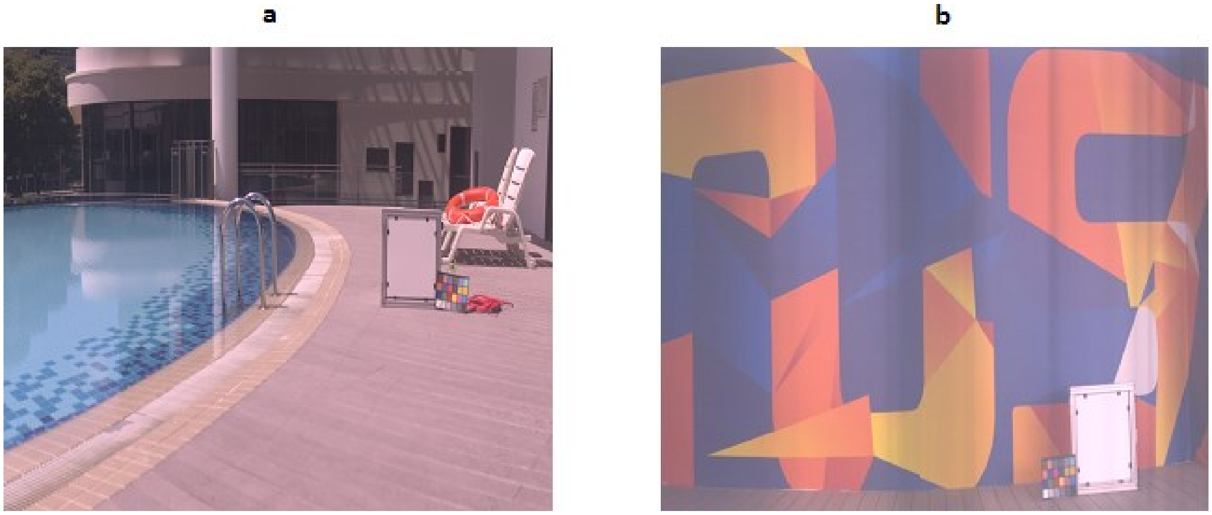


Figure 26 RGB representation after reduction, (a) scene07, (b) scene08

According to figure 26, we have no significant changes in the image structure after a high reduction rate, and the rare objects (colorbar) present in the two images are preserved. So, our reduction method can achieve high reduction rates without losing rare events.

3.3 Conclusion

Our goal was to develop a competitive dimensionality reduction algorithm adapted to hyperspectral images. All the scientific aspects which led to developing our algorithm were interpreted. The comparative study proved that our reduction algorithm is highly competitive. The method we developed has several advantages. First, it is simple and not time consuming, this feature is very attracting for space remote sensing hyperspectral applications where storage capacity is limited, and time is valuable. Second, the method is fully adapted to the nature of the data it reduces. The reduction algorithm treats the hyperspectral image as a cube formed by a certain number of spectrally adjacent images. The approach is image processing oriented, and the reduced data are intuitive images (explainable) which can be observed by the eye without any inverse processing technique. This particular trait is essential in several applications such as biomedical ones where the needed information is collected by observing the images. In other projection-oriented reduction techniques such as the PCA, the resulting reduced data is not at all intuitive when observed by the human eye. Such techniques require inverse processing before being ready to be used. Third, our method also treats the concept of rare events by local image processing. Preserving rare events, while reducing a hyperspectral image, is very important especially in detection applications. In such applications scientists look for ‘fingerprints’ which may be present in certain spectral bands which allow identifying and detecting objects. These ‘fingerprints’ or spectral signatures often appear in very small areas of the images (rare events). So, preserving these rare events is very essential in such applications. Fourth, the algorithm has multi-criteria which means that it evaluates scientific properties from multiple aspects, robustness to illuminations changes during acquisitions is one of its positive consequences.

The method we developed showed great potential; however, there is always room for improvement. Testing the algorithm on many hyperspectral data samples may reveal the need for more optimization tasks. And the comparative study can still be broadened to include more performance metrics.

Chapter 4

Hyperspectral imaging used for landmine detection

In this chapter we first describe several applications using hyperspectral imaging. Then, we describe some projects that have investigated the use of hyperspectral imaging in landmine detection and identification, to show the possible relevance of this technique in this particular application field. In addition, several state-of-the-art detection techniques are reviewed and explained in detail. Therefore, we propose to combine the most performant detection algorithm with both a rare event detection method and a suitable dimensionality reduction technique in order to improve the landmine detection performance. The performance of the proposed method is evaluated in both a real and a simulated landmine experiment.

4.1 Hyperspectral imaging applications

Today, the richness and accuracy of geo-localized information is a key factor in the development of decision support in many applications. Through processing methods, it is possible to highlight and interpret relevant information regarding observed phenomena, both locally and on a large scale. The result produced is then a digital thematic map. In the following sections, we give a brief survey of different application using hyperspectral imaging. Of course, the use of hyperspectral imaging is not limited to these applications, however the described ones can be considered the most relevant.

1 Precision Agriculture

The controlled management of farms allows both an increasing yield and reduced negative environmental impact. Hyperspectral airborne remote sensing helps to the achievement of this compromise [53]. It intervenes in:

- The estimation of harvests
- Early detection of diseases
- Soil characterization

- Controlled treatment of plots

2 Algae detection

The marine environment is a valuable indicator of the health of our environment. The study of algae provides a lot of useful information for assessing and correcting the impact of human activity. Multispectral imaging allows the study of this environment, sometimes difficult to access, by listing the different species of algae [54]. The campaigns studies carried out by the CEVA in 2007 and 2008 made it possible to observe precisely the distribution of four dominant species in the Jaudy mudflat (Côtes d'Armor) in Brittany.

3 Medical Sciences

For medical diagnostic applications, Hyperspectral imaging provides highly resolved means for imaging tissue at either a macroscopic or macroscopic level, and provides highly accurate spectral information about the patient, the tissue sampling, or the status of a disease[55]. Hyperspectral imaging can be used for the following applications:

- Imaging for non-invasive diagnosis.
- Optical biopsy
- Tissue Delineation
- Therapeutic Analysis

4 Remote sensing

It is also known as chemical detection, offering researchers and biologists to perform airborne and stationary spectral analysis for applications of remote sensing [56]. The airborne hyperspectral representation represents a technique of remote sensing established to capture critical important spectral data. A hyperspectral detector can simultaneously acquire accurate information for all wavelengths through the full available spectral range. With the creation of the hyperspectral cube, a data structure that includes all spatial and spectral information, researchers have in-depth spectral information that can be used for several purposes:

- Civil and Environmental Engineering

- Environmental Monitoring
- Pollution Detection
- Forest Management
- Mineral Exploration

4.2 Landmine detection projects using hyperspectral imaging

Our goal in this section is to describe some projects on landmine detection using hyperspectral imaging to clarify the importance of this technique in this field.

4.1.1 Canadian teleoperated landmine detection systems [57]

This project focuses on systems developed primarily for the detection of anti-personnel (AP) landmines. The Articulated Robotic Scanner (ARS) is a system application that utilizes generic robotics. It is a device capable of automatically moving a landmine detection sensor over natural soil surfaces, mimicking the behavior of an operator. Using an increased-mobility robotic device Platform, the proposed system integrates five separate innovations: two hyper-spectral technologies, Cameras, thermal and visual infrared, along with a scanning sensor imaging system. Mounted on a purpose built articulated robotic scanner, working in tandem with a robotic scanner. Confirmation sensor for nuclear imaging. The end goal is not only to operate in all environments and conditions, but also to extend the demining capabilities of the military leaders and humanitarian demining professional developers in situations where there is a high probability of accidents.

4.1.2 Joint Multispectral Sensor Program (JMSP)

The aim of the study [58] is to test the development of multi-spectral and hyperspectral imagers that can improve detection efficiency in accordance with target detection terms and procedures. Target needs to be detected in both sunlight and nighttime situations. Panchromatic or multispectral images in the VNIR and SWIR ranges provide this functionality during daylight. Even so, for military use, the MWIR and LWIR ranges are essential, especially for night operations. The authors concluded that thermal multispectral images provide a better target detection and false alarm rate compared to a single infrared (IR) sensor band after evaluating dual bands of the MWIR-and LWIR-range. Testing found that selected bands were suited for effective detection, with an acceptable range of 8 to 10.5 micrometers. The usage of LWIR with MWIR is greatly improved compared to the use of MWIR alone. The authors created a new hyperspectral image called SEBASS which runs from 2.9 to 5.2 microns and 7.8 to 13.4 microns thanks to the results obtained. The Aero Corporation also uses this camera to capture remote images in the two ranges.

4.1.3 NATO project

A collaboration with NATO has been carried out in the Netherlands to develop a remote landmines detector [59]. The main goal was to achieve minefield detection almost in real time throughout a conflict using a standard unmanned aerial vehicle (UAV) at 100 m above mean sea level. The researchers first described the accessible imaging technologies at the time: radar, radiometers of microwave, wavelengths visible, near, center, and long infrared. The writers then illustrated the key methods of signal processing at the moment for mine detection. The principal measures represented by image correction, detection of edge, segmentation, classification, and feature extraction. Based on various experimental results, authors have several recommendations:

1. Conventional medium-resolution radar imaging is less suitable for remote mine detection.

2. The Remote Mine Detection Microwave Detection Theory is promising.

3. Visible and near-infrared imagery characteristics are also requested. It is because these bands are often low cost, lightweight and imaging devices that have a high spatial resolution and can be used in real time.

4. A promising band for remote mine detection is the medium or long-wave infrared wavelength.

5. Because weather conditions (such as rain showers) can make it difficult to detect mine and minefield in mid- and longwave infrared wavelength bands, multiple wave longitude bands would best be combined.

4.1.4 TELOPS test to detect buried object using airborne thermal hyperspectral images

In 2015, the prospect of detecting buried objects using an airborne LWIR hyperspectral imager was demonstrated by a Canadian research company specializing in infrared and hyperspectral imaging called TELOPS [60]. They obtained thermal hyperspectral images of areas which contain historically buried man-made objects from an aircraft platform. They found that the disturbed soil right above a buried target is warmer than the neighboring undisturbed soil area [60]. Through comparing the emissivity data obtained through the distinction between temperature and emissivity, the buried target sites appear as part of the hottest ground within the scene, but further classification or additional information is needed to discriminate the targets.

4.1.5 Equinox Corporation fusion test

The combination of visible and SWIR samples could yield better results. A simple fusion of two bands of spectrum provides appropriate background segmentation of artifacts, regardless of the illumination conditions. In other words, it has been found that selecting a set of two or three spectral image bands is just as efficient in differentiating man-made objects from the background as using all spectral bands at once [61]. These fusions have the potential to use an integrated camera with visible and SWIR sensors and more sophisticated and advanced detection algorithms to detect mine-like objects in an image.

4.1.6 Hyperspectral Mine Detection program HMD

Recently the American army initiated the "Hyperspectral Mine Detection Phenomenology Program" (HMDP) project. Its main objective was to define spectral characteristics useful for identification of landmines [62]. They therefore collected high-quality hyperspectral signatures of surface materials and mines, assessed temporal impacts on buried landmines, and reported, under variations of regulated variables, a statistically significant collection of hyperspectral signatures of surface and buried landmines in natural soils. The results of HMDP project are presented in [63].

The researchers concluded that the findings are usually influenced by unregulated variables, primarily wind and rainfall. The mines influenced by further rainfall tend to show a noticeably different signature distribution from the past. It is also interesting that the temporal evolution of vegetation around landmines is too complex and makes it exceedingly difficult to describe the evolution of temporal signatures. The following general observations were made: (1) A light shower does not substantially reduce the signature; (2) The signature is reduced by half an inch of rain; (3) one inch of rain further decreases the signature but does not eliminate it.

4.1.7 Indian Test to detect landmines using infrared images

Researchers in India have proposed a hierarchical algorithm to detect landmines from infrared images consisting of preprocessing (contrast enhancement- filtering- smoothing), segmentation, extraction of features, and classification based on artificial neural network ANN [64]. The authors evaluated the algorithm into two types of soil on surface-laid mines: black cotton and sand. The picture is transformed to gray level during preprocessing. Contrast enhancement and noise reduction are the two most critical preprocessing steps. Segmentation is the process of grouping homogeneous pixels in an image, sharing common attributes such as color, intensity, or texture. The key three groups covering the numerous current image segmentation techniques are clustering, edge detection and threshold-based area growth [64]. So, extraction of features and further processes are applied to clusters considered to be mine-like. The mine is distinguished from the surrounding area using a Neural Network (NN) based algorithm. The writers used a tiny NN of 1 hidden layer and 4 neurons during the experiments. The results given on a simple dataset are good, but it is not anticipated that the algorithm will work well on another field or soil type as the data used during the learning process are not rich enough.

4.3 Target detection in hyperspectral imaging

In the literature, we find different algorithms used for target detection using hyperspectral images [65, 66]. Generally, we can classify those algorithms in two types: Supervised algorithms where we search for a known spectrum that characterizes the target; Unsupervised algorithms, known as anomaly detectors, where pixels different from their surrounding are marked as targets. The latter type could be used in a preliminary stage to reduce the working area before identifying the target using a supervised method. In this thesis, we will use some rare event detectors in the first step in order to detect the suspected pixels and then apply in the second step a target detection algorithm. We are taking into account different cases of full-pixel and subpixel targets.

Numerous algorithms were proposed for target detection in hyperspectral imaging. Some of these algorithms do not support the detection of several targets at the same time. So, the algorithm should be applied several times, each run searching for a specific target. But this will be a time-consuming process especially if the number of targets is high. Few algorithms have been extended to the multitarget case, like Constrained Energy minimization (CEM). This algorithm is able to estimate the abundance of a target in a hyperspectral image. Many algorithms are extended from CEM e.g. MultiCEM, SumCEM, Winner-take-all CEM (WTACEM) and others [67].

4.3.1 Target detection algorithms used for hyperspectral imaging

In this section, a brief review of the classification algorithm usually employed for target detection in Hyperspectral imagery is presented. Target detection in Hyperspectral imagery is not limited to these algorithms. However, these algorithms are the mostly used ones.

Generally, we can divide these algorithms into two main groups:

Algorithms based on a statistical model, where the spectral variability of the signatures is assumed to be following a probability density function (usually Gaussian). Therefore, the problem is posed as a hypothesis testing problem competing between the two hypotheses [68]:

H0: target absent (Background material) $x\sim N(\mu_b,\sigma^2 \Sigma_b)$

H1: target present $x\sim N({aS}_t, {\sigma^2 \Sigma}_t)$

Where μ_b is the background mean vector, S_t is the target spectrum, Σ_t and Σ_b are the variance-covariance matrices of background and target respectively, a and σ are scaling factors.

The other types of algorithms are based on a linear mixing model where each pixel is represented by a linear combination of signatures of pure materials called endmembers with an additional noise [68].

Equation 31

$$x=a_1 s_1+a_2 s_2+\cdots+a_m s_m+n$$

$$a_k\geq 0, a_1+a_2+\cdots+a_m=1$$

Where sk is the reflectance spectra of endmembers k, ak is the fill fraction (also named abundance factor) of k endmembers and n represents the noise.

4.3.2 Adaptive Coherence Estimator

This method is based on the statistical model. It is derived from the generalized maximum likelihood ratio test[69]:

Equation 26

$$\boldsymbol{L(x)}=\frac{\boldsymbol{f(x/H_1)}}{\boldsymbol{f(x/H_0)}}$$

Where f is the conditional l probability density function of the pixel x given hypothesis H. To simplify the derivation of the maximum likelihood estimation, a monotonic function (logarithmic) is used. By deriving this ratio and replacing the values of the scaling factors by their maximum likelihood estimates, we arrive to the following ACE detector:

Equation 33

$$\boldsymbol{y_{ACE}}=\frac{\boldsymbol{x^T\Sigma_b^{-1}S\left(S^T\Sigma_b^{-1}S\right)^{-1}S^T\Sigma_b^{-1}x}}{\boldsymbol{x^T\Sigma_b^{-1}x}}$$

Where x is a vector representing the reflectance spectrum of the pixel under test. S vector containing the signature of the target Σ_b is the variance-covariance matrix of the background.

4.3.3 Matched Filter (MF)

This is another algorithm based on the statistical model. It is derived in the same way as the adaptive coherence estimator but in this case, the background and target are considered to have the same covariance matrix. By expanding the GLRT, we arrive to the following detector[70]:

Equation 34

$$y_{MF=\frac{S^T\Sigma_b^{-1}x}{S^T\Sigma_b^{-1}S}}$$

S: Target spectrum x: pixel spectrum Σ_b: background variance-covariance matrix.

4.3.4 Constrained Energy minimization (CEM)

The CEM algorithm is based on finding a finite impulse response filter that maximizes the output for a target while minimizing the output power[71].
The output of the filter is given by:

Equation 35

$$y = w^T x = \sum_{i=1}^{L} w_i x_i$$

w_i are the coefficients of the filter and x is the vector spectrum of the pixel. The coefficients of the filter are calculated such that the output must be 1 if the target was introduced to the filter $S^T w = 1$. The average output power is given by:

Equation 36

$$\frac{1}{N}\sum_{i=1}^{L} y_i^2 = \frac{1}{N}\sum_{i=1}^{L}(x_i^T w)^T x_i^T w = w^T\left(\frac{1}{N}\sum_{i=1}^{L} x_i x_i^T\right)w = w^T R_{LxL} w$$

R_{LxL} is the autocorrelation matrix of the samples. The solution is found using Lagrangian method[72]:

Equation 37

$$y_{CEM} = \frac{S^T R_{LxL}^{-1} x}{S^T R_{LxL}^{-1} S}$$

4.3.5 Multiple Target CEM (MTCEM)

MTCEM is a CEM based algorithm used to detect multiple targets. Let us consider a matrix D= [S1, S2…Sp] that contains the signature of p targets. The main idea is to minimize the output energy with the constraint $D^T. w=\mathbf{1}$. Where **1** is a pixe1 column vector of ones. The solution is given in [71] by:

Equation 38

$$y_{MTCEM} = (R_{LxL}^{-1} D (D^T R_{LxL}^{-1} D)^{-1})^T . x$$

4.3.6 Winner take all CEM (WTACEM) and Sum CEM (SCEM)

Also, Winner take all CEM and Sum CEM multitarget detection techniques are based CEM. These algorithms work on running CEM each time for each target. Then the results of the detectors are summed up in SCEM or we take the maximum among other detectors in case of WTACEM. One advantage of WTACEM over SCEM is that if the results are noisy, the noise is not summed up in the final detector.

4.3.7 Orthogonal Subspace Projection (OSP)

This method is based on the linear unmixing model. It necessitates the knowledge of the target spectrum in addition to the spectral signature of the background materials. The signatures are treated as vectors. The idea is to calculate the projection of the pixel on the orthogonal subspace of the background materials in the direction of the target. This method is derived in [73] and the final detector is given as follows:

Equation 39

$$P_U^{\perp} = I - U(U^T U)^{-1} U^T$$

Equation 40

$$\boldsymbol{P_{OSP} = P_U^{\perp} d}$$

Equation 41

$$\boldsymbol{y_{OSP} = {P_{OSP}}^{T} x}$$

4.3.8 Spectral Angular Mapper (SAM)

In this method, the spectral signatures of the pixel and the targets are treated as vectors. The test pixel is more similar to the target if the angle between the vectors is smaller. The angle is calculated as follows[71]:

Equation 42

$$\boldsymbol{y_{SAM} = cos^{-1}\left(\frac{\vec{x}.\vec{S}}{\|\vec{x}\|\|\vec{S}\|}\right)}$$

x represents the spectrum of the pixel. S is the spectrum of the target.

4.3.9 Spectral Information divergence (SID)

This another technique based on the calculation of similarity between test pixel and target signature. The similarity is calculated as in information theory using the relative entropy formula[74]:

Equation 43

$$\boldsymbol{D(x||y) = \sum_{l=1}^{L} p_l log\left(\frac{p_l}{q_l}\right)} \ \ and\ \boldsymbol{p_l} = \frac{x_l}{\sum_{l=1}^{L} x_l}.$$

The detector is given by: $\boldsymbol{SID(x,y) = D(x||y) + D(y||x)}$

4.4 Rare event detection

Some rare objects or targets appear only in a limited number of bands[43, 75]. Therefore, we choose to analyze each pixel value in each band instead of the average value of a pixel in all bands, in order to be able to detect the presence of rare events even in few bands. When considering the average value of a pixel in all bands, small rare objects having a small difference with respect to their neighbors can be difficult to detect. Also, experimental results confirm that working on one pixel value for each band performs better than the average value in terms of detection of rare objects. So, we use three different similarity methods at the pixel level to detect rare objects that appear in the image.

We can define an anomaly or a rare event as an object in the scene that differs in a substantial way from the surrounding objects[43]. An example is given in Figure 27.

Figure 27 Image Containing a Rare Event

Pixel intensity (PI) represents the amount of gray level in the pixel, i.e., its brightness [43]. Using this method, we compute how much the intensity of a pixel is different from that of surrounding pixels. Also, PI is the square root of the squared error criterion [23]. The equation is given by:

Equation 44

$$PI = \left|(\boldsymbol{a_i} - \boldsymbol{\bar{\mu}})\right|$$

Where a_i is the pixel value and $\bar{\mu}$ is the average of the pixel values in a neighborhood of a_i.

Let us consider a hyperspectral image $I(x, y, \lambda)$ having n_λ spectral bands. In the context of our algorithm we consider the following steps:

- Scan each pixel in each band.
- Compute the similarity measure between pixel and neighborhood (figure 28).
- If the similarity measure is above a certain threshold, the pixel is rare and then we store its position.
- Finally, after all bands are scanned, we get a dataset containing the rare pixels positions.

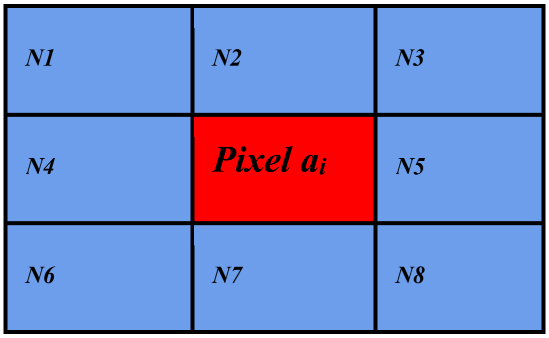


Figure 28 Representation of pixel Aij and its neighborhood Nk

When using this neighborhood, we are assuming that the size of rare objects is just a single pixel. In our scenario this is fine because the size of a landmine usually does not exceed one pixel. However, if the resolution of images increases and targets become larger than a single pixel, we can compute the similarity considering a larger window for the target and a subsequently larger neighborhood.

The above algorithm returns only rare objects in the image, without any indication concerning the type of each target. Therefore, we apply one of the best performing existing detection algorithms to the set of rare pixels to classify them.

there are several available algorithms for target detection in hyperspectral images. Several studies demonstrate that ACE is the best performing algorithm in terms of probability of detection and False Alarm rate [76, 77]. However, ACE has a high computational cost compared to other methods. In this study, ACE method is applied after the PI method. This will reduce the complexity of ACE and may improve its performance in terms of False alarm and probability of detection.

4.5 Material and Method

a. Simulated landmine experiment

Our dataset consists of 17 hyperspectral image scenes acquired using AVIRIS sensor and processed as in [78]. The scenes are originally composed of 224 bands between 394 nm and 2500 nm with spectral resolution of 10 nm. We applied several image preprocessing steps including the elimination of water absorption bands and low pass filtering to remove the bands that have a spectral response above 1nm. We simulated in different locations of the scenes the spectral signatures of the 6 types of landmines (figure 29) shown in figure 30. These signatures were collected using an ASD field spectroradiometer. Acquisition were made in our laboratory using a special source of illumination. This source is specified by the 70-watt quartz-tungsten-halogen light source with integrated reflector, creating stable illumination over the 350 to 2500 nm range. The targets were simulated in different proportions using the following formula:

Equation 45

$$\boldsymbol{PS = \alpha * T + (1 - \alpha)B}$$

where PS is the simulated spectrum in the image, T is the target reflectance spectrum represented as a vector, B is the reflectance of background material and α is the target fill fraction between 0.1 and 0.9.

Using these scenes, we tested the performance of ACE method and ACE method combined with the pixel intensity measure. We will evaluate the performance of these two algorithms based on three criteria:

Detection Rate (DR): which is the ratio between the number of detected targets and the number of targets in the scene. In the landmine scenario missing detection can be very dangerous, so the detection rate should ideally be one.

False positive rate (FPR): this is the number of false positives per unit area. This value counts the average number of non-target pixels detected as target per unit area (m^2). This value is dependent on the image size. To be more general, and to remove the effect of image resolution, we use also the **Relative False Positive Rate (RFPR)**. This value is the ratio between the number of false positives and the sum of true positives and false positives.

Equation 46

$$\text{FPR}=\frac{FP}{image\ resolution*(pixel\ size)^2}$$

Equation 47

$$\text{RFPR}=\left(\frac{FP}{TP+FP}\right)*100$$

True positive (TP): target pixel correctly detected as target

false positive (FP), non-target pixel wrongly detected as target

<u>Computation time (CT):</u> CPU time registered in seconds needed to execute the algorithms.

The measurement of the detection rate and FPR depends on the threshold chosen in the classification phase to discriminate between target and non-target pixels. When comparing different detection methods, the Receiver Operating Curve (ROC), visualizes the detection rate in terms of FPR for different choices of the threshold. However, in our scenario, due to the low presence of targets, ROC curve is not the ideal method of evaluating the detection performance. Since in our scenario missing a target can be very dangerous, the threshold is set in such a way as to detect all targets and then the FPR is calculated. Therefore, an algorithm is more efficient if it has a lower FPR and lower computational time when all targets are detected (DR=1).

Figure 29 6 different landmines

b. Real Landmine Experiment

This experiment consists in a scene presented in figure 31. An area containing soil, grass, wood, water and multiple material is prepared to be more realistic.

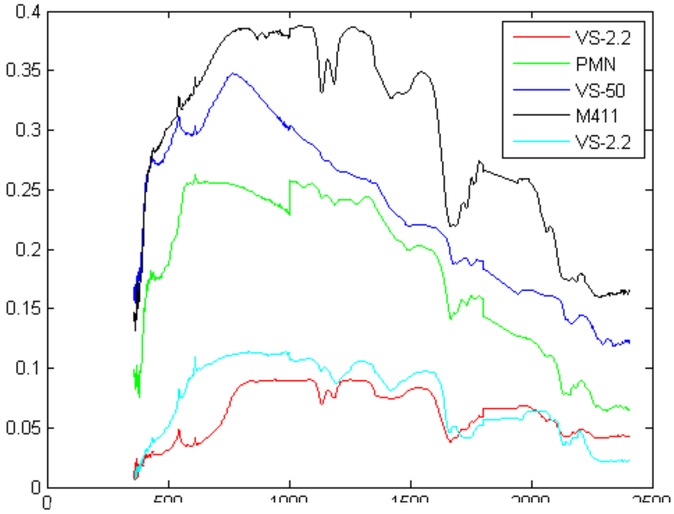


Figure 30 Landmine Spectra

Figure 31 presented scene

A CNC router is used to handle the spectrometer and scan the whole area. A Hamamatsu spectrometer is used with a spectral range between 350 and 1700 nm and a spectral resolution of 3nm.

In the reproduced scene 2 types of landmine are used (Figure 32): a plastic landmine and a metal landmine For a realistic simulation we used two containers with the same material as the real landmines (the metal landmine represents the cluster bomb and the plastic one represents the vs50 landmine) . In addition, the same explosive powder is used in the two types of landmines. This powder is a type of TNT explosive with nitrating mixture (conc. sulfuric & conc. nitric acid).

Figure 33 shows some landmine distribution cases, where landmines are deployed in different positions in order to mimic a realistic scenario. Using the explosive powder is important in our experience because in most cases, when the landmines stay a long time in the area, the powder material may seep from landmines.

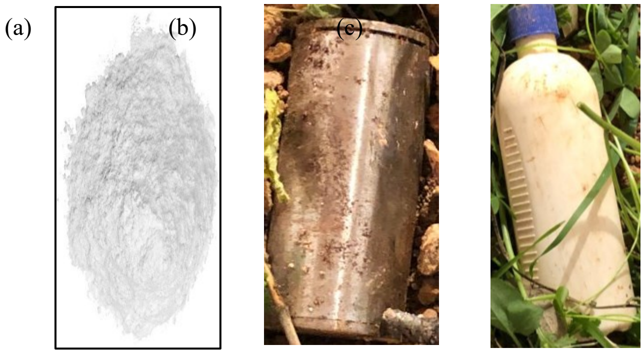


Figure 32 (a) explosive powder. (b)metal landmine. (c) plastic landmine

Figure 33 landmines distribution scenarios

4.6 Results and discussion

c. *Comparative study between ACE and PI+ACE*

In this part, we show the results obtained when applying the two used methods on all images that contain targets with different abundances. We show the average false alarm rate and computation time at full target detection rate. The experiments were carried out in MATLAB environment under Intel AMD operator (TM) 6276 with 64 CPU and 2.9 GHz processor frequency with 128 GB RAM. Comparisons are made between PI+ACE and ACE in the two experimental settings previously described.

Simulated landmine experiment

As discussed above, the decision threshold used to detect rare pixels and the other thresholds to identify the landmine types for the existing detection techniques are set to be the lowest value such that all targets are detected (DR=1) and then the FPR is registered. Therefore, a technique is said to be more efficient if it has lower FPR when all targets have been detected. In order to set the threshold, we use a dataset of 5 images for training and we use the remaining 12 images for testing. For each landmine abundance factor, we compute a different threshold. Hence, we measure DR and FPR on the testing dataset using the same threshold.

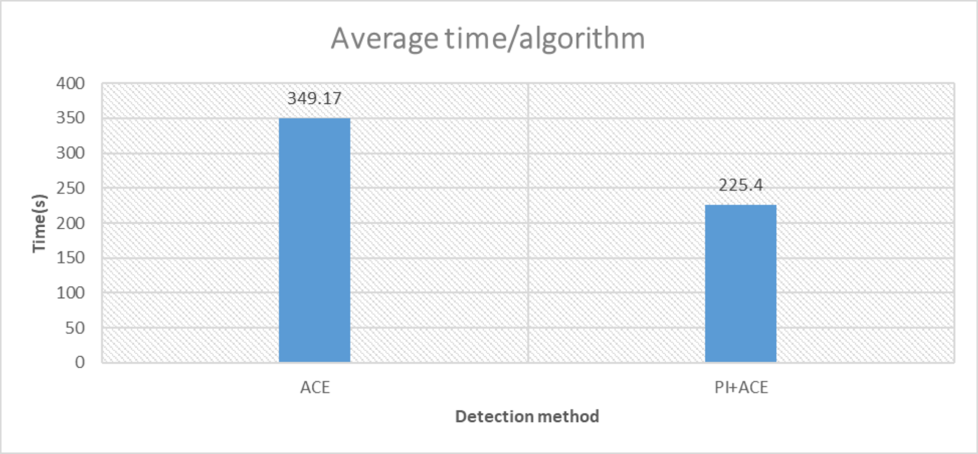


Figure 34 Average computational time /algorithm

In Figure 34, the average time needed by each algorithm to detect all the targets in the 12 images is shown. As we see in the Figure 34, to detect both types of targets using ACE we needed about 349.17 seconds while the computation time of PI+ACE is 225.4 seconds.

Table8 Average DR for each mine with 0.9-0.1 abundance Factor

Algorithm/target	**PMN**	**VS-2.2**	**VS-50**	**N°4**	**M411**	**TM-46**
ACE	1	0.9975	1	1	1	**1**
PI+ACE	1	1	1	1	1	**1**

Table 9 Average FPR/m2 and RFPR for each mine with 0.9-0.1 abundance Factor

Target	**PMN**	**VS-2.2**	**VS-50**	**N°4**	**M411**	**TM-46**
Algorithm	**ACE**					

FPR	0.00001	0.58790	0	0	0.4300	0
RFPR	0.00122%	72.25%	0	0	58.2%	0
Algorithm	**PI+ACE**					
FPR	0	0.00032	0	0	0.00058	0
RFPR	0	0.04%	0	0	0.0077%	0

In Tables 8 and 9 we show the detection performance of two tested algorithms, when the landmine abundance factor is between 0.9 and 0.1. As we see in Table 2, using the similarity measures before ACE, we can decrease the FPR from 0.58 and 0.43 to 0.00032 and 0.000058 in case of VS-2.2 and M411 mines, respectively. Similar results can be observed for RFPR. In the other types of mines, we measure 0 FPR when applying similarity measures before ACE, indicating that in these cases we expect the FPR to be less than 1/number of tested pixels. Overall, PI+ACE method has the lowest FPR with the lowest CT compared to ACE with DR=1. It is important to note that some targets have 0 FPR while other targets have high FPR. This due to the material of each target and their spectral similarity with the background pixels.

Concerning the detection rate, ACE has DR less than one concerning the VS-2.2 landmines where it can detect all the other types of landmines with DR=1. When we apply the similarity methods before ACE, we get DR=1 in all the types of landmines.

Real landmine experiment

Figure 35 shows the RGB representation of the scene acquired in the lab. The spatial resolution is 0.16m^2. The size of obtained image is 25*40 pixels.

Figure 35 RGB representation

Table 3 show the results of FPR and RFPR in the two cases when ACE is used and when PI+ACE is used. All the results in this table is with DR=1.

Table 10 Average FPR/m2 and RFPR for each mine

Target	**Metal Landmine**	**Plastic Landmine**	**Powder**
Algorithm	**ACE**		
FPR	0.009	0.013	0
RFPR	12.33%	18.1%	0
Algorithm	**PI+ACE**		
FPR	0	0	0
RFPR	0%	0%	0%

From table 10, we can observe that when we use PI+ACE both FPR and RFPR decrease for the two types of landmines and we are not able to measure any false positive. It is important to clarify that in this experiment we cannot measure any significant FPR less than 1/1000, since the tested image has a size of 1000 pixels.

d. Bands Selection

After verifying that PI+ACE method has the best performance concerning FPR, CT and DR, here we study the effect of net analyte signal (NAS) and multicriteria band selection methods on the performance of PI+ACE method.

NAS is defined as the part of an analyte spectrum that is singular to that analyte[79]. This method is supervised, takes the spectrum of the target and the image as input and returns the bands that represent the selected target.

Multicriteria method is an unsupervised method described in Chapter 3. It works on classifying the set of neighboring bands based on three criteria: Mutual Information (MI), fidelity, and the cross-correlation coefficient. If these three criteria between two successive bands are above a certain threshold at the same time, these two bands are classified in the same class: this procedure is repeated for all the bands. After classifying bands; the more informative (highest Entropy) band is selected from its class[38, 80].

We apply these two methods before applying PI+ACE. First, we apply the NAS method to visualize the representative bands for each landmine and determine the optimal range for the detection of all landmine types. Second, the multicriteria method is applied in order to obtain the most representative bands in the images. Therefore, we visualize how much we can reduce the image size without affecting the performance of PI+ACE: Figure 36 shows the FPR increase as a function of the reduction rate, considering the two reduction methods, on the simulated landmine dataset.

Table 11 shows the optimal bands for each types of landmines of the simulated landmine dataset. To cover all the types of landmines we need a spectral range between 900nm and 2500nm.

When multicriteria is used we obtain that the optimal range to detect all the landmines types is between 1100nm and 2250nm. The advantages of multicriteria over the NAS method is that we can reduce the image size whatever the types of targets present in the image are, achieving higher reduction rates.

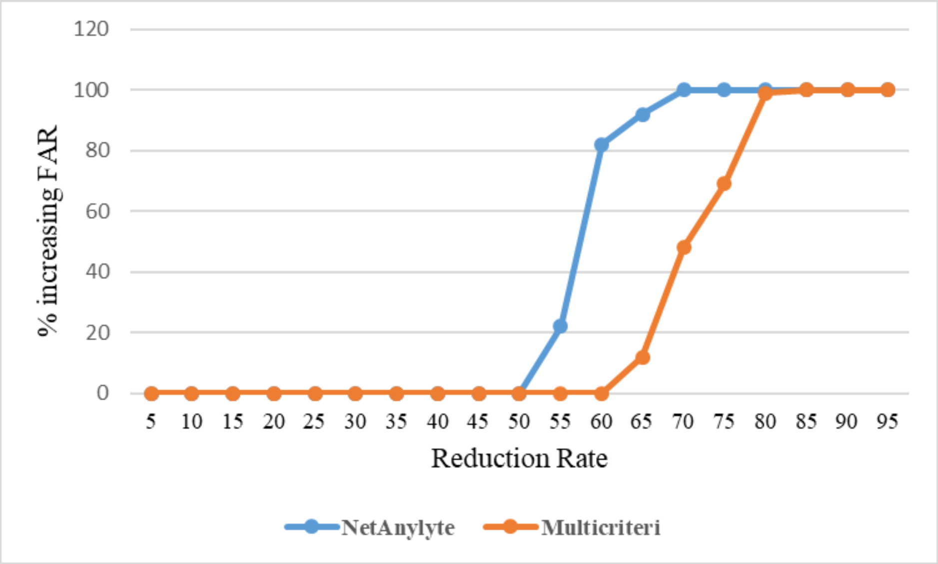


Figure 36 FPR increase (%) of PI+ACE as a function of reduction rate.

Table 11 optimal range for each landmine for the NAS signal method

landmine	**Optimal Range**
Vs22:	**909 to 1432 nm**
PMN	**685 to 1005nm**
N4	**909nm to 1177nm**
TM46	**1629nm to 2097nm**
M411	**928nm to 1502nm**
Vs50	**909nm to 1337nm**
All types	**900nm 2500nm**

Without affecting the performance of PI+ACE we can reduce the images up to 52% using the NAS method and up to 60% using the multicriteria method.

Using NAS method, Table 12 displays the main range for each landmine type in the real landmine case. As can be seen, the most informative frequency range for all the landmines types is between 380 and 1700nm. When multicriteria is applied the selected bands are between 911 and 1700nm.

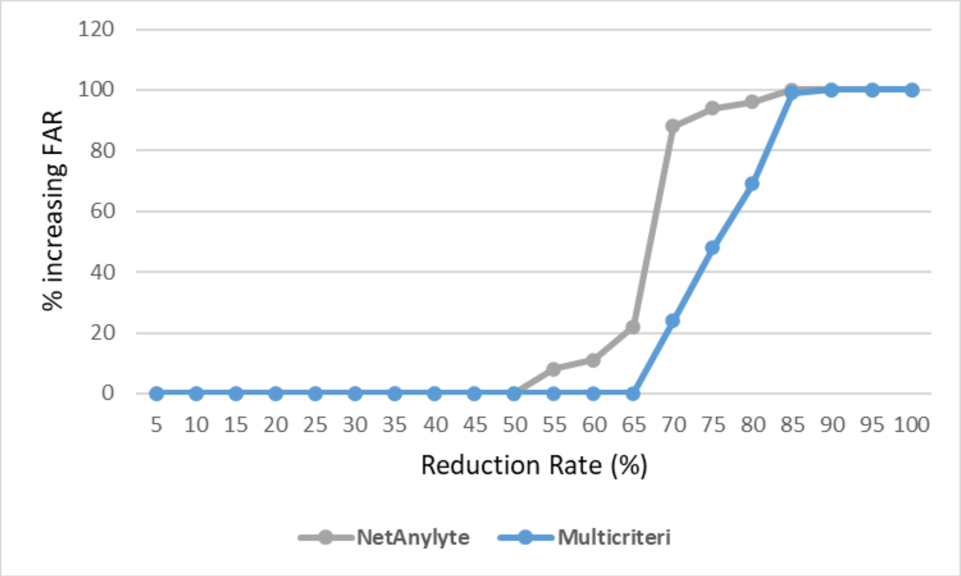


Figure 37 % of increasing of FPR of PI+ACE after Reduction

Figure 37 shows the FPR increase as a function of the reduction rate on the real landmine image. Using NAS method we can reduce the image up to 52% whereas using the multicriteria method we can reduce the image up to 66%.

Table 12 optimal range for each landmine for the NAS signal method

landmine	**Optimal Range**
Metal	**600 to 1700 nm**
Plastic	**380 to 1350nm**
Powder	**411 to 1120nm**
All types	**380 to 1700nm**

Concerning the execution time figure 38 shows the execution time of PI+ACE after maintaining the two bands selection methods.

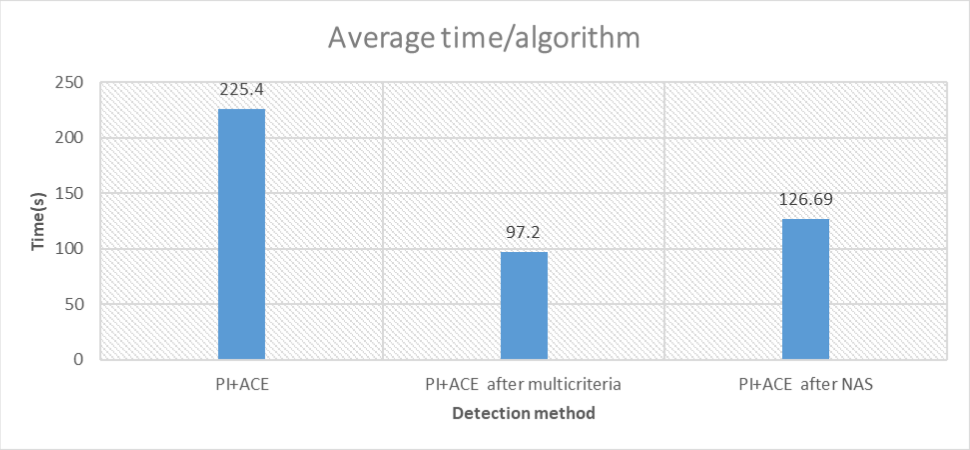


Figure 38 execution time of PI+ACE after the two bands selection methods

e. *Detection based on RGB bands*

Here we investigate whether an RGB image is sufficient to detect all the landmines or we need additional frequencies. Three RGB more informative bands are selected from the image based on the multicriteria method. Table 13 shows the detection results when we use only three RGB bands, for DR = 1.

Table 13 Average FPR/m2 and RFPR for RGB image

Target	**Metal Landmine**	**Plastic Landmine**	**Powder**
Algorithm	**ACE**		
FPR	0.885	0.915	0.867
RFPR	98.66%	98.84%	92.33%
Algorithm	**PI+ACE**		
FPR	0.866	0.903	0.867
RFPR	98.71%	98.68%	92.33%

As we see, using only RGB bands is not enough to detect all the landmines with low FPR and RFPR. In case of metal landmines, FPR is 0.885 and RFPR is 98.66%, this means that 98% of the image pixels are detected to have DR=1. Similar results can be observed with the plastic landmine and powder.

4.7 Conclusion

In this chapter we study the importance of hyperspectral imaging for landmine detection. First by explaining different application used hyperspectral image. In addition, we elaborate many projects work in the field of landmine detection in hyperspectral imaging. Moreover, different detection techniques are implemented in this chapter.

We proposed a pixel similarity method to improve landmine detection in hyperspectral images. The rationale is to select only high dissimilar pixels as potential targets, before applying a state-of-the-art target detection algorithm.

Results obtained on simulated images and realistic images demonstrate that by using the pixel similarity methods with ACE, we can reduce both the execution time of ACE and reduce false alarm at 100% detection rate.

In addition, two dimensionality reduction methods based on band selection are used before applying PI+ACE method: NAS and multicriteria. We can have up to a 35% reduction of the image without losing the performance of detection method (PI+ACE). The execution time of ACE also decreases by about 60%. Moreover, using pixel similarity after the band selection method, we can simultaneously decrease the image at the band and pixel levels without affecting the performance. Lastly, we show that using only RGB bands is not sufficient to achieve good target detection in the landmine scenario, confirming that the use of hyperspectral data is fundamental in this application.

Chapter 5

introduction to hyperspectral thermal infrared imaging in landmine detection

This chapter presents a brief introduction about thermal infrared hyperspectral imaging. We discuss the importance of this technique to detect buried objects, especially landmines.

5.1 Introduction to thermal infrared imaging

Emerging Defense and Security technologies include sensors with cutting-edge capabilities in terms of sensitivity. Among these sensors, the imaging spectrometer is an instrument that provides a large amount of rich information about the measured scene. Detection, identification, and quantification of chemical substances in the gaseous state are fundamental needs in several fields of application. Imaging spectrometers have unparalleled capacity to satisfy these applications' requirements. Instruments for long wave infrared (LWIR), short-wave infrared (SWIR) and thermal infrared (TIR) hyperspectral imaging have been used by Telops and DRDC [81]. These technologies provide high spectral resolution and high precision radiometric calibration.

Thermal IR imaging has become an advanced technology in many fields of application over the past decades, from electrical and mechanical inspections to building diagnostics, from the detection of optical gases to automation and safety. Up to date, cooled and uncooled IR detectors provide a high degree of thermal resolution and accuracy. Table 14 shows the range of each types in hyperspectral imaging [81].

Table 14 spectral range of each spectroscopy type

Types	range
VNIR, shortwave Infrared	1000 to 2500 nm
Midwave infrared (MWIR)	3500 to 5000nm
Thermal infrared (TIR or LWIR)	8000 to 12000 nm

In general, thermal imaging camera has two types, cooled thermal imager and uncooled thermal imaging. Thermal imaging cameras with a cooled detector deliver a range of advantages over thermal imaging cameras with an uncooled detector [82]. They are, however, more costly. A traditional cooled thermal imaging camera has an image sensor integrated with a cryocooler that lowers the temperature of the sensor to cryogenic temperatures. This decrease in the temperature of the sensor is required in order to minimize thermally induced noise to a degree below that of the imaged signal.

Over the last few decades, many hyperspectral thermal infrared imagers have been developed in a number of unrelated fields, such as chemistry, mapping, military, food quality control, agriculture and so on. In this segment, we are looking at the most used ones. The list is not exhaustive, as the number of manufacturers continues to grow as the application of this technology extends to new areas.

5.1.1 MIVIS

MIVIS (Multispectral Infrared and Visible Imaging Spectrometer) is a very promising whisk-broom system. MIVIS has 4 hyperspectral sensors , one of these sensors is the Thermal infrared range (10 bands from 8340 nm to 12420 nm).

The main instrument of the Sensytech company, MIVIS has only 20 tapes in the field of the visible. Nevertheless, its four spectrometers, covering the spectral range of the visible to the thermal infrared, can provide useful data for many applications. However, the noise is very irregular in different bands [15].

5.1.2 EPS-H

Environmental Protection System (EPS) is a sensor developed by GER Corporation. It is composed of multiple imagers covering the spectrum from visible 430 nm to thermal 12500 nm infrared [20]. Within each range the spectral resolution varies.

5.1.3 SPECIM

SPECIM (spectral imaging) presents its thermal hyperspectral cameras in the LWIR region 8 to 12 µm. Two camera models have been specially designed to meet diverse requirements in industrial, research and security applications. Spectral Cameras are pushbroom type line scan cameras that provide full, contiguous hyperspectral data for each pixel along the imaged line. To respond to a wide range of applications and requirements, SPECIM has developed 2 models of LWIR Spectral Cameras: HS (with uncooled detectors), and C (with cooled detector) [83]. It has 30 spectral bands and spectral sampling of 200 nm. With a good sensitivity and moderate spectral resolution, HS is suitable for many industrial, chemical, and military Imaging applications.

5.1.4 **MAG62**

The new MAG62 high resolution product from Magnity Shanghai Juge electronics is a high-quality thermal imager for all applications where superior image quality and measurement precision are both crucial. With advanced high resolution FPA detector, powerful low noise and high-speed circuits, as well as advanced image processing algorithms, MAG62 provides high quality images and accurate temperature readings. It is the perfect on-line imager for the most challenging vision and thermography applications. The spectral thermal range is from 7500 nm to 14000 nm with a spatial resolution of 0.3 m [84].

5.1.5 **HYPER-CAM LW (TELOPS)**

Hyper-Cam-LW (long wave) built by Telops, Quebec, Canada [85] . This remote sensing instrument integrates high spatial, spectral and temporal resolution to provide unprecedented efficiency. This is a flexible method for remote detection, recognition and quantification.

This instrument gives the full spectrum of each pixel in the image, each pixel having an instant field of view of 0.35 mrad. The instrument features two black body internal calibration mounted in front of the Hyper-Cam. Used to conduct a complete radiometric end-to - end calibration of infrared measurements. In it, Long wave IR version of the instrument has a high sensitivity over the 8–12 μm area [86].

5.1.6 **Itres TASI-600**

TASI-600 built by Itres Stuttgart, Germany. This Airborne Hyperspectral Thermal Imager has spectral range between 8-11.5 microns with 32 Spectral Channels and 40° FOV with 600 Spatial Imaging Pixels. This instrument Measures Apparent Emissivity with Custom diffraction-limited, high-performance optics. The main applications of this instrument represented by Mineral Composition / Stratigraphy & Structural Geology / Rock Types and buried landmine detection (www.Itres.com).

5.2 Thermal infrared spectrum basics

Figure 39 shows the basics of thermal infrared imaging. The region of 3 to 12 mm is generally used in terrestrial remote sensing for thermal infrared spectroscopy. The TIR is however described by different authors using slightly different ranges [87]. In all related remote sensing missions, data acquisitions are only made in the regions with the least spectral absorption defined as the atmospheric window.

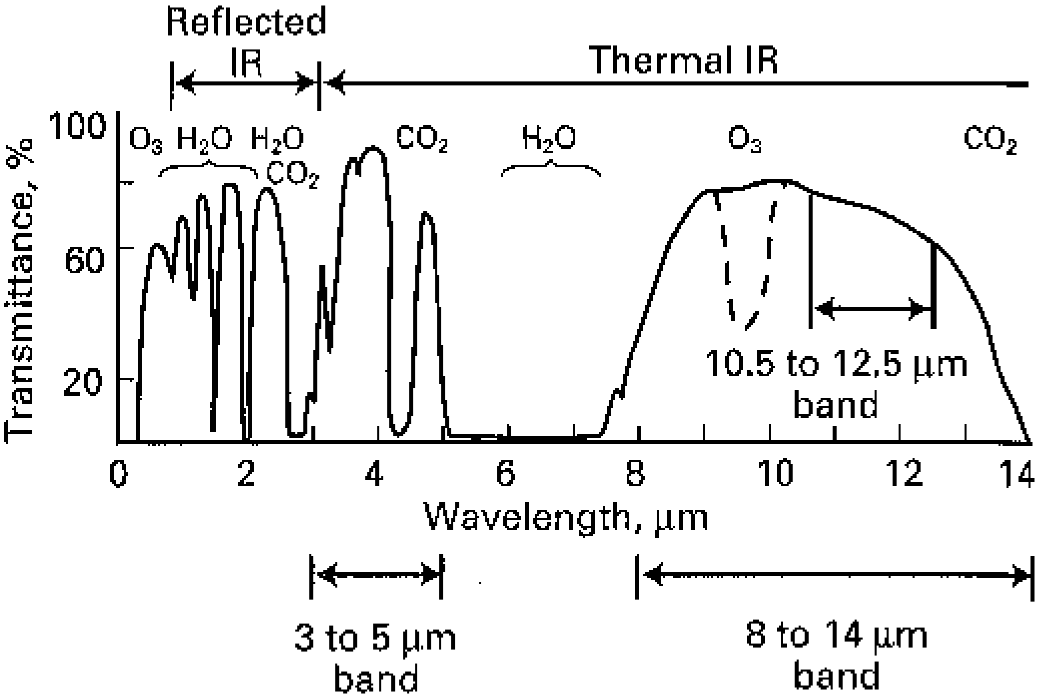


Figure 39 Basic understanding of thermal infrared spectrum or tropical atmospheric condition

However, atmospheric absorption and emission in these atmospheric windows should be taking into consideration. The narrow absorption band from 9 to 10 µm is caused by ozone and is omitted by most thermal IR satellite sensors. Moreover, it is worth noting that reflected sunlight can contaminate thermal IR signals recorded in the 3 - 5 µm windows during daytime.

5.3 Detection of buried landmine

In chapter 4, we study surface landmine detection in the VNIR range. However, often buried mines cannot be directly seen in the VNIR band due to the heavy soil absorption and vegetation cover [81]. Although the variations in the reflective spectrum between compact regions above mines and the background material regions can be indirectly detected because of the:

- Disrupted natural vegetation and surface soil
- Vegetative tension due to mine implantation while the soil appears to be undisturbed
- Vegetative tension from toxic vapors or particulates from mine spills or residues that contaminate the case

There are several projects and research work on the detection of buried landmines using thermal infrared hyperspectral imaging. A summary of these projects and of the results obtained is given in Table 15.

Table 15 summary of projects working on buried landmine

Project	Range types	Detection techniques	Results
FOI (MOMS) [88]	VNIR, SWIR, MWIR, LWIR, 3D LADAR.	Anomaly detection	For the automatic detection of open and semi-hidden mines, hyperspectral imaging is helpful. The sensor suite and algorithms are selected according to the environment and operating conditions.
NATO project [59]	VNIR, SWIR, MWIR, LWIR	image processing	Airborne mine detection radar is less appropriate. To overcome

			meteorological effects, band combinations are required. Algorithms and methods will be improved in parallel.
airborne sensors tests (NVESD) [89]	VNIR, SWIR, MWIR, LWIR	NFINDR and Rx with STD anomaly detection.	The use of proper algorithms gives strong detection in LWIR. The inclusion in anomaly detection of spatial pattern data enhances detection efficiency.
Joint Multispectral Sensor Program [58]	VNIR, SWIR, MWIR, LWIR	Fourier Transform	For nighttime target detection, thermal sensors are helpful. LWIR bands are better than MWIR bands
DARPA project to detect buried landmines [90]	MWIR and LWIR	spectral comparison	LWIR and MWIR are ideally suited to detect buried landmines
Landmines detection using SWIR bands (DRDC) [91]	SWIR	Low Cost Chromatography detector	Similarly, to VNIR bands, the use of SWIR is beneficial to detect surface-laid mines and recently buried landmines.

5.4 EXPERIMENTS

In this section, the results of two experiments are presented to show the importance of thermal infrared imaging in the detection of buried landmines. The main advantage of our study is that the previous projects work on some bands in thermal imaging, where we are working on the full range thermal imaging.

We use the USGS digital spectral library [92], to construct a thermal infrared image with a range between 3500 and 11000 nm using the Nicolet Fourier transform infrared spectrometer with a spectral resolution of 1nm . the construction was made pixel by pixels. In each pixel we put the signature of desired material in order to complete the scene. The first image represents the realistic scene acquired in the lab and described in Chapter 4. The second one represents an outdoor scene containing different materials.

Figure 40 indoor scene1

Figure 41 outdoor scene2

Figures 41 shows the indoor scene presented in Chapter four and figure 42 represents the outdoor scene. In the USGS library there are a lot of spectral signatures from visible to thermal infrared of many materials to construct the two images presented.

Indoor scene size: 25*40*1000 with pixels size =0.16m^2

Outdoor scene size: 88*130*1009 with pixel size = 1m^2

Figure 42 and 43 show the hyperspectral representations of the two images above. Each material has a spectral signature in the visible and in the thermal range. Therefore, we take the RGB bands to preview these two images.

Figure 42 spectral representation of scene 1

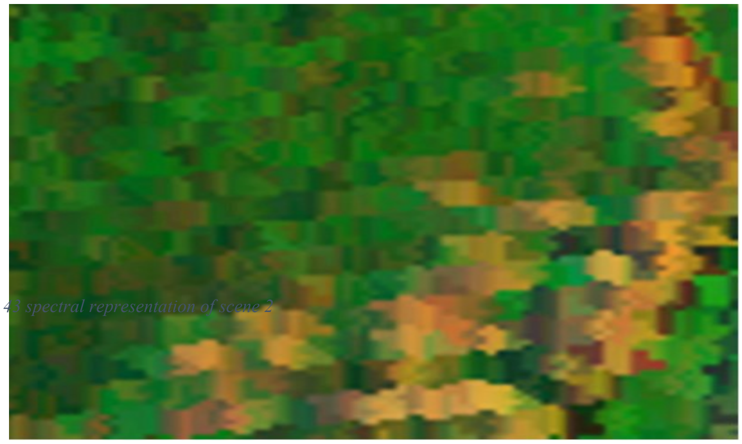


Figure 43 spectral representation of scene 2

In the two represented scenes, we simulate two targets: the first one represents a metal material and the other one represents a plastic material. Figure 44 represents the signature of each material.

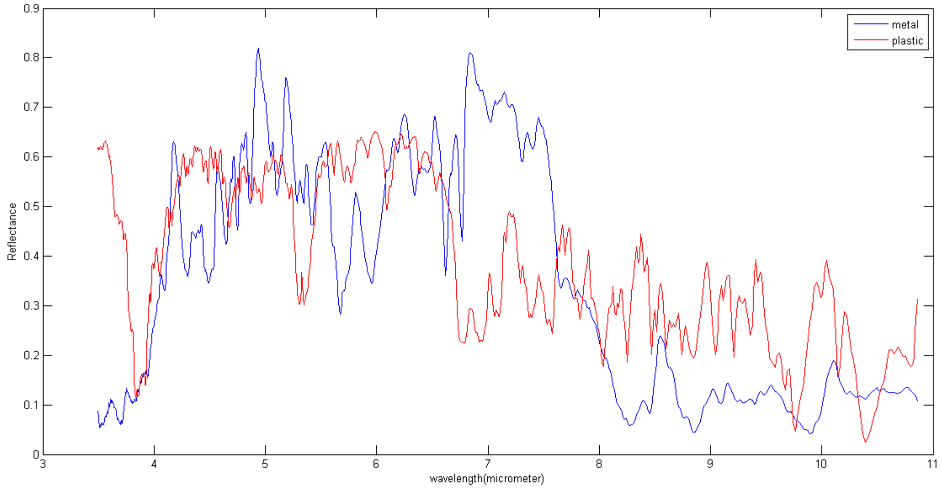


Figure 44 plastic (red) and metal (blue) signatures

The simulation of targets is done using equation (48).

Equation 48

$$\boldsymbol{x = \alpha b + (1 - \alpha) * \beta t}$$

Where α represents the mixing factor at the surface level and β is a factor taking into account the depth of the buried object . b represents the background signature and t represents the target signature.

Concerning β, this value is between 0 and 1 and represents the depth factor. In [93], the authors verify that the signature of a buried target is affected by depth. The shape of the signature remains the same but the thermal energy decrease. From the results in [93, 94], the overall positive value was changed from 46.50 to 38.00 by a change in the depth of the buried object from 1 to 10 cm. The overall positive value shifted from 30,00 to 29.00 by changing the depth from 40 cm to 50 cm. So, from the above study we extrapolate the following ranges: β=0.82 represents a depth value between 1 and 10cm and β=0.69 represents a depth value between 12 and 20cm, β=0.65 represents a depth value between 22 and 30cm, β=0.6 represents a depth value between 32 and 40cm.

5.5 Result and discussion

Here, we study different values of α and β to cover all the expected scenarios. PI+ACE method is applied to detect the two targets after band selection. Then, FPR and RFPR are calculated in each step. Two band selection methods are used in the experiments: multicriteria and NAS. 13 metal and plastic targets are simulated in scene1 using equation 48. We show only results for PI+ACE since this was the best performing method in the NVIR scenario.

5.5.1 Indoor scene Target detection using PI+ACE

Pixel reduction rate =47.23%.

β between 1 and 0.82	**Plastic**		**Metal**	
	FPR	RFPR	FPR	RFPR
(1-α)= 0.9	0	0	0	0
(1-α)= 0.7	0	0	0	0
(1-α)= 0.5	0	0	0	0
(1-α)= 0.4	0	0	0	0
(1-α)= 0.3	0	0	0	0
(1-α)= 0.2	0	0	0	0
(1-α)= 0.1	0	0	0	0

β between 0.8 and 0.69	**Plastic**		**Metal**	
	FPR	RFPR	FPR	RFPR
(1-α)= 0.9	0	0	0	0
(1-α)= 0.7	0	0	0	0

(1-α)= 0.5	0	0	0	0
(1-α)= 0.4	0	0	0	0
(1-α)= 0.3	0	0	0	0
(1-α)= 0.2	0	0	0	0
(1-α)= 0.1	0	0	0	0

β =0.65	**Plastic**		**Metal**	
	FPR	RFPR	FPR	RFPR
(1-α)= 0.9	0	0	0	0
(1-α)= 0.7	0	0	0	0
(1-α)= 0.5	0	0	0	0
(1-α)= 0.4	0	0	0	0
(1-α)= 0.3	0	0	0	0
(1-α)= 0.2	0	0	0	0
(1-α)= 0.1	0	0	0.03	31.5%

β =0.6	**Plastic**		**Metal**	
	FPR	RFPR	FPR	RFPR
(1-α)= 0.9	0	0	0	0
(1-α)= 0.7	0	0	0	0
(1-α)= 0.5	0	0	0	0

(1-α)= 0.4	0	0	0	0
(1-α)= 0.3	0	0	0	0
(1-α)= 0.2	0	0	0	0
(1-α)= 0.1	0.06	43.4%	0.075	48%

5.5.2 Outdoor scene Target detection using PI+ACE

13 metal and plastic targets are simulated in scene2 using equation 44. Pixel reduction rate =57.57%.

β between 1 and 0.82	**Plastic**		**Metal**	
	FPR	RFPR	FPR	RFPR
(1-α)= 0.9	0	0	0	0
(1-α)= 0.7	0	0	0	0
(1-α)= 0.5	0	0	0	0
(1-α)= 0.4	0	0	0	0
(1-α)= 0.3	0	0	0	0
(1-α)= 0.2	0	0	0	0
(1-α)= 0.1	0	0	0	0

β between 0.8 and 0.69	**Plastic**		**Metal**	
	FPR	RFPR	FPR	RFPR

(1-α)= 0.9	0	0	0	0
(1-α)= 0.7	0	0	0	0
(1-α)= 0.5	0	0	0	0
(1-α)= 0.4	0	0	0	0
(1-α)= 0.3	0	0	0	0
(1-α)= 0.2	0	0	0	0
(1-α)= 0.1	0	0	0	0

β =0.65	**Plastic**		**Metal**	
	FPR	RFPR	FPR	RFPR
(1-α)= 0.9	0	0	0	0
(1-α)= 0.7	0	0	0	0
(1-α)= 0.5	0	0	0	0
(1-α)= 0.4	0	0	0	0
(1-α)= 0.3	0	0	0	0
(1-α)= 0.2	0	0	0	0
(1-α)= 0.1	0	0	0.0006	38%

β =0.6	**Plastic**		**Metal**	
	FPR	RFPR	FPR	RFPR
(1-α)= 0.9	0	0	0	0
(1-α)= 0.7	0	0	0	0

(1-α)= 0.5	0	0	0	0
(1-α)= 0.4	0	0	0	0
(1-α)= 0.3	0	0	0	0
(1-α)= 0.2	0	0	0	0
(1-α)= 0.1	0.0011	50%	0.0012	51.8%

From the above results on the two tested scenes we see that a metal landmine can be detected without any FPR where β=0.69 (depth=20cm) and a plastic landmine can be detected without any FPR where β=0.65(depth=30cm). The FPR is still quite low for other values of β. Using PI, we can reduce the indoor scene about 47.23% and the outdoor scene 57.57%. It is important to note that the FPR in this experiment should be lower than the minimum value measurable by the experiments (1/number of tested pixels).

5.5.3 Bands selection for indoor scene

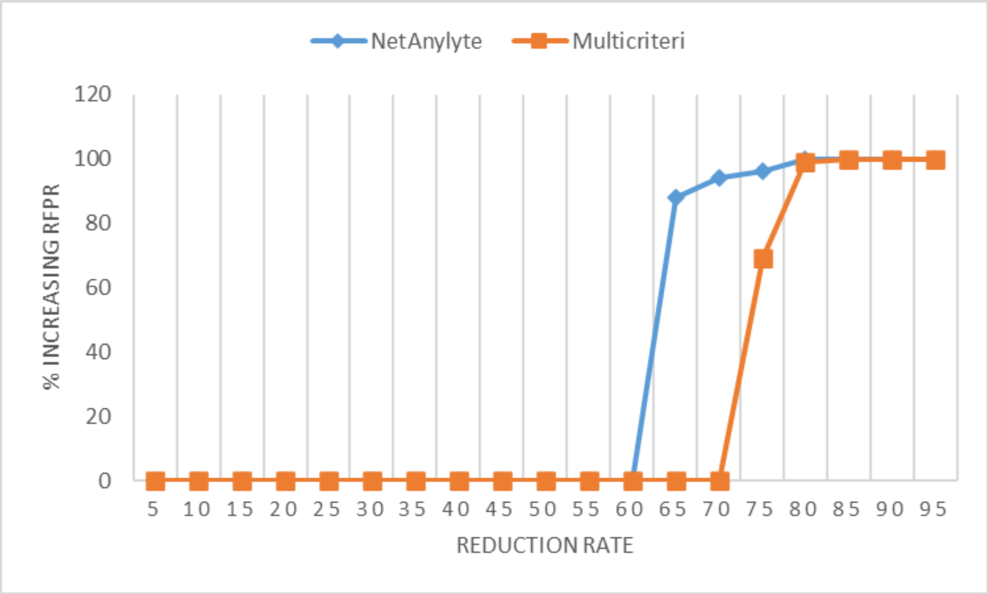


Figure 45 % increasing of FPR of PI+ACE in the two bands selection methods for scene 1

5.5.4 Bands selection for outdoor scene

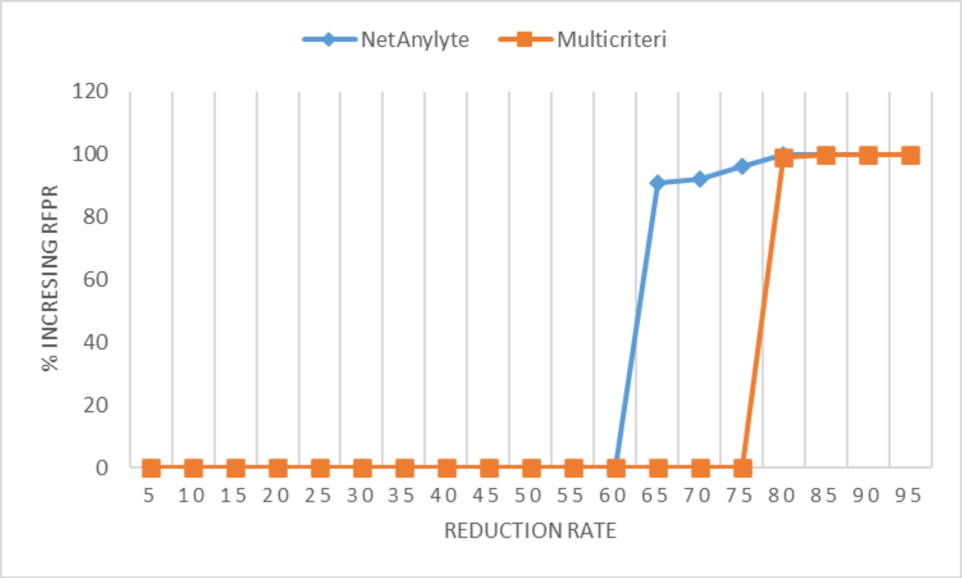


Figure 46 % increasing of FPR of PI+ACE in the two bands selection methods for scene 2

As we see in Figure 45 and 46, for the indoor scene we can reduce the image about 70% using the multicriteria method and 60% using the NAS method without affecting the performance of PI+ACE method. In addition, for the outdoor scene we can reduce the image up to 75% using the multicriteria method and up to 60% using the NAS method.

Table 16 shows the optimal bands for each target, based on the results presented in table 13.

Table 16 optimal bands for each target

target	**bands**
metal	6000--11000
plastic	5000--9000

It is important to note that for plastic landmines we do not need the whole used spectral range. The same is true also for metal landmines.

5.6 Conclusion

For buried explosive detection, thermal infrared imaging may have an important role. In this research, the impact of the depth of the buried item on the thermal signatures of both metal objects and plastic objects has been examined. The depth of the buried objects impacts the temperature amplitude and hence its thermal signature on the surface. PI+ACE method shows a good detection performance also in this scenario. In addition, two bands selection methods used to reduce the image at the band level have been investigated, where the multicriteria method shows a better performance than the NAS method in terms of reduction rate.

Chapter 6: Conclusion and perspectives

The main contribution of this thesis is that we have proved the applicability of hyperspectral imagery to detect mines laid on the surface using the VNIR hyperspectral imaging. Landmines embedded in the soil or in grass may still be found using the thermal infrared imaging. In addition, our goal was to develop a competitive dimensionality reduction algorithm adapted to hyperspectral images. Moreover, using pixel similarity measure lead to reduce hyperspectral image at the pixel level and then reduce the complexity of existing target detection method.

The landmines problem is expanding all over the world. Although the use of landmines must be prohibited immediately, there is a need to find a new solution capable of detecting landmines of different types and shapes, while being safe, fast and reliable at the same time. A good candidate for this purpose is the hyperspectral imaging technology. Thanks to its low cost, the most used technique until now is the metal detector. Nowadays, though, most landmines are made of plastic, which made their detection more difficult using the metal detector. One of the advantages of hyperspectral imaging is that landmines of any type of case are detected. The hyperspectral images are too costly but when installed on a UAV, they are comparable with other techniques especially when the time of detection is taken into consideration through comparison. The cost of hyperspectral cameras is expected to decline, with more companies manufacturing such devices as well as finding new technologies making the manufacture of optical instruments more affordable.

Previous tests have shown that landmines that have been buried recently can be detected using the VNIR band. The VNIR and SWIR fusion could also yield better results. Landmine burying changes the thermal characteristics of a certain type of soil in the upper level. It also changes the reflectivity of its surface and stresses the vegetation. Therefore, the detection of buried mines is possible by monitoring the difference of reflectivity between distorted soil, background and stressed and unstressed vegetation. Correspondingly the potential of detecting the anti-tank mine is higher than that of anti-personnel mines since the exploration of a wider area of surface (soil or vegetation) and the disruption of a greater volume of land is done. In addition, MWIR and LWIR strips are used to identify buried mines. Even if only SWIR and VNIR detect disturbances to soil from being buried.

The false alarm rate can be reduced using MWIR and LWIR hyperspectral imaging. The use of SWIR strips is however more frequent, since most manufactured illustrations operate on the VNIR and SWIR strips. After checking many hyperspectral imagers of different bands, it was found that with the use of proper algorithms, imagers in LWIR bands have the ability to detect buried landmines. Based on data availability, the algorithms can be monitored or unchecked. Note that the possibility of detecting landmines using other bands can not be eliminated. However, for each case, correct algorithms and thresholds must be used.

We started by presenting the two general approaches (the techniques by linear/non-linear projection and spectral band selection techniques) no unsupervised parametric and unsupervised image dimensionality reduction for large images. dimension. We recalled the principles of these methods. Let us recall that this thesis is part of unsupervised, non-parametric learning by preserving the rare event and that in this context, we are interested in the methods for automatic dimensionality reduction for large images.

The comparative study proved that our reduction algorithm is highly competitive. The method that we developed has several advantages. First, it is simple and not time consuming; this trait is very interesting for space applications, where storage capacity is limited and time is valuable. Second, the method is fully adapted to the nature of the data it reduces. The reduction algorithm treats the hyperspectral image as a cube formed by a certain number of spectrally adjacent images. The approach is image processing oriented, and the reduced data are intuitive images (explainable), which can be observed by the eye without any inverse processing technique. This particular trait is essential in several applications, such as biomedical ones, where the needed information is collected by observing the images. In other projection oriented reduction techniques, such as the PCA, the resulting reduced data is not at all intuitive when observed by the human eye. Such techniques require inverse processing before being ready to be used. Third, our method also treats the concept of rare events, in particular, by local image processing. Preserving rare events, while reducing a hyperspectral image, is very important, especially in detection applications. In such applications, scientists look for “fingerprints” that may be present in certain spectral bands, which allow for identifying and detecting objects. These “fingerprints” or spectral signatures often appear in very small areas of the images (rare events). Thus, preserving these rare events is essential in such applications. Fourth, the algorithm has multicriteria, which means that it evaluates scientific properties from multiple aspects; robustness to illuminations changes during acquisitions is one of its positive consequences. The method that we developed showed great potential; however, there is always room for optimization. Testing the algorithm on many hyperspectral data samples may reveal the need for more optimization tasks, the comparative study can still be broadened to include more performance metrics.

Results show that using pixel intensity before ACE lead to better performance of ACE in terms of elapsed time and FPR. In addition, using the developed band selection method before PI also reduce the complexity of target detection method without affecting the performance.

Results for thermal infrared imaging for indoor and outdoor scene prove that using TIR imaging can detect buried landmine more than the VNIR one. We can detect landmine using TIR imaging at 40cm depth.

To validate the algorithms proposed and developed in real case scenarios, a scan of real outdoor mineral fields using a UAV hyper-spectral imager should be carried out in a future work.

References

1. *Landmine monitor report* , http://www.the-monitor.org/media/3074086/Landmine-Monitor-2019-Report-Final.pdf. 2018.
2. Boutros-Ghali, B., *The land mine crisis: a humanitarian disaster.* Foreign Affairs, 1994: p. 8-13.
3. Zucchetti, M., et al. *Landmines Crisis, legacy, international and local action*. in *2017 First International Conference on Landmine: Detection, Clearance and Legislations (LDCL)*. 2017. IEEE.
4. Keeley, R., *Understanding landmines and mine action.* Mines Action Canada, 2003.
5. Laurie, G., *An Analysis of MRE Provisions in Recent MA-related Conventions.* Journal of Conventional Weapons Destruction. **14**(1): p. 20.
6. MacDonald, J., et al., *Alternatives for landmine detection*. 2003, RAND CORP SANTA MONICA CA.
7. Gooneratne, C., S. Mukhopahyay, and G.S. Gupta. *A review of sensing technologies for landmine detection: Unmanned vehicle based approach*. in *2nd International Conference on Autonomous Robots and Agents*. 2004.
8. Gunatilaka, A.H. and B.A. Baertlein, *Feature-level and decision-level fusion of noncoincidently sampled sensors for land mine detection.* IEEE transactions on pattern analysis and machine intelligence, 2001. **23**(6): p. 577-589.
9. Lucey, P.G., et al. *AHI: an airborne long-wave infrared hyperspectral imager*. in *Airborne Reconnaissance XXII*. 1998. International Society for Optics and Photonics.
10. ALDANA, L.F.R., *A survey of land mine detection technology.* International Joual of Remote Sensing, 2009. **30**(9).
11. Göth, A., I.G. McLean, and J. Trevelyan, *How do dogs detect landmines.* Odor detection: the theory and the practice, Geneva International Centre for Humanitarian Demining, 2003: p. 195-208.
12. Joseph, G., *Fundamentals of remote sensing*. 2005: Universities Press.
13. Porter, W.M. and H.T. Enmark. *A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS)*. in *Imaging Spectroscopy II*. 1987. International Society for Optics and Photonics.
14. Gomez, R.B., *Hyperspectral imaging: a useful technology for transportation analysis.* Optical Engineering, 2002. **41**(9): p. 2137-2143.
15. Cocks, T., et al. *The HyMapTM airborne hyperspectral sensor: the system, calibration and performance*. in *Proceedings of the 1st EARSeL workshop on Imaging Spectroscopy*. 1998.
16. Kruse, F.A., et al. *The AIG/HyVista 1999 USA HyMap group shoot: overview and analysis examples*. in *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120)*. 2000. IEEE.
17. Pearlman, J., et al. *Development and operations of the EO-1 Hyperion imaging spectrometer*. in *Earth Observing Systems V*. 2000. International Society for Optics and Photonics.

18. ESTEC, E.S.C.U., *Exploitation of CHRIS data from the PROBA mission for science and applications.* Experimenters' Handbook, 1999(4).
19. Koponen, S., et al., *Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data.* Remote Sensing of Environment, 2002. **79**(1): p. 51-59.
20. Rickard, L.J., et al. *HYDICE: An airborne system for hyperspectral imaging*. in *Imaging Spectrometry of the Terrestrial Environment*. 1993. International Society for Optics and Photonics.
21. Zhai, Y., et al., *A modified locality-preserving projection approach for hyperspectral image classification.* IEEE Geoscience and Remote Sensing Letters, 2016. **13**(8): p. 1059-1063.
22. der Maaten, V. and E. Postma, *Van den Herik, Dimensionality reduction a comparative review*. 2009, LJP Technical Report TiCC-TR 2009-005.
23. Wang, Z., et al., *Image quality assessment: from error visibility to structural similarity.* IEEE transactions on image processing, 2004. **13**(4): p. 600-612.
24. Chen, S., et al., *2D-LPP: A two-dimensional extension of locality preserving projections.* Neurocomputing, 2007. **70**(4-6): p. 912-921.
25. DeMers, D. and G.W. Cottrell. *Non-linear dimensionality reduction*. in *Advances in neural information processing systems*. 1993.
26. Cox, M.A. and T.F. Cox, *Multidimensional scaling*, in *Handbook of data visualization*. 2008, Springer. p. 315-347.
27. Hofmann, M., *Support vector machines-kernels and the kernel trick.* Notes, 2006. **26**(3).
28. Tenenbaum, J.B., V. De Silva, and J.C. Langford, *A global geometric framework for nonlinear dimensionality reduction.* science, 2000. **290**(5500): p. 2319-2323.
29. Weinberger, K.Q. and L.K. Saul, *Unsupervised learning of image manifolds by semidefinite programming.* International journal of computer vision, 2006. **70**(1): p. 77-90.
30. Chen, H., G. Jiang, and K. Yoshihira. *Robust nonlinear dimensionality reduction for manifold learning*. in *18th International Conference on Pattern Recognition (ICPR'06)*. 2006. IEEE.
31. Saul, L.K. and S.T. Roweis, *Think globally, fit locally: unsupervised learning of low dimensional manifolds.* Journal of machine learning research, 2003. **4**(Jun): p. 119-155.
32. Roweis, S. *Automatic alignment of hidden representations*. in *Sixteenth Annual Conference on Neural Information Processing Systems, Vancouver, Canada*. 2002.
33. Tipping, M.E. and C.M. Bishop, *Mixtures of probabilistic principal component analyzers.* Neural computation, 1999. **11**(2): p. 443-482.
34. Ghahramani, Z. and G.E. Hinton, *The EM algorithm for mixtures of factor analyzers*. 1996, Technical Report CRG-TR-96-1, University of Toronto.
35. Dempster, A.P., N.M. Laird, and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society: Series B (Methodological), 1977. **39**(1): p. 1-22.
36. Nadler, B., et al., *Diffusion maps, spectral clustering and reaction coordinates of dynamical systems.* Applied and Computational Harmonic Analysis, 2006. **21**(1): p. 113-127.
37. Belkin, M. and P. Niyogi. *Laplacian eigenmaps and spectral techniques for embedding and clustering*. in *Advances in neural information processing systems*. 2002.

38. Khoder, M., et al., *Multicriteria classification method for dimensionality reduction adapted to hyperspectral images.* Journal of Applied Remote Sensing, 2017. **11**(2): p. 025001.
39. Kuybeda, O., D. Malah, and M. Barzohar, *Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors.* IEEE Transactions on Signal Processing, 2007. **55**(12): p. 5579-5592.
40. Bioucas-Dias, J.M. and J.M. Nascimento, *Hyperspectral subspace identification.* IEEE Transactions on Geoscience and Remote Sensing, 2008. **46**(8): p. 2435-2445.
41. Chanussot, J., M.M. Crawford, and B.-C. Kuo, *Foreword to the special issue on hyperspectral image and signal processing.* 2010.
42. Acito, N., M. Diani, and G. Corsini, *A new algorithm for robust estimation of the signal subspace in hyperspectral images in the presence of rare signal components.* IEEE Transactions on Geoscience and Remote Sensing, 2009. **47**(11): p. 3844-3856.
43. Khoder, J. and R. Younes. *Proposal for preservation criteria to rare event. Application on multispectral/hyperspectral images*. in *2013 25th International Conference on Microelectronics (ICM)*. 2013. IEEE.
44. Mitchell, H.B., *Image fusion: theories, techniques and applications*. 2010: Springer Science & Business Media.
45. Pluim, J.P., J.A. Maintz, and M.A. Viergever, *Mutual-information-based registration of medical images: a survey.* IEEE transactions on medical imaging, 2003. **22**(8): p. 986-1004.
46. Eskicioglu, A.M. and P.S. Fisher, *Image quality measures and their performance.* IEEE Transactions on communications, 1995. **43**(12): p. 2959-2965.
47. Whitnall, C. and E. Oswald. *A comprehensive evaluation of mutual information analysis using a fair evaluation framework*. in *Annual Cryptology Conference*. 2011. Springer.
48. Christophe, E., D. Léger, and C. Mailhes, *Quality criteria benchmark for hyperspectral imagery.* IEEE Transactions on Geoscience and Remote Sensing, 2005. **43**(9): p. 2103-2114.
49. Fong, M., *Dimension reduction on hyperspectral images.* Univ. California, Los Angeles, CA, 2007.
50. Tsai, F. and W. Philpot, *Derivative Analysis of Hyperspectral Data. Remote Sensing Environment.* 1998.
51. Qian, S., *Dimensionality reduction of multidimensional satellite imagery.* SPIE Newsroom, 2011. **21**.
52. Nahr, S.T., P. Pahlavani, and M. Hasanlou, *Different optimal band selection of hyperspectral images using a continuous genetic algorithm.* The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2014. **40**(2): p. 249.
53. Li, X., et al., *Hyperspectral imaging and their applications in the nondestructive quality assessment of fruits and vegetables*, in *Hyperspectral imaging in agriculture, food and environment*. 2018, IntechOpen Limited London, UK. p. 27-63.
54. Barott, K., et al., *Hyperspectral and physiological analyses of coral-algal interactions.* PloS one, 2009. **4**(11): p. e8043.
55. Sun, D.-W., *Hyperspectral imaging for food quality analysis and control*. 2010: Elsevier.
56. Williams, D.J., et al. *Detection and identification of toxic air pollutants using airborne LWIR hyperspectral imaging*. in *Proc. SPIE*. 2004.

57. Faust*, A., et al., *Canadian teleoperated landmine detection systems. Part II: Antipersonnel landmine detection.* International journal of systems science, 2005. **36**(9): p. 529-543.
58. Eismann, M.T., et al. *Comparison of infrared imaging hyperspectral sensors for military target detection applications*. in *Imaging Spectrometry II*. 1996. International Society for Optics and Photonics.
59. Groot, J. and Y. Janssen, *REMOTE LAND MINE (FIELD) DETECTION. An Overview of Techniques (DETECTIE VAN LANDMIJNEN EN MIJNENVELDEN OP AFSTAND. Een Overzicht van de technieken)*. 1994, FYSISCH EN ELEKTRONISCH LAB TNO THE HAGUE (NETHERLANDS).
60. Gagnon, M.-A., et al. *Airborne thermal infrared hyperspectral imaging of buried objects*. in *Electro-Optical Remote Sensing, Photonic Technologies, and Applications IX*. 2015. International Society for Optics and Photonics.
61. Wolff, L.B., et al. *Image fusion of shortwave infrared (SWIR) and visible for detection of mines, obstacles, and camouflage*. in *Detection and Remediation Technologies for Mines and Minelike Targets VIII*. 2003. International Society for Optics and Photonics.
62. Smith, A.M., et al. *Hyperspectral mine detection phenomenology program*. in *Detection and Remediation Technologies for Mines and Minelike Targets IV*. 1999. International Society for Optics and Photonics.
63. Kenton, A.C., et al. *Detection of land mines with hyperspectral data*. in *Detection and Remediation Technologies for Mines and Minelike Targets IV*. 1999. International Society for Optics and Photonics.
64. Suganthi, G. and R. Korah, *Discrimination of mine-like objects in infrared images using artificial neural network.* Indian J. Appl. Res., 2014. **4**(12): p. 206-208.
65. Manolakis, D. and G. Shaw, *Detection algorithms for hyperspectral imaging applications.* IEEE signal processing magazine, 2002. **19**(1): p. 29-43.
66. Zou, H. and T. Hastie, *Regularization and variable selection via the elastic net.* Journal of the royal statistical society: series B (statistical methodology), 2005. **67**(2): p. 301-320.
67. Marwaha, R., et al., *Target detection algorithm for airborne thermal hyperspectral data.* The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2014. **40**(8): p. 827.
68. Broadwater, J. and R. Chellappa, *Hybrid detectors for subpixel targets.* IEEE transactions on pattern analysis and machine intelligence, 2007. **29**(11): p. 1891-1903.
69. Manolakis, D., et al., *Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms.* IEEE Signal Processing Magazine, 2013. **31**(1): p. 24-33.
70. Ren, H., et al. *Comparison between constrained energy minimization based approaches for hyperspectral imagery*. in *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*. 2003. IEEE.
71. Wang, L. and C. Zhao, *Hyperspectral Image Processing*. 2016: Springer.
72. Harsanyi, J.C. and C.-I. Chang, *Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach.* IEEE Transactions on geoscience and remote sensing, 1994. **32**(4): p. 779-785.
73. Chang, C.-I. *Spectral information divergence for hyperspectral image analysis*. in *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293)*. 1999. IEEE.

74. Yu, C., M. Song, and C.-I. Chang, *Band subset selection for hyperspectral image classification.* Remote Sensing, 2018. **10**(1): p. 113.
75. Guo, Q., R. Pu, and J. Cheng, *Anomaly detection from hyperspectral remote sensing imagery.* Geosciences, 2016. **6**(4): p. 56.
76. Camacho Velasco, A., C.A. Vargas García, and H. Arguello Fuentes, *A comparative study of target detection algorithms in hyperspectral imagery applied to agricultural crops in Colombia.* Tecnura, 2016. **20**(49): p. 86-99.
77. Basener, W.F., E. Nance, and J. Kerekes. *The target implant method for predicting target difficulty and detector performance in hyperspectral imagery*. in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVII*. 2011. International Society for Optics and Photonics.
78. Makki, I., et al. *RBF Neural Network for Landmine Detection in H Yperspectral Imaging*. in *2018 7th European Workshop on Visual Information Processing (EUVIP)*. 2018. IEEE.
79. Lorber, A., *Error propagation and figures of merit for quantification by solving matrix equations.* Analytical Chemistry, 1986. **58**(6): p. 1167-1172.
80. Khoder, J., *Nouvel algorithme pour la réduction de la dimensionnalité en imagerie hyperspectrale*. 2013.
81. McFee, J.E., et al. *Landmine detection using passive hyperspectral imaging*. in *Chemical and Biological Sensing VIII*. 2007. International Society for Optics and Photonics.
82. Wood, R.A. *Uncooled thermal imaging with monolithic silicon focal planes*. in *Infrared Technology XIX*. 1993. International Society for Optics and Photonics.
83. Hyvärinen, T., et al., *New Compact High Performance Cooled and Uncooled LWIR Hyperspectral Sensors.*
84. Chen, X., et al. *Color guided thermal image super resolution*. in *2016 Visual Communications and Image Processing (VCIP)*. 2016. IEEE.
85. Puckrin, E., et al. *Airborne measurements in the infrared using FTIR-based imaging hyperspectral sensors*. in *Electro-Optical Remote Sensing, Photonic Technologies, and Applications III*. 2009. International Society for Optics and Photonics.
86. Schlerf, M., et al., *A hyperspectral thermal infrared imaging instrument for natural resources applications.* Remote Sensing, 2012. **4**(12): p. 3995-4009.
87. Tristancho, J., et al., *Climatic chamber with centrifuge to simulate different weather conditions.* Geotechnical Testing Journal, 2012. **35**(1): p. 159-171.
88. Sjökvist, S., M. Uppsäll, and D. Letalick. *MOMS: a multi-optical approach for land mine detection*. in *Detection and Remediation Technologies for Mines and Minelike Targets XI*. 2006. International Society for Optics and Photonics.
89. Winter, E.M., et al. *Mine detection experiments using hyperspectral sensors*. in *Detection and Remediation Technologies for Mines and Minelike Targets IX*. 2004. International Society for Optics and Photonics.
90. Kenton, A.C., et al. *Joint spectral region buried land mine discrimination performance*. in *Detection and Remediation Technologies for Mines and Minelike Targets V*. 2000. International Society for Optics and Photonics.
91. McFee, J.E., Y. Das, and A.A. Faust. *Canadian landmine detection research program*. in *Detection and Remediation Technologies for Mines and Minelike Targets VIII*. 2003. International Society for Optics and Photonics.
92. Clark, R.N., et al., *USGS digital spectral library splib06a*. 2007, US Geological Survey.

93. Olowofela, J., et al., *Effect of Depth on the Thermal Signature of Buried Metallic Object.* Earth Science India, 2010. **3**.

94. Van Dam, R.L., et al. *Soil effects on thermal signatures of buried nonmetallic landmines*. in *Detection and Remediation Technologies for Mines and Minelike Targets VIII*. 2003. International Society for Optics and Photonics.