WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2006

Lossless compression of hyperspectral images

Sushil Kamalchand Jain
West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation

LOSSLESS COMPRESSION OF HYPERSPECTRAL IMAGES

by

Sushil Kamalchand Jain

**Thesis submitted to the College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Master of Science
in
Electrical Engineering**

**Dr. Donald Adjeroh, PhD., Chair
Dr. Xin Li, PhD.,
Dr. V. Jagannathan, PhD**

Lane Department of Computer Science and Electrical Engineering

**Morgantown, West Virginia
2006**

Keywords: Hyperspectral Images, Lossless Compression, Spatio-Spectral Predictor, Band Ordering, Minimum Spanning Tree.

Abstract

Lossless Compression of Hyperspectral Images.

Sushil K. Jain

Hyperspectral imaging is a useful tool in mineral exploration, environmental monitoring, agriculture, and resource management. The capture of images at small discrete frequencies is possible with better sensor technologies. This has resulted in the acquisition of large amount of data everyday. There is a need to compress this data due to restriction on the available bandwidth and storage space. In this thesis we study two aspects of lossless compression of hyperspectral images. Ordering and the prediction scheme are the two major aspects of hyperspectral imaging which have been studied to improve the performance of the compression system. Each module is studied separately and new techniques are suggested which lead to better performance compared to the state-of-the art technologies.

In the prediction module, we propose spatio-spectral prediction methods. Two non-linear spectral prediction methods have been proposed in this thesis. NPHI (Non-linear Prediction for Hyperspectral Images) is based on a band look-ahead technique wherein a reference band is included in the prediction of pixels in the current band. The prediction technique estimates the variation between the contexts of the two bands to modify the weights computed in the reference band to predict the pixels in the current band. EPHI (Edge-based Prediction for Hyperspectral Images) is the modified NPHI technique wherein an edge-based analysis is used to classify the pixels into edges and non-edges in order to perform the prediction of the pixel in the current band. Since the global structures of the bands remain the same, the edge feature calculated in the reference band should be similar to features present in the current band. The edge based information can be computed easily since at the decoder the reference band is decoded before the current band. Hyperspectral images have high spectral correlation between most of the bands, but some of the bands are corrupted by atmospheric absorption and hence perform poorly with spectral prediction. These bands are predicted using spatial predictor NP (North Pixel). The spatial predictor (NP) performs better than the standard spatial predictors on the bands corrupted by atmospheric absorption. Some of the bands have very low intensity values, these bands are best encoded without prediction since they have low self entropy and poor spectral correlation.

Three ordering methods have been proposed in this thesis. The first ordering method computes the local and global features in each band to group the bands. The bands in each group are ordered by estimating the compression ratios achieved between the entire band in the group and then ordering them using Kruskal's algorithm. The other two methods of ordering compute the compression ratios between b-neighbors to order the bands. The weighted graphs formed are then ordered based on two MST (Minimum Spanning Tree) algorithms. Prim's and Kruskal's MST algorithms have been modified to obtain the ordered bands for lossless compression of hyperspectral images.

Dedication

I dedicate my thesis to my father who I wish had been here to see me complete my masters and my mother without whom I would never had excelled in my studies.

Kamalchand Jain
Shashi Jain.

Acknowledgement

I would like to mention my gratitude and expresses my acknowledgements to all the people who have supported me. I would like to put this in words without missing a chance to mention how fortunate I am to have all these people around. I would like to thank all my professors, friends and family members without whom this difficult task wouldn't have been completed.

It was my honor to have Dr. Donald Adjeroh as my advisor and committee head. His knowledge in the field of data compression has always guided me in the right direction. His timely suggestions and observations have helped me to get in to the roots of the problem and find a systematic solution. His attention to detail, quest for excellence, love for perfection, ingenious ideas and understanding has always motivated me to give my best time and again.

I am very grateful that I had Dr. Xin Li, and Dr. V Jagannathan support, in addition, which was also a very valuable contribution towards my thesis. Their expert guidance and suggestions have given me good ideas and solutions for my thesis.

I would also like to extend my gratitude to all my colleagues who have been with me all the time and made the environment exactly suitable for research endeavors. I would like to specially thank all Lane Department of Computer Science and Electrical Engineering professors for conducting quality research and course work. Contribution by the department secretaries and systems staff has also played a key role in the completion of my thesis.

I feel that this thesis has no meaning to me without the mention of my father who passed away recently. I wish he had been with me to see me complete my master's. I would like to thank my mother and my sister, for their support and encouragement through out my master's thesis program.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Introduction to the Problem

Data compression can be classified into lossless compression and lossy compression. In lossless compression the original data can be retrieved from the compressed data without any distortion or change. Lossless compression yield lower compression ratios. Lossy compression retains the important part of the data while discarding the rest. The measure of the quality of the compressed data is based on the Mean Square Error (MSE) or Peak signal to Noise Ratio of the reconstructed data. Compression gains are dependent on the allowable rate-distortion factor for any given application for which the data is being compressed. Lossless compression is used in applications like medical imaging, hyperspectral imaging, compressing legal documents and for executables. Lossy compression is used in internet applications, cell phones, and entertainment archives.

The focus of this work is on hyperspectral image compression, in particular, on lossless compression of hyperspectral images. With the advent of satellite imaging, huge amounts of data are being captured everyday and transmitted back to earth for further processing. Improved sensor technology has allowed the researchers to capture images at small discrete frequencies resulting in increase in the number of bands captured for one scene. AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensors capture 224 bands in the range of 400nm to 2500nm in discrete steps of 10nm [2, 35]. In one day, 76 GB's of data is captured by the AVIRIS sensors [2, 35]. With constraints on available bandwidth and storage space there is a need to compress this data. The captured data is further processed to study the mineral deposits and atmospheric changes and require that the data should not be corrupted in any form at the time of compression and transmission.

Intense research in the field of lossless compression has been carried out for last many decades. Standard algorithms such as JPEG-LS (Joint Photographic Expert Group) [15] and CALIC (Context Based Adaptive Lossless Image Coder) [54] are widely used in

different applications; however, these algorithms perform poorly in terms of compression ratio on hyperspectral images [50, 20, 55, and 40]. Hyperspectral images have strong spectral correlation than spatial correlation, hence spectral prediction algorithms work better than the spatial prediction algorithms. The standard lossless image compression algorithms have been optimized to compress 8 bit images, while hyperspectral images have varying bit depth across the bands ranging from 6 bits per pixel to 16 bits per pixel. Hyperspectral imaging captures reflected light from different surfaces. This spectral information is then sampled by a 12-bit ADC and then represented with 16-bit precision [2, 39, 35]. Standard lossless image compression algorithms are designed to compress largely natural images having a bit depth of 8 bit and hence do not perform well on 16-bit hyperspectral images. The proposed 3-D CALIC [55, 20] algorithm was not optimized for 16 bit data and hence did not perform well on the hyperspectral data. M-CALIC which improved 3-D CALIC combined the previous bands to form a reference band before applying 3-D CALIC.

Hyperspectral images are affected by atmospheric absorption and sensor noise at the time of capture [2, 22]. Gases, water molecules and the suspended particles in the atmosphere interact with the solar energy and reduce the energy of light reaching the earth. Similarly the reflected light from the earth surface is scattered or absorbed by the atmosphere. This reduces the total incident energy on the sensors resulting in large variation of pixel values across the bands. At certain wavelengths the incident light on the sensors is almost completely absorbed or scattered by the atmosphere. These bands have high dominance of sensor noise over the incident light energy and hence have very little spectral information. Since the sensor noise dominates over the spectral information incident on the sensors, the captured images have very low pixel value range.

Hyperspectral images have same global structure across the bands but have different pixel intensities due to different absorption properties of the atmosphere and the region of interest. Since the global structure remains similar across all the bands, hyperspectral images visually look similar if they are normalized and displayed. Though the global structure remains same across the bands the variation in pixel intensities is different

locally. This is due to the absorption and reflection properties of the material which changes with change in wavelength of the incident light. There is a need to capture this variation at the time of prediction to reduce the prediction error residual.

Classification of pixels into edge and non-edge areas becomes easy in the case of hyperspectral images due to the a priori information obtained from the reference band. This process of classification becomes difficult in the case of spatial prediction. The pixel gradients in the bands is very high and sometimes in thousands. It becomes difficult to classify the pixels into edge and non-edge pixels as almost all the pixels will be classified as an edge region. Though this might not improve the performance of the system, it could become computationally intensive. As discussed earlier, the general structure in the bands remain same hence the edge areas and non-edge areas should be similar across the bands. This information can be utilized for classification of pixels instead of analyzing each pixel spatially. As discussed in [54] higher compression ratios are obtained if the pixels are classified correctly. The pixel correlation is higher along the edge then across the edge, hence the need to classify the pixels.

1.2 Thesis Objective

The basic objective of this thesis is to study the problem of lossless hyperspectral image compression. Lossless compression is a two step process. The first step is **prediction** wherein the pixel under consideration is predicted based on information obtained from its neighboring pixels. The second step is **encoding,** wherein the error residuals from prediction are encoded. The focus of this thesis is on designing a good prediction scheme so as to lower the prediction error residual and obtain a highly skewed error distribution. In particular, we consider the problem of designing the spectral predictor. There is a need to identify bands for different prediction schemes. Some bands perform better with spatial prediction, while others perform better without prediction. There is a need to identify such bands to gain higher compression ratios [39, 26]. Due to atmospheric absorption and scattering, some bands have higher sensor noise and less spectral

information. These bands have poor spectral correlation and need to be compressed using spatial predictors or encoded without prediction [2, 22, 39].

The other aspect of hyperspectral image compression which very few researches have addressed is ordering of bands. One popular algorithm is the method published by Tate [48] for ordering purposes. It has been shown that the performance in terms of compression ratio of a system can be improved if the bands are ordered prior to prediction and encoding [48, 49]. The ordering scheme proposed in [48] has an order of complexity $O(n^2)$ where n is the number of bands. In this scheme the compression ratio is computed between all the bands and then the bands are ordered to optimize the compression ratio. The $O(n^2)$ complexity only accounts for the ordering of the compression ratios computed between the bands. It does not account for the time required to compute the compression ratios between all the bands. A faster algorithm was proposed in [49] where the algorithm was 13 times faster than [48]. In this algorithm compression ratios were computed between 40 neighboring bands. There is a need to explore lower complexity algorithms for band ordering.

1.3 Research Contribution

Performance of a compression scheme depends on how effectively the prediction scheme can model the neighboring pixels in order to predict the current pixel. In spatial prediction the context is formed by the neighboring pixels in 2-D space while in hyperspectral images the context for prediction is formed by neighboring pixels in 3-D space. In this thesis, the context for prediction is formed by pixels in the current band and the pixels in a reference band. A good spectral predictor should be able to capture the change in pixel intensities between the contexts in the reference band and the current band. The variation in the pixel intensities between the contexts of the two bands can be used in predicting the pixel in current band. The method proposed in this thesis computes the change in pixel intensities between the two bands and utilizes the information to modify the weights computed in the reference band to predict the pixel in current band.

Since the pixels in the reference band have already been decoded, the weights for the pixel can be computed easily. These weights are then modified depending on the change in intensities. An error feedback mechanism is employed which computes the error between the calculated weights and the actual weights for the pixels in the context and utilizes this error to modify the weights for the current pixel. The prediction scheme is further modified to incorporate pixel classification into edge and non-edges. It has been proven in [54] that pixel classification improves the performance of the compression system.

The motivation behind the spectral prediction scheme comes from the fact that the general structure of the image captured remains similar in all the bands, though at different intensities. Hence assuming that a reference band has already been encoded we can calculate the coefficients of prediction accurately for the pixels in the reference band. Ideally these weights should apply to the pixels in the next band or the band being predicted without any need for modification. But due to change in intensity values caused by atmospheric absorption, atmospheric scattering, sensor noise, and different absorption properties of different materials, there is a need to modify these coefficients with the variation between the context of the two bands and also the scale difference between the two bands. The spectral prediction scheme proposed in this thesis is a 3-D context based scheme. The 3-D context is formed by 12 neighboring and causal pixels in the current band and the reference band. Hyperspectral images have different band intensities hence simple differences cannot be used in the calculations [50]. The spectral predictor proposed in this thesis computes ratios between pixels for the calculation of the weights and feedback error values. This eliminates the need to normalize the bands before prediction or to compute the scale factor between the bands. The prediction schemes discussed in [54, 33] have an error feedback mechanism which is used to modify the prediction error residual depending on the context classification. In this thesis, we propose an error feedback mechanism wherein we compute the error between the actual weights and the calculated weights for the pixels in the contexts. This error is used to modify the weights computed for the current pixel.

Prediction techniques which can classify the image area into edge and non-edge areas fare better than the techniques which do not perform classification [54, 33, 32, 51]. Pixel correlation is higher along the edge than across the edge. Hence prediction along the edge results in lower prediction error residuals. Edges can be classified using standard edge detection algorithms such as Canny edge detector, or by setting a threshold value wherein pixels are classified as edge pixels if the change in pixel value with the neighboring pixel is greater than the set threshold. In case of hyperspectral images a global threshold cannot be set as the intensity values vary from one band to another. Since the hyperspectral images have similar structure in all the bands the pixel classification in the reference band applies to the band being predicted. Since the reference band is decoded first it becomes easy to classify the pixels in the reference band and utilize them in the current band.

Higher compression ratios can be achieved if the bands are ordered before prediction [48, 49]. But there is a tradeoff between ordering complexity and compression ratios achieved. The algorithm proposed by Tate has a complexity in $O(n^2)$ and it gives the best result achievable with the given prediction method after ordering. The algorithm proposed in [49] is 13 times faster than [48] and the results are very close to the results obtained by [48]. In this thesis we have proposed three different methods for ordering. The first method computes the global and local features of the bands to cluster them into smaller groups. Then compression ratios between bands in each group are computed. The weighted graph formed by the compression ratios between the bands is then solved by Kruskal's algorithm for minimum spanning tree. In the second and third method the compression ratios are computed between the 10 neighboring bands and the weighted graph is solved by modified Prim's and Kruskal's algorithms..

1.4 Thesis Organization

In Chapter 2 gives an overview about the nature and physics of hyperspectral imaging. The details of AVIRIS sensors and AVIRIS data have been mentioned in this chapter. Also the motivation behind the proposed method has been discussed briefly in this

chapter. Different methods of lossless compression technique that have been proposed till date have been discussed. In the end ordering methods proposed for multispectral/hyperspectral images have been discussed.

In Chapter 3 the proposed method of lossless compression of hyperspectral images have been discussed. The two prediction method NPHI and EPHI have been discussed in details. Also we discuss the need for spatial predictors and encoding some bands without prediction in this chapter.

Chapter 4 describes the ordering methods for lossless compression of hyperspectral images. Three methods for ordering have been proposed in this thesis which has been described in this chapter. The first method of ordering is based on global and local features extracted from each band while the other two methods is based on the features estimated from the neighboring band.

Chapter 5 summarizes the results with a comparative study with the state-of-the-art methods. Effects of change in parameters on the proposed techniques have been discussed.

Chapter 6 concludes the thesis with suggestion for future work.

Chapter 2: Background

Hyperspectral imaging has been an area of active research and development for many years. Till recent past hyperspectral data was available only to researches and not for commercial use. Since the advent of commercial satellites the data has been available for commercial use. Hyperspectral imaging is a useful tool in mineral exploration, environmental monitoring, agriculture, and resource management. Capture of images at small discrete frequencies is possible with better sensor technologies. This has resulted in capture of large amount of data everyday. AVIRIS sensors capture 224 images/spectral bands in the range of 400nm to 2500nm with each band having a range of 10 nm. One AVIRIS scene consists of 512 lines and covers an area of 10 Km on ground. One scene having 512 lines yields 140 megabytes of data. AVIRIS sensors capture 76 gigabytes of data everyday. This data needs to be communicated back to the earth from the satellites for further processing. With limitation on available bandwidth and storage space there is a need to compress this data. Hyperspectral imaging applications require that the data be stored without any loss, hence the need for lossless compression of hyperspectral images.

2.1 Physics of Hyperspectral Imaging

Hyperspectral imaging is also termed as imaging spectroscopy. "*Imaging spectroscopy is defined as being the simultaneous acquisition of spatially co-registered images, in many, spectrally contiguous bands, in an internationally recognized system of units from a remotely operated platform (Schaepman, 2005)*"[13, 43].

Hyperspectral images are acquired using image spectrometers. The development of image spectrometers resulted from two fields, spectroscopy and remote sensing of earth. Spectroscopy is study of light emitted or reflected by the minerals and its variations in energy at different wavelengths. Hyperspectral remote sensing exploits the fact that all

materials reflect, absorb, and emit electromagnetic energy, at specific wavelengths, in distinctive patterns related to their molecular composition [30].

The Image spectrometer captures the emitted or reflected light from the earth surface. The optical dispersing element splits the incoming light into different narrow contiguous spectral bands which is then measured by separate detectors. Ideally the visible spectrum, infrared and shortwave infrared spectrum is captured by the spectrometers. The acquired signal by the detectors is then converted into a digital signal by an ADC. This signal is then stored for further processing.
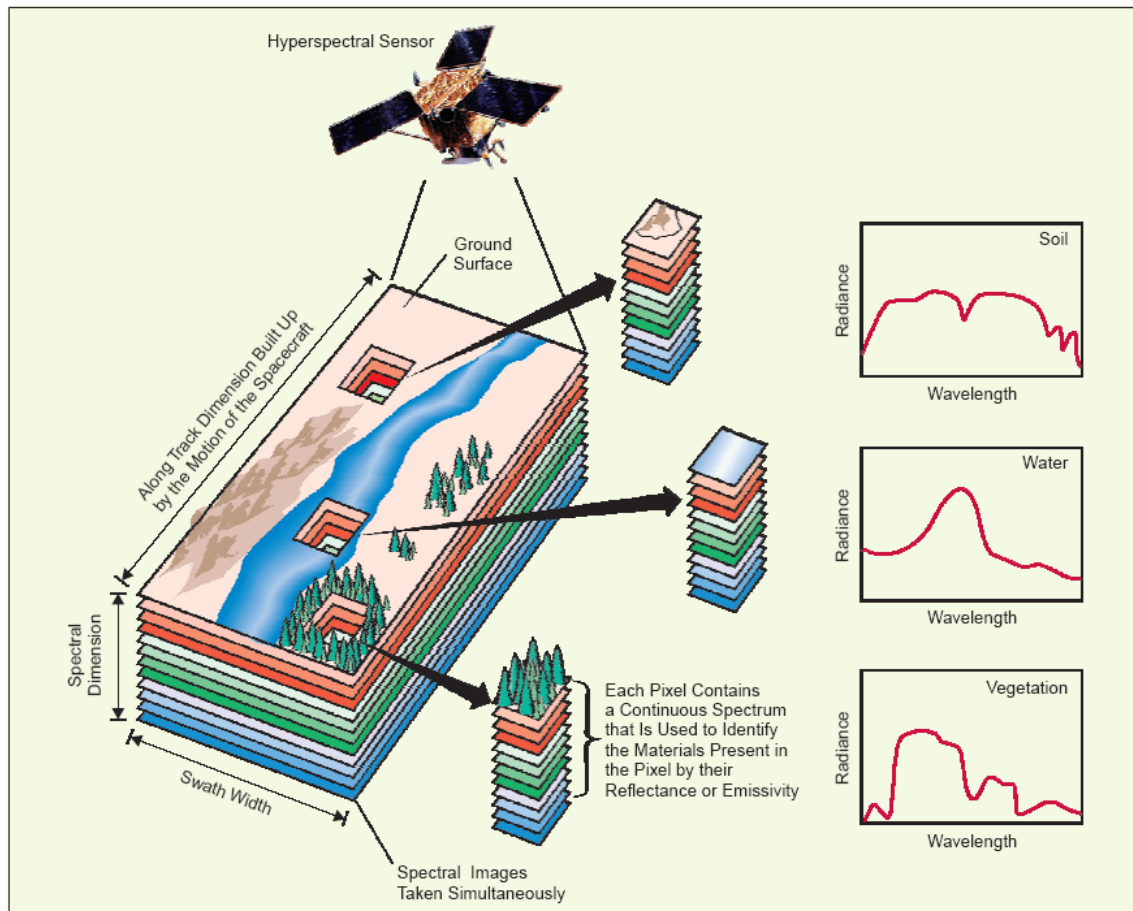
Figure 1: Image Spectroscopy [46]

Figure 1 taken from [46] shows the basic diagram for image spectrometer. The basic source of light energy is the sun hence the spectrum of light captured is dominated by the energy curves of sun. Different materials possess different absorption and reflectance

properties depending on there chemical composition. The materials absorb or reflect light with different intensities at different wavelengths. Every material has its own signature of absorption and reflectance property which can be seen in Figure 2. From Figure 2 it can be observed that vegetation has higher reflectance in infrared region than in visible spectrum. Plants have high content of chlorophyll which absorbs visible light while reflecting the infrared light. As compared to vegetation, soils have higher absorption property in the infrared region.
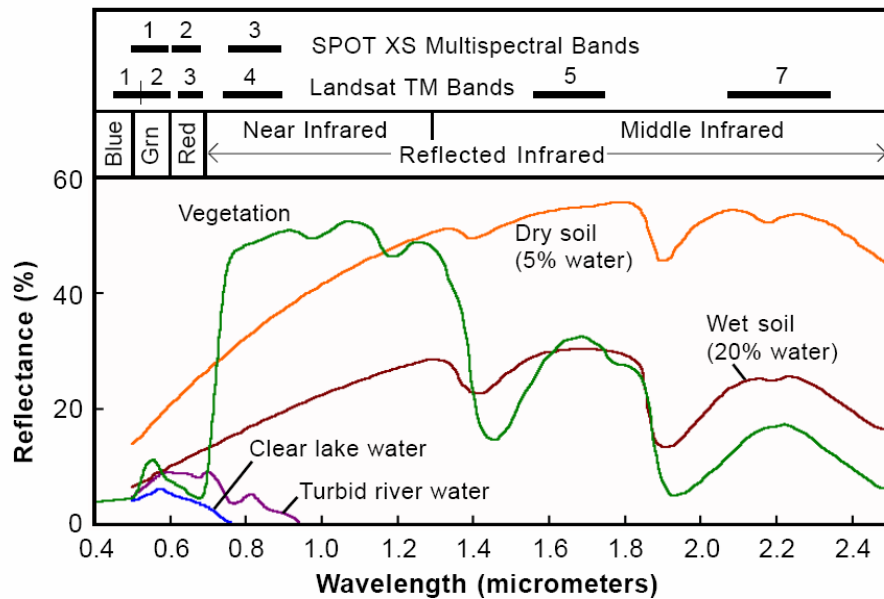
Figure 2: Spectral Absorption/Reflectance for different materials [22]

Figure 3 shows the mineral spectra for different minerals [22]. Different minerals have different absorption/reflectance properties for different spectral light. Hematite's have higher absorption property in the visible light spectrum due to ferric iron while calcite has higher absorption in the spectral range of 1.8 µm- 2.4 µm range due to carbonate ion content. Higher spectral energy absorption around 1.9 µm is seen in montmorillonite (clay) material due to water content.
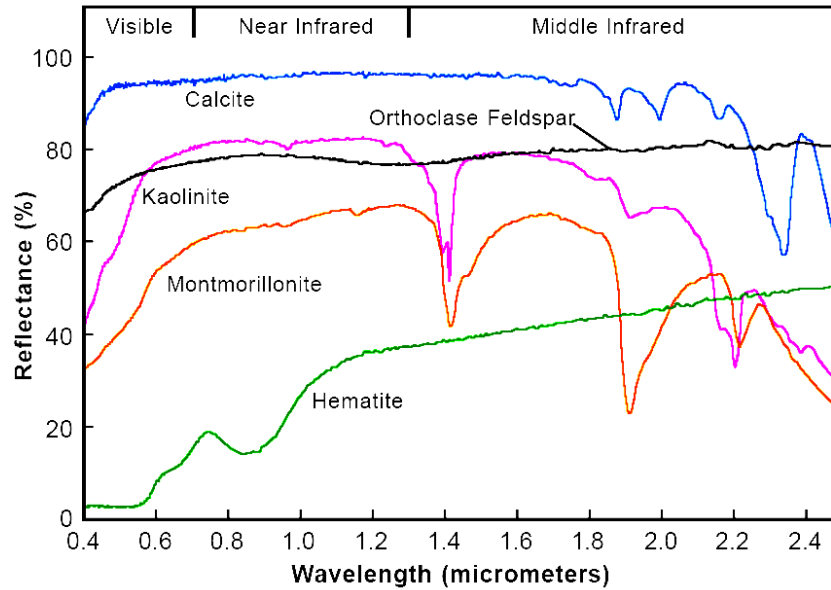
Figure 3: Mineral Spectra [22]

In Figure 4 it can be observed that there is higher absorption at certain spectral wavelength due to water and carbon dioxide which for essential component of earth's atmosphere. Hence higher absorption is seen in this region in the spectra plotted for hyperspectral imaging. Figure 4 shows absorption properties for gases and water present in earth's atmosphere. It can be observed that due to carbon dioxide and water high absorption of spectral energy is seen around 1.4 µm and 1.9 µm. Hence the interaction of light with the atmosphere reduces the energy of light in these spectral ranges resulting in capture of very little information of the area under study. Sensor noise dominates over the incoming spectral information and hence after spectral correction these bands have very low pixel intensity.
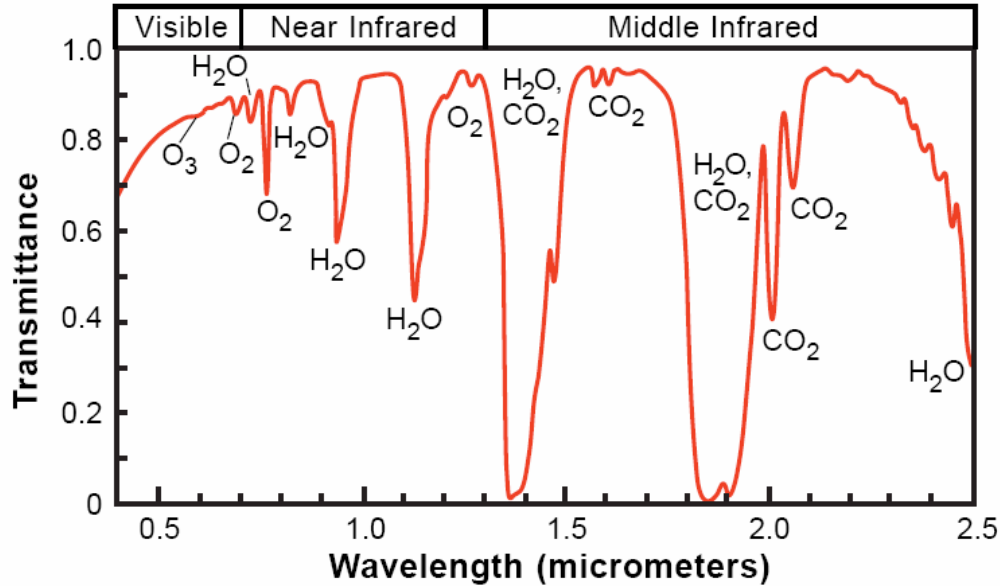
Figure 4: Spectral curves for atmospheric absorption

Spectral curve captured by AVIRIS sensors shown in Figure 5 show the dominance of atmospheric absorption in the 1.4 µm and 1.9 µm range corresponding to band 107 to 114 and 153 to 160. These bands have poor spectral correlation and hence best predicted by spatial predictors or best encoded without prediction. One of the reasons is that these bands have very low pixel intensities and low self entropy.

2.2 AVIRIS Sensors and AVIRIS Data

The main focus of this thesis is on lossless compression of hyperspectral images captured by AVIRIS (Airborne Visible Infrared Imaging Spectrometer) sensors. Hyperspectral images captured by the AVIRIS sensors have been widely used by researchers for the study of lossless compression of hyperspectral images. These images are easily available on the JPL (Jet Propulsion Laboratory) website [2]. Hyperspectral imaging is acquisition of images from the same area into narrow contiguous spectral bands, also termed as "Imaging Spectroscopy" [2, 23, and 22]. The main objective of the AVIRIS project is to identify, measure, and monitor constituents of the Earth's surface and atmosphere based on molecular absorption and particle scattering signatures. Research with AVIRIS data is

predominantly focused on understanding processes related to the global environment and climate change.

The AVIRIS instrument contains 224 different detectors which have wavelength sensitive range of 10 nm [2]. The detectors acquire images in the range of 400nm to 2500nm covering the visual, near infrared and short wave IR spectrum. The area covered on earth depends on the distance of the satellite from the earth. At a distance of 20km each pixel produced by the instrument covers an area of 20 m diameter hence covering an area of 11km for one scene while at a distance of 4 km each pixel covers an area of 4 m diameter covering the total area of 2km for one scene. AVIRIS has been flown on aircrafts which fly at an altitude of 4 km and 20 km and across two continents North America and Europe. Each scene captured by AVIRIS has 614 pixels per line, 512 lines and 224 bands after preprocessing. The size of each scene is around 140 megabytes. Each flight of AVIRIS captures 76 Giga bytes of data every day.
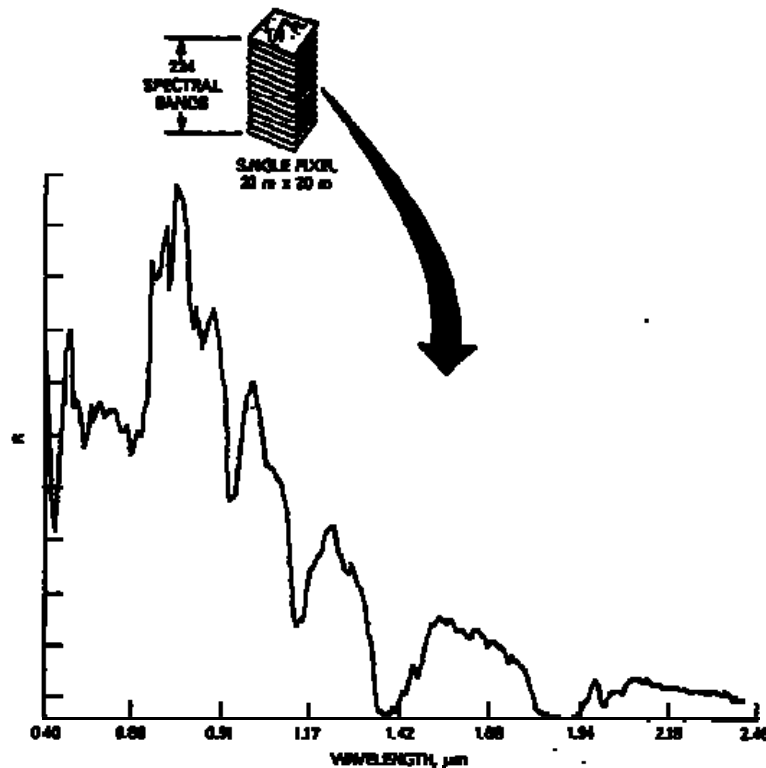
Figure 5: Single Pixel Spectrum from AVIRIS [2]

Figure 5 taken from [2] and it shows the single pixel spectrum of AVIRIS. Here the x-axis is the channel wavelength while y-axis is radiance. The spectrum of AVIRIS is dominated by the light curve of the sun and the absorption and scattering properties of the atmosphere. Sun has a black body curve which implies that the curve peaks in the green wavelength and diminishes for all other wavelengths [2]. The valleys in the curves are due to the absorption properties of the atmosphere. Different components in the atmosphere like nitrogen, oxygen, carbon dioxide, water vapor absorb light at different frequencies. The valley in Figure-5 between 1.4 µm and 1.9 µm is due to water vapor present in the atmosphere. The variation in the curve other than due to sun and atmosphere depict the information of the chemical composition of the material being studied. Like in the case of vegetation chlorophyll present in plants absorb the visible spectrum of light and reflects the infrared light. Hence pixels capturing vegetation will show a peak in infrared region.
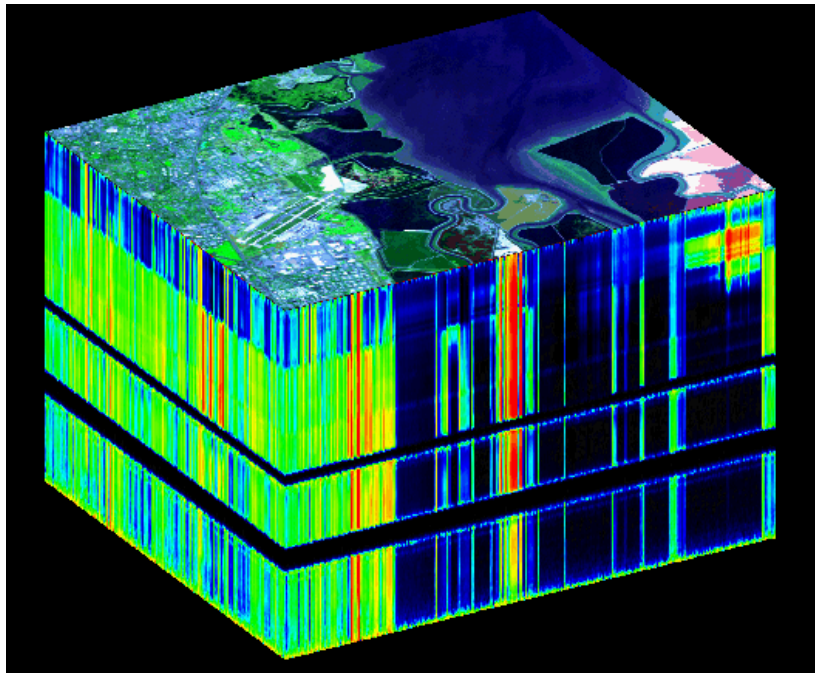
Figure 6: AVIRIS Moffet field Image Cube [2]

Figure 6 taken from [2] shows the AVIRIS image cube for the Moffet field where all the 224 bands are stacked over each other. The top of the stack is a false color image to show the image captured. To the right is the water body and to the left is the Moffet field airport. To the right of the stack at 700 nm which is the red part of the visible spectrum a red area is seen. This is due to the red shrimps present in the pond.

Jet Propulsion Lab (JPL) website has 5 different image sets captured by AVIRIS in 1997 for download [2]. Research on AVIRIS data has been carried out since last two decades. Prior to 1997 data acquired in 1989 by the AVIRIS instrument were available for research purposes. The 5 image set acquired by the AVIRIS instrument which have been made available for research and commercial use are Cuprite, Moffet Field, Jasper Ridge, Lunar Lake and Low Altitude. Each image is of different size and comprises of scenes of size 614×512 wherein the last scene has lesser number of rows in each image set. .
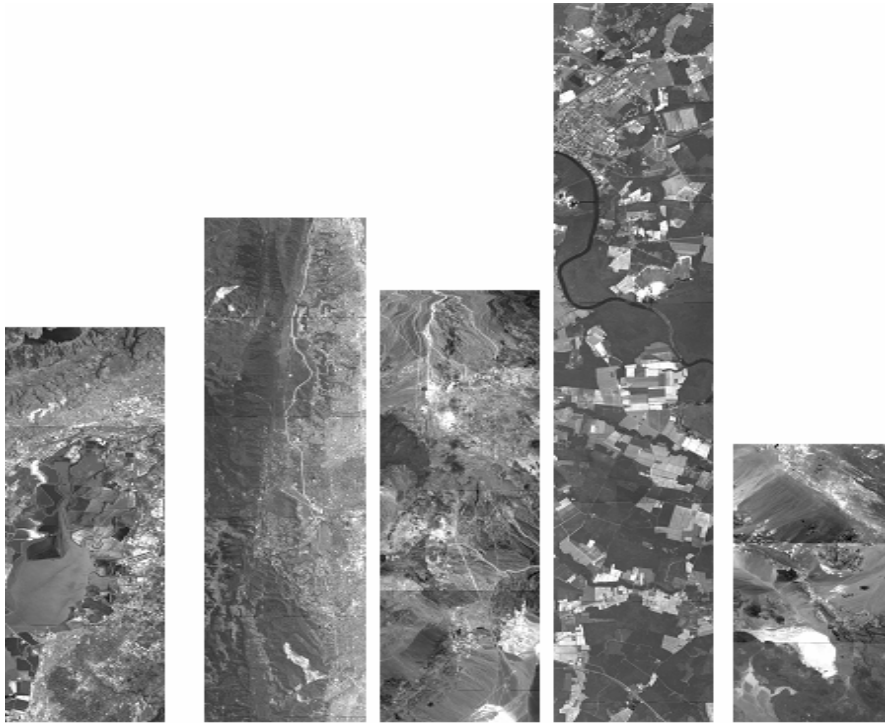
Figure 7: Moffet Field, Jasper Ridge, Cuprite, Low Altitude, and Lunar Lake

An image of each of the AVIRIS image has been shown in Figure 7. These images were acquired from [2]. The images have been rescaled but they depict actual size with each

other. It can be seen from the Figure 7 that low altitude is the largest of the images and lunar lake is the smallest in size.

2.3 Lossless Compression of Hyperspectral Images

Research in data compression has been mainly focused through three different approaches transform coding, prediction based and vector quantization based. These algorithms have been implemented successfully on different images in 2-D and 3-D domain.

- Transform Coding: The original data is transformed from spatial domain to frequency domain or wavelet domain depending on the transform used. Discrete Cosine Transform, Discrete Wavelet Transform, Discrete Fourier Transform and Discrete Sine Transform are the main compression tool in transform coding [16, 42, 45, 1, 3, 5]. The coefficients in transform domain are truncated for higher compression, but this results in lossy compression. For lossless compression the error between the retrieved data and the original data is sent with the truncated coefficients. Though the transform based compression schemes are fast they yield lower compression ratios.

- Prediction Based Compression: Main focus of research in data compression has been on prediction based techniques. The simplest prediction based technique is DPCM wherein the pixel is predicted based on the previous pixel in raster scan order. The difference between the two pixels is then encoded. Higher compression ratios can be achieved with prediction based schemes [21, 9, 20, 25, 55]. The focus of this thesis is on prediction based method.

- Vector Quantization: A code book of vector is formed from the data. The difference between the code book vector and the incoming vector is encoded. The code book vector needs to be sent with the encoded data [47, 31 11, 41, 26].

Prediction based compression scheme can be broken down into two steps: prediction based on the causal data set which has already been encoded, and encoding of the prediction error residual. Simple DPCM based technique gives low compression rates

hence adaptive DPCM was introduced [6, 44]. But Adaptive DPCM lacked the modeling ability and hence yielded low compression ratios. A good prediction scheme should be able to model the pixels into smooth areas, edge areas and textured area. Usually the smooth region yields lower prediction error residuals compared to edge area and textured area. Many algorithms have been proposed which model the local area for prediction. JPEG LS [15, 14] is the standard image compression algorithm which uses median based predictor MED (Median Edge Predictor) for prediction. CALIC [53] is also a know algorithm in the field of image compression which models the local area pixel. CALIC utilizes GAP (Gradient Adjust Predictor) for the purpose of prediction. 2-D compression algorithms have been applied to Hyperspectral Imaging but they failed as they did not consider the spectral correlation between the bands. 3-D implementation of CALIC [55] was proposed but it still did not yield higher compression ratios.

An interband linear predictor and a least square based approach for lossless compression of hyperspectral images was proposed in [39] in 2005. The linear prediction [LP] scheme computes the average difference between the context of the pixel $I(x, y, t)$ in the current band and the context of the pixel $I(x, y, t-1)$ in the reference band. If the deviation between the maximum and minimum value of the difference is greater than a threshold than the prediction is corrected by adding the average prediction error of previous two bands. The linear prediction scheme is not completely interband. It switches between interband predictor and intraband predictor for the band marked for the respective prediction. It was observed in [39] that the first 8 bands are best predicted by intraband predictor. Standard median predictor is used for intraband prediction.

The least square based prediction scheme proposed in [39] is fully interband. SLSQ (Spectral Oriented Least Square) utilizes 4 neighboring pixels from the current and the previous band, to optimize the prediction at each pixel in each band. Least square based approaches have been proposed for lossless compression of standard images [18]. They optimize the prediction parameters along the edge hence improving the prediction at edge pixels. Also the least square based approaches perform better along the edges than in smooth areas [33]. With SLSQ, the authors observed that with increase in context size

there is very little improvement in the final compression ratio. Hence they have fixed the context size to 4 to reduce the overall complexity of the algorithm. The SLSQ method involves 4 multiplications and 6 additions at each step of prediction.

Two methods heuristic and optimal have been proposed in [39] to improve the overall compression ratio. In HEU, all the bands in all 5 image sets are predicted by the intraband scheme. Bands which are better compressed using intraband scheme are marked as IB. To generalize the intrabands across all 5 image sets, bands which perform better in more than 15 scenes out of 26 scenes are marked as intraband. These bands are predicted using median predictor in all the scenes in the 5 image sets. In the optimal scheme (marked as OPT), the bands are predicted using both intraband and interband prediction scheme. The scheme giving better result is used to predict the respective band.

Clustering based approach has also been proposed in [26]. In [26] the LBG (Linde Buzo Gray) [19] algorithm was used to cluster the bands based on their spectral content. LBG is a vector quantization design technique used to develop the code book for compression. The first code vector is formed by taking average of the whole training set. Then at each step the code vector is split into two till the desired number of code vectors is not obtained. In [26] the LBG algorithm was used to cluster similar bands and not to form the code book for vector quantization. Thus in this case, the code vector forms the centroid of the cluster and the bands closest to a given centroid forms part of that cluster. A multiple linear regression scheme is used to design a linear predictor for each of the clusters and each band. Each cluster is encoded using a range coder whose model is dependent on the cluster. A two mode approach similar to [39] is implemented in [26]. All the bands are entropy coded without prediction and if the entropy of the band is better than the entropy with prediction than the bands are stored without prediction. In [26], using 20 previous bands for predicting, the LBG is set to form 16 clusters as it was found to give optimal results. [26] has the best result published, but its algorithmic complexity is very high due to the clustering and the use of linear regression in computing optimum coefficients for each band in each cluster.

In [50] two methods were proposed, one based on JPEG's 7 predictors and the other using correlation-based conditional averages. The motivation for this works comes from the fact that hyperspectral images have the same global structure. Hence the relationship of a pixel and its surrounding in one band should be similar to the relationship of a co-located pixel and its neighbors in the neighboring bands. Hence if a predictor is designed for a pixel in the previous band it should hold true for a co-located pixel in the current band. Seven JPEG predictors were used to predict the pixel in the reference band and the best predictor was selected to predict the pixel in the current band. As the reference band is used to calculate the optimal predictor no overhead is required to derive the predictor at the decoder. In the correlation based conditional average (CCAP) technique the author's use a context size of 4 since increasing the context size did not improve the result. The correlation is computed between the context of the pixel in the current band and context of the co-located pixel in the 5 previous bands. The context having the highest correlation with the context in the current band is used to predict the pixel in the current band. Since the images captured in each band may not be at the same pixel locations, (i.e. the images may be shifted by few pixels) the authors propose a search window of 3×3. The correlation coefficient is computed between the context of the current band and the context formed by the pixels in the search window in all 5 previous bands. Due to different dynamic range of each band, rather than using a linear predictor, a scaled prediction is computed $\hat{y} = y_1 \left(\dfrac{x}{x_1} \right)$, where y_1 is the adjacent pixel to the pixel in the current band, x_1 is the adjacent pixel to the collocated pixel in the reference band and x is the collocated pixel in the reference band.. To improve the robustness of the prediction, rather than one single prediction value, the average of prediction of all the pixels having correlation coefficient higher than 0.95 is computed. [50] has the best published results for Cuprite but on an average its performance was not better than the most published results.

Standard lossless compression schemes such as JPEG-LS, JPEG2000, and CALIC perform poorly on hyperspectral images. It was shown in [39] that some improvements can be made if the difference between bands is compressed using JPEG- LS or

JPEG2000. But this improvement is not significant enough as compared to already existing algorithms for lossless compression Hyperspectral images.

Two extensions of the original CALIC algorithm, namely 3D CALIC and M-CALIC were proposed in [55] and [20] respectively. In [55] correlation based approach similar to [50] is used to switch between interband mode and intraband mode. Correlation between the context of the pixel in the current band and the context of the co-located pixel in the previous band is computed. If the correlation between the two context is higher than 0.5 than interband method for prediction is applied otherwise the intraband method for prediction is used. Standard CALIC is used for intraband compression. In case of the interband prediction technique $\hat{y} = \alpha X + \beta$ is computed so as to minimize the $\left\| \hat{Y} - Y \right\|_2$.

Here α and β are calculated from the context formed in the current and previous band. In case of edges, horizontal and vertical gradients are computed in the reference band and the predicted value is calculated using the gradients from the current band and the reference band as was done in [50]. Similar to CALIC, 3D CALIC also operates in two modes binary and continuous. The selection of binary mode in 3D CALIC depends on the pixel values in both the reference band and the current band. If the binary pattern between the two contexts is different then the pixel in the current band is predicted using the binary mode in intraband CALIC. In case if the two binary patterns are same and the symbolization for the pixel in the current band and the reference band is same then the value of current pixel can be uniquely determined. In case the binary patterns are same but the symbolization for the pixels is different the mode is switched to the continuous mode. The context selection for biased error cancellation is dependent on the correlation coefficient. The higher the correlation coefficient the lower is the prediction error residual. Hence the correlation coefficient plays an important role in context formation for both biased error cancellation and encoding. An extension to 3D CALIC algorithm was proposed in [20]. In M-CALIC the reference band was computed by taking weighted averages of previous 2 bands. The reference band is computed as: $\hat{y}_\lambda = \gamma_1 + \gamma_2 x_{\lambda-1} + \gamma_3 x_{\lambda-2}$ .The coefficients γ_1, γ_2 and γ_3 are fixed to 8, 0.65, and 0.35 respectively. The other changes over 3-D CALIC proposed in M-CALIC were the

quantization parameters and model parameters. These changes improved the results of M-CALIC over 3-D CALIC scheme.

Vector quantization based lossless compression of AVIRIS image is discussed in [41]. In vector quantization based lossless compression scheme a codebook vector is formed form the image set to be compressed. Then the difference between the image vectors and the closest vector in the code book is entropy coded with the index of the vector in the code book. In case of hyperspectral images the codebook is formed spectrally and not spatially. In [41] the 224 bands are broken into 'b' blocks and codebooks are formed for each block such that the length of each vector in the codebook is 'v' and the product of 'v' and 'b' equals to 224. The k-means algorithm was then used to design the codebook, using 32 vectors in each codebook. With higher number of vectors in the codebook, the input vector becomes closer to the reference vector. Hence the difference image entropy will reduce but the entropy of the addresses will increase. Hence a tradeoff is needed between the size of the codebook and the achieved compression.

2.4 Ordering Techniques for Hyperspectral Data

In multispectral/hyperspectral imaging the previous image may not be always the best image to predict the current image. Tate [48] showed that with optimal ordering the performance in terms of compression ratio improves significantly. Ordering of bands can be based on the compression ratios achieved by compressing all the bands with each other or by clustering the bands based on there global or local features. If the compression of bands has been performed already, the complexity of ordering when all the bands are compared with each other is $O\left(n^2\right)$, where n is number of bands. This complexity can be reduced by clustering techniques wherein the features in each band is used to cluster the bands.

The problem of ordering of hyperspectral/multispectral bands for compression was addressed by Stephen Tate [48] in 1997. He proposed an algorithm which had an order of

time complexity of $O\left(n^2\right)$ after computing the compression obtained by compressing one band with the other. Thus, the time for estimating the compression is not included, which could be quite substantial given the size of the bands and the number of bands. In his algorithm he calculated the compressed size of the band depending on if the current pixel is available in the predictor band or not. He obtained two matrices which had entries of compressed sizes of the band given that the current pixel is available or not in the predictor band. He transforms the problem of optimal compression into problem of weighted directed graph. Here the weights for the edges between the bands come from the two matrices depending on which of the compressed size is less. For a sparse weighted graph the solution has an order of $O\left(|V|\log|E|+|E|\right)$ where V is the number of vertices in the graph and E is the number of edges in the graph [48]. For a dense weighted graph the solution has an order of complexity of $O\left(n^2\right)$. He further proposed that for larger set of data where N is very large the bands should be grouped in small sets so that if one band needs to be accessed then only few number of bands will be needed to be uncompressed, whereas in earlier case in an extreme case all the bands might be needed to be uncompressed to access a single band. In [48] it was shown that optimal ordering improved the performance of AVIRIS (1989) and CZCS data set but the improvement in TM and AVHRR dataset was not significant. This shows that improvement due to ordering is data dependent. Tate showed that picking 5 adjacent blocks and optimally ordering those significantly improved the results for AVIRIS data set.

In [28] five different ordering methods for lossless compression of hyperspectral images were discussed. The five ordering methods discussed in [28] were namely forward monotonic, reverse monotonic, best forward, best reverse and optimal order. Forward monotonic and reverse monotonic are the simplest of the band ordering techniques. The bands are predicted in either ascending order or descending order depending on the technique used. The current band is predicted by the closest neighbor in ascending or descending order. The measure of band correlation is computed by the calculating the PMSE between the two bands. Lower the PMSE between the two bands lesser number of bits will be required to compress the bands. In the best forward and best reverse ordering

PMSE is computed between the current band and all the bands that have been encoded prior to the current band. The band having the lowest PMSE with the current band is used to predict the band. The final ordering method discussed in [28] is the optimal ordering technique. In this method the PMSE is computed between all possible pairs of bands that can be formed. The order of complexity of this technique is $O(n^2)$ where n is number of bands. The optimal ordering method discussed in [28] is similar to that proposed in [48]. In [48] the weight of the edge is a directed graph computed by calculating the compression ratio between the pair of bands while in [28] PMSE is computed between the bands.

In [49] a correlation based band ordering technique was proposed. Instead of computing the PMSE or compression ratios between the bands, correlation between the bands is computed. To reduce the complexity of the ordering method the bands are down-sampled before the correlation is computed between the bands. A significant gain in computation time is obtained with slight loss in compression ratios. The weighted graph formed is solved using a minimum spanning tree algorithm; in this case, Prim's algorithm is used. Since the correlation coefficients are computed between the bands a maximum weight tree is formed instead of a minimum spanning tree. The results for the correlation based ordering is compared with optimal ordering proposed in [49] and it was observed that at a reduced ordering complexity the results were slightly poorer than obtained by optimal ordering. To reduce the overall complexity of ordering a new method was proposed in which the correlation is computed between the current band and 42 of its closest neighbors. The results though degraded slightly but significant reduction in computation time was obtained.

The performance of the ordering scheme depends on the method of prediction and the number of bands involved in ordering. In [49] the performance of the system reduced slightly when lesser number of bands was involved in ordering but significant improvement in computation time was obtained. The optimal ordering methods proposed in [48, 20] cannot be implemented onboard but they provide a limit on the achievable compression ratio for a compression method used. There is a need to study ordering

methods having lower computation time so that they can be implemented successfully on the satellites to improve on the obtained compression ratio performance.

Chapter 3: Prediction

3.1 Introduction

Two methods for lossless compression of hyperspectral images are proposed in this thesis. The first method predicts the band in ascending order, wherein the current band is predicted by the previous band. The encoder and decoder complexity is same for this method. The second method orders the band based on different features estimated between the bands. The complexity of the encoder and decoder is different; the complexity of the encoder is higher than the decoder. The decoder complexity remains the same in both the lossless compression techniques.

Table I, Table II and Table III give an idea about the nature of the hyperspectral images and the motivation to use spatio-spectral predictor for lossless compression of hyperspectral images. In Table I the average statistics for the different images in AVIRIS data have been shown. The data was collected for a small band size of 256×256. Table I shows the average variation of the bands. It can be seen from the table that the range of pixel intensities over the 224 bands in each image set is very large. It can be seen that the entropies of Moffet Field and Low Altitude is higher that the entropies of the other images sets in AVIRIS data. Also the standard deviation of these images is higher than the other images.

TABLE I AVERAGE IMAGE STATISTICS FOR AVIRIS DATA SET

	Mean	Median	Standard Deviation	Minimum	Maximum	Entropy
Cuprite	1796	1801	21	-17	8352	7.0879
Jasper Ridge	1072	1064	32	-17	7016	7.4674
Moffet Field	1232	1292	136	-18	7814	8.6798
Lunar Lake	2269	2262	55	-13	12127	7.3321
Low Altitude	1248	1244	68	-52	19011	8.1057

In Table II the effect of spatial prediction has been shown. It can be observed that the prediction error residual entropy of the bands has improved over the self entropy of the bands. Standard median predictor has been used to predict the bands.

TABLE II EFFECT OF SPATIAL PREDICTION ON AVIRIS DATA

	Mean	Median	Standard Deviation	Minimum	Maximum	Entropy
Cuprite	-0.8368	0	21.5711	-2291	5379	6.2093
Jasper Ridge	-0.8427	0	11.6909	-2077	3226	6.5403
Moffet Field	-1.7266	0	21.6527	-2544	3700	7.1742
Lunar Lake	-0.6776	0	34.3098	-2742	6848	5.7449
Low Altitude	0.8506	0	26.6915	-17901	18832	7.2264

Table III shows the results for spectral prediction on AVIRIS data set. The pixel in the current band is predicted by the co-located pixel in the reference band. It can be observed that with simple differences between the bands significant improvement in terms of bits per pixel has been obtained over the self entropy and entropy of prediction error residual obtained by spatial prediction. Also the range of error values has reduced significantly over the range of error values obtained with spatial predictors. Better de-correlation of the bands is obtained with simple differences. Higher compression ratios can be obtained with spectral predictors as compared to spatial predictors.

TABLE III EFFECT OF SPECTRAL PREDICTION ON AVIRIS DATA

	Mean	Median	Standard Deviation	Minimum	Maximum	Entropy
Cuprite	-4.1982	-4	2.3349	-2302	1877	5.2585
Jasper Ridge	-2.7301	-4	4.5049	-1877	1390	5.4281
Moffet Field	-3.3561	-10.125	24.2412	-2293	2111	6.2814
Lunar Lake	-4.8387	-8	6.2669	-3235	2502	5.3816
Low Altitude	-3.1420	-1	11.0795	-4930	3223	6.0803

In Figure 8 the entropies for the bands for different image sets have been plotted. Figure 8(a) shows the self entropy of the bands. It can be seen that the change in the entropy on an average remains the same across all the image sets. This is due to the dominance of the absorption properties of the atmosphere and spectral energy curve of the sun. The variation between the curves shows the effect of different materials captured at different wavelengths. In Figure 8(b) the entropies for prediction error residuals for spatial prediction on the bands in each image set have been plotted. It can be observed that the spatial predictors perform better on an average over the bands. Significant improvement is obtained in few of the bands.
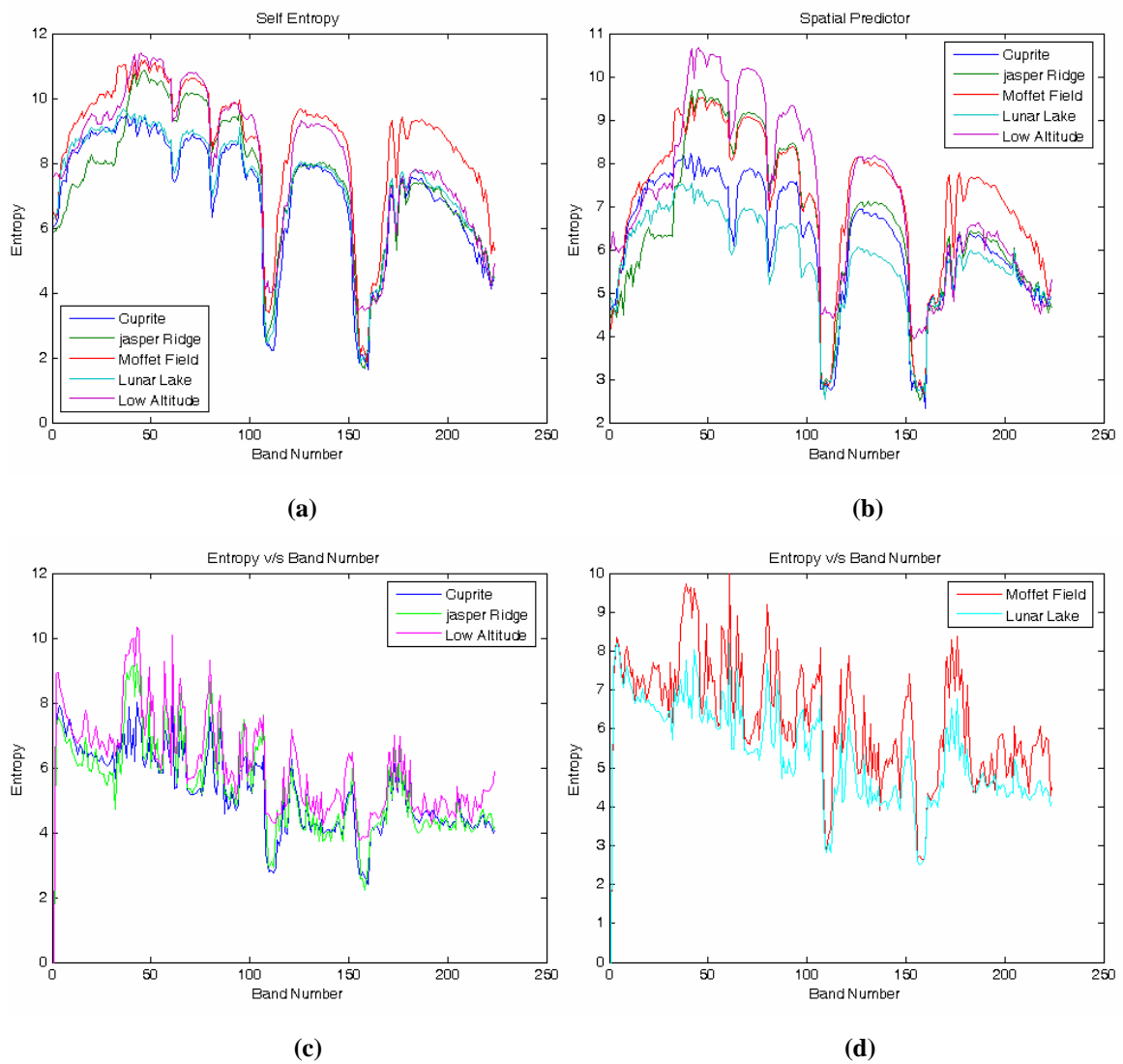

(a) (b)

(c) (d)

Figure 8: Entropy plots for different predictors; (a) Entropy of the bands, (b) Entropy for spatial prediction, (c) Entropy for Spectral Prediction for 3 bands , (d) Entropy for Spectral Prediction for 2 bands

In Figure 8 (c) & 8(d) the effect of spectral prediction on the AVIRIS images have been shown. It can be observed that on an average the entropies obtained with spectral prediction is better than the entropies obtained with self entropy of the bands and with spatial prediction. It is important to notice that the self entropy of the bands is lower than the entropies obtained with the two prediction methods obtained for bands in the range of 150 to 160. Also for bands in the range of 1 to 15, 105 to 115 the spatial predictors perform better than the spectral predictors. These bands have poor spectral correlation and hence perform poorly with spectral predictors. Two figures have been plotted for entropies obtained with spectral prediction since the plots were not distinguishable when plotted together.

Lossless compression of hyperspectral images is achieved using three different predictors namely, spectral predictor, spatial predictor, and encoding without prediction. Hyperspectral images have high spectral correlation as the images captured are of the same area but at different spectral frequency [39, 50]. Hence the global structure in each band is similar but they are at different intensity levels. In [39, 50] it has been shown that the performance of the spectral predictors is better than the standard lossless image compression algorithms. Hyperspectral images are affected by the atmospheric absorption and scattering. The light energy emitted by the sun is either absorbed or scattered by the atmospheric particles like gases and water. The bands affected by this phenomenon have low spectral correlation and dominance of sensor noise. These bands are best predicted by some spatial predictor or perform better when they are encoded without prediction.

The cost of computing the compression ratio for each band with three different prediction methods is very high, but significant improvement can be obtained in terms of compression ratio. Since the lossless compression method for hyperspectral images are designed to be implemented onboard on the satellites it is necessary to keep the computation time low. Though the memory space and computation time required by the spatial predictor is very less compared to the spectral predictor it still affects the overall computation time. It was observed that certain bands were best predicted by spectral

prediction and these bands remained fixed over all the scenes and image sets. Hence we can fix the bands which are to be predicted by one of the prediction method rather than predicting each band by three different methods. This reduces the overall computation time and memory requirement at cost of very little change in the compression ratio. Also we need not send the header information required to specify the type of prediction method applied to each band. Though this observation was made on 26 scenes across 5 different image sets it can be generalized for the AVIRIS data set. Similarly the information can be generalized for other data sets captured by different sensors but only in this case the predictor assignment will be data dependent as different number of bands at different frequency range is captured by different sensors. Atmospheric absorption is due to the gases and water molecule present. These elements affect the spectral energy is specific bands as seen in the five image sets acquired from the JPL website. Hence the prediction scheme can be generalized over all the hyperspectral images if the wavelengths of the bands are known.

The proposed spectral prediction scheme is non-linear as the prediction is made based on the locality of the pixel and its context. The spectral prediction scheme is a context based prediction scheme and requires 12 neighboring pixels in the current band and reference band for prediction. The prediction is performed in raster scan order. Most of the state-of-the-art algorithms have a feedback scheme to minimize the prediction error; similarly we calculate the error made in computing the weights for the prediction of pixels in the context and use these to modify the weights for the prediction of current pixel.

Figure 9 shows the block diagram for the prediction method proposed in this thesis. The initial block of edge analysis is present in the edge based prediction (EPHI) method proposed. Here the pixels in the context are classified into edge and non-edge pixels depending on the threshold set for that band. The c in $I(c,t)$ defines the context used. In case of edge-based prediction the pixels in the context will be reduced set of the original set depending on the edge classification. The block coefficients for pixels in the reference band compute the weights to predict $I(x, y, t-1)$. These weights are modified by α which is the variation between the two contexts. The weight error buffer computes the

difference between the ratio of the contexts, and the ratio of the weights from reference band and the actual weights of the current band, after each prediction. An average of the error computed is stored. The weight error buffer stores all the average differences computed for the pixels in the context. The average of these weights is used to modify the ratio computed between the two contexts to compensate for the scale factor. This factor is multiplied with the weights computed in the reference band to obtain the weights for the pixels in the current band.
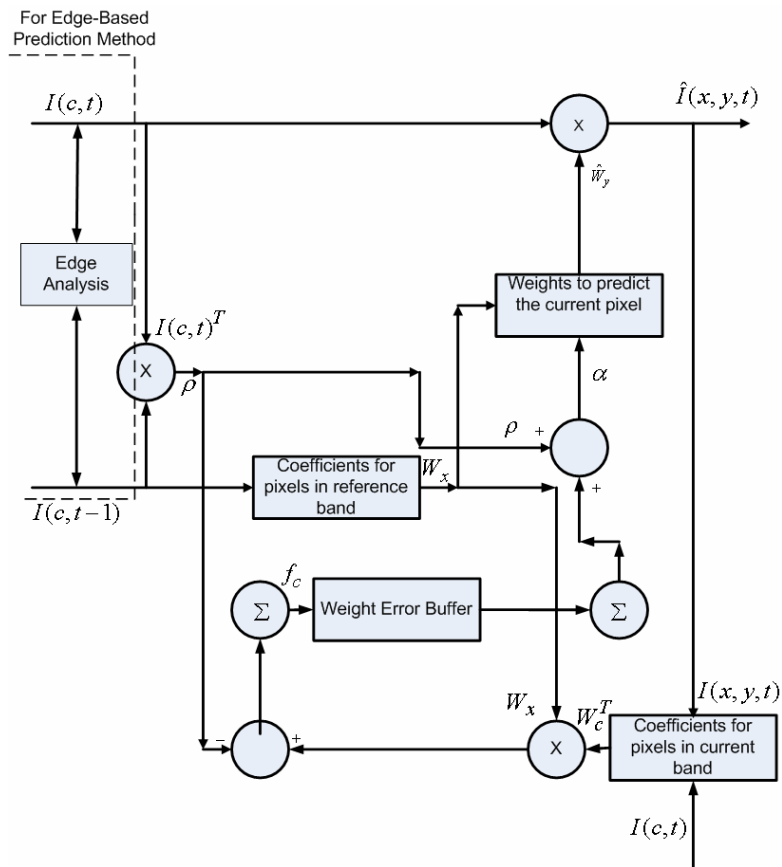
Figure 9: Block Diagram for the Prediction Method

3.2 Prediction Algorithm

In this section context based spectral prediction without edge based consideration is discussed. The spectral predictor utilizes 12 neighboring causal pixels in the current band and 12 neighboring causal pixels in the reference band. The coefficients calculated for the pixel in reference band are modified according to the variance and scale between the two contexts to obtain the weights to predict pixels in the current band. The variance and scale between the contexts can be obtained by taking the ratios between the two contexts. The weights calculated are fine tuned by a feedback mechanism wherein the error made in calculating the weights for the pixels in context is used to correct the weights calculated for the pixel being predicted.

3.2.1 Spectral Prediction - NPHI

Hyperspectral imaging is acquisition of images wherein different spectral energy arriving at the detector is stored in digital form. As the portion of area being acquired is same in all the spectral bands the global structure of the image in all the bands remains same. Every material displays different absorption and reflectance properties at different wavelengths hence the intensity of the acquired image varies from band to band. The average intensity variation across the bands is from 10 to 5000. The major reason for such large variation is due to the effect of atmospheric absorption on light energy at some spectral frequencies. Since the images have same global structure they generally have high spectral correlation than spatial correlation. Hence the spectral predictors perform better on most of the bands. Some of the bands affected by atmospheric absorption are best predicted by spatial predictors as they have poor spectral correlation.

Spectral prediction schemes proposed till date utilize one band-look ahead schemes [39], more than one band [26, 50] or average of bands [20] to predict the band under consideration. The proposed prediction scheme (NPHI) utilizes one band look-ahead method to predict the bands. The band decoded last is used to predict the next band. The spectral prediction scheme works on the basis that the variation of pixels in one band will

be similar to the variation in other bands since the bands have same general structure. Hence if we can compute the variation between the two bands it is possible to predict the pixel in current band and minimize the prediction error residual. As discussed in [54] two large a context size leads to context dilution and small context size does not hold enough information to predict the pixel in order to minimize the prediction error. Hence context size of 12 was fixed after testing partial image set for different context size.

Figure 10 and Figure 11 show the context formed to predict the pixel $I(x, y, t)$. Figure 10 is for the context in the current band, while Figure 11 is for the context in the reference band.

$I(x-2, y-2, t)$	$I(x-2, y-1, t)$	$I(x-2, y, t)$	$I(x-2, y+1, t)$	$I(x-2, y+2, t)$
$I(x-1, y-2, t)$	$I(x-1, y-1, t)$	$I(x-1, y, t)$	$I(x-1, y+1, t)$	$I(x-1, y+2, t)$
$I(x, y-2, t)$	$I(x, y-1, t)$	$I(x, y, t)$		

Figure 10: Prediction context formed in the current band

$I(x-2, y-2, t-1)$	$I(x-2, y-1, t-1)$	$I(x-2, y, t-1)$	$I(x-2, y+1, t-1)$	$I(x-2, y+2, t-1)$
$I(x-1, y-2, t-1)$	$I(x-1, y-1, t-1)$	$I(x-1, y, t-1)$	$I(x-1, y+1, t-1)$	$I(x-1, y+2, t-1)$
$I(x, y-2, t-1)$	$I(x, y-1, t-1)$	$I(x, y, t-1)$		

Figure 11: Prediction context formed in the reference band

In Figure 10 and Figure 11 x and y represents the spatial coordinates of the pixel while 't' represents the spectral position (i.e. the band number). Hence $I(x, y, t)$ represents the intensity of the pixel located at position (x, y) in the tth band.

As the dynamic range could vary significantly between bands, prediction using differences will not produce good results. Similarly, global normalization is not feasible due to the dynamic range of the images which introduces error in prediction. Different image range will lead to stretching of the image with lower range leading to higher errors in prediction. Also local normalization becomes computationally intensive over all the pixels in all the bands. Hence scale based prediction is used where ratios are calculated instead of linear prediction [50]. The prediction scheme works on the basis that even with sensor noise there is some correlation between bands. Hence if the pixel variation between two contexts can be computed efficiently then the prediction error magnitudes can be reduced. Ideally the coefficients calculated for $I(x, y, t-1)$ should apply for prediction of $I(x, y, t)$, but this is not the case as the two images have different dynamic range. Also since the images are captured at different spectral frequency there is slight variation in the pixel values. But this difference can be compensated for if we can capture the variation between the two bands.

Coefficients for prediction of $I(x, y, t-1)$ can be calculated as shown in (1) and (2). As the pixels have already been decoded the weights for prediction of $I(x, y, t-1)$ can easily be computed by taking the ratio of each pixel in the context with $I(x, y, t-1)$ as shown in (1).

$$W_x(x-i, y-j, t-1) = \frac{I(x, y, t-1)}{I(x-i, y-j, t-1)} \qquad i, j \in C \qquad (1)$$

$$I(x, y, t-1) = \frac{1}{n_c} \sum_i \sum_j \left(W_x(x-i, y-j, t-1) * I(x-i, y-j, t-1)\right) \qquad i, j \in C \qquad (2)$$

Here n_c is the context size W_x are the weights computed for the pixels used to predict $I(x, y, t-1)$. 'C' is the context used. Ideally the weights computed for the co-located pixel $I(x, y, t-1)$ in the reference band should be similar to the actual weights of the pixel $I(x, y, t)$ in the current band. Since the bands could be at different intensity

level and the light absorbed by different materials is different at different wavelengths there is slight variation between the weights computed for the reference bands for the pixel $I(x,y,t-1)$ and the actual weights for pixel $I(x,y,t)$ in the current band. The variation between the two bands can be computed from the two contexts. The ratio between the bands will help us in identifying the variation in the bands.

$$W_x \propto W_y \qquad\qquad\qquad (3)$$

Where, W_y is the actual weight to predict the pixel $I(x,y,t)$ in the current band. The proportionality can be made similar if the weights W_x computed in the reference band are modified by a factor which is dependent on the variation between the two contexts.

$$W_y \approx \alpha \cdot W_x \qquad\qquad\qquad (4)$$

In (4) α is the factor which should be estimated to make W_x as close as possible to W_y. The variation between the two bands can be estimated by taking the ratio between the two contexts.

$$\rho(x-i,y-j,t) = \frac{I(x-i,y-j,t-1)}{I(x-i,y-j,t)} \qquad\qquad i,j \in C \qquad\qquad (5)$$

Since the bands have different pixel intensity levels the scale factor between the two contexts affect the calculations of ρ. To nullify this, we can either compute the normalization factor between the bands or compute the value that needs to be added or subtracted from the ratios. The normalization will spread the ratio around the zero value which will make some of the ratios to be negative. Since this is a non-linear method of prediction, and also ratios are computed for all the calculations, the variance computed between the two contexts will be multiplied to the weights W_x of the reference band. The negative ratios will affect the weights and finally the prediction. One method to correct

the ratios from the pixels in the context is to normalize the two contexts before taking the ratios. However; as pointed out earlier, normalizing the two contexts leads to change in pixel gradient which will no more be proportional to the earlier pixel gradient for that context. The context with smaller range will be stretched in terms of pixel values with respect to the other context leading to errors in the prediction. Secondly if the two contexts are normalized independently they will lead to pixel values being normalized around zero leading to negative ratios. Here we avoid these problems by determining the error made in computing the weights after predicting each pixel and using this value to correct the ratios. Here the error computed for each pixel in the context will be used to correct the current ratios.

The following calculations are computed for the pixels in the context. After predicting each pixel actual weights are computed for those pixels. Similar to (1) and (2) we compute the actual weights for the pixel in the context of pixel $I(x, y, t)$.

$$W_c(x-i, y-j, t) = \frac{I_c(x, y, t)}{I_c(x-i, y-j, t)} \qquad i, j \in C \qquad (6)$$

$$I_c(x, y, t) = \frac{1}{n_c} \sum_i \sum_j \left(W_c(x-i, y-j, t) \cdot I_c(x-i, y-j, t) \right) \qquad i, j \in C \qquad (7)$$

Here the subscript 'C' implies that the calculations are computed for each pixel in the context. The ratio between the actual weights of pixel $I_c(x, y, t)$ in the current band and the weights of the pixel $I_c(x, y, t-1)$ in the reference band will give the estimate that was required to vary the weights W_x in the reference band to obtain the weights for the current band.

Since we are trying to estimate the scale factor between the two bands so that the ratios can be modified by this value we compute the difference between the ratios of the

weights to the ratio of the pixel in contexts. The final correction factor used to correct the ratio between the two contexts is then given by.

$$f_c = \frac{1}{n_c} \sum_i \sum_j \left(\frac{W_c(x-i, y-j, t)}{W_c(x-i, y-j, t-1)} - \rho_c(x-i, y-i, t) \right) \qquad i, j \in C \qquad (8)$$

f_c is computed for all the pixels after they have been predicted. f is formed by all the f_c within the context.

$$f(i, j, t) \leftarrow f_c \qquad (9)$$

Finally α is computed as:

$$\alpha(x-i, y-j, t) = \rho(x-i, y-j, t) + \frac{1}{n_c} \sum_i \sum_j \left(f(x-i, y-j, t) \right) \qquad i, j \in C \qquad (10)$$

Here f is the correction factor calculated for each pixel in the context to compensate for the scale factor which affects the estimate of variation ρ between the two bands. Our empirical observation shows that using the average of the correction factor gives better results than adding the individual correction factor to its respective pixel ratio.

The weights for the pixels in the context of $I(x, y, t)$ are calculated as shown.

$$\hat{W}_y(x-i, y-j, t) = \alpha(x-i, y-j, t) \cdot W_x(x-i, y-j, t-1) \qquad i, j \in C \qquad (11)$$

The prediction for pixel $I(x, y, t)$ in the current band is given by

$$\hat{I}(x, y, t) = \frac{1}{n_c} \sum_i \sum_j \left(\hat{W}_y(x-i, y-j, t) \cdot I(x-i, y-j, t) \right) \qquad i, j \in C \qquad (12)$$

The prediction error residual $e(x, y, t)$) for $I(x, y, t)$ is given by

$$e(x, y, t) = I(x, y, t) - \hat{I}(x, y, t) \tag{13}$$

3.2.2 Spatial Prediction

Hyperspectral images are affected by atmospheric absorption and scattering at certain spectral wavelengths. The major contribution to absorption of light energy at these spectral wavelengths is due to carbon dioxide and water present in the atmosphere. The bands affected by atmospheric absorption and scattering have poor spectral correlation. Simple spatial predictors like JPEG's 7 predictors and median predictors perform better than the spectral predictors. These bands have high correlation with the vertical neighbors of the current pixel.

$I(x-1, y-1, t)$	$I(x-1, y, t)$	$I(x-1, y+1, t)$
$I(x, y-1, t)$	$I(x, y, t)$	

Figure 12: Context for spatial prediction.

Pixel $I(x, y, t)$ is best predicted by pixel $I(x-1, y, t)$.

$$\hat{I}(x, y, t) = I(x-1, y, t) \tag{14}$$

$$e(x, y, t) = I(x, y, t) - \hat{I}(x, y, t) \tag{15}$$

The prediction scheme works better than the standard median predictor and JPEG LS for the given bands. In [39] bands 1 to 8 were identified as the bands that should be predicted

spatially. Empirically we found out that bands 1 to 8 and bands 107 to 114 should be predicted spatially and that the spatial prediction scheme gives the best results for these bands. These bands give the best results over all the image sets hence these bands were identified for spatial predictors. The bands are not easily identified hence they need to be predicted by both the spectral and spatial predictors to find the best prediction scheme. In case of AVIRIS images the bands predicted spatially remain same over different scenes in different images and hence can be incorporated in the compression scheme without any need of sending overhead bytes. Fig 13 shows Bands 1, 7, 109, and 112 in Jasper Scene 1 which were predicted spatially.
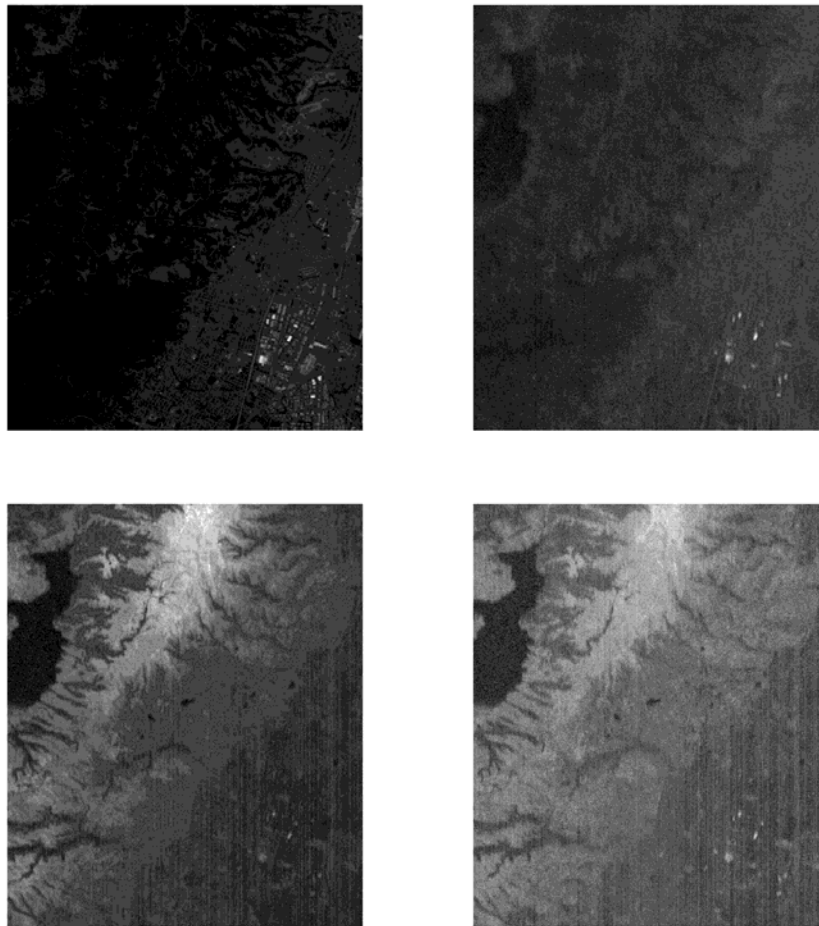
Figure 13: Jasper Scene 1- Top Row Bands (1, 7) Bottom Row Bands (109,112)

3.2.3 Encoding with no Prediction

Few of the bands have very low dynamic pixel range. The pixel value varies from -20 to 20. These bands have high sensor noise as the reflected light in these frequencies is absorbed by the atmosphere and the intensity captured by these sensors is superimposed by the sensor noise. These bands have no correlation with any other band and can be easily identified by there pixel value range. These bands do not have any spatial or spectral correlation and are best encoded without any prediction due to there low pixel values. In AVIRIS images these bands are in the range of 153 to 160. These bands are very noisy as seen in Figure14. The images displayed in Figure 14 are from Jasper Scene1 and the band numbers are 155 and 156.
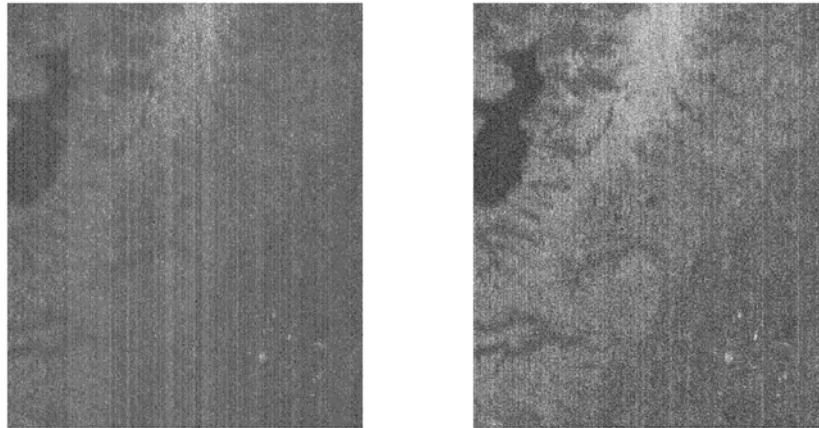
Figure 14: Jasper Scene 1 Band numbers 155 and 156

3.3 Edge Based Prediction - EPHI

Improvements in lossless compression have been possible with the use of edge-based approaches [18, 54, 33] over the DPCM-based techniques. The prediction scheme discussed in section 3.2 does not consider edges while calculating the coefficients for prediction. Pixels along the edges have higher correlation than pixels across the edge. In this section we modify the previous prediction scheme to incorporate the edge-based information. The spatial predictors still give better performance than the edge-based

scheme hence the bands that have been marked for spatial prediction and with encoding without prediction are not considered for edge-based prediction.

The main aspect of edge based techniques is to decide on the threshold value to classify edge and non-edge pixels. Different threshold values work better with different images. Different methods of edge classifications have been studied in the field of lossless compression. Canny edge detectors have performed well in classifying the edges and non-edges in natural images but the computation time required to perform the edge detection is high. Hyperspectral images have high pixel gradients and large texture like areas. Hence most of the detection algorithms identify almost all the pixels as edge pixels. Hence there is a need for an edge detection method which is computationally efficient.

Size of hyperspectral images is very large hence computing the edge gradient calculations for every pixel becomes computationally intensive. It has bee shown in [33] that computation time can be reduced at the expense of little loss in compression ratio when a global threshold is computed. Since the image size of hyperspectral image is very large computing edge thresholds locally become computationally very intensive, hence a global threshold is assigned to every band. In case of hyperspectral images each band has different range of pixel values. The average value of the bands varies from 10 to 5000. With such large variation it is difficult to assign one threshold value to a complete set of 224 bands. Empirically it was found that threshold value of 20% of the mean of each band gives the best results. Different tests were conducted varying the threshold value from 75 to 250 in steps of 25 and 10% to 24% of the mean of each band in steps of 2. The best results were obtained with a threshold value of 20% of mean of the band. The threshold being dependent on the average of each band works better than assigning a fixed threshold over all the bands. To save overhead cost of sending the threshold value for each band the threshold of reference band is used to predict that of current band. This did not affect the results considering that the global structure in all the bands is similar. Also the edge detection is performed on the reference band and the pixels marked as edge pixel in reference band are also marked as edge pixel in the current band.

$$\mu_{t-1} = \frac{1}{m*n} \sum_{y=1}^{m} \sum_{x=1}^{n} I(x,y,t-1) \qquad (16)$$

$$T = \frac{20*\mu_{t-1}}{100} \qquad (17)$$

where, T is the edge threshold.

μ_{t-1} is the average of all the pixel intensities in band $t-1$.

The main aspect of edge based technique is to efficiently pick edge pixel in a context. It might happen that two edges pass in the context and may contribute as edge pixels. But the pixel being predicted may have higher correlation with pixels forming an edge with it than pixels which are part of a different edge.

$I(x-2,y-2,t-1)$	$I(x-2,y-1,t-1)$	$I(x-2,y,t-1)$	$I(x-2,y+1,t-1)$	$I(x-2,y+2,t-1)$
$I(x-1,y-2,t-1)$	$I(x-1,y-1,t-1)$	$I(x-1,y,t-1)$	$I(x-1,y+1,t-1)$	$I(x-1,y+2,t-1)$
$I(x,y-2,t-1)$	$I(x,y-1,t-1)$	$I(x,y,t-1)$		
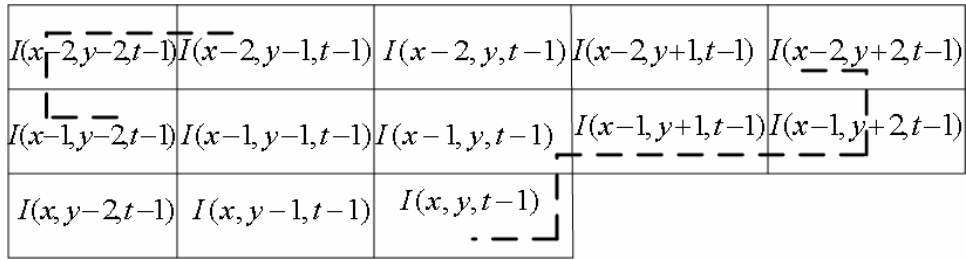
Figure 15: Context for Edge Pixels

Consider Figure 15, here $I(x,y,t-1)$ lies on a edge and the edge is formed by pixels $I(x-1,y+1,t-1)$, $I(x-1,y+2,t-1)$ and $I(x-2,y+2,t-1)$. The edge formed in the context is shown by the dotted lines. In the same context there is a second edge which is formed by pixels $I(x-1,y-2,t-1)$, $I(x-2,y-2,t-1)$ and $I(x-2,y-1,t-1)$. Since the two set of pixels are part of different edge structures they should not be combined to predict $I(x,y,t-1)$.

Consider the case shown in Figure 15 where, $I(x-1,y+1,t-1)$, $I(x-1,y+2,t-1)$ and $I(x-2,y+2,t-1)$ are part of the edge formed going through the pixel being predicted. $I(x-1,y-2,t-1)$, $I(x-2,y-2,t-1)$ and $I(x-2,y-1,t-1)$ forms a second

edge which is not connected to $I(x-1, y-2, t-1)$, $I(x-2, y-2, t-1)$ and $I(x-2, y-1, t-1)$. To avoid the second we start classifying the inner set of pixels as edges or non-edges. Here the inner set of pixels is formed by the four closest pixels to $I(x, y, t-1)$. It can be observed that only $I(x-1, y+1, t-1)$ is part of the edge. Hence all the pixels connected to $I(x-1, y+1, t-1)$ are tested for edge properties. Pixels $I(x-1, y+2, t-1)$ and $I(x-2, y+2, t-1)$ are identified as edge pixels. Since only connected pixels are tested for edge properties we save time as the edge calculations are computed only on the reduced set of pixels.

The calculations for the edge-based is similar to the context based method discussed in the previous section. In the case of edge based method the pixel classified as edges and non-edges for there respective context classification are used for the computation of the prediction error of the current pixel. Considering that the current pixel is classified as an edge pixel and the case studied in Figure 15 is taken. Then the calculation to predict $I(x, y, t)$ is as follows.

$$W_x(x-i, y-j, t-1) = \frac{I(x, y, t-1)}{I(x-i, y-j, t-1)} \qquad i, j \in E_p \qquad (18)$$

The weights in the reference band are computed for the pixels classified as edge pixels. Here E_p stand for the edge pixels in the context.

Since the edge calculations in the reference band holds true for the current band. Co-located pixels classified as edge pixels in the reference band are picked from the current band to compute the ratio between the two contexts

.

$$\rho(x-i, y-j, t) = \frac{I(x-i, y-j, t-1)}{I(x-i, y-j, t)} \qquad i, j \in E_p \qquad (19)$$

The calculations made for the correction factor to correct the ratios are computed only on the pixels classified as edge pixels.

$$W_{ep}(x-i, y-j, t) = \frac{I_{ep}(x, y, t)}{I_{ep}(x-i, y-j, t)} \qquad\qquad i, j \in E_p$$

(20)

$$f_{ep} = \frac{1}{n_{ep}} \sum_i \sum_j \left(\frac{W_{ep}(x-i, y-j, t)}{W_{ep}(x-i, y-j, t-1)} - \rho_{ep}(x-i, y-i, t) \right) \qquad i, j \in E_p \qquad (21)$$

Since the estimation for the correction factor is computed from only the edge pixels better approximation is obtained for the weights to predict $I(x, y, t)$

$$\alpha(x-i, y-j, t) = \rho(x-i, y-j, t) + \frac{1}{n_{ep}} \sum_i \sum_j \left(f(x-i, y-j, t) \right) \qquad i, j \in E_p \qquad (22)$$

The weights computed for the edge pixel $I(x, y, t)$ is given in (22).

$$\hat{W}_y(x-i, y-j, t) = \alpha(x-i, y-j, t) \cdot W_x(x-i, y-j, t-1) \qquad i, j \in E_p \qquad (23)$$

The predicted value for $I(x, y, t)$ is computed as shown.

$$\hat{I}(x, y, t) = \frac{1}{n_{ep}} \sum_i \sum_j \left(\hat{W}_y(x-i, y-j, t) \cdot I(x-i, y-j, t) \right) \qquad i, j \in E_p \qquad (24)$$

Here n_{ep} is number of pixels classified as edge pixels.

$$e(x, y, t) = I(x, y, t) - \hat{I}(x, y, t) \qquad\qquad (25)$$

Since only the edge pixels are involved in the calculations higher gains in compression ratios are achieved. Gains obtained are highest in the case of Moffet field since they have

very distinct edge. It can be observed from the results of M-CALIC that Moffet Field performs better with edge based algorithms than non-edge based algorithms.

3.4 Sign Bit Flipping

The sign bit flipping techniques to remap the prediction error residuals discussed in (26) is difficult to implement in the case of hyperspectral images due to the dynamic range of the pixel values.

$$
\begin{aligned}
&if\ \hat{y}(m,n) <= median \\
&\qquad if\ |e(m,n)| <= \hat{y}(m,n) \\
&\qquad\qquad if\ e(m,n) < 0 \qquad \hat{e}(m,n) = 2|e(m,n)| - 1 \\
&\qquad\qquad Else \qquad\qquad\qquad \hat{e}(m,n) = 2|e(m,n)| \\
&\qquad Else \\
&\qquad\qquad \hat{e}(m,n) = \hat{y}(m,n) + |e(m,n)| \\
&Else \\
&\qquad if\ |e(m,n)| <= \max - \hat{y}(m,n) \\
&\qquad\qquad if\ e(m,n) < 0 \qquad \hat{e}(m,n) = 2|e(m,n)| \\
&\qquad\qquad Else \qquad\qquad\qquad \hat{e}(m,n) = 2|e(m,n)| - 1 \\
&\qquad Else \\
&\qquad\qquad \hat{e}(m,n) = \max - \hat{y}(m,n) + |e(m,n)|
\end{aligned} \tag{26}
$$

Where, $e(m,n)$ is the prediction error being remapped?

$\quad \hat{y}(m,n)$ is the prediction for $y(m,n)$.

$\quad \hat{e}(m,n)$ is the remapped symbol.

$\quad$ Max is he maximum pixel intensity.

$\quad$ Median is the median of all the pixel intensities in a band.

In the case of standard images the minimum ad the maximum value that can be present in the image is know i.e. 0 and 255. But in case of Hyperspectral images the minimum and the maximum values are different in each image. The major problem with hyperspectral

images is that there are few pixels which have very high value compared to other pixels in the band. Due to this the last condition in the equation is never satisfied as the difference between the maximum pixel intensity and the median of the image pixel intensity is never less than the error value. Hence no gain in entropy is obtained after remapping. Hence simple sign bit flipping is used to remap the prediction error residue as shown in (27).

$$\hat{e}(m,n) = \begin{cases} 2*|e(m,n)| & if \ \ e(m,n) \leq 0 \\ 2*|e(m,n)| - 1 & if \ \ e(m,n) > 0 \end{cases} \tag{27}$$

Chapter 4: Band Ordering

4.1 Introduction

Lossless compression of hyperspectral imaging can be broken down into two parts; ordering of bands, and then prediction based on the ordering. In [48, 49] significant gains in compression ratio have been shown for the lossless compression of hyperspectral images after ordering. Ordering technique proposed in [48] has an order of complexity of $O(n^2)$ where n is the number of bands. This does not include the complexity of the compression algorithm used to compute the compression ratios between the bands. Due to high complexity of ordering it is not feasible to implement the ordering schemes onboard on satellites capturing hyperspectral images. There is a need to study different ordering scheme in order to reduce the overall complexity of ordering and obtain the compression ratios compared to optimal ordering. In this thesis we propose three different schemes of ordering.

4.2 Band Ordering Problem

Band ordering problem can be defined as, given a set of bands $B = \{b_1, b_2, ..., b_n\}$ in a hyperspectral image, the band ordering problem is to find a permutation $\Pi(B)$, such that predicting band b_i using $\Pi(b_i)$ for each band $b_i \in B$ will give the maximal compression. The permutation Π constitutes an ordering of the band in B.

The band ordering problem for multispectral/hyperspectral images was studied by Tate in [48]. He computed the compression ratios between all possible combinations of bands to obtain a weighted directed graph. The solution of a weighted directed graph to obtain a maximum weight spanning forest for a dense graph has an order of complexity of $O(n^2)$ where n is the number of bands. The computation time required to order the bands is also

dependent on the lossless compression algorithm used to compute the compression ratio between the bands. A faster algorithm was proposed in [49], instead of computing the compression ratios between bands, correlation between bands were computed to order the bands. To reduce the time required to compute the correlation coefficient between bands the images were down-sampled. The effect of this on the final performance of the system was very little. To further improve the performance of the ordering scheme in terms of time the correlation factor was computed between 42 neighboring bands.

In this thesis we have proposed 3 different ordering algorithms. The first algorithm computes the global and local features of the band to cluster the bands having similar features together. After the bands have been grouped, bands in each group are then ordered by computing the compression ratios between the bands and finding the spanning tree using Kruskal's algorithm. The second and third methods of ordering for hyperspectral images require the computation of compression ratios between the 10 neighboring bands. The weighted graph formed is the solved to obtain the spanning tree using Prim's algorithm and Kruskal's algorithm.

4.3 Minimum Spanning Tree

Given a connected, undirected graph, a spanning tree of that graph is a sub-graph which is a tree and connects all the vertices together [29]. A single graph can have many different spanning trees. Weights are assigned to each connection in the graph. A minimum spanning tree is formed by the spanning tree having the minimum weight compared to all other trees. In this thesis we form a maximum weight spanning tree as higher the compression ratios better is the performance of the system. Two MST algorithms have been discussed in this thesis namely Prim's algorithm and Kruskal's algorithm. Both the algorithms are greedy algorithms and run in polynomial time.

In Kruskal's algorithm the edges are sorted based on their weights. The first edge has the least weight while the last weight has the maximum weight. At each step the edge is connected to the tree if it does not form a cycle. Since independent trees are formed

Kruskal's algorithm forms a forest. Kruskal's algorithm forms a forest but the final tree might connect to form a single tree instead of a forest. Kruskal's algorithm has a worst case complexity of $O(E \log V)$ where E is number of edges and V is number of vertices.

Prim's algorithm the tree starts from an arbitrary root vertex r and grows until the tree spans all the vertices in V. At each step, a light edge connecting a vertex in E to a vertex V-E is added to the tree [7, 37, 36]. Prim's algorithm forms a single tree with a single root vertex. The run time for Prim's algorithm is $O(E \log V)$. Two different algorithms Prim's and Kruskal's form different spanning trees when tested on the same data, hence the two algorithms have been used to test which of them gives the best results.

4.4 Ordering based on Global and Local Features

We divide the problem of ordering into three parts: global grouping wherein global features of the bands are used to group the bands, local grouping wherein local features in each band is used to sub group the global groups; and ordering whereby the bands in the subgroups are ordered based on techniques similar to [48]. The major problem with ordering methods is the time required to compute the weights of the edges in a weighted graph. In [48] two compression methods were used to compute the weights for each edge in the graph. Though the complexity of the ordering scheme is $O(n^2)$ which implies that the run time to find the maximum weight spanning tree is $O(n^2)$. But the time required to compute the weights is very high. The total time required will be dependent on the computation time of the lossless compression algorithm.

In this thesis we propose an ordering scheme which groups the bands based on their global and local features. Bands in each group are then ordered using Kruskal's algorithm. The compression ratio between the bands is computed within each group hence the total computation time required to compute the weights is reduced as the total

combination of bands is reduced from n^2 to $k \times \left(\dfrac{n}{k}\right)^2$ where n is the total number of bands and k is the number of groups. Here the assumption is made that the bands are equally divided between the groups.

Figure 16 shows the ordering of bands. Here 4 groups are formed and in each group the bands are ordered. The root band in each group is predicted by the closest band in the n-dimensional space.
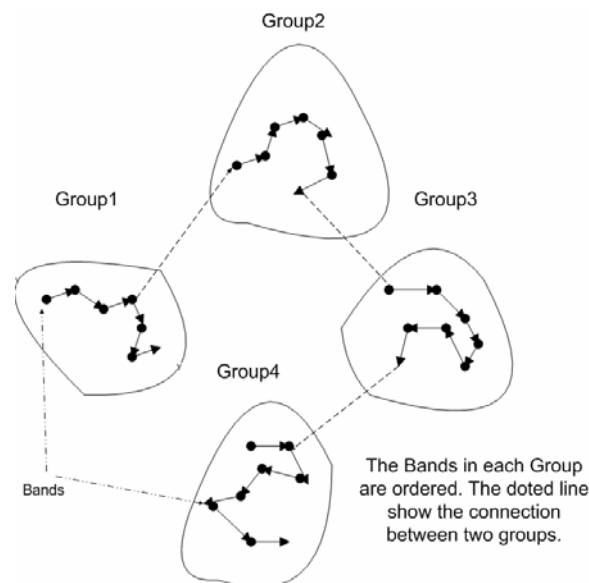
Figure 16: Ordering of Bands

4.4.1 Global Grouping

The global structure captured by the sensors in hyperspectral images is similar across all the bands. But the intensity level of the pixels varies from band to band. The average computed for all the pixels in a band vary from 10 to 5000 across the bands in the hyperspectral images. Hence the gradient of pixel with its neighbor in a band having an average of 10 will be different from the gradient of pixel with its neighbor in a band having an average of 5000. Though the band with average pixel value of 10 can be

scaled/normalized to 5000 but this will also rescale the gradients between the pixels by that scale factor, which will be higher than the gradients in band with an average of 5000. Hence to obtain higher compression gains it is advisable to predict a band with a band having similar intensity levels. Hence the motivation of this work comes from the fact that higher prediction gains can be obtained if bands with similar intensity levels predict each other.

Hyperspectral images are affected by sensor noise and absorption and scattering properties of the atmosphere [22]. The bands affected by noise have either lower intensity level or high variance. Hence these bands should be grouped together based on there average intensity level and the variance. It has been observed that images corrupted by noise have higher variance than normal images.

From the above two considerations it can be observed that the bands should be grouped based on there average intensity level and the standard deviation of the pixel intensities to achieve better prediction and higher compression ratios. Two parameters were computed for each band, average band intensity and standard deviation. This 2 dimensional data was then given to k-means to group the bands. Here any clustering algorithm can be used to group the bands. The number of clusters or groups was fixed to 7 which gave the best performance.

4.4.2 Local Grouping

Global Grouping will group images with similar intensity level and variation together. But this may not result in higher compression gains as intensity levels in different part of the image in the same group may not be similar enough to obtain good spatial and spectral predictions. Hence there is a need to further divide the images into smaller groups based on the features extracted from the bands. The time required to achieve the final grouping is reduced as now only the bands in the group need to be compared and not all the n bands. Global grouping is used to reduce the overall time required to sub-group. KL-distances is computed between the bands to obtain the final sub-group. To

reduce the overall time required to obtain the final grouping the number of sub-groups were calculated adaptively and instead of computing KL-distance between all the bands in the group, reference bands were generated which were used to compute the KL-distances. The distances were computed between the bands and the reference bands. The matrix obtained after computing all the distances is then given to a clustering algorithm. In this thesis k-means is used to group the bands.

To improve the performance of the system in terms of time the number of sub-groups that should be formed is pre-calculated. In this thesis we have proposed a technique which can be used to decide on the number of subgroups to be assigned to each global group. Ideally the sub-group should be assigned based on the number of bands in a group. But there is a possibility that a group may have higher number of bands and also the cluster density is very high. In this case it is not advisable to assign more number of sub-groups. Hence the assignment of the sub-group should be dependent on the cluster density of a group and the number of bands in each group. Hence poor cluster density but fewer numbers of bands will have lesser number of groups while a group with higher cluster density and more number of bands will also have lesser number of subgroups. The calculation for subgroups helps us in using the optimized k-means instead of the adaptive k-means.

The cluster density of the groups is calculated as follows.

$$C_d(i) = \frac{\sum_{j=1}^{h}\left(\left(\mu(i,j)-C_\mu(i)\right)^2+\left(\sigma(i,j)-C_\sigma(i)\right)^2\right)^{1/2}}{h} \tag{3}$$

Where, i stands for the i_{th} group and h is the number of bands in the group, $C_d(i)$ is the cluster density for the i_{th} group. In a two dimensional space the cluster center or the centroid will be represented by two points. $C_\mu(i)$ and $C_\sigma(i)$ are the cluster centers for the i_{th} group. In order to compare the cluster densities it is necessary to normalize them with

respect to the number of bands in the group. The normalized cluster density for the i_{th} band is as shown

$$\overline{C}_d(i) = \frac{C_d(i) * N_b(i)}{\sum_k \left(C_d(k) * N_b(k) \right)} \tag{4}$$

Here $\overline{C}_d(i)$ is he normalized cluster density and $N_b(i)$ is the normalized group size for the i_{th} group. k is the number of groups formed after the global grouping. To compute the number of sub-groups that can be assigned to each group it is necessary to fix the upper limit to the number of groups that can be formed for a hyperspectral data. Here z is the maximum number of groups that can be assigned to a hyper-spectral data.

$$G_s(i) = \frac{(z-k) * \overline{C}_d(i)}{\sum_k \overline{C}_d(k)} \tag{5}$$

Here $G_s(i)$ is the number of sub-groups assigned to each group. Ideally two images can be compared by calculating there MSE's or PMSE's but they are computationally time consuming. Lesser computational method to compare the images is to compare there Histograms. KL distance is a good tool to compare two distributions.

$$KL(p,q) = \frac{1}{2} \left(\sum_i p_i \log_2 \left(p_i \big/ q_i \right) + \sum_i q_i \log_2 \left(q_i \big/ p_i \right) \right) \tag{6}$$

Where p and q are the two histograms to be compared and i represents the bins in the histogram. For optimal grouping we should calculate KL distance between all the histograms in a group. But the computation of KL distances between all the bands will increase the complexity of the algorithm and will require more run time due to log calculation. To reduce the time complexity of the algorithm reference bands are picked from the group of bands. The number of reference band in a group is dependent on the

number of bands present. Hence for the largest group 3 reference bands are picked and for the smallest 1 reference band is picked. The reference bands are picked by sorting the bands in ascending order. Here the bands are placed in 2 dimensional space where the two dimensions are mean and standard deviation. The bands are sorted in ascending order. The first reference band is picked close to the median of the sorted order. The second reference band is picked by flipping the first reference band. The band closest to the flipped dimension is picked as the second reference band. The third reference band is picked close to the centroid of the group. The reason to obtain reference bands in this order is we need a good representation of the group. The first two reference bands are placed between the centroid and the first and last band. The third band is located close to the centroid. In case only one band is chosen as the reference band then the centroid is picked. While, if two reference bands are chosen then the first two reference bands are taken as the representative of the group. After computing the reference band, KL distances are computed between the reference bands and all the bands in the group. The distances are then grouped using the k-means algorithm. The numbers of clusters or sub-groups were pre-calculated as shown in equation (2, 3, and 5) for each group.

After the final grouping is obtained the bands are ordered by computing the compression ratios between all the bands in a sub-group. Here since the number of bands in a sub-group is less than the total number of bands in an image set the computation time required is much less than the algorithms proposed in [48, 49]. In optimal ordering [48] with n number of bands, n^2 compression ratios need to be computed. Assuming that there are P_i number of bands in the i_{th} group total number of compression ratios that need to be computed is given by $\sum_{i=1}^{m} P_i \left(P_i \right)^2$. Here m is the total number of groups. Since in each group the number of bands is fraction of the total number of bands the computation time required to obtain the compression ratios is reduced. After obtaining the compression ratios the weighted graph formed was solved using Kruskal's algorithm. Since the Kruskal's algorithm forms a forest there was a need to connect all the roots to form a single tree. To obtain this depth first search was used to connect the roots to the main tree.

The performance of the ordering algorithm was not as expected. High losses were incurred in predicting the root bands of each group. The loss incurred in the prediction of the root bands was higher than the gain obtained due to ordering. These bands showed poor performance with spatial prediction and prediction by bands from different groups. Here we tried to study the problem of ordering by computing the global and local features in the bands. The major aim was to reduce the time required to compute the compression ratio between all the bands by reducing the number of bands from n to fraction of it in each group.

4.5 Neighborhood-based Ordering

In hyperspectral images the correlation factor reduces with the increase in the distance between the current band and the prediction band. The absorption and reflection property of material remains similar in the neighboring bands. Hence the band under consideration is best predicted by the neighboring bands. The ordering method proposed in this thesis computes the compression ratio of the current band with limited number of the neighboring bands.
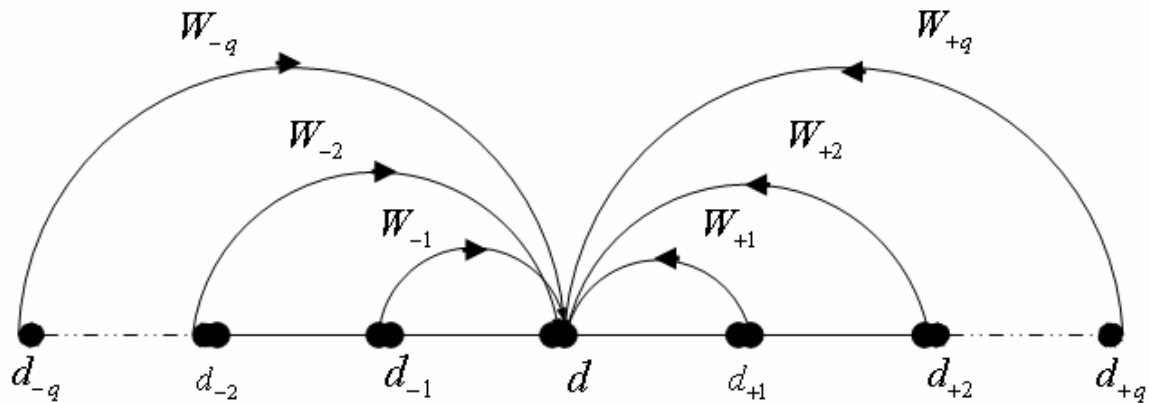

Figure 17 Ordering based on Neighbors

In Figure 17 band d is predicted by bands from d_{-q} to d_{+q} and the compression ratios computed between them is given by W_{-q} and W_{+q}. Here the limit to the number of

neighboring bands is set to q. Since only $2 \times q$ neighboring bands are used to predict the current band d. Hence overall $2 \times q \times n$ compression ratios need to be computed to order the bands as compared to n^2 in the case of optimal ordering proposed by [49] and $2 \times n^2$ in the method proposed by [48].

The ordering methods proposed have been inspired by Prim's algorithm and Kruskal's algorithm. The proposed algorithms are modified Kruskal's and Prim's algorithm. The main difference between the proposed method and the two standard algorithms is that they require a symmetric matrix wherein the connection between two nodes has the same weights in either direction. This means that the compression ratio computed from band b_1 to band b_2 is same as computed from band b_2 to band b_1. In real world case the compression ratios will never be the same. Hence the two algorithms were modified to incorporate this fact.

The first method of ordering is based on the Kruskal's greedy algorithm to find the minimum spanning tree from a weighted graph. In Kruskal's algorithm the weights for the edges are sorted in ascending order and the least edge weight forms the root node. Kruskal's algorithm computes a forest since many trees are built simultaneously till all the edges have been connected. The final result might have a single tree if all the edges are connected to each other or it will form a forest having many spanning trees in it. Similar to the Kruskal's algorithm we order the compression ratios computed between the bands. Since the compression ratios vary from small value to large it is not advisable to order the compression ratios, since it may not result in the best performance achievable. The compression ratios do not provide the gain made if a band is predicted by another band. Here the best compression ratio obtained will form the first edge but the improvement obtained in predicting the band may not be significant compared to other bands. Hence there is a need to order the gain made by the prediction of one band with another, than ordering the compression ratios.

Since the method is being compared with the results obtained by unordered bands we select the difference between the compression ratios achieved by predicting the band with its q neighbors with the compression ratio achieved for that band when the bands are unordered. In an unordered method the previous band is used to predict the current band. Now the weights of the edges is given by the gain made in compressing one band from another and the first edge will be for the bands having highest gain. Similar to Kruskal's the edges are ordered in ascending order depending on their weights. Greater improvement in performance is obtained since edges having higher gain have higher priority compared to other bands.

For each edge there is a pair of predicted and predictor band P_d^i and P_r^i respectively. The pair forms a union under 2 conditions.

- The band has not already been predicted.
- The two bands should not form a cycle.

The first condition can be checked by storing all the bands that have been predicted and there predictors. This data is also needed by the decoder to decode the bands correct order. A cycle is formed when two nodes of the same tree are connected to each other. Hence the band can be reached from two or more different paths. To avoid cycles we need to check if the two bands the predicted and predictor are not subset of the same tree. As independent trees are grown it is difficult to keep account for all the tree formation and there nodes. A look up table is kept which keeps the information of the bands being predicted and there predictor. So each time a union of two bands is formed the look up table is updated. Each time two new bands come in the look up table is parsed to find the bands that have already been predicted by P_d^i. Similarly for all the bands predicted by P_d^i we find the bands predicted by them and the process continues till all the bands connected to P_d^i have been found. P_r^i is compared with all the connected bands found from the look up table and if a match is found then the pair is discarded.

The ordering method discussed forms a forest which may result in many root nodes which have not been predicted. It is difficult to compress these bands spatially due to poor spatial correlation. Also the distance between the root nodes will decide the achievable compression ratio if one root node is predicted by another. Hence there is loss in performance due to poor performance of the root bands in terms of compression ratios. To avoid formation of multiple trees the next ordering method forms a single tree using Prim's algorithm. Similar to Prim's algorithm the first band forms the root node of the tree. The best prediction by the current band forms the next edge of the tree.

The complexity of both the above algorithms is $O(E \log V)$ where E forms the edges of the weighted graph and V forms the vertices of the weighted graph. Since only q neighboring bands are selected to predict the current band the overall computation time is reduced by $n/2q$ where n is the total number of bands and q is the number of neighboring bands used for prediction.

Chapter 5 Results

The main objective of this thesis is to develop a method that could achieve comparable or better compression ratios when compared with state of the art algorithms for lossless compression of hyperspectral images. Two different aspects of lossless compression of hyperspectral images have been studied in this thesis, namely the prediction scheme and the ordering method. Two different prediction methods have been proposed in this thesis. The first scheme performs a spatio-spectral prediction of the pixels in a given band, using information in causal spatio-spectral neighborhood. The second scheme classifies the pixels into edges and non-edge pixels, and then performs prediction based on this classification. Three different ordering schemes have been studied in this thesis; the first one orders the bands based on their global and local features, while the other two schemes order the bands based on the information computed from their b-neighboring bands.

5.1 Experimental Setup and Data Set

The experiments were performed on a machine with 1.53 GHz, AMD Athlon processor and 1 GB RAM using Matlab 7.

The JPL website has 5 different image sets captured by AVIRIS in 1997 for download [2]. Research on AVIRIS data has been carried out since last two decades. Prior to 1997 data acquired in 1989 by the AVIRIS instrument were available for research purposes. The 5 image set acquired by the AVIRIS instrument which have been made available for research and commercial use are Cuprite, Moffet Field, Jasper Ridge, Lunar Lake and Low Altitude. Each image is of different size and comprises of scenes of size 614×512 wherein the last scene has lesser number of rows in each image set. The size and number of scenes of each image set is given in Table IV

TABLE IV: DETAILS OF AVIRIS IMAGE DATA SET.

Image Set	Number of Scenes	Size of each scene	Size of the Last Scene	Total Image Size in Pixels	Total Image Size in MB's
Cuprite	5	614×512	614×158	614×2206	606.8
Moffet Field	4	614×512	614×495	614×2031	558.6
Lunar Lake	3	614×512	614×407	614×1431	393.6
Jasper Ridge	6	614×512	614×26	614×2586	711.3
Low Altitude	8	614×512	614×105	614×3689	1014.7

The parameters used in the prediction method were

Context Size – 12

Threshold Value – 20 % of the mean of each band for (EPHI)

Bands predicted spatially- 1 to 8 and 107 to 114

Bands encoded without prediction- 153-160

All other bands were spectrally predicted.

Parameters for Ordering

1. Feature based ordering

   Cluster Size for Global Grouping –7

   Maximum groups allowed 20

   Maximum Reference Bands used for local grouping-3

2. Neighborhood based ordering

   Number of Neighbors–10

The encoding scheme used in this thesis is the Matlab's arithmetic encoder. The arithmetic encoder in Matlab requires the probability of occurrence of each symbol. This information is required at the decoder. Hence the probability of each symbol is added to the header of the compressed file and is included at the time of computing the compression ratios.

5.2 Spatial Prediction Schemes

Before we study the performance of prediction schemes proposed in this thesis, it may be helpful to see the performance of standard lossless image compression algorithms on hyperspectral images. The performance of spatial predictors on hyperspectral images have already been shown in [39, 38, 55, 34, 20]. Spatial predictors do not perform well (in terms of compression ratios) on hyperspectral images. In [39] it was shown that differential JPEG LS and differential JPEG 2000 perform better in terms of compression ratios. In the differential schemes, the differences between the bands were compressed instead of the bands being compressed individually. Significant improvement in the performance was obtained when the difference between bands were compressed, hence the motivation for spectral predictors. TABLE V shows the performance for different standard lossless image compression algorithms in terms of compression ratio on hyperspectral images.

TABLE V: PERFORMANCE OF STANDARD LOSSLESS IMAGE COMPRESSION ALGORITHMS

	CALIC [20]		JPEG LS [38]		JPEG 2000 [38]		Differential JPEG LS [39]		Differential JPEG 2000 [39]	
	bpp	CR	bpp	CR	bpp	CR	bpp	CR	bpp	CR
Cuprite	6.99	2.29	7.66	2.09	8.38	1.91	5.50	2.91	5.48	2.92
Jasper Ridge	7.69	2.08	8	2.00	8.89	1.80	8.84	2.81	1.8	2.82
Moffet Field	6.53	2.45	8.38	1.91	8.99	1.78	5.63	2.84	5.65	2.83
Lunar Lake	6.48	2.47	8.04	1.99	8.79	1.82	5.46	2.93	5.44	2.94
Low Altitude	-	-	7.47	2.14	8.16	1.96	5.93	2.70	5.95	2.69
Average	6.90	2.32	7.62	2.03	8.65	1.85	5.63	2.84	5.63	2.84

Earlier images from 1989 flight were used for research in lossless compression of hyperspectral images. These images were replaced by images from the flight run of 1997. From TABLE V it can be observed that standard lossless image compression algorithms perform very poorly on hyperspectral images. The results have been quoted from the relevant sources as indicated in TABLE V. But mentioned said earlier in Chapter 4 on prediction, some of the bands are best predicted by spatial predictors. These bands have

poor spectral correlation and perform poorly with spectral predictors but give better compression ratios with simple median predictors or JPEG's 7 predictors.

5.3 Spatio-Spectral Predictor

Spatial predictors perform poorly in terms of compression ratios as compared to the spectral predictors as discussed in chapter on background and prediction. Comparing Table V and table VI it can be observed that significant improvements in compression ration can be obtained in lossless compression of hyperspectral images. AVIRIS hyperspectral images have 5 different sets of images and each image set has different number of scenes. Each image set has scenes of size 512×614 while the size of the last scene varies in all the image sets. The results shown in this thesis have been for the combined results of all the scenes for each image set.

Context based prediction has been classified into edge-based prediction (EPHI) and non-edge based prediction (NPHI). Slight improvement is obtained when the edge-based classification is used for prediction. A context size of 12 causal pixels was fixed after testing partial image set for different context size. The Figure 16 shows the effect of the change of context size for different image sets. The figures are based on image sizes of 256×256 from scene-1 of the respective image sets, and thus do not represent actual entropy for each band. The graphs basically show the change in the entropies with changing context size. It can be observed that after context size 12 the change in entropy is almost constant. Hence the context size of 12 was fixed considering the gain in compression ratio and computation time required.

In Figure 18(a) the effect of change in context size on entropy of prediction for the five image sets of AVIRIS is shown. It can be observed that the change in entropy after context size of 12 is not significant. In Figure 18(b) it can be observed that the change in entropy is of 0.02 from context size of 12 to 23. Hence with almost twice the number of computations required to compute the prediction error residual the change is not significant.
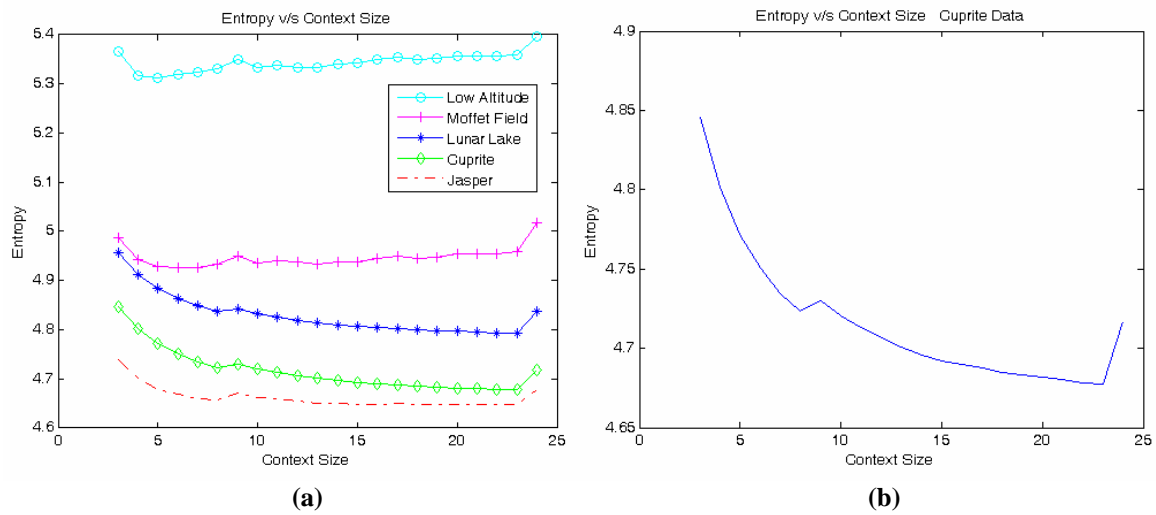
Figure 18: Effect of Context size on prediction error residual; (a) Entropies plotted for all the bands; (b) Plot for Cuprite

As discussed in Chapter 3 on prediction, some bands have poor spectral correlation and are best predicted by spatial predictors. The spatial predictor used in this case is the North Pixel (NP) Predictor, wherein the pixel north to the current pixel is used to predict the current pixel. Also some of the bands have low self entropy and are best encode without prediction. In Figure 19 the entropies obtained for the three prediction scheme on Jasper Scene 3 and Moffet Scene2 have been shown. It can be seen that bands 1 to 8 and 107 to 114 are best predicted by spatial predictor and bands 153 to 160 are best encoded without prediction. From the Figure 19 it can be observed that the self entropy of the images for bands 153 to 160 is lower than that obtained by the two prediction schemes. The two figures have been plotted for image set Jasper Ridge scene-3 and image set Moffet Field scene-2.
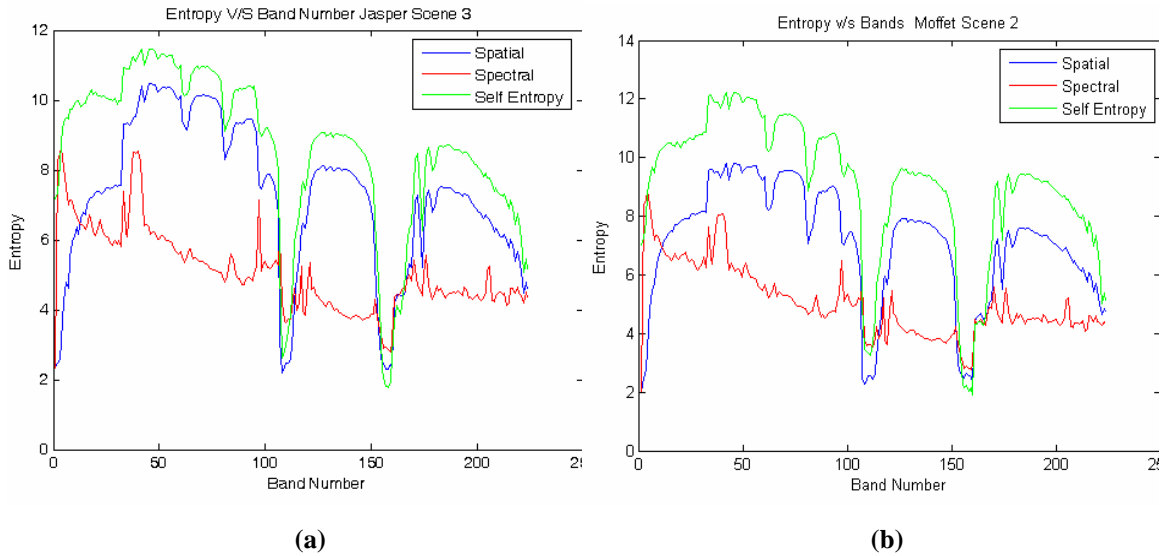
(a)   (b)

Figure 19 Entropy plots for different prediction schemes; (a) Jasper Scene 3; (b) Moffet Scene 2.

In Figure 20(a) the effect of threshold value on the prediction error residual is show. It can be observed that the curves become almost constant after .02 (20 % of the mean of the bands). Though for Lunar Lake and Cuprite the entropy does not vary much with threshold value, it changes significantly for Moffet Field and Jasper Ridge. In Figure 20(b) the variation of entropy against threshold value is shown for Jasper Ridge.
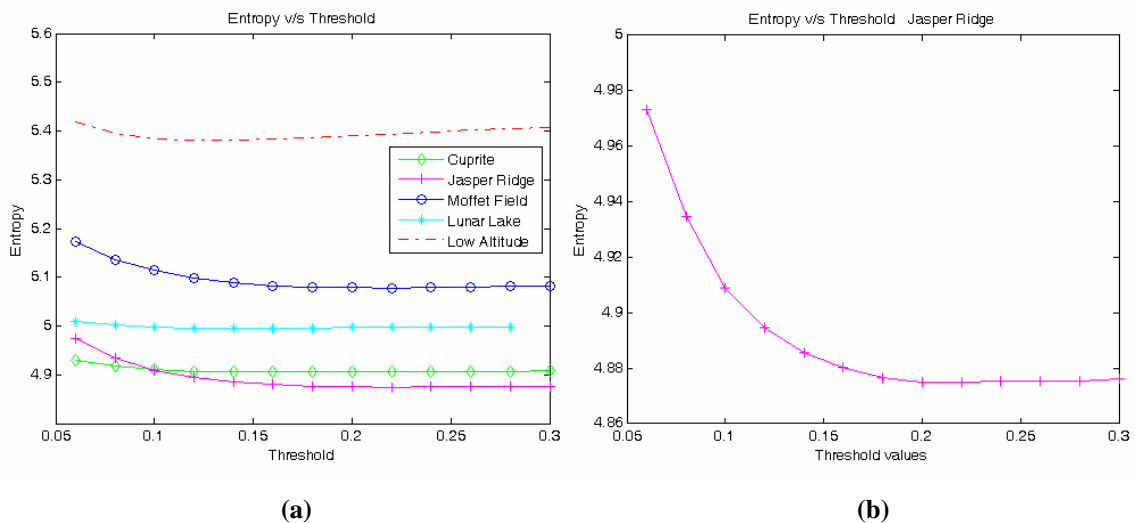

(a)   (b)

Figure 20 Variation of entropy of prediction residual with edge threshold; (a) for all the images; (b) Magnified plot for Jasper Ridge

Table VI shows the results obtained with the two proposed methods for prediction namely Non-Linear Prediction for Hyperspectral Images (NPHI) and Edge-Based Prediction for Hyperspectral Images (EPHI). It can be observed from Table VI that EPHI performs better than NPHI for all the image sets. Higher gains in terms of compression are obtained in EPHI over NPHI on Moffet Field and Low Altitude.

TABLE VI: RESULTS FOR THE PROPOSED PREDICTION METHODS

	NPHI		EPHI	
	Entropy in bpp	CR	Entropy in bpp	CR
Cuprite	4.7904	3.34	4.7619	3.36
Jasper Ridge	4.8930	3.27	4.8632	3.29
Moffet Field	4.9689	3.22	4.8780	3.28
Lunar Lake	4.7904	3.34	4.7619	3.36
Low Altitude	5.3156	3.01	5.1948	3.08
Average	**4.9383**	**3.24**	**4.8930**	**3.27**

Table VII shows comparative results with other state-of-the-art algorithms with comparable complexity. The results quoted in Table VII have been taken from there respective sources as indicated in the table. CCAP proposed in [50] performs better for Cuprite while LUT proposed in [27] performs better on Lunar Lake. From the Table VII it can be observed that on an average the proposed method in the thesis has best performance for edge based method. Average 1 mentioned in the table is for average over the first 4 image sets while the Average 2 mentioned includes results for Low Altitude. The results for [28] have been obtained after optimally ordering the bands.

TABLE VII: COMPARISON WITH OTHER PUBLISHED METHODS

	M-CALIC	SLSQ-OPT	[28] (Opt)	CCAP	3D CALIC	LUT	NPHI	EPHI
Cuprite	3.21	3.24	3.23	**3.53**	3.06	3.44	3.34	3.36
Jasper Ridge	3.17	3.23	3.18	2.56	3.08	3.23	3.27	**3.29**
Moffet Field	**3.38**	3.21	3.13	2.57	3.25	3.17	3.22	3.28
Lunar Lake	3.28	3.23	3.21	3.21	3.09	**3.40**	3.34	3.36
Low Altitude	-	3.04	3.03	2.81	-	-	3.01	**3.08**
Average 1	3.26	3.23	3.19	2.97	3.12	3.31	3.29	**3.32**
Average 2	-	3.19	3.16	2.97	-	-	3.24	**3.27**

5.4 Band Ordering

Higher compression ratios can be obtained if the bands are ordered before prediction. A popular ordering method for lossless compression of multispectral/hyperspectral images was proposed by Tate in [48]. The complexity of the algorithm as published in [48] is $O(n^2)$ where n is number of bands which in the case of AVIRIS images is 224. The algorithm is computationally intensive as it needs to calculate the compression ratios between all the bands. Hence the complexity of the algorithm will be dependent on the type of prediction method used and the compression methodology. In [49] a correlation based method was proposed where the correlation between all the bands is computed. Though the order of complexity remained the same the time required to compute the weights for the edges in the weighted graph reduced. Further in [49] a lesser complex method was proposed wherein the correlation was computed between the neighboring 40 bands. Though this reduced the compression ratios achieved.

Table VIII shows the results for the three ordering methods described in chapter 4 on ordering. We have included same results from Table V for comparison. The results for unsorted ordering perform the best between the three ordering methods proposed and have the second best results overall. In Table 5 the Ordering I method refers to the

ordering based on global and local features while ordering II refers to ordering based on Prim's algorithm and ordering III is based on Kruskal's algorithm. The results are based on EPHI edge based prediction.

TABLE VIII: COMPARISON OF COMPRESSION RATIOS FOR ORDERING METHOD

	M-CALIC	SLSQ-OPT	[28]	LUT	Ordering I	Ordering II	Ordering III
Cuprite	3.21	3.24	3.23	3.44	3.24	**3.38**	3.37
Jasper Ridge	3.17	3.23	3.18	3.23	3.27	**3.29**	3.27
Moffet Field	**3.38**	3.21	3.13	3.17	3.15	3.28	3.27
Lunar Lake	3.28	3.23	3.21	**3.40**	3.23	3.36	3.36
Average	3.26	3.23	3.19	3.31	3.22	**3.33**	3.32

Table IX shows the result when the independent bands are not included in the results. It can be seen that the ordering improves the results significantly. Here the root bands of each tree have not been included in the results. Significant gain in compression ratio is achieved as these bands perform poorly in terms of compression ratio when predicted spatially. Table IX therefore shows the improvements due to ordering and explain why the results shown in Table VIII show little improvements. An effective scheme to compress the root nodes is therefore required for overall improvement in compression performance.

TABLE IX: COMPRESSION RATIO FOR ORDERING WITH INDEPENDENT BAND EXCLUDED

	Ordering II	Ordering II w/o the root bands
Cuprite	3.38	3.41
Jasper Ridge	3.29	3.34
Moffet Field	3.28	3.33
Lunar Lake	3.36	3.40
Average	3.33	3.37

Chapter 6: Conclusion

Research in the field of lossless compression of hyperspectral images has been motivated by the need to conserve bandwidth and storage pace. Earlier the hyperspectral data was restricted to the research community. With the advent of commercial satellites huge amount of data is being captured. The basic area of study for hyperspectral images is mineral exploration and environmental monitoring. Same applications dependent on hyperspectral images need the data to be stored such that it can be recovered without any corruption. Hence the need arose for lossless compression of hyperspectral images. Since the compression algorithm needs to be implemented on-board on satellites the algorithm should have low complexity and should run in real time.

The prediction algorithm proposed in thesis method is based on one-band look-ahead hence does not require huge amount of memory space. A context based non-linear predictor has been proposed in this thesis which predicts the pixel in the current band based on the information of 12 causal pixels in the current band and 12 causal co-located pixel in the reference band. The context based non-linear predictor is modified to add the edge-based analysis which improves the results. The proposed prediction method has the best results compared to the state-of-art-methods having comparable complexity.

A major contribution to this thesis has also been the identification of bands for spatial prediction. Significant improvement is obtained when these bands are spatially predicted since these bands are affected by atmospheric absorption. Since we know the range of wavelengths in which atmospheric absorption we can generalize the algorithm for most of the datasets. Also significant improvement is obtained when bands dominated by sensor noise are encoded without prediction. These bands carry very little information and have poor spectral and spatial correlation. Since they have low pixel intensities they are best encoded without prediction.

Though the proposed ordering techniques do not provide significant compression, they open a new avenue for ordering of hyperspectral images. The ordering based on

neighboring bands has better results than the proposed prediction scheme but there is a need to study it in more detail to achieve significantly higher compression ratios.

In this thesis a context based spatio-spectral predictor was proposed. Though edge-based method has been proposed but the pixel classification was based on single threshold value. A better classification is needed to exploit the spectral information. Also bias error cancellation similar to those proposed in CALIC needs to be studied for the proposed method which might improve the results. The symbol remapping algorithm proposed in CALIC does not provide any gain in entropy due to the nature of the hyperspectral data. A detailed study needs to be carried out by comparing the difference between the two bands and the error residuals obtained. The change in pixel gradients is very high compared to the error values obtained hence the sign remapping technique does not perform well on the hyperspectral data.

A major avenue for improvement in the proposed technique is to model the error residuals for encoding. 16 bit predictor used either gave poorer results than the entropy of the prediction error residuals or gave no improvements in it. Hence there is a need to study more appropriate encoding techniques for the hyperspectral images.

References

[**1**] Ahmed, N., Natarajan, T., and Rao, K. R. "Discrete Cosine Transform," *IEEE Transaction on Computers*, vol. C-23, pp.90-93, 1974

[**2**] AVIRIS Data http://aviris.jpl.nasa.gov/html/aviris.freedata.html

[**3**] Baizert P., Pickering M., and Ryan M., "Compression of hyperspectral data by spatial/spectral discrete cosine transform," *Proceedings IEEE International Geoscience and Remote Sensing Symposium*, 4, pp. 1859–1861, 2001.

[**4**] Berman K. A., and Paul J. L., "*Algorithm: Sequential, Parallel and Distributed*" Course Technology, 2004.

[**5**] Bilgin A., Zweig G., and Marcellin M., "Three-dimensional image compression with wavelet transforms," *Applied Optics*, 39(11), pp. 1799–1814, 2000.

[**6**] Chung Y. S., and Kanefsky M, "On 2-D recursive LMS algorithm using ARMA prediction for ADPCM encoding of images," *IEEE Transaction on Image Processing*, 1, pp. 416–442, 1992.

[**7**] Cormen T.H, Leiserson C. E., Rivest R L, Stein C., "*Introduction to Algorithms*" The MIT Press; 2 ed, 2001.

[**8**] Cover T.M., and Thomas J.A.*, "Elements of Information Theory*", Wiley Series in Telecommunications, John Wiley and Sons Inc., 1991.

[**9**] Deng G., "A study of a lossless image compression algorithm using adaptive prediction and context-based entropy coding," *The IX European Signal Processing Conference*, pp. 1529-1532, Island of Rhodes, Greece, Sept.

[**10**] Green R., Eastwood M. L., & Williams O., "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)". *Remote Sensing of the Environment*, 65, 227-248., (1998).

[**11**] Gersho A., and Gray R. M., "*Vector Quantization and Signal Compression*," Kluwer Academic Publishers, 1991

[**12**] Gonzalez R., and Woods R., "Digital Image Processing", Prentice Hall, 2002.

[**13**] Hyperspectral Imaging http://en.wikipedia.org/wiki/Hyperspectral_imaging

[**14**] HP Labs LOCO-I/JPEG-LS. http://www.hpl.hp.com/research/info_theory/loco/

[**15**] ISO. Lossless and Near-Lossless Compression of Continuous-Tone Still Images, ISO/IEC 14 495-1 and ITU-T Recommendation T.87, 1999.

[**16**] ISO. JPEG2000 Image Coding System, ISO/IEC 15 444-1, 2000.

[**17**] Kruskal's Algorithmhttp://en.wikipedia.org/wiki/Kruskal%27s_algorithm

[**18**] Li X., and Orchard, M.T., "Edge-directed prediction for lossless compression of natural images" *IEEE transaction on Image Processing*, 10(6), 813-817, 2001

[**19**] Linde Y, Buzo A., and Gray R. M., "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications,* 702-710, 1980.

[**20**]. Magli E, Olmo E, and Quacchio E, "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC," *IEEE Geosciences. Remote Sensing. Letters.*, 1(1), 21–25, 2004.

[**21**] Marusic S., and Deng G., "Adaptive prediction using local area training,", *Signal Processing Image Communication*, Vol.17, pp. 367-372, 2002 .

[**22**]   MicroImages   Inc.   Introduction   to   Hyperspectral   Imaging http://www.microimages.com/getstart/pdf_new/hyprspec.pdf

[**23**] MicroImages Inc. Tutorial Analyzing Hyperspectral Images with tint maps. http://www.microimages.com/getstart/pdf_new/hypanly.pdf

[**24**] Mielikäinen J. and. Toivanen P, "Improved vector quantization for lossless compression of AVIRIS Images," in *Proceedings. Eur. Signal Processing Conference*, 2002, pp. 495–497.

[**25**] Mielikainen J., Toivanen P., and Kaarna A., "Linear prediction in lossless compression of hyperspectral images," *Optical Engineering*, 42(4), pp. 1013–1017, 2003

[**26**] Mielikainen J., and Toivanen P., "Clustered DPCM for the Lossless Compression of Hyperspectral Images" IEEE *Trans. Geosciences Remote Sensing*, 41(12), 2943–2946, 2003.

[**27**] Mielikainen J., "Lossless Compression of Hyperspectral Images Using Lookup Tables," *IEEE Signal Processing Letters*, 13(3), 2006

[**28**] Miguel, A.C., Askew, A.R., Chang, A., Hauck, S., Ladner, R.E.,and Riskin, E.A. "Reduced Complexity Wavelet-Based Predictive Coding of Hyperspectral Images for FPGA Implementation" *Proceedings Data Compression Conference*, 2004.

[**29**] Minimum Spanning tree http://en.wikipedia.org/wiki/Minimum_spanning_tree

[**30**] Manolakis D., and Shaw G., "Detection algorithms for hyperspectral imaging applications" *Signal Processing Magazine, IEEE,* 19(1). pp 29-43, 2002

[**31**] Netravali A., and Limb J. O., " Picture Coding: A review," Proceedings IEEE , vol, 68,pp. 366-406, 1980

[**32**] Netravali A.N., and Haskell B.G., "*Digital Pictures Presentation, Compression, and Standards*," New York, NY: Plenum, 1988

[**33**] Parthe, R. G., "Adaptive Edge Based Prediction for Lossless Image Compression," Master's Thesis, West Virginia University, Available: https://eidr.wvu.edu/etd/documentdata.eTD?documentid=4064

[**34**] Pickering M. R., and Ryan M. J., "Efficient Spatial–Spectral Compression of Hyperspectral Data" *IEEE Transaction on Geoscience and Remote Sensing*, 39(7), 2001

[**35**] Porter W., and Enmark H., "A system overview of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Proceedings SPIE*, 834, pp. 22–31, 1997.

[**36**] Prim R. C., "*Shortest connection networks and some generalizations*", Bell System Technical Journal, 36, 1389-1401, 1957.

[**37**] Prim's Algorithm http://en.wikipedia.org/wiki/Prim%27s_algorithm

[**38**] Rizzo F., Carpentieri B., Capocelli R.M., Motta G., and Storer J. A., "High Performance Compression of Hyperspectral Imagery with Reduced Search Complexity in the Compressed Domain" *Proceedings of the Data Compression Conference*, 2004.

[**39**] Rizzo F., Carpentieri B., Motta G., and Storer J. A., "Low-Complexity Lossless Compression of Hyperspectral Imagery via Linear Prediction" *IEEE Signal Processing Letters*. 12(2) , 2005.

[**40**] Roger R. E., and Cavenor M. C., "Lossless Compression of AVIRIS Images" *IEEE Transactions on Image Processing*, 5(5), 1996.

[**41**] Ryan M J., and Arnold J F., "The Lossless Compression of AVIRIS Images by Vector Quantization" *IEEE Trans. on Geosciences. Remote Sensing,* 35(3), 1997

[**42**] Said, A. and Pearlman, W. A. "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," *IEEE Transaction Circuits and Systems for Video Technology*, 6(3), pp. 43-250, June 1996

[**43**] Schaepman, M. (2005) Spectrodirectional Imaging: From Pixels to Processes. Inaugural address, Wageningen University, Wageningen (NL).

[**44**] Seemann T. and Tischer P. E., "Generalized locally adaptive DPCM," Technical Report No. 97/301, Monash University, Australia, 1997

[**45**] Shapiro, J. M. "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Transaction on. Signal Processing*, 41(12) , pp. 3445-3462, 1993

[**46**] Shaw G., Manolakis D., "Signal Processing for Hyperspectral Image Exploitation," *IEEE Signal Processing Magazine*, 19(1), pp. 12-16, 2002

[**47**] Slyz M. J., and Neuhoff D. L., "A Nonlinear VQ-Based Predictive Lossless Image Coder," *Data Compression Conference*, pp.304-310, 1994.

[**48**] S. Tate, "Band ordering in lossless compression of multispectral images", *IEEE Transaction on Computers,* **46(4)**, 1997**.**

[**49**] Toivanen P., Kubasova O., and Mielikainen J., "Correlation based band ordering heuristic for lossless compression of Hyperspectral sounder data", *IEEE Geosciences and Remote Sensing Lettrs,* **2(1)**, 2005**.**

[**50**] Wang H., Babacan S D., and Sayood K., "Lossless Hyperspectral Image Compression Using Context-based Conditional averages"*Proceedings Data Compression Conference*, 2005.

[**51**] Weinberger M., Seroussi G., Sapiro G., "LOCO-I: A Low Complexity, Context-Based, Lossless Image Compression Algorithm," *Proceedings of IEEE Data Compression Conference*, 1996

[**52**] Weinberger M., Seroussi G., Sapiro G., "The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS", Hewlett-Packard Laboratories Technical Report No. HPL-98-193R1, November 1998, revised October 1999. IEEE Trans. Image Processing, Vol. 9, pp.1309-1324, August 2000

[**53**] Wu X, "Lossless compression of continuous-tone images via context selection, quantization, and modeling", *IEEE Transaction on Image Processing*, 6(5), 656-664, 1997.

[**54**] Wu X., Memon N., "Context-Based, Adaptive, Lossless Image Coding" *IEEE Transactions on Communications,* 45 (4), 1997.

[**55**] Wu X., and Memon N., "Context-based lossless interband compression-extending CALIC," *IEEE Transaction Image Process.*, 9(6), 994–1001, 2000.