Automatic Denoising and Unmixing in Hyperspectral Image Processing

Honghong Peng

Follow this and additional works at: https://scholarworks.rit.edu/theses

Recommended Citation
Automatic Denoising and Unmixing in Hyperspectral Image Processing

by

Honghong Peng

M.S. Beijing Institute of Technology, 2002

A dissertation submitted in partial fulfillment of the requirements for the degree of Ph.D. in the Chester F. Carlson Center for Imaging Science of the College of Science Rochester Institute of Technology

March 28, 2014

Signature of the Author ____________________________________________________________

Accepted By ____________________________________________

Coordinator, Ph.D. Degree Program Date
The Ph.D. Degree Dissertation of Honghong Peng has been examined and approved by the dissertation committee as satisfactory for the dissertation requirement for the Ph.D. degree in Imaging Science.

Prof. Sohail A. Dianat,

Prof. Juan Cockburn,

Prof. John Kerekes,

Prof. Raghuveer Rao,

Prof. Harvey Rhody,

Prof. Eli Saber,

Date
Automatic Denoising and Unmixing in Hyperspectral Image Processing

by

Honghong Peng

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

March 28, 2014

Abstract

This thesis addresses two important aspects in hyperspectral image processing: automatic hyperspectral image denoising and unmixing. The first part of this thesis is devoted to a novel automatic optimized vector bilateral filter denoising algorithm, while the remainder concerns nonnegative matrix factorization with deterministic annealing for unsupervised unmixing in remote sensing hyperspectral images. The need for automatic hyperspectral image processing has been promoted by the development of potent hyperspectral systems, with hundreds of narrow contiguous bands, spanning the visible to the long wave infrared range of the electromagnetic spectrum. Due to the large volume of raw data generated by such sensors, automatic processing in the hyperspectral images processing chain is preferred to minimize human workload and achieve optimal result. Two of the mostly researched processing for such automatic effort are: hyperspectral
image denoising, which is an important preprocessing step for almost all remote sensing tasks, and unsupervised unmixing, which decomposes the pixel spectra into a collection of endmember spectral signatures and their corresponding abundance fractions. Two new methodologies are introduced in this thesis to tackle the automatic processing problems described above.

Vector bilateral filtering has been shown to provide good tradeoff between noise removal and edge degradation when applied to multispectral/hyperspectral image denoising. It has also been demonstrated to provide dynamic range enhancement of bands that have impaired signal to noise ratios. Typical vector bilateral filtering usage does not employ parameters that have been determined to satisfy optimality criteria. This thesis also introduces an approach for selection of the parameters of a vector bilateral filter through an optimization procedure rather than by *ad hoc* means. The approach is based on posing the filtering problem as one of nonlinear estimation and minimizing the Stein’s unbiased risk estimate (SURE) of this nonlinear estimator. Along the way, this thesis provides a plausibility argument with an analytical example as to why vector bilateral filtering outperforms band-wise 2D bilateral filtering in enhancing SNR. Experimental results show that the optimized vector bilateral filter provides improved denoising performance on multispectral images when compared to several other approaches.

The non-negative matrix factorization (NMF) technique and its extensions were developed to find component based, linear representations of non-negative multivariate data. They have been shown to provide more interpretable results with realistic non-negative constraints in unsupervised learning applications such as hyperspectral imagery unmixing, image feature extraction, and data mining. This thesis extends the NMF
method by incorporating a deterministic annealing optimization procedure, which will help solve the non-convexity problem in NMF and provide a better choice of sparseness constraints. The approach is based on replacing the difficult non-convex optimization problem of NMF with an easier one by adding an auxiliary convex entropy constraint term and solving this first. Experiment results with hyperspectral unmixing application show that the proposed technique provides improved unmixing performance compared to other state-of-the-art methods.
Acknowledgements

I could not finish this thesis without the help and support from a numerous of people.

First and foremost, I would like to thank my Ph.D advisor Dr. Raghuveer Rao for his generous advice, guidance and inspiration from the start to the very end. It has been my honor to have Dr.Rao as my mentor. His support and effort are greatly appreciated. My thanks also go to my co-adviser, Dr. Sohail A. Dianat, who was abundantly helpful and offered invaluable assistance, support and guidance.

I wish to express my gratitude to Dr. John Kerekes, Dr. Juan Cockburn, Dr. Harvey Rhody and Dr. Eli Saber for serving on my thesis committee and reviewing the thesis.

I also want to extend my gratitude to all the rest of the academic and support staffs of Chester F. Carlson Center for Imaging Science, especially Ms. Susan Chan, Ms. Cindy Schultz, Ms. Joyce French and Mr. Joe Pow, for their help and assistance.

Last but not the least, I am deeply indebted to my wife and my parents for their endless love and support. I would not have finished this degree without their encouragement. I dedicate this thesis to them.
# Table of Contents

Abstract ......................................................... I  
Acknowledgements ........................................... IV  
List of Tables .................................................. VIII  

Chapter 1. Introduction ........................................ 1  
    Motivation .................................................. 1  
    Statement of Work ........................................ 2  
    Organization of Dissertation .............................. 4  

Chapter 2. 2D Bilateral Filter Range and Distance Kernel Optimization by Risk Minimization ........................................... 5  
    2.1 Introduction ............................................ 5  
    2.2 Mean-Squared Error (MSE) and Stein’s Unbiased Risk Estimator (SURE) for Gray Image Denoising ..................... 7  
    2.3. Bilateral Filtering Parameter Optimization .................. 9  
    2.4. Experimental Results and Discussion ...................... 11  

Chapter 3. Multispectral Image Denoising with Optimized Vector Bilateral Filter ........................................... 15  
    3.1 Introduction ............................................. 15  
    3.2 Optimized Vector Bilateral Filter .......................... 17  
        3.2.1 Vector bilateral filter .................................. 17  
        3.2.2 Mean-Squared Error (MSE) and Stein’s Unbiased Risk Estimator (SURE) for multi-band image denoising .......... 24  
        3.2.3 Vector bilateral filtering parameter optimization .......... 25  
    3.3 Experimental Results and Discussion ...................... 29  
        3.3.1 Experiment design ..................................... 29  
        3.3.2. Experiment test case results and discussion .......... 32  

Chapter 4. Nonnegative Matrix Factorization with Deterministic Annealing for Unsupervised Unmixing of Hyperspectral Imagery ....................... 48  
    4.1 Introduction ............................................. 48  
    4.2. Nonnegative Matrix Factorization with Deterministic Annealing ................................. 49  
        4.2.1. Non-negative matrix factorization .......................... 49  
        4.2.2. NMF with Deterministic annealing ....................... 51  
    4.3. Experimental Results and Discussion ...................... 53  
        4.3.1 Synthetic Data ......................................... 55  
        4.3.2 Real world hyperspectral imagery ....................... 63  
    4.4 Experimental Results & Discussion for Combined Optimized Vector Bilateral Filter and DANMF ........................................... 66  
        4.4.1 Synthetic Data ......................................... 66
4.4.2 Real world hyperspectral imagery.............................................. 73

Chapter 5. Conclusion and Future Work ................................. 76
List of Figures

Figure. 1 2D Bilateral Filter Experiment results: ................................................................. 13
Figure. 2 2D bilateral filter \( \hat{R}_\theta \) vs. MSE plot ................................................................. 14
Figure. 3 Illustration of perfect edge image case ................................................................. 20
Figure. 4 Synthetic color dead leaves image examples ......................................................... 30
Figure. 5. Color dead leaves target moving average comparison result for 2D bilateral filter .......................................................................................................................... 33
Figure. 6. Color dead leaves target moving average comparison result for SURE-LET .. 33
Figure. 7. Color dead leaves target moving average comparison result for vector bilateral filter .......................................................................................................................... 34
Figure. 8 Hyperspectral image denoising performance trend ................................................. 35
Figure. 9. HYDICE image moving average comparison results ............................................ 37
Figure. 10 Denoising experiment results of color dead leaves image: ................................. 39
Figure. 11 Vector bilateral filter: \( \hat{R}_\theta \) vs. MSE plot ......................................................... 40
Figure. 12. Experiment results on HYDICE band image....................................................... 46
Figure. 13. Experiment results of HYDICE waste band image: .......................................... 47
Figure. 14 Synthetic data unmixing spectrum results ......................................................... 62
Figure. 15 HYDICE image unmixing Experiment results: .................................................. 65
Figure. 16 Synthetic data denoising and unmixing combined spectrum results .............. 73
Figure. 17 HYDICE image combined denoising and unmixing experiment results: ............ 75
List of Tables

Table.1 Experiment results for “pepper” image............................................................. 14
Table.2 Comparison on color dead leaves model (same noise level in each band)........ 34
Table.3 Comparison on HYDICE urban image (same noise level in each band)......... 38
Table.4 Comparison on color dead leaves model (color noise with noise correlation)... 41
Table.5 Comparison on HYDICE urban image ................................................................. 41
Table.6 Comparison on total correlated dead leaves image (increased number of bands
with fixed noise standard deviation of 30 for each band)........................................... 42
Table.7 Comparison on HYDICE urban image (increased number of bands with fixed
noise standard deviation of 30 for each band)............................................................. 43
Table.8 SAD comparision for synthetic data ................................................................. 55
Table.9 RMSE comparison for synthetic data ............................................................... 56
Table.10 RMSE comparision for synthetic data ............................................................ 66
Table.11 SAD comparision for synthetic data ............................................................... 67
Chapter 1. Introduction

Motivation

Automatic hyperspectral image processing has become more and more important due to the popularity of hyperspectral imaging systems that have been developed. Examples of such hyperspectral imaging systems are the Hyperspectral Digital Imagery Collection Experiment (HYDICE) [1] which images in 210 bands and the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) [2] which images in 224 bands. Both systems image in the electromagnetic spectrum range of 400 to 2500nm. The hyperspectral imaging systems provide users with unprecedented capability in conducting remote sensing applications such as target detection, target activities, surveillance, land use, agriculture assessment, ecological and environmental monitoring, mineral exploitation, change detection, man-made materials identification and detection, and ground-cover classification [3], [4].

Usually, huge volumes of raw data processing are involved in these applications. Concurrently in these applications, the hyperspectral image processing chain plays a vital role. The final results are heavily influenced by each step along the whole processing chain. Due to these two natures of hyperspectral image processing applications, it is critical to automat the processing steps to reduce human work load and achieve the best performance. Common hyperspectral image processing and analysis algorithm chains can be classified as follows [5], [6]: detect known or unknown objects in a given scenario; classify/segment the image into predominant material regions; estimate the materials and the corresponding area fractions that they
occupy within a pixel, i.e. hyperspectral unmixing. In this research, we focus on the
development of two important automatic processes for hyperspectral process chain:
hyperspectral image denoising and unsupervised hyperspectral unmixing.

**Statement of Work**

Hyperspectral image denoising is a fundamental step for almost all hyperspectral
image processing chains. As the same for most signal processing problems, noise
suppression of hyperspectral images in the first step is essential to the success of final
tasks. The unique challenge for hyperspectral image denoising is how to effectively
utilize the spatial & spectral correlation presented in the hyperspectral image
concurrently. Hyperspectral image noise can be modeled as stochastic Gaussian
distributed with possible intra-band correlation [24], while high intra-band correlation
is often observed in hyperspectral image. Considering these correlation characteristics
of both signal and noise, a joint denoising approach efficiently exploiting mutual
correlation between bands and pixels, i.e. multivariate approach, would utilize more
information and thus has advantage over denoising approaches that only process each
band image separately. Like most denoising problem, the performance of the
hyperspectral image denoising algorithm is greatly influenced by the parameters
tuning. As there are a lot of parameters involved in the most vector denoising
methods, it will be very challenging for human beings to find the optimal parameter
set. A parameter optimization solution based on Stein’s unbiased risk estimation
metric is proposed in this paper to address this problem.

Since the spatial resolution of any hyperspectral imaging system is finite, the pixel
spectra acquired by the system are very often the additive result of spatially mixed
spectra of different substances. Hyperspectral unmixing [56], [57] help to decompose the composite spectral image into a collection of constituent (end-member) spectra signatures, and their corresponding abundance fraction maps. Hyperspectral unmixing, as an important object analysis step for inputs in decision-making processes, is a vital part in the hyperspectral image processing chain for many environmental applications. The general hyperspectral unmixing problem is an ill-posed problem if both end-member spectra and abundance map are unknown. In most approaches, supervised unmixing is carried out with intensive interaction from the image analyst to determine the endmember spectra. To relieve the burden of human intervention, unsupervised unmixing is introduced to fully automate the unmixing process by having the machine perform both the endmember spectra and corresponding abundance map estimation simultaneously. Non-negative matrix factorization [46],[47] (NMF) has been used in the context of machine learning and factor analysis to solve problems with similar mathematical structure.

The major contributions of this thesis are:

1. Developed parameters optimization scheme for traditional 2D bilateral filter based on Stein’s unbiased risk estimation method (SURE). Under additive Gaussian noise assumption, demonstrated 2D bilateral filter parameter optimization can be achieved even without access to noise free ground truth image.

2. Created a new vector bilateral filter form for hyperspectral image denoising problem, which could ultilize full hyperspectral image in the denoising process without introducing cross-talk between band images. Also developeda systematic
formulation approach with matrix calculus for vector bilateral filter multi-dimentional parameters optimization based on SURE. With additive multi-dimentional Gaussian noise assumption, this parameter optimization approach is proved to be able to guide parameter optimization of proposed algorithm without ground truth knowledge. We have also extended dead leaves targets into color form to get more representitive quantified multi-channel image denoising experiment results.

3. Further extended the NMF approach with more sophisticated entropy constraint condition and introduced deterministic annealing optimizing strategy to better solve the hyperspectral image unmixing problem.

4. Demonstrate the system performance when combining the two approaches.

**Organization of Dissertation**

The thesis is divided into 5 chapters: Chapters 1 gives an introduction and statement of the work. Chapter 2 establishes 2D bilateral filter parameters optimization approach as a preparation for vector bilateral filter optimization and Chapter 3 addresses the automatic hyperspectral image denoising problem with optimized vector bilateral filter. Chapter 4 covers unsupervised hyperspectral image unmixing with deterministic annealing augmented Non-negative matrix factorization and also describes the system performance when the two solutions are combined together. Finally, Chapter 5 draws conclusions of the thesis work and discusses the direction of future work.
Chapter 2.  2D Bilateral Filter Range and Distance Kernel Optimization by Risk Minimization

2.1 Introduction

Noise is an inevitable part of real world signals. Before introducing the optimized vector bilateral filter for multispectral and hyperspectral images denoising, we first review previous research on optimized 2D bilateral filter for gray image denoising problem to lay down some foundation for further discussion.

The bilateral filter \[7\],\[8\],\[9\] finds application in image noise reduction. It is a nonlinear filter that takes both range (intensity) and distance (spatial proximity) metrics into account. The bilateral filter preserves localizations of the edges well while suppressing random noise. It is similar to a linear space-invariant filter except that, in addition to a linear convolution kernel that weights pixel values as a function of distance from the position in question, it also has a nonlinear kernel that weights pixel values as a function of their relative value with respect to that of the current pixel.

The input-output relationship for a bilateral filter is given by:

\[
I_{\text{out}}(s) = f_{2D-\text{bilateral}}(I_{\text{in}}(s), (\sigma_d, \sigma_s)) \\
= \sum_{p \in \Omega} g_d\left(|p-s|, \sigma_d\right) g_s\left(D(p,s), \sigma_s\right)I_{\text{in}}(p) / \sum_{p \in \Omega} g_d\left(|p-s|, \sigma_d\right) g_s\left(D(p,s), \sigma_s\right) 
\]

(1)

where \(p\) and \(s\) are 2-D vectors of the pixel coordinates, \(\Omega\) is the summation window, \(I_{\text{in}}(p)\) is the input image pixel value at position \(p\), \(I_{\text{out}}(s)\) is the output pixel value at
center position \( s \), \( D(p,s) \) is pixel dissimilarity, which in most 2D image cases is just defined as the difference of pixel values:

\[
D(p,s) = \left| I_{in}(p) - I_{in}(s) \right|
\]

(2)

\( g_d(x,\sigma_d) \) and \( g_s(x,\sigma_s) \) are the weight functions for geographical distance and pixel value difference respectively. Conventionally, they are both defined as Gaussian functions [7]:

\[
g_d(x,\sigma_d) = \frac{1}{\sqrt{2\pi\sigma_d}} e^{-\frac{x^2}{2\sigma_d^2}}, \quad g_s(x,\sigma_s) = \frac{1}{\sqrt{2\pi\sigma_s}} e^{-\frac{x^2}{2\sigma_s^2}}
\]

(3)

Any pixel that is far from the pixel being considered, in position or in value, will have a very small weight from either \( g_d \) or \( g_s \) and thus have a very small influence on the output. This contributes to edge preservation in denoising applications as follows.

The pixel values on one side of an edge differ more from values of pixels on the other side of the edge than from pixel values on the same side. Thus, pixels on any given side of an edge contribute to smoothing on the same side but not to smoothing on the other.

We risk edge smearing if \( \sigma_s \) and \( \sigma_d \) are too large and poor noise suppression if they are too small. There are thus optimum values combination that provides the best tradeoff between noise suppression and edge preservation. However, there is no work reported so far that addresses the problem of determining such optimal values for any individual image with a closed form solution, although some empirical study has been carried out in [10], and a complex parameter optimization procedure which involves pixel classification and exhaustive optimal parameters searching on representative training images set is described in [11]. This chapter provides an approach to
determining the parameters of a bilateral filter to achieve optimum tradeoff between
denoising and edge preservation. The approach is based on forming Stein’s unbiased
risk for the estimate of the true image through bilateral filtering.

This chapter is organized as follows. Section 2 defines Mean Squared Error
(MSE), estimation risk and provides Stein’s unbiased risk estimator for gray image
denoising. In Section 3, a parameter optimization procedure is proposed.
Experimental results demonstrating key features of the proposed approach are
presented in Section 4.

2.2 Mean-Squared Error (MSE) and Stein’s Unbiased Risk
Estimator (SURE) for Gray Image Denoising

Suppose there is an (noisy) observed 2D gray image signal:

$$I_{in}(s) = I_{real}(s) + n(s), \ s \in T$$  \hspace{1cm} (4)

where $I_{real}(s)$ is the deterministic 2D gray image signal and $n(s)$ is an independent and
identical distributed (i.i.d.) Gaussian noise with mean zero and covariance $\sigma^2$, and $T$ is the set of spatial indices of the whole image($T \equiv [T_1, T_2, \ldots T_{HL}]$). The total number of
pixels in the image is $HL$. Suppose $I_{out, \theta}(s)$ is an estimate of $I_{real}(s)$ obtained from

$I_{in}(s)$ as

$$I_{out, \theta}(s) = f(I_{in}(s), \theta) = f(I_{real}(s) + n(s), \theta), \ s \in T$$  \hspace{1cm} (5)
where \( f \) is an estimator (possibly nonlinear) of \( I_{\text{real}}(s) \) and \( \theta \) is a parameter vector associated with this estimator. The goodness of the estimator \( f \) can be measured using sample mean square error (MSE) measure expressed in L2 norm as:

\[
MSE = \frac{1}{HL} \sum_{s \in T} \left[ \left\| I_{\text{out},\theta}(s) - I_{\text{real}}(s) \right\|^2 \right]
\]  

(6)

The difficulty in applying this measurement metric to the observed noisy image is that the underlying image, \( I_{\text{real}}(s) \), is unknown. The \( MSE \) is a random variable depending on noise, and the expected value of \( MSE \) in (6) is referred to as Risk \( R_\theta \):

\[
R_\theta = E[MSE]
\]  

(7)

The problem of estimating Risk without access to ground truth image is circumvented to some extent with Stein’s Unbiased Risk Estimator (SURE) [12], [13]. With additive Gaussian noise hypothesis, SURE provides an analytical means for unbiased estimation of MSE. It is given by:

\[
\hat{R}_\theta = \frac{1}{HL} \sum_{s \in T} \left[ \left\| I_{\text{in}}(s) - I_{\text{out},\theta}(s) \right\|^2 \right] - \sigma^2 + 2 \sigma^2 \frac{1}{HL} \sum_{s \in T} \left[ \frac{\partial f(I_{\text{in}}(s), \theta)}{\partial I_{\text{in}}(s)} \right]
\]  

(8)

It is an unbiased estimator for the expectation of MSE in (6):

\[
R_\theta = E[MSE] = E[\hat{R}_\theta]
\]  

(9)

For image denoising purpose, we regard \( \hat{R}_\theta \) as a reliable estimate of the MSE for optimization, as the total number of pixels, \( HL \), in an image is usually a very large number. If we know \( \sigma^2 \) (or estimate it separately), then we can calculate \( \hat{R}_\theta \) and minimize it with respect to \( \theta \) to find the optimum parameters of the signal estimator in Eq. (5).
2.3. Bilateral Filtering Parameter Optimization

Now, suppose $I_{in}$ is the input 2D image and it is the sum of the original signal image and Gaussian noise. By combining Eqs. (1) and (8), we can derive the closed form expression of unbiased estimated risk by treating the output of 2D bilateral filtering as an estimate of the underlying noiseless image. In Eq. (9), the parameter $(\sigma_d, \sigma_s)$ corresponds to $\theta$. It is to be noted that $\sigma_s$ and $\sigma_d$ are the bilateral filter parameters targeted for optimizing, while $\sigma$ in Eq. (8) is the estimated noise variance. Therefore, the SURE for 2D bilateral filter is given by:

$$\hat{R}_\theta = \frac{1}{HL} \sum_{s \in T} \left[ \|I_{in}(s) - f_{2D\text{-bilateral}}(I_{in}(s), \theta)\|^2 - \sigma^2 \frac{\partial f_{2D\text{-bilateral}}(I_{in}(s), \theta)}{\partial I_{in}(s)} \right]$$

(10)

A closed-form expression for $\partial f_{2D\text{-bilateral}}(I_{in}(s), \theta) / \partial I_{in}(s)$ is given by:
The noise level $\sigma$ can be estimated efficiently [14],[15] as:

$$\sigma = 1.4826 \cdot \text{MAD}(\nabla I_m)$$

$$= 1.4826 \cdot \text{median}_I \left(\|\nabla I_m - \text{median}_I (|\nabla I_m|)\|\right)$$

(12)
where $MAD$ and $\text{median}_I$ denote the mean absolute deviation and median respectively. $\nabla I_{in}$ is the gradient of the input image. The constant 1.4826 comes from the knowledge that the $MAD$ of a zero-mean normal distribution with unit variance is $0.6745=1/1.4826$.

The optimization problem is posed as:

$$ (\sigma_d_{opt}, \sigma_s_{opt}) = \arg \min_{\sigma_d, \sigma_s} \hat{R}(\sigma_d, \sigma_s) $$  \hspace{1cm} (13)

This unconstrained single-parameter non-linear optimization can be solved using Newton–Raphson method or any other numerical method. In essence, solution to Eq. (13) maximizes the signal to noise ratio since the risk is an estimate of the noise power.

### 2.4. Experimental Results and Discussion

The proposed approach was tested on images with additive white Gaussian noise of different variances. The initial test was to verify if the solution provided by the minimization proposed in Eq. (13) results in maximization of the SNR at the corresponding bilateral filter output. For each of these images, $\Omega$ in Eq.(1) were chosen to provide the best visual tradeoff between blurring and noise-removal in the absence of range weighting (i.e. with $g_r$ and $g_d$ identically equal to unity). For example, a $11\times11$ neighborhood was found to work best for the “pepper” image of Figure.1.

The solutions to Eq.(13) were obtained using $\sigma$ estimated from Eq.(12). The estimated values were found to be close to the true values. In contrast to the conventional
wisdom in bilateral filtering of using $\sigma_s \approx \sigma$, the optimal solutions generally yielded $\sigma_{s, opt}$ that was a non-unity multiple of $\sigma$. For example, for additive noise with variance equal to 0.02, the solution yielded $\sigma_{s, opt} = 2.5 \sigma$ and $\sigma_{d, opt} = 1.5$.

In all cases, it was found that the signal to noise ratio was in fact maximized when $\sigma_s$ and $\sigma_d$ were set to $\sigma_{s, opt}$ and $\sigma_{d, opt}$ respectively. For example, when the bilateral filter was applied with $\sigma_s$ ranging from $1 \sigma$ to $5.5 \sigma$ and $\sigma_d$ ranging from 1 to 2.5 to the Peppers image with noise variance of 0.02, and the PSNR was calculated it yielded the results shown in Table.1. It is seen that when $\sigma_s$ equals 2.5$\sigma$ and $\sigma_d$ equals 1.5, the bilateral filter will output the best denoised image. We also calculated $\hat{R}_\theta$ for the same set of $\sigma_s$ and $\sigma_d$, and recorded them in Table.1. As expected, the optimal combination of $\sigma_s$ and $\sigma_d$ will generate the minimum $\hat{R}_\theta$, and as shown in Fig.2, the calculated $\hat{R}_\theta$ shows good match with MSE, either with real $\sigma$ or estimated $\sigma$. There is a maximum of 4.87 dB PSNR difference over the range of $\sigma_s$ and $\sigma_d$ settings.

The parameter optimization also leads to visual improvement in denoising results as shown in Fig.1. Compared to just setting $\sigma_s$ to $\sigma$ and $\sigma_d$ to 1, the optimized $\sigma_s$ and $\sigma_d$ provides a 4.87 dB improvement in PSNR for bilateral filter, which is perceived in improved denoised image quality as shown in Fig.1. Using the optimum value combination of $\sigma_s$ and $\sigma_d$ provide the best tradeoff between blurring and denoising.
Figure 1 2D Bilateral Filter Experiment results:

(a) Original “pepper” (b) noisy “pepper” image (c) 2D bilateral filtered “pepper” image with $\sigma_s = 1 \times \sigma$ and $\sigma_d = 1$ (d) 2D bilateral filtered “pepper” image with $\sigma_s = 2.5 \times \sigma$ and $\sigma_d = 1.5$ (e) 2D bilateral filtered “pepper” image with $\sigma_s = 5.5 \times \sigma$ and $\sigma_d = 2.5$
### Table 1: Experiment results for “pepper” image

<table>
<thead>
<tr>
<th>$(\sigma_d, \sigma_x$ in $\sigma$ unit)</th>
<th>PSNR(dB)</th>
<th>$\hat{R}_\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>20.02</td>
<td>1887.45</td>
</tr>
<tr>
<td>(1, 2.5)</td>
<td>24.25</td>
<td>1468.78</td>
</tr>
<tr>
<td>(1, 4)</td>
<td>24.68</td>
<td>1438.01</td>
</tr>
<tr>
<td>(1, 5.5)</td>
<td>24.59</td>
<td>1439.13</td>
</tr>
<tr>
<td>(1.5, 1)</td>
<td>20.79</td>
<td>1768.02</td>
</tr>
<tr>
<td>(1.5, 2.5)</td>
<td>24.89</td>
<td>1422.61</td>
</tr>
<tr>
<td>(1.5, 4)</td>
<td>24.57</td>
<td>1430.73</td>
</tr>
<tr>
<td>(1.5, 5.5)</td>
<td>24.15</td>
<td>1449.43</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>21.05</td>
<td>1728.61</td>
</tr>
<tr>
<td>(2, 2.5)</td>
<td>24.63</td>
<td>1428.74</td>
</tr>
<tr>
<td>(2, 4)</td>
<td>23.86</td>
<td>1462.35</td>
</tr>
<tr>
<td>(2, 5.5)</td>
<td>23.31</td>
<td>1493.69</td>
</tr>
<tr>
<td>(2.5, 1)</td>
<td>21.13</td>
<td>1715.63</td>
</tr>
<tr>
<td>(2.5, 2.5)</td>
<td>24.24</td>
<td>1445.10</td>
</tr>
<tr>
<td>(2.5, 4)</td>
<td>23.25</td>
<td>1497.40</td>
</tr>
<tr>
<td>(2.5, 5.5)</td>
<td>22.63</td>
<td>1538.41</td>
</tr>
</tbody>
</table>

![Graph](image_url)  

**Figure 2: 2D bilateral filter: $\hat{R}_\theta$ vs. MSE plot**
Chapter 3. Multispectral Image Denoising with Optimized Vector Bilateral Filter

3.1 Introduction

Noise is an inevitable part of most real world multispectral and hyperspectral images and it is well known that good denoising leads to performance improvement in problems such as classification, segmentation and object identification [16]-[19]. Various approaches [20]-[31] have been proposed for noise removal in such images. Recently, a vector formulation of a bilateral filter was provided in [32]. It was shown to possess several advantages, the most important being a weighting mechanism that tends to preserve edges, thus contributing to improved denoising vs. edge-preservation tradeoffs compared to other approaches such as, for example, those based on the wavelet transform.

A problem with the vector bilateral filter, however, is that the performance depends on the choice of the filter parameters. In [32] the parameters were chosen as functions of the estimated noise variance of the various PCA components of the noisy hyperspectral/multispectral input image. Although there is some basis for this approach, it is largely ad hoc. In this chapter, similar to the previous chapter, we formulate an optimization problem by adopting the perspective that the vector bilateral filtered output is essentially a nonlinear estimate of the underlying image. In the specific case of multivariate Gaussian noise – it does not have to be white – the risk (to be defined later) associated with a nonlinear estimator of the mean can itself be estimated using Stein’s unbiased risk estimation (SURE) [12]. We make this the
basis of an approach that estimates the parameters using the estimated SURE as an optimality criterion. It is to be noted that SURE-based optimization has been used in other denoising contexts including those founded on the wavelet transform [28]-[31]. In [28], a closed-form solution of Stein’s estimation of mean squared error (MSE) is deduced for a Maximum-a-Posteriori (MAP) multivariate denoising estimator with a Bernoulli–Gaussian prior, and parameters that correspond to minimum MSE are selected as optimal parameters. A more general framework that does not assume any prior model is proposed in [29],[30] for MAP multivariate denoising estimator. A multi-channel SURE with linear expansion of thresholds (SURE-LET) approach is proposed in [31] to simplify parameter optimization based on SURE. The denoising function is constructed as a linear expansion of thresholds and optimized linearly according to SURE without any prior model assumption. The proposed wavelet thresholding function in [31] is “point-wise” and depends on the coefficient vector that contains coefficients of every channel in the same location and their parent coefficients in the coarser wavelet sub-band. This approach is a state of the art denoising technique for multispectral images.

In Chapter 2, 2D bilateral filter parameters optimization is proposed [43], which only optimizes a pair of fixed parameters, the geographical distance parameter and pixel similarity differences parameter, for a scalar image. In this chapter, we extend the framework in Chapter 2 to the vector bilateral image filtering case to optimize an arbitrary number of parameters. The contribution of this chapter is in showing that (a) for multi-spectral images, the vector bilateral filter demonstrates superior performance over the component-wise bilateral approach, (b) the vector bilateral filter
parameter optimization can be cast into a SURE-based framework and (c) the optimized filter provides improved noise removal vs. edge preservation tradeoff compared to techniques in [31] and [43].

This chapter is organized as follows. In Section 2, the bilateral filter is extended to a vector form and its parameter optimization procedure is proposed based on MSE and Stein’s unbiased risk estimator for multi-band image denoising. Section 3 shows the experimental results.

3.2 Optimized Vector Bilateral Filter

3.2.1 Vector bilateral filter

Let $I_{in}$ be an acquired noisy $k$-band multispectral/hyperspectral image ($k=3$ for color image) with spatial dimension of $H$ rows and $L$ columns. Let

$$I_{in}(s) = I_{red}(s) + n(s)$$  \hspace{1cm} (14)

where $s$ is a 2-D vector denoting the pixel’s coordinates, $I_{in}(s)$ represents the $k$-dimensional vector often referred to as the spectral vector at position $s$, $I_{red}(s)$ denotes the actual image, and $n(s)$ is a (generally non-white) Gaussian additive noise vector whose covariance matrix is $\Gamma_n$.

For a pre-defined domain kernel $\Omega$, a $n \times n$ neighborhood centered at the target pixel at position $s$, the vector bilateral filter takes the form [32]:

\begin{align}
&
\end{align}
\[ I_{\text{out}}(s) = f_{\text{vec-bilateral}}(I_{\text{in}}(s), (\sigma_d, \Sigma)) \]
\[ = \sum_{p \in \Omega} g_d \left( |p-s|, \sigma_d \right) g_s \left( D(p,s), \Sigma \right) I_{\text{in}}(p) \]
\[ = \sum_{p \in \Omega} g_d \left( |p-s|, \sigma_d \right) g_s \left( D(p,s), \Sigma \right) I_{\text{in}}(p) \]

where \( I_{\text{in}}(s) \) and \( I_{\text{out}}(s) \) are the input and output images of the vector bilateral filter, \( p \) is a 2-D vector representing the pixel coordinates, \( D(p,s) = I_{\text{in}}(p) - I_{\text{in}}(s) \) is the pixel value difference, \( g_d(x, \sigma_d) \) and \( g_s(x, \Sigma) \) are the weight functions for geometric distance and pixel value difference respectively. \( g_d(x, \sigma_d) \) is defined as a Gaussian function:

\[ g_d(x, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{x^2}{2\sigma^2} \right) \]

It is very similar in form to the geometric distance weight function of a scalar bilateral filter for gray scale images. Except for the vectorized input and output, the major difference between the vector and scalar cases lies in the weight function for the pixel value difference metric, for which the authors proposed using [32].

\[ g_s(x, \Sigma) \equiv \frac{1}{(2\pi)^{N/2}} \left| \Sigma \right|^{1/2} \exp \left( -\frac{x^T \Sigma^{-1} x}{2} \right) \]

where \( \Sigma \) is a positive definite matrix that we will refer to as the correlation matrix or the range kernel matrix. From Eqs. (15) and (17) it is seen that the exponent of \( g_s(x, \Sigma) \) has the squared Mahalanobis distance between \( I_{\text{in}}(p) \) and \( I_{\text{in}}(s) \).

Compared to component-wise 2D bilateral filter, one important advantage of the proposed vector bilateral filter is that it can exploit correlation between bands more efficiently, and as shown below, this advantage increases as the number of bands
increases. Considering that edges are the salient structures in an image, we demonstrate this point here using an image that has a single sharp edge.

Assume we have a $k$-band multi-spectral image with similar perfect edge images in each band. The $ith$ band image, for any $i$ between 1 and $k$ is assumed to have the form shown in Figure. 3 with signal edge value $I_{e,i}$ and constant background value $I_{b,i}$.

In this way, the correlation between any two band images is 1.
Figure 3 Illustration of perfect edge image case

Note that the edge transition is assumed to occur at the same locations in all the bands. However, the edge height, determined by the difference between the top or signal value $I_{t,i}$ and the bottom or background value $I_{b,i}$, varies with the bands. Also assume that the signal image is contaminated with zero mean additive Gaussian white noise.
noise to form the actual observed multispectral image. The noise covariance matrix is \( \Gamma_n = \sigma^2 E_n \), where \( \sigma \) is the identical noise standard deviation for each channel and \( E_n \) is a \( k \times k \) identity matrix. Then for a center pixel with value \( I_c(s) = I_{i,j} + n(s)_j \), right on the edge position \( s \) in the observed \( ith \) band image, the corresponding denoising estimation result \( \hat{I}(s)_i \) can be expressed as:

\[
\hat{I}(s)_i = \frac{\sum_{j \in \Omega^-} W_j^- (I_{b,j} + n_{j,j}) + \sum_{i \in \Omega^+} W_i^+ (I_{i,j} + n_{i,j})}{\sum_{j \in \Omega^-} W_j^- + \sum_{i \in \Omega^+} W_i^+}
\]

where \( W_j^- \) is the positive weighting factor for pixels that belong to background kernel area \( \Omega^- \) with value of noise \( n_{j,j} \) plus background \( I_{b,j} \), and \( W_i^+ \) is the positive weighting factor for pixels that belong to signal kernel area \( \Omega^+ \) with the value of signal \( I_{i,j} \) plus noise \( n_{i,j} \). As assumed before, the entire signal has the same value, i.e. \( I(c, s)_j = I_{i,j} \). Accordingly, we can derive signal to error (SER) measure for this estimation as:

\[
SER_{edge} = \left| \frac{I_{i,j}}{I_{i,j} - \hat{I}(s)_i} \right| = \left| \frac{I_{i,j}}{I_{i,j} - \sum_{j \in \Omega^-} W_j^- (I_{b,j} + n_{j,j}) + \sum_{i \in \Omega^+} W_i^+ (I_{i,j} + n_{i,j})} \frac{\sum_{j \in \Omega^-} W_j^- + \sum_{i \in \Omega^+} W_i^+}{\sum_{j \in \Omega^-} W_j^- + \sum_{i \in \Omega^+} W_i^+} \right|
\]

As the noise is zero mean Gaussian noise and assuming the edge signal, \( I_{i,j} - I_{b,j} \), is significantly larger than the noise, we can further approximate the \( SER_{edge} \) as:
\[ SER_{\text{edge}} = \frac{I_{t,i} - I_{b,i}}{1 + \sum_{j \in \Omega} W_j^-} \left(1 + \sum_{i \in \Omega} W_i^+ \right) \]

where

\[ WR_i = \frac{W_i^+}{\sum_{j \in \Omega} W_j} \]

Thus, the \( SER_{\text{edge}} \) measure is positively related to each individual \( WR_i \). The larger each \( WR_i \) is, the better the noise filtering result. The ideal denoising filter weighting kernel for the surrounding pixels would intuitively be such that: the pixel with a similar value to center pixel should be assigned a higher weight while pixels with significant different values should be assigned lower weights. In our perfect edge case, the ratio of weights between signal area and background area is very similar for observed signal region and background noise region, and we can consider only one ratio to predict the estimation kernel performance. Higher value of this ratio will lead to better \( SER \) and filtering results.

If we only consider the pixel value weighting kernel (the fixed geometric weighting will not affect the final result), for the vector bilateral filter case, if we set \( \sum \) to be a diagonal matrix with parameter \( \sigma_{s,d} \) for the \( d \)th band, the filtering weight ratio is calculated as:
\[ WR_{\text{vec}} = \frac{\exp\left( \sum_{d=1}^{k} \frac{(n(s)_d - n_{i,d})^2}{2\sigma_{s,d}^2} \right)}{\sum_{j \in \Omega} \exp\left( \sum_{d=1}^{k} \frac{(I_{t,d} - I_{b,d} + n(s)_d - n_{j,d})^2}{2\sigma_{s,d}^2} \right)} \]

\[ = \frac{1}{\sum_{j \in \Omega} \prod_{d=1}^{k} \exp\left( -\frac{(I_{t,d} - I_{b,d})^2}{2\sigma_{s,d}^2} \right)} = J \prod_{d=1}^{k} \exp\left( \frac{(I_{t,d} - I_{b,d})^2}{2\sigma_{s,d}^2} \right) \tag{21} \]

where \( n(s)_d, n_{i,d}, n_{j,d} \) are the additive noise at the center pixel, the signal side pixel and background side pixel at band \( d \) image respectively, \( J \) is the number of pixels belong to background kernel area \( \Omega \). The approximation is made with high signal to noise ratio assumption. With previous hypothesis, if 2D bilateral filter is applied separately to \( i \)th band image, it will be a special case of vector bilateral filter, the corresponding \( WR_{i,2D} \) is simplified to \( J \exp\left( \frac{(I_{t,i} - I_{b,i})^2}{2\sigma_{s,i}^2} \right) \). As for the assumed ideal scenario, \( \exp\left( \frac{(I_{t,i} - I_{b,i})^2}{2\sigma_{s,i}^2} \right) \) is always larger than 1 for each band, the vector bilateral filter will always have a larger \( WR \) and hence outperform component wise 2D bilateral filter by utilizing more band images, this performance advantage will increase with band numbers. Since for multispectral images there is usually a strong degree of correlation of spatial edge features between the bands, the proposed vector bilateral filter should commonly benefit from this fact.
3.2.2 Mean-Squared Error (MSE) and Stein’s Unbiased Risk Estimator (SURE) for multi-band image denoising

We now address the problem of finding $\sigma_j$ and $\sum$ such that the vector bilateral filter offers optimum performance in terms of tradeoff between noise removal and edge-smearing across all spectral band images.

For an (noisy) observed $k$-dimensional data vector in Eq. (14), suppose $\mathbf{I}_{\text{out}, \theta}(s)$ is a denoised image obtained from $\mathbf{I}_{\text{in}}(s)$ as

$$\mathbf{I}_{\text{out}, \theta}(s) = f(\mathbf{I}_{\text{in}}(s), \theta) = f(\mathbf{I}_{\text{real}}(s) + n(s), \theta), \ s \in \mathcal{T}$$

(22)

where $f$ is a nonlinear estimator of $\mathbf{I}_{\text{real}}(s)$, $\theta$ is a parameter vector associated with this estimator, and $\mathcal{T}$ is the set of spatial indices of the whole image ($\mathcal{T} \triangleq [T_1, T_2 \cdots T_H]$). The total number of pixels in the image is $HL$. The quality of this denoising estimator $f$ is often evaluated using sample mean square error (MSE) measure expressed in L2 norm as:

$$MSE = \frac{1}{HL} \sum_{s \in \mathcal{T}} \left[ \| \mathbf{I}_{\text{out}, \theta}(s) - \mathbf{I}_{\text{real}}(s) \|^2 \right]$$

(23)

The difficulty in applying this measurement metric to the observed noisy image is that the underlying image, $\mathbf{I}_{\text{real}}(s)$, is unknown. The $MSE$ is a random variable depending on noise, and the expected value of $MSE$ in (23) is referred to as Risk $R_\theta$:

$$R_\theta = \mathbb{E}[MSE]$$

(24)

The problem of estimating Risk without access to ground truth image is circumvented to some extent with Stein’s Unbiased Risk Estimator (SURE) [12],

24
With additive multivariate Gaussian noise hypothesis, SURE provides an analytical means for unbiased estimation of MSE. It is given by:

$$\hat{R}_0 = \frac{1}{HL} \sum_{s \in \Gamma} \left[ \left\Vert I_{in}(s) - I_{out,\theta}(s) \right\Vert^2 - \text{Tr}(\Gamma_n) + 2 \frac{1}{HL} \sum_{s \in \Gamma} \text{Tr} \left( \Gamma_n J_f \left( I_{in}(s) \right) \right) \right]$$  \hspace{1cm} (25)

where $\Gamma_n$ is the noise covariance matrix and $J_f \left( I_{in}(s) \right)$ is the Jacobian matrix with respect to $I_{in}(s)$. The matrix element in the $i$-th row and $j$-th column of $J_f \left( I_{in}(s) \right)$ is given by:

$$J_f \left( I_{in}(s) \right)_{i,j} = \frac{\partial f(I_{in}(s), \theta)}{\partial I_{in,j}(s)}$$  \hspace{1cm} (26)

It is an unbiased estimator for the expectation of MSE in (23):

$$R_0 = \mathbb{E}[MSE] = \mathbb{E} \left[ \hat{R}_0 \right]$$  \hspace{1cm} (27)

For image denoising purpose, we regard $\hat{R}_0$ as a reliable estimate of the MSE for optimization, as the total number of pixels, $HL$, in an image is usually a very large number.

### 3.2.3 Vector bilateral filtering parameter optimization

By substituting the $f(I_{in}(s), \theta)$ in eq.(25) with the proposed vector bilateral function $f_{\text{vec-bilateral}}(I_{in}(s), (\sigma_d, \Sigma_d))$ in eq.(15), we obtain an expression of SURE for the proposed vector bilateral filter [33], [34]:

$$\hat{R}_{(\sigma_d, \Sigma_d)} = \frac{1}{HL} \sum_{s \in \Gamma} \left[ \left\Vert I_{in}(s) - I_{out,\sigma_d,\Sigma_d}(s) \right\Vert^2 - \text{Tr}(\Gamma_n) + 2 \frac{1}{HL} \sum_{s \in \Gamma} \text{Tr} \left( \Gamma_n J_{f_{\text{vec-bilateral}}} \left( I_{in}(s) \right) \right) \right]$$  \hspace{1cm} (28)
The parameter set \((\sigma_d, \Sigma_s)\) corresponds to \(\theta\) in Eq. (22). A closed form expression in terms of observed signal \(I_{in}(s)\) and parameters \((\sigma_s, \Sigma_s)\) for the gradient vector \(J_{vec-bilateral}(I_{in}(s))\) is derived as:

\[
J_{vec-bilateral}(I_{in}(s))_{i,j} = \frac{\partial f_{vec-bilateral,i}(I_{in}(s), \theta)}{\partial I_{in,j}(s)} = \frac{\sum_{p \in \Omega} g_d(|p-s|, \sigma_d) \exp\left( -\frac{1}{2} |I_{in}(s) - I_{in}(p)|^T \Sigma_s^{-1} (I_{in}(s) - I_{in}(p)) \right) I_{in,j}(s)}{\sum_{p \in \Omega} g_d(|p-s|, \sigma_d) \exp\left( -\frac{1}{2} |I_{in}(s) - I_{in}(p)|^T \Sigma_s^{-1} (I_{in}(s) - I_{in}(p)) \right)} \]

\[
= \frac{\sum_{p \in \Omega} A(p) I_{in,i}(s)}{\sum_{p \in \Omega} A(p)} \frac{\sum_{p \in \Omega} \delta_{i,j}}{\sum_{p \in \Omega} A(p)} \left( \sum_{p \in \Omega} A(p) \right)^2 \frac{\sum_{p \in \Omega} A(p) I_{in,i}(s)}{\sum_{p \in \Omega} A(p)} \frac{\sum_{p \in \Omega} A(p) I_{in,j}(s)}{\sum_{p \in \Omega} A(p)} \left( \sum_{p \in \Omega} A(p) \right)^2 \]

\[
= \frac{\sum_{p \in \Omega} A(p) \left( I_{in}(p) - I_{in}(s) \right)^T \left( \frac{1}{2} \Sigma_s^{-1} + \Sigma_s^{-1} \right) I_{in,j}(s) + \delta_{i,j}}{\sum_{p \in \Omega} A(p)} \frac{\sum_{p \in \Omega} A(p) I_{in,i}(s)}{\sum_{p \in \Omega} A(p)} \frac{\sum_{p \in \Omega} A(p) I_{in,j}(s)}{\sum_{p \in \Omega} A(p)} \left( \sum_{p \in \Omega} A(p) \right)^2 \]

(29)

where \(\delta_{i,j}\) is the delta function

\[
\delta_{i,j} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}
\]

(30)

and \(A(p)\) is defined as:
The noise covariance matrix $\Gamma_n$ can either be measured from sensor calibration data or be estimated efficiently with the median absolute deviation method [35], [36]. The diagonal terms of the noise covariance matrix $\Gamma_n(j, j)$ is estimated with [36]:

$$\hat{\Gamma}_n(j, j) = \left(1.4826 \text{median} \left( I_{w_j} - \text{median}(I_{w_j}) \right) \right)^2$$

For the off-diagonal terms of the estimated noise covariance matrix $\hat{\Gamma}_n(i, j)$:

$$\hat{\Gamma}_n(i, j) = \frac{1.4826}{4ab} \left( \text{median} \left( (aI_{w_i} + bI_{w_j}) - \text{median}(aI_{w_i} + bI_{w_j}) \right)^2 \right)$$

$$a = \left( \hat{\Gamma}_n(i, i) \right)^{-1/2}, b = \left( \hat{\Gamma}_n(j, j) \right)^{-1/2}$$

As $\hat{R}_s$ is analytically defined and can be computed numerically, we pose the bilateral filter parameter optimization problem as a constrained optimization defined as:

$$\begin{align*}
\sigma_{d, opt}, \Sigma_{s, opt} = \arg\min_{\sigma_d, \Sigma_s} \hat{R}_s(\sigma_d, \Sigma_s), \\
\text{s.t. } \sigma_d > 0, \Sigma_s \succeq 0
\end{align*}$$

This constrained non-linear optimization is solved numerically using the sequential quadratic programming (SQP) method. In essence, the solution to Eq.(34) maximizes the signal to noise ratio since the risk is an estimate of the noise power.

Our optimized vector bilateral filter algorithm is summarized below:

**Algorithm:** Optimized vector bilateral filter algorithm (OVBF)
Given the input vector image $I_{in}$

**Step 1.** Initialization: Estimate $\Gamma$, set initial parameters $\sigma, \Sigma$. Set start $t = 0$. Set maximum iteration number $t_{max}$ and stop threshold $\varepsilon$.

**Step 2.** Iteration: do

a) Calculate $R_t = \hat{R}_{(\sigma_{x,t}, \Sigma_{x,t})}$ by (28)

b) Calculate $I_{out}$ by (15)

c) $t = t+1$

d) Update $\sigma_{x,t}$ with SQP

e) Update $\Sigma_{x,t}$ with SQP

f) Calculate $R_{t+1} = \hat{R}_{(\sigma_{x,t+1}, \Sigma_{x,t+1})}$

while ($t < t_{max}$ && $|R_{t+1} - R_t|/ R_{t+1} > \varepsilon$)

**Step 3.** Output optimal $I_{out}$ with minimal $\hat{R}_{(\sigma_x, \Sigma_x)}$

The SQP calculation can be implemented with MATLAB function `fmincon`. As the calculation of vector bilateral filter output $I_{out}$ and risk estimation $\hat{R}_{(\sigma_x, \Sigma_x)}$ share common elements, their computation can be incorporated together to increase efficiency. For optimized MATLAB implementation, the computation of vector bilateral filter on a $256 \times 256 \times 3$ image takes 5s on a Dell XPS laptop with 2.4GHz Intel i7 processor, while the risk estimation takes additional 3s.
3.3 Experimental Results and Discussion

3.3.1 Experiment design

To compare and analyze the performance of the proposed approach, Monte Carlo simulation experiments were carried out on ground truth multispectral images which were contaminated with simulated multivariate zero mean additive Gaussian noise. The noise covariance matrices were varied to generate different test images. For each noise type, the same numerical experiment was repeated twenty times and the results were averaged to ensure a reliable quantitative assessment. The denoising performances were evaluated with the widely used PSNR metric

$$PSNR = 20 \log_{10} \left( \frac{2^{bitdepth} - 1}{\sqrt{MSE}} \right) dB.$$  \hspace{1cm} (35)

where $bitdepth$ refers to the maximum ground truth image bit depth across all bands, $MSE$ is defined in Eq.(23).

3.3.1.1. Experiment ground truth targets

Instead of randomly selecting specific natural scene images as ground truth images, we employed the dead leaves model to generate synthetic images that are more representative of the overall natural image statistics. The original dead leaves model is proposed in [37], where it has been demonstrated that specific dead leaves models can reproduce most known statistics of natural images [38]-[40]. We generated the multispectral image test target with modified code for grayscale dead leaves target\(^1\).

---

\(^1\) Available at http://www.mathworks.com/matlabcentral/fileexchange/16201-toolbox-image/content/toolbox_image/compute_dead_leaves_image.m, Courtesy of Gabriel Peyré
Circle objects with uniformly random distributed size and gray level in each band are arbitrarily placed and occluded in a fixed size blank image with a looking up manner. By this means, the combination space of contrast and texture size change is well covered with enough samples. In our experiment, the PSNR results are also averaged over twenty synthetic implementations of the color dead leaves model to reinforce complete coverage of the sampling space. Four examples of the synthetic color dead leaves images with size of $256 \times 256 \times 3$ are shown in Fig.4.

*Figure. 4 Synthetic color dead leaves image examples*

We have also conducted experiments on several real world multispectral images extracted from Hyperspectral Digital Imagery Collection Experiment (HYDICE) set
3.3.1.2. **Comparisons Method**

Two denoising algorithms are used to compare with the proposed optimized vector bilateral filter [34]:

1) The Vector SURE-LET Multichannel image denoising algorithm [31], for which good results have been reported and hence is a good reference for evaluation. The general MAP multivariate SURE denoising estimator proposed in [30] reported slighter better performance than the SURE-LET implementation with average SNR advantage around 0.2 and maximum SNR difference around 0.8, but the neither the code or the test data are available. Considering the small difference in performance shown in the results of [30], these two algorithms can be regarded to be close in performance. To be fair, we would like to mention the SURE-LET implementation has the potential of additional performance gain with undecimated wavelet transform.

2) The optimal 2D bilateral filter [43] applied separately to each channel of the multispectral image. This reference will demonstrate the advantage of proposed vector bilateral filter when the number of available bands increases.

---

2 Available at: [http://bigwww.epfl.ch/demo/suredenoising-color/index.html#soft](http://bigwww.epfl.ch/demo/suredenoising-color/index.html#soft), Courtesy of Florian Luisier et al.’s
3.3.2. Experiment test case results and discussion

3.3.2.1. White noise image test case

To illustrate how the proposed method compares to other methods for different values of noise power, one hundred fifty 256×256×3 8 bit color dead leaves targets are generated, and corrupted with zero-mean white Gaussian noise of different uniform noise covariance matrices \( \Gamma_{\sigma} = \sigma^2 I_3 \), where \( \sigma \) changes from 20 to 100 and \( I_3 \) denotes the 3×3 identity matrix. The obtained average results are provided in Table. 2. For color dead leaves targets, as for each test instances, the test target images are generated differently as shown in Figure. 4, the test results will have big variation. An optimal test set number needs to be decided to get representative statistical results while avoiding wasting test time. The selected test set number of 150 is decided with experiment.

For Table. 2, sigma = 50 test cases, we repeated 200 times of test, and calculate the moving average of 5, 20 and 150 samples respectively for all three test methods, the averaging trend lines figures are shown below (only first 50 moving average results are shown as moving average of 150 samples can only be carried out on first 50 sample points). Although the 20 moving average shows very close trend to stable output, there are still some small variation. The moving average of 150 samples shows good stable result. Similar trend is also observed for sigma = 100 test cases. So we conduct experiments for color dead leaves test with 150 times average.
Figure 5. Color dead leaves target moving average comparison result for 2D bilateral filter

Figure 6. Color dead leaves target moving average comparison result for SURE-LET
Figure. 7. Color dead leaves target moving average comparison result for vector bilateral filter

Table. 2 Comparison on color dead leaves model (same noise level in each band)

<table>
<thead>
<tr>
<th>Standard deviation of noise $\sigma$ per channel</th>
<th>Input PSNR (dB)</th>
<th>Component wise Optimal 2D bilateral filter PSNR[43]</th>
<th>Vector SURE-LET PSNR [31]</th>
<th>Proposed Optimal Vector Bilateral Filter PSNR [34]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma = 20$</td>
<td>22.11</td>
<td>28.87</td>
<td>26.76</td>
<td>32.74</td>
</tr>
<tr>
<td>$\sigma = 30$</td>
<td>18.59</td>
<td>25.62</td>
<td>23.97</td>
<td>27.62</td>
</tr>
<tr>
<td>$\sigma = 40$</td>
<td>16.08</td>
<td>23.24</td>
<td>22.97</td>
<td>24.49</td>
</tr>
<tr>
<td>$\sigma = 50$</td>
<td>14.15</td>
<td>21.89</td>
<td>21.60</td>
<td>22.72</td>
</tr>
<tr>
<td>$\sigma = 100$</td>
<td>8.13</td>
<td>18.34</td>
<td>18.26</td>
<td>18.45</td>
</tr>
</tbody>
</table>

Judging from quantitative PSNR measure, the proposed vector bilateral filter outperforms the comparison methods clearly throughout most of the noise strength.
range. In the case of medium noise strength (\( \sigma \) ranges from 20 to 50), the advantage is around 1 dB gain. With severe noise presence (\( \sigma \) up to 100), the advantage narrows to around 0.1dB. This narrowing difference trend can be explained by the following observation: if the noise variance is really large, the contamination will strongly demolish weak part of the original signal to an unrecoverable point across all band images, the original PSNR will become so low that signal can hardly be distinguished from noise with any denoising method. Hence the denoising algorithm will have very similar performance. This trend is illustrated in Figure. 8. As the results are based on the average of many different texture image instances, i.e. different combination of dead leaf patterns and noise realizations, this figure should well represent the denoising performance of different methods on typical noisy texture images. And it evidently shows the recognizable advantage of proposed method in medium and low noise situation.

![Denoising performance trend](image)

*Figure. 8 Hyperspectral image denoising performance trend*
One real color image rendering example is provided in Figure 10 for visual inspection. An instance of the color dead leaves model is corrupted with medium noise strength of $\sigma$ equals 30. It is shown in Figure 10 that the propose vector bilateral filter preserved small details better. More small circle objects can be discerned from the denoising results compared to other methods. And the boundaries of the vector bilateral filter rendering are more naturally smoothed, while ringing effect can be inspected in the vector SURE-LET results.

The accuracy of derived $\hat{\mathbf{R}}_n$ is also verified on the same signal and noisy image examples shown in Figure 11. For each iteration in the parameter optimization process, we computed derived $\hat{\mathbf{R}}_n$ with real noise covariance matrix and estimated noise covariance matrix respectively. Corresponding $MSE$ values are also calculated for the denoising results at each iteration. As shown in figure, the calculated $\hat{\mathbf{R}}_n$ with either real noise covariance matrix or estimated noise covariance matrix is very close to the $MSE$ value. Thus the derived $\hat{\mathbf{R}}_n$ is a good indicator for $PSNR$ optimization, which in this case improves from 24.08 dB to 27.83 dB.

Similar test is carried out on a multispectral subset of eleven band images (ranging from 450nm to 650nm with 20 nm interval in the visible wavelength), which is extracted from the clear 8 bit target HYDICE image and corrupted with zero-mean additive white Gaussian noise of different noise covariance matrices $\Gamma_{\chi} = \sigma^2 I_{11}$. $\sigma$ varies from 20 to 100 and $I_{11}$ denotes the $11 \times 11$ identity matrix. The results are shown in Table 3. For this test, we average the test results over 20 noise instances. As the target signal image is fixed and only the added noise instance changes, we
decided from experiment that 20 is a sufficient number to get stable results. For Table. 3, sigma = 100 test cases, we repeated 150 times of test, and calculate the moving average of 5, 20 and 100 samples respectively for three test cases, as shown in Figure. 9 below, the test results will not change much no matter how many times of test is carried out to get the final average result. As the target clear image is fixed (unlike color dead leaves test case), only additive noise instances changes do not cause much change in denoising results. The variance of individual samples is small in scale, and 20 is more than sufficient to get stable test results. We also verified with sigma = 50 cases, and observed similar trend.

![Moving average number comparison](image)

*Figure. 9. HYDICE image moving average comparison results*
Table. 3 Comparison on HYDICE urban image (same noise level in each band)

<table>
<thead>
<tr>
<th>Standard deviation of noise σ per channel</th>
<th>Input PSNR (dB)</th>
<th>Component wise Optimal 2D bilateral filter[43]</th>
<th>Vector SURE-LET [31]</th>
<th>Proposed Optimal Vector Bilateral Filter [34]</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ = 20</td>
<td>22.11</td>
<td>30.18</td>
<td>27.83</td>
<td><strong>32.91</strong></td>
</tr>
<tr>
<td>σ = 30</td>
<td>18.59</td>
<td>27.33</td>
<td>24.41</td>
<td><strong>30.73</strong></td>
</tr>
<tr>
<td>σ = 40</td>
<td>16.08</td>
<td>25.13</td>
<td>21.97</td>
<td><strong>29.42</strong></td>
</tr>
<tr>
<td>σ = 50</td>
<td>14.15</td>
<td>23.33</td>
<td>20.07</td>
<td><strong>28.57</strong></td>
</tr>
<tr>
<td>σ = 100</td>
<td>8.13</td>
<td>17.55</td>
<td>14.10</td>
<td><strong>25.78</strong></td>
</tr>
</tbody>
</table>

The quantitative results show same trend as synthetic color image results. The advantage of the vector bilateral filter is stronger in this HYDICE image as more band images are utilized.
Figure 10 Denoising experiment results of color dead leaves image:

(a) Original signal color image (b) Noisy image (Initial PSNR = 18.59 dB) (c) Result of proposed optimal vector bilateral filter (PSNR = 27.62 dB) (d) Result of component wise optimal 2D bilateral filtered (PSNR = 25.62 dB) (e) Result of vector SURE-LET (PSNR = 23.97 dB)
3.3.2.2. Colored noise test case

As the noise in the multispectral image is not always white, we also compare all the proposed algorithms for zero mean additive colored Gaussian noise with covariance matrix:

\[
\Gamma_{\nu,\delta} = \begin{bmatrix}
1 & 0.5 & 0.5 \\
0.5 & 1 & 0.5 \\
0.5 & 0.5 & 1
\end{bmatrix}
\]

is added to targeting color dead leaves images. The comparison is shown in Table. 4. The same conclusions as in the white Gaussian noise case can be drawn.
Table. 4 Comparison on color dead leaves model (color noise with noise correlation)

<table>
<thead>
<tr>
<th>Noise covariance matrix $\Gamma_n$</th>
<th>Input PSNR (dB)</th>
<th>Component-wise Optimal 2D bilateral filter [43]</th>
<th>Vector SURE-LET [31]</th>
<th>Proposed Optimal Vector Bilateral Filter [34]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Gamma_{n,4}$</td>
<td>18.6</td>
<td>24.98</td>
<td>24.79</td>
<td>28.71</td>
</tr>
</tbody>
</table>

Repeat the test on target HYDICE image with noise covariance matrix:

$$
\Gamma_{n,5} = \begin{bmatrix}
1 & 0.5 & L & 0.5 \\
0.5 & 1 & L & 0.5 \\
M & M & 0 & M \\
0.5 & 0.5 & K & 1 \\
\end{bmatrix}_{11 \times 11}
$$

The result is shown in Table. 5.

Table. 5 Comparison on HYDICE urban image (color noise with noise correlation)

<table>
<thead>
<tr>
<th>Noise covariance matrix $\Gamma_n$</th>
<th>Input PSNR (dB)</th>
<th>Component-wise Optimal 2D bilateral filter [43]</th>
<th>Vector SURE-LET [31]</th>
<th>Proposed Optimal Vector Bilateral Filter [34]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Gamma_{n,5}$</td>
<td>18.6</td>
<td>27.32</td>
<td>24.32</td>
<td>29.81</td>
</tr>
</tbody>
</table>
3.3.2.3. Increased number of bands case

To show how the number of bands will impact the performance of different algorithms on complex image, two experiments are conducted. In the first test, one grayscale dead leaves image is created and duplicated with random scale to multiple bands to be used as target signal image. Gaussian white noise with same noise standard deviation of 30 for each band is added to generate the final test image. In this way, the signal is totally correlated. The results are shown in Table. 6.

Table. 6 Comparison on total correlated dead leaves image (increased number of bands with fixed noise standard deviation of 30 for each band)

<table>
<thead>
<tr>
<th>Number of bands</th>
<th>Input PSNR (dB)</th>
<th>Component wise Optimal 2D bilateral filter[43]</th>
<th>Vector SURE-LET [31]</th>
<th>Proposed Optimal Vector Bilateral Filter [34]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>18.59</td>
<td>25.25</td>
<td>27.42</td>
<td>28.91</td>
</tr>
<tr>
<td>5</td>
<td>18.59</td>
<td>25.25</td>
<td>28.85</td>
<td>30.16</td>
</tr>
<tr>
<td>7</td>
<td>18.59</td>
<td>25.25</td>
<td>29.78</td>
<td>30.84</td>
</tr>
<tr>
<td>9</td>
<td>18.59</td>
<td>25.25</td>
<td>30.45</td>
<td>31.31</td>
</tr>
<tr>
<td>11</td>
<td>18.59</td>
<td>25.25</td>
<td>30.81</td>
<td>31.76</td>
</tr>
</tbody>
</table>

As predicted by theory, the vector bilateral filter shows significant performance improvement (almost 4 dB) as the band numbers increases (from 3 to 11), while the performance of 2D bilateral remains unchanged. It is also noticeable that the vector
SURE-LET approach also considerably benefits from the increasing band number in this idea case.

In the second test, real world situation is examined. For the target HYDICE image, the first 3, 5, 7, 9 and 11 band images are chosen as the target signal images. The correlation between band images are measured by the average correlation coefficients defined as:

\[
C_{\text{mean}} = \text{mean} \left( \sum_{i=1}^{k} \sum_{j=1, j \neq i}^{k} C(i, j) \right)
\]

where \( C(i, j) \) is correlation coefficient between band \( i \) and \( j \). Gaussian white noise with same noise standard deviation of 30 for each band is added for test. In our test signal image, no matter how many band images are selected, there is always considerable amount of correlation between the signal bands images, the average correlation coefficients are all above 99%.

| Table. 7 Comparison on HYDICE urban image (increased number of bands with fixed noise standard deviation of 30 for each band) |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| Number of bands   | Input PSNR (dB)   | Component wise Optimal 2D bilateral filter[43] | Vector SURE-LET [31] | Proposed Optimal Vector Bilateral Filter [34] |
| 3                 | 18.59             | 27.34             | 24.39             | **30.22**         |
| 5                 | 18.59             | 27.34             | 24.40             | **30.43**         |
| 7                 | 18.59             | 27.33             | 24.40             | **30.52**         |
| 9                 | 18.59             | 27.33             | 24.41             | **30.65**         |
| 11                | 18.59             | 27.33             | 24.41             | **30.73**         |
From the results provided in Table. 7, while other methods show little or no improvement, the proposed vector bilateral filter shows again best performance and noticeable performance improvement (about 0.5 dB) with the increase of the number of bands (from 3 to 11 bands). One band image example (at wavelength 490 nm) is also shown in Figure. 12 for visual comparison. It clearly shows that as the total number of bands rises from 3 to 11, for the same individual band, the denoising result is sharper in edges and more small details (Figure. 12.(d)) are recovered. As very often real world multi-spectral images show strong correlation between bands and majority of multi-spectral image bands has high SNR, these two experiments should reflect the nature of real world multi-spectral image denoising well and demonstrate the strong advantage of the proposed vector bilateral filter in denoising multispectral images, especially for hyperspectral images with large number of bands.
(a) Original signal band image at 490nm

(b) Noisy image (PSNR = 18.59 dB)

(c) Optimal vector bilateral filter on 3 band multispectral image (PSNR = 30.22 dB)

(d) Optimal vector bilateral filter on 11 band multispectral image (PSNR = 30.73 dB)

(e) Component wise optimal 2D bilateral filtered on 3 band multispectral image

(f) Component wise optimal 2D bilateral filtered on 11 band multispectral image
3.3.2.5. Waste band image recovery case

We also demonstrate the performance of proposed vector bilateral filter using the whole HYDICE urban hyper spectral image of 210 bands, in which many band images are considered as noisy band images with little signal and strong noise due to atmosphere absorption. The results for noisy band image at wavelength 1946.23 nm (band 151) are shown in Figure. 13. It is clearly shown in visual comparison that the optimized vector bilateral filter successfully recovers important signal features such as edge and prominent small features.
Figure. 13. Experiment results of HYDICE waste band image:

(a) Waste band image at 1946.23 nm (band 151) (b) Result of proposed optimal vector bilateral filter (c) Result of component wise optimal 2D bilateral filtered (only band 151 image is used for denoising) (d) Result of vector SURE-LET
Chapter 4. Nonnegative Matrix Factorization with Deterministic Annealing for Unsupervised Unmixing of Hyperspectral Imagery

4.1 Introduction

A primary problem in unsupervised learning tasks is to find a suitable factorization of the target data matrix. Well known techniques such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) have been developed for this purpose. Recently, a Non-Negative Factorization (NMF) approach was provided [46]-[48]. It was shown to work better in applications where the target data matrix and the linear decompositions are all expected to be non-negative. The major advantage comes from the embedded strong non-negative constraint vs. weak orthogonality and statistical independence constraints of PCA and ICA techniques.

Nevertheless, NMF suffers from two drawbacks, the non-convexity of the objective function and lack of explicit sparseness constraint. Various auxiliary constraints and optimization procedures [49]-[54] have been developed to address these two problems. But no simple general form extension has been proposed to solve these two problems simultaneously.

In this chapter, we propose to tackle the two problems simultaneously with an additive gradually diminishing convex entropy constrain term, i.e. the deterministic annealing [55] procedure, for NMF optimization process. By doing so, we transfer the initial non-convexity optimization problem of NMF into a convex optimization
problem, hence avoiding trapped into shallow local optima in the starting stage as the original NMF often does. At the same time, the added convex entropy constraint also serves as a natural measurement of sparseness with clear physical meaning. And after the proposed NMF optimization procedure reaches near global optima, the fading constraint term helps to find the real solution to the original optimization problem. The contribution here is in showing that deterministic annealing algorithm can be cast into NMF framework and the resulting extended NMF provides superior performance compared to other competing NMF techniques.

4.2. Nonnegative Matrix Factorization with Deterministic Annealing

4.2.1. Non-negative matrix factorization

Very often, we have a large number of non-negative \( d \times 1 \) data vectors \( v \) collected from experiment observations (e.g. hyperspectral imagery). In a matrix factorization framework, these data vectors are arranged into the columns of a \( d \times w \) matrix \( V \), where \( w \) is the number of examples in the data set. Non-negative matrix factorization (NMF) finds non-negative matrix factors such that:

\[
V = \begin{bmatrix} v_1 & v_2 & \cdots & v_w \end{bmatrix} \approx \begin{bmatrix} m_1 & m_2 & \cdots & m_r \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_r \end{bmatrix} = MS (37)
\]

where \( M \) is an \( d \times r \) matrix containing the \( d \times 1 \) basis vectors \( m \) in its columns, \( S \) is an \( r \times w \) matrix holding the \( 1 \times w \) hidden components coefficients \( s \) that controls the
weighting of corresponding basis vector for every data vector in its rows, and \( r \) is a predetermined number that represents how many estimated hidden components exist and is usually chosen to be much smaller than \( d \) or \( w \). Each data vector \( \mathbf{v} \) is approximated by an additive only linear combination of the basic vectors \( \mathbf{m} \), weighted by the coefficients of \( \mathbf{s} \). As only a few basis vectors (\( r \) is small) are used to represent many data vectors, good approximation can only be achieved if the basis vectors discover the latent structure in the data.

To find the optimal approximation, we need to define a cost function that quantifies approximation quality. The most common cost function is the squared error (Euclidean distance) function:

\[
C(M,S) = \frac{1}{2} \| \mathbf{V} - \mathbf{MS} \|_F^2
\]  

Although the function \( C(M,S) \) is convex in \( M \) only or \( S \) only, it is not convex in both factors together. Consequently, a global minimum solution is hard to achieve. Furthermore, NMF suffers from lack of a unique solution. This can be verified with any nonnegative invertible matrix \( \mathbf{D} \) by considering \( \mathbf{MS} = (\mathbf{MD})(\mathbf{D}^{-1}\mathbf{S}) \). As a sparse representation of the data by a limited number of hidden components is very common in many fields such as hyperspectral imagery, statistics, biology etc., various auxiliary sparseness constraints on \( \mathbf{s} \) has been attached to the cost function to help limit the solution space. The cost function with such auxiliary sparseness constraint takes the form as:

\[
C(M,S) = \frac{1}{2} \| \mathbf{V} - \mathbf{MS} \|_F^2 + \lambda \, f(S) \\
\lambda \in \mathbb{R}^+, f(S) \in \mathbb{R}^+
\]  

(39)
where $\lambda$ is a constant scalar that weights the sparseness constraint measure $f(S)$ as a regularization term.

4.2.2. NMF with Deterministic annealing

Many forms of sparseness constraint $f(S)$ have been proposed, such as $L_1$ and $L_{1/2}$ norm of $S$. But entropy of $S$ as a natural measure of sparseness with clear physical meaning has never been explored as sparseness constraint. According to information theory, the entropy of $S$ will reach minimum when the uncertainty associated with $S$ is removed and the sparsest structure of $S$ is achieved.

Without loss of generality, we can constrain the columns of $S$ to sum to unity under the transformation $M \rightarrow MA$, $S \rightarrow A^t S$, where $A$ is a diagonal matrix. In this way, every coefficient element in $S$ can be regarded as a probability measure that ranges from zero to one. The corresponding entropy sparseness constraint function becomes:

$$f_{\text{entropy}}(S) = \| S \otimes \log(S) \| = \sum_{i=1}^{r} \sum_{j=1}^{w} -S_{i,j} \log(S_{i,j})$$

where $\otimes$ denotes element wise multiplication. Due to the nature of entropy, this constraint measurement will reach zero only when there is only one unity in each column of $S$, i.e. the sparsest configuration of hidden components is reached.

We propose a new cost function with entropy constraint as:

$$C(M,S,t) = \frac{1}{2} \| V - MS \|_F^2 + T(t) f_{\text{entropy}}(S)$$

where $t \in \mathbb{R}^+, T(t) \in \mathbb{R}^+$.

In addition to the new entropy sparseness constraint, we also introduce a flexible weighting parameter $T(t)$ which will gradually cool down to zero with the increment of optimization iteration count $t$. By setting $T(0)$ high in the initial iteration, the
entropy constraint term dominates the cost function and the problem can be solved more easily due to the convexity. As the optimization iteration continuous, \( T(t) \) decreases to zero, and the original cost function in (38) dominates and leads to a nonrandom (hard) solution for the primary problem in (37). Thus, the proposed cost function will solve the non-convexity and sparseness constraint problems in the original NMF simultaneously. This cost function form can be regarded as a special realization of deterministic annealing process described in [55], which mimics the annealing process in statistical physics that avoids shallow local minima with beginning high “temperature” and reaches global minima with final low “temperature”.

The new cost function can be optimized with a multiplicative iteration algorithm similar to [47], we derive it on the traditional gradient descent algorithm basis. The gradient of (41) regarding \( M \) and \( S \) are:

\[
\frac{\partial C(M, S, t)}{\partial M} = -VS^T + MSS^T
\]

\[
\frac{\partial C(M, S, t)}{\partial S} = -M^TV + M^TMS - T(t)(\log(S) + 1)
\]  

(42)

The corresponding update rules with traditional element wise gradient decent algorithm are:

\[
M \leftarrow M + \eta_M \otimes (VS^T - MSS^T)
\]

\[
S \leftarrow S + \eta_S \otimes (M^TV - M^TMS + T(t)(\log(S) + 1))
\]

(43)

where \( \otimes \) stands for element wise multiplication, and \( \eta_M, \eta_S \) are step length matrixes.

If we define step lengths matrixes as in [47]:

\[
\eta_M = M \odot (MSS^T), \eta_S = S \odot (M^TMS)
\]

(44)
where \( \odot \) denotes element wise division. Substitute (44) into (43), then the multiplicative update rules are derived as:

\[
M \leftarrow M \odot (VS^T) \odot (MSS^T) \\
S \leftarrow S \odot (M^TV + T(t)(\log(S) + 1)) \odot (M^TMS)
\]  

(45)

When \( T(t) \) diminishes to near zero, these multiplicative update rules are reduced to the original update rule form in [47], which can guarantee the non-negativity of the results and the monotonically non-increasing of Euclidean distance cost function.

Our deterministic annealing based NMF algorithm is summarized below.

**Algorithm:** Deterministic annealing based NMF (DANMF)

Given the observation matrix \( V \) and number of hidden components \( r \)

**Step 1.** Initialization: Randomly set elements of \( M \) and \( S \) in the interval \([0, 1]\).

Rescale each column of \( S \) to unit norm. Set start heating parameter \( T(0) = T_0, t = 0 \).

Set cooling speed parameter \( \alpha \). Set maximum iteration time \( t_{max} \) and stop threshold \( \varepsilon \).

**Step 2.** Iteration: do

a) Calculate \( T(t) = \exp(-t/\alpha) \) \( T_0 \)

b) Calculate \( C_t = C(M,S,t) \) by (41)

c) Update \( M \) by (45)

d) Update \( S \) by (45)

e) Calculate \( C_{t+1} = C(M,S,t+1) \) by (41)

f) \( t = t+1 \)

\[ \text{while}(t < t_{max} \&\& |C_{t+1} - C_t| / C_{t+1} > \varepsilon) \]

**Step 3.** Output results

---

**4.3. Experimental Results and Discussion**
We demonstrate the performance of our DANMF method with hyperspectral unmixing on both synthetic and real world hyperspectral imagery. Two alternative NMF algorithms are used for comparison: The \( L_{1/2} \)NMF [54], for which good results have been reported and hence a good reference for evaluation, and the original NMF [47] as baseline.

Due to the imaging system resolution limitation, the observed data vector from one pixel of hyperspectral imagery may contain a mixture of substance spectra. Therefore, an important preprocessing step, hyperspectral unmixing, is often adopted to decompose a mixed pixel spectral vector into a set of constituent spectra, referred to as \textit{endmember signatures}, and their corresponding \textit{abundance} coefficients [56], [57].

Assuming an additive only linear mixture model and the abundance sum-to-one constraint (ASC), as most hyperspectral unmixing technique do, NMF is a natural solution for this unmixing problem.

The unmixing performances was evaluated with the Spectral Angle Distance(SAD) and Root Mean Squared Error(RMSE) metrics. The SAD is defined as:

\[
SAD_i = \arccos \left( \frac{\hat{m}_i^T \hat{m}_i}{\|\hat{m}_i\|_2 \|m_i\|_2} \right)
\] (46)

where \( m_i \) is the ground truth \( k^{th} \) endmember and \( \hat{m}_i \) is the corresponding NMF estimation. We use \textit{rad} unit for the SAD results. RMSE for related abundance estimation \( \hat{s}_i \) is defined as:

\[
RMSE_i = \left( \frac{1}{w} \|s_i - \hat{s}_i\|_2 \right)^{1/2}
\] (47)
4.3.1 Synthetic Data

To quantitatively compare and analyze the performance of the proposed approach, Monte Carlo simulation experiments were carried out. A 100x100x111 synthetic hyperspectral images were created with six spectral signatures randomly chosen from the United States Geological Survey (USGS) digital spectral library [58] as the endmember signatures (111 bands were pick from 474 bands of original spectra with 4 band skipping from band 20 to band 460). Ground truth abundance coefficients are generated according to Dirichlet distribution under the non-negative and sum-to-one constraint. To simulate the system noise, a multivariate zero mean Gaussian noise was added to the mixture data to achieve a 20dB SNR. The same numerical experiment is repeated five times and the results are averaged to ensure a reliable quantitative assessment. The experiment results are shown in Table. 8 (in rad unit) and Table. 9.

Table. 8 SAD comparison using synthetic data (in rad unit)

<table>
<thead>
<tr>
<th>End member</th>
<th>DANMF</th>
<th>$L_{1/2}$NMF [9]</th>
<th>NMF [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Actinolite</td>
<td>0.3182</td>
<td><strong>0.1859</strong></td>
<td>0.2161</td>
</tr>
<tr>
<td>2. Andradite</td>
<td>0.2645</td>
<td><strong>0.2479</strong></td>
<td>0.2485</td>
</tr>
<tr>
<td>3. Carnallite</td>
<td><strong>0.2954</strong></td>
<td>0.3485</td>
<td>0.3004</td>
</tr>
<tr>
<td>4. Diaspora</td>
<td><strong>0.1949</strong></td>
<td>0.3333</td>
<td>0.2696</td>
</tr>
<tr>
<td>5. Erionite</td>
<td>0.3841</td>
<td>0.4359</td>
<td><strong>0.3609</strong></td>
</tr>
<tr>
<td>6. Halloysite</td>
<td><strong>0.1733</strong></td>
<td>0.4621</td>
<td>0.3903</td>
</tr>
</tbody>
</table>
Table. 9 RMSE comparison with synthetic data

<table>
<thead>
<tr>
<th>End member</th>
<th>DANMF</th>
<th>( L_{1/2\text{NMF}} ) [9]</th>
<th>NMF [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Actinolite</td>
<td>0.2686</td>
<td>0.1669</td>
<td>0.3158</td>
</tr>
<tr>
<td>2. Andradite</td>
<td>0.1121</td>
<td>0.1129</td>
<td>0.1757</td>
</tr>
<tr>
<td>3. Carnallite</td>
<td>0.1251</td>
<td>0.2417</td>
<td>0.2070</td>
</tr>
<tr>
<td>4. Diaspora</td>
<td>0.0610</td>
<td>0.1427</td>
<td>0.2420</td>
</tr>
<tr>
<td>5. Erionite</td>
<td>0.1162</td>
<td>0.1928</td>
<td>0.1755</td>
</tr>
<tr>
<td>6. Halloysite</td>
<td>0.1208</td>
<td>0.2271</td>
<td>0.2966</td>
</tr>
</tbody>
</table>

The spectrum comparison of all six elements are shown as below:
Endmember Spectra 4 Comparison

Reflectance

Wavelength (μm)

Ground truth
Proposed DANMF
LHALF NMF
Original NMF
Endmember Spectra 5 Comparison

Reflectance

Wavelength (μm)

(c)
Figure 14 Synthetic data unmixing spectrum results

DANMF shows most similar shape to ground truth spectrum in most cases which coincides with numerical results.
4.3.2 Real world hyperspectral imagery

We also tested out method on real-world Urban HYDICE hyperspectral image. The image is of size $307 \times 307 \times 210$ with spatial resolution of 2m and spectral resolution of 10nm in the 400nm and 2500nm range. After removing low SNR bands (channels 1-4, 76, 87, 101-111, 136-153, and 198-210), only 162 bands remain. A color rendering of the hyperspectral image is shown in Figure 15. Judging from the color image, there are four distinct targets of interest: tree, roof, grass and asphalt. Figure 15 shows the DANMF estimated abundance map images corresponding to these four targets. The brightness of a pixel denotes the abundance of the end-member under consideration. The DANMF gives out good estimation of the four targets according to the original color rendering. Even small objects like car roofs are presented correctly.
Figure 15 HYDICE image unmixing Experiment results:

(a) Original color rendering image (b) Tree abundance result from DANMF (c) Tree abundance result from $L_{1/2}$NMF (d) Roof abundance result from DANMF (e) Roof abundance result from $L_{1/2}$NMF (f) Grass abundance from DANMF (g) Grass abundance result from $L_{1/2}$NMF (h) Asphalt abundance from DANMF (i) Asphalt abundance result from $L_{1/2}$NMF
4.4 Experimental Results & Discussion for Combined Optimized Vector Bilateral Filter and DANMF

It is reasonable to assume that applying optimized vector bilateral filter before DANMF will help improve performance in many cases. We have repeated previous experiments to verify this assumption.

4.4.1 Synthetic Data

By applying the optimized vector bilateral filter before applying DANMF on the same synthetic hyperspectral image as used in previous test, we have the following results as shown in Table. 10:

*Table. 10 RMSE comparison for synthetic data*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Actinolite</td>
<td>0.2527</td>
<td><strong>0.1934</strong></td>
<td>0.2438</td>
</tr>
<tr>
<td>2. Andradite</td>
<td><strong>0.2422</strong></td>
<td>0.2645</td>
<td>0.2999</td>
</tr>
<tr>
<td>3. Carnallite</td>
<td><strong>0.3253</strong></td>
<td>0.3483</td>
<td>0.2868</td>
</tr>
<tr>
<td>4. Diaspora</td>
<td><strong>0.1962</strong></td>
<td>0.3268</td>
<td>0.2606</td>
</tr>
<tr>
<td>5. Erionite</td>
<td>0.4029</td>
<td>0.4322</td>
<td><strong>0.3515</strong></td>
</tr>
<tr>
<td>6. Halloysite</td>
<td><strong>0.3690</strong></td>
<td>0.4931</td>
<td>0.3943</td>
</tr>
</tbody>
</table>
Table. 11 SAD comparison for synthetic data (in \textbf{rad} unit)

<table>
<thead>
<tr>
<th>End member</th>
<th>DANMF</th>
<th>L1/2NMF \cite{9}</th>
<th>NMF \cite{2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Actinolite</td>
<td>0.1299</td>
<td>0.2978</td>
<td>0.3239</td>
</tr>
<tr>
<td>2. Andradite</td>
<td>0.0722</td>
<td>0.2070</td>
<td>0.2067</td>
</tr>
<tr>
<td>3. Carnallite</td>
<td>0.1203</td>
<td>0.2376</td>
<td>0.1398</td>
</tr>
<tr>
<td>4. Diaspora</td>
<td>0.0544</td>
<td>0.0844</td>
<td>0.1303</td>
</tr>
<tr>
<td>5. Erionite</td>
<td>0.1584</td>
<td>0.1853</td>
<td>0.2526</td>
</tr>
<tr>
<td>6. Halloysite</td>
<td>0.1921</td>
<td>0.3027</td>
<td>\textbf{0.1564}</td>
</tr>
</tbody>
</table>

It is shown that, as a general trend, all three approaches benefit in performance by applying prior optimized vector bilateral filter. SAD performance benefits the most. Correspondingly, the spectrum retrieved from all three approaches in general match closer to ground truth spectrum, Spectrum comparison figures are shown as below:
Endmember Spectra 3 Comparison

![Graph comparing different endmember spectra with various markers and line styles.](c)
Figure 16 Synthetic data denoising and unmixing combined spectrum results

4.4.2 Real world hyperspectral imagery

For previous real world hyperspectral imagery test, we manually disregarded bands with strong noise. In this experiment, we did not remove these bands and apply DANMF without and with prior optimized vector bilateral filter. The image results are shown in following figures.
Figure. 17 HYDICE image combined denoising and unmixing experiment results:

(a) Tree abundance result from OVBF & DANMF combination (b) Tree abundance result from DANMF only (c) Roof abundance result from OVBF & DANMF combination (d) Roof abundance result from DANMF only (e) Grass abundance from OVBF & DANMF combination (f) Grass abundance result from DANMF only (g) Asphalt abundance from OVBF & DANMF combination (h) Asphalt abundance result from DANMF only

It is shown that without prior OVBF, the final results could be very noisy and have little value for some endmembers. The prior OVBF can significantly boost the final results. But still, the manual noise band removal process will deliver best results compared to the OVBF&DANMF combination without any intervening. It could be explained by very strong noisy bands in the test image.
Chapter 5. Conclusion and Future Work

In this dissertation, first a method to find the optimum parameters of the range and distance kernel of a 2D bilateral filter was proposed, and its effectiveness was demonstrated using a synthesized noisy image. The PSNR measured from the artificially generated experimental image shows that the proposed bilateral filter parameter optimization approach can lead to non-trivial improvements in PSNR. Qualitatively, using the optimum parameter provided by the approach results in a good tradeoff between blurring and denoising thus leading to preservation of edges while suppressing noise. The optimization procedure can be implemented numerically using any of a number of algorithms.

Consecutively in this dissertation, we proposed an optimized vector bilateral filter approach based on Stein’s principle for denoising multispectral images corrupted by additive Gaussian noise. The basis for the approach is the viewpoint that the output of the bilateral filter is a nonlinear estimate of the underlying noiseless image. Our experiments show that this method leads to improved performance in both quantitative and visual quality measures compared with the wavelet based multispectral image denoising approach as well as the component wise optimized 2D bilateral filter technique. Furthermore, the proposed method demonstrates amplified performance advantage over component wise approach as the number of bands increases.

For the future work, additional Cramér-Rao bound analysis should be applied to this research. This work can also potentially be extended to other bilateral filter types [11] and the non-Gaussian noise case [44], [45] in future research.
In this dissertation, we have also proposed an extended NMF approach based on deterministic annealing. The basis for the approach is the viewpoint that a diminishing entropy sparseness constraint will help avoid local minima. Our experiments show that this method leads to improved performance in both quantitative and visual inspection measures compared with other NMF techniques. This work can be further extended to other applications such as automatic image feature extraction. Graph cut and genetic algorithms can also be explored in future research regarding hyperspectral image unsupervised unmixing problem.

Finally, we examined the system performance by combining optimized vector bilateral filter and DANMF. Unsurprisingly, it is shown that, in general, better system DANMF results can be achieved with prior applied optimized vector bilateral filter.
Bibliography


