Image Fusion in Remote Sensing

and Quality Evaluation of Fused Images

Frosti Pálsson

Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree
in
Electrical and Computer Engineering

Advisors
Professor Jóhannes R. Sveinsson
Professor Magnús Örn Úlfarsson

PhD Committee
Professor Jóhannes R. Sveinsson
Professor Magnús Örn Úlfarsson
Professor Jón Atli Benediktsson

Opponents
Professor Andrea Garzelli
Professor Paul Scheunders

Faculty of Electrical and Computer Engineering
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, September 2017

Image Fusion in Remote Sensing
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in
Electrical and Computer Engineering

Faculty of Electrical and Computer Engineering
School of Engineering and Natural Sciences
University of Iceland
Hjarðarhagi 2-6
107, Reykjavik
Iceland

Telephone: 525-4000

Bibliographic information:
Frosti Pálsson, 2017, *Image Fusion in Remote Sensing*,
PhD dissertation, Faculty of Electrical and Computer Engineering, University of Iceland

ISBN 978-9935-9383-1-2

Printing: Háskólaprent
Reykjavik, Iceland, September 2017

Abstract

In remote sensing, acquired optical images of high spectral resolution have usually a lower spatial resolution than images of lower spectral resolution. This is due to physical, cost and complexity constraints. To make the most of the available imagery, many image fusion techniques have been developed to address this problem. Image fusion is an ill-posed inverse problem where an image of low spatial resolution and high spectral resolution is enhanced in spatial-resolution by using an auxiliary image of high spatial resolution and low spectral resolution. It is assumed that both images display the same scene and are properly co-registered. Thus, the problem is essentially to transfer details from the higher spatial resolution auxiliary image to the upscaled lower resolution image in a manner that minimizes the spatial and spectral distortion of the fused image. The most common image fusion problem is pansharpening, where a multispectral (MS) image is enhanced using wide-band panchromatic (PAN) image. A similar problem is the enhancement of a hyperspectral (HS) image by either a PAN image or an MS image. As there is no reference image available, the reliable quantitative evaluation of the quality of the fused image is a difficult problem. This thesis addresses the image fusion problem in three different ways and also addresses the problem of quantitative quality evaluation. The contributions of the thesis are the following.

- A method for the general fusion problem. It based on developing a pixel-wise *maximum a posteriori* (MAP) estimator, which is derived using a forward model for the image to be enhanced, and the assumption of the joint Gaussianity of the estimated image and the auxiliary image. By employing principal component analysis (PCA), the dimensionality of the problem can be reduced significantly, enabling efficient and robust estimation.

- A pre-processing method to enhance the fusion quality of the two largest classes of pansharpening methods, i.e., the component substitution (CS) methods and the multi-resolution analysis (MRA) methods. These methods can be described using general detail-injection schemes where details obtained from a PAN image are modulated and added to the interpolated bands of the MS image. By using a Wiener filter based deblurring on the interpolated bands, the fusion quality of CS methods, in particular, can be enhanced.

- A method based on training a three dimensional convolution neural network (3D-CNN) to fuse MS and HS images. To make the problem computationally feasible, dimensionality reduction of the HS image using PCA is a key ingredient in the method.

- Finally, the problem of quantitative quality evaluation is addressed. It is shown that the verification of the consistency property of the fused image can be used

instead of the conventional synthesis at reduced resolution protocol that is widely used. The main argument for using consistency is that when using the synthesis protocol, the fusion of images of reduced resolution does not give a good estimate of the fusion at the full resolution scale.

Ágrip

Í fjarkönnun hafa myndir með háa rófsupplausn lægri rúmupplausn en myndir með lægri rófsupplausn vegna eðlisfræðilegra og kostnaðarlegra takmarkana. Til að auka upplýsingamagn slíkra mynda hafa verið þróaðar fjölmargar sambræðsluaðferðir á síðustu tveimur áratugum. Myndsambræðsla er illa framsett andhverft vandmál (e. inverse problem) þar sem rúmupplausn myndar af hárri rófsupplausn er aukin með því að nota upplýsingar frá mynd af hárri rúmupplausn og lægri rófsupplausn. Það er gert ráð fyrir að báðar myndir sýni nákvæmlega sama landsvæði. Þannig er vandamálið í eðli sínu að flytja fíngerða eiginleika myndar af hærri rúmupplausn yfir á mynd af lægri rúmupplausn sem hefur verið brúuð upp í stærð hinnar myndarinnar, án þess að skerða gæði rófsupplýsinga upphaflegu myndarinnar. Algengasta myndbræðsluvandamálið í fjarkönnun er svokölluð panskerpun (*e. pansharpening*) þar sem fjölrásamynd *(e. multispectral image*) er endurbætt í rúmi með svokallaðri víðbandsmynd (*e. panchromatic image*) sem hefur aðeins eina rás af hárri upplausn. Annað svipað vandamál er sambræðsla háfjölrásamyndar (*e. hyperspectral image*) og annaðhvort fjölrásamyndar eða víðbandsmyndar. Þar sem myndsambræðsla er andhverft vandmál er engin háupplausnar samanburðarmynd tiltæk, sem gerir mat á gæðum sambræddu myndarinnar að erfiðu vandamáli. Í þessari ritgerð eru kynntar þrjár aðferðir sem taka á myndsambræðlsu og einnig er fjallað um mat á gæðum sambræddra mynda, þá sérstaklega panskerptra mynda. Framlög ritgerðarinnar eru eftifarandi:

- Aðferð fyrir sambræðslu á sínu almennasta formi. Hún er byggð á því að þróa *maximum a posteriori* (MAP) metil fyrir hverja myndeiningu, sem er leiddur út frá einföldu líkani fyrir mynd sem skal skerpa, og þeirri staðhæfingu að metna háupplausnarmyndin og utanaðkomandi háupplausnarmyndin séu sameiginlega normaldreifðar. Með því að beita meginþáttagreiningu (PCA) á líkanið, er hægt að lækka vídd vandamálsins umtalsvert og þar með gera aðferðinna afar skilvirka og ónæma fyrir suði.

- Aðferð til að bæta sambræðslugetu aðferða sem byggja á að setja þætti í staðinn (*e. component substitution methods*) og fjölupplausnaraðferða (*e. multi resolution analysis methods*), sem eru tveir meginflokkar aðferða í panskerpun. Þessum aðferðum er hægt að lýsa með einföldum líkönum þar sem fíngerðir eiginleikar PAN myndarinnar eru mótaðir með svokölluðum innspýtingar-stuðli og lagðir saman við bönd MS myndarinnar sem hefur verið brúuð (*e. interpolated*) upp í stærð víðbandsmyndarinnar. Með því að nota Wienersíu til eyða óskerpu (*e. blurring*) brúuðu fjölrásamyndarinnar er hægt að bæta getu þessara aðferða umtalsvert, og þá sérstaklega þáttarstaðgengdaraðferða.

- Aðferð sem byggist á að þjálfa þrívítt földunartauganet (*e. 3D convolutional neural network*) með leiðbeindum lærdómi, til að bræða saman fjöl- og háfjöl-

rásamyndir. Til að gera aðferðina reiknilega framkvæmanlega er notuð megin-þáttagreining á HS myndinni.

- Að lokum er fjallað um mat á gæðum sambræddra mynda. Þar er rökstutt að staðfesting á samkvæmnieiginleika (*e. consistency property*) sambræddra mynda sé betri aðferð en að notast við uppbyggingu við skerta upplausn (*e. synthesis at reduced resolution*) sem er hin hefðbundna aðferð. Megingalli hennar er að þar er verið að meta sambræðslu mynda sem eru skertar útgáfur af frumgögnunum og gefur því ekki eins gott mat á sambræðslugetu mismundandi aðferða og staðfesting á samkvæmni.

Acknowledgments

First and foremost, I would like to thank my supervisors, professors Jóhannes and Magnús, for their guidance, support and last but not least, patience. I would also like to thank professor Jón Atli Benediktsson for his support. I thank my opponents, professor Andrea Garzelli and professor Paul Scheunders, for their insightful comments and observations during my defense. Finally, I want to thank the Research Fund of University of Iceland for Doctoral Studies for their financial support.

I have had the opportunity to work with many people during the last years and I have enjoyed their company and friendship. Thanks go to Jakob, Behnood, Nicola, Kyriaki, Gabriele, Pedram and Fadi.

At last, I would like to thank my family for their great support through the years; my brother Burkni, my parents, my daughter Lena, and my girlfriend, Yvonne.

Contents

List of Figures

List of Tables

Abbreviations

CS Component Substitution

DWT Discrete Wavelet Transforsm

ERGAS Erreur Relative Globale Adimensionelle de Synthese

HS Hyperspectral Image

MAP Maximum a Posteriori

MRA Multiresolution Analysis

MS Multispectral Image

MTF Modulation Transfer function

PAN Panchromatic Image

PCA Principal Component Analysis

QNR Quality No Reference

SAM Spectral Angle Mapper

UDWT Undecimated Discrete Wavelet Transform

UIQI Universal Image Quality Index

Notations

$\boldsymbol{X}$	all matrices (images) are in upper case bold
$\boldsymbol{x}$	all vectors are in lower case bold
$\mathbf{Z}$	denotes the fused (estimated) image.
$\boldsymbol{Y}$	denotes the MS image
$\boldsymbol{Y}_k$	denotes the kth band of an image
$\tilde{\boldsymbol{Y}}$	denotes the MS image, interpolated to the PAN scale
$\boldsymbol{X}$	denotes the higher resolution image, i.e. PAN for pansharpening, MS image for MS/HS fusion
$\boldsymbol{H}$	spatial degradation operator
$\boldsymbol{\mathcal{G}}$	Matrix in wavelet domain
g_k	injection gain for CS/MRA methods
w_k	intensity image coefficients for CS methods
N	number of pixels in a reflectance band
N_b	number of bands in an MS/HS image

Chapter 1

Introduction

In this chapter, an introduction to pixel level fusion of optical remote sensing imagery is given. There are three kinds of fusion that are discussed in this thesis. Firstly, the fusion of panchromatic (PAN) images and multi-spectral (MS) images, which is usually referred to as pansharpening. Secondly, the fusion of PAN images and hyperspectral (HS) images, referred to as hypersharpening and thirdly, there is the fusion of MS images and HS images, which will be referred to as MS/HS fusion. In all three cases, the objective of the fusion is to enhance a low spatial resolution multi-band image using a high spatial resolution single band PAN image or a high resolution MS image. The thesis does not only focus on new methods for image fusion, but also on pre-processing of the data in order to enhance existing methods and the quantitative quality evaluation of the fused product. The chapter starts with an introduction to the types of images involved. Then a short discussion on the image fusion problem in general as an inverse problem. Next, an overview of the three fusion cases is given and then the quantitative evaluation of the fused images is discussed. Finally, the thesis contributions and organization is summarized along with a list of publications.

1.1 The data - MS, HS and PAN images

A typical modern optical earth imaging satellite has two kinds of sensors mounted on the same platform, i.e., an MS sensor and a PAN sensor. The PAN sensor is sensitive to light from a wide range of wavelengths, i.e., a single band, while the MS sensor is sensitive to light from much narrower ranges of wavelengths, e.g., from the visible red, green and blue (RGB) wavelengths and also near infrared (NIR). Since the MS sensor captures light from narrower ranges of wavelengths than the PAN sensor, while having the same instantaneous field of view (IFOV), it has a lower spatial resolution in order to preserve the signal-to-noise ratio (SNR) [1]. This is the primary reason that the MS sensor captures images that have a lower spatial resolution than the PAN sensor. Another factor is the cost and complexity of a high resolution MS sensor and there is also the issue of the great amount of data that needs to be transferred back to Earth.

In Earth imaging satellites such as IKONOS or QuickBird, the MS sensor has 4 bands, RGB and NIR, and the PAN sensor has typically 4 times higher spatial resolution than the MS sensor. Newer platforms such as the recently launched WorldView-3 satellite, have MS sensors with up to 8 bands of 1.24m resolution and PAN sensors of 0.31m resolution. The resolution is typically given as ground sample distance (GSD), which is the distance between two adjacent pixel centers. The larger the GSD, the less the spatial resolution. The spectral resolution of a sensor is measured by how narrow its spectral bandwidth is, i.e., the higher number of distinct spectral bands, the higher the

spectral resolution. This also means, that spectral and spatial resolution of a sensor are complementary, i.e., the higher the spectral resolution of a given sensor, the lower the spatial resolution and vice versa. Figure 1.1 shows the relative spectral response of the IKONOS satellite's MS and PAN sensors. There are 4 MS bands, which together span a range of wavelengths which is narrower than that of the PAN sensor.

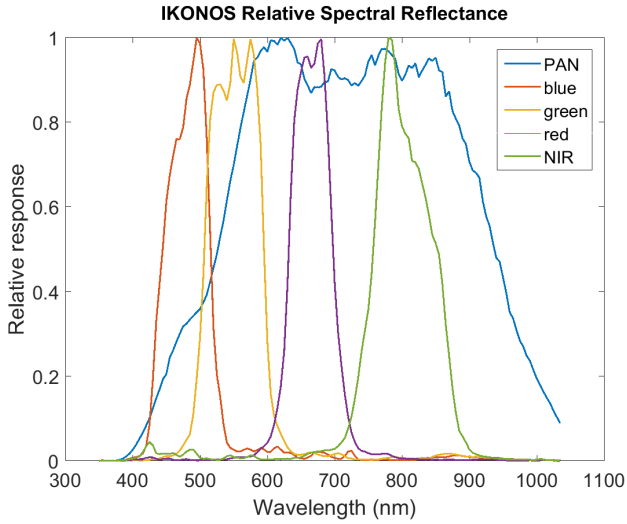


Figure 1.1: Relative spectral response of the IKONOS MS sensor and PAN sensor.

Another type of instrument is the HS sensor. HS images are often acquired using an airborne sensor, while space borne sensors are also used. A typical HS sensor has hundreds of spectral bands. Due to the high spectral resolution, an HS sensor has typically much lower spatial resolution than PAN and MS sensors.

1.2 Image Fusion in Remote Sensing

Image fusion is an ill-posed inverse problem. The forward model is typically formulated such that the observed data equals some system matrix times the ideal high resolution image one wants to estimate plus some zero-mean Gaussian noise. This is basically an under-determined regression problem since the signal to be estimated has a significantly higher dimension than the observed data. Three fusion cases are considered in the thesis.

1. Pansharpening, i.e., fusion of a low spatial resolution MS image and high spatial resolution PAN image.
2. Hypersharpening, i.e., the fusion of a low spatial resolution HS image and a high spatial resolution PAN image.

3. MS/HS fusion, i.e., the fusion of a low spatial resolution HS image and high spatial resolution MS image.

In pansharpening, the higher resolution image has a lower spectral resolution than, e.g., in MS/HS fusion. In hypersharpening, the lower spatial resolution HS image has a higher spectral resolution than the MS image in the pansharpening case. The only difference between these problems is the dimension of the observed data and the various approaches to solve the problem just impose different structure or conditions on the signal to be estimated. This is important, since every fusion algorithm is essentially solving an under-determined inverse problem.

In the following discussion, it will be shown that the main difference between the three fusion cases and the approaches used to solve them, lies in the structure imposed on the high resolution image to be estimated.

1.3 Pansharpening

Since a typical Earth imaging satellite has both MS and PAN sensors mounted on the same platform, these images are acquired simultaneously such that both images show the same scene. Consequently, there is no need to finely co-register the MS and PAN images. To make the most of the available data, the MS image is typically enhanced using the PAN image such that the resulting MS image has the same spatial resolution as the PAN image. This process of fusing the MS and PAN images is called pansharpening. The pansharpened image should, ideally, have the spectral resolution of the MS image and the spatial resolution of the PAN image.

One of the first uses of pansharpening was in the 1987 paper [2], where the intensity-hue-saturation (IHS) transform was used to fuse SPOT-HRV PAN images and Landsat TM MS images. In [3], the high pass filtering (HPF) method was used for the same purpose.

Today, pansharpening is an important technique and is used to produce the imagery seen in the popular Google Maps/Earth™ and Microsoft Bing Maps™ products. There are also several applications within the field of remote sensing that benefit from pansharpened imagery, e.g., change-detection [4] and classification [5].

Since the late eighties, a large number of pansharpening methods have been proposed using many and diverse techniques and there have been several attempts to categorize the methods into various groups such as in [1,6–12]. For example, in [10], pansharpening methods are categorized into spectral and spatial approaches and methods based on the super resolution paradigm (spectral-spatial). In [9], methods are divided into 5 groups, i.e., component substitution (CS), numerical and statistical methods, modulation based methods, multiresolution analysis (MRA) and hybrid techniques. In [11], the methods are broadly assigned to either linear combination approximation (LCA)

methods and spatial filter approximation (SFA) methods. In [8], the methods are categorized into CS and MRA methods. From the above discussion, it is clear that there is no consensus on how to categorize the various pansharpening methods. In the thesis, methods will be categorized into the following 4 families of methods:

1. Component Substitution (CS) methods.
2. Multiresolution Analysis (MRA) methods.
3. Model-based methods.
4. Deep Learning (DL) based methods.

Model-based methods are methods that basically use the explicit image formation (forward model) to solve the pansharpening problem, which is an under-determined problem that needs regularization for optimal results. DL based methods are, e.g., methods based on convolutional neural networks. While there are currently not many papers in the literature where DL is used for image fusion, it is expected that there will be an increase in the number of such methods in the coming years due to the efficiency of DL based methods. In the next 4 subsections, each category of methods will be covered in more detail.

1.3.1 CS Methods

For the CS family of methods, the basic principle is a spectral transformation of the MS image that has been interpolated to the size of the PAN image, e.g., using principal component analysis (PCA) [13], and then one of the components in the new space is replaced by the PAN image which has been histogram adjusted to have the same mean and variance as the component being replaced. Finally, the inverse spectral transform gives the fused image. A requirement for this scheme to work is that the substituted component and the PAN image need to be highly correlated. However, it has been shown [14] that CS methods can be described using a general injection scheme where the forward and inverse transformations do not have to be computed explicitly. The CS framework basically consists of two steps. The first step is detail extraction where details are extracted from the PAN image by subtracting from it a linear combination of the interpolated MS image bands. Thus, the first step determines the coefficients of the linear combination. The second step is detail injection, where the extracted details are injected in the interpolated bands of the MS image. The details are modulated by an injection gain that is usually band specific. The main difference between CS methods is indeed how the injection gains are computed. In Figure 1.2 a flowchart is shown to describe the CS detail injection scheme. The detail injection scheme is given by

$$\mathbf{Z}_k = \tilde{\mathbf{Y}}_k + g_k(\mathbf{X} - \mathbf{I}_{\mathrm{LP}}), \quad k = 1, \ldots, N_b, \tag{1.1}$$

where $\mathbf{Z}_k$ is the estimated kth band of the fused image, $\tilde{\mathbf{Y}}_k$ is the kth band of the observed MS image interpolated to the size of the PAN image, $\mathbf{X}$ is the PAN image

itself, g_k is the kth injection gain, N_b is the number of MS bands, and

$$\mathbf{I}_{\mathrm{LP}} = \sum_{i=1}^{N_b} \omega_i \tilde{\mathbf{Y}}_i, \tag{1.2}$$

where the weights, $\mathbf{w} = \{\omega_1, \ldots, \omega_{N_b}\}$, give the relative overlap of the spectral coverage of the MS sensor with the spectral coverage of the PAN sensor. These weights can be obtained from sensor specifications or they can be estimated, e.g, as in [15]. The interpolation used is a 23-tap near ideal low pass filter [16].

Common CS methods include the intensity-hue-saturation (IHS) methods [11, 17, 18], methods based on the Gram-Schmidt (GS) process [19, 20], the band-dependent spatial-detail (BDSD) method [15], the PCA method [21,22] and the Brovey transform method [23]. A good overview of CS (and MRA) methods can be found in [16].

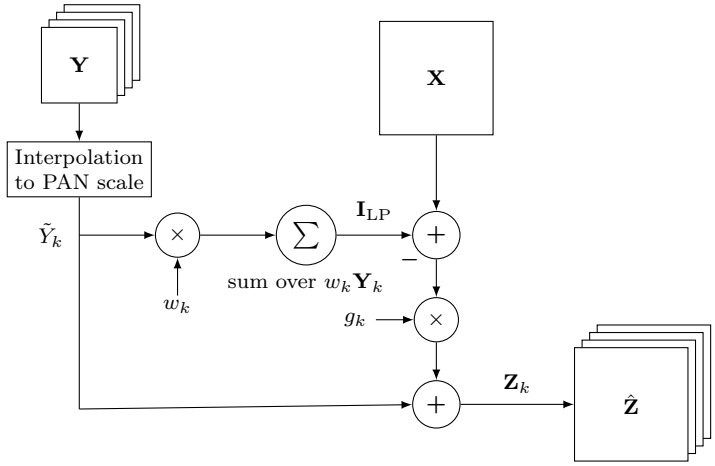


Figure 1.2: Flowchart of the CS injection scheme. The MS image $\mathbf{Y}$ is interpolated to the scale of the PAN image $\mathbf{X}$. An intensity image $\mathbf{I}_{\mathrm{LP}}$ is computed as a linear combination (with coefficients w_k) of the bands of the interpolated MS image, $\tilde{\mathbf{Y}}$, and subtracted from the PAN image to obtain the injection details. The details are then modulated by the injection gain g_k and added to the interpolated bands $\tilde{\mathbf{Y}}_k$ to obtain the fused bands $\mathbf{Z}_k$.

1.3.2 MRA Methods

The MRA family is often based on performing a multiscale decomposition of the PAN image (such as the discrete wavelet transform (DWT) or the undecimated wavelet transform (UDWT) [24]), and injecting the details obtained into the interpolated MS data. As with the CS family of methods, it has been shown that there exists a general detail injection framework for these methods [25, 26] and also that the multiscale decomposition can be replaced by a single unique low pass filter. The MRA framework

is very similar to the CS one, but the critical difference lies in the detail extraction step. For MRA methods, the details are extracted by subtracting from the PAN image a low pass filtered version of the PAN image. The main difference between MRA methods lies in how the low pass filtering is performed and in how the injection gains are computed.

The general detail injection scheme for MRA methods is given by

$$\mathbf{Z}_k = \tilde{\mathbf{Y}}_k + g_k(\mathbf{X} - \mathbf{X}_L), \quad k = 1, \ldots, N_b, \tag{1.3}$$

where $\mathbf{X}_L$ is the low pass filtered PAN image, i.e., such that $\mathbf{X} - \mathbf{X}_L$ gives the details of the PAN image. The low pass filter should ideally be matched to the MTF of the PAN sensor. A good discussion of how the injection gains are chosen for these methods and also of the filters used, is given in [16].

Examples of methods in the MRA category are the high pass filtering (HPF) method [27], the additive wavelet luminance proportion (AWLP) [28] and the additive á trous wavelet transform (ATWT) [29] methods and the indusion [30] method.

Another important class of MRA methods are methods based on pyramidal decompositions of the PAN image based on Gaussian low pass filters [31]. The corresponding differential representation of the Gaussian pyramidal levels is called Laplacian pyramid and there are several methods in the literature based on pyramidal decompositions such as [32] and [33]. Given the above discussion it is evident that the terms used in [11], i.e., linear combination approximation (LCA) methods and spatial filter approximation (SFA) methods are indeed the same as the definition for CS and MRA methods used here, i.e., CS methods can be considered as LCA methods and MRA methods are SFA methods. For more information on MRA methods see, e.g., [16, 34].

Since the detail extraction step in the CS and MRA methods is fundamentally different in the sense, that for CS methods, the details are dependent on the bands of the interpolated MS image, while for MRA methods, the extracted details are obtained by low pass filtering of the PAN image itself, it is evident that the CS and MRA methods have different spectral and spatial qualities. In [35], it is argued that CS methods generally produce results that have a slightly lower spectral/spatial quality than MRA methods, but are more tolerant to mis-registration between the MS and PAN image and aliasing artifacts originating from the modulation transfer function (MTF) of the MS bands being too broad during sampling. In Figure 1.3 a flowchart is shown to describe the MRA injection scheme.

In the thesis, the comparison methods that are mostly used for experiments are CS and MRA methods and the difference between how CS and MRA methods operate and handle mis-registration etc. will be discussed in more detail in subsequent chapters.

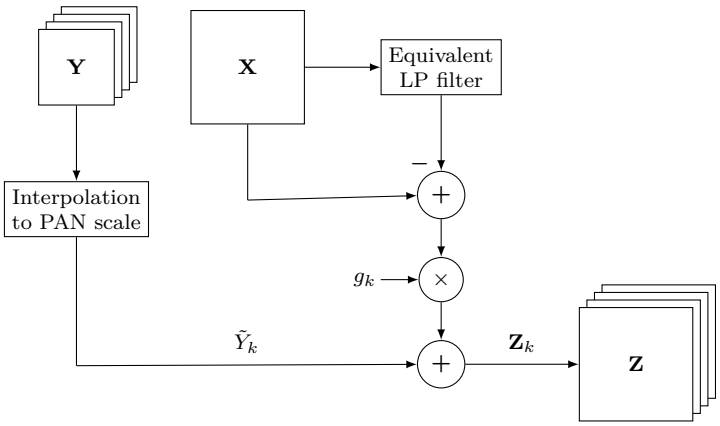


Figure 1.3: Flowchart of the MRA injection scheme. As with the CS scheme, the first step is the interpolation of the MS image $\mathbf{Y}$ to the scale of the PAN image $\mathbf{X}$ to obtain $\tilde{\mathbf{Y}}$. The injection details are obtained by subtracting from the PAN image a low pass filtered version of the PAN image. These details are then modulated by the injection gains g_k and added to the interpolated bands of the MS image, $\tilde{\mathbf{Y}}_k$, to obtain $\mathbf{Z}_k$.

1.3.3 Model-Based Methods

While all image fusion methods are based on a model, it is an under-determined inverse problem. The class of model-based methods contains many diverse approaches. In the discussion below the following categorization of model-based methods will be used:

1. Sparse representation based methods, including compressed sensing.
2. Variational methods.

Below, a short overview of each of those approaches is given along with references to recent or notable works in the literature.

Sparse Representation and Compressed Sensing based Methods

In recent years, the theory of sparse representations and compressed sensing has been steadily growing and several pansharpening methods based on these paradigms have been proposed such as [36–43].

Most model-based methods are based on an observational model where the observed MS image is given by a degradation of a high resolution MS image plus some additive zero-mean Gaussian noise. Another assumption often used, is that the PAN image can be approximated by a linear combination of the bands of the higher resolution MS image. It is this higher resolution MS image that needs to be estimated.

A sparse representation [44–47] means that, e.g., natural images can be modeled as a linear combination of elements from a dictionary or atoms. The representation is called sparse if most of the coefficients of the linear combination are zero, i.e., a few elements are being used from the whole dictionary. The first pansharpening method based on compressed sensing [48, 49] was proposed in [43]. Compressed sensing theory is related to the sparse representation in the sense that when a signal has a sparse representation, compressed sensing allows you to restore the signal using far fewer measurements than required by the Shannon sampling theorem [50]. In [43], the degradation of the high resolution MS image and the construction of the high resolution PAN image is modeled as a linear sampling process, described by a matrix which is considered as the compressed sensing measure matrix. Thus, the pansharpening problem is converted to a signal restoration problem with sparsity regularization.

In [36], a simple additive model is used, where the high resolution MS image is given as the sum of the interpolated MS image (to the scale of the PAN image) plus a detail image. The detail image is estimated using sparse representation theory where the algorithm is based on orthogonal matching pursuit (OMP) [51].

In [52], the pansharpening problem is formulated as an under-determined inverse problem regularized by the ℓ_1-norm of the coefficients of overcomplete multi-scale transforms which all are tight-frames. Two approaches are compared for sparsity promoting ℓ_1-norm regularization, the analysis and the synthesis approach.

Variational Methods

Variational methods are usually formulated using a data-fidelity term, i.e., the ℓ_2-norm of the difference between the observed MS image and the downgraded high resolution MS image to be estimated, according to the image formation model. Since this is a typical ill-posed inverse problem, some regularization is needed for optimal results and thus the objective function to be minimized consists of the data-fidelity term plus one or more regularization terms, where at least one of them involves the PAN image. A common assumption is that that the sum of the bands of the high resolution MS image gives the PAN image. This can also be incorporated into the data-fidelity term as hard constraint. In [53], the data-fidelity term includes both the fidelity between the observed MS and degraded high resolution MS image and also the fidelity between a linear combination of high resolution MS bands and the observed PAN image. The regularization term used was the total variation [54] norm of the high resolution MS image. This objective function was minimized using a Minorize-Maximization (MM) [55] algorithm. In [56], the assumption that the PAN can be approximated by a linear combination of the bands of the fused image was included as a soft constraint along with a TV norm of the fused image.

In [57], the authors use a more complex scheme where they develop a variational model based on both spatial and spectral sparsity priors. They use a weighted TV vector norm to align spatial information in the fused image and propose two regularizers

that promote a soft constrain on the linear dependence between the bands of the high resolution MS image and the PAN image. The former regularizer estimates the coefficients of the linear combination directly by regression and the second regularizer is based on principal component pursuit (PCP) to obtain the underlying low rank structure of the MS image. Both methods employ a split augmented Lagrangian shrinkage algorithm [58]. Other variational methods include [59–66]

1.3.4 Deep Learning Based Methods

During the past few years, an emerging set of methods called Deep Learning (DL) has been making impressive improvements over traditional signal processing methods in areas such as image classification [67,68], speech recognition [69,70], natural language analysis and understanding [71] and various other fields such as analyzing data from particle accelerators [72]. The main ingredient of DL is the artificial neural network (NN) [73–75]. One of the most powerful new tools of DL is the convolutional neural network (CNN) [76–79] which is inspired by how the animal visual cortex works on a simple level. Not many papers can be found in the literature on image fusion using DL. In [80], a deep NN is trained using patches of the intensity image, i.e., a linear combination of the bands of the interpolated MS image, as input, and corresponding patches of the observer PAN image as targets. In [81], a two dimensional (2D) CNN is trained using patches of the stacked and spatially downgraded interpolated MS and PAN image as inputs and corresponding patches from the observed MS image as targets.

1.4 Hypersharpening and MS/HS fusion

The fusion of an HS image and a PAN image is referred to as hypersharpening [82]. This is essentially the same problem as the pansharpening problem, the main difference being that an HS image has a much higher number of spectral bands than an MS image, often spanning a larger part of the electromagnetic spectrum than the PAN image.

Another important problem is the fusion of MS and HS images. HS images have usually lower resolution than MS images and hence the HS image is sharpened using the spatial information from the MS image. The only difference between pansharpening and MS/HS fusion is that both observed images are of a higher dimension. So this is the same kind of inverse problem as pansharpening and hypersharpening. In the next two sections, methods developed for hypersharpening and MS/HS fusion are discussed in more detail.

1.4.1 Hypersharpening

With the increasing availability of HS sensors and amount of acquired HS imagery, this kind of fusion is becoming increasingly relevant. Since the hypersharpening problem

is intrinsically the same as pansharpening, most of the currently available methods for pansharpening can be adapted to work with HS images instead of MS images without difficulties. Methods for hypersharpening include those in [83–93]. A good review on this topic is given in [94].

1.4.2 MS/HS Fusion

HS images have usually lower resolution than MS images and thus the HS image is sharpened using the spatial information from the MS image. Due to the multi-band nature of the higher resolution MS image, adapting existing pansharpening methods to solve this problem requires more work than for the hypersharpening case. Since both source images are multi-band images, methods derived from, e.g., spectral unmixing and matrix factorization techniques can be developed that exploit the rich spatial and spectral information present in the MS and HS images, respectively. Methods that perform MS/HS fusion include [83,95–112]. The following approaches are relevant and this thesis makes contributions to approaches 3. and 4.

1. Extend existing pansharpening methods. Break the MS/HS problem into a set of sub-problems equal to the number of MS bands and use existing pansharpening methods, such as the CS and MRA methods.

2. Model-based methods which use spectral unmixing.

3. Model-based methods using Bayesian inference, i.e., MAP estimation.

4. DL based methods

Extend existing pansharpening methods

The MS/HS fusion problem can be broken into a number of sub-problems where bands of the HS image that are closest to the the bands of the MS image in terms of wavelength are assigned to each MS band. Thus, the fusion problem basically becomes which HS bands to assign to each MS band and then fuse the assigned HS bands and MS band using an existing pansharpening method.

In [104], this approach was used to make a generalized framework for using existing methods, however the band assignment was done manually. Also, HS bands outside the bandwidth of MS bands were processed using ratio image-based spectral resampling (RIBSR) to enhance the results. In [108], a similar approach was used, i.e., manual assignment of HS bands to MS bands and fusion using an the method proposed in [26]. In [112], a band assignment algorithm based on minimizing the SAM [113] metric was proposed for MRA and CS methods.

Model-based methods which use spectral unmixing

Spectral unmixing [114–117] is a technique based on a mixture model to extract so-called endmembers (spectra) and their abundances in an HS image. Thus an HSI can be decomposed into a product of two non-negative matrices, where one contains the endmember spectra and the other consists of the abundance maps of those endmembers. A typical unmixing MS/HS fusion method, tries to estimate the high resolution abundances and the high spectral resolution endmember spectra.

In [106], an MS/HS fusion method is proposed that uses coupled non-negative matrix factorization, i.e., the HS and MS images are unmixed into abundance and endmember matrices using a non-negative matrix factorization (NMF) unmixing [118] method, while taking into account the sensor observational models. The high resolution HS image is obtained by a product of the high resolution abundance map matrix obtained from the MS image and the endmember matrix obtained from the observed HS image.

In [95], the fusion is performed by estimating the high resolution abundances from the MS image and the endmember spectra from the HS image. The fusion is formulated as an ill-posed inverse problem where sparse coding is employed as a regularizer. The dictionary is obtained from unrelated high resolution MS or PAN images.

Model-based methods using Bayesian inference

MAP estimation MS/HS methods are typically based on treating the observed images and high resolution HS image to be estimated as random vectors and developing an MAP estimator for the conditional probability density function (PDF) of the high resolution fused image, given the observed MS and HS images. Using the Bayes rule and assuming independence of the observed data given the high resolution image, this conditional PDF is expressed as the product of the conditional PDF of MS data given the high resolution image and the conditional PDF of the high resolution fused image given the MS image. The former PDF is obtained by the observational model, i.e., the observed HS image is the degradation and downsampling of the high resolution HS image, and the latter PDF is obtained by assuming that the MS image and high resolution HS image are jointly Gaussian. Another important assumption to be made is that the pixel-wise conditional covariances are constant. In [119], this kind of estimator was developed to estimate a high resolution HS image, given an observed HS and either observed MS image or a PAN image. In [120], this estimation was performed in the wavelet domain, where the á trous wavelet transform was used.

DL based methods

The MS/HS fusion problem can be solved in a similar manner as the pansharpening problem using a CNN. One of this thesis contributions is a method that performs the estimation of the first few PCs of the fused image by training it on data at a lower resolution level. The fused image is yielded by performing the inverse PCA transform.

The basic assumption being made is that the spectral singular vectors for the low resolution and high resolution HS images are the same.

1.5 Quantitative quality evaluation of fused images

Since pansharpening/hypersharpening is an inverse problem, there is no high resolution reference image available. The evaluation and comparison of the various image fusion methods is often problematic due to the fact that there is no single accepted protocol for quantitative quality evaluation of the fused image and also because of the lack of standardized data sets that are available to the whole community of researchers. Even though quality assessment of the pansharpened images is a much debated topic, there are surprisingly few papers in the literature on the subject, especially regarding quality assessment of full scale imagery. Papers on the topic include [121–133]. Regarding evaluation protocols, there are two main approaches:

1. Reduced resolution evaluation, where observed MS/HS image is used as the reference.
2. Full resolution evaluation. Use of metrics that attempt to evaluate the fusion quality at the full resolution scale without using a reference image.

The former approach or protocol was first used in [121] but is generally attributed to L. Wald [122] and called Wald's protocol. The second method consists of using metrics such as the Quality with No Reference (QNR) [134] metrics or the metrics proposed in [124]. In the following two subsections each protocol is discussed in more detail.

1.5.1 Reduced Resolution Evaluation

In [122], there are three properties stated which the fused image should have. The first one, consistency property, states that the fused image reduced to the resolution scale of the MS image should be as identical to the observed MS image as possible. The second property, called the synthesis property, states that the the fused image should be as identical as possible to the MS image acquired if the sensor had the resolution of the fused image. This is not possible to verify since there is no high resolution reference image and thus the reduced resolution protocol is to downgrade the observed data spatially by the resolution factor between the low and high resolution images and perform the fusion at this lower (reduced) resolution scale. The fused image obtained should have the same resolution as the observed MS image which is used as the reference image. When spatially downgrading the observed images in order to verify both the consistency and synthesis properties, it is important to use filters that are matched to the shape of the sensor's MTF, to ensure correct results [26, 125, 134].

Since the observed MS image can be used as a reference image, it is possible to use well-established metrics such as ERGAS [135], spectral angle mapper (SAM) [113], universal image quality index (UIQI) [136], its extension Q4/Q2n [137, 138] and Qavg [136]. These metrics are described in more detail in Appendix A.

Synthesis property

The synthesis property [122] dictates that the fused image should be as much identical as possible to a higher resolution image acquired by the MS sensor, i.e., if it had the same spatial resolution as the PAN sensor. Due to the inherent lack of the high resolution MS image, it is not possible to verify the synthesis property at this resolution scale. What needs to be done, is to decimate the MS and PAN images such that the fused image has the same resolution as the observed MS image. This means that the observed MS image can then be used as the reference image and thus well-established quantitative metrics such as ERGAS and SAM can be used. The decimation of the source images should ideally be performed using decimation filters tuned to the MTF of the MS and PAN sensors. This is the most commonly used method to quantitatively evaluate pansharpened images. A figure that shows how the evaluation is performed is shown in Figure 1.4.

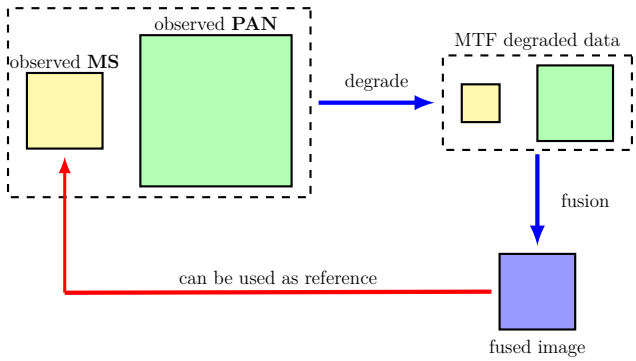


Figure 1.4: Protocol for the verification of the synthesis property.

For this method to be reliable for assessing the performance of the evaluated methods at full scale, it must be assumed that the performance of the pansharpening methods being evaluated is mostly invariant to the resolution scale of the source images.

Consistency property

The consistency property [122] dictates that the fused image, when spatially decimated to the scale of the observed MS image, ideally using decimation filters tuned to the

MTF of the MS sensor, should be as identical as is possible to the observed MS image. The outline for this protocol is shown in Fig. 1.5. When the fused image has been decimated to the scale of the observed MS image, it can be evaluated using well-known quantitative metrics using the observed MS image as the reference image. As was mentioned in the introduction, the verification of this property is considered to be a necessary condition [125] and that its fulfillment does not have to imply correct fusion. However, it is demonstrated in the experiments section of Chapter 5 that evaluation of pansharpening methods using this property or protocol gives results that are highly correlated with evaluation using the synthesis property w.r.t. the ranking of methods and especially for CS methods.

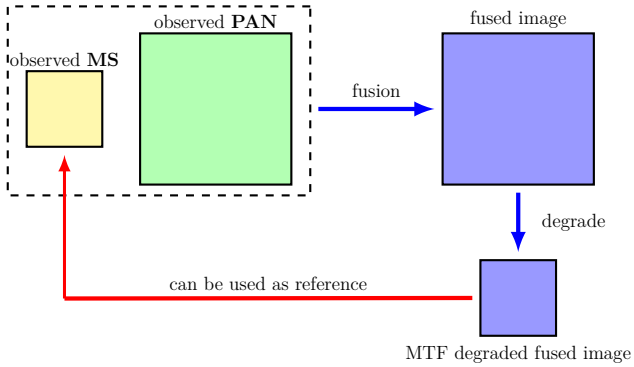


Figure 1.5: Protocol for the verification of the consistency property.

1.5.2 Full Resolution Evaluation

In order to quantitatively measure the quality of the fused image at the full resolution scale it is obviously not possible to use any reference image and thus metrics that do not take a reference need to be used, such as the QNR metric [134], the metrics proposed in [125] which take into account the MTF of the sensor, the metrics proposed in [124] and the joint quality metric (JQM) proposed in [133], which uses the structural similarity index (SSIM) [139]. It has been demonstrated [140], that the spatial metric in [124] is not reliable and tends to show opposite trends compared to reduced resolution evaluation using metrics with reference. Finally, there are the information-level metrics proposed in [141] which require land-cover ground reference.

The QNR metric is the preferred metric when evaluating pansharpened images without a reference, at the full scale. Since there is no reference image, some assumption need to be made regarding the fusion process. The metric is composed of two sub-metric or indices. The first one, D_λ, attempts to measure spectral distortion while the latter, D_S, attempts the measure spatial distortion. The idea behind the D_λ metric is that

the fused image should have the same interband UQUI scores as the the observed image. Any deviation is considered distortion. it is given by

$$D_\lambda = \sqrt[p]{\frac{1}{N_b(N_b-1)} \sum_{l=1}^{N_b} \sum_{\substack{r=1 \\ r\neq l}}^{N_b} \left| Q(\mathbf{Z}_l, \mathbf{Z}_r) - Q(\mathbf{Y}_l, \mathbf{Y}_r) \right|^p}, \tag{1.4}$$

where Q is the UIQI index and the constant p is usually chosen to be 1.

The idea behind the spatial distortion metric D_S is of a similar nature. The assumption being made is that the UIQI values obtained for the bands of the MS image and the spatially degraded PAN image should stay the same for the fused image and the PAN image. It is given by

$$D_\mathrm{S} = \sqrt[q]{\frac{1}{N_b} \sum_{l=1}^{N_b} \left| Q(\mathbf{Z}_l, \mathbf{X}) - Q(\mathbf{Y}_l, \tilde{\mathbf{X}}) \right|^q}, \tag{1.5}$$

where $\tilde{\mathbf{X}}$ is the spatially degraded PAN image such it has the same size as the observed MS image and q is typically chosen to be 1.

Finally, the QNR index is given by

$$\mathrm{QNR} = (1 - D_\lambda)^\alpha (1 - D_\mathrm{S})^\beta, \tag{1.6}$$

where the constants α and β control the relative relevance of the one's complements of the spectral and spatial indices. These constants are usually chosen to be 1. According to (1.6) the highest value of QNR is 1 when the spectral and spatial indices are zero and the lowest value of QNR is zero.

The D_S index has a certain flaw which is that if the UIQI values are low, the index will give a low value, indicating little spectral distortion. A common problem with the QNR index in general is that it often gives images that have a low spatial fidelity a high score. Another weakness of the QNR is metric is that for the D_λ index, no justification is given why the interband UIQI scores should be the same across different resolution scales. For a typical pansharpening method, the spatial details are obtained from an auxiliary sensor and then extracted and injected into the interpolated MS image. Often, especially when using imagery from newer sensors such as, e.g., WorldView2, there are MS bands that are well outside the wavelength span of the PAN band. Sharpening these bands should give a higher D_λ value, indicating spectral distortion.

1.6 Data Sets

In this section, the data sets used for all experiments in this thesis are described. A total of 5 data sets are used, 2 real MS/PAN data sets, one simulated MS/PAN data

set and finally two real hyperspectral data sets.

The first real data set is a rural WorldView-2 (WV2) image of the Agricultural University of Iceland at Hvanneyri in the western part of Iceland. The data set is large, the PAN image is 4096×4096 pixels and the 8-band MS image is 1024×1024 pixels. The resolution of the PAN and MS images are 0.46m and 1.84m, respectively. The radiometric resolution of both images is 11 bits per pixel.

The second real data set is a QuickBird (QB) image of the Mississippi State University campus. This is also a relatively large data set; the PAN image is 2048×2048 pixels with 0.65m resolution and the MS image is of size 512×512 pixels each, with 2.6m resolution. As for the WV2 data, the radiometric resolution is 11 bits per pixel for both the MS and PAN image. There is a small mis-alignment between the PAN and MS images, which can affect the performance of the pansharpening algorithms tested.

The simulated or semi-real data set is an image depicting a part of Toulouse, France. It is simulated using images obtained from an airborne sensor, according to early specification for the Pléiades 1 MS and PAN sensors. The PAN image is a 0.7 resolution 1024×1024 pixels image while the 4-band MS image has 2.8m resolution and 256×256 pixels. As for the QB data set, there is a small misalignment between the MS and PAN images.

The HS image used for experiments in Chapter 2 was acquired using the AVIRIS airborne sensor and has a spatial resolution of 1.5m. It depicts the Cuprite mining district in Nevada, US. The spectral resolution is 219 spectral reflectance bands spanning a wavelength range of $400 - 2500$nm.

In Chapter 4, an HS image of a part of the city center of Pavia in Italy is used [1]. It was acquired by the ROSIS sensor and its spatial dimensions in pixels are 512 by 512 pixels. The spectral resolution is 102 spectral reflectance bands spanning a wavelength range of $430 - 860$nm.

1.7 Thesis Contribution and Organization

The thesis covers topics related to pansharpening and in particular MS/HS fusion. Two MS/HS fusion methods are proposed. One method is based on MAP estimation in the wavelet domain of PCA transformed data. The second method is based on DL, i.e., using a three dimensional CNN to estimate the fused HSI image. Two other topics are deblurring pre-processing for MRA and CS methods where the goal is to extract additional spectral information from the MS image to enhance the fused image without much computational overhead. The final topic is the quantitative quality evaluation of fused images, where it is demonstrated that the consistency property can give a good estimate of the performance of pansharpening methods at the full resolution scale.

[1] The ROSIS data was kindly provided by Prof. P. Gamba from the University of Pavia, Italy.

In Chapter 2, a model-based Bayesian method is proposed that can be used for pansharpening and also the fusion of MS and HS images. It is based on the wavelet-based Bayesian fusion method proposed in [120] where the authors base their method on an observational model in the wavelet domain, linking the the observed HS image to the spatially degraded high resolution version which has been corrupted by additive Gaussian noise. Since the spatial degradation operator is not precisely known and to avoid having to solve a deconvolution problem, it is implicitly incorporated in the model by a-priori interpolating the observed HS image to the same scale as the MS image [120] and making the assumption that it conserves the correlation structure of the data. The main assumption behind the estimation is that the desired high resolution HS image and the MS image are jointly normal, which applies in the wavelet domain as well. A MAP estimator is derived for the detail wavelet coefficients, where independence of the coefficients is used to globally estimate the pixel-wise covariance matrices assuming that the degradation process preserves the correlation structure of the data. This enables the pixel-wise MAP estimation of all the bands of the HS image. Finally, the inverse wavelet transform produces the estimated fused image. The novelty of the method is the application of the PCA transform [13] to the model in [120] and subsequently assuming that the spatial loadings of the desired high resolution HS image are jointly normal with the MS image in the wavelet domain. This greatly simplifies the estimation since it is only necessary to estimate the first few PCs. Finally, the desired fused HS image is reconstructed via the inverse wavelet and PCA transforms.

In Chapter 3, a pre-processing step is proposed for the CS and MRA general injection frameworks that performs MTF matched deblurring on the interpolated MS image into which the details are injected. Deblurring, in itself, is an important topic in image processing, especially image restoration. It is demonstrated that the deblurring using a Wiener filter [142] can greatly enhance the CS family of methods and also MRA methods based on the ARSIS concept [26], according to quantitative quality metrics, both using reduced resolution evaluation and also full scale QNR metrics.

In Chapter 4, a method for MS/HS fusion using both 2D- and 3D-CNNs is developed. While DL methods have been used to solve the pansharpening problem [80, 81], these methods are extended to handle the fusion of multi-band images, i.e., MS and HS images. Co-registered MS and HS data is difficult to obtain and thus the high resolution MS image is simulated using the HS image which is then spatially degraded. In order to make the problem more computationally tractable, dimensionality reduction of the HS image using PCA is used prior to fusion. The training of the CNN is based on Wald's protocol [122], i.e., the input data are spatially degraded in order to be able to use the PCA transformed observed data as the training target. The method is based on two assumptions. The first one is that the CNN can be trained using data that has lower resolution than the data used for the actual prediction, i.e., the relationship learned between the input and target data at a lower resolution scale, also applies for higher resolution input. The second assumption is that the spectral singular vectors obtained for the lower resolution data are the same as for the higher resolution data. The method is compared to the methods in [83, 120] and it is demonstrated that the proposed methods gives good results.

Chapter 5 is devoted to a comparison of the synthesis and consistency properties [122] for quantitative quality evaluation of fused images. It is argued that verifying the consistency property using well known metrics, is more reliable for estimating the performance of pansharpening methods at the full resolution scale, than using the traditional synthesis property, i.e., Wald's protocol. A number of experiments are performed using 3 data sets and 18 CS and MRA methods.

Finally, in Chapter 6, conclusions are summarized and future work is detailed.

1.7.1 List of publications

Chapter 2 is based on the following publications:

[a] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model based PCA/wavelet fusion of multispectral and hyperspectral images," in *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 1532–1535.

[b] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model-based fusion of multi- and hyperspectral images using PCA and wavelets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2652–2663, May 2015.

Chapter 3 is based on the following publications:

[a] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "MTF-deblurring preprocessing for CS and MRA pansharpening methods," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 1104–1107.

[b] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model based pansharpening method based on tv and mtf deblurring," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 33–36.

[c] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "MTF-based deblurring using a Wiener filter for CS and MRA pansharpening methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2255–2269, June 2016.

Chapter 4 is based on the following publication:

[a] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 639–643, May 2017.

Chapter 5 is based on the following publication:

[a] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1247–1259, March 2016.

Chapter 2

Model-Based Fusion of Multi- and Hyperspectral Images using PCA and Wavelets

In this chapter a method for MS/HS fusion as well as pansharpening is developed. It is based on the method proposed in [120], which assumes that the PAN/MS image and the HS image are jointly normal in order to develop an MAP estimator for the details of the fused image in the wavelet domain. The proposed method adds dimensionality reduction via PCA prior to the fusion and the MAP estimation is performed in the PCA transformed wavelet domain. Experiments are performed using real MS and PAN images and a simulated data set where the MS image is simulated using the HS image which is then spatially decimated. The results indicate that the dimensionality reduction not only significantly reduces the computational complexity of the method in [120], but also improves the results.

2.1 Introduction

In this chapter, a method that can be used for pansharpening/hypersharpening and also for MS/HS fusion is proposed. It is based on the wavelet-based maximum-a-posteriori (MAP) estimation fusion method proposed in [120]. The method is based on an observational model in the wavelet domain which links the observed HS image to the spatially degraded high resolution version, which has been corrupted by additive Gaussian noise. Because the spatial degradation operator is not known and to avoid solving a deconvolution problem, the observed HS image is interpolated to the same scale as the MS image prior to the estimation, where it is assumed that the interpolation operator preserves the correlation structure of the observed HS image.

The main assumption in [120] is that the desired high resolution HS image we want to estimate and the MS image are jointly normal, which also applies in the wavelet domain. Based on this, an MAP estimator is derived for the detail wavelet coefficients. Another assumption that simplifies things considerably, is the independence of the wavelet coefficients which enables the use of a global estimate of the pixel-wise covariances. This makes the derivation of the MAP estimator straight-forward, and it can be used for the estimation of the detail wavelet coefficients for all the bands of the desired high resolution HS image. The inverse wavelet transform then yields the final estimated fused HS image. The drawback of the method is, however, that every band needs to be estimated in the wavelet domain and when using a redundant, i.e., over-complete wavelet transform, this can computationally be quite costly.

The novelty of the method proposed here is the application of the PCA transform [13] to the model in [120]. This yields spatial loadings and singular vectors, i.e., principal components (PCs), where it is assumed that the spatial loadings and the bands of MS image are jointly normal in the wavelet domain, and that the PCs are the same for the observed HS and high resolution HS image we want to estimate. The benefits of the PCA transform is that only the first few r spatial loadings need to be estimated. This greatly decreases the computational cost of the method, without sacrificing quality of the fused image. In fact, the by using the PCA transform, the proposed method gives better results than the method in [120], as will be demonstrated by experiments. The estimated high resolution fused HS image is then reconstructed using the inverse wavelet transform and then the inverse PCA transform. Furthermore, the proposed method is even more tolerant to noise in the HS image than the original method.

Examples of pansharpening are given using a real WorldView 2 data set and both hypersharpening and HS/MS fusion using simulated MS/HS images based on an AVIRIS data set, with and without additive noise in the observed HS/MS images. Due to scarcity of co-registered PAN and HS images and also MS and HS images, the PAN and MS images are simulated using the HS image.

The organization of this chapter is as follows. In Section 2.2, the proposed method is derived. In Section 2.3, parameter selection, i.e., the choice of the number of PCs to be estimated, r, is discussed and in Section 2.4, results using real and simulated data are presented and finally in Section 2.5, the conclusion is drawn. The proposed method is referred to as PWMBF (PCA/wavelet model-based fusion) and the method in [120] will be referred to as MAP.

2.2 Derivation of the Method

2.2.1 Observational Model

The method is based on the observational model in [120] which is a standard model for pansharpening

$$\tilde{\mathbf{Y}} = \mathbf{H}\mathbf{Z} + \mathbf{N}_1, \tag{2.1}$$

where $\tilde{\mathbf{Y}}$ is the spatially interpolated (using bicubic filters) $N \times P$ observed low resolution HS image with P spectral reflectance bands with N pixels each, $\mathbf{H}$ is the spatial degradation operator, $\mathbf{Z}$ is the $N \times P$ unknown high resolution HS image and $\mathbf{N}_1$ is zero-mean Gaussian noise. Here, it is assumed that $\mathbf{Z}$ and $\mathbf{X}$ are jointly Gaussian, where $\mathbf{X}$ is the $N \times Q$ observed MS image with Q spectral reflectance bands.

The blurring operator $\mathbf{H}$ is based on the sensor's point spread function (PSF) followed by downsampling. Since we do not know the exact form of $\mathbf{H}$ and to simplify things considerably, we will assume that the observed data $\mathbf{Y}$ has the same number of pixels as $\mathbf{Z}$ and $\mathbf{X}$ such that $\mathbf{H}$ acts as a spatial degradation operator without downsampling. Later on in the derivation, we will make further assumptions regarding $\mathbf{H}$ to simplify the estimation.

2.2.2 PCA Transform of the Observational Model

The motivation for using the PCA transform for dimensional reduction of the HS image, is that such images contain a lot of spectral redundancy that can be easily exploited, i.e., most of the data resides on a manifold of much lower dimension than the actual image. The low rank representation obtained using only a few PCs can account for nearly all the variance contained in the image. Noise tolerance is another motivation. By discarding noisy spatial loadings, significant denoising of the HS image can be achieved. By only having to estimate the first few PCs instead of every spectral reflectance band can significantly reduce the computational complexity of the method. PCA transforming the observational model (2.1) yields

$$\boldsymbol{B}\boldsymbol{U}_Y^T = \mathbf{H}\mathbf{G}\boldsymbol{U}_Z^T + \mathbf{N}_2, \tag{2.2}$$

where $\boldsymbol{B}\boldsymbol{U}_Y^T$ and $\mathbf{G}\boldsymbol{U}_Z^T$ are the PCA transforms of $\tilde{\mathbf{Y}}$ and $\mathbf{Z}$, respectively, obtained by singular value decomposition, i.e., $\tilde{\mathbf{Y}} = \boldsymbol{V}_Y\mathbf{S}_Y\boldsymbol{U}_Y^T$, where $\boldsymbol{B} = \boldsymbol{V}_Y\mathbf{S}_Y$ is $N \times P$, and similarly $\mathbf{Z} = \boldsymbol{V}_Z\mathbf{S}_Z\boldsymbol{U}_Z^T$, with $\mathbf{G} = \boldsymbol{V}_Z\mathbf{S}_Z$ is $N \times P$, $\boldsymbol{U}$ is $P \times P$ and $\mathbf{N}_2$ is Gaussian noise. The matrices $\boldsymbol{B}$ and $\mathbf{G}$ are the loading matrices associated with the PC matrix $\boldsymbol{U}$. In the following discussion, the columns of the loading matrices $\boldsymbol{B}$ and $\mathbf{G}$, will be referred to spatial loadings. A key point here is that it is assumed that the singular spectral vectors (PCs) $\boldsymbol{U}$ are identical for both $\tilde{\mathbf{Y}}$ and $\mathbf{Z}$, i.e., $\boldsymbol{U}_Y^T = \boldsymbol{U}_Z^T$, implying spectral consistency between the different resolution scales.

The justification for this assumption is the consistency property developed in [122,123]. This property has two aspects, namely, consistency and synthesis. The consistency property states that the original MS image can be obtained by degradation of the fused image. This property also holds for the PCs of both the observed data and the desired high-resolution fused image.

The synthesis property requires that the inter-band mutual relations should be preserved. By assuming that the same linear combination of the spatial loadings is needed to reconstruct both the original MS image and the fused image, it is essentially being implied that the synthesis property holds. This can be also be interpreted as the degradation operator W preserves the correlation structure of the data. Hence, $\boldsymbol{U} = \boldsymbol{U}_Y = \boldsymbol{U}_Z$ and thus $\boldsymbol{U}$ appears in the decomposition for both $\tilde{\mathbf{Y}}$ and $\mathbf{Z}$.

This assumption simplifies the estimation enormously and helps to ensure spectral consistency between the two resolution scales. Multiplying (2.2) with $\boldsymbol{U}$ from the right yields

$$\boldsymbol{B} = \mathbf{H}\mathbf{G} + \mathbf{N}_3, \tag{2.3}$$

where $\mathbf{N}_3$ is again zero-mean Gaussian noise and since PCA is a linear transformation, it still holds that $\mathbf{G}$ and $\mathbf{X}$ are jointly normal. Figure 2.1 shows the first 4 spatial loadings and the corresponding singular spectral vectors (PCs) for the WorldView 2 data set.

2.2.3 2D UDWT of the Model

As was done in [120], the estimation is performed in the wavelet domain. The wavelet transform decomposes an image into a low-pass or coarse approximation and high-pass detail bands at different resolution scales and orientations. The 2D DWT has the nice property of being orthogonal so the number of wavelet coefficients for a given image is the same as the number of pixels. However, it suffers from not being shift-invariant. Instead of using the orthogonal DWT, the UDWT [24] is a much better choice since it is shift-invariant. However, unlike the orthogonal DWT, it is $3L + 1$ redundant, where L is the number of wavelet decomposition levels. The 2D UDWT decomposes an image into a low-pass approximation and high-pass detail bands. There are 3 detail bands, one for each orientation, for each level of decomposition. Each band has the same number of coefficients as there are pixels in the transformed image.

The main motivation for wavelet transforming the model in (2.3) is that it reduces the spatial correlation of the pixels, simplifying the estimation and making it more robust. And since the spatial detail is encoded into the detail sub-bands, it is only necessary to estimate those, leaving the low-pass approximation unchanged.

Applying the 2D UDWT on the loadings in $\boldsymbol{B}$ and $\mathbf{G}$ in (2.3) gives

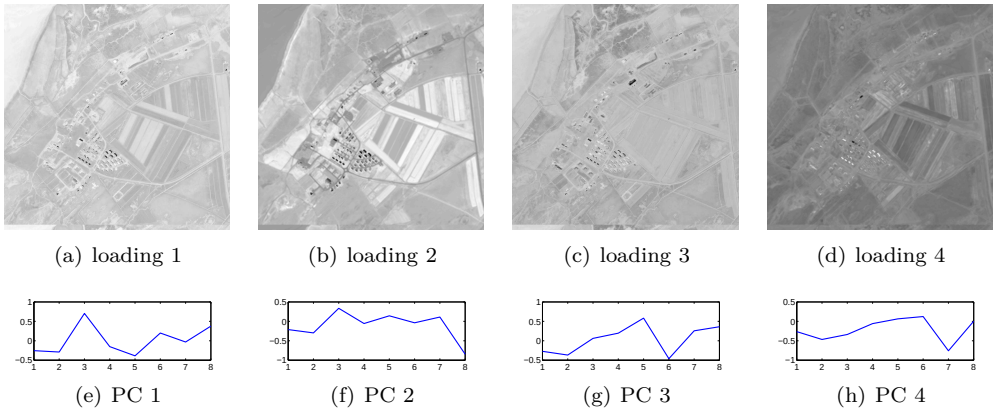
$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{G}} + \boldsymbol{\mathcal{N}}, \tag{2.4}$$

where $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{\mathcal{G}}$ are the 2D UDWT coefficients of the spatial loadings of $\boldsymbol{B}$ and $\mathbf{G}$, respectively, $\boldsymbol{\mathcal{N}}$ is Gaussian noise, independent of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{B}}$ and, $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{X}}$ are jointly Gaussian, where $\boldsymbol{\mathcal{X}}$ are the 2D UDWT coefficients of $\mathbf{X}$. $\boldsymbol{\mathcal{H}}$ is obtained from $\boldsymbol{H}$ using $\boldsymbol{\mathcal{H}} = \boldsymbol{\Phi}\boldsymbol{H}\boldsymbol{\Phi}^T$ where $\boldsymbol{\Phi}$ is the 2D UDWT operator.

The matrices $\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{X}}$ denote the 2D UDWT coefficients and only the detail coefficients are estimated, i.e., the sub-bands (orientations) LH, HL and HH at all levels of decomposition. Here, L denotes low-pass filtering and H denotes high-pass filtering. Thus, LH denotes low-pass filtering by rows and high-pass filtering by columns.

The reason for only estimating the detail sub-bands is that the coarse approximation LL contains an approximation of the signal itself and this band has in general a non-zero mean while the zero-mean detail bands contain deviations (details) from the LL band at various resolution levels and orientations. The assumption that is being made here is that the spatial loadings of the interpolated observed image $\tilde{\mathbf{Y}}$ and the high resolution estimated image $\mathbf{Z}$ have the same representation in the LL band of the wavelet decompositions, i.e., the difference between $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{\mathcal{G}}$ resides almost exclusively in the detail sub-bands.

Each sub-band is estimated independently. With L levels of decomposition, each column of $\boldsymbol{\mathcal{B}}$, $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{X}}$ is arranged as follows.



(a) loading 1 (b) loading 2 (c) loading 3 (d) loading 4

(e) PC 1 (f) PC 2 (g) PC 3 (h) PC 4

Figure 2.1: The first 4 columns of $\hat{\mathbf{G}}$, i.e., spatial loadings for the World-View 2 data set and the associated spectral singular vectors (PCs), i.e., the columns of $\boldsymbol{U}$.

$$\boldsymbol{\mathcal{X}}_{(i)} = [LL^{(L)}, LH^{(1)}, HL^{(1)}, HH^{(1)} \\ , \cdots, LH^{(L)}, HL^{(L)}, HH^{(L)}]^T,$$

where $LL^{(L)}$ denotes the low-pass approximation at level L, $LH^{(j)}, HL^{(j)}$ and $HH^{(j)}$ denote the detail sub-bands at decomposition level j. Each sub-band has N coefficients, making the total number of coefficients $(3L+1)N$.

2.2.4 MAP Estimation

The main novelty of the method is that instead of estimating the wavelet coefficients of every loading, i.e., all the columns of $\boldsymbol{\mathcal{G}}$, it is only necessary to estimate the first r loadings, where r is the only tuning parameter of the method.

Since the the first few PCs explain almost all the variance of the data, the remaining spatial loadings are not so important and can be taken from $\boldsymbol{\mathcal{B}}$, which contains in its columns the spatial loadings of the interpolated observed data. This is especially useful when working with HS images since they are highly spectrally redundant. Another useful benefit of this approach is that if the HS is noisy, the first estimated high resolution spatial loadings will be mostly noise free, since the details come from the relatively clean MS/PAN image. If the HS image is extremely noisy, the remaining spatial loadings can be discarded. Noting that each column $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{\mathcal{G}}$ can be processed independently for each resolution level and orientation, $\mathbf{b}, \boldsymbol{g}$ and $\boldsymbol{x}$ denote a vectorized detail sub-band of the respective 2D UDWT transforms.

Now the MAP estimator of $\boldsymbol{g}$ given the observed vectors $\mathbf{b}$ and $\boldsymbol{x}$ can be derived and

is given as

$$\hat{\boldsymbol{g}} = \arg\max_{\boldsymbol{g}} p(\boldsymbol{g}|\mathbf{b}, \boldsymbol{x}), \tag{2.5}$$

i.e., the MAP estimator is the random vector $\hat{\boldsymbol{g}}$ that maximizes the conditional probability function of $\boldsymbol{g}$ given that the observed random vectors $\mathbf{b}$ and $\boldsymbol{x}$ are known. Applying the Bayes rule gives

$$p(\boldsymbol{g}|\mathbf{b}, \boldsymbol{x}) = \frac{p(\mathbf{b}, \boldsymbol{x}|\boldsymbol{g})p(\boldsymbol{g})}{p(\mathbf{b}, \boldsymbol{x})}. \tag{2.6}$$

According to the observational model (2.4), since the noise $\boldsymbol{\epsilon}$ is independent of both $\boldsymbol{x}$ and $\boldsymbol{g}$, it follows that $\boldsymbol{x}$ and $\mathbf{b}$ are independent given $\boldsymbol{g}$ and

$$p(\boldsymbol{g}|\mathbf{b}, \boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{g})p(\mathbf{b}|\boldsymbol{g})p(\boldsymbol{g})}{p(\mathbf{b}, \boldsymbol{x})}. \tag{2.7}$$

Applying the Bayes rule on the conditional probability distribution function (pdf) $p(\boldsymbol{x}|\boldsymbol{g})$ gives $p(\boldsymbol{x}|\boldsymbol{g}) = p(\boldsymbol{g}|\boldsymbol{x})p(\boldsymbol{x})/p(\boldsymbol{g})$ and (2.7) can be formulated as

$$p(\boldsymbol{g}|\mathbf{b}, \boldsymbol{x}) = \frac{p(\boldsymbol{g}|\boldsymbol{x})p(\boldsymbol{x})p(\mathbf{b}|\boldsymbol{g})p(\boldsymbol{g})}{p(\mathbf{b}, \boldsymbol{x})p(\boldsymbol{g})}. \tag{2.8}$$

In the above expression, $p(\boldsymbol{g})$ can be eliminated and the terms $p(\boldsymbol{x})$ and $p(\boldsymbol{x}, \mathbf{b})$ can be dropped, since they are not a function of $\boldsymbol{g}$. The MAP estimator for $\boldsymbol{g}$ becomes

$$\hat{\boldsymbol{g}} = \arg\max_{\boldsymbol{g}} p(\mathbf{b}|\boldsymbol{g})p(\boldsymbol{g}|\boldsymbol{x}). \tag{2.9}$$

The conditional pdf, $p(\mathbf{b}|\boldsymbol{g})$, is given by the observational model (2.4)

$$p(\mathbf{b}|\boldsymbol{g}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C_{n,n}}|}} \exp\big\{ -\frac{1}{2}(\mathbf{b} - \boldsymbol{\mathcal{H}g})^T \boldsymbol{C}_{n,n}^{-1}(\mathbf{b} - \boldsymbol{\mathcal{H}g})\big\}, \tag{2.10}$$

where $\boldsymbol{C_{n,n}}$ is the noise covariance of a detail band. It is important to note that, while for the orthogonal DWT, the noise covariance is equal to the input image (band) noise covariance, while for the UDWT, it is scaled by a factor depending on the wavelet filters and level of decomposition.

The second conditional pdf $p(\boldsymbol{g}|\boldsymbol{x})$ is given by

$$p(\boldsymbol{g}|\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C_{g|x}}|}} \exp\big\{ -\frac{1}{2}(\boldsymbol{g} - \boldsymbol{\mu_{g|x}})^T \boldsymbol{C}_{\boldsymbol{g}|\boldsymbol{x}}^{-1}(\boldsymbol{g} - \boldsymbol{\mu_{g|x}})\big\}, \tag{2.11}$$

where $\boldsymbol{\mu_{g|x}}$ and $\boldsymbol{C_{g|x}}$ denote the conditional mean and covariance, respectively, given by

$$\boldsymbol{\mu_{g|x}} = E[\boldsymbol{g}] + \boldsymbol{C_{g,x}}\boldsymbol{C}_{\boldsymbol{x},\boldsymbol{x}}^{-1}(\boldsymbol{x} - E[\boldsymbol{x}]) \tag{2.12}$$

$$\boldsymbol{C_{g|x}} = \boldsymbol{C_{g,g}} - \boldsymbol{C_{g,x}}\boldsymbol{C}_{\boldsymbol{x},\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{g},\boldsymbol{x}}^{T}, \tag{2.13}$$

where $E[\cdot]$ is the expectation operator and $\boldsymbol{C}_{\cdot,\cdot}$ denotes the covariance matrix. Note that $E[\boldsymbol{g}] = 0$ and $E[\boldsymbol{x}] = 0$ since $\boldsymbol{g}$ and $\boldsymbol{x}$ are detail UDWT sub-bands.

Solving (2.9) using (2.10) and (2.11) the MAP estimator for $\boldsymbol{g}$ is given as

$$\hat{\boldsymbol{g}} = \boldsymbol{\mu}_{\boldsymbol{g}|\boldsymbol{x}} + \boldsymbol{C}_{\boldsymbol{g}|\boldsymbol{x}}\boldsymbol{\mathcal{H}}^T[\boldsymbol{\mathcal{H}}\boldsymbol{C}_{\boldsymbol{g}|\boldsymbol{x}}\boldsymbol{\mathcal{H}}^T + \boldsymbol{C}_{\boldsymbol{n},\boldsymbol{n}}]^{-1}(\mathbf{b} - \boldsymbol{\mathcal{H}}\boldsymbol{\mu}_{\boldsymbol{g}|\boldsymbol{x}}). \tag{2.14}$$

Solving this requires knowledge of $\boldsymbol{\mathcal{H}}$, the conditional covariance $\boldsymbol{C}_{\boldsymbol{g}|\boldsymbol{x}}$ and the mean $\boldsymbol{\mu}_{\boldsymbol{g}|\boldsymbol{x}}$. The former is not always known and its presence requires solving a deconvolution problem, and the latter, i.e., the conditional covariance, is not possible to compute exactly. However, there are simplifications [120] that can be made to address this problem.

By assuming that $\boldsymbol{H}$ preserves the correlation structure of the data between resolution levels, its operation on the MS image can be approximated by smoothing until the spatial resolution of the HS image is obtained. This also applies to the mean and covariances in (2.14), so $\boldsymbol{\mathcal{H}}\boldsymbol{\mu}_{\boldsymbol{g}|\boldsymbol{x}}$ is replaced by $\tilde{\boldsymbol{\mu}}_{\mathbf{b}|\tilde{\boldsymbol{x}}}$ and $\boldsymbol{\mathcal{H}}\boldsymbol{C}_{\boldsymbol{g}|\boldsymbol{x}}$ by $\boldsymbol{C}_{\mathbf{b}|\tilde{\boldsymbol{x}}}$ and similarly, $\boldsymbol{C}_{\boldsymbol{g}|\boldsymbol{x}}\boldsymbol{\mathcal{H}}^T$ is replaced by $\boldsymbol{C}_{\mathbf{b}|\tilde{\boldsymbol{x}}}$, where $\boldsymbol{g}$ is replaced by $\mathbf{b}$ and $\boldsymbol{x}$ is replaced by its smoothed version $\tilde{\boldsymbol{x}}$. Accordingly, the term $\boldsymbol{\mathcal{H}}\boldsymbol{C}_{\boldsymbol{g}|\boldsymbol{x}}\boldsymbol{\mathcal{H}}^T$ in (2.14) is the covariance of data that have been smoothed twice. However, as was argued in [120], all the covariance should be calculated at the same resolution scale.

In summary, the effect of $\boldsymbol{\mathcal{H}}$ in (2.14) can be approximated by calculating the mean and covariances at the lower resolution scale. This enables the elimination of $\boldsymbol{\mathcal{H}}$ from the estimation in (2.14).

Another simplification is that the $\boldsymbol{g}$ can be estimated pixel-wise. This is due to the assumption that the wavelet coefficients in $\boldsymbol{g}$ are spatially independent. As in [120], it is also assumed that the pixel-wise covariances are constant over the detail sub-bands. This gives the estimate $\boldsymbol{C}_{\boldsymbol{u}_n,\boldsymbol{v}_n} = (1/N)\sum_{n=1}^{N}\boldsymbol{u}_n\boldsymbol{v}_n^T$. Pixel-wise estimation is indicated by the subscript n.

The pixel-wise estimator for $\boldsymbol{g}$ is given as

$$\hat{g}_n = \mu_{g_n|\boldsymbol{x}_n} + C_{b_n|\tilde{\boldsymbol{x}}_n}[C_{b_n|\tilde{\boldsymbol{x}}_n} + C_{n_n,n_n}]^{-1}(b_n - \tilde{\mu}_{b_n|\tilde{\boldsymbol{x}}_n}). \tag{2.15}$$

Assuming that the pixel-wise covariance structure is preserved between resolution levels gives

$$\mu_{g_n|\boldsymbol{x}_n} = \boldsymbol{C}_{b_n,\tilde{\boldsymbol{x}}_n}\boldsymbol{C}^{-1}_{\tilde{\boldsymbol{x}}_n,\tilde{\boldsymbol{x}}_n}\boldsymbol{x}_n \tag{2.16}$$

$$\tilde{\mu}_{b_n|\tilde{\boldsymbol{x}}_n} = \boldsymbol{C}_{b_n,\tilde{\boldsymbol{x}}_n}\boldsymbol{C}^{-1}_{\tilde{\boldsymbol{x}}_n,\tilde{\boldsymbol{x}}_n}\tilde{\boldsymbol{x}}_n \tag{2.17}$$

$$C_{b_n|\tilde{\boldsymbol{x}}_n} = C_{b_n,b_n} - \boldsymbol{C}_{b_n,\tilde{\boldsymbol{x}}_n}\boldsymbol{C}^{-1}_{\tilde{\boldsymbol{x}}_n,\tilde{\boldsymbol{x}}_n}\boldsymbol{C}^{T}_{b_n,\tilde{\boldsymbol{x}}_n}. \tag{2.18}$$

Note that the derivation above applies to a single detail sub-band of the estimated image.

2.2.5 Computation of $\hat{\mathbf{Z}}$.

In order to obtain $\hat{\mathbf{Z}}$, i.e., the estimated high resolution image, the detail sub-bands of the 2D UDWT coefficients of the first r-loadings of $\boldsymbol{\mathcal{G}}$ are estimated using the pixel-wise MAP estimator (2.15), which gives $\hat{\boldsymbol{\mathcal{G}}}$. The next step is using the inverse 2D UDWT transform to obtain $\hat{\mathbf{G}}$. There are 2 reconstruction scenarios possible:

1. If the low resolution data are relatively clean, i.e., noise free, the remaining $P - r$ columns of $\hat{\mathbf{G}}$ are taken from $\boldsymbol{B}$, i.e., the number of PCs is not reduced. However, the first r spatial loadings (columns) of $\hat{\mathbf{G}}$ have the same spatial resolution as $\mathbf{X}$ while the remaining spatial loadings have the spatial resolution of $\tilde{\mathbf{Y}}$. Then $\hat{\mathbf{Z}}$ is given by $\hat{\mathbf{Z}} = \hat{\mathbf{G}}\boldsymbol{U}^T$.

2. In the case of noisy data, performing reduced PCA can greatly reduce the noise of the estimate. This means that only the estimated r columns of $\hat{\mathbf{G}}$ and $\boldsymbol{U}$ are retained. The estimated high resolution image is then given by $\hat{\mathbf{Z}} = \hat{\mathbf{G}}_r\boldsymbol{U}_r^T$, where $\mathbf{G}_r$ and $\boldsymbol{U}_r^T$ are the reduced spatial loading matrices and PCs, respectively.

2.3 Tuning Parameter Selection

The method has a single tuning parameter, r, which is the number of spatial loadings to estimate. Obviously, the choice of r is very important for the quantitative quality of the fused image. If it is chosen too small, the quantitative quality will decrease. If it is chosen too large, the computational time will increase, without increasing the quantitative quality of the fused image. The easiest way for choosing a good value for r is to make a scree plot of the MS/HS image. Such a plot shows the magnitude of the eigenvalues associated with the PCs or it can also be normalized such that the sum of the eigenvalues is 1. Thus, the plot shows the relative proportion of each component of the total variance of the data. Figure 2.2 shows a scree plot for the 8 MS band WV2 data set which is used in our experiments. Considering the plot, it is evident that choosing r larger than 4 is not going to give much improvement.

2.4 Experimental Setup and Results

There are 3 types of experiments performed using different types of source images. The first experiment demonstrates the performance of the proposed method for pan-sharpening, using the high resolution WorldView 2 data set, where the MS image has 8 bands instead of the usual 4 bands, and where the spectral range of the PAN image is less than that of the MS bands, making the problem more challenging.

In the second experiment, the fusion of a HS image and a PAN image is considered using a simulated PAN image derived from an AVIRIS HS image.

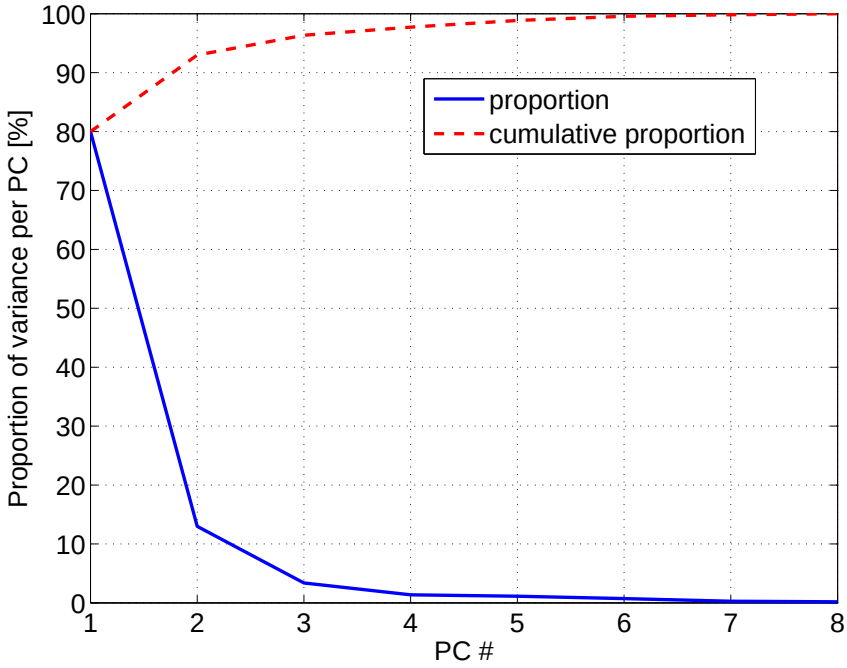


Figure 2.2: Scree plot for the WorldView 2 MS image. The blue line shows the proportional contribution of each PC to the total variance and the red dotted line shows the cumulative variance of the PCs.

Finally, in the third experiment, MS/HS fusion is investigated. Since co-registered MS and HS images were not available, the MS image is simulated using the HS image.

The experiment results demonstrate that the proposed method can handle images with any number of bands and gives excellent results w.r.t. the comparison methods while requiring significantly less computational resources in the case of MS/HS fusion.

In each experiments, both relatively noise free data, i.e., the original data and also data that have been corrupted by a zero-mean Gaussian noise are used. This applies to the lower resolution images. This is done to demonstrate the high noise tolerance of the proposed method.

The quantitative quality assessment is performed at the reduced scale, i.e., the fused image is evaluated using synthesis at the reduced scale. Since the MTF for the HS sensor is not known, it is degraded using bicubic filters. This may not be an optimal choice of filter compared to MTF matched filters, however, it still gives a good idea of the relative performance of the proposed method vs. the methods used for comparison.

The ERGAS [135], SAM [113], Q_{avg} and signal-to-noise-ratio (SNR) metrics are used

for the quantitative quality evaluation of the fused image.

2.4.1 Comparison methods

For the pansharpening and hypersharpening experiments, the proposed method is compared to the MAP method in [120], and the PCA [22] and additive wavelet luminance proportion (AWLP) [28] methods. For MS/HS fusion, the only comparison method is the MAP method.

The PCA method is based on the PCA transformation of the MS image that has been interpolated to the scale of the PAN image. The first PC is substituted for the PAN image, which has been histogram adjusted to match the mean and variance of the first PC. The fused image is then obtained via the inverse PCA transform. This method is one of the first CS methods.

The AWLP method belongs is a typical MRA method and those methods, as well as the CS methods, can be described using a simple detail injection framework [16]. The k-th band of the fused image (P bands) is given by

$$\mathbf{Z}_k = \tilde{\mathbf{Y}}_k + \boldsymbol{\gamma}_k(\mathbf{X} - \mathbf{X}_L), \quad k = 1, \ldots, N_b,$$

where $\tilde{\mathbf{Y}}_k$ is the spatially upsampled k-th band of the observed MS image (using bicubic interpolation), the injection gain $\boldsymbol{\gamma}_k$ is given by

$$\boldsymbol{\gamma}_k = \frac{\tilde{\mathbf{Y}}_k}{\frac{1}{P}\sum_{k=1}^{P} \mathbf{Y}_k}.$$

In the preceding equation, $\mathbf{X}$ is the PAN image and $\mathbf{X}_L$ is the low pass filtered PAN image and all the operations are point-twise. The details of the PAN image are given by the term $(\mathbf{X} - \mathbf{X}_L)$. The Starck & Murtagh filter [148] is used as the low-pass filter, which is a common choice for this method.

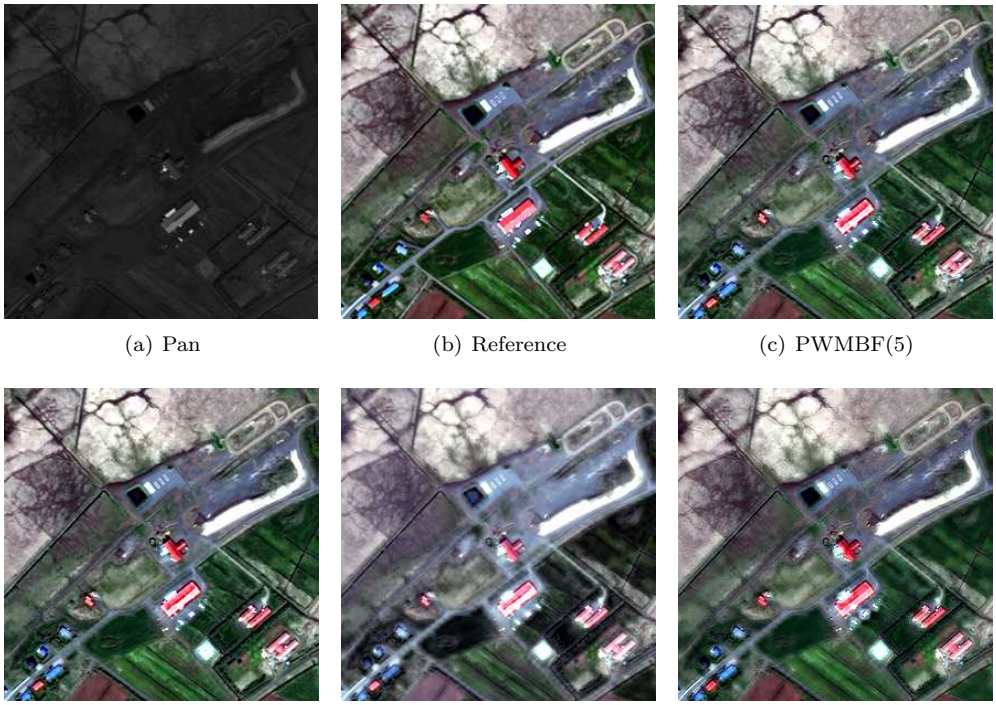
2.4.2 Pansharpening using WorldView 2 data

For this experiment, a WorldView 2 data set, depicting a rural area in the western part of Iceland, is used.

Two examples are given. In the first one, the original data which have been degraded is fused and in the second one, zero-mean Gaussian noise is added to the MS image, prior to fusion, such the SNR becomes 20dB.

The results are summarized in Table 2.1 and the fused images using the clean and noisy data are shown in Figures 2.3 and 2.4, respectively.

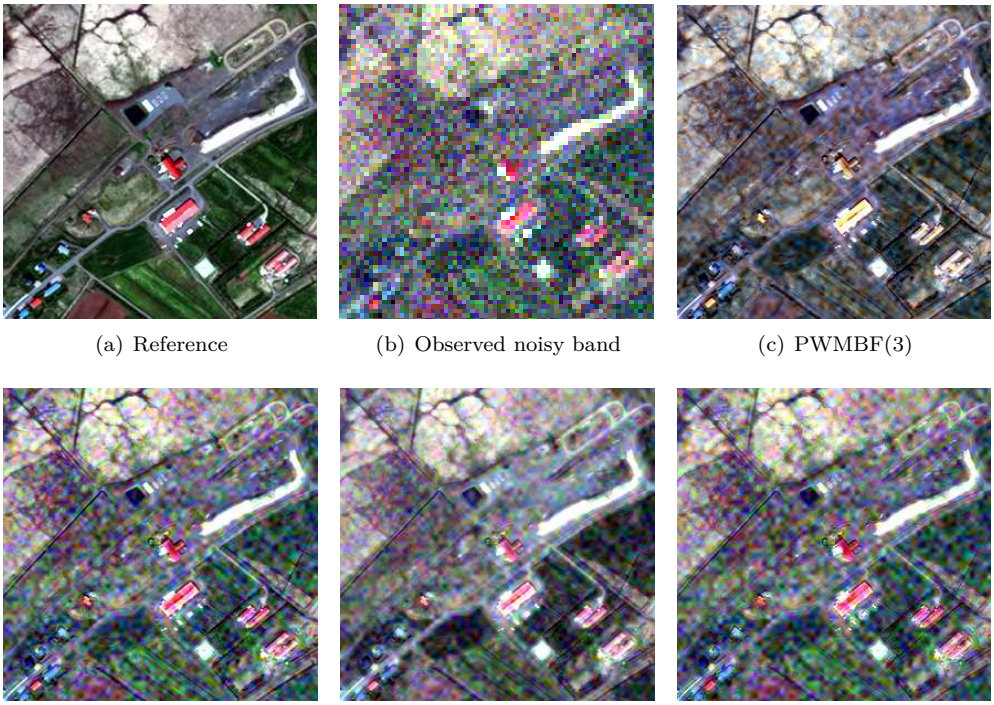
Considering Table 2.1, it is evident that the best results according to the quantitative quality metrics are obtained by the proposed method using only 5 PCs. The MAP



(a) Pan (b) Reference (c) PWMBF(5)

(d) MAP (e) PCA (f) AWLP

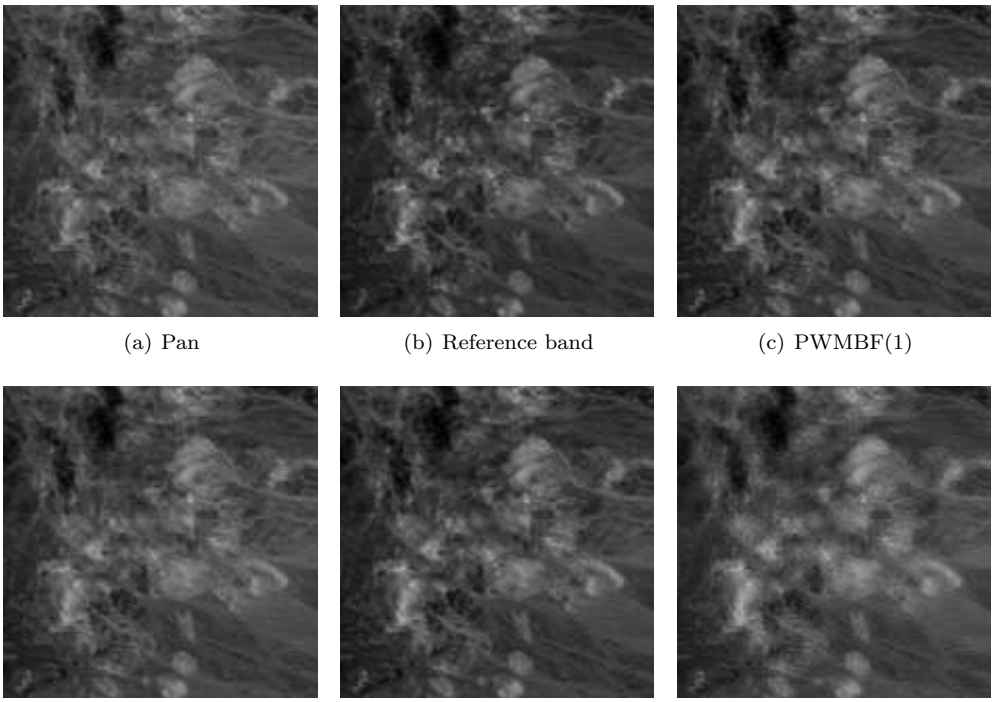
Figure 2.3: Fusion results for a portion of the noise free WorldView 2 data set for all the methods and including the PAN and reference MS image.



(a) Reference (b) Observed noisy band (c) PWMBF(3)

(d) MAP (e) PCA (f) AWLP

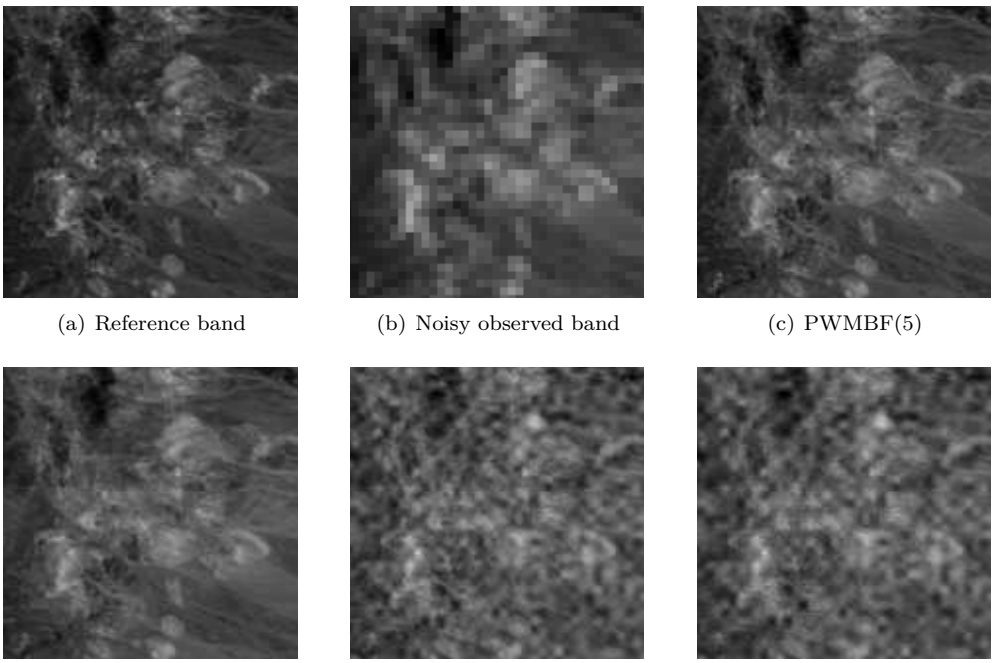
Figure 2.4: Fusion results for a portion of the WorldView 2 data set with added noise for all the methods and including noisy observed MS image and the reference.



(a) Pan (b) Reference band (c) PWMBF(1)

(d) MAP (e) PCA (f) AWLP

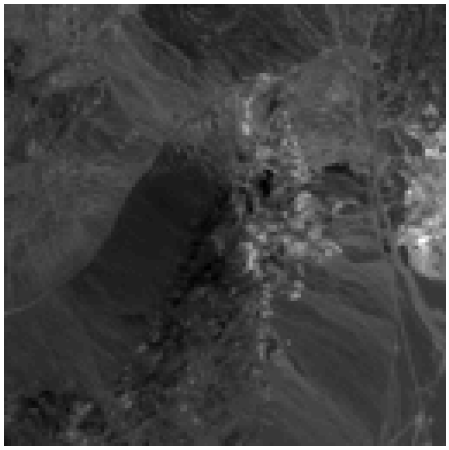
Figure 2.5: PAN/HS fusion results for a portion of the noise free AVIRIS data set for all the methods including the PAN and the reference image.



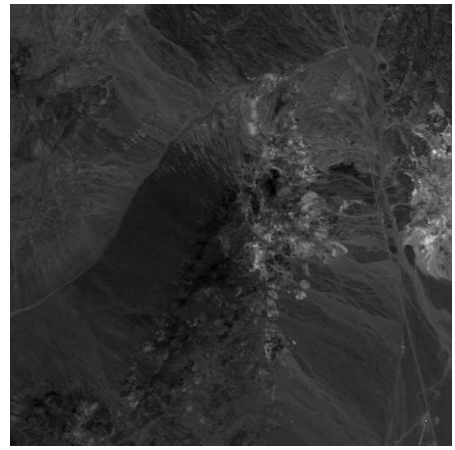
(a) Reference band (b) Noisy observed band (c) PWMBF(5)

(d) MAP (e) PCA (f) AWLP

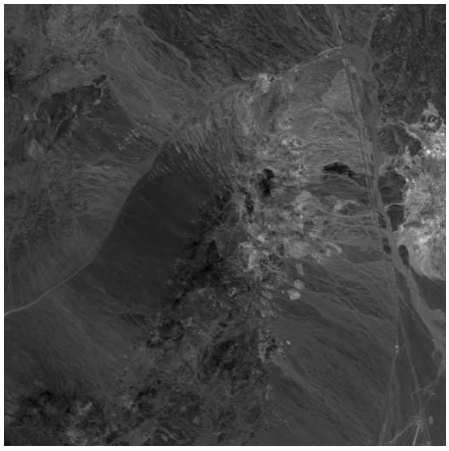
Figure 2.6: PAN/HS fusion results for a portion of the AVIRIS data set with added noise for all the methods including the noisy observed and reference bands.



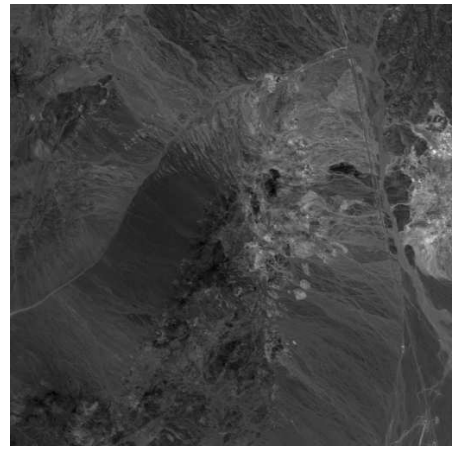
(a) Observed HS



(b) Reference HS

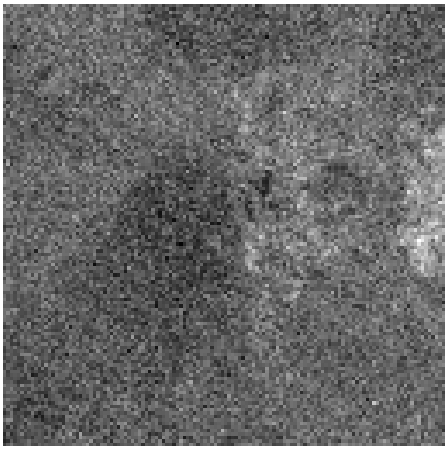


(c) PWMBF(4)

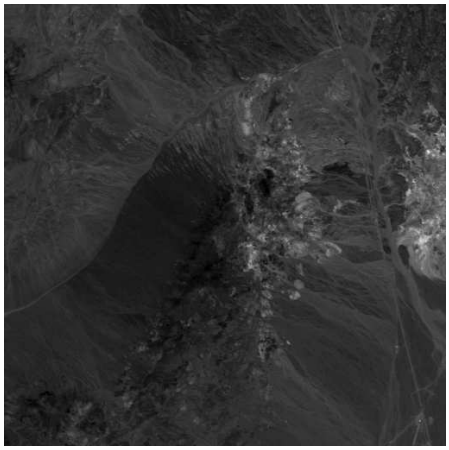


(d) MAP

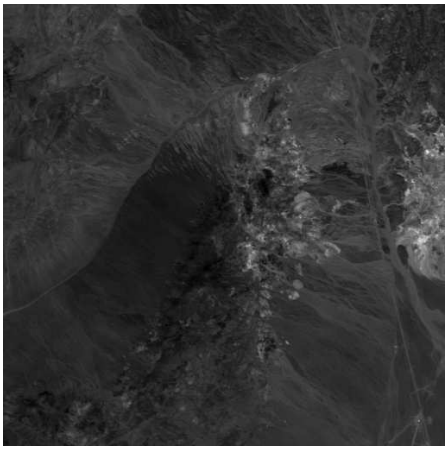
Figure 2.7: MS/HS fusion results for band 7 using the AVIRIS data, including the observed and reference bands.



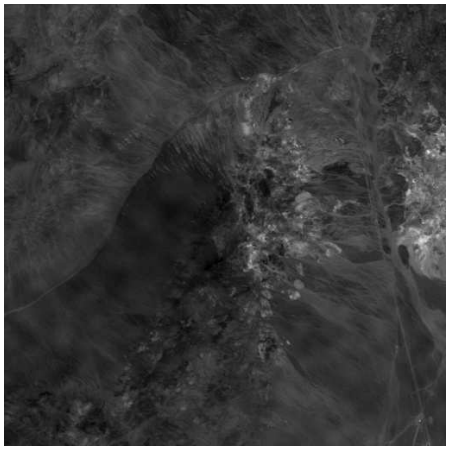
(a) Observed noisy HS



(b) Reference HS



(c) PWMBF(5)



(d) MAP

Figure 2.8: MS/HS fusion results for band 7 using the AVIRIS data with added noise, including the noisy and reference bands.

Table 2.1: Results for the PAN/MS fusion using the WorldView 2 data set. The numbers in parenthesis for the proposed method indicate the number of estimated PCs. The last column is the CPU time in sec. and the best results for each metric are highlighted using bold typeface.

Method	ERGAS	SAM	Qavg	SNR	CPU
PWMBF(4)	1.342	**1.767**	0.941	24.829	7.070
PWMBF(5)	**1.342**	1.768	**0.943**	**24.832**	8.597
MAP	1.353	1.780	0.941	24.807	12.244
PCA	3.408	2.536	0.808	17.227	4.675
AWLP	1.560	1.811	0.918	24.037	0.298
Noisy MS image : SNR=20 dB					
PWMBF(2)	2.578	3.984	**0.801**	21.313	4.082
PWMBF(3)	**2.228**	**3.379**	0.785	**22.040**	5.580
PWBBF(5)	2.405	4.027	0.753	21.355	9.009
MAP	3.014	5.043	0.687	20.190	12.305
PCA	4.321	5.383	0.656	16.078	4.672
AWLP	3.164	5.106	0.670	19.792	0.298

method comes close to giving similar results. The AWLP method gives third best results and the PCA method gives the worst results.

Using the noisy data, the proposed method using only 3 PCs (reduced PCA) gives significantly better results in terms of the metrics than the other methods. The decrease in dB compared to the results using clean data is only 2.8. As before, the MAP method gives the second best results and the PCA method gives the worst results.

In Fig. 2.3, a small subset of the fused image (using clean data) by all methods is shown as an RGB image. The proposed and MAP methods give very similar looking results, while the PCA method shows significant spectral distortion. The AWLP results seem to have less detail.

In Fig. 2.4, a small subset of the fused image (using noisy data) by all methods is shown as an RGB image. The proposed method using only 3 PCs seems to show considerable spectral distortion even though according to the quantitative quality metrics, it gives by far the best results. The apparent spectral distortion could be due to a few outliers, i.e., pixels with a large deviation from the reference. The AWLP and MAP methods seem to give similar looking results.

2.4.3 Hypersharpening using AVIRIS data

In this experiment, the fusion of a high resolution PAN image and low resolution HS image, i.e., hypersharpening, is investigated. Due to the lack of real co-registered PAN

and HS images, the PAN image has been simulated by averaging bands from the HS image. The HS image used in this experiment was acquired using the AVIRIS ariborne sensor and has a spatial resolution of 1.5m. It depicts the Cuprite mining district in Nevada, US. The spectral resolution is 219 spectral reflectance bands spanning a wavelength range of $400 - 2500$nm.

The PAN image is simulated by averaging bands 4 to 39 of the original AVIRIS HS image. This gives a PAN image that corresponds to the wavelength range of approximately $430 - 930$nm, which is typical for a PAN sensor. The HS image is also reduced to this wavelength range, giving 56 bands. To obtain the lower resolution HS image, the AVIRIS image is degraded spatially using bicubic interpolation such that the pixel ratio between the PAN and HS image becomes 16, i.e., 1 pixel of the HS image corresponds to 4×4 pixels of the PAN image. As in the previous experiment, fusion without and with added zero-mean Gaussian noise is considered.

The quantitative evaluation results for this experiment are summarized in Table 2.2 and for visual inspection, band 7 is shown in Figures 2.5 and 2.6 for all the methods using clean data and noisy data, respectively.

Table 2.2: Results for the PAN/HS fusion using the AVIRIS data set. The numbers in parenthesis for the proposed method indicate the number of estimated PCs. The last column is the CPU time in sec. and the best results for each metric are highlighted using bold typeface.

Method	ERGAS	SAM	Qavg	SNR	CPU
PWMBF(1)	**0.338**	**0.424**	**0.989**	**40.510**	1.467
PWMBF(2)	0.437	0.517	0.984	38.293	1.801
PWMBF(3)	0.445	0.550	0.984	37.877	2.121
MAP	0.471	0.578	0.982	37.503	18.770
PCA	0.430	0.424	0.987	36.825	7.866
AWLP	0.635	0.426	0.968	32.447	0.873
Noisy HS image : SNR=20 dB					
PWMBF(4)	0.525	0.712	0.977	36.449	2.431
PWMBF(5)	**0.522**	**0.711**	**0.976**	**36.478**	2.757
PWMBF(6)	0.529	0.738	0.975	36.306	3.035
MAP	0.791	1.375	0.936	32.274	17.970
PCA	2.557	4.749	0.639	21.758	8.006
AWLP	2.615	4.751	0.620	21.458	0.845

With no added noise, the proposed method gives best results based on the metrics with only one estimated PC. Estimating more PCs gives worse results. Interestingly, the PCA method gives the second best results. However, given the scree plot of the observed HS image in Figure 2.9 it is evident that the first PC explains roughly 95% of the variance in the data. The MAP method gives the third best results and the

AWLP method gives the worst results. Figure 2.5 shows a portion of the band of the fused image, including the PAN and reference image. Visual inspection shows that the proposed method gives the best details and is closest to the reference. The AWLP image (f) shows a significant loss of details.

The results using the noisy HS image demonstrate again that the proposed method significantly outperforms the other methods. The results show a drop in dB of only 4 for the proposed approach, compared to 5.2 dB, 15 dB and 11 dB for the MAP, PCA and AWLP methods, respectively. The PCA and AWLP methods perform poorly in this experiment since they are not noise resistant.

In [120], the authors chose to add the noise to the spatially interpolated observed HS image, which is good for demonstrating the noise tolerance of the method, but hardly practical in reality. It is more realistic to add the to the observed image prior to its interpolation. The interpolation of the noise makes it more coarse, i.e., it is not longer on the pixel scale, hence shrinkage of the estimated detail wavelet coefficients is not longer as effective for denoising. In the next experiment, this effect will be shown more prominently.

As was described before, in the proposed method, only the first r high resolution spatial loadings are estimated. In the presence of noise, there is the option of simply discarding the remaining noisy loadings of the interpolated HS image. This is a much more effective strategy of eliminating the interpolated noise and is the primary reason for the high noise tolerance of the proposed method.

2.4.4 MS & HS fusion using AVIRIS data

In this final experiment, MS/HS fusion is investigated. The 4-band MS image is simulated using the same 56 band HS image as in the previous experiment. Each band of the MS image is obtained by averaging consecutive 14 bands of the HS image. Finally, the HS image is spatially degraded in the same way as before, in order to be able to use the original observed HS image as the reference. The results of the quantitative evaluation of the fused HS image without and with added noise, are summarized in Table 2.3 and a visual comparison of the results is shown in Figures 2.7 and 2.8, respectively.

The results of the experiment are summarized in Table 2.3. The upper half of the table shows the results obtained using clean data. The proposed method gives best results with 4 estimated spatial loadings. The MAP method gives almost as good results according to the quantitative metrics. However, it takes much longer time to complete. In this experiment the proposed method is more than 7 times faster and it also uses far less memory resources, since the number of decomposition levels for the 2D UDWT is set equal to 4 which means it is 13 fold redundant. Figure 2.7 shows a portion of a band number 7, including the observed and reference bands. The fused bands using both methods look very similar here and it is difficult to distinguish

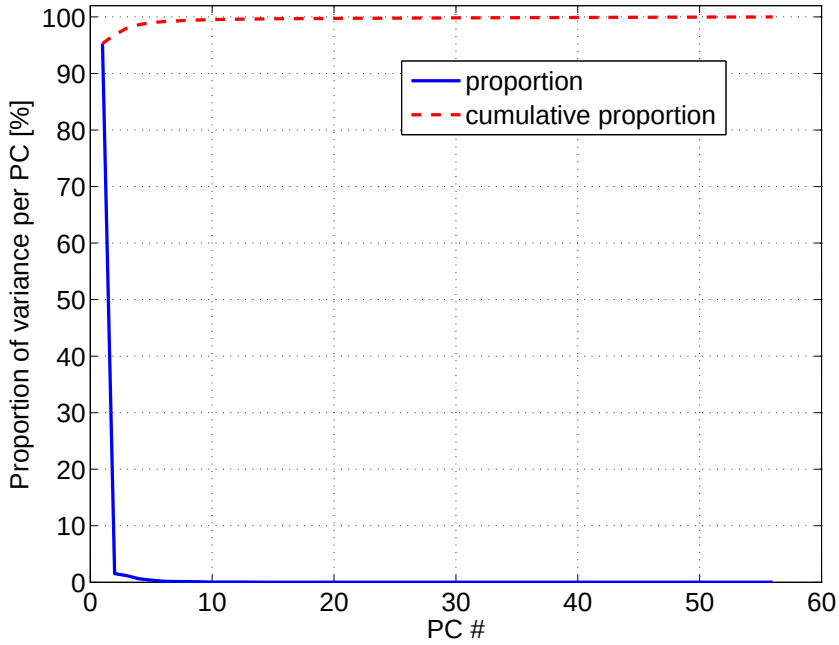


Figure 2.9: Scree plot for the observed AVIRIS HS image without added noise. The blue line shows the proportional contribution of each PC to the total variance and the red dotted line shows the cumulative variance of the PCs. Almost all the variance of the HS image is explained by the first PC.

between them, and both are very close to the reference image.

The results using the noisy data are summarized in the lower half of Table 2.3. Again, the noise tolerance of the proposed method is very good. It gives better results than the MAP method for this experiment. The drop in SNR dB because of the added noise is only 6.4 for the proposed method while it is 13.7dB for the MAP method. Visual depiction of the results for band 7 of the fused image is shown in Figure 2.8. The difference between the images obtained by the two methods is significant. The image obtained using the MAP method has a mottled appearance due to the coarse interpolated noise, which resides primarily in the LL UDWT sub-band and cannot be removed by simple shrinkage of the detail coefficients.

2.5 Conclusions

In this chapter, a method is proposed for image fusion that can handle PAN, MS and HS images. It is based on the MAP estimation of the UDWT coefficients for the PCs of the fused image. The proposed approach extends the WBBF method in [120]

Table 2.3: Results for the MS/HS fusion using the AVIRIS data set. The numbers in parenthesis for the proposed method indicate the number of estimated PCs. The last column is the CPU time in sec. and the best results for each metric are highlighted using bold typeface.

Method	ERGAS	SAM	Qavg	SNR	CPU
PWMBF(3)	0.190	0.252	0.996	46.696	2.837
PWMBF(4)	**0.172**	**0.231**	**0.997**	**47.650**	3.271
PWMBF(5)	0.174	0.234	0.997	47.552	3.711
MAP	0.184	0.244	0.996	47.152	24.382
Noisy HS image : SNR=20 dB					
PWMBF(4)	0.307	0.468	**0.991**	41.172	3.287
PWMBF(5)	**0.305**	**0.467**	0.990	**41.279**	3.717
PWMBF(6)	0.316	0.496	0.989	40.826	4.092
MAP	0.673	1.250	0.948	33.427	24.420

by employing a PCA transform of the observational model. The experiment results presented show that the proposed method gives excellent results in terms of established quantitative quality metrics, which are better than for the WBBF method and the other methods used for comparison in the experiments, especially when the observed data are noisy, while using only a fraction of the computational and memory resources of the WBBF method. The proposed method is, therefore, able to handle much larger data sets while at the same time being up to orders of magnitude faster.

Chapter 3

MTF Based Deblurring Using a Wiener Filter for CS and MRA Pansharpening Methods

CS and MRA methods can generally be described using simple detail injection frameworks which are quite similar. In this chapter, a simple pre-processing step based on deblurring the interpolated MS image is introduced, where the deblurring is based on the classical Wiener filter in the frequency domain. This can significantly enhance the quality of many of the methods tested, especially methods from the CS family, due to how the details are obtained from the PAN image. Experiments using two real data sets and a simulated data set reveal that the proposed approach gives good results for all CS methods and some MRA methods which are based on the undecimated wavelet transform.

3.1 Introduction

Most model-based pansharpening methods are based on minimizing a cost function that consists of a data fidelity term, i.e., the mean squared error between the blurred and downsampled desired high resolution MS image and the observed MS image, and some regularization terms. The iterative algorithm that minimizes this cost function, usually contains a Landweber iteration [149] term which performs deconvolution or deblurring, based on the blurring or degradation operator assumed in the model. For example of a typical model-based method see, e.g., [150].

An optical instrument, such as an MS sensor, is often described by its so-called point spread function (PSF). The modulus of the Fourier transform of the PSF, is called the modulation transfer function (MTF) and it measures the ability of the optical instrument to discern fine details based on contrast as a function of its spatial frequency. Today, most model-based pansharpening methods match the model's blurring operator to the MTF if the MS sensor [26, 125].

In this chapter, a simple pre-processing step for the CS and MRA methods is proposed. This involves performing MTF matched deblurring (deconvolution) on the interpolated MS image. The CS and MRA methods are based on similar general injection frameworks [16] (see sections 1.3.1 and 1.3.2, respectively) that add or inject details, obtained from the PAN image, into the MS bands that have been interpolated to the scale of the PAN image. This can greatly enhance both the details of the interpolated image and also enhance its spectral fidelity. In the field of image processing, deblurring

(deconvolution) is an important topic, especially for image restoration. The most well-known deblurring methods are the the Wiener filter [142] and the Richardson-Lucy deconvolution method [151]. While, the Wiener filter is optimal in the MSE sense, it can produced ringing artifacts around high contrast edges. This is also known as the Gibbs phenomenon [152] and it occurs because of the lack of high spatial frequencies in the blurred image. Several deblurring methods have been proposed to tackle this problem including [153–155].

When the blurring kernel is precisely known, many advanced deblurring methods based on wavelets and total variation [54] work well, especially on synthetic images. This is the case with the methods found in [154]. However, these methods did, in fact, not give better results for the proposed approach than the Wiener filter.

Another factor worth considering is that the CS and MRA methods, are usually very computationally efficient, making them practical, and thus it is important for the proposed approach to be computationally efficient as well. When processing large images, the processing time for the deblurring step alone, using the methods in [154], is often orders of magnitude longer than for the pansharpening process itself.

The proposed deblurring method based on the Wiener filter is very fast and gives good results in terms of quantitative quality metrics. However, it seems, that it benefits CS methods more than MRA methods for reasons that will be discussed later in Section III of the chapter. For CS methods, the deblurring pre-processing can give significant increases according to metrics for spectral and spatial quality of the fused image, and also according to visual inspection.

The deblurring kernel used for the Wiener filter is matched to the MTF of the MS imaging sensor. To decrease the ringing artifacts caused by the deblurring, homogenous areas in the image are masked prior to the deblurring operation and restored to their original state post deblurring. This technique is simple and effective and can greatly enhance the final results according visual inspection.

The CS and MRA methods are often computationally efficient which means they can handle very large images, and many of the recent methods in these categories give state-of-the-art results. By considering the general detail injection schemes for CS and MRA method, it is evident that details extracted from the PAN image are being injected or added to the interpolated MS image. The main idea behind the proposed method is to extract more spectral details from the interpolated MS image which is the spectral basis for the fused image. These details that are sometimes not fully visible in the PAN image.

To test the proposed approach, two real data sets and one semi-simulated data set are used. The CS and MRA methods used in the study are from the MATLAB [156] toolbox in [16]. All methods are evaluated using the synthesis property at the reduced scale [122] using well established quantitative quality metrics that take a reference and also using the QNR [134] indices, at the full resolution scale.

This chapter is organized as follows. Section 3.2 outlines the proposed method and in Section 3.3, the experimental results are presented and discussed, using the 3 data sets described above and 17 CS and MRA methods. Finally, in Section 3.4, conclusions are drawn.

3.2 The Proposed Method

Recall that the CS and MRA general injection schemes are given by

$$\mathbf{Z}_k = \tilde{\mathbf{Y}}_k + g_k(\mathbf{X} - \mathbf{I}_{\text{LP}}), \quad k = 1, \ldots, N_b, \tag{3.1}$$

$$\mathbf{Z}_k = \tilde{\mathbf{Y}}_k + g_k(\mathbf{X} - \mathbf{X}_L), \quad k = 1, \ldots, N_b, \tag{3.2}$$

where (3.1) is the CS injection scheme and (3.2) is the MRA scheme, respectively. The basic idea is to replace the interpolated kth band of $\tilde{\mathbf{Y}}_k$ with a deblurred version, $\mathbf{Y}_{\text{MTF}_k}$, which is obtained using a Wiener filter in the frequency domain.

The blurring filters used in this work are 2D Gaussian low-pass filters whose gain at the Nyquist frequency is matched to the gain of the MTF of the MS sensor at the same spatial frequency. The computation of the Wiener [142] filter is straight-forward since the blurring kernel is precisely known. The deblurred kth band of the interpolated MS image, $\tilde{\mathbf{Y}}_k$, in the frequency domain is given by

$$\boldsymbol{\mathcal{Y}}_{\text{MTF}_k} = \frac{\boldsymbol{\mathcal{M}}_k^*}{|\boldsymbol{\mathcal{M}}_k|^2 + \frac{1}{\text{SNR}}} \cdot \boldsymbol{\mathcal{Y}}_{\boldsymbol{k}}. \tag{3.3}$$

where $\boldsymbol{\mathcal{Y}}_{\boldsymbol{k}}$ is the 2D Fourier transform of $\tilde{\mathbf{Y}}_k$, $\boldsymbol{\mathcal{M}}_k$ is the 2D Fourier transform of the blurring kernel for the kth MS band, $\frac{1}{\text{SNR}}$ is the noise-to-signal ratio, i.e., $\frac{\sigma_k^2}{\boldsymbol{\mathcal{Y}}_{xx}}$, where $\boldsymbol{\mathcal{Y}}_{xx}$ is spectral density of $\boldsymbol{\mathcal{Y}}_{\text{MTF}_k}$, which is not known, and σ_k^2 is the power spectral density of the zero mean Gaussian additive noise of $\boldsymbol{\mathcal{Y}}_{\text{MTF}_k}$. $\boldsymbol{\mathcal{M}}_k^*$ is the complex conjugate of $\boldsymbol{\mathcal{M}}_k$ and $|\boldsymbol{\mathcal{M}}_k|^2$ is the squared magnitude of $\boldsymbol{\mathcal{M}}_k$, i.e. the 2D Fourier coefficients of $\mathbf{M}_k$. Since $\boldsymbol{\mathcal{Y}}_{xx}$ is unknown, the term $\frac{1}{\text{SNR}}$ is often approximated by a small constant [157].

The deblurred band $\mathbf{Y}_{\text{MTF}_k}$ is then obtained by the inverse 2D Fourier transform of $\boldsymbol{\mathcal{Y}}_{\text{MTF}_k}$. The division multiplication and division is performed in an element-wise manner.

Ringing artifacts can be expected due to non-periodic boundaries when using frequency domain methods such as the Wiener filter. To avoid these, edge tapering is used on $\tilde{\mathbf{Y}}_k$. This means that it is blurred using $\mathbf{M}_k$, and then the blurred version is replaced with the original except for an area at the edges which has width equal to the kernel size.

The biggest drawback of frequency domain deblurring is the Gibbs phenomenon which shows up as ripples or ringing artifacts around sharp high contrast discontinuities in the image to be deblurred. These ripples become especially apparent when there are sharp edges surrounded by homogenous areas. To decrease this ringing, we used a simple deringing technique which is applied as a post-processing step after the deblurring operation. We use a sliding window standard deviation filter to obtain a map of the standard deviation of the 3 by 3 pixel neighborhood of every pixel of the interpolated MS band $\tilde{\mathbf{Y}}_k$. By selecting pixels which have a standard deviation less than some threshold, one can obtain a binary image that shows areas with low standard deviation in a 3 by 3 pixel neighborhood. By substituting these pixels in the deblurred image $\mathbf{Y}_{\mathrm{MTF}_k}$ with pixels from the original image $\tilde{\mathbf{Y}}_k$, ringing artifacts in homogenous areas, where they are most visible, can be eliminated. In the experiments the standard deviation threshold was set equal to 4.

When we have obtained the MTF-deblurred kth band of the interpolated MS image, denoted $\mathbf{Y}_{\mathrm{MTF}_k}$, using the Wiener filter, the respective CS and MRA injection schemes become

$$\mathbf{Z}_k = \mathbf{Y}_{\mathrm{MTF}_k} + g_k(\mathbf{X} - \mathbf{I}_{\mathrm{MTF}}), \quad k = 1, \ldots, N_b \tag{3.4}$$

$$\mathbf{Z}_k = \mathbf{Y}_{\mathrm{MTF}_k} + g_k(\mathbf{X} - \mathbf{X}_L), \quad k = 1, \ldots, N_b, \tag{3.5}$$

where the only difference is that we have replaced $\tilde{\mathbf{Y}}_k$ with $\mathbf{Y}_{\mathrm{MTF}_k}$ in (3.1) and (3.2), respectively, and now the intensity image $\mathbf{I}_{\mathrm{MTF}}$ is a linear combination of the bands of $\mathbf{Y}_{\mathrm{MTF}_k}$ instead of the bands of $\tilde{\mathbf{Y}}_k$.

As will be demonstrated in the experiments section, it is evident that the proposed method works better in general for CS methods than for the MRA methods. This is because of the difference between the CS and MRA injection schemes. The CS injection framework for the proposed method, after rearranging terms, is given by

$$\mathbf{Z}_k = \mathbf{Y}_{\mathrm{MTF}_k} - g_k\mathbf{I}_{\mathrm{LP}} + g_k\mathbf{X}, \quad k = 1, \ldots, N_b.$$

For CS methods the injection gains g_k are usually larger than 0.5 and often close to 1. This means that halos and blurs associated with the deblurred image (ringing artifacts) are partially canceled out according to the term $\mathbf{Y}_{\mathrm{MTF}_k} - g_k\mathbf{I}_{\mathrm{LP}}$. Now consider the MRA scheme from (3.5). It is evident that details extracted from the PAN image by removing its low frequency component are added directly to the deblurred interpolated MS image $\mathbf{Y}_{\mathrm{MTF}_k}$. Thus, if there are considerable ringing artifacts in $\mathbf{Y}_{\mathrm{MTF}_k}$, they become much more prominent and hence the degradation in quality for MRA methods which are based on (3.5). Another way of expressing this difference, is that for the MRA methods, some spatial frequencies might get injected twice, since the deblurred image $\mathbf{Y}_{\mathrm{MTF}}$ contains higher frequencies than the interpolated MS image $\tilde{\mathbf{Y}}$. Therefore, the reason why the proposed method works better in general for CS methods is exactly the same that makes CS methods more effective when original data are affected by aliasing or mis-registration problems [35].

3.3 Experiment Results

This section is divided into to subsections. In the first subsection, experiment results using three data sets are presented and discussed. The data sets used for the experiments are the WV2, QB and simulated Pléiades data sets that were described in Section 1.6. The performance of all methods used is evaluated using the synthesis property at the reduced scale (see Section 1.5.1 and at the full resolution scale using the QNR metrics (see Section 1.5.2). The results are presented where only the original data has been used alone and where the proposed pre-processing step has been used. The relative gains (in percentages) in individual quality metrics are also shown for each method.

In the second subsection, some interesting observations from the results are discussed in more detail and also a related approach found in [26].

The method implementations used in this chapter come from a comprehensive Matlab toolbox[1], that was published along with paper [16]. From this toolbox, 8 CS methods and 9 MRA methods are used in the experiments, i.e., all CS and MRA methods that take the interpolated MS image is input. An overview of the methods is given in Appendix A.

In Figure 3.1, the interpolated MS images for all three data sets are shown. The yellow rectangle depicts the region of interest that is later used in figures 5.3, 5.4 and 3.4, to show in more detail the comparison between the unmodified pansharpening methods and the modified versions using the proposed method.

3.3.1 Experiment results for all data sets

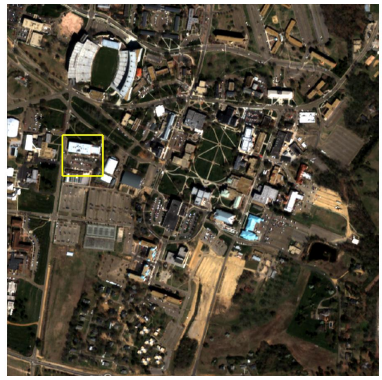
Experiment results using WV2 data set

The results for all methods evaluated for the WV2 data set using the synthesis property at the degraded scale is summarized in Table 3.1. The column of the table are grouped by the quality metric used and for each metric there is another group of 3 columns labeled "Interp.", which are the results obtained using unmodified data, "Deblurring", which gives the results using the proposed approach, and "Change (%)", which gives the relative change in percentages of the metric score after the deblurring operation. A positive change means that the metric score has improved, while a negative value means that the deblurring operation has degraded the fusion quality according to the metric. The results using the synthesis QNR metrics for the WV2 data set are summarized in 3.2 and are presented in a similar manner.

The CS methods are in the upper part of the table while the MRA methods are in the lower part. The EXP method is the interpolated or deblurred MS image itself, i.e.,

[1]The toolbox is available from:
`http://www.openremotesensing.net/index.php/codes/11-pansharpening`

(a) WorldView-2 (4096 × 4096 × 8)



(b) QuickBird (2048 × 2048 × 4)

(c) Pléiades (1024 × 1024 × 4)

Figure 3.1: The interpolated MS images from all three data sets. The yellow rectangles show regions of interest that will be used later to visually demonstrate the proposed method at the full resolution scale.

no pansharpening has been performed. The two methods of each family showing the biggest gains in metric scores are indicated using a bold typeface. For the ERGAS and SAM metrics, lower values are better while for the Q4/Q8 metric, higher values are better with 1 being optimal.

According to Table 3.1, all CS methods except BDSD [15] show gains in all metrics when using the proposed approach. However, only 2 MRA methods show improvement, i.e., the ATWT-M2 [25, 28] and ATWT-M3 [25, 28] methods. This is not surprising, since in Section III, the differences between the two injection schemes were detailed. Generally, MRA methods do not benefit from the proposed method due their inability to cancel out artifacts produced by the deblurring operation. Of the CS methods, the GSA [20] and PRACS [158] methods show the greatest overall improvement. The reason behind the quite significant improvement of the ATWT-M2 and -M3 methods due to the proposed approach,is given in the last subsection in this

section.

Table 3.2 summarizes the results obtained using the synthesis QNR metrics. Note that for the D_λ and D_S indices, lower values are better, while for the combined QNR metric, higher values are better with 1 being optimal.

The results for the synthesis QNR metrics are summarized in Table 3.2. Now the trend is different from the results obtained using the synthesis property at the degraded scale. All methods except the MRA methods, ATWT-M2 and ATWT-M3, show improvements for the combined QNR metrics. This seems odd, since those two methods showed the greatest improvement according to Table 3.1. Of the CS methods, the BDSD and GSA method show the greatest improvement while for the MRA methods, the AWLP and MTF-GLP-CBD [32] methods benefit the most using the proposed approach. Visual inspection of the images obtained by all MRA methods, except ATWT-M2 and ATWT-M3, show that they are indeed corrupted by artifacts for reasons detailed above. Hence, this indicates that the QNR metrics are not reliable in some cases. However, it is apparent from the Table, that the MRA methods are generally not as much improved by the proposed approach as the CS methods.

Figure 5.3 depicts a small region of images obtained by the most improved methods for both synthesis at the reduced scaled and the synthesis QNR metrics. In the first row of the figure, the PAN image, interpolated MS image and the deblurred interpolated MS image are shown. In the second and third row, the most improved images from the CS methods are shown and the last two rows show images obtained using the most improved MRA methods. It is clear, especially for the CS methods, that the images obtained using the proposed method are more detailed, have richer colors and are sharper. A good example is the building in the lower right corner with the striped roof. The images obtained using the proposed approach resolve the stripes much better.

Figure 3.4 shows in a visual manner the results obtained using the ERGAS metric for synthesis at the reduced scale for all the data sets used in the experiments. The most important observation is that the results are consistent across different data sets. The CS methods are generally enhanced, with the exception of BDSD (at the reduced scale). However, at full scale, the BDSD method is indeed greatly enhanced. Also, there is the trend discussed before, the only MRA methods that benefit from the proposed method according to synthesis at the reduced scale are the ATWT-M2 and -M3 MRA methods. Why BDSD is not enhanced according to synthesis at the reduced scale and why the two ATWT methods are the only MRA enhanced methods, will be discussed in detail at the end of this section.

Table 3.1: WV2 - synthesis at the degraded scale

	ERGAS			SAM			Q8		
Method	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)
EXP	2.390	2.090	12.552	2.036	1.842	9.551	0.802	0.832	3.676
PCA	4.313	4.247	1.549	5.587	5.409	3.181	0.784	0.796	**1.608**
IHS	2.523	2.482	1.625	2.604	2.427	6.798	0.676	0.667	-1.449
Brovey	2.250	2.182	3.018	2.036	1.842	**9.551**	0.737	0.730	-0.882
BDSD	1.522	1.687	-10.869	2.216	2.344	-5.763	0.890	0.865	-2.898
GS	2.355	2.207	6.296	2.890	2.643	8.552	0.829	0.839	1.267
GSA	1.499	1.339	**10.669**	1.902	1.731	8.956	0.911	0.916	0.571
PRACS	1.829	1.569	**14.203**	1.939	1.766	8.908	0.885	0.899	**1.581**
HPF	1.693	1.669	1.406	2.036	1.957	3.866	0.891	0.892	0.146
SFIM	1.568	1.511	3.660	1.935	1.820	5.968	0.892	0.894	0.179
ATWT	1.678	1.857	-10.712	2.045	2.048	-0.132	0.897	0.890	-0.703
AWLP	1.854	2.078	-12.095	2.313	2.430	-5.059	0.895	0.885	-1.073
ATWT-M2	2.138	1.685	**21.209**	2.303	2.035	**11.645**	0.854	0.892	**4.378**
ATWT-M3	2.049	1.676	**18.220**	2.382	2.119	**11.014**	0.862	0.881	**2.169**
MTF-GLP	1.677	1.824	-8.734	2.042	2.044	-0.073	0.900	0.894	-0.678
MTF-GLP-HPM	1.466	1.632	-11.295	1.920	1.857	3.286	0.902	0.897	-0.643
MTF-GLP-CBD	1.493	1.533	-2.693	1.896	1.814	4.346	0.912	0.911	-0.132

Experiment results for the QuickBird (QB) data set

Tables 3.3 and 3.4 summarize the results for the QB data set for the synthesis property at the reduced scale and synthesis QNR metrics, respectively. Apparently, the metric scores are generally worse for this data set than for the WV2 data set and this is due to a small misalignment between the PAN and MS images. However, the results show the same trend as for the previous experiment.

Considering the synthesis results in Table 3.3, all CS methods are improved except BDSD. The two most enhanced CS methods are the PCA and GS methods, closely followed by GSA and again the most enhanced MRA methods are the ATWT-M2 and ATWT-M3 methods. The other MRA methods show little or negative improvement.

The results obtained using the synthesis QNR metric is given in Table 3.4. Again the same behavior is observed as seen in the previous experiment. All CS methods benefit from the deblurring and all MRA methods except the ATWT-M2 and -M3 methods, which show negative improvement. The most improved CS methods according to QNR metrics are the PCA and GSA methods, while the MTF-GLP-HPM [159] and SFIM [160] methods display the greatest gains in the MRA family.

Figure 5.4 shows the results obtained using imagery at the full scale, where the focus is on the region shown in Figure 3.1 (b). This figure is arranged in the same manner as Figure 5.3. The region depicts a building with an orange stripe and a parking lot with cars of various colors. The results obtained using the proposed method look sharper and the colors are more enhanced and look less washed out. This is especially apparent in the orange stripe on the building and for the red cars in the parking lot. The deblurred MS image (c) shows this clearly. Hence, these details are not obtained from the PAN image. This highlights well the benefits of the proposed approach. I.e., small details like cars are better resolved and the colors are more concentrated. In Figure 3.4 (b), the results for the ERGAS metric using synthesis at the reduced scale are summarized visually, and the same trend as before for the WV2 data set is clearly visible.

Experiment results for the Pléiades (PL) data set

As with the previous data set there is a small misalignment between the MS and PAN image, which is reflected in somewhat worse scores for the quality metrics. The results using the synthesis property at the degraded scale and the synthesis QNR metrics are summarized in Tables 3.5 and 3.6, respectively. Again the same pattern is observed, i.e., for the synthesis at reduced scale, all CS methods show improvement except the BDSD method and the only improved MRA methods are the ATWT-M2 and -M3 methods.

Table 3.6 summarizes the results using the QNR metrics. For this data set, the pattern is somewhat different from the previous data sets regarding the MRA methods. As before, all CS methods are improved using the proposed approach, especially the

Table 3.2: WV2 - synthesis QNR metrics

	D_λ			D_S			QNR		
Method	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)
EXP	0.002	0.008	-320.000	0.061	0.063	-3.279	0.937	0.929	-0.854
PCA	0.056	0.057	-2.334	0.139	0.128	7.692	0.813	0.822	1.107
IHS	0.084	0.069	17.696	0.094	0.082	13.256	0.830	0.855	3.026
Brovey	0.082	0.064	22.722	0.097	0.086	11.740	0.829	0.856	3.331
BDSD	0.050	0.038	24.851	0.079	0.055	**30.847**	0.875	0.910	**3.979**
GS	0.045	0.032	**28.508**	0.090	0.090	-0.222	0.869	0.881	1.323
GSA	0.056	0.036	**36.542**	0.088	0.063	**28.311**	0.861	0.904	**4.958**
PRACS	0.012	0.008	30.000	0.054	0.044	18.097	0.935	0.948	1.390
HPF	0.058	0.050	**14.580**	0.073	0.062	14.560	0.873	0.891	2.062
SFIM	0.061	0.053	14.029	0.074	0.064	14.189	0.869	0.887	2.071
ATWT	0.073	0.064	11.433	0.093	0.080	14.270	0.841	0.861	**2.378**
AWLP	0.049	0.042	13.292	0.061	0.049	**19.275**	0.893	0.911	1.936
ATWT-M2	0.040	0.049	-25.063	0.052	0.062	-20.309	0.911	0.892	-2.130
ATWT-M3	0.052	0.059	-13.628	0.048	0.065	-36.211	0.903	0.880	-2.548
MTF-GLP	0.072	0.063	11.961	0.093	0.079	14.162	0.842	0.862	**2.386**
MTF-GLP-HPM	0.074	0.066	11.471	0.093	0.080	14.054	0.840	0.860	2.356
MTF-GLP-CBD	0.056	0.048	**15.330**	0.080	0.068	**14.981**	0.868	0.888	2.223

Table 3.3: QuickBird - synthesis at the degraded scale

	ERGAS			SAM			Q4		
Method	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)
EXP	4.283	3.768	12.035	4.369	3.773	13.641	0.780	0.839	7.565
PCA	3.937	2.743	**30.314**	5.153	3.695	**28.292**	0.819	0.898	**9.645**
IHS	3.909	3.480	10.987	5.174	4.499	13.045	0.803	0.828	3.115
Brovey	3.578	3.173	11.323	4.369	3.773	13.641	0.841	0.866	2.950
BDSD	3.004	3.462	-15.259	4.506	5.013	-11.249	0.892	0.871	-2.321
GS	3.782	3.245	**14.216**	4.942	4.031	**18.440**	0.830	0.867	**4.469**
GSA	3.034	2.674	11.849	4.358	3.740	14.168	0.881	0.905	2.747
PRACS	2.959	2.615	11.639	4.064	3.563	12.327	0.894	0.919	2.830
HPF	3.041	3.118	-2.529	4.314	4.094	5.109	0.883	0.889	0.725
SFIM	3.040	3.192	-4.994	4.275	3.781	11.542	0.889	0.893	0.495
ATWT	2.921	3.383	-15.805	4.285	4.405	-2.810	0.889	0.881	-0.922
AWLP	2.819	3.436	-21.897	3.877	4.202	-8.391	0.902	0.889	-1.453
ATWT-M2	3.848	2.897	**24.699**	4.976	4.037	**18.867**	0.820	0.892	**8.776**
ATWT-M3	3.691	2.943	**20.275**	5.117	4.267	**16.597**	0.838	0.888	**5.994**
MTF-GLP	2.902	3.405	-17.333	4.216	4.562	-8.221	0.889	0.879	-1.092
MTF-GLP-HPM	2.893	3.502	-21.042	4.155	3.867	6.927	0.896	0.884	-1.373
MTF-GLP-CBD	2.912	3.422	-17.525	4.210	4.577	-8.705	0.889	0.879	-1.147

Table 3.4: QuickBird - synthesis QNR metrics

	D_λ			D_S			QNR		
Method	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)
EXP	0.002	0.014	-652.632	0.064	0.067	-5.172	0.934	0.920	-1.584
PCA	0.046	0.009	**80.477**	0.091	0.040	**56.092**	0.867	0.951	**9.723**
IHS	0.092	0.043	53.855	0.166	0.143	13.993	0.757	0.821	8.398
Brovey	0.068	0.026	62.003	0.129	0.115	10.939	0.812	0.862	6.208
BDSD	0.029	0.022	24.830	0.071	0.023	**67.227**	0.901	0.955	5.946
GS	0.069	0.027	61.705	0.148	0.124	16.240	0.793	0.853	7.544
GSA	0.071	0.016	**77.606**	0.126	0.064	49.723	0.812	0.922	**13.527**
PRACS	0.026	0.016	40.154	0.050	0.029	42.600	0.925	0.956	3.339
HPF	0.057	0.039	**31.095**	0.077	0.071	8.269	0.870	0.893	2.574
SFIM	0.057	0.040	28.496	0.079	0.065	**17.871**	0.869	0.897	**3.256**
ATWT	0.074	0.057	23.822	0.108	0.097	10.093	0.826	0.852	3.161
AWLP	0.061	0.051	16.749	0.088	0.084	4.670	0.857	0.870	1.553
ATWT-M2	0.034	0.057	-68.639	0.030	0.084	-175.000	0.937	0.864	-7.750
ATWT-M3	0.042	0.053	-23.821	0.032	0.065	-104.389	0.927	0.886	-4.455
MTF-GLP	0.077	0.058	23.958	0.109	0.100	8.775	0.822	0.848	3.089
MTF-GLP-HPM	0.076	0.053	30.423	0.108	0.090	**16.156**	0.825	0.862	**4.486**
MTF-GLP-CBD	0.060	0.042	**30.795**	0.092	0.082	11.304	0.853	0.880	3.141

BDSD and GSA methods. However, for the MRA methods, all methods are improved for this data set and the most improved MRA methods in terms of the QNR metric is the ATWT-M2 method. According to the table, all methods are improved spectrally according to the D_λ sub-index. However, the two ATWT-M methods are the only MRA methods that show a decrease in the D_S spatial index, which is odd, since these methods are greatly enhanced spatially by the proposed approach.

Figure 3.4 shows the region depicted in Figure 3.1 (c) of the images obtained by the most enhanced methods according to both synthesis at the reduced scaled and synthesis QNR metrics. It is arranged in the same way as for the previous data sets. The GSA, GS and BDSD methods represent the CS methods and the ATWT-M2, HPF and SFIM methods represent the MRA methods. The image shows a V shaped intersection of streets and a triangular building, some cars and there is also a large tree present. the enhanced images look noticeably sharper and the colors are more concentrated, especially for small objects like cars. The enhancement of the ATWT-M2 image is especially impressive. This method benefits greatly from the proposed approach. Another observation is that aliasing artifacts are clearly present in the images obtained using the proposed approach for the HPF and SFIM methods. This was discussed at the end of Section III.

3.3.2 Discussion about the experiment results

The fact that the CS method BDSD is not enhanced by the proposed approach when using synthesis at the degraded scale, is perhaps the most obvious inconsistency in the presented results using all the data sets. However, visual inspection of the results using this method for full scale imagery clearly shows enhanced results. This is due to the implementation of the method. It is based on the joint estimation of the injection gains g_k and the weights of the linear combination of the bands of the interpolated MS image, i.e., the intensity image $\mathbf{I}_{\text{LP}}$, (which is estimated individually for each band, rather than for all bands) such that the MSE between the MTF degraded version of $\mathbf{X}_k$ and the bi-cubically downsampled interpolated MS image is minimized. The fact that the MSE is calculated using the bi-cubically downsampled interpolated MS image, instead of the actual observed MS image is the reason that the proposed method does not work well when using reduced scale imagery, since the already spatially degraded MS image is degraded again and interpolated, when using synthesis at the reduced scale.

Another observation is that only two MRA methods are enhanced, according to the quality metrics, by the proposed approach when using synthesis at the reduced scale. These are the ARSIS (Amélioration de la Résolution Spatiale par Injection de Structures) ATWT-M2 and ATWT-M3 methods [25, 28]. The reason for this, is that while all the other MRA methods are implemented to adhere to MRA injection framework, these methods are based on a non-decimated wavelet transform (ATWT). They are based on the ATWT approach as presented in [28] and the detail injection models proposed in [25]. ATWT-M3 is also proposed in [26].

The general approach is as follows. Each interpolated/deblurred MS band and the PAN image are decomposed into approximation and detail coefficients. The detail coefficients of the PAN image are transformed using an affine transformation whose parameters are estimated from the wavelet coefficients according to a simple model. The next step is to replace the detail coefficients of the interpolated/deblurred MS image with the transformed PAN details. The fused image is then obtained using the inverse transform. For a ratio of 4 between the MS and PAN image, 2 levels of decompositions are used. This means that the approximation coefficients of the interpolated/deblurred MS image are obtained by low-pass filtering it twice. This is the exact reason for the efficiency of the proposed approach in enhancing this type of method.

Yet another interesting observation regarding the results obtained using the QNR synthesis metric is that for the MRA methods, the results are often complementary to what was obtained using synthesis at the reduced scaled. This means that all MRA methods are enhanced by the proposed approach except for the two ARSIS methods, i.e., ATWT-M2 and ATWT-M3. It must be noted, however, that this was not the case for the Pléiades data set. The reason for this great discrepancy between the synthesis QNR results and synthesis at the reduced scale are not clear. However, by inspecting the results in tables 3.2 and 3.4, it is clear that it the D_{S} index score of the composite QNR index that contributes mostly the the decrease of the QNR score for those methods. This is odd, since these methods are most spatially enhanced by the proposed approach. Perhaps, the QNR synthesis metric does not give a good idea of the spatial quality of the fused image.

Finally, a method is proposed in [26] which shares some similarities with the proposed approach. In [26] the authors propose that the synthesized image should have an MTF that is the same as the MS MTF for low frequencies (relative spatial frequencies below 0.125) but for higher frequencies it should have the MTF of the PAN sensor. This is implemented by deconvolution of the observed MS image with the MS MTF and then uspsampling the deconvolved MS image using spline filters and finally convolving the interpolated image with the PAN MTF.

3.4 Conclusions

In this chapter, a deblurring approach for the enhancement of CS and MRA methods was proposed. The CS and MRA methods are usually simple, fast and effective and the majority of the methods can be described by a general detail injection scheme that is easy to implement. However, many CS and MRA methods produce images where spectral and or spatial distortion is present. The approach proposed here aims to increase the quality of the fused image produced by the CS and MRA methods, by applying a Wiener filter based deblurring pre-processing to the interpolated MS image that is then fused with the PAN image. The deblurring is performed using a Wiener filter, which only requires knowledge of the MTF blurring kernel and an estimate of the noise-to-signal ratio of the original MS image, which can be approximated by a small

Table 3.5: Pléiades - synthesis at the degraded scale

	ERGAS			SAM			Q4		
Method	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)
EXP	5.855	5.312	9.272	4.663	4.406	5.524	0.632	0.728	15.344
PCA	5.425	5.081	6.342	5.865	5.652	3.640	0.784	0.846	**7.880**
IHS	4.743	4.261	10.153	4.896	4.844	1.060	0.810	0.852	5.235
Brovey	4.780	4.231	**11.486**	4.663	4.406	5.524	0.806	0.858	6.424
BDSD	3.225	4.630	-43.601	4.745	5.739	-20.965	0.940	0.893	-5.030
GS	4.719	4.119	**12.722**	4.759	4.572	3.927	0.807	0.873	**8.189**
GSA	3.249	3.144	3.244	5.336	4.624	**13.346**	0.941	0.943	0.287
PRACS	3.250	2.964	8.789	5.084	4.683	**7.897**	0.931	0.945	1.504
HPF	3.603	3.363	6.664	4.312	4.572	-6.044	0.897	0.932	3.904
SFIM	3.601	3.407	5.366	4.276	4.200	1.768	0.896	0.930	3.863
ATWT	3.226	3.376	-4.647	4.174	4.772	-14.317	0.924	0.939	1.646
AWLP	3.532	3.922	-11.032	4.719	5.621	-19.103	0.918	0.930	1.296
ATWT-M2	5.040	3.621	**28.156**	5.047	4.320	**14.392**	0.739	0.889	**20.371**
ATWT-M3	4.604	3.511	**23.729**	5.159	4.445	**13.839**	0.805	0.899	**11.676**
MTF-GLP	3.107	3.223	-3.737	4.159	4.716	-13.395	0.931	0.946	1.600
MTF-GLP-HPM	3.084	3.379	-9.578	4.174	4.201	-0.649	0.932	0.943	1.105
MTF-GLP-CBD	3.156	4.274	-35.437	5.221	5.477	-4.911	0.943	0.912	-3.278

Table 3.6: Pléiades - synthesis QNR metrics

	D_λ			D_S			QNR		
Method	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)	Interp.	Deblurring	Change (%)
EXP	0.001	0.012	-1816.667	0.154	0.132	14.137	0.845	0.858	1.467
PCA	0.033	0.023	29.518	0.148	0.133	9.817	0.824	0.847	2.730
IHS	0.032	0.036	-12.658	0.089	0.077	12.867	0.883	0.890	0.827
Brovey	0.023	0.021	9.292	0.097	0.081	16.770	0.882	0.900	2.029
BDSD	0.036	0.011	**69.553**	0.053	0.038	**29.401**	0.913	0.952	**4.284**
GS	0.012	0.010	22.764	0.103	0.082	20.117	0.886	0.909	2.596
GSA	0.072	0.030	**57.660**	0.077	0.056	**28.202**	0.856	0.916	**6.936**
PRACS	0.034	0.016	52.047	0.039	0.032	17.010	0.928	0.952	2.542
HPF	0.036	0.028	**22.928**	0.044	0.017	**60.998**	0.921	0.955	**3.701**
SFIM	0.040	0.032	**19.500**	0.043	0.019	56.221	0.918	0.950	3.386
ATWT	0.047	0.039	15.699	0.037	0.024	33.880	0.919	0.938	2.057
AWLP	0.032	0.028	12.893	0.040	0.021	46.851	0.930	0.952	2.355
ATWT-M2	0.023	0.043	-91.593	0.102	0.040	**61.057**	0.878	0.919	**4.684**
ATWT-M3	0.033	0.043	-30.887	0.073	0.033	54.897	0.897	0.926	3.199
MTF-GLP	0.046	0.039	16.164	0.039	0.025	34.286	0.917	0.937	2.159
MTF-GLP-HPM	0.052	0.045	13.269	0.040	0.028	29.397	0.910	0.928	1.967
MTF-GLP-CBD	0.059	0.048	18.399	0.048	0.032	33.264	0.896	0.921	2.869

(a) PAN

(b) Interpolated MS

(c) Deblurred MS

(d) GSA

(e) PRACS

(f) BDSD

(g) Deblurred GSA

(h) Deblurred PRACS

(i) Deblurred BDSD

(j) ATWT-M2

(k) ATWT

(l) MTF-GLP-CBD

(m) Deblurred ATWT-M2

(n) Deblurred ATWT

(o) Deblurred MTF-GLP-CBD

Figure 3.2: Visual inspection at the full scale of a sub-region of the most improved methods for the WV2 data set according to the synthesis property at reduced scale and QNR metrics. The 3 most improved methods from each family are shown. The first row shows the PAN image, the interpolated MS image and the MTF deblurred version of the interpolated image. The next 2 rows show results obtained by 3 CS methods using the interpolated image and then the deblurred image. Similarly, the last two rows show 3 MRA methods. The "Deblurred" prefix denotes images obtained using the proposed method.

constant. Experiments using three data sets from 3 different sensors demonstrate that the proposed approach can significantly improve the images produced by CS methods and only in some cases, MRA methods.

(a) PAN

(b) Interpolated MS

(c) Deblurred MS

(d) PCA

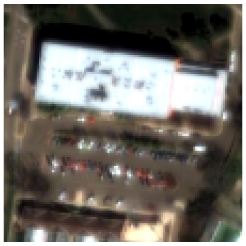
(e) GS

(f) GSA

(g) Deblurred PCA

(h) Deblurred GS

(i) Deblurred GSA



(j) ATWT-M2

(k) ATWT-M3

(l) MTF-GLP-HPM

(m) Deblurred ATWT-M2

(n) Deblurred ATWT-M3

(o) Deblurred MTF-GLP-HPM

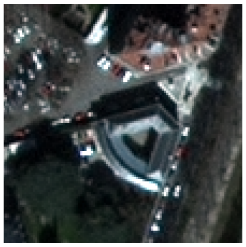
Figure 3.3: Visual inspection at the full scale of a sub-region of the most improved methods for the QuickBird data set according to the synthesis property at reduced scale and QNR metrics. The 3 most improved methods from each family are shown. The first row shows the PAN image, the interpolated MS image and the MTF deblurred version of the interpolated image. The next 2 rows show results obtained by 3 CS methods using the interpolated image and then the deblurred image. Similarly, the last two rows show 3 MRA methods. The "Deblurred" prefix denotes images obtained using the proposed method.

(a) PAN

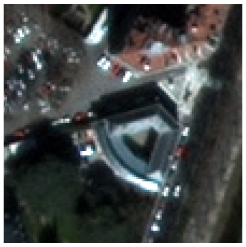
(b) Interpolated MS

(c) Deblurred MS

(d) GS

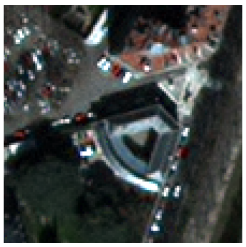


(e) GSA

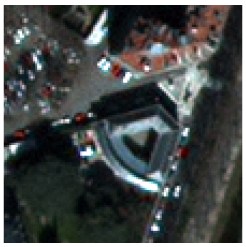


(f) BDSD

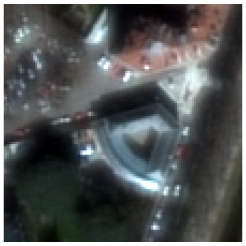
(g) Deblurred GS



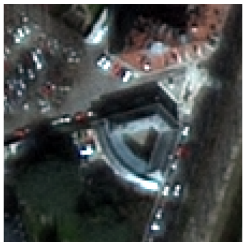
(h) Deblurred GSA



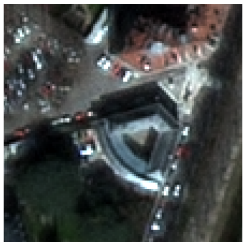
(i) Deblurred BDSD



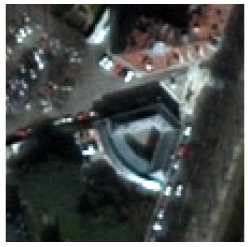
(j) ATWT-M2



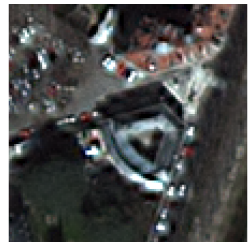
(k) HPF



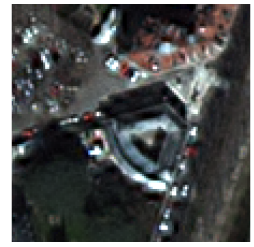
(l) SFIM



(m) Deblurred ATWT-M2

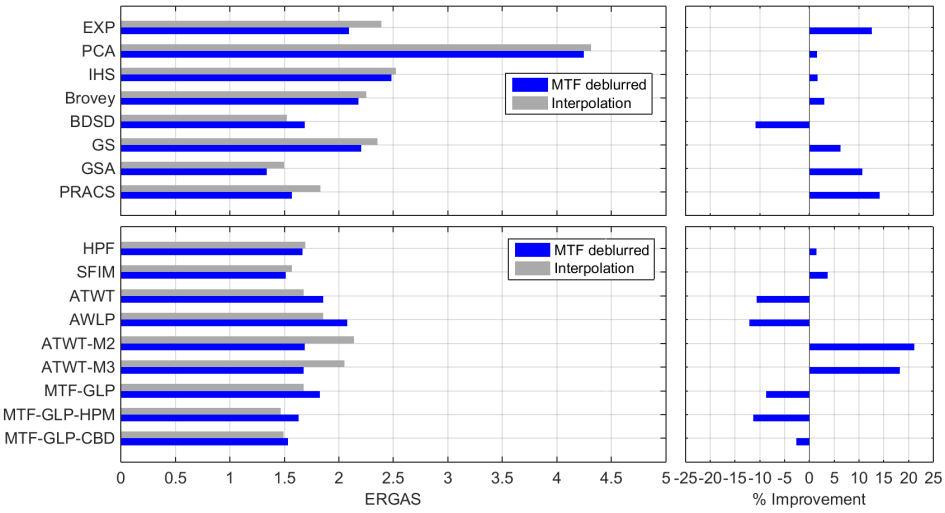


(n) Deblurred HPF

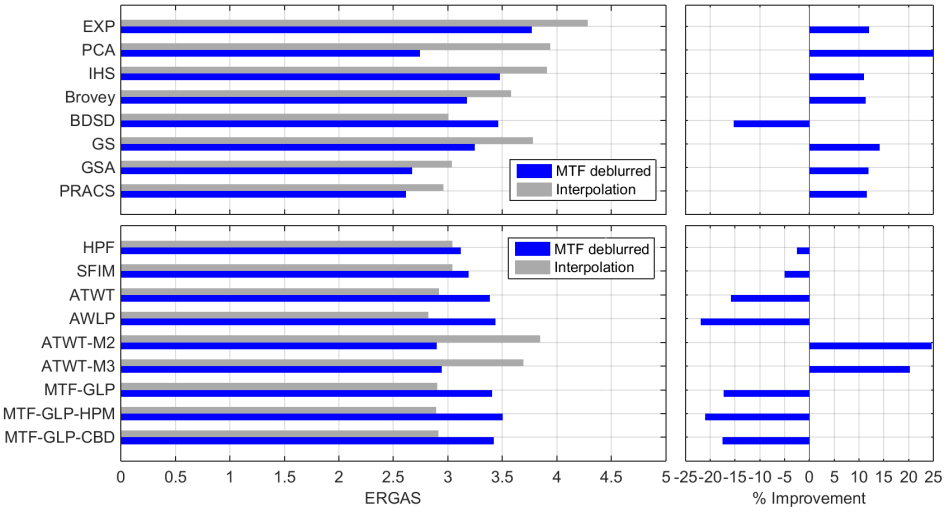


(o) Deblurred SFIM

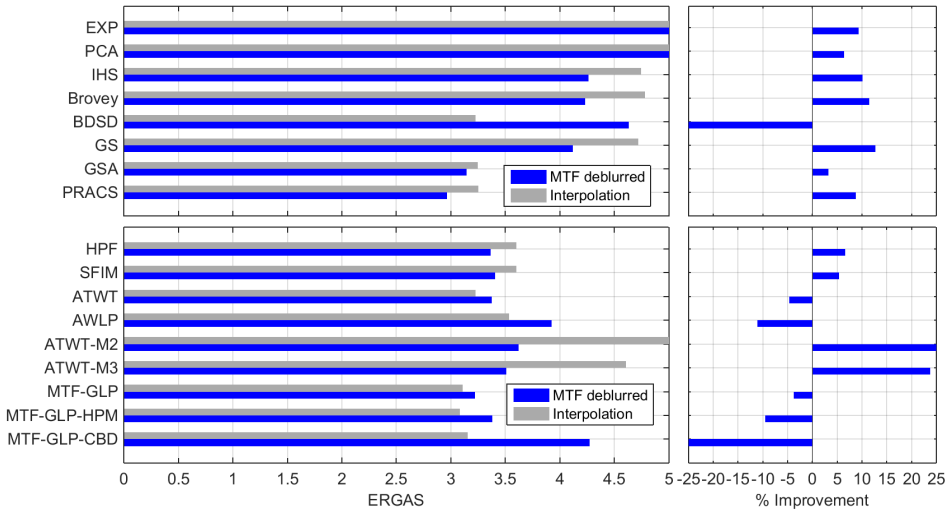
Figure 3.4: Visual inspection at the full scale of a sub-region of the most improved methods for the Pléiades data set according to the synthesis property at reduced scale and QNR metrics. The 3 most improved methods from each family are shown. The first row shows the PAN image, the interpolated MS image and the MTF deblurred version of the interpolated image. The next 2 rows show results obtained by 3 CS methods using the interpolated image and then the deblurred image. Similarly, the last two rows show 3 MRA methods. The "'Deblurred"' prefix denotes images obtained using the proposed method.



(a) WorldView 2 - ERGAS



(b) QuickBird- ERGAS



(c) Pléiades - ERGAS

Figure 3.4: ERGAS metric results from all data sets using synthesis at the reduced scale. The trends are very similar for all data sets, i.e., CS methods are enhanced with the exception of BDSD and MRA methods are not generally enhanced, with the exception of the ATWT-M2 and ATWT-M3 methods.

Chapter 4

Multispectral and Hyperspectral Image Fusion Using a 3D-Convolutional Neural Network

In this chapter a method for MS/HS fusion based on a 3D-convolutional neural network (3D-CNN) is developed. An important component of the method is the dimensionality reduction of the HS image via PCA prior to the training of the 3D-CNN. This step significantly reduces the computational cost of the method and makes it more resistant to noise in the HS image.

4.1 Introduction

The method developed here is based on training a 3D-CNN using supervised learning to learn filters that are used to fuse MS and HS images. Since the training is supervised, a target high resolution image is required. Since a high resolution target image is not available, the input data need to be spatially decimated (low-pass filtered and downsampled) so the observed HS image can be used as the target image. By doing this, it is assumed that the relationship between the input and target data, learned by the 3D-CNN at a lower resolution scale, also applies for a higher resolution scale.

An important component of the approach presented here, is the dimensionality reduction of the HS image using PCA [13] prior to the fusion stage. The main idea behind the dimensionality reduction is to significantly decrease the computational overhead of the method. However, it requires the assumption that the spectral singular vectors of the observed HS image are identical to the spectral singular vectors of the desired high resolution HS image. Indeed, the method proposed in [83] used PCA for dimensionality reduction of the HS image prior to fusion, in the same manner as the approach proposed here, and it was shown to give better results than the method in [120] which did not use any dimensionality reduction. By comparing the method proposed here to the conventional methods given in [83, 120], it is demonstrated that it gives better results according to three quantitative quality metrics.

The comparison methods [83,120] are very sensitive to the choice of decimation filters used for the MS image and an advantage of the proposed method is that the 3D-CNN learns the decimation filter in an automatic manner. This means that it is less sensitive to the decimation filters used to prepare the input samples for the 3D-CNN.

The proposed method also produces images that are free of artifacts, such as halos and ringing artifacts, often seen when using conventional methods.

The outline of this chapter is as follows. In Section 4.2, CNNs are briefly explained. In Section 4.3, the proposed method is described in detail. In Section 4.4, experimental results are presented and discussed, and finally, in Section 4.5, the conclusion is drawn.

4.2 Convolutional Neural Networks

The main ingredient of a CNN is a so-called convolutional layer. The convolutional layer consists of a number of hidden layers that contain a number of neurons. The central idea behind CNNs is the concept of a local receptive field [161], which is associated with each neuron of a hidden layer. The input to the convolutional layer is an image which has one or more channels, e.g., an RGB image. Each neuron in a hidden layer receives input from a rectangular subset of the input image, which is the neuron's receptive field.

The neurons of a hidden layer provide a complete tiling of the input image. This is achieved by sliding the receptive field over the input image and after each shift, it becomes the receptive field of the next neuron and so on. What is important here, is that all the neurons in the hidden layer share their weights and bias and therefore different neurons can detect the same feature at different locations in the input image.

The shared weights are called a filter or a kernel and the output of the hidden layer is basically the convolution of this filter with the input image and the resulting image is called a feature map. A single convolutional layer can have many featuremaps, i.e., hidden layers and thus it can learn several filters that detect distinct features in the input image. Between convolutional layers there usually pooling layers, which perform non-linear sub-sampling based on a function such as maximum value, i.e., max-pooling, which is the most common pooling layer. The pooling layers in a CNN simplify or downsample the feature maps and this reduces the number of parameters in the network, i.e., weights and biases, that need to be learned.

The most important benefit of the CNN architecture is that much fewer parameters need to be learned than for a conventional fully connected NN. This is due to the local receptive fields of the neurons, that provide shift invariance through their shared weights (and biases). The primary benefit is that much deeper networks can be constructed and they can learn much faster, without sacrificing performance.

For a 3D-CNN, which has 3D receptive fields and thus 3D filters, the output of the nth feature map $\mathbf{y}^n$ at location $\{i, j, k\}$ is given by

$$y^n_{i,j,k} = \sigma(b^n + (\mathbf{H}^n * \mathbf{x})_{i,j,k}),$$

where $*$ denotes 3D-convolution, b^n and $\mathbf{H}^n$ are the shared bias and filter (shared weights), respectively, σ denotes the non-linear activation function and $\mathbf{x}$ is the input.

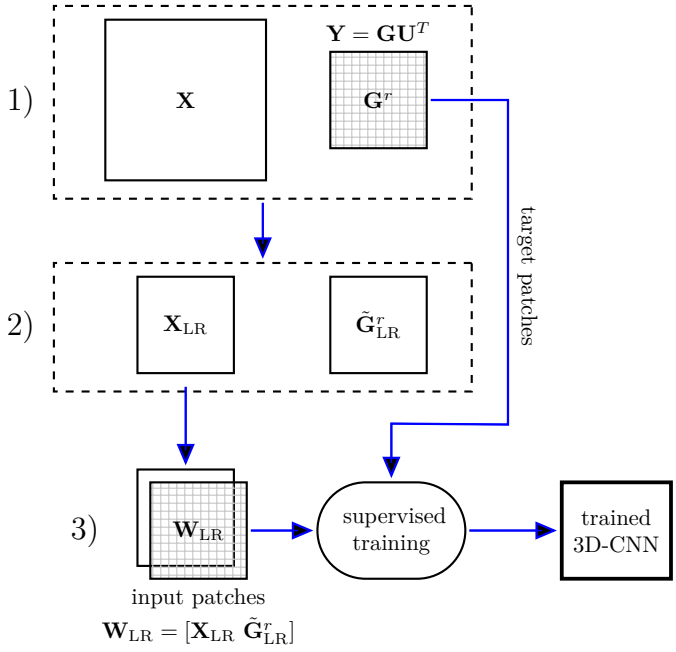


Figure 4.1: General outline of the training part of the algorithm. The steps labeled 1), 2) and 3), correspond to similarly labeled steps in the text.

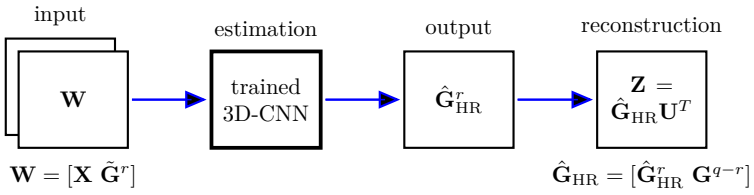


Figure 4.2: General outline of estimation part of the algorithm. The trained CNN is fed the entire input data at its full resolution and yields the high resolution spatial loadings, which are used to reconstruct the estimated high resolution HS image via the inverse PCA transform.

4.3 Proposed Method

This section is divided into two subsection. In the first subsection, the proposed method is described in detail and in the second subsection, the chosen architecture of the 3D-CNN is described.

4.3.1 General Outline of the Method

In the following, the observed MS image is denoted by $\mathbf{X}$ and is of dimension $M \times N \times P$, where P is the number of spectral bands. The observed $m \times n \times q$ HS image is denoted by $\mathbf{Y}$. Figure 4.1 depicts the training phase of the 3D-CNN which is detailed below.

To simplify the notation, the same symbols are used for 3D-matrices, i.e., images, and normal matrices. The implicit reshaping of a 3D image into a matrix with vectorized images (bands) in the columns is assumed. A tilde above a symbol denotes interpolation (upsampling followed by spatial filtering), a hat above a symbol denotes an estimate, and concatenation/stacking of matrices/images is denoted by square brackets, e.g., $[\mathbf{X}\ \mathbf{Y}]$.

1. Dimensionality reduction of $\mathbf{Y}$ using singular value decomposition gives

   $$\mathbf{Y} = \mathbf{V}\mathbf{D}\mathbf{U}^T = \mathbf{G}\mathbf{U}^T,$$

   where the $mn \times q$ matrix $\mathbf{G} = \mathbf{V}\mathbf{D}$ contains the spatial loadings and the $q \times q$ matrix $\mathbf{U}$ contains the spectral singular vectors. The first r columns of $\mathbf{G}$ are used to form an $m \times n \times r$ image, $\mathbf{G}^r$.

2. The MS image $\mathbf{X}$ is spatially decimated, using a bicubic filter, by the resolution factor between the MS and HS images to yield $\mathbf{X}_{\mathrm{LR}}$, which is of dimension $m \times n \times P$. Similarly, $\mathbf{G}^r$, is spatially decimated and then interpolated using a bicubic filter to the dimension of $\mathbf{X}_{\mathrm{LR}}$, giving $\tilde{\mathbf{G}}^r_{\mathrm{LR}}$.

3. The decimated MS image,$\mathbf{X}_{\mathrm{LR}}$, and the decimated and interpolated loadings, $\tilde{\mathbf{G}}^r_{\mathrm{LR}}$, are stacked to obtain the $m \times n \times (r+P)$ input image $\mathbf{W}_{\mathrm{LR}} = [\mathbf{X}_{\mathrm{LR}}\ \tilde{\mathbf{G}}^r_{\mathrm{LR}}]$. The target data are the first r bands of $\mathbf{G}$, i.e., $\mathbf{G}^r$. $\mathbf{W}_{\mathrm{LR}}$ and $\mathbf{G}^r$, are randomly divided into a number of matching overlapping patches of size 7×7 pixels, of depth $r + P$ and r, for inputs and targets, respectively.

The fusion part of the method is depicted in Fig. 4.2. The trained 3D-CNN can accept the entire input data at once, without having to break it down into patches, since it has learned all its filters. The input $\mathbf{W}$ to the trained 3D-CNN consists of the stacked MS image, $\mathbf{X}$, and the first r spatial loadings of $\mathbf{G}$, which have been interpolated to the size of $\mathbf{X}$, i.e., $\mathbf{W} = [\mathbf{X}\ \tilde{\mathbf{G}}^r]$. The output of the 3D-CNN is the estimated high resolution loadings, $\hat{\mathbf{G}}^r_{\mathrm{HR}}$.

The final step is the reconstruction of the estimated high resolution HS image, $\mathbf{Z}$, via

$$\mathbf{Z} = \hat{\mathbf{G}}_{\text{HR}}\mathbf{U}^T = [\hat{\mathbf{G}}^r_{\text{HR}}\ \tilde{\mathbf{G}}^{q-r}]\mathbf{U}^T,$$

where $\hat{\mathbf{G}}_{\text{HR}} = [\hat{\mathbf{G}}^r_{\text{HR}}\ \tilde{\mathbf{G}}^{q-r}]$ and $\tilde{\mathbf{G}}^{q-r}$ are the remaining $q-r$ interpolated spatial loadings obtained from the observed HS image, $\mathbf{Y}$, and $\mathbf{U}$ are the spectral singular vectors of $\mathbf{Y}$.

There are two options for the reconstruction of the estimated fused image. The first option is the one described above, where the first r loadings in $\tilde{\mathbf{G}}$ are replaced by the high resolution estimate $\hat{\mathbf{G}}^r_{\text{HR}}$. If the HS image is noisy, a second option is to retain only the first r PCs, i.e., performing the inverse PCA transform using the reduced $\hat{\mathbf{G}}^r_{\text{HR}}$ and $\mathbf{U}^r$ matrices, yielding $\mathbf{Z} = \hat{\mathbf{G}}^r_{\text{HR}}\mathbf{U}^{rT}$, where $\mathbf{U}^r$ denotes the reduced matrix $\mathbf{U}$.

4.3.2 3D-CNN Architecture

For this work, a 3D-CNN architecture was chosen, since an HS image has two spatial dimensions and one spectral dimension, and a 3D-CNN learns spectral-spatial features.

If the input image for a convolutional layer has dimension $M \times N \times P$ and the dimension of the filter is $i \times j \times k$, then the resulting feature map will be of dimension $M - i + 1 \times N - j + 1 \times P - k + 1$. In order to preserve the dimensions of the input image as it is passed through the convolutional layers, and to avoid boundary artifacts due to the convolution operations, it is necessary to zero-pad the input image by $(i-1)/2$ zeros at each end of the first dimension, $(j-1)/2$ for the second dimension, and $(k-1)/2$ for the third dimension.

In the experiments, the 3D-CNN used has 3 convolutional layers with 32, 64 and r filters, respectively, where r is the number of spatial loadings. The corresponding filter sizes for the convolutional layers are $3 \times 3 \times 3$, for the first two layers and $1 \times 1 \times 1$ for the last layer. A given convolutional layer is preceded by a zero-padding layer, as described above, and followed by a Gaussian noise regularization layer. However, last convolutional layer is not followed by such a layer. The Gaussian noise layers add zero-mean Gaussian noise to the output of the previous layer and can be considered as a form of random data augmentation [162]. These regularization layers help to avoid over-fitting in the network. Regarding the activation functions, the first two layers have have rectified linear unit (ReLU) activation functions, i.e., $\sigma(x) = \max(x, 0)$, while the output layer has linear activation.

When the 3D-CNN has been trained, i.e., it has learned all the filters, the entire input image can be predicted at once, without having to break it down into patches. However, this can be very memory consuming if the input image is large, and, therefore, PCA dimensionality reduction helps to significantly reduce the memory overhead in the fusion process. The layers of the CNN are summarized in Table 4.1.

Table 4.1: 3D-CNN architecture. The numbers in parenthesis following zero-padding layers indicate the number of zeros added to each dimension. The numbers in parenthesis after convolution3D, indicate number of filters and the filter size of each dimension. r indicates the number of PCs. Finally, the number following Gaussian noise denotes the noise variance.

layer #	Type	Activation
1	zero-padding3D (1,1,1)	none
2	convolution3D (**32**,3,3,3)	ReLU
3	Gaussian noise (0.5)	none
4	zero-padding3D (1,1,1)	none
5	convolution3D (**64**,3,3,3)	ReLU
6	Gaussian noise (0.5)	none
7	convolution3D (**r**,1,1,1)	none

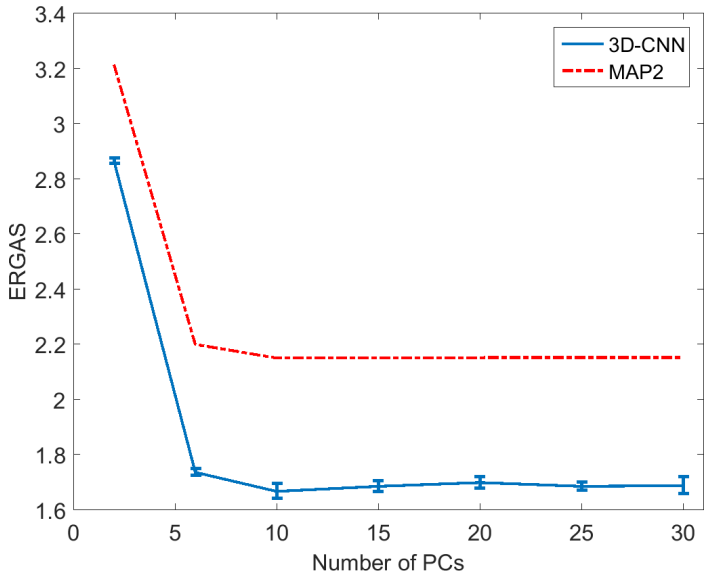


Figure 4.3: Performance in terms of ERGAS of the proposed and MAP2 methods, as a function of the number of PCs. Six trials were performed for the 3D-CNN method. The mean is shown and the standard deviation is displayed using errorbars.

4.4 Experiments

4.4.1 Simulated Data Set

The HS image used for the experiments is the ROSIS Pavia center data set. Due to the presence of a blank strip along the left side, only 480 pixels out of 512 will be used along the row dimension.

To simulate the MS image using the HS image, bands corresponding to the wavelength of the R, G, B and NIR bands of the IKONOS MS sensor are averaged, weighted by the sensors relative spectral reflectance profiles. The observed HS image used in the fusion is obtained by spatially decimating the original observed HS image by a factor of 4 for each spatial dimension, using a bicubic decimation filter. The images to be fused are then an MS image of dimension 512×480 pixels with 4 spectral bands, and an HS image of dimension of 128×120 pixels with 102 spectral bands.

The method was implemented in the Python programming language using the Keras DL library which runs on top of the Theano backend [1] and the computations were performed using an Intel i5-2400 CPU@3.1 GHz with 16GB of RAM.

4.4.2 Results

The training data are prepared as follows. The decimated simulated MS image and the decimated and interpolated spatial loadings obtained from the HS image via SVD, are stacked and 8192 randomly overlapping patches of spatial size 7×7 pixels are selected from the stacked image. The target data are the corresponding patches from the loadings of the HS image. This procedure is shown in Figure 4.1.

The mean squared error (MSE) between the target patch and the estimated patch is used as the network objective and the optimizer used is the adaptive moment estimation (ADAM) [163] optimizer with default parameters as given in [163]. The number of training epochs is set equal to 50, which guarantees full convergence of the objective function during training. The batch size is equal to 5, which gives better results and faster convergence than higher values. Finally, the variance for the Gaussian noise regularization layer is equal to 0.5.

The comparison methods are the method in [120], and the extended version of that method in [83]. These methods will be referred to as MAP1 and MAP2, respectively. Both methods are based on maximum a posteriori (MAP) estimation of wavelet coefficients The difference between MAP1 and MAP2 is that MAP2 performs the estimation in the PCA subspace, in a similar manner to the proposed method.

The first experiment investigates the effect of the number of PCs used, i.e., the parameter r, on the performance of the proposed and MAP2 methods. The performance

[1] `http://keras.io`, `https://github.com/Theano`

is measured by the ERGAS metric and the following values for r are considered: 2, 6, 10, 15, 20, 25 and 30. The results are shown in Figure 4.3 and it shows that for both methods, 10 PCs give optimal results, as measured by the ERGAS metric. In all the following experiments, r will be chosen equal to 10 for these methods.

The second experiment is quantitative quality evaluation of the fused HS image for all methods in terms of the ERGAS [135], SAM [113] and SSIM [139] quantitative quality metrics. The experiment is two-fold, i.e., without and with zero-mean Gaussian noise (SNR=20dB) added to the HS image. The results are summarized in Table 4.2 where the upper half of the table shows results without additive noise and the lower half shows the results when using an HS image contaminated with noise. In both cases, the proposed method significantly outperforms the MAP1 and MA2 methods according to all three quality metrics. The MAP2 method outperforms the MAP1 method and its noise tolerance is similar to the proposed method. The MAP1 method, which does not use PCA prior to the fusion, performs significantly worse than the other methods in the presence of noise as was expected.

Finally, it is clear that the proposed method cannot compete with the comparison methods when it comes to computational time. However, these results were obtained using a conventional CPU but by using a powerful GPU, the training time can be reduced more than a hundred-fold, making the proposed method competitive in terms of computational time. Fig. 4.5 depicts a small portion of the 102th band of the interpolated, reference, and estimated HS image for all three methods. Visual inspection shows that the proposed method gives the best results.

In the third experiment, the performance of all methods, in terms of the ERGAS metric, is investigated when the SNR varies due to additive Gaussian noise, from 10 to 30 dB, in the increments of 5 dB. The results of this experiment are shown graphically in Figure 4.4. Unsurprisingly, the proposed method performs best, with the MAP2 performing second best and the MAP1 method performs significantly worse than the other two methods.

In final experiment, the performance of all three methods with respect to the decimation filters used, is investigated. There are three types of filters considered: bicubic, bilinear and nearest neighbor decimation filters. The performance is evaluated using the ERGAS and SAM metrics since they are more sensitive than the SSIM metric. The results for this experiment are summarized in Table 4.3. As was expected, bicubic decimation gives the best results for all three methods. Choosing a bilinear decimation filter degrades the performance of the methods, however, the proposed method and MAP2 are less affected than the MAP1 method. Finally, choosing a nearest neighbor decimation filter degrades the performance significantly more for the MAP1 and MAP2 methods, than for the proposed method, when compared to the results obtained using the bicubic decimation filter.

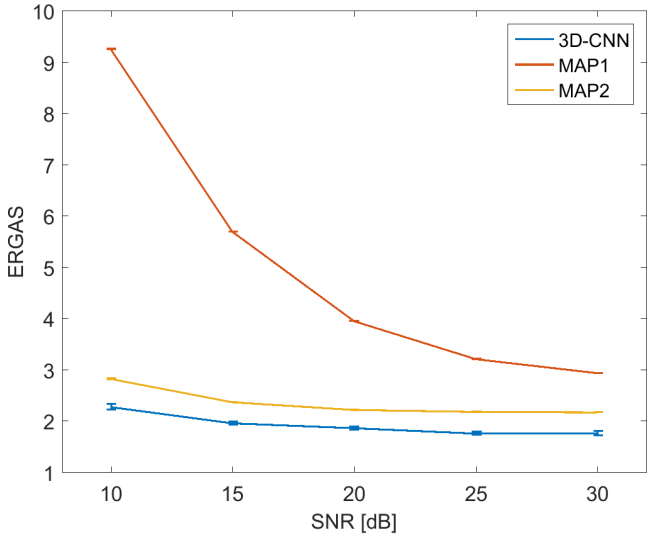


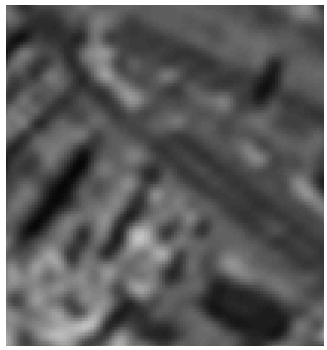
Figure 4.4: Noise resistance of the proposed method vs comparison methods. For each value of SNR, 6 trials were conducted and the graph shows the mean and standard deviation of the trials as a function of the SNR.

Table 4.2: Quantitative quality evaluation results, without and with additive Gaussian noise (SNR=20dB). For the proposed and MAP1 methods, 10 PCs were used in reduced PCA. The CPU time is given in seconds. Six trials were performed for the noisy case, and also for the proposed method without noise.

Method	ERGAS	SAM	SSIM	CPU time
MAP1 [120]	2.806	3.711	0.971	45
MAP2 [83]	2.17	3.26	0.978	9
3D-CNN	**1.676±0.02**	**2.730±0.02**	**0.988±1.14e-4**	978±8
Noisy HS image (SNR=20dB)				
MAP1 [120]	3.95±0.004	7.42±0.005	0.89±6.98e-4	45±0.4
MAP2 [83]	2.23±0.002	3.46±0.003	0.98±6.06e-5	9±0.4
3D-CNN	**1.79±0.05**	**3.03±0.02**	**0.99±3.3e-4**	989±10

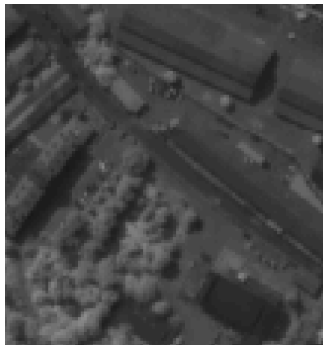
Table 4.3: Performance of all methods w.r.t. to the interpolation filter used. Bicubic, bilinear and nearest neighbor interpolation is considered. One trial was performed for the proposed method.

	Bicubic		Bilinear		Nearest	
Method	ERGAS	SAM	ERGAS	SAM	ERGAS	SAM
MAP1	2.806	3.711	3.080	4.721	5.680	5.501
MAP2	2.170	3.260	2.233	3.468	5.234	5.193
3D-CNN	1.676	2.730	2.069	3.022	3.104	3.858

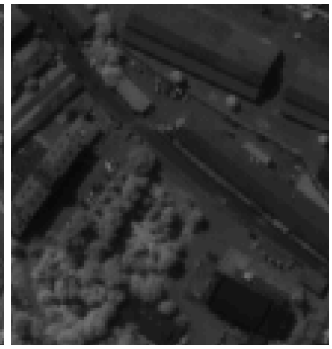


(a) Interpolated HS image

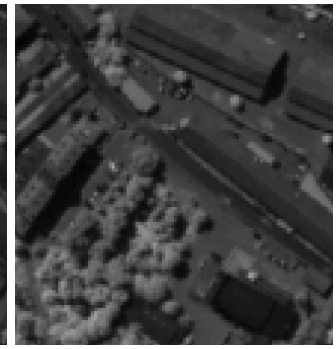
(b) Reference



(c) MAP1



(d) MAP2



(e) 3D-CNN

Figure 4.5: A subset of band 102 of the HS image is shown. (a) shows the interpolated HS image band, (b) is the reference band, (c) shows the image obtained using the MAP1 method, (d) shows the image obtained using the MAP2 method and (e) shows the image obtained using the proposed method.

4.5 Conclusions

In this chapter, a new method for the fusion of MS and HS images using a 3D-CNN was proposed. While a method based on DL tools such as CNNs has been proposed for pansharpening, the proposed method expands the pansharpening method to handle MS/HS fusion in an efficient manner. Also, a novelty is the use of a 3D-CNN which can exploit the spatial-spectral nature of the HS image. An important component of the method is the dimensionality reduction via PCA of the HS image, prior to the fusion. This decreases the computational cost significantly while having no impact on the quality of the fused image. In the presence of noise, the dimensionality reduction can considerably improve the results. The proposed method was compared to two methods based on MAP estimation and several experiments using a simulated data set demonstrated that the proposed method gives excellent results and is also highly tolerant to noise in the HS image.

Chapter 5

Quantitative Quality Evaluation of Pansharpened Imagery: Consistency vs. Synthesis

Since pansharpening is an inverse problem, there is no high resolution reference image available. This makes the quantitative quality evaluation of fused images a difficult problem and there is no single universally accepted method or protocol for such evaluation. Two important properties where stated in [122] that the fused image must have. The first one is called the consistency property and it states that the fused image when decimated to the size of the MS image should be as identical as possible to the MS image. The second property, called the synthesis property, states that the fused image should be as identical as possible to a higher resolution MS image acquired by the MS sensor if it had the resolution of the PAN image. This is not possible to verify unless the source images are decimated prior to fusion such the MS image can be used as the reference. This is the most common method for the quantitative evaluation of pansharpened images. In this chapter it is investigated whether the verification of the consistency property, referred to here as the consistency protocol can be a good indicator of the performance of a pansharpening method. Three experiments are performed. The first one is evaluation of two real data sets using the synthesis, consistency and QNR metrics for a total of 18 methods, including 7 CS methods and 11 MRA methods. In the second experiment, the correlation between the relative method ranking according to synthesis and consistency is investigated, for all methods, for CS methods and for MRA methods. In the final experiment, the sensitivity of the various protocols with respect to resolution scale is investigated. The result suggest that the consistency protocol can be used as a reliable indicator of fusion performance when comparing pansharpening methods.

5.1 Introduction

Pansharpening is a typical inverse problem and there is no single universally accepted protocol for the quantitative evaluation of the quality of pansharpened images. Another problem that makes the comparison of fusion methods often difficult, is the lack of standardized data sets for evaluation. This lack of a single robust quality evaluation protocol and standardized evaluation data sets is a problem that needs to be addressed by the remote sensing community. Even though quality assessment of the pansharpened images is a debated topic, there are surprisingly few papers in the literature on

the subject, especially regarding quality assessment of full scale imagery. Papers on the topic include [121–133].

For quantitative evaluation of fused images, there are two methods that are mostly used, i.e, the synthesis property at the reduced resolution and the QNR metric at the full resolution scale. These were detailed in sections 1.5.1 and 1.5.2, respectively. The former property or protocol is based on reducing the resolution of the observed images, such that the fused image has the same dimensions as the observed MS image, which is then used as the reference image. Obviously, a lot of information is lost during this process and it can be argued that most pansharpening methods perform differently at different resolution levels. The fusion task being evaluated is actually far from being the same as the original one, i.e., at the full scale. Is the synthesis property/protocol a good choice for comparing pansharpening methods?

Another important question that is addressed in this chapter is whether the consistency property (see Subsection 1.5.1) can be used as quantitative measure of the quality of the fused image. The important difference between the evaluation protocols based on the synthesis and consistency properties, is that for the former, the degradation of the images is performed prior to fusion, but for consistency, the fused image is degraded such that it has the same dimension as the observed MS image, which is the reference. As with the synthesis protocol, one can argue that a lot of information is lost in the degradation of the fused image and thus it can not give a measure of the fusion performance. However it is important to note that it is stated in [122] that it is necessary for any pansharpening method to produce images that are as identical to the observed MS image as possible when downgraded to the resolution scale of the observed image. This means that the consistency property should be fulfilled in the ideal case. Therefore the main motive of this chapter is to investigate if the discrepancy between the downgraded fused image and observed MS image, as measured by the consistency protocol, can give a good measure of the relative performance of different pansharpening methods.

Finally, it also investigated how well the results obtained by the QNR metrics correlate with the results, i.e., method ranking, obtained using the synthesis and consistency protocols. Also, the fused images produced by methods selected as the best performing by the QNR metrics are visually compared to the images produced by the best performing methods according to synthesis and consistency. The motive here is to check whether the QNR metrics are indeed reliable for evaluation of full scale imagery.

Another motivation for using the consistency protocol instead of the synthesis protocol for assessing the performance of pansharpening methods, are methods based on DL such as the method presented in the previous chapter, where a CNN is trained using data at the degraded scale. Obviously, the trained CNN will give very good results according to the synthesis metrics at the reduced scale, since it was trained using the very same protocol. This will cause a bias in favor of methods based on supervised training compared to conventional methods, when evaluating them using the synthesis metrics at the reduced scale.

Experiments are performed using two real data sets. The first data set is a rural WorldView-2 data set and the second data set is an urban QuickBird data set. In the experiments, the two data sets are evaluated using the synthesis protocol at the reduced scale, consistency and QNR metrics for the full scale. A total of 18 CS and MRA methods are evaluated, 7 CS methods and 11 MRA methods. As was discussed in chapter 3, there is a significant difference between the CS and MRA methods in the way the details are obtained from the PAN image. Interestingly, verification of the consistency protocol reflects these differences, especially if there is a slight misalignment between the MS and PAN image. Generally, MRA methods score better than the CS methods when evaluated by the consistency protocol. The exact reason for this will be discussed in the experiments section of this chapter.

Each family of methods is evaluated and the best methods according to each evaluation protocol are visually inspected. The correlation between the order or ranking of methods according to quantitative quality as measured by the three protocols is also investigated. Finally, the resolution scale invariance of the synthesis at reduced scale metrics and the consistency metrics is investigated. It is important that both properties show a high degree of resolution scale invariance. For all experiments, decimation filters which have been matched to the MTF of the MS and PAN sensors are used. This should increase the validity of the results.

The outline of this chapter is as follows. In Section 5.2, the experiment results are presented and discussed and in Section 5.3, the conclusion is drawn.

5.2 Experiment Results

The method implementations used in this chapter come from a comprehensive Matlab toolbox[1], that was published along with paper [16]. There are 7 CS methods and 11 MRA methods used in the experiments. Since these two classes of methods extract details from the PAN image in different ways, the results of the experiments will be summarized for all methods and for each class of methods separately.

The experiments are comprised of three different studies. In the first one, each data set is evaluated using all methods and using the synthesis, consistency and QNR metrics. The objective of this study is to see which methods give the best results according to the three different quantitative evaluation protocols or metrics. Finally, a visual inspection of the best overall, CS and MRA method for each data set is given.

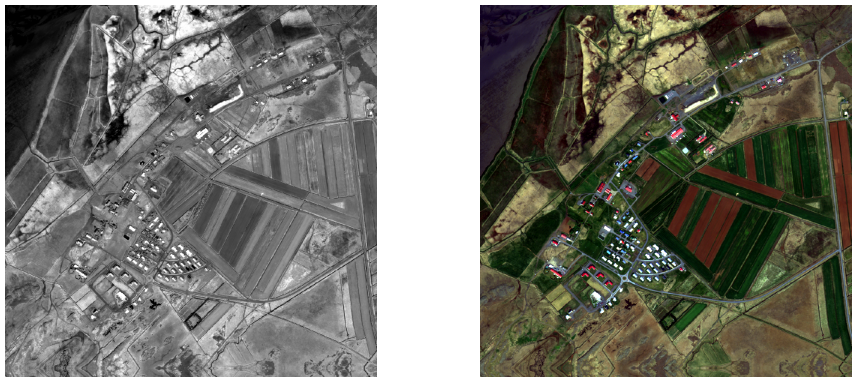
The objective of the second study is to investigate the how well the ranking of methods according to the consistency protocol correlates with the ranking of methods according to both synthesis and QNR metrics. The correlation of the synthesis and QNR metrics is also investigated. For the synthesis and consistency we use the ERGAS, SAM,

[1]The toolbox is available from:
`http://www.openremotesensing.net/index.php/codes/11-pansharpening`

Q4/Q8 and Q metrics (see Appendix A) and only the QNR score and not its sub-indices D_λ and D_S. The objectives of this study is two-fold. The first is to see how well the consistency and synthesis metrics correlate with the QNR metrics at full scale and secondly, to investigate how well the ranking of methods according to the consistency and synthesis metrics are correlated. This is important since if the consistency protocol yielded results that had no, or little correlation with the synthesis metrics, it would not be desirable to investigate it further as an option to evaluate pansharpening performance.

In the third and final study, it is examined how well the ranking of the methods being evaluated is conserved across different resolution scales for both consistency and synthesis metrics. The resolution scale invariance property is important for the synthesis protocol, since it is assumed that the evaluation of a method at a lower resolution scale give a good idea of its performance at the full resolution scale. This property is also import when verifying the consistency property. A good scale invariance of the ranking of methods should indicate that the quantitative result obtained for a lower resolution scale give a good idea or estimate of its performance at a higher resolution scale.

In all the experiments, decimation filters are tuned to the MTF of the respective sensors and for interpolation a near ideal 23 tap filter is used [16]. The use of MTF matched filters should enhance the validity of the experiment results.

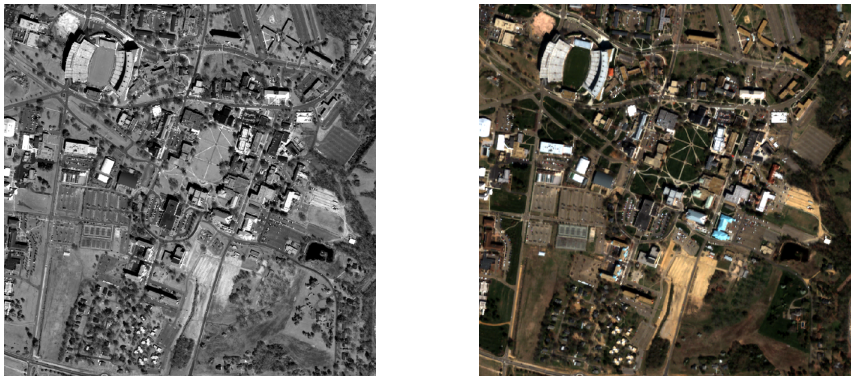


(a) PAN (b) MS

Figure 5.1: The rural WorldView-2 data set. The PAN image (a) is 4096×4096 pixels and the MS image (b) is $1024 \times 1024 \times 8$ pixels. The MS image is shown as an RGB image using bands 5, 3 and 2.

5.2.1 Synthesis, Consistency and QNR Results.

In this subsection the results of the quantitative evaluation of both WV2 and QB data sets are presented, using synthesis, consistency and QNR metrics. The results



(a) PAN (b) MS

Figure 5.2: The urban Quickbird data set. The PAN image (a) is 2048×2048 pixels and the MS image (b) is $512 \times 512 \times 4$ pixels.

are shown in Tables 5.1 and 5.2 for the WV2 and QB data sets, respectively. The best results for each metric in each family of methods and for each protocol have been highlighted using bold typeface.

Note that the ERGAS, SAM and nMSE metrics, lower values are better with 0 being the optimal value, whereas for the Q4/Q8 and Q metrics, higher values are better with 1 being optimal. The nMSE is given in percentage (%). For the QNR metric, lower values are better for the D_λ and D_S indices with 0 being optimal and for the joint QNR score, higher is better with 1 being optimal.

In Table 5.1, the results are summarized for the WV2 data set. The metric scores indicate good performance for most of the methods. The best CS method according to synthesis metrics is the GSA method, the best CS method according to consistency is PRACS and finally, the best CS method according to the QNR metrics is BDSD. For MRA methods the best method according to synthesis and consistency is MTF-GLP-CBD, while the best method according to QNR is ATWT-M2. For both CS and MRA, the MTF-GLP-CBD method gives the best results according to both synthesis and consistency while the overall best method according to QNR is BDSD. Full scale RGB images produced by the methods that scored best for each protocol and method family are shown in Fig. 5.3. Visual inspection of the results reveals that the GSA method gives the best image, with the MTF-GLP-CBD image being very close in quality and detail. The PRACS image looks a little blurry but has very little spectral distortion. The ATWT-M2 image looks noticeably worst since it is very blurry compared to the others.

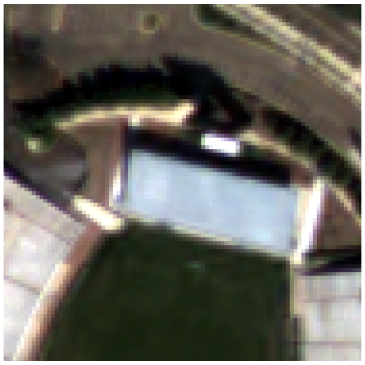
The QB results are summarized in Table 5.2. The scene depicted in this data set has more details such as roads and buildings than the WV2 scene, making it more challenging. There is also a slight misalignment between the MS and PAN images. For

(a) EXP (b) GSA

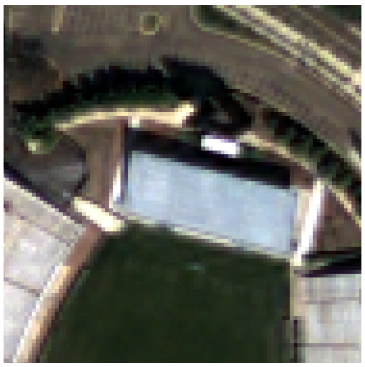
(c) PRACS (d) MTF-GLP-CBD

(e) BDSD (f) ATWT-M2

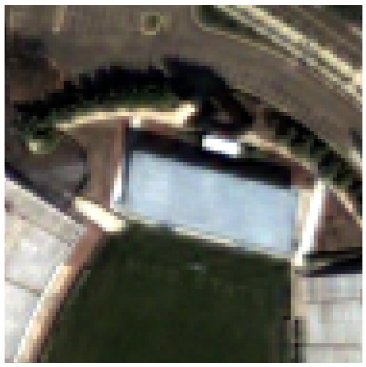
Figure 5.3: A small portion of the urban WorldView-2 data set, displaying best results from Table 5.1.(a) Interpolated MS image (b) Best CS result according to synthesis, (c) best CS result according to consistency, (d) best MRA and best result overall according to synthesis and consistency, (e) best CS (and overall) result according to QNR, and (f) best MRA result according to QNR.



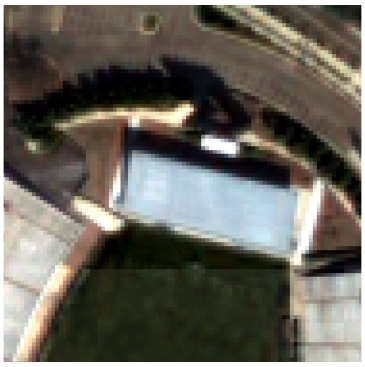
(a) PRACS



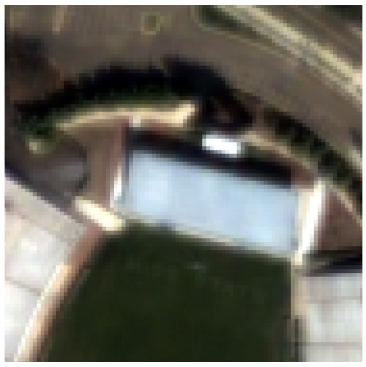
(b) AWLP



(c) MTF-GLP-CBD



(d) BDSD



(e) ATWT-M2

Figure 5.4: A small portion of the urban Quickbird data set, displaying best results from Table 5.2. (a) Best CS results according to synthesis and consistency, (b) best MRA results according to synthesis, (c) best MRA (and overall) result according to consistency, (d) best CS (and overall) result according to QNR and finally, (e) best MRA result according to QNR.

the CS family of methods, the best performing methods are the PRACS method, for both synthesis and consistency and the BDSD method according to the QNR metrics. For the MRA methods, AWLP performs best according to synthesis, while the MTF-GLP-CBD method performs best according to consistency, as for the previous data set, and finally, the ATWT-M2 is again the best MRA method according to the QNR metrics. Finally, the overall best methods are AWLP for synthesis, MTF-GLP-CBD for consistency and BDSD for QNR. A visual comparison of the best CS and MRA results for each protocol is shown in Fig. 5.4. It seems as there is some kind of a black artifact present in the BDSD image, in the upper part of the stadium. The ATWT-M2 image looks blurry, as for the WorldView-2 data set.

Interestingly, for both data sets, the best CS and MRA methods are the same according to consistency, while according to synthesis metrics, the best methods are not the same for each data set. For the WV2 data set, the MTF-GLP-CBD is the best MRA method and overall best, according to both synthesis and consistency metrics. However, the PRACS method which is the best CS method according to consistency produces an image, that under visual inspection looks a little blurry. On the other hand, the PRACS method is the best CS method according to both protocols for the QB image. The best method according to the QNR metric, for both data sets is the BDSD method. This method produces highly detailed images, and in some cases, as with the WV2 image, the detail injection seems to be too strong. But for the MRA methods, the best method according to QNR is the ATWT-M2 method which produces very blurry images. This seems to be a contradiction and is probably related to the way the CS and MRA methods work.

5.2.2 Method Ranking Correlation: Synthesis vs. QNR and Consistency

In this experiment, the ranking of the methods according to specific metrics is compared when evaluated using the consistency, synthesis and QNR metrics. One would expect the ranking of methods according to ERGAS using synthesis or consistency to be highly correlated. It is the correlation between the method ranking that is being assessed in this experiment.

As in the previous experiment, both data sets are evaluated using all three methods, i.e., synthesis, consistency and QNR metrics. The procedure is best described using an example. Table 5.3 summarized the results of this experiment for all methods and both data sets. The correlation coefficient for ERGAS in the top left cell of the table is calculated as follows. First the results table is sorted according the ERGAS value. The next step is to add an column of integers for the ranking, i.e., the topmost method with the lowest ERGAS scores is assigned 1 and the next method 2 and so on. Thus the ranking is described using a sequence $\{1, 2, 3, \cdots, 18\}$ (using all 18 methods). The next step is to sort this table according to the consistency ERGAS score, and then computing the correlation coefficient between the original sequence (best synthesis ERGAS score) and the altered sequence based on the consistency ERGAS metric. The

Table 5.1: Evaluation for the WorldView-2 rural data set using synthesis, consistency and QNR metrics. Best results within each family of methods for each metric are shown using bold typeface.

CS methods

	Synthesis					Consistency					QNR		
Method	ERGAS	SAM	Q4	Q	nMSE (%)	ERGAS	SAM	Q4	Q	nMSE (%)	D_λ	D_s	QNR
PCA	4.313	5.587	0.784	0.743	21.616	3.864	4.681	0.863	0.869	17.314	0.066	0.121	0.821
IHS [17]	2.523	2.604	0.676	0.819	7.535	2.125	1.186	0.771	0.889	5.343	0.066	0.080	0.859
Brovey	2.250	2.036	0.737	0.846	5.992	1.847	0.444	0.821	0.910	4.080	0.056	0.084	0.865
BDSD	1.605	2.237	0.874	0.889	3.004	1.190	1.385	0.958	0.956	1.652	0.020	**0.015**	**0.966**
GS	2.355	2.890	0.829	0.881	6.469	1.730	1.558	0.885	0.947	3.487	0.034	0.091	0.878
GSA	**1.499**	**1.902**	**0.911**	**0.913**	**2.619**	0.577	0.492	0.981	0.986	0.388	0.037	0.067	0.898
PRACS	1.829	1.939	0.885	0.889	3.901	**0.466**	**0.454**	**0.991**	**0.992**	**0.254**	**0.006**	0.040	0.954
mean	2.339	2.742	0.814	0.854	7.305	1.686	1.457	0.896	0.936	4.645	0.041	0.071	0.892

MRA methods

	Synthesis					Consistency					QNR		
Method	ERGAS	SAM	Q4	Q	nMSE (%)	ERGAS	SAM	Q4	Q	nMSE (%)	D_λ	D_s	QNR
HPF	1.693	2.036	0.891	0.899	3.337	0.347	0.437	0.993	0.995	0.141	0.037	0.054	0.912
SFIM	1.568	1.935	0.892	0.903	2.865	0.384	0.440	0.990	0.994	0.170	0.039	0.055	0.909
Indusion	2.434	2.358	0.813	0.836	6.892	0.928	0.541	0.966	0.968	1.001	0.041	0.033	0.927
ATWT	1.678	2.045	0.897	0.904	3.271	0.378	0.453	0.991	0.993	0.166	0.050	0.070	0.883
AWLP	1.854	2.313	0.895	0.901	3.992	0.401	0.496	0.992	0.993	0.187	0.034	0.047	0.921
ATWT-M2	2.138	2.303	0.854	0.850	5.346	0.652	0.705	0.983	0.985	0.496	**0.023**	0.039	**0.939**
ATWT-M3	2.049	2.382	0.862	0.851	4.904	0.652	0.720	0.983	0.984	0.495	0.031	**0.037**	0.933
MTF-GLP	1.677	2.042	0.900	0.903	3.269	0.360	0.446	0.993	0.993	0.151	0.049	0.070	0.884
MTF-GLP-HPM-PP	1.813	2.251	0.877	0.873	3.820	0.773	0.753	0.971	0.972	0.694	0.077	0.089	0.841
MTF-GLP-HPM	**1.466**	1.920	0.902	0.907	**2.497**	0.396	0.432	0.992	0.993	0.180	0.050	0.070	0.884
MTF-GLP-CBD	1.493	**1.896**	**0.912**	**0.914**	2.598	**0.327**	**0.427**	**0.994**	**0.995**	**0.125**	0.036	0.060	0.906
mean	1.806	2.135	0.881	0.886	3.890	0.509	0.532	0.986	0.988	0.346	0.042	0.057	0.903

Table 5.2: Evaluation for the Quickbird urban data set using synthesis, consistency and QNR metrics. Best results within each family of methods for each metric are shown using bold typeface.

CS methods

	Synthesis					Consistency					QNR		
Method	ERGAS	SAM	Q4	Q	nMSE (%)	ERGAS	SAM	Q4	Q	nMSE (%)	D_λ	D_s	QNR
PCA	3.937	5.153	0.819	0.824	8.529	1.691	1.282	0.969	0.972	1.570	0.031	0.066	0.905
IHS	3.909	5.174	0.803	0.817	8.502	2.463	1.682	0.918	0.926	3.374	0.070	0.138	0.802
Brovey	3.578	4.369	0.841	0.853	7.100	2.309	1.130	0.941	0.948	2.939	0.049	0.104	0.852
BDSD	3.004	4.506	0.892	**0.898**	4.970	1.702	2.117	0.975	0.974	1.596	**0.016**	**0.010**	**0.974**
GS	3.782	4.942	0.830	0.838	7.876	2.361	1.588	0.937	0.945	3.057	0.053	0.123	0.830
GSA	3.034	4.358	0.881	0.881	5.065	1.528	1.228	0.974	0.975	1.281	0.060	0.102	0.844
PRACS	**2.959**	**4.064**	**0.894**	0.894	**4.821**	**1.157**	**1.118**	**0.986**	**0.987**	**0.737**	0.023	0.043	0.935
mean	3.458	4.652	0.851	0.858	6.695	1.887	1.449	0.957	0.961	2.079	0.043	0.084	0.877

MRA methods

	Synthesis					Consistency					QNR		
Method	ERGAS	SAM	Q4	Q	nMSE (%)	ERGAS	SAM	Q4	Q	nMSE (%)	D_λ	D_s	QNR
HPF	3.041	4.314	0.883	0.885	5.093	0.865	1.068	0.992	0.992	0.413	0.045	0.060	0.898
SFIM	3.040	4.275	0.889	0.890	5.086	0.956	1.152	0.991	0.991	0.493	0.044	0.063	0.896
Indusion	4.303	4.548	0.791	0.802	10.190	1.806	1.152	0.964	0.965	1.790	0.039	0.026	0.936
ATWT	2.921	4.285	0.889	0.890	4.701	0.862	1.079	0.991	0.991	0.409	0.061	0.085	0.859
AWLP	**2.819**	**3.877**	**0.902**	**0.903**	**4.388**	0.837	1.016	0.992	0.992	0.387	0.055	0.070	0.879
ATWT-M2	3.848	4.976	0.820	0.817	8.154	1.418	1.567	0.979	0.979	1.106	**0.023**	**0.027**	**0.950**
ATWT-M3	3.691	5.117	0.838	0.831	7.502	1.414	1.606	0.979	0.979	1.100	0.029	0.024	0.948
MTF-GLP	2.902	4.216	0.889	0.889	4.639	0.839	1.044	0.991	0.992	0.388	0.064	0.086	0.855
MTF-GLP-HPM-PP	3.301	4.809	0.870	0.866	5.995	1.694	1.718	0.969	0.969	1.571	0.094	0.114	0.803
MTF-GLP-HPM	2.893	4.155	0.896	0.896	4.607	0.923	1.106	0.991	0.991	0.463	0.061	0.086	0.858
MTF-GLP-CBD	2.912	4.210	0.889	0.889	4.670	**0.819**	**1.032**	**0.992**	**0.992**	**0.370**	0.050	0.072	0.882
mean	3.243	4.435	0.869	0.869	5.911	1.130	1.231	0.985	0.985	0.772	0.051	0.065	0.888

correlation coefficient gives an estimate of how much the relative ranking of methods according to ERGAS has changed under the two evaluation protocols. This is repeated for all metrics for synthesis and consistency. For the QNR metric, only the combined QNR index is used and not the D_λ and D_S sub-metrics. A high value in the table indicates that the relative ranking is similar. A low value indicates that the relative ranking has changed significantly and a negative number can mean that the metrics being compared demonstrate an opposite trend. The experiment is repeated 3 times, i.e., for all 18 methods, for the 7 CS methods and finally, the 11 MRA methods.

A general trend seen in the table, is that the relative ranking of methods according to synthesis and consistency is highly correlated and especially so for the CS methods. Another striking result is the low correlation of the ranking of methods according to the QNR metrics when computed for all methods for both synthesis and consistency metrics. However, the correlation is high for the CS methods but for the MRA methods, the ranking shows an opposite trend, i.e., the correlation coefficient becomes negative. Another observation is that the correlation between synthesis and consistency is generally higher for the WV2 data set. It is mostly the very high correlation between the CS methods that is the main contributing factor. The same applies for the QNR metric vs. synthesis and consistency, i.e., the correlation is higher for the WV2 data set. This might be caused by the difference in type of scenery and also by the slight misalignment between the MS and PAN images for the QB data set.

5.2.3 Scale Invariance of the Synthesis and Consistency Properties

In this experiment, it is investigated how well the synthesis and consistency protocols conserve relative method ranking across different resolution scales. When evaluating methods using the synthesis protocol, an important assumption being made is that the methods being evaluated should behave similarly at the full resolution scale. This means that if a particular method being evaluated is better than a comparison method according to the synthesis metrics at the reduced scale, it should also be better if the methods were evaluated at the full scale, given that a reference image is available, which is of course not the case. Hence, the assumption. A similar assumption needs to be made for the consistency property, since the evaluated image is on the same resolution scale as the observed MS image. A good scale invariance strengthens the argument for using the consistency property as a measure of pansharpening performance at the full scale.

Three resolution scales are considered, where the MS and PAN images are reduced in scale by a factor of two for each step. The important difference between the two protocols is that the MS and PAN images are degraded in resolution prior to fusion using the synthesis protocol, while for the consistency protocol, the degradation is done after fusion. This means that for the WorldView-2 data set, where the PAN image is 4096×4096 pixels and the MS image is $1024 \times 1024 \times 8$ pixels, they are reduced to 1024×1024 pixels and $256 \times 256 \times 8$ pixels for the first resolution scale,

Table 5.3: Correlation of method ranking according to synthesis vs. consistency and vs. the QNR metric. The number in parentheses indicates the number of methods evaluated.

All methods (18)						
	WV2			QB		
	Syn vs. Con	Syn vs. QNR	Con vs. QNR	Syn vs. Con	Syn vs. QNR	Con vs. QNR
ERGAS	0.637	0.232	0.344	0.744	-0.106	0.207
SAM	0.763	0.203	0.172	0.763	-0.007	-0.005
Q8	0.707	0.179	0.426	0.657	0.057	0.333
Q	0.814	0.158	0.344	0.651	0.071	0.263
mean	0.730	0.193	0.321	0.704	0.004	0.200
CS methods (10)						
	Syn vs. Con	Syn vs. QNR	Con vs. QNR	Syn vs. Con	Syn vs. QNR	Con vs. QNR
ERGAS	0.857	0.857	0.893	0.571	0.571	0.643
SAM	0.821	0.607	0.286	0.714	0.464	0.143
Q8	0.964	0.750	0.786	0.893	0.714	0.857
Q	0.964	0.857	0.893	0.786	0.750	0.714
mean	0.902	0.768	0.714	0.741	0.625	0.589
MRA methods (8)						
	Syn vs. Con	Syn vs. QNR	Con vs. QNR	Syn vs. Con	Syn vs. QNR	Con vs. QNR
ERGAS	0.455	-0.636	-0.218	0.791	-0.636	-0.300
SAM	0.646	-0.518	-0.146	0.746	-0.518	-0.164
Q8	0.455	-0.600	-0.091	0.582	-0.600	-0.073
Q	0.682	-0.627	-0.164	0.655	-0.518	-0.182
mean	0.559	-0.595	-0.155	0.693	-0.568	-0.180

Table 5.4: Correlation between 3 different resolution scales for the synthesis and consistency properties using the WorldView-2 data set. The number in parentheses indicates the number of methods evaluated.

All methods (18)								
	Synthesis				Consistency			
Resolution level	ERGAS	SAM	Q8	Q	ERGAS	SAM	Q8	Q
1 vs. 2	0.994	0.994	0.969	0.971	0.986	0.996	0.994	0.977
2 vs. 4	0.973	0.961	0.990	0.965	0.938	0.946	0.951	0.944
1 vs. 4	0.979	0.948	0.959	0.955	0.905	0.934	0.940	0.905
mean	0.982	0.968	0.972	0.964	0.943	0.959	0.961	0.942
CS methods (10)								
	Synthesis				Consistency			
Resolution level	ERGAS	SAM	Q8	Q	ERGAS	SAM	Q8	Q
1 vs. 2	1.000	0.964	1.000	0.964	1.000	1.000	1.000	1.000
2 vs. 4	0.964	1.000	1.000	1.000	0.964	0.964	0.857	0.964
1 vs. 4	0.964	0.964	1.000	0.964	0.964	0.964	0.857	0.964
mean	0.976	0.976	1.000	0.976	0.976	0.976	0.905	0.976
MRA methods (8)								
	Synthesis				Consistency			
Resolution level	ERGAS	SAM	Q8	Q	ERGAS	SAM	Q8	Q
1 vs. 2	1.000	1.000	0.936	0.927	0.982	0.982	0.991	0.900
2 vs. 4	0.973	0.918	0.991	0.946	0.964	0.927	1.000	0.955
1 vs. 4	0.973	0.918	0.946	0.955	0.946	0.882	0.991	0.818
mean	1.000	0.945	0.958	0.942	0.964	0.930	0.994	0.891

Table 5.5: Correlation between 3 different resolution scales for the synthesis and consistency properties using the QuickBird data set. The number in parentheses indicates the number of methods evaluated.

All methods (18)								
	Synthesis				Consistency			
Resolution level	ERGAS	SAM	Q8	Q	ERGAS	SAM	Q8	Q
1 vs. 2	0.994	0.994	0.969	0.971	0.986	0.996	0.994	0.977
2 vs. 4	0.973	0.961	0.990	0.965	0.938	0.946	0.951	0.944
1 vs. 4	0.979	0.948	0.959	0.955	0.905	0.934	0.940	0.905
mean	0.982	0.968	0.972	0.964	0.943	0.959	0.961	0.942
CS methods (7)								
	Synthesis				Consistency			
Resolution level	ERGAS	SAM	Q8	Q	ERGAS	SAM	Q8	Q
1 vs. 2	0.964	0.964	0.929	1.000	0.964	1.000	1.000	0.964
2 vs. 4	0.893	0.964	0.857	0.929	0.893	0.893	0.893	0.750
1 vs. 4	0.857	0.929	0.714	0.929	0.857	0.893	0.893	0.607
mean	0.905	0.952	0.833	0.952	0.905	0.929	0.929	0.774
MRA methods (11)								
	Synthesis				Consistency			
Resolution level	ERGAS	SAM	Q8	Q	ERGAS	SAM	Q8	Q
1 vs. 2	0.964	0.918	0.964	0.955	0.918	0.946	0.973	0.991
2 vs. 4	0.955	0.964	0.946	0.927	0.700	0.691	0.955	0.818
1 vs. 4	0.936	0.882	0.855	0.873	0.864	0.755	0.882	0.827
mean	0.952	0.921	0.921	0.918	0.827	0.797	0.936	0.879

yielding a fused image of $1024 \times 1024 \times 8$ pixels. For the next two resolution scales, the images are further reduced by factor of two, giving a fused image of $256 \times 256 \times 8$ pixels at the last resolution scale. For the consistency property, starting at the full scale, yields a degraded fused image of $1024 \times 1024 \times 8$ pixels, then $512 \times 512 \times 8$ pixels and finally $256 \times 256 \times 8$ pixels. The QB data set is smaller by a factor of two, so the fused image at the last resolution scale is $128 \times 128 \times 8$ pixels, which is rather small.

The relatively large size of the data sets make this study more reliable, especially for the WV2 data. The correlation between the method rankings for consecutive resolution scales is computed as well as the correlation between method ranking for the first and last resolution scale. This is computed for both data sets and for both synthesis and consistency. The different resolution scales are denoted by 1, 2 and 4, respectively. The results for this study are summarized in Tables 5.4 and 5.5, for each data set, respectively. As before, the experiment is performed using all the methods, CS methods only and MRA methods only.

Considering Table 5.4, it is evident that for the WV2 data set, synthesis has a slightly higher mean correlation for all the 3 different scales using all methods. However, consistency has a higher correlation between the first 2 resolution scales. For CS methods, the correlation is similar for the two protocols except for the Q8 metric, which has a significantly lower correlation for consistency than synthesis. This could be explained by the relatively compressed range of values for this metric for consistency, i.e., the relative variation between methods is smaller than that for synthesis and thus the relative method ranking is more prone to change. Again, the correlation between scales 1 and 2 for CS methods using consistency is very high or 100% for all metrics. Finally, for MRA methods, synthesis has a higher correlation and the correlation for the Q metric for consistency is surprisingly low between scales 1 and 4.

The results using the QB data set are summarized in Table 5.5. The trends are similar as for the WV2 data set. Using all methods, synthesis has a slightly higher mean correlation for all resolution scales, however the correlation for synthesis between the first two scales is slightly higher. For CS methods only, there is some drop in correlation for both protocols. It is notable how much the correlation drops for the Q8 metric for synthesis and the Q metric for consistency. As with the previous data set, synthesis has higher correlation for MRA methods.

The results of this experiment indicate that the correlation between resolution scales for the relative ranking of methods for the 4 metrics is very high in general and both protocols are largely scale invariant in this respect.

5.3 Conclusions

There is currently no standardized and universally accepted way for doing quantitative evaluation of pansharpened imagery. This applies especially for quantitative evaluation of pansharpened imagery at the full resolution scale. Pansharpening is an

inverse problem and there is no reference available. The preferred way of assessing the performance of pansharpening methods is to use the synthesis property, also referred to as Wald's protocol. This involves spatial degradation of the source images prior to the fusion such that the observed MS image can be used as a reference. The consistency property and its verification, which has been referred to here as the consistency protocol, has been considered as a necessary property that a pansharpening method must fulfill. The consistency protocol is fusing the source images at the full resolution and then spatially decimating the fused image such that it can be compared to the observed MS image. The question that has been addressed in this chapter is whether the verification of the consistency property, i.e., using the consistency protocol for quantitative evaluation of pansharpened images, can give a better idea of the relative performance of pansharpening methods at the full resolution scale than the synthesis protocol and the QNR metrics.

Three experiments where performed. The first one involved the quantitative evaluation of 2 real data sets using all three protocols and investigating which methods gave the best results according to each protocol. The methods used in all experiments are 7 CS methods and 11 MRA methods. For the first data set, the synthesis and consistency protocols agreed on the best overall and MRA method, while for the second data set, they agreed on the best CS method, which was the same CS method that gave best results for the first data set according to consistency. The best methods in either category of methods were the same according to consistency for both data sets. According to the QNR metrics, the same CS and MRA methods were ranked best for both data sets but judging by visual inspection, the images produced by those methods were far from being optimal. Another interesting observation from this experiment is the lack of any agreement between the QNR metrics at full scale and the synthesis and consistency protocols.

The second experiment was to determine the correlation between the relative ranking of pansharpening methods between the synthesis and consistency protocols. The experiment was performed for all evaluation methods, on the CS and MRA methods. The correlation between synthesis and consistency turned out to be quite good and also the correlation between synthesis/consistency and the QNR metrics for CS methods. For MRA methods, the correlation was negative, implying a negative trend for the metrics.

The final experiment was performed to investigate the scale invariance of the synthesis and consistency protocols. Scale invariance is an important assumption for both protocols. For the synthesis protocol, the data need to be spatially decimated prior to fusion, thus the assessment is performed on a lower resolution scale. For the consistency protocol the fused image is spatially decimated after the fusion. The experiment verified that both protocols are to a large extent invariant to resolution scale with respect to the relative ranking of methods. Especially, for methods belonging to the CS family.

The experiments suggest that the consistency protocol can indeed be used to give

an idea of the performance of pansharpening methods at the full scale and is highly correlated with the synthesis protocol. Another important result is that the QNR metrics do not seem to be reliable for evaluating pansharpend images based on visual inspection.

Chapter 6

Conclusions

In this chapter, the conclusions and main contributions of the thesis are given. Future research topics are also discussed.

6.1 Main Contributions

In this thesis, two new methods for image fusion in remote sensing have been presented, a new method to enhance the results obtained by CS methods and in some cases MRA methods was presented and finally, the problem of quantitative quality evaluation of pansharpened images was addressed.

6.1.1 Model-Based Fusion of Multi- and Hyperspectral Images using PCA and Wavelets

A method that performs MS/HS fusion as well as pan/hyper-sharpening was presented. It is based on a forward model for the high spectral resolution observed image and the assumption that the high resolution fused image and the auxiliary higher resolution observed image are jointly Gaussian. By assuming that the spectral singular vectors of the observed and fused high spectral resolution images are identical, the model can be transformed using PCA. This decreases the dimensionality of the problem considerably, since only the first few PCs need to be estimated. Next, the model is transformed into the wavelet domain using the UDWT. This decreases the correlation of data, making pixel-wise estimation practical. A pixel-wise MAP estimator is then derived for the detail UDWT coefficients of the PCs of the high resolution fused image. After the estimation of the UDWT coefficients, there are two options for the reconstruction of the fused image. The first one is to retain all PCs, both sharpened and unsharpened. This is preferred approach if there is little noise in the lower resolution image. If it is very noisy, then the second approach is to retain only the sharpened PCs and thus perform a reduced inverse PCA transform to obtain the fused image.

The proposed method is shown to give better results than the comparison methods for pansharpening, hypersharpening and MS/HS fusion, and to be significantly more tolerant to noise.

6.1.2 MTF Based Deblurring Using a Wiener Filter for CS and MRA Pansharpening Methods

The CS and MRA families of pansharpening methods can be generally described using simple detail injection schemes. In these schemes, details obtained from the PAN image

are added to the interpolated MS band, modulated by a injection gain. In Chapter 3, a method was presented that extracts more spectral and spatial detail from the interpolated MS bands, using Wiener filter deblurring, where the kernel is based on the MTF of the MS sensor. A simple method to suppress frequency deblurring artifacts was also presented.

A critical difference between the CS and MRA schemes is how the details are extracted from the PAN image. For CS methods, a low-pass approximation of the PAN image which is a linear combination of the interpolated MS bands is subtracted from the PAN image, while for MRA methods, a low-pass filtered PAN image is subtracted from the PAN image to obtain the high-pass details. This makes the proposed method generally not suitable for MRA methods. However, a certain class of MRA methods can greatly benefit from the proposed deconvolution pre-processing. Experiments were performed on 3 data sets using a number of CS and MRA methods. The results indicate that the proposed method can enhance CS methods in general, and some MRA methods, according to quantitative quality evaluation and visual inspection.

6.1.3 Multispectral and Hyperspectral Image Fusion Using a 3D-Convolutional Neural Network

A method based on using a 3D-CNN was presented. Since the training of the 3D-CNN is supervised, a target high resolution image is required. For image fusion, a high resolution target image is never available, thus the input data need to be spatially decimated (low-pass filtered and downsampled) such that the observed HS image can be used as the target image. Here it is assumed that the relationship between the input and target data, learned by the 3D-CNN at a lower resolution scale, also applies for a higher resolution scale.

An important step in the method is to use PCA for dimension reduction of the HS image prior to training of the 3D-CNN. By doing this, the dimension of the HS image can be greatly reduced which results in a reduction in computational complexity. This step requires the assumption that the spectral singular vectors of the HS image at the higher resolution scale are identical to those obtained for the lower resolution scale. This assumption was also used in Chapter 2.

Experiments demonstrated that the proposed method gives substantially better results than the methods from Chapter 2.

6.1.4 Quantitative Quality Evaluation of Pansharpened Imagery: Consistency vs. Synthesis

The quantitative quality evaluation of pansharpened images is difficult due to the inherent lack of a high resolution reference image. During the past two decades, many methods and quality metrics have been proposed to address this problem. Today, there

are two protocols that are mostly used. The former is based on the synthesis property and requires that the observed data be downgraded such that the fused image has the same size as the observed MS image which is the reference image. The latter approach is to evaluate the fused image at its full resolution using the QNR metrics.

It was investigated whether the verification of the consistency property can be useful for quantitative evaluation of pansharpened images. The property states that the bands of the fused image when degraded (using MTF matched low-pass Gaussian filters) should be as identical as is possible to the bands of the observed MS image. The main argument against using the synthesis property is that the fusion problem being evaluated is far from being the same as that for the full resolution images. Experiments using three data sets and a number of CS and MRA methods revealed that the verification of the consistency property can indeed be used to compare different pansharpening methods and while the ranking of the methods according to consistency is highly correlated with the ranking obtained using synthesis at the reduced resolution scale, it is less sensitive to the type of data set used. It was also shown that the consistency property is largely scale-invariant as is the case with synthesis.

6.2 Further work

The work presented in this thesis can be further improved in future research. Below is a list of possible topics for future research, detailed for each topic of the thesis.

Model-Based Fusion of Multi- and Hyperspectral Images using PCA and Wavelets

- The choice of the blurring filters used to downgrade the PAN image is important. Choosing this filter in some optimal manner could considerably enhance the method, since it is basically an MRA method.
- Using coupled NMF decomposition on the MS and HS images to yield endmembers and abundances could replace the PCA transform.

MTF Based Deblurring Using a Wiener Filter for CS and MRA Pansharpening Methods

- While CS methods are not affected by the ringing artifacts due to the Wiener filter, it makes the method unusable for most MRA methods. Investigation of better deblurring methods and/or better deringing methods could improve the proposed method.
- One deringing method that gives good results is to use a guided filter [164] on the deblurred MS image, with the PAN image as the guide. This can significantly

reduce the ringing artifacts, while at the same time, preserve the most important features of the deblurred image.

Multispectral and Hyperspectral Image Fusion Using a 3D-Convolutional Neural Network

- There are several hyper-parameters that have a large impact on the quality of the fused image that could be chosen better for optimal results. These are, e.g., number of convolutional layers, number of filters in each layer, batch size and the sizes of filters.
- Implementing a network loss function that takes spatial structure into account, such as the structural similarity index (SSIM) [139].

Quantitative Quality Evaluation of Pansharpened Imagery: Consistency vs. Synthesis

- One interesting property of the consistency protocol is its sensitivity to misalignment between the PAN and MS image. When the images are not properly aligned or co-registered, it will obviously affect the fused image. The synthesis protocol does not seem to be sensitive to small misalignments, since the images are degraded in resolution prior to the fusion. Also, in the case of misalignment, its effects are different depending on whether the fused image comes from a CS or an MRA method. This is something that could be investigated in a future work.

Appendix

A

Evaluation Metrics and Comparison Methods

A.1 ERGAS, SAM, Q and Q4/Q8 metrics

In this appendix we denote the fused and reference images by $\mathbf{Z}$ and $\mathbf{Y}$, respectively. It is assumed that there are N pixels in each image and the number of bands is N_b. The n-th pixel of the l-th band of $\mathbf{Z}$ is denoted by $\mathbf{Z}_{n,l}$.

A.1.1 ERGAS

ERGAS [135] calculates the amount of spectral/spatial distortion in the fused image based on the MSE and is given by

$$\text{ERGAS} = 100R\sqrt{\frac{1}{N_b}\sum_{l=1}^{N_b}\left(\frac{\sqrt{\sum_{n=1}^{N}(\mathbf{Z}_{n,l}-\mathbf{Y}_{n,l})^2/N}}{\sum_{n=1}^{N}\mathbf{Z}_{n,l}/N}\right)^2},$$

where R is the ratio of pixels in the PAN image to the MS image.

A.1.2 Spectral Angle Mapper (SAM)

SAM [113] calculates the spectral similarity between two vectors as an angle. The value of SAM for the entire image is the average of all the angles for each pixel.

$$\text{SAM}(\mathbf{Z},\mathbf{Y}) = \arccos\left(\frac{1}{N}\sum_{n=1}^{N}\frac{\sum_{l=1}^{N_b}\mathbf{Z}_{n,l}\mathbf{Y}_{n,l}}{\sum_{l=1}^{N_b}\mathbf{Z}_{n,l}^2\sum_{l=1}^{N_b}\mathbf{Y}_{n,l}^2}\right).$$

A.1.3 Universal Image Quality Index

Q [136] models spectral distortion as a combination of three factors, i.e., loss of correlation, luminance distortion and contrast distortion. Q is given by

$$Q(\mathbf{Z},\mathbf{Y}) = \frac{4\sigma_{\mathbf{ZY}}\bar{\mathbf{Z}}\bar{\mathbf{Y}}}{(\sigma_{\mathbf{Z}}^2+\sigma_{\mathbf{Y}}^2)[(\bar{\mathbf{Z}})^2+(\bar{\mathbf{Y}})^2]},$$

where

$$\bar{\mathbf{Z}} = \frac{1}{LN}\sum_{l=1}^{N_b}\sum_{n=1}^{N}\mathbf{Z}_{n,l},$$
$$\bar{\mathbf{Y}} = \frac{1}{LN}\sum_{l=1}^{N_b}\sum_{n=1}^{N}\mathbf{Y}_{n,l},$$
$$\sigma^2_{\mathbf{Z}} = \frac{1}{LN-1}\sum_{l=1}^{N_b}\sum_{n=1}^{N}(\mathbf{Z}_{n,l}-\bar{\mathbf{Z}})^2,$$
$$\sigma^2_{\mathbf{Y}} = \frac{1}{LN-1}\sum_{l=1}^{N_b}\sum_{n=1}^{N}(\mathbf{Y}_{n,l}-\bar{\mathbf{Y}})^2,$$
$$\sigma_{\mathbf{ZY}} = \frac{1}{LN-1}\sum_{l=1}^{N_b}\sum_{n=1}^{N}(\mathbf{Z}_{n,l}-\bar{\mathbf{Z}})(\mathbf{Y}_{n,l}-\bar{\mathbf{Y}}).$$

This index is calculated using a sliding window. A common size is 32 by 32 pixels but it depends on the size of the images. For very large images larger window, e.g., 64 by 64 pixels, might be needed.

A.1.4 Q4 and Q8 metrics

These metrics are hyper-complex vector extensions of the Q metric given above. Each pixel is represented using quaternions and octonions for the Q4 (4 bands) and Q8 (8 bands) metrics, respectively. In [138], a general formulation is given for images with 2^n bands where n is an integer.

A.1.5 QNR metric

This metric consists of two separate metrics or indices called D_λ, which is supposed to measure spectral distortion and D_{S}, which is supposed to measure spatial distortion, respectively. the UIQI scores between bands (inter-band) of the observed MS image. What this means is that the inter-band UIQI scores of the observed MS image should ideally be the same as the inter-band UIQI scores for the pansharpened image. The D_λ metric is given by

$$D_\lambda = \sqrt[p]{\frac{1}{N_b(N_b-1)}\sum_{l=1}^{N_b}\sum_{\substack{r=1\\ r\neq l}}^{N_b}\left|Q(\mathbf{Z}_l,\mathbf{Z}_r)-Q(\mathbf{Y}_l,\mathbf{Y}_r)\right|^p}, \tag{A.1}$$

where Q is the UIQI index and the constant p is usually chosen to be 1. The D_{S} metric is given by

$$D_{\mathrm{S}} = \sqrt[q]{\frac{1}{N_b}\sum_{l=1}^{N_b}\left|Q(\mathbf{Z_l},\mathbf{X}) - \mathbf{Q}(\mathbf{Y_l},\widetilde{\mathbf{X}})\right|^{\mathbf{q}}}, \tag{A.2}$$

where $\widetilde{\mathbf{X}}$ is the spatially degraded PAN image such it has the same size as the observed MS image and q is typically chosen to be 1. Finally, the QNR index is given by

$$\mathrm{QNR} = (1 - D_\lambda)^\alpha (1 - D_{\mathrm{S}})^\beta. \tag{A.3}$$

The constants α and β control the relative relevance of the one's complements of the D_λ and D_{S} indices and they are usually chosen to be 1. The maximum value of QNR is 1 when D_λ and D_{S} are zero and the minimum value of QNR is zero.

A.2 A Short Overview of MTF filters and Comparison Methods

In this appendix, we give information on the MTF filters for the QB and WV2 sensors, the formulas used to calculate the ERGAS, SAM and Q metrics and finally a short overview of all methods used for comparison in experiments.

A.2.1 Sensor MTF specification

The MTF filters used in this work are Gaussian low pass filters and the amplitude of the MTF at the Nyquist frequency is given in Table A.1

Table A.1: MTF amplitude at the Nyquist frequency for the QB and WV2 sensors. The last 4 bands are WV2 specific.

	Sensor	
Band	QB	WV2
PAN	0.15	0.11
Blue	0.34	0.35
Green	0.32	0.35
Red	0.3	0.35
NIR	0.22	0.35
Red edge	NA	0.35
Coastal	NA	0.35
Yellow	NA	0.35
NIR2	NA	0.27

A.2.2 Overview of Pansharpening Methods

Table A.2 shows an overview of all the CS and MRA methods used in the experiments. For a detailed overview see [16].

Table A.2: Overview of all the methods used in the experiments.

CS methods			
Method	Name	ω_k	Gain (g_k)
IHS [165]	Intensity-Hue-Saturation	$1/N$	1
GIHS [17, 18]	Generalized Intensity-Hue-Saturation	$\omega_l \geq 0$	$\left(\sum_{l=1}^{L} \omega_l\right)^{-1}$
BROVEY [23]	Brovey Transform	$1/N$	$\frac{\mathbf{Z_k}}{\mathbf{J}_L}$
PCA [22]	Principal Component Analysis	First L columns of PCA matrix	First L rows of PCA matrix
GS [19]	Gram-Schmidt spectral sharpening	$1/N$	$\frac{\sigma_{\mathbf{J}_L,\mathbf{Z_k}}}{\sigma^2_{\mathbf{J}_L}}$
GSA [20]	Adaptive GS	eq. (1.2)	$\frac{\sigma_{\mathbf{J}_L,\mathbf{Z_k}}}{\sigma^2_{\mathbf{J}_L}}$
BDSD [15]	Band Dependent Spatial Detail	see [16]	see [16]
PRACS [158]	Partial Replacement Adaptive CS	eq. (1.2)	see [16]
MRA methods			
Method	Name	WT/filter	Gain (g_k)
HPF [27]	High-Pass Filtering	ATWT w/ box filter	1
HPM [166]	High-Pass Modulation	ATWT w/ box filter	$\frac{\mathbf{Z_k}}{\mathbf{X}_{\mathrm{LP}}}$
SFIM [160]	Smoothing Filter based Intensity Modulation	ATWT w/ box filter	$\frac{\mathbf{Z_k}}{\mathbf{X}_{\mathrm{LP}}}$
Indusion [30]	Wavelet decomposition pyramids	DWT using CDF bi-orthogonal filter	1
MTF-GLP [32]	Gaussian Laplacian Pyramids using MTF filters	GLP using MTF filter	1
MTF-GLP-CBD [32]	MTF-GLP with Context Based Decision	GLP using MTF filter	$\frac{\sigma_{\mathbf{X}_{\mathrm{LP}},\mathbf{Z_k}}}{\sigma^2_{\mathbf{X}_{\mathrm{LP}}}}$
ATWT [29]	Additive A-trous Wavelet Transform	ATWT using S&M filter [148]	$\frac{\mathbf{Z_k}}{\mathbf{X}_{\mathrm{LP}}}$
ATWT-M2 [25]	Additive A-trous Wavelet Transform	ATWT using S&M filter [148]	see [25]
ATWT-M3 [25]	Additive A-trous Wavelet Transform	ATWT using S&M filter [148]	see [25]
MTF-GLP-HPM [159]	MTF-GLP-HPM	GLP using MTF filter	$\frac{\mathbf{Z_k}}{\mathbf{X}_{\mathrm{LP}}}$
AWLP [28]	Additive Wavelet Luminance Proportional	ATWT using S&M filter [148]	$\frac{\mathbf{Z_k}}{\frac{1}{N}\sum_{l=1}^{L} \mathbf{Y}_l}$

Bibliography

[1] Israa Amro, Javier Mateos, Miguel Vega, Rafael Molina, and Aggelos K Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–22, Sept. 2011. [Cited on pages 1 and 3]

[2] R Welch and Manfred Ehlers, "Merging multiresolution spot hrv and landsat tm data," *Photogrammetric Engineering and Remote Sensing*, vol. 53, pp. 301–303, 1987. [Cited on page 3]

[3] Pat S Chavez Jr, "Comparison of the spectral information content of landsat thematic mapper and spot for three different sites in the phoenix, arizona region," *Photogrammetric Engineering and Remote Sensing*, vol. 54, no. 12, pp. 1699–1708, 1988. [Cited on page 3]

[4] F. Bovolo, L. Bruzzone, L. Capobianco, A. Garzelli, S. Marchesi, and F. Nencini, "Analysis of the effects of pansharpening in change detection on VHR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 1, pp. 53–57, Jan. 2010. [Cited on page 3]

[5] F. Palsson, J. R. Sveinsson, J. A. Benediktsson, and H. Aanaes, "Classification of pansharpened urban satellite images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 281 –297, Feb. 2012. [Cited on page 3]

[6] Thierry Ranchin, Bruno Aiazzi, Luciano Alparone, Stefano Baronti, and Lucien Wald, "Image fusion—the ARSIS concept and some successful implementation schemes," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 1, pp. 4–18, 2003. [Cited on page 3]

[7] Zhijun Wang, Djemel Ziou, Costas Armenakis, Deren Li, and Qingquan Li, "A comparative analysis of image fusion methods," *IEEE transactions on geoscience and remote sensing*, vol. 43, no. 6, pp. 1391–1402, 2005. [Cited on page 3]

[8] Bruno Aiazzi, Stefano Baronti, Franco Lotti, and Massimo Selva, "A comparison between global and context-adaptive pansharpening of multispectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 302–306, 2009. [Cited on pages 3 and 4]

[9] Christine Pohl and John van Genderen, "Structuring contemporary remote sensing image fusion," *International Journal of Image and Data Fusion*, vol. 6, no. 1, pp. 3–21, 2015. [Cited on page 3]

[10] Andrea Garzelli, "A review of image fusion algorithms based on the super-resolution paradigm," *Remote Sensing*, vol. 8, no. 10, pp. 797, 2016. [Cited on page 3]

[11] L Leung, J Liu, and J Zhang, "An improved adaptive intensity hue saturation method for the fusion of remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 985–989, May 2014. [Cited on pages 3, 5, and 6]

[12] Timo Bretschneider and Odej Kao, "Image fusion in remote sensing," in *Proceedings of the 1st Online Symposium of Electronic Engineers*, 2000, pp. 1–8. [Cited on page 3]

[13] I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2005. [Cited on pages 4, 17, 22, and 67]

[14] Te-Ming Tu, Shun-Chi Su, Hsuen-Chyun Shyu, and Ping S. Huang, "A new look at IHS-like image fusion methods," *Information Fusion*, vol. 2, no. 3, pp. 177 – 186, Sept. 2001. [Cited on page 4]

[15] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 1, pp. 228–236, Jan. 2008. [Cited on pages 5, 48, and 105]

[16] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G.A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, May 2015. [Cited on pages 5, 6, 30, 43, 44, 47, 81, 82, 104, and 105]

[17] Te-Ming Tu, Shun-Chi Su, Hsuen-Chyun Shyu, and Ping S Huang, "A new look at IHS-like image fusion methods," *Information fusion*, vol. 2, no. 3, pp. 177–186, Sept. 2001. [Cited on pages 5, 87, and 105]

[18] Te-Ming Tu, P.S. Huang, Chung-Ling Hung, and Chien-Ping Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 309–312, Oct. 2004. [Cited on pages 5 and 105]

[19] C.A. Laben and B.V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," Jan. 4 2000, US Patent 6,011,875. [Cited on pages 5 and 105]

[20] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007. [Cited on pages 5, 48, and 105]

[21] P. S. Chavez Jr. and A. W. Kwarteng, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm Eng. Remote Sens.*, vol. 55, no. 3, pp. 329–348, Mar. 2004. [Cited on page 5]

[22] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set,"

Photogramm Eng. Remote Sens., vol. 58, no. 5, pp. 561–567, Jan. 1992. [Cited on pages 5, 30, and 105]

[23] Alan R Gillespie, Anne B Kahle, and Richard E Walker, "Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques," *Remote Sensing of Environment*, vol. 22, no. 3, pp. 343 – 365, Sept. 1987. [Cited on pages 5 and 105]

[24] P. Dutilleux, "An implementation of the "algorithme à trous" to compute the wavelet transform," in *Wavelets*, Jean-Michel Combes, Alexander Grossmann, and Philippe Tchamitchian, Eds., Inverse Problems and Theoretical Imaging, pp. 298–304. Springer Berlin Heidelberg, 1989. [Cited on pages 5 and 24]

[25] Thierry Ranchin, Lucien Wald, et al., "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 1, pp. 49–61, Jan. 2000. [Cited on pages 5, 48, 55, and 105]

[26] P. Massip, P. Blanc, and L. Wald, "A method to better account for modulation transfer functions in ARSIS-based pansharpening methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 800–808, Mar. 2012. [Cited on pages 5, 10, 12, 17, 43, 47, 55, and 56]

[27] M. Shensa, "The discrete wavelet transform: wedding the a trous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992. [Cited on pages 6 and 105]

[28] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1204–1211, May 1999. [Cited on pages 6, 30, 48, 55, and 105]

[29] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 930–934, May 2014. [Cited on pages 6 and 105]

[30] M.M. Khan, J. Chanussot, L. Condat, and Annick Montanvert, "Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 98–102, Jan. 2008. [Cited on pages 6 and 105]

[31] P.J. Burt and E.H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, Apr. 1983. [Cited on page 6]

[32] B Aiazzi, L Alparone, S Baronti, A Garzelli, and M Selva, "MTF-tailored multiscale fusion of high-resolution ms and pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006. [Cited on pages 6, 49, and 105]

[33] J Lee and C Lee, "Fast and efficient panchromatic sharpening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 155–163, Jan. 2010. [Cited on page 6]

[34] L. Alparone, S. Baronti, B. Aiazzi, and A. Garzelli, "Spatial methods for multispectral pansharpening: Multiresolution analysis demystified," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2563–2576, May 2016. [Cited on page 6]

[35] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone, "A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 446–453, June 2011. [Cited on pages 6 and 46]

[36] M. R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot, "A pansharpening method based on the sparse representation of injected details," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 180–184, Jan 2015. [Cited on pages 7 and 8]

[37] M. Cheng, C. Wang, and J. Li, "Sparse representation based pansharpening using trained dictionary," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 293–297, Jan 2014. [Cited on page 7]

[38] X. X. Zhu, C. Grohnfeldt, and R. Bamler, "Exploiting joint sparsity for pansharpening: The j-sparsefi algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2664–2681, May 2016. [Cited on page 7]

[39] V. Patel and K. P. Upla, "Compressed sensing based pansharpening technique with learned dictionary," in *Signal Propagation and Computer Technology (ICSPCT), 2014 International Conference on*, July 2014, pp. 201–204. [Cited on page 7]

[40] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "Two-step sparse coding for the pan-sharpening of remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 5, pp. 1792–1805, May 2014. [Cited on page 7]

[41] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2827–2836, May 2013. [Cited on page 7]

[42] M. Ghahremani and H. Ghassemian, "A compressed-sensing-based pansharpening method for spectral distortion reduction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2194–2206, April 2016. [Cited on page 7]

[43] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 738–746, Feb 2011. [Cited on pages 7 and 8]

[44] David L Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006. [Cited on page 8]

[45] E. Candes and T. Tao, "The Dantzig selector: statistical estimation when p is much larger then n," *The Annals of Statisitcs*, vol. 35, no. 6, 2007. [Cited on page 8]

[46] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006. [Cited on page 8]

[47] David L Donoho, Michael Elad, and Vladimir N Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on information theory*, vol. 52, no. 1, pp. 6–18, 2006. [Cited on page 8]

[48] David L Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006. [Cited on page 8]

[49] Yaakov Tsaig and David L Donoho, "Extensions of compressed sensing," *Signal processing*, vol. 86, no. 3, pp. 549–571, 2006. [Cited on page 8]

[50] Claude Elwood Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949. [Cited on page 8]

[51] Yagyensh Chandra Pati, Ramin Rezaiifar, and PS Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44. [Cited on page 8]

[52] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Pansharpening via sparsity optimization using overcomplete transforms," in *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, July 2013, pp. 856–859. [Cited on page 8]

[53] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 318–322, Jan. 2014. [Cited on page 8]

[54] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992. [Cited on pages 8 and 44]

[55] James M Ortega and Werner C Rheinboldt, *Iterative solution of nonlinear equations in several variables*, vol. 30, Siam, 1970. [Cited on page 8]

[56] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model based pansharpening method based on tv and mtf deblurring," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 33–36. [Cited on pages 8 and 18]

[57] X. He, L. Condat, J. M. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 4160–4174, Sept 2014. [Cited on page 8]

[58] Magnus R Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969. [Cited on page 9]

[59] C. Ballester, V. Caselles, L. Igual, and J. Verdera, "A variational model for p+xs image fusion.," *International Journal of Computer Vision*, pp. 43–58, 2006. [Cited on page 9]

[60] Z. m. Zhou, Y. x. Li, H. q. Shi, N. Ma, C. He, and P. Zhang, "A fast variational fusion approach for pan-sharpening," in *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*, Oct 2010, pp. 1110–1113. [Cited on page 9]

[61] Z. m. Zhou, Z. j. Wu, J. Wang, P. l. Yang, and L. Jiang, "Panchromatic and multi-spectral image fusion using ihs and variational models," in *Image and Signal Processing (CISP), 2012 5th International Congress on*, Oct 2012, pp. 1077–1080. [Cited on page 9]

[62] Z. m. Zhou, P. l. Yang, Y. x. Li, W. Yin, and L. Jiang, "Joint ihs and variational methods for pan-sharpening of very high resolution imagery," in *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, July 2013, pp. 2597–2600. [Cited on page 9]

[63] J. Duran, A. Buades, B. Coll, C. Sbert, and G. Blanchet, "Pansharpening by a nonlocal channel-decoupled variational method," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 4339–4343. [Cited on page 9]

[64] P. Liu, L. Xiao, and S. Tang, "A new geometry enforcing variational model for pan-sharpening," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. PP, no. 99, pp. 1–14, 2016. [Cited on page 9]

[65] Z. Li and H. Leung, "Fusion of multispectral and panchromatic images using a restoration-based method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1482–1491, May 2009. [Cited on page 9]

[66] G. Vivone, M. Simões, M. Dalla Mura, R. Restaino, J. M. Bioucas-Dias, G. A. Licciardi, and J. Chanussot, "Pansharpening based on semiblind deconvolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1997–2010, April 2015. [Cited on page 9]

[67] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105. [Cited on page 9]

[68] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9. [Cited on page 9]

[69] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012. [Cited on page 9]

[70] Tara N Sainath, Abdel-rahman Mohamed, Brian Kingsbury, and Bhuvana Ramabhadran, "Deep convolutional neural networks for lvcsr," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 8614–8618. [Cited on page 9]

[71] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, no. Aug, pp. 2493–2537, 2011. [Cited on page 9]

[72] Claire Adam-Bourdarios, Glen Cowan, Cecile Germain, Isabelle Guyon, Balazs Kegl, and David Rousseau, "Learning to discover: the higgs boson machine learning challenge," *URL http://higgsml. lal. in2p3. fr/documentation*, 2014. [Cited on page 9]

[73] Warren S McCulloch and Walter Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943. [Cited on page 9]

[74] Frank Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, pp. 386, 1958. [Cited on page 9]

[75] Geoffrey E Hinton, "Connectionist learning procedures," *Artificial intelligence*, vol. 40, no. 1, pp. 185–234, 1989. [Cited on page 9]

[76] B Boser Le Cun, John S Denker, D Henderson, Richard E Howard, W Hubbard, and Lawrence D Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems*. Citeseer, 1990. [Cited on page 9]

[77] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. [Cited on page 9]

[78] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997. [Cited on page 9]

[79] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014. [Cited on page 9]

[80] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 1037–1041, May 2015. [Cited on pages 9 and 17]

[81] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa, "Pan-sharpening by convolutional neural networks," *Remote Sensing*, vol. 8, no. 7, pp. 594, 2016. [Cited on pages 9 and 17]

[82] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hyper-sharpening: A first approach on sim-ga data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3008–3024, 2015. [Cited on page 9]

[83] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model-based fusion of multi- and hyperspectral images using PCA and wavelets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2652–2663, May 2015. [Cited on pages 10, 17, 18, 67, 73, and 75]

[84] G.A Licciardi, M.M. Khan, and J. Chanussot, "Fusion of hyperspectral and panchromatic images: A hybrid use of indusion and nonlinear pca," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept. 2012, pp. 2133–2136. [Cited on page 10]

[85] G.A Licciardi, M. M. Khan, J. Chanussot, A Montanvert, L. Condat, and C. Jutten, "Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear pca band reduction," in *2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2011, pp. 1783–1786. [Cited on page 10]

[86] L. Capobianco, A Garzelli, F. Nencini, L. Alparone, and S. Baronti, "Spatial enhancement of hyperion hyperspectral data through ali panchromatic image," in *2007 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2007, pp. 5158–5161. [Cited on page 10]

[87] A Garzelli, L. Capobianco, L. Alparone, B. Aiazzi, S. Baronti, and M. Selva, "Hyperspectral pansharpening based on modulation of pixel spectra," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Jun. 2010, pp. 1–4. [Cited on page 10]

[88] D. M. McKeown, S. D. Cochran, S. J. Ford, J. C. McGlone, J. A. Shufelt, and D. A. Yocum, "Fusion of hydice hyperspectral data with panchromatic imagery for cartographic feature extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1261–1277, May 1999. [Cited on page 10]

[89] Y. Zhao, J. Yang, and J. C. W. Chan, "Hyperspectral imagery super-resolution by spatial-spectral joint nonlocal similarity," *IEEE Journal of Selected Topics*

in Applied Earth Observations and Remote Sensing, vol. 7, no. 6, pp. 2671–2679, June 2014. [Cited on page 10]

[90] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2827–2836, May 2013. [Cited on page 10]

[91] G Licciardi, Miguel Angel Veganzones, G Vivone, L Loncan, and J Chanussot, "Impact of hybrid pansharpening approaches applied to hyperspectral images," *HAL Id: hal-01121191*, 2015. [Cited on page 10]

[92] M. M. Khan, J. Chanussot, and L. Alparone, "Hyperspectral pansharpening using qnr optimization constraint," in *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Aug 2009, pp. 1–4. [Cited on page 10]

[93] D. Picone, R. Restaino, G. Vivone, P. Addesso, and J. Chanussot, "Pansharpening of hyperspectral images: Exploiting data acquired by multiple platforms," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2016, pp. 7220–7223. [Cited on page 10]

[94] Laetitia Loncan, Luis B de Almeida, José M Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A Licciardi, Miguel Simoes, et al., "Hyperspectral pansharpening: a review," *IEEE Geoscience and remote sensing magazine*, vol. 3, no. 3, pp. 27–46, 2015. [Cited on page 10]

[95] Z. H. Nezhad, A. Karami, R. Heylen, and P. Scheunders, "Fusion of hyperspectral and multispectral images using spectral unmixing and sparse coding," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2377–2389, June 2016. [Cited on pages 10 and 11]

[96] Y. Zhang, Y. Wang, Y. Liu, C. Zhang, M. He, and S. Mei, "Hyperspectral and multispectral image fusion using cnmf with minimum endmember simplex volume and abundance sparsity constraints," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 1929–1932. [Cited on page 10]

[97] H Pande, Poonam S Tiwari, and Shashi Dobhal, "Analyzing hyper-spectral and multi-spectral data fusion in spectral domain," *Journal of the Indian Society of Remote Sensing*, vol. 37, no. 3, pp. 395–408, Sept. 2009. [Cited on page 10]

[98] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3658–3668, July 2015. [Cited on page 10]

[99] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1212–1226, May 1999. [Cited on page 10]

[100] Richard B Gomez, Amin Jazaeri, and Menas Kafatos, "Wavelet-based hyperspectral and multispectral image fusion," in *Aerospace/Defense Sensing, Simulation, and Controls*. International Society for Optics and Photonics, 2001, pp. 36–42. [Cited on page 10]

[101] Y. Kim, J. Choi, D. Han, and Y. Kim, "Block-based fusion algorithm with simulated band generation for hyperspectral and multispectral images of partially different wavelength ranges," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2997–3007, June 2015. [Cited on page 10]

[102] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of multispectral and hyperspectral images with unknown sensor spectral response," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 698–702. [Cited on page 10]

[103] N. Yokoya, J. Chanussot, and A. Iwasaki, "Hyperspectral and multispectral data fusion based on nonlinear unmixing," in *2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2012, pp. 1–4. [Cited on page 10]

[104] Z. Chen, H. Pu, B. Wang, and G. M. Jiang, "Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 8, pp. 1418–1422, Aug 2014. [Cited on page 10]

[105] C. Grohnfeldt, X. X. Zhu, and R. Bamler, "Jointly sparse fusion of hyperspectral and multispectral imagery," in *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, July 2013, pp. 4090–4093. [Cited on page 10]

[106] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, Feb 2012. [Cited on pages 10 and 11]

[107] N. Yokoya, T. Yairi, and A. Iwasaki, "Hyperspectral, multispectral, and panchromatic data fusion based on coupled non-negative matrix factorization," in *2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2011, pp. 1–4. [Cited on page 10]

[108] D. Sylla, A. Minghelli-Roman, P. Blanc, A. Mangin, and O. Hembise Fanton d'Andon, "Fusion of multispectral images by extension of the pan-sharpening arsis method," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 5, pp. 1781–1791, May 2014. [Cited on page 10]

[109] Yifan Zhang and Mingyi He, "Multi-spectral and hyperspectral image fusion using 3-d wavelet transform," *Journal of Electronics (China)*, vol. 24, no. 2, pp. 218–224, 2007. [Cited on page 10]

[110] Harry N Gross and John R Schott, "Application of spectral mixture analysis and image fusion techniques for image sharpening," *Remote Sensing of Environment*, vol. 63, no. 2, pp. 85–94, 1998. [Cited on page 10]

[111] Gary D Robinson, Harry N Gross, and John R Schott, "Evaluation of two applications of spectral mixing models to image fusion," *Remote Sensing of Environment*, vol. 71, no. 3, pp. 272–281, 2000. [Cited on page 10]

[112] D. Picone, R. Restaino, G. Vivone, P. Addesso, M. Dalla Mura, and J. Chanussot, "Multispectral and hyperspectral data fusion based on SAM minimization band assignment approach," in *2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2016. [Cited on page 10]

[113] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Summaries of the 3rd Annual JPL Airborne Geoscience Workshop*, 1992, vol. 1, pp. 147–149. [Cited on pages 10, 13, 29, 74, and 101]

[114] Nirmal Keshava, "A survey of spectral unmixing algorithms," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 55–78, 2003. [Cited on page 11]

[115] Michael E Winter, "N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*. International Society for Optics and Photonics, 1999, pp. 266–275. [Cited on page 11]

[116] José MP Nascimento and José MB Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005. [Cited on page 11]

[117] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Total variation and ℓ_q based hyperspectral unmixing for feature extraction and classification," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 437–440. [Cited on page 11]

[118] Xuesong Liu, Wei Xia, Bin Wang, and Liming Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 757–772, 2011. [Cited on page 11]

[119] R.C. Hardie, M.T. Eismann, and G.L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1174–1184, Sept. 2004. [Cited on page 11]

[120] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based bayesian fusion of multispectral and hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3834–3843, Nov. 2009. [Cited on pages 11, 17, 21, 22, 24, 27, 30, 39, 40, 67, 73, and 75]

[121] Curtis K Munechika, James S Warnick, Carl Salvaggio, and John R Schott, "Resolution enhancement of multispectral image data to improve classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 59, no. 1, pp. 67–72, Jan. 1993. [Cited on pages 12 and 80]

[122] Lucien Wald, Thierry Ranchin, Marc Mangolini, et al., "Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, Jun. 1997. [Cited on pages 12, 13, 17, 18, 23, 44, 79, and 80]

[123] L. Wald, *Data Fusion: Definitions and Architectures : Fusion of Images of Different Spatial Resolutions*, Les Presses de l'École des Mines, 2002. [Cited on pages 12, 23, and 80]

[124] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *International Journal of Remote Sensing*, vol. 19, no. 4, pp. 743–757, Nov. 1998. [Cited on pages 12, 14, and 80]

[125] M.M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3880 –3891, Nov. 2009. [Cited on pages 12, 14, 43, and 80]

[126] A. Medina, J. Marcello, D. Rodriguez, F. Eugenio, and J. Martin, "Quality evaluation of pansharpening techniques on different land cover types," in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2012, pp. 5442–5445. [Cited on pages 12 and 80]

[127] Qian Du, N.H. Younan, R. King, and V.P. Shah, "On the performance evaluation of pan-sharpening techniques," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 518–522, Oct. 2007. [Cited on pages 12 and 80]

[128] Wenjing Pei, Guian Wang, and Xianchuan Yu, "Performance evaluation of different references based image fusion quality metrics for quality assessment of remote sensing image fusion," in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2012, pp. 2280–2283. [Cited on pages 12 and 80]

[129] D. Rodriguez-Esparragon, J. Marcello-Ruiz, A. Medina-Machin, F. Eugenio-Gonzalez, C. Gonzalo-Martin, and A. Garcia-Pedrero, "Evaluation of the performance of spatial assessments of pansharpened images," in *2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2014, pp. 1619–1622. [Cited on pages 12 and 80]

[130] Veeraraghavan Vijayaraj, C.G. O'Hara, and N.H. Younan, "Pansharpening and image quality interface," in *2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Sept. 2004, vol. 4, pp. 2558–2560. [Cited on pages 12 and 80]

[131] L. Santurri, B. Aiazzi, S. Baronti, and R. Carla, "Influence of spatial resolution on pan-sharpening results," in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2012, pp. 5446–5449. [Cited on pages 12 and 80]

[132] B. Aiazzi, L. Alparone, S. Baronti, and I. Pippi, "Quality assessment of decision-driven pyramid-based fusion of high resolution multispectral with panchromatic image data," in *IEEE/ISPRS Joint Workshop 2001 Remote Sensing and Data Fusion over Urban Areas*, 2001, pp. 337–341. [Cited on pages 12 and 80]

[133] G. Palubinskas, "Quality assessment of pan-sharpening methods," in *2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2014, pp. 2526–2529. [Cited on pages 12, 14, and 80]

[134] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetric Engineering and Remote Sensing*, vol. 74, no. 2, pp. 193–200, Feb. 2008. [Cited on pages 12, 14, and 44]

[135] L. Wald, "Quality of high resolution synthesized images: Is there a simple criterion?," in *In Proceedings of the third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*, Thierry Ranchin and Lucien Wald, Eds. 2000, SEE/URISCA. [Cited on pages 13, 29, 74, and 101]

[136] Zhou Wang and A.C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, Mar. 2002. [Cited on pages 13 and 101]

[137] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 313 – 317, Oct. 2004. [Cited on page 13]

[138] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 662–665, Oct. 2009. [Cited on pages 13 and 102]

[139] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004. [Cited on pages 14, 74, and 100]

[140] Claire Thomas and Lucien Wald, "Comparing distances for quality assessment of fused images," in *26th EARSeL Symposium*. Millpress, 2007, pp. 101–111. [Cited on page 14]

[141] Xin Huang, Dawei Wen, Junfeng Xie, and Liangpei Zhang, "Quality assessment of panchromatic and multispectral image fusion for the zy-3 satellite: From an information extraction perspective," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 753–757, Apr. 2014. [Cited on page 14]

[142] Norbert Wiener, *Extrapolation, interpolation, and smoothing of stationary time series*, vol. 2, MIT press Cambridge, MA, 1949. [Cited on pages 17, 44, and 45]

[143] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model based PCA/wavelet fusion of multispectral and hyperspectral images," in *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 1532–1535. [Cited on page 18]

[144] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "MTF-deblurring preprocessing for CS and MRA pansharpening methods," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 1104–1107. [Cited on page 18]

[145] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "MTF-based deblurring using a Wiener filter for CS and MRA pansharpening methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2255–2269, June 2016. [Cited on page 18]

[146] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 639–643, May 2017. [Cited on page 18]

[147] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1247–1259, March 2016. [Cited on page 19]

[148] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996. [Cited on pages 30 and 105]

[149] Louis Landweber, "An iteration formula for fredholm integral equations of the first kind," *American journal of mathematics*, pp. 615–624, 1951. [Cited on page 43]

[150] F. Palsson, J.R. Sveinsson, M.O. Ulfarsson, and J.A. Benediktsson, "Model based pansharpening method based on TV and MTF deblurring," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 33–36. [Cited on page 43]

[151] William Hadley Richardson, "Bayesian-based iterative method of image restoration∗," *J. Opt. Soc. Am.*, vol. 62, no. 1, pp. 55–59, Jan 1972. [Cited on page 44]

[152] J Willard Gibbs, "Fourier's series," *Nature*, vol. 59, pp. 200, 1898. [Cited on page 44]

[153] Jian-Feng Cai, Stanley Osher, and Zuowei Shen, "Linearized bregman iterations for frame-based image deblurring," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 226–252, 2009. [Cited on page 44]

[154] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010. [Cited on page 44]

[155] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838–1857, 2011. [Cited on page 44]

[156] The Mathworks, Inc., Natick, Massachusetts, *MATLAB version 8.5.0.197613 (R2015a)*, 2015. [Cited on page 44]

[157] Azriel Rosenfeld and Avinash C. Kak, *Digital Picture Processing*, Academic Press, Inc., Orlando, FL, USA, 2nd edition, 1982. [Cited on page 45]

[158] Jaewan Choi, Kiyun Yu, and Yongil Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 295–309, Jan. 2011. [Cited on pages 48 and 105]

[159] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, 2003.*, May 2003, pp. 90–94. [Cited on pages 51 and 105]

[160] JG Liu, "Smoothing filter-based intensity modulation: a spectral preserve image fusion technique for improving spatial details," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3461–3472, 2000. [Cited on pages 51 and 105]

[161] David H Hubel and Torsten N Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, pp. 106–154, 1962. [Cited on page 68]

[162] A. Graves, A. r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6645–6649. [Cited on page 71]

[163] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. [Cited on page 73]

[164] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013. [Cited on page 99]

[165] W. J. Carper, T. M. Lillesand, and P. W. Kiefer, "The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, Jan. 1990. [Cited on page 105]

[166] Robert A Schowengerdt, *Remote sensing: models and methods for image processing*, Academic press, 2006. [Cited on page 105]