

Pansharpening and Classification of Pansharpened Images

Frosti Pálsson

A Thesis Presented in Partial Fulfillment of the
Requirements for the Degree
Master of Science in Electrical and Computer Engineering at the
University of Iceland
2013



Thesis Committee:

Professor Jóhannes R. Sveinsson, advisor

Associate Professor Magnús Örn Úlfarsson, co-advisor

Professor Anna Soffía Hauksdóttir

Háskóli Íslands / University of Iceland

Verkfræði- og náttúruvísindasvið / School of Engineering and Natural Sciences

VR-II, Hjardarhaga 2-6, IS-107 Reykjavík, Iceland

Phone + 354 525 4000

verkognatt@hi.is

www.hi.is

Abstract

Satellites provide very valuable data about the Earth, e.g., for environmental monitoring, weather forecasting, map-making and military intelligence. But satellites are expensive, both to build and operate. Therefore it is important to make the best use of the data obtained from available satellites, e.g., by combining the output from different sensors.

A good example of this is the fusion of multispectral satellite images of low spatial and high spectral resolution with panchromatic images of high spatial and low spectral resolution. This kind of image fusion is called pansharpening and it is the topic of the thesis.

The thesis is comprised of three parts. In the first part, an observational model for the pansharpening process is derived. A new pansharpening method is developed that involves solving an ill-posed inverse problem dictated by the observational model. The solution is based on a convex optimization problem that is regularized by total variation. The performance of the method is evaluated using quantitative quality metrics for pansharpening using two well-known datasets and the results are compared to existing state-of-the-art methods. The proposed method is shown to give excellent results.

In the second part of the thesis, the solution of the inverse problem for pansharpening using sparsity regularization is investigated. These methods exploit the sparsity of the coefficients of multi-scale overcomplete transforms. Methods based on the two main paradigms of sparsity optimization, i.e., the analysis and synthesis formulations, are derived and the two approaches are then compared in a number of experiments.

In the final part, the classification of pansharpened remote sensing imagery is addressed using two kinds of unsupervised classifiers. Images produced using several pansharpening methods are classified and the results compared and analyzed with the complementary nature of the spectral and spatial quality of the pansharpened images in mind. Furthermore, it is investigated how the use of techniques based on mathematical morphology can be used to increase the classification accuracy.

Contents

Contents	iii
List of Figures	vi
Abbreviations	xi
Notation	xiii
I Pansharpening Using Total Variation Regularization	1
1 Introduction	3
2 Overview of Pansharpening Paradigms	7
2.1 MRA methods	8
2.2 CS methods	8
2.2.1 PCA method	8
2.2.2 IHS fusion	9
3 Image Formation Model	11

4	Majorization-Minimization	15
5	Total Variation Regularization	17
6	Quantitative Quality Metrics	21
6.1	Metrics that need reference	22
6.1.1	Root Mean Square Error (RMSE)	22
6.1.2	Relative average spectral error (RASE)	22
6.1.3	ERGAS	22
6.1.4	Universal Image Quality Index : UIQI	23
6.1.5	Q4	23
6.1.6	Spectral Angle Mapper (SAM)	24
6.1.7	Spectral Information Divergence (SID)	24
6.1.8	Correlation coefficient (CC)	25
6.1.9	Spatial	25
6.2	Metrics without reference	25
6.2.1	QNR spectral metric	25
6.2.2	QNR spatial metric	26
7	Experimental Results	27
7.1	Datasets	27
7.2	Comparison to other methods	28
7.3	Choice of parameters	28
7.4	Quantitative Quality Assessment Results	29
7.5	Visual Comparison	31

CONTENTS	v
8 Conclusions	35
II Pansharpening using Analysis and Synthesis based Regularization of Overcomplete Transforms	37
9 Introduction	39
10 Sparse Representations and Overcomplete Transforms	43
11 Solutions of the Analysis and Synthesis Problems	47
11.1 Analysis Prior Formulation	47
11.2 Synthesis Prior Formulation	49
12 Experiments & Results	53
13 Conclusions	63
III Classification of Pansharpened Satellite Images	65
14 Introduction	67
15 Image Fusion	71
15.1 IHS fusion	71
15.2 Brovey Fusion	72
15.3 Model Based Fusion	72
15.4 MRA Method	74
15.5 PCA Method	74

16 Mathematical Morphology	77
17 Classification Methods	81
17.1 Support Vector Machine	81
17.2 Random Forest	82
18 Datasets and Experimental Results	85
18.1 First Dataset	87
18.1.1 Three Band Data (RGB)	89
18.1.2 Four Band Data (RGB and NIR)	90
18.2 Second Dataset	94
18.2.1 Three Band Data (RGB)	96
18.2.2 Four Band Data (RGB and NIR)	97
18.3 Third Dataset	100
18.3.1 Three Band Data (RGB)	101
18.3.2 Four Band Data (RGB and NIR)	102
19 Conclusions	107
IV Papers	117

List of Figures

2.1	Basic wavelet based pansharpening.	8
2.2	PCA based pansharpening.	9
3.1	Relative Spectral Response of the IKONOS sensor for the panchromatic, R,G,B and N bands.	12
4.1	Majorization-Minimization	16
7.1	Pansharpening results for the Pléiades data. Note that only a subset of the pansharpened images is shown.	31
7.2	Pansharpening results for the full scale Quickbird data. Note that only a subset of the pansharpened images is shown. The reference image shown is the upscaled MS image.	31
7.3	This image shows the Pleiades dataset and the pansharpened image produced using the proposed method. The images have been adjusted for better contrast and gamma corrected.	32
7.4	This image shows the Quickbird dataset and the pansharpened image produced using the proposed method. The images have been adjusted for better contrast and gamma corrected.	33
11.1	Soft thresholding	50

12.1 Results for Pleiades dataset with the NSCT for both analysis and synthesis for the ERGAS, SAM and Qave metrics	54
12.2 Results for Pleiades dataset with the UDWT for both analysis and synthesis for the ERGAS, SAM and Qave metrics	55
12.3 Results for Pleiades dataset with the FDCT for both analysis and synthesis for the ERGAS, SAM and Qave metrics	56
12.4 Results for Quickbird dataset with the NSCT for both analysis and synthesis for the ERGAS, SAM and Qave metrics	58
12.5 Results for Quickbird dataset with the UDWT for both analysis and synthesis for the ERGAS, SAM and Qave metrics	59
12.6 Results for Quickbird dataset with the FDCT for both analysis and synthesis for the ERGAS, SAM and Qave metrics	60
15.1 Relative Spectral Response of the IKONOS sensor for the panchromatic, R,G,B and N bands.	73
15.2 Basic wavelet based pansharpening.	74
15.3 PCA based pansharpening.	75
16.1 Erosion and dilation of the Lena image using a disk of radius 10 as the SE.	78
16.2 Opening and Closing of the Lena image using a disk of radius 10 as the SE.	78
16.3 Opening and closing by reconstruction of the Lena image using a disk of radius 10 as the SE.	79
16.4 MPs based on successive Opening by reconstruction and closing by reconstruction operations using a disk shaped SE of increasing radius.	79
18.1 Training map (a) and reference map (b) for dataset 1.	90
18.2 Training map (a) and reference map (b) for dataset 2.	94

-
- 18.3 The panchromatic images and the classification maps obtained using all 4 bands for each dataset. The classification images were obtained using the configuration that yielded the highest accuracy. (a) Original IKONOS panchromatic image for dataset 1. (b) RF classification map for dataset 1 (OA: 68.73%). (c) Original IKONOS panchromatic image for dataset 2. (d) RF classification map for dataset 2 (OA: 89.46%). (e) Original QuickBird panchromatic image for dataset 3. (f) RF classification map for dataset 3 (OA: 94.79%). 103

x

Abbreviations

DMP	Derivative of the Morphological Profile.
DWT	Discrete Wavelet Transform.
FDCT	Fast Discrete Curvelet Transform
MM	Majorization-Minimization (in Part I).
MM	Mathematical Morphology (In Part III).
MS	Multispectral image.
MP	Morphological Profile.
NSCT	Non-Subsampled Contourlet Transform.
PAN	Panchromatic image.
TV	Total Variation.
UDWT	Undecimated Discrete Wavelet Transform.

Notation

A	Matrices are written using upper case bold font.
a	Vectors are written using lower case bold font.
a	Scalars are denoted lower case.
\mathbf{A}^T	The transpose of a matrix.
\mathbf{a}^T	The transpose of a vector.
$\ \cdot\ _1$	The ℓ_1 -norm.
$\ \cdot\ _2$	The ℓ_2 -norm.
I	The identity matrix.
\mathbf{x}_k	The k-th iterate of \mathbf{x} .

Part I

Pansharpening Using Total Variation Regularization

Introduction

Satellites provide very valuable data about the Earth, e.g., for environmental monitoring, weather forecasting, map-making and military intelligence. But satellites are expensive, both to build and operate. Therefore we are obliged to make the best use of the data obtained from available satellites, e.g., by combining the output from different sensors.

A good example of this is the fusion of multispectral satellite images of low spatial and high spectral resolution with panchromatic images of high spatial and low spectral resolution.

A multispectral sensor captures light of several distinct wavelengths or bands for each spatial element or pixel. Hence, a multispectral image can be considered to consist of several layers where each layer contains the recorded intensity of light of a specified wavelength, or rather, a tight band of wavelengths. On the other hand, a panchromatic sensor captures the total intensity of light from a broad continuous range of wavelengths.

Due to cost and complexity issues, the multispectral sensor has much smaller aperture than the panchromatic sensor thus reducing the spatial resolution of the sensed multispectral image [1]. For a typical modern multispectral satellite sensor, this ratio is 1 to 16, i.e., a single multispectral image pixel translates to 4 by 4 panchromatic pixels.

Since the multispectral image contains information about light of specific and narrow bands of wavelengths, it has a high spectral resolution while the panchromatic image has low spectral resolution since it represents light of a broad spectrum of wavelengths. On the other hand, the multispectral image has low spatial resolution while the panchromatic image has a high spatial resolution.

Pansharpening is the fusion of the images captured by the multispectral and panchromatic sensors. The output is an image that has the high spectral resolution of the multispectral image and also the high spatial resolution of the panchromatic image. One can say that the spatial resolution of the multispectral image has been increased using the information contained in the panchromatic image. This means that the pansharpened image has the same number of pixels as the panchromatic image and also the same number of bands as the multispectral image, hence pansharpening can be regarded as an image or sensor fusion process.

Pansharpening is a typical under-determined inverse problem. If e.g., the multispectral image is $256 \times 256 \times 4$ pixels then the corresponding panchromatic image is 1024×1024 pixels. These two images are the observed data. The pansharpened multispectral image that we would like to estimate has $1024 \times 1024 \times 4$ pixels. The number of pixels that have to be estimated is 3.2 times the number of pixels that make up the observed data.

In part I we begin by developing a model that describes the pansharpening process using a classical observational model. The key element of this model is the model matrix \mathbf{M} which will end up having 3.2 times more columns than rows so the underlying system of equations that needs to be solved is under-determined.

To solve this ill-posed problem additional constraints are needed, i.e., the solution needs regularization. We investigate the use of total variation (TV) [2] regularization to solve this inverse problem. The TV norm has been shown to be very effective for denoising signals, especially those that are piece-wise smooth. Most natural images are piece-wise smooth so this choice of regularization is well justified.

In order to evaluate the quantitative quality of pansharpened imagery, various quality metrics have been proposed. There are two aspects to the quality of a pansharpened image. One is the spectral quality, sometimes referred to as spectral consistency and the other is the spatial quality, i.e. how much spatial detail has been transferred from the PAN image. This is sometimes measured as the correlation of the panchromatic image with the bands of the pansharpened image. In this study, we evaluate the proposed methods using a number of spectral and spatial quality metrics and also compare our results with several other pansharpening methods that represent the main paradigms in the field.

In the second part of this thesis, the pansharpening inverse problem is tackled in a different way. Instead of using the TV norm, we will investigate sparsity regularization techniques. Sparse representations of signals and systems, and methods that exploit sparsity are currently a very active research topic in modern signal-processing and related fields.

There are two dominant paradigms that are studied in this work. These are the so called analysis and synthesis approaches which are two different formulations for sparsity regularization of the coefficients of overcomplete transforms. This involves solving an optimization problem where the non-sparsity of the coefficients of the transformed solution is penalized.

The analysis formulation sparsifies the forward coefficients as the signal is analyzed by the forward transform operator, while the synthesis formulation is based upon sparsifying the linear combination of columns of the reconstruction operator, as the signal is reconstructed from the coefficients.

These two methods are compared using three different multi-scale overcomplete transforms to see if there is potential for sparsity regularization in pansharpening based on the previously defined model.

In the final part of the thesis, the classification of pansharpened imagery is studied. Unsupervised classification is a very important topic in remote sensing but very little has been published in the literature regarding the benefits to classification when using pansharpened images instead of the low-resolution MS image.

The classification of pansharpened images produced using several different methods will be studied extensively in a number of experiments, with particular attention to the effects of spectral and spatial quality on the classification accuracy. Furthermore, the extraction of additional spatial information using techniques derived from mathematical morphology that can be used to increase the classification accuracy, is investigated. These techniques are based on the so called Morphological Profile and its derivative.

Overview of Pansharpening Paradigms

Over the years, a number of different pansharpening methods have been proposed and many of these methods are either based on Multiresolution Analysis (MRA) or Component Substitution (CS). There are also hybrid methods based on both MRA and CS. Then there are methods based on variational techniques, such as the P+XS method. The MRA methods are usually based on methods such as the Undecimated Discrete Wavelet Transform (UDWT) or other kinds of pyramidal or multi-scale representations. Typically, the MS image is up-scaled and a multiresolution representation is calculated for this image along with the PAN image. The idea is to use some kind of injection model to replace or enhance the detail of the MS image with details from the PAN image.

The CS methods typically make use of transformations such as Principal Component Analysis (PCA) or a spectral transformation such as Intensity Hue Saturation (IHS) transformation, where a component derived from the MS image is substituted for a component derived from the PAN image and then the fused or pansharpened image is obtained from the inverse transformation. There are also methods that do not fit either of these categories such as the method proposed in this letter.

The remainder of this chapter is devoted to give a short overview of those different methodologies.

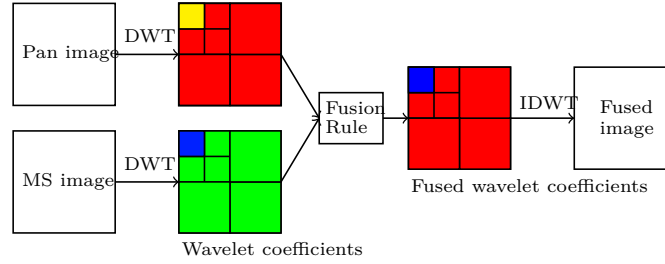


Figure 2.1: Basic wavelet based pansharpening.

2.1 MRA methods

The multi-resolution analysis approach to pansharpening is widely used and there are numerous techniques today based on it. The basic idea is to take the discrete wavelet transform (DWT) [3] of both the MS and Pan images. The next step is to retain the approximation coefficients for the MS image but replace the detail coefficients with those from the Pan image. Instead of just replacing the coefficients, some fusion rule can be used. This is shown in figure 2.1.

The main drawback to this approach is that there will be substantial artifacts in the fused image due to the nature of the DWT, i.e., it is not shift-invariant and lacks directionality, thus the spatial quality of the final fused image is reduced. One approach to overcome this problems is to use the undecimated DWT [4] or other overcomplete transforms such as the curvelet transform or the non-subsampled contourlet transform.

2.2 CS methods

2.2.1 PCA method

A good example of a this kind of method is the PCA method of pansharpening [5]. First, the mutually correlated bands of the upscaled MS image (i.e., the MS image is upsampled so it has same size as PAN image) are transformed using PCA into a set of uncorrelated components whose number is the same as number of bands in the MS image. The first principal component has the highest variance and is similar to the Pan image itself. The next step is to replace this component with the actual Pan image and finally take the inverse transform to get the fused image. Figure 2.2 illustrates this.

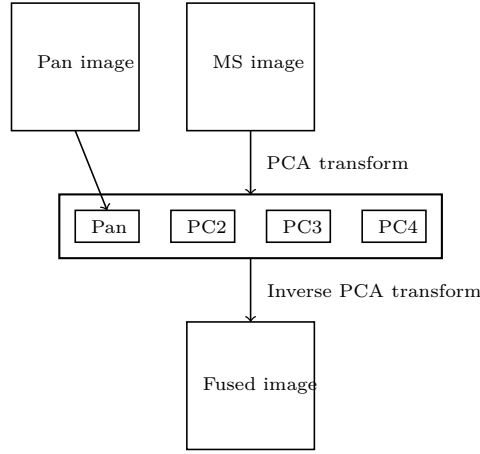


Figure 2.2: PCA based pansharpening.

2.2.2 IHS fusion

The IHS method [6–9] is a frequently used method, especially when working with LANDSAT and SPOT imagery. The basic idea is to first transform the MS image into intensity (I), hue (H) and saturation (S) components (IHS colorspace)

$$\begin{bmatrix} I \\ H \\ S \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{-\sqrt{2}}{6} & \frac{-\sqrt{2}}{6} & \frac{2\sqrt{2}}{6} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

The next step is to scale the Pan image so that it has the same mean and variance as the intensity component of the MS image

$$P = \frac{\sigma_I}{\sigma_P}(P - \mu(P)) + \mu(I).$$

The intensity component is then replaced with the appropriately scaled Pan image and finally the inverse IHS transformation is taken to get the fused image

$$\begin{bmatrix} F(R) \\ F(G) \\ F(G) \end{bmatrix} = \begin{bmatrix} R + P - I \\ G + P - I \\ B + P - I \end{bmatrix}.$$

The IHS method produces images that typically have high spatial fidelity but suffer from spectral distortion.

Image Formation Model

In this chapter an image formation model for pansharpening will be developed. This model can also be called an observational model since it relates the pansharpened image to the observed data. This model is not new and has been used in several methods [10] [11].

The model is derived from two assumptions about the true pansharpened image. The first assumption is that downgrading (decimation) the pansharpened image gives the observed low resolution MS image. This is based on the spectral consistency property proposed by Lucien Wald [12], which states that the degraded pansharpened image should give an image similar to the original MS image. However, the exact type of filter used in the degradation is not specified.

The exact choice of this filter is important and it will affect the quality of the fused image, both the spectral and spatial quality, as will be shown later. However, to keep the derivation of the model simple, a 4 tap moving average filter will be used in this chapter, while the actual implementation of algorithm based on this model will use a more advanced filter.

The other assumption is that a linear combination of the bands of the pansharpened image gives the observed PAN image. This is based on the fact that for a typical 4 band MS sensor, the wavelengths of the R, G, B and NIR bands are within the range of wavelengths of the PAN sensor. The relative spectral response of the IKONOS sensor is shown in Figure 3.1. In the figure it is seen that the PAN band overlaps the R, G, B and NIR bands.

The main criticism of this assumption is that it is an over-simplification and does not hold if there are bands in the MS image that do not overlap with

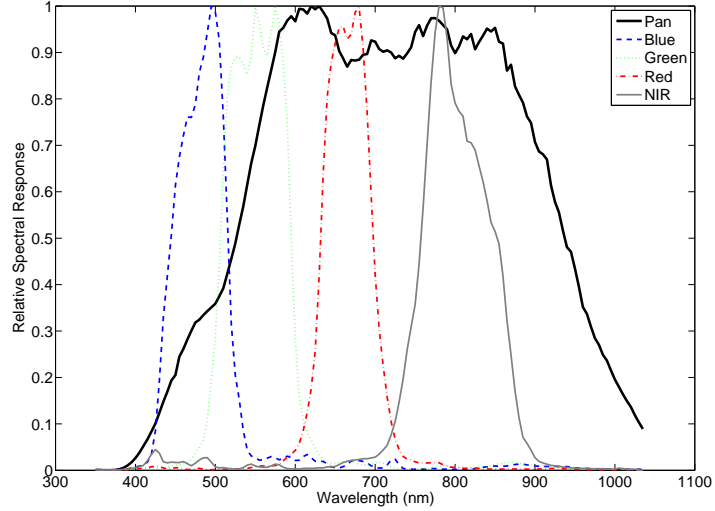


Figure 3.1: Relative Spectral Response of the IKONOS sensor for the panchromatic, R, G, B and N bands.

the PAN band, as is the case with new sensors which have more than 4 MS bands. Also, that the constraint this imposes on the pansharpened image, is too limiting for the eventual solution. However, when there is overlap for all MS bands, this assumption can be justified as a reasonable approximation, and by carefully selecting the coefficients of the linear combination, very good results can be achieved.

Thus, the model is based on the following assumptions

- Decimation of the fused image gives the observed MS image.
- Linear combination of the bands of the fused image approximates the PAN image.

The dataset consists of a high spatial resolution PAN image, denoted by \mathbf{y}_{PAN} and the low spatial resolution MS image, denoted by \mathbf{y}_{MS} . The PAN image has dimensions four times larger than the MS image so the ratio in pixels is 1 to 16. The MS image contains 4 bands, RGB and near-infrared (NIR).

The PAN image is of dimension $M \times N$ and the MS image is of dimension $m \times n$ where $m = M/4$ and $n = N/4$. The vectors \mathbf{y}_{PAN} and \mathbf{y}_{MS} are the vectorized PAN and MS images, respectively and \mathbf{y}_{MS} is arranged as $\mathbf{y}_{\text{MS}} =$

$[1, \dots, mn, 1, \dots, mn, 1, \dots, mn, 1, \dots, mn]^T$ where mn is the number of pixels in each band. The vectorized pansharpened image \mathbf{x} has this same arrangement where there are MN pixels in each band.

The assumption that the low spatial resolution MS image can be described as a degradation (decimation) of the pansharpened image \mathbf{x} is described in matrix notation as $\mathbf{y}_{\text{MS}} = \mathbf{M}_1 \mathbf{x} + \boldsymbol{\epsilon}$ where

$$\mathbf{M}_1 = \frac{1}{16} \mathbf{I}_4 \otimes \left((\mathbf{I}_n \otimes \mathbf{1}_{4 \times 1}^T) \otimes (\mathbf{I}_m \otimes \mathbf{1}_{4 \times 1}^T) \right) \quad (3.1)$$

is a decimation matrix of size $4mn \times 4MN$, \mathbf{I}_4 is an identity matrix of size 4 by 4, \otimes is the Kronecker product and $\boldsymbol{\epsilon}$ is zero mean Gaussian noise.

The second assumption, that the PAN image is a linear combination of the bands of the pansharpened image with some additive Gaussian noise is written in matrix notation as $\mathbf{y}_{\text{PAN}} = \mathbf{M}_2 \mathbf{x} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon}$ is zero mean Gaussian noise and

$$\mathbf{M}_2 = [\omega_1 \mathbf{I}_{MN}, \omega_2 \mathbf{I}_{MN}, \omega_3 \mathbf{I}_{MN}, \omega_4 \mathbf{I}_{MN}], \quad (3.2)$$

where $\omega_1, \dots, \omega_4$ are constants that sum to one. These constants determine the weight of each band in the PAN image.

Now \mathbf{M}_1 and \mathbf{M}_2 have the same number of columns and thus we can combine the expressions for \mathbf{y}_{MS} and \mathbf{y}_{PAN} into a single equation which is the classical observational model,

$$\mathbf{y} = \mathbf{M} \mathbf{x} + \boldsymbol{\epsilon}, \quad (3.3)$$

where $\mathbf{y} = [\mathbf{y}_{\text{MS}}^T, \mathbf{y}_{\text{PAN}}^T]^T$ and $\mathbf{M} = [\mathbf{M}_1^T, \mathbf{M}_2^T]^T$.

Majorization-Minimization

Majorization-Minimization (MM) is an important technique in optimization theory. The idea behind MM can be traced back to the work of J. M. Ortega and W. C. Rheinboldt in 1970 [13] while it appeared more fully developed under the name MM in the paper [14] by Hunter et al. MM can be used to turn a difficult optimization problem into a sequence of easier ones, by the use of surrogate functions. A surrogate function is a function that majorizes (minorizes) the objective function being minimized (maximized). In the following discussion it is assumed that we are minimizing the objective function and hence the surrogate function will be called a majorizer.

If we denote the objective function to be minimized by $J(\mathbf{x})$ and suppose that this function is not necessarily convex nor differentiable, it is obvious that the minimization of $J(\mathbf{x})$ is a difficult problem using conventional convex optimization techniques.

The central idea of MM is that if we have a guess for the minimum of $J(\mathbf{x})$, assuming that J is vector valued, denoted by the vector \mathbf{x}_k , we would like to find a new \mathbf{x}_{k+1} that further minimizes $J(\mathbf{x})$ therefore getting closer to the actual minimum of $J(\mathbf{x})$, i.e., we want to find a \mathbf{x}_{k+1} such that $J(\mathbf{x}_{k+1}) < J(\mathbf{x}_k)$.

But how do we find the new vector \mathbf{x}_{k+1} ? This is done by finding a function that majorizes $J(\mathbf{x})$, denote this function by $Q(\mathbf{x})$. For $Q(\mathbf{x})$ to be a majorizer of $J(\mathbf{x})$ it has to be chosen such that $Q(\mathbf{x}) \geq J(\mathbf{x})$ for all \mathbf{x} and such that $Q(\mathbf{x}_k) = J(\mathbf{x}_k)$. The next step is to find \mathbf{x}_{k+1} by minimizing $Q(\mathbf{x})$.

For this method to be effective, the function $Q(\mathbf{x})$ should be a function that can be easily minimized. This is repeated until the minimum of $J(\mathbf{x})$ is found.

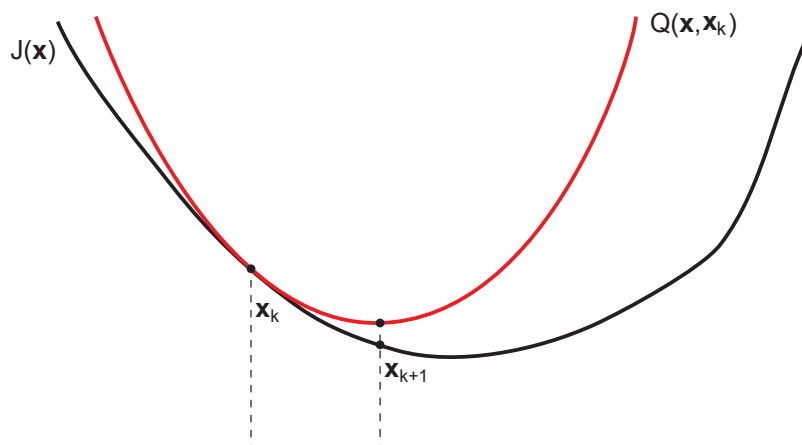


Figure 4.1: Majorization-Minimization

Since the majorizer function $Q(\mathbf{x})$ will be different for each iteration, it will be denoted by $Q_k(\mathbf{x})$.

The steps required to find the minimum of $J(\mathbf{x})$ can be summarized as

1. Set $k = 0$. Initialize \mathbf{x}_k .
2. Choose $Q_k(\mathbf{x})$ such that
 - (a) $Q_k(\mathbf{x}) \geq J(\mathbf{x}) \forall \mathbf{x}$
 - (b) $Q_k(\mathbf{x}_k) = J(\mathbf{x}_k)$
3. Set \mathbf{x}_{k+1} as minimizer of $Q_k(\mathbf{x})$.
4. Set $k = k + 1$ and go to step 2.

An iteration of the MM algorithm is shown in Figure 4.1.

Total Variation Regularization

Total Variation (TV) denoising is an important technique in the literature and the use of TV for regularization is currently an actively researched topic. TV regularization was first proposed by Rudin et al. in [15]. It encourages noise removal while preserving edges. By penalizing solutions with high TV, the estimate is forced to be smooth without necessarily penalizing sharp boundaries since the estimated boundary or edge is decided by the observed data. The TV for 1D discrete signal \mathbf{x} is defined

$$\text{TV}(\mathbf{x}) = \sum_i |x_{i+1} - x_i|. \quad (5.1)$$

From this it is seen that the TV is the sum of all variations of the signal, hence the name. For a 2D signal \mathbf{x} the TV norm is defined by

$$\text{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h \mathbf{x})^2 + (\Delta_i^v \mathbf{x})^2}, \quad (5.2)$$

where $\Delta_i^h \mathbf{x}$ and $\Delta_i^v \mathbf{x}$ are the horizontal and vertical first order difference at pixel i , respectively. The TV norm is clearly not differentiable. Using matrix notation one can write

$$\text{TV}(\mathbf{x}) = \left\| \sqrt{(\mathbf{D}_h \mathbf{x})^2 + (\mathbf{D}_v \mathbf{x})^2} \right\|_1, \quad (5.3)$$

where the matrices \mathbf{D}_h and \mathbf{D}_v are defined such that when multiplied by a vectorized image they give the first order differences in the horizontal direction and vertical direction, respectively.

Using the above, the TV of the 4-band MS image can be written

$$\text{TV}(\mathbf{x}) = \left\| \sqrt{(\mathbf{D}_H \mathbf{x})^2 + (\mathbf{D}_V \mathbf{x})^2} \right\|_1, \quad (5.4)$$

where \mathbf{x} is the vectorized 4 band MS image, $D_H = (\mathbf{I}_4 \otimes \mathbf{D}_h)$, $D_V = (\mathbf{I}_4 \otimes \mathbf{D}_v)$. The cost function of the TV regularized problem can now be formulated as

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda \text{TV}(\mathbf{x}). \quad (5.5)$$

Minimizing this cost function is difficult because the TV functional is not differentiable. However, MM techniques can be used to replace this difficult problem with a sequence of easier ones

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} Q(\mathbf{x}, \mathbf{x}_k), \quad (5.6)$$

where \mathbf{x}_k is the current iterate and $Q(\mathbf{x}, \mathbf{x}_k)$ is a function that majorizes the cost function $J(\mathbf{x})$. This means that $Q(\mathbf{x}, \mathbf{x}_k) \geq J(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{x}_k$ and $Q(\mathbf{x}, \mathbf{x}_k) = J(\mathbf{x})$ for $\mathbf{x} = \mathbf{x}_k$. By iteratively solving (5.6), \mathbf{x}_k will converge to the global minimum of $J(\mathbf{x})$. (The cost function (5.5) was minimized in [16] using a different algorithm.)

A majorizer for the TV term [17] can be written using matrix notation as

$$Q_{\text{TV}}(\mathbf{x}, \mathbf{x}_k) = \mathbf{x}^T \mathbf{D}^T \mathbf{W}_k \mathbf{D} \mathbf{x} + c \quad (5.7)$$

where we define

$$\boldsymbol{\omega}_k = \left(2\sqrt{(\mathbf{D}_H \mathbf{x}_k)^2 + (\mathbf{D}_V \mathbf{x}_k)^2} \right)^{-1} \quad (5.8)$$

and $\mathbf{W}_k = \text{diag}(\boldsymbol{\omega}_k, \boldsymbol{\omega}_k)$ and the matrix \mathbf{D} is defined as $\mathbf{D} = [\mathbf{D}_H^T \ \mathbf{D}_V^T]^T$. When the data fidelity term, $\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2$, is expanded, it is evident that we end up with the term $\mathbf{M}^T \mathbf{M}$ in the solution. In order to avoid having to find the inverse of this term later in the solution, it is necessary to introduce the function

$$Q_{\text{DF}}(\mathbf{x}, \mathbf{x}_k) = (\mathbf{x} - \mathbf{x}_k)^T (\alpha \mathbf{I} - \mathbf{M}^T \mathbf{M}) (\mathbf{x} - \mathbf{x}_k), \quad (5.9)$$

where $\alpha > \max \text{eig}(\mathbf{M}^T \mathbf{M})$. This ensures that $Q_{\text{DF}}(\mathbf{x}, \mathbf{x}_k)$ is positive definite and $Q(\mathbf{x}, \mathbf{x}_k)$ is still a majorizer for $J(\mathbf{x})$. The function to minimize becomes

$$Q(\mathbf{x}, \mathbf{x}_k) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + Q_{\text{DF}}(\mathbf{x}, \mathbf{x}_k) + \lambda Q_{\text{TV}}(\mathbf{x}, \mathbf{x}_k). \quad (5.10)$$

Differentiating this function w.r.t. \mathbf{x} , setting to zero and subsequently solving for \mathbf{x} gives

$$(\alpha \mathbf{I} + \lambda \mathbf{D}^T \mathbf{W}_k \mathbf{D}) \mathbf{x} = \mathbf{b}, \quad (5.11)$$

where

$$\mathbf{b} = \alpha \mathbf{x}_k + \mathbf{M}^T (\mathbf{y} - \mathbf{M} \mathbf{x}_k). \quad (5.12)$$

The matrix inversion lemma gives

$$(\alpha \mathbf{I} + \lambda \mathbf{D}^T \mathbf{W}_k \mathbf{D})^{-1} = \frac{1}{\alpha} (\mathbf{I} - \mathbf{D}^T (\frac{\alpha}{\lambda} \mathbf{W}_k^{-1} + \mathbf{D} \mathbf{D}^T)^{-1} \mathbf{D}). \quad (5.13)$$

This gives $\mathbf{x} = \frac{1}{\alpha}(\mathbf{b} - \mathbf{D}^T \mathbf{z})$ where we have defined

$$\mathbf{z} = (\frac{\alpha}{\lambda} \mathbf{W}_k^{-1} + \mathbf{D}\mathbf{D}^T)^{-1} \mathbf{D}\mathbf{b}. \quad (5.14)$$

From (11.13) we have $(\frac{\alpha}{\lambda} \mathbf{W}_k^{-1} + \mathbf{D}\mathbf{D}^T) \mathbf{z} = \mathbf{D}\mathbf{b}$ and by adding $c\mathbf{z}$, where the constant $c > \max \text{eig}(\mathbf{D}\mathbf{D}^T)$, and by subtracting $\mathbf{D}\mathbf{D}^T \mathbf{z}$ from both sides of this equation we obtain

$$(\frac{\alpha}{\lambda} \mathbf{W}_k^{-1} + c\mathbf{I}) \mathbf{z} = \mathbf{D}\mathbf{b} + (c\mathbf{I} - \mathbf{D}\mathbf{D}^T) \mathbf{z}, \quad (5.15)$$

which leads to the update rule

$$\mathbf{z}_{k+1} = (\frac{\alpha}{\lambda} \mathbf{W}_k^{-1} + c\mathbf{I})^{-1} (\mathbf{D}\mathbf{b} + (c\mathbf{I} - \mathbf{D}\mathbf{D}^T) \mathbf{z}_k) \quad (5.16)$$

$$\mathbf{x}_{k+1} = \frac{1}{\alpha} (\mathbf{b} - \mathbf{D}^T \mathbf{z}_{k+1}). \quad (5.17)$$

We can see that finding the inverse in (5.16) is trivial since it is a diagonal matrix.

The implementation of the above algorithm becomes straightforward once it is realized that all the matrix multiplications involving the operators \mathbf{D} , \mathbf{D}^T , \mathbf{M} and \mathbf{M}^T can be implemented as simple operations on multispectral images. This significantly reduces computation time and memory requirements due to the enormous size of the matrices involved.

However, there is one important thing to note. Bearing in mind the structure of the model matrix \mathbf{M} , it is clear that the multiplication with \mathbf{M}^T in (5.12) indicates nearest neighbor interpolation (the inverse of decimation) of an MS image. By using bilinear decimation and interpolation instead gives better results, both according to quality metrics and visual inspection.

Quantitative Quality Metrics

For the evaluation of the quality of the pansharpened image, visual inspection alone is obviously not sufficient. Many different metrics have been proposed to evaluate or measure the quantitative quality of pansharpened images. There are metrics to evaluate both spectral and spatial quality. This is further complicated by the fact that the spectral and spatial quality tend to be complementary. Images of very high spatial fidelity are often lacking in spectral quality and vice versa. This complementarity is probably exaggerated by the design of the metrics themselves.

However, accurate quantitative quality assessment is difficult since a reference high resolution MS image is not available. There exist datasets where a simulated reference image is available but this is very rare.

There are two approaches to this problem. One is to use metrics that do not need a reference image. Another approach is to make use of Wald's consistency property [12] which states that the pansharpened image degraded to the resolution of the MS image should be very similar to the original MS image. The spectral metrics without reference attempt to measure this spectral consistency.

There is a similar property related to the spatial consistency, i.e., the idea is that the inter correlation of the bands of the MS image and the degraded PAN image should be the same as the correlation between the bands of the pansharpened image and the PAN image.

This approach is often not very accurate since the choice of filters used for the decimation is important and not all researchers agree on the exact choice of these filters. However, recent papers in the literature propose to use the Modulation

Transfer Function (MTF) of the sensor itself to design these filters. The PAN and MS images are degraded in resolution such that the resulting pansharpened image has the same size as the original MS image and which can then be used as the reference.

In light of the above discussion, the types of metrics used can be assigned to two categories. Metrics that need reference and those without reference. And within these categories there are metrics that attempt to measure spectral and spatial quality.

The remainder of this chapter is a description of the different metrics used for quality assessment used in this thesis.

6.1 Metrics that need reference

6.1.1 Root Mean Square Error (RMSE)

The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_x \sum_i (X_i(x) - Y_i(x))^2}{n \times m \times d}},$$

where X is the MS image, Y is the pansharpened image, x is the pixel and i is the band number. Finally, n is the number of rows, m is the number of columns and d is the number of bands.

6.1.2 Relative average spectral error (RASE)

RASE [18] computes the average performance in terms of the RMSE of the bands in the pansharpened image.

$$\text{RASE} = \frac{100}{M} \sqrt{\frac{1}{N} \sum_{i=1}^N \text{RMSE}^2(B_i)}.$$

6.1.3 ERGAS

ERGAS [19] is an acronym in French for "Erreur relative globale adimensionnelle de synthese" which translates to "relative dimensionless global error in

synthesis". This metric calculates the amount of spectral distortion in the fused image and is given by

$$\text{ERGAS} = 100 \frac{h}{l} \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{\text{RMSE}(n)}{\mu(n)} \right)^2},$$

where N is the number of bands, RMSE is the root mean square error, $\frac{h}{l}$ is the ratio of pixels in the Pan image to the MS image and $\mu(n)$ is the mean of the n -th band.

6.1.4 Universal Image Quality Index : UIQI

Qave [20] is a metric that attempts to model the spectral distortion as a combination of three factors. These factors are loss of correlation, luminance distortion and contrast distortion. The metric is given by

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]},$$

where $x = \{x_i | i = 1, 2, \dots, N\}$ and $y = \{y_i | i = 1, 2, \dots, N\}$ are the original MS and fused image vectors, respectively, and

$$\begin{aligned} \bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i & \bar{y} &= \frac{1}{N} \sum_{i=1}^N y_i \\ \sigma_x^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 & \sigma_y^2 &= \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 \\ \sigma_{xy} &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \end{aligned}$$

This metric is used by the QNR spectral and spatial metrics described in the next section and also by the Qave metric which is defined as the mean UIQI between the bands of the pansharpened image and the reference image.

6.1.5 Q4

The Q4 metric is an extension of the UIQI metric to images with 4 bands. This is done by considering the 4 values for each pixel as a quaternion, hence this metric can be viewed as the hypercomplex extension of UIQI. Further details can be found in [21]. It is given by

$$Q4 = \frac{|\sigma_{\mathbf{z}_1 \mathbf{z}_2}| \cdot |\mathbf{z}_1| \cdot |\mathbf{z}_2|}{(\sigma_{\mathbf{z}_1}^2 + \sigma_{\mathbf{z}_2}^2)(|\mathbf{z}_1|^2 + |\mathbf{z}_2|^2)}, \quad (6.1)$$

where $\mathbf{z}_1 = a_1 + \mathbf{i}b_1 + \mathbf{j}c_1 + \mathbf{k}d_1$ and $\mathbf{z}_2 = a_2 + \mathbf{i}b_2 + \mathbf{j}c_2 + \mathbf{k}d_2$ are quaternions representing the 4-band reference MS image and the 4-band pansharpened image. The operator $|\cdot|$ is the hypercomplex modulus, $|\mathbf{z}| = \sqrt{a^2 + b^2 + c^2 + d^2}$ and $\sigma_{\mathbf{z}_1 \mathbf{z}_2}$ is the hypercomplex covariance between \mathbf{z}_1 and \mathbf{z}_2 .

Both the UIQI and Q4 metrics use a sliding window for the calculations. The size used in this work is 64 by 64 pixel window for high resolution images and 16 by 16 pixel window for low resolution images, as was done in [22].

6.1.6 Spectral Angle Mapper (SAM)

The Spectral Angle Mapper (SAM) [23] is a metric that calculates the spectral similarity between two spectral vectors as a spectral angle,

$$\cos \alpha = \frac{\sum_{i=1}^N x_i y_i}{\sqrt{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i^2}},$$

where N is the number of bands and $\mathbf{x} = (x_1, x_2, \dots, x_N)$ and $\mathbf{y} = (y_1, y_2, \dots, y_N)$ are two spectral vectors at some pixel location in the original MS image and the fused image, respectively. The value of SAM for the entire image is the average of all the α values for every pixel.

6.1.7 Spectral Information Divergence (SID)

SID [24] originates from information theory. Each pixel spectrum is viewed as a random variable and SID measures the difference or discrepancy of the probabilistic behaviors between two spectral vectors, taken from the MS image and final fused image, respectively. Let the vector $\mathbf{x} = (x_1, \dots, x_N)^T$ be taken from the MS image and the vector $\mathbf{y} = (y_1, \dots, y_N)^T$ be taken from the final pansharpened image. The range of these vectors needs to be normalized to the range $[0, 1]$ so we define

$$p_j = \frac{x_j}{\sum_{i=1}^N x_i} \quad q_j = \frac{y_j}{\sum_{i=1}^N y_i},$$

where N is the number of bands. SID is defined by

$$\text{SID}(x, y) = D(x||y) + D(y||x),$$

where $D(x||y)$ is called the relative entropy and is given by $D(x||y) = \sum_{i=1}^L p_i \log(\frac{p_i}{q_i})$. Similar for the term $D(y||x)$.

6.1.8 Correlation coefficient (CC)

The correlation coefficient (CC) between the original MS image (X) and the pansharpened image (Y) is defined as

$$CC(X, Y) = \frac{\sum_{mn} (X_{mn} - \bar{X})(Y_{mn} - \bar{Y})}{\sqrt{(\sum_{mn} (X_{mn} - \bar{X})^2)(\sum_{mn} (Y_{mn} - \bar{Y})^2)}},$$

where \bar{X} and \bar{Y} are the mean values of the corresponding images.

6.1.9 Spatial

The spatial metric used in Part III is based on computing the correlation coefficient between the high-frequency data of each MS band and the high frequency data of the Pan image. To extract the high-frequency data of a band, it is convoluted with the following mask

$$\text{mask} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

The metric value is the average of the CCs for each band of the MS image.

6.2 Metrics without reference

6.2.1 QNR spectral metric

The QNR D_λ spectral distortion metric is given by

$$D_\lambda = \frac{1}{N_B(N_B - 1)} \sum_{l=1}^{N_B} \sum_{r=1, r \neq l}^{N_B} \left| Q(F_l, F_r) - Q(\hat{F}_l, \hat{F}_r) \right|, \quad (6.2)$$

where N_B is the number of bands, F is the pansharpened image and \hat{F} is the MS image and Q is the Universal Image Quality Index (UIQI) [25].

6.2.2 QNR spatial metric

The QNR D_s spatial distortion is given by

$$D_s = \frac{1}{N_B} \sum_{l=1}^{N_B} \left| Q(F_l, P) - Q(\hat{F}_l, \hat{P}) \right|, \quad (6.3)$$

where P is the PAN image and \hat{P} is the degraded PAN image of same size as the MS image. UIQI is a metric that attempts to model the spectral distortion as a product of three factors, i.e., loss of correlation, luminance distortion and contrast distortion.

Experimental Results

7.1 Datasets

We present experimental results using simulated Pléiades data [22] and Quickbird data [22]. The Pléiades dataset consists of a 0.8m resolution PAN image of dimension 1024 by 1024 pixels and a 3.2m resolution 256 by 256 pixel MS image. There is also a simulated high-resolution MS image of same size and resolution as the PAN image. This makes it possible to use the metrics that use a reference for this dataset.

The other dataset is a high resolution Quickbird image of the Mississippi State University campus. The PAN image is 2048 by 2048 pixels and of resolution 0.7m and the MS image is 512 by 512 pixels and of resolution 2.8m. For this dataset there is no high resolution reference image available.

In order to be able to use the low spatial resolution MS image as a reference, we perform experiments using degraded MS and PAN images. This gives a pansharpened image of the same size as the original low spatial resolution MS image. The images are degraded using bilinear interpolation which is consistent with the use of bilinear decimation and interpolation in the implementation of the proposed method. We also present results using the Quickbird dataset at full scale using QNR metrics only.

7.2 Comparison to other methods

We compare the proposed method to three state-of-the-art pansharpening methods. The method called UDWT [26] uses the undecimated wavelet transform to decompose the PAN image and the bicubic expanded MS image into their coarse (LL) and detail parts (HH). The LL part of the PAN decomposition is replaced with the LL part of the MS image and finally the inverse transform yields the pansharpened image. This method produces images with excellent spatial quality. Because of its simplicity there are no tuning parameters other than number of levels of decomposition, which was set to 2 and the type of filter, where we use the Daubechies filter of length 2.

The Principal Component Analysis (PCA) method [5] is a method which belongs to the CS family of methods and finally the P+XS [11] method is based on a variational model. The PCA method implementation we used has no parameters to tune and the number of components is fixed at 4.

The P+XS method is more complex and there are several parameters to tune. However, most of them were left at their default values. The parameters that matter most are the weights for each multispectral band in the fused image and the regularization term λ . We used the same weights as for the proposed method and we chose $\lambda = 5$ which was shown to give optimal results based on sensitivity analysis.

7.3 Choice of parameters

There are several parameters that affect the performance of the proposed method. The constants α and c affect the convergence of the algorithm and in all the experiments they were chosen as 0.75 and 8, respectively. Sensitivity analysis of λ shows that for imagery at full scale, the performance of the algorithm is very good for small values of λ . Based on this we choose $\lambda = 2$ for imagery at full scale. When using degraded imagery, the ERGAS metric, which is based on the mean square error, has a global minimum at $\lambda = 1.5$, which is the value we choose.

The constants ω_i that determine the weights of the bands of the pansharpened image are important for spectral quality. Choosing the value of 0.25 for each ω_i gives good results, but there is room for improvement, since in reality the PAN image is not the average of the bands. These constants were selected for each dataset by doing a grid search over the parameter space, evaluating each possible combination of the parameters using the ERGAS metric. The values chosen for the Pléiades dataset are 0.30, 0.26, 0.22 and 0.22 for the red, green,

blue and NIR bands, respectively and for the Quickbird dataset as 0.21, 0.21, 0.21 and 0.37, respectively.

7.4 Quantitative Quality Assessment Results

The results of the quantitative quality evaluations of the different pansharpening algorithms are shown in Tables 7.1-7.3. The last row in these tables, called Expanded is where the expanded or upsampled (using bicubic interpolation) MS image is used as the fused image. This image is used as a baseline reference, i.e., this is the most basic method for pansharpening.

The results for the Pléiades data are shown in Table 7.1. Since a high resolution MS image is available for this dataset, the pansharpened image can be evaluated at full scale using all the metrics discussed in Chapter 6. The table shows that the proposed method gives best results in every metric using reference, while the P+XS and UDWT methods give slightly better results for QNR spectral and spatial metrics. The high values of the CC, Qave and Q4 metrics indicate good spatial detail as well as good spectral quality. The CPU time in seconds for each algorithm, in the same order as they appear in the table, is 1.8, 0.2, 83.4 and 15.4, respectively. The computer specification is i5-2400, 3.1 Ghz CPU with 12GB of memory. All methods are implemented in Matlab.

The results for the degraded Quickbird data are shown in Table 7.2. Note how similar values are for the QNR spectral and spatial metrics. This is very probably a result of the degradation process. The proposed method gives the best results for all metrics except the QNR spectral metric, where surprisingly, PCA gives slightly better result. The CPU time for the experiments in Table 7.2 is 0.3, 0.1, 20.3 and 3.2, respectively and in the same order as before.

Finally, the results using the Quickbird dataset at full scale are shown in Table 7.3. At this scale, no reference image is available and thus we only use the QNR metrics to assess the quality of the pansharpened image. The proposed method is shown to give better results for both the spectral and spatial QNR metrics than the other methods. The CPU time in seconds for the experiments in Table 7.3 is 6.1, 0.9, 335.2 and 57.7, respectively.

In Table 7.3 we see that the best results for QNR D_λ are for the expanded image. This seems to indicate that the QNR metrics are not very reliable.

Table 7.1: Comparison of pansharpening methods using the simulated Pléiades data at full scale. The best results are denoted using bold font.

	SAM	ERGAS	CC	Qave	Q4	D_λ	D_s
Reference	0	0	1	1	1	0.011	0.040
UDWT [26]	5.178	4.424	0.947	0.913	0.890	0.044	0.020
PCA [5]	6.964	7.681	0.853	0.678	0.670	0.073	0.201
P+XS [11]	5.108	4.965	0.951	0.912	0.894	0.017	0.024
Proposed	4.213	3.119	0.972	0.960	0.945	0.039	0.044
Expanded	4.905	6.345	0.869	0.792	0.799	0.002	0.122

Table 7.2: Comparison of pansharpening methods using the degraded Quickbird data. The best results are denoted using bold font.

	SAM	ERGAS	CC	Qave	Q4	D_λ	D_s
Reference	0	0	1	1	1	0.020	0.029
UDWT [26]	5.038	3.411	0.933	0.885	0.912	0.100	0.102
PCA [5]	4.821	4.835	0.930	0.784	0.818	0.071	0.097
PXS [11]	4.513	2.982	0.950	0.914	0.931	0.117	0.113
Proposed	3.520	2.448	0.965	0.942	0.960	0.073	0.086
Expanded	4.439	4.444	0.893	0.813	0.812	0.006	0.119

Table 7.3: Comparison of pansharpening methods using the Quickbird data at full scale. The best results are denoted using bold font. Only metrics that don't need reference are used.

	D_λ	D_s
UDWT [26]	0.048	0.055
PCA [5]	0.093	0.113
P+XS [11]	0.060	0.079
Proposed	0.027	0.042
Expanded	0.002	0.045



Figure 7.1: Pansharpening results for the Pléiades data. Note that only a subset of the pansharpened images is shown.

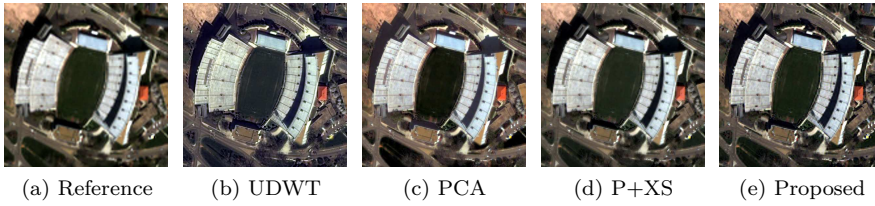


Figure 7.2: Pansharpening results for the full scale Quickbird data. Note that only a subset of the pansharpened images is shown. The reference image shown is the upscaled MS image.

7.5 Visual Comparison

A visual comparison of the results for the Pléiades data using all methods is shown in Fig. 7.1. Only a subset of the data is shown in order to display more detail. The images have been gamma corrected in order to make them more visually pleasing. Visual inspection shows that all the methods show good level of detail while the PCA results look very spectrally distorted. In Figure 7.3, the Pleiades dataset is shown along with the pansharpened image using the proposed method and in Figure 7.4, the Quickbird dataset is shown.

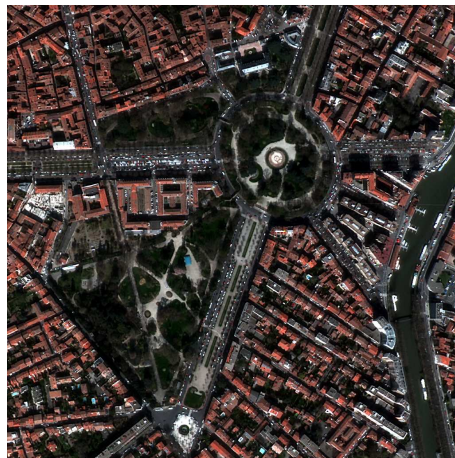
In Figure 7.2, a subset of the pansharpened images using the Quickbird data at full scale is shown for all methods. Again the PCA method shows considerable spectral distortion while the proposed and P+XS method give best results. The UDWT image shows some noticeable spectral distortion while the spatial resolution is excellent which is the primary strength of this method.



(a) Pleiades PAN image (1024×1024)



(b) Pleiades MS image (256×256)

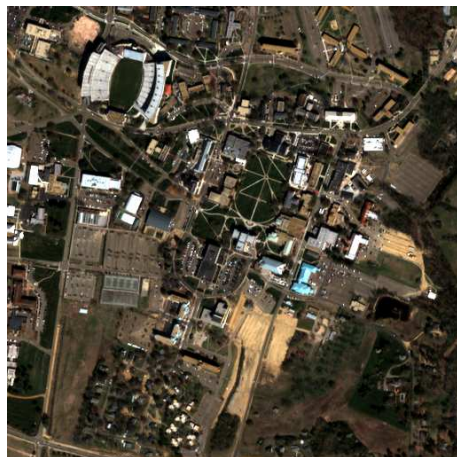


(c) Pansharpened image using the proposed method (1024×1024)

Figure 7.3: This image shows the Pleiades dataset and the pansharpened image produced using the proposed method. The images have been adjusted for better contrast and gamma corrected.



(a) Quickbird PAN image (2048×2048)



(b) Quickbird MS image (512×512)



(c) Pansharpened image using the proposed method (2048×2048)

Figure 7.4: This image shows the Quickbird dataset and the pansharpened image produced using the proposed method. The images have been adjusted for better contrast and gamma corrected.

Conclusions

In this work we have proposed a new pansharpening method based on an observational model which is regularized using TV. The fact that the model is applied to the whole dataset results in computations involving very large matrices. However, since these operations can be implemented as operations on images, the resulting algorithm is quite fast.

We performed experiments using datasets from two different sensors, Pléiades and Quickbird and compared our method to three other well known state-of-the-art methods, representing three important paradigms in pansharpening, and evaluated the results using a number of quantitative quality metrics.

The results show that the proposed method consistently outperformed the other methods used in virtually all experiments for all quality metrics used. Visual inspection shows that the method produces pansharpened images that have excellent spatial and spectral quality. The results obtained using the Pleiades dataset are among the best in the literature.

Part II

Pansharpening using Analysis and Synthesis based Regularization of Overcomplete Transforms

Introduction

Pansharpening is the process of synthesizing a high resolution multispectral (MS) image from a low spatial resolution multispectral image and a high spatial resolution panchromatic (PAN) image. As was shown in Part I of the thesis, the pansharpening process can be described using the classic observational model

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \boldsymbol{\epsilon}, \quad (9.1)$$

where \mathbf{y} is the observed data, \mathbf{M} is the model matrix previously defined in part I of the thesis, \mathbf{x} is the pansharpened image we are estimating and $\boldsymbol{\epsilon}$ is zero mean Gaussian noise. This is an ill-posed image restoration problem (inverse problem) where we want to restore an underlying image that has been degraded and transformed in some way as described by the linear operator \mathbf{M} . The typical way to solve this kind of problem is by minimizing cost function of the form

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda R(\mathbf{x}) \quad (9.2)$$

where the first term is a data fidelity term, λ is the regularization parameter and the function $R(\mathbf{x})$ is the regularizer or penalty function.

In the past decade, regularization methods based on the ℓ_1 -norm of the coefficients of the transform of the underlying image have gained much attention. The key idea behind these methods is that the image being estimated has a sparse approximation under the given transform, which can be a basis, a frame or some general overcomplete dictionary. Here, the focus will be on transforms which are overcomplete, so-called tight-frames [27].

Overcomplete systems have been shown to be superior to the orthonormal basis when it comes to sparsely approximating signals, since a signal is more likely to

have a sparse approximation under an overcomplete system. We will study three kinds of overcomplete systems. These are the Undecimated Discrete Wavelet Transform (UDWT) [28], the Non-Subsampled Contourlet Transform (NSCT) [29] and the Fast Discrete Curvelet Transform (FDCT) [30]. What is important here is that these transform are all tight-frame systems.

If we denote the system/transform using the linear operator \mathbf{W} , then for a tight-frame system we have $\mathbf{W}^T \mathbf{W} = \mathbf{I}$, where the operator \mathbf{W} is called the analysis (forward) operator since the signal under study \mathbf{x} is analyzed by $\boldsymbol{\omega} = \mathbf{W}\mathbf{x}$. The coefficients $\boldsymbol{\omega}$ are the projections of \mathbf{x} on the rows of \mathbf{W} .

The signal can be restored from the coefficients $\boldsymbol{\omega}$ by the synthesis operator \mathbf{W}^T , as a linear combination of some atom signals (the columns of \mathbf{W}^T), i.e., $\mathbf{x} = \mathbf{W}^T \boldsymbol{\omega}$. The exact nature of the system \mathbf{W} is not the topic here, suffice it to say, we are interested in these systems or transforms since they are known to give sparse representations of many types of natural signals.

A sparse representation is a representation that accounts for all or most of the information of a signal using a linear combination of a small number of elementary signals called atoms. These atoms come from a so called dictionary, here the dictionary is the columns of \mathbf{W}^T . Usually, this dictionary is overcomplete, which means that the dimension of the dictionary, i.e., the number of atoms, is considerably greater than the dimension of the signal space.

Today, one can identify two main paradigms in ℓ_1 -norm sparsity optimization. One is the use of an analysis prior and the other is the use of a synthesis prior.

The analysis prior formulation is given as

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1, \quad (9.3)$$

where \mathbf{W} is the transform (analysis) operator. In this setting the regularization term enforces a solution whose forward transform coefficients $\mathbf{W}\mathbf{x}$ are sparse. The prior information utilized about the coefficients is that they are sparse.

On the other hand, the synthesis prior formulation is given by

$$J(\boldsymbol{\omega}) = \|\mathbf{y} - \mathbf{M}\mathbf{W}^T \boldsymbol{\omega}\|_2^2 + \lambda \|\boldsymbol{\omega}\|_1, \quad (9.4)$$

where $\boldsymbol{\omega}$ are the transform coefficients. We see that the underlying image \mathbf{x} is being restored or synthesized by $\mathbf{x} = \mathbf{W}^T \boldsymbol{\omega}$ in the data fidelity term.

It is important to note that for orthonormal transforms like the critically sampled discrete wavelet transform (DWT), we have $\mathbf{W}^T \mathbf{W} = \mathbf{W}\mathbf{W}^T = \mathbf{I}$ and thus the analysis and synthesis formulations are the same.

Another key aspect of the tight-frame system \mathbf{W} is that it is a Multiresolution Analysis (MRA) system. In short, this means that \mathbf{W} decomposes the signal

under study into a space that is composed of a sequence of nested subspaces. We have a coarse representation (the low pass coefficients) and a hierarchy of detail or fine scale representations (the high-pass coefficients).

The high-pass wavelet coefficients are sparse while the low-pass coefficients of most natural signals are far from sparse. This means that in the implementation of these methods, the low-pass coefficients are not included in the penalty term. In all our experiments, the same value of the regularization parameter λ is used for all the high-pass subbands of the transform coefficients.

The focus of the work presented here is the comparison of the analysis and synthesis approaches for sparsity optimization for the solution of the inverse problem of pansharpening. These two approaches give different results for many types of inverse problems. For some problems, the analysis formulation gives better results while for other problems, the synthesis formulation is better. In the application of denoising, the analysis and synthesis methods are known to give significantly different results with the analysis method giving better results [31].

Experiments are performed using simulated Pleiades data [22] and Quickbird data [22]. The Pleiades dataset includes a high resolution reference image which is very useful when evaluating the quality of the pansharpened image using quality metrics. In order to evaluate the results for the Quickbird image, the data is degraded using bilinear interpolation such that the resulting pansharpened image is of the same size as the original MS image, which can then be used as the reference MS image. This method has the obvious drawback that information is lost during the degradation of the data, but it can still give a good indication of the performance of the method being evaluated.

The metrics used in the experiments are ERGAS [26], SAM [26] and Qavg [32], which all need a reference image. Qave is defined as the mean UIQI between the bands of the pansharpened image and the reference image. While there are metrics that do not need a reference image such as the QNR [33] spectral and spatial metrics, they usually do not give a good estimate of the mean square error (MSE).

The outline of the work is as follows. Chapter 10 gives an overview of sparse representations and overcomplete transforms. In Chapter 11 the algorithms based on the analysis and synthesis priors are derived, Chapter 12 describes the experiments and their results and finally, conclusions are drawn in Chapter 13.

Sparse Representations and Overcomplete Transforms

As shown previously we can state the process of pansharpening as an ill posed inverse problem. To keep the discussion simple we will neglect the presence of noise. This problem is stated as

$$\mathbf{y} = \mathbf{M}\mathbf{x}, \quad (10.1)$$

where \mathbf{y} is the observed data, the MS and PAN image in this case, \mathbf{M} is the linear operator that models the process and \mathbf{x} is the high resolution MS image we are seeking. As was shown in Part I of the thesis, the matrix \mathbf{M} has more columns than rows, i.e., we have $\mathbf{M} \in \mathbb{R}^{m \times n}$ where $n > m$. The exact ratio is $n = 3.2m$, to be specific. According to fundamental results in linear algebra, this system of equations is under-determined and has infinitely many solutions.

We want to narrow the number of solutions to one and therefore we need more information, i.e., some additional constraints. The choice of this constraint affects not only the nature of the desired solution but also the computational complexity and feasibility of the required computations. E.g., if we want to obtain the solution with the smallest energy or ℓ_2 -norm, there exists a closed form solution, the minimum norm solution given by

$$\mathbf{x} = \mathbf{M}^+ \mathbf{y} = \mathbf{M}^T (\mathbf{M}\mathbf{M}^T)^{-1} \mathbf{y}, \quad (10.2)$$

where $\mathbf{M}^+ = \mathbf{M}^T (\mathbf{M}\mathbf{M}^T)^{-1}$ is the Moore-Penrose pseudoinverse.

If we want to find the most sparse solution to the under-determined system $\mathbf{y} = \mathbf{A}\mathbf{x}$ we need to solve

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \quad (10.3)$$

where $\|\mathbf{x}\|_0$ is the ℓ_0 -norm of \mathbf{x} . This norm is simply the number of non-zero elements of \mathbf{x} and hence is a measure of sparsity. This problem is in general NP-hard and therefore not computationally feasible. However, it has been shown that under certain conditions [34] the solution is unique and can be found by using the ℓ_1 -norm instead of the difficult ℓ_0 -norm. Since the ℓ_1 -norm is convex, the problem can be solved using standard convex optimization tools and even linear programming.

Multi-scale transforms such as the critically sampled DWT is an example of a representation that is sparse. This transform can be described by a linear operator \mathbf{W} . The DWT is orthonormal which means that the dimension of the transformed signal is the same as the dimension of the signal itself and we have $\mathbf{W}^{-1} = \mathbf{W}^T$. The atoms of the dictionary are the columns of \mathbf{W} and the number of atoms is the same as the number of elements in the signal.

One overcomplete version of DWT is UDWT. In the overcomplete case, the analysis operator \mathbf{W} has more rows than columns while the synthesis operator \mathbf{W}^T has more columns than rows. In contrast with the orthonormal case, the columns of \mathbf{W}^T no longer form an orthonormal basis but an overcomplete frame.

Overcomplete transforms are sometimes referred to as redundant transforms, since the \mathbf{W} has more rows than columns. As an example, the 2-level UDWT of an image with N pixels contains $7N$ coefficients. The redundancy is 7-fold. The inverse or synthesis transform is a projection from \mathbb{R}^{7N} -the transform space to \mathbb{R}^N -the image space.

The transform space of the synthesis matrix $\mathbf{W}^T \in \mathbb{R}^{7N \times N}$ is composed of the N -dimensional range space and the $6N$ -dimensional null space of the transform which is orthogonal to the range space. Changes in the coefficients within the range space result in changes in the synthesized image while changes in the coefficients that lie in the null space produce no changes in the synthesized image because it is orthogonal to the range space.

This means that an infinite number of different configurations of the coefficients in the null space produce the same image. This is where sparsity comes in. The best configuration of the coefficients is the one where as few coefficients as possible contain the energy of the image and most of the coefficients are zero. Overcomplete systems allow for sparse representations of the coefficients and this is what we are exploiting in this part of the thesis.

It has been shown repeatedly that for a number of applications in signal and image processing such as compression, coding and denoising, the best exploitation of transform sparsity leads to better and more practical solutions of those problems [34]. The sparse representations result in faster and more efficient computations.

The main difference between the two approaches to ℓ_1 -norm sparsity regularization, i.e., analysis and synthesis approaches, is to be found in the way the sparsity of the coefficients of the overcomplete transforms is exploited. Considering the analysis formulation

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}\mathbf{x}\|_1, \quad (10.4)$$

we see that the forward projection of the signal on the basis elements (rows of \mathbf{W}) is being sparsified, i.e., the assumption is that the analyzed coefficients can be sparsely approximated. The important thing here is that the optimization is performed in the signal or image-space where \mathbf{x} resides. The problem solved in Part I of the thesis is of the exactly same form. There, the operator \mathbf{W} is the TV operator. Piece-wise smooth signals are sparse under this operator and thus the best solution is the one where $\text{TV}(\mathbf{x})$ is sparse.

The synthesis approach is formulated as

$$J(\boldsymbol{\omega}) = \|\mathbf{y} - \mathbf{M}\mathbf{W}^T\boldsymbol{\omega}\|_2^2 + \lambda\|\boldsymbol{\omega}\|_1. \quad (10.5)$$

Here, the representation being sparsified is the linear combination of the atoms (columns of \mathbf{W}^T). There are essentially an infinite number of representations that give rise to the same reconstructed signal $\mathbf{W}^T\boldsymbol{\omega}$, i.e., the probability of finding a sparse representation is high. The optimization is being performed in the coefficient space, i.e., in the same space as the transformed signal $\boldsymbol{\omega}$ resides in.

Many papers have been written dealing with the nature of these two formulations, such as [31], [35]. This is an active research topic and is still open to debate.

Solutions of the Analysis and Synthesis Problems

11.1 Analysis Prior Formulation

The cost function using the analysis prior is stated as

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1, \quad (11.1)$$

where λ is the regularization parameter and \mathbf{W} is the wavelet analysis (forward) operator. Minimization of this cost function is problematic because after we have differentiated it and solved for \mathbf{x} we will end up having in the solution the inverse of a term containing the matrix $\mathbf{M}^T\mathbf{M}$. This matrix is of enormous size and inverting it is not practical.

However, using Majorization-Minimization techniques this can be avoided by adding the term $Q_{\mathbf{M}^T\mathbf{M}}(\mathbf{x}, \mathbf{x}_k) = (\mathbf{x} - \mathbf{x}_k)^T(\alpha\mathbf{I} - \mathbf{M}^T\mathbf{M})(\mathbf{x} - \mathbf{x}_k)$ to the objective function (11.1). The constant α is chosen as $\alpha > \max \text{eig}(\mathbf{M}^T\mathbf{M})$ such that $(\mathbf{x} - \mathbf{x}_k)^T(\alpha\mathbf{I} - \mathbf{M}^T\mathbf{M})(\mathbf{x} - \mathbf{x}_k)$ is positive.

The cost function to minimize becomes

$$Q(\mathbf{x}, \mathbf{x}_k) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + Q_{\mathbf{M}^T\mathbf{M}}(\mathbf{x}, \mathbf{x}_k) + \|\mathbf{W}\mathbf{x}\|_1 \quad (11.2)$$

$$= \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + (\mathbf{x} - \mathbf{x}_k)^T(\alpha\mathbf{I} - \mathbf{M}^T\mathbf{M})(\mathbf{x} - \mathbf{x}_k) + \|\mathbf{W}\mathbf{x}\|_1. \quad (11.3)$$

Differentiation of the first term gives

$$\frac{\partial}{\partial \mathbf{x}} \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 = -2\mathbf{M}^T\mathbf{y} + 2\mathbf{M}^T\mathbf{M}\mathbf{x}. \quad (11.4)$$

Differentiation of the second term gives

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_k)^T (\alpha \mathbf{I} - \mathbf{M}^T \mathbf{M})(\mathbf{x} - \mathbf{x}_k) = 2\alpha \mathbf{x} - 2\alpha \mathbf{x}_k - 2\mathbf{M}^T \mathbf{M} \mathbf{x} + 2\mathbf{M}^T \mathbf{M} \mathbf{x}_k. \quad (11.5)$$

The term $-2\mathbf{M}^T \mathbf{M} \mathbf{x}$ will cancel out the same term in (11.4). Differentiation of the third term gives

$$\frac{\partial}{\partial \mathbf{x}} \lambda \|\mathbf{W} \mathbf{x}\|_1 = \lambda \mathbf{W}^T \mathbf{\Lambda} \mathbf{W} \mathbf{x}, \quad (11.6)$$

where $\mathbf{\Lambda}$ is defined as the diagonal matrix $\mathbf{\Lambda} = \text{diag}(1./|\mathbf{W} \mathbf{x}|)$ where $./$ denotes element-wise division and $|\cdot|$ denotes element-wise absolute value.

Adding all terms, setting to zero and rearranging gives

$$\alpha \mathbf{x} + \frac{\lambda}{2} \mathbf{W}^T \mathbf{\Lambda} \mathbf{W} \mathbf{x} = \alpha \mathbf{x}_k + \mathbf{M}^T \mathbf{y} - \mathbf{M}^T \mathbf{M} \mathbf{x}_k \quad (11.7)$$

$$(\alpha \mathbf{I} + \frac{\lambda}{2} \mathbf{W}^T \mathbf{\Lambda} \mathbf{W}) \mathbf{x} = \alpha \mathbf{x}_k + \mathbf{M}^T (\mathbf{y} - \mathbf{M} \mathbf{x}_k) \quad (11.8)$$

Now define $\mathbf{b} = \alpha \mathbf{x}_k + \mathbf{M}^T (\mathbf{y} - \mathbf{M} \mathbf{x}_k)$ and we have

$$\mathbf{x} = (\alpha \mathbf{I} + \frac{\lambda}{2} \mathbf{W}^T \mathbf{\Lambda} \mathbf{W})^{-1} \mathbf{b}. \quad (11.9)$$

The matrix inversion lemma gives

$$(\alpha \mathbf{I} + \frac{\lambda}{2} \mathbf{W}^T \mathbf{\Lambda} \mathbf{W})^{-1} = \frac{1}{\alpha} \left(\mathbf{I} - \mathbf{W}^T \left(\frac{2\alpha}{\lambda} \mathbf{\Lambda}^{-1} + \mathbf{W} \mathbf{W}^T \right)^{-1} \mathbf{W} \right). \quad (11.10)$$

Substituting this into (11.9) gives

$$\mathbf{x} = \frac{\mathbf{b}}{\alpha} - \frac{1}{\alpha} \mathbf{W}^T \left(\frac{2\alpha}{\lambda} \mathbf{\Lambda}^{-1} + \mathbf{W} \mathbf{W}^T \right)^{-1} \mathbf{W} \mathbf{b}. \quad (11.11)$$

Define a new variable \mathbf{z} as

$$\mathbf{z} = \left(\frac{2\alpha}{\lambda} \mathbf{\Lambda}^{-1} + \mathbf{W} \mathbf{W}^T \right)^{-1} \mathbf{W} \mathbf{b} \quad (11.12)$$

$$= \left(\frac{2\alpha}{\lambda} \text{diag}(|\mathbf{W} \mathbf{x}|) + \mathbf{W} \mathbf{W}^T \right)^{-1} \mathbf{W} \mathbf{b} \quad (11.13)$$

so that now we have

$$\mathbf{x} = \frac{\mathbf{b}}{\alpha} - \frac{1}{\alpha} \mathbf{W}^T \mathbf{z}. \quad (11.14)$$

This suggests the iteration

$$\mathbf{z}^{(k+1)} = \left(\frac{2\alpha}{\lambda} \text{diag}(|\mathbf{W} \mathbf{x}^{(k)}|) + \mathbf{W} \mathbf{W}^T \right)^{-1} \mathbf{W} \mathbf{b} \quad (11.15)$$

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{b}}{\alpha} - \frac{1}{\alpha} \mathbf{W}^T \mathbf{z}^{(k)}. \quad (11.16)$$

There is a problem with this formulation. It is evident that the term $\mathbf{W}\mathbf{W}^T$ will require solving a very large system of equations in the matrix inversion term in (11.15). However, a little trick can be used to arrive at a feasible solution. To find \mathbf{z} in (11.13) we need to solve

$$\left(\frac{2\alpha}{\lambda}\mathbf{\Lambda}^{-1} + \mathbf{W}\mathbf{W}^T\right)\mathbf{z} = \mathbf{W}\mathbf{b}. \quad (11.17)$$

By subtracting the term $\mathbf{W}\mathbf{W}^T\mathbf{z}$ from both sides and adding $c\mathbf{z}$ to both sides we arrive at

$$\left(\frac{2\alpha}{\lambda}\text{diag}(|\mathbf{W}\mathbf{x}^{(k)}|) + c\mathbf{I}\right)\mathbf{z} = \mathbf{W}\mathbf{b} + (c\mathbf{I} - \mathbf{W}\mathbf{W}^T)\mathbf{z}, \quad (11.18)$$

which suggests the iteration

$$\mathbf{z}^{(k+1)} = \left(\frac{2\alpha}{\lambda}\text{diag}(|\mathbf{W}\mathbf{x}^{(k)}|) + c\mathbf{I}\right)^{-1} (\mathbf{W}\mathbf{b} + (c\mathbf{I} - \mathbf{W}\mathbf{W}^T)\mathbf{z}^{(k)}). \quad (11.19)$$

This is trivial to compute because the matrix $\left(\frac{2\alpha}{\lambda}\text{diag}(|\mathbf{W}\mathbf{x}^{(k)}|) + c\mathbf{I}\right)$ is diagonal. We have now shown that the analysis problem (11.1) can be solved using the following iterative algorithm

$$\mathbf{z}^{(k+1)} = \left(\frac{2\alpha}{\lambda}\text{diag}(|\mathbf{W}\mathbf{x}^{(k)}|) + c\mathbf{I}\right)^{-1} (\mathbf{W}\mathbf{b} + (c\mathbf{I} - \mathbf{W}\mathbf{W}^T)\mathbf{z}^{(k)}), \quad (11.20)$$

$$\mathbf{x}^{(k+1)} = \frac{1}{\alpha}\mathbf{b} - \frac{1}{\alpha}\mathbf{W}^T\mathbf{z}^{(k)}, \quad (11.21)$$

where $\alpha > \max \text{eig}(\mathbf{M}^T\mathbf{M})$ and $c > \max \text{eig}(\mathbf{W}\mathbf{W}^T)$.

To initialize the algorithm, $\mathbf{z}^{(0)}$ and $\mathbf{x}^{(0)}$ are simply chosen as zeros.

11.2 Synthesis Prior Formulation

The formulation using the synthesis prior is

$$J(\boldsymbol{\omega}) = \|\mathbf{y} - \mathbf{M}\mathbf{W}^T\boldsymbol{\omega}\|_2^2 + \lambda\|\boldsymbol{\omega}\|_1, \quad (11.22)$$

where \mathbf{W}^T is the synthesis (inverse) transform operator and \mathbf{x} is reconstructed by $\mathbf{W}^T\boldsymbol{\omega}$.

In order to solve this problem we will first look at a simpler problem, considering

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1. \quad (11.23)$$

What is important here is that the variables are in fact uncoupled, meaning we can expand (11.23) as

$$J(\mathbf{x}) = (y_1 - x_1)^2 + \lambda|x_1| + (y_2 - x_2)^2 + \lambda|x_2| + \cdots + (y_N - x_N)^2 + \lambda|x_N|. \quad (11.24)$$

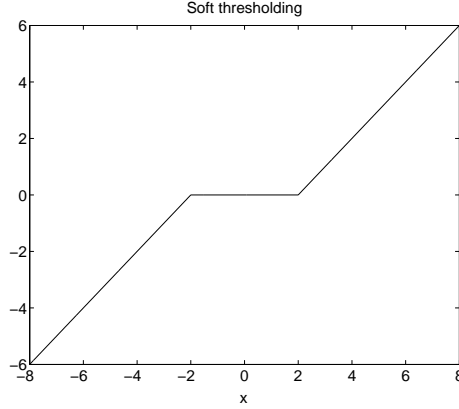


Figure 11.1: Soft thresholding

This means that in order to minimize $J(\mathbf{x})$ we can minimize each $(y_i - x_i)^2 + \lambda|x_i|$ term individually for $1 \leq i \leq N$, thus the problem is reduced to the minimization of a scalar function

$$f(x) = (y - x)^2 + \lambda|x|. \quad (11.25)$$

We proceed to finding the derivative w.r.t. x ,

$$f'(x) = -2(y - x) + \lambda \text{sign}(x), \quad (11.26)$$

setting to zero gives

$$y = x + \frac{\lambda}{2} \text{sign}(x). \quad (11.27)$$

The graph of this function is shown in Figure 11.1. This is called soft-thresholding and it is defined as

$$\text{soft}(x, T) = \begin{cases} x + T & x \leq -T \\ 0 & |x| < T \\ x - T & x \geq T \end{cases} \quad (11.28)$$

It has now been shown that the minimization of (11.23) is given by $\mathbf{x} = \text{soft}(\mathbf{y}, \lambda/2)$. Now we can consider the minimization of (11.22). Similar to what was done in the analysis case we begin by finding a majorizer for (11.22) that is easy to minimize.

$$Q_k(\boldsymbol{\omega}, \boldsymbol{\omega}_k) = \|\mathbf{y} - \mathbf{M}\mathbf{W}^T \boldsymbol{\omega}\|_2^2 + (\boldsymbol{\omega} - \boldsymbol{\omega}_k)^T (\alpha \mathbf{I} - \mathbf{W}\mathbf{M}^T \mathbf{M}\mathbf{W}^T) (\boldsymbol{\omega} - \boldsymbol{\omega}_k) + \|\boldsymbol{\omega}\|_1, \quad (11.29)$$

where $\alpha \geq \text{maxeig}(\mathbf{W}\mathbf{M}^T \mathbf{M}\mathbf{W}^T)$ in order to keep Q_k non-negative. Expanding the function gives

$$Q_k(\mathbf{x}, \mathbf{x}_k) = \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{M}\mathbf{W}^T \mathbf{x} + \mathbf{x}^T \mathbf{W}\mathbf{M}^T \mathbf{M}\mathbf{W}^T \mathbf{x} \quad (11.30)$$

$$\begin{aligned} &+ (\boldsymbol{\omega} - \boldsymbol{\omega}_k)^T (\alpha \mathbf{I} - \mathbf{W}\mathbf{M}^T \mathbf{M}\mathbf{W}^T) (\boldsymbol{\omega} - \boldsymbol{\omega}_k) + \|\boldsymbol{\omega}\|_1 \\ &= \alpha \|\boldsymbol{\omega}_k + \frac{1}{\alpha} \mathbf{W}\mathbf{M}^T (\mathbf{y} - \mathbf{M}\mathbf{W}^T \boldsymbol{\omega}_k) - \boldsymbol{\omega}\|_2^2 + \lambda \|\boldsymbol{\omega}\|_1 + C, \end{aligned} \quad (11.31)$$

where C is a constant w.r.t. \mathbf{x} . We observe that minimizing this function is the same as minimizing

$$Q_k(\mathbf{x}, \mathbf{x}_k) = \left\| \boldsymbol{\omega}_k + \frac{1}{\alpha} \mathbf{W} \mathbf{M}^T (\mathbf{y} - \mathbf{M} \mathbf{W}^T \boldsymbol{\omega}_k) - \boldsymbol{\omega} \right\|_2^2 + \lambda \|\boldsymbol{\omega}\|_1, \quad (11.32)$$

which has the same form as (11.23), therefore $\boldsymbol{\omega}^{(k+1)}$ is obtained by the soft-thresholding equation

$$\boldsymbol{\omega}^{(k+1)} = \text{soft} \left(\boldsymbol{\omega}_k + \frac{1}{\alpha} \mathbf{W} \mathbf{M}^T (\mathbf{y} - \mathbf{M} \mathbf{W}^T \boldsymbol{\omega}_k), \frac{\lambda}{2\alpha} \right). \quad (11.33)$$

To initialize the algorithm $\boldsymbol{\omega}^{(0)}$ is calculated by taking the given transform, e.g. UDWT, NSCT or FDCT, of \mathbf{x} that has been filled with zeros. When then the algorithm has converged, the estimated pansharpened image is reconstructed by using the inverse transform on the solution of (11.33), i.e., $\hat{\mathbf{x}} = \mathbf{W}^T \boldsymbol{\omega}$.

Experiments & Results

Experiments are performed using three overcomplete transforms, i.e., the Undecimated Discrete Wavelet Transform (UDWT) [28], the Non-Subsampled Contourlet Transform (NSCT) [29] and the Fast Discrete Curvelet Transform (FDCT) [30]. Two levels of decomposition are used in all experiments.

For the both dataset, results are evaluated for 41 values of λ ranging from 0 to 20 at intervals of 0.5. We use the ERGAS, SAM and Qave metrics to evaluate the quality of the pansharpened images.

The results for the Pleiades dataset are shown in Figures 12.1-12.3 and the results for the Quickbird dataset are shown in Figures 12.4-12.6. For the ERGAS and SAM metrics, lower values are better, while for the Qave metric, higher values are better with 1 being optimal.

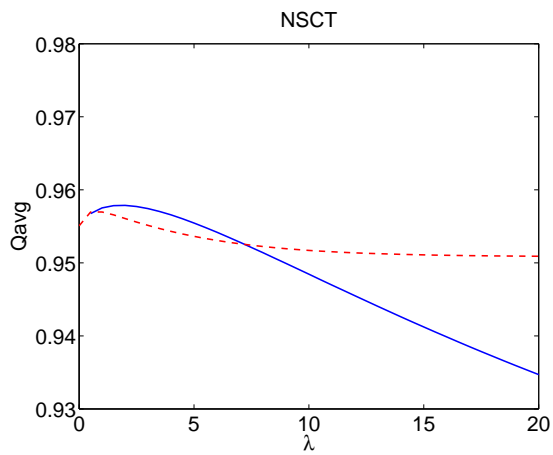
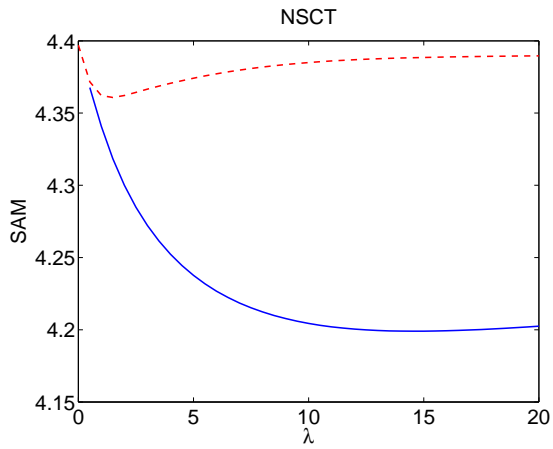
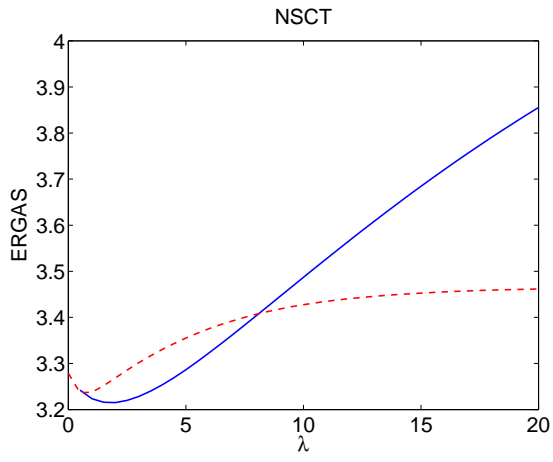


Figure 12.1: Results for Pleiades dataset with the NSCT for both analysis and synthesis for the ERGAS, SAM and Q_{ave} metrics

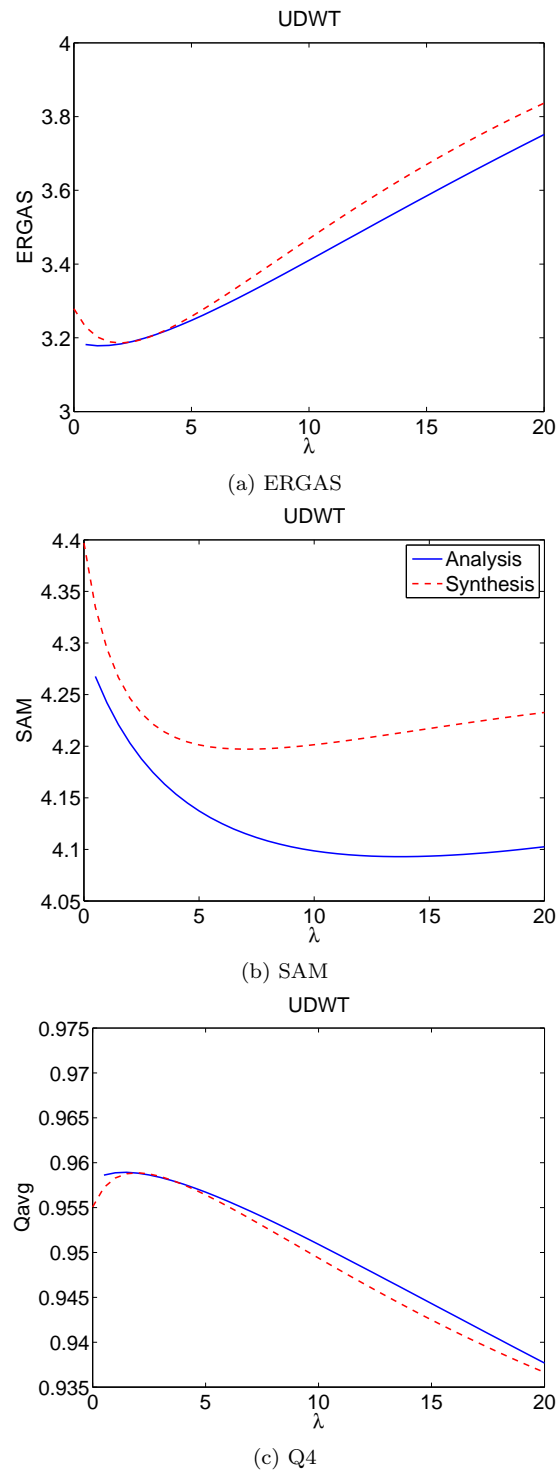
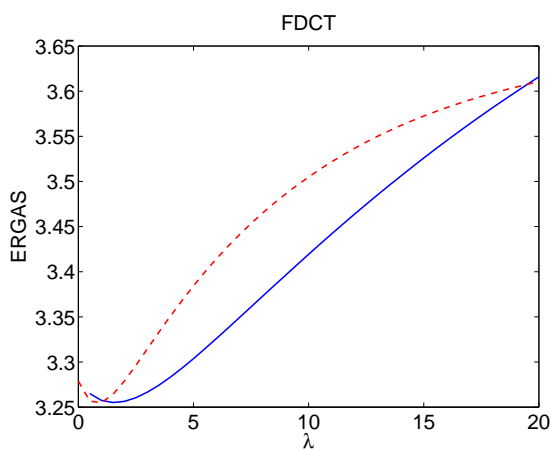
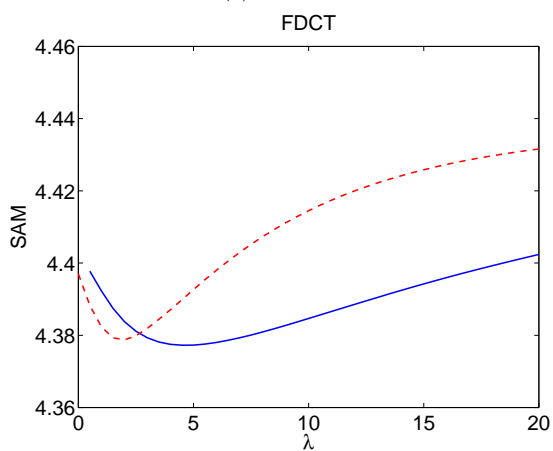


Figure 12.2: Results for Pleiades dataset with the UDWT for both analysis and synthesis for the ERGAS, SAM and Qave metrics



(a) ERGAS



(b) SAM

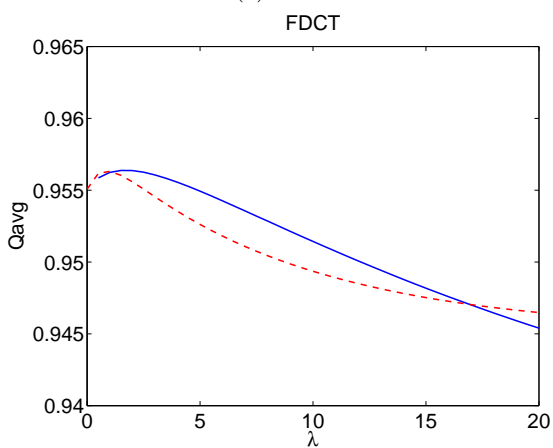
(c) Q_4

Figure 12.3: Results for Pleiades dataset with the FDCT for both analysis and synthesis for the ERGAS, SAM and Q_{ave} metrics

In Figures 12.1-12.3 the results for the Pleiades dataset are presented. Beginning with the NSCT in Figure 12.1 (a) we see that the analysis method gives better results for the ERGAS metric. The same is true for the SAM metric in Figure 12.1 (b). Here, the relative difference for the SAM metric is much greater than for the ERGAS metric. We also see that the first value for the analysis results are missing. This is because the algorithm does not converge for $\lambda = 0$. This applies to all the figures.

In Figure 12.1 (c) the results for the Qave metric are shown. For this metric, higher is better. Again the analysis method gives better results although the difference is quite small.

The results for the UDWT are shown in Figure 12.2. Things are similar to the NSCT results. The results for the Qave metric are the same for both methods.

Figure 12.3 shows the results for the FDCT. These results are very different from what we saw using the other transforms. Now the synthesis method gives better results for the ERGAS metric while the results for the other metrics are practically the same.

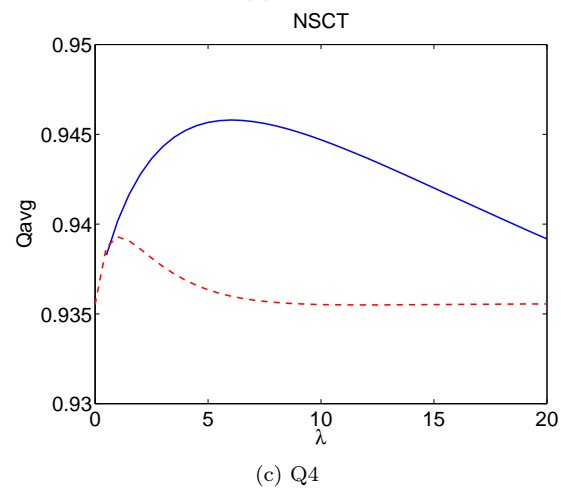
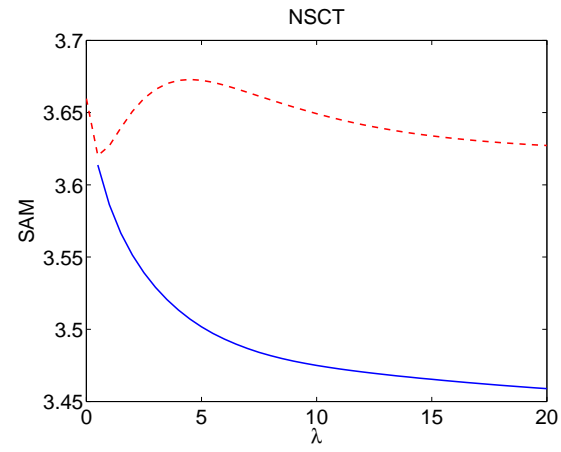
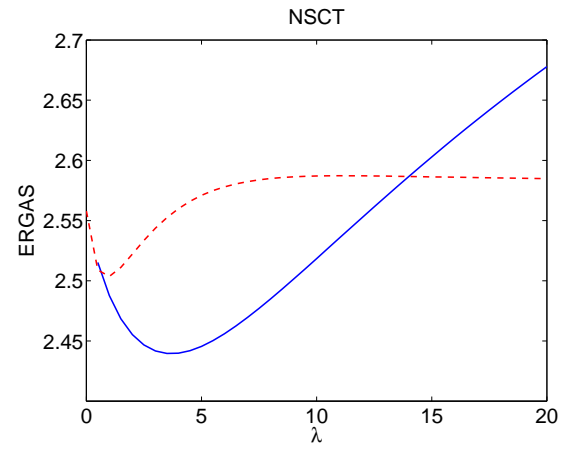


Figure 12.4: Results for Quickbird dataset with the NSCT for both analysis and synthesis for the ERGAS, SAM and Qave metrics

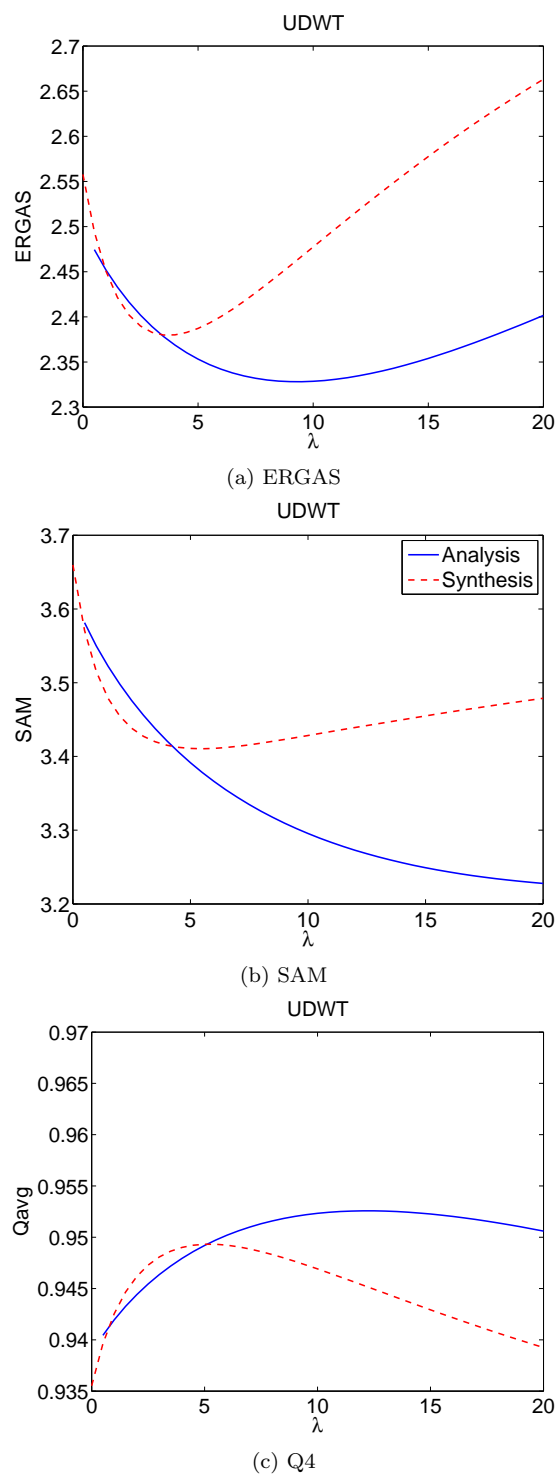
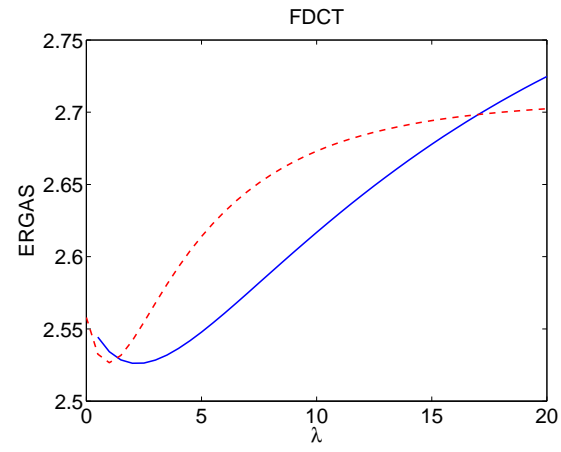
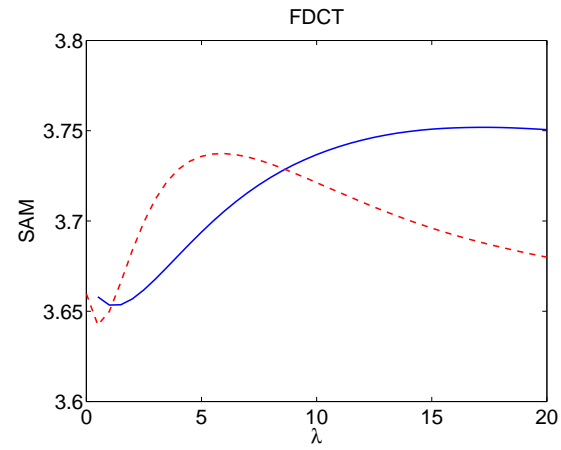


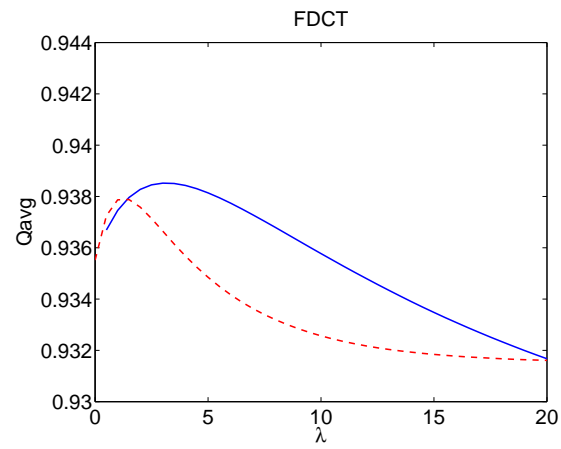
Figure 12.5: Results for Quickbird dataset with the UDWT for both analysis and synthesis for the ERGAS, SAM and Qave metrics



(a) ERGAS



(b) SAM



(c) Q4

Figure 12.6: Results for Quickbird dataset with the FDCT for both analysis and synthesis for the ERGAS, SAM and Qave metrics

In Figures 12.4-12.6 the results for the Quickbird dataset are presented. Although the curves look different from the Pleiades results, they exhibit the same general behavior as we saw in the Pleiades case. For the NSCT and UDWT the analysis method gives better results than synthesis, while for the FDCT the difference between the two methods is small, with synthesis giving better results for the SAM metrics.

A notable difference between the two datasets is that in Figures 12.4 (b) and 12.5 (b), the SAM metric does not have a global minimum for the range of λ chosen, i.e., $0 \leq \lambda \leq 20$. However, we can see that the curve is relatively flat at the largest value of λ thus the global minimum should not be significantly smaller than the value at $\lambda = 20$.

In Table 12.1, the best results for each metric in each experiment are shown. The table gives a general idea of the relative performance of the methods. For the NSCT and UDWT, the analysis approach gives better results in all experiments, while for the FDCT, the two approaches give very similar results. Overall, the UDWT gives best results.

Table 12.1: Best results obtained for each experiments. Bold font indicates best result obtained for analysis vs. synthesis.

Pleiades						
	NSCT		UDWT		Curvelet	
	Analysis	Synthesis	Analysis	Synthesis	Analysis	Synthesis
ERGAS	3.215	3.237	3.178	3.186	3.255	3.255
SAM	4.199	4.361	4.093	4.197	4.377	4.379
Qavg	0.958	0.957	0.959	0.959	0.956	0.956
Mississippi						
ERGAS	2.440	2.504	2.328	2.380	2.526	2.527
SAM	3.459	3.620	3.228	3.411	3.653	3.642
Qavg	0.946	0.939	0.953	0.949	0.939	0.938

Conclusions

In this part of the thesis we have proposed two kinds of methods to solve the ill-posed inverse problem of pansharpening based on exploiting the sparsity of overcomplete multi-scale transforms. These transforms are the UDWT, NSCT and the FDCT.

One implementation was based on the analysis approach while the other was based on the synthesis approach. For overcomplete transforms these two approaches give different results while for orthogonal transforms they are the same. Experiments were performed for two datasets and three quality metrics, a total of 36 combinations.

The results, as measured by the three metrics used, indicate that the analysis approach gives better performance than the synthesis approach. In 34 out of 36 experiments, analysis gives better results. However, the relative difference between the two approaches never exceeded 5.7%, with the average relative difference being 0.83% for the Pleiades dataset and 1.78% for the Quickbird dataset.

The assumption in the model matrix \mathbf{M} in (9.1), that a linear combination of bands gives the PAN image (see eq. (3.2)), is one possible reason for apparent similarity of the analysis and the synthesis method. This assumption sets strong constraints on the solution and thus limits the effectiveness of the sparsity regularization.

Thus, with this in mind, the most important conclusion to draw here is that we have reached the end of the road with the current model, used in parts I and II of this thesis. Future work is to relax the linear assumption constraint and then revisit the regularization methods discussed here and in Part I.

Part III

Classification of Pansharpened Satellite Images

Introduction

Several pansharpening techniques have been proposed during the past two decades [36]. For the majority of those techniques there is a compromise between the desired spatial enhancement and the spectral consistency. Achieving good spatial resolution usually compromises the spectral consistency and vice versa. Spectral consistency is an important property since spectral distortion can produce unreliable results in many applications such as classification.

There are few papers that address classification of pansharpened data. Pansharpened images are commonly not used directly for classification, rather, pansharpening is used to improve the image visually.

The classification of high resolution urban remote sensing imagery is a challenging research problem. Here we look at the classification of such data by both considering the classification of panchromatic imagery (single data channel) and spectral image (multiple data channels) obtained by image fusion. Obviously, both the spatial and spectral quality of the pansharpened image are inherently important in order to be able to correctly classify the pixels. Low spatial quality means missing details and low spectral quality, i.e., spectral distortion, can result in mis-classified pixels.

In this part of the thesis, we will investigate the classification of images that have been pansharpened using several different methods. Some of these methods have good spectral consistency while lacking spatial quality and for some methods, the opposite is true. We will also, try to assess the relative importance of spectral consistency vs spatial quality for classification.

In order to estimate the spectral and spatial quality of a pansharpened image, several metrics have been formulated, such as ERGAS [19], SAM [23], SID [24],

Qave [20], RASE [18] and more (see Chapter 6). To gain more insight into what qualities of the pansharpened image are important for classification, we will not only classify the multispectral pansharpened image but also add the original Pan image as well as morphological profiles (MP) and their derivatives (DMP), as proposed in [37, 38].

Panchromatic images are characterized by high spatial resolution. This high spatial resolution allows the identification of small structures in a dense urban area. However, the analysis of a scene by considering the value of a single pixel only will produce very poor classification results compared to the fine resolution. To solve this problem, some local spatial information is needed.

An interesting approach to provide local spatial information is based on the theory of *Mathematical Morphology* [39], which provides tools to analyze spatial relationship between pixels. The Morphological Profile (MP) was proposed in [37, 38] for segmentation of high-resolution satellite images. An MP is made up of an Opening Profile (OP) and a Closing Profile (CP) and we will also use the Derivative of the Morphological Profile (DMP) [38, 40]. Only the panchromatic data are used to build the morphological profile and its derivative.

For classification purposes, the MP and the DMP are regarded as feature vectors, where each class has a typical MP-DMP. Hence each MP (or DMP) is considered as a channel of a multispectral image. In this way, classification methods applied to multispectral images can be applied [38, 40].

Of the pansharpening methods used, one is our model-based method derived from the imaging sensor's physical properties. This method has the nice feature of being spectrally consistent [41] by design. The Intensity-Hue-Saturation (IHS) method has been widely used [6–9]. Another popular method is the Brovey [5] sharpening method. The IHS and Brovey methods suffer from spectral distortions. An important class of pansharpening methods are those based on multi-resolution analysis (MRA). In our experiments we will be using two such methods, one based on the Discrete Wavelet Transform (DWT) [3] and another based on the undecimated DWT [4]. Finally, we will use a pansharpening method based on Principal Component Analysis (PCA).

For the experiments, we have three different datasets along with ground truth data. There are 2 different IKONOS images of an urban area (Reykjavik, Iceland) with multispectral images with four bands (R, B, G and Near Infrared (NIR)) and panchromatic images which are of higher resolution (4 by 4 to 1 pixel) as compared to the multispectral images. Third experimental dataset is a QuickBird image of an urban area (Rome, Italy) [42] where the ratio of the pan and MS images is the same as for the IKONOS images and the MS image has also four bands. Thus we have data from two different sensors.

The work flow of the experiments is as follows. We begin by pansharpening

the data from each dataset using the different methods. For all datasets we classify the pan image and the MS data separately. The next step is to add the Pan image to the MS data and classify the resulting multi-channel image. Finally, we add an MP and a DMP. As noted before, the MP or DMP can be thought of as multispectral images on their own, where, of course, the different bands do not contain any spectral information, only spatial information. The classification is performed using Support Vector Machines (SVM) and Random Forests (RF).

The outline of this part of the thesis is as follows. In the next chapter, we will give a brief overview of the different pansharpening methods. Chapter 16 gives a brief review on mathematical morphology and Chapter 17 reviews the two types of classification methods used, SVM and RF, respectively. Experimental set up and results are given in Chapter 18 and finally, conclusions are drawn in Chapter 19.

Image Fusion

Image fusion can be done at several levels, e.g., at pixel, feature and decision levels. Here, we will only be concerned with pixel level fusion. The next sections give a brief overview of the pansharpening methods used in the experiments.

15.1 IHS fusion

The IHS method [6–9] is a frequently used method, especially when working with LANDSAT and SPOT imagery. The basic idea is to first transform the MS image into intensity (I), hue (H) and saturation (S) components (IHS colorspace)

$$\begin{bmatrix} I \\ H \\ S \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{-\sqrt{2}}{6} & \frac{-\sqrt{2}}{6} & \frac{2\sqrt{2}}{6} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

The next step is to scale the Pan image so that it has the same mean and variance as the intensity component of the MS image

$$P = \frac{\sigma_I}{\sigma_P}(P - \mu(P)) + \mu(I).$$

The intensity component is then replaced with the appropriately scaled Pan image and finally the inverse IHS transformation is taken to get the fused image

$$\begin{bmatrix} F(R) \\ F(G) \\ F(G) \end{bmatrix} = \begin{bmatrix} R + P - I \\ G + P - I \\ B + P - I \end{bmatrix}.$$

The IHS method produces images that have high spatial resolution and low spectral quality.

15.2 Brovey Fusion

The Brovey transform is a simple method for pansharpening and similar to the IHS method, produces images that have good spatial resolution and poor spectral quality. Each channel of the fused image is computed as

$$F(n) = NPan \circ \frac{MS(n)}{\sum_{k=1}^N MS(k)}$$

where \circ means element-wise multiplication, $F(n)$ and $MS(n)$ are the n -th channels of the fused and MS images, respectively, and N is the number of channels in the MS image. So N is either 3 or 4.

15.3 Model Based Fusion

Several fusion methods have been proposed that take into account the physics of the imaging sensor [41, 43]. The model based image fusion method [41] is derived from a model of the imaging sensor's underlying physics. This approach ensures spectral consistency of the resulting fused image. Spectral consistency can be measured as the cross correlation between the low resolution RGB image and the downsampled fused image. In our case the low resolution spectral image consists of four bands: R,B,G and Near Infrared (NIR).

The method of image fusion presented in [41] creates a high resolution multi band image $[R^{high}, G^{high}, B^{high}, N^{high}]$ from a high resolution panchromatic image P^{high} and a low resolution multi band image $[R^{low}, G^{low}, B^{low}, N^{low}]$. The first part of the method calculates a correlation matrix, Σ , between the five channels, $[R, G, B, N, P]$. This is done by considering the inner product, or spectral overlap, of the respective channels. The relative spectral response of the IKONOS sensor can be seen in Figure 15.1.

The high resolution image is then grouped into blocks, where each block exactly corresponds to a low resolution pixel $[R_i^{low}, G_i^{low}, B_i^{low}, N_i^{low}]$. Assuming a perfect overlap between low and high resolution pixels, spectral consistency, which

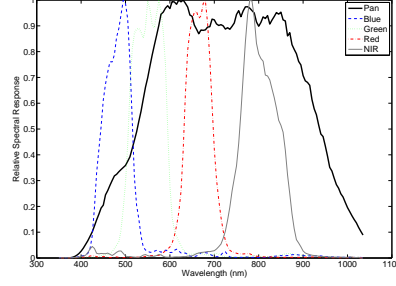


Figure 15.1: Relative Spectral Response of the IKONOS sensor for the panchromatic, R,G,B and N bands.

is enforced, has the form

$$\begin{bmatrix} R_i^{low} \\ G_i^{low} \\ B_i^{low} \\ N_i^{low} \end{bmatrix} = \frac{1}{16} \sum_{j=1}^n \begin{bmatrix} R_{ij}^{high} \\ G_{ij}^{high} \\ B_{ij}^{high} \\ N_{ij}^{high} \end{bmatrix}, \quad (15.1)$$

where j runs over the n high resolution pixels corresponding to the low resolution pixel i . It is then assumed that these blocks follow a multivariate normal distribution, with a variance of Σ , and a mean value computed via conditional means utilizing that the P_{ij}^{high} and $[R_i^{low}, G_i^{low}, B_i^{low}, N_i^{low}]$ are known. This is used in a Bayesian setting, where this normal distribution is used as the data term.

To complete the Bayesian setting, a prior (or model of a likely image) is also needed. For that, a piecewise smooth model is used. The model is specifically formulated by penalizing deviation between neighboring high resolution pixel values, *except* for when the panchromatic image P has edges. The data term and the prior are combined into a Markov Random Field (MRF) framework. The additional constraint on this MRF framework is that the resulting $[R^{high}, G^{high}, B^{high}, N^{high}]$, must be spectrally consistent. That is accomplished by observing that (15.1), describes a hyperplane (i.e., a linear constraint), and then only allowing solutions located on that hyperplane. The resulting optimization problem is a large sparse least squares function - which has the nice quality that it is convex.

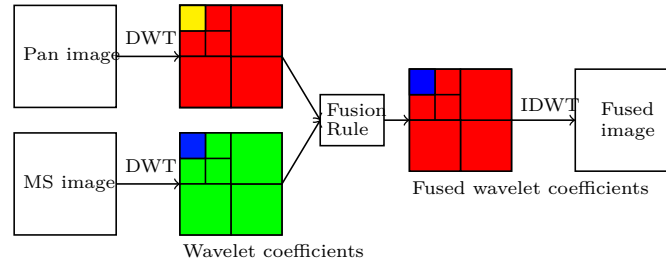


Figure 15.2: Basic wavelet based pansharpening.

15.4 MRA Method

The multi-resolution analysis approach to pansharpening is widely used and there are numerous techniques today based on it. The basic idea is to take the discrete wavelet transform (DWT) [3] of both the MS and Pan images. The next step is to retain the approximation coefficients for the MS image but replace the detail coefficients with those from the Pan image. Instead of just replacing the coefficients, some fusion rule can be used. This is shown in figure 15.2.

The main drawback to this approach is that there will be substantial artifacts in the fused image due to the nature of the DWT, i.e., it is not shift-invariant and lacks directionality so there is bound to be some aliasing, thus the spatial quality of the final fused image is reduced. One approach to overcome this problems is to use the undecimated DWT [4].

15.5 PCA Method

Similar to the MRA method, the PCA method for pansharpening [5] is a so called spatial detail injection method. First, the mutually correlated bands of the MS image are transformed using PCA into a set of independent components whose number is the same as number of bands in the MS image. The first principal component has the highest variance and is similar to the Pan image itself. The next step is to replace this component with the actual Pan image and finally take the inverse transform to get the fused image. Figure 15.3 illustrates this.

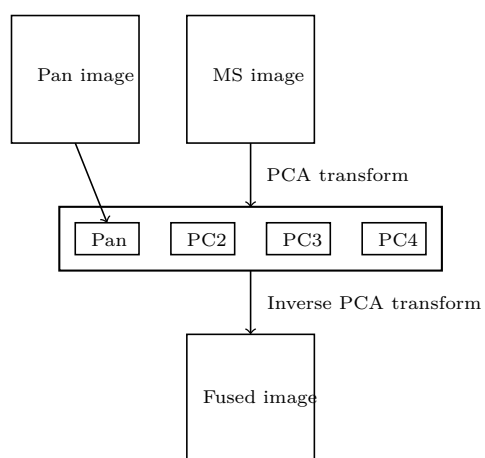


Figure 15.3: PCA based pansharpening.

Mathematical Morphology

Mathematical Morphology (MM) is a theory that provides mathematical tools to analyze spatial relationship between pixels [37, 39, 44]. Matheron and Serra originally introduced MM to study binary porous media in the 1960s [45].

When working with MM, one is usually interested in specific objects or structures in the image. These objects of interest are viewed as subsets of the image. After these objects have been identified, several sets of known size and shape (such as square, line or disk) can be used to characterize their morphology.

These simple sets are called *Structuring Elements* (SEs). An SE always has an origin (usually its symmetric center) that allows one to position the SE at a given pixel of the image.

The most fundamental morphological operations are *erosion* and *dilation*. The eroded value at a given pixel x is the minimum pixel value over the SE and the dilated value at a given pixel x is the maximum pixel value over the SE. Erosion and dilation of the Lena image is shown in Figure 16.1.

Two of the most common operations of MM are *opening* and *closing* operations. The morphological opening of an image is an erosion followed by dilation while the morphological closing of an image is the reverse: dilation followed by erosion. Both operations are done using the same SE. The morphological opening of an image removes light features that are smaller than the SE whereas morphological closing removes dark features smaller than the SE.

While these operations are useful, they have the drawback of not being connected filters. What this means is that they do not preserve shapes and thus



Figure 16.1: Erosion and dilation of the Lena image using a disk of radius 10 as the SE.

introduce shape noise. So instead we will use the morphological reconstruction operations. Opening and closing of the Lena image is shown in Figure 16.2.



Figure 16.2: Opening and Closing of the Lena image using a disk of radius 10 as the SE.

Opening by reconstruction of an image removes unconnected light features while *closing by reconstruction* of an image removes unconnected dark features. For both operations, shapes are preserved and the structures present after the transformation are of a size greater than or equal to the SE used. This effectively means that less shape noise is generated. Opening and closing by reconstruction of the Lena image is shown in Figure 16.3.

Since objects in a given image may be of varying sizes, the use of a single SE might prove to be inefficient. To solve this, a multiscale approach can be taken, where SEs of the same shape but of several different sizes are used.

A Morphological Profile (MP) [37, 44] is a $2n + 1$ -dimensional vector that consists of the input image, n openings and n closings (reconstruction operations)



Figure 16.3: Opening and closing by reconstruction of the Lena image using a disk of radius 10 as the SE.

generated with an SE of fixed shape but of varying size. The MP can be defined as

$$MP(x) = [CP_n(x), \dots, I(x), \dots, OP_n(x)]$$

where CP and OP are closing profile and opening profile of n -dimension, constructed using an SE of fixed shape but of varying size and $I(x)$ is the original panchromatic image. An n dimensional profile has n different sizes in total. This kind of structure is called a *granulometry*. The number of openings/closings and the corresponding sizes of the SE depend on the size distribution of structures of interest in the panchromatic image. Finally, since these are operations performed on the panchromatic image, the MPs contain no spectral information, i.e., they only contain spatial information. Two different MPs are shown in Figure 16.4.

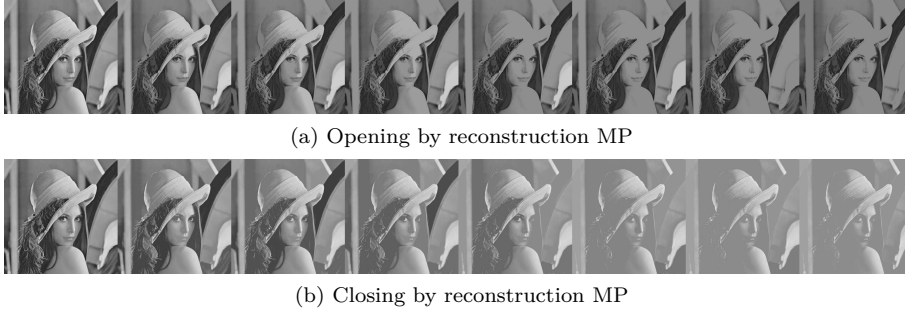


Figure 16.4: MPs based on successive Opening by reconstruction and closing by reconstruction operations using a disk shaped SE of increasing radius.

The *Derivative of the Morphological Profile* (DMP) [40] is defined as a $2n$ -dimensional vector equal to the discrete derivative of the MP and is given by

$$DMP_i(x) = |MP_{i-1}(x) - MP_i(x)|.$$

The information provided by the DMP is both spatial and radiometric. For a given pixel, the shape of the DMP can give an idea of the neighborhood of the pixel, e.g., whether it belongs to darker or lighter structure than the surrounding pixels. Finally, the amplitude of the DMP gives information about the local contrast of the structure.

From the above it can be seen that an MP (DMP) can be viewed as a multiband image, where the different openings and closings (or their derivatives) make up the different bands. In a classification setting, inclusion of these profiles effectively increases the number of features available for classification.

Classification Methods

17.1 Support Vector Machine

One of the most promising of recent developments in the field of machine learning and pattern recognition is the Support Vector Machine (SVM) [46–48]. These are supervised learning methods that are widely used for classification and regression. When given a set of training samples where each sample is marked as belonging to one of two classes, the SVM algorithm builds a model that can predict to which class new data points will belong to.

The SVM model can be viewed as a representation of the data points as points in space where the separate categories or classes are divided by a gap that is as wide as possible. The SVM constructs a hyperplane in a space that has a high or infinite dimension. For a good separation, the distance from the training data points of each class to the hyperplane should be maximized. When this distance or margin is maximized we call the resulting linear classifier, a maximum margin classifier.

For this linear classifier to be able to solve non-linear problems, the non-linear data points are mapped into a higher dimensional space in such a way that linear classification in this new space is equivalent to non-linear classification in the original space. This is what is known as the 'kernel trick'. By using non-linear kernel-functions, this approach can separate complex (e.g., multi-modal) class distributions in high dimensional feature spaces. A commonly used kernel-function is the Gaussian Radial Basis function and it is used for the experiments presented here.

The SVM is in fact a binary classification strategy so some kind of multiclass decomposition is required to handle multiclass problems. A single multiclass problem is reduced into multiple binary classification problems. Two common methods are the One-Against-All (OAA) and One-Against-One (OAO).

For a multiclass problem with k -classes, $k(k-1)/2$ binary SVMs are trained. For OAO, classification is done by max-wins voting strategy where every individual binary classifier assigns the instance to one of the two classes thus increasing the vote count for that class by one.

Since SVM does not provide class-labels as classifier output but instead provides the distance from each data point to the hyperplane, the sign of the distance is used to determine a class. Finally, the instance classification is determined by the class with the most votes. A detailed discussion on SVM is for example given in [49].

17.2 Random Forest

Random Forest (RF) [50] is a relatively recent ensemble method for classification and regression. Ensemble classifiers get their name from the fact that several classifiers, i.e., an ensemble of classifiers, are trained and their individual results are then combined through a voting process. Many such methods have been proposed [51–53]. These methods are usually based on the techniques of boosting [54, 55] or bagging [56]. Boosting uses an iterative re-training procedure where the samples that have been incorrectly classified are given increased weight with each iteration. Bagging (or bootstrap aggregation) is based on training an ensemble of classifiers using samples that have been bootstrapped from a training set. This has been shown to reduce the variance of the classification.

Trees are good examples of high-variance and low-bias procedures so bagging seems to work very well for them. The idea is to average many noisy but relatively unbiased models in order to reduce the variance. Since trees are inherently noisy, averaging is of great benefit. Also, when grown sufficiently deep, they have a low bias.

Boosting is in general a much more computationally costly method than bagging but it is considerably more accurate. The RF algorithm uses an improved method of bootstrapping as bagging and has been shown to be comparable to boosting in terms of accuracy while being free from the drawbacks of boosting. Thus RF is a much faster algorithm than boosting.

During the training process, the RF algorithm grows many CART-like trees [57], with each individual tree trained on bootstrapped samples of the original train-

ing set. In order to determine a split for each node, the algorithm only searches across a randomly selected subset of the input variables.

In order to classify an object from an input vector, the input vector is run down each tree in the forest. Each tree gives a classification or a unit vote for a particular class and the forest chooses the classification having the most votes.

The most important user-tunable parameter in the RF algorithm is the number of variables used for a split of a tree. A common value for this parameter is the square root of the number of inputs. By limiting this parameter, the amount of computation can be greatly reduced while the correlation between the trees in forest is minimized. Thus reducing the error rate. Since each tree only uses a portion of the available input variables, the algorithm is much faster than a conventional bagging algorithm with a similar tree-like classifier. Another nice feature is that the trees do not need to be pruned, hence reducing the computational costs even further.

The computational time of the RF algorithm [57] has been shown to be

$$cT\sqrt{MN}\log(N)$$

where c is some constant, T is the number of trees in the forest, M is the number of variables and N is the number of samples in the dataset. This shows that the algorithm is not very computationally intensive but on the other hand it can require a considerable amount of memory as it needs to store an N by T matrix while running.

RF as a classifier has many nice properties [57] such as excellent accuracy, it scales up very well, it can handle thousands of variables and a lot of missing data and it is insensitive to noise in the training labels. It also gives an unbiased estimate of the test set error as trees are added to the ensemble and finally it does not overfit.

Datasets and Experimental Results

There are three datasets used the experiments. Two of these are images of Reykjavik, Iceland, acquired by the IKONOS Earth imaging satellite on 9th August 2001. Each dataset consists of a low-resolution (4m) multispectral image with four bands R,G,B and NIR and a high-resolution panchromatic image of resolution 1m. The ratio between the panchromatic and low-resolution images is 16 or 4 by 4 pixels.

The third dataset is an image of Rome, Italy acquired by the QuickBird satellite. As with the other datasets, it consists of a low-resolution (2.4m) multispectral image with the four bands R, G, B and NIR and a high-resolution panchromatic image of resolution 0.6m. The ratio between the Pan and MS image is again 4.

These images were then fused using the methods in Table 18.1, and the resulting high-resolution multispectral image was used for classification.

Ground truth data for 6 classes (*Large buildings, small buildings, large roads, streets, open areas and shadows*) was available for the IKONOS datasets and extensive ground-truth data with 9 classes was available for the QuickBird dataset.

Two different classification algorithms were used, 1) SVM and 2) RF. For the SVM part, the libSVM¹ library was used. The Gaussian Radial Basis function was chosen for the kernel and the training parameters were found using a grid-search.

¹Chih-Chung Chang and Chih-Jen Lin - LIBSVM – A Library for Support Vector Machines
- <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

Table 18.1: Notation for the different pansharpening methods used in experiments

Name	Type
MBF	Model Based Fusion
Brovey	Brovey fusion
IHS	IHS fusion
DWT	MRA based fusion using the DWT
UDWT	MRA based fusion using the UDWT
PCA	PCA based fusion

Table 18.2: The Morphological Profiles (MP) used in the experiments.

Dataset	Design	SEs (radius of disks)
1	3 openings, 3 closings	2,4,6
2,3	11 openings, 11 closings	1,2,3,4,6,8,10,12,17,25,30

For the RF part, the randomforest-Matlab² package, developed by Abhishek Jaiaantilal was used. The number of trees was chosen to be 200, all other options used the default values.

For the datasets, two different MPs (and their derivatives, DMPs) were constructed. Details on these profiles are shown in Table 18.2.

By including these profiles in the classification process, the number of features is increased by their respective lengths, i.e., by 6 and 22, respectively.

For each dataset and pansharpening method, five different experiments were conducted. The first one was to simply classify the Pan image. The second experiment was to classify the pansharpened MS image. The third type of experiment was to classify the pansharpened MS image along with the Pan image. In the fourth experiment, the pansharpened MS image with an MP was classified and finally, the MP was replaced with the DMP.

These experiments were done using both a three band (RGB) pansharpened MS image and the four band (RGBNIR) image. Since the original IHS method can only handle the RGB bands, this method was omitted when using the four

²Abhishek Jaiaantilal, randomforest-Matlab, Random Forest (Regression, Classification and Clustering) implementation for MATLAB (and Standalone) - <http://code.google.com/p/randomforest-matlab/>

band MS images.

Each MS image produced by the different pansharpening methods was evaluated using a number of quality metrics, 7 metrics for spectral quality and one metric to evaluate the spatial quality. A summary of these metrics can be found in Chapter 6.

There two different approaches to evaluating these metrics. One to compare the upsampled original MS image with final pansharpened image. Another way of doing this is to fuse the low-resolution MS image with a downsampled (degraded) Pan image and then comparing the final result with the original MS image. We chose to use the former approach.

For the metrics CC, ERGAS, RASE, RMSE, SAM and SID lower is better, while for the Qave and spatial metric, closer to one is better.

All classification accuracies are presented using 3 different methods. The Overall Accuracy (OA) is simply the percentage of correctly classified pixels. The Average Accuracy (AA) is the mean of class-specific accuracy for all the classes. Finally, the Kappa Coefficient (Kappa) is the percentage of agreement, i.e., correctly classified pixels, corrected by the number of agreements that one would expect purely by chance alone.

In the next three subsections a short description of the datasets is given and the experimental results for each dataset are presented. The OA is used in the discussion.

18.1 First Dataset

This dataset is composed of a high resolution panchromatic image (976×640 pixels) and a low resolution multispectral image consisting of 4 bands: R, G, B and NIR. As stated above, the ratio between the panchromatic and low-resolution images is 16 or 4 by 4 pixels. The panchromatic image can be seen in Figure 18.3 (a).

It is important to note that the training set was chosen in such a way as to be not very representative and thus making the classification a much more challenging task. Information on the size of the training and test sets is shown in Table 18.3 and the training and validation sets are shown in Figure 18.1 (a) and (b), respectively.

The spectral and spatial quality metrics for all the pansharpening methods are shown in Table 18.4 (three bands) and Table 18.5 (four bands), respectively.

Table 18.3: Training and test set details for the first dataset.

Class	Train	Test
Small buildings	1526	34155
Open areas	7536	43867
shadows	1286	25806
Large buildings	2797	30916
Large roads	3336	39202
Streets	5616	35147
Total	22097	209093

Table 18.4: Quality metrics for the three band (RGB) data using first dataset

Metric	MBF	Brovey	IHS	DWT	UDWT	PCA
CC	0.002	0.014	0.007	0.033	0.004	0.027
ERGAS	2.359	8.846	8.904	5.636	6.502	9.436
Qave	0.728	0.680	0.713	0.985	0.688	0.673
RASE	8.623	33.256	32.996	22.111	24.400	35.325
RMSE	0.013	0.051	0.051	0.034	0.038	0.054
SAM	1.072		2.081	1.407	2.105	2.470
SID	0.001	0.0	0.005	0.018	0.003	0.004
spatial	0.501	0.981	0.985	0.847	0.998	0.969

Since there are so many different metrics, it is not easy to order the different methods by performance, but it is evident while studying the three band data in Table 18.4 that the MBF method produces by far the best spectrally consistent results while its spatial performance is rather low.

The DWT method gives the second best results when it comes to spectral quality while the spatial quality is the second worst. The IHS, Brovey and PCA methods produce similar results of low spectral quality and rather good spatial quality, while the UDWT method gives better spectral quality and the best spatial quality of all the methods. The same can be said about the four band data in Table 18.5.

The results from the classification experiments are presented in Table 18.6 and Table 18.7, for three band (RGB) and four band (RGBN) data, respectively. Note that the IHS method is not available for the four band data.

In the following discussion we use the OA and only refer to the results achieved using the RF classifier. The accuracies achieved using the two classifiers are

Table 18.5: Quality metrics for the four band (RGBNIR) data using first dataset

Metric	MBF	Brovey	DWT	UDWT	PCA
CC	0.007	0.026	0.033	0.067	0.233
ERGAS	2.518	7.426	5.636	6.387	8.212
Qave	0.976	0.946	0.985	0.858	0.949
RASE	10.072	27.575	22.111	25.261	31.763
RMSE	0.016	0.043	0.034	0.039	0.049
SAM	1.103		1.407	4.472	2.461
SID	0.007	0.0	0.018	0.013	0.004
spatial	0.522	0.920	0.847	0.998	0.808

somewhat similar for the first dataset, with the SVM classifier usually trailing behind the RF, but for the other datasets, RF gives significantly better results than SVM in every single case.

18.1.1 Three Band Data (RGB)

By studying Table 18.6 we see that the classification of the Pan image alone gives OA of 37.2%. When classifying the pansharpened image alone the MBF data give the best accuracy, 56.57% closely followed by the PCA and UDWT data. The Brovey results are the worst in terms of accuracy at 50.93% with IHS only slightly better. For this experiment we see that difference between best and worst accuracy is 5.64%, so all the results are relatively similar.

When the Pan image is added to classification process, the results in terms of accuracy change somewhat drastically. Now the data giving best accuracies are the UDWT, at 63.20%, while the MBF data are second best at 59.77%. Again the accuracies for the Brovey data are at the bottom (50.70%) with IHS only slightly better. The difference in accuracy between best and worst is now 12.5%, more than twice the difference from the previous experiment.

It is also interesting to note that the Brovey, IHS and PCA data perform slightly worse in this experiment than in the previous one. It seems that adding the spatial information from the Pan image into the classification process is not of any noticeable benefit for these data. The average gain in accuracy, compared to the previous experiment, for all the methods is 2.37%.

The next two experiments add the MP followed by the DMP so we have richer spatial information available for classification. The results are very similar for these two last experiments so we will only comment on the results obtained by including the MP. The MP for this dataset was chosen to be short; we have only three different sizes for the SEs. This dataset did not benefit from longer

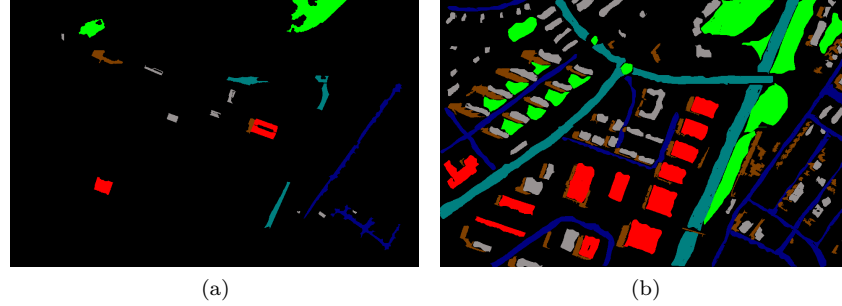


Figure 18.1: Training map (a) and reference map (b) for dataset 1.

profiles. This is probably a result from the fact that the training set was chosen in a way not to be representative, thus making the classification difficult.

Again the UDWT method produces the best accuracies, 67.87% with MBF coming second at 65.72%. The Brovey and IHS data give the worst accuracies, 61.21% and 61.17%, respectively. The DWT data 65.12% and the PCA data 63.51%. The difference between best and worst accuracies is now 6.66% and it is clear that the further inclusion of more spatial information decreases this difference compared to the previous experiment (Pan and MS). The most spectrally inconsistent data, gain the most from the MP/DMP. For example, we see a gain of 10.51% accuracy for the Brovey data, while the gain for the UDWT data is only 4.67%. The average gain in accuracy by including the MP is 7.28% and 6.78% by including the DMP, a difference of 0.5% favoring the MP.

18.1.2 Four Band Data (RGB and NIR)

Table 18.7 presents the results for the experiments using all four bands of the fused image. By comparing this table to the previous one, we see that there are some important differences. It is clear that the benefit of adding the NIR band is very different for the various pansharpening methods. If we look at the classification of MS data only, it is evident that the best results are achieved by the UDWT data, 63.77%, an increase of 7.85% from the three band experiment.

Surprisingly, the accuracy for the Brovey data now comes second at 58.94%, an increase of 8.01% compared the the three-band experiment. In the same manner, the DWT data show a modest increase of accuracy, 3.73% to 58.09% and the PCA data goes from 55.24% to 58.72%, an increase of only 3.48%. The best spectrally consistent MBF data only shows an increase of accuracy of 2.01%, to 58.58%, coming third in this experiment.

Clearly, the biggest increase in accuracy is for the Brovey data which also is the least spectral consistent. It is also noticeable that the UDWT data perform best now. While both the MBF and DWT data give better spectral quality, the UDWT data scores highest in spatial quality, at 0.998 according to Table 18.5.

When the Pan data are included, there are not as big gains overall in accuracy as in the three band case. This can easily be explained by the presence of the NIR band in the MS data, so the addition of spatial data from the Pan image is relatively smaller. The most notable result is that all data gain in accuracy except the PCA data. In the three band case we saw very big gains in accuracy for the UDWT in this experiment, but in this case the gain in accuracy is only 1.58%. The average gain in accuracy for all data is now 1.35%. Also, similar to the three band case, the UDWT data give the best accuracies and the MBF data come second.

By looking at the last two experiments (MP/DMP) in Table 18.7 we see a similar trend as for the three band data but obviously the gains in accuracy are not quite as big, now that we have increased the amount of MS data. The average gain in accuracy by including the MP is 6.34% and the gain achieved by including the DMP is 4.92%, giving a difference of 1.42%, which is almost three times more than for the three-band data.

Table 18.6: Classification accuracies (%) for all experiments using first dataset with three input channels (RGB)

Dataset1		RF		SVM											
Features		Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT
P	AA	AA	49.30	49.30	49.30	49.30	49.30	49.30	AA	48.98	48.98	48.98	48.98	48.98	48.98
	OA	OA	48.45	48.45	48.45	48.45	48.45	48.45	OA	47.98	47.98	47.98	47.98	47.98	47.98
	K	K	37.32	37.32	37.32	37.32	37.32	37.32	K	36.80	36.80	36.80	36.80	36.80	36.80
RGB	AA	AA	55.78	51.80	53.60	54.09	56.27	55.89	AA	53.40	49.91	52.45	51.78	55.54	52.44
	OA	OA	56.57	50.93	52.94	54.36	55.96	55.92	OA	53.12	48.85	51.79	51.38	55.25	52.20
	K	K	47.52	40.53	43.07	44.84	46.76	46.77	K	43.61	38.00	41.65	41.38	45.93	42.39
PRGB	AA	AA	59.41	51.59	53.17	59.53	55.55	62.74	AA	59.76	50.06	52.53	58.42	55.61	60.58
	OA	OA	59.77	50.70	52.45	59.55	55.24	63.20	OA	59.90	48.94	51.80	58.38	55.25	60.93
	K	K	51.39	40.26	42.51	51.13	45.90	55.49	K	51.59	38.11	41.67	49.77	45.94	52.82
RGB+MP	AA	AA	64.61	61.60	61.04	64.28	63.13	67.10	AA	67.19	60.50	60.87	65.66	64.56	66.65
	OA	OA	65.72	61.21	61.17	65.12	63.51	67.87	OA	67.98	60.07	60.57	66.20	64.60	67.53
	K	K	58.50	53.04	52.99	57.76	55.83	61.10	K	61.30	51.72	52.38	59.14	57.23	60.75
RGB+DMP	AA	AA	64.68	60.84	61.12	64.09	63.30	66.95	AA	65.40	57.79	59.72	63.96	62.13	64.59
	OA	OA	65.37	60.21	60.67	64.53	63.09	67.70	OA	66.33	57.33	59.48	64.33	62.19	65.81
	K	K	58.11	51.87	52.45	57.09	55.40	60.92	K	59.31	48.41	51.07	56.93	54.31	58.68

Table 18.7: Classification accuracies (%) for all experiments using first dataset with four input channels (RGBNIR)

Dataset1	RF			SVM		
Features	Accuracy	MBF	Brovey	DWT	PCA	UDWT
P	AA	49.30	49.30	49.30	49.30	49.30
	OA	48.45	48.45	48.45	48.45	48.45
	K	37.32	37.32	37.32	37.32	37.32
RGBN	AA	57.83	58.93	57.55	58.64	63.12
	OA	58.58	58.94	58.09	58.72	63.77
	K	49.94	50.36	49.32	50.07	56.20
PRGBN	AA	60.75	59.14	60.69	58.68	64.25
	OA	61.23	59.15	60.95	58.71	64.78
	K	53.12	50.60	52.79	50.07	57.40
RGBN+MP	AA	67.04	65.77	66.50	66.59	68.02
	OA	67.65	66.25	66.99	66.88	68.73
	K	60.86	59.15	60.04	59.94	62.15
RGBN+DMP	AA	64.48	65.32	64.55	65.00	67.08
	OA	65.20	65.70	65.16	65.42	67.93
	K	57.90	58.49	57.84	58.14	61.18
	AA	48.98	48.98	48.98	48.98	48.98
	OA	47.98	47.98	47.98	47.98	47.98
	K	36.80	36.80	36.80	36.80	36.80
	AA	57.89	59.03	56.31	58.65	62.07
	OA	58.19	59.17	56.50	58.58	62.75
	K	49.58	50.68	47.55	49.98	55.04
	AA	61.34	59.00	60.03	58.60	63.06
	OA	61.63	59.13	60.14	58.52	63.48
	K	53.66	50.62	51.90	49.91	55.91
	AA	67.08	66.12	66.16	65.62	67.32
	OA	67.91	66.70	66.84	66.13	68.28
	K	61.23	59.76	59.92	59.07	61.66
	AA	65.78	64.87	64.72	64.72	65.35
	OA	66.69	65.43	65.34	65.25	66.56
	K	59.76	58.24	58.15	57.98	59.60

Table 18.8: Training and test set details for the second dataset

Class	Training	Test
Houses	1863	6213
Open Areas	6068	28144
Shadows	2619	10610
Large buildings	5599	29768
Large Roads	2489	12051
Streets	4103	11940
Total	22741	98726

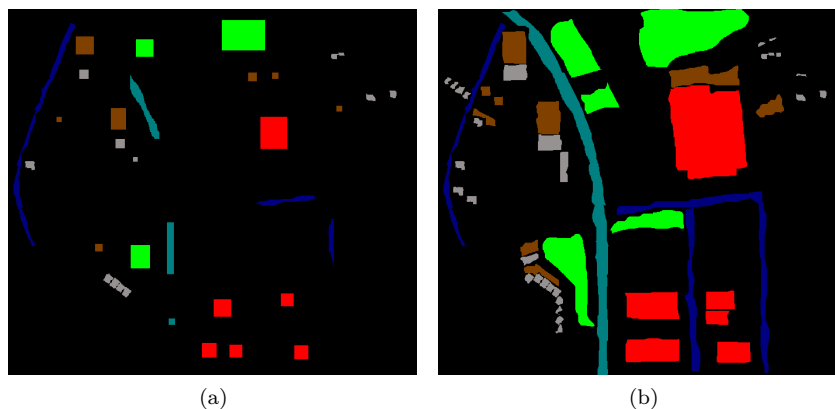


Figure 18.2: Training map (a) and reference map (b) for dataset 2.

18.2 Second Dataset

The second dataset is also an IKONOS image of Reykjavik, Iceland. The high resolution panchromatic image is 628×700 pixels and the lower-resolution multispectral image consists of R,G,B and NIR bands. The size ratio is again 16 or 4 by 4 pixels. The panchromatic image is shown in Figure 18.3 (c). Information on the size of the training and test sets is shown in Table 18.8 and the training and validation sets are shown in Figure 18.2 (a) and (b), respectively.

The training set has been chosen to be more representative than that for the previous dataset hence the accuracy is generally much better than for the first dataset. The MP/DMP is now of length 23.

Table 18.9: Quality metrics for the three band (RGB) data using second dataset

RGB	MBF	Brovey	IHS	DWT	UDWT	PCA
bandCorrs	0.003	0.009	0.004	0.056	0.006	0.038
ERGAS	2.766	10.306	10.523	6.657	7.504	8.034
Qave	0.727	0.673	0.708	0.980	0.704	0.701
RASE	10.199	37.296	37.003	26.151	29.155	30.723
RMSE	0.010	0.035	0.035	0.025	0.028	0.029
SAM	1.082	NA	2.754	1.481	1.824	2.099
SID	0.003	0.0	0.002	0.032	0.007	0.003
spatial	0.699	0.990	0.995	0.907	0.999	0.972

Table 18.10: Quality metrics for the four band (RGBNIR) data using second dataset

dataset2-RGBN	MBF	Brovey	DWT	UDWT	PCA
bandCorrs	0.014	0.081	0.056	0.083	0.288
ERGAS	3.105	9.375	6.657	7.544	7.372
Qave	0.969	0.924	0.980	0.900	0.961
RASE	11.912	33.797	26.151	29.627	28.571
RMSE	0.011	0.032	0.025	0.028	0.027
SAM	1.287	NA	1.481	3.300	2.123
SID	0.009	0.0	0.032	0.018	0.003
spatial	0.718	0.970	0.907	0.999	0.904

The spectral and spatial quality metrics for all the pansharpening methods are shown in Table 18.9 (three bands) and Table 18.10 (four bands), respectively. The MBF method produces the best results in terms of spectral quality while having the worst spatial quality. However, the spatial quality of 0.699 is significantly higher than for the previous dataset.

If we try to order the methods from best to worst according to spectral quality (based on the average of the metrics where lower is better), we have the ranking: MBF (2.344), DWT (5.734), UDWT (6.421), PCA (6.821), IHS (8.387) and Brovey (9.529). This should give a reasonable idea how these methods rank spectrally. Spatially, they rank, from best to worst: UDWT (0.999), IHS (0.995), Brovey (0.990), PCA (0.972), DWT (0.907) and MBF (0.699).

These results reflect the complementary nature of the spectral and spatial quality. The best spectrally performing method has also the lowest spatial quality. The best spatially performing method, UDWT, has rather low spectral quality.

The values for the four-band data are similar. It is worth noting though, that in general the Qave values for three-band data are lower (mean : 0.749) than for the four-band data (mean : 0.947).

The accuracies from the classification experiments are presented in Table 18.6 and Table 18.7, for three band (RGB) and four band (RGBN) data, respectively. As before, we will use the OA in the discussion and omit the SVM results.

18.2.1 Three Band Data (RGB)

Classification accuracy for the Pan data alone is 62.63% which is considerably better than for the previous dataset. The next experiment is classification of the pansharpened MS data. The average accuracy of all data is 75.41%. The data giving the best accuracy is now the UDWT at 79.29%, the PCA data come second with 77.41%, followed by the MBF data at 75.84%, the DWT data with 74.59% and finally the IHS and Brovey data with 73.73% and 71.63%, respectively. Compared to the first dataset, we see now that the MBF data goes from giving best accuracy to being third. The method with the best spatial detail, UDWT, gives the best result even though the spectral quality is lower than that of MBF and DWT.

By adding the Pan image to the classification we now see the difference between data produced by the different methods become even more pronounced. The average accuracy for this experiment is 77.31% which is only slightly higher than for the MS data alone (75.41%) but as we have seen before, the accuracy for the IHS, Brovey and PCA methods lowers a little bit compared to the previous experiment. This means greater increase for the MBF and wavelet based methods. The accuracy of the MBF data increases by 2.92%, the DWT data see an increase of 5.33% and finally for the UDWT data there is an increase of 4.90%. The data that show an increase in accuracy here are those which give the best spectral quality in Table 18.9. The data produced by the other methods do not benefit at all from the added spatial information in the Pan image.

In the final two experiments for this dataset, an MP and an DMP are included. By looking at Table 18.11 one sees that the increase in accuracy is, in every case, more for the DMP than for the MP, so we will only comment on the DMP experiment, since the behavior is very similar.

The average gain in accuracy by including the DMP in the classification is 10.92% (7.83% for the MP). This is significantly higher than for the first dataset. Since these two IKONOS images are somewhat similar, the rather large differences in the experiments are probably due to the way the training sets were chosen.

The data that gain most in terms of accuracy in this experiment are those with low spectral quality, namely the IHS, Brovey and PCA data. The gains are 16.59%, 14.13% and 11.24%, respectively. The smallest gain, 5.12%, is indeed for the UDWT data, which have the best spatial detail.

18.2.2 Four Band Data (RGB and NIR)

The inclusion of the NIR band in the pansharpened data has a similar effect as was seen for the first dataset. The experiment results are shown in Table 18.12. Basically, we see the same kind of behavior as before.

Since we now have four bands of MS data, the classification accuracy is better than for the three-band case. The average accuracy is 80.86% so the NIR band brings in an average increase of 5.12%. The UDWT data gives yet again the best results 85.31%, followed by Brovey with 80.53%, the PCA with 80.02%, DWT with 79.96% and finally the MBF with 78.46%. The Brovey method comes second here, which was also the case in this experiment for the first dataset. It is also interesting to see how well the UDWT data perform here, almost 5% better than the second best.

Adding the Pan data to the classification does not give much change in terms of accuracy, the average gain in accuracy is only 0.89%. This is also similar to what was seen in the first dataset. The biggest gains in terms of accuracy are for the Brovey and DWT data, 1.89% and 1.98%, respectively. For the UDWT data there is only gain in accuracy of 0.37%, so it is clear that these data don't gain anything from the spatial information in the Pan data, which may be explained by the high spatial quality of the UDWT data.

Again, there is more gain in accuracy when adding the DMP to the classification process. But unlike the three-band case, the gains are not nearly as great, the average being 7.15%. This is consistent with the results from the previous dataset. The MBF data gain the most, 8.70% while the UDWT data gain the least, 3.79. The explanation for this is probably the low spatial quality of the MBF data and the high spatial quality of the UDWT data.

Table 18.11: Classification accuracies (%) for all experiments using second dataset with three input channels (RGB)

Dataset2		RF										SVM									
Features		Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT						
P	AA	53.46	53.46	53.46	53.46	53.46	53.46	53.46	AA	48.28	48.28	48.28	48.28	48.28	48.28						
	OA	62.63	62.63	62.63	62.63	62.63	62.63	62.63	OA	60.73	60.73	60.73	60.73	60.73	60.73						
	K	51.75	51.75	51.75	51.75	51.75	51.75	51.75	K	48.09	48.09	48.09	48.09	48.09	48.09						
RGB	AA	71.21	68.01	69.39	69.39	70.26	72.26	75.23	AA	61.43	53.96	55.79	56.99	57.92	60.39						
	OA	75.84	71.63	73.73	73.73	74.59	77.41	79.29	OA	70.96	64.22	65.99	68.00	69.87	70.08						
	K	69.50	63.93	66.59	66.59	67.81	71.22	73.80	K	62.93	53.55	56.17	58.82	61.04	61.84						
PRGB	AA	74.41	67.49	69.19	69.19	74.86	71.94	79.88	AA	64.63	54.02	55.86	61.81	58.55	65.42						
	OA	78.76	70.79	73.29	73.29	79.92	76.89	84.19	OA	73.22	64.13	66.07	72.52	69.69	75.16						
	K	73.15	62.93	66.08	66.08	74.48	70.60	79.95	K	65.90	53.49	56.28	64.75	60.93	68.29						
RGB+MP	AA	81.64	80.70	80.82	80.82	81.16	81.50	82.62	AA	80.05	76.28	76.44	78.47	78.31	80.37						
	OA	85.27	84.56	84.60	84.60	85.03	85.16	86.16	OA	83.48	80.17	80.31	82.02	81.75	84.10						
	K	81.36	80.45	80.50	80.50	81.04	81.20	82.48	K	79.08	74.98	75.16	77.27	76.91	79.85						
RGB+DMP	AA	84.59	82.94	83.04	83.04	83.96	83.87	84.97	AA	80.01	77.86	77.94	79.04	79.07	80.07						
	OA	88.75	87.37	87.42	87.42	88.40	88.13	89.31	OA	82.97	81.11	81.21	82.24	82.02	83.31						
	K	85.68	83.93	83.97	83.97	85.22	84.88	86.38	K	78.49	76.20	76.32	77.58	77.31	78.91						

Table 18.12: Classification accuracies (%) for all experiments using second dataset with four input channels (RGBNIR)

Dataset2		RF		SVM									
Features		Accuracy	MBF	Brovey	DWT	PCA	UDWT	Accuracy	MBF	Brovey	DWT	PCA	UDWT
P	AA	53.46	53.46	53.46	53.46	53.46	53.46	AA	48.28	48.28	48.28	48.28	48.28
	OA	62.63	62.63	62.63	62.63	62.63	62.63	OA	60.73	60.73	60.73	60.73	60.73
	K	51.75	51.75	51.75	51.75	51.75	51.75	K	48.09	48.09	48.09	48.09	48.09
RGB	AA	74.92	76.00	75.50	75.50	75.08	81.13	AA	69.98	67.35	67.20	68.04	74.80
	OA	78.46	80.53	79.96	79.96	80.02	85.31	OA	76.01	75.42	75.32	75.01	81.45
	K	72.84	75.31	74.60	74.60	74.58	81.34	K	69.56	68.60	68.43	68.15	76.36
PRGBN	AA	76.78	76.14	77.27	77.27	75.23	81.53	AA	72.05	67.49	69.93	68.13	76.83
	OA	80.35	80.69	81.94	81.94	80.05	85.68	OA	77.50	75.53	76.80	75.11	82.58
	K	75.19	75.51	77.05	77.05	74.63	81.83	K	71.49	68.74	70.46	68.27	77.85
RGBN+MP	AA	82.08	81.81	81.54	81.54	81.79	82.85	AA	80.86	78.85	79.60	79.99	82.10
	OA	85.65	85.33	85.21	85.21	85.34	86.07	OA	83.67	82.64	82.56	82.85	85.46
	K	81.83	81.44	81.28	81.28	81.44	82.39	K	79.37	78.04	77.99	78.35	81.58
RGBN+DMP	AA	85.02	84.12	84.68	84.68	84.04	85.50	AA	80.76	79.77	80.34	80.48	81.15
	OA	89.05	88.54	89.02	89.02	88.38	89.46	OA	83.49	82.88	83.25	83.40	83.95
	K	86.07	85.41	86.01	86.01	85.20	86.59	K	79.18	78.40	78.87	79.04	79.74

Table 18.13: Training and validation set details for the third dataset

Class	Training	Test
Buildings	10872	54360
Blocks	6589	32945
Roads	9811	49056
Light Train	964	4818
Vegetation	4177	20884
Trees	5451	27253
Bare Soil	4876	24381
Soil	903	4515
Tower	2875	14375
Total	46518	232587

18.3 Third Dataset

The third dataset is a QuickBird image of Rome, Italy. The Pan image is 972×1188 pixels and the spatial resolution is 0.6m, while the low-resolution four-band MS image has spatial resolution of 0.24m. The ratio of high-resolution Pan pixels to the low-resolution MS pixels is 4, same as for the IKONOS data. The panchromatic image is shown in Figure 18.3 (e). There are several things that make this dataset challenging. First, this image is taken when the satellite was far from being at the nadir. As a result the buildings in the images have a very 'oblique' look. Second, this is a dense urban area with primarily three classes of buildings: 1) small buildings (1 – 2 floors), 2) blocks, large buildings with 3 – 5 floors and 3) towers, tall rectangular buildings with more than 5 floors. There are 9 different classes. Information on the classes and number of training and validation samples is shown in Table 18.13.

The quality metrics for this dataset are shown in Table 18.14 (three-band) and Table 18.15 (four-band). The rank of the methods is similar as for the previously discussed datasets.

The results of the classification experiments are presented in Table 18.16 for the three-band data and in Table 18.17 for the four-band data.

Table 18.14: Quality metrics for the three band (RGB) data using third dataset

RGB	MBF	Brovey	IHS	DWT	UDWT	PCA
bandCorrs	0.029	0.003	0.018	0.049	0.020	0.090
ERGAS	1.562	6.068	6.235	3.779	4.291	4.484
Qave	0.745	0.709	0.729	0.991	0.732	0.717
RASE	5.645	26.367	25.894	14.474	15.626	17.206
RMSE	22.900	106.968	105.049	58.718	63.393	69.804
SAM	0.677	0.0	1.967	1.012	1.821	1.776
SID	0.001	0.001	0.010	0.029	0.009	0.012
spatial	0.587	0.967	0.968	0.771	0.995	0.889

Table 18.15: Quality metrics for the four band (RGBNIR) data using third dataset

RGBN	MBF	Brovey	DWT	UDWT	PCA
bandCorrs	0.026	0.113	0.049	0.077	0.330
ERGAS	1.652	4.711	3.779	4.108	3.827
Qave	0.992	0.970	0.991	0.965	0.962
RASE	6.593	20.444	14.474	15.807	15.625
RMSE	26.746	82.941	58.718	64.127	63.388
SAM	0.731	0.0	1.012	2.390	1.638
SID	0.009	0.001	0.029	0.013	0.006
spatial	0.609	0.924	0.771	0.995	0.766

18.3.1 Three Band Data (RGB)

Classification of the Pan image alone gives the relatively low accuracy of 41.03%. The reason for the low AA and Kappa values for this experiment is that several classes had accuracy of zero or close to zero.

Classification of the MS data gives average accuracy of 65.07% which is a big improvement in accuracy compared to the previous experiment. Now the IHS data give the best accuracy 66.86%, MBF data give 65.59%, Brovey 65.45%, then PCA with 64.78%, UDWT with 64.68% and finally DWT with 63.06%. The difference here between the best and worst accuracy is 3.80%. Interestingly, the least spectrally consistent data give the best classification results here.

Adding the Pan image to the classification brings the average accuracy up to 69.62%, which does not seem much. Closer inspection shows that only the MBF and the wavelet based data show improvement in accuracy, 8.22% for MBF, 9.42% for DWT and 12.04% for UDWT. The accuracy for the other data actually decreases by almost a percentage. The difference in accuracy between the best (UDWT, 76.72%) and worst (Brovey, 64.64%) is 12.08%, which is a

much larger difference than for the previous experiment.

Adding the MP and DMP brings big gains in accuracy for this dataset. The MP and the DMP improve the average accuracy by 22.64% and 22.79%, respectively. If we look at the DMP, the average accuracy is 92.42%, which is very good. Of all the different pansharpening methods, UDWT gives yet again the best results, followed closely by the MBF at 93.39%. IHS is again at the bottom with 91.69%. The difference between best and worst accuracy is 3.78%.

18.3.2 Four Band Data (RGB and NIR)

With the added NIR band, the average accuracy for classification of the MS data alone is now 76.05%. This is an improvement of 10.98% over the three band experiment. The UDWT data give the best accuracy, 78.47%, followed by the MBF at 76.34%, DWT at 75.32%, Brovey at 75.07% and finally PCA at 75.04%. The difference between best and worst accuracy is 3.40%.

Adding the Pan data to the classification does not improve the accuracy significantly. The average accuracy increases by a meager 0.83% up to 76.88%. The Brovey and PCA data actually decrease in accuracy by a small amount. The UDWT data give the best accuracy, 79.69%, MBF comes second with 77.92%. The difference between the best and worst accuracy is 4.75%.

The inclusion of the MP and DMP does not help as much as in the three-band case in terms of accuracy. The respective gains of the MP and DMP are now 17.27% and 17.34%. The UDWT data give the highest accuracy of 94.79%, followed closely by MBF at 94.17%, PCA at 94.16%, Brovey at 94.09% and DWT at 93.90%. The difference between the best and worst accuracy is 0.89%.

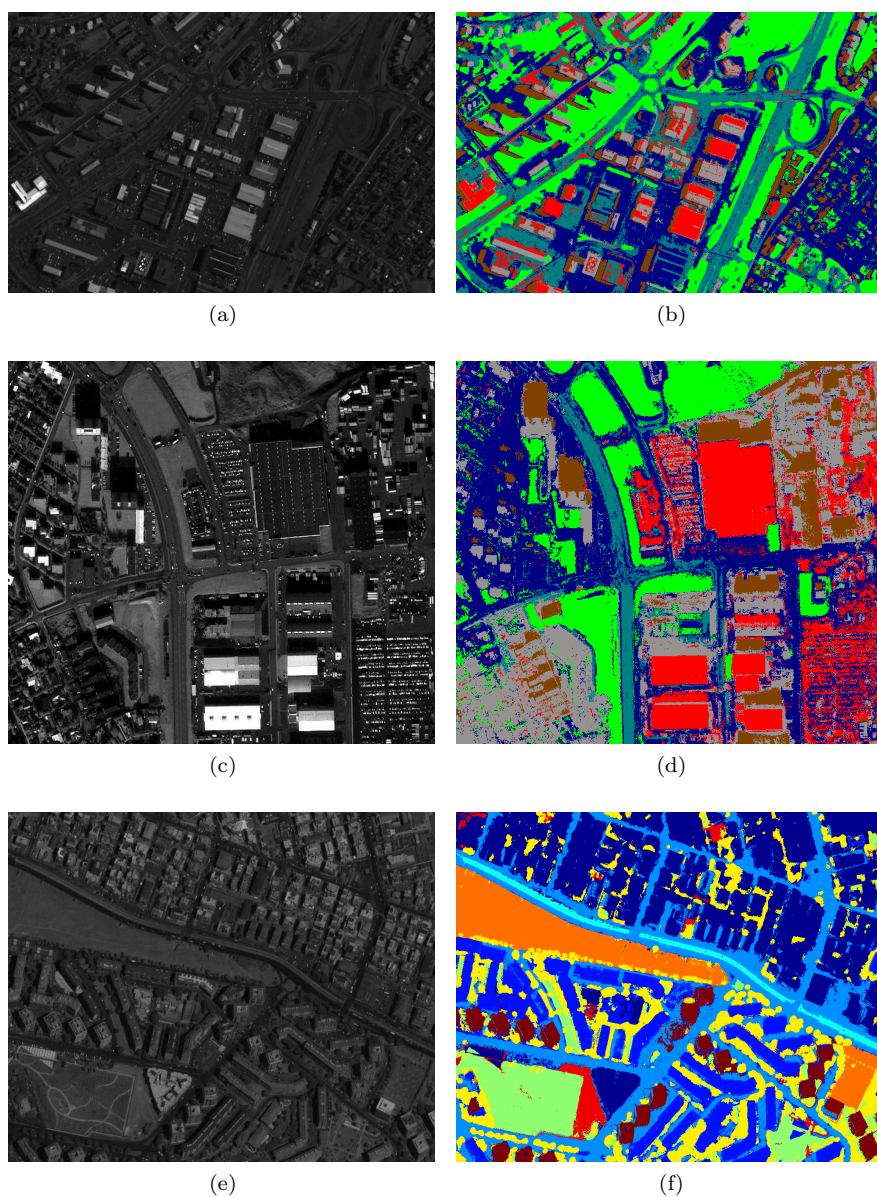


Figure 18.3: The panchromatic images and the classification maps obtained using all 4 bands for each dataset. The classification images were obtained using the configuration that yielded the highest accuracy. (a) Original IKONOS panchromatic image for dataset 1. (b) RF classification map for dataset 1 (OA: 68.73%). (c) Original IKONOS panchromatic image for dataset 2. (d) RF classification map for dataset 2 (OA: 89.46%). (e) Original QuickBird panchromatic image for dataset 3. (f) RF classification map for dataset 3 (OA: 94.79%).

Table 18.16: Classification accuracies (%) for all experiments using third dataset with three input channels (RGB)

Dataset3		RF		SVM											
Features		Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT
P															
	AA	26.76	26.76	26.76	26.76	26.76	26.76	26.76	AA	25.73	25.73	25.73	25.73	25.73	25.73
	OA	41.03	41.03	41.03	41.03	41.03	41.03	41.03	OA	40.56	40.56	40.56	40.56	40.56	40.56
RGB	K	27.77	27.77	27.77	27.77	27.77	27.77	27.77	K	27.11	27.11	27.11	27.11	27.11	27.11
	AA	59.59	57.44	60.92	58.02	58.02	59.23	60.47	AA	44.70	40.86	46.11	41.71	45.29	43.37
	OA	65.59	65.45	66.86	63.06	63.06	64.78	64.68	OA	56.84	56.71	58.92	53.44	56.13	54.59
PRGB	K	58.80	58.59	60.32	55.75	55.75	57.91	57.70	K	47.79	47.44	50.13	43.51	47.09	44.79
	AA	68.20	57.01	60.56	66.78	66.78	58.74	71.50	AA	52.58	41.00	46.09	51.55	45.33	56.63
	OA	73.81	64.64	66.04	72.48	72.48	64.04	76.72	OA	65.00	56.78	58.94	64.09	56.21	68.83
RGB+MP	K	68.65	57.69	59.41	67.04	67.04	57.07	72.15	K	57.60	47.51	50.15	56.44	47.16	62.32
	AA	92.67	91.10	90.83	92.02	92.02	89.68	93.34	AA	88.70	84.86	84.79	87.82	82.82	90.06
	OA	93.30	91.89	91.59	92.59	92.59	90.31	93.91	OA	89.11	86.48	86.19	88.57	83.73	90.54
RGB+DMP	K	92.04	90.37	90.01	91.19	91.19	88.49	92.77	K	87.06	83.89	83.56	86.41	80.61	88.76
	AA	92.82	91.34	91.08	92.33	92.33	90.02	93.78	AA	91.78	89.52	89.42	91.35	87.73	92.99
	OA	93.39	91.98	91.69	92.76	92.76	90.45	94.23	OA	91.92	89.99	89.76	91.46	87.84	93.21
	K	92.14	90.46	90.13	91.40	91.40	88.64	93.14	K	90.41	88.10	87.83	89.86	85.55	91.94

Table 18.17: Classification accuracies (%) for all experiments using third dataset with four input channels (RGBNIR)

Dataset:3		RF		SVM										
Features P	Accuracy		MBF	Brovey	DWT	PCA	UDWT	Accuracy		MBF	Brovey	DWT	PCA	UDWT
	AA	26.76	26.76	26.76	26.76	26.76	26.76	AA	26.76	26.76	26.76	26.76	26.76	26.76
	OA	41.03	41.03	41.03	41.03	41.03	41.03	OA	41.03	41.03	41.03	41.03	41.03	41.03
	K	27.77	27.77	27.77	27.77	27.77	27.77	K	27.77	27.77	27.77	27.77	27.77	27.77
RGB	AA	71.54	70.00	70.18	70.17	74.07	AA	60.58	56.58	57.01	57.02	61.30		
	OA	76.34	75.07	75.32	75.04	78.47	OA	68.81	67.71	68.19	67.14	71.15		
	K	71.71	70.17	70.48	70.17	74.25	K	62.43	60.91	61.53	60.40	65.19		
PRGBN	AA	73.59	69.90	72.17	70.08	75.46	AA	63.26	56.61	61.18	57.03	65.02		
	OA	77.92	74.99	76.84	74.94	79.69	OA	69.75	67.73	69.16	67.15	72.18		
	K	73.59	70.08	72.27	70.06	75.70	K	63.58	60.93	62.79	60.41	66.50		
RGBN+MP	AA	93.43	93.42	93.09	93.72	94.08	AA	90.37	89.71	89.97	90.15	91.08		
	OA	94.07	94.01	93.82	94.19	94.64	OA	90.51	89.94	90.26	90.27	91.45		
	K	92.96	92.89	92.67	93.10	93.64	K	88.73	88.05	88.42	88.44	89.85		
RGBN+DMP	AA	93.70	93.69	93.41	93.86	94.40	AA	93.21	92.88	92.85	93.07	93.86		
	OA	94.17	94.09	93.90	94.16	94.79	OA	93.19	92.85	92.97	93.06	93.99		
	K	93.08	92.98	92.75	93.07	93.81	K	91.92	91.52	91.66	91.76	92.86		

Conclusions

In this work we have attempted to assess the relative importance of the spectral and spatial quality for classification of pansharpened satellite urban imagery. We used two different classifiers, SVM and RF and we considered 6 different pansharpening methods with varying degrees of spectral and spatial quality.

Of those methods, there are two that can be said to represent the opposite ends of the quality spectrum. The MBF method produces images of very high spectral quality while the spatial quality is considerably lacking. On the other hand, the multi-resolution analysis based UDWT method produces images of spectral quality that is relatively low but have outstanding spatial quality, producing images of almost the same detail as the Pan image itself.

A number of experiments were done using three different datasets produced by two different satellite sensors, IKONOS and QuickBird. All the different experiments for a single dataset were done using pansharpened images of three channels (RGB) and four channels (RGB and NIR), respectively. The different experiments were classification of the Pan image alone and MS image alone, Pan image and MS image together, and, finally, of the Pan image together with the MS image and an MP or a DMP.

The results of the classification experiments show that the UDWT method gives best results in terms of accuracy in 22 out of 24 experiments using the RF as classifier and in 19 out of 24 experiments using the SVM classifier. The MBF method gives the best results in terms of accuracy in 1 experiments out of 24 using the RF classifier and in 4 out of 24 experiments using the SVM classifier. The RF classifier produced better results in terms of accuracy than the SVM method in almost every experiment, except for about a fourth of the experiments for the first dataset.

In general, the results obtained for each dataset showed the same patterns. In the case of the three-band data (RGB), classification of the pansharpened MS image alone, gave relatively similar accuracies for all the pansharpening methods. The difference between the best and the worst accuracies for the first dataset was 5.64%, 5.78% for the second dataset and 3.8% for the third dataset. The MBF data gave best accuracies for the first dataset, The UDWT data gave the best accuracies for the second dataset and surprisingly, for the QuickBird dataset, the IHS data gave the best accuracy for both type of classifiers in this experiment.

The gap in performance between data produced by the different pansharpening methods was greatly increased with the addition of the Pan image to the classification process. The difference between the best and the worst accuracies are now 12.5%, 13.4% and 12.08% for the first, second and third dataset, respectively. The UDWT data gave the best accuracies for this experiment for all datasets and both classifiers.

For the four-band data (RGB and NIR) there were substantial differences compared to the three-band case. The UDWT data gave the best accuracies in every experiment using the RF classifier. For the SVM classifier, the only exception were the MP/DMP experiments for the first dataset, where the MBF data gave best results.

The most notable things concerning the classification of the MS data alone using all the four bands, are how much higher accuracy the UDWT data gives compared to the rest of the methods, and secondly, how well the Brovey data perform compared to the more spectrally consistent methods, coming second place for the first two datasets and third in the last dataset. It has to be mentioned though, that the performance of all the data with the exception of the UDWT data, are very similar for this experiment.

Adding the Pan image to the classification using all four channels does not change much in terms of accuracy. For all datasets, the general behavior is that the MBF, DWT and UDWT data gain 1%-2% in accuracy while the other data usually give lower accuracy than for the previous experiment for the MS data only.

Including the MP/DMP in the classification increased the accuracy by a considerable amount for all datasets. The most notable increase in accuracy being for the third dataset, an average increase of 22.7% (using RGB data) compared to the PRGB experiment. The first dataset showed the least benefits from including the MP/DMP, the probable reason being how unrepresentative the training set was chosen. The three band data gained more accuracy than the four band data in the experiments using the MP/DMP, since the added spatial information was relatively less in the four band case.

The conclusion to draw here seems to be that while spectral consistency is important, spatial quality seems to be more important for classification purposes, given that the spectral consistency is acceptable. Here, "acceptable" means significantly better spectral consistency than that of the IHS and Brovey methods.

We also demonstrated that adding spatial information (Pan image, MP/DMP) to the classification of data with very good spatial quality resulted in relatively greater gains in accuracy than for data with good spectral quality but lacking somewhat in spatial quality. Often the classification accuracy actually decreased when including the Pan image in the classification, for the least spectrally consistent data. The conclusion being that if the data has bad spectral quality, adding spatial data will not help at all. This behavior was seen in every dataset when working with three band data.

The UDWT is the pansharpening method that produced the data achieving the highest classification accuracy in almost every experiment performed. The images produced by this method ranked third in terms of spectral quality for all three datasets, but in every case the spatial quality was excellent (as good as it can get). The method giving the best spectral quality, by far, did worse than expected in our experiments and we draw the conclusion that the reason for this, is the relatively low spatial quality of the fused images produced by that method.

Hence, the main conclusion is that while the spectral consistency and the spatial quality of pansharpened images tend to be complementary in nature, i.e., gains in one quality often compromise the other, for classification purposes, it seems better to aim for as good spatial quality as is possible, given that the spectral quality stays above some acceptable minimum.

Since all the datasets used the experiments were of urban areas, future work is to see if the above conclusions also hold for rural imagery.

Bibliography

- [1] Gary A Shaw and Hsiao-hua K Burke, "Spectral imaging for remote sensing," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 3–28, 2003.
- [2] 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.
- [3] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 3, pp. 1204 –1211, may 1999.
- [4] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 40, no. 10, pp. 2300 – 2312, oct 2002.
- [5] 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, 1992.
- [6] 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," *Photogrammetric Engineering and Remote Sensing*, vol. 56, no. 4, pp. 459–467, 1990.
- [7] P. Chavez and J. Bowell, "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.
- [8] K. Edwards and P. Davis, "The use of intensity-hue-saturation transformation for producing color shaded-relief images," *Photogrammetric Engineering and Remote Sensing*, vol. 60, no. 11, pp. 1369–1373, 1994.

- [9] 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," *Geoscience and Remote Sensing Letters, IEEE*, vol. 1, no. 4, pp. 309–312, oct. 2004.
- [10] 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.
- [11] 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.
- [12] L. Wald, *Data Fusion: Definitions and Architectures : Fusion of Images of Different Spatial Resolutions*, Les Presses de l'École des Mines, 2002.
- [13] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1970.
- [14] David R Hunter and Kenneth Lange, "Quantile regression via an mm algorithm," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 60–77, 2000.
- [15] 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.
- [16] F. Palsson, J.R. Sveinsson, M.O. Ulfarsson, and J.A. Benediktsson, "A new pansharpening method using an explicit image formation model regularized via total variation," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, 2012, pp. 2288–2291.
- [17] J.M. Bioucas-Dias, M.A.T. Figueiredo, and J.P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006.*, May 2006, vol. 2, p. II.
- [18] Thierry RANCHIN and Lucien WALD, "Fusion of high spatial and spectral resolution images: The arsis concept and its implementation," *Photogrammetric Engineering and Remote Sensing*, vol. 66, no. 1, pp. 49–61, 2000.
- [19] 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.
- [20] Zhou Wang and A.C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

- [21] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *Geoscience and Remote Sensing Letters, IEEE*, vol. 1, no. 4, pp. 313 – 317, Oct. 2004.
- [22] M.M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 11, pp. 3880 –3891, Nov. 2009.
- [23] Robert H. Yuhas, Alexander F. Goetz, and Joe W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," in *Summaries 3rd Annual JPL Airborne Geoscience Workshop*, 1992, vol. 1, pp. 147–149.
- [24] Chein-I Chang, "Spectral information divergence for hyperspectral image analysis," in *Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium, 1999.*, 1999, vol. 1, pp. 509–511.
- [25] Zhou Wang and A.C. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81 –84, march 2002.
- [26] 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.
- [27] Ingrid Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *Information Theory, IEEE Transactions on*, vol. 36, no. 5, pp. 961–1005, 1990.
- [28] Mark J Shensa, "The discrete wavelet transform: wedding the a trous and mallat algorithms," *Signal Processing, IEEE Transactions on*, vol. 40, no. 10, pp. 2464–2482, 1992.
- [29] Arthur L Da Cunha, Jianping Zhou, and Minh N Do, "The nonsubsampling contourlet transform: theory, design, and applications," *Image Processing, IEEE Transactions on*, vol. 15, no. 10, pp. 3089–3101, 2006.
- [30] Emmanuel Candes, Laurent Demanet, David Donoho, and Lexing Ying, "Fast discrete curvelet transforms," *Multiscale Modeling & Simulation*, vol. 5, no. 3, pp. 861–899, 2006.
- [31] Ivan W Selesnick and Mário AT Figueiredo, "Signal restoration with over-complete wavelet transforms: comparison of analysis and synthesis priors," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2009, pp. 74460D–74460D.
- [32] F. Palsson, J.R. Sveinsson, and M.O. Ulfarsson, "A new pansharpening algorithm based on total variation," Accepted for publication in *IEEE Geoscience and Remote Sensing Letters*, 2013.

- [33] 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.
- [34] Alfred M Bruckstein, David L Donoho, and Michael Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.
- [35] Michael Elad, Peyman Milanfar, and Ron Rubinstein, "Analysis versus synthesis in signal priors," *Inverse problems*, vol. 23, no. 3, pp. 947, 2007.
- [36] L.. Alparone, L.. Wald, J.. Chanussot, C.. Thomas, P.. Gamba, and L.M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 10, pp. 3012–3021, oct. 2007.
- [37] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.
- [38] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, 2003.
- [39] P. Soille, *Morphological Image Analysis, Principles and Applications*, Springer, 2nd edition, 2003.
- [40] M. Fauvel, J. A. Palmason, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Classification of remote sensing imagery with high spatial resolution," in *Image and Signal Processing for Remote Sensing XI 06 (SPIE)*, Bruges, Belgium, September 2005.
- [41] H. Aanaes, J.R. Sveinsson, A.A. Nielsen, T. Bovith, and J.A Benediktsson, "Model-based satellite image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1336–1346, 2008.
- [42] W. J. Emery F. Pacifici, M. Chini, "A neural network approach using multi-scale textural metrics from very high resolution panchromatic imagery for urban land-use classification," *Remote Sensing of Environment*, vol. 113, no. 6, pp. 1276–1292, June 2009.
- [43] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multi-spectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 5, pp. 1301–1312, may 2008.

- [44] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2042–2055, 2002.
- [45] G. Matheron, *Éléments pour une théorie des milieux poreux*, Masson, Paris, 1967.
- [46] Mathieu Fauvel, Jon Atli Benediktsson, Jocelyn Chanussot, and Johannes R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, Part 2, pp. 3804–3814, Nov. 2008.
- [47] L. S. Davis C. Huang and J. R. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of Remote Sensing*, vol. 23, no. 4, pp. 725–749, Feb. 2002.
- [48] G. M. Foody and A. Mathur, "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a svm," *Remote Sensing of Environment*, vol. 103, no. 2, pp. 179–189, Jul. 2006.
- [49] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
- [50] Leo Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [51] P.H. Swain J.A. Benediktsson, "Consensus theoretic classification methods," *IEEE Transactions on Systems, Man, and Cybernetics*, , no. 22, pp. 688–704, 2001.
- [52] P. Salamon L.K. Hansen, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 12, pp. 993–1001, 1990.
- [53] L.I. Kuncheva, "Fuzzy versus nonfuzzy in combining classifiers designed by boosting," *IEEE Transactions on Fuzzy Systems*, , no. 11, pp. 1214–1219, 1990.
- [54] R.E. Schapire Y. Freund, "Experiments with a new boosting algorithm," in *Machine Learning. Proceedings of the Thirteenth International Conference.*, 1996, pp. 148–156.
- [55] R.E. Schapire, "A brief introduction to boosting," in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence.*, 1999, pp. 1401–1406.
- [56] Leo Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, August 1996.

- [57] Leo Breiman, “Rf/tools—a class of two eyed algorithms,” in *SIAM Workshop*, 2003, <http://oz.berkeley.edu/users/breiman/siamtalk2003.pdf>.

Part IV

Papers

Conference Paper 1

F. Palsson, J.R. Sveinsson, J.A. Benediktsson, and H. Aanaes, “Image fusion for classification of high resolution images based on mathematical morphology,” in *2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2010, pp. 492–495.

IMAGE FUSION FOR CLASSIFICATION OF HIGH RESOLUTION IMAGES BASED ON MATHEMATICAL MORPHOLOGY

Frosti Palsson[†], Johannes R. Sveinsson[†], Jon Atli Benediktsson[†] and Henrik Aanaes[‡]

[†]University of Iceland

Faculty of Electrical and Computer Engineering
Hjardarhagi 2-6, 107 Reykjavik, Iceland
and

[‡]Department of Informatics and Mathematical Modelling
Technical University of Denmark, 2800 Lyngby, Denmark.

ABSTRACT

Classification of high resolution urban remote sensing imagery is addressed. The classification is done by both considering the panchromatic imagery and the multi-spectral image obtained using the spectrally consistent fusion method introduced in [1]. The data are classified using support vector machines (SVM). To further enhance the classification accuracy, mathematical morphology is used to derive local spatial information from the panchromatic data. In particular we use the Morphological Profile (MP) in classification of satellite imagery as was proposed in [2, 3]. We also use the derivative of the MP (DMP). In the majority of the image fusion (pansharpening) techniques proposed today, there is a compromise between the spatial enhancement and the spectral consistency. By comparing classification results obtained by using our model based scheme [1] to results obtained using the IHS and Brovey fusion methods, we find that spectrally consistent data give better results when it comes to classification.

Index Terms— Classification, Pansharpening, Morphological Profile

1. INTRODUCTION

Satellites provide very valuable data about the Earth, e.g., for environmental monitoring, weather forecasting, map-making and military intelligence. But satellites are expensive, both to build and operate. This implies that we are obliged to make the best use of the data obtained from existing satellites, e.g., by combining the output from different sensors. Good example of this is the merging of satellite images of low spatial and high spectral resolution with images of high spatial and low spectral resolution. This is also known as image fusion (pansharpening).

In this paper we work with pixel-level satellite image fusion derived directly from a model of the imaging sensor. This method has the nice feature of being spectrally consistent by design [1].

Since the problem of satellite image fusion is an ill-posed problem, regularization is necessary for its solution. For this purpose, the framework for pixel neighborhood regularization was presented in [1]. In [1], the low-resolution RGB image was only used in the image fusion. Here we add the near infrared (NIR) band of the low-resolution (spectral) image to the image fusion method in [1].

For our experiments, we use an IKONOS image of an urban area (Reykjavik, Iceland) with a multi-spectral image with four bands (R,B,G and NIR) and a panchromatic image which is of higher resolution (4 by 4 to 1 pixel) as compared to the multi-spectral image. The resulting fused image is then going to be used for classification.

The classification of high resolution urban remote sensing imagery is a challenging research problem. In this paper we look at the classification of such data by both considering the classification of panchromatic imagery (single data channel) and spectral image (multiple data channels) obtained by the extended fusion method in [1].

Panchromatic images are characterized by a high spatial resolution. This high spatial resolution allows the identification of small structures in a dense urban area. However, the analysis of a scene by considering the value of a pixel only will produce very poor classification results compared to the fine resolution. To solve this problem, some local spatial information is needed.

An interesting approach to provide local spatial information is based on the theory of *Mathematical Morphology* [4], which provides tools to analyze spatial relationship between pixels. The morphological profile (MP) was proposed in [2, 3] for segmentation of high-resolution satellite images. An MP is made up of an opening profile (OP) and a closing profile (CP) and we will also use the derivative of the morphological

This work was sponsored by the University of Iceland Research Fund.

profile (DMP). Only the panchromatic data are used to build the morphological profile and its derivative.

For classification purposes, the MP and the DMP are regarded as feature vectors, where each class has a typical MP-DMP. Hence each MP (or DMP) is considered as a channel of a multi-spectral image. This way, classification methods applied to multi-spectral images can be applied [5, 6]. The classification will be performed using support vector machines (SVM). We will also add the 4-channel fused image as new element in the feature vectors obtained using the MP-DMP on the IKONOS panchromatic image.

In the next section, spectrally consistent images fusion is discussed. Section 3 gives a brief review on mathematical morphology and Section 4 reviews SVM. Experimental set up and results are given in Sections 5 and 6, respectively and finally, conclusions are drawn in Section 7.

2. IMAGE FUSION

Image fusion can be done at several levels: pixel, feature, object and decision levels, for example. Here we will only be concerned with pixel level fusion. The image fusion method [1] we will be using is derived from a model of the imaging sensor's underlying physics. This approach ensures spectral consistency of the resulting fused image. This property is very important since spectral distortion can result in unreliable results for applications.

Spectral consistency can be measured as the cross correlation between the low resolution RGB image and the appropriately down-sampled fused image. In our case the low resolution spectral image consists of four bands: R,B,G and NIR.

Several pansharpening techniques have been proposed during the past two decades. In the majority of those techniques there is a compromise between the desired spatial enhancement and the spectral consistency. The intensity-hue-saturation (IHS) method has been widely used. Another popular method is the Brovey [5] sharpening method. Both suffer from spectral distortions.

In this paper we compare classification results obtained with our model-based, spectrally consistent scheme to results obtained by using IHS and Brovey image fusion.

3. MATHEMATICAL MORPHOLOGY

Mathematical Morphology (MM) is a theory that provides mathematical tools to analyze spatial relationship between pixels. It was originally introduced by Matheron and Serra in the 1960s to study binary porous media.

When working with MM, one is usually interested in specific objects in the image. These objects of interest are viewed as subsets of the image. After these objects have been identified, several sets of known size and shape (such as square, line or disk) can be used to characterize their morphology.

These simple sets are called *Structuring Elements* (SEs). An SE always has an origin (generally its symmetric center) that allows the positioning of the SE at a given pixel of the image.

The two most fundamental morphological operations are *erosion* and *dilation*. The eroded value at a given pixel x is the minimum pixel value over the SE. The dilated value at a given pixel x is the maximum pixel value over the SE.

Two of the most common operations of MM are the morphological *opening* and *closing* operations. The morphological opening of an image is an erosion followed by dilation while the morphological closing of an image is the reverse: dilation followed by erosion. Both operations are done using the same SE. It can be said that morphological opening of an image removes light features that are smaller than the SE whereas closing removes dark features smaller than the SE. While these operations are useful, they have the drawback of not being connected filters so they do not preserve shapes. So instead we will use the morphological reconstruction operations.

Opening by reconstruction of an image removes unconnected light features while *closing by reconstruction* of an image removes unconnected dark features. For both operations, shapes are preserved and structures present after transformation are of a size greater than or equal to the SE. This effectively means that less shape noise is generated.

Since objects in a given image may be of varying sizes, the use of a single SE might be prove to be inefficient. To solve this, a multiscale approach can be taken, where SEs of same shape but of different size are used.

A Morphological Profile (MP) is a $2n + 1$ -dimensional vector that consists of the input image, n openings and n closings (reconstruction operations) generated with an SE of fixed shape but of varying size. The MP can be defined as

$$MP(x) = [CP_n(x), \dots, I(x), \dots, OP_n(x)]$$

where CP and OP are closing profile and opening profile of n -dimension, constructed using an SE of fixed shape but of varying size. There should be n different sizes in total.

The *Derivative of the Morphological Profile* (DMP) is defined as a $2n$ -dimensional vector equal to the discrete derivative of the MP and is given by

$$DMP_i(x) = MP_{i-1}(x) - MP_i(x).$$

The information provided by the DMP is both spatial and radiometric. For a given pixel, the shape of the DMP can give an idea of the neighborhood of the pixel, e.g., whether it belongs to darker or lighter structure than the surrounding pixels. Finally, the amplitude of the DMP gives information about the local contrast of the structure.

4. SUPPORT VECTOR MACHINES

One of the most promising of recent developments in the field of machine-learning and pattern-recognition are Support Vector Machines (SVM). These are supervised learning methods that are widely used for classification and regression. When given a set of training data points (examples) where each data point is marked as belonging to one of two classes, the SVM algorithm builds a model that can predict to which class new data points will belong to.

The SVM model can be viewed as a representation of the data points as points in space where the separate categories or classes are divided by a gap that is as wide as possible. The SVM constructs a hyperplane in a space that has a high or infinite dimension. For good separation, the distance from the training data points of each class to the hyperplane should be maximized. When this distance or margin is maximized we call the resulting linear classifier, a maximum margin classifier.

For this linear classifier to be able to solve non-linear problems, the non-linear data points are mapped into a higher dimensional space in such a way that linear classification in this new space is equivalent to non-linear classification in the original space. This is what is known as the 'kernel trick'. By using non-linear kernel-functions, this approach can separate complex (e.g., multi-modal) class distributions in high dimensional feature spaces. A commonly used kernel-function is the Gaussian Radial Basis function and it is used for the experiments in this paper.

Since the SVM is in fact a binary classification strategy, some kind of multiclass decomposition is required to handle multiclass problems. A single multiclass problem is reduced into multiple binary classification problems. Two common methods are the One-Against-All (OAA) and One-Against-One (OAO). OAO is probably the widest used.

For a multiclass problem with k -classes, $k(k-1)/2$ binary SVMs are trained. For OAO, classification is done by max-wins voting strategy where every individual binary classifier assigns the instance to one of the two classes thereby increasing the vote count for that class by one.

Since SVM does not provide class-labels as classifier output but instead provides the distance from each data point to the hyperplane, the sign of the distance is used to determine a class. Finally, the instance classification is determined by the class with the most votes. A detailed discussion on SVM is given in [7].

5. DATASET AND EXPERIMENTAL SETUP

The data set used in this paper consists of two images of Reykjavik, Iceland. These images were acquired by the IKONOS Earth imaging satellite on 9th August 2001. One is a low-resolution (4m) multi-spectral image that consists of four bands R,G,B and NIR. The other is a high-resolution

panchromatic image of resolution 1m. The ratio between the panchromatic and low-resolution images is 16 or 4 by 4 pixels. These images were then fused by using the model based method described in [1], and the resulting high-resolution multi-spectral image was used for classification.

An MP was constructed, consisting of 3 openings by reconstruction and 3 closings by reconstruction where the SE was a disc of sequentially increasing radius of 3, 6 and 9 pixels. Finally the discrete derivative (DMP) of the MP was calculated and used.

Ground truth data for 6 classes (*Large buildings, Small buildings, Residential Lawns, Streets, Open Areas and Shadows*) was used for training and validation. Details about the training and validation sets can be seen in Table 1.

Table 1. Training and validation set details.

Class		Samples	
No.	Name	Train	Test
1	Large Buildings	1526	34155
2	Small Buildings	7536	43867
3	Residential Lawns	1286	25806
4	Streets	2797	30916
5	Open Areas	3336	39202
6	Shadows	5616	35147
Total		22097	209093

For the classification, the libSVM¹ library was used. It is a freeware and is regularly updated. The Gaussian Radial Basis function was chosen for the kernel and the training parameters were found using a grid-search.

6. EXPERIMENT RESULTS

Classification based on the panchromatic (PC) image only, gave a relatively low test accuracy of 46.23%. The test accuracy improved significantly when the R,G,B bands were added, improving to 58.78%. Further adding the NIR band did not change much in terms of accuracy, bringing the over all test accuracy to 60.32%.

To further increase the classification accuracy, a morphological profile (MP), or rather, a derivative of the morphological profile (DMP) was also added. MPs and DMPs of two different lengths were used in classification. The DMPs gave slightly better results than the MPs. The best results were achieved by using a DMP of length 6 with the SEs being disks of radius 3, 6 and 9 pixels. The same DMP was used in all experiments presented here. By adding the DMP the total number of features available for classification was increased by 6, resulting in the overall accuracy of 66.58% for the R,G,B bands along with the panchromatic image and

¹Chih-Chung Chang and Chih-Jen Lin - LIBSVM – A Library for Support Vector Machines - <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

Table 2. Training and test accuracies (%) for all experiments

Data	Panchromatic		3 bands+PC+DMP		4 bands+PC+DMP		IHSFUSE+PC+DMP		BROVERY+PC+DMP	
Features	1		10		11		10		10	
Class no	train	test	train	test	train	test	train	test	train	test
1	58.06	29.32	94.10	51.41	94.69	52.13	98.82	50.50	98.75	52.24
2	89.97	65.92	99.80	87.11	99.76	86.33	99.81	50.09	99.50	51.04
3	87.25	77.70	97.28	83.03	97.59	82.95	95.72	84.03	89.27	83.12
4	56.60	25.88	95.07	33.00	95.17	34.69	92.67	41.15	93.21	40.36
5	21.01	19.01	81.95	77.56	84.05	78.28	83.03	71.65	84.65	73.51
6	83.32	63.25	91.45	60.89	91.72	62.11	91.58	62.80	91.54	61.91
Ave	66.03	46.85	93.28	65.50	93.83	66.08	93.61	60.04	92.82	60.36
OA	71.28	46.23	93.85	66.58	94.29	67.11	93.98	59.20	93.79	59.66

67.11% by adding the NIR band. The gains in test accuracy are 7.80% and 6.79%, respectively.

These results clearly demonstrate the value of the local spatial information obtained by using the derivative of the morphological profile in the classification.

Finally the same experiments were repeated (without using the NIR band) using the multi-spectral image obtained from the IHS and Brovey fusion methods.

As can be seen from Table 2, the IHS and Brovey fusion methods did not give quite as good results for the classification as the spectrally consistent model based approach.

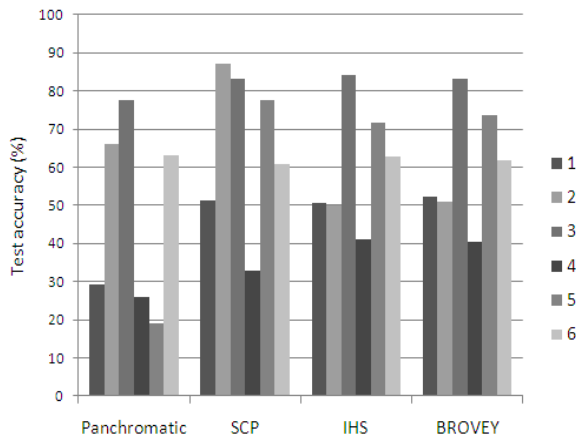


Fig. 1. The distribution of the test accuracies for all 6 classes. SCP stands for spectrally consistent pansharpening.

7. CONCLUSIONS

Three different methods of image fusion were used to produce high resolution multi-spectral images that were then classified using an SVM. Morphological profiles were added as feature vectors to aid in the classification process.

The model-based, spectrally consistent fusion method introduced in [1] was shown to give better results than the IHS and Brovey methods which are not spectrally consistent.

The use of the morphological profiles and their derivative to provide local spatial information resulted in improved classification accuracy.

8. REFERENCES

- [1] H. Aanaes, J.R. Sveinsson, A.A. Nielsen, T. Bovith, and J.A. Benediktsson, "Model-Based Satellite Image Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1336–1346, 2008.
- [2] P. Soille and M. Pesaresi, "Advances in Mathematical Morphology Applied to Geoscience and Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2042–2055, 2002.
- [3] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.
- [4] P. Soille, *Morphological Image Analysis, Principles and Applications*, Springer, 2nd edition, 2003.
- [5] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and Feature Extraction for Remote Sensing Images From Urban Areas Based on Morphological Transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, 2003.
- [6] M. Fauvel, J. A. Palmason, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Classification of remote sensing imagery with high spatial resolution," in *Image and Signal Processing for Remote Sensing XI 06 (SPIE)*, Bruges, Belgium, September 2005.
- [7] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.

Conference Paper 2

F. Palsson, J.R. Sveinsson, M.O. Ulfarsson, and J.A. Benediktsson, “A new pansharpening method using an explicit image formation model regularized via total variation,” in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012, pp. 2288–2291.

A NEW PANSHARPENING METHOD USING AN EXPLICIT IMAGE FORMATION MODEL REGULARIZED VIA TOTAL VARIATION

Frosti Palsson, Johannes R. Sveinsson, Magnus O. Ulfarsson and Jon A. Benediktsson

Faculty of Electrical and Computer Engineering
University of Iceland
Hjardarhagi 2-6. IS-107 Reykjavik, Iceland

ABSTRACT

In this paper we present a new method for the pansharpening of multi-spectral satellite imagery. This method is based on a simple explicit image formation model which leads to an ill posed problem that needs to be regularized for best results. We use both Tikhonov (ridge regression) and Total Variation (TV) regularization. We develop the solutions to these two problems and then we address the problem of selecting the optimal regularization parameter λ . We find the value of λ that minimizes Stein's unbiased risk estimate (SURE). For ridge regression this leads to an analytical expression for SURE while for the TV regularized solution we use Monte Carlo SURE where the estimate is obtained by stochastic means. Finally, we present experiment results where we use quality metrics to evaluate the spectral and spatial quality of the resulting pansharpened image.

Index Terms— Pansharpening, Total variation, SURE.

1. INTRODUCTION

Due to limitations of the imaging sensor found in satellites such as IKONOS and Quickbird, images of high spectral resolution have lower spatial resolution than images of low spectral resolution. Pansharpening is the process of fusing a multi-spectral (MS) image of low spatial resolution with a Panchromatic (PAN) image of high spatial resolution to obtain a MS image of high spatial resolution.

One area of application where the pansharpened image is useful is, i.e., classification, since the added spectral resolution increases the classification accuracy [1]. Pansharpening is in general an ill-posed problem that needs regularization for optimal results.

We begin by formulating an explicit image formation model as was given in [2] for the pansharpening process. While the formulation in [2] is based on dividing the image data into small patches and solving the problem for each patch since the matrices involved appear unfeasibly large, we

use the whole dataset. This involves very large matrices but since they are very sparse this is not a problem.

We use two different regularization methods, ridge regression (Tikhonov regularization) and Total Variation (TV). The solution to the ridge regression problem is easy to obtain analytically but solving the problem using TV regularization is more challenging since one has to use majorization-minimization (MM) techniques and in the end obtain the solution in an iterative manner.

There are many papers on TV based denoising in the literature and our work is based on the results in [3],[4]. The algorithm can be easily modified to use the ℓ_1 -norm for the data fidelity term, instead of the more commonly used ℓ_2 -norm. Our experiments indicate that the ℓ_1 -norm based solutions give better results.

To select the optimal value of the regularization parameter λ , we use Stein's unbiased risk estimate (SURE) [5],[6]. This makes it possible to calculate an estimate of the risk (or MSE) of our estimate of the pansharpened image.

For the ridge regression case, it is straightforward to obtain a closed form equation for SURE. However, for the TV case, it is more difficult to get a closed form solution because the problem is more complex and we are dealing with an iterative algorithm. Instead of trying to find the SURE estimate by analytical means, we chose to implement what is called Monte Carlo SURE (MCSURE) [7].

Calculation of SURE involves finding the trace of a very large matrix, which is essentially the divergence of the algorithm or operator. The idea of MCSURE is to estimate the diagonal of this matrix using stochastic methods. The algorithm is probed with additive noise and the difference or response signal is then used to estimate the divergence. This estimate of SURE is quite good as we will demonstrate, and does not incur great computational cost.

Finally, we present experimental results using both cases of regularization and choice of ℓ_1 - or ℓ_2 -norm as the data fidelity term. There are a number of metrics [1] available to estimate both the spatial and spectral quality of pansharpened imagery and we will use them to get some quantitative quality measurements on the results obtained.

This work was supported by the Research Fund of the University of Iceland.

2. IMAGE FORMATION MODEL

The data consists of a high spatial resolution panchromatic (PAN) image \mathbf{y}_{PAN} and a low spatial resolution multispectral (MS) image \mathbf{y}_{MS} . The PAN image has dimensions four times larger than the MS image thus the ratio in pixels is 1 to 16. The MS image contains 4 bands, RGB and near-infrared (NIR).

The PAN image is of dimension $M \times N$ and the MS image is of dimension $m \times n$ where $m = M/4$ and $n = N/4$. The vectors \mathbf{y}_{PAN} and \mathbf{y}_{MS} are the vectorized PAN and MS images, respectively and \mathbf{y}_{MS} is arranged as $\mathbf{y}_{\text{MS}} = \{\underbrace{1, \dots, mn}_{\text{R}}, \underbrace{1, \dots, mn}_{\text{G}}, \underbrace{1, \dots, mn}_{\text{B}}, \underbrace{1, \dots, mn}_{\text{NIR}}\}$ where mn is the number of pixels in each band. The vectorized fused image \mathbf{x} has this same arrangement where there are MN pixels in each band.

The image formation model is based on the results in [2] but as stated before we do not divide the problem into small patches, but instead work with the whole image, so very large sparse matrices as involved.

There are two key observations that define the model. The first observation is that the low resolution MS image can be described as a degradation (downsampling) of the fused image we seek. We write this in matrix notation as $\mathbf{y}_{\text{MS}} = \mathbf{M}_1 \mathbf{x} + \epsilon$ where \mathbf{M}_1 is a decimation matrix of size $4mn \times 4NM$ and ϵ is a Gaussian noise.

The second observation is that the PAN image is a linear combination of the bands of the fused image we seek with some additive Gaussian noise. This can be written in matrix notation as $\mathbf{y}_{\text{PAN}} = \mathbf{M}_2 \mathbf{x} + \epsilon$ where ϵ is a Gaussian noise and $\mathbf{M}_2 = \{\frac{1}{4}\mathbf{I}, \frac{1}{4}\mathbf{I}, \frac{1}{4}\mathbf{I}, \frac{1}{4}\mathbf{I}\}$ where \mathbf{I} is an identity matrix of size MN by MN .

\mathbf{M}_1 and \mathbf{M}_2 have the same number of columns and we can combine the expressions for \mathbf{y}_{MS} and \mathbf{y}_{PAN} into a single equation

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \epsilon, \quad (1)$$

where $\mathbf{y} = \begin{pmatrix} \mathbf{y}_{\text{MS}} \\ \mathbf{y}_{\text{PAN}} \end{pmatrix}$ and $\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{pmatrix}$ and ϵ is a Gaussian noise. The dimension of \mathbf{M} is $4mn + MN$ by $4MN$ and this is clearly an under determined system since there are more columns than rows and thus some kind of regularization is needed.

3. REGULARIZATION AND SELECTION OF OPTIMAL REGULARIZATION PARAMETER

3.1. Tikhonov Regularization

Here we seek the solution of the minimization problem

$$\arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda \|\mathbf{D}\mathbf{x}\|_2^2, \quad (2)$$

where \mathbf{D} is the Tikhonov matrix and λ is the regularization parameter. This matrix is often chosen as the identity matrix

but here it is chosen to be a difference operator so that it will enforce a smooth solution. This matrix \mathbf{D} is the same matrix as we will be using to calculate the TV and it is defined as $\mathbf{D} = \begin{pmatrix} \mathbf{D}_h \\ \mathbf{D}_v \end{pmatrix}$ where the matrix \mathbf{D}_h is defined such that when it is multiplied by a vectorized image it gives the first order differences in the horizontal direction and similarly the matrix \mathbf{D}_v gives the first order differences in the vertical dimension.

It is easy to show that the solution to (2) is given by

$$\hat{\mathbf{x}} = (\mathbf{M}^T \mathbf{M} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{M}^T \mathbf{y}. \quad (3)$$

To find the optimal value of the regularization parameter λ we are going to use SURE which we denote by R_λ to emphasize that it is a function of λ . It is given by

$$R_\lambda = \|\hat{\mathbf{y}}_\lambda - \mathbf{y}\|^2 + 2\sigma^2 \text{tr} \left(\frac{\partial \hat{\mathbf{x}}_\lambda}{\partial \mathbf{y}^T} \right), \quad (4)$$

where $\hat{\mathbf{y}}_\lambda = \mathbf{M}\hat{\mathbf{x}}_\lambda$ and $\hat{\mathbf{x}}_\lambda$ is the estimate of the fused image given by (3) and $\text{tr}(\cdot)$ is the trace operator. Now that we have a closed form expression for $\hat{\mathbf{x}}_\lambda$ it is easy to show that $\frac{\partial \hat{\mathbf{x}}_\lambda}{\partial \mathbf{y}^T} = \mathbf{M}(\mathbf{M}^T \mathbf{M} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{M}^T$ and we can write

$$R_\lambda = \|\mathbf{M}\hat{\mathbf{x}}_\lambda - \mathbf{y}\|^2 + 2\sigma^2 \text{tr}(\mathbf{M}(\mathbf{M}^T \mathbf{M} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{M}^T). \quad (5)$$

Unfortunately the inverse of the matrix $\mathbf{M}^T \mathbf{M} + \lambda \mathbf{D}^T \mathbf{D}$ is not sparse. Even though we have a closed form expression for the SURE estimate, it is going to be impossible to calculate it by using (5). Instead, we will use MCSURE to estimate the diagonal of the matrix $\mathbf{M}(\mathbf{M}^T \mathbf{M} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{M}^T$.

3.2. Total variation (TV) regularization

TV is a commonly used regularization in image denoising. By penalizing solutions with high total variation we are forcing the estimate to be smooth without necessarily penalizing sharp boundaries since the estimated boundary or edge is decided by the observed data. The solution we present here is based on the work described in [3],[4].

The TV of an image can defined as

$$\text{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h \mathbf{x})^2 + (\Delta_i^v \mathbf{x})^2}, \quad (6)$$

where $\Delta_i^h \mathbf{x} = x_i - x_{j_i}$ is the first order horizontal difference of pixel i and $\Delta_i^v \mathbf{x} = x_i - x_{k_i}$ is the first order vertical difference of pixel i . This can be written more compactly in matrix notation as

$$\text{TV}(\mathbf{x}) = \|\sqrt{(\mathbf{D}_h \mathbf{x})^2 + (\mathbf{D}_v \mathbf{x})^2}\|_1, \quad (7)$$

where \mathbf{x} is the vectorized image and the matrices \mathbf{D}_h and \mathbf{D}_v are defined exactly as in the Tikhonov case. Now we need to solve the optimization problem formulated as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda \text{TV}(\mathbf{x}), \quad (8)$$

where $\text{TV}(\mathbf{x})$ is given in (7).

This is a difficult problem to solve directly, because the $\text{TV}(\mathbf{x})$ functional is not differentiable and the above objective function is not necessarily convex. In order to be able to solve this problem we use majorization-minimization (MM) techniques to replace this difficult problem with a sequence of easier ones. Details on how the majorizing function is determined and the solution is found can be found in [3] and [4].

The solution to this problem is given by the simple iteration

$$\mathbf{x}^{k+1} = (\mathbf{M}^T \mathbf{M} + \lambda \mathbf{D}^T \mathbf{W}^k \mathbf{D})^{-1} \mathbf{M}^T \mathbf{y}, \quad (9)$$

where

$$\mathbf{W}^k = \frac{1}{2\sqrt{(\mathbf{D}_h \mathbf{x}^k)^2 + (\mathbf{D}_v \mathbf{x}^k)^2}}. \quad (10)$$

To replace the ℓ_2 -norm of the data fidelity term in (8) with the ℓ_1 -norm we need to introduce a weighting term \mathbf{W}_F for the data fidelity term. This changes the iteration equation (9) to

$$\mathbf{x}^{k+1} = (\mathbf{M}^T \mathbf{W}_F^k \mathbf{M} + \lambda \mathbf{D}^T \mathbf{W}^k \mathbf{D})^{-1} \mathbf{M}^T \mathbf{W}_F^k \mathbf{y}, \quad (11)$$

where the weights \mathbf{W}_F^k are given by

$$\mathbf{W}_F^k = g(|\mathbf{y} - \mathbf{x}^k|) \quad (12)$$

and the function g is

$$g(x) = \begin{cases} |x|^{-1} & \text{if } |x| > \epsilon_F \\ \epsilon_F^{-1} & \text{if } |x| \leq \epsilon_F, \end{cases} \quad (13)$$

and ϵ_F is a constant chosen to be very small.

3.3. Monte Carlo SURE

The last term in (4) is sometimes referred to as the divergence term, because it is the divergence of our operator. To estimate this term, we adopt the stochastic method presented in [7]. The central idea is that if you have an operator denoted by $f_\lambda(\mathbf{y})$ that is continuous and bounded and we also require that

$$\text{div}_y \{f_\lambda(\mathbf{y})\} = \sum_{k=1}^N \frac{\partial f_\lambda(\mathbf{y})}{\partial y_k} \quad (14)$$

then we can estimate the divergence using the following algorithm.

1. For $\lambda = \lambda_0$, evaluate $f_\lambda(\mathbf{y})$.
2. Perturb $\mathbf{z} = \mathbf{y} + \delta \mathbf{b}_T$, where \mathbf{b}_T is a unit variance random vector and δ is a constant chosen sufficiently small.
3. Evaluate $f_\lambda(\mathbf{z})$ for $\lambda = \lambda_0$.
4. Compute $\text{div}_y \{f_\lambda(\mathbf{y})\} = \frac{1}{\delta} \mathbf{b}_T^T (f_\lambda(\mathbf{z}) - f_\lambda(\mathbf{y}))$.

Then we can calculate SURE using

$$R_\lambda = \|\hat{\mathbf{y}}_\lambda - \mathbf{y}\|^2 + 2\sigma^2 \text{tr}(\text{div}_y \{f_\lambda(\mathbf{y})\}), \quad (15)$$

where the parameter σ^2 is estimated as the variance of the difference of our estimate $\hat{\mathbf{y}} = \mathbf{M}\hat{\mathbf{x}}$ and the data \mathbf{y} .

4. EXPERIMENT RESULTS

In our experiments we used an image of Reykjavik, Iceland, acquired by the IKONOS satellite on 9th of August 2001. This dataset consists of a panchromatic image of dimension 976 by 640 pixels and a multispectral image of dimension 244 by 160 pixels. The MS image has four bands, RGB and NIR.

The optimal value of the regularization parameter λ was found by using MCSURE to estimate the MSE for a range of values of λ and selecting the value that minimized the MCSURE estimate. Once the optimal value of λ was found for each type of regularization, we proceeded to pansharpen the data with the chosen values. Figure 1 shows the pansharpened results for a subset of the final image, along with the PAN and MS images for the same area.

Table 1 shows several spectral and a spatial quality metrics evaluated for the proposed method and for several other methods such as the Brovey, PCA and two wavelet methods. Details on these methods and experiments are given in [1]. All these metric are for spectral quality except spatial which measures spatial properties. Lower numbers are better for the bandCorrs, ERGAS, RASE, RMSE, SAM and SID metrics, while values close to 1 are best for the QAVE and spatial metrics.

The proposed method gives good results in comparison to the other methods. It gives good spectral consistency coupled with good spatial detail of 96.2% for the TV regularized image and 98.2% for the Tikhonov regularized image. We also see that though the wavelet method based on the 2D-DWT does better spectrally than the proposed method for some metrics, the spatial fidelity is very low at 85%.

Table 1 suggests that the results obtained using Tikhonov regularization have higher spatial metric than the TV results and that the spectral quality is slightly worse than the TV results gives. Here it has to be kept in mind the the Tikhonov regularization blurs the resulting image a little and for larger values of λ , the Tikhonov results are better spectrally and worse spatially than the TV results. We were able to get better spectral quality than is presented here, but at the expense of some spatial detail.

The different metrics were computed by comparing the original MS image with the pansharpened imaged by upsampling the low spatial resolution MS image to make it the same size as the pansharpened one. Some authors [2] downsample the PAN and MS images to produce a fused image of same dimension as the original MS image. By using downsampling we feel that too much information was lost compared to using upsampling as we did when the metrics were calculated.



Fig. 1. Experiment results using optimal value of λ . The first image is a 200 by 200 pixel subset of the PAN image. The second image shows the MS image corresponding to the same area, the third image shows the pansharpened image using TV regularization ($\lambda = 0.063$) and the last image shows the pansharpened image using the Tikhonov regularization ($\lambda = 0.0169$).

Table 1. Various spectral and spatial metrics. The proposed methods are denoted by the TV and Tikh. (Tikhonov) columns.

Metric	TV	Tikh.	Wavelet	PCA	CWT	Brovey
bandCorrs	0.04	0.04	0.03	0.23	0.06	0.03
ERGAS	5.94	6.42	5.64	8.21	6.40	7.43
QAVE	0.95	0.95	0.98	0.95	0.86	0.95
RASE	21.98	23.71	22.11	31.76	25.31	27.58
RMSE	0.03	0.04	0.03	0.05	0.04	0.04
SAM	2.31	2.46	1.41	2.46	4.41	NA
SID	0.002	0.004	0.02	0.00	0.01	0.00
spatial	0.96	0.98	0.85	0.81	0.997	0.92

5. CONCLUSIONS

In this paper we have proposed a method for pansharpening images based on both Tikhonov and TV regularization. We have also shown that the optimal regularization parameter can be found via Monte Carlo SURE. The experiment results indicate that the proposed method produces images with excellent spatial detail and spectral quality.

The optimal value of the regularization parameter λ chosen by MCSURE gives results that minimize the MSE. Low values of λ produce results with up to 99% of the spatial detail found in the PAN image while higher values produce results of better spectral quality but lower spatial detail. There is a tradeoff between desired spatial and spectral quality.

In [1] we evaluated several different pansharpening methods using classification accuracy. The results indicate that the spectral quality is not very important when it comes to classification if it is adequate to begin with. Spatial fidelity is very important in classification and hence the results presented here are very good.

6. REFERENCES

- [1] 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.
- [2] 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.
- [3] P. Rodriguez and B. Wohlberg, "An iteratively reweighted norm algorithm for total variation regularization," in *Proceedings of the Fortieth Asilomar Conference on Signals, Systems and Computers, ACSSC '06.*, 11 2006, pp. 892–896.
- [4] J.M. Bioucas-Dias, M.A.T. Figueiredo, and J.P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006.*, May 2006, vol. 2, p. II.
- [5] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.
- [6] V. Solo, "A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems," in *Proceedings of the International Conference on Image Processing, 1996.*, Sep 1996, vol. 3, pp. 89–92 vol.3.
- [7] S. Ramani, T. Blu, and M. Unser, "Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1540–1554, Sept. 2008.

Conference Paper 3

F. Palsson, J.R. Sveinsson, M.O. Ulfarsson, and J.A. Benediktsson, “SAR Image Denoising Using Total Variation Based Regularization with SURE-based Optimization of the Regularization Parameter,” in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012, pp. 2160–2163.

SAR IMAGE DENOISING USING TOTAL VARIATION BASED REGULARIZATION WITH SURE-BASED OPTIMIZATION OF THE REGULARIZATION PARAMETER

Frosti Palsson, Johannes R. Sveinsson, Magnus O. Ulfarsson and Jon A. Benediktsson

Faculty of Electrical and Computer Engineering
University of Iceland
Hjardarhagi 2-6. IS-107 Reykjavik, Iceland

ABSTRACT

Images obtained using Synthetic Aperture Radar (SAR) are corrupted by speckle. Speckle noise results from the chaotic interference of backscattered electromagnetic waves and makes the analysis, interpretation and classification of SAR images difficult. In this paper, we present a denoising algorithm based on Total Variation (TV) regularization. While this kind of denoising algorithm is not new, we propose to select the regularization parameter by minimizing the estimate of the mean square error (MSE) between the denoised image and the clean image. We do not have to know the clean image because we use a statistically unbiased MSE estimate - Stein's Unbiased Risk Estimate (SURE), that depends on the observed image and the estimated image. However, since it is difficult to derive SURE analytically for this kind of problem, we estimate SURE using stochastic methods. We present results using both a simulated image and real SAR image.

Index Terms— SAR, TV, denoising, speckle, SURE

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is an important technique for obtaining images of the Earth's surface, especially because of its ability to operate under all weather conditions. However, the images obtained are degraded by speckle which makes them difficult to analyze, interpret and classify.

Speckle is a chaotic phenomenon resulting from the interference of backscattered signals. Speckle is not the result of a stochastic process, but it can typically be modeled as multiplicative i.i.d. Rayleigh noise [1]. A logarithmic transformation changes the multiplicative noise model to an additive noise model.

Since the speckle obscures the scene content of the SAR image, some kind of speckle removal is necessary before further processing of the image. Many speckle suppressing methods have been proposed. A commonly used method is the Lee filter which is based on local variance statistics. Another important class of methods is based on the concept of wavelet coefficient thresholding [2].

In this paper we propose a speckle suppressing method which is based on Total Variation (TV) denoising. This leads to an optimization problem regularized by the TV of the image. Using TV for despeckling of SAR images is not new, but we propose to select the optimal value of the regularization parameter by minimizing an estimate of the mean square error (MSE) between the true noise free image and the estimated image. We don't have access to the true underlying image, but we have at our disposal, a statistically unbiased estimate of the MSE, Stein's unbiased risk estimate (SURE) [3], which depends on the noisy image and the estimated image. To find the optimal regularization weight λ , we compute SURE for a range of values of λ and select the value of λ that minimizes the SURE estimate.

Due to difficulties deriving a formula for calculating SURE analytically, we use a method called Monte-Carlo SURE (MCSURE) [4] to estimate the SURE by stochastic means.

The idea of MCSURE is to estimate the diagonal of this matrix using stochastic methods. The algorithm is probed with additive noise and the difference or response signal is then used to estimate the divergence. This estimate of SURE is quite good as we will demonstrate, and does not incur great computational cost.

Finally, we present experiment results for both a simulated SAR image and a real SAR image. Our experiments show that the MCSURE method gives a very good estimate of the true MSE and that the TV based denoising performs well compared to existing methods.

2. TOTAL VARIATION DENOISING

For an SAR image that has been digitized, let \mathbf{y} denote the observed intensity of the image and we can write the pixel level of the image as

$$\mathbf{y} = \mathbf{x}\mathbf{e}, \quad (1)$$

where \mathbf{x} is the desired image and \mathbf{e} is the Rayleigh i.i.d. multiplicative noise. By using the logarithm transformation we can convert this into an additive model

$$\tilde{\mathbf{y}} = \tilde{\mathbf{x}} + \tilde{\mathbf{e}}, \quad (2)$$

where $\tilde{\mathbf{y}} = \ln(|\mathbf{y}|)$ and $\tilde{\mathbf{e}}$ is approximately Gaussian additive noise. Now that we have an additive model we can proceed to formulate the denoising problem as

$$\hat{\mathbf{x}} = \arg \min_{\tilde{\mathbf{x}}} \frac{1}{p} \|\tilde{\mathbf{y}} - \tilde{\mathbf{x}}\|_p^2 + \lambda \text{TV}(\tilde{\mathbf{x}}), \quad (3)$$

where $\text{TV}(\mathbf{x})$ is the TV of the image and p is either 1 for the ℓ_1 -norm or $p = 2$ for the ℓ_2 -norm and λ is the regularization parameter. It controls the relative weight of the data fidelity and regularization terms.

This is a commonly used regularization in image denoising. The TV term in (3) penalizes solution with large total variation and hence the solution is forced to be smooth. This does not mean that sharp boundaries or edges in the image are heavily penalized since the estimated edge is decided by the observed data. Also, smooth edges have the same total variation as sharp edges. TV is defined as the sum of all first order differences in the image in the horizontal and vertical directions and thus it can be formulated as

$$\text{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h \mathbf{x})^2 + (\Delta_i^v \mathbf{x})^2}, \quad (4)$$

where $\Delta_i^h \mathbf{x} = x_i - x_{j_i}$ is the first order horizontal difference of pixel i and $\Delta_i^v \mathbf{x} = x_i - x_{k_i}$ is the first order vertical difference of pixel i . This can be written more compactly in matrix notation as

$$\text{TV}(\mathbf{x}) = \left\| \sqrt{(\mathbf{D}_h \mathbf{x})^2 + (\mathbf{D}_v \mathbf{x})^2} \right\|_1, \quad (5)$$

where the matrix \mathbf{D}_h is defined such that when it is multiplied by a vectorized image it gives the first order differences in the horizontal direction and similarly the matrix \mathbf{D}_v gives the first order differences in the vertical dimension. The solution to (3) is well known and the solution we present here is based on the work described in [5],[6].

2.1. ℓ_2 -norm solution

We begin by finding the solution to (3) with the ℓ_2 -norm as the data fidelity term. Immediately one can see that this is a difficult problem to solve using conventional means. Firstly, the TV functional is not differentiable and secondly, the objective function does not necessarily have to be convex. So instead of trying to solve (3) directly, we use Majorization-Minimization (MM) [5],[6] techniques and minimize a majorizer function of (3). Without going into details on exactly how the solution to (3) is found, when $p = 2$, it is given as the iteration

$$\tilde{\mathbf{x}}^{k+1} = (\mathbf{I} + \mathbf{D}^T \mathbf{W}^k \mathbf{D})^{-1} \tilde{\mathbf{y}}, \quad (6)$$

where

$$\mathbf{W}^k = \frac{\lambda}{2\sqrt{(\mathbf{D}_h \tilde{\mathbf{x}}^k)^2 + (\mathbf{D}_v \tilde{\mathbf{x}}^k)^2}}. \quad (7)$$

and $\mathbf{D} = \begin{pmatrix} \mathbf{D}_h \\ \mathbf{D}_v \end{pmatrix}$. It must be noted that when implementing this iterative algorithm it is necessary to use some means to avoid dividing by zero in (7). This can be done by adding a small constant to the square root in (7) or by using the method described in [5].

2.2. ℓ_1 -norm solution

By replacing the ℓ_2 -norm on the data fidelity term with the ℓ_1 -norm, the algorithm becomes more robust with respect to outliers. Since speckle is impulsive in nature, using the ℓ_1 -norm is justified. To solve (3) with the ℓ_1 -norm on the data fidelity term we need to use a technique similar to Iteratively Reweighted Least Squares (IRLS) where instead of solving

$$\|\mathbf{y} - \mathbf{x}\|_1, \quad (8)$$

we replace the ℓ_1 -norm by a weighted ℓ_2 -norm problem and solve it in an iterative manner [5]. The data fidelity term in (3) is replaced by

$$\frac{1}{2} \|\mathbf{W}_F^{1/2} (\tilde{\mathbf{y}} - \tilde{\mathbf{x}})\|_2^2, \quad (9)$$

and the iterative solution (6) is now replaced by

$$\tilde{\mathbf{x}}^{k+1} = (\mathbf{W}_F^k + \lambda \mathbf{D}^T \mathbf{W}^k \mathbf{D})^{-1} \mathbf{W}_F^k \tilde{\mathbf{y}}, \quad (10)$$

where the weights \mathbf{W}_F^k are given by

$$\mathbf{W}_F^k = g(|\tilde{\mathbf{y}} - \tilde{\mathbf{x}}^k|) \quad (11)$$

and the function g is

$$g(x) = \begin{cases} |x|^{-1} & \text{if } |x| > \epsilon_F \\ \epsilon_F^{-1}, & \text{if } |x| \leq \epsilon_F, \end{cases} \quad (12)$$

where ϵ_F is a constant chosen to be small.

3. STEIN'S UNBIASED RISK ESTIMATE (SURE) AND MONTE-CARLO SURE

To find the value of λ that minimizes the MSE between the true desired texture information we seek $\tilde{\mathbf{x}}$ and our estimate $\hat{\mathbf{x}}$ we do not need to know $\tilde{\mathbf{x}}$. Instead we will compute a statistically unbiased estimate of the MSE denoted by R_λ that is given by

$$R_\lambda = \|\tilde{\mathbf{y}} - \hat{\mathbf{x}}\|^2 + 2\sigma^2 \text{tr} \left(\frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}^T} \right), \quad (13)$$

where $\tilde{\mathbf{y}}$ is our log transformed observed data and $\hat{\mathbf{x}}$ is our estimate. This remarkable result is called Stein's unbiased risk estimate (SURE). The last term of (13), which is also called the divergence term because it is the divergence of our denoising operator, is going to be problematic. This is because we

do not have a closed form expression for $\hat{\mathbf{x}}$ for the ℓ_1 -norm solution. Also, the analytical solution for TV denoising using the ℓ_2 norm for 1D signals was derived in [7]. Using that solution we know that computing SURE will involve finding the diagonal of the inverse of a matrix whose number of elements is the number of pixels in the image, squared.

To overcome this, we will use an estimate of SURE called Monte-Carlo SURE (MCSURE) which is based on estimating the diagonal of our operator using a stochastic method. The central idea is that if you have an operator denoted by $f_\lambda(\mathbf{y})$ that is continuous and bounded and we also require that

$$\text{div}_y \{f_\lambda(\mathbf{y})\} = \sum_{k=1}^N \frac{\partial f_\lambda(\mathbf{y})}{\partial y_k} \quad (14)$$

then we can estimate the divergence using the following algorithm.

1. For $\lambda = \lambda_0$, evaluate $f_\lambda(\tilde{\mathbf{y}})$.
2. Perturb $\mathbf{z} = \tilde{\mathbf{y}} + \delta \mathbf{b}_T$, where \mathbf{b}_T is a unit variance random vector and δ is a constant chosen sufficiently small.
3. Evaluate $f_\lambda(\mathbf{z})$ for $\lambda = \lambda_0$.
4. Compute $\text{div}_y \{f_\lambda(\tilde{\mathbf{y}})\} = \frac{1}{\delta} \mathbf{b}_T^T (f_\lambda(\mathbf{z}) - f_\lambda(\tilde{\mathbf{y}}))$.

Then we can calculate SURE using

$$R_\lambda = \|\hat{\mathbf{y}}_\lambda - \tilde{\mathbf{y}}\|^2 + 2\sigma^2 \text{tr}(\text{div}_y \{f_\lambda(\tilde{\mathbf{y}})\}), \quad (15)$$

where the parameter σ^2 is the variance of the noise and remains to be estimated.

4. EXPERIMENT RESULTS

In our experiments we used both a simulated SAR image and a real SAR image. The simulated image is based on the well known Lena image. To this image we added multi-look [8] Rayleigh distributed noise. We did experiments for several different numbers of looks, namely, 1, 2, 4, 9 and 25. The Lena image is 512 by 512 pixels and it was scaled such that the image intensity was in the range 0 to 1.

The real image is a 256×256 , 8-bit single-look ERS-1 SAR image. It shows a portion of the glacier Kotlujokull in the southern part of Iceland. The image is from the C-band with VV-polarization.

For the simulated image, we began by finding the optimal λ for each kind of norm and each number of looks using MCSURE. Once the value was found that minimized the MCSURE estimate, the image was denoised using this value. The experiment was repeated 10 times and the resulting PSNR is the average. The results for the simulated image are shown in Table 1.

As previously stated, we used a log transform on the images prior to processing. After denoising, we used the exponential function to obtain the final image. We did the experiments with both the ℓ_1 -norm and the ℓ_2 -norm as the data fidelity term. We found that when using the ℓ_1 -norm, the MCSURE estimate was very close to the real MSE for all values of λ used and for all different number of looks.

When using the ℓ_2 -norm, the optimal results were achieved using relatively smaller values of λ than was required when using ℓ_1 -norm. For small values of λ and large values of noise (1 and 2 looks) the MCSURE estimate diverges somewhat from the true MSE. Due to this, in experiments using few looks (1 and 2) the value of λ that minimized the MCSURE estimate was not the same as the value that minimized the true MSE when using the ℓ_2 -norm. However, the corresponding maximal deviation of the Peak Signal to Noise Ratio (PSNR) is equal to 2% which is a relatively small error.

The results also indicate that better results are achieved using ℓ_1 -norm when the noise is heavy but when the noise is less, i.e., the number of looks is more than 4, the ℓ_2 norm gives better results. This is not unexpected since the noise is impulsive in nature and the ℓ_1 -norm is more robust to outliers. A plot of the MSE and the MCSURE estimate for both norms and 2-looks is shown in Figure 2.

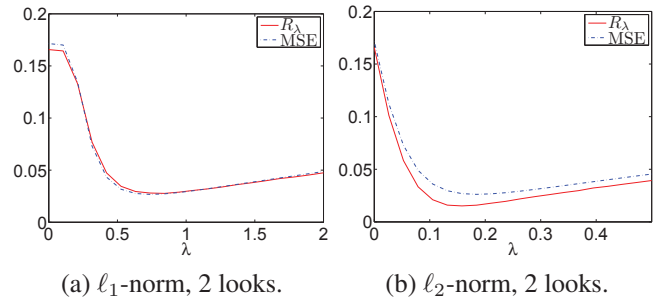


Fig. 2. The MCSURE estimate is close to the true MSE. For more than 2 looks, the two curves become difficult to distinguish on the plot.

Table 1. PSNR (in dB) for noisy and denoised simulated image

Looks	1	2	4	9	25
Noise	11.65	14.68	17.68	21.20	25.64
ℓ_1 -TV	22.47	25.26	26.58	28.19	28.06
ℓ_2 -TV	20.07	24.34	27.46	29.56	32.20

From Table 1 it is seen that the ℓ_1 -norm results are better in terms of the PSNR for 1 and 2 looks while the ℓ_2 -norm gives better results for the remaining number of looks. The results shown here can be directly compared to Table 1 in [2], where the same simulated image was used and the number of

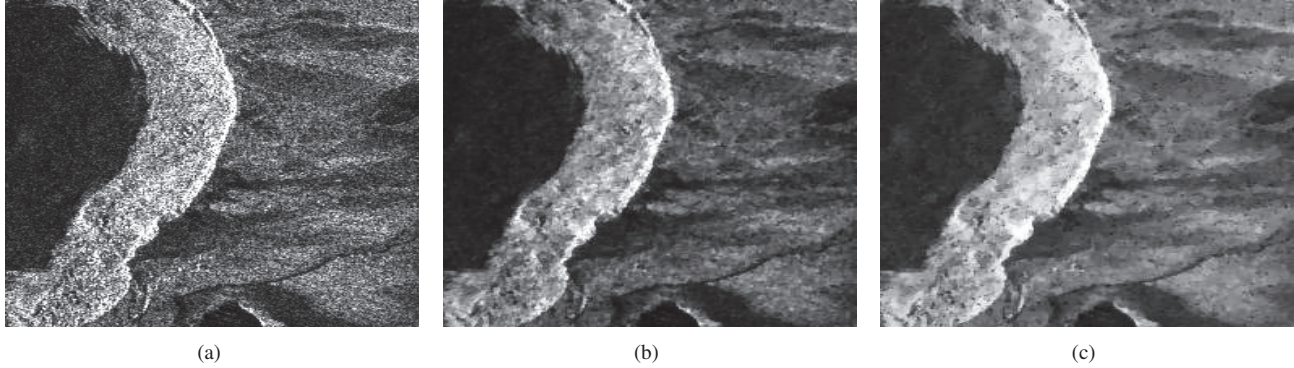


Fig. 1. Experiments results using real SAR image; (a) the original image, (b) ℓ_1 -norm denoised image and (c) ℓ_2 denoised image.

looks are the same. It turns out that the results presented here are significantly better in terms of the PSNR.

For the real SAR image, prior to denoising, the variance of the noise has to be estimated. The method we used here to estimate the variance of the noise is to calculate the variance of the difference of the noisy image and a denoised image where a relatively large value of λ is used to ensure that the denoised image is sufficiently smooth. The optimal value of the regularization was then found by minimizing the MCSURE estimate. The logarithmic transform was used prior to denoising and the exponential transform was used after denoising. The results for the real SAR image are shown in Figure 1.

5. CONCLUSION

In this paper we have proposed to denoise images corrupted by multiplicative noise by using TV regularization using both ℓ_1 - and ℓ_2 -norms as the data fidelity terms. Speckle noise can be modeled as a multiplicative Rayleigh i.i.d. noise and a logarithmic transform gives an additive noise model. While this is not new, the novelty here is using Monte-Carlo SURE to find the optimal value of the regularization parameter.

The experiment results for a simulated image indicate that MCSURE gives a very good estimate of the true MSE between the noise free image and the denoised image, and that using the ℓ_1 -norm as the data fidelity terms gives better results for noise of high variance (fewer looks) while the ℓ_2 -norm gives better results when the noise is less (higher number of looks). Finally we presented results for a real single look SAR image.

6. REFERENCES

- [1] F.T. Ulaby and C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*, Artech House Remote Sensing Library. Artech House, 1989.
- [2] J.R. Sveinsson, M.O. Ulfarsson, and J.A. Benediktsson, "Speckle reduction of SAR images using SURE-based adaptive sigmoid thresholding in the wavelet domain," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2009*, July 2009, vol. 4, pp. IV-462 –IV-465.
- [3] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.
- [4] S. Ramani, T. Blu, and M. Unser, "Monte-carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1540 –1554, sept. 2008.
- [5] P. Rodriguez and B. Wohlberg, "An iteratively reweighted norm algorithm for total variation regularization," in *Proceedings of the Fortieth Asilomar Conference on Signals, Systems and Computers, ACSSC '06.*, 11 2006, pp. 892 –896.
- [6] J.M. Bioucas-Dias, M.A.T. Figueiredo, and J.P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006 Proceedings.*, May 2006, vol. 2, p. II.
- [7] V. Solo, "Selection of regularisation parameters for total variation denoising," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999.*, Mar 1999, vol. 3, pp. 1653 –1655 vol.3.
- [8] Wu Yiyong, Niu Ruixin, and Peng Hailiang, "A new approach to SAR image speckle filtering," in *Microwave Conference Proceedings, APMC '97, 1997 Asia-Pacific*, Dec 1997, vol. 3, pp. 929 –932 vol.3.

Conference Paper 4

F. Palsson, J.R. Sveinsson and M.O. Ulfarsson, "Pansharpening Via Sparsity Optimization Using Overcomplete Transforms," To be presented at *2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*,

PANSHARPENING VIA SPARSITY OPTIMIZATION USING OVERCOMPLETE TRANSFORMS

Frosti Palsson, Johannes R. Sveinsson, Magnus O. Ulfarsson and Jon A. Benediktsson

University of Iceland
Faculty of Electrical and Computer Engineering
Hjardarhagi 2-6. IS-107 Reykjavik, Iceland

ABSTRACT

In this paper we consider pansharpening of multispectral satellite imagery based on solving an under-determined inverse problem regularized by the ℓ_1 -norm of the coefficients of overcomplete multi-scale transforms which all are tight-frame systems. There are two main approaches in sparsity promoting ℓ_1 -norm regularization, the analysis and the synthesis approach. We perform a number of experiments using two real and well known datasets where the focus is the comparison of the two approaches. One dataset includes a high resolution reference image while the other needs to be degraded prior to pansharpening in order to use the original multispectral image as the reference. Experiments are performed for a range of values for the regularization parameter, where each resulting pansharpened image is evaluated using three quality metrics. The behavior of those metrics as a function of the regularization parameter is compared for the analysis and synthesis formulations and it is shown that analysis gives better results.

1. INTRODUCTION

Pansharpening is the process of synthesizing a high spatial resolution multispectral (MS) image from a low spatial resolution multispectral image and a high spatial resolution panchromatic (PAN) image.

During the past two decade many diverse methods for pansharpening have been proposed, where one can discern three main paradigms. These are methods based on multi-resolution analysis (MRA), methods based on component substitution and methods based on variational techniques. Here the approach is based on modeling the pansharpening process using the classic observational model

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{n}, \quad (1)$$

where \mathbf{y} is the observed data, \mathbf{M} is the model matrix, \mathbf{x} is the pansharpened image we are estimating and \mathbf{n} is zero mean Gaussian noise. The model is well known and the exact structure of \mathbf{M} was derived in [1]. It is based on the assumptions that a decimation of the pansharpened image gives the

MS image and that a linear combination of the bands of the pansharpened image gives the PAN image. The matrix \mathbf{M} has more columns than rows and therefore this is an under-determined inverse problem. The classical approach to solving this kind of problem is to minimize a cost function of the form

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda R(\mathbf{x}) \quad (2)$$

where the first term is a data fidelity term, λ is the regularization parameter and the function $R(\mathbf{x})$ is the regularizer or penalty function.

Sparsity promoting optimization techniques have been a source of much interest and research in recent years. Today, one can identify two main paradigms in ℓ_1 -norm sparsity optimization [2]. One is the use of an analysis prior and the other is the use of a synthesis prior.

The analysis prior formulation is given as

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda \|\mathbf{A}\mathbf{x}\|_1, \quad (3)$$

where \mathbf{A} is the analysis operator. This formulation derives its name from the fact that the signal \mathbf{x} is being analyzed by applying linear filters (rows of \mathbf{A}) to it. On the other hand, the synthesis prior formulation is given by

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{S}\boldsymbol{\omega}\|_2^2 + \lambda \|\boldsymbol{\omega}\|_1, \quad (4)$$

where \mathbf{S} is the synthesis operator. Now, \mathbf{x} is being restored or synthesized as linear combination of the columns of \mathbf{S} where the coefficients of this linear combination are the vector $\boldsymbol{\omega}$. Hence, this approach is called synthesis.

Here we will be concerned with over-complete transforms, i.e., wavelet and directional wavelet transforms which are tight-frame systems. If we denote the transform operator by \mathbf{W} where $\mathbf{W} \in \mathbb{R}^{M \times N}$ and $M \gg N$, the analysis operator is $\mathbf{A} = \mathbf{W}$ while the synthesis operator is the adjoint operator, i.e., $\mathbf{S} = \mathbf{W}^T$.

For tight-frame systems we have $\mathbf{S}\mathbf{A} = \mathbf{W}^T\mathbf{W} = \mathbf{I}$ and the signal \mathbf{x} can be perfectly reconstructed from the analyzed signal by $\mathbf{x} = \mathbf{W}^T(\mathbf{W}\mathbf{x})$. Over-complete systems give different results for the two approaches while for orthonormal systems we have $\mathbf{W}^T\mathbf{W} = \mathbf{W}\mathbf{W}^T = \mathbf{I}$ and thus the analysis and synthesis cases are the same.

Since wavelet and directional wavelet transforms are multi-resolution analysis (MRA) systems, they decompose the signal under study into a low-pass representation and a hierarchy of high-pass representations at different scales. While the high-pass coefficients are sparse, the low-pass coefficients of most natural signals are far from sparse. Therefore, in the implementation of the methods used in this study, the low-pass coefficients are not included in the penalty term. In all experiments the same value of the regularization parameter λ is used for all the high-pass subbands.

The focus of the work presented here is the comparison of these two different methods for sparsity optimization for the solution of the inverse problem of pansharpening. These two approaches give different results for many types of inverse problems. For some problems, e.g. denoising, the analysis formulation gives better results [3] while for other problems, the synthesis formulation is better.

Experiments are performed using simulated Pleiades data [4] and Quickbird data [4]. The Pleiades dataset includes a high resolution reference image which is very useful when evaluating the quality of the pansharpened image using quality metrics. In order to evaluate the results for the Quickbird image, the data is degraded using bilinear interpolation such that the resulting pansharpened image is of the same size as the original MS image, which can then be used as the reference MS image. This method has the obvious drawback that information is lost during the degradation of the data, but it can still give a good indication of the performance of the method being evaluated.

The metrics used in the experiments are ERGAS [5], SAM [5] and Qavg [1], which all need a reference image. While there are metrics that do not need a reference image such as the QNR [6] spectral and spatial metrics, they usually do not give a good estimate of the mean square error (MSE).

2. SOLUTIONS OF THE ANALYSIS AND SYNTHESIS PROBLEMS

2.1. Analysis prior formulation

The optimization problem objective function using the analysis prior is stated as

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1, \quad (5)$$

where λ is the regularization parameter and \mathbf{W} is the transform operator. The algorithm to solve this problem is well known [3] and is given as

$$\mathbf{b}^{(i)} = \mathbf{x}^{(i)} + \frac{1}{\alpha} \mathbf{M}^T (\mathbf{y} - \mathbf{M}\mathbf{x}^{(i)}) \quad (6)$$

$$\mathbf{z}^{(i+1)} = \left(\mathbf{c}\mathbf{z}^{(i)} + \mathbf{W}(\mathbf{b}^{(i)} - \mathbf{W}^T \mathbf{z}^{(i)}) \right) ./ \left(\frac{2\alpha}{\lambda} |\mathbf{W}\mathbf{x}^{(i)}| + \mathbf{c} \right) \quad (7)$$

$$\mathbf{x}^{(i+1)} = \mathbf{b}^{(i)} - \mathbf{M}^T \mathbf{z}^{(i+1)}, \quad (8)$$

where the constants α and c are chosen as $\alpha > \text{maxeig}(\mathbf{M}^T \mathbf{M})$ and $c > \text{maxeig}(\mathbf{W}\mathbf{W}^T)$. The operator $./$ denotes element-wise division.

2.2. Synthesis prior formulation

The formulation using the synthesis prior is

$$J(\boldsymbol{\omega}) = \|\mathbf{y} - \mathbf{M}\mathbf{W}^T \boldsymbol{\omega}\|_2^2 + \lambda \|\boldsymbol{\omega}\|_1, \quad (9)$$

where \mathbf{W}^T is the adjoint operator (synthesis) and \mathbf{x} is reconstructed by $\mathbf{W}^T \boldsymbol{\omega}$. The solution to this problem is the iterated soft-thresholding algorithm [3] :

$$\boldsymbol{\omega}^{(i+1)} = \mathbf{S}_\lambda \left(\boldsymbol{\omega}^{(i)} + \frac{1}{\alpha} \mathbf{W}\mathbf{M}^T (\mathbf{y} - \mathbf{M}\mathbf{W}^T \boldsymbol{\omega}^{(i)}), \frac{\lambda}{2\alpha} \right), \quad (10)$$

where $\mathbf{S}_\lambda(\cdot, \lambda)$ is the element-wise soft-thresholding operator with threshold λ and for convergence the constant α needs to be chosen such that $\alpha \geq \text{maxeig}(\mathbf{W}\mathbf{M}^T \mathbf{M}\mathbf{W}^T)$.

3. DATASETS & EXPERIMENTAL RESULTS

3.1. Datasets

The simulated Pleiades dataset [4] consists of a 0.8m resolution PAN image of dimension 1024 by 1024 pixels and a 3.2m resolution 256 by 256 pixel MS image. There is also a high-resolution MS image of same size and resolution as the PAN image which means that no degradation of the data is needed.

The Quickbird dataset [4] is a high resolution image of the Mississippi State University campus. The PAN image is 2048 by 2048 pixels and of resolution 0.7m and the MS image is 512 by 512 pixels and of resolution 2.8m.

3.2. Experiments & Results

Experiments are performed using three overcomplete transforms, i.e., the Undecimated Discrete Wavelet Transform (UDWT) [7], the Non-Subsampled Contourlet Transform (NSCT) [8] and the Fast Discrete Curvelet Transform (FDCT) [9]. Two levels of decomposition are used in all experiments.

For the both dataset, results are evaluated for 41 values of λ ranging from 0 to 20 at intervals of 0.5. We use the ERGAS [5], SAM [5] and Qave [1] metrics to evaluate the quality of the pansharpened images.

The results for the Pleiades dataset are shown in Figure 1. The UDWT, FDCT and NSCT results are shown in the top, middle and bottom row, respectively. The columns of the figure array represent the ERGAS, SAM and Qave results, respectively. The results for the Quickbird dataset are shown in Figure 2 using the same layout. For the ERGAS and SAM metrics, lower values are better, while for the Qave metric, higher values are better with 1 being optimal.

In Table 1, the best results for each metric in each experiment are shown. The table gives a general idea of the relative performance of the methods. For the NSCT and UDWT,

Table 1: Best results obtained for each experiments. Bold font indicates best result obtained for analysis vs. synthesis.

Pleiades						
	NSCT		UDWT		Curvelet	
	Analysis	Synthesis	Analysis	Synthesis	Analysis	Synthesis
ERGAS	3.215	3.237	3.178	3.186	3.255	3.255
SAM	4.199	4.361	4.093	4.197	4.377	4.379
Qavg	0.958	0.957	0.959	0.959	0.956	0.956
Mississippi						
ERGAS	2.440	2.504	2.328	2.380	2.526	2.527
SAM	3.459	3.620	3.228	3.411	3.653	3.642
Qavg	0.946	0.939	0.953	0.949	0.939	0.938

the analysis approach gives better results in all experiments, while for the FDCT, the two approaches give very similar results. Overall, the UDWT gives best results.

4. CONCLUSIONS

In this paper we have compared the analysis and the synthesis approaches for ℓ_1 -norm sparsity regularization for the inverse model of pansharpening, where we have used three different over-complete transforms which all are tight-frame systems. Experiments were performed for two datasets and three quality metrics, a total of 36 combinations.

The results, as measured by the three metrics used, indicate that the analysis approach gives better performance than the synthesis approach. In 34 out of 36 experiments, analysis gives better results. However, the relative difference between the two approaches never exceeded 5.7%, with the average relative difference being 0.83% for the Pleiades dataset and 1.78% for the Quickbird dataset.

The assumption in the model matrix \mathbf{M} in (1), that a linear combination of bands gives the PAN image, is one possible reason for apparent similarity of the analysis and the synthesis method. This assumption sets strong constraints on the solution and thus limits the effectiveness of the sparsity regularization.

5. ACKNOWLEDGMENT

The authors would like to thank Jocelyn Chanussot, GIPSA-Lab, Grenoble Institute of Technology, for kindly providing the Quickbird and the Pleiades dataset. This work was supported by the Research Fund of the University of Iceland.

6. REFERENCES

- [1] F. Palsson, J.R. Sveinsson, and M.O. Ulfarsson, "A new pansharpening algorithm based on total variation," Accepted for publication in *IEEE Geoscience and Remote Sensing Letters*, 2013.
- [2] Michael Elad, Peyman Milanfar, and Ron Rubinstein, "Analysis versus synthesis in signal priors," *Inverse problems*, vol. 23, no. 3, pp. 947, 2007.
- [3] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," *In Proceedings of SPIE*, vol. 7446 (Wavelets XIII), August 2009.
- [4] 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.
- [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.
- [6] 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.
- [7] Mark J Shensa, "The discrete wavelet transform: wedding the a trous and mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, 1992.
- [8] Arthur L Da Cunha, Jianping Zhou, and Minh N Do, "The nonsubsampling contourlet transform: theory, design, and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089–3101, 2006.
- [9] Emmanuel Candes, Laurent Demanet, David Donoho, and Lexing Ying, "Fast discrete curvelet transforms," *Multiscale Modeling & Simulation*, vol. 5, no. 3, pp. 861–899, 2006.

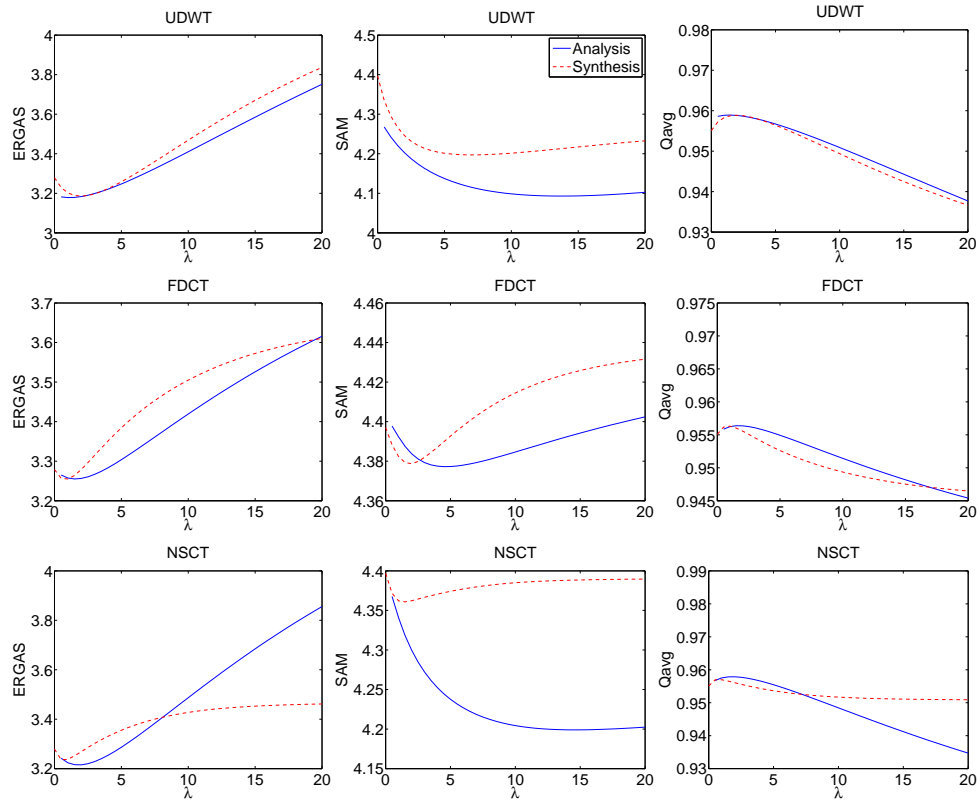


Fig. 1: Results for the Pleiades data.

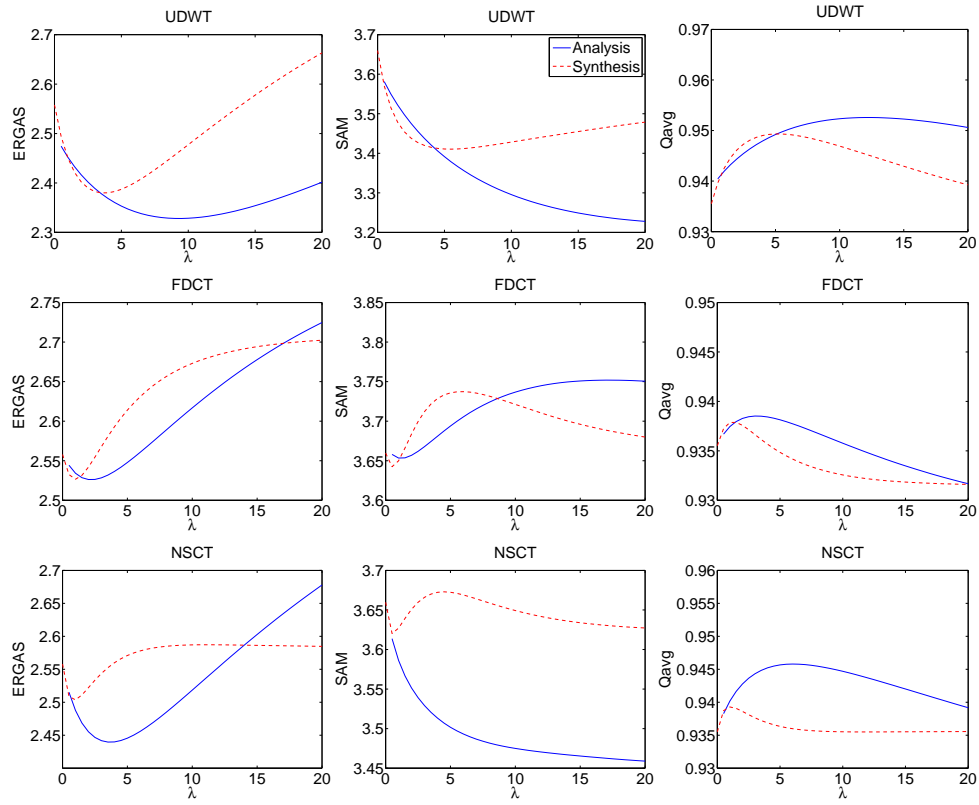


Fig. 2: Results for the degraded Quickbird data.

Journal Paper 1

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.

Classification of Pansharpened Urban Satellite Images

Frosti Pálsson, Johannes R. Sveinsson, *Senior Member, IEEE*, Jon Atli Benediktsson, *Fellow, IEEE*, and Henrik Aanæs

Abstract—The classification of high resolution urban remote sensing imagery is addressed with the focus on classification of imagery that has been pansharpened by a number of different pansharpening methods. The pansharpening process introduces some spectral and spatial distortions in the resulting fused multispectral image, the amount of which highly varies depending on which pansharpening technique is used. In the majority of the pansharpening techniques that have been proposed, there is a compromise between the spatial enhancement and the spectral consistency. Here we study the effects of the spectral and spatial distortions on the accuracy in classification of pansharpened imagery. We also study the performance in terms of accuracy of the various pansharpening techniques during classification with spatial information, obtained using mathematical morphology (MM). MM is used to derive local spatial information from the panchromatic data. Random Forests (RF) and Support Vector Machines (SVM) will be used as classifiers. Experiments are done for three different datasets that have been obtained by two different imaging sensors, IKONOS and QuickBird. These sensors deliver multispectral images that have four bands, R, G, B and near infrared (NIR). To further study the contribution of the NIR band, experiments are done using both the RGB bands and all four bands, respectively.

Index Terms—Classification, mathematical morphology, morphological profile, pansharpening, spatial consistency, spectral consistency.

I. INTRODUCTION

SATELLITES provide very valuable data about the Earth, e.g., for environmental monitoring, weather forecasting, map-making and military intelligence. But satellites are expensive, both to build and operate. This implies that it is important to make the best use of the data obtained from available satellites, e.g., by combining the output from different sensors. A good example of this is the merging of satellite images of low spatial and high spectral resolution with images of high spatial and low spectral resolution. This is also known as image fusion (pansharpening). An ideal pansharpened image should have the same spatial resolution as the panchromatic (Pan) image and be absolutely spectrally consistent with the multispectral (MS) image.

Several pansharpening techniques have been proposed during the past two decades [1]. For the majority of those techniques there is a compromise between the desired spatial enhancement and the spectral consistency. Achieving good spatial resolution usually compromises the spectral consistency and vice versa. Spectral consistency is an important property since spectral distortion can produce unreliable results in many applications such as classification.

There are few papers that address classification of pansharpened data. Pansharpened images are commonly not used directly for classification, rather, pansharpening is used to improve the image visually.

The classification of high resolution urban remote sensing imagery is a challenging research problem. Here we look at the classification of such data by both considering the classification of panchromatic imagery (single data channel) and spectral image (multiple data channels) obtained by image fusion. Obviously, both the spatial and spectral quality of the pansharpened image are inherently important in order to be able to correctly classify the pixels. Low spatial quality means missing details and low spectral quality, i.e., spectral distortion, can result in mis-classified pixels.

In this paper, we will investigate the classification of images that have been pansharpened using several different methods. Some of these methods have good spectral consistency while lacking spatial quality and for some methods, the opposite is true. We will also, try to assess the relative importance of spectral consistency vs spatial quality for classification.

In order to estimate the spectral and spatial quality of a pansharpened image, several metrics have been formulated, such as ERGAS [2], SAM [3], SID [4], Qave [5], RASE [6] and more (see Appendix A). To gain more insight into what qualities of the pansharpened image are important for classification, we will not only classify the multispectral pansharpened image but also add the original Pan image as well as morphological profiles (MP) and their derivatives (DMP), as proposed in [7], [8].

Panchromatic images are characterized by high spatial resolution. This high spatial resolution allows the identification of small structures in a dense urban area. However, the analysis of a scene by considering the value of a single pixel only will produce very poor classification results compared to the fine resolution. To solve this problem, some local spatial information is needed.

An interesting approach to provide local spatial information is based on the theory of *Mathematical Morphology* [9], which provides tools to analyze spatial relationship between pixels. The Morphological Profile (MP) was proposed in [7], [8] for segmentation of high-resolution satellite images. An MP is made up of an Opening Profile (OP) and a Closing Profile (CP) and we will also use the Derivative of the Morphological

Manuscript received June 30, 2011; revised September 14, 2011; accepted September 19, 2011. Date of publication December 07, 2011; date of current version February 29, 2012. This work was supported by the Research Fund of the University of Iceland.

F. Pálsson, J. R. Sveinsson, and J. A. Benediktsson are with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (corresponding author, e-mail: sveinss@hi.is).

H. Aanæs is with the Department of Informatics and Mathematical Modeling, Technical University of Denmark, 2800 Lyngby, Denmark.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2011.2176467

Profile (DMP) [8], [10]. Only the panchromatic data are used to build the morphological profile and its derivative.

For classification purposes, the MP and the DMP are regarded as feature vectors, where each class has a typical MP-DMP. Hence each MP (or DMP) is considered as a channel of a multispectral image. In this way, classification methods applied to multispectral images can be applied [4], [15].

Of the pansharpening methods used, one is our model-based method derived from the imaging sensor's physical properties. This method has the nice feature of being spectrally consistent [11] by design. The Intensity-Hue-Saturation (IHS) method has been widely used [12]–[15]. Another popular method is the Brovey [5] sharpening method. The IHS and Brovey methods suffer from spectral distortions. An important class of pansharpening methods are those based on multi-resolution analysis (MRA). In our experiments we will be using two such methods, one based on the Discrete Wavelet Transform (DWT) [16] and another based on the undecimated DWT [17]. Finally, we will use a pansharpening method based on Principal Component Analysis (PCA).

For our experiments, we have three different datasets along with ground truth data. There are 2 different IKONOS images of an urban area (Reykjavik, Iceland) with multispectral images with four bands (R, B, G and Near Infrared (NIR)) and panchromatic images which are of higher resolution (4 by 4 to 1 pixel) as compared to the multispectral images. Third experimental dataset is a QuickBird image of an urban area (Rome, Italy) [18] where the ratio of the pan and MS images is the same as for the IKONOS images and the MS image has also four bands. Thus we have data from two different sensors.

The work flow of our experiments is as follows. We begin by pansharpening the data from each dataset using the different methods. For all datasets we classify the pan image and the MS data separately. The next step is to add the Pan image to the MS data and classify the resulting multi-channel image. Finally, we add an MP and a DMP. As noted before, the MP or DMP can be thought of as multispectral images on their own, where, of course, the different bands do not contain any spectral information, only spatial information. The classification is performed using Support Vector Machines (SVM) and Random Forests (RF).

The outline of the paper is as follows. In the next section, we will give a brief overview of the different pansharpening methods. Section III gives a brief review on mathematical morphology and Section IV and V review SVM and RF, respectively. Experimental set up and results are given in Section VI and finally, conclusions are drawn in Section VII. Appendix A includes a brief discussion on the different quality metrics for estimating spectral consistency and spatial quality.

II. IMAGE FUSION

Image fusion can be done at several levels, e.g., at pixel, feature and decision levels. Here, we will only be concerned with pixel level fusion. The next subsections give a brief overview of the pansharpening methods used in this paper.

A. IHS Fusion

The IHS method [12]–[15] is a frequently used method, especially when working with LANDSAT and SPOT imagery. The

basic idea is to first transform the MS image into intensity (I), hue (H) and saturation (S) components (IHS colorspace)

$$\begin{bmatrix} I \\ H \\ S \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{-\sqrt{2}}{6} & \frac{-\sqrt{2}}{6} & \frac{2\sqrt{2}}{6} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

The next step is to scale the Pan image so that it has the same mean and variance as the intensity component of the MS image

$$P = \frac{\sigma_I}{\sigma_P} (P - \mu(P)) + \mu(I).$$

The intensity component is then replaced with the appropriately scaled Pan image and finally the inverse IHS transformation is taken to get the fused image

$$\begin{bmatrix} F(R) \\ F(G) \\ F(B) \end{bmatrix} = \begin{bmatrix} R + P - I \\ G + P - I \\ B + P - I \end{bmatrix}.$$

The IHS method produces images that have high spatial resolution and low spectral quality.

B. Brovey Fusion

The Brovey transform is a simple method for pansharpening and similar to the IHS method, produces images that have good spatial resolution and poor spectral quality. Each channel of the fused image is computed as

$$F(n) = N \text{Pan} \circ \frac{\text{MS}(n)}{\sum_{k=1}^N \text{MS}(k)}$$

where \circ means element-wise multiplication, $F(n)$ and $\text{MS}(n)$ are the n -th channels of the fused and MS images, respectively, and N is the number of channels in the MS image. So N is either 3 or 4.

C. Model Based Fusion

Several fusion methods have been proposed that take into account the physics of the imaging sensor [11], [19]. The model based image fusion method [11] is derived from a model of the imaging sensor's underlying physics. This approach ensures spectral consistency of the resulting fused image. Spectral consistency can be measured as the cross correlation between the low resolution MS image and the downsampled fused image. In our case the low resolution spectral image consists of four bands: R, B, G and Near Infrared (NIR).

The method of image fusion presented in [11] creates a high resolution multi band image $[R^{high}, G^{high}, B^{high}, N^{high}]$ from a high resolution panchromatic image P^{high} and a low resolution multi band image $[R^{low}, G^{low}, B^{low}, N^{low}]$. The first part of the method calculates a correlation matrix, Σ , between the five channels, $[R, G, B, N, P]$. This is done by considering the inner product, or spectral overlap, of the respective channels. The relative spectral response of the IKONOS sensor can be seen in Fig. 1.

The high resolution image is then grouped into blocks, where each block exactly corresponds to a low resolution

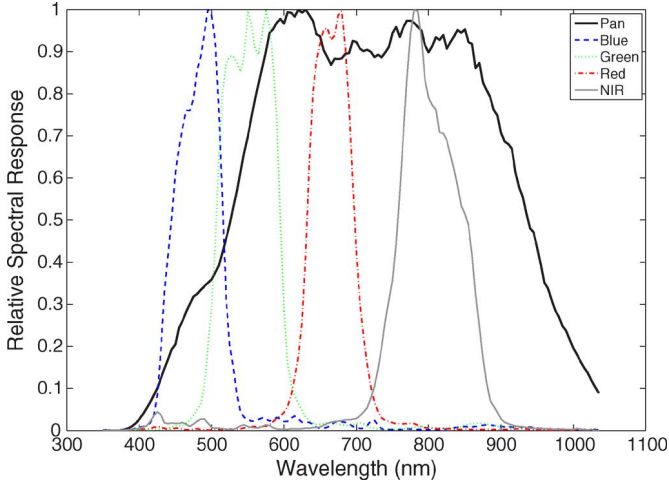


Fig. 1. Relative Spectral Response of the IKONOS sensor for the panchromatic, R, G, B and N bands.

pixel $[R_i^{low}, G_i^{low}, B_i^{low}, N_i^{low}]$. Assuming a perfect overlap between low and high resolution pixels, spectral consistency, which is enforced, has the form

$$\begin{bmatrix} R_i^{low} \\ G_i^{low} \\ B_i^{low} \\ N_i^{low} \end{bmatrix} = \frac{1}{16} \sum_{j=1}^n \begin{bmatrix} R_{ij}^{high} \\ G_{ij}^{high} \\ B_{ij}^{high} \\ N_{ij}^{high} \end{bmatrix} \quad (1)$$

where j runs over the n high resolution pixels corresponding to the low resolution pixel i . It is then assumed that these blocks follow a multivariate normal distribution, with a variance of Σ , and a mean value computed via conditional means utilizing that the P_{ij}^{high} and $[R_i^{low}, G_i^{low}, B_i^{low}, N_i^{low}]$ are known. This is used in a Bayesian setting, where this normal distribution is used as the data term.

To complete the Bayesian setting, a prior (or model of a likely image) is also needed. For that, a piecewise smooth model is used. The model is specifically formulated by penalizing deviation between neighboring high resolution pixel values, *except* for when the panchromatic image P has edges. The data term and the prior are combined into a Markov Random Field (MRF) framework. The additional constraint on this MRF framework is that the resulting $[R^{high}, G^{high}, B^{high}, N^{high}]$, must be spectrally consistent. That is accomplished by observing that (1), describes a hyperplane (i.e., a linear constraint), and then only allowing solutions located on that hyperplane. The resulting optimization problem is a large sparse least squares function—which has the nice quality that it is convex.

D. MRA Method

The multi-resolution analysis approach to pansharpening is widely used and there are numerous techniques today based on it. The basic idea is to take the discrete wavelet transform (DWT) [16] of both the MS and Pan images. The next step is to retain the approximation coefficients for the MS image but replace the detail coefficients with those from the Pan image. Instead of just replacing the coefficients, some fusion rule can be used. This is shown in Fig. 2.

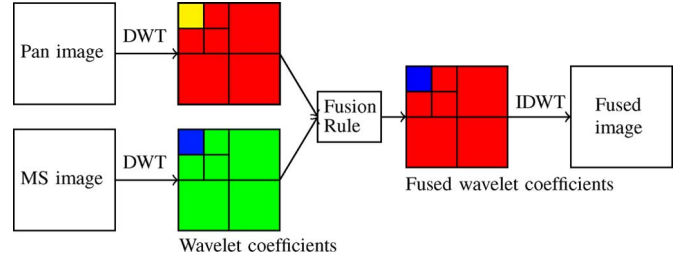


Fig. 2. Basic wavelet based pansharpening.

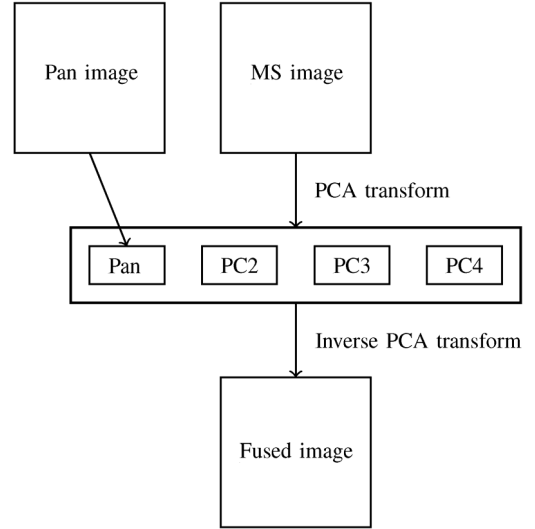


Fig. 3. PCA based pansharpening.

The main drawback to this approach is that there will be substantial artifacts in the fused image due to the nature of the DWT, i.e., it is not shift-invariant and lacks directionality so there is bound to be some aliasing, thus the spatial quality of the final fused image is reduced. One approach to overcome this problems is to use the undecimated DWT [17].

E. PCA Method

Similar to the MRA method, the PCA method for pansharpening [20] is a so called spatial detail injection method. First, the mutually correlated bands of the MS image are transformed using PCA into a set of independent components whose number is the same as number of bands in the MS image. The first principal component has the highest variance and is similar to the Pan image itself. The next step is to replace this component with the actual Pan image and finally take the inverse transform to get the fused image. Fig. 3 illustrates this.

III. MATHEMATICAL MORPHOLOGY

Mathematical Morphology (MM) is a theory that provides mathematical tools to analyze the spatial relationship between pixels [7], [9], [21]. Matheron and Serra originally introduced MM to study binary porous media in the 1960s [22].

When working with MM, one is usually interested in specific objects or structures in the image. These objects of interest are viewed as subsets of the image. After these objects have been

identified, several sets of known size and shape (such as square, line or disk) can be used to characterize their morphology.

These simple sets are called *Structuring Elements* (SEs). An SE always has an origin (usually its symmetric center) that allows one to position the SE at a given pixel of the image.

The most fundamental morphological operations are *erosion* and *dilation*. The eroded value at a given pixel x is the minimum pixel value over the SE and the dilated value at a given pixel x is the maximum pixel value over the SE.

Two of the most common operations of MM are *opening* and *closing* operations. The morphological opening of an image is an erosion followed by dilation while the morphological closing of an image is the reverse: dilation followed by erosion. Both operations are done using the same SE. The morphological opening of an image removes light features that are smaller than the SE whereas morphological closing removes dark features smaller than the SE. While these operations are useful, they have the drawback of not being connected filters. What this means is that they do not preserve shapes and thus introduce shape noise. So instead we will use the morphological reconstruction operations.

Opening by reconstruction of an image removes unconnected light features while *closing by reconstruction* of an image removes unconnected dark features. For both operations, shapes are preserved and the structures present after the transformation are of a size greater than or equal to the SE used. This effectively means that less shape noise is generated.

Since objects in a given image may be of varying sizes, the use of a single SE might prove to be inefficient. To solve this, a multiscale approach can be taken, where SEs of the same shape but of several different sizes are used.

A Morphological Profile (MP) [7], [21] is a $2n + 1$ -dimensional vector that consists of the input image, n openings and n closings (reconstruction operations) generated with an SE of fixed shape but of varying size. The MP can be defined as

$$MP(x) = [CP_n(x), \dots, I(x), \dots, OP_n(x)]$$

where CP and OP are closing profile and opening profile of n -dimension, constructed using an SE of fixed shape but of varying size and $I(x)$ is the original panchromatic image. An n dimensional profile has n different sizes in total. This kind of structure is called a *granulometry*. The number of openings/closings and the corresponding sizes of the SE depend on the size distribution of structures of interest in the panchromatic image. Finally, since these are operations performed on the panchromatic image, the MPs contain no spectral information, i.e., they only contain spatial information.

The *Derivative of the Morphological Profile* (DMP) [10] is defined as a $2n$ -dimensional vector equal to the discrete derivative of the MP and is given by

$$DMP_i(x) = |MP_{i-1}(x) - MP_i(x)|.$$

The information provided by the DMP is both spatial and radiometric. For a given pixel, the shape of the DMP can give an idea of the neighborhood of the pixel, e.g., whether it belongs to darker or lighter structure than the surrounding pixels. Finally,

the amplitude of the DMP gives information about the local contrast of the structure.

From the above it can be seen that an MP (DMP) can be viewed as a multiband image, where the different openings and closings (or their derivatives) make up the different bands. In a classification setting, inclusion of these profiles effectively increases the number of features available for classification.

IV. SUPPORT VECTOR MACHINES

One of the most promising of recent developments in the field of machine learning and pattern recognition are Support Vector Machines (SVM) [23]–[25]. These are supervised learning methods that are widely used for classification and regression. When given a set of training samples where each sample is marked as belonging to one of two classes, the SVM algorithm builds a model that can predict to which class new data points will belong to.

The SVM model can be viewed as a representation of the data points as points in space where the separate categories or classes are divided by a gap that is as wide as possible. The SVM constructs a hyperplane in a space that has a high or infinite dimension. For a good separation, the distance from the training data points of each class to the hyperplane should be maximized. When this distance or margin is maximized we call the resulting linear classifier, a maximum margin classifier.

For this linear classifier to be able to solve non-linear problems, the non-linear data points are mapped into a higher dimensional space in such a way that linear classification in this new space is equivalent to non-linear classification in the original space. This is what is known as the ‘kernel trick’. By using non-linear kernel-functions, this approach can separate complex (e.g., multi-modal) class distributions in high dimensional feature spaces. A commonly used kernel-function is the Gaussian Radial Basis function and it is used for the experiments in this paper.

The SVM is in fact a binary classification strategy so some kind of multiclass decomposition is required to handle multiclass problems. A single multiclass problem is reduced into multiple binary classification problems. Two common methods are the One-Against-All (OAA) and One-Against-One (OAO).

For a multiclass problem with k -classes, $k(k - 1)/2$ binary SVMs are trained. For OAO, classification is done by max-wins voting strategy where every individual binary classifier assigns the instance to one of the two classes thus increasing the vote count for that class by one.

Since SVM does not provide class-labels as classifier output but instead provides the distance from each data point to the hyperplane, the sign of the distance is used to determine a class. Finally, the instance classification is determined by the class with the most votes. A detailed discussion on SVM is for example given in [26].

V. RANDOM FOREST

Random Forest (RF) [27] is a relatively recent ensemble method for classification and regression. Ensemble classifiers get their name from the fact that several classifiers, i.e., an ensemble of classifiers, are trained and their individual results are then combined through a voting process. Many such methods

have been proposed [28]–[30]. These methods are usually based on the techniques of boosting [31], [32] or bagging [33]. Boosting uses an iterative re-training procedure where the samples that have been incorrectly classified are given increased weight with each iteration. Bagging (or bootstrap aggregation) is based on training an ensemble of classifiers using samples that have been bootstrapped from a training set. This has been shown to reduce the variance of the classification.

Trees are good examples of high-variance and low-bias procedures so bagging seems to work very well for them. The idea is to average many noisy but relatively unbiased models in order to reduce the variance. Since trees are inherently noisy, averaging is of great benefit. Also, when grown sufficiently deep, they have a low bias.

Boosting is in general a much more computationally costly method than bagging but it is considerably more accurate. The RF algorithm uses an improved method of bootstrapping as bagging and has been shown to be comparable to boosting in terms of accuracy while being free from the drawbacks of boosting. Thus RF is a much faster algorithm than boosting.

During the training process, the RF algorithm grows many CART-like trees [34], with each individual tree trained on bootstrapped samples of the original training set. In order to determine a split for each node, the algorithm only searches across a randomly selected subset of the input variables.

In order to classify an object from an input vector, the input vector is run down each tree in the forest. Each tree gives a classification or a unit vote for a particular class and the forest chooses the classification having the most votes.

The most important user-tunable parameter in the RF algorithm is the number of variables used for a split of a tree. A common value for this parameter is the square root of the number of inputs. By limiting this parameter, the amount of computation can be greatly reduced while the correlation between the trees in forest is minimized. Thus reducing the error rate. Since each tree only uses a portion of the available input variables, the algorithm is much faster than a conventional bagging algorithm with a similar tree-like classifier. Another nice feature is that the trees do not need to be pruned, hence reducing the computational costs even further.

The computational time of the RF algorithm [34] has been shown to be $cT\sqrt{MN} \log(N)$ where c is some constant, T is the number of trees in the forest, M is the number of variables and N is the number of samples in the dataset. This shows that the algorithm is not very computationally intensive but on the other hand it can require a considerable amount of memory as it needs to store an N by T matrix while running.

RF as a classifier has many nice properties [34] such as excellent accuracy, it scales up very well, it can handle thousands of variables and a lot of missing data and it is insensitive to noise in the training labels. It also gives an unbiased estimate of the test set error as trees are added to the ensemble and finally it does not overfit.

VI. DATASETS AND EXPERIMENTAL RESULTS

There are three datasets used in this paper. Two of these are images of Reykjavik, Iceland, acquired by the IKONOS Earth imaging satellite on 9th August 2001. Each dataset consists of a

TABLE I
NOTATION FOR THE DIFFERENT PANSHARPENING METHODS
USED IN EXPERIMENTS

Name	Type
MBF	Model Based Fusion
Brovey	Brovey fusion
IHS	IHS fusion
DWT	MRA based fusion using the DWT
UDWT	MRA based fusion using the UDWT
PCA	PCA based fusion

TABLE II
MORPHOLOGICAL PROFILES (MP) USED IN THE EXPERIMENTS

Dataset	Design	SEs (radius of disks)
1	3 openings, 3 closings	2,4,6
2,3	11 openings, 11 closings	1,2,3,4,6,8,10,12,17,25,30

low-resolution (4m) multispectral image with four bands R,G,B and NIR and a high-resolution panchromatic image of resolution 1 m. The ratio between the panchromatic and low-resolution images is 16 or 4 by 4 pixels.

The third dataset is an image of Rome, Italy acquired by the QuickBird satellite. As with the other datasets, it consists of a low-resolution (2.4 m) multispectral image with the four bands R, G, B and NIR and a high-resolution panchromatic image of resolution 0.6 m. The ratio between the Pan and MS image is again 16.

These images were then fused using the methods in Table I, and the resulting high-resolution multispectral image was used for classification.

Ground truth data for 6 classes (*Large buildings, small buildings, large roads, streets, open areas and shadows*) was available for the IKONOS datasets and extensive ground-truth data with 9 classes was available for the QuickBird dataset.

Two different classification algorithms were used, 1) SVM and 2) RF. For the SVM part, the libSVM¹ library was used. The Gaussian Radial Basis function was chosen for the kernel and the training parameters were found using a grid-search.

For the RF part, the randomforest-Matlab² package, developed by Abhishek Jaialtilal was used. The number of trees was chosen to be 200, all other options used the default values.

For the datasets, two different MPs (and their derivatives, DMPs) were constructed. Details on these profiles are shown in Table II.

By including these profiles in the classification process, the number of features is increased by their respective lengths, i.e., by 6 and 22, respectively.

For each dataset and pansharpening method, five different experiments were conducted. The first one was to simply classify

¹Chih-Chung Chang and Chih-Jen Lin—LIBSVM—A Library for Support Vector Machines—<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

²Abhishek Jaialtilal, randomforest-Matlab, Random Forest (Regression, Classification and Clustering) implementation for MATLAB (and Stand-alone)—<http://code.google.com/p/randomforest-matlab/>.

TABLE III
TRAINING AND TEST SET DETAILS
FOR THE FIRST DATASET

Class	Train	Test
Small buildings	1526	34155
Open areas	7536	43867
shadows	1286	25806
Large buildings	2797	30916
Large roads	3336	39202
Streets	5616	35147
Total	22097	209093

the Pan image. The second experiment was to classify the pansharpened MS image. The third type of experiment was to classify the pansharpened MS image along with the Pan image. In the fourth experiment, the pansharpened MS image with an MP was classified and finally, the MP was replaced with the DMP.

These experiments were done using both a three band (RGB) pansharpened MS image and the four band (RGBNIR) image. Since the original IHS method can only handle the RGB bands, this method was omitted when using the four band MS images.

Each MS image produced by the different pansharpening methods was evaluated using a number of quality metrics, 7 metrics for spectral quality and one metric to evaluate the spatial quality. A summary of these metrics can be found in Appendix A.

There are two different approaches to evaluating these metrics. One is to compare the upsampled original MS image with the final pansharpened image. Another way of doing this is to fuse the low-resolution MS image with a downsampled (degraded) Pan image and then comparing the final result with the original MS image. We chose to use the former approach.

For the metrics CC, ERGAS, RASE, RMSE, SAM and SID lower is better, while for the Qave and spatial metric, closer to one is better.

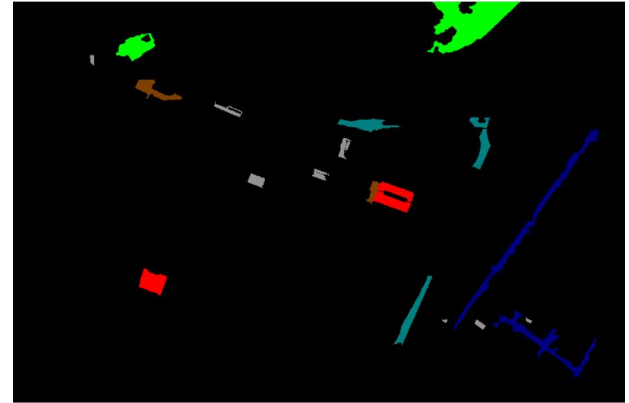
All classification accuracies are presented using 3 different methods. The Overall Accuracy (OA) is simply the percentage of correctly classified pixels. The Average Accuracy (AA) is the mean of class-specific accuracy for all the classes. Finally, the Kappa Coefficient (Kappa) is the percentage of agreement, i.e., correctly classified pixels, corrected by the number of agreements that one would expect purely by chance alone.

In the next three subsections a short description of the datasets is given and the experimental results for each dataset are presented. The OA is used in the discussion.

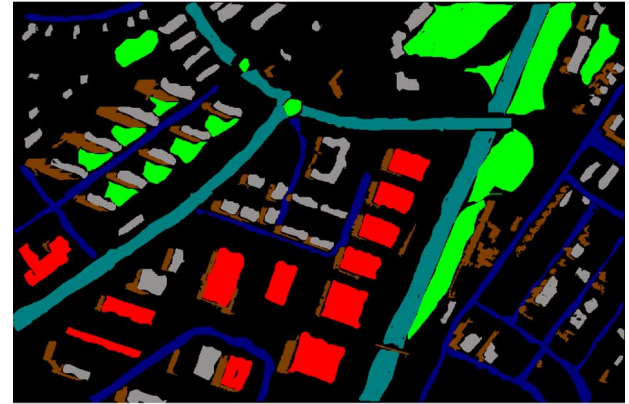
A. First Dataset

This dataset is composed of a high resolution panchromatic image (976×640 pixels) and a low resolution multispectral image consisting of 4 bands: R, G, B and NIR. As stated above, the ratio between the panchromatic and low-resolution images is 16 or 4 by 4 pixels. The panchromatic image can be seen in Fig. 6(a).

It is important to note that the training set was chosen in such a way as to be not very representative and thus making the classification a much more challenging task. Information on the size



(a)



(b)

Fig. 4. Training map (a) and reference map (b) for dataset 1.

TABLE IV
QUALITY METRICS FOR THE THREE BAND (RGB) DATA USING FIRST DATASET

Metric	MBF	Brovey	IHS	DWT	UDWT	PCA
CC	0.002	0.014	0.007	0.033	0.004	0.027
ERGAS	2.359	8.846	8.904	5.636	6.502	9.436
Qave	0.728	0.680	0.713	0.985	0.688	0.673
RASE	8.623	33.256	32.996	22.111	24.400	35.325
RMSE	0.013	0.051	0.051	0.034	0.038	0.054
SAM	1.072		2.081	1.407	2.105	2.470
SID	0.001	0.0	0.005	0.018	0.003	0.004
spatial	0.501	0.981	0.985	0.847	0.998	0.969

TABLE V
QUALITY METRICS FOR THE FOUR BAND (RGBNIR) DATA
USING FIRST DATASET

Metric	MBF	Brovey	DWT	UDWT	PCA
CC	0.007	0.026	0.033	0.067	0.233
ERGAS	2.518	7.426	5.636	6.387	8.212
Qave	0.976	0.946	0.985	0.858	0.949
RASE	10.072	27.575	22.111	25.261	31.763
RMSE	0.016	0.043	0.034	0.039	0.049
SAM	1.103		1.407	4.472	2.461
SID	0.007	0.0	0.018	0.013	0.004
spatial	0.522	0.920	0.847	0.998	0.808

of the training and test sets is shown in Table III and the training and validation sets are shown in Fig. 4(a) and (b), respectively.

TABLE VI
CLASSIFICATION ACCURACIES (%) FOR ALL EXPERIMENTS USING FIRST DATASET WITH THREE INPUT CHANNELS (RGB)

DatasetI	RF							SVM						
Features	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT
P	AA	49.30	49.30	49.30	49.30	49.30	49.30	AA	48.98	48.98	48.98	48.98	48.98	48.98
	OA	48.45	48.45	48.45	48.45	48.45	48.45	OA	47.98	47.98	47.98	47.98	47.98	47.98
	K	37.32	37.32	37.32	37.32	37.32	37.32	K	36.80	36.80	36.80	36.80	36.80	36.80
RGB	AA	55.78	51.80	53.60	54.09	56.27	55.89	AA	53.40	49.91	52.45	51.78	55.54	52.44
	OA	56.57	50.93	52.94	54.36	55.96	55.92	OA	53.12	48.85	51.79	51.38	55.25	52.20
	K	47.52	40.53	43.07	44.84	46.76	46.77	K	43.61	38.00	41.65	41.38	45.93	42.39
PRGB	AA	59.41	51.59	53.17	59.53	55.55	62.74	AA	59.76	50.06	52.53	58.42	55.61	60.58
	OA	59.77	50.70	52.45	59.55	55.24	63.20	OA	59.90	48.94	51.80	58.38	55.25	60.93
	K	51.39	40.26	42.51	51.13	45.90	55.49	K	51.59	38.11	41.67	49.77	45.94	52.82
RGB+MP	AA	64.61	61.60	61.04	64.28	63.13	67.10	AA	67.19	60.50	60.87	65.66	64.56	66.65
	OA	65.72	61.21	61.17	65.12	63.51	67.87	OA	67.98	60.07	60.57	66.20	64.60	67.53
	K	58.50	53.04	52.99	57.76	55.83	61.10	K	61.30	51.72	52.38	59.14	57.23	60.75
RGB+DMP	AA	64.68	60.84	61.12	64.09	63.30	66.95	AA	65.40	57.79	59.72	63.96	62.13	64.59
	OA	65.37	60.21	60.67	64.53	63.09	67.70	OA	66.33	57.33	59.48	64.33	62.19	65.81
	K	58.11	51.87	52.45	57.09	55.40	60.92	K	59.31	48.41	51.07	56.93	54.31	58.68

TABLE VII
CLASSIFICATION ACCURACIES (%) FOR ALL EXPERIMENTS USING FIRST DATASET WITH FOUR INPUT CHANNELS (RGBNIR)

DatasetI	RF						SVM					
Features	Accuracy	MBF	Brovey	DWT	PCA	UDWT	Accuracy	MBF	Brovey	DWT	PCA	UDWT
P	AA	49.30	49.30	49.30	49.30	49.30	AA	48.98	48.98	48.98	48.98	48.98
	OA	48.45	48.45	48.45	48.45	48.45	OA	47.98	47.98	47.98	47.98	47.98
	K	37.32	37.32	37.32	37.32	37.32	K	36.80	36.80	36.80	36.80	36.80
RGBN	AA	57.83	58.93	57.55	58.64	63.12	AA	57.89	59.03	56.31	58.65	62.07
	OA	58.58	58.94	58.09	58.72	63.77	OA	58.19	59.17	56.50	58.58	62.75
	K	49.94	50.36	49.32	50.07	56.20	K	49.58	50.68	47.55	49.98	55.04
PRGBN	AA	60.75	59.14	60.69	58.68	64.25	AA	61.34	59.00	60.03	58.60	63.06
	OA	61.23	59.15	60.95	58.71	64.78	OA	61.63	59.13	60.14	58.52	63.48
	K	53.12	50.60	52.79	50.07	57.40	K	53.66	50.62	51.90	49.91	55.91
RGBN+MP	AA	67.04	65.77	66.50	66.59	68.02	AA	67.08	66.12	66.16	65.62	67.32
	OA	67.65	66.25	66.99	66.88	68.73	OA	67.91	66.70	66.84	66.13	68.28
	K	60.86	59.15	60.04	59.94	62.15	K	61.23	59.76	59.92	59.07	61.66
RGBN+DMP	AA	64.48	65.32	64.55	65.00	67.08	AA	65.78	64.87	64.72	64.72	65.35
	OA	65.20	65.70	65.16	65.42	67.93	OA	66.69	65.43	65.34	65.25	66.56
	K	57.90	58.49	57.84	58.14	61.18	K	59.76	58.24	58.15	57.98	59.60

The spectral and spatial quality metrics for all the pansharp-ening methods are shown in Table IV (three bands) and Table V (four bands), respectively. Since there are so many different met-rics, it is not easy to order the different methods by performance, but it is evident while studying the three band data in Table IV that the MBF method produces by far the best spectrally consis-ent results while its spatial performance is rather low.

The DWT method gives the second best results when it comes to spectral quality while the spatial quality is the second worst.

The IHS, Brovey and PCA methods produce similar results of low spectral quality and rather good spatial quality, while the UDWT method gives better spectral quality and the best spatial quality of all the methods. The same can be said about the four band data in Table V.

The results from the classification experiments are presented in Table VI and Table VII, for three band (RGB) and four band (RGBN) data, respectively. Note that the IHS method is not available for the four band data.

In the following discussion we use the OA and only refer to the results achieved using the RF classifier. The accuracies achieved using the two classifiers are somewhat similar for the first dataset, with the SVM classifier usually trailing behind the RF, but for the other datasets, RF gives significantly better results than SVM in every single case.

1) *Three Band Data (RGB)*: By studying Table VI we see that the classification of the Pan image alone gives OA of 48.45%. When classifying the pansharpened image alone the MBF data give the best accuracy, 56.57% closely followed by the PCA and UDWT data. The Brovey results are the worst in terms of accuracy at 50.93% with IHS only slightly better. For this experiment we see that difference between best and worst accuracy is 5.64%, so all the results are relatively similar.

When the Pan image is added to classification process, the results in terms of accuracy change somewhat drastically. Now the data giving best accuracies are the UDWT, at 63.20%, while the MBF data are second best at 59.77%. Again the accuracy for the Brovey data is at the bottom (50.70%) with IHS only slightly better. The difference in accuracy between best and worst is now 12.5%, more than twice the difference from the previous experiment.

It is also interesting to note that the Brovey, IHS and PCA data perform slightly worse in this experiment than in the previous one. It seems that adding the spatial information from the Pan image into the classification process is not of any noticeable benefit for these data. The average gain in accuracy, compared to the previous experiment, for all the methods is 2.37%.

The next two experiments add the MP followed by the DMP so we have richer spatial information available for classification. The results are very similar for these two last experiments so we will only comment on the results obtained by including the MP. The MP for this dataset was chosen to be short; we have only three different sizes for the SEs. This dataset did not benefit from longer profiles. This is probably a result from the fact that the training set was chosen in a way not to be representative, thus making the classification difficult.

Again the UDWT method produces the best accuracies, 67.87% with MBF coming second at 65.72%. The Brovey and IHS data give the worst accuracies, 61.21% and 61.17%, respectively. The DWT data 65.12% and the PCA data 63.51%. The difference between best and worst accuracies is now 6.66% and it is clear that the further inclusion of more spatial information decreases this difference compared to the previous experiment (Pan and MS). The most spectrally inconsistent data, gain the most from the MP/DMP. For example, we see a gain of 10.51% accuracy for the Brovey data, while the gain for the UDWT data is only 4.67%. The average gain in accuracy by including the MP is 7.28% and 6.78% by including the DMP, a difference of 0.5% favoring the MP.

2) *Four Band Data (RGB and NIR)*: Table VII presents the results for the experiments using all four bands of the fused image. By comparing this table to the previous one, we see that there are some important differences. It is clear that the benefit of adding the NIR band is very different for the various pansharpening methods. If we look at the classification of MS data only, it is evident that the best results are achieved by the UDWT data, 63.77%, an increase of 7.85% from the three band experiment.

Surprisingly, the accuracy for the Brovey data now comes second at 58.94%, an increase of 8.01% compared to the three-band experiment. In the same manner, the DWT data show a modest increase of accuracy, 3.73% to 58.09% and the PCA data goes from 55.24% to 58.72%, an increase of only 3.48%. The best spectrally consistent MBF data only shows an increase of accuracy of 2.01%, to 58.58%, coming third in this experiment.

Clearly, the biggest increase in accuracy is for the Brovey data which also is the least spectral consistent. It is also noticeable that the UDWT data perform best now. While both the MBF and DWT data give better spectral quality, the UDWT data scores highest in spatial quality, at 0.998 according to Table V.

When the Pan data are included, there are not as big gains overall in accuracy as in the three band case. This can easily be explained by the presence of the NIR band in the MS data, so the addition of spatial data from the Pan image is relatively smaller. The most notable result is that all data gain in accuracy except the PCA data. In the three band case we saw very big gains in accuracy for the UDWT in this experiment, but in this case the gain in accuracy is only 1.58%. The average gain in accuracy for all data is now 1.35%. Also, similar to the three band case, the UDWT data give the best accuracies and the MBF data come second.

By looking at the last two experiments (MP/DMP) in Table VII we see a similar trend as for the three band data but obviously the gains in accuracy are not quite as big, now that we have increased the amount of MS data. The average gain in accuracy by including the MP is 6.34% and the gain achieved by including the DMP is 4.92%, giving a difference of 1.42%, which is almost three times more than for the three-band data.

B. Second Dataset

The second dataset is also an IKONOS image of Reykjavik, Iceland. The high resolution panchromatic image is 628×700 pixels and the lower-resolution multispectral image consists of R, G, B and NIR bands. The size ratio is again 16 or 4 by 4 pixels. The panchromatic image is shown in Fig. 6(c). Information on the size of the training and test sets is shown in Table VIII and the training and validation sets are shown in Fig. 5(a) and (b), respectively.

The training set has been chosen to be more representative than that for the previous dataset hence the accuracy is generally much better than for the first dataset. The MP/DMP is now of length 23.

The spectral and spatial quality metrics for all the pansharpening methods are shown in Table IX (three bands) and Table X (four bands), respectively. The MBF method produces the best results in terms of spectral quality while having the worst spatial quality. However, the spatial quality of 0.699 is significantly higher than for the previous dataset.

If we try to order the methods from best to worst according to spectral quality (based on the average of the metrics where lower is better), we have the ranking: MBF (2.344), DWT (5.734), UDWT (6.421), PCA (6.821), IHS (8.387) and Brovey (9.529). This should give a reasonable idea how these methods rank spectrally. Spatially, they rank, from best to worst: UDWT (0.999), IHS (0.995), Brovey (0.990), PCA (0.972), DWT (0.907) and MBF (0.699).

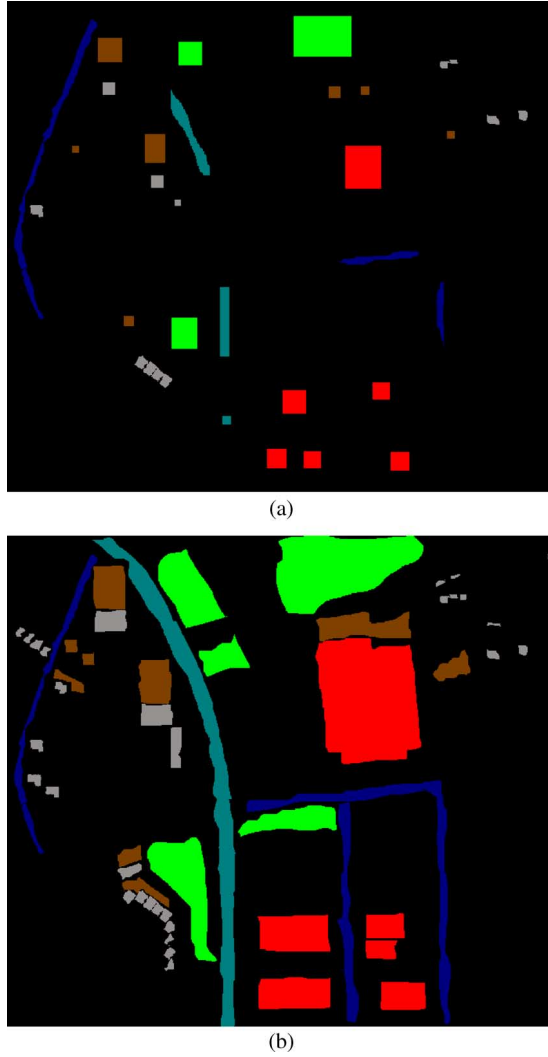


Fig. 5. Training map (a) and reference map (b) for dataset 2.

TABLE VIII
TRAINING AND TEST SET DETAILS FOR THE SECOND DATASET

Class	Training	Test
Houses	1863	6213
Open Areas	6068	28144
Shadows	2619	10610
Large buildings	5599	29768
Large Roads	2489	12051
Streets	4103	11940
Total	22741	98726

These results reflect the complementary nature of the spectral and spatial quality. The best spectrally performing method has also the lowest spatial quality. The best spatially performing method, UDWT, has rather low spectral quality.

The values for the four-band data are similar. It is worth nothing though, that in general the Qave values for three-band data are lower (mean: 0.749) than for the four-band data (mean: 0.947).

TABLE IX
QUALITY METRICS FOR THE THREE BAND (RGB) DATA
USING SECOND DATASET

RGB	MBF	Brovey	IHS	DWT	UDWT	PCA
bandCorrs	0.003	0.009	0.004	0.056	0.006	0.038
ERGAS	2.766	10.306	10.523	6.657	7.504	8.034
Qave	0.727	0.673	0.708	0.980	0.704	0.701
RASE	10.199	37.296	37.003	26.151	29.155	30.723
RMSE	0.010	0.035	0.035	0.025	0.028	0.029
SAM	1.082	NA	2.754	1.481	1.824	2.099
SID	0.003	0.0	0.002	0.032	0.007	0.003
spatial	0.699	0.990	0.995	0.907	0.999	0.972

TABLE X
QUALITY METRICS FOR THE FOUR BAND (RGBNIR) DATA
USING SECOND DATASET

dataset2-RGBN	MBF	Brovey	DWT	UDWT	PCA
bandCorrs	0.014	0.081	0.056	0.083	0.288
ERGAS	3.105	9.375	6.657	7.544	7.372
Qave	0.969	0.924	0.980	0.900	0.961
RASE	11.912	33.797	26.151	29.627	28.571
RMSE	0.011	0.032	0.025	0.028	0.027
SAM	1.287	NA	1.481	3.300	2.123
SID	0.009	0.0	0.032	0.018	0.003
spatial	0.718	0.970	0.907	0.999	0.904

The accuracies from the classification experiments are presented in Table XI and Table XII, for three band (RGB) and four band (RGBN) data, respectively. As before, we will use the OA in the discussion and omit the SVM results.

1) *Three-Band Data (RGB)*: Classification accuracy for the Pan data alone is 62.63% which is considerably better than for the previous dataset. The next experiment is classification of the pansharpened MS data. The average accuracy of all data is 75.41%. The data giving the best accuracy is now the UDWT at 79.29%, the PCA data come second with 77.41%, followed by the MBF data at 75.84%, the DWT data with 74.59% and finally the IHS and Brovey data with 73.73% and 71.63%, respectively.

Compared to the first dataset, we see now that the MBF data goes from giving best accuracy to being third. The method with the best spatial detail, UDWT, gives the best result even though the spectral quality is lower than that of MBF and DWT.

By adding the Pan image to the classification we now see the difference between data produced by the different methods become even more pronounced. The average accuracy for this experiment is 77.31% which is only slightly higher than for the MS data alone (75.41%) but as we have seen before, the accuracy for the IHS, Brovey and PCA methods lowers a little bit compared to the previous experiment. This means greater increase for the MBF and wavelet based methods. The accuracy of the MBF data increases by 2.92%, the DWT data see an increase of 5.33% and finally for the UDWT data there is an increase of 4.90%. The data that show an increase in accuracy here are those which give the best spectral quality in Table IX. The data produced by the other methods do not benefit at all from the added spatial information in the Pan image.

TABLE XI
CLASSIFICATION ACCURACIES (%) FOR ALL EXPERIMENTS USING SECOND DATASET WITH THREE INPUT CHANNELS (RGB)

Dataset2	RF							SVM						
Features	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT
P	AA	53.46	53.46	53.46	53.46	53.46	53.46	AA	48.28	48.28	48.28	48.28	48.28	48.28
	OA	62.63	62.63	62.63	62.63	62.63	62.63	OA	60.73	60.73	60.73	60.73	60.73	60.73
	K	51.75	51.75	51.75	51.75	51.75	51.75	K	48.09	48.09	48.09	48.09	48.09	48.09
RGB	AA	71.21	68.01	69.39	70.26	72.26	75.23	AA	61.43	53.96	55.79	56.99	57.92	60.39
	OA	75.84	71.63	73.73	74.59	77.41	79.29	OA	70.96	64.22	65.99	68.00	69.87	70.08
	K	69.50	63.93	66.59	67.81	71.22	73.80	K	62.93	53.55	56.17	58.82	61.04	61.84
PRGB	AA	74.41	67.49	69.19	74.86	71.94	79.88	AA	64.63	54.02	55.86	61.81	58.55	65.42
	OA	78.76	70.79	73.29	79.92	76.89	84.19	OA	73.22	64.13	66.07	72.52	69.69	75.16
	K	73.15	62.93	66.08	74.48	70.60	79.95	K	65.90	53.49	56.28	64.75	60.93	68.29
RGB+MP	AA	81.64	80.70	80.82	81.16	81.50	82.62	AA	80.05	76.28	76.44	78.47	78.31	80.37
	OA	85.27	84.56	84.60	85.03	85.16	86.16	OA	83.48	80.17	80.31	82.02	81.75	84.10
	K	81.36	80.45	80.50	81.04	81.20	82.48	K	79.08	74.98	75.16	77.27	76.91	79.85
RGB+DMP	AA	84.59	82.94	83.04	83.96	83.87	84.97	AA	80.01	77.86	77.94	79.04	79.07	80.07
	OA	88.75	87.37	87.42	88.40	88.13	89.31	OA	82.97	81.11	81.21	82.24	82.02	83.31
	K	85.68	83.93	83.97	85.22	84.88	86.38	K	78.49	76.20	76.32	77.58	77.31	78.91

TABLE XII
CLASSIFICATION ACCURACIES (%) FOR ALL EXPERIMENTS USING SECOND DATASET WITH FOUR INPUT CHANNELS (RGBNIR)

Dataset2	RF						SVM					
Features	Accuracy	MBF	Brovey	DWT	PCA	UDWT	Accuracy	MBF	Brovey	DWT	PCA	UDWT
P	AA	53.46	53.46	53.46	53.46	53.46	AA	48.28	48.28	48.28	48.28	48.28
	OA	62.63	62.63	62.63	62.63	62.63	OA	60.73	60.73	60.73	60.73	60.73
	K	51.75	51.75	51.75	51.75	51.75	K	48.09	48.09	48.09	48.09	48.09
RGB	AA	74.92	76.00	75.50	75.08	81.13	AA	69.98	67.35	67.20	68.04	74.80
	OA	78.46	80.53	79.96	80.02	85.31	OA	76.01	75.42	75.32	75.01	81.45
	K	72.84	75.31	74.60	74.58	81.34	K	69.56	68.60	68.43	68.15	76.36
PRGBN	AA	76.78	76.14	77.27	75.23	81.53	AA	72.05	67.49	69.93	68.13	76.83
	OA	80.35	80.69	81.94	80.05	85.68	OA	77.50	75.53	76.80	75.11	82.58
	K	75.19	75.51	77.05	74.63	81.83	K	71.49	68.74	70.46	68.27	77.85
RGBN+MP	AA	82.08	81.81	81.54	81.79	82.85	AA	80.86	78.85	79.60	79.99	82.10
	OA	85.65	85.33	85.21	85.34	86.07	OA	83.67	82.64	82.56	82.85	85.46
	K	81.83	81.44	81.28	81.44	82.39	K	79.37	78.04	77.99	78.35	81.58
RGBN+DMP	AA	85.02	84.12	84.68	84.04	85.50	AA	80.76	79.77	80.34	80.48	81.15
	OA	89.05	88.54	89.02	88.38	89.46	OA	83.49	82.88	83.25	83.40	83.95
	K	86.07	85.41	86.01	85.20	86.59	K	79.18	78.40	78.87	79.04	79.74

In the final two experiments for this dataset, an MP and an DMP are included. By looking at Table XI one sees that the increase in accuracy is, in every case, more for the DMP than for the MP, so we will only comment on the DMP experiment, since the behavior is very similar.

The average gain in accuracy by including the DMP in the classification is 10.92% (7.83% for the MP). This is significantly higher than for the first dataset. Since these two IKONOS images are somewhat similar, the rather large differences in the experiments are probably due to the way the training sets were chosen.

The data that gain most in terms of accuracy in this experiment are those with low spectral quality, namely the IHS, Brovey and PCA data. The gains are 16.59%, 14.13% and 11.24%, respectively. The smallest gain, 5.12%, is indeed for the UDWT data, which have the best spatial detail.

2) *Four-Band Data (RGB and NIR)*: The inclusion of the NIR band in the pansharpened data has a similar effect as was seen for the first dataset. The experiment results are shown in Table XII. Basically, we see the same kind of behavior as before.

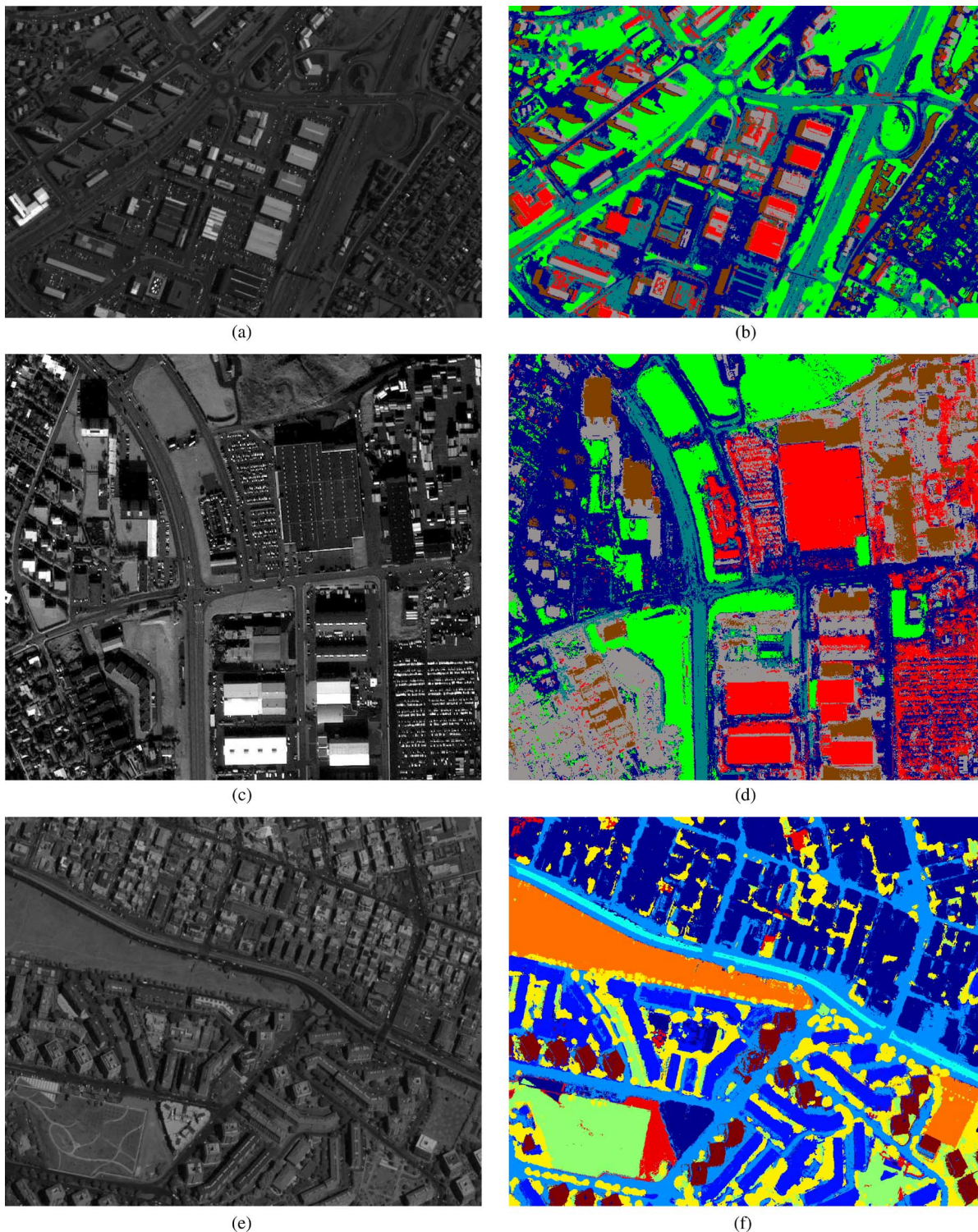


Fig. 6. The panchromatic images and the classification maps obtained using all 4 bands for each dataset. The classification images were obtained using the configuration that yielded the highest accuracy. (a) Original IKONOS panchromatic image for dataset 1. (b) RF classification map for dataset 1 (OA: 68.73%). (c) Original IKONOS panchromatic image for dataset 2. (d) RF classification map for dataset 2 (OA: 89.46%). (e) Original QuickBird panchromatic image for dataset 3. (f) RF classification map for dataset 3 (OA: 94.79%).

Since we now have four bands of MS data, the classification accuracy is better than for the three-band case. The average accuracy is 80.86% so the NIR band brings in an average increase of 5.12%. The UDWT data gives yet again the best results 85.31%, followed by Brovey with 80.53%, the PCA with 80.02%, DWT with 79.96% and finally the MBF with 78.46%.

The Brovey method comes second here, which was also the case in this experiment for the first dataset. It is also interesting to see how well the UDWT data perform here, almost 5% better than the second best.

Adding the Pan data to the classification does not give much change in terms of accuracy, the average gain in accuracy is

TABLE XIII
TRAINING AND VALIDATION SET DETAILS
FOR THE THIRD DATASET

Class	Training	Test
Buildings	10872	54360
Blocks	6589	32945
Roads	9811	49056
Light Train	964	4818
Vegetation	4177	20884
Trees	5451	27253
Bare Soil	4876	24381
Soil	903	4515
Tower	2875	14375
Total	46518	232587

only 0.89%. This is also similar to what was seen in the first dataset. The biggest gains in terms of accuracy are for the MBF and DWT data, 1.89% and 1.98%, respectively. For the UDWT data there is only gain in accuracy of 0.37%, so it is clear that these data do not gain anything from the spatial information in the Pan data, which may be explained by the high spatial quality of the UDWT data.

Again, there is more gain in accuracy when adding the DMP to the classification process. But unlike the three-band case, the gains are not nearly as great, the average being 7.15%. This is consistent with the results from the previous dataset. The MBF data gain the most, 8.70% while the UDWT data gain the least, 3.79. The explanation for this is probably the low spatial quality of the MBF data and the high spatial quality of the UDWT data.

C. Third Dataset

The third dataset is a QuickBird image of Rome, Italy. The Pan image is 972×1188 pixels and the spatial resolution is 0.6 m, while the low-resolution four-band MS image has spatial resolution of 0.24 m. The ratio of high-resolution Pan pixels to the low-resolution MS pixels is 16, same as for the IKONOS data. The panchromatic image is shown in Fig. 6(e). There are several things that make this dataset challenging. First, this image is taken when the satellite was far from being at the nadir. As a result the buildings in the images have a very ‘oblique’ look. Second, this is a dense urban area with primarily three classes of buildings: 1) small buildings (1–2 floors), 2) blocks, large buildings with 3–5 floors and 3) towers, tall rectangular buildings with more than 5 floors. There are 9 different classes. Information on the classes and number of training and validation samples is shown in Table XIII.

The quality metrics for this dataset are shown in Table XIV (three-band) and Table XV (four-band). The rank of the methods is similar as for the previously discussed datasets.

The results of the classification experiments are presented in Table XVI for the three-band data and in Table XVII for the four-band data.

1) *Three-Band Data (RGB)*: Classification of the Pan image alone gives the relatively low accuracy of 41.03%. The reason for the low AA and Kappa values for this experiment is that several classes had accuracy of zero or close to zero.

TABLE XIV
QUALITY METRICS FOR THE THREE BAND (RGB) DATA USING THIRD DATASET

RGB	MBF	Brovey	IHS	DWT	UDWT	PCA
bandCorrs	0.029	0.003	0.018	0.049	0.020	0.090
ERGAS	1.562	6.068	6.235	3.779	4.291	4.484
Qave	0.745	0.709	0.729	0.991	0.732	0.717
RASE	5.645	26.367	25.894	14.474	15.626	17.206
RMSE	22.900	106.968	105.049	58.718	63.393	69.804
SAM	0.677	0.0	1.967	1.012	1.821	1.776
SID	0.001	0.001	0.010	0.029	0.009	0.012
spatial	0.587	0.967	0.968	0.771	0.995	0.889

TABLE XV
QUALITY METRICS FOR THE FOUR BAND (RGBNIR) DATA
USING THIRD DATASET

RGBN	MBF	Brovey	DWT	UDWT	PCA
bandCorrs	0.026	0.113	0.049	0.077	0.330
ERGAS	1.652	4.711	3.779	4.108	3.827
Qave	0.992	0.970	0.991	0.965	0.962
RASE	6.593	20.444	14.474	15.807	15.625
RMSE	26.746	82.941	58.718	64.127	63.388
SAM	0.731	0.0	1.012	2.390	1.638
SID	0.009	0.001	0.029	0.013	0.006
spatial	0.609	0.924	0.771	0.995	0.766

Classification of the MS data gives average accuracy of 65.07% which is a big improvement in accuracy compared to the previous experiment. Now the IHS data give the best accuracy 66.86%, MBF data give 65.59%, Brovey 65.45%, then PCA with 64.78%, UDWT with 64.68% and finally DWT with 63.06%. The difference here between the best and worst accuracy is 3.80%. Interestingly, the least spectrally consistent data give the best classification results here.

Adding the Pan image to the classification brings the average accuracy up to 69.62%, which does not seem much. Closer inspection shows that only the MBF and the wavelet based data show improvement in accuracy, 8.22% for MBF, 9.42% for DWT and 12.04% for UDWT. The accuracy for the other data actually decreases by almost a percentage. The difference in accuracy between the best (UDWT, 76.72%) and worst (Brovey, 64.64%) is 12.08%, which is a much larger difference than for the previous experiment.

Adding the MP and DMP brings big gains in accuracy for this dataset. The MP and the DMP improve the average accuracy by 22.64% and 22.79%, respectively. If we look at the DMP, the average accuracy is 92.42%, which is very good. Of all the different pansharpening methods, UDWT gives yet again the best results, 94.23%, followed closely by the MBF at 93.39%. PCA is at the bottom with 90.45%. The difference between best and worst accuracy is 3.78%.

2) *Four-Band Data (RGB and NIR)*: With the added NIR band, the average accuracy for classification of the MS data alone is now 76.05%. This is an improvement of 10.98% over the three band experiment. The UDWT data give the best accuracy, 78.47%, followed by the MBF at 76.34%, DWT at 75.32%, Brovey at 75.07% and finally PCA at 75.04%. The difference between best and worst accuracy is 3.40%.

TABLE XVI
CLASSIFICATION ACCURACIES (%) FOR ALL EXPERIMENTS USING THIRD DATASET WITH THREE INPUT CHANNELS (RGB)

Dataset3	RF							SVM						
Features	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT	Accuracy	MBF	Brovey	IHS	DWT	PCA	UDWT
P	AA	26.76	26.76	26.76	26.76	26.76	26.76	AA	25.73	25.73	25.73	25.73	25.73	25.73
	OA	41.03	41.03	41.03	41.03	41.03	41.03	OA	40.56	40.56	40.56	40.56	40.56	40.56
	K	27.77	27.77	27.77	27.77	27.77	27.77	K	27.11	27.11	27.11	27.11	27.11	27.11
RGB	AA	59.59	57.44	60.92	58.02	59.23	60.47	AA	44.70	40.86	46.11	41.71	45.29	43.37
	OA	65.59	65.45	66.86	63.06	64.78	64.68	OA	56.84	56.71	58.92	53.44	56.13	54.59
	K	58.80	58.59	60.32	55.75	57.91	57.70	K	47.79	47.44	50.13	43.51	47.09	44.79
PRGB	AA	68.20	57.01	60.56	66.78	58.74	71.50	AA	52.58	41.00	46.09	51.55	45.33	56.63
	OA	73.81	64.64	66.04	72.48	64.04	76.72	OA	65.00	56.78	58.94	64.09	56.21	68.83
	K	68.65	57.69	59.41	67.04	57.07	72.15	K	57.60	47.51	50.15	56.44	47.16	62.32
RGB+MP	AA	92.67	91.10	90.83	92.02	89.68	93.34	AA	88.70	84.86	84.79	87.82	82.82	90.06
	OA	93.30	91.89	91.59	92.59	90.31	93.91	OA	89.11	86.48	86.19	88.57	83.73	90.54
	K	92.04	90.37	90.01	91.19	88.49	92.77	K	87.06	83.89	83.56	86.41	80.61	88.76
RGB+DMP	AA	92.82	91.34	91.08	92.33	90.02	93.78	AA	91.78	89.52	89.42	91.35	87.73	92.99
	OA	93.39	91.98	91.69	92.76	90.45	94.23	OA	91.92	89.99	89.76	91.46	87.84	93.21
	K	92.14	90.46	90.13	91.40	88.64	93.14	K	90.41	88.10	87.83	89.86	85.55	91.94

TABLE XVII
CLASSIFICATION ACCURACIES (%) FOR ALL EXPERIMENTS USING THIRD DATASET WITH FOUR INPUT CHANNELS (RGBNIR)

Dataset3	RF						SVM					
Features	Accuracy	MBF	Brovey	DWT	PCA	UDWT	Accuracy	MBF	Brovey	DWT	PCA	UDWT
P	AA	26.76	26.76	26.76	26.76	26.76	AA	26.76	26.76	26.76	26.76	26.76
	OA	41.03	41.03	41.03	41.03	41.03	OA	41.03	41.03	41.03	41.03	41.03
	K	27.77	27.77	27.77	27.77	27.77	K	27.77	27.77	27.77	27.77	27.77
RGB	AA	71.54	70.00	70.18	70.17	74.07	AA	60.58	56.58	57.01	57.02	61.30
	OA	76.34	75.07	75.32	75.04	78.47	OA	68.81	67.71	68.19	67.14	71.15
	K	71.71	70.17	70.48	70.17	74.25	K	62.43	60.91	61.53	60.40	65.19
PRGBN	AA	73.59	69.90	72.17	70.08	75.46	AA	63.26	56.61	61.18	57.03	65.02
	OA	77.92	74.99	76.84	74.94	79.69	OA	69.75	67.73	69.16	67.15	72.18
	K	73.59	70.08	72.27	70.06	75.70	K	63.58	60.93	62.79	60.41	66.50
RGBN+MP	AA	93.43	93.42	93.09	93.72	94.08	AA	90.37	89.71	89.97	90.15	91.08
	OA	94.07	94.01	93.82	94.19	94.64	OA	90.51	89.94	90.26	90.27	91.45
	K	92.96	92.89	92.67	93.10	93.64	K	88.73	88.05	88.42	88.44	89.85
RGBN+DMP	AA	93.70	93.69	93.41	93.86	94.40	AA	93.21	92.88	92.85	93.07	93.86
	OA	94.17	94.09	93.90	94.16	94.79	OA	93.19	92.85	92.97	93.06	93.99
	K	93.08	92.98	92.75	93.07	93.81	K	91.92	91.52	91.66	91.76	92.86

Adding the Pan data to the classification does not improve the accuracy significantly. The average accuracy increases by a meager 0.83% up to 76.88%. The Brovey and PCA data actually decrease in accuracy by a small amount. The UDWT data give the best accuracy, 79.69%, MBF comes second with

77.92%. The difference between the best and worst accuracy is 4.75%.

The inclusion of the MP and DMP does not help as much as in the three-band case in terms of accuracy. The respective gains of the MP and DMP are now 17.27% and 17.34%.

The UDWT data give the highest accuracy of 94.79%, followed closely by MBF at 94.17%, PCA at 94.16%, Brovey at 94.09% and DWT at 93.90%. The difference between the best and worst accuracy is 0.89%.

VII. CONCLUSIONS

In this paper we have attempted to assess the relative importance of the spectral and spatial quality for classification of pansharpened satellite urban imagery. We used two different classifiers, SVM and RF and we considered 6 different pansharpening methods with varying degrees of spectral and spatial quality.

Of those methods, there are two that can be said to represent the opposite ends of the quality spectrum. The MBF method produces images of very high spectral quality while the spatial quality is considerably lacking. On the other hand, the multi-resolution analysis based UDWT method produces images of spectral quality that is relatively low but have outstanding spatial quality, producing images of almost the same detail as the Pan image itself.

A number of experiments were done using three different datasets produced by two different satellite sensors, IKONOS and QuickBird. All the different experiments for a single dataset were done using pansharpened images of three channels (RGB) and four channels (RGB and NIR), respectively. The different experiments were classification of the Pan image alone and MS image alone, Pan image and MS image together, and, finally, of the Pan image together with the MS image and an MP or a DMP.

The results of the classification experiments show that the UDWT method gives best results in terms of accuracy in 22 out of 24 experiments using the RF as classifier and in 19 out of 24 experiments using the SVM classifier. The MBF method gives the best results in terms of accuracy in one experiment out of 24 using the RF classifier and in 4 out of 24 experiments using the SVM classifier. The RF classifier produced better results in terms of accuracy than the SVM method in almost every experiment, except for about a fourth of the experiments for the first dataset.

In general, the results obtained for each dataset showed the same patterns. In the case of the three-band data (RGB), classification of the pansharpened MS image alone, gave relatively similar accuracies for all the pansharpening methods. The difference between the best and the worst accuracies for the first dataset was 5.64%, 5.78% for the second dataset and 3.8% for the third dataset. The MBF data gave best accuracies for the first dataset, The UDWT data gave the best accuracies for the second dataset and surprisingly, for the QuickBird dataset, the IHS data gave the best accuracy for both type of classifiers in this experiment.

The gap in performance between data produced by the different pansharpening methods was greatly increased with the addition of the Pan image to the classification process. The difference between the best and the worst accuracies are now 12.5%, 13.4% and 12.08% for the first, second and third dataset, respectively. The UDWT data gave the best accuracies for this experiment for all datasets and both classifiers.

For the four-band data (RGB and NIR) there were substantial differences compared to the three-band case. The UDWT

data gave the best accuracies in every experiment using the RF classifier. For the SVM classifier, the only exception were the MP/DMP experiments for the first dataset, where the MBF data gave best results.

The most notable things concerning the classification of the MS data alone using all the four bands, are how much higher accuracy the UDWT data gives compared to the rest of the methods, and secondly, how well the Brovey data perform compared to the more spectrally consistent methods, coming second place for the first two datasets and third in the last dataset. It has to be mentioned though, that the performance of all the data with the exception of the UDWT data, are very similar for this experiment.

Adding the Pan image to the classification using all four channels does not change much in terms of accuracy. For all datasets, the general behavior is that the MBF, DWT and UDWT data gain 1%–2% in accuracy while the other data usually give lower accuracy than for the previous experiment for the MS data only.

Including the MP/DMP in the classification increased the accuracy by a considerable amount for all datasets. The most notable increase in accuracy being for the third dataset, an average increase of 22.7% (using RGB data) compared to the PRGB experiment. The first dataset showed the least benefits from including the MP/DMP, the probable reason being how unrepresentative the training set was chosen. The three band data gained more accuracy than the four band data in the experiments using the MP/DMP, since the added spatial information was relatively less in the four band case.

The conclusion to draw here seems to be that while spectral consistency is important, spatial quality seems to be more important for classification purposes, given that the spectral consistency is acceptable. Here, “acceptable” means significantly better spectral consistency than that of the IHS and Brovey methods.

We also demonstrated that adding spatial information (Pan image, MP/DMP) to the classification of data with very good spatial quality resulted in relatively greater gains in accuracy than for data with good spectral quality but lacking somewhat in spatial quality. Often the classification accuracy actually decreased when including the Pan image in the classification, for the least spectrally consistent data. The conclusion being that if the data has bad spectral quality, adding spatial data will not help at all. This behavior was seen in every dataset when working with three band data.

The UDWT is the pansharpening method that produced the data achieving the highest classification accuracy in almost every experiment presented in this paper. The images produced by this method ranked third in terms of spectral quality for all three datasets, but in every case the spatial quality was excellent (as good as it can get). The method giving the best spectral quality, by far, did worse than expected in our experiments and we draw the conclusion that the reason for this, is the relatively low spatial quality of the fused images produced by that method.

Hence, the main conclusion is that while the spectral consistency and the spatial quality of pansharpened images tend to be complementary in nature, i.e., gains in one quality often compromise the other, for classification purposes, it seems better to

aim for as good spatial quality as is possible, given that the spectral quality stays above some acceptable minimum.

Since all the datasets used in this paper were of urban areas, future work is to see if the above conclusions also hold for rural imagery.

APPENDIX

SUMMARY OF QUALITY METRICS

ERGAS:

ERGAS [2] is an acronym in French for “Erreur relative globale adimensionnelle de synth se” which translates to “relative dimensionless global error in synthesis”. This metric calculates the amount of spectral distortion in the fused image and is given by

$$\text{ERGAS} = 100 \frac{h}{l} \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{\text{RMSE}(n)}{\mu(n)} \right)^2},$$

where N is the number of bands, RMSE is the root mean square error, h/l is the ratio of pixels in the Pan image to the MS image and $\mu(n)$ is the mean of the n -th band.

Universal Image Quality Index: Qave:

Qave [5] is a metric that attempts to model the spectral distortion as a combination of three factors. These factors are loss of correlation, luminance distortion and contrast distortion. The metric is given by

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]},$$

where $x = \{x_i | i = 1, 2, \dots, N\}$ and $y = \{y_i | i = 1, 2, \dots, N\}$ are the original MS and fused image vectors, respectively, and

$$\begin{aligned} \bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i & \bar{y} &= \frac{1}{N} \sum_{i=1}^N y_i \\ \sigma_x^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \\ \sigma_y^2 &= \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 \\ \sigma_{xy} &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \end{aligned}$$

Relative Average Spectral Error (RASE):

RASE [6] computes the average performance in terms of the RMSE of the bands in the pansharpened image.

$$\text{RASE} = \frac{100}{M} \sqrt{\frac{1}{N} \sum_{i=1}^N \text{RMSE}^2(B_i)}.$$

Spectral Angle Mapper (SAM):

The Spectral Angle Mapper (SAM) [3] is a metric that calculates the spectral similarity between two spectral vectors as a spectral angle,

$$\cos \alpha = \frac{\sum_{i=1}^N x_i y_i}{\sqrt{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i^2}},$$

where N is the number of bands and $\mathbf{x} = (x_1, x_2, \dots, x_N)$ and $\mathbf{y} = (y_1, y_2, \dots, y_N)$ are two spectral vectors at some pixel location in the original MS image and the fused image, respectively. The value of SAM for the entire image is the average of all the α values for every pixel.

Spectral Information Divergence (SID):

SID [4] originates from information theory. Each pixel spectrum is viewed as a random variable and SID measures the difference or discrepancy of the probabilistic behaviors between two spectral vectors, taken from the MS image and final fused image, respectively. Let the vector $\mathbf{x} = (x_1, \dots, x_N)^T$ be taken from the MS image and the vector $\mathbf{y} = (y_1, \dots, y_N)^T$ be taken from the final pansharpened image. The range of these vectors needs to be normalized to the range $[0, 1]$ so we define

$$p_j = \frac{x_j}{\sum_{i=1}^N x_i} \quad q_j = \frac{y_j}{\sum_{i=1}^N y_i},$$

where N is the number of bands. SID is defined by

$$\text{SID}(x, y) = D(x||y) + D(y||x),$$

where $D(x||y)$ is called the relative entropy and is given by $D(x||y) = \sum_{i=1}^L p_i \log(p_i/q_i)$. Similar for the term $D(y||x)$.

Correlation Coefficient (CC):

The correlation coefficient (CC) between the original MS image (X) and the pansharpened image (Y) is defined as

$$\text{CC}(X, Y) = \frac{\sum_{mn} (X_{mn} - \bar{X})(Y_{mn} - \bar{Y})}{\sqrt{(\sum_{mn} (X_{mn} - \bar{X})^2) (\sum_{mn} (Y_{mn} - \bar{Y})^2)}},$$

where \bar{X} and \bar{Y} are the mean values of the corresponding images.

Root Mean Square Error (RMSE):

The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_x \sum_i (X_i(x) - Y_i(x))^2}{n \times m \times d}}$$

where X is the MS image, Y is the pansharpened image, x is the pixel and i is the band number. Finally, n is the number of rows, m is the number of columns and d is the number of bands.

Spatial:

The spatial metric used in this paper is based on computing the correlation coefficient between the high-frequency data of each MS band and the high frequency data of the Pan image. To extract the high-frequency data of a band, it is convoluted with the following mask

$$\text{mask} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

The metric value is the average of the CCs for each band of the MS image.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments.

REFERENCES

- [1] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.
- [2] L. Wald, "Quality of high resolution synthesized images: Is there a simple criterion," in *Proc. 3rd Conf. Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images*, T. Ranchin and L. Wald, Eds., 2000, SEE/URISCA, Nice, France.
- [3] A. F. H. Goetz, R. H. Yuhas, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm," *Summaries 3rd Annu. JPL Airborne Geoscience Workshop*, vol. 1, pp. 147–149, 1992.
- [4] C.-I. Chang, "Spectral information divergence for hyperspectral image analysis," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp., IGARSS'99*, 1999, vol. 1, pp. 509–511.
- [5] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [6] T. Ranchin and L. Wald, "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, 2000.
- [7] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, 2001.
- [8] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, 2003.
- [9] P. Soille, *Morphological Image Analysis, Principles and Applications*, 2nd ed. New York: Springer, 2003.
- [10] M. Fauvel, J. A. Palmason, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Classification of remote sensing imagery with high spatial resolution," in *Image and Signal Processing for Remote Sensing XI 06 (SPIE)*, Bruges, Belgium, Sep. 2005.
- [11] H. Aanaes, J. R. Sveinsson, A. A. Nielsen, T. Bovik, and J. A. Benediktsson, "Model-based satellite image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1336–1346, 2008.
- [12] 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, 1990.
- [13] P. Chavez and J. Howell, "Comparison of the spectral information content of Landsat thematic mapper and spot for three different sites in the Phoenix, Arizona region," *Photogramm. Eng. Remote Sens.*, vol. 54, no. 12, pp. 1699–1708, 1988.
- [14] K. Edwards and P. Davis, "The use of intensity-hue-saturation transformation for producing color shaded relief images," *Photogramm. Eng. Remote Sens.*, vol. 60, no. 11, pp. 1369–1373, 1994.
- [15] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for Ikonos imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, Oct. 2004.
- [16] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.
- [17] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on over-sampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.
- [18] W. J. Emery, F. Pacifici, and M. Chini, "A neural network approach using multi-scale textural metrics from very high resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276–1292, Jun. 2009.
- [19] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.
- [20] 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, 1992.
- [21] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2042–2055, 2002.
- [22] G. Matheron, *Éléments pour une Théorie des Milieux Poreux*. Paris, France: Masson, 1967.
- [23] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pt. 2, pp. 3804–3814, Nov. 2008.
- [24] L. S. Davis, C. Huang, and J. R. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, Feb. 2002.
- [25] G. M. Foody and A. Mathur, "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM," *Remote Sens. Environ.*, vol. 103, no. 2, pp. 179–189, Jul. 2006.
- [26] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
- [27] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [28] J. A. Benediktsson and P. H. Swain, "Consensus theoretic classification methods," *IEEE Trans. Syst., Man, Cybern.*, no. 22, pp. 688–704, 2001.
- [29] P. Salamon and L. K. Hansen, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 12, pp. 993–1001, 1990.
- [30] L. I. Kuncheva, "Fuzzy versus nonfuzzy in combining classifiers designed by boosting," *IEEE Trans. Fuzzy Syst.*, no. 11, pp. 1214–1219, 1990.
- [31] R. E. Schapire and Y. Freund, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Machine Learning*, 1996, pp. 148–156.
- [32] R. E. Schapire, "A brief introduction to boosting," in *Proc. 16th Int. Joint Conf. Artificial Intelligence*, 1999, pp. 1401–1406.
- [33] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, Aug. 1996.
- [34] L. Breiman, "RF/tools—A class of two eyed algorithms," in *SIAM Workshop*, 2003 [Online]. Available: <http://oz.berkeley.edu/users/breiman/siamtalk2003.pdf>.



Frosti Pálsson is currently working towards the M.S. degree in electrical and computer engineering at the University of Iceland, Reykjavik.

His research interests include remote sensing, signal processing and image analysis.



Johannes R. Sveinsson (SM'02) received the B.S. degree from the University of Iceland, Reykjavik, and the M.S. and Ph.D. degrees from Queen's University, Kingston, ON, Canada, all in electrical engineering.

He is currently the Head and Professor with the Department of Electrical and Computer Engineering, University of Iceland, he was with the Laboratory of Information Technology and Signal Processing from 1981 to 1982 and, from November 1991 to 1998, the Engineering Research Institute and the Department

of Electrical and Computer Engineering as a Senior Member of the research staff and a Lecturer, respectively. He was a Visiting Research Student with the Imperial College of Science and Technology, London, U.K., from 1985 to 1986. At Queen's University, he held teaching and research assistantships. His current research interests are in systems and signal theory.

Dr. Sveinsson received the Queen's Graduate Awards from Queen's University.



Jón Atli Benediktsson (F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, in 1984, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1987 and 1990, respectively.

He is currently Pro Rector for Academic Affairs and Professor of Electrical and Computer Engineering at the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has published extensively

in those fields.

Prof. Benediktsson is the President of the IEEE Geoscience and Remote Sensing Society for 2011–2012. He was Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGARS) from 2003 to 2008. He received the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, Dr. Benediktsson was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award, in 2000, he was granted the IEEE Third Millennium Medal, in 2004, he was a co-recipient of the University of Iceland's Technology Innovation Award, in 2006 he received the yearly research award from the Engineering Research Institute of the University of Iceland, and in 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society. He is a member of Societas Scinetiarum Islandica and Tau Beta Pi.



Henrik Aanaes received the Ph.D. degree from the Department of Informatics, Technical University of Denmark, in 2003.

He is now an Associate Professor (Lektor) of computer science at the Department of Informatics, Technical University of Denmark. His main interests include view geometry, image registration and classification as well as sensor fusion, areas in which he has published substantially. His main area of application of this theory is within mapping and remote sensing.

Journal Paper 2

Accepted for publication in *IEEE Geoscience and Remote Sensing Letters*.

A New Pansharpening Algorithm Based on Total Variation

Frosti Palsson, *Member, IEEE*, Johannes R. Sveinsson, *Senior Member, IEEE*
and Magnus O. Ulfarsson, *Member, IEEE*

Abstract—In this letter we present a new method for the pansharpening of multispectral satellite imagery. Pansharpening is the process of synthesizing a high resolution multispectral image from a low spatial resolution multispectral image and a high resolution panchromatic image. The method uses total variation to regularize an ill-posed problem dictated by a widely used explicit image formation model. This model is based on the assumptions that a linear combination of the bands of the pansharpened image gives the panchromatic image and that a decimation of the pansharpened image gives the original multispectral image. Experimental results are based on two real datasets and the quantitative quality of the pansharpened images is evaluated using a number of spatial and spectral metrics, some of which have been recently proposed and do not need a reference image. The proposed method compares favorably to other well known methods for pansharpening and produces images of good spatial and spectral quality.

Index Terms—Image fusion, remote sensing, total variation, pansharpening.

I. INTRODUCTION

PANSHARPENING is the process of fusing a multispectral (MS) image of low spatial resolution with a Panchromatic (PAN) image of high spatial resolution to obtain a MS image of high spatial and spectral resolution.

One area of application where the pansharpened image is useful is classification, since the added spatial resolution increases the classification accuracy [1]. Another application is feature extraction.

Over the years, a number of different pansharpening methods have been proposed and many of these methods are either based on Multiresolution Analysis (MRA) or Component Substitution (CS). There are also hybrid methods based on both MRA and CS. The MRA methods are usually based on methods such as the Undecimated Discrete Wavelet Transform (UDWT) or other kinds of pyramidal representations. Typically, the MS image is upsampled and a multiresolution representation is calculated for this image along with the PAN image. The idea is to use some kind of injection model to replace or enhance the detail of the MS image with details from the PAN image.

The CS methods typically make use of transformations such as Principal Component Analysis (PCA) or a spectral transformation such as Intensity Hue Saturation (IHS) transformation, where a component derived from the MS image is substituted

for a component derived from the PAN image and then the fused or pansharpened image is obtained from the inverse transformation. There are also methods that do not fit either of these categories such as the method proposed in this letter.

We begin by formulating an observational model widely used for the pansharpening process and is for example given in [2]. The model is based on the assumptions that the decimated pansharpened image gives the multispectral image and that the PAN image is a linear combination of the bands of the multispectral image.

The approach we take here is to apply the model to the whole dataset instead of dividing the data into small patches and applying the model to these separately, as was done in [2]. This involves very large matrices but since these operators represent simple operations on images, this makes the implementation of the algorithm fast and straightforward.

Pansharpening is in general an ill-posed problem that needs regularization for optimal results. We have chosen to use Total Variation (TV) regularization which is justified by the fact that the TV of the solution is reduced without penalizing discontinuities, thus resulting in a solution that is essentially free of noise while preserving the fine detail of the panchromatic image. This type of regularization is well known in image denoising and there exist many algorithms to solve this kind of problem.

The algorithm presented here uses Majorization-Minimization (MM) techniques to obtain the solution in an iterative manner. It is related to the work presented in [3] and is shown to be a 1D-version of Chambolle's algorithm. The novelty of the method presented here is the use of the TV regularization in combination with the observational model.

In Section II we derive the observational model. In Section III we derive the TV regularized method. Section IV gives a short overview of the quantitative quality metrics used and Section V describes the datasets and presents the experimental results. Conclusions are drawn in Section VI.

II. OBSERVATIONAL MODEL FOR PANSHARPENING

The dataset consists of a high spatial resolution PAN image \mathbf{y}_{PAN} and the low spatial resolution MS image \mathbf{y}_{MS} . The PAN image has dimensions four times larger than the MS image thus the ratio in pixels is 1 to 16. The MS image contains 4 bands, RGB and near-infrared (NIR).

The PAN image is of dimension $M \times N$ and the MS image is of dimension $m \times n$ where $m = M/4$ and $n = N/4$. The vectors \mathbf{y}_{PAN} and \mathbf{y}_{MS} are the vectorized

All the authors are with the Department of Electrical and Computer Engineering, University of Iceland, Reykjavik, E-mail: (frostip@gmail.com, sveinsson@hi.is, mou@hi.is)

This work was funded by the Research Fund of the University of Iceland.

PAN and MS images, respectively and \mathbf{y}_{MS} is arranged as $\mathbf{y}_{\text{MS}} = \underbrace{[1, \dots, mn]}_{\text{R}}, \underbrace{[1, \dots, mn]}_{\text{G}}, \underbrace{[1, \dots, mn]}_{\text{B}}, \underbrace{[1, \dots, mn]}_{\text{NIR}}]^T$ where mn is the number of pixels in each band. The vectorized pansharpened image \mathbf{x} has this same arrangement where there are MN pixels in each band.

There are two assumptions that define the model. The first is that the low spatial resolution MS image can be described as a degradation (decimation) of the pansharpened image \mathbf{x} . We write this in matrix notation as $\mathbf{y}_{\text{MS}} = \mathbf{M}_1 \mathbf{x} + \epsilon$ where

$$\mathbf{M}_1 = \mathbf{I}_4 \otimes \left((\mathbf{I}_n \otimes \mathbf{1}_{4 \times 1}^T) \otimes (\mathbf{I}_m \otimes \mathbf{1}_{4 \times 1}^T) \right) \quad (1)$$

is a decimation matrix of size $4mn \times 4MN$, \mathbf{I}_4 is an identity matrix of size 4 by 4, \otimes is the Kronecker product and ϵ is zero mean Gaussian noise.

The second assumption is that the PAN image is a linear combination of the bands of the pansharpened image with some additive Gaussian noise. This can be written in matrix notation as $\mathbf{y}_{\text{PAN}} = \mathbf{M}_2 \mathbf{x} + \epsilon$ where ϵ is zero mean Gaussian noise and

$$\mathbf{M}_2 = [\omega_1 \mathbf{I}_{MN}, \omega_2 \mathbf{I}_{MN}, \omega_3 \mathbf{I}_{MN}, \omega_4 \mathbf{I}_{MN}], \quad (2)$$

where $\omega_1, \dots, \omega_4$ are constants that sum to one. These constants determine the weight of each band in the PAN image.

Now \mathbf{M}_1 and \mathbf{M}_2 have the same number of columns and thus we can combine the expressions for \mathbf{y}_{MS} and \mathbf{y}_{PAN} into a single equation which is the classical observational model,

$$\mathbf{y} = \mathbf{M} \mathbf{x} + \epsilon, \quad (3)$$

where $\mathbf{y} = [\mathbf{y}_{\text{MS}}^T, \mathbf{y}_{\text{PAN}}^T]^T$ and $\mathbf{M} = [\mathbf{M}_1^T, \mathbf{M}_2^T]^T$.

III. TV REGULARIZATION

TV regularization encourages noise removal while preserving edges. By penalizing solutions with high TV, the estimate is forced to be smooth without necessarily penalizing sharp boundaries since the estimated boundary or edge is decided by the observed data.

One can define the TV of the MS image as

$$\text{TV}(\mathbf{x}) = \left\| \sqrt{(\mathbf{D}_H \mathbf{x})^2 + (\mathbf{D}_V \mathbf{x})^2} \right\|_1, \quad (4)$$

where \mathbf{x} is the vectorized 4 band MS image, $\mathbf{D}_H = (\mathbf{I}_4 \otimes \mathbf{D}_h)$, $\mathbf{D}_V = (\mathbf{I}_4 \otimes \mathbf{D}_v)$ and the matrices \mathbf{D}_h and \mathbf{D}_v are defined such that when multiplied by a vectorized image they give the first order differences in the horizontal direction and vertical direction, respectively. The cost function of the TV regularized problem can be formulated as

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M} \mathbf{x}\|_2^2 + \lambda \text{TV}(\mathbf{x}). \quad (5)$$

Minimizing this cost function is difficult because the TV functional is not differentiable. However, MM techniques can be used to replace this difficult problem with a sequence of easier ones

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} Q(\mathbf{x}, \mathbf{x}_k), \quad (6)$$

where \mathbf{x}_k is the current iterate and $Q(\mathbf{x}, \mathbf{x}_k)$ is a function that majorizes the cost function $J(\mathbf{x})$. This means that $Q(\mathbf{x}, \mathbf{x}_k) \geq J(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{x}_k$ and $Q(\mathbf{x}, \mathbf{x}_k) = J(\mathbf{x})$ for $\mathbf{x} = \mathbf{x}_k$. By

iteratively solving (6), \mathbf{x}_k will converge to the global minimum of $J(\mathbf{x})$.

A majorizer for the TV term [4] can be written using matrix notation as

$$Q_{\text{TV}}(\mathbf{x}, \mathbf{x}_k) = \mathbf{x}^T \mathbf{D}^T \mathbf{\Lambda}_k \mathbf{D} \mathbf{x} + c \quad (7)$$

where we define

$$\mathbf{w}_k = \left(2\sqrt{(\mathbf{D}_H \mathbf{x}_k)^2 + (\mathbf{D}_V \mathbf{x}_k)^2} \right)^{-1} \quad (8)$$

and $\mathbf{\Lambda}_k = \text{diag}(\mathbf{w}_k, \mathbf{w}_k)$ and the matrix \mathbf{D} is defined as $\mathbf{D} = [\mathbf{D}_H^T, \mathbf{D}_V^T]^T$. When the data fidelity term, $\|\mathbf{y} - \mathbf{M} \mathbf{x}\|_2^2$, is expanded, it is evident that we end up with the term $\mathbf{M}^T \mathbf{M}$ in the solution. In order to avoid having to find the inverse of this term later in the solution, it is necessary to introduce the function

$$Q_{\text{DF}}(\mathbf{x}, \mathbf{x}_k) = (\mathbf{x} - \mathbf{x}_k)^T (\alpha \mathbf{I} - \mathbf{M}^T \mathbf{M}) (\mathbf{x} - \mathbf{x}_k), \quad (9)$$

where $\alpha > \max \text{eig}(\mathbf{M}^T \mathbf{M})$. This ensures that $Q_{\text{DF}}(\mathbf{x}, \mathbf{x}_k)$ is positive definite and $Q(\mathbf{x}, \mathbf{x}_k)$ is still a majorizer for $J(\mathbf{x})$. The function to minimize becomes

$$Q(\mathbf{x}, \mathbf{x}_k) = \|\mathbf{y} - \mathbf{M} \mathbf{x}\|_2^2 + Q_{\text{DF}}(\mathbf{x}, \mathbf{x}_k) + \lambda Q_{\text{TV}}(\mathbf{x}, \mathbf{x}_k). \quad (10)$$

Differentiating this function w.r.t. \mathbf{x} , setting to zero and subsequently solving for \mathbf{x} gives

$$(\alpha \mathbf{I} + \lambda \mathbf{D}^T \mathbf{\Lambda}_k \mathbf{D}) \mathbf{x} = \mathbf{b}, \quad (11)$$

where

$$\mathbf{b} = \alpha \mathbf{x}_k + \mathbf{M}^T (\mathbf{y} - \mathbf{M} \mathbf{x}_k). \quad (12)$$

The matrix inversion lemma gives

$$(\alpha \mathbf{I} + \lambda \mathbf{D}^T \mathbf{\Lambda}_k \mathbf{D})^{-1} = \frac{1}{\alpha} (\mathbf{I} - \mathbf{D}^T (\frac{\alpha}{\lambda} \mathbf{\Lambda}_k^{-1} + \mathbf{D} \mathbf{D}^T)^{-1} \mathbf{D}). \quad (13)$$

This gives $\mathbf{x} = \frac{1}{\alpha} (\mathbf{b} - \mathbf{D}^T \mathbf{z})$ where we have defined

$$\mathbf{z} = (\frac{\alpha}{\lambda} \mathbf{\Lambda}_k^{-1} + \mathbf{D} \mathbf{D}^T)^{-1} \mathbf{D} \mathbf{b}. \quad (14)$$

From (14) we have $(\frac{\alpha}{\lambda} \mathbf{\Lambda}_k^{-1} + \mathbf{D} \mathbf{D}^T) \mathbf{z} = \mathbf{D} \mathbf{b}$ and by adding $c \mathbf{z}$, where the constant $c > \max \text{eig}(\mathbf{D} \mathbf{D}^T)$, and by subtracting $\mathbf{D} \mathbf{D}^T \mathbf{z}$ from both sides of this equation we obtain

$$(\frac{\alpha}{\lambda} \mathbf{\Lambda}_k^{-1} + c \mathbf{I}) \mathbf{z} = \mathbf{D} \mathbf{b} + (c \mathbf{I} - \mathbf{D} \mathbf{D}^T) \mathbf{z}, \quad (15)$$

which leads to the update rule

$$\mathbf{z}_{k+1} = (\frac{\alpha}{\lambda} \mathbf{\Lambda}_k^{-1} + c \mathbf{I})^{-1} (\mathbf{D} \mathbf{b} + (c \mathbf{I} - \mathbf{D} \mathbf{D}^T) \mathbf{z}_k) \quad (16)$$

$$\mathbf{x}_{k+1} = \frac{1}{\alpha} (\mathbf{b} - \mathbf{D}^T \mathbf{z}_{k+1}). \quad (17)$$

We can see that finding the inverse in (16) is trivial since it is a diagonal matrix.

The implementation of the above algorithm becomes straightforward once it is realized that all the matrix multiplications involving the operators \mathbf{D} , \mathbf{D}^T , \mathbf{M} and \mathbf{M}^T can be implemented as simple operations on multispectral images. This significantly reduces computation time and memory requirements due to the enormous size of the matrices involved.

However, there is one important thing to note. Bearing in mind the structure of the model matrix \mathbf{M} , it is clear that

the multiplication with M^T in (12) indicates nearest neighbor interpolation (the inverse of decimation) of an MS image. By using bilinear decimation and interpolation instead gives better results, both according to quality metrics and visual inspection.

IV. QUANTITATIVE QUALITY METRICS

The quantitative quality assessment of pansharpened imagery is often difficult since a reference high resolution MS image is not available. When there is no reference image available, the PAN and MS images are degraded in resolution such that the resulting pansharpened image has the same size as the original MS image and can thus be used as a reference. However, there are metrics which do not need a reference image, such as the Quality Without Reference (QNR) spectral and spatial metrics.

When there is no reference image available it is difficult to make objective comparisons, however, a possible solution is to make use of Wald's consistency property [5] which states that the pansharpened image degraded to the resolution of the MS image should be very similar to the original MS image. The spectral metrics without reference attempt to measure this spectral consistency.

In this letter we use the SAM [6] (Spectral Angle Mapper), ERGAS (relative dimensionless global error synthesis) [7], Q4 [8], CC and Qave as metrics that need a reference image and we use the QNR D_λ spectral [9] and QNR D_s spatial [9] metrics for quality assessment without reference. Details on ERGAS, SAM, CC and Qave can be found in [1].

The QNR D_λ spectral distortion metric is given by

$$D_\lambda = \frac{1}{N_B(N_B - 1)} \sum_{l=1}^{N_B} \sum_{r=1, r \neq l}^{N_B} \left| Q(F_l, F_r) - Q(\hat{F}_l, \hat{F}_r) \right|, \quad (18)$$

where N_B is the number of bands, F is the pansharpened image and \hat{F} is the MS image and Q is the Universal Image Quality Index (UIQI) [10]. Similarly, the QNR D_s spatial distortion is given by

$$D_s = \frac{1}{N_B} \sum_{l=1}^{N_B} \left| Q(F_l, P) - Q(\hat{F}_l, \hat{P}) \right|, \quad (19)$$

where P is the PAN image and \hat{P} is the degraded PAN image of same size as the MS image. UIQI is a metric that attempts to model the spectral distortion as a product of three factors, i.e., loss of correlation, luminance distortion and contrast distortion.

The Q4 metric is an extension of the UIQI metric to images with 4 bands. This is done by considering the 4 values for each pixel as a quaternion, hence this metric can be viewed as the hypercomplex extension of UIQI. Further details can be found in [8].

Both the UIQI and Q4 metrics use a sliding window for the calculations. We use 64 by 64 pixel window for high resolution images and 16 by 16 pixel window for low resolution images, as was done in [11].

The other metrics used in this letter are CC and Qave, CC is simply the average correlation coefficient between the bands of the pansharpened image and the reference image, while

Qave is defined as the mean UIQI between the bands of the pansharpened image and the reference image.

V. EXPERIMENTAL RESULTS

A. Datasets

We present experimental results using simulated Pléiades data [11] and Quickbird data [11]. The Pléiades dataset consists of a 0.8m resolution PAN image of dimension 1024 by 1024 pixels and a 3.2m resolution 256 by 256 pixel MS image. There is also a simulated high-resolution MS image of same size and resolution as the PAN image. This makes it possible to use the metrics that use a reference for this dataset.

The other dataset is a high resolution Quickbird image of the Mississippi State University campus. The PAN image is 2048 by 2048 pixels and of resolution 0.7m and the MS image is 512 by 512 pixels and of resolution 2.8m. For this dataset there is no high resolution reference image available.

In order to be able to use the low spatial resolution MS image as a reference, we perform experiments using degraded MS and PAN images. This gives a pansharpened image of the same size as the original low spatial resolution MS image. The images are degraded using bilinear interpolation which is consistent with the use of bilinear decimation and interpolation in the implementation of the proposed method. We also present results using the Quickbird dataset at full scale using QNR metrics only.

B. Comparison to other methods

We compare the proposed method to three state of the art pansharpening methods. The method called UDWT [1] uses the undecimated wavelet transform to decompose the PAN image and the bicubic expanded MS image into their coarse (LL) and detail parts (HH). The LL part of the PAN decomposition is replaced with the LL part of the MS image and finally the inverse transform yields the pansharpened image. This method produces images with excellent spatial quality. Because of its simplicity there are no tuning parameters other than number of levels of decomposition, which was set to 2 and the type of filter, where we use the Daubechies filter of length 2.

The Principal Component Analysis (PCA) method [12] is a method which belongs to the CS family of methods and finally the P+XS [13] method is based on a variational model. The PCA method implementation we used has no parameters to tune and the number of components is fixed at 4.

The P+XS method is more complex and there are several parameters to tune. However, most of them were left at their default values. The parameters that matter most are the weights for each multispectral band in the fused image and the regularization term λ . We used the same weights as for the proposed method and we chose $\lambda = 5$ which was shown to give optimal results based on sensitivity analysis.

C. Choice of parameters

There are several parameters that affect the performance of the proposed method. The constants α and c affect the convergence of the algorithm and in all the experiments they

were chosen as 0.75 and 8, respectively. Sensitivity analysis of λ shows that for imagery at full scale, the performance of the algorithm is very good for small values of λ . Based on this we chose $\lambda = 2$ for imagery at full scale. When using degraded imagery, the ERGAS metric, which is based on the mean square error, has a global minimum at $\lambda = 1.5$, which is the value we choose.

The constants ω_i that determine the weights of the bands of the pansharpened image are important for spectral quality. Choosing the value of 0.25 for each ω_i gives good results, but there is room for improvement, since in reality the PAN image is not the average of the bands. These constants were selected for each dataset by doing a grid search over the parameter space, evaluating each possible combination of the parameters using the ERGAS metric. The values chosen for the Pléiades dataset are 0.30, 0.26, 0.22 and 0.22 for the red, green, blue and NIR bands, respectively and for the Quickbird dataset as 0.21, 0.21, 0.21 and 0.37, respectively.

D. Quantitative Quality Assessment Results

Tables I-III show the results of the quantitative quality evaluations of the different pansharpening algorithms. The last row in these tables, called Expanded is where the expanded or upsampled (using bicubic interpolation) MS image is used as the fused image. This image is used as a baseline reference, i.e., this is the most basic method for pansharpening.

Table I shows the results for the Pléiades data. Since a high resolution MS image is available for this dataset, the pansharpened image can be evaluated at full scale using all the metrics discussed in Section IV. The table shows that the proposed method gives best results in every metric using reference, while the P+XS and UDWT methods give slightly better results for QNR spectral and spatial metrics. The high values of the CC, Qave and Q4 metrics indicate good spatial detail as well as good spectral quality. The CPU time in seconds for each algorithm, in the same order as they appear in the table, is 1.8, 0.2, 83.4 and 15.4, respectively. The computer specification is i5-2400, 3.1 Ghz CPU with 12GB of memory. All methods are implemented in Matlab.

Table II shows the results for the degraded Quickbird data. Note how similar values are for the QNR spectral and spatial metrics. This is very probably a result of the degradation process. The proposed method gives the best results for all metrics except the QNR spectral metric, where surprisingly, PCA gives slightly better result. The CPU time for the experiments in Table II is 0.3, 0.1, 20.3 and 3.2, respectively and in the same order as before.

Finally, Table III shows the results using the Quickbird dataset at full scale. At this scale, no reference image is available and thus we only use the QNR metrics to assess the quality of the pansharpened image. The proposed method is shown to give better results for both the spectral and spatial QNR metrics than the other methods. The CPU time in seconds for the experiments in Table III is 6.1, 0.9, 335.2 and 57.7, respectively.

In Table III we see that the best results for QNR D_λ are

TABLE I: Comparison of pansharpening methods using the simulated Pléiades data at full scale. The best results are denoted using bold font.

	SAM	ERGAS	CC	Qave	Q4	D_λ	D_s
Reference	0	0	1	1	1	0.011	0.040
UDWT [1]	5.178	4.424	0.947	0.913	0.890	0.044	0.020
PCA [12]	6.964	7.681	0.853	0.678	0.670	0.073	0.201
P+XS [13]	5.108	4.965	0.951	0.912	0.894	0.017	0.024
Proposed	4.213	3.119	0.972	0.960	0.945	0.039	0.044
Expanded	4.905	6.345	0.869	0.792	0.799	0.002	0.122

for the expanded image. This seems to indicate that the QNR metrics are not very reliable..

TABLE II: Comparison of pansharpening methods using the degraded Quickbird data. The best results are denoted using bold font.

	SAM	ERGAS	CC	Qave	Q4	D_λ	D_s
Reference	0	0	1	1	1	0.020	0.029
UDWT [1]	5.038	3.411	0.933	0.885	0.912	0.100	0.102
PCA [12]	4.821	4.835	0.930	0.784	0.818	0.071	0.097
PXS [13]	4.513	2.982	0.950	0.914	0.931	0.117	0.113
Proposed	3.520	2.448	0.965	0.942	0.960	0.073	0.086
Expanded	4.439	4.444	0.893	0.813	0.812	0.006	0.119

TABLE III: Comparison of pansharpening methods using the Quickbird data at full scale. The best results are denoted using bold font. Only metrics that don't need reference are used.

	D_λ	D_s
UDWT [1]	0.048	0.055
PCA [12]	0.093	0.113
P+XS [13]	0.060	0.079
Proposed	0.027	0.042
Expanded	0.002	0.045

E. Visual Comparison

A visual comparison of the results for the Pléiades data using all methods is shown in Fig. 1. Only a subset of the data is shown in order to display more detail. The images have been gamma corrected in order to make them more visually pleasing. Visual inspection shows that all the methods show good level of detail while the PCA results look very spectrally distorted.

Figure 2 shows a subset of the pansharpened images using the Quickbird data at full scale for all methods. Again the PCA method shows considerable spectral distortion while the proposed and P+XS method give best results. The UDWT image shows some noticeable spectral distortion while the spatial resolution is excellent which is the primary strength of this method.

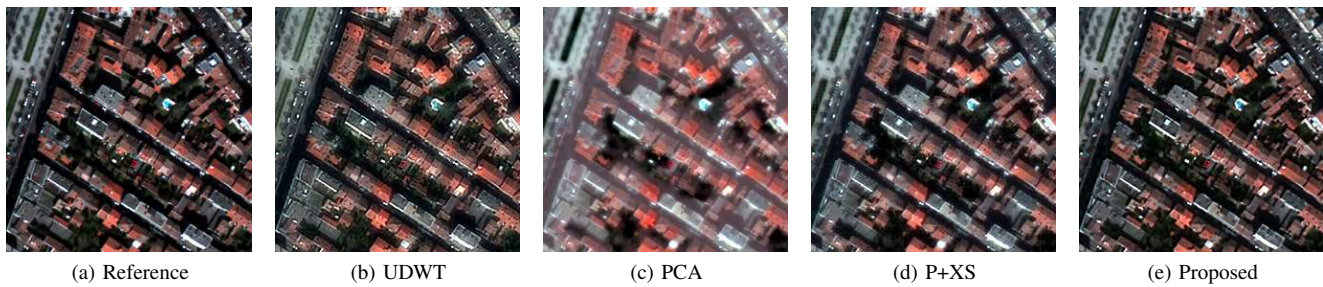


Fig. 1: Pansharpening results for the Pléiades data. Note that only a subset of the pansharpened images is shown.



Fig. 2: Pansharpening results for the full scale Quickbird data. Note that only a subset of the pansharpened images is shown.

VI. CONCLUSIONS

In this letter we have proposed a new pansharpening method based on an observational model which is regularized using TV. The fact that the model is applied to the whole dataset results in computations involving very large matrices. However, since these operations can be implemented as operations on images, the resulting algorithm is quite fast. We performed experiments using datasets from two different sensors, Pléiades and Quickbird and compared our method to three other well known state of the art methods, representing three important paradigms in pansharpening, and evaluated the results using a number of quantitative quality metrics. The results show that the proposed method consistently outperformed the other methods used in virtually all experiments for all quality metrics used. Visual inspection shows that the method produces pansharpened images that have excellent spatial and spectral quality.

ACKNOWLEDGMENT

The authors would like to thank Jocelyn Chanussot, GIPSA-Lab, Grenoble Institute of Technology, for kindly providing the Quickbird and the Pléiades dataset. The authors would also like to thank the Editor and anonymous reviewers for their constructive comments that greatly improved this letter.

REFERENCES

- [1] 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.
- [2] 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.
- [3] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," *In Proceedings of SPIE*, vol. 7446 (Wavelets XIII), August 2009.
- [4] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006*, vol. 2, May 2006, p. II.
- [5] L. Wald, *Data Fusion: Definitions and Architectures : Fusion of Images of Different Spatial Resolutions*. Les Presses de l'École des Mines, 2002.
- [6] 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 3rd Annual JPL Airborne Geoscience Workshop*, vol. 1, 1992, pp. 147–149.
- [7] 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"*, T. Ranchin and L. Wald, Eds. SEE/URISCA, 2000.
- [8] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *Geoscience and Remote Sensing Letters, IEEE*, vol. 1, no. 4, pp. 313–317, Oct. 2004.
- [9] 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.
- [10] Z. Wang and A. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81–84, march 2002.
- [11] M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 11, pp. 3880–3891, Nov. 2009.
- [12] 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, 1992.
- [13] 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.