A Variational Pan-Sharpening with Local Gradient Constraints

Xueyang Fu[12], Zihuang Lin[1], Yue Huang[1], Xinghao Ding[1*]

[1]Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University, China

[2]School of Information Science and Technology, University of Science and Technology of China, China

[*]Corresponding author: dxh@xmu.edu.cn

Abstract

Pan-sharpening aims at fusing spectral and spatial information, which are respectively contained in the multi-spectral (MS) image and panchromatic (PAN) image, to produce a high resolution multi-spectral (HRMS) image. In this paper, a new variational model based on a local gradient constraint for pan-sharpening is proposed. Different with previous methods that only use global constraints to preserve spatial information, we first consider gradient difference of PAN and HRMS images in different local patches and bands. Then a more accurate spatial preservation based on local gradient constraints is incorporated into the objective to fully utilize spatial information contained in the PAN image. The objective is formulated as a convex optimization problem which minimizes two least-squares terms and thus very simple and easy to implement. A fast algorithm is also designed to improve efficiency. Experiments show that our method outperforms previous variational algorithms and achieves better generalization than recent deep learning methods.

1. Introduction

Remote sensing images have become widely used in many practical applications, such as environmental monitoring, object positioning and classification. Due to physical constraints, satellites such as IKONOS, QuickBird-2, WorldView-2 and WorldView-3 capture two images of the same scene at the same time, where one image called panchromatic (PAN) image is of high spatial resolution, and the other called multi-spectral (MS) image is of low spatial resolution but it contains good spectral content. In order to obtain the high resolution multi-spectral (HRMS) images, pan-sharpening techniques which refer to fuse the low resolution spectral information with the spatial structure in the PAN image have been developed.

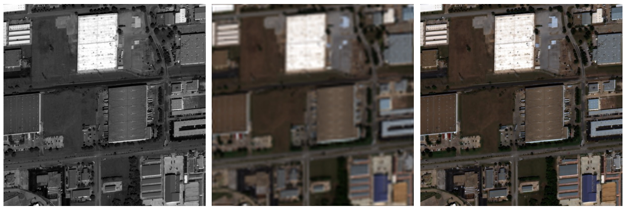

(a) PAN image (b) MS image (c) Fused result

Figure 1: An example of our proposed method. The fused result has rich details with promising spectral preservation.

1.1. Related works

In the past decades, many pan-sharpening methods have been proposed. Among these existing methods, the most common methods include the intensity hue-saturation technique (IHS) [8], the principal component analysis (PCA) [21] and Brovey transform [15]. These methods are popular due to their relatively fast computation. But they usually suffer from spectral distortion while increasing spatial resolution of fused results.

Beside component substitution, the multi-resolution analysis (MRA) method is another popular pan-sharpening method in which the PAN image and MS image are decomposed into other planes by using some multi-resolution tools, e.g., decimated wavelet transform (DWT) [22], a trous wavelet transform (ATWT) [26] and Laplacian pyramid (LP) [7]. The MRA method can sharpen MS image effectively. However, this may cause some local dissimilarities because the high frequencies extracted from the PAN image are not exactly to those of the HRMS images.

Recently, in the light of the strong nonlinear mapping ability of deep learning, researchers have begun exploring the deep convolutional neuron network based methods. Although these methods [17, 23, 30] obtain excellent per-

This work was supported in part by the National Natural Science Foundation of China under Grants 61571382, 81671766, 61571005, 81671674, 61671309 and U1605252, in part by the Fundamental Research Funds for the Central Universities under Grants 20720160075 and 20720180059, in part by the CCF-Tencent open fund, and the Natural Science Foundation of Fujian Province of China (No.2017J01126).

formance, they require substantial computational resources and training data, of which the latter is not easy obtained in the pan-sharpening area since there is no true ground-truth. Since all deep learning methods use synthetic data for training, their generalization performance for real-world data and new satellite is limited.

So from a practical perspective, variational methods are reconsidered to pan-sharpening field. These methods [4, 5, 10, 13, 18, 32] achieve pan-sharpening by modeling the relationship between PAN, MS and HRMS images into an objective function with some prior knowledge, which is universal and independent on specific training data. The first variational pan-sharpening method P+XS technique [5] preserves spectral well, but produces blurring effects. To struggle against the blurred edges, a large part of methods introduce a high-pass filter to describe structural similarity while minimizing spectral distortion, such as guided filter-based fusion (GDF) [13], Bayesian nonparametric dictionary learning (BNDL) [12] and satellite image registration and fusion (SIRF) [10]. However, they still suffer some degradation due to indecent structural constraints.

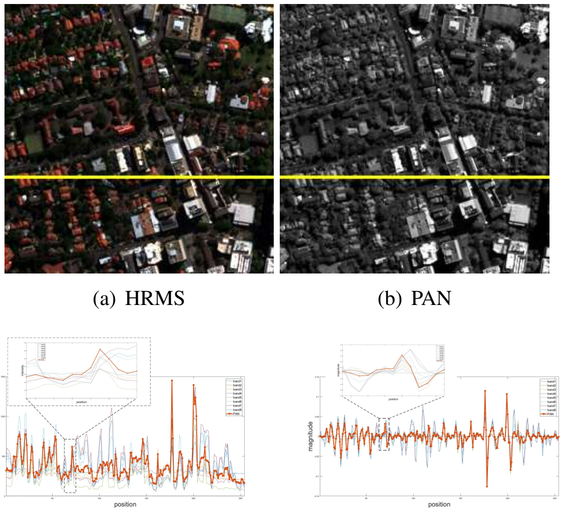
1.2. Our contributions

In this paper, for the pan-sharpening problem, we focus on spatial improvement by considering local gradient constraints while keeping the spectral information as undistorted as possible. To improve spatial resolution, recent variational methods make assumptions based on the gradient difference of PAN and HRMS images. For example, MBF [4] assumes this relationship follows the Gaussian distribution while PHLP [18] considers it obeys the Laplacian distribution through statistical experiments.

All the previous methods assume that the gradient difference of PAN and HRMS is global linear. However, we find that the relationship is not consistent in different local image patches. To verify this viewpoint, we randomly scan one line from a 8-band image and present the difference of gradient values among the PAN image and each band of HRMS image in Figure 2(d). Obviously, it is unreasonable to model the gradient relationships between PAN and HRMS images with only a global linear function. Therefore, to avoid global constraints from limiting the modeling flexibility, a new variational model based on local gradient constraints is proposed. We formulate our objective function to consider the following two aspects:

- Spectral preservation: we assume that the down-sampled HRMS image should be close to the original MS image, which aims at preserving the exact spectral information without introducing false information.

- Spatial improvement: a simple yet effective local linear regression model is proposed to constraint the gradient difference of PAN and HRMS images, so as to effectively utilize the spatial information of PAN image. To the best of our knowledge, this is the first variational model for the specific pan-sharpening problem that based on the local constraint.



(a) HRMS (b) PAN

(c) Intensity scanning (d) Gradient scanning

Figure 2: 1D signals of intensity and gradient values in one line of HRMS and PAN images.

We show that by using our local gradient constraints, a simple least-square term, which is easy to optimize, is sufficient to model spatial preservation. To optimize the objective function, a fast iterative shrinkage-thresholding algorithm (FISTA) is designed. Experiments show that our method has a great advantage over non deep learning methods, both subjectively and objectively. Moreover, since we adopt the universal local constraint, our proposed method has a better generalization ability than deep learning based methods that adopt the supervised learning strategy.

2. Motivation

At first, some auxiliary notations and definitions are introduced to simplify our analysis, which will be used in the following paper. The satellite typically captures two kinds of images including a PAN image and a corresponding MS image which has B bands (*e.g.*, $B = 8$ for WorldView-2 satellite). We denote the observed PAN image as $P \in \mathbb{R}^{M\times N}$ and $\mathbf{P} \in \mathbb{R}^{M\times N\times B}$ represents P that expanded to B bands. The corresponding MS image is denoted by $\mathbf{M} \in \mathbb{R}^{\frac{M}{c}\times\frac{N}{c}\times B}$ and the pan-sharpened HRMS image is denoted by $\mathbf{X} \in \mathbb{R}^{M\times N\times B}$, where c is a reduction ratio.

Since both the PAN and MS images are taken from the same scene, the spatial structure of them should have a strong similarity. The PAN image contains abundant spatial information, which makes it play a significant role in improving the spatial resolution of the MS image. The first P + XS method [5] assumes that the PAN image can be modeled as a global linear combination among all bands of the HRMS image, *i.e.*,

$$\sum\nolimits_{b=1}^{B} w_b X_b = P + \varepsilon. \tag{1}$$

However, even for the same object, different sensor has different response. In other words, differences in intensity of the HRMS image and the PAN image may be very large, as shown in Figure 2(c). To avoid this drawback, recent approaches ensure the consistency of the high-pass filtered components of PAN image and HRMS image. This requirement, which enforces structure similarity rather than intensity similarity, is based on the following assumption:

$$\sum\nolimits_{b=1}^{B} w_b \nabla X_b = \nabla P + \varepsilon, \tag{2}$$

where ∇ represents the gradient. To enforce spatial resolution, previous variational pan-sharpening methods often use the ℓ_2 norm [4] to enforce spatial resolution, or switch to ℓ_1 [25] when sparsity is desired. In SIRF [10], group sparsity is encouraged by introducing the $\ell_{2,1}$ norm. However, according to empirical image statistics, assuming the error ε obeys Gaussian (ℓ_2) or Laplacian (ℓ_1) assumptions are not as appropriate as heavy-tailed distribution such as a hyper-Laplacian [20]. Thus, the PHLP method [18] which adopts the $\ell_{1/2}$ penalty on the gradients of the reconstruction error is introduced to enforce structural preservation.

Although spatial improvement is achieved by using different sparse assumptions, we argue that modeling based on equation (2) is not appropriate. First, equation (2) is built from a global perspective, which is a relative rough assumption. As shown in Figure 2(d), setting the weight w as a global parameter cannot well model the local relationship between ∇X_b and ∇P. Second, most previous methods simply set w as $1/B$, which further reduce the modeling flexibility. Thus, based on the above analysis, we propose a new local linear model to better describe the relationship between ∇X and ∇P at each band:

$$\nabla x_i = a_k \nabla p_i + c_k, \forall i \in \omega_k, \tag{3}$$

where ω_k represents an image block centered at location k. For a random pixel $i \in \omega_k$, ∇x_i and ∇p_i are the intensity of ∇X and ∇P at location i, a_k and c_k are the linear coefficients which are constants in the local area ω_k. We easily find that equation (2) is a special form of our model with $a_k = 1/w_b$ and $c_k = \varepsilon/w_b$. Thus, our model equation (3) can be seen as a general form of previous methods.

To get a_k and c_k, we minimize this objective function:

$$\min_{a_k, c_k} \sum\nolimits_{i \in \omega_k} (\nabla x_i - a_k \nabla p_i - c_k)^2. \tag{4}$$

Let the derivative of the equation (4) be zero, we can get:

$$\begin{aligned} a_k &= \frac{\frac{1}{|\omega|}\sum_{i\in\omega_k} \nabla x_i \nabla p_i - \mu(\nabla x_k)\mu(\nabla p_k)}{\sigma^2(\nabla p_k) + \varepsilon}, & (5) \\ c_k &= \mu(\nabla x_k) - a_k \mu(\nabla p_k), & (6) \end{aligned}$$

where μ and σ^2 are the mean and variance, respectively. ε is a very small parameter to prevent the denominator from being zero. Note that when $\varepsilon \to 0$, a_k can be rewritten to:

$$\begin{aligned} a_k &= \frac{cov(\nabla x_k, \nabla p_k)}{\sigma^2(\nabla p_k)} = \frac{cov(\nabla x_k, \nabla p_k)}{\sigma(\nabla p_k)\sigma(\nabla x_k)} \cdot \frac{\sigma(\nabla x_k)}{\sigma(\nabla p_k)} \\ &= \rho(\nabla p_k, \nabla x_k) \cdot \frac{\sigma(\nabla x_k)}{\sigma(\nabla p_k)}, \end{aligned} \tag{7}$$

where the $cov()$ is the covariance and ρ is the correlation coefficient. Equation (7) means when p_k contains structures that do not exist in x_k, $\rho(\nabla p_k, \nabla x_k)$ is very small and a_k tends to zero, and our model can greatly reduce the effect of ∇p_k on ∇x_k and vice versa. Moreover, shown in Figure 3, both a and c can be positive or negative. This implies that our assumption is much more robust than the equation (2) which hypothesizes the coefficients are global constants.

However, a pixel i is involved in all the overlapping windows ω_k that covers i, so the value of ∇x_i in equation (3) is not identical when it is computed in different windows. Thus, we follow the strategy of the guided filter [16] to average coefficients of all windows overlapping i by

$$\nabla x_i = \bar{a}_i \nabla p_i + \bar{c}_i, \tag{8}$$

where $\bar{a}_i = \frac{1}{|\omega|}\sum_{k\in\omega_i} a_k$ and $\bar{c}_i = \frac{1}{|\omega|}\sum_{k\in\omega_i} c_k$.

To clearly illustrate different value of a and c, we calculate the value of a and c on 65,536 pixels by using equations (5) and (6), the results are shown in Figure 3. We found a and c have different values at different local patches, which demonstrates that the previous global assumption is inaccurate by simply setting w as a positive value that equals $1/B$. We therefore believe that our local linear model (3) is more reasonable than the global one (2) for the spatial preservation.

3. Modeling

Most previous pan-sharpening methods firstly up-sample the multi-spectral image to obtain low resolution multi-spectral (LRMS) image **M** to the same size with the PAN image **P**, and then propose a spectral preservation based on the LRMS image to obtain the HRMS image **X**. However, on the one hand, the up-sampling approach will introduce incorrect information, and different up-sampling approaches may also affect the results. On the other hand, the objective function should be added more prior regularization according to the specific up-sampling way. Therefore, instead

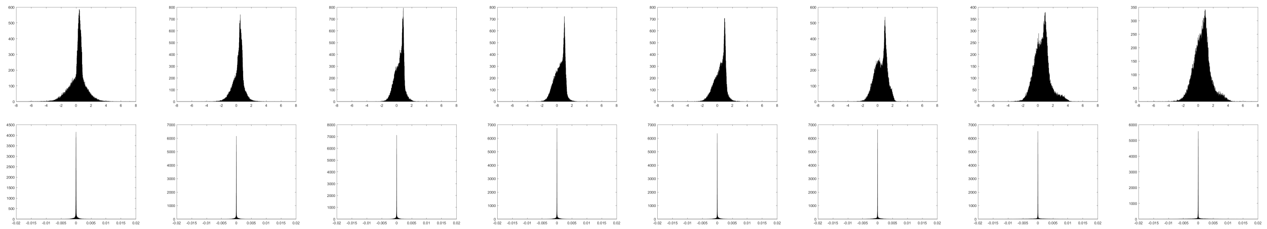


Figure 3: According to the equation (3), we show statistical values of calculated a and c of all 65,536 pixels of each band from a WorldView-2 image of size 256 × 256. As can be seen, both a and c have various sign and magnitudes.

of up-sampling MS image, we argue that the down-sampled HRMS image should be consistent with the MS image. The following equation is presented for spectral preservation:

$$f_1(\mathbf{X}, \mathbf{M}) = \frac{1}{2} \|\psi \mathbf{X} - \mathbf{M}\|_2^2 , \tag{9}$$

where ψ denotes a down-sampling operator.

Based on the above analysis in section 2, the spatial preservation term can be written as follows:

$$\begin{aligned} & f_2(\mathbf{X}, \mathbf{A}, \mathbf{C}, \mathbf{P}) \\ & = \sum_{b=1}^{B} \sum_{k} \sum_{i \in \omega_k} (\nabla x_{b,i} - a_{b,k} \nabla p_i - c_{b,k})^2 , \end{aligned} \tag{10}$$

where $\mathbf{A}$ and $\mathbf{C}$ are matrix form of a and c. Note that since each pixel of ∇x at location i has its own coefficients a_i and c_i, for the entire image, equation (10) can be written in matrix form via the ℓ_2 regularization:

$$f_2(\mathbf{X}, \mathbf{A}, \mathbf{C}, \mathbf{P}) = \frac{1}{2} \|\nabla \mathbf{X} - \mathbf{A} \cdot \nabla \mathbf{P} - \mathbf{C}\|_2^2 , \tag{11}$$

where $\cdot$ is the element-wise multiplication. Thus, the final objective function composed of the two energy functions can be rewritten to

$$\min_{\mathbf{X}, \mathbf{A}, \mathbf{C}} \mathcal{L} = \min_{\mathbf{X}, \mathbf{A}, \mathbf{C}} f_1(\mathbf{X}, \mathbf{M}) + \lambda f_2(\mathbf{X}, \mathbf{A}, \mathbf{C}, \mathbf{P}), \tag{12}$$

where λ is a regularization parameter.

Compared with previous model, the proposed one has several advantages. First, the down-sampled way can make better use of the spectral information of the observed LRMS image, which reduces the spectral distortion. Second, the local regularization term preserves the differences in different bands and patches simultaneously, making the gradient constraints more refined. Last but not least, both f_1 and f_2 are least-square terms, which make it easy to optimize.

4. Optimization

In this section, our goal is to minimize the objective function (12). We first use Bregman iteration to solve the model. By decomposing the problem into three sub-problems, each problem can be solved in a closed form. We summarize our optimization for pan-sharpening in **Algorithm 1**.

Update for X: first, we introduce an auxiliary variable $\mathbf{X}_g = \mathbf{A} \cdot \nabla \mathbf{P} + \mathbf{C}$, then the objective function (12) becomes:

$$\mathcal{L} = \frac{1}{2} \|\psi \mathbf{X} - \mathbf{M}\|_2^2 + \frac{\lambda}{2} \|\nabla \mathbf{X} - \mathbf{X}_g\|_2^2 . \tag{13}$$

This is a simple least-square optimization problem. Since ψ can not be written in matrix form, the FISTA framework [6] is applied to optimize the model to separate ψ. Under the FISTA framework, the objective function is split into the following iterative procedure:

$$\mathbf{Y} = \mathbf{Y}^j - \psi^{-1}(\psi \mathbf{X} - \mathbf{M})/L, \tag{14}$$

where ψ^{-1} denotes the inverse operator of ψ, j is the jth iteration. L is the Lipschitz constant for $\psi^{-1}(\psi \mathbf{X} - \mathbf{M})$. Then $\mathbf{X}$ can be obtained by solving the following function:

$$\begin{aligned} \mathbf{X}^j & = \underset{\mathbf{X}}{argmin} \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_2^2 + \frac{\lambda}{2} \left\| \nabla \mathbf{X} - \mathbf{X}_g^j \right\|_2^2 \\ & = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(\mathbf{Y}) + \lambda \left(\mathcal{F}(\nabla_x)^* \mathcal{F}(\mathbf{X}_g^j) + \mathcal{F}(\nabla_y)^* \mathcal{F}(\mathbf{X}_g^j) \right)}{\mathcal{F}(\mathbf{1}) + \lambda \left(\mathcal{F}(\nabla_x)^* \mathcal{F}(\nabla_x) + \mathcal{F}(\nabla_y)^* \mathcal{F}(\nabla_x) \right)} \right) , \end{aligned} \tag{15}$$

where $\mathcal{F}$ is the FFT operator and $\mathcal{F}()^*$ denotes the complex conjugate. ∇_x and ∇_y denote the horizontal and vertical differential operators, respectively. $\mathcal{F}(\mathbf{1})$ is the Fourier Transform of the delta function. All operations in equation (15) are component-wise. Then the step size t and auxiliary variable $\mathbf{Y}$ is updated:

$$\begin{aligned} t^{j+1} & = (1 + \sqrt{1 + 4(t^j)^2})/2, & (16) \\ \mathbf{Y}^{j+1} & = \mathbf{X}^j + \frac{t^j - 1}{t^{j+1}} (\mathbf{X}^j - \mathbf{X}^{j-1}). & (17) \end{aligned}$$

Update for a and c: with $\mathbf{X}$, we update a and c according to equations (5) and (6):

$$\overline{a}_{b,i}^j = \frac{1}{|\omega|} \sum\nolimits_{k \in \omega_i} a_{b,k}^j , \tag{18}$$

$$\overline{c}_{b,i}^j = \frac{1}{|\omega|} \sum\nolimits_{k \in \omega_i} c_{b,k}^j , \tag{19}$$

Update for $\mathbf{X}_g$: the $\mathbf{X}_g$ can be directly updated with fixed a and c:

$$\mathbf{X}_g^{j+1} = \mathbf{A}^j \cdot \nabla \mathbf{P} + \mathbf{C}^j. \tag{20}$$

Algorithm 1

```
Input: $L$, $\lambda$, $t^1 = 1$, $\mathbf{P}$, $\mathbf{Y}^0$, $\mathbf{M}$.
for $j = 1$ to Max-Iteration do
    $\mathbf{Y} = \mathbf{Y}^j - \psi^T(\psi\mathbf{X} - \mathbf{M})/L$
    $\mathbf{X}^j = \underset{\mathbf{X}}{argmin} \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_2^2 + \lambda \left\|\nabla\mathbf{X} - \mathbf{X}_g^j\right\|_2^2$
    $t^{j+1} = (1 + \sqrt{1 + 4(t^j)^2})/2$
    $\mathbf{Y}^{j+1} = \mathbf{X}^j + \frac{t^j - 1}{t^{j+1}}(\mathbf{X}^j - \mathbf{X}^{j-1})$
    $a_{d,k}^j = \frac{cov(\nabla x_{d,k}^j, \nabla p_{d,k})}{\sigma^2(\nabla p_{d,k}) + \varepsilon}$
    $c_{d,k}^j = \mu(\nabla x_{d,k}^j) - a_{d,k}^j \mu(\nabla p_k)$
    $\mathbf{X}_g^{j+1} = \mathbf{A}^j \cdot \nabla\mathbf{P} + \mathbf{C}^j$
end for
Output: the HRMS image $\mathbf{X}$.
```

5. Experiments

To demonstrate the effectiveness of proposed algorithm, we compare our method with five conventional pan-sharpening methods: AWLP [24], BDSD [14], Indusion [19], MTF-GLP [3], PRACS [11], as well as two variational pan-sharpening methods: SIRF [10], PHLP [18]. For fair comparison, we adjust parameters of each approach to get their best performances. For visual convenience, we only present the RGB bands of fused images but conduct experiments in all spectral bands.

5.1. Evaluation at lower scale

Due to the lack of HRMS images of the same scene, Walds synthesis protocol [28] is used in the simulated experiments. On the basis of this protocol, pan-sharpening is conducted on the degraded data, and the original MS image is regarded as a ground truth which is used to compare with the pan-sharpened image.

To evaluate different methods at lower scale, we introduce both qualitative results and quantitative metrics for assessing the fused images. Quantitative metrics including spectral angle mapper (SAM) [31], universal image quality index averaged over the bands (QAVE) [29] and 8-band extension of Q8 [2], relative dimensionless global error in synthesis (ERGAS) [27] and the spatial correlation coefficient (SCC) [33]. These metrics are used to measure the distortion of the spectral information and spatial structures.

For quantitative evaluation, we list the mean and standard deviation across 225 images with different methods in Table 1. The best results are boldfaced and the last row of the table indicate the ideal value. It can be seen that our proposed approach significantly exceeds all conventional and variation methods, which we believe that our local constraint is more reasonable than others before.

For qualitative analysis, Figure 4 presents the visual results of each methods while the corresponding residuals are shown in Figure 5. Even though all the fused images provide clear versions of the target image by visually, we can still find several subtle discrepancies from residuals. It can be seen that BDSD suffers severe spectral and spatial distortion, followed by PRACS. In the case of Indusion, strong artifacts introduced by the decimation can be noticed. AWLP and MTF-GLP have different levels of spatial distortion. SIRF performs poorly in keeping some spectral features and PHLP leads to over-blurred result. Our method achieves proper trade-off between spectral information and sharp edges preservation.

5.2. Evaluation at the original scale

Since the PAN images are down-sampled in the simulated experiment, we apply these methods at the original scale of PAN images as a complement. Since there are unavailable ground truth images, we adopt LRMS images as spectral reference and PAN images as spatial reference. Furthermore, we use the reference-free measurement QNR [1] to assess the pan-sharpened images. The QNR index is composed by two components: spectral distortion index D$_\lambda$ and spatial distortion index D$_s$.

Performance indexes in Table 2 are obtained by calculating means and standard deviations over 200 test images. We highlight the optimal value and find that PRACS presents the best D$_\lambda$ index while our method has the best performance in terms of D$_s$ and QNR metrics.

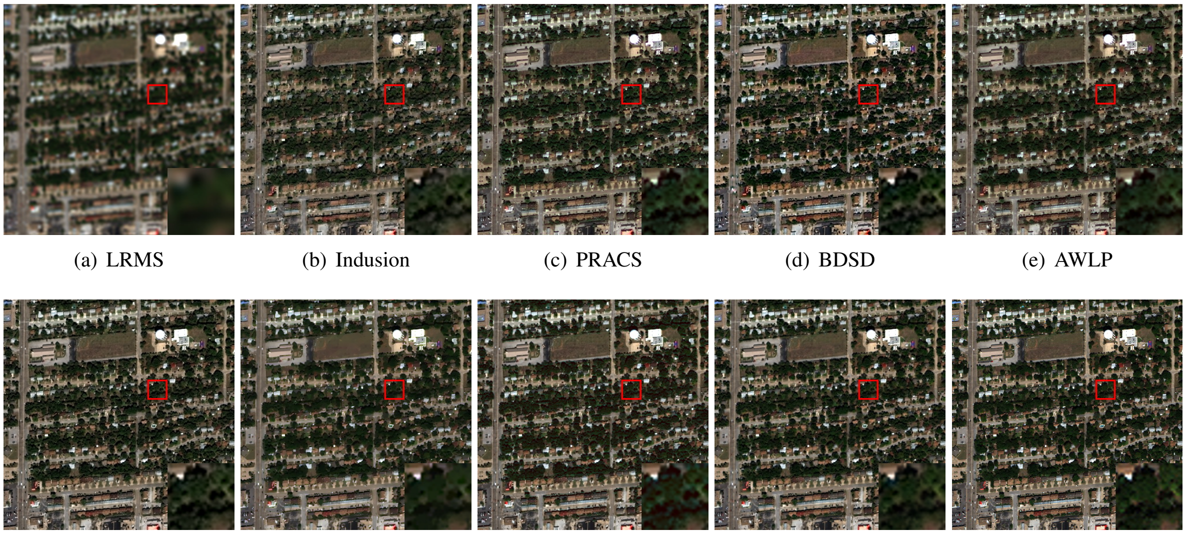
From a qualitative point of view, we scale up the small region of parking lot in Figure 6. There is obvious spectral distortion in bright area produced by Indusion, PRACS, MTF-GLP, PHLP and SIRF. BDSD and AWLP exist a slight degree of color variation compared to the LRMS image. Only our proposed method not only makes full use of the position information provided by PAN image, but also prevents the spectral content from distorting. Since we lack of the ground truth, the residuals to the LRMS image are shown in Figure 7. The LRMS image loses many high-resolution spatial details but contains abundant spectral information. Thus, the smooth regions of residuals should tend to be gray while edges of structures should show apparently. Again, we observe that our model also performs well in dealing with original scale images.

5.3. Comparison with deep learning based methods

Due to the powerful non-linear modeling ability, deep learning technology has been explored to handle pan-sharpening. Therefore, we also compare our model with two recent deep learning based methods, *i.e.*, PNN [23] and PanNet [30]. These two methods are designed in a supervised fashion to learn the mapping function from labeled data. To evaluate both pan-sharpening performance and gen-

Table 1: Quality metrics of different methods on 225 satellite images from WorldView-3.

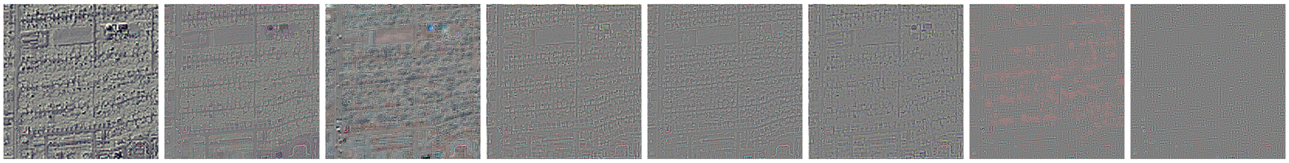
Algorithm	Q8	QAVE	SAM	ERGAS	SCC
Indusion [19]	0.799 ± 0.017	0.799 ± 0.015	6.385 ± 1.544	4.340 ± 0.699	0.825 ± 0.026
PRACS [11]	0.836 ± 0.023	0.822 ± 0.025	6.675 ± 1.628	3.834 ± 0.718	0.835 ± 0.040
BDSD [14]	0.871 ± 0.010	0.867 ± 0.013	7.158 ± 1.909	3.631 ± 0.621	0.856 ± 0.032
AWLP [24]	0.849 ± 0.028	0.844 ± 0.029	6.219 ± 1.487	3.697 ± 0.697	0.865 ± 0.029
MTF-GLP [3]	0.871 ± 0.023	0.858 ± 0.030	6.639 ± 1.723	3.494 ± 0.723	0.857 ± 0.047
PHLP [18]	0.859 ± 0.013	0.835 ± 0.011	5.748 ± 0.926	3.747 ± 0.590	0.845 ± 0.024
SIRF [9, 10]	0.863 ± 0.013	0.859 ± 0.002	6.140 ± 1.416	3.564 ± 0.553	0.866 ± 0.019
Proposed	**0.891** ± **0.023**	**0.890** ± **0.023**	**5.460** ± **1.309**	**3.172** ± **0.603**	**0.891** ± **0.027**
ideal value	1	1	0	0	1



(a) LRMS (b) Indusion (c) PRACS (d) BDSD (e) AWLP

(f) MTF-GLP (g) PHLP (h) SIRF (i) Proposed (j) Ground truth

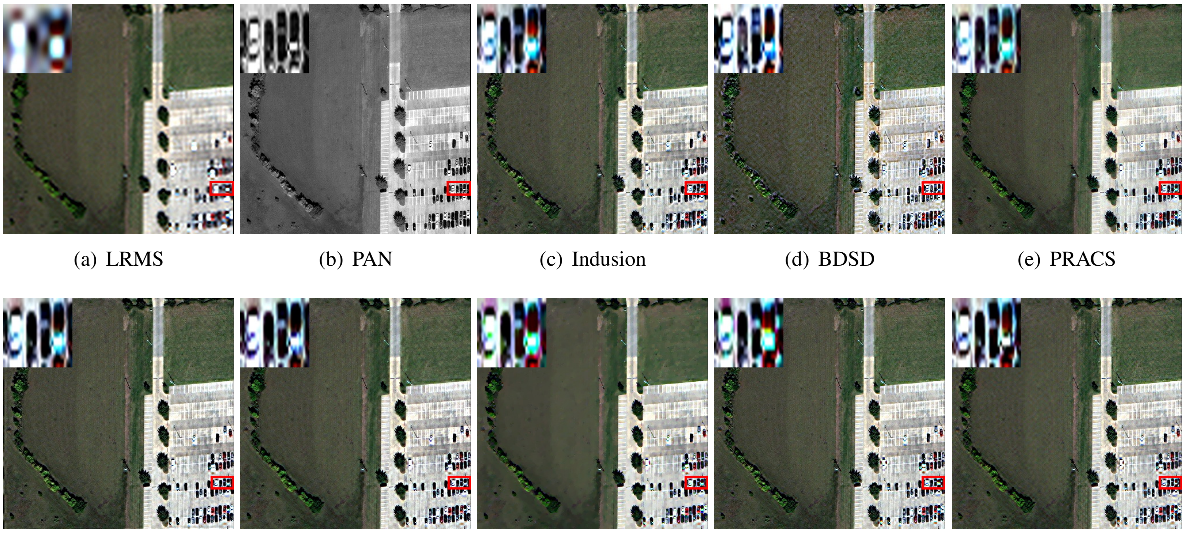
Figure 4: Comparison with different methods (source: WorldView-3). The size of PAN is 400 × 400.



(a) Indusion (b) PRACS (c) BDSD (d) AWLP (e) MTF-GLP (f) PHLP (g) SIRF (h) Proposed

Figure 5: The residuals between the HRMS image reconstructions and the ground truth from Figure 4.

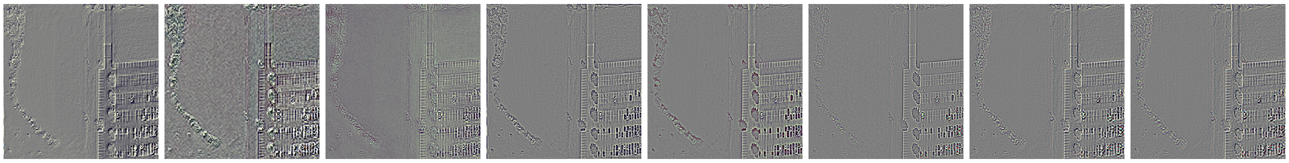
eralization ability, the compared models of both PNN and PanNet are **only** trained on WorldView-3. While for testing, we use the data from both WorldView-2 and WorldView-3. Since PNN and PanNet are trained on WorldView-3, they have good visual quality on the testing image that from the same satellite, as shown in Figure 8. However, the generalization ability of PNN and PanNet is limited due to the supervised learning strategy. As shown in Figure 9, the residuals of PNN and PanNet contain more detail and spectral information. This is because once the training is finished, the network parameters of PNN and PanNet will be fixed and cannot adapt to the new type of data. On the contrary, our model adopts the local constraint, which is a universal regularization and is independent of data. This makes our model has a better generalization ability than PNN and PanNet. This advantage is further proved in Table 3.



(a) LRMS (b) PAN (c) Indusion (d) BDSD (e) PRACS

(f) MTF-GLP (g) AWLP (h) PHLP (i) SIRF (j) Proposed

Figure 6: The fusion results at the original scale (source: WorldView-3). The size of PAN is 400×400.

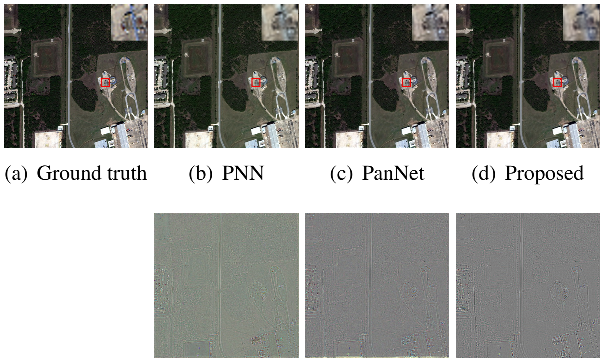


(a) Indusion (b) BDSD (c) PRACS (d) MTF-GLP (e) AWLP (f) PHLP (g) SIRF (h) Proposed

Figure 7: The residuals to the LRMS image from Figure 6. Note that ideal residuals should have smooth regions close to gray while the edges of structures should be apparent.

Table 2: Quality metrics evaluated at original scales on 200 satellite images from WorldView-3.

Algorithm	D_λ	D_s	QNR
BDSD [14]	0.079 ± 0.035	0.128 ± 0.034	0.803 ± 0.048
Indusion [19]	0.055 ± 0.023	0.073 ± 0.018	0.876 ± 0.034
PRACS [14]	**0.019 ± 0.006**	0.103 ± 0.021	0.880 ± 0.021
AWLP [24]	0.065 ± 0.026	0.108 ± 0.018	0.835 ± 0.037
MTF-GLP [3]	0.049 ± 0.018	0.072 ± 0.018	0.883 ± 0.032
SIRF [9, 10]	0.070 ± 0.027	0.088 ± 0.027	0.849 ± 0.047
PHLP [18]	0.029 ± 0.020	0.077 ± 0.019	0.896 ± 0.035
Proposed	0.030 ± 0.012	**0.050 ± 0.015**	**0.922 ± 0.024**
ideal value	0	0	1



(a) Ground truth (b) PNN (c) PanNet (d) Proposed

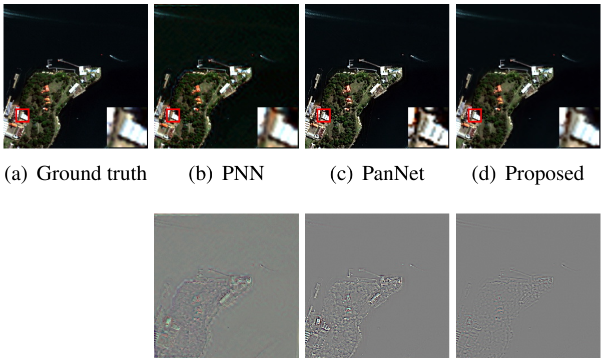
(e) \|(a) - (b)\| (f) \|(a) - (c)\| (g) \|(a) - (d)\|

Figure 8: Visual comparison with deep learning. PNN and PanNet are trained and tested on WorldView–3 images.

We also test on data at original scales. As shown in Figure 10, both PNN and PanNet have spectral distortion even though their models are trained on the same data source, *i.e.*, WorldView-3. This spectral distortion is more obvious when PNN and PanNet are directly tested on new satellite, *i.e.*, WorldView-2, as shown in Figure 11. On the contrary, our model achieves a good trade-off between spatial and spectral preservation. The corresponding quantitative results are shown in Table 4, which further proves the generalization ability of our model to original scale images.

Table 3: Quality metrics of different methods on WorldView-3 and WorldView-2 satellites. "-WV3" and "-WV2" indicates testing on WorldView-3 and WorldView-2, respectively.

Algorithm	Q8	QAVE	SAM	ERGAS	SCC
PNN-WV3 [23]	0.882 ± 0.005	0.891 ± 0.003	4.752 ± 0.870	3.277 ± 0.473	0.915 ± 0.009
PanNet-WV3 [30]	**0.925 ± 0.005**	**0.928 ± 0.010**	**4.128 ± 0.787**	**2.469 ± 0.347**	**0.943 ± 0.018**
Proposed-WV3	0.891 ± 0.023	0.890 ± 0.023	5.460 ± 1.309	3.172 ± 0.603	0.891 ± 0.027
PNN-WV2 [23]	0.694 ± 0.251	0.710 ± 0.246	4.696 ± 1.535	4.720 ± 0.751	0.904 ± 0.015
PanNet-WV2 [30]	0.723 ± 0.179	0.728 ± 0.180	4.091 ± 2.090	5.569 ± 0.876	0.845 ± 0.032
Proposed-WV2	**0.775 ± 0.189**	**0.774 ± 0.196**	**2.940 ± 1.585**	**3.598± 0.587**	**0.915± 0.019**
ideal value	1	1	0	0	1



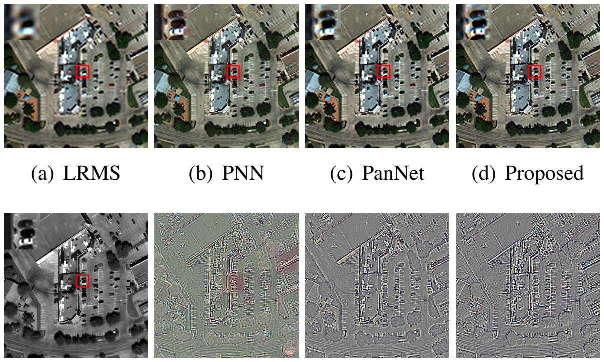
(a) Ground truth (b) PNN (c) PanNet (d) Proposed

(e) |(a) - (b)| (f) |(a) - (c)| (g) |(a) - (d)|

Figure 9: Visual comparison with deep learning. PNN and PanNet are trained on WorldView–3 images and tested on WorldView–2 images to evaluate generalization ability.

Table 4: Quality metrics evaluated at original scales on WorldView-3 and WorldView-2 satellites.

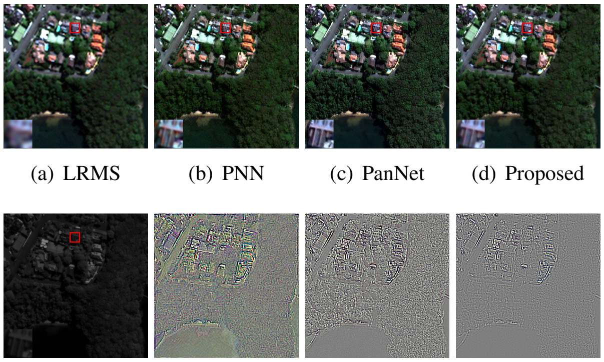
Algorithm	D$_\lambda$	D$_s$	QNR
PNN-WV3 [23]	0.036 ± 0.008	0.087 ± 0.021	0.880 ± 0.022
PanNet-WV3 [30]	**0.023 ± 0.008**	0.071 ± 0.013	0.908 ± 0.015
Proposed-WV3	0.030 ± 0.012	**0.050 ± 0.015**	**0.922 ± 0.024**
PNN-WV2 [23]	0.054 ± 0.055	0.035 ± 0.033	0.915 ± 0.080
PanNet-WV2 [30]	0.091 ± 0.079	0.125 ± 0.113	0.803 ± 0.161
Proposed-WV2	**0.011 ± 0.005**	**0.035 ± 0.015**	**0.954 ± 0.018**
ideal value	0	0	1



(a) LRMS (b) PNN (c) PanNet (d) Proposed

(e) PAN (f) |(a) - (b)| (g) |(a) - (c)| (h) |(a) - (d)|

Figure 10: Comparison with deep learning on a original s-cale WorldView-3 image.



(a) LRMS (b) PNN (c) PanNet (d) Proposed

(e) PAN (f) |(a) - (b)| (g) |(a) - (c)| (h) |(a) - (d)|

Figure 11: Comparison with deep learning on a original s-cale WorldView-2 image.

6. Conclusion

We propose a pan-sharpening method that incorporates a local constraint for image spatial preservation. Firstly, we show our local penalty can outperform global one through statistical verification. Secondly, based on this local constraint, we build a new variational model for pan-sharpening. We also derive an simple optimization algorithm to efficiently solve the proposed model. The experiment proves that our model can achieve better spectral and spatial preservation compared with other methods. Moreover, due to the proposed universal local constraint, our model has a better generalization ability than recent deep learning based method. Since our method does not require training phase, it has enough flexibility to directly deal with new satellites and achieve satisfactory performance.

References

[1] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva. Multispectral and panchromatic data fusion assessment without reference. *Photogrammetric Engineering & Remote Sensing*, 74(2):193–200, 2008. 5

[2] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini. A global quality measurement of pan-sharpened multispectral imagery. *IEEE Geoscience Remote Sensing Letters*, 1(4):313–317, 2004. 5

[3] 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 Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021, 2007. 5, 6, 7

[4] H. A. Aly and G. Sharma. A regularized model-based optimization framework for pan-sharpening. *IEEE Transactions on Image Processing*, 23(6):2596–2608, 2014. 2, 3

[5] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé. A variational model for p+ xs image fusion. *International Journal of Computer Vision*, 69(1):43–58, 2006. 2, 3

[6] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009. 4

[7] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983. 1

[8] W. J. Carper. The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogramm. Eng. Remote Sens.*, 56(4):457–467, 1990. 1

[9] C. Chen, Y. Li, W. Liu, and J. Huang. Image fusion with local spectral consistency and dynamic gradient sparsity. In *CVPR*, pages 2760–2765, 2014. 6, 7

[10] C. Chen, Y. Li, W. Liu, and J. Huang. Sirf: simultaneous satellite image registration and fusion in a unified framework. *IEEE Transactions on Image Processing*, 24(11):4213–4224, 2015. 2, 3, 5, 6, 7

[11] J. Choi, K. Yu, and Y. Kim. A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1):295–309, 2011. 5, 6

[12] X. Ding, Y. Jiang, Y. Huang, and J. Paisley. Pan-sharpening with a bayesian nonparametric dictionary learning model. In *AISTATS*, pages 176–184, 2014. 2

[13] F. Fang, F. Li, C. Shen, and G. Zhang. A variational approach for pan-sharpening. *IEEE Transactions on Image Processing*, 22(7):2822–2834, 2013. 2

[14] A. Garzelli, F. Nencini, and L. Capobianco. Optimal mmse pan sharpening of very high resolution multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1):228–236, 2008. 5, 6, 7

[15] A. R. Gillespie, A. B. Kahle, and R. E. Walker. Color enhancement of highly correlated images. ii. channel ratio and chromaticity transformation techniques. *Remote Sensing of Environment*, 22(3):343–365, 1987. 1

[16] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2013. 3

[17] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang. A new pan-sharpening method with deep neural networks. *IEEE Geoscience Remote Sensing Letters*, 12(5):1037–1041, 2015. 1

[18] Y. Jiang, X. Ding, D. Zeng, Y. Huang, and J. Paisley. Pan-sharpening with a hyper-laplacian penalty. In *ICCV*, pages 540–548, 2015. 2, 3, 5, 6, 7

[19] M. M. Khan, J. Chanussot, L. Condat, and A. Montanvert. Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique. *IEEE Geoscience Remote Sensing Letters*, 5(1):98–102, 2008. 5, 6, 7

[20] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *NIPS*, pages 1033–1041, 2009. 3

[21] P. Kwarteng and A. Chavez. Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens*, 55:339–348, 1989. 1

[22] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989. 1

[23] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa. Pan-sharpening by convolutional neural networks. *Remote Sensing*, 8(7):594, 2016. 1, 5, 8

[24] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez. Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods. *IEEE Transactions on Geoscience and Remote Sensing*, 43(10):2376–2385, 2005. 5, 6, 7

[25] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson. A new pansharpening algorithm based on total variation. *IEEE Geoscience Remote Sensing Letters*, 11(1):318–322, 2014. 3

[26] M. J. Shensa. The discrete wavelet transform: wedding the a trous and mallat algorithms. *IEEE Transactions on signal processing*, 40(10):2464–2482, 1992. 1

[27] L. Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002. 5

[28] L. Wald, T. Ranchin, and M. Mangolini. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering & Remote Sensing*, 63(6):691–699, 1997. 5

[29] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002. 5

[30] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley. PanNet: A deep network architecture for pan-sharpening. In *ICCV*, 2017. 1, 5

[31] R. H. Yuhas, A. F. Goetz, and J. W. Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *Summaries of the Third Annual JPL Airborne Geoscience Workshop; AVIRIS Workshop: Pasadena, CA, USA*, pages 147–149, 1992. 5

[32] D. Zeng, Y. Hu, Y. Huang, Z. Xu, and X. Ding. Pan-sharpening with structural consistency and $\ell_{1/2}$ gradient prior. *Remote Sensing Letters*, 7(12):1170–1179, 2016. 2

[33] J. Zhou, D. Civco, and J. Silander. A wavelet transform method to merge landsat tm and spot panchromatic data. *International Journal of Remote Sensing*, 19(4):743–757, 1998. 5