MULTIRESOLUTION IMAGE FUSION USING NONSUBLAEMEED CONTOURLET TRANSFORM

Asst Prof. Pravin J Barad¹

¹Electronics and Communication Engineering, Veerayatan Group of Institutions FOE & FOM, Haripar, Mandvi, Gujarat, India

ABSTRACT

A novel image fusion strategy is presented for panchromatic (PAN) high resolution image and multispectral image (MS) in nonsubsampled contourlet transform (NSCT) domain. The NSCT can give an asymptotic optimal representation of edges and contours in image by virtue of its characteristics of good multi resolution, shift invariance, and high directionality. We obtain a high spatial resolution (HR) and high MS image using the available high spectral but resolution MS image and the PAN image. Since we need to predict the missing high resolution pixels in each of the MS images the problem is ill-posed and is solved using maximum a posteriori (MAP) approach. We first obtain an initial approximation to the HR fused image by learning the edges from the PAN image using the NSCT. Markov random field (MRF) prior term is used for regularization to obtain smoothness in the image. We then optimize the cost function which is formed using the data fitting term and the prior term and obtain the fused image, in which the edges correspond to those in the initial HR approximation. The procedure is repeated for each of the MS images. The advantage of the proposed method lies in the use of simple gradient based optimization for regularization purposes while preserving the discontinuities and color components.

Keywords: Fusion, Multiresolution, NSCT, Panchromatic (PAN), Multispectral (MS), MAP, MRF.

1. INTRODUCTION

Image fusion is a process by combining two or more source images from different modalities or instruments into a single image with more information. Image fusion has become a new and promising research area such as remote sensing, medical imaging, machine vision, and military applications [1], [2]. Fusion of images coming from different sensors (visible and infrared, or panchromatic and multispectral, satellite images) is referred as Multimodal Image Fusion. Multiresolution fusion is a method of combining a high spatial resolution Panchromatic (PAN)
image and a low spatial but high spectral resolution Multispectral (MS) image in order to obtain a high spatial and spectral resolution MS image. PAN images of high spatial resolution can provide detailed geometric information, such as shapes, features, and structures of objects on the earth’s surface. While MS images with usually lower resolution are used to obtain spectral information i.e. color information necessary for environmental applications. Multiresolution image fusion methods can be broadly classified into two - spatial domain fusion and transform domain fusion. The fusion methods such as averaging, Brovey method, principal component analysis (PCA) and intensity hue saturation (IHS) based methods fall under spatial domain approaches [3]. [4]. The classical fusion methods are PCA, IHS transformed. The disadvantage of spatial domain approaches is that they produce color distortion in the fused image. Spatial distortion can be very well handled by transform domain approaches on image fusion. Moreover, separable wavelets can capture only limited directional information, and thus cannot represent the directions of the edge accurately. Contourlet transform (CT) [5], [6], developed by Do and Vetterli, is a powerful image fusion method. Compared with the wavelet transforms, the CT can represent edges and other singularities along curves much more efficiently. However, the CT lacks the shift invariance, which is desirable in many image applications such as image enhancement, image denoising and image fusion. NSCT [7], [8], proposed by Cunha, inherits the perfect properties of the CT, and meanwhile possesses the shift invariance. The main difference between NSCT and other multiscale directional systems is that the NSCT allows for different and flexible number of directions at each scale. In this paper based on the above analysis, an approach for QuickBird MS image and PAN image fusion using NSCT is presented. We obtain a high spatial resolution and high spectral resolution MS image using the available high spectral but low spatial resolution MS image and the PAN image by learning the edges using NSCT.

A canny edge detector is used to locate the edge pixels in the learned image which are retained as the edge pixels in the final fused image. We then use a maximum a posteriori (MAP) MRF approach to obtain the final solution. The optimization carried out only on those pixels which do not belong to the edge pixels as the edge pixels correspond to those obtained using the NSCT based learning.

2. NSCT BASED EDGE LEARNING

Different from the CT, the Multiresolution decomposition step of NSCT is realized by shift-invariant filter banks satisfying Bozout identical eq. (1). The NSCT is the shift-invariant version of the CT, and provides not only Multiresolution analysis but also provides localization, multi-scaling, multi-directionality, and anisotropy. The three level NSCT decomposition is shown in “Fig. 1”. The core of the NSCT is designing the needed nonseparable two-channel nonsubsampled filter Bank (NSFB) that lead to an NSCT with better frequency selectivity and regularity when compared to the corresponding CT.

\[ H_0(Z)H_1(Z) + G_0(Z)G_1(Z) = 1 \]  

In NSCT a two-dimensional (2-D) filter is represented by its Z-transform H(z) where \( Z = [Z_1, Z_2]^T \). Evaluated on the unit sphere, a filter is denoted by \( H(e^{j\omega}) \), where \( e^{j\omega} = [e^{j\omega_1}, e^{j\omega_2}]^T \). If \( m = [m_1, m_2]^T \) is a 2-D vector, then \( Z^m = z_1^{m_1}z_2^{m_2} \) whereas if M is a 2x2 matrix, then \( Z^m = z_1^mz_2^m \) the columns of M. Here we often deal with zero-phase 2-D filters. On the unit sphere, such filters can be written as polynomials in \( \cos \omega = [\cos \omega_1, \cos \omega_2]^T \). Thus we write \( F(x_1, x_2) \) for a zero-phase filter in which \( x_1 \) and \( x_2 \) denote \( \cos \omega_1 \) and \( \cos \omega_2 \) respectively.
The structure consists a bank of filters that splits the 2-D frequency plane in the subbands. NSCT can thus be divided into parts. First nonsubsampled pyramid (NSP) before you begin to structure that gives the multiscale property. Second nonsubsampled directional filter bank (NSDFB) structure that ensure directionality.

We upsample all filters by a quincunx matrix for the next level by $Q = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

3. PROPOSED APPROACH

The method for the proposed Multiresolution fusion is explained using the block diagram shown in “Fig. 2” which gives the fusion process low resolution MS image and PAN image giving the fused MS image as the result. The initial HR approximation MS is obtained using PAN and low resolution(LR) MS images and is 4 times the size of LR MS image, i.e. decimation factor $q=4$. Learning is done by coping the NSCT coefficients of the PAN image that correspond to the high frequency details(fourth level) to fourth level of the unknown fused MS images shown in “Fig. 3”.

The initial approximation of the fused MS image is then obtained by taking the inverse NSCT. We use this approximation to extract the edges in the fused image and to estimate the aliasing/decimation as well as the MRF prior model parameter. The decimation block has the inputs as the LR MS and initial HR approximation and gives decimation matrix coefficients as the output. Finally, the cost
function consisting of data fitting term and the prior term is minimized using the gradient descent optimization.

4. **FORWARD MODEL**

Since the problem is in restoration, it needs a forward model to represent the image formation process. There are p observed low resolution MS images \(Y_n, \quad n=1,2,\ldots,p\), each captured with a different spectral band, of size \(N \times N\) pixels and a single PAN image \(Z\) captured with a high spatial resolution of size \(qN \times qN\). This is also the size of each of the fused images \(Z_n, \quad n=1,2,\ldots,p\). Here, \(q\) represents the decimation factor (aliasing factor) and is an integer. After lexicographically ordering the LR MS image and HR MS image respectively, as \(Y\) and \(Z\), their size becomes \(N^2 \times 1\) and \(qN^2 \times 1\), respectively. The image formation model for each of the LR MS images can be expressed as

\[
Y_n = D_n Z_n + \eta
\]

(2)

where \(\eta\) represents the additive noise and \(D\) is the decimation matrix of size \(N^2 \times qN^2\). We use the fixed decimation matrix as used by the super-resolution community[9], which can be written as

\[
D = \frac{1}{q^2} \begin{bmatrix}
11 \ldots 1100 & \ldots 000 \ldots \\
0011 \ldots 11000 \ldots & \\
0 \ldots 001 \ldots 10 \ldots & \\
00 \ldots 011 \ldots 10 \ldots & \\
000 \ldots 0011 \ldots 1 
\end{bmatrix}
\]

5. **MRF PRIOR MODEL**

The proposed method is clearly ill-posed problem, and hence, we need a proper regularization. We use a MRF[10] based prior term for capturing the spatial correlation among pixels in the HR MS image to obtain the fused image for each of the MS band. It is well known that MRFs are the most general models used as priors during regularization when solving ill-posed problems. MRF theory provides a basis for modeling contextual constraints in image processing and
interpretation. Since the method does not directly operate on PAN pixel values as most of the other methods do, spectral distortion is minimum, and the spatial properties are better preserved in the fused image as the MRF parameters are learned at every pixel. An MRF prior for the unknown image can be described by using a energy function \( U(z) \) expressed in the Gibbsian density given by

\[
P(Z_n) = \frac{1}{Z_n\theta} \exp (-U(Z_n))
\]  

(3)

Here \( Z_n \) is the unknown image to be estimated, and represents the normalizing constant known as partition function. One can choose \( U(z) \) as a quadratic form with a single global parameter, assuming that the images are globally smooth. However, a more efficient model would be to choose \( U(z) \) such that only the homogeneous regions are smooth and that discontinuities are preserved in the form of edges and is written as

\[
U(Z) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left[ (Z_{i,j} - Z_{i-1,j})^2 + (Z_{i,j} - Z_{i-1,j-1})^2 + (Z_{i,j} - Z_{i-1,j+1})^2 + (Z_{i,j} - Z_{i+1,j})^2 + (Z_{i,j} - Z_{i+1,j-1})^2 + (Z_{i,j} - Z_{i+1,j+1})^2 + (Z_{i,j} - Z_{i,j+1})^2 + (Z_{i,j} - Z_{i,j+1})^2 \right]
\]  

(4)

In order to preserve discontinuities, as well as reconstruct the smooth areas well, we use an MRF as a prior as in eq. (4). This enables us to capture the smooth regions, as well as the sudden intensity variation due to sharp edges.

6. MAP ESTIMATION

Using MAP estimation we obtain the high-spatial and high-spectral resolution MS images and is done for all MS images separately. The MRF model on the fused image serves as the prior for
MAP estimation [11],[12]. In order to use MAP estimation to obtain the fused MS image \( Z_n \), given the MS and PAN observation, we need to obtain the estimate as 

\[
P(Z_n|Y_n) = \frac{P(Z_n Y_n)}{P(Y_n)}
\]

Since denominator is not a function of \( Z \) above eq. (5) can be written as,

\[
P(Z_n|Y_n) = P(Z_n Y_n)
\]

Now taking the log and considering that the random variables are independent, we can write

\[
\hat{Z} = \text{argmax}_{Z_n} \left[ \log P \left( \frac{Z_n}{Y_n} \right) + \log P(Z_n) \right]
\]

The final cost function to be minimized can be expressed as

\[
Z_n = \text{argmin}_{Z_n} \left[ \frac{||Y_n - D_n Z_n||^2}{2\sigma^2} + U(Z_n) \right]
\]

The above cost function is quadratic therefore, the most computationally efficient way to optimize it would be the gradient descent method for the sake of simplicity.

7. EXPERIMENTAL RESULTS

We present the results of the proposed method for fusion. The experiments are conducted on real images captured using Quickbird satellite. The original PAN image and the MS images are of size 512×512 and 128×128, respectively. In order to make the quantitative comparison possible, we down sampled the images by a factor of 4 and conducted the experiments on PAN and MS images of size 128×128 and 32×32 respectively. The size of fused MS images is 128×128. We compare the performance of the proposed method with other methods on the basis of quality of images in terms of quantitative measures. “Fig. 4,7” shows the results of fusion for Band-1 to Band-4 using different approaches. From these results we conclude that the proposed method is better when compared to the other MS fusion approaches. After learning the edges of PAN and MS images we do the optimization by a proper regularization method. We use MRF-MAP regularization method for optimization. There are many metrics that analyze the spectral quality[13]. Spectral Angle Mapper(SAM) compares each pixel in the image with every end member for each class.

\[
\cos(\alpha) = \frac{\sum_{i=1}^{N} (B_i B_i)}{\sqrt{\sum_{i=1}^{N} A_i A_i \sum_{i=1}^{N} B_i B_i}}
\]

Universal Image Quality Index(Q-average) models any distortion as a combination of three different factors: loss of correlation, luminance distortion, and contrast distortion.

\[
Q_{avg} = \frac{4\sigma_{xy} \bar{x} \bar{y}}{\left(\sigma_x^2 + \sigma_y^2\right) \left[\bar{x}^2 + \bar{y}^2\right]}
\]
The relative average spectral error (RASE) characterizes the average performance of the method of image fusion in the spectral bands.

\[
RASE = \frac{100}{M} \sqrt{\frac{1}{N} \sum_{i=1}^{N} RMSE^2(B_i)} \tag{11}
\]

We also used root mean squared error (RMSE) and correlation coefficient (CC) to analyze and compare the spectral quality. The CC between the original MS image and the final fused image is defined as

\[
CC(A, B) = \frac{\sum_{mn}(A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\sum_{mn}(A_{mn} - \bar{A})^2 \sum_{mn}(B_{mn} - \bar{B})^2}} \tag{12}
\]

\[
RMSE = \sqrt{\frac{\sum_x \sum_i (A_i(x) - F_i(x))^2}{(n \times m \times d)^2}} \tag{13}
\]

“Fig. 4,6” shows comparison of various fusion techniques from which it can be seen that the edges are better preserved by NSCT approach and also the color distortion is minimum. “Fig. 5,7” gives the comparison of original image, initial approximate image and the optimized image from which it can be said that the proposed method is better in preserving the edges with minimum color distortion. “Table 1,2” gives the quantitative measures for various fusion techniques.

**Figure 4:** Fused images by various multiresolution methods for image 1
Figure 5: Color Image comparison of Original, Learned, and Optimised image with RGB as 4,3,2 bands respectively for image 1

### TABLE 1: Spectral Quantitative Measures For Image 2

<table>
<thead>
<tr>
<th>Parameters</th>
<th>CC</th>
<th>Qavg</th>
<th>RASE</th>
<th>RMSE</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Original HIS</td>
<td>0.1035</td>
<td>0.916</td>
<td>0.2721</td>
<td>0.05</td>
<td>5.73</td>
</tr>
<tr>
<td>Egde HIS</td>
<td>0.3185</td>
<td>0.8978</td>
<td>0.2928</td>
<td>0.0695</td>
<td>6.054</td>
</tr>
<tr>
<td>Adaptive HIS</td>
<td>0.0883</td>
<td>0.9284</td>
<td>0.1568</td>
<td>0.0372</td>
<td>4.739</td>
</tr>
<tr>
<td>PCS</td>
<td>0.1319</td>
<td>0.8928</td>
<td>0.2064</td>
<td>0.0490</td>
<td>5.4293</td>
</tr>
<tr>
<td>Wavelet</td>
<td>0.1576</td>
<td>0.9185</td>
<td>0.1842</td>
<td>0.0437</td>
<td>5.665</td>
</tr>
<tr>
<td>CT</td>
<td>0.1898</td>
<td>0.8905</td>
<td>0.2016</td>
<td>0.0492</td>
<td>6.488</td>
</tr>
<tr>
<td>NSCT</td>
<td>0.0867</td>
<td>0.9210</td>
<td>0.1572</td>
<td>0.0383</td>
<td>5.36</td>
</tr>
<tr>
<td>Optimised Band1</td>
<td>0.5244</td>
<td>0.9137</td>
<td>0.3128</td>
<td>0.0682</td>
<td>0.0951</td>
</tr>
<tr>
<td>Optimised Band1</td>
<td>0.4671</td>
<td>0.9137</td>
<td>0.3128</td>
<td>0.0196</td>
<td>0.1876</td>
</tr>
<tr>
<td>Optimised Band1</td>
<td>0.4886</td>
<td>0.9137</td>
<td>0.3128</td>
<td>0.01034</td>
<td>0.1817</td>
</tr>
<tr>
<td>Optimised Band1</td>
<td>0.5520</td>
<td>0.9137</td>
<td>0.3128</td>
<td>0.0189</td>
<td>0.1358</td>
</tr>
</tbody>
</table>

### TABLE 2: Spectral Quantitative Measures For Image 2

<table>
<thead>
<tr>
<th>Parameters</th>
<th>CC</th>
<th>Qavg</th>
<th>RASE</th>
<th>RMSE</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Original HIS</td>
<td>0.0729</td>
<td>0.894</td>
<td>0.1523</td>
<td>11.3038</td>
<td>4.462</td>
</tr>
<tr>
<td>Egde HIS</td>
<td>0.2033</td>
<td>0.8887</td>
<td>0.1627</td>
<td>1407715</td>
<td>5.0471</td>
</tr>
<tr>
<td>Adaptive HIS</td>
<td>0.0847</td>
<td>0.9025</td>
<td>0.1238</td>
<td>10.9719</td>
<td>4.1360</td>
</tr>
<tr>
<td>PCS</td>
<td>0.3912</td>
<td>0.8878</td>
<td>0.1507</td>
<td>13.6845</td>
<td>4.2925</td>
</tr>
<tr>
<td>Wavelet</td>
<td>0.0679</td>
<td>0.8830</td>
<td>0.1473</td>
<td>13.3730</td>
<td>4.7187</td>
</tr>
<tr>
<td>CT</td>
<td>0.0729</td>
<td>0.8400</td>
<td>0.1421</td>
<td>18.4101</td>
<td>5.3370</td>
</tr>
<tr>
<td>NSCT</td>
<td>0.1739</td>
<td>0.8800</td>
<td>0.1221</td>
<td>10.9202</td>
<td>4.963</td>
</tr>
<tr>
<td>Optimised Band1</td>
<td>0.5244</td>
<td>0.8913</td>
<td>0.5459</td>
<td>0.6120</td>
<td>0.0898</td>
</tr>
<tr>
<td>Optimised Band1</td>
<td>0.7861</td>
<td>0.8913</td>
<td>0.5459</td>
<td>0.1946</td>
<td>0.1109</td>
</tr>
<tr>
<td>Optimised Band1</td>
<td>0.8013</td>
<td>0.8913</td>
<td>0.5459</td>
<td>0.2613</td>
<td>0.1034</td>
</tr>
<tr>
<td>Optimised Band1</td>
<td>0.6411</td>
<td>0.8913</td>
<td>0.5459</td>
<td>0.3338</td>
<td>0.1064</td>
</tr>
</tbody>
</table>
8. CONCLUSION

An approach for merging high spatial resolution PAN image and high spectral resolution MS image was proposed after analyzing the theory of NSCT. We take full advantage of NSCT, including good multiscaling, localization, anisotropy, shift-invariance, and multidirectional decomposition. The image is first decomposed into multiscale and multidirection subbands by NSCT and then fused. Finally the fused image is reconstructed by inverse transform. Since the final solution is obtained using canny edge detector and MRF prior, the suggested method gives finer details present in different directions with minimum spectral distortion. The results demonstrate that the proposed technique yields better solution as compared to those obtained using the recent approaches. Experimental results show that proposed method performs better in preserving the edges than that of the other traditional image fusion methods in multiresolution image fusion field.

At the same time this new method may have wide application in many fields such as remote sensing, battlefield surveillance, target tracking, machine vision and medical diagnosis. In future results can be improved by using different prior term for regularization and the final cost function can also be solved with different computationally texting optimizing techniques like graph cuts.
REFERENCES