PANSHARPENING WITH A DECISION FUSION BASED ON THE LOCAL SIZE INFORMATION

Bin Luo, Muhammad Murtaza Khan, Thibaut Bienvenu & Jocelyn Chanussot

GIPSA-Lab. Department of Images and Signals, Grenoble Institute of Technology, France
bin.luo, muhammad-murtaza.khan, jocelyn.chanussot@gipsa-lab.inpg.fr, thibaut.bienvenu@gmail.com

ABSTRACT

Pansharpening may be defined as the process of synthesizing multispectral images at a higher spatial resolution. Different pansharpening methods produce images with different characteristics. In the 2006 IEEE Data Fusion Contest, À-trous Wavelet Transform based pansharpening (AWLP) and Context Adaptive (CBD) pansharpening methods were declared as joint winners. While assessing the quantitative quality of the pansharpened images, it was observed that the two methods outperform each other depending upon the local content of the scene. Hence, it is interesting to develop a method which could produce results locally approximately similar to the best method, among the two pansharpening methods. In this paper we propose a method which selects either of the two methods for performing pansharpening on local regions, based upon the size of the objects. The results obtained demonstrate that the proposed method produces images with quantitative results approximately similar to the method which is better among the AWLP and CBD pansharpening methods.

1. INTRODUCTION

Satellites provide both multispectral (MS) and panchromatic (Pan) images. The Pan and MS images captured by satellites are not at the same resolution. The MS images have a high spectral and low spatial resolution while the Pan image has a high spatial resolution, but no spectral diversity. Images with both high spectral and spatial resolution cannot be captured by satellites due to the fact that in order to maintain the same signal to noise ratio for the narrow spectral band MS images the aperture of the sensor has to be opened for a longer duration. Since the satellite moves during this duration the image captured is blurred, thus resulting in low resolution MS images. Generally, satellite images with both high spectral and spatial resolution are required for improving the results of automated tasks like classification, segmentation and object detection [1]. The process of pansharpening helps in synthesizing MS images having both high spectral and spatial resolution. For example, the Pan image captured by the IKONOS sensor has a spatial resolution of 1 m where as the MS image has a resolution of 4m. The pansharpening process synthesizes a MS image at the spatial resolution of 1m.

Pansharpening methods can be divided into Multiresolution Analysis (MRA) based methods and Component Substitution (CS) based methods. CS methods are based on the spectral transformation of MS data and the inverse transformation using the high resolution Pan image, which results in the high resolution MS image [2]. MRA based methods employ filters to obtain multi level projections of the MS & Pan images and use the high frequency information extracted from the Pan image during the pansharpening process [3, 4, 5, 6, 7]. In the recently conducted Data Fusion Contest [8] it was observed that the MRA methods quantitatively outperformed the CS methods. Among the different MRA methods tested, the AWLP [3, 6] and ATWT-CBD [5] pansharpening methods resulted in images having good quantitative characteristics. It was observed that the results obtained by the AWLP and ATWT-CBD methods are either the best or second best. Thus an attempt can be made to use the size of the objects for determining which of the two methods should be locally used for pansharpening. It was observed while testing on local regions, *i.e.* context adaptivity, that AWLP method generally works better for objects of small size in the image *i.e.* roads, cars *etc.* where as the ATWT-CBD method is better for pansharpening larger objects in the image, such as industrial buildings, fields, *etc.*.

The second section of this paper presents a brief introduction to AWLP and CBD pansharpening methods. The third section describes briefly how to determine the relative size of objects in a satellite image. The fourth section presents the proposed pansharpening method in the context of using different pansharpening methods based upon the relative size of objects. The fifth section presents the results obtained for the simulated Pléiades sensor images and the final section presents the conclusions and perspectives.

2. EXISTING PANSHARPENING METHODS

2.1. À-trous Wavelet Pansharpening (AWLP)

This pansharpening method is based on the use of undecimated Wavelet transform to extract the high frequency

details from the Pan image and the addition of these details in the upscaled low resolution MS image. A version of this method, for three MS bands, was proposed in [3] and its extension to multi-band MS images, having more than three bands, was proposed in [6]. The method does not incorporate any specific detail injection model. All the details extracted from the high resolution Pan image are added to the MS image with a detail injection gain value of 1. *i.e.* the fused method can be simply defined as:

$$\hat{M}_{AWLP}(x) = \tilde{M}(x) + \ddot{P}(x). \quad (1)$$

Where $\tilde{M}$ is the upscaled low resolution MS image, $\ddot{P}$ represents the high frequency details extracted from the high resolution Pan image and $\hat{M}$ represents the pansharpened image.

2.2. Context adaptive À-trous Wavelet Pansharpening (ATWT-CBD)

In [5], Aiazzi *et al.* proposed a context adaptive multiresolution fusion process called Context Driven Fusion. The authors proposed the use of a Laplacian pyramid for the extraction of high frequency component from the Pan image and the addition of these details to the low resolution MS images using a context driven model. However, instead of using the Laplacian pyramid, the À-trous Wavelet transform can be used during the pansharpening process. The Fusion process can be described as follows: i) Upscale the low resolution MS images (noted as $\tilde{M}(x)$); ii) Extract the high frequency details from the high resolution Pan image (noted as $\ddot{P}(x)$); iii) The fused image $\hat{M}_{AWCBD}(x)$ is defined by:

$$\hat{M}_{AWCBD}(x) = \tilde{M}(x) + \mathbf{1}_{L_{\tilde{M},\tilde{P};B}>\theta}(x) \times G_{\tilde{M},\tilde{P};B}(x) \times \ddot{P}(x), \quad (2)$$

where $\tilde{P}(x)$ is the upscaled low resolution Pan image. The indicator function $\mathbf{1}_{L_{\tilde{M},\tilde{P};B}>\theta}$ is defined as:

$$\mathbf{1}_{L_{\tilde{M},\tilde{P};B}>\theta}(x) = \begin{cases} 1, & L_{\tilde{M},\tilde{P};B}(x) > \theta \\ 0, & L_{\tilde{M},\tilde{P};B}(x) \leq \theta \end{cases} \quad (3)$$

θ is a constant threshold. $L_{\tilde{M},\tilde{P};B}(x)$ is the local correlation computed on the neighborhood B_x defined around the pixel x between $\tilde{M}$ and $\tilde{P}$, i.e.

$$L_{\tilde{M},\tilde{P};B}(x) = \frac{\sum_{y \in B_x} (\tilde{M}(y) - \mu_{\tilde{M};B}(x))(\tilde{P}(y) - \mu_{\tilde{P};B}(x))}{|B_x| \sigma_{\tilde{M};B}(x) \sigma_{\tilde{P};B}(x)},$$

$\mu_{\tilde{M};B}(x)$ (resp. $\mu_{\tilde{P};B}(x)$), $\sigma_{\tilde{M};B}(x)$ (resp. $\sigma_{\tilde{P};B}(x)$) are the mean and standard deviation computed on B_x of the image $\tilde{M}$ (resp. $\tilde{P}$). In practice, B_x is a square window of size s centred at x. $G_{\tilde{M},\tilde{P};B}(x)$ is the local gain computed on B_x and is defined as:

$$G_{\tilde{M},\tilde{P};B}(x) = \frac{\sigma_{\tilde{M};B}(x)}{\sigma_{\tilde{P};B}(x)}$$

The method estimates a detail injection gain for the addition of the details if the local correlation between the low resolution MS and Pan images is high. In addition, one should notice that the detail injection model is local and hence context adaptive. However, the model can be determined on fixed size windows or local regions. Thus, the proposed method does not take into account the size of the object or region while determining the model.

It was observed while testing both the AWLP and ATWT-CBD methods that the quantitative performances varied with regions, based upon the size of the object in the region of interest. Hence, an attempt was made to locally select which of the two methods should be used based upon the size of objects.

3. RELATIVE SIZE OF OBJECTS

In [9], the authors propose a method based on the topographic map of the image to estimate the local scale of each pixel in the case of gray scale remote sensing images. The idea is that, for each pixel, the most contrasted shape containing it is extracted, and the scale of this shape defines the characteristic scale of this pixel. The topographic map [10], which can be obtained by Fast Level Set Transformation (FLST) [11], represents an image by an inclusion tree of the shapes (which are defined as the connected components of level sets). For each pixel x, there is a branch of shapes $f_i(x)$ ($f_{i-1} \subset f_i$) containing it. Note $I(f_i)$ the gray level of the shape $f_i(x)$, $S(f_i)$ its area and $P(f_i)$ its perimeter. The contrast of the shape $f_i(x)$ is defined as $C(f_i) = |I(f_{i+1}) - I(f_i)|$.

The most contrasted shape $f_{\hat{i}}(x)$ of a given pixel is defined as the shape containing this pixel, of which the contrast is the most important, i.e.

$$f_{\hat{i}}(x) = \arg\max_j \{C(f_j(x)\} \quad (4)$$

Since the optical instruments always blur remote sensing images, several shapes with very low contrasts can belong to the same structure. In order to deal with the blur, the authors of [9] propose a geometrical criterion to cumulate the contrasts of the shapes corresponding to one given structure. The idea is that the difference of the areas of two succesive shapes (for example f_i and f_{i+1}) corresponding to one given structure is proportional to the perimeter of the smaller shape, i.e. $S(f_{i+1}) - S(f_i) \sim \lambda P(f_i)$, where λ is a constant. It is shown in [9] that the most contrasted shapes extracted in an image form a segmentation of this image. For a given pixel, the segmented region obtained by the method of [9] is actually the most significant structure containing this pixel. The scale $E(x)$ of the pixel x is then defined as the size of the segmented region containing it, *i.e.*

$$E(x) = S(f_{\hat{i}}(x)). \quad (5)$$

4. PROPOSED PANSHARPENING METHOD

It was observed that the performance of AWLP approach is better than ATWT-CBD on an image containing mainly small scale regions (for example images of dense urban area, *etc.*), because AWLP better preserves small details. However, the details $\ddot{P}$ extracted from images containing mainly large scale homogeneous regions (for example images with fields, *etc.*) could be noisy. Therefore, on such images, the pansharpened results obtained by AWLP will be noisy. Contrarily, ATWT-CBD adds the details only in those regions where the upscaled MS image is well correlated with the upscaled low resolution Pan image. This ensures better performances of ATWT-CBD on the images containing large scale regions. This observation will be confirmed by the experiments presented in Section 5.

A natural solution for taking advantage of these two approaches is to select for each pixel the optimal approach according to the scale of the most significant structure containing this pixel, which is the main idea of our proposed approach. We associate to each pixel the scale of the most significant region containing it computed on the high resolution Pan image P according to Equation (5), noted as $E_P(x)$. With the help of this value, we incorporate Equations (2) and (1). If the scale $E_P(x)$ of the most significant region containing x is large (*i.e.* larger than a threshold γ), we use ATWT-CBD to fuse $\tilde{M}(x)$ and $\tilde{P}(x)$ on this pixel in order to reduce the noise introduced by AWLP. Otherwise if the scale of the region is small (*i.e.* $E_P(x) \leq \gamma$), we use AWLP in order to preserve the details. Therefore the fused image is finally defined by:

$$\hat{M}(x) = \mathbf{1}_{E_P(x)\leq\gamma}\hat{M}_{AWLP}(x)+(1-\mathbf{1}_{E_P(x)\leq\gamma})\hat{M}_{AWCBD}(x), \quad (6)$$

where $\hat{M}_{AWLP}(x)$ and $\hat{M}_{AWCBD}(x)$ are computed respectively by Equation (1) and Equation (2).

5. EXPERIMENTS AND RESULTS

5.1. Data sets

For comparing the results of the proposed method with existing pansharpening methods (AWLP and ATWT-CBD) both qualitative and quantitative analysis were performed on simulated Pléiades data. Since, the Pléiades data sets used is simulated the high resolution MS image is available and hence the quantitative analysis can be performed at the desired high resolution. The high resolution MS and Pan images along with the low resolution MS image are provided by CNES (the French Space Agency) at 0.8m and 3.2m resolutions, respectively. The 1024 x 1024 pixel test images were selected from the images taken over the cities of Toulouse and Strasbourg in France. The Toulouse image represents an urban region having small objects and fine details whereas, the Strasbourg image represents countryside having large objects. Two such regions for each image were used for testing and verification however, due to constraints of space only one result for each region is presented in this paper.

5.2. Quantitative evaluations

Quantitative quality assessment is generally performed using a reference MS image to assess the quality of the pansharpened images, as defined by Wald's synthesis property [4]. Relative Global Error of Synthesis (ERGAS), Spectral Angle Mapper (SAM) and Quaternion Index (Q4) [8] can be used to determine global quality of pansharpened images. We have used these indexes for quality assessment as they were used in the 2006 IEEE Data Fusion Contest [8].

5.3. Parameters and Results

For comparison, we performed AWLP, ATWT-CBD and the proposed method for the pansharpening of the data sets. The threshold of the local correlations for the ATWT-CBD method is $\theta = 0.5$ (see Equation (2)). According to [5], the result of pansharpening is not very sensitive to the choice of θ if it is less than 0.8. The neighbourhood B_x (see Equation (3)) for computing the local correlations and variances is defined by a 16×16 analysis window. This choice is based on the experiments using the ATWT-CBD method with different window sizes varying from 12 to 32. When the window size is too small, the considered context is not enough for reducing the noise. If the window size is too large, the improvement of the result is not significant, while the computation time increases a lot. The threshold of the local object size is $\gamma = 256$ (see Equation (6)). If the object is smaller than the neighbourhood B_x, it is considered as a small object. We therefore use AWLP method on this object in order to preserve more details. Otherwise, we use ATWT-CBD method in order to reduce the noise.

The results obtained for the Strasbourg and Toulouse Pléiades images are presented in Tables 1& 2, respectively. From Table 1 it can be seen that for all the indexes determined the ATWT-CBD produces the best results. However from Table 2, it is observed that, for SAM and Q4 indexes, AWLP produces better results, while for ERGAS, ATWT-CBD produces better results. This indicates that AWLP produces better results for urban images with fine details whereas ATWT-CBD performs better for objects with a large size. The proposed method which selects among the two methods produces results that are closer to the best results among the two methods. This means that the results of SAM, ERGAS and Q4 are closer to the ATWT-CBD method for the Strasbourg image and are closer to the AWLP method for the Toulouse image. The ERGAS value of the proposed method is even better than both the ATWT-CBD and AWLP methods. Thus, it can be concluded that the proposed selection method results in images which are quantitatively closer to the best result among the two methods.

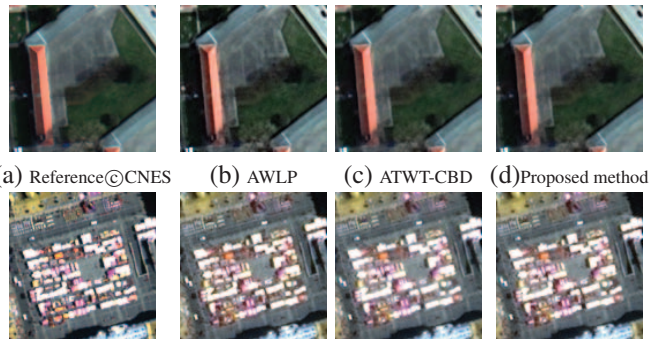
Approach	AWLP	ATWT-CBD	proposed method
ERGAS	2.2853	**1.8735**	2.0283
SAM	2.6383	**2.2914**	2.3573
Q4	0.8771	**0.8968**	0.8929

Table 1. Quantitative evaluation of the results obtained on the Strasbourg image by the three pansharpening approaches.

Approach	AWLP	ATWT-CBD	proposed method
ERGAS	3.4911	3.5539	**3.4762**
SAM	5.3153	**4.4998**	4.768
Q4	**0.9378**	0.9301	0.9353

Table 2. Quantitative evaluation of the results obtained on the Toulouse image by the three pansharpening approaches.

In Figure 1, we zoom at two regions of the pansharpening results obtained by AWLP, ATWT-CBD and the proposed method on the two simulated Pléiades images. For the visualization, only the red, green and blue channels are shown. It can be observed that, for the Strasbourg image, the AWLP method introduces spectral distortions on the building roofs and the fields, since their sizes are large. While the results of the ATWT-CBD method and the proposed method are quite consistent to the reference. On the Toulouse image, the results of the AWLP and the proposed method well preserve the details of the small objects, such as the cars. While the result of the ATWT-CBD method is more blurred.



(a) Reference©CNES (b) AWLP (c) ATWT-CBD (d)Proposed method

(e) Reference©CNES (f) AWLP (g) ATWT-CBD (h)Proposed method

Fig. 1. Zoom (128×128) at the reference MS images and the pansharpening results on **(a)-(d)** the Strasbourg image and **(e)-(h)** the Toulouse image.

6. CONCLUSION

À-trous Wavelet Fusion (AWLP) along with Laplacian Pyramid based Context Decision Fusion(CBD) were rated as the joint winners of the 2006 IEEE Data Fusion Contest . It was observed that instead of the Laplacian Pyramid the À-trous Filter can be used along with the CBD method thus producing a variant called ATWT-CBD. We observed that generally AWLP method produced better quantitative results for urban areas with fine details whereas the ATWT-CBD method quantitatively performed better when tested with country side scenes with large objects. Hence, a merging of the two algorithms was proposed in this paper to make use of AWLP method for pansharpening of finer details whereas ATWT-CBD for large objects in the same scene. The proposed algorithm was tested on Pléiades simulated data sets and the results obtained for the two different scenarios were closer to the best results from either of the two methods. The extension of this work to IKONOS and QuickBird images is actually investigated, as well as other decision fusion rules on the object size.

7. REFERENCES

[1] P. Sirguey, R. Mathieu, Y. Arnaud, M. M. Khan, and J. Chanussot, "Improving MODIS Spatial Resolution for Snow Mapping Using Wavelet Fusion and ARSIS Concept," *IEEE Geosci. Rem. Sens. Lett.*, vol. 5, pp. 78–82, 2008.

[2] P. S. Chavez, J. Stuart, C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photo. Eng. Rem. Sens.*, vol. 57, pp. 259–303, 1991.

[3] J. Núnez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Rem. Sens.*, vol. 37, no. 3, pp. 1204–1211, 1999.

[4] Lucien Wald, *Data Fusion: Definitions and Architectures — Fusion of images of different spatial resolutions*, ENSMP, 2002.

[5] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-Driven Fusion of High Spatial and Spectral Resolution Images Based on Oversampled Multiresolution Analysis," *IEEE Trans. Geosci. Rem. Sens.*, vol. 40, no. 10, pp. 2300–2312, 2002.

[6] X. Otazu, M. González-Audícana, O. Fors, and J. Núnez, "Introduction of Sensor Spectral Response Into Image Fusion Methods. Application to Wavelet-Based Methods," *IEEE Trans. Geosci. Rem. Sens.*, vol. 43, pp. 2376–2385, 2005.

[7] M. M. Khan, J. Chanussot, and L. Alparone, "Pansharpening of hyperspectral images using spatial distortion optimization," in *IEEE ICIP*, Cairo Egypt, 2009.

[8] 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. Rem. Sens.*, vol. 45, pp. 3012–3021, 2007.

[9] B. Luo, J.-F. Aujol, and Y. Gousseau, "Local scale measure from the topographic map and application to remote sensing images," *SIAM Multiscale Modeling and Simulation*, vol. 8, no. 1, pp. 1–29, 2009.

[10] V. Caselles, B. Coll, and J.-M. Morel, "Topographic maps and local contrast changes in natural images," *Int. J. Comp. Vision*, vol. 33, no. 1, pp. 5–27, 1999.

[11] P. Monasse and F. Guichard, "Fast computation of a contrast-invariant image representation," *IEEE Trans. on Image Processing*, vol. 9, no. 5, pp. 860–872, may 2000.