A New Evaluation Protocol for Image Pan-sharpening Methods

Miloud Chikr El-Mezouar, Nasreddine Taleb
Djillali Liabes University of Sidi Bel Abbes
Dept. of Electronics, RCAM Laboratory,
Sidi Bel Abbes, Algeria
{chikrelmezouar , ne_taleb}@univ-sba.dz

Kidiyo Kpalma, Joseph Ronsin
UEB INSA IETR Groupe Image et Télédétection
CS 70839 35708
Rennes Cedex 7, France
{Kidiyo.Kpalma , Joseph.Ronsin }@insa-rennes.fr

Abstract—**Pan-sharpening consists in fusing the spatial and spectral characteristics of panchromatic and multispectral (MS) images to get synthesized MS images. When such a fusion technique is proposed, it is delicate and important to evaluate its results. Generally, to evaluate the pan-sharpening methods both spectrally and spatially, a variety of quality indexes are available. Although, spectral indexes play a more important role than spatial ones to assess the fusion methods, spatial quality is important too. In this paper, a new protocol is proposed to evaluate pan-sharpening methods. This evaluation, by considering both spectral and spatial indexes, facilitates, reduces and even avoids any visual analyses, and allows automatic classification when comparing fusion methods.**

Keywords-image fusion; quality assessment; pan-sharpening methods evaluation; spectral indices; spatial indices

I. INTRODUCTION

Many quality indexes are available to evaluate pan-sharpening methods. Assessment, between each fused and reference Multispectral (MS) band, is possible by using the correlation coefficient (CC), the bias, the difference in variance (VAR), the standard deviation of the differences on a pixel basis (SD), and the correlation between high frequencies (sCC). Moreover, a global evaluation can be done using the average root mean square error, the universal quality index (Q4), the spectral angle mapper (SAM) and the relative dimensionless global error in synthesis (ERGAS) [1]. For the whole data set, SAM, ERGAS and Q4 are used. These indexes can be classified into two groups: spectral and spatial. Often, the pan-sharpened results are assessed using indexes based mainly on the spectral similarity. Hence, the presented metrics results tend to favor those methods improving the spectral quality. However, spatial quality is important too. The authors in [1] recommended the use of 9 indexes in pan-sharpening evaluation, where only one index is considered to assess spatial quality.

At first two fusion methods are considered to show the importance of the choice of the quality indexes: the standard Principal Component Analysis (PCA) based pan-sharpening [2] and wavelet (WAV) [3] fusion method. PCA is a commonly used technique for spectral transformation of the original data yielding uncorrelated principal components by a linear combination. It is assumed that the first PC image with the largest variance contains the major information from the original image and hence would be an ideal choice to replace the high spatial resolution PAN image. The PAN image is histogram matched with the first PC before the substitution. The remaining PCs, considered to have band-specific information, are unaltered. Inverse PCA is performed on the modified PAN image and the PCs to obtain a high-resolution pan-sharpened image [2]. Wavelet-based fusion schemes are extensions of the high-pass filter method, which makes use of the idea that spatial detail is contained in high frequencies. In the wavelet-based fusion schemes, detail information is extracted from the PAN image using wavelet transforms and injected into the MS image. Various models exist for injection information, with the simplest model being by substitution [3].

The obtained results of the PCA and wavelet fusion methods are shown in Fig. 1. Moreover, the most important indexes, as recommended in [1], are given in table 1. When comparing the PCA and wavelet fusion methods based only on the metrics of table 1, it appears that PCA is more valuable than wavelet, however, visual results, given in Fig. 1, are not coherent with this conclusion. The original MS and panchromatic (PAN) images are given in upper left and upper right corners of Fig. 1, respectively. The fused images, obtained using PCA and wavelet are shown in the lower left and lower right corners of Fig. 1, respectively.

TABLE I. METRICS COMPARISON BETWEEN PCA AND WAVELET FUSION METHODS.

	CC	**VAR**	**SD**	**sCC**	**Q_4**	**ERGAS**	**SAM**
PCA	**0.95**	0.48	**0.05**	**0.86**	**0.88**	**1.71**	**2.51**
WAV	0.90	**0.34**	0.07	0.82	0.81	2.01	3.11

This work was supported by the 'National Research Program' under projet: PNR N° 15/u22/1408.

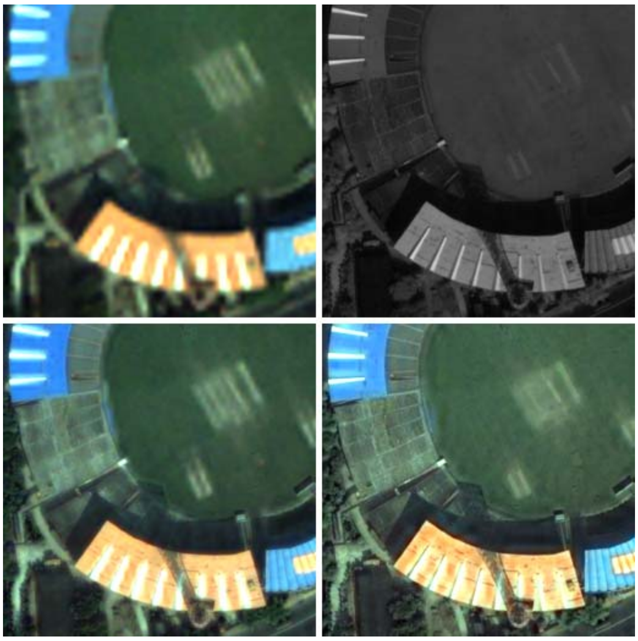

Figure 1. Top left: MS, top right: PAN, bottom left: PCA pan-sharpened, bottom right: wavelet pan-sharpened.

Visually it appears that, in addition to preserving the original colors, wavelet results are more accurate for spatial information representation. Hence for assessing fusion techniques, the choice of indexes consistent with the visual results is significant. In this paper, we propose an evaluation protocol to assess pan-sharpening methods. Both spectral and spatial qualities are considered. This protocol will be helpful when a classification of several techniques is to be accomplished.

II. Protocol Description

In general, most authors present their quantitative evaluation of the various methods in tables. Then they must add text to classify these methods. Our objective is to find a way to give, in the same table, both metrics and ranking methods in order to facilitate the author's explanation and reader's clearness. This method can also be used during experimentations of fusion techniques for assessment of their results with other approaches. Thus it can reduce the visual evaluation or help to reduce it.

To assess M methods, let K_1 and K_2 be, respectively, the numbers of the spectral and the spatial metrics. The experiments are achieved on N images. For each image, two tables of size (K_1xM) and (K_2xM), corresponding to spectral and spatial metrics respectively, are used for presenting the results. As shown in Fig. 2, each table component is represented as $C_m^n(k)$, where $m \in \{1,2,\cdots,M\}$, $n \in \{1,2,\cdots,N\}$, $k \in \{1,2,\cdots,K_1\}$ for spectral metrics and $k \in \{1,2,\cdots,K_2\}$ for spatial metrics.

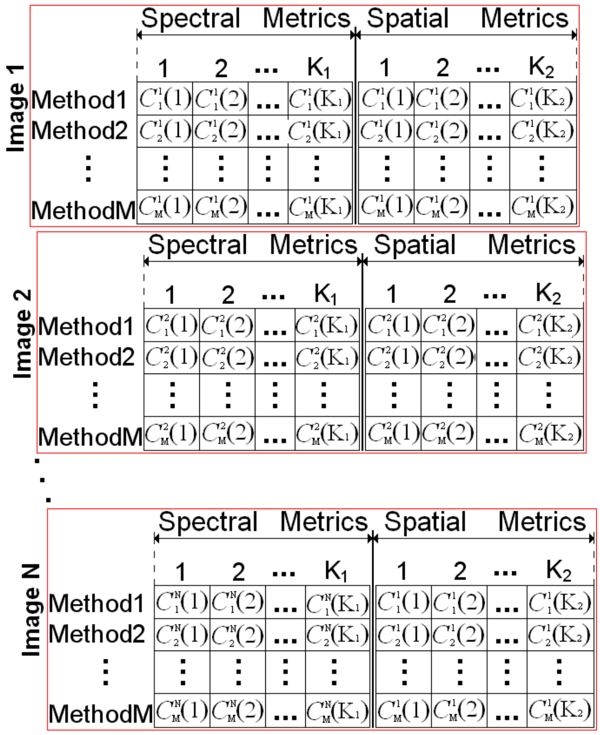

Figure 2. Spectral and spatial tables structure.

For the i[th] image and k[th] metric, a column $v^i(k)$ is constructed as $v^i(k) = \{C_1^i(k), C_2^i(k), \cdots, C_M^i(k)\}$, where the M components correspond to the M values obtained for the M methods to be evaluated, respectively.

At first, for each metric, the methods are sorted into two classes: satisfying and no-satisfying methods, based on a fixed metric threshold. As a simple case, the metric threshold can be chosen as being the mean of column $v^i(k)$ i.e. (μ_{ik}). However, for more reliability, we consider also the standard deviation of column $v^i(k)$ i.e. (σ_{ik}). Each metric has an ideal value which is considered in the classification of the methods. Generally, the ideal values are zero (0) or one (1), depending on the type of the metric. Hence, the way of combining σ_{ik} with μ_{ik} depends on the ideal value of the corresponding metric. For each metric, a threshold (θ_{ik}) is defined as:

$$\theta_{ik} = \mu_{ik} \pm \alpha\sigma_{ik}, \qquad (1)$$

where α is a value to be chosen experimentally. In this equation, $\alpha\sigma_{ik}$ is added to μ_{ik} if the optimal value of the corresponding metric is 1, and is subtracted from μ_{ik} if the optimal value is 0.

Then we use a simple statistical concept, based on logical values (0,1), to decide if a method is satisfactory or not. Using the i[th] image, a method (m) is considered satisfactory in term of the k[th] metric if its value is greater than the corresponding threshold and “1” is assigned to this method, else “0” will be assigned to it. The obtained results are expressed as:

$$B_m^i(k) = \begin{cases} 1 & \text{if } C_m^i(k) \geq \theta_{ik} \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Hence, for N images, one obtains N spectral tables and N spatial tables of values 1 and 0. Then both N tables are summed. This will produce two tables corresponding to the spectral and spatial metrics. Each component (A(m,k)) of these resulting tables is expressed as:

$$A(m,k) = \sum_{i=1}^{N} B_m^i(k). \tag{3}$$

After that, the columns of the spectral table and the spatial table are summed too. This will produce two columns characterizing the spectral and spatial indices: QI_{spec} and QI_{spat}. Each component of these two columns can be expressed as:

$$\sum_{k=1}^{K} A(m,k) = \sum_{k=1}^{K}\sum_{i=1}^{N} B_m^i(k). \tag{4}$$

Hence:

$$QI_{spec} = \left\{ \sum_{k=1}^{K_1}\sum_{i=1}^{N} B_1^i(k), \sum_{k=1}^{K_1}\sum_{i=1}^{N} B_2^i(k), \cdots, \sum_{k=1}^{K_1}\sum_{i=1}^{N} B_M^i(k) \right\}, \tag{5}$$

and

$$QI_{spat} = \left\{ \sum_{k=1}^{K_2}\sum_{i=1}^{N} B_1^i(k), \sum_{k=1}^{K_2}\sum_{i=1}^{N} B_2^i(k), \cdots, \sum_{k=1}^{K_2}\sum_{i=1}^{N} B_M^i(k) \right\}. \tag{6}$$

The maximum value of each column is reached when a method is satisfactory for all the metrics. This maximum value is NxK_1 and NxK_2 for the spectral and spatial cases, respectively. Thus, the various methods can be evaluated using, independently, the two columns after they are normalized. Nevertheless, the combination of the spectral and spatial results needs a normalization step, where each value of the spectral (or spatial) column is normalized by dividing it by the corresponding maximum value. Hence, to simplify the assessment, a global measure of quality index (QI_{glob}), resulting from the combination of the spectral (QI_{spec}) and the spatial (QI_{spat}) columns, can be expressed by a linear relation as:

$$QI_{glob} = a\frac{QI_{spec}}{NK_1} + b\frac{QI_{spat}}{NK_2} \tag{7}$$

where a and b are values to be adjusted experimentally, so that a+b=1 and $0<QI_{glob}<1$. Clearly, the higher the QI_{glob}, the better is the quality, and the lower the QI_{glob}, the worse is the quality.

Figure 3. Set of the ten images used for testing the proposed protocol.

III. Experimental Results

Experiments were conducted to evaluate the performance of the proposed protocol using Quickbird images downloaded from the landcover.org site. Ten images, shown in Fig. 3, containing forests, buildings, and roads, are used for the evaluation purpose. The protocol is based on the spectral indexes: CC, VAR, SD, Q4, ERGAS and SAM, and on the spatial indexes: sCC, Zhou spatial CC (ZCC) [1] and true edge (TE) [4].

Some of the popular pan-sharpening methods are the Fast Intensity Hue Saturation (FIHS), the Generalized IHS (GIHS), the Spectral Adjust IHS (SAIHS) [5], wavelet [3], PCA and NonSubsampled Contourlet Transform (NSCT) [2]. These techniques are implemented to test the proposed protocol. Tables 2 and 3 present the results found for spectral and spatial metrics, respectively, obtained for the first of the ten images.

It is clear that considering only spectral indexes, will lead to conclude that the PCA based method is the best. However, the authors in [2] have demonstrated that NSCT is more efficient than PCA: this is not consistent with the obtained results of table 2. Hence, the use of spatial indexes seems to be mandatory, however, when looking at table 3, it appears that, in term of sCC, PCA is good; in term of ZCC, SAIHS is good; and finally, in term of TE, FIHS is the best. In most of the literature, there is a consensus that the IHS-based fusion method is the best in preserving spatial information [6]. This is consistent with the TE and ZCC metrics. However and up to now, as can be seen in table 3, no ranking of the different methods can be envisaged yet when considering all metrics.
The visual analysis of an obtained result, corresponding to tables 2 and 3 and shown in Fig. 4, states that the wavelet and NSCT are the best methods.

TABLE II. SPECTRAL COMPARISON OF PAN-SHARPENING METHODS.

	CC	**VAR**	**SD**	**Q_4**	**ERGAS**	**SAM**
FIHS	0.70	0.63	0.11	0.63	2.82	3.40
GIHS	0.83	0.66	0.08	0.73	2.15	3.23
SAIHS	0.82	0.57	0.08	0.69	2.20	3.28
PCA	**0.95**	0.48	**0.05**	**0.88**	**1.71**	**2.51**
WAV	0.90	**0.34**	0.07	0.81	2.01	3.11
NSCT	0.91	0.35	0.07	0.82	1.96	3.03

In this case, where visual results do not correspond to metrics, it is necessary to have a protocol making the classification of methods reliable and a lot easier. Applying the proposed protocol produces table 4, which shows the two columns NV_{spec} and NV_{spac}, obtained by fixing α to 0.5.

The global measure NV_{glob} is then computed assuming that spatial indexes are as important as the spectral ones, so that a=b=0.5. Then, it is obvious that the NSCT based fusion method is the best. In accordance with visual analysis, wavelet is ranked in second position better than PCA, which is in fourth place. Thus, this proposed classification is reliable, in accordance with the visual evaluation, automatic and easy to apply.

TABLE III. SPPATIAL COMPARISON OF PAN-SHARPENING METHODS.

	sCC	**ZCC**	**TE**
FIHS	0.80	0.97	**0.80**
GIHS	0.79	0.98	0.67
SAIHS	0.76	**0.99**	0.64
PCA	**0.86**	0.86	0.42
WAV	0.82	0.95	0.65
NSCT	0.82	0.95	0.73

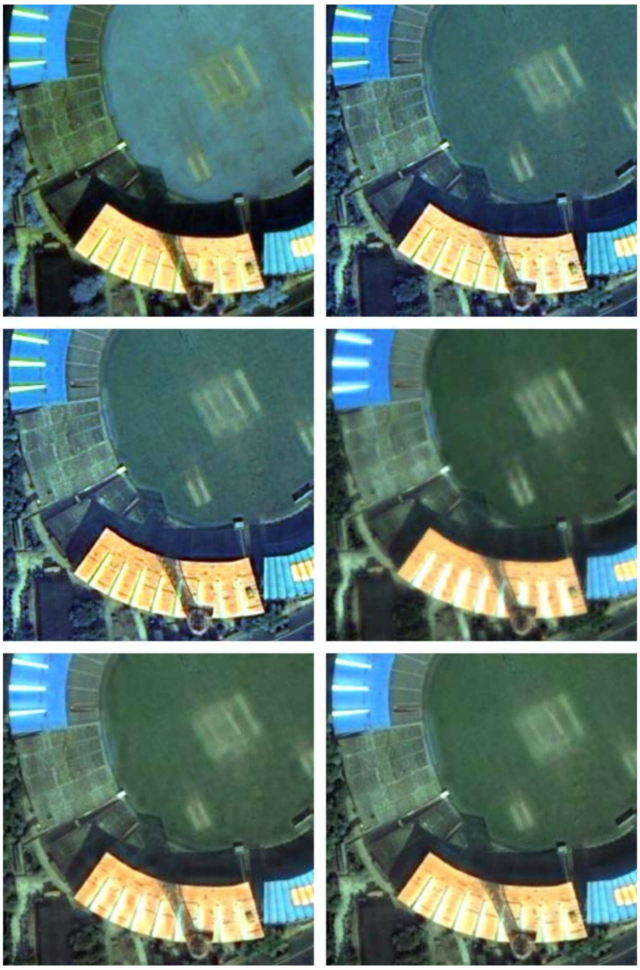

Figure 4. Pan-sharpened images: top left: FIHS, top right: GIHS, middle left: SAIHS, middle right: PCA, bottom left: wavelet, bottom right: NSCT.

TABLE IV. PROPOSED PROTOCOL ASSESSMENT RESULTS

	NVspec	**NVspac**	**NVglob**	**Rank**
FIHS	0.333	0.025	0.179	5
GIHS	0.333	0.025	0.179	5
SAIHS	0.333	0.075	0.204	3
PCA	0.133	0.263	0.198	4
WAV	**0.700**	0.388	0.544	2
NSCT	**0.700**	**0.538**	**0.619**	1

IV. CONCLUSION

The comparison of various pan-sharpening methods is not an easy task if the aspect of spatial characteristic is not considered. Moreover, the use of multiple indexes in the evaluation process enforces the final results. The proposed protocol can make the comparison of quality metrics a lot easier. Hence, more indexes can be integrated into this protocol to assess different methods making comparison a lot more accurate. The experiments conducted in this study show that using the proposed protocol can facilitate the visual evaluation of the results, by making assessing methods automatic and more reliable.

ACKNOWLEDGMENT

Imagery courtesy of Space Imaging LLC.

REFERENCES

[1] C. Thomas and L. Wald, “Comparing distances for quality assessment of fused products,” in Proc. 26th EARSeL Annu. Symp. New Develop. Challenges Remote Sens., Warsaw, Poland, May 2006. Z. Bochenek, Ed., Rotterdam, The Netherlands: Balkema, 2007, pp. 101 – 111.

[2] V. P. Shah, N. H. Younan, and R. L. King, “An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets,” IEEE Trans. Geosci. Remote Sens., vol. 46, pp. 1323 – 1335, May 2008.

[3] K. Amolins, Y. Zhang, and P. Dare, “Wavelet based image fusion techniques-An introduction, review and comparison,” ISPRS J. Photogramm. and Remote Sens., vol. 62, no. 4, pp. 249 – 263, Sep. 2007.

[4] Pradhan, P.S., King, R.L., Younan, N.H., and Holcomb, D.W.: “Estimation of the number of decomposition levels for a waveletbased multiresolution multisensor image fusion ,” IEEE Trans. Geosci. Remote Sens., vol. 44, pp. 3674 – 3683, May 2006.

[5] M. Chikr El-Mezouar, N. Taleb, K. Kpalma, and J. Ronsin, “An IHS-Based Fusion for Color Distortion Reduction and Vegetation Enhancement in IKONOS Imagery,” IEEE Trans. Geosci. Remote Sens., vol. 49, pp. 1590 – 1602, May 2011.

[6] T-M. Tu, Y-H. Lee, C-P. Chang, and P. S. Huang, “Adjustable intensity-hue-saturation and Brovey transformation fusion technique for IKONOS/QuickBird imagery,” Optical Engineering, vol. 44 (11), pp. 116201-1–116201-10, 2005.