Multispectral and Panchromatic Data Fusion Assessment Without Reference

Luciano Alparone, Bruno Aiazzi, Stefano Baronti, Andrea Garzelli, Filippo Nencini, and Massimo Selva

Abstract

This paper introduces a novel approach for evaluating the quality of pansharpened multispectral (MS) imagery without resorting to reference originals. Hence, evaluations are feasible at the highest spatial resolution of the panchromatic (PAN) sensor. Wang and Bovik's image quality index (QI) provides a statistical similarity measurement between two monochrome images. The QI values between any couple of MS bands are calculated before and after fusion and used to define a measurement of spectral *distortion. Analogously, QI values between each MS band and the PAN image are calculated before and after fusion to yield a measurement of* spatial *distortion. The rationale is that such QI values should be unchanged after fusion, i.e., when the spectral information is translated from the coarse scale of the MS data to the fine scale of the PAN image. Experimental results, carried out on very high-resolution Ikonos data and simulated Pléiades data, demonstrate that the results provided by the proposed approach are consistent and in trend with analysis performed on spatially degraded data. However, the proposed method requires no reference originals and is therefore usable in all practical cases.*

Introduction

Remote sensing image fusion techniques aim at integrating the information conveyed by data acquired with different spatial and spectral resolution from satellite or aerial platforms (Wald, 1999). The main goal is photo analysis, but also automated tasks such as feature extraction and segmentation/classification have been found to benefit from fusion (Colditz *et al.*, 2006; Bruzzone *et al.*, 2006). A variety of image fusion techniques are devoted to merge multispectral (MS) and panchromatic (PAN) images, which exhibit complementary characteristics of spatial and spectral resolutions (Wang *et al.*, 2005). Such an application of data fusion is often called *pansharpening*. Injection in the resampled MS images of spatial details extracted from the PAN image has been found to be adequate for preserving the spectral characteristics (Chavez Jr. *et al.*, 1991). Multiresolution analysis, based on undecimated wavelets decompositions and Laplacian pyramids, has proven itself effective to implement fusion at different resolutions (Núñez *et al.*, 1999; Aiazzi *et al.*, 2002a).

Quantitative results of data fusion are provided thanks to the availability of reference originals obtained either by simulating the target sensor by means of high-resolution data from an airborne platform (Laporterie-Déjean *et al.*, 2005), or by degrading all available data to a coarser resolution and carrying out fusion from such data. In practical cases, this strategy is not feasible. The underlying assumption, however, is that fusion performances are invariant to scale changes (Wald *et al.*, 1997). Hence, algorithms optimized to yield best results at coarser scales, i.e., on spatially degraded data, should still be optimal when the data are considered at finer scales, as it happens in practice. This assumption may be reasonable in general, but unfortunately may not hold for very high-resolution data, especially in a highly detailed urban environment, unless the spatial degradation is performed by using lowpass filters whose frequency responses match the shape of the modulation transfer functions (MTF) of the sensor (Aiazzi *et al.*, 2006).

In this work, we present a global index capable of measuring the quality of pansharpened MS images and working at the full scale without performing any preliminary degradation of the data. The spatial and spectral distortions are separately calculated from the fused MS image, the source MS image, and the PAN image. A combination of the spectral and spatial distortion indices may be performed to obtain a unique quality index. The rationale is that the interrelationships between any couple of spectral bands and between each band and the PAN image should be unchanged after fusion. Changes in the former are responsible for *spectral* distortion. Changes in the latter indicate *spatial* distortion. The underlying assumption of inter-scale preservation of cross-similarity is demonstrated by the fact that the true high-resolution MS data, whenever available, exhibit spectral and spatial distortions that are both zero, within the approximations of the model, and definitely lower than those attained by any fusion method. A thorough experimental section highlights the assets of the proposed approach compared with the two main approaches in the literature.

The paper is organized as follows. The two major protocols for MS + PAN image fusion assessment and the related statistical indices are reviewed next, followed by a description of how the mutual relationships among the

Luciano Alparone is with the Department of Electronics & Telecommunications, University of Florence, 3 Via Santa Marta, 50139 Florence, Italy (alparone@lci.det.unifi.it).

Bruno Aiazzi, Stefano Baronti, and Massimo Selva are with the Institute of Applied Physics "Nello Carrara" IFAC-CNR, 10 Via Madonna Del Piano, 50019 Sesto F.no (Florence), Italy.

Andrea Garzelli and Filippo Nencini are with the Department of Information Engineering, University of Siena, 56 Via Roma, 53100 Siena, Italy.

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 2, February 2008, pp. 193–200.

0099-1112/08/7402–0000/$3.00/0

original and fused data can be combined to yield separate spectral and spatial distortion indices that may be merged into a unique quality index. Extensive experimental results are then reported and discussed on two very high-resolution datasets, followed by conclusions.

Quality Assessment of Fusion Products

Quality assessment of pansharpened MS images is a much-debated problem (Chavez Jr. *et al.*, 1991; Wald *et al.*, 1997; Zhou *et al.*, 1998; Alparone *et al.*, 2004). So far, if quality is evaluated at the highest resolution of the PAN image, the measured spectral and spatial quality follow opposite trends, with the paradox that the least spectrally distorted fused image is that obtained when no spatial enhancement is introduced. Since the reference for spectral quality is the original MS image to be fused, while the reference for spatial quality is the PAN image, spectral and spatial quality are in counter-trend because all previous methods perform a direct comparison between the fused data and the references, i.e., between the MS data before and after fusion. To overcome this inconvenience, it is necessary to conceive new distortion measurements that do not depend on the true high-resolution (unavailable) data, but would yield best performance (zero distortions) if such data were hypothetically available.

Wald's Protocol

As there is no obvious reference available at high spatial resolution, original PAN and MS images are spatially degraded down to a lower resolution in order to compare fused product to the sole genuine references, constituted by the original (un-degraded) MS data set. This is the way to check the *synthesis* property (Wald *et al.*, 1997; Ranchin *et al.*, 2003); any synthetic image should be as identical as possible to the image that the corresponding sensor would observe with the highest spatial resolution, if existent. These authors also recommend checking another property called the *consistency* property; it states that any synthetic image once degraded to its original resolution should be as close as possible to the original image. In other words, a spatial degradation of the fused image should lead to the original image or close. Consistency, however, is a necessary condition, and its fulfillment may not imply a correct fusion. Therefore, the verification of the synthesis property only is more discriminating, and assessments usually focus on that. The original formulation of the protocol leaves the choice of the degradation filter as an open concern. A subsequent work has demonstrated that the spatial degradation of the dataset should be performed by using lowpass filters, whose frequency responses match the shapes of the MTFs of the different channels of the sensor (Aiazzi *et al.*, 2006). The main limitation of Wald's protocol is that fusion is actually assessed at a spatial scale different from that of the user application. Depending on the original resolution and/or landscape complexity, conclusions drawn at the coarser scale may not match user's expectation and thus the concept of *quality* at full scale (Zhang, 2004).

Zhou's Protocol

As an alternative to Wald's protocol, the problem of measuring the quality of fusion may be approached at the full spatial scale without any degradation. The spectral and spatial distortions are separately evaluated from the available data, i.e., from the original low-resolution MS bands and high-resolution PAN image.

According to the protocol described by Zhou *et al.* (1998), the spectral distortion is calculated for each band as the average absolute difference between the fused bands and the resampled input bands. Thus, any spatial enhancement of the (resampled) MS bands is supposed to produce a loss of spectral quality. Such a spectral distortion index is analogous to the correlation coefficient (CC) between fused and resampled original MS bands, introduced by Chavez Jr. *et al.* (1991) as a spectral quality index.

The spatial quality index of Zhou's protocol is given by the CC between the spatial details of each of the fused MS bands and those of the PAN image. Such details are extracted by means of a Laplacian filter, and the outcome spatial CC (SCC) should in principle be as close to one as possible, though no evidence of that is given either in the original paper (Zhou *et al.*, 1998) or in subsequent ones.

Objective Quality/Distortion Measurements

When spatially degraded MS images are processed for pansharpening, and therefore reference MS images are available for comparisons, assessment of *fidelity* to the reference usually requires computation of a number of different statistical indices. Examples are CC between each band of the fused and reference MS images, bias in the mean, root mean square error (RMSE), and average angular error, which measures the spectral distortion introduced by the fusion process.

Spectral Angle Mapper

Spectral Angle Mapper (SAM) denotes the absolute value of the spectral angle between two vectors, v and $\hat{v}$,

$$\mathrm{SAM} = \arccos\left(\frac{\langle v,\hat{v}\rangle}{\|v\|_2 \cdot \|\hat{v}\|_2}\right). \tag{1}$$

A resulting value of Equation 1 equal to zero denotes absence of spectral distortion, but possible radiometric distortion (the two pixel vectors are parallel but have different lengths). SAM is measured in either degrees or radians and is usually averaged over the whole image to yield a global measurement of spectral distortion.

ERGAS

Ranchin and Wald (2000) proposed an error index that offers a global picture of the quality of a fused product. This error is called ERGAS, the French acronym for *relative dimensionless global error in synthesis*, and is given by:

$$\mathrm{ERGAS} \triangleq 100\frac{d_h}{d_l}\sqrt{\frac{1}{L}\sum_{l=1}^{L}\left(\frac{\mathrm{RMSE}(l)}{\mu(l)}\right)^2} \tag{2}$$

where d_h/d_l is the ratio between pixel sizes of PAN and MS, e.g., 1:4 for Ikonos and QuickBird data, $\mu(l)$ is the mean (average) of the l^{th} band, and L is the number of bands. This index measures a distortion, and thus must be as small as possible.

Q4

An image quality index suitable for MS images having four spectral bands was recently proposed by Alparone *et al.* (2004) for assessing pansharpening methods. The quality index Q4 is a generalization to four-band images of the Q index (Wang and Bovik, 2002), which can be applied only to monochrome images. Q4 is obtained through the use of CC between hypercomplex numbers, or *quaternions*, representing spectral pixel vectors. Q4 is made of three different factors: the first is the modulus of the hypercomplex CC between the two spectral pixel vectors and is sensitive both to loss of correlation and to spectral distortion between the

two MS data sets. The second and third terms, respectively, measure contrast changes and mean bias on all bands simultaneously. The modulus of the hypercomplex CC measures the alignment of spectral vectors. Therefore, its low value may detect when radiometric distortion is accompanied by spectral distortion. Thus, both radiometric and spectral distortions may be encapsulated in a unique parameter. All statistics are calculated as averages on $N \times N$ blocks, either $N = 16$ or $N = 32$. Eventually, Q4 is averaged over the whole image to yield the *global* score index. The highest value of Q4, attained if and only if the test MS image is equal to the reference, is one; the lowest value is zero.

A Novel Quality Index Without Reference

The proposed approach to quality evaluation of pansharpened MS images relies on checking two main properties of the fused data. Following an idea also proposed by Pradhan *et al.* (2006), *spectral* quality of the fused data means that the similarity relationships between any couple of bands are unchanged after fusion. Analogously, but differently from (Pradhan *et al.*, 2006), *spatial* quality measures the extent to which the relationships between each MS band and the PAN image are preserved after fusion.

The problem is to find a mathematical function suitable for measuring such similarity relationships between two (monochrome) images and either two spectral bands or one spectral band and the PAN image. The simplest function is CC, as suggested by Pradhan *et al.* (2006), who compare the interband correlation matrices before and after fusion. However, experiments carried out during the present research on full resolution and spatially degraded MS data have shown that CCs are unchanged with scale only for couples of bands that are strongly correlated. Instead, CCs between any visible band and the near-infrared (NIR) band deviate by more than 20 percent from one scale to another. The reason of that is the use of *global* interband CC to measure the correlation of spatially non-stationary images. If the global CC is replaced with the spatially averaged *local* CC, e.g., calculated on small blocks covering the same areas before and after fusion, the similarity is preserved almost perfectly on the same MS dataset at two different resolutions.

However, the interband CC, being equal to the normalized cross-covariance, is insensitive to possible changes in the mean level and contrast originated by fusion. This means that if fusion abnormally changes, the local mean and/or the local contrast across bands, the CC will not measure such a distortion. Therefore, the plain CC should be integrated by measurements of differences in local mean and in local contrast, such that they are in principle invariant with the scale. As an example, if two original MS bands have CC = 90 percent, ratio of means equal to 0.8, and ratio of average local standard deviations (contrast) equal to 0.7, we wish that after fusion CC is still 90 percent, mean ratio is 0.8 and contrast ratio is 0.7. Even if the contrast of one band is significantly different before and after fusion, the contrast invariance property requires that the contrasts of all bands are multiplied by the same factor after fusion. By jointly considering the geometric and radiometric similarity of couples of MS bands, i.e., average local CC and average difference in local mean and in local contrast, it has been found that the cross-similarity relationships, on which the proposed approach relies, are fully captured by the quality index (QI), introduced by Wang and Bovik (2002) and suitable for monochrome images.

The Quality Index *Q*

The Q index, defined in Wang and Bovik, (2002) for a reference original image x and an image y to be tested, is stated as:

$$Q(x,y) \triangleq \frac{4\sigma_{xy} \cdot \bar{x} \cdot \bar{y}}{(\sigma_x^2 + \sigma_y^2)\,(\bar{x}^2 + \bar{y}^2)} \tag{3}$$

in which σ_{xy} denotes the covariance between x and y, $\bar{x}$ and $\bar{y}$ are the means, σ_x^2 and σ_y^2 the variances of x and y, respectively. Equation 3 may be equivalently rewritten as the product of three factors:

$$Q(x,y) \triangleq \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \times \frac{2 \cdot \bar{x} \cdot \bar{y}}{\bar{x}^2 + \bar{y}^2} \times \frac{2 \cdot \sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2} \tag{4}$$

The first one is the correlation coefficient (CC) between x and y. The second one is always ≤ 1, from Cauchy-Schwartz inequality, and is sensitive to bias in the mean of y with respect to x. The third term is also ≤ 1 and accounts for relative changes in contrast between x and y. Relative change means that, if the two contrasts σ_x and σ_y are both multiplied by the same constant, the contrast factor will be unchanged. Apart from CC which ranges in $[-1, 1]$, being equal to 1 *iff* $x = y$, and equal to -1 *iff* $y = 2\bar{x} - x$, i.e., y is the negative of x, all the other terms range in [0,1], if $\bar{x}$ and $\bar{y}$ are non-negative. Hence, the dynamic range of Q is $[-1,1]$ as well, and the best value $Q = 1$ is achieved *iff* $x = y$ for all pixels. To increase the discrimination capability of the three factors in Equation 4, all statistics are calculated on suitable $N \times N$ image blocks and the resulting values of Q averaged over the whole image to yield a unique global score.

Spectral Distortion Index

A spectral distortion index can be derived from the difference of inter-band QI values calculated from the fused MS bands, indicated as $\{\hat{G}_l\}_{l=1}^{L}$, and from the low-resolution MS bands, $\{\tilde{G}_l\}_{l=1}^{L}$. The terms $Q(\hat{G}_l, \hat{G}_r)$ and $Q(\tilde{G}_l, \tilde{G}_r)$ can be grouped into two $L \times L$ matrices. The two matrices are symmetrical and the values on the main diagonal are all equal to one.

A spectral distortion index, referred to as D_λ, is calculated as:

$$D_\lambda \triangleq \sqrt[p]{\frac{1}{L(L-1)} \sum_{l=1}^{L} \sum_{\substack{r=1 \\ r \neq l}}^{L} \left| Q\left(\hat{G}_l, \hat{G}_r\right) - Q\left(\tilde{G}_l, \tilde{G}_r,\right) \right|^p} \tag{5}$$

with p being a positive integer exponent chosen to emphasize large spectral differences: for $p = 1$, all differences are equally weighted; as p increases, large components are given more relevance. The index (Equation 5) is proportional to the p-norm of the difference matrix, being equal to 0, if and only if the two matrices are identical. If negative values of $Q\left(\hat{G}_l, \hat{G}_r\right)$ and $Q\left(\tilde{G}_l, \tilde{G}_r\right)$, originated by anti-correlated bands, are clipped below zero; then, Equation 5 is always lower than or equal to one.

Spatial Distortion Index

A spatial distortion index is calculated as:

$$D_s \triangleq \sqrt[q]{\frac{1}{L} \sum_{l=1}^{L} \left| Q\left(\hat{G}_l, P\right) - Q(\tilde{G}_l, \tilde{P}) \right|^q} \tag{6}$$

in which P is the PAN image, and $\tilde{P}$ a spatially degraded version of the PAN image obtained by filtering with a lowpass filter having normalized frequency cutoff at the resolution ratio between MS and PAN, followed by decimation. Analogously, D_s is proportional to the q-norm of the difference vector, where q may be chosen so as to emphasize

higher difference values. The index D_s attains its minimum (equal to zero), when the two vectors are identical. Analogously to Equation 5, also Equation 6 is upper bounded by one if clipping below zero of Q values is enabled.

Jointly Spectral and Spatial Quality Index

The use of two separate indices may be not sufficient to establish the ranking of performances of fusion methods. In fact, D_λ and D_s respectively measure changes in spectral behavior occurring between the resampled original and the fused images and discrepancies in spatial details originated by fusion.

To trade off the above trends, let us introduce a single index, namely QNR, i.e., *Quality with No Reference*, which is the product of the one's complements of the spatial and spectral distortion indices, each raised to a real-valued exponent that attributes the relevance of spectral and spatial distortions to the overall quality. The two exponents jointly determine the non-linearity of response in the interval [0,1], same as a *gamma* correction, to achieve a better discrimination of the fusion results compared:

$$\text{QNR} \triangleq (1 - D_\lambda)^\alpha \cdot (1 - D_s)^\beta. \tag{7}$$

Thus, the highest value of QNR is one and is obtained when the spectral and spatial distortions are both zero. The main advantage of the proposed index is that, in spite of the lack of a reference data set, the global quality of a fusion product can be assessed at the full scale of PAN.

Experimental Results

The goal of this section is validating the proposed approach and the related indices, compared with Wald's and Zhou's protocols and related indices. Extensive tests have been carried out on two different very-high resolution MS + PAN datasets, covering approximately the same area, but acquired by different instruments, of which reference MS data are available for objective measurement. Several fusion methods have been characterized by spectral and spatial quality measured according to the different protocols. The following fusion methods have been utilized in the experiments:

- Method based on Generalized Laplacian Pyramid (GLP) (Aiazzi *et al.*, 2002a) with Context-Based Decision (CBD) injection model and MTF adjustment (Aiazzi *et al.*, 2006);
- GLP-based method with the Enhanced Spectral Distortion Minimizing injection model and MTF adjustment (GLP-ESDM) (Aiazzi *et al.*, 2006);
- Generalized Intensity-Hue-Saturation method with Spectral Adjustment (GIHS-SA) (Tu *et al.*, 2004);
- GIHS method with injection model based on Genetic Algorithms (GIHS-GA) (Garzelli and Nencini, 2006a; Garzelli and Nencini, 2006b);
- Gram-Schmidt spectral sharpening method (GS) (Laben and Brower, 2000), as implemented in ENVI (ENVI, 2004).

The *void* fusion method, corresponding to plain (bicubic) resampling of the MS dataset at the scale of PAN is also included in the comparisons and will be referred as EXP. Note that in the proposed approach, the original low-resolution MS data and not their interpolated versions (as in Zhou's Protocol) are taken as reference of spectral quality. Hence, the spectral distortion of the resampled MS data (EXP) may not be identically zero, as it happens with Zhou's Protocol, because such a distortion depends on the interpolation filter that is generally non-ideal.

Description of the Datasets

From 1999 to 2002, CNES (Centre National d'Études Spatiales: French Space Agency) led a research program aimed at identifying and evaluating panchromatic/multispectral fusion methods on infra-metric spatial resolution images. For this study, images have been acquired from an airborne platform over nine sites covering various landscapes (rural, urban, or coastal). Five methods, selected as the most efficient, were compared on this data set. The evaluation was both qualitative with visual experts opinions and quantitative with statistical criteria (Laporterie-Déjean *et al.*, 2005).

Four MS bands have been acquired by the airborne *Pelican* instrument with 60 cm sampling distance. Since the panchromatic camera of the future Pléiades instrument was under development, the high-resolution panchromatic image was obtained by: (a) averaging the green and red channels, (b) applying the nominal MTF of the panchromatic camera, (c) resampling the outcome to 80 cm, (d) adding the instrument noise, and (e) recovering the ideal image by means of inverse filtering and wavelet denoising. The MS bands with spatial resolution four times lower than that of the PAN image were simulated by proper lowpass filtering and decimation. The frequency response of the lowpass filter was designed so as to match the MTF of each spectral channel of the spaceborne instrument. The size of the simulated panchromatic image is approximately 5,000 × 20,000 pixels; the four *reduced* MS bands are 1,250 × 5,000 pixels. Of the six scenes made available by CNES, only that portraying the city of Toulouse was used for evaluations.

The main advantage of the simulated Pléiades data set is that fusion performances can be objectively evaluated at the same spatial scale of the final user's product, unlike in fusion assessments carried out on spatially degraded versions of commercial data. Perhaps one disadvantage is that the simulated PAN image is *narrowband*, i.e., does not comprise the NIR wavelengths, analogously to SPOT-5, while the instrument which will be launched will have spectral response in the interval 500 to 850 nm, analogously to other instruments (Ikonos and QuickBird). The bandwidth of PAN is likely to influence the performances of fusion algorithms.

To overcome this limitation, a second data set from a commercial instrument was used as well. The proposed method has been assessed on very high-resolution image data collected by the spaceborne Ikonos MS scanner, again on the city of Toulouse, France. The four MS bands of Ikonos span the visible and NIR wavelengths and are non-overlapped, with the exception of B1 and B2: B1 = 440 to 530 nm, B2 = 520 to 600 nm, B3 = 630 to 700 nm, and B4 = 760 to 850 nm. The bandwidth of PAN embraces the interval 450 to 950 nm. The data set has been radiometrically calibrated from digital counts and geo-coded to 4 m (MS) and 1 m (PAN) GSD. A square region covering 4.2 km^2 was analyzed. The original PAN image is of size 2,048 × 2,048 pixels and the original MS image of size 512 × 512 pixels. The MS and panchromatic image were downsized by four to allow quantitative evaluations to be performed. Although fusion is assessed at a scale different from that of the user's product (2.8 m instead of 0.7 m), the data set is significant because the panchromatic image has not been synthesized from the MS bands, and its bandwidth comprises also the NIR wavelengths.

Comparison of Protocols and Indices on Ikonos Dataset

The first experiment aims at demonstrating that the results of the proposed approach, which does not require reference originals, is in accordance with other score indices that require reference originals. To this purpose, according to the protocol proposed by Wald *et al.*, (1997), the Ikonos data set has been spatially degraded by four, by means of Gaussian-like lowpass filters matching the shapes of the MTFs of the different spectral channels (B, G, R, NIR, and PAN) of the instrument (Cook *et al.*, 2001).

For SAM, ERGAS, and Q4, statistics have been calculated between 4 m fused and original MS data. All indices that do not require reference originals are calculated from fused bands, $\hat{G}_k$, input bands G_k, bicubically resampled input bands expand $\{\tilde{G}_k\}$, PAN image P and lowpass filtered and decimated PAN image, $\tilde{P}$. They are the proposed D_λ and D_s, and the outcome QNR, Zhou's spectral distortion index on the k^{th} band,

$$D_{Zhou}(k) \triangleq \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left|\hat{G}_k(i,j) - \text{expand}\left\{\tilde{G}_k\right\}(i,j)\right|, \qquad (8)$$

and its average over all spectral bands, $\overline{D}_{Zhou}$, as well as SCC (CC between highpass details of fused band and PAN) both band-by-band and average.

The numerical values reported in Table 1 evidence that while the spectral distortion D_λ of the resampled low-resolution MS (EXP) is almost zero (it would be zero if interpolation were ideal), both the proposed spectral and spatial distortions of the reference (REF) are very close to zero. They are not identically equal to zero because all filters used for downscaling the data are Gaussian approximations of the unknown true MTFs of the MS and PAN instruments. The outcome QNR value of the 4 m reference original (REF) is close to one and far greater than those of any fusion method. QNR values, calculated without reference original, are substantially in accordance with the other scores, which require reference originals. Discrepancies are due to the fact that the exponents α and β in Equation 7 are taken both equal to one. Thus, spectral and spatial quality are given the same importance. Instead, SAM is supposed to measure mainly spectral distortion, while in ERGAS and Q4, it is impossible to quantify the sensitiveness to spectral and spatial distortion separately. This explains why the resampled image without enhancement (EXP), which has spectral distortion close to zero, exhibits better QNR than other methods providing a spatial enhancement. Eventually, for the first six indices in Table 1, GLP-CBD attains scores better than those of the other methods, closely followed by GIHS-GA.

The behavior of Zhou's indices, summarized by their band-averaged values $\overline{D}_{Zhou}$ and $\overline{SCC}$ reported in the last two columns of Table 1 is puzzling. According to the protocol stated by Zhou *et al.*, (1998), the true 4 m MS data (REF) are the most spectrally distorted among all entries and have also the least spatial quality, with the exception of the unfused data (EXP). This paradox occurs because there is no evidence, in general, that $\overline{D}_{Zhou}$ should be zero for the best fusion results, because the actual value is $\overline{D}_{Zhou} = 23.14$ for REF. Analogously, $\overline{SCC}$ should match the value attained by REF (i.e., $\overline{SCC} = 0.815$, in the present case) and not be as close to one as possible, as claimed. The ranking of methods according to Zhou's Protocol is almost opposite to those obtained from the proposed indices and from the three indices following Wald's Protocol.

The second experiment aims at comparing the numerical values of indices at the full scale, i.e., 1 m for Ikonos data. This is the case in which assessments will be performed in practice. Only the PAN image must be lowpass filtered and decimated by four along both rows and columns to obtain the degraded version $\tilde{P}$ used in Equation 6. Again, a Gaussian model of the MTF of the PAN channel, having amplitude equal to 0.17 at the Nyquist frequency (Cook *et al.*, 2001; Aiazzi *et al.*, 2006) has been used. Fusion results are shown in Plate 1. The numerical values of indices (SAM, ERGAS, and Q4 are missing because the true reference data at 1 m are unavailable) are reported in Table 2. Distortions D_λ and D_s are slightly lower than those in the previous case; hence quality (QNR) is higher. The performance ranks are substantially in trend with those at the coarser scale. A notable exception is that GIHS-SA and GLP-ESDM have been swapped from 4 m to 1 m scale. Again, Zhou's indices $\overline{D}_{Zhou}$ and $\overline{SCC}$, which roughly follow the same trend exhibited at 4 m, are not in agreement with D_λ and D_s, respectively. The completely different rankings of

TABLE 1. GLOBAL DISTORTION/QUALITY INDICES OF FUSED IKONOS DATA AT 4 M SCALE: SAM (DEGREES), ERGAS, AND Q4 ($N = 32$) NEED 4 M REFERENCE MS BANDS (REF); D_λ, D_s, QNR, $\overline{D}_{Zhou}$, AND $\overline{SCC}$ DO NOT. THE LAST TWO INDICES ARE AVERAGE VALUES OVER THE FOUR MS BANDS

Ikonos: 4 m	SAM	ERGAS	Q4	D_λ ($p = 1$)	D_s ($q = 1$)	QNR ($\alpha = \beta = 1$)	$\overline{D}_{Zhou}$	$\overline{SCC}$
REF	0	0	1	0.015	0.012	0.972	23.14	0.815
GLP-CBD	3.214	3.041	0.899	0.052	0.058	0.892	16.96	0.933
GLP-ESDM	4.968	3.977	0.885	0.198	0.124	0.701	17.58	0.959
GIHS-GA	3.335	3.391	0.882	0.060	0.073	0.870	14.17	0.958
GIHS-SA	4.366	3.788	0.873	0.202	0.153	0.674	21.18	0.996
GS	4.786	4.716	0.780	0.089	0.067	0.849	12.75	0.948
EXP	4.968	6.070	0.535	0.002	0.285	0.712	0	0.409

TABLE 2. GLOBAL DISTORTION/QUALITY INDICES OF FUSED IKONOS DATA AT 1 M SCALE: SAM (DEGREES), ERGAS, AND Q4 ($N = 32$) NEED 1 M REFERENCE MS BANDS, WHICH ARE UNAVAILABLE; D_λ, D_s, QNR, $\overline{D}_{Zhou}$, AND $\overline{SCC}$ DO NOT. THE LAST TWO INDICES ARE AVERAGE VALUES OVER THE FOUR MS BANDS

Ikonos: 1 m	SAM	ERGAS	Q4	D_λ ($p = 1$)	D_s ($q = 1$)	QNR ($\alpha = \beta = 1$)	$\overline{D}_{Zhou}$	$\overline{SCC}$
REF (UNAV.)	–	–	–	–	–	–	–	–
GLP-CBD	–	–	–	0.039	0.045	0.917	21.73	0.913
GLP-ESDM	–	–	–	0.154	0.083	0.775	22.41	0.917
GIHS-GA	–	–	–	0.021	0.069	0.911	17.80	0.993
GIHS-SA	–	–	–	0.113	0.122	0.777	22.02	0.998
GS	–	–	–	0.035	0.087	0.881	17.43	0.990
EXP	–	–	–	0.002	0.152	0.845	0	0.159

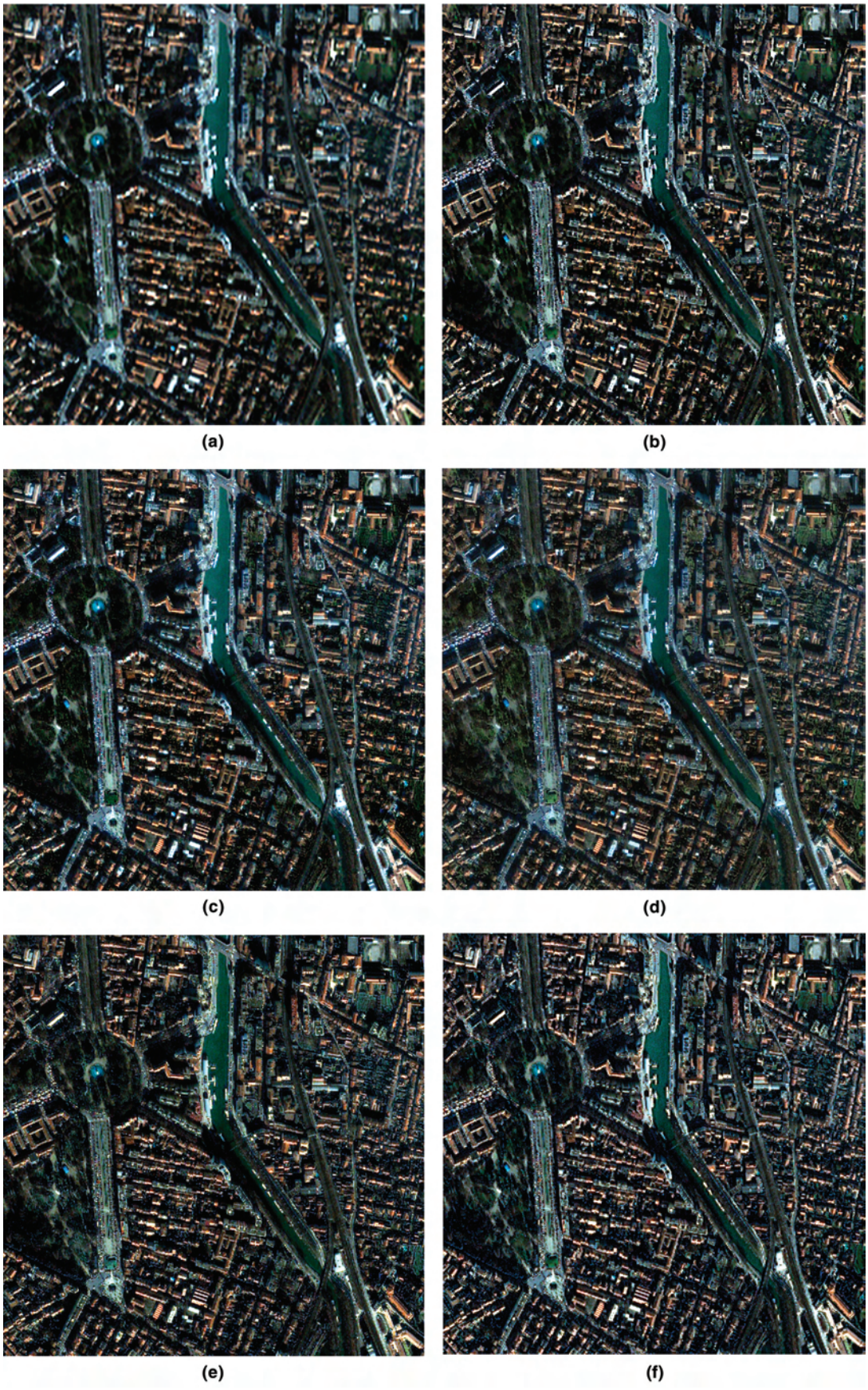

Plate 1. True color (3-2-1 MS bands as R-G-B display channels) of 1 m Ikonos data of Toulouse: (a) Bicubically resampled 4 m MS, (b) GLP-CBD fusion, (c) GHIS-GA fusion, (d) GS fusion, (e) GIHS-SA fusion, and (f) GLP-ESDM fusion.

fusion methods at 1 m scale obtained according to either the proposed protocol or Zhou's Protocol cannot be matched by objective evaluations because true 1 m MS data are unavailable. Therefore, extreme care must be taken of the choice of the evaluation protocol, to assess a new fusion method at the full scale of PAN.

Comparison of Protocols and Indices on Pléiades Dataset

The analysis carried out on Ikonos data has the disadvantage that the validation of the proposed indices has been performed at a scale (4 m) different from that used in practical applications (1 m). Therefore, one might object that the behavior of the indices at the full scale is unpredictable. The Pléiades dataset, having true MS images at 80 cm scale, is ideal for demonstrating that the proposed indices retain their confidence also at the lowest scale presently achievable by modern instruments.

A further concern is the lowpass filter used to spatially degrade the PAN image for calculating Equation 6. Since the equivalent MTF, (i.e., the MTF after restoration and denoising) of the Pléiades PAN channel is unavailable, we tried four different (separable) digital filters and evaluated the results of spatial distortion D_s obtained on the reference 80 cm MS data (the spectral distortion D_λ is unaffected by filtering the PAN image). Obviously, the best filter will yield the minimum of D_s on REF. The first filter is the Ikonos PAN filter with 0.17 amplitude at Nyquist. The second filter is the five-coefficients cubic spline wavelet filter by Starck and Murtagh (1994), widely used for image fusion (Núñez *et al.*, 1999), whose equivalent filter for decimation by four[1] exhibits amplitude at Nyquist equal to 0.19. The third filter is that of the red channel of Ikonos, which exhibits an amplitude equal to 0.29 at Nyquist. The fourth filter is non-Gaussian and almost ideal (trapezoidal) with amplitude equal to 0.5 at Nyquist, and is described by Aiazzi *et al.* (2002a) and reviewed by Aiazzi *et al.* (2006) in the context of MTF-adjusted multiresolution analysis filters. The numerical values of spatial distortion reported in Table 3 suggest that the lowpass filter used to smooth the PAN image for calculating D_s is noncrucial, provided that its frequency response match the shape of a likely optical MTF, having amplitude at Nyquist around 0.2. The simple and widespread cubic spline filter is definitely recommended to this purpose.

The values of all indices have been calculated on the Pléiades image of Toulouse and are reported in Table 4. This time, the resampled version EXP has been obtained by means of the almost ideal 23-coefficient interpolation filter described by Aiazzi *et al.* (2002a). What immediately stands out is that the underlying assumption of the proposed approach that the true reference data should exhibit spectral and spatial distortions that are both zero is even better verified on the Pléiades dataset at 80 cm scale. Perhaps the impressively low D_s also depends on the simulated Pan, but the low value of D_λ (about one-half of that on Ikonos) confirms the trend experienced on Ikonos that distortions (both reference-based and not) are lower, and hence quality is higher, at higher resolution. Concerning the ranking of methods, GLP-ESDM stands out as the second-most performing, after GLP-CBD. The explanation of that relies in the narrow-band PAN image, i.e., not comprising the NIR wavelengths and was noticed also by Laporterie-Déjean *et al.* (2005), who evaluated fusion algorithms working on the Pléiades datasets; the best performing method was very similar to GLP-ESDM, or better to its original version, called GLP-SDM (Aiazzi *et al.*, 2002b). In order to provide a fair comparison, GS and GIHS-SA, which implicitly use models of the spectral response of PAN far different from that of the simulated Pléiades PAN, were forced to follow the latter. Hence, their spectral distortions D_λ is surprisingly low. The outcome QNR, however, is in trend with the objective evaluations expressed by Q4 and ERGAS. Eventually, the unexplainable and mismatched values of D_{Zhou} and of $\overline{SCC}$ corroborate the conclusion that such indices, and the related evaluation protocol may be highly misleading if they are used for assessments of pansharpening products and methods.

TABLE 3. SPECTRAL DISTORTION D_s OF REFERENCE 80 CM PLÉIADED MS DATA VERSUS AMPLITUDE AT NYQUIST OF PAN FILTER

Amplitude at Nyquist	0.17	0.19	0.29	0.5
D_s	0.004	0.004	0.014	0.05

Conclusions

A new set of quality indices of pansharpened MS images has been developed, based on the evidence that the similarity relationships, measured by Wang and Bovik's Quality Index, between couples of bands and between each band and the PAN image should be unchanged from one scale to another, that is, before and after fusion is accomplished. Thus, the original MS and PAN data can be used to measure the spectral and spatial distortion, without resorting to spatial degradation of the complete MS + PAN dataset to a coarse scale. Only the PAN image must be degraded to the scale of MS. The proposed indices feature adjustable parameters, p and q for the definition of spectral and spatial distortions, respectively; α and β for the synthesis of the cumulative quality index QNR from such distortions. The tuning of such parameters has not been discussed in this work and is left to the user's applications. The experimental results, carried out on simulated Pléiades data and on Ikonos data by means of a number of fusion methods, demonstrate three main points: (a) *Consistency*, which means that true high-resolution MS data, whenever available, yield spectral and spatial distortions close to zero and far lower than those attainable by any fusion method; (b) *Congruence*, that is, the results provided by the proposed method are in trend with analysis performed by means of objective quality indices requiring reference originals, and (c) *Feasibility*: the method is directly usable in all practical cases; a model of the MTF of the PAN channel is required for lowpass filtering the PAN image, but is not crucial. Two interesting conclusions can be drawn from the experimental analysis. The quality of results of the various fusion methods depends on the spatial scale and on the imaging instrument; the performance ranks of fusion methods at full scale may not be the same as that at degraded scale. Unlike in a previous approach (Zhou *et al.*, 1998), specifically proposed to validate a new fusion algorithm, but generally usable to assess pansharpened images at the full spatial scale, in the present paper evidence has been given, on both true and simulated datasets, that measurements of inter-scale preservation of inter-band similarities is the key to achieve a blind, yet reliable evaluation protocol at full scale of pansharpening methods and products.

[1]If a signal is lowpass filtered and decimated by two, and the outcome is filtered again with the same filter and decimated by two again, this procedure is equivalent to filtering the original signal with an equivalent filter given by the linear convolution of the original filter by its upsampled version (coefficients are interleaved with zeroes) and decimating by four the outcome.

TABLE 4. GLOBAL DISTORTION/QUALITY INDICES OF FUSED PLÉIADES DATA AT 80 CM SCALE: SAM (DEGREE), ERGAS, AND Q4 ($N = 32$) NEED 80 CM REFERENCE MS DATA, WHICH ARE AVAILABLE, THE DATA BEING SIMULATED; $D\lambda$, D_S, QNR, $\overline{D}_{Zhou}$, AND $\overline{SCC}$ DO NOT. THE LAST TWO INDICES ARE AVERAGE VALUES OVER THE FOUR MS BANDS

PLEIADES:	SAM	ERGAS	Q4	D_λ ($p = 1$)	D_s ($q = 1$)	QNR ($\alpha = \beta = 1$)	$\overline{D}_{Zhou}$	$\overline{SCC}$
REF	0	0	1	0.009	0.004	0.986	30.59	0.843
GLP-CBD	3.673	2.778	0.965	0.045	0.034	0.921	25.70	0.963
GLP-ESDM	4.547	2.964	0.959	0.090	0.056	0.860	26.94	0.979
GIHS-GA	4.296	3.061	0.958	0.095	0.053	0.856	29.12	0.997
GIHS-SA	5.645	5.450	0.827	0.030	0.145	0.830	32.92	0.994
GS	5.574	5.114	0.844	0.030	0.127	0.848	25.99	0.987
EXP	4.547	6.079	0.780	0.000	0.150	0.850	0	0.270

Acknowledgments

The authors are grateful to the anonymous reviewers, whose insightful comments have notably improved the clarity of presentation of the paper. The Ikonos and simulated Pléiades datasets have been kindly made available by CNES. This work has been supported in part by MIUR under PRIN Cofin, 2005 "Development and Validation of Innovative Data-fusion Techniques for Environmental Remote Sensing."

References

Aiazzi, B., L. Alparone, S. Baronti, and A. Garzelli, 2002a. Context-driven fusion of high spatial and spectral resolution data based on oversampled multiresolution analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2300–2312.

Aiazzi, B., L. Alparone, S. Baronti, I. Pippi, and M. Selva, 2002b. Generalised Laplacian pyramid-based fusion of MS + P image data with spectral distortion minimisation, *ISPRS International Archives Photogrammetry and Remote Sensing*, 34(3B-W3):3–6.

Aiazzi, B., L. Alparone, S. Baronti, A. Garzelli, and M. Selva, 2006. MTF-tailored multiscale fusion of high-resolution MS and PAN imagery, *Photogrammetric Engineering & Remote Sensing*, 72(5):591–596.

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

Bruzzone, L., L. Carlin, L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, 2006. Can multiresolution fusion techniques improve classification accuracy?, *Image and Signal Processing for Remote Sensing XII*, Proceedings of SPIE, Vol. 6365.

Chavez Jr., P.S., S.C. Sides, and J.A. Anderson, 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic, *Photogrammetric Engnineering & Remote Sensing*, 57(3):295–303.

Colditz, R.R., T. Wehrmann, M. Bachmann, K. Steinnocher, M. Schmidt, G. Strunz, and S. Dech, 2006. Influence of image fusion approaches on classification accuracy: A case study, *International Journal of Remote Sensing*, 27(15):3311–3335.

Cook, M.K., B.A. Peterson, G. Dial, F. Gerlach, K. Hutchins, R. Kudola, and H.S. Bowen, 2001. IKONOS technical performance assessment, *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII* (S.S. Shen and M.R. Descour, editors), SPIE, 4381:94–108.

Garzelli, A., and F. Nencini, 2006a. Fusion of panchromatic and multispectral images by genetic algorithms, *Proceedings of the IEEE International Geoscience And Remote Sensing Symposium*.

Garzelli, A., and F. Nencini, 2006b. PAN-sharpening of very high-resolution multispectral images using genetic algorithms, *International Journal of Remote Sensing*, 27(15):3273–3292.

Laben, C.A., and B.V. Brower, 2000. *Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-sharpening*, U.S. Patent No. 6,011,875, Eastman Kodak Company.

ENVI, 2004. *User Manual, version 4.1*, Research Systems, Inc.

Laporterie-Déjean, F., H. de Boissezon, G. Flouzat, and M.-J. Lefévre-Fonollosa, 2005. Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR images, *Information Fusion*, 6(3):193–212.

Núñez, J., X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, 1999. Multiresolution-based image fusion with additive wavelet decomposition, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1204–1211.

Pradhan, P.S., R.L. King, N.H. Younan, and D.W. Holcomb, 2006. Estimation of the number of decomposition levels for a wavelet-based multiresolution multisensor image fusion, *IEEE Transactions on Geoscience and Remote Sensing*, 44(12):3674–3683.

Ranchin, T., B. Aiazzi, L. Alparone, S. Baronti, and L. Wald, 2003. Image fusion – The ARSIS concept and some successful implementation schemes, *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1–2):4–18.

Ranchin, T., and L. Wald, 2000. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation, *Photogrammetric Engineering & Remote Sensing*, 66(1):49–61.

Starck, J.L., and F. Murtagh, 1994. Image restoration with noise suppression using the wavelet transform, *Astronomy and Astrophysics*, 288:342–350.

Tu, T.-M., P.S. Huang, C.-L. Hung, and C.-P. Chang, 2004. A fast Intensity-Hue-Saturation fusion technique with spectral adjustment for IKONOS imagery, *IEEE Geoscience and Remote Sensing Letters*, 1(4):309–312.

Wald, L., 1999. Some terms of reference in data fusion, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1190–1193.

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

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

Wang, Z., D. Ziou, C. Armenakis, D. Li, and Q. Li, 2005. A comparative analysis of image fusion methods, *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1391–1402.

Zhang, Y., 2004. Understanding image fusion, *Photogrammetric Engineering & Remote Sensing*, 70(6):657–661.

Zhou, J., D.L. Civco, and J.A. Silander, 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data, *International Journal of Remote Sensing*, 19(4):743–757.