

# Quality assessment by region in spot images fused by means dual-tree complex wavelet transform

Diego Renza , Estibaliz Martinez, Agueda Arquero

*Computer School, Dept. of Architecture and Technology of Informatic Systems, Polytechnic University of Madrid, Campus Montegancedo, Boadilla del Monte, 28660 Madrid, Spain*

---

## Abstract

This work is motivated in providing and evaluating a fusion algorithm of remotely sensed images, i.e. the fusion of a high spatial resolution panchromatic image with a multi-spectral image (also known as pansharpening) using the dual-tree complex wavelet transform (DT-CWT), an effective approach for conducting an analytic and oversampled wavelet transform to reduce aliasing, and in turn reduce shift dependence of the wavelet transform. The proposed scheme includes the definition of a model to establish how information will be extracted from the PAN band and how that information will be injected into the MS bands with low spatial resolution. The approach was applied to Spot 5 images where there are bands falling outside PAN's spectrum. We propose an optional step in the quality evaluation protocol, which is to study the quality of the merger by regions, where each region represents a specific feature of the image. The results show that DT-CWT based approach offers good spatial quality while retaining the spectral information of original images, case SPOT 5. The additional step facilitates the identification of the most affected regions by the fusion process.

*Keywords:* Spot image; Image fusion; DT-CWT; Pansharpening; Quality assessment

## 1. Introduction

The term “image fusion” usually implies the integration of images acquired by multiple sensors with the intention of providing a better perspective of a scene that contains more content. For merging remotely sensed images, particularly one multi-spectral (MS) image and one panchromatic (PAN) image (also known as pansharpening), fusion algorithms should aim to integrate information from images of different spectral and spatial resolution, leading to obtain a single image that includes the best features of each one. The main restriction is to minimize the degradation of the spectral information from original multi-spectral image. In remote sensing, the fusion schemes can be classified in feature space, spatial domain and scale-

space techniques (Schowengerdt, 2007). The first category exploits the complementary characteristics of spatial and spectral resolutions of data (e.g. intensity-hue-saturation (IHS), principal component analysis (PCA), etc.). The second group of methods injects high frequency components from the PAN band over the MS bands (e.g. high-pass filter (HPF)).

In the third category, scale-space fusion algorithms have been found to be promising and their results can be used for quantitative studies of multi-spectral properties (Gonzalez-Audicana et al., 2004). Indeed, image fusion methods based on injecting high frequency components taken from the PAN band into MS resampled bands have shown a superior ability to translate the spectral information from MS image to the new fused bands (Wang et al., 2005), i.e. introducing minimal spectral distortions, as in the following cases: discrete wavelet transform (DWT) (Garguet-Duport et al., 1996), Laplace pyramids (Wang et al., 2005), AWT (à trous wavelet transform)

*E-mail addresses:* d.renza@alumnos.upm.es (D. Renza), emartinez@fi.upm.es (E. Martinez), aarquero@fi.upm.es (A. Arquero).

(Gonzalez-Audicana et al., 2004), among others. Therefore, multi-resolution analysis (MRA) has been recognized as one of the most efficient tools for the implementation of image fusion in different spatial resolution images.

Regarding critically subsampled transforms such as DWT, in some cases there may be aliasing, and in turn shift dependence of the wavelet transform. The latter means that small changes in the input signal (time or space) can cause major variations in the distribution of energy between wavelet transform coefficients at different scales and possibly large changes in reconstructed waveforms (Kingsbury, 1999); this explains the superiority of fusion schemes based on the AWT over algorithms based on DWT.

Likewise, spatial distortions may occur in MRA-based fusion approaches, resulting in translations or blurring of the contours and textures, hence, MRA-based mergers require the definition of a model to establish how information will be extracted from the high resolution PAN band and how that information will be injected into the MS bands. The aim is that fused bands should be as identical as possible to bands that the corresponding sensors would observe with the highest resolution, i.e. PAN resolution (Aiazzi et al., 2008).

With this goal, several methods have been proposed including histogram matching between PAN and each of the MS bands (Li et al., 2002; Nuñez et al., 1999; Wang et al., 2005), injection of weighted details according to context-driven method (Aiazzi et al., 2002) or window spectral response (WiSpeR) method which takes into account the relative spectral responses of PAN and MS bands in the definition of a set of weights governing the injection of high-pass wavelet planes (Otazu et al., 2005).

In the latter case, something to keep in mind is the spectral range that covers both the PAN and each of the MS bands that will be merged. Taking the case of Spot 5 sensor, where the bands cover the following wavelengths: PAN (0.48–0.71  $\mu\text{m}$ ), green (0.50–0.59  $\mu\text{m}$ ), red (0.61–0.68  $\mu\text{m}$ ) and near-infrared (NIR) (0.78–0.89  $\mu\text{m}$ ) it is clear that NIR spectral response falls completely outside the scope of PAN spectral response, see Fig. 1 which shows spectral sensitivities of SPOT 5<sup>1</sup>. Therefore, wisper method would not apply, because the weights to inject high pass details depend on the intersection between PAN and band-*i* spectral response, which is zero for NIR band.

According to the above, it would be appropriate to use a non-critically subsampled wavelet transform with an injection model, such as “dual-tree complex wavelet transform” (DT-CWT) (Selesnick et al., 2005) described in Section 2. DT-CWT has been used previously for fusing Quickbird images (Ioannidou and Karathanassi, 2007) combining the PAN details with the MS low frequency component (approximation), applied only to the RGB bands. That work proposed to decompose the PAN and the MS bands

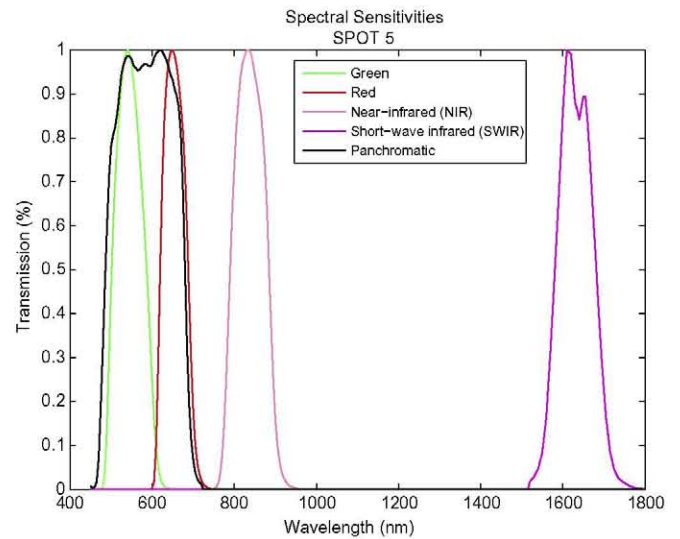


Fig. 1. SPOT 5 – spectral sensitivities.

using DT-CWT, where the wavelet coefficients of the PAN decomposition and the scaling coefficients of the MS decomposition are used to obtain the new fused image by means DT-CWT synthesis. This means that a model to define the injection of high frequency components into spectral bands does not exist yet, an aspect that becomes critical when some spectral bands fall outside the PAN spectrum, causing spectral distortions. In SPOT 5, where there are bands falling outside PAN’s spectrum this aspect can be advantageous while in Quickbird with a wide PAN spectral response, this added value would not be very significant.

The rest of this paper is organized as follows. In Section 3 we describe an injection model to be applied in conjunction with multi-resolution analysis based on DT-CWT. Section 4, describes the application and the results of the approach applied to Spot 5 images of Madrid, Spain. Here, the results of this methods are compared with alternative fusion approaches. Section 5 introduces the quality assessment by region. The conclusions are presented in Section 6.

## 2. MRA with DT-CWT

Multi-resolution analysis (MRA) gives a simple and fast method to analyze a signal at different resolutions; particularly, the wavelet transform uses two types of functions: scaling and wavelet, which together with their translations and dilations can see the different components of the signal. Scaling function gives a series of pictures of the signal, each at a resolution differing by a given factor from the previous resolution, while the differences of information between the signal seen at two successive resolutions are encoded by the wavelet function. Hence, the signal is studied at a coarse resolution (associated to a scaling function) to get an overall picture and at higher resolutions (associated to a wavelet function) to see an increasing level of details (Burke, 1998).

<sup>1</sup> <http://www.spotimage.com/web/fr/152-resolution-et-modes-spectraux.php> (accessed January 2011).

To achieve DWT in practice, there is a simple and fast method, the Mallat's algorithm (Mallat, 2009), which decomposes a signal into components differing in size by a factor of two. A wavelet coefficient  $w_{j,n}$  (or scaling coefficient  $a_{j,n}$ ) can be obtained by integrating the product of the wavelet function  $\psi(t)$  (or scaling function  $\phi(t)$ ) with the signal ( $n$  and  $j$  refer to the index of shifts and dilations). In practice one uses a simpler scheme, convolving the signal with both a short high-pass digital filter associated with the wavelet function and a short low-pass digital filter associated with the scaling function. The filter bank for DWT, hereafter referred to as a tree, corresponds to MRA of a signal  $x$  implemented by means of a filter bank that includes two analysis-filter types:  $h_{a_i}$  (low-pass filter) and  $g_{a_i}$  (high-pass filter); wavelet and scaling coefficients can be obtained as shown Eqs. (1) and (2), respectively:

$$a_{j,n} = \langle x, \phi_{j,n} \rangle = \sum_i h_{a_{i-2n}} a_{j-1,i}, \quad (1)$$

$$w_{j,n} = \langle x, \psi_{j,n} \rangle = \sum_i g_{a_{i-2n}} a_{j-1,i}. \quad (2)$$

In this case, the low-pass filtering followed by a down-sampling operator deliver coarser approximations of signal  $x$ , i.e. scaling coefficients  $a_{j,n}$ ; while the high-pass filtering followed by a downsampling operator deliver wavelet coefficients  $w_{j,n}$ . The reconstruction of the output signal,  $y$ , is possible through the synthesis filters  $\{\tilde{h}_{a_i}\}$  and  $\{\tilde{g}_{a_i}\}$ , as shown in Eq. (3):

$$y(n) = \langle x, \phi_{j-1,n} \rangle = \sum_i \tilde{h}_{a_{n-2i}} w_{j,i} + \sum_i \tilde{g}_{a_{n-2i}} w_{j,i}. \quad (3)$$

Discrete wavelet transform (DWT) is implemented using Mallat's algorithm, so after each filter there is a down-sampling stage, which discards every second wavelet coefficient at each decomposition level, therefore DWT is a critically sampled transform. That can lead to aliasing and a subsequent shift dependence of the wavelet transform, which means that small shifts in the input signal (time or space) can cause major variations in the distribution of energy between wavelet transform coefficients at different scales (Kingsbury, 1999). This can lead to small shifts in the input waveform causing large changes in the wavelet coefficients, large variations in the distribution of energy at different scales and possibly large changes in reconstructed waveforms. Consequently, shift-invariance is desirable for several applications such as, pattern recognition, image fusion, motion estimation or edge detection (Kingsbury, 1999; Liang and Parks, 1996; Mallat, 2009).

The dual-tree complex wavelet transform (DT-CWT) has as its main objective the reduction of the aliasing of the DWT and obtaining an analytic transform (i.e. its Fourier transform is zero for negative frequencies) to minimize the shift dependence, for which it's possible to use complex wavelets that are characterized by the use of a complex-type filter bank. With this objective, the DT-CWT proposes to replace the complex component by two real filter banks; to be of complex-type, the response of

the combination of these two filters should be analytic and they must form a "Hilbert pair" (90° out of phase with each other).

To make the DT-CWT analytic, or nearly so, two real DWT (corresponding to the two filter banks) are used, the first delivers the real DWT transform while the second delivers the imaginary part. Each DWT, can be implemented using Mallat's algorithm (Mallat, 2009). This algorithm is a discrete wavelet transform with a fast implementation, which obtains from a discrete signal input ( $N$  samples) the coefficients that represent the new domain. The DT-CWT is characterized by good shift invariance, good directional selectivity in 2D and 3D, near-perfect reconstruction with short support filters, limited redundancy – 2:1 in 1D, 4:1 in 2D and low computation requirements (compared with undecimated DWT) (Selesnick et al., 2005).

DT-CWT provides a time-frequency analysis of the signal by measuring its frequency content (controlled by the scale factor  $j$ ) at different times (controlled by the time shift  $n$ ). The algorithm decomposes the input signal  $x(n)$  in two signals or subbands,  $a_{j,n}$  and  $w_{j,n}$  that represent the low frequency component (approximation) and the high frequency component (details), respectively. DT-CWT with the complex wavelet function,  $\psi_a(x) + j\psi_b(x)$  and the complex scaling function  $\phi_a(x) + j\phi_b(x)$  decomposes an image  $I(x,y)$  in one complex scaling subband and six complex wavelet subbands (for each level). The wavelet subbands are oriented in six dimensions  $\Theta = \{\pm 15^\circ, \pm 45^\circ, \pm 75^\circ\}$  (Selesnick et al., 2005).

### 3. Image fusion with DT-CWT

The benefit of a fused image is that the single resulting image potentially contains both the high-spatial resolution and the spectral information. Hence, the result of image fusion is a new image which is more suitable for human and machine perception or further image-processing tasks such a classification, segmentation, feature extraction or object recognition.

The proposed DT-CWT-based image fusion approach is summarized in Fig. 2. The first step is to resample the MS image so that their bands have the same pixel size as the PAN band. Let  $MS$  denote the multi-spectral image,  $R$  denote the scale ratio between MS bands and PAN band; and let  $f$  denote the 2D interpolation function, then the new interpolated image is,  $MSr = f(MS, R)$ .

These bands, along with the PAN band are decomposed by means of DT-CWT. For each band, one complex-valued approximation band ( $A_j$ ) and six complex-valued directional detail subbands ( $W_j^\theta$ ) are generated in each scale of the DT-CWT (Celik and Ma, 2010):

$$A(x,y) = \sum_{n \in \mathbb{Z}^2} a_{j_0,n} \phi_{j_0,n}(x,y), \quad (4)$$

$$W(x,y) = \sum_{\theta \in \Theta} \sum_{j \geq j_0} \sum_{n \in \mathbb{Z}^2} w_{j,n}^\theta \psi_{j,n}^\theta(x,y). \quad (5)$$

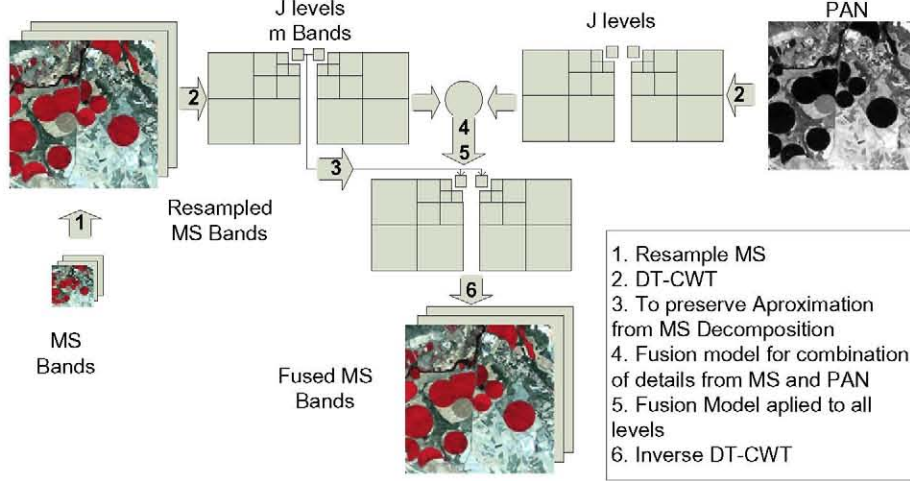


Fig. 2. Image fusion with DT-CWT. Overview.

After having the coefficients associated with MRA of the two input images, it is necessary to build a new image that contains both a high-spatial resolution coupled with multi-spectral information. Since spectral information is present only in the MS image, the approximation coefficients to be retained should be precisely those of the latter, without taking into account the corresponding PAN coefficients. Mathematically, using the GIF model (Wang et al., 2005) we have:

$$\begin{bmatrix} A_{F_1} \\ A_{F_2} \\ \vdots \\ A_{F_m} \end{bmatrix} = \begin{bmatrix} A_{MSr_1} \\ A_{MSr_2} \\ \vdots \\ A_{MSr_m} \end{bmatrix}. \quad (6)$$

And for detail coefficients:

$$\begin{bmatrix} W_{F_1}^\theta \\ W_{F_2}^\theta \\ \vdots \\ W_{F_m}^\theta \end{bmatrix} = \alpha \begin{bmatrix} W_{MSr_1}^\theta \\ W_{MSr_2}^\theta \\ \vdots \\ W_{MSr_m}^\theta \end{bmatrix} + \beta \begin{bmatrix} W_P^\theta \\ W_P^\theta \\ \vdots \\ W_P^\theta \end{bmatrix}. \quad (7)$$

Here,  $F$  denotes the new fused image,  $MSr$  the resampled multi-spectral image,  $A$  the approximation coefficients,  $W$  the wavelet coefficients (in direction  $\theta$ ),  $-m$  represents the spectral bands,  $\alpha$  and  $\beta$  denote the weight variables for the MS and PAN details, respectively.

For the spatial information, the first approach is to retain only the detail coefficients of the PAN band, this implies that the weight variables  $\alpha = 0$  and  $\beta = 1$  in Eq. (7), resulting in the panchromatic details injected in each band of the new fused image.

Since the high-frequency components of the MS image include not only spatial information but also spectral information, the high-frequency components should not be taken directly from PAN band. Therefore PAN detail coefficients will be injected into MS multiplied by a given

weight, so we will show how to inject the high frequency information (details) extracted from the PAN into resampled versions of the MS.

Given that the detail information may include coefficients either MS or PAN, it is necessary to define a model for comparison in order to generate a fusion map to determine which coefficients are extracted from the MS and which of the PAN. Map-fusion elements, which are nothing more than a logical matrix, will depend on which coefficient best represents the signal at a specific point.

The fusion map can be given by similarity measures such as correlation. In this case, it is necessary to calculate the degree of correlation between each MS and PAN detail bands, and also from a broader perspective to determine the correlation between same-level MS and PAN approximations. Therefore, we calculate the local correlation  $\rho$  (within a neighborhood) between the MS and PAN approximation for each scale, i.e.

$$\rho = \frac{\sum_p \sum_q [(C_{pq} - C'_{pq})(D_{pq} - D'_{pq})]}{\sqrt{\sum_{p,q} (C_{pq} - C'_{pq})^2 \sum_{p,q} (D_{pq} - D'_{pq})^2}}, \quad (8)$$

where  $'$  denotes the mean,  $C$  and  $D$  are the MS and PAN approximations in a given scale,  $p$  and  $q$  the size of the window (neighborhood). Accordingly, to obtain the fusion map, the local correlation value will be compared against a threshold ( $\tau$ ) determined by the overall correlation ( $P$ ) between each MS band and the PAN, which gives an overview of the spectral content of the PAN.

$$\tau = 1 - P. \quad (9)$$

Finally the fusion map,  $G^{\text{corr}}$ , is given by

$$G^{\text{corr}} = \begin{cases} 1, & \text{if } \rho \geq \tau, \\ 0, & \text{else.} \end{cases} \quad (10)$$

Since the map is calculated using the approaches of each level, it will be applied to all six DT-CWT subbands.

The map fusion gives us the locations where PAN details will be injected. After having defined the map fusion elements, it's necessary to determine how to combine them, therefore we must determine the weights  $\alpha$  and  $\beta$  in Eq. (7). Regarding  $\alpha$ , this weight is used where PAN details are not injected, in these locations MS details will be retained, so  $\alpha$  is given by

$$\alpha = 1 - G^{\text{corr}}. \quad (11)$$

The local correlation coefficient takes into account the degree of spectral content of the PAN with respect to MS allowing you to define what factors are to be removed from the PAN. For the injection of these details the gain factor  $\beta$  is defined locally by the ratio of standard deviation of MS-bands and PAN approximation for each scale:

$$\beta = \min \left\{ \frac{\sigma_{A_{MSm}}}{\sigma_{A_P}}, \xi \right\} * G^{\text{corr}}, \quad (12)$$

where  $\xi$  is a constant used to avoid numerical instabilities, and in this work  $\xi = 2.5$  as suggested by Aiazzi et al. (2002). The advantage of this model is to adjust the injection of details into MS bands with the objective of minimizing the spectral distortion. This minimization is equivalent to applying a radiometric correction on information from the PAN to be injected into the MS bands.

After having defined the new coefficients, it is necessary to return to the image domain. This is achieved by applying the inverse DT-CWT to these new coefficients, with the same number of decomposition levels. The full fusion process is showed in Fig. 3.

### 3.1. Quality assessment

Quality assessment implies a comparison between the fused product, its properties or some derived quantities, and a reference. A good quality is obtained if the product is close to this reference. The major problem here is the selection of the reference. If it does not exist yet, it should be constructed. Then, the comparison may be performed using qualitative (e.g. visual analysis) and quantitative

criteria. Since in practice the image does not exist it is necessary to create it from the original PAN image, with high spatial resolution  $h$  ( $PAN_h$ ) and MS image, with low spatial resolution  $l$  ( $MS_l$ ).

For quantitative evaluation we follow the protocol proposed by Thomas and Wald (2005). This protocol allows to assess the quality of any fused product, provided the aim of the user is to handle fused images of higher spatial resolution  $h$  that are as identical as possible to the image  $MS_h$  that the corresponding sensor would observe with the highest spatial resolution  $h$ , if existent. The first to take into account, is that the protocol considers the mono-modality (spectral band) and the multi-modality (multi-spectral) aspects. The mono-modality aspect considers each modality separately, i.e. each fused band is compared to the corresponding reference band and the assessment is done independently. The multi-modality aspect considers the whole set of modalities and assesses whether its multi-modality properties are close to those of the reference ensemble. Likewise, this protocol takes into account two properties that should be checked for individual modality as well as for the multi-modality set:

1. *The consistency property:* any fused image  $FUS_h$  once degraded to its original resolution  $l$ , should be as identical as possible to the original image ( $MS_l$ ).
2. *The synthesis property:* any synthetic image  $FUS_h$  should be as identical as possible to the image  $MS_h$  that the corresponding sensor would observe with the highest spatial resolution  $h$ , if existent.

The protocol comprises five operations. See Fig. 4.

*First operation.* Perform the fusion process on the data sets  $PAN_h$  and  $MS_l$ , obtaining a new image,  $FUS_h$ .

*Second operation.* Resample the set  $FUS_h$  down to spatial resolution  $l$ . Check the consistency property by comparing  $MS_l$  and  $FUS_l$ .

*Third operation.* Resample the data sets  $PAN_h$  and  $MS_l$  down to respectively  $PAN_l$  and  $MS_v$ , where  $v = l/h$ .

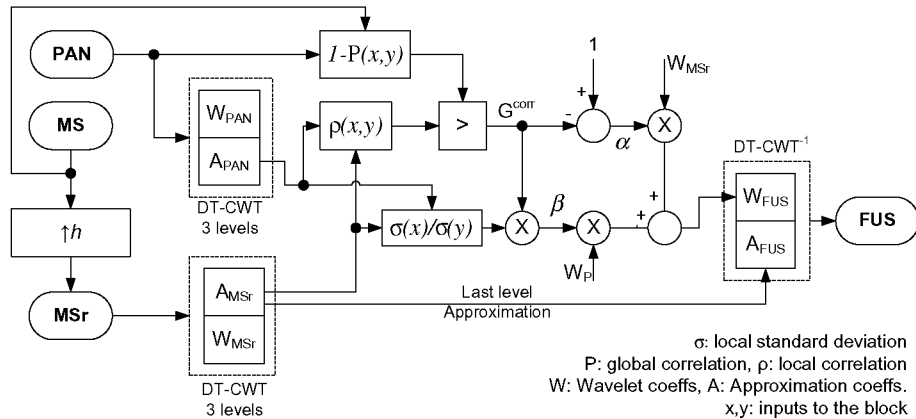


Fig. 3. Image fusion with DT-CWT. Full outline.

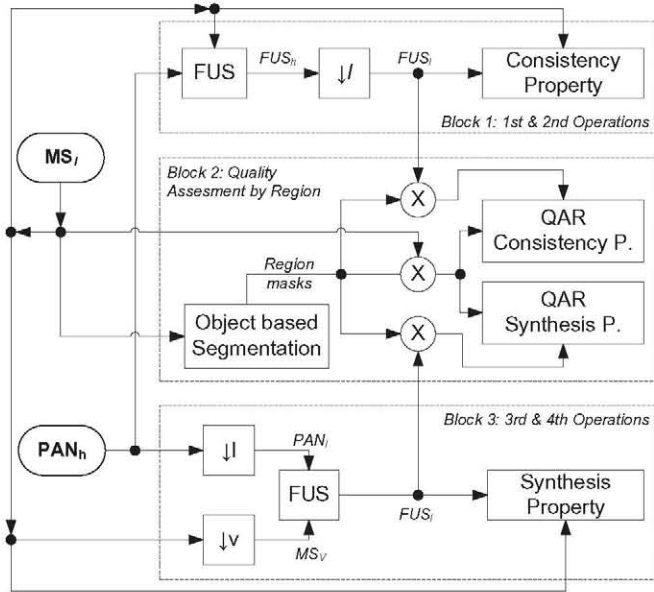


Fig. 4. Quality assessment for image fusion.

Perform the fusion process on these new data sets, obtaining an image  $FUS_i$ .

*Fourth operation.* Check the synthesis property by comparing  $MS_i$  and  $FUS_i$ .

*Fifth operation.* Assuming that the synthesis quality assessed with  $FUS_i$  is equivalent to that of  $FUS_h$ , completes the quality assessment.

For assessment of the fusion process results we use the following metrics (Alparone et al., 2007; Thomas and Wald, 2007; Wald, 2002):

*Multi-modality:* (1) relative dimensionless global error (ERGAS), used to estimate the overall spectral quality of fused images. (2) Spectral angle mapper (SAM), which determines the degree of spectral similarity of an image against a known or reference image, expressed in terms of the average angle between the two spectra.

*Mono-modality:* (3) Variance difference (VD) for estimating the change in variance during the enhancement of the spatial resolution. (4) Standard deviation difference (SDD), which provides a global indication of the level of error at any pixel. (5) Correlation coefficient (CC), which shows the similarity in small size structures between the original and fused images.

## 4. Results

### 4.1. Datasets

MS and PAN Spot 5 images were used in this study. The images have a spatial resolution of 10 and 2.5 m, respectively. Two sets, 26.2 km<sup>2</sup> in area, was extracted from scene 033-268 acquired on June 28, 2005 by the Spot Image corporation. The scene includes partially degraded natural mediterranean forest, vegetation crop areas and urban areas. The two sets of images are displayed in Figs. 5 and 6.

MS SPOT 5 image includes the following bands, B1: green, B2: red, B3: NIR (near infrared) and B4: SWIR (short-wave infrared). In this work we used only the first 3 bands since the SWIR band has lower resolution than the other three. Because the blue band is absent, it is common to display SPOT images in false color. A typical false-color composition put NIR band in the red channel, making vegetated areas appear red; red band will be in the green channel and green band in the blue channel. So, targets with high near-infrared reflectance appear red, those with a high red reflectance appear green, and those with a high green reflectance appear blue, thus giving us a “false” presentation of the targets relative to the color we normally perceive them to be.

The idea was to compare the scheme shown in this paper with alternative fusion schemes aimed to minimizing spectral distortion; these fusion approaches were applied to bands 1–3 (green, red, NIR) and PAN band. The used fusion-approaches were: (i) One classical fusion technique, standard IHS merger (Gonzalez-Audicana et al., 2004). (ii) DT-CWT basic model (DT-B), i.e.  $\alpha = 0$  and  $\beta = 1$  (Ioannidou and Karathanassi, 2007). (iii) DT-CWT basic model with previous histogram matching (HM) between PAN and each of the MS bands (DT-HM). (iv) DT-CWT modified injection (DT-MI), i.e.  $\alpha$  and  $\beta$  defined as in Eqs. (11) and (12). DT-CWT based models were applied with three decomposition levels.

### 4.2. Visual analysis

An apparent improvement in spatial quality compared to the original PAN image is perceptible in the resulting fused images shown in Figs. 7 and 8. Visually, it is possible to see slight variations between the resulting images, but in general all algorithms meet the requirement of improving the spatial quality. Display in false-color allows visual evaluation of the spectral quality. For example, when the fusion process kept properly spectral information, vegetated areas in fused images will take a similar red tone of the original image; otherwise it will degrade into lighter

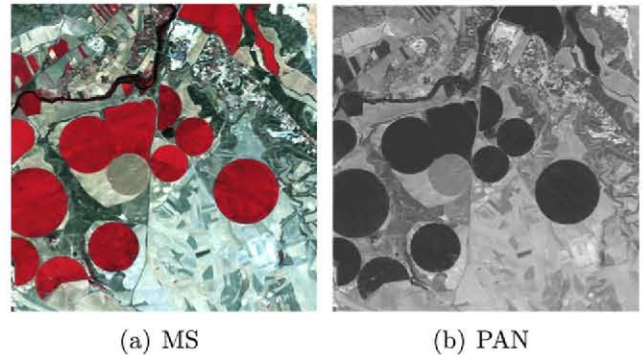


Fig. 5. Image Set 1. (a) MS (512 × 512 pixels false color NIR, red and green composition). (b) PAN (2048 × 2048). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

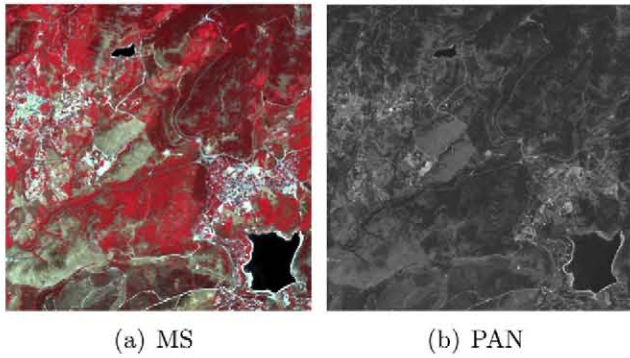


Fig. 6. Image Set 2. (a) MS ( $512 \times 512$  pixels false color NIR, red and green composition). (b) PAN ( $2048 \times 2048$ ). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

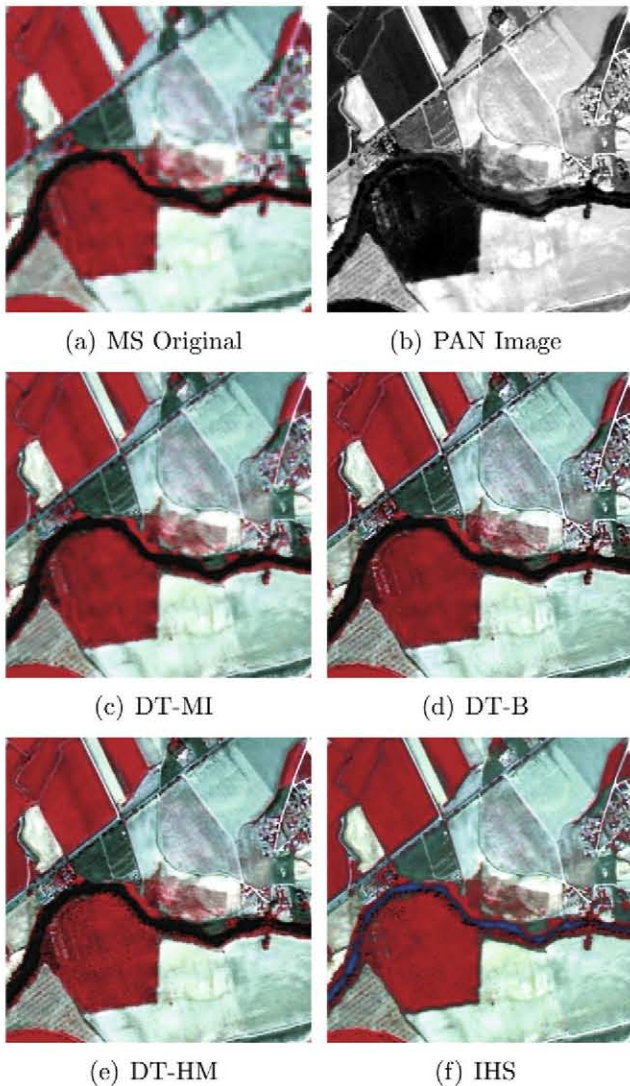


Fig. 7. Image Set 1. A fragment of the original MS (10 m), PAN (2.5 m) and fused images (2.5 m) ( $400 \times 400$  pixels false color NIR, red and green composition). The upper left corner of the full scene is placed at 489330 E and 4444560 N (UTM geographic coordinates, zone 30, WGS-84). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

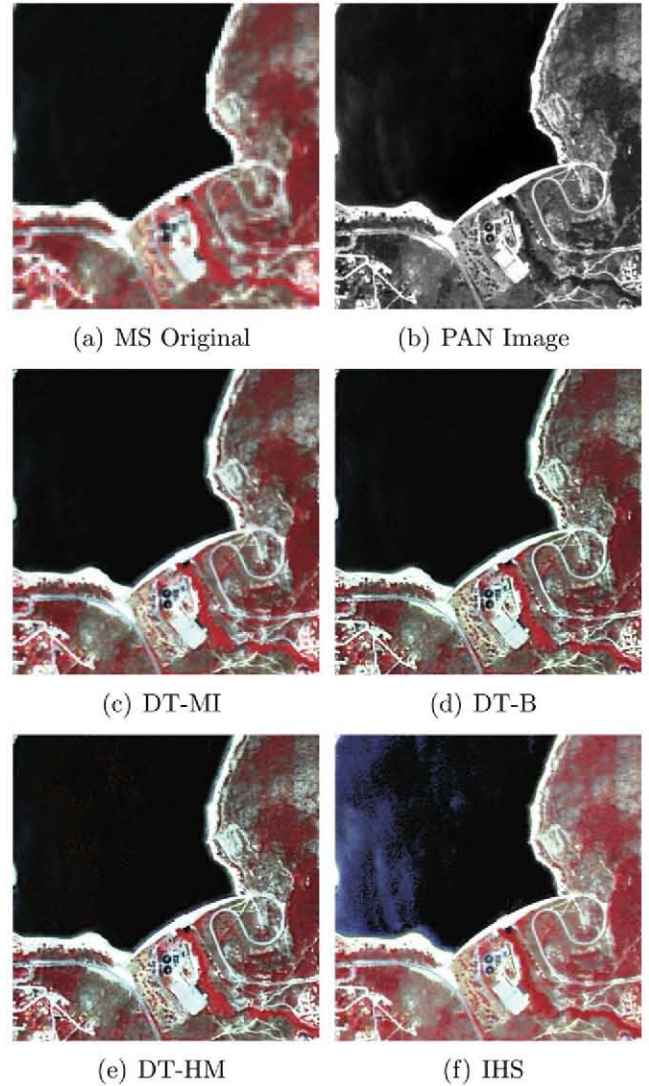


Fig. 8. Image Set 2. A fragment of the original MS (10 m), PAN (2.5 m) and fused images (2.5 m) ( $400 \times 400$  pixels false color NIR, red and green composition). The upper left corner of the full scene is placed at 410525 E and 4512525 N (UTM geographic coordinates, zone 30, WGS-84). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(pink) or darker hues when spectral quality decreases. By comparing the resulting images, is also seen that fused images retain a high degree of spectral fidelity, demonstrated by the red tones of vegetation in these images. However, traditional fusion approaches focused on the preservation of spectral information such as histogram-matching based and IHS-based methods have a significant color distortion, especially noticeable in the river (Fig. 7(e) and (f)) and in water reservoir (Fig. 8(e) and (f)). Therefore, visually the best spectral results were obtained with DT-B and DT-MI schemes.

Likewise, visual inspection of the two cases also guess a good performance for fusion algorithms based on DT-CWT, as well as a trade-off between spatial and spectral quality of the schemes DT-B (Figs. 7(d) and 8(d)) and DT-MI (Figs. 7(c) and 8(c)).

### 4.3. Consistency property (first and second operation)

According to the above, the source  $PAN_h$  and  $MS_l$  images are fused to obtain a new multi-spectral fused image with spatial resolution  $h$ ,  $FUS_h$ . To evaluate the consistency property,  $FUS_h$  is resampled down to spatial resolution  $l$  obtaining  $FUS_l$ . Then the quality can be assessed comparing  $FUS_l$  against the source, MS image ( $MS_l$ ). See block 1 in Fig. 4. The results are showed in Tables 1 and 3.

### 4.4. Synthesis property (third and fourth operation)

Although verification of the property of consistency shows results that support the quality of the methods, it should be noted that verification of this property is a necessary but not sufficient condition, therefore to verify compliance with the second property is required. To check the synthesis property, PAN image is degraded to the low spatial resolution  $l$  and the MS image is degraded to the spatial resolution  $l(l/h)$ . The fusion process is applied to new degraded images and the quality can be assessed using as a reference image the original MS image. See block 3 in Fig. 4. The results are showed in Tables 2 and 4.

### 4.5. Analysis of results

The ERGAS values in Tables 1–4 were acceptable ( $<3$ ) (Wald, 2002) for DT-MI case in both images. If we combine these results with the angles obtained by the SAM metric, preservation of spectral information is confirmed, taking into account that this parameter is calculated for three bands (NIR–red–green). Indeed, these metrics corroborate the observed spectral distortions for the histogram-matching-based and IHS-based approaches (Figs. 7(e) & (f) and 8(e) & (f)), being the worst the IHS-based approach, according to visual assessment. Similarly, one can conclude that the best spectral results are obtained for the scheme based on DT-CWT with modified injection (DT-MI).

VD, an indicator of spatial quality, suggests a slight disadvantage for modified injection method (DT-MI) compared to the direct-injection method (DT-B), which should result in a slightly higher spatial-information preservation. This index confirms the visual difference between Figs. 7(c) & (d) and 8(c) & (d) where you can see that the DT-MI images, Figs. 7(c) and 8(c) are slightly blurred compared with Figs. 7(d) and 8(d), so the above results imply a trade-off between spatial and spectral quality.

SDD shows similar results for the four schemes, with acceptable values (nearly 0). In this case, the best results are obtained for the modified injection model (DT-MI).

As for the overall correlation between the reference image and the merged image, again the best results are obtained with the DT-MI fusion scheme. One thing to highlight is the CC value for the NIR band, which, as highlighted above, falls outside the scope of the PAN. In particular, DT-B, DT-HM and IHS schemes have a CC value for the NIR band much lower than the corresponding value to R and G bands while in DT-MI scheme, CC values for all three bands are similar.

By comparing the outcomes of the consistency property with the results of the synthesis property, one can see that these results have been consistent with each other, i.e. the above analysis can be inferred both based on the consistency tables as in synthesis tables, confirming that the synthesis quality assessed with  $FUS_l$  is equivalent to that of  $FUS_h$ , i.e. the fifth operation. However, in the outcomes presented in Figs. 7 and 8 you can see significant spectral distortions particularly noticeable in the fusion method based on IHS and especially in regions such as bodies of water.

The spectral distortion for the IHS method can be corroborated with consistency and synthesis tables (Tables 1–4), particularly with ERGAS index. However, these differences are more apparent in consistency tables, for example in the case of Image Set 1 show a ratio of 1:2 for ERGAS index between DT-MI and IHS-based methods. Anyway, the big difference in spectral quality dis-

Table 1  
Image Set 1. Consistency property metrics for fused images showed in Fig. 7. The full scenes have a size of  $2048 \times 2048$  pixels. Bands: NIR (N), red (R) and green (G).

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
ERGAS ( $<3$ )		1.0524	0.8973	0.8127	1.5859
SAM (0)		0.0111	0.0107	0.0117	0.0038
VD (0)	N	-0.0368	0.0125	0.0093	0.2704
	R	-0.0068	-0.0032	-0.0032	0.0208
	G	0.0060	0.0032	0.0034	0.0106
SDD (0)	N	0.0424	0.0254	0.0075	0.0614
	R	0.0474	0.0437	0.0426	0.0634
	G	0.0357	0.0362	0.0360	0.0652
CC (1)	N	0.9743	0.9905	0.9992	0.9478
	R	0.9938	0.9947	0.9950	0.9889
	G	0.9942	0.9940	0.9941	0.9805



Table 2

Image Set 1. Synthesis property metrics for fused images showed in Fig. 7. The full scenes have a size of  $2048 \times 2048$  pixels. Bands: NIR (N), red (R) and green (G).

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
ERGAS (<3)		1.9579	1.6164	1.4000	1.8703
SAM (0)		0.0313	0.0261	0.0244	0.0225
VD (0)	N	-0.1920	0.1319	0.1483	0.3083
	R	-0.0109	0.0151	0.0186	0.0793
	G	0.0178	0.0193	0.0332	0.0688
SDD (0)	N	0.1010	0.0706	0.0575	0.0784
	R	0.0711	0.0663	0.0583	0.0756
	G	0.0561	0.0562	0.0520	0.0700
CC (1)	N	0.8670	0.9241	0.9508	0.9086
	R	0.9861	0.9878	0.9906	0.9844
	G	0.9856	0.9855	0.9876	0.9775

Table 3

Image Set 2. Consistency property metrics for fused images showed in Fig. 8. The full scenes have a size of  $2048 \times 2048$  pixels. Bands: NIR (N), red (R) and green (G).

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
ERGAS (<3)		2.4246	2.2051	1.9543	3.0999
SAM (0)		0.0284	0.0302	0.0303	0.0104
VD (0)	N	-0.0252	0.0124	0.0245	0.4060
	R	0.0113	0.0088	0.0030	0.0748
	G	-0.0566	0.0285	0.0171	0.0259
SDD (0)	N	0.0803	0.0703	0.0152	0.1174
	R	0.1074	0.1084	0.1072	0.1260
	G	0.1011	0.0815	0.0812	0.1278
CC (1)	N	0.9357	0.9498	0.9977	0.8534
	R	0.9536	0.9528	0.9539	0.9347
	G	0.9171	0.9437	0.9443	0.8616

Table 4

Image Set 2. Synthesis property metrics for fused images showed in Fig. 8. The full scenes have a size of  $2048 \times 2048$  pixels. Bands: NIR (N), red (R) and green (G).

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
ERGAS (<3)		3.9761	3.5508	2.7871	3.7376
SAM (0)		0.0631	0.0654	0.0500	0.0570
VD (0)	N	-0.1054	0.0402	0.1834	0.4008
	R	0.0729	0.1090	0.1220	0.2680
	G	-0.2251	0.1518	0.1588	0.2170
SDD (0)	N	0.1712	0.1512	0.0831	0.1512
	R	0.1587	0.1559	0.1392	0.1596
	G	0.1463	0.1155	0.1049	0.1366
CC (1)	N	0.7197	0.7644	0.9279	0.7347
	R	0.8961	0.8986	0.9194	0.8931
	G	0.8433	0.8821	0.9032	0.8311

played by the visual results can not be captured totally by the metrics showed in these tables. The reason is that we evaluate the whole image, therefore, major differences in

a given region can be minimized by small differences in another region. Accordingly, we propose an optional step in the quality assessment protocol, which is to study the

quality of the merger by regions, where each region represents a specific feature of the image. This will be the subject of the next section.

## 5. Quality assessment by region (QAR)

In order to apply more localized quality indices, is proposed in this paper divide the image into regions that are representative of a specific feature of the image, i.e. the quality indices can be applied individually to a segment that contains only one coverage such as vegetation. So, you can apply the above quality indices but for each of the regions defined in the image, thus facilitating the analysis of the impact of fusion methods in different coverages.

With this goal, the first step was to segment the original MS image in a small number of representative segments of the existing land covers. For each segment binary masks are extracted, which will be applied to both the reference image (source MS) and the resulting fused images. See block 2 in Fig. 4. After extracting the image regions corresponding to each segment, we proceed to apply the quality indexes.

The segmentation algorithm used was described in our previous work (Sanchez et al., 2010). The method involves obtaining the objects by means an unsupervised method based on a self-calibrating multi-band region growing approach. An adaptive Canny edge detection algorithm is applied to obtain a calibration map. Next, the starting points of the segmentation, often referred to as seed pixels, have to be identified. The seeds are selected according to a scan of the image by rows without taking into account the edge information. The regions are built around these pixels by joining the similar neighboring pixels to them. To compute the similarity the Euclidean distance is used. Once the regions have been obtained, a merging process is performed in order to eliminate small regions. The automatically generated map prior to segmentation, is a calibration edge map that can be used to obtain the optimal region maps by means a disparity measure. To obtain the disparity measure, the maps of distances associated with edges of each region maps from the segmentation process are compared with the map of distances from the calibration edge map. Finally, the obtained segments are classified and labeled by a minimum distance supervised method.

### 5.1. QAR in Image Set 1

After applying the segmentation process on the Image Set 1 (Fig. 5(a)), we obtained the results shown in Fig. 9, which shows the segmented image and masks for each of the segments associated with the land cover types.

The masks for the six segments allowed to extract regions of interest in the original MS image and the fused images, thus applying the quality indices only in pixels associated with the respective segment. In this way we obtained the results shown in Tables 5 and 6. The Table 5 shows the results of the property of consistency, while the Table 6 shows the fulfillment of the synthesis property.

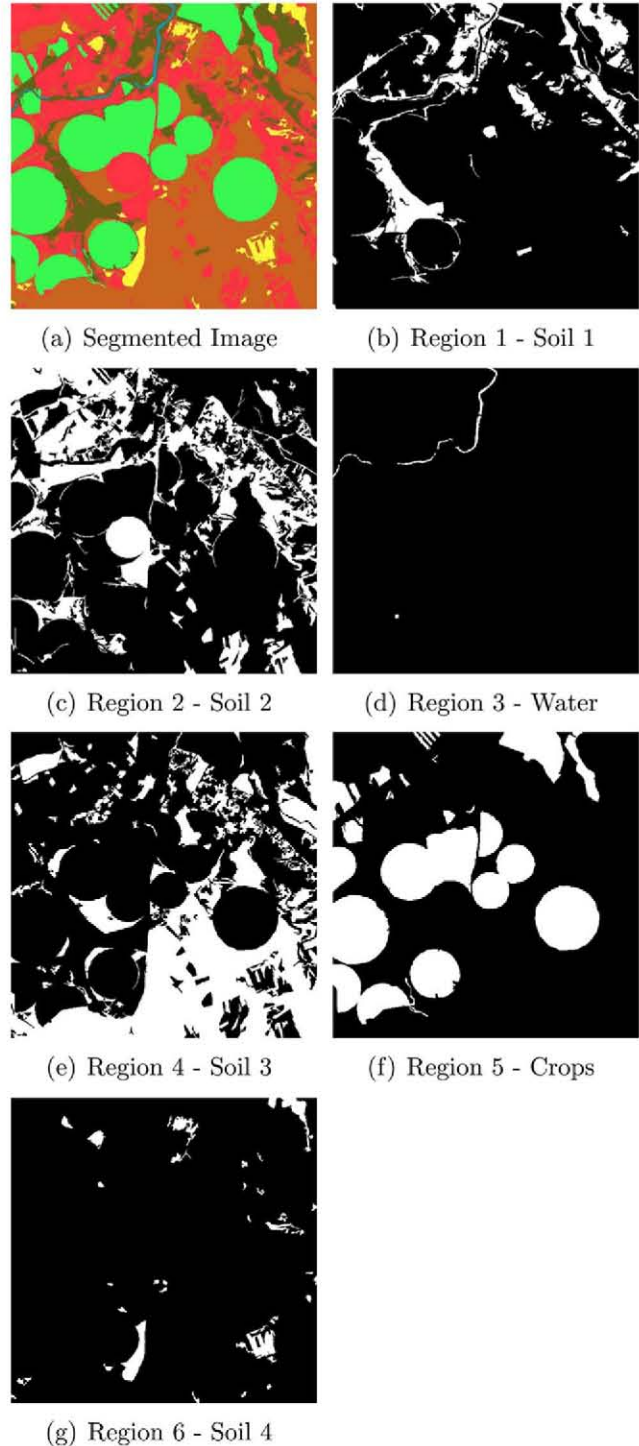


Fig. 9. Image Set 1. Segmented image and regions.

The tables show both individual modality as multi-modality set.

### 5.2. QAR in Image Set 2

Quality assessment by region in Image Set 2 (Fig. 6(a)) began with the segmentation process, resulting in the segmented image and masks showed in Fig. 10. The land cover

Table 5  
Image Set 1. Consistency property metrics for regions showed in Fig. 9.

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
<i>Region 1</i>					
ERGAS (<3)		1.6028	1.3778	1.178	2.6142
SAM (0)		0.0182	0.0179	0.0186	0.0056
VD (0)	N	-0.0863	0.0046	0.0188	0.2866
	R	-0.1025	-0.0903	-0.0774	0.1185
	G	-0.0505	-0.0543	-0.0571	0.1986
SDD (0)	N	0.0718	0.0482	0.0132	0.0835
	R	0.0684	0.0651	0.0624	0.0786
	G	0.0499	0.0503	0.0508	0.0808
CC (1)	N	0.9077	0.9560	0.9967	0.8586
	R	0.9391	0.9442	0.9483	0.9103
	G	0.9401	0.9394	0.9384	0.8256
<i>Region 2</i>					
ERGAS (<3)		1.3209	1.0586	0.9659	1.6911
SAM (0)		0.0133	0.0123	0.0153	0.0040
VD (0)	N	-0.1999	0.0002	0.0162	0.2054
	R	-0.1087	-0.0756	-0.0679	0.0221
	G	-0.0305	-0.0327	-0.0404	-0.0951
SDD (0)	N	0.0583	0.0287	0.0091	0.0683
	R	0.0558	0.0515	0.0497	0.0663
	G	0.0432	0.0436	0.0438	0.0671
CC (1)	N	0.9088	0.9747	0.9974	0.8455
	R	0.9294	0.9385	0.9423	0.8918
	G	0.9336	0.9324	0.9321	0.8455
<i>Region 3</i>					
ERGAS (<3)		3.4231	3.6589	1.4767	14.0082
SAM (0)		0.0572	0.0621	0.0228	0.0157
VD (0)	N	-0.1218	-0.2020	-0.0039	-0.3058
	R	-0.2441	-0.2532	-0.2426	-0.7757
	G	-0.0097	-0.0231	-0.0080	-3.5945
SDD (0)	N	0.1901	0.2029	0.0426	0.1316
	R	0.0707	0.0730	0.0766	0.1315
	G	0.0401	0.0412	0.0413	0.1470
CC (1)	N	0.8895	0.8807	0.9940	0.9591
	R	0.8303	0.8199	0.7997	0.5332
	G	0.8837	0.8777	0.8766	0.5723
<i>Region 4</i>					
ERGAS (<3)		0.8128	0.6455	0.5901	1.1235
SAM (0)		0.0077	0.0072	0.0088	0.0026
VD (0)	N	-0.1436	0.0035	0.0120	0.1077
	R	-0.0846	-0.0574	-0.0599	-0.0451
	G	-0.0236	-0.0378	-0.0301	-0.1645
SDD (0)	N	0.0368	0.0186	0.0070	0.0451
	R	0.0332	0.0298	0.0298	0.0435
	G	0.0266	0.0276	0.0270	0.0440
CC (1)	N	0.9299	0.9802	0.9972	0.8787
	R	0.9505	0.9594	0.9594	0.9122
	G	0.9639	0.9614	0.9628	0.9102

Table 5 (continued)

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
<i>Region 5</i>					
ERGAS (<3)		0.8588	0.8684	0.7833	1.4702
SAM (0)		0.0092	0.0100	0.0083	0.0039
VD (0)	N	-0.0485	-0.0149	0.0127	0.4930
	R	-0.1081	-0.1283	-0.0985	-0.1353
	G	-0.0878	-0.0968	-0.0883	-0.2840
SDD (0)	N	0.0214	0.0204	0.0038	0.0490
	R	0.0484	0.0493	0.0467	0.0564
	G	0.0271	0.0277	0.0274	0.0515
CC (1)	N	0.9541	0.9575	0.9985	0.7091
	R	0.9572	0.9564	0.9599	0.9429
	G	0.9622	0.9609	0.9615	0.8797
<i>Region 6</i>					
ERGAS (<3)		1.0477	0.8310	0.7105	1.2932
SAM (0)		0.0109	0.0088	0.0112	0.0048
VD (0)	N	-0.2158	-0.0571	0.0173	-0.0004
	R	-0.0993	-0.0339	-0.0611	-0.0124
	G	0.0357	0.0005	0.0100	-0.0022
SDD (0)	N	0.0489	0.0303	0.0100	0.0512
	R	0.0422	0.0376	0.0370	0.0510
	G	0.0321	0.0307	0.0303	0.0470
CC (1)	N	0.9103	0.9615	0.9957	0.8854
	R	0.9390	0.9492	0.9520	0.9058
	G	0.9310	0.9378	0.9390	0.8545

Table 6  
Image Set 1. Synthesis property metrics for regions showed in Fig. 9.

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
<i>Region 1</i>					
ERGAS (<3)		2.9259	2.5002	2.1964	3.3144
SAM (0)		0.0475	0.0403	0.0396	0.035
VD (0)	N	-0.486	0.0629	0.0983	0.3673
	R	-0.2146	-0.1264	-0.0555	0.2777
	G	-0.0774	-0.0296	0.0498	0.3482
SDD (0)	N	0.1587	0.1223	0.1012	0.1131
	R	0.0987	0.0928	0.0829	0.0941
	G	0.0756	0.0741	0.0706	0.0878
CC (1)	N	0.6297	0.7089	0.7977	0.7229
	R	0.8814	0.8885	0.9071	0.8678
	G	0.8645	0.8661	0.8737	0.7865
<i>Region 2</i>					
ERGAS (<3)		2.3284	1.7722	1.6124	1.9010
SAM (0)		0.0346	0.0271	0.0281	0.0228
VD (0)	N	-0.9596	0.0461	0.1671	0.1855
	R	-0.2614	-0.0636	-0.0118	0.1899
	G	-0.0686	0.0216	0.0874	0.0818
SDD (0)	N	0.1218	0.0729	0.0668	0.0798
	R	0.0819	0.0737	0.0652	0.0751
	G	0.0664	0.0645	0.0604	0.0711

(continued on next page)

Table 6 (continued)

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
CC (1)	N	0.7313	0.8330	0.8536	0.7879
	R	0.8613	0.8724	0.8973	0.8530
	G	0.8464	0.8479	0.8630	0.8102
<i>Region 3</i>					
ERGAS (<3)		12.1761	14.3279	10.0491	19.3044
SAM (0)		0.2216	0.2570	0.1845	0.1826
VD (0)	N	0.2948	0.2805	0.0679	0.2297
	R	-0.6163	-1.5941	-0.7570	-0.7389
	G	-0.0031	-0.3437	-0.0283	-1.4923
SDD (0)	N	0.4352	0.4067	0.3460	0.3822
	R	0.1012	0.1349	0.1026	0.1130
	G	0.0620	0.0638	0.0626	0.1077
CC (1)	N	0.2738	0.3723	0.5930	0.4612
	R	0.7135	0.6731	0.7293	0.6596
	G	0.7213	0.7559	0.7191	0.5713
<i>Region 4</i>					
ERGAS (<3)		1.4804	1.1608	1.0974	1.2703
SAM (0)		0.0206	0.0177	0.0187	0.0132
VD (0)	N	-0.6227	0.1288	0.1268	0.1332
	R	-0.2170	-0.1139	-0.0615	-0.0254
	G	-0.0666	-0.1249	-0.0138	-0.1311
SDD (0)	N	0.0772	0.0483	0.0490	0.0513
	R	0.0507	0.0462	0.0410	0.0475
	G	0.0426	0.0440	0.0395	0.0461
CC (1)	N	0.7629	0.8599	0.8557	0.8414
	R	0.8942	0.9055	0.9231	0.8945
	G	0.9094	0.9075	0.9199	0.8986
<i>Region 5</i>					
ERGAS (<3)		1.8898	1.8857	1.4173	2.1348
SAM (0)		0.0321	0.0268	0.0185	0.0273
VD (0)	N	-1.2536	-0.4468	0.0804	0.0642
	R	-0.3611	-0.6989	-0.3384	-0.5885
	G	-0.2608	-0.4752	-0.2675	-0.5379
SDD (0)	N	0.0855	0.0651	0.0320	0.0685
	R	0.0829	0.0981	0.0787	0.1083
	G	0.0484	0.0538	0.0476	0.0666
CC (1)	N	0.5804	0.6531	0.8905	0.4985
	R	0.8952	0.8890	0.9046	0.8424
	G	0.8923	0.8885	0.8966	0.8276
<i>Region 6</i>					
ERGAS (<3)		1.7603	1.3563	1.4320	1.5063
SAM (0)		0.0261	0.0180	0.0248	0.0149
VD (0)	N	-0.7159	0.1016	0.1933	0.2243
	R	-0.2171	0.0361	0.0411	0.2081
	G	-0.1293	-0.0782	-0.0164	0.0543
SDD (0)	N	0.0878	0.0547	0.0651	0.0550
	R	0.0634	0.0549	0.0505	0.0553
	G	0.0491	0.0476	0.0447	0.0503
CC (1)	N	0.7796	0.8637	0.8000	0.8580
	R	0.8718	0.8879	0.9053	0.8813
	G	0.8519	0.8568	0.8692	0.8288

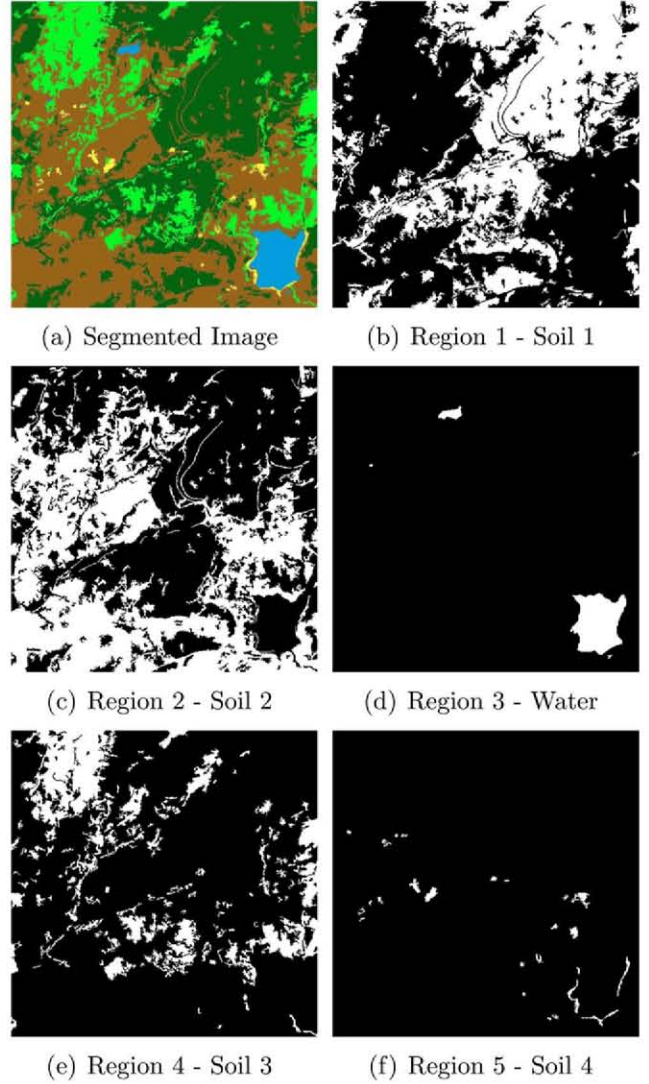


Fig. 10. Image Set 2. Segmented image and regions.

types associated with the five segments are displayed in the same image. With these masks, we extracted regions of interest for the original MS image and the fused images. The quality indices were applied only in pixels corresponding to each segment. The obtained results for consistency and synthesis are shown in Tables 7 and 8, respectively, showing both individual modality as multi-modality set.

### 5.2.1. Analysis of results for QAR

Analyzing the results by region, is of great interest the region 3, both for the set of images 1 as 2. These two cases involve bodies of water, corresponding to a river in the Image Set 1 and a reservoir in Image Set 2. Looking at the mask of Class 3 in Image Set 1, we see that it contains mainly thin features (river), features that may be affected by the spatial degradation prior to fusion in synthesis property evaluation, degradation applied to the PAN and MS images, i.e. as this operation involves a low-pass filtering may affect the evaluation of the spectral quality of the resulting images. This difference can be seen clearly

Table 7  
Image Set 2. Consistency property metrics for regions showed in Fig. 10.

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
<i>Region 1</i>					
ERGAS (<3)		2.0397	2.0089	1.6903	2.7454
SAM (0)		0.0265	0.0289	0.0282	0.0098
VD (0)	N	-0.0101	-0.0243	0.033	0.4687
	R	-0.0832	-0.1288	-0.1028	-0.2721
	G	-0.311	-0.1282	-0.1173	-0.638
SDD (0)	N	0.0637	0.0661	0.0126	0.0991
	R	0.0958	0.1027	0.0970	0.1074
	G	0.0817	0.0662	0.0640	0.1042
CC (1)	N	0.9093	0.9030	0.9965	0.7473
	R	0.8862	0.8727	0.8848	0.8742
	G	0.7956	0.8507	0.8596	0.7201
<i>Region 2</i>					
ERGAS (<3)		2.5662	2.2117	1.9538	2.7123
SAM (0)		0.0301	0.0303	0.0329	0.0108
VD (0)	N	-0.2324	-0.0458	0.0523	0.3663
	R	-0.0355	-0.0246	-0.0463	0.1964
	G	-0.1537	0.0252	-0.0020	0.1691
SDD (0)	N	0.0969	0.0752	0.0167	0.1043
	R	0.1037	0.1021	0.1029	0.1101
	G	0.1068	0.0855	0.0861	0.1109
CC (1)	N	0.8142	0.8750	0.9939	0.7166
	R	0.8867	0.8893	0.8891	0.8606
	G	0.8334	0.8820	0.8819	0.7894
<i>Region 3</i>					
ERGAS (<3)		3.0565	3.4535	2.2522	11.5368
SAM (0)		0.0176	0.0355	0.0145	0.0077
VD (0)	N	-0.4613	-0.7058	-0.1364	-2.2761
	R	-0.2808	-0.2847	-0.2792	-3.4765
	G	-0.2494	-0.2077	-0.2028	-7.3220
SDD (0)	N	0.1636	0.2004	0.0856	0.3445
	R	0.1096	0.1088	0.1117	0.3438
	G	0.0764	0.0649	0.0662	0.3230
CC (1)	N	0.8804	0.8447	0.9593	0.7737
	R	0.9117	0.9135	0.9080	0.7896
	G	0.8810	0.9125	0.9085	0.7490
<i>Region 4</i>					
ERGAS (<3)		2.4645	2.3832	2.1288	3.4574
SAM (0)		0.0310	0.0324	0.0314	0.0112
VD (0)	N	-0.0179	-0.0088	0.0539	0.4909
	R	-0.1572	-0.2081	-0.1771	-0.2465
	G	-0.3804	-0.1506	-0.1455	-0.5792

Table 7 (continued)

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
SDD (0)	N	0.0618	0.0609	0.0124	0.1014
	R	0.1223	0.1290	0.1230	0.1100
	G	0.1006	0.0816	0.0794	0.1043
CC (1)	N	0.8678	0.8709	0.9948	0.5538
	R	0.8349	0.8218	0.8349	0.8752
	G	0.7440	0.8085	0.8182	0.7557
<i>Region 5</i>					
ERGAS (<3)		2.5633	2.0238	1.9263	3.8701
SAM (0)		0.0248	0.0216	0.0300	0.0075
VD (0)	N	-0.5704	-0.1475	-0.0485	-0.0520
	R	-0.1938	-0.1448	-0.2258	0.0244
	G	-0.2261	-0.0265	-0.0981	-0.0098
SDD (0)	N	0.1112	0.0695	0.0323	0.1163
	R	0.0902	0.0845	0.0923	0.1152
	G	0.1024	0.0833	0.0875	0.1200
CC (1)	N	0.7918	0.8957	0.9762	0.6880
	R	0.8633	0.8765	0.8601	0.7464
	G	0.8229	0.8682	0.8603	0.7240

Table 8  
Image Set 2. Synthesis property metrics for regions showed in Fig. 10.

Index (ideal value)		Fusion approach			
		DT-B	DT-HM	DT-MI	IHS
<i>Region 1</i>					
ERGAS (<3)		3.3079	3.2232	2.4099	3.4584
SAM (0)		0.0544	0.0586	0.0448	0.0515
VD (0)	N	-0.1247	-0.1745	0.1991	0.3988
	R	-0.112	-0.2362	-0.0333	-0.112
	G	-0.7769	-0.2257	-0.0531	-0.5385
SDD (0)	N	0.1375	0.1395	0.0698	0.1327
	R	0.1387	0.1456	0.1258	0.1396
	G	0.1177	0.0923	0.0816	0.1125
CC (1)	N	0.6010	0.5991	0.8837	0.5221
	R	0.7641	0.7577	0.7978	0.7613
	G	0.6607	0.7232	0.7626	0.6494
<i>Region 2</i>					
ERGAS (<3)		4.1773	3.4553	2.7779	3.3470
SAM (0)		0.0678	0.0652	0.0532	0.0590
VD (0)	N	-0.9631	-0.2085	0.2897	0.3420
	R	-0.0548	0.0977	0.1114	0.4334
	G	-0.4988	0.1900	0.1999	0.4279
SDD (0)	N	0.1879	0.1449	0.0886	0.1326
	R	0.1507	0.1430	0.1295	0.1405
	G	0.1537	0.1185	0.1089	0.1236
CC (1)	N	0.4874	0.5722	0.8037	0.5313
	R	0.7625	0.7700	0.8106	0.7588
	G	0.7133	0.7568	0.7952	0.7174

(continued on next page)

Table 8 (continued)

Index (ideal value)	Fusion approach				
	DT-B	DT-HM	DT-MI	IHS	
<i>Region 3</i>					
ERGAS (<3)	6.1566	9.3408	5.4559	11.4491	
SAM (0)	0.0451	0.1013	0.0417	0.0335	
VD (0)	N	-2.6930	-5.9987	-2.0385	-4.0109
	R	-0.5516	-0.7277	-0.5819	-3.7185
	G	-0.4633	-0.4267	-0.3634	-7.0872
SDD (0)	N	0.3626	0.6048	0.3105	0.4605
	R	0.1703	0.1803	0.1684	0.3647
	G	0.1182	0.0987	0.0956	0.3227
CC (1)	N	0.7956	0.6521	0.8145	0.7535
	R	0.8138	0.8142	0.8227	0.7631
	G	0.7414	0.8203	0.8244	0.7200
<i>Region 4</i>					
ERGAS (<3)	4.1223	3.9460	3.0592	4.0399	
SAM (0)	0.0748	0.0768	0.0559	0.0698	
VD (0)	N	-0.2591	-0.1656	0.3379	0.3804
	R	-0.2678	-0.4377	-0.1359	-0.1514
	G	-0.9781	-0.3102	-0.0741	-0.4725
SDD (0)	N	0.1456	0.1412	0.0751	0.1334
	R	0.1786	0.1864	0.1607	0.1642
	G	0.1455	0.1153	0.1024	0.1203
CC (1)	N	0.3461	0.3572	0.7790	0.2384
	R	0.6652	0.6667	0.7097	0.6991
	G	0.5890	0.6463	0.6848	0.6461
<i>Region 5</i>					
ERGAS (<3)	4.6489	3.7729	3.3884	4.7074	
SAM (0)	0.0723	0.0633	0.0526	0.0478	
VD (0)	N	-1.9696	-0.2004	0.1855	0.1239
	R	-0.5059	-0.0862	-0.2363	0.2995
	G	-0.8433	0.0942	0.0128	0.3133
SDD (0)	N	0.2082	0.1368	0.1108	0.1260
	R	0.1471	0.1288	0.1287	0.1173
	G	0.1677	0.1264	0.1214	0.1207
CC (1)	N	0.5557	0.5995	0.6830	0.6005
	R	0.6882	0.7006	0.7245	0.7058
	G	0.6484	0.6781	0.7142	0.6790

between consistency and synthesis results obtained for region 3.

Taking the region 3 in the Image Set 1, the DT-MI and IHS-based fusion methods and reviewing ERGAS index is observed that in the metrics for synthesis property evaluation, there is a ratio of 1:2, while in the metrics of consistency, the ratio is close to 1:10. The latter describes a more significant spectral differences in the river shown in Fig. 7 for the two methods mentioned.

As for the region 3 of Image Set 2, significant distortions are seen again, particularly for the method based on IHS. In this case the ERGAS index for DT-MI and IHS-based fusion methods show a ratio close to 1:2 in synthesis metrics, while the consistency metrics ratio is close to 1:5.

Again the consistency metrics confirm that they are needed, showing results according with visual inspection and describing the reservoir spectral differences shown in Fig. 8.

## 6. Conclusion

Quality assessment by region was proposed, facilitating the analysis of the impact of fusion methods in different coverages and confirming the need to evaluate the properties of synthesis and consistency. DT-CWT being an over-sampled transform reduce aliasing and shift dependence, so DT-CWT-based fusion method offers good spatial quality while retaining the spectral information of original images, in particular sensors which have spectral bands that fall outside the PAN spectrum, case Spot 5 sensor. Both visual inspection and quality indexes were used, to confirm the good overall results for the DT-CWT-based fusion schemes and the trade-off between spectral quality and spatial quality.

This finding is supported by a case study with typical Spot 5 images (MS and PAN). Four approaches were used for the injection of PAN image details into the MS image, with the aim of minimize the degradation of the spectral information from original multi-spectral image: IHS (standard IHS merger Gonzalez-Audicana et al., 2004), DT-B (DT-CWT basic model Ioannidou and Karathanassi, 2007), DT-HM (DT-B with previous histogram matching between PAN and each of the MS bands), DT-MI (proposed method defining an injection model). Regarding spatial quality of the fused images, the list from best to worst is as follows: DT-MI, DT-B, DT-HM and IHS. To note is the superior ability of injecting high frequency components taken from the PAN band into MS resampled bands, and in particular MRA-based mergers. Also, the definition of an injection model can imply a trade-off between spectral quality and spatial quality.

Since the main restriction was to minimize the degradation of the spectral information from original multi-spectral image, it is interesting to compare the spectral quality results. From best to worst the results were: DT-MI, DT-B, DT-HM and IHS. Even the quality metrics are better for DT-HM compared with DT-B, the visual analysis show radiometric distortions for DT-HM. These radiometric distortions are higher for IHS-based fusion specially for some land covers. The superiority in spectral results of DT-MI lies in the injection model, which improves the spectral fidelity for bands with spectral response outside the scope of PAN spectral response.

Future work could use the directional information given by the DT-CWT, while taking into account the computational cost required to process wavelet subbands in six directions for each level. The proposed approach can be applied and evaluated in another areas, such as medical imaging, where lower time image capture and image sizes may allow applications in (near) real time.

## Acknowledgment

The authors would like to thank Dr. Nick Kingsbury (from the University of Cambridge) for the DT-CWT toolbox.

## References

- Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Transactions on Geoscience and Remote Sensing* 40 (10), 2300–2312, 2002.
- Aiazzi, B., Baronti, S., Selva, M. Image fusion through multiresolution oversampled decompositions Chapter 2, 2008, in: *Image Fusion – Algorithms and Applications*. Academic Press, pp. 27–66, 2002.
- Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., Bruce, L. Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data-fusion contest. *IEEE Transactions on Geoscience and Remote Sensing* 45 (10), 3012–3021, 2007.
- Burke, B. *The World According to Wavelets – The Story of a Mathematical Technique in the Making*. second ed.. A.K. Peters Ltd., 1998.
- Celik, T., Ma, K.-K. Unsupervised change detection for satellite images using dual-tree complex wavelet transform. *IEEE Transactions on Geoscience and Remote Sensing* 48 (3), 1199–1210, 2010.
- Garguet-Duport, B., Girel, J., Chassery, J., Pautou, G. The use of multiresolution analysis and wavelets transform for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering and Remote Sensing*, 1996.
- Gonzalez-Audicana, M., Saleta, J., Catalan, R., Garcia, R. Fusion of multispectral and panchromatic images using improved ihs and pca mergers based on wavelet decomposition. *IEEE Transactions on Geoscience and Remote Sensing* 42 (6), 1291–1299, 2004.
- Ioannidou, S., Karathanassi, V. Investigation of the dual-tree complex and shift-invariant discrete wavelet transforms on Quickbird image fusion. *IEEE Geoscience and Remote Sensing Letters* 4 (1), 166–170, 2007.
- Kingsbury, N. Image processing with complex wavelets. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 357 (1760), 2543–2560, 1999.
- Liang, J., Parks, T. A translation-invariant wavelet representation algorithm with applications. *IEEE Transactions on Signal Processing* 44 (2), 225–232, 1996.
- Li, S., Kwok, J.T., Wang, Y. Using the discrete wavelet frame transform to merge landsat tm and spot panchromatic images. *Information Fusion* 3 (1), 17–23, 2002.
- Mallat, S. *A Wavelet Tour of Signal Processing*. third ed.. Academic Press, 2009.
- Nuñez, J., Otazu, X., Fors, O., Prades, A., Pala, V., Arbiol, R. Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Transactions on Geoscience and Remote Sensing* 37 (3), 1204–1211, 1999.
- Otazu, X., Gonzalez-Audicana, M., Fors, O., Nuñez, J. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *IEEE Transactions on Geoscience and Remote Sensing* 43 (10), 2005.
- Sanchez, J., Martinez, E., Arquero, A., Renza, D. Automatic image segmentation optimized by bilateral filtering, in: *CIARP*, pp. 303–310, 2010.
- Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*. third ed.. Academic Press, 2007.
- Selesnick, I., Baraniuk, R., Kingsbury, N. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine* 22 (6), 123–151, 2005.
- Thomas, C., Wald, L. Assessment of the quality of fused products, in: *24th EARSeL Symposium on New Strategies for European Remote Sensing*. Balkema, The Netherlands, pp. 317–325, 2005.
- Thomas, C., Wald, L. Comparing distances for quality assessment of fused images, in: *26th Symposium of the European Association of Remote Sensing Laboratories*. Millpress, The Netherlands, pp. 101–111, 2007.
- Wald, L. Assessing the quality of synthesized images. Definitions and Architectures – Fusion of Images of Different Spatial Resolutions. Presses de l’Ecole, Ecole des Mines de Paris, France, Chapter 8 of *Data Fusion*, pp. 143–163, 2002.
- Wang, Z., Ziou, D., Armenakis, C., Li, D., Li, Q. A comparative analysis of image fusion methods. *IEEE Transactions on Geoscience and Remote Sensing* 43 (6), 1391–1402, 2005.