International Journal of Remote Sensing

Feature space and metric measures for fusing multisensor images

S. H. Chen a, B. Su a, H. Zhang a

a Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Online Publication Date: 01 June 2008


To link to this article: DOI: 10.1080/01431160701644253
URL: http://dx.doi.org/10.1080/01431160701644253

Please scroll down for article

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
A multivalued wavelet transform (MWT) is proposed to fuse multisensor images in feature space. First, feature space is constructed using image-derived features, and then the MWT is introduced. The multisensor images are then fused in the MWT domain using a voting and electing fuser based on the cross-feature scale guideline and the posterior probability of the MWT coefficient. The performance of the MWT is estimated using metric measures regarding various aspects of image quality. A fusion experiment using Thematic Mapper (TM) multispectral and SPOT panchromatic images of south China demonstrates that MWT outperforms smoothing filter-based intensity modulation (SFM) in terms of the fidelity to spectral properties and the injection of salient information. The experimental results confirm that the MWT is a superior fusion method for enhancing spatial quality of multispectral images with their spectral properties reliably preserved.

1. Introduction

The objectives of fusing multisensor images are mainly concerned with compounding not only the photointerpretation (Tapiador and Casanova 2002) and geoapplications but also the explanation of potentialities and the classification information of image sets into one composite image. The image sets do not offer such accuracy on their own and relevant preponderance by themselves, while the composite image is better for post-processing, such as thematic map production, ground cover identification and automatic object tracing. The formal framework has attracted significant research efforts because (i) a tradeoff exists between spatial and spectral resolution and signal-to-noise ratio (SNR) by way of physical and technological constraints (Scheunders 2004), (ii) the fusion method can reduce the complexity of the data-filtered procedure because only a fraction of images provides complementary or useful information (Scheunders 2002), and (iii) several sensing modalities are used under a wide range of operating conditions through the availability of plentiful image data from operational earth observation satellites (Pohl and van Genderen 1998).

Various theories and approaches have been put forward for image fusion. From overviews of the literature, fusion methods can be classified into pixel-, feature- and decision-level methods (Pohl and van Genderen 1998). As far as implementation and calculation methods are concerned, pixel-level methods are arithmetic and contribution methods; for example, weighting average, Brovey transform (Gillespie...
et al. 1987), addition and multiplication (Pohl and van Genderen 1998), smoothing filter-based intensity modulation (SFM) (Liu 2000), and high-pass filtering (Schowengerdt 1980). Pixel-level methods can be used to increase the information content associated with each pixel in an image through multiple image combination (Du et al. 2003) at the cost of engendering undesirable side-effects such as reduced contrast (Li et al. 2001). Most transform and substitution methods can be accommodated as feature-level methods, such as intensity-hue-saturation transform (Chavez and Bowell 1988), principal component analysis (Zhou et al. 1998), pyramid class (Akerman 1992) with its derivatives (Wald 1999), and the wavelet transform family (Mallat 1989) with its descendants (Chibani and Houacine 2002), fuzzy logic (Zhao et al. 2005, Liu et al. 2006), and spatial modelling (Babu Madhavan et al. 2006). Feature-level methods can improve the spatial resolution and preserve the spectral content of the multispectral image while introducing unexpected artefacts and redundant distortions into the result (Li et al. 2004). The group of decision-level methods is generally grounded in statistical methodologies. It comprises, but is not limited to, Dempster–Shafer evidence theory (Shafer 1979), Markov random field model (Solberg et al. 1996), Bayesian (Hauter et al. 1997), artificial neural networks (Li et al. 2002), and support vector machine (Li et al. 2004). These methods combine information to reinforce common interpretations and resolve differences by applying decision rulers at the expense of attending to one thing and losing another in terms of harmonizing the loss of information during the extraction process (Li et al. 2004) and a better understanding of the observed objects (Shen 1990).

Following a thorough investigation into these three groups of fusion techniques, only a singular feature of the input images has been used, namely the grey values. It is well known that there are many other features of images, such as contrast, and if these features along with the original image constitute a multivalued image as input, then conventional image processing and analysis techniques would benefit from the deduce-then-generalize architecture. Quality assessment tools are divided into subjective thematic analysis and objective quantities. Although both of these aim to maintain the image quality and analyse the executive mode of fusion methods, they lose sight of some critical aspects of the image, such as the structure, texture, scale and reliability.

This paper propounds a multivalued wavelet transform (MWT) in feature space in collaboration with a voting and electing fuser for producing the fused image under the guidance of the posterior probability of the wavelet coefficient. Quality certification of the fused image from a group of novel protocols corroborates that this method can perfect the tradeoff preservation (Wang et al. 2005).

2. Generation of fused images

2.1 Feature space

A remote sensing image is the carrier of information by sampling the real valued function of space–time about the observed Earth’s surface. The digital number values of a remote sensing image have various meanings, including fractal geometry (Liu and Li 1997), raggedness of ground surface (Liu 2000), inner specialties (Eskicioglu and Fisher 1995), definition and contrast (Lu and Healy 1994), and edge- and boundary-dependent shape segmentation (Nikolov et al. 2000). They are displayed by the grey values, the abstracted spectral reflectance, statistical elements
(e.g. mean and variance), the mutual relationship between neighbourhood pixels, and grey values of the same object, respectively. In the following text, these statistical attributes of the original image \( I \) are dissected into seven representative features with pseudo-formulae.

1. **Setover**. Setover \( S \) is an important connection between the specific observation of grey-value fluctuation and the usual intensity stability. It balances the total oscillation around the centre by the absolute bias between each grey value and the mean \( \mu_I \). Simultaneously, it improves the confidence and sensitivity to locate abnormality by removing \( \mu_I \).

\[
S = |I - \mu_I|
\]

2. **Visibility**. Visibility \( V \) is defined as inspired from the human visual system (Li et al. 2002) with \( \mu_I \) and the standard deviation \( \sigma_I \). Each of its elements is of the contributive rate scaling local variety. It is equivalent to the deep projection of the corresponding setover onto \( \sigma_I \).

\[
V = \left( \frac{I - \mu_I}{\sigma_I} \right)^2
\]

3. **Flat**. The grey values of a remote sensing image indirectly memorize the reflectance of the scanned groundcover by a surveying device. To eliminate the possible influence of sunshine, namely the average intensity, flat \( F \) is defined by each grey value being divided by \( \mu_I \).

\[
F = \frac{I}{\mu_I}
\]

4. **Gradient**. Gradient \( G \) is pictured by the spatial frequency (Eskicioglu and Fisher 1995) following from the fact that the relationship between contiguous grey values usually implies change. It is the way in which the grey values switch to their neighbours and estimates the overall activity level of the image.

\[
G = \left| \frac{\partial I}{\partial m} + \frac{\partial I}{\partial n} \right|
\]

where \( m \) and \( n \) denote the row and column, respectively, of the image \( I \).

5. **Contrast**. Contrast \( C \) is another ratio of the difference between the grey values of the current pixel and the background to \( \mu_I \) for magnifying the maximum likelihood of variation-dependent identification. Between the contrast and the visibility of an image, a high correlation exists (Li et al. 2002).

\[
C = \left| \frac{I - \mu_I}{\mu_I} \right|
\]

6. **Definition**. To find out where there is change, definition \( D \) is defined with the minimum \( m_I \) of all grey values, the current grey value, and the total deflection \( \delta_I \). Definition predicates that the more abrupt the change is, the clearer the
Curvature. Curvature ($U$) is a ruler of the extent of deflection, and it is rewarded by increasing the accuracy of smoothness or roughness recognition. However, it is also an indicator of salient information that will guide the variation finder (Chakraborty et al. 1995).

\[
U = \frac{|I - m_i|}{m_a - m_i}
\]

(8)

where $m_a$ is the maximum grey value.

All of the features are cognate with each other, in other words, when one is high or goes down, so the others appear. Subsequently, a feature vector formed from the above seven features can be considered as a paradigm in a mathematical structure called feature space. It is evident that this representation space is beneficial to image processing and analysis technologies by heightening the precision of significance by replacing the original image with the feature vector as follows:

\[
I_0 = [S \ V \ F \ G \ C \ D \ U]
\]

(9)

2.2 A multivalued wavelet transform (MWT)

In 1992, Mallat and Zhong designed a fast algorithm for the orthogonal wavelet transform (OWT) of a discrete signal $f_0(x)$ having finite energy by level filtering with a brace of low-pass $h(n)$ and high-pass filters $g(n)$. The down-sampled multi-resolution analysis does not preserve the translation invariance, that is a translation of the original signal does not necessarily imply a translation of the corresponding wavelet coefficients. To preserve this property, a stationary wavelet transform was introduced (Garzelli 2002). A redundant wavelet transform (RWT) can be finished by using an à trous (holes) algorithm (Holdschneider et al. 1989) as:

\[
f_j(x) = \sum_n h(n)f_{j-1}(x + n2^{j-1}); j = 1, \ldots, J
\]

(10)

\[
\omega_j(x) = f_{j-1}(x) - f_j(x) = \sum_n g(n)f_{j-1}(x + n2^{j-1}); j = 1, \ldots J
\]

(11)

The original signal can be reconstructed by adding the set of wavelet coefficients for all scales with the last approximation scale $f_J(x)$ (Chibani and Houacine 2002) as:

\[
f_0(x) = f_J(x) + \sum_{j=1}^J \omega_j(x)
\]

(12)
For an image, the OWT and the RWT can be finished by a separable convolution to the rows and columns, respectively. The MWT used in this paper can be performed by applying the RWT to each feature of $I_0$ as

$$I_{j-1}(x, y) = \sum_n h(n) I_{j-1}(x + n2^{j-1}, y)$$

$$I_j(x, y) = \sum_n h(n) I_{j-1}(x, y + n2^{j-1})$$

$$\omega_j(x, y) = I_{j-1}(x, y) - I_j(x, y); \ j = 1, \cdots, J$$

The original feature vector $I_0$ can be rebuilt perfectly as

$$I_0(x, y) = I_j(x, y) + \sum_{j=1}^J \omega_j(x, y)$$

### 2.3 Fusion scheme

In general, fusion schemes consist of three grades: an elementary, intermediate and advanced format. An elementary format makes the combination effective and real time whereas it may falsely judge interference as information; for example, pixel-, window- and region-measured energy activity (Zhang and Blum 1999), and the consistency verification method (Pajares and de la Cruz 2004). The intermediate format relies on statistical characteristics of wavelet coefficients to establish a connection between probability and spatial accuracy improvement at the expense of intensive time; for example, the correlation-guided match (Li et al. 2003), texture- and fission-instructed scheme (Viveros et al. 2002), context-based injection model (Aiazzi et al. 2002), the static fusion scheme (Loog and van Ginneken 2004), a real joint log-likelihood (Hallouli et al. 2002), the plausibility measure (Luo and Li 1994), majority voting and complete agreement (Petrakos et al. 2001), and the consensus voting and rejection schemes (Benediktsson and Kanellopoulos 1999). Advanced format is complicatedly mode-oriented; for example, semantic shot classification (Xu et al. 2003), fuzzy logic (Zhao et al. 2005, Liu et al. 2006), point and chain representation (Nikolov et al. 2000), and posterior knowledge (Lallier and Farooq 2000).

In this section, a Gaussian distribution-steering voting and consensus-supervising electing fuser is designed to construct the fused coefficients. This fuser begins with the hypothesis that for every coefficient there is a posterior probability indicating the probability of the presence or absence of the information to be fused (Loog and van Ginneken 2004). Subsequently, wavelet coefficients are divided by the posterior probabilities to obtain the underlying wavelet coefficients, which are compared to harvest $7*J$ binary images denominated voting maps. Afterwards, for each position, the number of times it has been detected as a maximum in any of the individual voting maps is stored. If the position has been voted at least 3, it is set to one, otherwise zero, in the resultant image named electing map that guides max and average scheme to yield the fused coefficients. Some elementary and intermediate formats (i.e. voting, averaging, maximum, and comparison) are accorded with this fuser. For fusing one multispectral ($T$) image and one panchromatic ($P$) image, the $T$ image is first resampled to the pixel size of the $P$ image. The fuser that produces the fused image ($F$) is summarized as follows:
1. **Voting stage.** In images, spatially local decision of pixels within a small observation window is possible (Loog and van Ginneken 2004), namely posterior probability. Because of its powerful flexibility and advantageous analytical tractability, the Gaussian model is used to estimate the posterior probability $p^j_{#i}(x, y)$ of the wavelet coefficient $\omega^j_{#i}(x, y)$ of the $i$th feature of the image $# (P$ or $T$) at scale $j$.

$$p^j_{#i}(x, y) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\omega^j_{#i}(x, y) - \mu)^2}{2\sigma^2}\right) d\omega^j_{#i}(x, y)$$

(15)

The mean and standard deviation of the wavelet coefficients within the small window ($W$) with size 3 x 3 in this paper centred on $\omega^j_{#i}(x, y)$ are denoted by $\mu$ and $\sigma$. $\overline{\omega}^j_{#i}(x, y)$ is the underlying wavelet coefficient.

2. **Electing stage.** For the majority of no-grouping schemes (Zhang and Blum 1999), these coefficients at interscale are not associated with each other. Conversely, if these coefficients at the same scale in different features and at interscale in the same feature are constrained together to take judgment, this case is designated as a cross-feature scale architecture. Let $S_T$ and $S_P$ represent the feature sequences derived from the $T$ and $P$ images. Let $\Pi$ and $E$ represent the voting map and the electing map. Let $J$ and $P_N$ represent the RWT decomposition level and the approximation component of the $P$ image. A schematic diagram of the proposed fusion method is shown in figure 1.

3. **Fusing stage.** Generally, the average scheme leads to the stabilization of the fusion result and incurs the problem of possible pattern cancellation due to opposite contrast in different input images, and this problem can be avoided using the selection scheme (Pajares and de la Cruz 2004). Only the approximation component ($T_N$) of the $T$ image is reserved for inverse RWT.

$$\omega^j_E(x, y) = \begin{cases} 
\omega^j_P(x, y) & E(x, y) > 0 \\
\omega^j_T(x, y) & E(x, y) < 0 \\
\left(\omega^j_P(x, y) + \omega^j_T(x, y)\right)/2 & E(x, y) = 0
\end{cases}$$

(17)

where $E(x, y)$ denotes the value of the electing map at position $(x, y)$.

![Figure 1. Schematic diagram of the proposed fusion method.](image-url)
The voting and electing fuser can be used to fuse such images so that the resolution ratio between the panchromatic image and the multispectral image is less than 1:5. If not, the influence that the multispectral image is interpolated to the same dimensions as the panchromatic image is not negligible.

3. Metric measures

Many objective yardsticks are defined to estimate the performances of different fusion techniques. However, three facets deserve particular attention regarding these objective yardsticks: (i) the correlation coefficient does not allow a subtle discrimination of possible fusion artefacts; the root mean square error may be scarcely accompanied by the visual appearance (Aiazzi et al. 2002); (ii) only the microcosmic similarity, not the macroscopic similarity, between images has received attention and only the information at the original scale, not the multiscale information, of the fused image has been taken into account; (iii) the multistructure correlation entropy, the all-round definition, and the spectral relationship have been overlooked.

In this section, four indices are proposed to quantify the resemblance between images and determine how much valuable information has been transmitted from source images to the resultant image. Let \(T(x, y), F(x, y)\) and \(P(x, y)\) stand for the multispectral image with size \(M \times N\), the fused image and the panchromatic image with size \(B \times C\), respectively.

1. The definition appraises the details from \(P(x, y)\) into \(T(x, y)\) and the spatial information loss in \(F(x, y)\) with respect to \(P(x, y)\) by the spatial frequency entropy \((E)\) of the image sequence \({\{K^w}\}\). \(K^w\) is obtained by \(w\)-level low-pass filtering the \(K\) image using a sliding window with weights:

\[
\begin{bmatrix}
\frac{1}{8(\sqrt{2}+1)} & \frac{\sqrt{2}}{8(\sqrt{2}+1)} & \frac{1}{8(\sqrt{2}+1)} \\
\frac{\sqrt{2}}{8(\sqrt{2}+1)} & 0.5 & \frac{\sqrt{2}}{8(\sqrt{2}+1)} \\
\frac{1}{8(2+1)} & \frac{\sqrt{2}}{8(\sqrt{2}+1)} & \frac{1}{8(\sqrt{2}+1)}
\end{bmatrix}
= \begin{bmatrix}
0.0518 & 0.0732 & 0.0518 \\
0.0732 & 0.5 & 0.0732 \\
0.0518 & 0.0732 & 0.0518
\end{bmatrix}
\]

\(K\) signifies \(T(x, y), F(x, y)\) or \(P(x, y)\). The maximal \(w\) is \(M_K\) (min\{\(M, N\)\}−2 for \(T(x, y)\) or min\{\(B, C\)\}−2 for \(F(x, y)\) and \(P(x, y)\)).

\[
d^w_K(x, y) = \left| \frac{\partial K^w(x, y)}{\partial x} + \frac{\partial K^w(x, y)}{\partial y} \right|
\]

\(d^w_K(x, y)\) is the spatial frequency map of \(K^w(x, y)\). Its probability density function is \({\{p^w_K(0), \cdots, p^w_K(L)\}}\) with \(L = \max\{d^w_K(x, y)\}\), and the definition is founded on \(E\) as:

\[
E_K = - \sum_{w=1}^{M_K} \sum_{g=0}^{L} p^w_K(g) \ln p^w_K(g)
\]

\[\text{definition} = \frac{|E_F - E_T| - |E_F - E_P|}{|E_F - E_T|}\]
2. The spectrum inspects the colour maintenance percentage of $F(x,y)$ to $T(x,y)$ in terms of maximizing the universal image quality index (UIQI) (Wang et al. 2005) with an alterable window. The length $(l) \times$ width $(w)$ of the window range $[2,...,M] \times [2,...,N]$. Above all, $F(x,y)$ is shrunk to $F'(x,y)$ with size $M \times N$. Let $UIQI(x,y)$ denote the UIQI value of $F'(x,y)$ and $T(x,y)$ from the $x$th to the $(x+l)$th row and from the $y$th to the $(y+w)$th column.

$$spectrum = \frac{\sum_{l=2}^{M-1} \sum_{w=2}^{N-1} \left[ \sum_{x=0}^{M-l-1} \sum_{y=0}^{N-w-1} UIQI(x, y)/(M-l+1)/(N-w+1) \right]}{MN} \ (20)$$

3. From the perspective of image formation (Tsai 2004), the multiconfigurable analogy overhauls the detail similarity in accordance with a series of correlation matrices $\{A_w\}_{2 \leq w \leq W} = \min\{B,C\} - 2$. Let $a_{xy}^w$ denote the correlation coefficient value of $F(x,y)$ and $P(x,y)$ from the $x$th to the $(x+w)$th row and from the $y$th to the $(y+w)$th column.

$$A_w = \begin{bmatrix}
a_{00}^w & \cdots & a_{0C-w}^w \\
\vdots & \ddots & \vdots \\
a_{B-w0}^w & \cdots & a_{B-wC-w}^w
\end{bmatrix} \quad \text{(21)}$$

$$analogy = \frac{1}{W-1} \sum_{w=2}^{W} \left[ \sum_{x=0}^{B-w-1} \sum_{y=0}^{C-w-1} a_{xy}^w \right] / [(B-w+1)(C-w+1)]$$

4. The correlation entropy of the stripe structure can collate $F(x,y)$ and $P(x,y)$ from the combination of similarity and information standpoint with the aid of the matrices below:

$$R_1 = \begin{bmatrix}1 & r_{00} & \cdots & r_{B-1B-1} \\
0 & 1 & \cdots & r_{B-2B-1} \\
0 & 0 & \cdots & r_{B-3B-1} \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1\end{bmatrix}_{(B+1) \times (B+1)} \quad \text{(22)}$$

$$R_2 = \begin{bmatrix}1 & 0 & \cdots & 0 \\
\bar{r}_{00} & 1 & \cdots & 0 \\
\bar{r}_{11} & \bar{r}_{12} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\bar{r}_{C-1C-1} & \bar{r}_{C-2C-1} & \cdots & 1\end{bmatrix}_{(C+1) \times (C+1)}$$

$\{r_k\}_{0 \leq k \leq l \leq B-1}$ and $\{\bar{r}_k\}_{0 \leq k \leq l \leq C-1}$ indicate the correlation coefficient values of $P(x,y)$ and $F(x,y)$ from the $k$th to the $l$th row and from the $k$th to the $l$th column, respectively. The less the value of $|R_1| + |R_2|$, the more dissimilar are the two images, and the more overlap information there is. 

$H$
can be built on the eigenvalue set \( \{ \lambda_i \} \) \( 1 \leq i \leq C + B + 2 \) of \(|R_1| + |R_2|\) as (Wang et al. 2003):

\[
\text{entropy} = - \sum_{i=1}^{B+C+2} \frac{\lambda_i}{B+C+2} \ln \frac{\lambda_i}{B+C+2}
\]  

(23)

These four indices attach image understanding, image analysis and image processing to image valuation. In the sense that the values of these indices may necessarily conform to the goodness of the image, they can cater for various assessment needs by affording a relationship between the microscopic properties of the image and the macroscopic indices; in other words, the larger the values of these indices, the better the fusion result, and vice versa.

4. Experiment and comparison

To verify the proposed fusion method, termed MQS (multivalued wavelet feature space) fuser, for spatial resolution improvement of multispectral images, experiments were conducted fusing three TM images of 300 m resolution with one SPOT image of 10 m resolution from Dapengwan, Shenzhen, China.

4.1 Fusing TM and SPOT images

In this section we describe how three TM images (TM3=Red, TM4=Near-infrared, TM5=Infrared) with 171\( \times \)171 pixels and one SPOT image with 5120\( \times \)5120 m\(^2\) were used to validate the MQS fuser. For comparison purposes, we also performed fusion using the SFM method. For the MQS fuser, the three TM images were interpolated to 10 m pixel size in advance, the SPOT image, the TM image and their feature sequences were decomposed with RWT into three levels, and then the voting and electing fuser was fulfilled from the first to the third level. Figure 2(a), 2(b), 2(c) and 2(d) show the original TM image as a colour composite (where TM3, TM4 and TM5 are coded in blue, green and red, respectively), the SPOT image, and the fused images produced from the SFM and MQS fusers, respectively.

The threshold \( \varepsilon \) (section 2.3) measures the similarity between the underlying transform coefficients. At sample locations where the difference between the underlying transform coefficients is lower than \( \varepsilon \), the source images are approximately similar, while at sample locations where the difference is higher than \( \varepsilon \), the source images are distinctly different. For simplicity, in this paper we used the value of \( \varepsilon = 1.5 \).

4.1.1 Visual inspection. Compared visually with the original TM image, the spatial discernment of the fused images for the pair of fusers is undoubtedly better. Some small features, such as edges and lines, that are not interpretable in the original TM image can be identified individually in each of the fused images. Other large features, such as lakes, rivers and blocks, are much sharper than those in the original TM image. These signify that both fusers can assimilate spatial information from the SPOT image.

Figure 2(c) shows less retained colours than figure 2(d), and recovery of the original colours is necessary for correct thematic mapping (Chibani and Houacine 2002). For instance, in figure 2(c), all of the green colours shown in the lower left part of figure 2(a) disappear, while in figure 2(d), some colours have been retained.
Second, with regard to clarity, a field of ‘spider-web’ shape in the left-of-centre part of figure 2(c) displays a ‘salt-and-granule’ face, whereas the one in figure 2(d) has the same appearance as that in figure 2(b). Therefore, figure 2(d) is better than figure 2(c) in terms of spectral preservation and spatial sharpness.

The main reasons for the superiority of the MQS fuser may be that the linkage between image and features allows most colour information to be preserved, whereas the voting and electing fusion scheme agglomerates complementary information while preventing unusable false information.

4.1.2 Objective comparison. Table 1 presents an objective comparison using metric measures between MQS and SFM. For convenience, we converted the fused images and the original TM image into grey images. Based on these grey images and the SPOT image, each item shown in table 1 was obtained. From table 1, it can be seen that the value for each item of figure 2(d) is higher than that of figure 2(c) except for
spectrum, and this may be because SFM modulates the intensity of the original TM image by using factors recording high frequencies from the SPOT image. The small discrepancy in all of the items is due to the relatively small percentage of the difference between the information content of the fused images and their entire information content. On the whole, the SFM fuser twists the spectral component slightly higher and accurately conveys less visual information from input images into a single image than the MQS fuser.

The essential reasons for MQS to operate properly are perhaps as follows: (i) real salient information has continuous behaviour in both feature and scale, and this means that a strong feature at one position on a particular scale in one feature will also be discovered at the same or a neighbouring position on the same scale in other feature; and (ii) MQS adopts the minority obeying majority fusion principle.

5. Conclusion

In this paper, an MWT was designed to decompose multisensor images in feature space with seven clusters of wavelet planes. The posterior probabilities of these wavelet coefficients were used to assist the voting and electing fuser under the direction of the cross-feature scale fusion strategy to produce the wavelet representation of the fused image. In addition, four indexes were proposed to evaluate the fused images of different fusion methods.

TM and SPOT panchromatic images are fused by this proposed method. The fusion result is compared with that of the traditional SFM method by visual inspection and statistical analysis, and the experimental results confirm the effectiveness of the proposed method in fusing multisensor images both spectrally and spatially. The joint use of feature space, the MQS fuser and metric measures can be considered a general approach to optimize the spatial resolution of multispectral images and estimate the fused images.

Acknowledgements

We thank the anonymous reviewers for detailed and useful comments on this paper. We thank Yan Sun of Huazhong University of Science and Technology, China, for providing many of the image samples used in this paper.

References


BABU MADHAVAN, B., SASAGAWA, T., TACHIBANA, K. and MISHRA, K.K., 2006, A high level data fusion and spatial modelling system for change detection analysis using high

<table>
<thead>
<tr>
<th>Definition</th>
<th>Spectrum</th>
<th>Analogy</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFM</td>
<td>0.8645</td>
<td>0.9203</td>
<td>0.8173</td>
</tr>
<tr>
<td>MQS</td>
<td>0.8732</td>
<td>0.9132</td>
<td>0.8905</td>
</tr>
</tbody>
</table>


Evaluating multisensor image fusion in feature space


