MTF-tailored multi-resolution analysis for fusion of multi-spectral and panchromatic images

B. Aiazzi[1], L. Alparone[2], S. Baronti[1], A. Garzelli[3], F. Nencini[3] & M. Selva[1]
[1] *Institute of Applied Physics "Nello Carrara" of the National Research Council IFAC–CNR, Florence, Italy*
[2] *Department of Electronics & Telecommunications, University of Florence, Florence, Italy*
[3] *Department of Information Engineering, University of Siena, Siena, Italy*

Keywords: modulation transfer function, image fusion, multi-resolution analysis, pan-sharpening

ABSTRACT: This work presents a multi-resolution framework for merging a multi-spectral image having an arbitrary number of bands with a higher-resolution panchromatic observation. The fusion method relies on the *generalised Laplacian pyramid* (GLP), which is a multi-scale oversampled structure. The goal is to selectively perform injection of spatial-frequencies from an image to another with the constraint of thoroughly retaining the spectral information of the coarser data. The novel idea is that a model of the modulation transfer functions (MTF) of the multi-spectral scanner is exploited to design the GLP *reduction* filter. Thus, the inter-band structure model (IBSM), which is calculated at the coarser scale, where both MS and Pan data are available, can be extended to the finer scale, without the drawback of the poor enhancement occurring when MTFs are assumed to be ideal filters. Experiments carried out on QuickBird data demonstrate that a superior spatial enhancement, besides the spectral quality typical of injection methods, is achieved by means of the MTF-adjusted fusion.

1 INTRODUCTION

Image fusion techniques, originally devised to allow integration of different information sources, may take advantage of the complementary spatial/spectral resolution characteristics typical of remotesensing imagery. When exactly three multi-spectral (MS) bands are concerned, the most straightforward fusion method is to resort to an intensity-hue-saturation (IHS) transformation. This procedure is equivalent to *inject*, i.e. add, the difference between the sharp panchromatic (Pan) image and the smooth intensity into the resampled MS bands (Tu *et al*. 2001). Since the Pan image, histogram-matched to the intensity component, does not generally have same local radiometry as the latter, when the fusion result is displayed in colour composition, large spectral distortion (colour changes) may be noticed. When more than three spectral bands are available, IHS fusion may be applied to three consecutive spectral components at a time, or better the IHS transformation may be replaced with principal component analysis (PCA). Generally speaking, if the spectral responses of MS bands are not perfectly overlapped with the Pan bandwidth, as it happens to Ikonos-2 and QuickBird, IHS- and PCA-based methods may yield very poor results in terms of spectral fidelity.

To definitely overcome this inconvenience, methods based on injecting spatial details only, taken from the Pan image without resorting to IHS transformation, have been introduced and have demonstrated superior performances. Multi-resolution analysis (MRA) provides effective tools, like wavelets and pyramids (Aiazzi *et al*. 2002), to carry out image merge tasks. However, in the case of high-pass detail injection, *spatial* distortions, typically *ringing* or *aliasing* effects, originating shifts or blur of contours and textures, may occur. These drawbacks, which may be as much annoying as spectral distortions, are emphasized by mis-registration between MS and Pan data, especially if the MRA underlying detail injection is not shift-invariant (Aiazzi *et al.* 2002; Gonzáles Audícana *et al.* 2004).

Goal of an advanced fusion method is to increase spectral information, by *unmixing* the coarse MS pixels through the sharp Pan image. The task requires the definition of a model establishing how the missing high-pass information to be injected is extracted from the Pan image. It may be accomplished either in the domain of *approximations* between each of the resampled MS bands and a low-pass version of the Pan image having the same spatial frequency content as the MS bands, or in that of medium frequency *details*, in both cases by measuring local matching (Ranchin *et al.* 2003).

Quantitative results of data merge 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, or by degrading all available data to a coarser resolution and carrying out merge from such data. In practical cases this strategy is not feasible. The idea behind, however, is that algorithms optimised 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 does not hold for very high resolution data, especially when a highly detailed urban environment is concerned. The reason of this behavior lies in the characteristics of the *modulation transfer function* (MTF) of the imaging system. Any inter-scale injection model should take into account that the MTF of real systems is generally bell-shaped and its magnitude value at the cutoff Nyquist frequency is far lower than 0.5, to prevent aliasing. Furthermore, the MTFs of the MS sensors may be significantly different from one another in terms of decay rate, and especially are different from that of the Pan sensor. Hence, models empirically optimised at a coarser scale on data degraded by means of digital filters that are close to be ideal, may yield little enhancement when are utilized at the finer scale. In this work, an assumed model of MTF is exploited to improve results of fusion schemes based on MRA, provided that the latter is not critically-subsampled.

2 MTF-TAILORED MULTI-RESOLUTION ANALYSIS

Figure 1(a) shows the theoretical MTF of an imaging system. As a first approximation, the MTF is equal to the optical point spread function (PSF) of the imaging system. In principle, two spectral

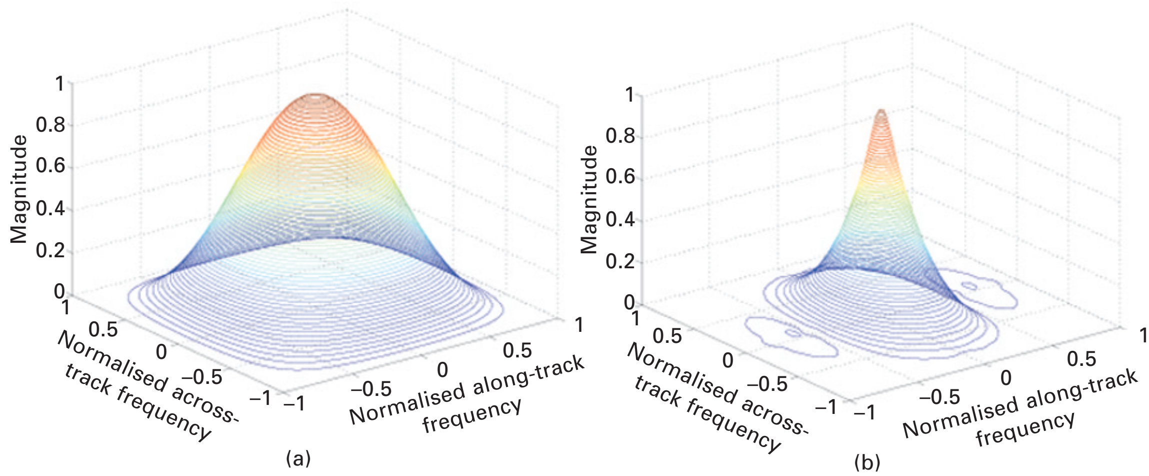

Figure 1. (a) ideal (isotropic) MTF with magnitude equal to 0.5 at cut-off Nyquist frequency; (b): typical (anisotropic) MTF with magnitude 0.2 at cut-off Nyquist frequency (across-track). All frequency scales are normalised to sampling frequency, or Nyquist rate (twice the Nyquist frequency).

replicas originated by 2D sampling of the radiance signal with sampling frequency along- and across-track equal to the Nyquist rate, should cross each other at the Nyquist frequency (half of Nyquist rate) with magnitude values equal to 0.5. However, the scarce selectivity of the response prevents from using a Nyquist frequency with magnitude equal to 0.5. As a trade-off between maximum spatial resolution and minimum aliasing of the sampled signal, the Nyquist frequency is usually chosen such that the corresponding magnitude value is around 0.2. This situation is depicted in Fig. 1(b) portraying the true MTF of an MS channel, which also depends on the platform motion (narrower along track), on atmospheric effects and on the finite sampling. A different situation occurs for the MTF of the Pan channel, whose extent is mainly dictated by diffraction limits (at least for instruments with resolution around 1 m). In that case the cut-off magnitude may be even lower (e.g. 0.1) and the appearance of the acquired images is rather blurred. However, whereas the enhancing Pan image that is commonly available has been already processed for MTF restoration, the MS bands cannot be pre-processed analogously because of SNR constraints. In fact, restoration implies a kind of inverse filtering, that has the effect of increasing the noisiness of the data.

Eventually, the problem may be stated in the following terms: an MS band resampled at the finer scale of the Pan image lacks high spatial frequency components, that may be inferred from the Pan image via a suitable inter-scale injection model. If the high-pass filter used to extract such frequency components from the Pan image is taken such as to approximate the complement of the MTF of the MS band to be enhanced, then the high frequency components, that are present in the MS band but have been damped by its MTF, can be restored. Otherwise, if spatial details are extracted from the Pan image by using a filter having normalised frequency cut-off at exactly the scale ratio between Pan and MS (e.g. 1/4 for 1 m Pan and 4 m MS), such frequency components will not be injected. This occurs with critically sub-sampled wavelet decompositions, whose filters are constrained to cut-off at exactly an integer fraction (usually a power of two) of the Nyquist frequency of Pan data, corresponding to the scale ratio between Pan and MS.

An attractive characteristic of the redundant pyramid and wavelet decompositions proposed by the authors (Aiazzi *et al*. 2002; Garzelli & Nencini 2005) is that the low-pass reduction filter used to analyse the Pan image may be easily designed such that it matches the MTF of the band into which the details extracted will be injected. Figure 2(a) shows examples for three values of magnitude cut-off. The resulting benefit is that the restoration of spatial frequency content of the MS bands is provided by the multi-resolution analysis of Pan through the injection model.

A forerunner of this rationale was the work by Núñez *et al*. (1999), even if no considerations on MTF were made. Figure 2(b) shows that the Gaussian-like frequency response of the cubic-spline

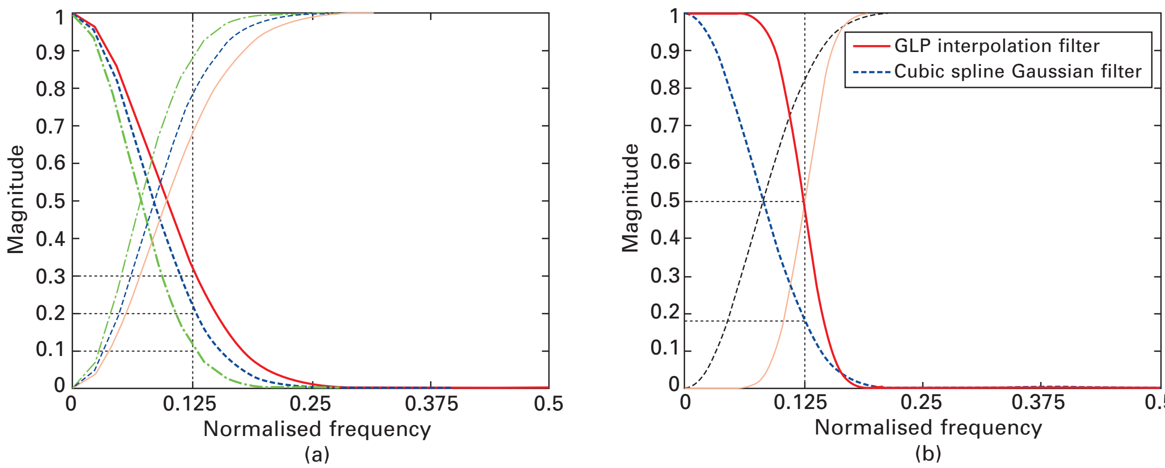

Figure 2. 1D frequency responses of equivalent filters for separable 2D 1:4 multi-resolution analysis (low-pass and high-pass filters generate approximation and detail, respectively). (a): sample MTF-adjusted Gaussian-shaped filters with magnitudes 0.1, 0.2 and 0.3 at cut-off Nyquist frequency; (b): GLP-generating filters (solid) and cubic spline filter for ATWT generation (dotted).

wavelet filter used to generate the "à trous" wavelet transform (ATWT) matches the shape of the the MTF of a typical V-NIR band, with cut-off magnitude value equal to 0.185. The complementary high-pass filter, yielding the detail level to be injected for 1:4 fusion, retains a greater amount of spatial frequency components than an ideal filter, such that used by the standard GLP and ATWT (Aiazzi *et al.* 2002; Garzelli & Nencini 2005), thereby resulting in a greater spatial enhancement.

3 GLP-BASED FUSION SCHEME

Figure 3 shows the flowchart of general case of MS + Pan fusion with 1 : p scale ratio, based on the GLP (Aiazzi *et al.* 2002). In this context, emphasis will be given to the detail injection model, indicated as inter-band structure model (IBSM), for which two improvements of the solutions introduced by Garzelli & Nencini (2005) and by Aiazzi *et al.* (2002) will be described.

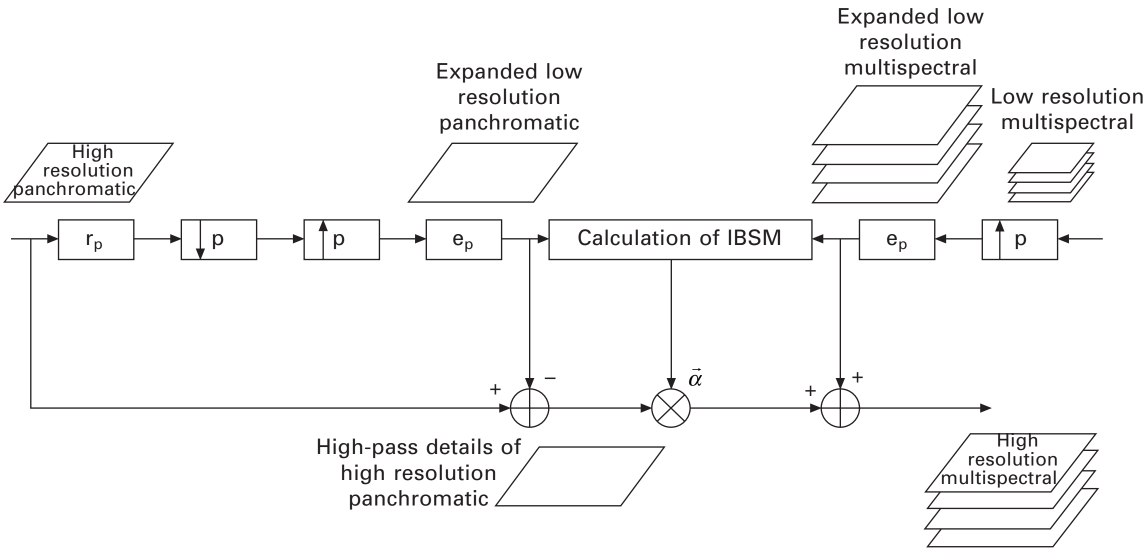

Figure 3. Flowchart of multi-resolution fusion based on generalised Laplacian pyramid and suitable for merging multi-spectral and panchromatic image data, whose integer scale ratio is $p > \cdot \downarrow p$ denotes *down-sampling* by p; $\uparrow p$ denotes *up-sampling* by p; r_p is the MTF-matching p-reduction low-pass filter with cutoff at $1/p$ of the spectrum extent; e_p is *ideal* p-expansion low-pass filter with $1/p$ cutoff.

3.1 *Spectral Distortion Minimizing (SDM) model*

Given two spectral vectors $\mathbf{v}$ and $\hat{\mathbf{v}}$ both having L components, in which $\mathbf{v} = \{v_1, v_2, \ldots, v_L\}$ is the original spectral pixel vector $v_l = G^{(l)}(i, j)$ while $\hat{\mathrm{v}} = \{\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_L\}$ is the distorted vector obtained by applying fusion to the coarser resolution MS data, i.e. $\hat{v} = \hat{G}^{(l)}(i, j)$, the spectral angle mapper (SAM) denotes the absolute value of the spectral angle between the two vectors:

$$\mathrm{SAM}(\mathbf{v}, \hat{\mathbf{v}}) \triangleq \arccos\left(\frac{<\mathbf{v}, \hat{\mathbf{v}}>}{\|\mathbf{v}\|_2 \cdot \|\hat{\mathbf{v}}\|_2}\right). \tag{1}$$

The *Spectral Distortion Minimizing* (SDM) model is both space- and spectrally-varying; stated with a vector notation, $\vec{\alpha}_S(i, j) = \{\alpha_S^{(l)}(i, j), l = 1, \ldots, L\}$. Let $\vec{\tilde{G}}(i, j) \triangleq \{\tilde{G}^{(l)}(i, j), l = 1, \ldots, L\}$ denote the pixel vector of the expanded MS image; let also $\vec{W}_0^{MS}(i, j) \triangleq \vec{\alpha}_S(i, j) \cdot W_0^{(P)}(i, j)$ denote the MS detail vector to be injected. In order to minimize the SAM distortion between resampled MS bands and fused products, the injected detail vector at pixel position (i, j) must be parallel to the resampled MS vector, i.e. to $\vec{\tilde{G}}(i, j)$. At the same time each component $\alpha^{(l)}(i, j)$

should be designed so as to minimize the radiometric distortion when the detail component $W_0^{(l)}(i,j)$ is injected into $\tilde{G}^{(l)}(i,j)$. Starting from the vector merge relationship

$$\hat{\vec{\tilde{G}}}(i,j) = \vec{\tilde{G}}(i,j) + \vec{\alpha}_S(i,j) \cdot W_0^{(P)}(i,j) \tag{2}$$

let us define the lth components of $\vec{\alpha}_S(i,j)$ as

$$\alpha_S^{(l)}(i,j) = \beta(i,j) \cdot \frac{\tilde{G}^{(l)}(i,j)}{\tilde{G}_2^{(P)}(i,j)}, l = 1, \ldots, L \tag{3}$$

in which $\tilde{G}_2^{(P)}(i,j)$ denotes the approximation of the Pan image produced by the equivalent low-pass filter and the coefficient $\beta(i, j)$, which does not depend on the band index l, is defined in such a way that the length of the fused vector is statistically close to that of the (unavailable) true high-resolution vector, as the ratio of local standard deviations of average of resampled MS bands and low-pass approximation of Pan:

$$\beta(i,j) = \sqrt{\frac{\operatorname{var}\left[\frac{1}{L}\sum_{l=1}^{L}\tilde{G}^{(l)}\right](i,j)}{\operatorname{var}[\tilde{G}_2^{(P)}](i,j)}}. \tag{4}$$

From (2) and (3) it stems that

$$\vec{\tilde{G}} \times \hat{\vec{\tilde{G}}} = \vec{\tilde{G}} \times [\vec{\tilde{G}} + \vec{W}_0^{MS}] = \vec{\tilde{G}} \times [\vec{\tilde{G}} + \vec{\alpha}_S \cdot W_0^{(P)}] = \vec{\tilde{G}} \times \vec{\tilde{G}} \cdot \left[1 + \beta \cdot \frac{W_0^{(P)}}{\tilde{G}_2^{(P)}}\right] = 0 \tag{5}$$

in which $\times$ stands for vector product and indexes (i, j) have been omitted. Eq. (5) states that the spectral angle (SAM) is unchanged when a vector pixel in the expanded MS image, $\vec{\tilde{G}}(i,j)$, is enhanced to yield the fused product, $\hat{\vec{\tilde{G}}}(i,j)$, because the upgrade $\vec{W}_0^{MS}(i,j)$ is always parallel to $\vec{\tilde{G}}(i,j)$.

3.2 *Context-Based Decision (CBD) model*

The CBD injection model rules the transformation in the detail from the Pan image to the target lth MS band. It may be stated as

$$W_0^{(l)}(i,j) = \alpha_C^{(l)}(i,j) \cdot W_0^{(P)}(i,j), l = 1, \ldots, L. \tag{6}$$

The space-varying model $\alpha_C^{(l)}(i,j)$ is calculated between the lth MS band resampled to the scale of the Pan image, $\tilde{G}^{(l)}$, and the approximation of the Pan image at the resolution of the MS bands, $\tilde{G}_2^{(P)}$,

$$\alpha_C^{(l)}(i,j) = \min\left\{\frac{\rho_{l,P}(i,j)}{\bar{\rho}_{l,P}} \cdot \frac{\sigma_{\tilde{G}^{(l)}}(i,j)}{\sigma_{\tilde{G}_2^{(P)}}(i,j)}, c\right\} \tag{7}$$

in which $\rho_{l,P}(i, j)$ is the *local* linear correlation coefficient between $\tilde{G}^{(l)}$ and $\tilde{G}_2^{(P)}$ calculated on a $N \times N$ sliding window centered on pixel (i, j); $\bar{\rho}_{l,P}$ is the *global* one and is constant throughout.

The CBD model is uniquely defined by the window size N depending on the spatial resolutions and scale ratio of the images to be merged, as well as on the landscape characteristics (typically, $7 \leq N \leq 11$). A clipping constant c was introduced to avoid numerical instabilities ($2 \leq c \leq 3$).

4 EXPERIMENTAL RESULTS

4.1 *Quality assessment of fusion products*

Quality assessment of Pan-sharpened MS images is a hard task (Chavez Jr *et al.* 1991; Wald *et al.* 1997). Even when spatially degraded MS images are processed for Pan-sharpening, and therefore reference MS images are available for comparisons, assessment of *fidelity* to the reference usually requires computation of a number of different indexes. Examples are correlation coefficient (CC) between each band of the fused and reference MS images, bias in the mean, root mean square error (RMSE) and SAM, which measures the spectral distortion introduced by the fusion process.

Wald *et al.* (1997) proposed an error index that offers a global picture of the quality of fused data, denoted as ERGAS, which means *relative dimensionless global error in synthesis*, and given by:

$$\mathrm{ERGAS} \triangleq 100\,\frac{d_h}{d_l}\sqrt{\frac{1}{L}\sum_{l=1}^{L}\left(\frac{\mathrm{RMSE}(l)}{\mu(l)}\right)^2} \quad (8)$$

where dh/dl is the ratio between pixel sizes of Pan and MS, e.g. 1/4 for Ikonos and QuickBird data, $\mu(k)$ is the mean (average) of the lth band, and L is the number of bands.

Another quality index, namely Q4, suitable for images having four spectral bands, was recently proposed by the authors for quality assessment of Pan-sharpened MS imagery (Alparone *et al.* 2004).

4.2 *Performance comparison of image fusion algorithms*

The proposed fusion procedures have been assessed on very high-resolution image data collected on June 23 2002 at 10:25:59 GMT+2 by the QuickBird spaceborne MS scanner on the urban area of Pavia, Italy. The four MS bands span the visible and near-infrared (NIR) wave-lengths and are spectrally disjointed. The Pan band embraces the interval 450 ÷ 900 nm. All data are resampled to uniform ground resolutions of 2.8 m and 0.7 m GSD for MS and Pan. The MS and Pan data were spatially degraded by 4, to yield 2.8 m P and 11.2 m MS, and used to synthesize the four bands at 2.8 m. Thus, the true MS data at 2.8 m are available for objective distortion measurements. Both an ideal filter and MTF-matched filters were used for pre-filtering MS data before decimation. In the former case the fusion algorithm uses a conventional MRA. In the latter, the reduction filter of GLP matches the MTF of each MS band, as shown in Sect. 2. Pan data were always degraded with the ideal filter. A comparison of GLP-based methods with improved SDM (see Subsect. 3.1) and modified CBD (see Subsect. 3.2) models was carried out with two methods: multi-resolution IHS (Núñez *et al.* 1999) with additive model (AWL), based on ATWT, and high-pass filtering (HPF) technique (Chavez Jr *et al.* 1991), with either 5×5 or 7×7 box filter for ideally- or MTF-degraded MS bands.

The parameters in Table 1 measuring the global distortion of pixel vectors, either radiometric (ERGAS, which should be as low as possible) or spectral (SAM), and both radiometric and spectral (Q4, which should be as close as possible to one) will give a comprehensive measurement of fusion performance. EXP denotes the case in which the degraded MS data are resampled through the 23-taps expansion filter of GLP and no injection of details is made. Depending on whether degradation is ideal or MTF-based, the quality of all fusion products, including the resampled MS bands (EXP) significantly changes, being lower in the latter case, because the startup MS data are poorer in both spectral and geometrical content. Only AWL is little penalised in the latter case, because its MRA implicitly embeds the MTF model. The SAM attained by CBD is lower than that of SDM (identical to that of resampled MS data), thanks to the *unmixing* capabilities of the former compared to the latter.

All methods accounting for MTF exhibit a performance increment with respect to plain resampling that is greater than that noticed when degradation is ideal and methods do not need MTF adjustment.

Eventually, Fig. 4 displays true colour compositions of the resampled 2.8 m MS bands and of

Table 1. Average cumulative quality indexes between 2.8 m MS spectral vectors and those obtained from fusion of 11.2 m MS with 2.8 m Pan. MS bands degraded with either ideal or MTF filters.

Ideal	EXP	AWL	SDM	CBD	HPF (5 × 5)
Q4	0.756	0.827	0.878	0.862	0.848
SAM (deg.)	2.14°	2.59°	2.14°	1.90°	2.51°
ERGAS	1.760	1.611	1.470	1.695	2.012
MTF	**EXP**	**AWL**	**SDM**	**CBD**	**HPF (7 × 7)**
Q4	0.641	0.815	0.819	0.834	0.781
SAM (deg.)	2.46°	2.53°	2.46°	2.13°	2.81°
ERGAS	1.980	1.647	1.608	1.589	2.132

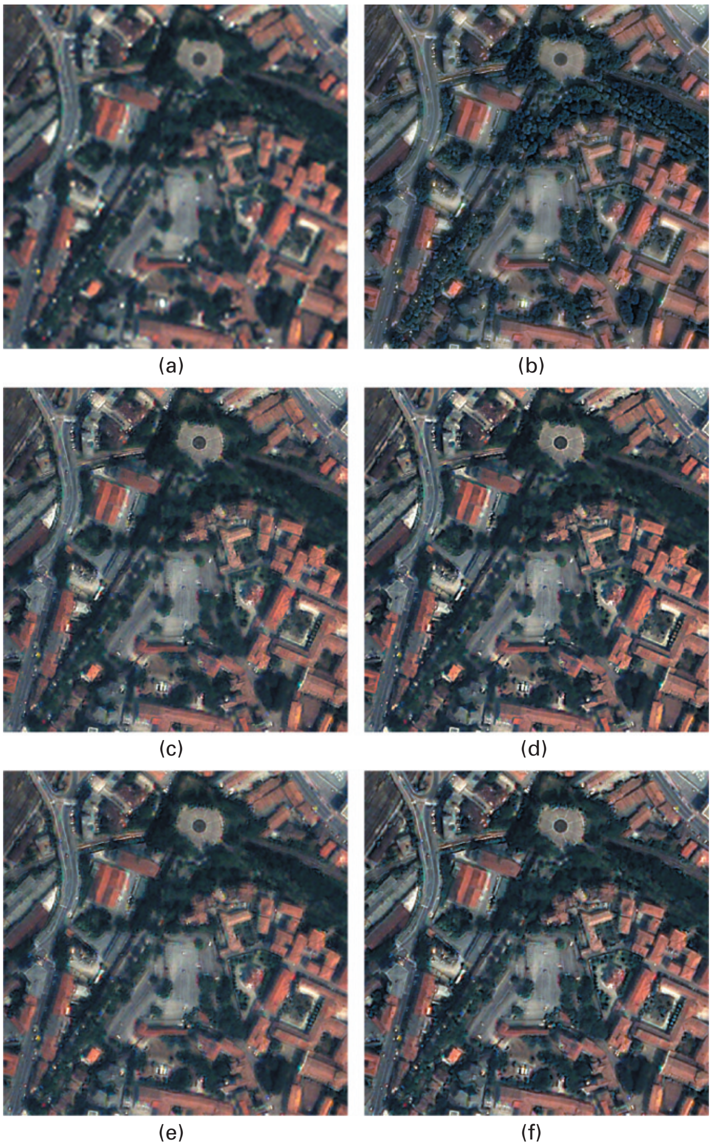

Figure 4. Examples of full-scale spatial enhancement of fusion algorithms displayed as 512 × 512 true color compositions at 0.7 m pixels spacing. (a): original MS bands (2.8 m) resampled to the scale of Pan image (0.7 m); (b): AWL method; (c): GLP-CBD without MTF adjustment; (d): GLP-CBD with MTF adjustment; (e): GLP-SDM without MTF adjustment; (f): GLP-SDM with MTF adjustment.

the spatially enhanced bands, all at 0.7 m. The SDM and CBD models are applied to GLP, achieved either without or with MTF-matched filters. A visual inspection highlights that all the spectral signatures of the original MS data are carefully incorporated in the sharpened bands. Thanks to the two injection models, the texture of the canopies, which is highlighted by the Pan image, but mostly derives from the NIR band, which is outside the visible wave-lengths, appears to be damped in the SDM and CBD fusion products. AWL, which implicitly accounts for the MTF in the MRA, is geometrically rich and detailed, but unlikely over-enhanced, especially on vegetated areas. Visual analysis reveals that the methods adopting MTF-tailored MRA yield sharper and cleaner geometrical structures.

REFERENCES

Aiazzi, B., Alparone, L., Baronti, S. & Garzelli, A. 2002. 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.

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

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

Garzelli, A. & Nencini, F. 2005. Interband structure modeling for Pan-sharpening of very high resolution multispectral images. *Information Fusion* 6(3) 213–224.

Gonzáles Audícana, M., Saleta, J. L., García Catalán, R. & García, R. 2004. 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.

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

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

Tu, T.-M., Su, S.-C., Shyu, H.-C. & Huang, P. S. 2001. A new look at IHS-like image fusion methods. *Information Fusion* 2 (3) 177–186.

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