

Pyramid-Based Multi-sensor Image Data Fusion with Enhancement of Textural Features

B. Aiazzi[°], L. Alparone^{*}, S. Baronti[°],
V. Cappellini^{*}, R. Carlà[°], L. Mortelli^{*}

[°]IROE “Nello Carrara” - CNR

Via Panciatichi, 64, 50127 Firenze, Italy

E-Mail: baronti@iroe.fi.cnr.it

^{*}Dip. Ing. Elettronica, University of Florence

Via S. Marta, 3, 50139 Firenze, Italy

E-Mail: alparone@cosimo.die.unifi.it

Abstract. In this work, a multi-resolution procedure based on a generalized *Laplacian pyramid* (GLP) with a *rational* scale factor is proposed to merge image data of any resolution and represent them at any scale. The GLP-based data fusion is shown to be superior to those of a similar scheme based on the discrete *wavelet transform* (WT) according to a set of parameters established in the literature. The pyramid-generating filters can be easily designed for data of any resolutions, differently from the WT, whose filter-bank design is non-trivial when the ratio between the scales of the images to be merged is not a power of two. Remotely sensed images from Landsat TM and from Panchromatic SPOT are fused together. Textured regions are enhanced without losing their spectral signatures, thereby expediting automatic analyses for contextual interpretation of the environment.

1 Multi-sensor Image Data Fusion

The availability of data from many sensors with different characteristics makes data fusion a topic of ever increasing relevance in the field of digital image processing. The main goal of pixel level algorithms [1] is to combine the original images from different sensors in order to synthesize a new set of data whose spatial and/or spectral resolution results to be enhanced, or to concentrate significant features of the various bands in a single image, thus compressing information and enhancing contrast and texture. In some cases processing is made with the main objective of extracting significant features [2] by maximizing the spatial contrast on the basis of the whole data set: distortion measures are not considered. In other applications, such as classification, merging algorithms are requested to maintain the spectral characteristics of the original data as much as possible [3] to avoid misinterpretation and introduction of undesired effects.

Approaches based on *principal component analysis* (PCA), on transformation of the original data in the *hue intensity saturation* (HIS) color space (three bands at a time), and on *high pass filtering* (HPF) [3] have been investigated

in the literature to achieve the latter objective: HPF resulted far more efficient than the other algorithms in preserving the spectral features of the enhanced bands. Therefore, such *space-frequency* image representations as *discrete wavelet transform* (WT) and *Laplacian pyramid* (LP) have been recently investigated for image fusion aimed at contrast enhancement [2].

Multi-spectral Earth observations from space exhibit limited spatial resolutions, differently from broad-spectrum imaging sensors, that may be inadequate to specific identification tasks. A typical example of such a situation is represented by Landsat Thematic Mapper (TM) multi-spectral imaging sensor, which has a $30m \times 30m$ ground resolution in seven spectral bands and by the SPOT panchromatic (PAN) sensor, which provides single-band observations on a broad wavelength interval, with a $10m \times 10m$ pixel size. Data fusion of Landsat-TM and SPOT-PAN images have been previously considered in the literature [3] due to their availability and their complementary spatial/spectral features. This paper reports about a *pyramid-based* approach to data fusion of Landsat-TM and SPOT-PAN images, with images previously registered on a common cartographic base, each at its own scale ($30m$ and $10m$, respectively). The proposed algorithm is a variant of the high-pass filter (HPF) method by Chavez *et al.* [3], recognized as one of the most efficient. Its generalization is achieved in a pyramid framework, since a generalized pyramid is an efficient structure by which both the high-pass filtering and the contrast enhancement algorithms can be easily implemented. Images are available at several different spatial scales. The *expansion/reduction* filters can be easily designed to cope with data of any resolutions from different sensors. Once new data from different sensors will be available on the selected test site (e.g., SAR data, digitized aerial photographs, hyperspectral high resolution aircraft data, data from *new-generation* satellites), they will be easily merged with the existing ones in order to assess any advantages occurring from a cooperative analysis based on multiple imaging sources. The algorithm is assessed in terms of both objective scores and visual quality. Spectral feature preservation of Landsat images is evaluated. The performance of the merging procedure is previously discussed and assessed in a comparison with an analogous scheme based on the WT [2], recently established in the literature. The pyramid algorithm is found to be superior on the basis of both subjective and objective criteria.

2 Multi-scale Image Analysis

2.1 Wavelet transform

The wavelet transform provides a multi-resolution representation of continuous and discrete signals [4]. When it is applied to a sequence of discrete data $f(n)$, the original signal can be considered as the coefficients of the projection of a continuous function into the highest resolution subspace: the coefficients relative to the lower resolution subspace and to its orthogonal complement can be obtained through the subsampling of the discrete convolution of $f(n)$ and the coefficients of the impulse response of two digital filters $H(\omega)$ and $G(\omega)$,

respectively low-pass and high-pass [4]. The two outcome sequences represent a smoothed version of $f(n)$ and a detail signal, respectively: the latter, being the output of a high-pass filter, highlights the points in which rapid changes of the signal occur. In a similar manner, the higher resolution data can be retrieved from the lower resolution projections by up-sampling and low-pass filtering. Therefore, the wavelet representation is closely related to a two-channel sub-band decomposition scheme.

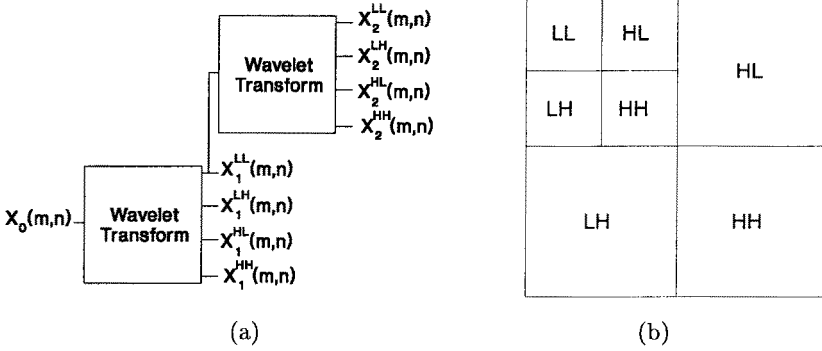


Fig. 1. Two-level 7 sub-bands wavelet transform scheme (a) and associated spatial frequency sub-bands (b).

If two dimension signals are dealt with, a wavelet representation can be obtained by separately processing rows and columns of the array. Let $X_0(m,n)$ be the original image with dimensions $M \times N$, and let $X_1^{LL}(m,n)$, $m = 0, \dots, M/2 - 1$, $n = 0, \dots, N/2 - 1$ be the lower resolution subsequence obtained by low-pass filtering rows and columns; analogously, let $X_1^{LH}(m,n)$, $X_1^{HL}(m,n)$, and $X_1^{HH}(m,n)$, $m = 0, \dots, M/2 - 1$, $n = 0, \dots, N/2 - 1$, be the sub-sequences obtained by the combination of low-pass and high-pass filtering along the rows or the columns. Since high-pass filtering highlights edges in an image, $X_1^{HL}(m,n)$ and $X_1^{LH}(m,n)$ will contain information about vertical and horizontal contours, respectively. With analogous considerations $X_1^{HH}(m,n)$ highlights diagonal details. Further splitting of $X_1^{LL}(m,n)$ yields a multi-level decomposition: the signals $X_2^{LL}(m,n)$, $X_2^{LH}(m,n)$, $X_2^{HL}(m,n)$ and $X_2^{HH}(m,n)$, $m = 0, \dots, M/2^2 - 1$, $n = 0, \dots, N/2^2 - 1$, are produced at the second level of the decomposition and general expressions for the higher levels can easily be derived. Figure 1 shows the scheme for a two-level decomposition yielding a seven sub-bands representation: in the figure the Wavelet Transform block denotes the one-level four sub-bands separable splitting. The low-frequency coefficients $X_i^{LL}(m,n)$, are further decomposed, thus yielding a wavelet space-frequency representation, in which the wavelet coefficients may be accommodated into sub-bands based on their content of spatial frequencies.

2.2 Laplacian pyramid

The Gaussian pyramid (GP) is a multi-resolution image representation obtained through a recursive *reduction*, i.e. low-pass filtering and decimation. Let $G_0(m, n)$, $m = 0, \dots, M-1$, and $n = 0, \dots, N-1$, $M = u \times 2^K$, $N = v \times 2^K$, be the input image. The GP [5] is defined with a decimation factor of 2 ($\downarrow 2$) as

$$\begin{aligned} G_k(m, n) &= \text{reduce}_2[G_{k-1}](m, n) \\ &\triangleq \sum_{i=-L_r}^{L_r} \sum_{j=-L_r}^{L_r} r_2(i) \times r_2(j) G_{k-1}(2m+i, 2n+j) \end{aligned} \quad (1)$$

for $k = 1, \dots, K$, for $m = 0, \dots, M/2^k - 1$, $j = 0, \dots, N/2^k - 1$, in which k identifies the level of the pyramid. The 2D reduction (low-pass) filter is given as the outer product of a linear symmetric odd-sized kernel $\{r_2(i)\}$ which should cut-off at one half of the signal bandwidth, to prevent aliasing.

From the GP, the LP is defined, for $k = 0, \dots, K-1$, as

$$L_k(m, n) \triangleq G_k(m, n) - \text{expand}_2[G_{k+1}](m, n) \quad (2)$$

in which $\text{expand}_2[G_{k+1}]$ denotes that the $(k+1)^{\text{st}}$ level of the GP is expanded by a factor 2 to match the size of the underlying k^{th} level:

$$\text{expand}_2[G_{k+1}](m, n) \triangleq \sum_{\substack{i=-L_e \\ (j+n) \bmod 2=0 \\ (i+m) \bmod 2=0}}^{L_e} \sum_{\substack{j=-L_e \\ (j+n) \bmod 2=0 \\ (i+m) \bmod 2=0}}^{L_e} e_2(i) \times e_2(j) G_{k+1}\left(\frac{i+m}{2}, \frac{j+n}{2}\right) \quad (3)$$

for $m = 0, \dots, M/2^k - 1$, $n = 0, \dots, N/2^k - 1$, and $k = 0, \dots, K-1$. The 2D low-pass filter for *expansion* is given as outer product of a linear symmetric odd-sized kernel $\{e_2(i)\}$, which again should cut-off at one half of the bandwidth. Summation terms are taken to be null for noninteger values of $(i+m)/2$ and $(j+n)/2$, corresponding to interleaving zeroes introduced by up-sampling ($\uparrow 2$).

2.3 Generalized LP with a rational scale factor

The expressions found for (1) and (3) may be generalized to comprise reduction and expansion factors different from 2 [7]. Reduction by q is defined as:

$$\text{reduce}_q[G_k](m, n) \triangleq \sum_{i=-L_r}^{L_r} \sum_{j=-L_r}^{L_r} r_q(i) \times r_q(j) G_k(qm+i, qn+j) \quad (4)$$

The reduction (low-pass) filter $\{r_q(i)\}$ should be designed to cut-off at one q^{th} of the signal bandwidth. Expansion by a factor p is defined as:

$$\text{expand}_p[G_k](m, n) \triangleq \sum_{\substack{i=-L_e \\ (j+n) \bmod p=0 \\ (i+m) \bmod p=0}}^{L_e} \sum_{\substack{j=-L_e \\ (j+n) \bmod p=0 \\ (i+m) \bmod p=0}}^{L_e} e_p(i) \times e_p(j) G_k\left(\frac{i+m}{p}, \frac{j+n}{p}\right) \quad (5)$$

The expansion filter $\{e_p(i)\}$ should cut-off at one p^{th} of the signal bandwidth. Summation terms are null for noninteger values of $(i + m)/p$ and $(j + n)/p$.

If $p/q > 1$, p, q integers, is the scale factor between two images to be merged, (1) modifies into the cascade of an expansion by q and a reduction by p

$$G_{k+1} = reduce_{p/q}[G_k] \triangleq reduce_p\{expand_q[G_k]\} \quad (6)$$

while (3) becomes an expansion by p followed by a reduction by q .

$$expand_{p/q}[G_k] \triangleq reduce_q\{expand_p[G_k]\} \quad (7)$$

When (4) is cascaded to (5), convolution can be skipped after up-sampling in (6), as well as before down-sampling in (7).

The Generalized Laplacian Pyramid (GLP) with p/q scale factor between adjacent layers, \tilde{L}_k , can thus be defined as:

$$\tilde{L}_k(m, n) \triangleq G_k(m, n) - expand_{p/q}\{reduce_{p/q}[G_k]\}(m, n) \quad (8)$$

The filter design usually is a tradeoff between selectivity (sharp cutoff) and computational cost. Filters with different characteristics have to be designed to cope with bandwidth requirements of data fusion algorithms. In particular for a p/q scale ratio, only low-pass filters with $1/p$ and $1/q$ normalized frequency cut-offs are needed. Instead, the WT requires also a high-pass filter (i.e. a complete filter-bank) which must generally be re-designed for every value of p/q .

3 Multi-resolution Data Fusion Schemes

The idea of the wavelet-based image fusion algorithm developed by Li *et al.* [2] is to merge couples of sub-bands of corresponding frequency content on the basis of an activity measure locally computed on 2×2 blocks of coefficients. The fused image is produced by taking the inverse transform of the blocks of coefficients chosen as the more *active* between the two images.

The block diagram reported in Figure 2 describes the data fusion algorithm in the general case of two image data sets, preliminarily registered on the same cartographic base, whose scale ratio is p/q . Let S_1 be the data set constituted by a single image having smaller scale and S_2 the data set made up of several multi-spectral observations with larger scale. The goal is to obtain a set of as many multi-spectral images as S_2 , each having same spatial resolution as S_1 . The upgrade of S_2 to the resolution of S_1 is the zero-mean GLP (8) of S_1 , computed for $k = 0$. The high-pass component from S_1 is added to each of the expanded images of S_2 to yield an enhanced set of multi-spectral observations, S_3 .

4 Experimental Results

Figures 3(a) and (b) show Band 6 (thermal infrared) and Band 5 (near infrared) of a Landsat TM image portraying a zone of the Elba island, in Tuscany, Italy.

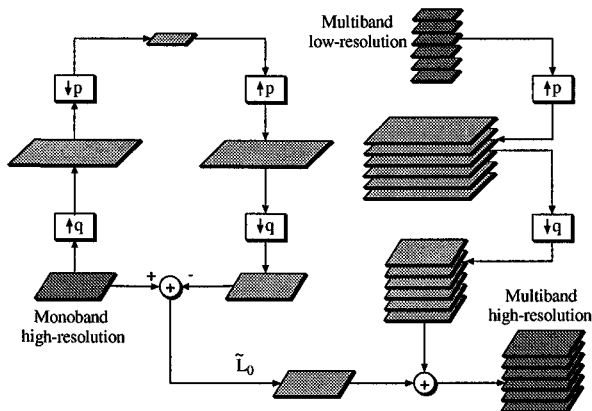


Fig. 2. Outline of data fusion procedure for two images with a p/q scale ratio.

Due to SNR constraints, TM Band 6 is actually sensed with a ground resolution of $120m/pel$ and resampled in order to match the size of the other bands ($30m/pel$). Figures 3(c) and (d) show fusion results of the two algorithms. The results of Fig. 3(c) have been obtained through the FIR implementation of a cubic spline WT [4,6]. Although the wavelet-fused image looks sharper, artifacts are perceivable around edges, due to ringing effects.

Spectral feature preservation is evaluated by taking the pixel differences between any of the merged images and a linearly resampled version of Band 6 (both integer valued). These differences are expected to be either zero or very small on homogeneous areas, and relevant on contours or highly textured areas. The standard deviation of such differences and the number of pixels in which

Table 1. Std. devs. (STD) of the differences obtained by subtracting merged images from expanded TM bands. Percentage of pixels ($P \pm 1$) whose absolute differences are equal to either one or zero. Results reported for wavelet (WT) and pyramid (GLP).

TM Band:	WT: STD	GLP: STD	WT: $P \pm 1$	GLP: $P \pm 1$
1	2.93	2.65	40.91 %	53.63 %
2	2.69	2.32	41.89 %	49.35 %
3	2.62	2.48	41.47 %	53.94 %
4	2.24	2.22	40.96 %	64.09 %
5	2.27	2.33	41.66 %	64.08 %
7	2.39	2.34	41.63 %	65.67 %

they are equal or very close to zero represent two figures of merit for image data fusion [3]. The former should be as close to zero as possible. Due to roundoff to integers, pixel differences are taken to be null if their absolute values do not

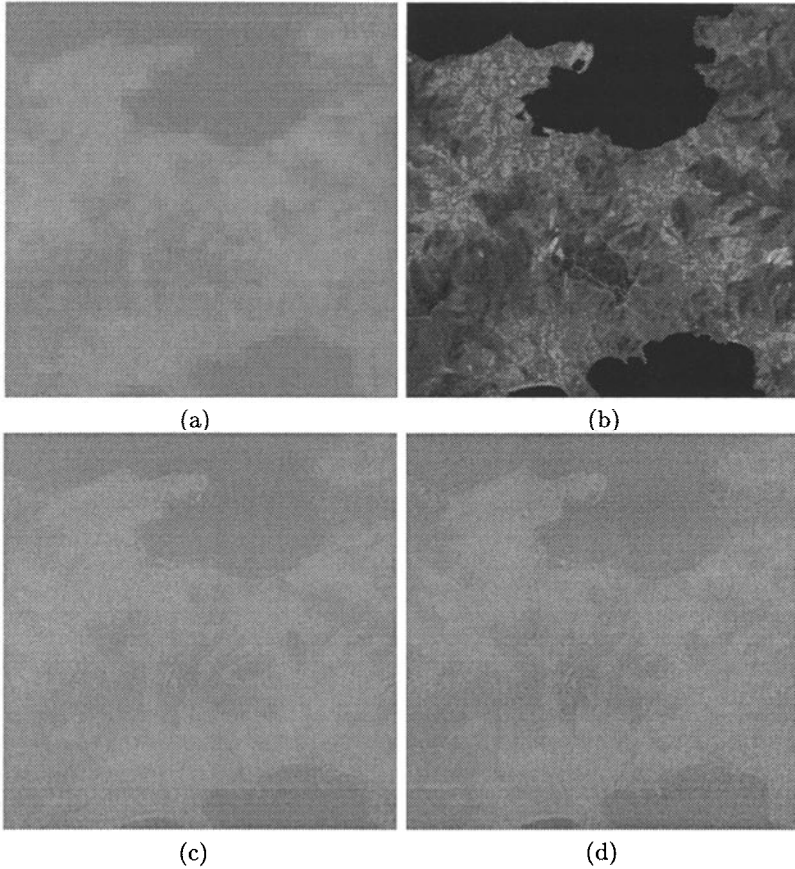


Fig. 3. Fusion of Landsat TM Band 6 (a) with Band 5 (b): (c) wavelet scheme; (d) pyramid scheme. Both images are 256×256 details.

exceed unity. Table 1 reports the scores of each TM band. It is apparent that the pixel percentages are far larger for the pyramid scheme. Standard deviations are slightly smaller for the pyramid, with the only exception of Band 5. The values of the parameters reported have been optimized over the pyramid-generating filter (15 taps) and are steady. The results of Table 1 are better than those reported in [3], thanks to the multi-resolution framework which allows a better filter design.

SPOT Panchromatic and Landsat TM data were available for the test area of Metaponto, in Southern Italy. The images were registered on the same cartographic base, each maintaining its own scale. The p/q ratio is 3. Original SPOT-PAN and TM Band-5 images are shown in Figure 4, together with the enhanced TM Band-5 version, in order to visually assess the quality of the results. Contours and textures are highlighted. The local average level is carefully preserved. Such a feature is important in determining spectral signatures, and its alteration may be responsible for misclassification and misinterpretation.

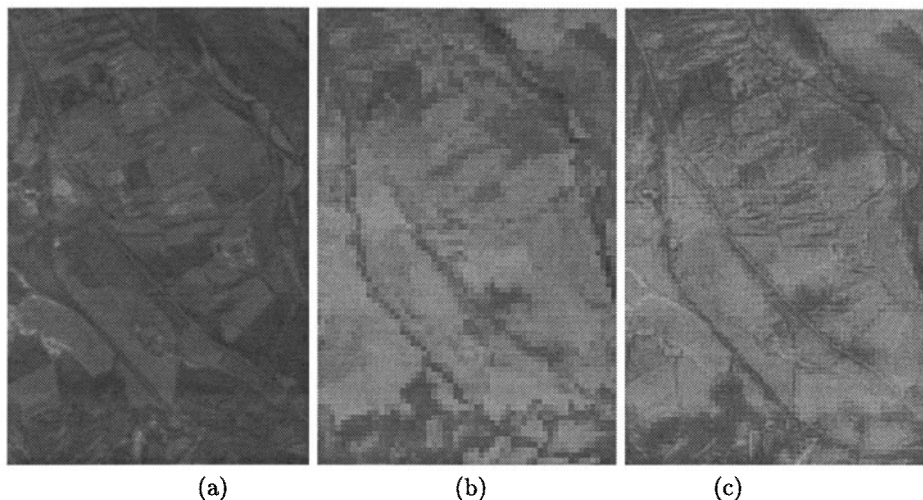


Fig. 4. 256×192 detail of the SPOT-PAN image (a), ground resolution $10m$, and TM-5 image (b) of the test area: resolution is $30m$ and a magnification by 3 has been applied for displaying. TM-5 image pyramid-fused with SPOT-PAN (c). Performance parameters, as defined in Table 1, are $STD = 4.05$ and $P \pm 1 = 36.44\%$.

Acknowledgments

This work was carried out under grants of ASI -Italian Space Agency- within a joint project on multisource classification, and of CNR -National Research Council- in the framework of the nationwide project on Cultural Heritage.

References

1. R. C. Luo and M. G. Kay, "Multisensor integration and fusion in intelligent systems," *IEEE Trans. Systems, Man, and Cybernetics*, **19**(5), 901-931 (1989).
2. H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *CVGIP: Graphical Models and Image Processing*, **57**(3), 235-245 (1995).
3. P. S. Chavez Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogram. Engin. Remote Sensing*, **57**(3), 295-303 (1991).
4. S. Mallat, "A Theory for Multiresolution Signal Decomposition: the Wavelet Representation," *IEEE Trans. Pattern Anal. Machine Intell.*, **11**(7), 674-693 (1989).
5. P. J. Burt, "The pyramid as a structure for efficient computation," in *Multiresolution Image Processing and Analysis*, A. Rosenfeld (Ed.), Berlin, Springer-Verlag (1984).
6. M. Unser and A. Aldroubi, "Polynomial Splines and Wavelets- A Signal Processing Perspective", in *Wavelets- A Tutorial in Theory and Applications*, C. K. Chui (Ed.), Academic Press, 91-122 (1992).
7. M. G. Kim, I. Dinstein, and L. Shaw, "A Prototype Filter Design Approach to Pyramid Generation," *IEEE Trans. Pattern Anal. Machine Intell.*, **15**(12), 1233-1240 (1993).