# GENERALISED LAPLACIAN PYRAMID-BASED FUSION OF MS + P IMAGE DATA WITH SPECTRAL DISTORTION MINIMISATION

B. Aiazzi[a], L. Alparone[b], S. Baronti[a], I. Pippi[a], M. Selva[a]

[a] Institute of Applied Physics "N. Carrara" IFAC–CNR, Florence, Italy - {aiazzi baronti pippi selva}@ifac.cnr.it
[b] Department of Electronics & Telecommunications, University of Florence, Florence, Italy - alparone@lci.det.unifi.it

**KEY WORDS:** Algorithms, Distortion, Fusion, High resolution, Multiresolution, Multispectral, Imagery, Urban.

**ABSTRACT:**

This work presents a general and formal solution to the problem of fusion of a multi-spectral (MS) image with a higher-resolution panchromatic (P) observation. The method relies on the *generalised Laplacian pyramid*, which is an over-sampled structure obtained by subtracting from an image its low-pass version. The goal is to selectively performs spatial-frequencies spectrum substitution from an image to another with the constraint of thoroughly retaining the spectral information of the coarser MS data. To this end, a vector injection model has been defined: at each pixel, the detail vector to be added is always parallel to the MS approximation. Furthermore, its components are scaled by factors measuring the ratio of local gains between the MS and P data. Quantitative results are presented and discussed on simulated SPOT 5 data of an urban area ($2.5m$ P, $10m$ XS) obtained from the MIVIS airborne imaging spectrometer.

## 1. INTRODUCTION

The ever increasing availability of space-borne sensors imaging in a variety of ground scales and spectral bands, makes fusion of multisensor data a discipline to which more and more general formal solutions to a number of application cases are demanded. Space-borne imaging sensors allow a global coverage of the Earth surface. Musti-spectral (MS) space observations, however, may exhibit limited ground resolutions, that may be inadequate to specific identification tasks.

Since the *high-pass filtering* (HPF) technique (Chavez *et al.*, 1991), fusion methods based on injecting high-frequency components into resampled versions of the MS data have demonstrated a superior performance (Wald *et al.*, 1997). HPF consists of an injection of high frequency components taken from a high-resolution panchromatic (P) observation into a bicubically resampled version of the low-resolution MS image. The frequency selection is obtained by taking the difference between the P image and its low-pass version achieved through a local pixel averaging, i.e. a box filtering. The rationale of spectrum substitution was formally developed in a multiresolution framework by employing the *discrete wavelet transform* (DWT) (Ranchin and Wald, 2000), and Laplacian pyramids (LP) (Aiazzi *et al.*, 2002).

According to he basic principle of DWT image fusion (Li *et al.*, 1995), couples of subbands of corresponding frequency content are merged. The fused image is synthesized by taking the inverse transform. Fusion based on the undecimated "a trous" wavelet was also proposed (Núñez *et al.*, 1999). Unlike the DWT, which is *critically sub-sampled*, the "à trous" wavelet and the LP are *over-sampled*. The LP can be *generalised* to deal with scales whose ratios are integer or fractional (GLP) (Aiazzi *et al.*, 1999).

Data fusion based on multiresolution analysis, however, requires the definition of a proper model establishing how the missing high-pass information to be injected into the MS bands is extracted from the P band (Ranchin and Wald, 2000). Such a model can be global over the whole image or depend on spatial context (Aiazzi *et al.*, 2002). Goal of the model is to make the fused bands the most similar to what the MS sensor would image if it had the same resolution as the broad-band one.

## 2. LAPLACIAN PYRAMIDS

The Laplacian pyramid (LP) is derived from the Gaussian pyramid (GP), which is a multi-scale representation obtained through a recursive *reduction* (low-pass filtering and decimation).

Let $G_0(i,j)$, $i = 0, \cdots, M-1$, and $j = 0, \cdots, N-1$, $M = u \times 2^K$, $N = v \times 2^K$, be a grey-scale image. The octave GP is defined as

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

for $k = 1, \cdots, K$, $i = 0, \cdots, M/2^k - 1$, and $j = 0, \cdots, N/2^k - 1$; in which $k$ identifies the level of the pyramid, $K$ being the top, or *root*, or *base-band approximation*, of size $u \times v$. The separable 2-D reduction *low-pass* filter stems from a linear symmetric 1-D kernel, generally odd-sized, i.e. $\{r_2(n),\ n = -L_r, \cdots, L_r\}$ which *should* have the $-3\ dB$ cut-off at one half of the band-width of the signal, to minimise the effects of *aliasing* (Vaidyanathan, 1992).

From the GP, the *enhanced* LP (ELP) (Aiazzi *et al.*, 1997) is defined, for $k = 0, \cdots, K-1$, as

$$L_k(i,j) \triangleq G_k(i,j) - \mathrm{expand}_2[G_{k+1}](i,j) \quad (2)$$

in which $expand_2[G_{k+1}]$ denotes the $(k+1)$st GP level expanded by 2 to match the underlying $k$th level:

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

for $i = 0, \cdots, M/2^k - 1$, $j = 0, \cdots, N/2^k - 1$, and $k = 0, \cdots, K-1$. The 2-D low-pass filter for expansion is given as the outer product of a linear symmetric odd-sized kernel $\{e_2(n),\ n = -L_e, \cdots, L_e\}$, which *must* cut off at one half

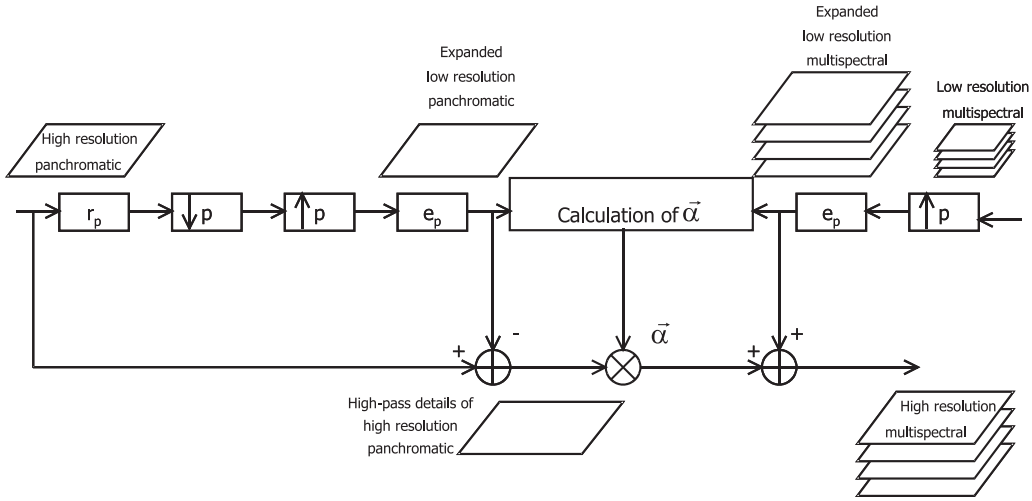


Figure 1. Flowchart of GLP fusion suitable for MS and P data, whose scale ratio is $p > 1$. $\downarrow p$: *down-sampling* by $p$; $\uparrow p$: *up-sampling* by $p$; $r_p$ is $p$-reduction low-pass filter with cut-off at $1/p$ of spectrum extent; $e_p$ is $p$-expansion low-pass filter with $1/p$ cut-off as well.

of the band-width of the signal to reject the *spectral images* introduced by up-sampling by 2 ($\uparrow$ 2) (Vaidyanathan, 1992). Summation terms are taken to be null for non-integer values of $(i+m)/2$ and $(j+n)/2$, corresponding to interleaving zeroes. The base-band approximation is added to the band-pass ELP, i.e $L_K(m,n) \equiv G_K(m,n)$, to yield a complete image description.

The attribute *enhanced* depends on the expansion filter being forced to be *half-band*, i.e. an interpolator by 2, and chosen independently of the reduction filter, which may be *half-band* as well, or not. The ELP outperforms the standard LP for image compression (Aiazzi *et al.*, 1997), thanks to the fact ythat its layers are uncorrelated with one another. The filter choice stems from a trade-off between selectivity (sharp cut-off) and computational cost (Aiazzi *et al.*, 2002). In particular, the absence of ripple, is the most favourable feature. The ELP can be easily *generalised* to deal with scales whose ratios are integer or even fractional numbers (GLP) (Aiazzi *et al.*, 1999).

## 3. PYRAMID-BASED FUSION WITH SPECTRAL DISTORTION MINIMISATION

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

$$SAM(\mathbf{v}, \hat{\mathbf{v}}) \triangleq \arccos\left(\frac{< \mathbf{v}, \hat{\mathbf{v}} >}{||\mathbf{v}||_2 \cdot ||\hat{\mathbf{v}}||_2}\right) \tag{4}$$

The SAM distortion (4) can be measured in degrees or radians.

Fig. 1 shows the flowchart of a GLP-based scheme suitable for spectral distortion-minimising fusion of MS + P data, whose scale ratio is an integer $p$. Let $G^{(P)}(i,j)$ be the data set constituted by a single P image having smaller scale, i.e. finer resolution, and size $Mp \times Np$. Let also $\{G^{(l)}(i,j),\ l = 1, \cdots, L\}$ be the data set made up of the $L$ bands of an MS image. Such bands have scale larger by a factor $p$, i.e. coarser resolution, and thus size $M \times N$. The goal is to obtain a set $\{\hat{G}^{(l)}(i,j),\ l = 1, \cdots, L\}$ of MS bands each having same spatial resolution as P. The upgrade of each band $G^{(l)}$ to yield the spatial resolution of P is the level $k = 0$ of the zero-mean GLP of the P image, i.e. $L_0^{(P)}$. First, the bands $\{G^{(l)}(i,j),\ l = 1, \cdots, L\}$ are interpolated by $p$ to match the finer scale. A new data set, $\{\tilde{G}^{(l)},\ l = 1, \cdots, L\}$, is thus produced. They constitute the low-pass component to which details are added in order to yield a spatially enhanced set of MS observations. Then, the high-pass component from P, $L_0^{(P)}(i,j)$, is weighted by a scaling factor and added to $\{\tilde{G}^{(l)}(i,j),\ l = 1, \cdots, L\}$ to yield $\{\hat{G}^{(l)}(i,j),\ l = 1, \cdots, L\}$.

Crucial point is the definition of a local gain (LG), by which high-pass details at pixel $(i,j)$ are to be weighted before being injected into the resampled multi-spectral bands. Such a gain is chosen to be both space- and spectrally-varying; stated with a vector notation, $\vec{\alpha}(i,j) = \{\alpha_l(i,j),\ l = 1, \cdots, L\}$. Let $\vec{\tilde{G}}(i,j) \triangleq \{\tilde{G}^{(l)}(i,j),\ l = 1, \cdots, L\}$ denote the pixel vector of the expanded MS image; let also $\vec{D}(i,j) \triangleq \vec{\alpha}(i,j) \cdot L_0^{(P)}(i,j)$ 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 $\vec{D}^{(l)}(i,j)$ is injected into $\tilde{G}^{(l)}(i,j)$. Starting from the vector merge relationship

$$\vec{\hat{G}}(i,j) = \vec{\tilde{G}}(i,j) + \vec{\alpha}(i,j) \cdot L_0^{(P)}(i,j) \tag{5}$$

let us define the *l*th components of LG as

$$\alpha_l(i,j) = \frac{\tilde{G}^{(l)}(i,j)}{\tilde{G}_1^{(P)}(i,j)},\ \ l = 1, \cdots, L \tag{6}$$

in which $\tilde{G}_1^{(P)}(i,j)$ denotes the expanded version of the reduced P image. ¿From (5) and (6) it stems that

$$\begin{aligned} \vec{\hat{G}} \times \vec{\tilde{G}} &= \vec{\tilde{G}} \times \left[\vec{\tilde{G}} + \vec{D}\right] = \vec{\tilde{G}} \times \left[\vec{\tilde{G}} + \vec{\alpha} \cdot L_0^{(P)}\right] \\ &= \vec{\tilde{G}} \times \vec{\tilde{G}} \cdot \left[1 + \frac{L_0^{(P)}}{\tilde{G}_1^{(P)}}\right] = 0 \end{aligned} \tag{7}$$

in which $\times$ stands for vector product and the indexes $(i, j)$ have been omitted throughout. Eq. (7) 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, $\vec{\hat{G}}(i, j)$, because the upgrade $\vec{D}(i, j)$ is always parallel to $\vec{\tilde{G}}(i, j)$.

Although one level of decomposition ($K = 1$) with $p = 4$ is capable to produce $1 : 4$ fusion, for computational convenience, $K = 2$ and $p = 2$ are preferable, since less data are processed at the second level thanks to decimation after the first one.

## 4. RESULTS AND DISCUSSION

A MIVIS hyper-spectral image portraying the urban areas of Viareggio, in Italy, was available in $102$ spectral bands at $2.5m$ resolution. The MIVIS bands were used to synthesize the XS bands of SPOT 5: B1=($500 - 590nm$), B2=($610 - 700nm$) and B3=($790 - 890nm$), as well as an overlapped P band, all at a $2.5m$ resolution. The simulated XS bands were low-pass filtered and decimated by $4$, to yield $10m$ XS, and used, together with the P at $2.5m$, to synthesize back $2.5m$ B1, B2 and B3.

The reasons underlying such a procedure is twofold: all the images are spatially registered, being acquired simultaneously from the airborne platform. Thus, no geometric corrections, which are likely to affect the results of data fusion, need to be preliminarily carried out. Secondly, the true XS data at $2.5m$ are available for objective comparisons, which are particularly significant for analyses of urban areas (Couloigner *et al.*, 1998; Terretaz, 1998).

Fig. 2(a)-(f) shows $2.5m$ P, $10m$ XS expanded to $2.5m$ scale by means of the same pyramid filter (23 coefficients) and its spatially enhanced versions achieved by means of the proposed GLP-based fusion scheme with spectral distortion minimisation (GLP-SDM), and of the former GLP-base fusion scheme with context-based decision (GLP-CBD) (Aiazzi *et al.*, 2002), working with decision threshold $\theta = 0.3$ and local statistics calculated on $9 \times 9$ windows. The HPF-fused version ($5 \times 5$ box filtering) and the original $2.5m$ XS are also reported for reference. As it appears, apart from HPF (see Fig. 2(e)), all the fused images are hardly distinguishable from the original. HPF reveals heavy spatial, spectral and radiometric distortions. Since HPF relies on undecimated multiresolution analysis, its poor performance is due to the *box* filter having little frequency selection, as well as to the absence of an injection model.

| $1:4$ | GLP-SDM | GLP-CBD | HPF | EXP |
|---|---|---|---|---|
| B1 | 0.990 | 0.982 | 0.958 | 0.875 |
| B2 | 0.994 | 0.988 | 0.970 | 0.860 |
| B3 | 0.973 | 0.959 | 0.937 | 0.834 |

Table 1. CC between $2.5m$ XS images and those obtained from $10m$ XS by means of $1 : 4$ fusion with $2.5m$ P. EXP denotes plain resampling without detail injection.

| $1:4$ | GLP-SDM | GLP-CBD | HPF | EXP |
|---|---|---|---|---|
| RMSE | 4.93 | 5.97 | 12.21 | 15.72 |
| SAM | $3.19^\circ$ | $3.96^\circ$ | $4.66^\circ$ | $3.19^\circ$ |

Table 2. Average radiometric and spectral distortions between $2.5m$ XS spectral vectors and those obtained from fusion of $10m$ XS with $2.5m$ P. EXP = resampling without injection.

Table 1 reports correlation coefficients (CC) between $2.5m$ reference originals and fused XS bands. This parameter measures how the shape of the fused image reflects that of the original. CC, however, is insensitive to a constant gain and bias between the images. Although GLP-SDM exploits a vector model designed to minimise SAM, the CC of each spectral component are higher than those of the other methods. Table 2 reports RMSE and SAM (4) between $2.5m$ original and fused pixel vectors, averaged over the XS bands. Notice that spectral distortion minimisation is achieved with a radiometric distortion lower by about $20\%$ with respect to GLP-fusion without SAM adjustment.

## REFERENCES


Aiazzi, B., L. Alparone, F. Argenti & S. Baronti, 1999. Wavelet and pyramid techniques for multisensor data fusion: a performance comparison varying with scale ratios. In: Serpico, S. B. (Ed.), Image and Signal Processing for Remote Sensing V., Proc. SPIE Vol. 3871, EUROPTO Series, pp. 251–262.

Aiazzi, B., L. Alparone, S. Baronti & F. Lotti, 1997. Lossless image compression by quantization feedback in a content-driven enhanced Laplacian pyramid. *IEEE Transactions on Image Processing*, 6(6), pp. 831–843.

Aiazzi, B., L. Alparone, S. Baronti & A. Garzelli, 2002. Context-driven fusion of high spatial and spectral resolution data based on oversampled multiresolution analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 40(9).

Couloigner, I., T. Ranchin, V. P. Valtonen & L. Wald, 1998. Benefit of the future SPOT-5 and of data fusion to urban road mapping. International Journal of Remote Sensing, 19(8), pp. 1519–1532.

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

Li, H., B. S. Manjunath & S. K. Mitra, 1995. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3), pp. 235–245.

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

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

Terretaz, P., 1998. Comparison of different methods to merge SPOT P and XS data: evaluation in an urban area. In: Gudmandsen, P. (Ed.), Future Trends in Remote Sensing. Balkema, Rotterdam, The Netherlands, pp. 435–443.

Vaidyanathan, P. P., 1992. *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, NJ.

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

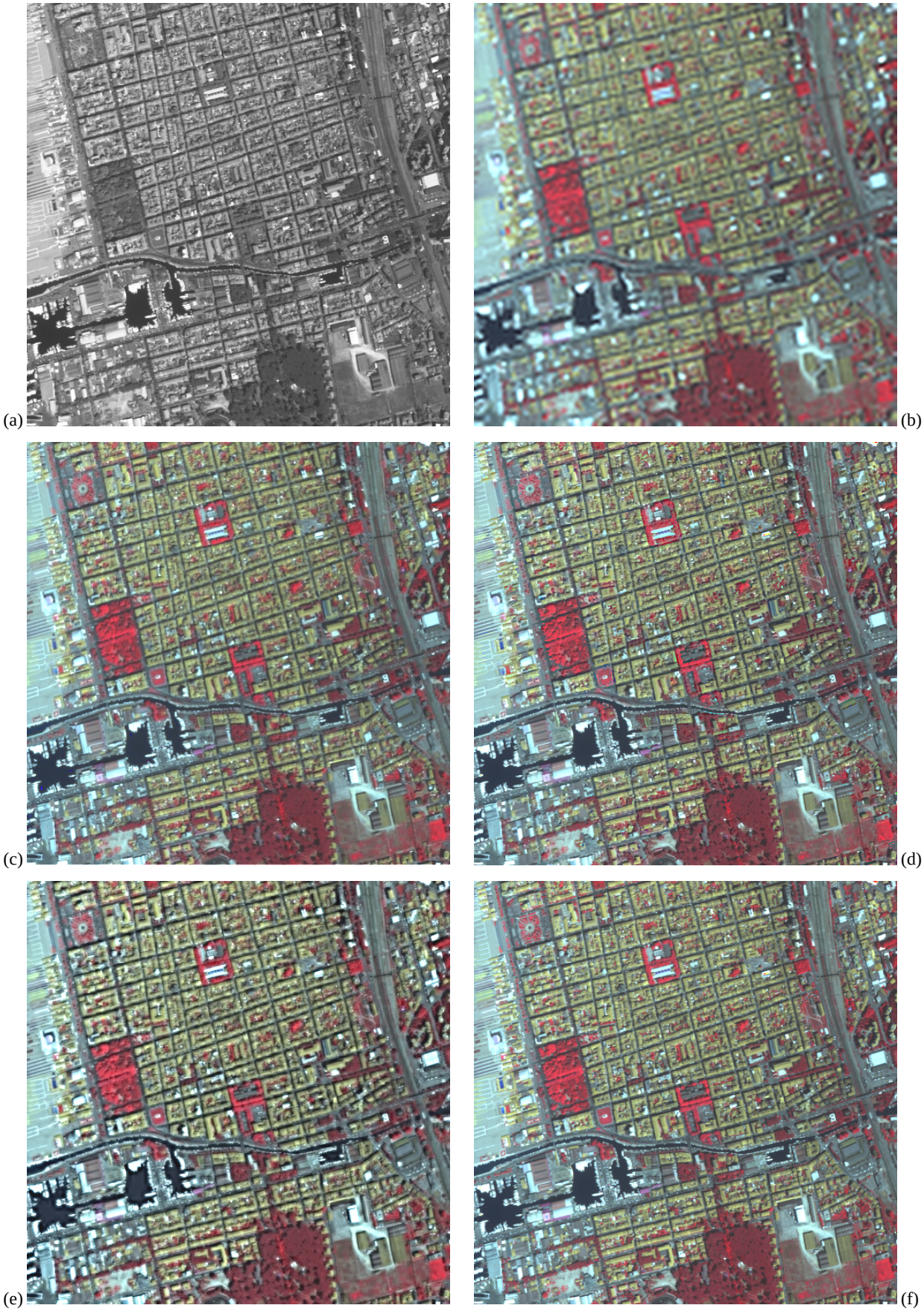

Figure 2. (a): $512 \times 512$ P channel of SPOT 5 ($2.5m$) synthesized as (B1+B2+B3)/3; (b): $10m$ XS expanded by 4 (B3-B2-B1 as R-G-B); (c): $1:4$ GLP fusion with spectral distortion minimization (GLP-SDM); (d): $1:4$ GLP fusion with context-based decision (GLP-CBD) ($\theta = 0.3$); (e): HPF fusion ($5 \times 5$ box filter); (f): true $2.5m$ XS.