



PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE @ DIRECT®

Pattern Recognition 37 (2004) 1855–1872

**PATTERN
RECOGNITION**

THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

A wavelet-based image fusion tutorial

Gonzalo Pajares*, Jesús Manuel de la Cruz

Dpto. Arquitectura de Computadores y Automática, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

Received 27 November 2002; received in revised form 22 March 2004; accepted 22 March 2004

Abstract

The objective of image fusion is to combine information from multiple images of the same scene. The result of image fusion is a new image which is more suitable for human and machine perception or further image-processing tasks such as segmentation, feature extraction and object recognition. Different fusion methods have been proposed in literature, including multiresolution analysis. This paper is an image fusion tutorial based on wavelet decomposition, i.e. a multiresolution image fusion approach. We can fuse images with the same or different resolution level, i.e. range sensing, visual CCD, infrared, thermal or medical. The tutorial performs a synthesis between the multiscale-decomposition-based image approach (Proc. IEEE 87 (8) (1999) 1315), the ARSIS concept (Photogramm. Eng. Remote Sensing 66 (1) (2000) 49) and a multisensor scheme (Graphical Models Image Process. 57 (3) (1995) 235). Some image fusion examples illustrate the proposed fusion approach. A comparative analysis is carried out against classical existing strategies, including those of multiresolution.

© 2004 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Image fusion; Wavelets; Multiresolution

1. Introduction

With the availability of multisensor data in many fields, such as remote sensing, medical imaging or machine vision, sensor fusion has emerged as a new and promising research area. It is possible to have several images of the same scene providing different information although the scene is the same. This is because each image has been captured with a different sensor. If we are able to merge the different information to obtain a new and improved image, we have a fused image and the method is called a fusion scheme. The following three examples clarify these assertions; a detailed explanation of such examples is given in Section 4:

(1) Due to the limited depth-of-focus of optical lenses in CCD devices, it is often not possible to get an image that contains all relevant objects “in focus”. To achieve all objects “in focus”, a fusion process is required so that all focused objects are selected.

(2) In the remote sensing field, the color information is provided by three sensors covering the red, green and blue spectral wavelengths. These sensors have a low number of pixels (low spatial resolution) and the small objects and details (cars, small lines, etc.) are hidden. Such small objects and details can be observed with a different sensor (panchromatic), which have a high number of pixels (high spatial resolution) but without color information. With a fusion process a unique image can be achieved containing both: high spatial resolution and color information.

(3) In medical imaging, we can have a positron emission tomography and a magnetic resonance images from the brain of the same patient. The first is a functional image displaying the brain activity, but without anatomical information. On the contrary, the second provides anatomical information but without functional activity. Moreover, although the two images come exactly from the same brain area, the positron emission tomography has less pixels than the magnetic resonance, i.e. we say that the first has less spatial resolution than the second. The goal of fusion scheme for the proposed example is

* Corresponding author. Tel.: +34-1-3-94-44-77; fax: +34-1-3-94-46-87.

E-mail address: pajares@dacya.ucm.es (G. Pajares).

to achieve a unique image with functional and anatomical information and with the best resolution.

In general, the problem that image fusion tries to solve is to combine information from several images (sensors) taken from the same scene in order to achieve a new fused image, which contains the best information coming from the original images. Hence, the fused image has better quality than any of the original images.

The wavelets-based approach is appropriate for performing fusion tasks for the following reasons:

- (1) It is a multiscale (multiresolution) approach well suited to manage the different image resolutions. In recent years, some researchers [4–8] have studied multiscale representation (pyramid decomposition) of a signal and have established that multiscale information can be useful in a number of image processing applications including the image fusion.
- (2) The *discrete wavelets transform* (DWT) allows the image decomposition in different kinds of coefficients preserving the image information.
- (3) Such coefficients coming from different images can be appropriately combined to obtain new coefficients, so that the information in the original images is collected appropriately.
- (4) Once the coefficients are merged, the final fused image is achieved through the *inverse discrete wavelets transform* (IDWT), where the information in the merged coefficients is also preserved.

The key step in image fusion based on wavelets is that of coefficient combination, namely, the process of merge the coefficients in an appropriate way in order to obtain the best quality in the fused image. This can be achieved by a set of strategies. The most simple is to take the average of the coefficients to be merged, but there are other merging strategies with better performances, which are exhaustively studied in Section 3 of this paper.

Three previous wavelets-based works: Zhang and Blum (ZB) [1], Ranchin and Wald (RW) [2] and Li et al. (LMM) [3] are the basis for this tutorial. ZB establishes a categorization of multiscale decomposition (MSD), RW introduces the ARSIS (Amélioration de la Résolution Spatiale par Injection de Structures) concept, making use of a multiscale method for the description and modeling of the missing information between images to be fused. LMM performs extensive experiments with several sets of images including the fusion of multifocus images. In ZB, the objective of the fusion is to achieve a high-quality digital camera image from several degraded images. In RW, the goal of the fusion is to achieve high spatial resolution together with a high-quality spectral content from two kinds of remote sensing images: (1) images with high quality in the spectral content but low quality in the spatial resolution and (2) images with high spatial resolution but with a unique spectral band. This is also the objective in Garguet-Dupont et al. (GGCP) [9].

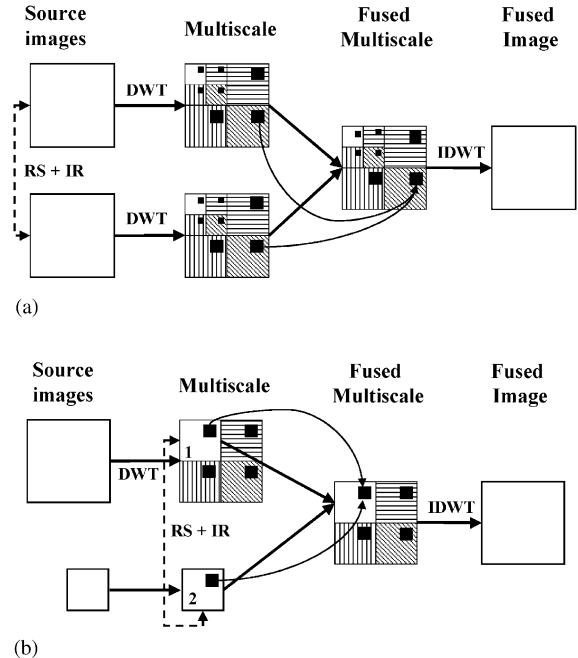


Fig. 1. Block diagrams of generic fusion schemes where the input images have identical (a) and different resolutions (b).

Fig. 1 illustrates two diagrams for generic MSD approaches. In Fig. 1(a) the source images must have identical spatial resolutions. Hence, if their resolutions are different, an image resampling (RS) followed by an image registration (IR) strategies are previously required (details about them are given below). The DWT is applied to both images and a decomposition of each original image is achieved. This is represented in the multiscale illustration where different bars (horizontal, vertical, diagonal and none) represent also different coefficients (as we will see later). There are two decomposition levels, as it is shown in the left upper subimage (this will be also detailed later). The different black boxes, associated to each decomposition level, are coefficient corresponding to the same image spatial representation in each original image, i.e. the same pixel or pixels positions in the original images. Only coefficients of the same level and representation are to be fused, so that the fused multiscale coefficients can be obtained. This is displayed in the diagonal details where the curved arrows indicate that both coefficients are merged to obtain the new fused multiscale coefficient. This is applicable to the remainder coefficients. Once the fused multiscale is obtained, through the IDWT, the final fused image is achieved. In Fig. 1(b) there are two sources images with different resolution levels, the DWT is only applied to the image with the higher spatial resolution. We obtain a multiscale image representation for such image. The image with the smaller resolution is not transformed. At this stage if subimage 1 and image 2 have different

spatial resolutions, as before, a previous RS followed by an IR strategies are required for such images. Now, only a unique type of coefficients belonging to the multiscale representation of the higher-resolution image and the original pixels of the smaller image are to be fused (see curved arrows). A fused multiscale representation is obtained and as before, through the IDWT the final fused image is achieved.

Image resampling (RS): As required by the wavelet transform, the coefficients must be merged or superimposed when the images are in the same scale. This means that the images must be re-scaled when their scales do not match. For example, one of the first steps for registration SPECT with MRI or CT images is to expand the 64×64 SPECT image to the 256×256 or even to 512×512 matrix, the usual size of the MRI or CT images, respectively. This is carried out by a well-known interpolation technique (nearest neighbor, bilinear, bicubic, etc.).

Image registration (IR): In image fusion, it is essential that the image information from all the constituent images be adequately aligned and registered prior to combining the images, ensuring that the information from each sensor is referring to the same physical structures in the environment [3]. This is a key issue in image fusion, as a misalignment produces severe edge artifacts in the combined images. This is particularly significant in images where the edges are abundant. There are several approaches that have been investigated for alignment of multiple digital images [10–13]. For the task of IR, a set of control points can be derived from the extraction of features in the image [12].

This paper is organized as follows. In Section 2, we give details about the DWT and IDWT for achieving the different multiresolution levels required for the fusion process and to obtain the final fused image, respectively. We include some pedagogical examples. In Section 3, we describe different methods for merging the coefficients obtained during the DWT decomposition process. In Section 4, we illustrate the fusion scheme with some real examples. In Section 5, a comparative analysis is carried out of the different wavelets families and classical fusion methods. Finally, in Section 6, the conclusions are presented.

2. The wavelet transform: a review

We start this section by introducing the specific concepts related to the wavelet transform, so that the reader can understand the basic concepts associated with this transform. We start the development based on the Haar wavelet transform as the simplest representation and then we extend the concepts to more complex representations. We include some pedagogical examples in order to enhance understanding of the wavelet transform for image multiresolution level.

2.1. One-dimensional (1-D) wavelet transform: from Haar to more complex representations

According to Ref. [14] we can think of images as piecewise-constant functions on the half-open interval $[0, 1)$. To do so, the *vector space* concept from linear algebra is introduced. A one-pixel image is just a function that is constant over the entire interval $[0, 1)$. Let V^0 be the vector space of all these functions. A two-pixel image has two constant pieces over the intervals $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$. The space containing all these functions is called V^1 . If we continue in this manner, the space V^j will include all piecewise constant functions defined on the interval $[0, 1)$ with constant pieces over each of the 2^j equal subintervals. Now every 1-D image with 2^j is a vector or an element in V^j . As these vectors are all functions defined on the unit interval, every vector in V^j is also contained in V^{j+1} . Thus, the spaces V^j are nested, that is

$$V^0 \subset V^1 \subset V^2 \subset \dots \tag{1}$$

Now we need to define a basis for the spaces V^j . The basis functions for the spaces V^j are called *scaling* functions $\phi(x)$. A simple basis for V^j is given by the set of scaled and translated functions

$$\begin{aligned} \phi_k^j(x) &= \phi(2^j x - k), \quad k = 0, \dots, 2^j - 1 \\ \text{where } \phi(x) &= \begin{cases} 1 & \text{for } 0 \leq x < 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

The next step is to choose an inner product defined on the vector spaces V^j such as the standard inner product. We can now define a new vector space W^j as the orthogonal complement of V^j in V^{j+1} . That is, W^j is the space of all functions in V^{j+1} that are orthogonal to all functions in V^j under the chosen inner product. Informally, W^j contains the details in V^{j+1} that cannot be represented in V^j . A collection of linearly independent functions $\psi_k^j(x)$ spanning W^j are called *wavelets*. These basis functions have two important properties:

- the basis functions ψ_k^j of W^j , together with the basis functions ϕ_k^j of V^j , form a basis for V^{j+1} ,
- every basis function ψ_k^j of W^j is orthogonal to every basis function ϕ_k^j of V^j under the chosen inner product.

The wavelets defined in Refs. [15,16,13] are

$$\psi_k^j(x) = a^{-j/2} \psi(a^{-j}(x - kba^j)). \tag{3}$$

The Haar wavelet transform is given by $a = 2$ and $b = 1$. The compact support of ψ_k^j is then $[2^j k, 2^j(k + 1)]$.

Haar described the following function as one that provides an orthonormal basis. The wavelet analysis of a continuous variable is a step function

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2}, \\ -1 & \text{if } \frac{1}{2} \leq x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Since $\psi_k^j(x)$ forms an orthonormal set, the wavelet coefficients of the signal $f(x)$ can be calculated by the inner product [17,18]

$$a_k^j = \langle f(x), \psi_k^j(x) \rangle = \int f(x) \psi_k^j(x) dx. \tag{5}$$

For the wavelet expansion, signal $f(x)$ can be reconstructed via

$$f(x) = \langle a_k^j, \psi_k^j(x) \rangle = \sum_{k,j} a_k^j \psi_k^j(x). \tag{6}$$

The multiresolution formulation needs two closely related basic functions. In addition to the mother wavelet $\psi(x)$, we will need another basic function, called the scaling function $\phi(x)$. $\phi(x)$ can be expressed in terms of a weighted sum of shifted $\phi(2x)$ as [16]

$$\phi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} l(n) \phi(2x - n), \tag{7}$$

where $l(n)$'s are the scaling (lowpass) coefficients, and the $\sqrt{2}$ maintains the norm of the scaling function with the scale of two. The scaling coefficients $l(n)$ must satisfy

$$\sum_{n \in \mathbb{Z}} l(n) = \sqrt{2} \quad \text{and} \tag{8}$$

$$\sum_{n \in \mathbb{Z}} l(n) l(n - 2k) = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

The dilation equation (7) is fundamental to the theory of scaling functions. The mother wavelet $\psi(x)$ is related to the scaling function via

$$\psi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} h(n) \phi(2x - n), \tag{9}$$

where $h(n)$'s are the wavelet (highpass) coefficients. They are required by orthogonality to be related to the scaling coefficients by $h(n) = (-1)^n l(1 - n)$.

The proofs for the above equations can be found in Refs. [15,16].

The mother wavelet $\psi(x)$ is good at representing the detail and high-frequency parts of a signal. The scaling function $\phi(x)$ is good at representing the smooth and low-frequency parts of the signal.

In most practical applications, one never explicitly calculates the scaling function $\phi(x)$ and wavelet $\psi(x)$, but performs the transform using the scaling coefficients $l(n)$ and the wavelet coefficients $h(n)$. In forward wavelet analysis, a J -level discrete decomposition can be written as

$$f(x) = \sum_{n \in \mathbb{Z}} c_n^0 \phi(x - n) = \sum_k c_k^J \phi_k^J(x) + \sum_{j=1}^J \sum_k d_k^j \psi_k^j(x), \tag{10}$$

where coefficients c_n^0 are given, and coefficients c_k^j and d_k^j at resolution j are related to the coefficients c_k^{j-1} at level $j - 1$

by the following recursive equations [17,19]:

$$c_k^j = \sum_{n \in \mathbb{Z}} c_k^{j-1} l(n - 2k) \quad \text{and} \tag{11}$$

$$d_k^j = \sum_{n \in \mathbb{Z}} c_k^{j-1} h(n - 2k)$$

for $j = 1, 2, \dots, J$.

In the equation expansion (10), the first summation gives a function that is a low resolution or coarse approximation of $f(x)$, which represents the smooth part of $f(x)$. For each increasing level j in the second summation a higher or fine-resolution function is added, which represents the detail part of $f(x)$. Eq. (11) shows that the scaling and wavelet coefficients at different levels of scale can be obtained by convolving the expansion coefficients at scale $j - 1$ by the time-reversed recursion coefficients $l(-n)$ and $h(-n)$, then down-sampling (taking every other term) to give the expansion coefficients at the next level of j .

In backward wavelet synthesis, a reconstruction of the original fine scale coefficients of the signal can be made from a combination of the scaling coefficients and wavelet coefficients at a coarse resolution. Because all of these functions are orthonormal, we have

$$c_k^j = \sum_{n \in \mathbb{Z}} c_n^{j+1} l(k - 2n) + \sum_{n \in \mathbb{Z}} d_n^{j+1} h(k - 2n). \tag{12}$$

The synthesis operation of Eq. (12) is equivalent to upsampling the coefficients c_n^{j+1} and d_n^{j+1} (inserting a zero between each of the original terms) in the coarser level $j + 1$, and then convolving with $l(n)$ and $h(n)$, individually, to obtain the scaling coefficients in the finer level j . The synthesis process can be recursively continued to the original level. The analysis and synthesis procedures lead to the pyramid-structured wavelet decomposition [20] which is explained in the next section.

2.2. The 2-D wavelet transform implementation for multiresolution decomposition images

The analysis and synthesis procedures lead to the pyramid-structured wavelet decomposition [20]. Once again, following Ref. [14], the 1-D multiresolution wavelet decomposition can be easily extended to two dimensions by introducing separable 2-D scaling and wavelet functions as the tensor products of their 1-D complements. Hence, we obtain

$$\begin{aligned} \phi_{LL}(x, y) &= \phi(x)\phi(y), & \psi_{LH}(x, y) &= \phi(x)\psi(y), \\ \psi_{HL}(x, y) &= \psi(x)\phi(y), & \psi_{HH}(x, y) &= \psi(x)\psi(y). \end{aligned} \tag{13}$$

The 2-D wavelet analysis operation consists in filtering and down-sampling horizontally using the 1-D lowpass filter L (with impulse responses $l(i)$) and highpass filter H (with impulse responses $h(j)$) to each row in the image $I(x, y)$,

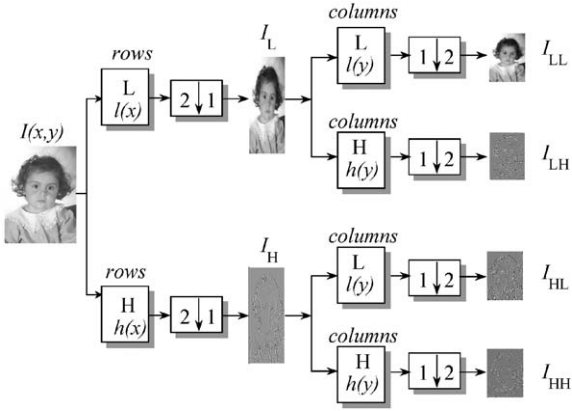


Fig. 2. One stage of 2-D DWT multiresolution image decomposition (forward wavelet analysis).

producing the coefficient matrices $I_L(x, y)$ and $I_H(x, y)$. Vertically filtering and down-sampling follows, using the lowpass and highpass filters L and H to each column in $I_L(x, y)$ and $I_H(x, y)$ and produces four subimages $I_{LL}(x, y)$, $I_{LH}(x, y)$, $I_{HL}(x, y)$ and $I_{HH}(x, y)$ for one level of decomposition. $I_{LL}(x, y)$ is a smooth subimage corresponding to the low-frequency band of the MSD and can be considered as a smoothed and subsampled version of the original image $I(x, y)$, i.e. it represents the coarse approximation of $I(x, y)$. $I_{LH}(x, y)$, $I_{HL}(x, y)$ and $I_{HH}(x, y)$ are detail subimages, which represent the horizontal, vertical and diagonal directions of the image $I(x, y)$.

Fig. 2 depicts one stage in a multiresolution pyramid decomposition of the input image $I(x, y)$, where the different symbols are explained in Fig. 5. In order to illustrate the examples of this section, we have used the Haar wavelet transform, although any other set of wavelets could be used. Hence, $L \equiv (1/\sqrt{2})[1, 1]$ and $H \equiv (1/\sqrt{2})[1, -1]$.

The detailed 2-D pyramid decomposition algorithm, can be expressed as follows: Let $I(x, y)$ be the original image of size $M \times N$, $l(i)$ the analysis lowpass coefficients of a specific wavelet basis, $i = 0, 1, 2, \dots, N_l - 1$, where N_l is the support length of the filter L , $h(j)$ the analysis highpass coefficients of a specific wavelet basis, $j = 0, 1, 2, \dots, N_h - 1$, where N_h is the support length of the filter H . Then,

$$\begin{aligned}
 I_L(x, y) &= \frac{1}{N_l} \sum_{i=0}^{N_l-1} l(i) \cdot I((2x + i) \bmod M, y), \quad I_H(x, y) \\
 &= \frac{1}{N_h} \sum_{j=0}^{N_h-1} h(j) \cdot I((2x + j) \bmod M, y) \quad (14)
 \end{aligned}$$

for $x = 0, 1, 2, \dots, M/2 - 1$ and $y = 0, 1, 2, \dots, N - 1$.

$$I_{LL}(x, y) = \frac{1}{N_l} \sum_{i=0}^{N_l-1} l(i) \cdot I_L(x, (2y + i) \bmod N), \quad I_{LH}(x, y)$$

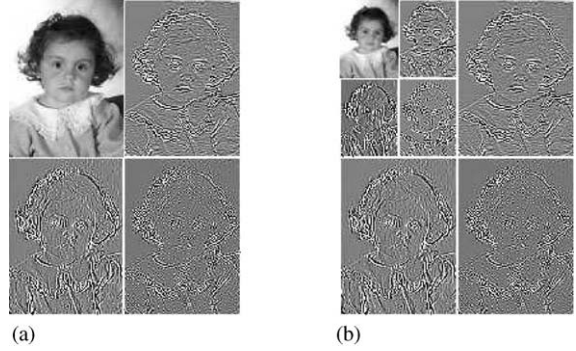


Fig. 3. A representation of (a) one-level and (b) two-level image decomposition.

$$= \frac{1}{N_h} \sum_{j=0}^{N_h-1} h(j) \cdot I_L(x, (2y + j) \bmod N), \quad (15)$$

$$\begin{aligned}
 I_{HL}(x, y) &= \frac{1}{N_l} \sum_{i=0}^{N_l-1} l(i) \cdot I_H(x, (2y + i) \bmod N), \quad I_{HH}(x, y) \\
 &= \frac{1}{N_h} \sum_{j=0}^{N_h-1} h(j) \cdot I_H(x, (2y + j) \bmod N) \quad (16)
 \end{aligned}$$

for $x = 0, 1, 2, \dots, M/2 - 1$ and $y = 0, 1, 2, \dots, N/2 - 1$.

The 2-D pyramid algorithm can iterate on the smooth subimage $I_{LL}(x, y)$ to obtain four coefficient matrices in the next decomposition level and so on. This is illustrated in Figs. 3(a) and (b) which correspond to one- and two-level image decompositions, respectively.

Some wavelet-based applications do not require all coefficients, only the most relevant. So an additional procedure can be carried out to eliminate non-significant coefficients by thresholding, since these have a magnitude close to zero. After thresholding, only the desired coefficients remain. The threshold value can be chosen as in [21], $T = \sigma\sqrt{2 \log n}/\sqrt{n}$ where σ is the standard deviation of the coefficients and n is the total size of samples. Another possibility is to fix T in order to replace a percentage of the coefficients with the smallest magnitude to zero. Obviously, the cancellation of coefficients implies a loss of information.

The inverse 2-D wavelet transform can be implemented using a backward 2-D pyramid algorithm. The 2-D wavelet synthesis operation consists in up-sampling and filtering vertically using the 1-D synthesis lowpass filter \tilde{L} (with impulse responses $\tilde{l}(i)$) and highpass filter \tilde{H} (with impulse responses $\tilde{h}(j)$) for each column in the subimage. Horizontal up-sampling and filtering then follows, using the lowpass \tilde{L} and highpass filter \tilde{H} , for each row of the reversed image. Fig. 4 shows one stage in a wavelet reconstruction.

The notation used in Figs. 2 and 4 is given in Fig. 5.

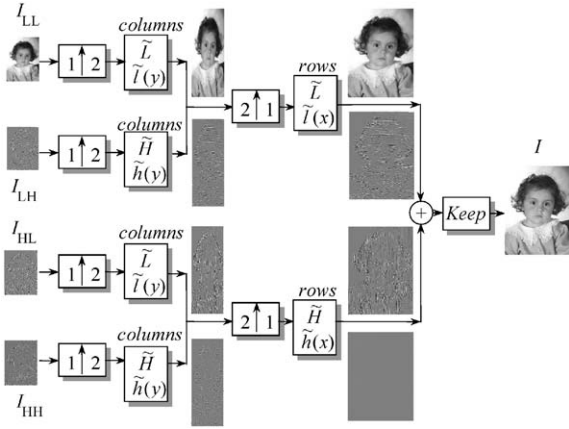


Fig. 4. One stage of 2-D DWT multiresolution image reconstruction (backward wavelet synthesis).

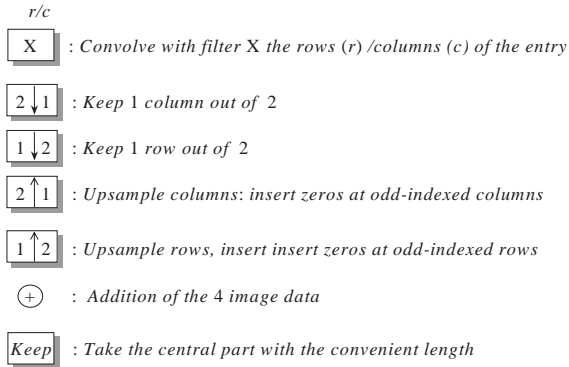


Fig. 5. Notations used in Figs. 2 and 4.

The detailed 2-D wavelet synthesis procedure for each of the four decomposed subimages $I_{LL}(x, y)$, $I_{LH}(x, y)$, $I_{HL}(x, y)$ and $I_{HH}(x, y)$ is individually described as follows.

Let $R \times C$ be the size of a subimage to restore, $\tilde{l}(i)$ the synthesis lowpass coefficients of a specific wavelet basis, $i=0, 1, 2, \dots, N_{\tilde{l}}$, where $N_{\tilde{l}}$ is the support length of the filter \tilde{L} , $\tilde{h}(j)$ the synthesis highpass coefficients of a specific wavelet basis, $j=0, 1, 2, \dots, N_{\tilde{h}}$, where $N_{\tilde{h}}$ is the support length of the filter \tilde{H} .

(1) *The synthesis of the smooth subimage $I_{LL}(x, y)$:* Upsample $I_{LL}(x, y)$ by a factor of 2 along the y -axis and filter with lowpass filter \tilde{L} : Let $I_{S_y}(x, 2y) = I_{LL}(x, y)$ and $I_{S_y}(x, 2y + 1) = 0$, for $x = 0, 1, 2, \dots, R - 1$ and $y = 0, 1, 2, \dots, C - 1$.

$$I_{S_y}^{-1}(x, y) = \frac{1}{N_{\tilde{l}}} \sum_{i=0}^{N_{\tilde{l}}-1} \tilde{l}(i) I_{S_y}(x, (y + i) \bmod 2C). \quad (17)$$

Upsample $I_{S_y}^{-1}$ by a factor of 2 along the x -axis and filter with lowpass filter \tilde{L} : Let $I_{S_x}^{-1}(2x, y) = I_{S_y}^{-1}(x, y)$ and $I_{S_x}^{-1}(2x +$

$1, y) = 0$, for $x = 0, 1, 2, \dots, R - 1$ and $y = 0, 1, 2, \dots, 2C - 1$:

$$I_S(x, y) = \frac{1}{N_{\tilde{l}}} \sum_{i=0}^{N_{\tilde{l}}-1} \tilde{l}(i) I_{S_x}^{-1}((x + i) \bmod 2R, y). \quad (18)$$

(2) *The synthesis of the horizontal detail subimage $I_{LH}(x, y)$:* Upsample $I_{LH}(x, y)$ by a factor of 2 along the y -axis and filter with highpass filter \tilde{H} : Let $I_{H_y}(x, 2y) = I_{LH}(x, y)$ and $I_{H_y}(x, 2y + 1) = 0$, for $x = 0, 1, 2, \dots, R - 1$ and $y = 0, 1, 2, \dots, C - 1$:

$$I_{H_y}^{-1}(x, y) = \frac{1}{N_{\tilde{h}}} \sum_{j=0}^{N_{\tilde{h}}-1} \tilde{h}(j) I_{H_y}(x, (y + j) \bmod 2C). \quad (19)$$

Upsample $I_{H_y}^{-1}$ by a factor of 2 along the x -axis and filter with lowpass filter \tilde{L} : Let $I_{H_x}^{-1}(2x, y) = I_{H_y}^{-1}(x, y)$ and $I_{H_x}^{-1}(2x + 1, y) = 0$, for $x = 0, 1, 2, \dots, R - 1$ and $y = 0, 1, 2, \dots, 2C - 1$:

$$I_H(x, y) = \frac{1}{N_{\tilde{l}}} \sum_{i=0}^{N_{\tilde{l}}-1} \tilde{l}(i) I_{H_x}^{-1}((x + i) \bmod 2R, y). \quad (20)$$

The vertical and diagonal detail subimages $I_{HL}(x, y)$ and $I_{HH}(x, y)$ can be reconstructed in a similar way to $I_{LH}(x, y)$ with corresponding filtering operations (\tilde{L} in column first, and then \tilde{H} in row for $I_{HL}(x, y)$; \tilde{H} in both column and row for $I_{HH}(x, y)$) to obtain the restored images $I_V(x, y)$ and $I_D(x, y)$, respectively.

The analysis and synthesis are illustrated with the two pedagogical examples (Fig. 6).

3. Merging the DWT coefficients

The content of this section is based on Ref. [1], as it is an excellent reference in this issue. Figs. 2 and 3(a) show that after one stage of processing, one image is decomposed into four frequency bands: low–low (LL), low–high (LH), high–low (HL) and high–high (HH). Fig. 3(b) shows a second stage of decomposition. Thus, a DWT with N decomposition levels will have $M = 3N + 1$ such frequency bands. The DWT will have a pyramid hierarchy [1]. The sizes of frequency bands will decrease as the decomposition goes on. Fig. 1, in the *Fusion* block, shows the 2-D structures of a multiscale transform with two decomposition levels. A given colored square corresponds to the same decomposition level. The black small squares in different frequency bands correspond to the same group of pixels, which indicate the spatial localization of the transform. These are the pixels to be fused in order to obtain the corresponding fused structure. For a transform with K levels of decomposition, there is always only one low-frequency band LL^K . The rest of bands are high-frequency bands in a given decomposition level.

As shown in the *Fusion* block, Fig. 1, we can have images at different decomposition levels, but only the fusion at the same resolution level is allowed. Now, the key issue is how

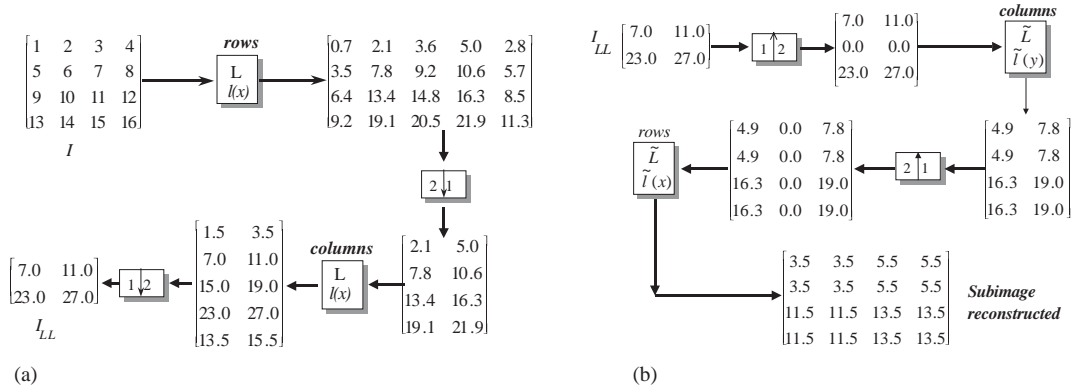


Fig. 6. Pedagogical examples: (a) analysis; (b) synthesis.

to form the fused representation at a given resolution level. In Ref. [1], the processing to achieve this goal is called the fusion rule. Some alternatives are given in this section. These include the choice of an activity-level measurement, coefficient grouping and coefficient combining (including consistency verification) methods.

As mentioned before, each step of the fusion is carried out with a set of frequency bands at the same resolution level. Following Ref. [1], to simplify the description of the different alternatives available in forming a fusion rule, we consider two source images X and Y , and the fused image Z . The extension to cases with more than two images is an easy task. Generally, an image I has its MSD representation denoted as D_I and the activity level as A_I . Thus, we shall encounter D_X, D_Y, D_Z, A_X and A_Y . Let $p = (m, n, k, l)$ indicate the index corresponding to a particular MSD coefficient, where m and n indicate the spatial position in a given frequency band, k the decomposition level, and l the frequency band of the MSD representation. Thus, $D_I(p)$ and $A_I(p)$ are the MSD value and activity level of the corresponding coefficient, respectively.

3.1. Activity-level measurement

The activity level of an MSD coefficient reflects the local energy in the space spanned by the term in the expansion corresponding to this coefficient. There are three categories of methods for computing the activity level $A_I(p)$ at position p : coefficient-based, window-based and region-based measures [1]. The coefficient-based activity (CBA) measures consider each coefficient separately. The activity level is described by the absolute value or square of the corresponding coefficient in the MSD representation:

$$A_I(p) = |D_I(p)| \quad \text{or} \quad A_I(p) = (D_I(p))^2. \quad (21)$$

The window-based activity (WBA) measures employ a small (typically 3×3 or 5×5) window centered at the current coefficient position. We also have several

alternatives here. One option is the weighted average method (WA-WBA)

$$A_I(p) = \sum_{s \in S, t \in T} w(s, t) |D_I(m + s, n + t, k, l)|, \quad (22)$$

where $w(s, t)$ is a weight and $\sum_{s \in S, t \in T} w(s, t) = 1$, S and T are sets of horizontal and vertical indexes that describe the current window, the sums over s and t range over all samples in the window.

Another option is the rank filter method (RF-WBA). Define $\text{Rank}(i)(Q)$ to pick the i th largest value in the set Q . Then RF-WBA uses $A_I(p) = \text{Rank}(i)(Q)$ where $Q = |D_I(m + s, n + t, k, l)|$ and S and T are as above.

A popular choice is to use RF-WBA to pick the maximum value, i.e. let $i = 1$. In this way, a high activity value indicates the presence of a prominent feature in the local area. To reduce the influence of impulse noise, we may also let $i = 2$ or 3 . One might also consider more general nonlinear processing of the coefficients in the window to compute activity that combines ranking and averaging.

Another option is the spatial frequency method (SF-WBA) [22,23], for an $M \times N$ image window block, the spatial frequency is defined as

$$A_I(p) = \sqrt{R^2 + C^2},$$

$$R = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=2}^N [D_I(m, n, k, l) - D_I(m, n-1, k, l)]^2},$$

$$C = \sqrt{\frac{1}{MN} \sum_{n=1}^N \sum_{m=2}^M [D_I(m, n, k, l) - D_I(m-1, n, k, l)]^2}, \quad (23)$$

where R and C are the row and column frequencies, respectively.

An additional option is to measure the activity by computing statistical properties in the window (ST-WBA), such as mean or standard deviation values. Both take into account image textures corresponding to the associated coefficients.

The regions used in region-based activity (RBA) measurement are similar to windows with odd shapes. Region segmentation can be performed on the same band using the edge information and a labeling algorithm. The output is a labeled image in which each different value represents a different region. Due to the spatial localization property, any region R^k in the low-frequency band has a corresponding group of coefficients in each high-frequency band, as illustrated in Fig. 1. We define the group of coefficients in all high-frequency bands corresponding to region R^k as $C(R^k)$. The activity level of the region R^k in the image I , $A_I(R^k)$ is computed as follows:

$$A_I(R^k) = \frac{1}{N_k} \sum_{p \in C(R^k)} A_I(p), \quad (24)$$

where N_k is the total number of the coefficients in $C(R^k)$. $A_I(p)$ is obtained by CBA measurement and expression (24) is considered as an energy, which is useful for texture analysis [24].

After all the region activity levels are obtained, the activity level of each MSD coefficient is determined as follows. Notice that the spatial location of each coefficient in the MSD representation will be either on an edge or inside a region. For the coefficient whose spatial location is on an edge, its activity level will be measured by the CBA or the RF-WBA method. For the coefficient whose spatial location is in a region, it will take the activity level of this region as its own activity measure. A slight generalization of Eq. (24) would compute activity over the region using a ranking or averaging or some combination of these.

3.2. Coefficient grouping method

We notice that after the MSD process, each coefficient will have a set of corresponding coefficients in other frequency bands and other decomposition levels, as illustrated in Fig. 1 by the black squares.

All coefficients with the black mark in Fig. 1 relate to the same group of pixels in the source image. For the majority of image fusion approaches, when determining the composite MSD representation, these coefficients are not associated with each other. This is called no-grouping schemes in Ref. [1]. If the corresponding coefficients in the same decomposition scale are jointly constrained to take the same decision, this is called a single-scale grouping scheme in Ref. [1]. This is a more restrictive case. The most restrictive case is to consider all the corresponding MSD samples together and ensure that they are all fused under the same method. This is called a multiscale grouping scheme in Ref. [1].

3.3. Coefficient combining method

When combining the source MSD representations to produce the composite MSD representation, there are at least two alternatives: averaging and selection. Averaging leads to a stabilization of the fusion result, while introducing the problem of possible pattern cancellation due to opposite object contrast in different input images. This can be avoided by the application of a selective combination scheme. In this selective scheme, the most salient image edge is chosen for the composite wavelet frame sequence. It is also possible to combine both schemes.

3.3.1. Selection

Except for the LL band, which has all positive transform values, all the other bands contain transform values that are fluctuating around zero. The larger transform values in these bands correspond to sharper brightness changes and thus to the salient features in the image such as edges, lines, and region boundaries. Therefore, a good integration rule is the choose-max (CM) scheme, which means just pick the coefficient with the larger activity level and discard the other. If Z is the fused image, this can be described as $D_Z(p) = D_i(p)$, where $i = X$ or Y depending on which source image satisfies

$$A_i(p) = \max(A_X(p), A_Y(p)). \quad (25)$$

Absolute values $|A_X(p)|$ and $|A_Y(p)|$ could be also considered in Eq. (25) as in Ref. [25].

Consistency verification attempts to exploit the idea that it is very likely that a good fusion method will compute neighboring coefficients in the composite MSD in a similar manner. In Ref. [3], the maximum absolute value within a window is chosen as an activity measure associated with the central MSD coefficient. In this way, a high activity value indicates the presence of a dominant feature in the local area. A binary decision map of the same size of the wavelet transform is then created to record the selection results based on a maximum selection rule. This binary map is subject to a consistency verification. Specifically, if the center composite MSD coefficient comes from image X while the majority of the surrounding coefficients come from image Y, the center sample is then changed to come from image Y. In Ref. [3], 3×3 or 5×5 windows are used for the neighborhood. This method can be viewed as a window-based verification (WBV). In the implementation, a majority filter (which outputs 1 if the count of 1's outnumbers the count of 0's, and outputs 0 otherwise) is applied to the binary decision map; the map is then negated, and is followed by the application of a majority filter. The resulting map is negated again. A fused image is finally obtained based on the new binary decision map. This selection scheme helps to ensure that the dominant features are incorporated as completely as possible into the fused images.

3.3.2. General weighted average

Another combining scheme is the weighted average (WA). At each position p , the composite MSD is obtained by

$$D_Z(p) = w_X(p)D_X(p) + w_Y(p)D_Y(p). \quad (26)$$

The weights w_X and w_Y may depend on the activity levels of the source MSD coefficients and the similarity between the source images at the current position. Suppose $A_X(p) < A_Y(p)$. One popular way to determine w_X and w_Y is suggested in Ref. [7]. At first, a match measure $M_{XY}(p)$ is defined as a normalized correlation averaged over a neighborhood of p ,

$$M_{XY}(p) = \frac{\sum_{s \in S, t \in T} w(s, t) D_X(m+s, n+t, k, l) D_Y(m+s, n+t, k, l)}{A_X^2(p) + A_Y^2(p)}, \quad (27)$$

where $w(s, t)$, A_X and A_Y are as defined in Eq. (22). If M_{XY} is smaller than a threshold α , then $w_X = 0$ and $w_Y = 1$, else if $M_{XY} \geq \alpha$ then

$$w_X = \frac{1}{2} - \frac{1}{2} \left(\frac{1 - M_{XY}}{1 - \alpha} \right) \quad \text{and} \quad w_Y = 1 - w_X. \quad (28)$$

Note that when the weight is zero, this means the substitution of an image by another.

Sometimes the degree of belief in a given event can be used to determine the different weights [26]. Indeed, when pieces of information issued from several sensors have to be combined, each of them can be represented as a degree of belief. The degrees of belief generally take their values in a real closed interval and are normalized in different ways, depending on the chosen mathematical framework. They are probabilities in data fusion methods based on probability and Bayesian theory, memberships degrees to a fuzzy set in fuzzy set theory or possibility distributions in possibility theory.

In Ref. [27], instead of the WA the following scheme is proposed:

$$D_Z(p) = \sqrt{D_X(p)D_Y(p)} + b, \quad (29)$$

where the constant b is a bias in order to modify the final result.

3.3.3. Adaptive weighted average

Some times the weight determination should be computed according to the application field. We need to determine the degree to which each pixel in the images to be fused is of interest. To fulfill this task, one must have a precise definition of the term interest. For example, in natural thermal images it could be useful to distinguish unnatural from natural objects. So, things that are of interest and unnatural will be either cooler or warmer than their background. Vehicles that have been functioning will be much warmer than the trees in which they will hide or cooler if they have

been stationary for a long time. This will lead to intensity levels in the produced image that will be darker or lighter. Consequently, the simplest way to emphasize these pixels is to weight them according to an adaptive weighted average (AWA) scheme, computing the weights as follows:

$$W_X(p) = |D_X(p) - \bar{D}_X(p)|^a, \quad (30)$$

where $D_X(p)$ is the MSD representation at $p(m, n, k, l)$ for an object of interest and $\bar{D}_X(p)$ is the average value for the MSDs computed over an area surrounding p , greater than the object (could be the whole image). Thus, the warmer and cooler pixels will be assigned larger weights. The a exponent allows the weight distribution to be modified.

On the contrary, in visual images the interest is centered on the detection of changes in contrast between regions on the edges separating these regions. A good method should produce large coefficients on the edges. This can be carried out by computing the variance in a neighborhood ($N \times M$) centered at $p(m, n, k, l)$

$$W_X(p) = \left[\frac{1}{NM} \sum_{s \in S, t \in T} (D_X(m+s, n+t, k, l) - \bar{D}_X(m+s, n+t, k, l))^2 \right]^a, \quad (31)$$

where the parameter a has the same function as above.

Instead of using the variance it is possible to use

$$W_X(p) = [|D_X(p) - \bar{D}_X(p)|]^a, \quad (32)$$

where the mean value $\bar{D}_X(p)$ is computed over the neighborhood ($N \times M$) centered at $p(m, n, k, l)$.

Some images have temporal intensity variations, instead of spatial ones. From temporal variations a weight can be computed as follows:

$$\tilde{D}_X(p, t) = \left| \frac{D_X(p, t) + \beta \tilde{D}_X(p, t-1)}{1 + \beta} \right|, \quad (33)$$

Initially, an average is calculated for each $p(m, n, k, l)$ at time t . The variable parameters β and a need to be determined in a heuristic fashion according to the specific application.

3.3.4. Other coefficient combining strategies

Some other coefficient combining methods are considered in this section, although they are basically derived from the above.

(a) *Fusion by energy comparison (EC)*: the MSDs coefficients are computed as follows [23,28]:

$$D_Z(p) = \begin{cases} D_i(p) & \text{if } A_i(R^k) > A_j(R^k) + T \quad \forall i \neq j, \\ D_j(p) & \text{if } A_i(R^k) < A_j(R^k) - T \quad \forall i \neq j, \\ \frac{1}{P} \sum_{i=1}^P D_i(p) & \text{if } |A_j(R^k) - A_l(R^k)| \leq T \quad \forall j, l, \end{cases} \quad (34)$$

where T is a threshold value, $A_i(R^k)$ is the activity level of the region R^k in the images i, j, l computed as in Eq. (24), P is the number of incoming images to be fused and the subindices i, j, l correspond to a given image of the P images.

(b) *Region-based fusion by multiresolution feature detection (RFD)*: the MSDs coefficients are computed as follows [28]:

$$D_z(p) = \frac{1}{Q(p)} \sum_{i=1}^P q_i(p) D_i(p)$$

$$\text{where } Q(p) = \begin{cases} \sum_{i=1}^P q_i(p) & \text{if } \sum_{i=1}^P q_i(p) \neq 0, \\ 1 & \text{if } \sum_{i=1}^P q_i(p) = 0 \end{cases} \quad (35)$$

and $q_i(p)$ is the mask matrix obtained as the result of binary morphological open and close operations applied to binary matrix $\bar{q}_i(p)$ defined by

$$\bar{q}_i(p) = \begin{cases} 1 & \text{if } |D_i(p)| > T, \\ 0 & \text{if } |D_i(p)| \leq T, \end{cases} \quad (36)$$

where T is a threshold value.

The principle of this method is to detect any irregularities (usually representing defects) from the images and preserve them in the fused image. This method was developed in the inspection field as described in Ref. [28].

(c) *Background elimination (BE)*: combining scheme based on a window approach is the direct BE scheme [25]. At each $p(m, n, k, l)$, the composite MSD can be obtained by

$$D_z(p) = D_x(p) - \mu_x + D_y(p) - \mu_y + \frac{\mu_x + \mu_y}{2}, \quad (37)$$

where μ_x and μ_y are the mean values in a given window. This fusion mechanism ensures that all the background information presented in the input images gets transferred into the fused image.

(d) *Variance area based (VA)*: the variance of each image patch over a window ($N \times M$) is computed as an activity measure associated with each $p(m, n, k, l)$ centered in the window. If the activity measures at the corresponding $p(m, n, k, l)$ locations are close to each other, the average of the two is considered as the new value; otherwise the larger value is chosen [3].

4. Fusion applications

In order to illustrate the fusion process, we propose three fusion examples according to the scheme in Fig. 1: multifocus CCD visual images, multispectral (MS) and panchromatic remote sensing images and functional and anatomical medical images.

4.1. Multifocus image fusion

Due to the limited depth-of-focus of optical lenses (especially those with long focal lengths) it is often not possible to get an image that contains all relevant objects “in focus”. One possibility to overcome this problem is to take several pictures with different focus points and combine them together into a single frame that finally contains the focused regions of all input images. This example follows the scheme given in Fig. 1(a).

- From the CCD visual images X and Y in Figs. 7(a) and (b), respectively, we can see that the left and right toy cars are out of focus. Apply the DWT to both X and Y to obtain approximation and details coefficients. Such coefficients are selected by CM and AWA, respectively, obtaining a fused multiscale image representation. Applying the IWDT to this multiscale image, the fused image in Fig. 7(c) is obtained.

4.2. Panchromatic and MS image fusion

Remote sensing images can be acquired in two different modes: either the panchromatic (P) mode with high spatial resolutions (HR) of 10 m pixel (SPOT), 5.8 m pixel (IRS) or even 1 m pixel (IKONOS), either the MS mode with much lower spatial resolution (LR) of 30 m pixel (LANDSAT TM), 23.5 m pixel (IRS), 20 m pixel (SPOT) or 4 m pixel (IKONOS). The P images are characterized by very high spatial information content well suited for intermediate scale mapping applications and urban analysis. The MS images provide the essential spectral information for smaller scale thematic mapping applications such as land use surveys. In order to take benefit of the high spatial information content of the P images and the essential spectral information of lower resolution MS images, fusion of these two types of images can be performed in order to produce pseudo-HRMS images.

The example shown is the fusion of the following images: three spectral bands (MS1 \equiv red, MS2 \equiv green, MS3 \equiv blue), i.e. with color information, coming from a remote IKONOS sensor with LR (4 m pixel) and a panchromatic band (P) image, i.e. without color information, with high spatial resolution (1 m pixel). The goal is to achieve a fused image containing both color information and high spatial resolution. This method is based on Refs. [4,29,30].

Figs. 8(a)–(c) show the MS1, MS2 and MS3 bands, respectively. Fig. 8(d) displays the MS color image resulting from the three MS i ($i = 1, 2, 3$) band combination. Fig. 8(e) shows the panchromatic band (P).

- Match the histogram of P according to the three MS i bands to obtain the corresponding PSM i at 1 m pixel of spatial resolution.
- Apply *DWT* twice (see Fig. 3 for details in the decomposition levels) to each PSM i to obtain the

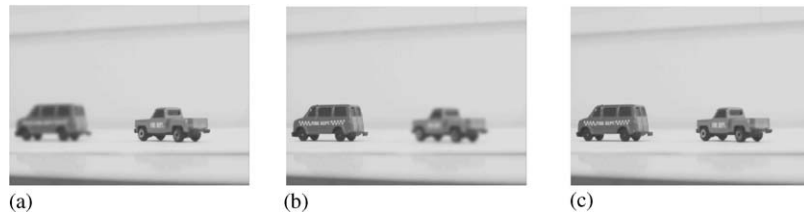


Fig. 7. Original CCD visual images with the (a) left and (b) right toy cars out of focus, respectively; (c) resulting fused image from (b) and (c) with the two toy cars in focus.

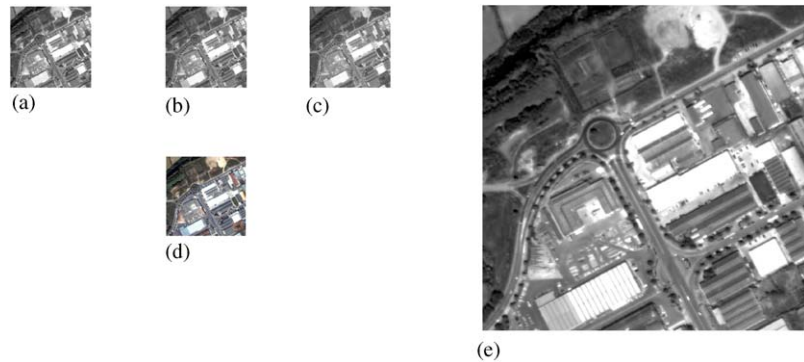


Fig. 8. (a) MS1, (b) MS2, and (c) MS3 spectral bands at 4 m of resolution; (d) color image obtained by the (a), (b) and (c) band combination; (e) panchromatic image of high resolution (1 m) without color information.

approximation PSA_{i_2} coefficients at the second resolution level and the PSD_{ijk} details coefficients $i = 1, 2, 3$ (bands), $j = h, v, d$ (horizontal, vertical and diagonal, respectively) and $k = 1, 2$ (decomposition level). The spatial resolution levels are 2 m pixel and 4 m pixel for $k = 1, 2$, respectively.

- Select and merge the corresponding PSA_{i_2} with MS_i by WA by taking a zero weight value for PSA_{i_2} and the unitary value for MS_i to obtain the fused scale $PSAF_i$. The WA carried out is a mere substitution of PSA_{i_2} with MS_i , i.e. in this particular example $MS_i \equiv PSAF_i$. This example follows the scheme given in Fig. 1(b).
- Apply IDWT to each $PSAF_i$ (approximation fused coefficients) with the corresponding detail coefficients PSD_{ij_2} to obtain PSA_{i_1} at 2 m pixel of spatial resolution. Apply once again IDWT to each PSA_{i_1} with the corresponding detail coefficients PSD_{ij_1} to obtain PMS_i at 1 m pixel of resolution, i.e. $PMS_1 \equiv$ red, $PMS_2 \equiv$ green, $PMS_3 \equiv$ blue.
- Combine the three spectral bands PMS_i at 1 m pixel of resolution to obtain the fused image, Fig. 9. This image includes the high resolution of the input P image and the color information of the MS information in MS_i .

4.3. Medical application

With the development of new imaging methods in medical diagnostics the need arises for a meaningful (and



Fig. 9. Fused MS image at 1 m of resolution (i.e. color image at high resolution).

spatially correct) combination of all available image data sets. Examples of imaging devices include computer tomography (CT), magnetic resonance imaging (MRI) or the newer positron emission tomography (PET).

Different medical imaging techniques may provide scans with complementary and occasionally conflicting information. The combination of images can often lead to additional clinical information not apparent in the separate images. The goal of image fusion is to impose a structural anatomical framework in functional images. Often in a functional image there simply is not enough anatomical detail to determine the position of a tumor or other lesion.

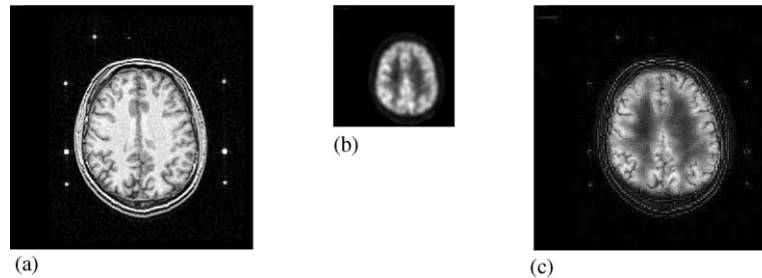


Fig. 10. (a) MRI and (b) PET images; (c) fused image from (a) and (b).

Generally, functional images have low spatial resolution and anatomical images have high spatial resolution. So, with anatomical images a lesion can be detected with an accuracy of millimeters. On the contrary, with the functional images this is not possible, but they have the ability to detect lesions before the anatomy is damaged. The fusion of both types of images could avoid undesired effects.

The following example illustrates the fusion of an MRI image, Fig. 10(a) with a PET image, Fig. 10(b). The MRI image has double the resolution of the PET.

- Apply DWT to the MRI to obtain the approximation MR_A and details MR_{Ij} coefficients $j = h, v, d$ (horizontal, vertical and diagonal). The PET should be image co-registered with respect to MR_A so that both images represent the same brain region.
- Resample and register (RS+IR) the PET image with respect to MR_A, by using the landmarks as control points.
- Select and merge MR_A and PET to obtain MR_A-PET by WA with weights of 0.4 and 0.6, respectively. This example follows the scheme given in Fig. 1(b).
- Apply IWDWT with MR_A-PET as the approximation coefficients and the MR_{Ij} as the detail coefficients.

A fused image at the resolution of MRI including the anatomical information of MRI and the functional information from PET, Fig. 10(c). In the fused image, the relative position of the functional information with respect to the anatomical landmark is clearly displayed.

5. Comparative analysis and performance evaluation

Although there have been as many attempts as there have been fusion algorithms, as yet no universally accepted standard has emerged for evaluating image fusion performance. In this work, we use both qualitative and quantitative methods. The qualitative methods are acceptance and verification tests which are accepted or rejected by a possible user, which determine visually the relative perceived image quality based on the contribution that the fusion makes to its

specific problem. This is the case for medical image fusion, where the corresponding professional compares the results against other non-imaging data.

The problem with defining a quantitative measure lies in the difficulty of defining an ideal composite image based on multisensor images or images taken at different times [3]. Nevertheless, we have used the following quantitative performance measure for the purpose of comparing different fusion methods.

5.1. Multifocus image fusion

We design the test strategy so that a correct focused image is captured and then we generate two images with different blurred objects by applying a lowpass mean filter to the desired regions containing the objects. The target image is the original image. We have used more than two images (up to five) and the performance of the different strategies is similar to that obtained with two images, so the results in this section are obtained by using two input images.

We compare different merging methods, different resolution levels and different wavelets families. As a quality measure we use an objective fidelity criterion [31], the root-mean-square error of the difference image between the ideal image and the fused image given by

$$e_{rms} = \left[\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [I_{pr}(x, y) - I_{fd}(x, y)]^2 \right]^{1/2}, \quad (38)$$

where I_{pr} is the perfect result and I_{fd} is the fused image.

The number of multifocus images (couples) is 15, five of them downloaded from the web and the remainder 10 self-generated by focusing–defocusing; Fig. 8 is a representative example. All results are averaged between the 15 images.

We have designed a test strategy according to the following steps:

(1) *Perform a comparative study for each wavelets family*: We have considered seven decomposition levels

Table 1
Results for the wavelets Biorthogonal family

Factors (\tilde{N}, N)															
bior	(1, 1)	(3, 1)	(1, 3)	(2, 2)	(3, 3)	(1, 5)	(2, 4)	(4, 4)	(3, 5)	(5, 5)	(2, 6)	(3, 7)	(2, 8)	(6, 8)	(3, 9)
Level	5	6	6	7	7	7	5	5	6	7	5	4	6	6	5
APX	AWA	AWA	AWA	AWA	WA	WA	AWA	AWA	WA	AWA	AWA	AWA	AWA	WA	WA
DET	CM	CM	CM	CM	CM	CM	CM	CM	CM	CM	CM	CM	CM	CM	CM
RMS	3.21	3.09	0.93	1.24	1.98	3.55	3.12	6.57	8.54	7.89	6.76	5.88	11.24	10.28	13.66

Table 2
Results for pyramidal families

Pyramid	LAP	FSD	COP	GRP	RAP	MOP
Level	4	5	4	6	6	7
LP	WA	WA	WA	WA	WA	WA
HP	CM	CM	CM	CM	CM	CM
RMS	0.62	2.89	1.23	2.41	2.12	1.95

and have implemented the following coefficient combining methods: CM, WA, AWA, EC, RFD, BE and VA. These methods will be analyzed later. We have also implemented the following activity-level measurement methods: CBA, WBA, RF-WBA, SF-WBA, ST-WBA and RBA. We have verified that similar final results are obtained with CBA, WBA and RF-WBA but the results obtained with SF-WBA and ST-WBA are worse than the previous. The worst results are obtained with RBA, as it requires good region segmentation and for the images used it has been ineffective. Therefore, in our experiments we have used CBA.

The wavelets families analyzed are (see Refs. [15,18]): Daubechies (db N , $N = 1..45$), Coiflets (coif N , $N = 1..5$), Symlets (sym N , $N = 2..45$), Biorthogonal (bior (\tilde{N}, N) with (\tilde{N}, N) the given in Table 1), Reverse Biorthogonal (rbior (\tilde{N}, N) with (\tilde{N}, N) as in bior), Mexican hat (mesh N with number of voices $N = 1..4$) and Morlet (morl N with number of voices $N = 2..4$). We firstly select the member of each family with the best results and this member is used as the representative member for this family so that it can be compared against the remaining representative wavelets family members and also against other methods.

Table 1 gives only the best results obtained for the Biorthogonal family. The pairs of factors (\tilde{N}, N) , first row, are ordered in the second row according to the increasing size of the filters. In the third row is the decomposition level number. In the fourth and fifth rows appear the combining methods for the approximation coefficients (APX) and details coefficients (DET), respectively. Finally, in the sixth row the root-mean-square error, between the fused image and the target image, is given.

We select the bior1.3 (the best of this family) as the representative member for this family in order to compare it against other representative family members and with other

classical fusion methods. The members selected as the best for each family are then given in Table 3.

(2) *Comparison of the wavelets families representative members against other fusion methods:* Now we perform a comparison of the wavelets families against the following classical fusion methods: Laplacian pyramid (LAP) [7], filter-subtract-decimate hierarchical pyramid (FSD) [33], contrast pyramid (COP) [34], gradient pyramid (GRP) [34], ratio pyramid (RAP) [35], morphological pyramid (MOP) [36] and neural networks (NN) (including consistency verification as in the original reference) [37]. We have used Rockinger's MATLAB toolbox [38] as the reference for the implementation of the above methods. The lowpass filter used in the pyramidal methods is the following 5-separable filter, which is used in Ref. [33]: [1, 4, 6, 4, 1]/16.

Table 2 gives the results obtained for the above pyramidal methods, first row. In the second row is the decomposition level. The third and fourth rows show the combining method used for the lowpass and highpass filters, respectively. Finally, in the fifth row the root-mean-square error is given between the fused and target images.

Table 3 and Fig. 11 give the root-mean-square error values for the different methods. For each family, the member which has obtained the best results has been selected (bior(1,3) for the Biorthogonal family). Similarly, for the remainder of the families, the selected members are given in the first row of Table 3. The best results for pyramidal methods are collected from Table 2 and then we have added the NN approach.

5.2. Panchromatic and MS image fusion

The approach described in Section 4.2 is based on the wavelet method in Ref. [4], which is exactly the ARSIS

Table 3
Results for different methods

Method	LAP	FSD	COP	GRP	RAP	MOP	NN	db4, size 8	coif1, size 6	bior(1,3), size 6	sym4, size 8	rbior(4,4), size 4	mesh3, size 6	mor13, size 6
RMS	0.62	2.89	1.23	2.41	2.12	1.95	2.02	0.99	1.45	0.93	3.12	1.24	2.11	1.98

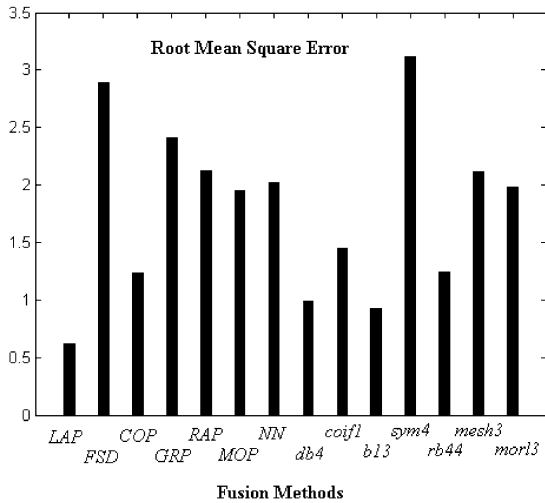


Fig. 11. Root-mean-square error for different fusion methods.

hybrid Model 2 described in Ref. [2] (ARS). We have emphasized on results obtained with different wavelets families because other comparative results are well reported in Refs. [2,4,38]. As in Section 5.1, for this purpose we have used the MATLAB Wavelets toolbox [18] and we have considered the set of families included within it. Note that this approach does not require activity level measurement and the coefficient combination is a mere substitution.

In order to assess the quality of the resulting fused images we have followed the method described in Refs. [2,39] applied to a set of 22 images coming from the IKONOS satellite. All the results are averaged for the 22 images. The results are compared against the following methods: local mean matching (LMM), local mean and variance matching (LMVM), high-frequency addition (HFA) in Refs. [40–42] and entropy data fusion (EDF) in Ref. [43]. Some of these methods require the selection of a window size; hence we have chosen the best size reported in Ref. [38] which is 3×3 .

With regard to other classical methods, such as principal component analysis (PCA), intensity-hue-saturation (IHS) or Brovey, we have found it irrelevant to include the comparative results because they are similar to the results reported in Refs. [38,44], although in our experiments, the correlation values between the different image channels are higher than those obtained in Ref. [38]. Indeed, we have found that the averaged correlation, computed by us, between the panchro-

matic and the XS3 channels is 0.11 compared with the 0.04 obtained in Ref. [38], for XS1, and XS2 is 0.68 and 0.62, respectively. This implies that the low correlation value for the XS3 channel precludes a satisfactory fusion with IHS and PCA. This is a well-known problem in the computer vision community and has been reported as due to the near infrared reflectance in the panchromatic band.

With testing purposes for the ARS method and following Refs. [3,4] we use the first property described in Ref. [39], which implies that any synthetic image, once degraded to its original resolution, should be as identical as possible to the original image. Hence, in our IKONOS images, the synthetic images are 1 m of resolution, so these images once degraded to 4 m should be as close as possible to the original MSi spectral bands. For the second and third properties we address the reader to Refs. [2,39] because they are properties referred to images obtained with better resolutions in the sensor and the IKONOS sensor does not provide better image resolutions than those used in our experiments. We also use the difference between the means of the original image and the synthetic image, the correlation coefficient between the original and synthetic images and the standard deviations of the difference images.

For each wavelets family, we obtain the member of the family giving the best results and then this member is selected for a global comparison with all the others. The best results are obtained for the member whose values are closest to the ideal case, i.e. differences in the mean zero, correlation coefficient one and low standard deviations. We use the 22 test images to select each member.

The same set of wavelets families as those studied in the above experiments have been analyzed here.

Table 4 and Fig. 12 give the statistics results on the differences between the original and synthesized images, for the Biorthogonal family, the pairs of factors (\tilde{N}, N) have been ordered according to the increasing size of the filters. This is because we have verified that the filter size is decisive for all families. Indeed, we have obtained the best results with sizes ranging from 6 to 10, since sizes above 10 are worse. Low sizes, such as 2 for db1 (Haar), bior(1,1) or sym1, introduce some kind of distortion around the edges. All values in Table 4 have been averaged taking into account all spectral bands for the 22 images because we have not found significant differences for each individual band.

Table 5 shows the member of each family that has achieved the best results against each one of the remaining classical methods.

Table 4

Averaged Statistics on the differences between the original images and the synthesized images for the wavelets Biorthogonal family

Factors (\tilde{N}, N)	(1, 1)	(3, 1)	(1, 3)	(2, 2)	(3, 3)	(1, 5)	(2, 4)	(4, 4)	(3, 5)	(5, 5)	(2, 6)	(3, 7)	(2, 8)	(6, 8)	(3, 9)
bior size	2	4	6	6	8	10	10	10	12	12	14	16	18	18	20
Mean	7.32	7.40	5.11	4.03	9.54	11.28	12.55	13.57	13.45	14.01	18.91	22.13	33.12	45.87	66.32
Std.	6.78	6.77	7.98	6.06	10.21	10.67	11.02	11.03	14.56	15.78	16.34	19.67	19.54	25.89	32.21
Corr	0.91	0.92	0.96	0.97	0.95	0.94	0.93	0.93	0.91	0.90	0.89	0.89	0.87	0.86	0.87

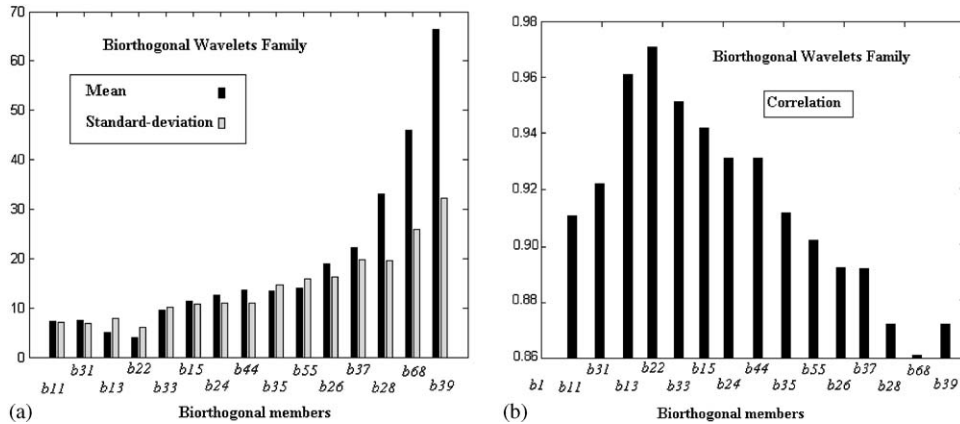


Fig. 12. Averaged statistics on the differences between the original images and the synthesized images for the wavelets Biorthogonal family: (a) mean and standard deviation; (b) correlation coefficient.

Table 5

Averaged Statistics on the differences between the original images and the synthesized images for different fusion methods

	Fusion methods										
	LMM	LMVM	HFA	EDF	db4	coif1	bior(2,2)	sym5	rbio(3,3)	mesh3	mor13
Mean	3.56	4.01	5.96	6.07	5.65	4.76	4.03	7.20	6.89	8.56	9.02
Std.	4.01	4.71	7.01	8.99	6.56	6.22	6.06	9.65	9.36	12.23	12.45
Corr	0.99	0.98	0.96	0.95	0.96	0.97	0.97	0.92	0.94	0.93	0.93

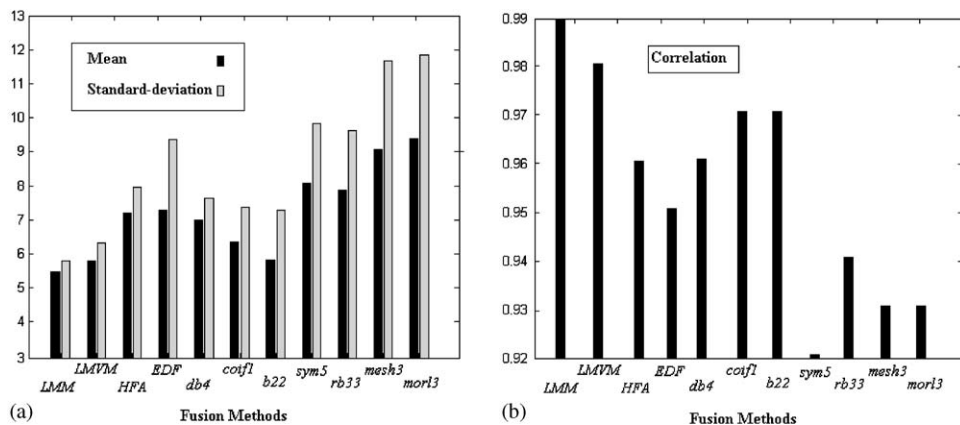


Fig. 13. Averaged statistics on the differences between the original images and the synthesized images for different fusion methods: (a) mean and standard deviation; (b) correlation coefficient.

Table 5 and Fig. 13 give the statistical results for the methods shown in the first row of the table. Here is the best member for each wavelets family. As above, these values are averaged over the 22 images and for all spectral bands.

6. Conclusions

This paper is a tutorial providing guidelines about the use of wavelets in image fusion. There are a lot of fusion possibilities and only a few application examples can be shown. For additional examples see Refs. [45], [46] or [47]. We have chosen the most representative examples, so that some of them can be easily extended for performing a wavelet-based fusion for any new application. We provide the theory for wavelets so that the paper is self-contained in wavelets fusion approaches.

The number of fusion possibilities with different wavelets families, resolution levels and coefficient combinations becomes very high. We give quantitative results for different fusion methods, where the wavelets families are compared with each other and with other fusion strategies. From Tables 1 to 5 and Figs. 11 to 13 the following conclusions can be inferred:

- (1) From Table 1, the best results are obtained with bior1.3 with a decomposition level of 6. Also AWA or WA for APX and CM for DET are the best coefficient combining methods. This is in accordance with Ref. [32]. The level number 6 appears to be the most appropriate. The previous general conclusions are applicable to the remaining wavelets families.
- (2) From Table 2, the best results are obtained with LAP and once again with WA and CM for the lowpass and highpass filters, respectively. This conclusion about WA and CM agrees with the results obtained for the different wavelets families. This assertion is based on the fact that the APX and DET coefficients come from lowpass and highpass filtering, respectively.
- (3) From Table 3 and Fig. 11 the best results are obtained for LAP followed by bior(1,3) and db4. It is important to note that LAP requires a decomposition level less than any of the wavelets families. The methods coif1 and rbior(4,4) obtain acceptable results. One important conclusion is that the results concerning wavelets families are obtained with filter sizes ranging from 4 to 8, but there is no direct relation between the size and the results, as can be seen from the sizes shown in Table 3. Indeed, a good result is obtained with size 8 (db4) but the worst result is also obtained with size 8 (sym4) and the best result is obtained with size 6 (bior(1,3)).
- (4) From the results in Table 4 and Fig. 12, we can verify that the best results are obtained with filter sizes of 6, i.e. bior2.2 and bior1.3 followed by filters with sizes of 10 (bior1.5, bior2.4 and bior4.4). As the sizes increase the results are worse. The above is extensible to the remaining wavelets families. Filters with low sizes achieve better results than those with high ones.
- (5) From the results in Table 5 and Fig. 13, we can infer that the best results are obtained with methods LMM and LMVM, followed by bior(2,2) and coif1. Both members have filters of size 6. As the filter size increases the results are worse than the above. This occurs for db4, rb33 (rbio(3,3)), sym5 with sizes 8, 8 and 10, respectively. The worst results are obtained with mesh3 and morl3, although they both have been generated to be of size 6. We have carried out experiments increasing the decomposition level without any improvement, and instead obtaining worse results. This is in disagreement with Ref. [44] where levels 3 and 5 are used.

From the above conclusions, we have the behavior and performance of the reported methods. Now, the following discussion about the use of wavelets can help the reader to take a decision. The wavelet-based methods, mentioned in the above conclusions, achieve similar results than the classical methods. Nevertheless, their worst reported performance is due to the presence of a high number of edges in the source images. This has been addressed and studied by us in multifocus images [48]. When the images are smooth, without abrupt intensity changes, the wavelets work appropriately, improving the results of the mentioned classical methods. This has been verified with smooth images and also with the medical images, where no significant changes are present. The type of images (visual CCD, remote sensing, medical) is irrelevant.

The main handicap is found in the decomposition level for multiresolution approaches. Indeed, a decomposition of 6 requires that the source images have sizes greater than 2048×2048 , so that the merging can be carried out to a resolution of 32×32 in wavelets with decomposition level of 6, otherwise is ineffective. In classical approaches, with decomposition level of 4, the sizes can be of 512×512 pixels.

Although this tutorial is a pixel-level-based approach, it can be easily extended to the other three fusion-level schemes, namely [3]: signal, feature and symbol. Nevertheless, special processing should be required such as edge/region detection or object recognition [49].

Acknowledgements

The authors wish to acknowledge Dr. L. Jañez Head of the Instituto Complutense de Imagen y Telemedicina, E. Ortiz co-worker in the same Institution, and Dr. Carreras Head of PET Institute, for his support in the medical image fusion applications. They have provided us with the medical images shown in this work. The constructive recommendations provided by the reviewers are also gratefully acknowledged.

References

- [1] Z. Zhang, R.S. Blum, A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application, *Proc. IEEE* 87 (8) (1999) 1315–1326.
- [2] T. Ranchin, L. Wald, Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation, *Photogramm. Eng. Remote Sensing* 66 (1) (2000) 49–61.
- [3] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, *Graphical Models Image Process.* 57 (3) (1995) 235–245.
- [4] A. Rosenfeld, M. Thurston, Edge and curve detection for visual scene analysis, *IEEE Trans. Comput.* 20 (1971) 562–569.
- [5] D. Marr, *Vision*, Freeman, San Francisco, CA, 1982.
- [6] P.J. Burt, E. Adelson, The Laplacian pyramid as a compact image code, *IEEE Trans. Commun.* 31 (1983) 532–540.
- [7] E.H. Adelson, C.H. Anderson, J.R. Bergen, P.J. Burt, J. Ogden, Pyramid methods in image processing, *RCA Eng.* 29 (6) (1984) 33–41.
- [8] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer, Norwell, MA, 1994.
- [9] B. Garguet-Dupont, J. Girel, J. Chassery, J.G. Pautou, The use of multiresolution analysis and wavelets transform for merging SPOT panchromatic and multispectral image data, *Photogramm. Eng. Remote Sensing* 62 (9) (1996) 1057–1066.
- [10] J.B.A. Maintz, M.A. Viergever, A survey of medical image registration, *Med. Image Anal.* 2 (1) (1998) 1–36.
- [11] R.J. Althof, M.G.J. Wind, J.T. Dobbins, A rapid and automatic image registration algorithm with subpixel accuracy, *IEEE Trans. Med. Imaging* 16 (3) (1997) 308–316.
- [12] C. Schmid, R. Mohr, C. Bauckhage, Evaluation of interest points, *Int. J. Comput. Vision* 37 (2) (2000) 151–172.
- [13] J.L. Starck, F. Murtagh, A. Bijaoui, *Image Processing and Data Analysis: the Multiscale Approach*, Cambridge, University Press, Cambridge, 2000.
- [14] E.J. Stollnitz, T.D. DeRose, D.H. Salesin, Wavelets for computer graphics: a primer, part 1, *IEEE Comput. Graphics Appl.* 15 (3) (1995) 76–84.
- [15] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
- [16] I. Daubechies, Orthonormal bases of compactly supported wavelets, *Commun. Pure Appl. Math.* 41 (1988) 909–996.
- [17] D.M. Tsai, B. Hsiao, Automatic surface inspection using wavelet reconstruction, *Pattern Recognition* 34 (2001) 1285–1305.
- [18] M. Misiti, Y. Misiti, G. Oppenheim, J. Poggi, *Wavelet Toolbox for Use with MATLAB: User's Guide*, The MathWorks, Natick, NA, 2000.
- [19] C.S. Burrus, R.A. Gopinath, H. Guo, *Introduction to Wavelets and Wavelet Transforms: a Primer*, Prentice-Hall, Upper Saddle River, NJ, 1998.
- [20] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (1989) 674–693.
- [21] E. Morales, F.Y. Shih, Wavelet coefficients clustering using morphological operations and pruned quadrees, *Pattern Recognition* 33 (2000) 1611–1620.
- [22] A.M. Eskicioglu, P.S. Fisher, Image quantity measures and their performance, *IEEE Trans. Commun.* 43 (12) (1995) 2959–2965.
- [23] S. Li, J.T. Kwok, Y. Wang, Combination of images with diverse focuses using the spatial frequency, *Inf. Fusion* 2 (2001) 169–176.
- [24] A. Mojsilovic, M.V. Popovic, A.N. Neskovic, A.D. Popovic, Wavelet image extension for analysis and classification of infarcted myocardial tissue, *IEEE Trans. Biomed. Eng.* 44 (9) (1997) 856–866.
- [25] V. Petrovic, C. Xydeas, Multiresolution image fusion using cross band selection, *Proc. SPIE* 39 (1999) 319–326.
- [26] I. Bloch, Information combination operators for data fusion: a comparative review with classification, *IEEE Trans. Systems Man Cybernet.* 26 (1) (1996) 52–67.
- [27] R. Welch, M. Ehlers, Merging multiresolution SPOT HRV and landsat TM Data, *Photogramm. Eng. Remote Sensing* 53 (3) (1987) 301–303.
- [28] B.J. Matuszewski, L.K. Shark, J.P. Smith, M.R. Varley, Automatic fusion of multiple non-destructive testing images and CAD models for industrial inspection, *Proceedings of the IEE Seventh International Conference on Image Processing and its Applications, IPA'99, Manchester, 1999*, pp. 661–665.
- [29] D.A. Yocky, Image merging and data fusion by means of the discrete two-dimensional wavelet transform, *J. Opt. Soc. Am. A: Opt., Image Sci. Vision* 12 (9) (1995) 1834–1841.
- [30] J. Núñez, X. Otazu, O. Fors, A. Prades, Simultaneous image fusion and reconstruction using wavelets applications to SPOT + LANDSAT images, *Vistas Astron.* 41 (3) (1997) 351–357.
- [31] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA, 1993.
- [32] S.G. Nikolov, D.R. Bull, C.N. Canagarajah, M. Halliwell, P.N.T. Wells, Fusion of 2-D images using their multiscale edges, *Proceedings of the International Conference on Pattern Recognition, 2000*, pp. 1051–1054.
- [33] H. Greenspan, C.H. Anderson, S. Akber, Image enhancement by nonlinear extrapolation in frequency space, *IEEE Trans. Image Process.* 9 (6) (2000) 1035–1048.
- [34] O. Rockinger, Image sequence fusion using a shift-invariant wavelet transform, *Proceedings of the IEEE International Conference on Image Processing, Vol. III, 1997*, pp. 288–291.
- [35] A. Toet, Image fusion by a ratio of low-pass pyramid, *Pattern Recognition Lett.* 9 (4) (1989) 245–253.
- [36] A. Toet, A morphological pyramidal image decomposition, *Pattern Recognition Lett.* 9 (4) (1989) 255–261.
- [37] S. Li, J.T. Kwok, Y. Wang, Multifocus image fusion using artificial neural networks, *Pattern Recognition Lett.* 23 (2002) 985–997.
- [38] Rockinger's MATLAB toolbox in www.rockinger.purespace.de.
- [39] L. Wald, T. Ranchin, M. Mangolini, Fusion of satellite images of different spatial resolutions, *Photogramm. Eng. Remote Sensing* 63 (6) (1997) 691–699.
- [40] D.P. Filiberti, S.E. Marsh, R.A. Schwengerdt, Synthesis of imagery with high spatial and spectral resolution from multiple image sources, *Opt. Eng.* 33 (8) (1994) 2520–2528.
- [41] S. de Bethune, F. Muller, M. Binard, Adaptive intensity matching filters: a new tool for multi-resolution data fusion, *Proceedings of the Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar, RTO-NATO Organization, 1997*, pp. 671–680.
- [42] J. Hill, C. Diemer, O. Stöver, T. Udelhoven, A local correlation approach for the fusion of remote sensing data

- with different spatial resolutions in forestry applications, *Proceedings of the International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 7-4-3, 1999, pp. 781–789.
- [43] R.A. Showengerdt, Reconstruction of multispatial, multispectral image data using spatial frequency contents, *Photogramm. Eng. Remote Sensing* 46 (10) (1980) 1325–1334.
- [44] F.J. Tapiador, J.L. Casanova, An algorithm for the fusion of images based on Jaynes' maximum entropy method, *Int. J. Remote Sensing* 23 (4) (2002) 777–785.
- [45] A. Toet, L. Ruyven, J. Velaton, Merging thermal and visual images by a contrast pyramid, *Opt. Eng.* 28 (7) (1989) 789–792.
- [46] J. Zhou, D.L. Civco, J.A. Silander, A wavelet transform method to merge Landsat TM and SPOT panchromatic data, *Int. J. Remote Sensing* 19 (4) (1998) 743–757.
- [47] S. Li, J.T. Kwok, Y. Wang, Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images, *Inf. Fusion* 3 (2002) 17–23.
- [48] M. Santos, G. Pajares, M. Portela, J.M. de la Cruz, in: F.J. Perales, A.J.C. Campilho, N. Pérez de la Blanca, A. Sanfeliu (Eds.), *A new Wavelets Image Fusion Strategy*, Lecture Notes in Computer Science, Pattern Recognition and Image Analysis, Vol. 2652, Springer, Berlin, 2003, pp. 919–926.
- [49] G. Pajares, J.M. de la Cruz, *Visión por Computador: Imágenes Digitales y Aplicaciones*, RA-MA, Madrid, 2001.

About the Author—GONZALO PAJARES G. Pajares Received M.Sc and Ph.D. degrees in Physics from UNED (distance University from Spain) (1987, 1995) discussing a thesis on the application of pattern recognition techniques to stereovision. Since 1990 he worked at Indra in critical real-time software development. He also was working in Indra Space developing remote sensing applications. He joined the Complutense University in 1995 as an associated professor and from 2004 as a professor at full time. The areas covered are: Robotics and computer visual perception, including image processing and computer vision tasks. His current research interests include robotics vision systems, medical imaging and remote sensing applications in registering, fusion, change detection and superresolution in video sequences.

About the Author—JESÚS MANUEL DE LA CRUZ J. M. de la Cruz received M.Sc degree in Physics and Ph.D. from the Complutense University in 1979 and 1984, respectively. From 1985 to 1990 he was with the Department of Automatic Control, UNED (Distance University of Spain), and from October 1990 to 1992 with the Department of Electronic, University of Santander. In October 1992, he joined the Department of Computer Science and Automatic Control of the Complutense University where he is a Professor. His current research interest include robotics vision systems, fusion sensors and applications of automatic control to robotics and flight control.