Probability, Networks and Algorithms

PNA

A general framework for multiresolution image fusion: from pixels to regions

G. Piella

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

A General Framework for Multiresolution Image Fusion: from Pixels to Regions

Gemma Piella

CWI

P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT

This paper presents an overview on image fusion techniques using multiresolution decompositions. The aim is two-fold: (i) to reframe the multiresolution-based fusion methodology into a common formalism and, within this framework, (ii) to develop a new region-based approach which combines aspects of both object and pixel-level fusion. To this end, we first present a general framework which encompasses most of the existing multiresolution-based fusion schemes and provides freedom to create new ones. Then, we extend this framework to allow a region-based fusion approach. The basic idea is to make a multiresolution segmentation based on all different input images and to use this segmentation to guide the fusion process. Performance assessment is also addressed and future directions and open problems are discussed as well.

1 Introduction

In this section, we introduce the concept of image fusion. We discuss the motivation for fusion and its benefits, and the various issues that need to be addressed when designing a fusion scheme. We also give some application examples and review some of the most important fusion techniques used in practice.

1.1 Problem statement: Needs for image fusion

Extraordinary advances in sensor technology, microelectronics and communications have brought a need for processing techniques that can effectively combine information from different sources into a single composite for interpretation. In image-based application fields, image fusion[1] has emerged as a promising research area.

Image fusion provides the means to integrate multiple images into a composite image that is more suitable for the purposes of human visual perception and computer-processing tasks such as segmentation, feature extraction and target recognition. For example, the fusion of visual and infrared images in an airborne sensor can aid pilots navigate in poor weather conditions, and the fusion of computer tomography and magnetic resonance images may facilitate medical diagnosis.

1.2 Concepts of image fusion

According to the International Society of Information Fusion (ISIF),

> *"information fusion encompasses the theory, techniques and tools conceived and employed for exploiting the synergy in the information acquired from multiple sources such that the resulting decision or action is in some sense better than would be possible if any of these sources were used individually without such synergy exploitation."*

---

[1] Terminologies such as fusion, integration or merging, are often used interchangeably in the literature.

The first part of the definition points out the importance of the architecture and of the mathematical tools in information fusion systems; the second part emphasizes the importance of fusion goals and performance in relation with the application considered. Note that this definition covers different methodologies in information fusion. Of these, image fusion is just a particular one.

In this paper we are concerned with the fusion of visual information. Indeed, as many sources produce images, image processing has become one of the most important domains for fusion. Image fusion can be broadly defined as the process of combining multiple input images into a smaller collection of images, usually a single one, which contains the 'relevant' information from the inputs, in order to enable a good understanding of the scene, not only in terms of position and geometry, but more importantly, in terms of semantic interpretation. In this context, the word 'relevant' should be considered in the sense of 'relevant with respect the task the fused images will be subject to', in most cases high level tasks such as interpretation or classification. In the sequel, we will refer to this 'relevant' information as *salient* information. The images to be combined will be referred to as *input* or *source* images, and the resultant combined image (or images) as *fused* image.

The actual fusion process can take place at different levels of information representation. A common categorization is to distinguish between pixel, feature and symbol level [55, 72], although indeed these levels can be combined themselves [19, 29, 55]. Image fusion at pixel level means fusion at the lowest processing level referring to the merging of measured physical parameters [23, 106]. It generates a fused image in which each pixel is determined from a set of pixels in the various sources. Fusion at feature level requires first the extraction (e.g., by segmentation procedures) of the features contained in the various input sources [8, 25]. Those features can be identified by characteristics such as size, shape, contrast and texture [59]. The fusion is thus based on those extracted features and enables the detection of useful features with higher confidence. Fusion at symbol level allows the information to be effectively combined at the highest level of abstraction [24, 41]. The input images are usually processed individually for information extraction and classification. This results in a number of symbolic representations which are then fused according to decision rules which reinforce common interpretation and resolve differences. The choice of the appropriate level depends on many different factors such as data sources, application and available tools. At the same time, the selection of the fusion level determines the necessary pre-processing involved. For instance, fusing data at pixel level requires co-registered images at subpixel accuracy because the existing fusion methods are very sensitive to misregistration.

Currently, it seems that most image fusion applications employ pixel-based methods. The advantage of pixel fusion is that the images used contain the original information. Furthermore, the algorithms are rather easy to implement and time efficient. As we observed before, an important pre-processing step in pixel-fusion methods is image registration, which ensures that the information from each source is referring to the same physical structures in the real world. Throughout this paper, it will be assumed that all source images have been registered. Comprehensive reviews on image registration can be found in [9, 50, 94, 103].

1.3 Objectives, requirements and challenges of pixel-based image fusion

It is the aim of image fusion to integrate complementary and redundant information from multiple images to create a composite that contains a 'better' description of the scene than any of the individual source images.processing task applied following fusion. By integrating information, image fusion can reduce dimensionality. This results in a more efficient storage and faster interpretation of the output. By using redundant information, image fusion may improve accuracy as well as reliability, and by using complementary information, image fusion may improve interpretation capabilities with respect to subsequent tasks. This leads to more accurate data, increased utility and robust performance.

Considering the objectives of image fusion and its potential advantages, some generic requirements can be imposed [79]:

- The fusion algorithm should not discard any salient information contained in the input images.

- The fusion algorithm should not introduce any artifacts or inconsistencies which can distract or mislead a human observer or any subsequent image processing steps.

- The fusion algorithm must be reliable, robust and have, as much as possible, the capability to tolerate imperfections such as noise or misregistrations.

Clearly, a choice of which information is salient has to be made. Here again, the knowledge about the input data and the application plays a crucial role. However, a fusion approach which is independent of the modalities of the inputs and produces a fused image which appears natural is highly desirable.

The requirements listed above are often very difficult to achieve and even more difficult to assess. The problem of evaluating image fusion methods lies in the variety of different application requirements and the lack of a clearly defined ground-truth. The topic of performance evaluation will be discussed in Section 5 in more detail.

To illustrate some of the challenges we have to face when developing a fusion algorithm, consider the registered source images Fig. 1(a) and Fig. 1(b) depicting the same scene. While in the visual image (Fig. 1(a)) it is hard to distinguish the person in camouflage from the background, this person is clearly observable in the infrared (IR) image (Fig. 1(b)). In contrast, the easily discernible background in the visual image, such as the fence, is nearly imperceptible in the IR image. How to combine both images in a unique composite which represents the overall scene better than any of the two individual images? We sum up explicitly some of the difficulties that we encounter:

- Complementary information: some image features[2] appear in one source but not in the other, e.g., the man in Fig. 1(b) or the fence in Fig. 1(a).

- Common but contrast reversal information: there are various objects and regions that occur in both images but with opposite contrast, e.g., part of the roof of the house or the bushes at the left lower corner. Thus, the direct approach of adding and averaging the source images is not satisfactory.

- Disparity between sensors: input images come from different types of sensors which have different dynamic range and different resolution. Moreover, they may not be equally reliable. If possible, such disparities have to be taken into account when comparing the content of the information in the images.

This is, by no means, an exhaustive list of problems that could arise. Moreover, we should also be aware of the inherent difficulties present in any image acquisition and analysis task: presence of noise, sensor calibration or hardware limitations, to name a few.

1.4 Application fields

Image fusion is widely recognized as a valuable tool for improving overall system performance in image-based application areas such as defense surveillance, remote sensing, medical imaging and computer vision. We describe some application fields in more detail:

Defense systems

Historically, this seems to be the first application area. It covers subareas such as detection, identification and tracking of targets [3, 7, 92], mine detection [63, 64], tactical situation assessment [80, 99], and person authentication [65].

Fig. 2 illustrates how information from visible and IR wavelength images can improve situational awareness in a typical pilotage scene. Note that the IR image (Fig. 2(b)) contains much of the road network details while the visual image (Fig. 2(a)) provides horizon information and additional building and vegetation details. Note also that the light spots appear only in the visual image and are here perceived as small dark blobs. Of different nature is the glare effect on the IR image. This is due to common scanner interference and is usually perceived as a ripple effect. The resulting composite image in Fig. 2(c) contains the most salient information from each sensor.

Geoscience

This field concerns the earth study with satellite and aerial images (remote sensing) [72]. The main problem is the interpretation and classification of images [21, 49, 102]. The fused image allows the detection of roads, airports, mountainous areas, etc. In remote sensing applications, there is often a difference in spatial or wavelength resolution between the images produced by the different sensors. A typical example is the merging of a high-resolution SPOT Panchromatic image with Landsat

---

[2]Image features are patterns in the image that arise due to objects and materials in the scene, environmental factors and the sensing process.

(a)

(b)

Figure 1: *Example of source images to be fused: (a) visual image; (b) infrared image. These images are courtesy of Alex Toet, from TNO Human Factors Institute, The Netherlands.*
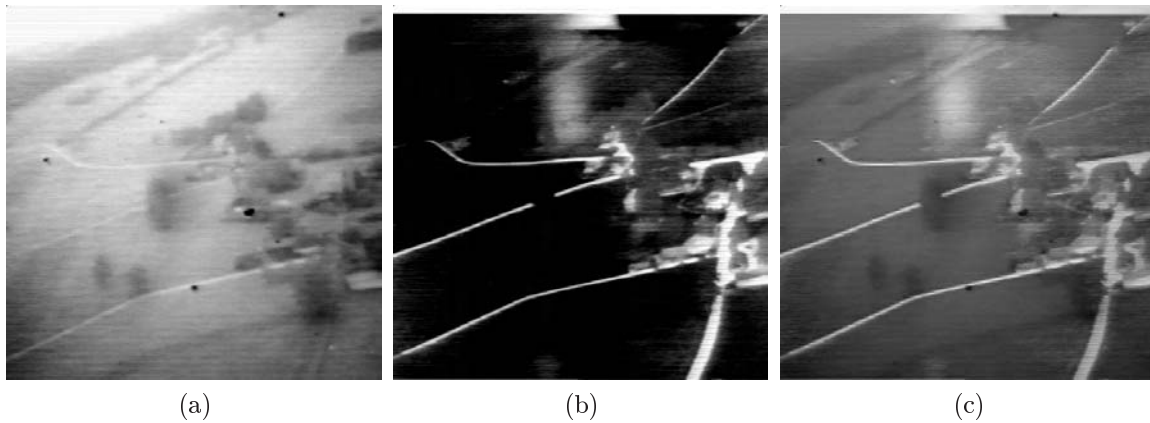
Figure 2: *Fusion of visual and IR images: (a) visual image; (b) IR image; (c) fused image. The fused image has been obtained by a multiresolution-based fusion strategy (Section 3). Here, a discrete wavelet transform (2 levels, Daubechies (2,2)) and a maximum selection rule were employed.*
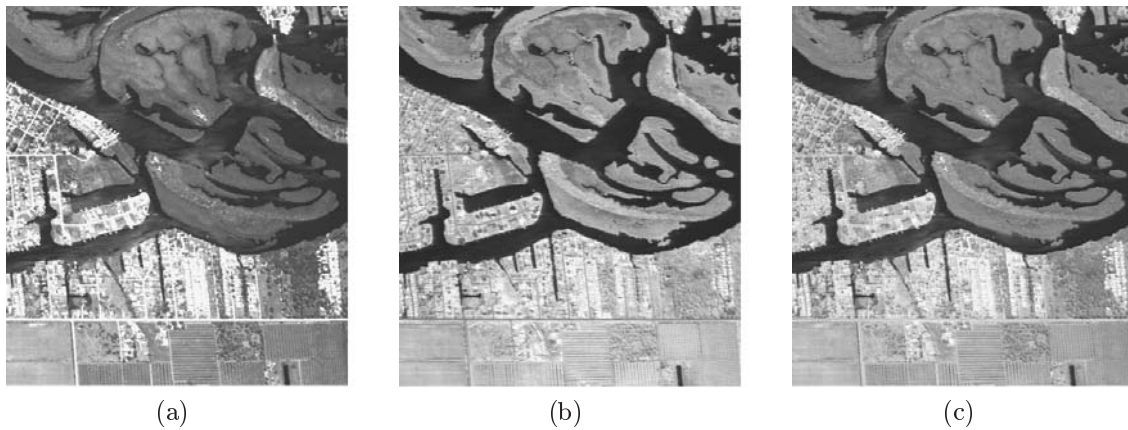

Figure 3: *Fusion of multispectral images: (a) image from band 1; (b) image from band 2; (c) fused image. The fused image has been obtained by a multiresolution-based fusion strategy (Section 3). In this case, a shift-invariant discrete wavelet transform (2 levels, Daubechies (2,2)) and a maximum selection rule were employed.*

Thematic Mapper multispectral images. The Landsat spectral bands allow for the classification of objects and areas in the scene, while the high spatial resolution resolution SPOT band locates more precisely the observed objects. One major challenge is to preserve the higher spatial resolution of the SPOT band (or from the available set of sources in the general case) without destroying the spectral information content provided by the Landsat bands (or by the lower resolution sources in the general case) [20, 75, 76].

Fig. 3 exemplifies the fusion of two bands of a multispectral scanner. Band 1 penetrates water and is useful for mapping along coastal areas, for soil-vegetation differentiation and for distinguishing forest types. In Fig. 3(a), buildings, roads and different agricultural zones are clearly discernible. Band 2 is more convenient for highlighting green vegetation and for detecting water-land interfaces. In Fig. 3(b), the bay is sharply delineated. The combined image (Fig. 3(c)) contributes to a better understanding of the objects observed and allows a more accurate identification.

Medical imaging

The fusion of multimodal images can be very useful for clinical applications such as diagnosis, modeling of the human body or treatment planning [39, 61, 62, 103, 110]. The next example illustrates the usage of fusion in radiotherapy and skull surgery. Here, the information provided by magnetic resonance imaging (MRI) and X-ray computed tomography (CT) is complementary. Normal and pathological soft tissues are better visualized by MRI (Fig. 4(a)), while the structure of tissue bone is
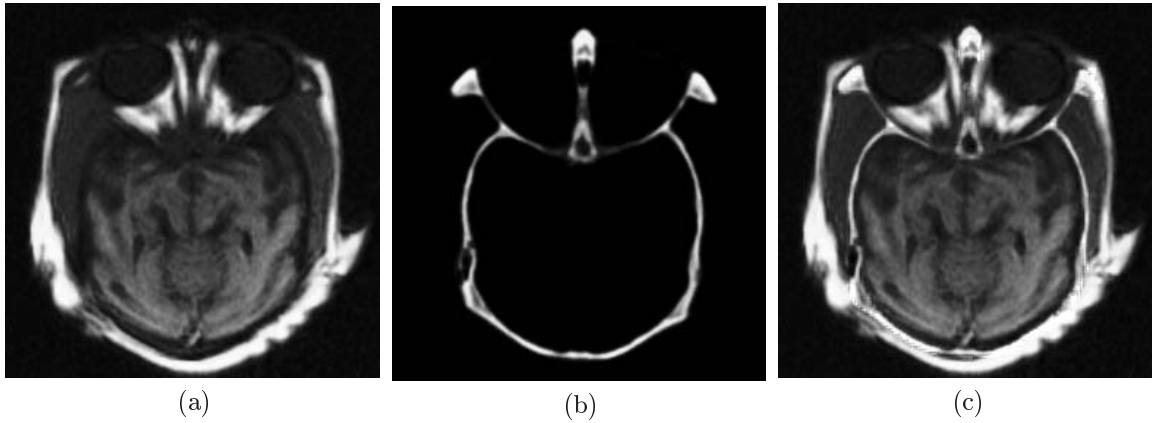
(a) (b) (c)

Figure 4: *Fusion of MRI and CT images: (a) MRI image; (b) CT image; (c) fused image. The fused image has been obtained by a multiresolution-based fusion strategy (Section 3). In this case, a morphological pyramid (2 levels, opening-closing and closing-opening filters) and a maximum selection rule were employed.*
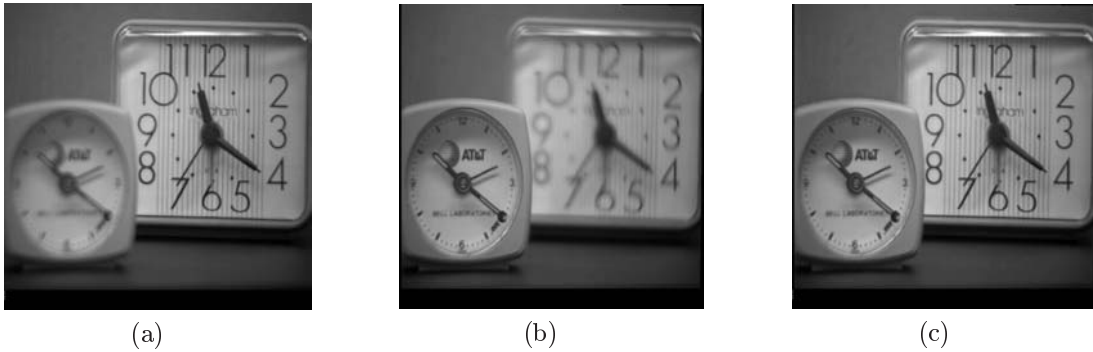

(a) (b) (c)

Figure 5: *Fusion of out-of-focus images: (a) image with focus on the right; (b) image with focus on the left; (c) fused image. The fused image has been obtained by a multiresolution-based fusion strategy (Section 3). Here, a Laplacian pyramid (3 levels) and a maximum selection rule were employed.*

better visualized by CT (Fig. 4(b)). The fused image, depicted in Fig. 4(c), not only provides salient information from both images simultaneously, but also the relative position of the soft tissue with respect to the bone structure. This can be useful for physicians in medical diagnosis.

Robotics and industrial engineering

Here, fusion is commonly used to identify the environment in which the robot or intelligent system evolves [1, 15]. It is also employed for navigation in order to avoid collisions and keep track of the trajectory [42, 67]. Image fusion is also employed in industry, for example, for the monitoring of factories or production lines [54], or for quality and defect inspection of products [77].

Fig. 5 shows how fusion can be used to extend the effective depth of field[3] of a vision system. Due to the limited depth of field of optical lenses, it is often not possible to get an image with all objects in focus. One way to overcome this problem is to take several recordings with different focus points and combine them into a single composite which contains the focused regions of all input images. This could be useful, for example, in digital camera design or in industrial inspection applications where the need to visualize objects at very short distances complicates the preservation of the depth of field.

---

[3]The depth of field is the range of distance from a camera that is acceptably sharp in the image obtained by that camera.

1.5 Fusion techniques

There are various techniques for image fusion, even at the pixel level [72]. The selection of the appropriate one depends strongly on the type of application. Here, we outline some of the most commonly used techniques in pixel-level fusion. We have grouped them into four major categories; however, this is a rather loose classification since these categories do overlap in various ways.

Weighted combination

A simple approach for fusion consists of synthesizing the fused image by averaging corresponding pixels of the image sources. An 'optimal' weighting can be determined, for example, by a principal component analysis of the correlation or covariance matrix of the sources [78]. The weightings for each input are obtained from the eigenvector corresponding to the largest eigenvalue. Variations of this method and other arithmetic signal combinations are numerous [5, 48].

Color space fusion

Image fusion by color transformations takes advantage of the possibility of representing data in different color channels. The simplest technique is to map the data from a sensor to a particular color channel. Many different band combinations and color spaces can be applied [101, 107]. The challenge is to generate an intuitive meaningful color fused image. Moreover, pseudocolor mappings can help identify sensor-specific details in a fused image display. That is, the use of color can be used to identify which sensor gave rise to the features appearing in the fused image [101]. The benefits of false-color imagery relative to monochromatic and unfused imagery in tasks such as detection and localization of targets have been studied in [100].

Optimization approach

In this approach, the methods are based on an a priori model of the real scene and the fusion task is expressed as an optimization problem. In Bayesian optimization, the goal is to find the fused image which maximizes the a posteriori probability. Some examples of probabilistic fusion schemes can be found in [41, 84]. In the Markov random field approach, the input images are modeled as Markov random fields to define a cost function which describes the fusion goal [4]. A global optimization strategy such as simulated annealing is employed to minimize this cost function.

Biologically-based approaches

One of the most popular instances of fusion in a biological system is the visual system of the rattlesnakes [69]. These vipers possess organs which are sensitive to thermal radiation. The IR signals provided by these organs are combined by bimodal neurons with the visual information obtained from the eyes. Inspired by this real-life example, several researchers have used neural networks to model multisensor image fusion [31, 107].

Another biologically-inspired fusion method is the approach based on multiresolution (MR) decompositions [2, 14, 16, 23, 51, 97, 114]. It is motivated by the fact that the human visual system is primarily sensitive to local contrast changes, i.e. edges, and MR decompositions provide a convenient spatial-scale localization of these local changes. The basic strategy of a generic MR fusion scheme is to use specific fusion rules to construct a combined MR representation from the MR representations of the different input sources. The fused image is then obtained by performing the inverse decomposition process.

Henceforth, we confine our discussion to multiresolution image fusion approaches. In particular, we focus on pixel and feature-level MR fusion schemes where the output is a single fused image which is constructed primarily for display on a computer monitor.

The rest of the report is organized as follows. In Section 2 we review the basics of MR decomposition theory. In Section 3 we present a general framework for pixel-based MR fusion. Within this framework, we describe some of the existing schemes in literature and show fused image examples of existing as well as new fusion schemes. In Section 4 we extend the previous framework and propose a region-based MR fusion strategy. We illustrate different ways of using the region information and present some experimental results using the region approach. In Section 5 we briefly discuss the topic of performance assessment. Finally, in Section 6, we present conclusions and suggest directions for further work.

It is to be noted that the fusion framework in Section 3 has been partially inspired by the MR

fusion methodology proposed by Zhang and Blum in [114]. The authors proposed also a region-based fusion algorithm [113]. Our approach, however, is different from theirs in several aspects.

2   Multiresolution decomposition schemes: an overview

A multiresolution decomposition scheme decomposes the signal being analyzed into several components, each of which captures information present at a given scale. The notion of resolution (or scale) relates to the size of the details that can be represented. These concepts are very useful in image processing for various reasons: (*i*) real-world objects usually consist of structures at different scales; (*ii*) there is strong evidence that the human visual system processes information in a multiresolution fashion; (*iii*) multiresolution methods lend themselves to effective designs of reduced-complexity algorithms.

In the following, we discuss in more detail the general concept of a decomposition system with perfect reconstruction and we explain how concatenation of such systems can lead to multiresolution decompositions. While there exists several types of multiresolution decompositions, we mainly concentrate on two well-known special classes: pyramids and wavelets. These are studied within the axiomatic framework proposed by Heijmans and Goutsias in [35, 38].

2.1   Decomposition systems with perfect reconstruction

When analyzing a signal, it is often useful to decompose it into different parts. Those parts can then be analyzed separately which may facilitate subsequent processing tasks. Of particular interest is the case where the signal is decomposed in such a way that no information is removed and the original signal can exactly be recovered (*perfect reconstruction*) from its constituting parts.

The idea of a decomposition system with perfect reconstruction is to obtain a more convenient representation (*analysis*) of the signal such that no information is lost, i.e., the signal can be recovered through some reconstruction process (*synthesis*). Fig. 6 depicts a general scheme for the decomposition of an input signal $x^{(0)} \in V_0$ into two components $(x^{(1)}, y^{(1)}) \in V_1 \times W_1$. Here, $x^{(1)}$ and $y^{(1)}$ can be interpreted as the *approximation* and *detail* signals of $x^{(0)}$, respectively. In other words, $x^{(1)}$ is a sort of 'simplification' of $x^{(0)}$, inheriting many of its properties, whereas $y^{(1)}$ is a kind of 'refinement' that contains the information that has been discarded in the simplification process.
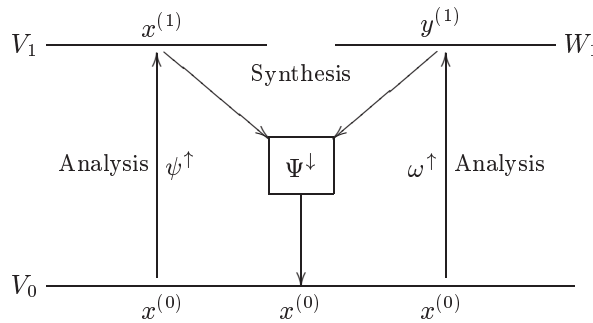

Figure 6: *A signal decomposition scheme with perfect reconstruction.*

The operators $\psi^{\uparrow} \colon V_0 \mapsto V_1$, $\omega^{\uparrow} \colon V_0 \mapsto W_1$ are called *analysis operators* and the operator $\Psi^{\downarrow} \colon V_1 \times W_1 \mapsto V_0$ is called the *synthesis operator*. The assumption that no information is lost by the decomposition is expressed by the requirement that $\Psi^{\downarrow}$ is the left inverse of $\Psi^{\uparrow} = (\psi^{\uparrow}, \omega^{\uparrow})$, i.e.,

$$\Psi^{\downarrow}\left(\psi^{\uparrow}(x^{(0)}), \omega^{\uparrow}(x^{(0)})\right) = x^{(0)}, \text{ for } x^{(0)} \in V_0 .$$

This condition will be referred to as the *perfect reconstruction condition*.

There are several ways of decomposing signals. Which one is appropriate depends on the application and the signal to be analyzed. We can adapt the decomposition to the signal being decomposed by selecting the appropriate analysis and synthesis operators.

In various signal and image applications, the decomposition $x^{(0)} \mapsto (x^{(1)}, y^{(1)})$ is only a first step toward an analysis of $x^{(0)}$. Subsequent steps comprise a decomposition of $x^{(1)}$ into $x^{(2)}$ and $y^{(2)}$, of $x^{(2)}$ into $x^{(3)}$ and $y^{(3)}$, and so forth. By concatenating several systems of the form depicted in Fig. 6 we obtain a *multilevel decomposition system*. If the higher levels are obtained by means of some spatial filtering (e.g., linear or morphological) of the lower level signals, possibly followed by a sampling step, then we call the system a *multiresolution* or *multiscale* decomposition scheme.

To formalize this procedure, assume that there exists a sequence of signal spaces V_k, $k \geq 0$, and detail spaces W_k, $k \geq 1$. At each level $k \geq 0$ we have two analysis operators, $\psi_k^\uparrow \colon V_k \mapsto V_{k+1}$ and $\omega_k^\uparrow \colon V_k \mapsto W_{k+1}$, and a synthesis operator $\Psi_k^\downarrow \colon V_{k+1} \times W_{k+1} \mapsto V_k$, satisfying the perfect reconstruction condition:

$$\Psi_k^\downarrow \left(\psi_k^\uparrow(x), \omega_k^\uparrow(x) \right) = x, \quad \text{for } x \in V_k. \tag{2.1}$$

A given input signal $x^{(0)} \in V_0$ can be decomposed by the recursive scheme

$$x^{(0)} \rightarrow \{y^{(1)}, x^{(1)},\} \rightarrow \{y^{(1)}, y^{(2)}, x^{(2)}\} \rightarrow \cdots \rightarrow \{y^{(1)}, \ldots, y^{(K-1)}, y^{(K)}, x^{(K)}\}, \tag{2.2}$$

where

$$\begin{cases} x^{(k+1)} = \psi_k^\uparrow(x^{(k)}) \\ y^{(k+1)} = \omega_k^\uparrow(x^{(k)}) \end{cases} \qquad k = 0, 1, \ldots, K-1. \tag{2.3}$$

Here, $x^{(k+1)}$ is an approximation of $x^{(k)}$, but can also be regarded as a 'k+1-th order' coarse approximation of the original signal $x^{(0)}$. In contrast, the detail signal $y^{(k+1)}$ contains information about $x^{(k)}$ that is not present in the simplified component $x^{(k+1)}$.

Note that, because of the perfect reconstruction condition, the original signal $x^{(0)}$ can be perfectly reconstructed from $x^{(K)}$ and $y^{(1)}, y^{(2)}, \ldots, y^{(K)}$ by means of the backward recursion:

$$x^{(k)} = \Psi_k^\downarrow(x^{(k+1)}, y^{(k+1)}), \qquad k = K-1, K-2, \ldots, 0. \tag{2.4}$$

Fig. 7 illustrates the analysis and synthesis schemes for the particular case where $K = 3$.
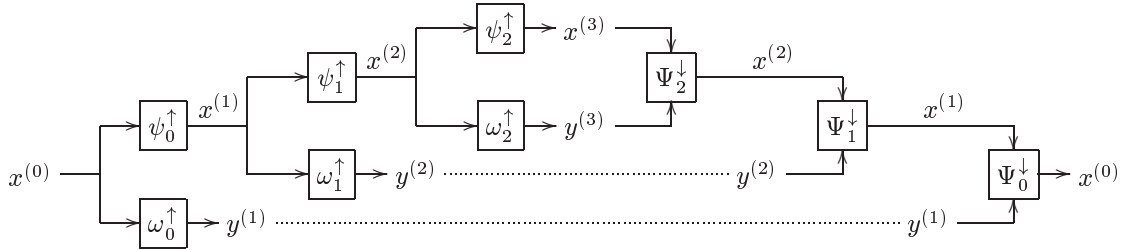

Figure 7: *A 3-level decomposition system: analysis and synthesis.*

2.2 The pyramid transform

The pyramid transform is characterized by the assumption that

$$\Psi_k^\downarrow(x, y) = \psi_k^\downarrow(x) + y, \quad \text{for } x \in V_{k+1}, y \in W_{k+1}, \tag{2.5}$$

where $W_{k+1} \subseteq V_k$ and $\psi_k^\downarrow \colon V_{k+1} \mapsto V_k$. The perfect reconstruction condition in (2.1) can be reformulated as

$$\psi_k^\downarrow \psi_k^\uparrow(x) + \omega_k^\uparrow(x) = x, \quad \text{for } x \in V_k.$$

Thus, $\omega_k^\uparrow(x) = x - \psi_k^\downarrow \psi_k^\uparrow(x)$ is the error of the synthesis operator $\psi_k^\downarrow$ when reconstructing x from the approximation $\psi_k^\uparrow(x)$. In this case, the recursive analysis scheme in (2.2) is given by

$$\begin{cases} x^{(k+1)} = \psi_k^\uparrow(x^{(k)}) \\ y^{(k+1)} = x^{(k)} - \psi_k^\downarrow(x^{(k+1)}) \end{cases} \qquad k = 0, 1, \ldots, K-1, \tag{2.6}$$

and the synthesis step in (2.4) is

$$x^{(k)} = \psi_k^{\downarrow}(x^{(k+1)}) + y^{(k+1)}, \qquad k = K-1, K-2, \dots, 0. \tag{2.7}$$

We refer to the decomposition process

$$x^{(0)} \mapsto \{y^{(1)}, \dots, y^{(K-1)}, y^{(K)}, x^{(K)}\}$$

by means of (2.6) as the *pyramid transform* of $x^{(0)}$, and to the process of synthesizing $x^{(0)}$ by means of (2.7) as the *inverse pyramid transform*. A block diagram illustrating the pyramid transform and its inverse is shown in Fig. 8.
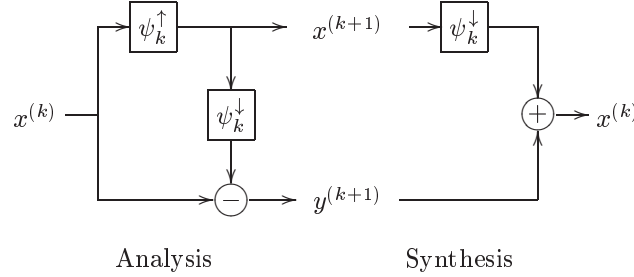

Analysis   Synthesis

Figure 8: *Pyramid transform (analysis) and its inverse (synthesis).*

We call the sequence

$$\{x^{(0)}, x^{(1)}, \dots, x^{(K)}\}$$

the *approximation pyramid*. For the particular case in which the analysis operators $\psi_k^{\uparrow}$ are linear, we also refer to it as the low-pass pyramid. Likewise, we call the sequence

$$\{y^{(1)}, y^{(2)}, \dots, y^{(K)}\}$$

the detail pyramid, and refer to it as the high-pass pyramid for the linear case.

The axiomatic pyramid approach described above encompasses several existing pyramid techniques. As an explicit example, we derive the well-known Laplacian pyramid introduced by Burt and Adelson in [12].

2.1 Example (Laplacian or Burt-Adelson pyramid). Let us consider that all spaces V_k are identical, namely $\ell^2(\mathbb{Z})$, the space of real-valued sequences $(\dots, x(-1), x(0), x(1), \dots)$ with $\sum_{n=-\infty}^{\infty} |x(n)| < \infty$. Consider also that at every level k the same analysis and synthesis operators $\psi^{\uparrow}$, $\psi^{\downarrow}$ are used. In particular, let us choose $\psi^{\uparrow}$ as a linear filter followed by a dyadic downsampling, i.e.,

$$\psi^{\uparrow}(x)(n) = (h * x)(2n) = \sum_{l=-\infty}^{\infty} h(l)x(2n - l)$$

and $\psi^{\downarrow}$ as a dyadic upsampling followed by a linear filter, i.e.,

$$\psi^{\downarrow}(x)(n) = (\tilde{h} * \bar{x})(n) = \sum_{l=-\infty}^{\infty} \tilde{h}(l)x(n - 2l).$$

Here, $\bar{x}$ denotes the upsampling of x, i.e., $\bar{x}(2n) = x(n)$ and $\bar{x}(2n + 1) = 0$ for every $n \in \mathbb{Z}$, and $h, \tilde{h} \in \ell^2(\mathbb{Z})$ are convolution kernels corresponding respectively to a smoothing and an interpolation filter. In [12] some design criteria are proposed for these filters. A particular solution is

$$h(-2) = h(2) = -\frac{1}{8} \qquad h(-1) = h(1) = \frac{1}{4} \qquad h(0) = \frac{3}{4} \qquad h(n) = 0 \quad \text{for other } n,$$

$$\tilde{h}(-1) = \tilde{h}(1) = \frac{1}{2} \qquad \tilde{h}(0) = 1 \qquad \tilde{h}(n) = 0 \quad \text{for other } n.$$

Or equivalently,

$$\psi^\uparrow(x)(n) = \tfrac{1}{8}\big(-x(2n-2) + 2x(2n-1) + 6x(2n) + 2x(2n+1) - x(2n+2)\big)$$

$$\begin{cases} \psi^\downarrow(x)(2n) = x(n) \\ \psi^\downarrow(x)(2n+1) = \tfrac{1}{2}\big(x(n) + x(n+1)\big). \end{cases}$$

Burt and Adelson called the sequence $\{x^{(k)}\}$, $k = 0, \ldots, K$, of approximation signals the Gaussian pyramid and the sequence $\{y^{(k)}\}$, $k = 1, \ldots, K$, of detail signals the Laplacian pyramid. Fig. 9 shows an example of the Gaussian (left) and Laplacian (right) pyramids involving sampling. The bottom or zero level of the Gaussian pyramid is equal to the original image $x^{(0)}$ which is depicted at the bottom left of Fig. 9. This is blurred (low-pass filtered) and subsampled to obtain the next level $x^{(1)}$, which is then filtered and subsampled in the same way to obtain $x^{(2)}$. Further repetitions of this filtering/subsampling procedure generate the remaining levels of the pyramid. As a low-pass filter, an approximation to a Gaussian filter is often used (hence the name Gaussian pyramid). The Laplacian pyramid can then be derived by interpolating each level $k > 0$ of the Gaussian pyramid and subtracting it from its predecessor. The resulting pyramid is depicted on the right part of Fig. 9. Now, each level contains only those image features lost from one level to the next, and therefore contains only details within a restricted range of sizes.

Obviously, the choice of different analysis and synthesis operators results in different kinds of pyramids. In particular, nonlinear pyramids have attracted a great deal of attention. A well-known example is the morphological pyramid [35, 96], where the analysis and synthesis operators are based on morphological operators such as erosions, dilations, openings and closings [83, 87]. Another instance of nonlinear pyramids is the ratio-of-low-pass pyramid [95]. Here, the ratio (rather than the standard difference) of the successive low-pass filtered signals is computed. In fact, the operator '+' used in (2.5) can be replaced by any invertible operation (see [35] for details).

Observe that a signal representation obtained by means of a pyramid transform (i.e., detail signals along with the coarsest approximation) is overcomplete in the sense that it produces more samples than the original signal. This is a direct consequence of the fact that the detail signal $y^{(k)}$ 'lives' at the same resolution[4] level as $x^{(k-1)}$.

2.3 The wavelet transform

This section describes another interesting family of decomposition systems, the *wavelet decomposition*. Again we follow the exposition by Heijmans and Goutsias [38] and adopt their notation.

A general wavelet decomposition has the structure depicted in Fig. 6, but in addition to the perfect reconstruction condition (2.1), i.e.,

$$\Psi^\downarrow\big(\psi^\uparrow(x), \omega^\uparrow(x)\big) = x, \quad \text{for } x \in V_0, \tag{2.8}$$

it satisfies the additional constraints

$$\psi^\uparrow\big(\Psi^\downarrow(x,y)\big) = x, \quad \text{for } x \in V_1, y \in W_1 \tag{2.9}$$

$$\omega^\uparrow\big(\Psi^\downarrow(x,y)\big) = y, \quad \text{for } x \in V_1, y \in W_1, \tag{2.10}$$

which guarantee that the decomposition is non-redundant. Note that (2.8)-(2.10) imply that the analysis operator $\Psi^\uparrow = (\psi^\uparrow, \omega^\uparrow)$ and the synthesis operator $\Psi^\downarrow$ are inverses of each other. Concatenation of a series of analysis steps yields a multiresolution decomposition called the *wavelet transform*.

Often, e.g. in the linear case, the synthesis operator $\Psi^\downarrow$ is of the special form

$$\Psi^\downarrow(x,y) = \psi^\downarrow(x) + \omega^\downarrow(y), \quad x \in V_1, y \in W_1. \tag{2.11}$$

In this case, we speak of an *uncoupled wavelet decomposition* and conditions (2.8)-(2.10) become

$$\psi^\downarrow\psi^\uparrow(x) + \omega^\downarrow\omega^\uparrow(x) = x, \quad x \in V_0 \tag{2.12}$$

$$\psi^\uparrow\big(\psi^\downarrow(x) + \omega^\downarrow(y)\big) = x, \quad x \in V_1, y \in W_1 \tag{2.13}$$

$$\omega^\uparrow\big(\psi^\downarrow(x) + \omega^\downarrow(y)\big) = y, \quad x \in V_1, y \in W_1. \tag{2.14}$$

---

[4]Here, the term 'resolution' refers to the size of the signal.
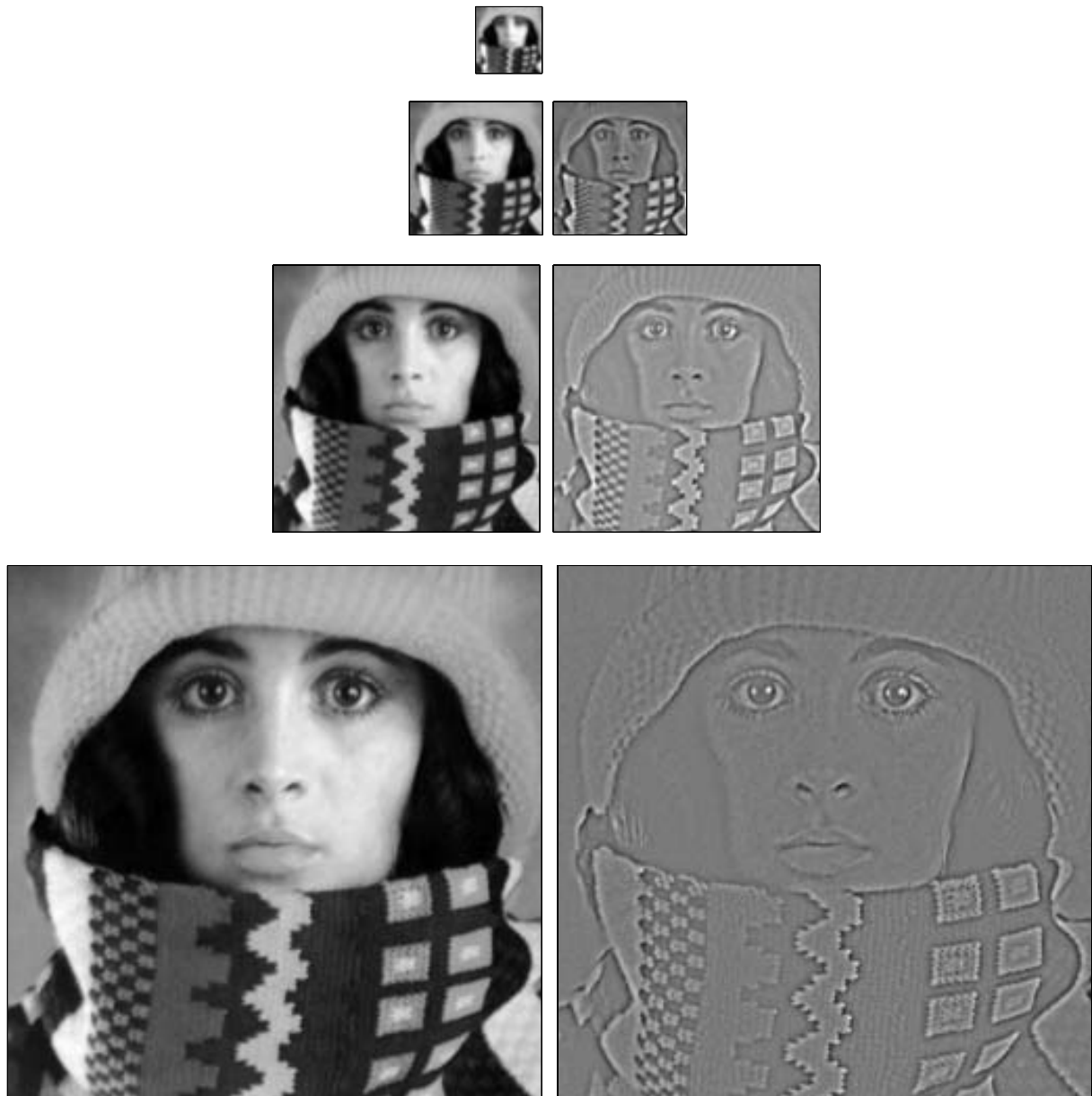
Figure 9: *Example of a Gaussian (left) and a Laplacian (right) pyramid. For the Gaussian pyramid (from bottom to top) images $x^{(0)}$, $x^{(1)}$, $x^{(2)}$ and $x^{(3)}$ are depicted. For the Laplacian (from bottom to top) images $y^{(1)}$, $y^{(2)}$ and $y^{(3)}$ are shown. The coarse approximation image $x^{(3)}$ in combination with the detail images $y^{(1)}$, $y^{(2)}$, $y^{(3)}$ provide an alternative (but redundant) representation of the original image $x^{(0)}$.*

We refer to $\psi^{\downarrow}$ as the *signal synthesis operator*, and to $\omega^{\downarrow}$ as the *detail synthesis operator*. Fig. 10 diagrams the corresponding synthesis part of an uncoupled wavelet decomposition for the multilevel case where $K = 3$.
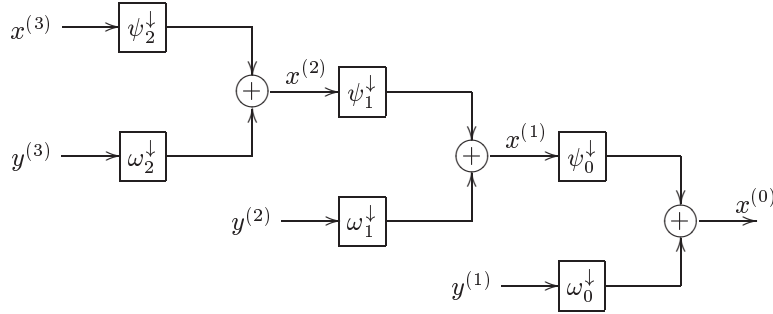

Figure 10: *Synthesis scheme of a 3-level uncoupled wavelet decomposition system.*

It was explained by Heijmans and Goutsias [38] how existing linear wavelets or filter banks can fit into this abstract wavelet scheme. As a way of illustration, we consider a one-dimensional dyadic biorthogonal wavelet transform. This can be implemented as a two-channel perfect reconstruction filter bank as depicted in Fig. 11. Here, the input signal $x^{(k)}$ is decomposed into an approximation
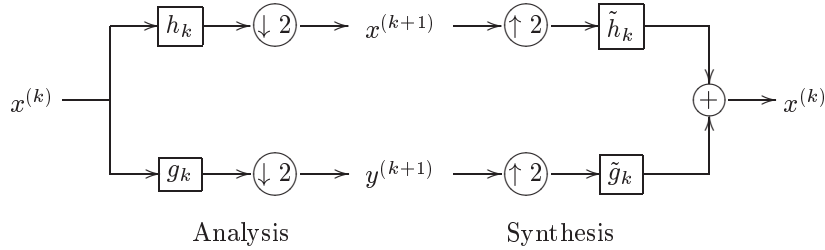

Figure 11: *Biorthogonal wavelet transform with low-pass and high-pass analysis filters h_k, g_k, respectively, and low-pass and high-pass synthesis filters $\tilde{h}_k$, $\tilde{g}_k$, respectively.*

signal $x^{(k+1)}$ and a detail signal $y^{(k+1)}$ by filtering with h_k and g_k, respectively, and downsampling. Thus, a one-dimensional discrete wavelet transform (DWT) is given by

$$x^{(k+1)}(n) = (h_k * x^{(k)})(2n), \quad y^{(k+1)}(n) = (g_k * x^{(k)})(2n).$$

Synthesis is achieved by upsampling signals $x^{(k+1)}$ and $y^{(k+1)}$, filtering with $\tilde{h}_k$ and $\tilde{g}_k$, respectively, and adding the respective outputs. Thus, the inverse DWT is given by

$$x^{(k)}(n) = (\tilde{h}_k * \bar{x}^{(k+1)})(n) + (\tilde{g}_k * \bar{y}^{(k+1)})(n).$$

For perfect reconstruction, the analysis and synthesis filters need to satisfy specific constraints known as biorthogonality conditions [56, 104]:

$$\sum_{l=-\infty}^{\infty} \tilde{h}_k(l) h_k(2n - l) = \sum_{l=-\infty}^{\infty} \tilde{g}_k(l) g_k(2n - l) = \delta(n) \tag{2.15}$$

$$\sum_{l=-\infty}^{\infty} \tilde{h}_k(l) g_k(2n - l) = \sum_{l=-\infty}^{\infty} \tilde{g}_k(l) h_k(2n - l) = 0. \tag{2.16}$$

One can easily establish the relation between these conditions and (2.12)-(2.14). Note, however, that the expressions in (2.12)-(2.14) (and more general (2.8)-(2.10)) are formulated in operator terms, and

do not require any sort of linearity assumption or inner product. This allows a broad class of nonlinear wavelet decomposition schemes [28, 30, 32, 36–38].

Under some additional conditions on the filters, the linear wavelet transform described above is orthonormal. Orthonormality implies that the energy of the signal is preserved under transformation. If these conditions are met, the synthesis filters are a reflected version of the analysis filters, and the high-pass filters are modulated versions of the low-pass filters, namely,

$$\tilde{h}_k(n) = h_k(-n), \quad \tilde{g}_k(n) = g_k(-n), \quad g_k(n) = (-1)^n h_k(M - n)$$

where M is an integer delay. Such filters are often known as quadrature mirror filters (QMF), conjugate quadrature filters or power-complementary filters.

One drawback of the (discrete) wavelet transform and, to a lesser extent, also for the pyramid transform, is that it generally yields a shift-variant signal representation. This means that a simple shift of the input signal may lead to complete different transform coefficients. The lack of translation invariance can be avoided if the outputs of the filter banks are not decimated. The resulting undecimated wavelet transform [56] yields a redundant MR representation where the approximation and detail signals have all the same size as the original signal.

Most wavelet and pyramid transforms have been designed in the one-dimensional case. By successive application of such one-dimensional transforms on the rows and the columns (or vice versa) of an image, one obtains a so-called *separable* two-dimensional transform. This construction is illustrated in Fig. 12 for the wavelet transform. At each level k, the input $x^{(k)}$ is decomposed into a coarse approximation $x^{(k+1)}$ and three detail signals $y^{(k+1)} = \{y^{(k+1)}(\cdot|1), y^{(k+1)}(\cdot|2), y^{(k+1)}(\cdot|3)\}$, corresponding to the horizontal, vertical and diagonal directions. Fig. 13 shows a 2-level wavelet decomposition computed
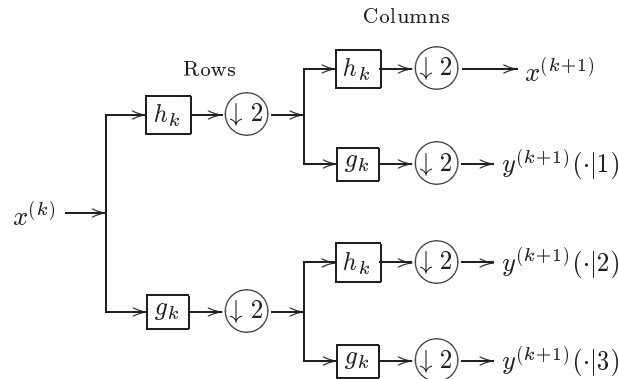
Figure 12: *Separable two-dimensional wavelet transform.*

in this way.

Non-separable transforms can also be constructed [38, 46, 86]. Although they provide decompositions with more general properties, they have been used less often in image applications due to the lack of general tools for their design.

2.4 Other multiresolution decompositions

In the previous sections we have focused on pyramids and wavelets because they are quite standard. However, there are many alternatives; for example, the wavelet packet representation introduced by Coifman and Wickerhauser [108], the local basis decompositions [56] or the multiwavelet construction [34, 89]. We briefly discuss some of these decompositions.

2.4.1 Wavelet packets

With a straightforward generalization of the two-channel filter bank presented in Section 2.3, we can obtain an even sparser representation of a signal. Instead of dividing only the approximation spaces

Figure 13: *Example of a 2-level discrete wavelet transform. In the upper-left quarter, the second level is displayed. Starting from the to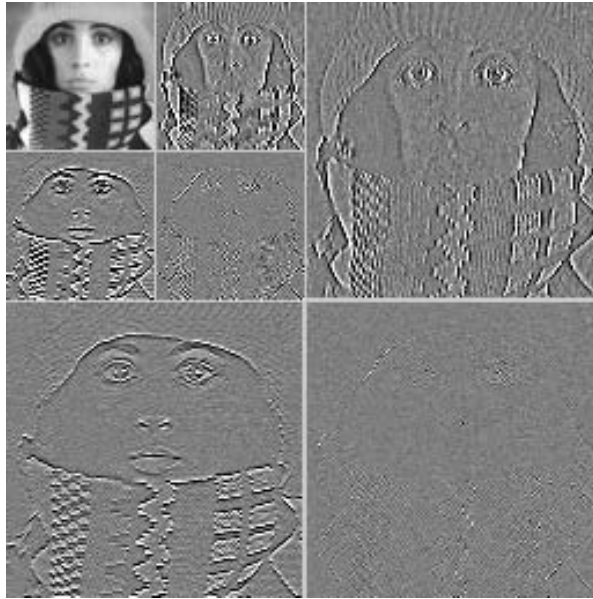p left and going clockwise: approximation, vertical, diagonal and horizontal detail images. The upper-right, bottom-right and bottom-left quarters show respectively the vertical, diagonal and horizontal first-level details.*
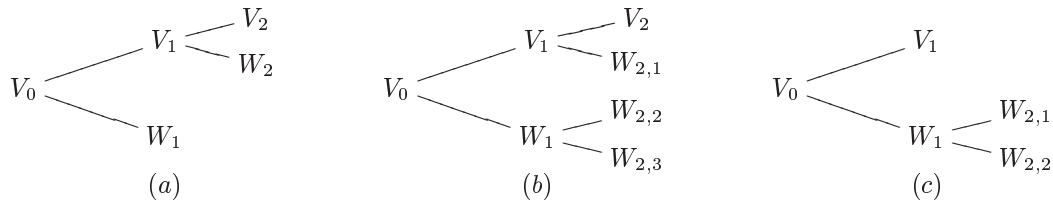

Figure 14: *Examples of tree-structure filter banks of depth 2: (a) only low-pass is split (standard wavelet); (b) full tree; (c) only high-pass is split.*

V_k to derive the approximation and detail spaces V_{k+1} and W_{k+1}, we divide the detail spaces as well. The recursive splitting of spaces can be represented in a binary tree as illustrated in Fig. 14.

At each node of the tree, we have the option to split or not. This allows the construction of an arbitrary dyadic tree structure. Each structure is associated with a function basis known as a *wavelet packet basis*. One possibility is the logarithmic tree, with low-pass iterations only (see Fig. 14(a)). In fact, the standard wavelet basis is an example of a wavelet packet basis of V_0, obtained by choosing such a logarithmic binary tree. Here, the frequency axis is decomposed in dyadic intervals whose sizes have an exponential growth as shown in the time-frequency plane[5] of Fig. 15. When the scale decreases, the frequency support of the wavelet is shifted toward high frequencies. The time resolution increases but the frequency resolution decreases. Thus, standard wavelet bases are suited for signals where high-frequency components have shorter duration than low-frequency components.

One can also construct wavelet packets corresponding with bases that have a better frequency resolution. They generalize the fixed dyadic construction of Fig. 15 by decomposing the frequency axis in intervals of varying sizes (see the tiling example of Fig. 16).

Clearly, an arbitrary tree-structure filter bank offers a very flexible frequency signal decomposition.

---

[5]In the time-frequency plane, a decomposition basis function is symbolically represented by a rectangle which indicates the time and frequency domains where the energy of this basis is concentrated.
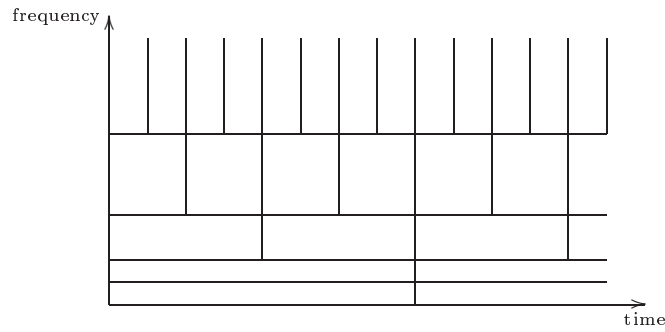
Figure 15: *Time-frequency plane of a standard wavelet basis. The frequency axis is decomposed in increasing dyadic intervals.*
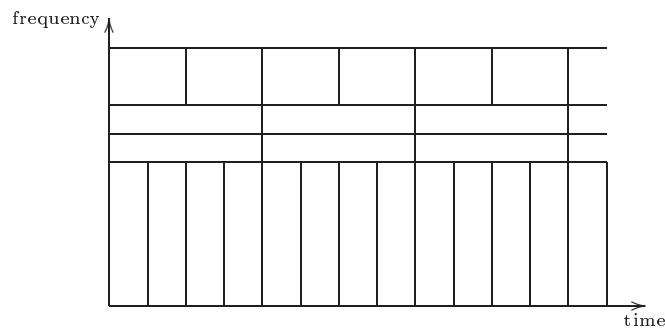

Figure 16: *A wavelet packet divides the frequency axis in separate intervals of varying sizes. A tiling is obtained by translating in time the wavelet packet basis covering each frequency interval.*

Given a signal (or class of signals) and a fixed set of filters, we can obtain the 'best' (according to some criterion) tree decomposition. The best wavelet packet basis algorithm described in [108] is often used to select a 'best' basis that minimizes a concave cost function.

From Fig. 16, we can see that wavelet packet bases are particularly well adapted to decompose signals that have different behavior in different frequency intervals. A best wavelet basis can thus be interpreted as a 'best' frequency segmentation. Note, however, that time-frequency tiles at a particular frequency have the same resolution. Thus, if the signal has properties that vary in time, it would be more appropriate to decompose the signal in a block basis that segments the time axis instead of the frequency axis. The local bases described next are examples of such a basis.

2.4.2   Local basis

A local basis divides the time axis into intervals of varying sizes as illustrated in Fig. 17. Of particular interest are the cosine bases [56], which are obtained by designing smooth windows that cover each time interval and multiplying them by cosine functions of different frequencies.

Similarly to wavelet packets, a local cosine tree can be constructed by recursively dividing spaces built with local cosine bases. As in the wavelet packet case, this offers the possibility of choosing a 'best' basis for a given signal. A best local cosine basis adapts the time segmentation to the variations of the signal time-frequency structures. In comparison with wavelet packets, we gain time adaptation but we lose frequency flexibility (since the frequency axis is being split with constant bandwidth).

Note that wavelet packets and local cosines are dual families of bases. Wavelet packets segment the frequency axis and are uniformly translated in time whereas local cosines divide the time axis and are uniformly translated in frequency. By combining the two dual concepts one can obtain arbitrary tilings of the time-frequency plane.
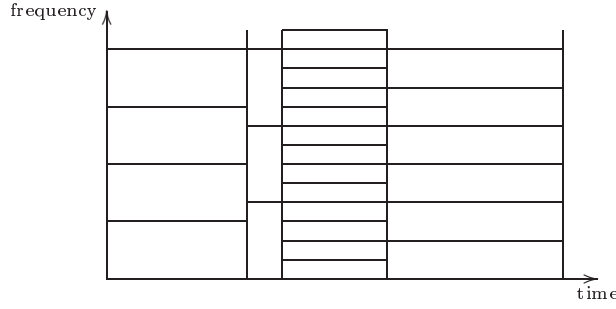
Figure 17: *A local basis divides the time axis in separate intervals of varying sizes. A tiling is obtained by translating in frequency the local basis covering each time interval.*

2.4.3 Multiwavelets

Multiwavelet decompositions offer more design flexibility by introducing (at each level) several analysis and synthesis operators. Multiwavelets have some advantages over scalar wavelets in relation to properties which are known to be important in signal processing such as short support, orthogonality, symmetry, and vanishing moments. A scalar wavelet, except for the Haar system, cannot posses all these properties at the same time [26]. In contrast, a multiwavelet system can simultaneously provide perfect reconstruction, orthogonality, linear-phase symmetry and a high order of approximation (vanishing moments) [88]. The main drawback, however, is that they are implemented with more complicated filter banks than the standard wavelet transforms.

The standard wavelet transform is a scalar transform: it uses one analysis operator $\psi_k^\uparrow$ for the computation of the approximation signal $x^{(k+1)}$, and one analysis operator $\omega_k^\uparrow$ for the computation of the detail signal $y^{(k+1)}$. A multiwavelet transform generalizes the scalar wavelet transform: the approximation signal $x^{(k+1)}$ is generated by M analysis operators $\boldsymbol{\psi}_k^\uparrow = (\psi_{k,1}^\uparrow, \ldots, \psi_{k,M}^\uparrow)$ and the detail signal $y^{(k+1)}$ by M analysis operators $\boldsymbol{\omega}_k^\uparrow = (\omega_{k,1}^\uparrow, \ldots, \omega_{k,M}^\uparrow)$.

In the same way that the scalar wavelet transform may be obtained by iterating a two-channel filter bank on its low-pass output (see Section 2.3), a multiwavelet transform can be obtained by a matrix-valued filter bank with 'coefficients' that are $M \times M$ matrices. Each input sample $\boldsymbol{x}(n)$ is a vector with M components The resulting two-channel $M \times M$ matrix filter bank operates on M input data streams, filtering them into $2M$ output streams.

Let us illustrate this for $M = 2$. We start with a signal $x^{(0)}$ which is pre-processed (e.g, by repeating each sample) to produce the sequences $x_1^{(0)}$ and $x_2^{(0)}$. Their coarse approximation is computed with the low-pass branch of the multiwavelet filter bank:

$$\begin{pmatrix} x_1^{(1)}(n) \\ x_2^{(1)}(n) \end{pmatrix} = \sum_{l=-\infty}^{\infty} \boldsymbol{h}(l) \begin{pmatrix} x_1^{(0)}(2n-l) \\ x_2^{(0)}(2n-l) \end{pmatrix},$$

where each 'coefficient' $\boldsymbol{h}(l)$ is a 2×2 matrix. Analogously, the details are computed with the high-pass branch of the multiwavelet filter bank:

$$\begin{pmatrix} y_1^{(1)}(n) \\ y_2^{(1)}(n) \end{pmatrix} = \sum_{l=-\infty}^{\infty} \boldsymbol{g}(l) \begin{pmatrix} x_1^{(0)}(2n-l) \\ x_2^{(0)}(2n-l) \end{pmatrix}.$$

This one-level decomposition is shown in Fig. 18. Full multiwavelet decomposition of the signal $x^{(0)}$ is obtained by iterative filtering of the approximation coefficients $\boldsymbol{x}^{(k)}(n) = \left(x_1^{(k)}(n), x_2^{(k)}(n)\right)^T$ (where T denotes transpose). The original signal can be reconstructed from the multiwavelet coefficients by means of the synthesis equation:

$$\boldsymbol{x}^{(k)} = \sum_{l=-\infty}^{\infty} \tilde{\boldsymbol{h}}(l)\boldsymbol{x}^{(k+1)}(n-2l) + \sum_{l=-\infty}^{\infty} \tilde{\boldsymbol{g}}(l)\boldsymbol{y}^{(k+1)}(n-2l).$$

Here, $\boldsymbol{x}^{(k+1)} = (x_1^{(k+1)}, x_2^{(k+1)})^T$, $\boldsymbol{y}^{(k+1)} = (y_1^{(k+1)}, y_2^{(k+1)})^T$ and $\tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}}$ are the synthesis multifilters.
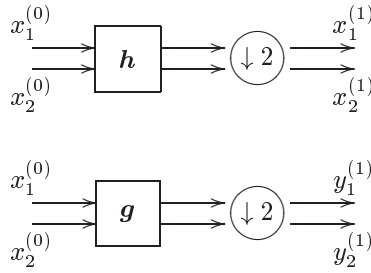
$$x_1^{(0)} \quad \boxed{\boldsymbol{h}} \quad \downarrow 2 \quad x_1^{(1)}$$
$$x_2^{(0)} \qquad\qquad\qquad x_2^{(1)}$$

$$x_1^{(0)} \quad \boxed{\boldsymbol{g}} \quad \downarrow 2 \quad y_1^{(1)}$$
$$x_2^{(0)} \qquad\qquad\qquad y_2^{(1)}$$

Figure 18: *Multiwavelet analysis filter bank with $M = 2$. An input $x^{(0)}$ is vectorized into $\boldsymbol{x}^{(0)} = (x_1^{(0)}, x_2^{(0)})^T$. Each row of the multifilter $\boldsymbol{h}$ (and the same applies to $\boldsymbol{g}$) is a combination of two ordinary filters, one operating on $x_1^{(0)}$ and other operating on $x_2^{(0)}$.*

As in the scalar case, the filters need to satisfy the biorthogonality conditions:

$$\sum_{l=-\infty}^{\infty} \tilde{\boldsymbol{h}}(l)\boldsymbol{h}(2n-l) = \sum_{l=-\infty}^{\infty} \tilde{\boldsymbol{g}}(l)\boldsymbol{g}(2n-l) = \delta(n)\,\mathsf{I} \qquad (2.17)$$

$$\sum_{l=-\infty}^{\infty} \tilde{\boldsymbol{h}}(l)\boldsymbol{g}(2n-l) = \sum_{l=-\infty}^{\infty} \tilde{\boldsymbol{g}}(l)\boldsymbol{h}(2n-l) = \mathsf{O}. \qquad (2.18)$$

Here, I, O denote the 2×2 identity and null matrices respectively.

Note that we have M times as many filters as in the classical wavelet case. One input signal x produces $2M$ outputs from the analysis bank. This means that the signal x has to be 'vectorized' so that M input streams $(x_1, \dots, x_M)$ go together. The most obvious way to get M inputs from a given signal is to repeat the signal so that M identical streams go into the multifilter bank. This produces an overcomplete representation. A different way is to pre-process the given scalar signal so that if the data enters at rate r, pre-processing yields M streams at rate r/M for input to the multifilter, which then produces $2M$ output streams (M for the approximation and M for the detail), each at rate $r/2M$. In either case, the pre-processing should be reversible so that after the synthesis step, post-processing can recover the original signal x.

2.4.4 Steerable pyramid

The steerable pyramid is an overcomplete, linear, multiresolution and multiorientation image decomposition where the analysis and synthesis operators are (in the simplest case) derivative operators with different supports and orientations. The associated filters are such that the resulting transform is self-inverting (i.e., the synthesis filters are just a reflected version of the analysis filters) and, moreover, it is translation and rotation invariant. In the following, we briefly describe the construction and implementation of the steerable pyramid; a more detailed account can be found in [86].

The steerable pyramid transform is implemented as a filter bank consisting of polar-separable filters. For simplicity, we consider the filters in the frequency domain. Thus, a filter $h(n, m)$ in the spatial domain is expressed as $H(u, v)$ in the frequency domain, where u and v are the frequency variables corresponding to the two spatial directions. More precisely, $H(u, v) = \sum_{n=\infty}^{\infty} \sum_{m=\infty}^{\infty} h(n, m)e^{-iun}e^{-ivm}$. For future convenience, we define $\boldsymbol{u} = (u, v)$ and $H^*(\boldsymbol{u}) = H(-\boldsymbol{u})$.

The frequency tiling of a one-level steerable pyramid decomposition is shown in Fig. 19 for the case of two orientation bands (i.e., $P = 2$). The corresponding diagram for this decomposition is depicted in Fig. 20. The filters B_p, $p = 1, 2$, are oriented band-pass filters, H_1 is a narrowband low-pass filter, G_0 is a non-oriented high-pass filter and H_0 is a low-pass filter.

As illustrated in Fig. 19, the band-pass filters together act as a circular symmetric band-pass filter. The low-pass filter H_1 passes the low-frequency components that fall inside the central core of that circular filter, while the high-pass filter G_0 passes the high frequency information that falls outside. In this way, the entire signal, regardless of its frequency, is passed to one of the output channels.

The system diagram for the steerable pyramid (both analysis and synthesis) is depicted in Fig. 21. Initially, the image is separated by the pre-processing filters H_0 and G_0 into a low and a high-pass
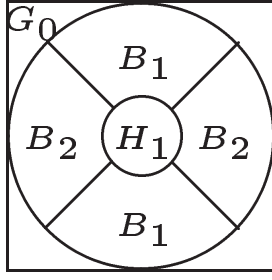
Figure 19: *Frequency tiling after 1-level decomposition of a steerable pyramid transform with $P = 2$. The frequency plane has been decomposed into a low-pass band (after filtering by H_1), two oriented band-pass components (after filtering by B_p, $p = 1, 2$), and a high-pass band (after filtering by G_0). The depicted squared region corresponds to the frequency range $[-\pi, \pi] \times [-\pi, \pi]$.*
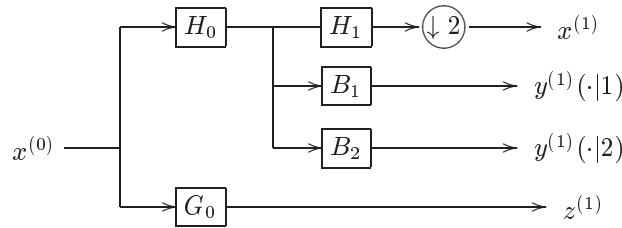

Figure 20: *First level of a steerable transform with $P = 2$.*

subbands. We denote the high-pass signal by $z^{(1)}$. The low-pass branch is then divided into a set of P oriented detail images and one approximation image. The detail images $y^{(1)}(\cdot|p)$, $p = 1, \dots, P$, are obtained using the band-pass filters $B_1, \dots, B_P$; while the approximation signal $x^{(1)}$ is obtained using a low-pass filter H_1 followed by a dyadic downsampling. The process of splitting into P details and one approximation is iterated on the approximation image (thus, filters H_0 and G_0 are not used in the successive levels).

In order to ensure that the transform is invertible as well as jointly invariant in orientation and position, the filters must satisfy specific radial (scale) and angular (orientation) frequency constraints [85]. The radial frequency constraints are:

1. Band limiting to prevent aliasing in the subsampling operation:

$$H_1(\boldsymbol{u}) = 0 \qquad \text{for } \|\boldsymbol{u}\| > \pi/2 \,,$$

where $\|\boldsymbol{u}\| = \sqrt{u^2 + v^2}$.

2. Flat system response to avoid amplitude distortion:

$$\left|H_0(\boldsymbol{u})\right|^2 \left(\left|H_1(\boldsymbol{u})\right|^2 + \sum_{p=1}^{P} \left|B_p(\boldsymbol{u})\right|^2\right) + \left|G_0(\boldsymbol{u})\right|^2 = 1 \,. \tag{2.19}$$

3. Recursion. The low-pass branch of the system must be unaffected by the iteration process:

$$\left|H_1(\boldsymbol{u}/2)\right|^2 \left(\left|H_1(\boldsymbol{u})\right|^2 + \sum_{p=1}^{P} \left|B_p(\boldsymbol{u})\right|^2\right) = \left|H_1(\boldsymbol{u}/2)\right|^2 \,. \tag{2.20}$$

A sufficient condition for (2.20) to hold is that the decomposition/reconstruction filter bank has unitary gain for low frequencies:

$$\left|H_1(\boldsymbol{u})\right|^2 + \sum_{p=1}^{P} \left|B_p(\boldsymbol{u})\right|^2 = 1 \,.$$

In this case, (2.19) implies that the pre and post-processing steps must also have a unitary gain, that is,

$$|H_0(\boldsymbol{u})|^2 + |G_0(\boldsymbol{u})|^2 = 1 .$$

Typically, $H_0(\boldsymbol{u}) = H_1(\boldsymbol{u}/2)$, so that the initial low-pass shape is the same as that used within the iteration. Thus, during the iteration $H_1(\boldsymbol{u}/2)$ plays the role of the initialization filter $H_0(\boldsymbol{u})$.

The angular constraint on the band-pass filters B_p requires these filters to form a steerable basis[6] [86]. In the simple case where the basis functions of the decomposition are directional derivative operators, the angular constraint can be expressed as

$$B_p(\boldsymbol{u}) = B(\boldsymbol{u})\big(- i \cos(\theta - \theta_p)\big)^{P-1} ,$$

where $i = \sqrt{-1}$, $\theta = \arctan(v/u)$, $\theta_p = \pi\frac{p-1}{P}$ for $p = 1, \dots, P$, and

$$B(\boldsymbol{u}) = \left(\sum_{p=1}^{P} |B_p(\boldsymbol{u})|^2 \right)^{1/2} .$$

The pyramid can be designed to produce any number of orientation bands P, resulting in an overcomplete transform by a factor of $4P/3$. The overcompleteness limits its efficiency but increases its applicability for many image processing tasks such as orientation and phase analysis, texture synthesis and noise removal.
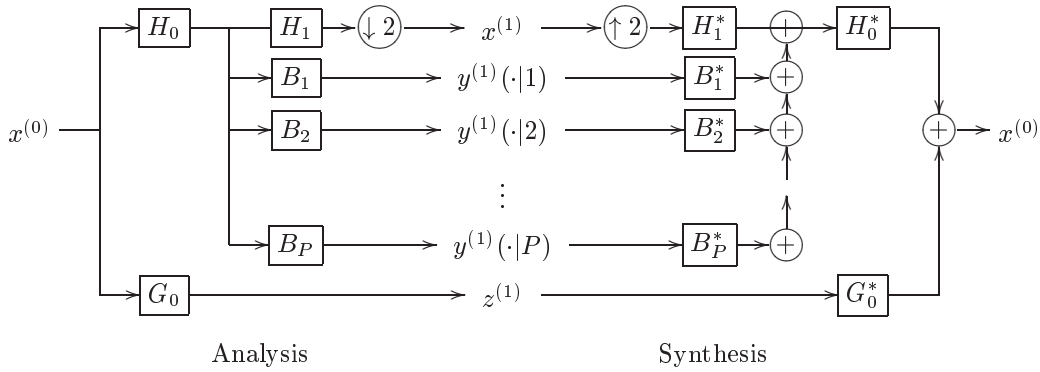
Figure 21: *Steerable transform (analysis) and its inverse (synthesis).*

2.4.5 Gradient pyramid

A gradient pyramid [11] is obtained by applying a gradient operator to each level of the Gaussian pyramid $\{x^{(k)}\}$, $k = 0, \dots, K$. Each image $x^{(k)}$ is filtered by a set of four oriented gradient filters g_p, $p = 1, \dots, 4$. The resulting filtered subbands correspond to the detail images

$$y^{(k+1)}(\cdot|p) = g_p * x^{(k)} , \qquad p = 1, \dots, P ,$$

representing the horizontal, vertical and the two diagonals directions. To reconstruct the original image from this gradient decomposition, a Laplacian pyramid is constructed as intermediate result. First, a (derivative) synthesis filter $\tilde{g}_p$ is applied to $y^{(k+1)}(\cdot|p)$, $p = 1, \dots, 4$. A Laplacian pyramid $\{y_L^{(k+1)}\}$, can then be obtained by summing up, at each level, the filtered resulting images, i.e.,

$$y_L^{(k+1)} = \sum_{p=1}^{4} \tilde{g}_p * y^{(k+1)}(\cdot|p) .$$

---

[6]A set of filters forms a steerable basis if they are rotated versions of each other and a version of the filters at any orientation may be synthesized as a linear combination of the basis filters. The simplest example of a steerable basis is a set of P directional derivatives of order $P - 1$.

Fig. 22 illustrates one level of the gradient pyramid transform. Here, h is the low-pass filter used to construct the Gaussian pyramid $\{x^{(k)}\}$, $k = 0, \dots, K$, and $\tilde{h}$ its corresponding synthesis filter. In [11], the filter h it is assumed to be of the form $h = \dot{h} * \dot{h}$ with $\dot{h}$ being a 3×3 binomial filter. Then, it can be shown that perfect reconstruction is possible by taking

$$g_p = (1 + \dot{h}) * d_p \qquad \text{and} \qquad \tilde{g}_p = -\frac{1}{8} d_p \,,$$

where d_p, $p = 1, \dots, 4$ are the derivative filters:

$$d_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad d_2 = \begin{pmatrix} 1 & -1 \end{pmatrix}$$
$$d_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad d_4 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \,.$$
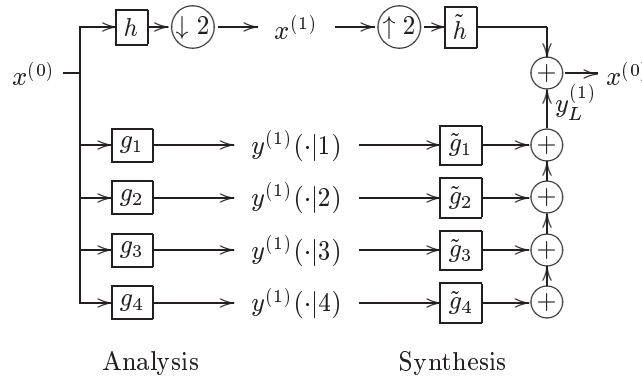

Figure 22: *Gradient pyramid transform (analysis) and its inverse (synthesis).*

2.5 Lifting

In this subsection, we describe a general and flexible technique to modify a given decomposition system into another one, possibly with some improved characteristics. This technique, called *lifting*, was developed by Sweldens in the context of wavelets [27,90,91]. The formulation we give here is based on the standard lifting approach proposed by Sweldens (see [38] for a more abstract formulation).

A general lifting scheme starts with an invertible transform of the input signal $x^{(0)} \in V_0$ into two parts, the *approximation signal* $x^{(1)} \in V_1$, and the *detail signal* $y^{(1)} \in W_1$. In general, this partition is the outcome of a particular wavelet transform. The most simple case is the one where the domain of the signal is subdivided into two disjoint subsets. In the one-dimensional case, this subdivision often comprises the even and odd samples, e.g., $x^{(1)}(n) = x^{(0)}(2n)$, $y^{(1)}(n) = x^{(0)}(2n+1)$. This latter decomposition, known in signal processing as polyphase decomposition, is sometimes called the *lazy wavelet transform*.

Two types of lifting schemes can be distinguished: *prediction lifting* and *update lifting*. We treat both cases separately.

Prediction lifting

The detail signal $y^{(1)}$ is predicted using information contained in the approximation signal $x^{(1)}$ and is replaced by the prediction error

$$y'^{(1)} = y^{(1)} - P(x^{(1)}) \,,$$

where $P : V_1 \mapsto W_1$ represents the prediction operator. The prediction error $y'^{(1)}$ becomes the new detail signal. Clearly, the original signal $x^{(0)}$ can be reconstructed from $x^{(1)}$ and $y'^{(1)}$ by

$$x^{(0)} = \Psi^{\downarrow}(x^{(1)}, y^{(1)}) = \Psi^{\downarrow}\left(x^{(1)}, y'^{(1)} + P(x^{(1)})\right) \,.$$

Update lifting

The approximation signal $x^{(1)}$ is updated using information contained in the detail signal $y^{(1)}$:

$$x'^{(1)} = x^{(1)} + U(y^{(1)})\,.$$

Here $U : W_1 \mapsto V_1$ represents the update operator. Generally, the update operator is chosen in such a way that the resulting signal $x'^{(1)}$ satisfies a certain constraint. For example, one might require that the mapping $x^{(0)} \mapsto x'^{(1)}$ preserves a given signal attribute such as the average or some local maximum. As before, the original signal can be easily reconstructed from $x'^{(1)}$ and $y^{(1)}$:

$$x^{(0)} = \Psi^{\downarrow}(x^{(1)}, y^{(1)}) = \Psi^{\downarrow}(x'^{(1)} - U(y^{(1)}), y^{(1)})\,.$$

Thus, an existing decomposition system with perfect reconstruction can be modified by an arbitrary prediction or update lifting step. Perfect reconstruction is guaranteed by the very structure of this scheme and does not require any particular assumptions on the lifting operators P and U. Moreover, the operators '+' and '−' used in the above expressions, can be replaced by any pair of invertible operators. This flexibility has challenged researchers to develop various nonlinear wavelet transforms [18,33,71], including morphological ones [38].

Fig. 23 illustrates a prediction-update lifting scheme. At analysis, the prediction operator P acting on x is used to modify y, resulting in a new detail signal y'. Subsequently, the update operator U acting on y' is used to modify x, yielding a new approximation signal x'. At synthesis, the signals x and y are reconstructed by reversing the lifting steps.
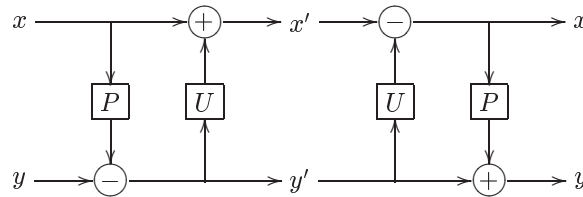
Figure 23: *Classical lifting scheme.*

Obviously, we can concatenate any number and type of lifting steps in order to modify a given decomposition system. In practice, these lifting steps are chosen in such a way that the resulting decomposition is an 'improvement' of the original one. Here, the word 'improvement' can have various meanings. For example, the lifted wavelet may have more vanishing moments than the original one, or it may be better able to decorrelate the signals within a given class, etc. In the context of linear wavelets, it has been shown that any system using finite impulse response (FIR) filters can be decomposed into elementary lifting steps [27].

2.6 Multiresolution decompositions and their application to image processing

MR decompositions are particularly useful in image and video processing and in computer vision. We conclude this section by giving some examples of image applications that can benefit from the use of MR transforms.

Compression

MR decompositions allow for efficient representation capturing the essence of the image with only a small set of significant coefficients. This is based on the fact that most images have correlation both in space and frequency. Thus, the new representations are sparse in the sense that most detail images contain few significant pixels (little significant detail) and therefore they can be represented with a small number of bits. Moreover, because the human visual system tends to be more sensitive to errors in low-frequency image components than in high-frequency ones, detail images at higher levels can be omitted and still one can obtain a good approximation of the original image. In this way, we can achieve high compression ratios without noticeable degradation.

Coarse-fine search techniques

Suppose we need to locate a large complex pattern within an image. Rather than attempting to convolve the whole pattern with the image, one may perform an approximate search by convolving a reduced-resolution pattern with a reduced-resolution version of the image. This serves to roughly locate possible occurrences of the target pattern with a minimum of computational effort. Next, higher-resolution copies of the pattern and image are used to refine the position estimates. Computation is kept to a minimum by restricting the search to neighborhoods of the points identified at the coarser resolution.

Image enhancement

Image enhancement is another area where MR decompositions can be used to reduce random noise in a degraded image while sharpening details of the image itself. This application is based on the fact that images and noise possess rather distinct properties in the transform domain. The detail coefficients in each level of the MR representation are passed through some kind of thresholding function where small values (which are likely to include most of the noise) are set to zero, while larger values (which include prominent image features) are retained. The final enhanced image is obtained by reconstructing the levels of the processed MR representation.

Image fusion

Since the essential goal of fusion is to preserve image features from the sources, a plausible approach is to transform the images into representations that decompose the images into relevant features such as edges, and perform fusion in this domain. A MR representation facilitates this type of analysis because it decomposes an image into different scales while preserving locality in space. Fusion using MR decompositions will be described comprehensively in Section 3.

2.7 Notation

Before continuing with the next section, we fix some notation and expressions for MR image decompositions. Given an input image $x^{(0)}$ and analysis and synthesis operators which satisfy (2.1) for $k = 0, \ldots, K - 1$, we can represent $x^{(0)}$ as a sequence of detail images at all levels along with the coarsest approximation image. Henceforth, the MR decomposition of an image $x^{(0)}$ is denoted by y and it is assumed to be of the form:

$$y = \{y^{(1)}, y^{(2)}, \ldots, y^{(K)}, x^{(K)}\}. \tag{2.21}$$

Here $x^{(K)}$ represents the approximation image at the highest level (lowest resolution) of the MR structure, while images $y^{(k)}$, $k = 1, \ldots, K$, represent the detail images at level k. The detail at level k will, in general, comprise various frequency or orientation bands, depending on the type of MR transform that has been used. We assume henceforth that $y^{(k)}$ is composed of P detail images, i.e., $y^{(k)} = \{y^{(k)}(\cdot|1), \ldots, y^{(k)}(\cdot|P)\}$.

Let $I_x^{(k)}$ and $I_y^{(k)}(p)$ denote the domain of $x^{(k)}$ and $y^{(k)}(\cdot|p)$ respectively. We use the vector coordinate $\boldsymbol{n} = (n, m)$ to index the location of the coefficient. Then, $x^{(k)}(\boldsymbol{n})$, where $\boldsymbol{n} \in I_x^{(k)}$, represents the approximation coefficient at location $\boldsymbol{n}$ within level k. Similarly, $y^{(k)}(\boldsymbol{n}|p)$, where $\boldsymbol{n} \in I_y^{(k)}(p)$, represents the detail coefficient at location $\boldsymbol{n}$ within level k and band p. Note that $I_x^{(k)}$ is not necessarily equal to $I_y^{(k)}(p)$. In the pyramid case, for example, $y^{(k)}$ represents a detail image of the same size as $x^{(k-1)}$, while in the standard wavelet transform $y^{(k)}(\cdot|p)$ is a detail image of the same dimensions as $x^{(k)}$. Note also that in most cases $I_y^{(k)}(p)$ does not depend on p.

For convenience, we will sometimes denote the approximation image $x^{(k)}$ by $y^{(k)}(\cdot|0)$. In this way, we can use the general expression of $y^{(k)}(\cdot|p)$ to refer both to the detail images (for $p = 1, \ldots, P$) and the approximation image (for $p = 0$). If no confusion is possible, we will use the shorthand notation $(\cdot)$ to denote $(\boldsymbol{n}|p)$; e.g., we will write $y^{(k)}(\cdot)$ rather than $y^{(k)}(\boldsymbol{n}|p)$.

3 The general pixel-based MR fusion scheme

The basic idea underlying the MR-based image fusion approach is to perform a MR transform on each source image and, following some specific fusion rules, construct a composite MR representation from

these inputs. The fused image is obtained by applying the inverse transform on this composite MR representation. This process is illustrated in Fig. 24.
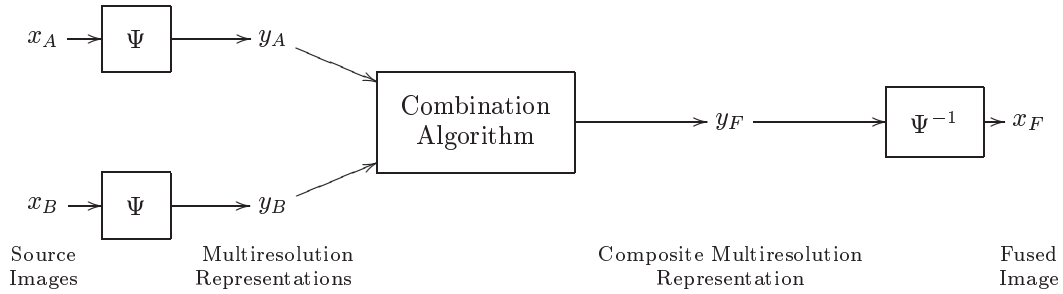

Figure 24: *Multiresolution image fusion scheme. Left: multiresolution transform Ψ of the sources; middle: combination in the transform domain; right: inverse multiresolution transform Ψ^{-1} of the composite representation.*

At the MR decomposition (analysis) stage, the data is transformed into a convenient representation which, besides scale or resolution, may also involve orientation or wavelength or some other physical parameters. At the combination stage, the actual fusion of the (transformed) data takes place. This involves identifying the salient information and transferring it into the fused image. This process, i.e., the way to combine the data, is governed by a number of rules called the *fusion rules*. The result is a composite MR representation from which the output fused image is obtained by application of the inverse MR transform (synthesis).

In the literature one finds several variants of the MR fusion scheme. In what follows, we present a general framework which encompasses most of them. Section 3.1 describes the various modules the framework consists of, while Section 3.2 discusses some of the possible alternatives for their construction. Within the framework, some of the existing algorithms proposed in literature are reviewed in Section 3.3. Examples of such schemes as well as other implementation alternatives are given in Section 3.4.

3.1   The general framework

In Fig. 25 we show a more detailed version of the fusion scheme of Fig. 24, in which the combination algorithm has been specified. In our framework, the combination algorithm consists of four modules: the *activity* and *match* measures extract information from the MR decompositions y_S, which is then used by the *decision* and *combination* map to compute the MR decomposition y_F of the fused image. Below, we give a short description of each of the building blocks. Note, however, that some of them, such as the 'match block' are optional. Furthermore, the routines comprised by the various blocks can be chosen in a variety of ways (see Section 3.2).

MR analysis (Ψ)

This block computes a MR decomposition of the input sources. We assume that the same type of transform Ψ is applied to all sources x_S, $S \in \mathcal{S}$, where $\mathcal{S}$ is the index set of source images. Thus, for every input x_S we obtain its MR representation $y_S = \Psi(x_S)$, with y_S having the form defined in (2.21). That is,

$$y_S = \{y_S^{(1)}, y_S^{(2)}, \ldots, y_S^{(K)}, y_S^{(K)}(\cdot|0)\},$$

where $y_S^{(K)}(\cdot|0)$ corresponds to the approximation image at the coarsest level K and $y_S^{(k)} = \{y_S^{(k)}(\cdot|p)\}$, $p = 1, \ldots, P$, to the detail images at level k.

Activity level measure

The degree to which each coefficient in y_S is salient (i.e, of interest for a task at hand) will be expressed by the so-called *activity level*. The activity function block associates to every band image $y_S^{(k)}(\cdot|p)$ an activity level $a_S^{(k)}(\cdot|p)$, which reflects the local activity of the image. Broadly speaking, the activity level $a_S^{(k)}(\boldsymbol{n}|p)$ of a sample $\boldsymbol{n}$ will be high if the average energy (or some other measure)
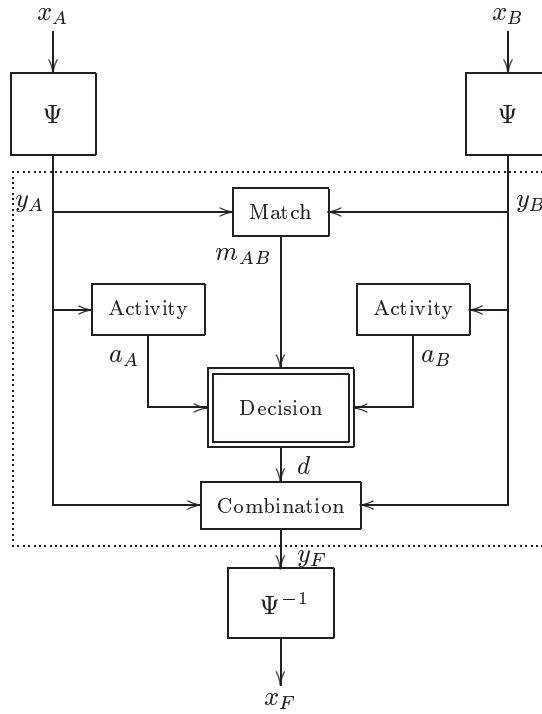
Figure 25: *Generic pixel-based MR fusion scheme with two input sources x_A and x_B, and one output fused image x_F.*

of $y_S^{(k)}(\cdot|p)$ is high in the vicinity of $\boldsymbol{n}$. Thus, one may consider that the activity level results from some sort of energy calculation over a local neighborhood of coefficients.

Match measure

This measure is supposed to quantify the degree of 'similarity' between the sources. More precisely, the match value $m_{AB}^{(k)}(\cdot)$ reflects the resemblance between the inputs $y_A^{(k)}(\cdot)$ and $y_B^{(k)}(\cdot)$.

By analyzing the match measure, one can determine where the sources differ and to which extent, and use this information to combine them in an appropriate way. For example, if the match measure at a given position is low (i.e, the sources are distinctly different at that position), the coefficient from the source decomposition with the highest activity level is taken as the composite coefficient. On the other hand, if the match value is high (i.e., the sources are similar at that position), the coefficients from the different sources are averaged to yield the composite coefficient.

Decision map

This block is the core of the combination algorithm. Its output governs the actual combination of the coefficients of the MR decompositions of the various sources.

For each level k, orientation band p and sample position $\boldsymbol{n}$, the decision process assigns a value $\delta = d^{(k)}(\boldsymbol{n}|p)$ which is then used for the computation of the composite $y_F^{(k)}(\boldsymbol{n}|p)$ (see combination map below). Most often, the decision is computed independently at each level, band and position. However, one may also take into account spatial, inter- and intra-scale dependencies between the samples, thus exploiting the idea that coefficients in the composite should not be computed independently. For instance, one may require neighboring coefficients in the same level and/or orientation to take the same decision.

Combination map

This module describes the actual combination of the transform coefficients of the sources. For each level k, orientation band p and sample position $\boldsymbol{n}$, the combination map yields the composite coefficient $y_F^{(k)}(\boldsymbol{n}|p)$. For simplicity, consider two sources and assume that every composite coefficient is

'assembled' from the source coefficients at the corresponding level, band and position. More precisely,

$$y_F^{(k)}(\cdot) = C^{(k)}\left(y_A^{(k)}(\cdot), y_B^{(k)}(\cdot), d^{(k)}(\cdot)\right), \tag{3.1}$$

where $C^{(k)} : \mathbb{R}^3 \mapsto \mathbb{R}$ is the combination map at level k.

MR synthesis (Ψ^{-1})

Finally, the fused image is obtained by applying the inverse transformation on the composite MR decomposition y_F:

$$x_F = \Psi^{-1}(y_F), \tag{3.2}$$

where Ψ^{-1} is the inverse MR transform. Strictly speaking, Ψ^{-1} needs only to be the left inverse of Ψ restricted to the source images domain, i.e., $\Psi^{-1}\Psi(x_S) = x_S$ for all $S \in \mathcal{S}$.

3.2 Choosing the building blocks

From the previous description, one can see that the parameters and functions comprised by the different blocks can be chosen in several ways. In the following, we discuss some of these alternatives. For convenience in the exposition, we have grouped the MR analysis and synthesis blocks, and have left for the end the discussion of the decision map.

3.2.1 MR analysis and synthesis

As we have seen in Section 2, the MR representation y_S comprises information at different scales. High levels contain coarse scale information while low levels contain finer details. Such a representation is suitable for image fusion, not only because it enables one to consider and fuse image features separately at different scales, but also because it produces large coefficients near edges, thus revealing salient information [60].

A large part of research on MR image fusion has focused on choosing an appropriate representation which facilitates the selection and combination of salient features. Basically, the issues to be addressed are the specific type of MR decomposition (pyramid, wavelet, linear, morphological, ...) and the number of decomposition levels.

In Section 2 we have presented a comprehensive discussion of various MR decompositions schemes, in particular pyramids and wavelets, as well as methods for their construction. Some aspects to take into account when choosing between different types of decompositions are:

- Invariance: do we need shift or rotation invariance?

- Redundancy: is it advantageous to have an overcomplete representation?

- Discrimination of structures: are we interested in a specific scale or orientation?

- Sensitivity to noise: how robust is the transform?

Suppose we choose linear wavelets. Even in this case we still have a lot of freedom in the choice of the specific filters which implement the wavelet transform. Some aspects we may consider are:

- Frequency characteristic of the basis function: low-pass or band-pass, maxima and minima location, etc.

- Regularity: smoothness is important in order to avoid discontinuities in the filtered image.

- Symmetry: typically, linear phase is desirable (to avoid phase distortion in the reconstruction), which implies the use of biorthogonal filters in order to obtain regular symmetrical FIR filters.

- Edge behavior: for example, antisymmetric filters give higher signal-to-noise ratios near edges.

- Filter length: this concerns a trade-off between processing speed and good spatial localization.

These considerations are in fact common to all applications that make use of MR representations. Focusing on image fusion and studying the existing literature, we draw the following conclusions:

- For image fusion purposes with no data compression objectives, the role of non-redundant representations is uncertain. In general, sampling causes a deterioration in the quality of the fused image by introducing heavier blocking effects than would have obtained by using decompositions without sampling.

- Shift and rotation-invariance properties are often required. For many applications, the fusion result should not depend on the location or orientation of the objects in the input sources. Shift and rotation dependency are especially undesirable considering misregistration problems or when used for image sequence fusion.

- In linear approaches, the specific filter used has little influence on the fusion result; shorter filters lead to slightly sharper fusion results.

Another parameter which will influence performance is the number of decomposition levels. To perform a consistent fusion of objects at arbitrary scales, the decomposition over a large number of scales may appear necessary. However, using more decomposition levels does not necessarily produce better results. Increasing the analysis depth may produce low-resolution bands where neighboring features overlap. This gives rise to discontinuities in the composite representation and thus, introduces distortions, such as blocking effects or 'ringing' artifacts, into the fused image. The required analysis depth is primarily related to the spatial extent of the relevant objects in the source images. In general, it is not possible to compute the optimal analysis depth, but as a rule of thumb, the larger the objects of interest are, the higher the number of decompositions levels should be.

3.2.2 Activity level measure

The meaning of 'saliency' (and thus the computation of the 'activity level') depends on the nature of the source images as well as on the particular fusion application. For example, when combining images having different foci, a desirable activity level measure would provide a quantitative value that increases when features are better in focus. In this case, a suitable measure is one that put emphasis on contrast differences. Since contrast information is partially captured in the decomposition by the magnitude of high-frequency components (details), a good choice is the absolute value of the detail coefficients or some other function that operates on their amplitude, e.g.,

$$a_S^{(k)}(\boldsymbol{n}|p) = \sum_{\Delta\boldsymbol{n}\in\mathcal{W}^{(k)}(p)} w^{(k)}(\Delta\boldsymbol{n}|p) \big| y_S^{(k)}(\boldsymbol{n}+\Delta\boldsymbol{n}|p) \big|^{\gamma}, \qquad \gamma \in \mathbb{R}_+ , \tag{3.3}$$

where $\mathcal{W}^{(k)}(p)$ is a finite window at level k and orientation p, and $w^{(k)}(\cdot|p)$ are the window's weights. Generally, based on the fact that the human visual system is primarily sensitive to local contrast changes (i.e., edges), most fusion algorithms compute the activity level as some sort of energy calculation. In the simplest case, the activity level is just the absolute value of the coefficient, that is,

$$a_S^{(k)}(\cdot) = \big| y_S^{(k)}(\cdot) \big| . \tag{3.4}$$

Alternatively, the contrast of the component with its neighbors, e.g.,

$$a_S^{(k)}(\boldsymbol{n}|p) = \frac{\big| y_S^{(k)}(\boldsymbol{n}|p) \big|}{\sum\limits_{\Delta\boldsymbol{n}\in\mathcal{W}^{(k)}(p)} w^{(k)}(\Delta\boldsymbol{n}|p) \big| y_S^{(k)}(\boldsymbol{n}+\Delta\boldsymbol{n}|p) \big|} ,$$

or some other linear or nonlinear criteria can provide that measure. For instance, to reduce the influence of impulsive noise, one may consider

$$a_S^{(k)}(\boldsymbol{n}|p) = \text{median}_{\Delta\boldsymbol{n}\in\mathcal{W}^{(k)}(p)} \left(\big| y_S^{(k)}(\boldsymbol{n}+\Delta\boldsymbol{n}|p) \big| \right) .$$

In practice, the window $\mathcal{W}^{(k)}(p)$ over which the function operates is small, typically including only the sample itself (*sample-based* operation), or a 3 by 3, or 5 by 5 window centered on the sample (*area-based* operation). However, other size and shape templates have also been used. Increasing the size of the neighborhood from the simple sample-based case, adds robustness to the fusion system as it provides a smooth activity level function. However, larger templates cause problems at lower resolution levels when their size exceeds the size of the most salient features.

3.2.3 Match measure

The match or similarity between the transform coefficients of the source images is usually expressed in terms of a local correlation measure. Alternatively, the relative amplitude of the coefficients or some other criteria can be used. In the following expression, the match value between $y_A^{(k)}(\cdot)$ and $y_B^{(k)}(\cdot)$ is defined as a normalized correlation averaged over a neighborhood of the samples:

$$m_{AB}^{(k)}(\boldsymbol{n}|p) = \frac{2\displaystyle\sum_{\Delta\boldsymbol{n}\in\mathcal{W}^{(k)}(p)} w^{(k)}(\Delta\boldsymbol{n}|p) y_A^{(k)}(\boldsymbol{n}+\Delta\boldsymbol{n}|p) y_B^{(k)}(\boldsymbol{n}+\Delta\boldsymbol{n}|p)}{\displaystyle\sum_{\Delta\boldsymbol{n}\in\mathcal{W}^{(k)}(p)} w^{(k)}(\Delta\boldsymbol{n}|p)\left(\left|y_A^{(k)}(\boldsymbol{n}+\Delta\boldsymbol{n}|p)\right|^2 + \left|y_B^{(k)}(\boldsymbol{n}+\Delta\boldsymbol{n}|p)\right|^2\right)} \,, \tag{3.5}$$

where $\mathcal{W}^{(k)}(p)$ is the window at level k and orientation p, and $w^{(k)}(\cdot|p)$ its corresponding weights.

3.2.4 Combination map

A simple choice for $C^{(k)}$ is a linear mapping, e.g.,

$$C^{(k)}(y_1, y_2, \delta) = w_A(\delta) y_1 + w_B(\delta) y_2 \,, \tag{3.6}$$

where the weights $w_A(\delta)$, $w_B(\delta)$ depend on the decision parameter δ.

Nonlinear mappings are another option. Some well-known nonlinear mapping techniques are multidimensional scaling [47], Sammon's mapping [81] and self-organizing maps [43]. These mappings techniques have been regularly used for visualization of high-dimensional data sets [58].

In this paper, we restrict ourselves to linear combination maps as in (3.6), yet with possibly more than two input sources. Thus, the composite coefficients $y_F^{(k)}(\cdot)$ are obtained by an *additive or weighted combination*:

$$y_F^{(k)}(\cdot) = \sum_{S\in\mathcal{S}} w_S\big(d^{(k)}(\cdot)\big) y_S^{(k)}(\cdot) \,. \tag{3.7}$$

For the particular case where only one of the coefficients $y_S^{(k)}(\cdot)$ has a weight distinct from zero, that is, only one of the sources contributes to the composite, we talk about *selective combination* or *combination by selection*.

3.2.5 Decision map

The construction of the decision map is a key point because its output $d^{(k)}$ governs the combination map $C^{(k)}$ and therefore, it is the decision map that actually determines the combination of the various MR decompositions y_S or, in another words, the construction of the composite y_F.

In our case, where we assume a weighted combination such as in (3.7), the decision map controls the values of the weights to be assigned to each of the source coefficients. Indeed, specifying the decision $\delta = d^{(k)}(\cdot)$ is in practice equivalent to specifying the weights $w_S(\delta)$. For this reason, very often the combination and decision maps are 'grouped' together by expressing the composite coefficients in terms of the parameters or functions the decision is based on. The problem of 'how to compute $d^{(k)}(\cdot)$' is translated to the problem of 'how to compute $w_S(d^{(k)}(\cdot))$'. A natural approach is to assign to each coefficient a weight that depends increasingly on the activity level. In general, the resulting weighted average (performed by the combination map) leads to a stabilization of the fusion result, but it introduces the problem of contrast reduction in case of opposite contrast in different source images. This can be avoided by using a selective rule where the most salient component, i.e., the one with largest activity level, is chosen for the composite. In this case, after the combination map we get

$$y_F^{(k)}(\cdot) = y_M^{(k)}(\cdot) \qquad \text{with } M = \arg\max_{S\in\mathcal{S}}\left(a_S^{(k)}(\cdot)\right). \tag{3.8}$$

In other words, the decision process 'decides' that the most salient coefficient (among the various $y_S^{(k)}(\cdot)$, $S\in\mathcal{S}$) is the best choice for the composite coefficient $y_F^{(k)}(\cdot)$, and 'tells' the combination process to select it, i.e,

$$w_S(d^{(k)}(\cdot)) = \begin{cases} 1 & \text{if } S = \arg\max_{S'\in\mathcal{S}}\left(a_{S'}^{(k)}(\cdot)\right) \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

This selective combination is also known in the literature as a *'choose max'* selection process or *maximum selection* rule. For the case of two input sources, (3.8) can be written as

$$
y_F^{(k)}(\cdot) = \begin{cases} y_A^{(k)}(\cdot) & \text{if } a_A^{(k)}(\cdot) > a_B^{(k)}(\cdot) \\ y_B^{(k)}(\cdot) & \text{otherwise.} \end{cases} \tag{3.10}
$$

The selective combination strategy works well under the assumption that at each image location, only one of the source images provides the most useful information. This assumption is sometimes not valid, and a weighted combination may appear a better option. Alternatively, a match measure can be used to decide how to combine the coefficients. For instance,

$$
y_F^{(k)}(\cdot) = \begin{cases} y_A^{(k)}(\cdot) & \text{if } m_{AB}^{(k)}(\cdot) \le T \text{ and } a_A^{(k)}(\cdot) > a_B^{(k)}(\cdot) \\ y_B^{(k)}(\cdot) & \text{if } m_{AB}^{(k)}(\cdot) \le T \text{ and } a_A^{(k)}(\cdot) \le a_B^{(k)}(\cdot) \\ \dfrac{y_A^{(k)}(\cdot) + y_B^{(k)}(\cdot)}{2} & \text{otherwise;} \end{cases}
$$

for some threshold T. Thus, at sample locations where the source images are distinctly different, the combination process selects the most salient component, while at sample locations where they are similar, the process averages the source components. In this manner, averaging reduces noise and provides stability where source images contain similar information, whereas selection retains salient information and reduce artifacts due to opposite contrast.

In the examples presented so far, the decision is taken for each coefficient, without reference to the others. This may degrade the fusion result since there is the possibility of feature cancellation when the inverse transform is applied to obtain the fused image. Having into account the spatial, inter- and intra-scale dependencies between the coefficients may provide a partial solution to this problem. Note that by construction, each coefficient of a MR decomposition has a set of 'family-related' components in other orientation bands and other levels: they represent the same (or nearby) spatial location in the original image. It seems reasonable then to consider all (or a set of) these coefficients when determining the composite MR representation. A simple instance is

$$
y_F^{(k)}(\boldsymbol{n}|p) = \begin{cases} y_A^{(k)}(\boldsymbol{n}|p) & \text{if } \sum_{p'=1}^{p'=P} a_A^{(k)}(\boldsymbol{n}|p') > \sum_{p'=1}^{p'=P} a_B^{(k)}(\boldsymbol{n}|p') \\ y_B^{(k)}(\boldsymbol{n}|p) & \text{otherwise,} \end{cases}
$$

where intra-scale dependencies are used. In this particular example, the decision is made globally for a group of samples: for all bands p, all samples in the same level k and position $\boldsymbol{n}$ are assigned the same decision.

Another possibility is to exploit spatial redundancy between neighboring samples. One may assume that spatially close samples are likely to belong to the same image feature and thus, they should be computed in the same way. An illustrative example is the consistency verification method proposed by Li *et al.* [51]. This method consists in applying a majority filter to a preliminary decision map $\tilde{d}^{(k)}(\cdot|p)$. It is then the filtered decision map which determines the combination of the images $y_S^{(k)}(\cdot|p)$. For example, if according to the preliminary decision map $\tilde{d}^{(k)}(\cdot|p)$, the composite $y_F^{(k)}(\cdot)$ should come from $y_A^{(k)}$, while the majority of the surrounding composite coefficients should come from $y_B^{(k)}$, the decision $\tilde{d}^{(k)}(\cdot)$ is changed so that the composite coefficient $y_F^{(k)}(\cdot)$ comes from $y_B^{(k)}$.

This kind of decision methods attempt to exploit the fact that significant image features tend to be stable with respect to variations in space, scale and orientation. Thus, when comparing the corresponding image features in multiple source images, considering the dependencies of the transform coefficients may provide a more robust fusion strategy.

3.2.6 Combination of approximation images vs. combination of detail images

Because of their different physical meaning, the approximation and detail images are usually treated by the combination algorithm through different procedures.

For the detail images $y_S^{(k)}$, a general observation is that relevant perceptual information relates to the 'edge' information that is present in each of the detail coefficients $y_S^{(k)}(\cdot)$. Detail coefficients having

large absolute values correspond to sharp intensity changes and hence to salient features in the image such as edges, lines and region boundaries.

The nature of the approximation coefficients, however, is different. The approximation image $y_S^{(K)}(\cdot|0)$ is a coarse representation of the original image x_S and may have inherited some of its properties such as the mean intensity or texture information. Thus, coefficients $y_S^{(K)}(\boldsymbol{n}|0)$ with high magnitudes do not necessarily imply salient features. In this case, an activity level $a_S^{(K)}(\cdot|0)$ based, for example, on entropy, variance or texture criteria, may be a better alternative than an activity level based on energy such as in (3.3).

In many approaches, the composite approximation coefficients of the highest decomposition level, representing the mean intensity, are taken to be a weighted average of the approximation of the sources:

$$y_F^{(K)}(\boldsymbol{n}|0) = \frac{\sum\limits_{S \in \mathcal{S}} y_S^{(K)}(\boldsymbol{n}|0)}{|S|}\,, \qquad (3.11)$$

where $|S|$ is the number of sources. The logic behind this combination relies on the assumptions that the sources x_S are contaminated by additive Gaussian noise and that, provided that K is high enough, the relevant features have already been captured by the details $y_S^{(k)}(\cdot|p)$. Thus, the approximation images $y_S^{(K)}(\cdot|0)$ of the various sources contain mostly noise and averaging them reduces the variance of the noise while ensuring that an appropriate mean intensity is maintained.

A popular way to construct the composite y_F is to use (3.11) for the approximation coefficients and the selective combination in (3.8) for the detail coefficients. For the simple case where $a_S^{(k)}(\cdot) = |y_S^{(k)}(\cdot)|$, and we have two input sources, we can express the combination algorithm as

$$y_F^{(K)}(\boldsymbol{n}|0) \;\; = \;\; \frac{y_A^{(K)}(\boldsymbol{n}|0) + y_B^{(K)}(\boldsymbol{n}|0)}{2} \qquad (3.12)$$

$$y_F^{(k)}(\boldsymbol{n}|p) \;\; = \;\; \begin{cases} y_A^{(k)}(\boldsymbol{n}|p) & \text{if } |y_A^{(k)}(\boldsymbol{n}|p)| > |y_B^{(k)}(\boldsymbol{n}|p)| \\ y_B^{(k)}(\boldsymbol{n}|p) & \text{otherwise.} \end{cases} \qquad p = 1,\ldots,P \qquad (3.13)$$

Note that other factors may be incorporated for the fusion rules. In particular, if some prior knowledge is available, all the fusion blocks can use such information to improve fusion performance. For instance, when combining the source coefficients, the weights assigned to them may depend not only on the activity level and match measure, but may also reflect some a-priori knowledge of a specific type, giving preference to certain levels k, spatial positions $\boldsymbol{n}$ or some input sources.

Finally, we want to remark that the decision on which techniques to use is very much driven by the application. At the same time, the characteristics of the resultant fused image depend strongly on the applied pre-processing and the chosen fusion techniques. The different options we have presented are neither exhaustive nor mutually exclusive and they should just be considered as practically important examples.

3.3 Overview of some existing fusion schemes

In the literature one finds several MR fusion approaches which fit into our general scheme. In this section, we review some of them. The reader may also get an impression of the evolution of MR-based schemes during the past fifteen years.

The first MR image fusion approach proposed in the literature is due to Burt [10]. His implementation used a Laplacian pyramid and a sample-based maximum selection rule with $a_S^{(k)}(\cdot) = |y_S^{(k)}(\cdot)|$. Thus, each composite coefficient $y_F^{(k)}(\cdot)$ is obtained by

$$y_F^{(k)}(\cdot) = \begin{cases} y_A^{(k)}(\cdot) & \text{if } \left|y_A^{(k)}(\cdot)\right| > \left|y_B^{(k)}(\cdot)\right| \\ y_B^{(k)}(\cdot) & \text{otherwise}. \end{cases}$$

Toet [95, 98] presented a similar algorithm but using the ratio-of-low-pass pyramid. His approach is motivated by the fact that the human visual system is based on contrast, and therefore, a fusion

technique which selects the highest local luminance contrast is likely to provide better details to a human observer. Another variation of this scheme is obtained by replacing the linear filters by morphological ones [61, 96].

Burt and Kolczynski [14] proposed to use a gradient pyramid (hence $P = 4$) together with a combination algorithm that is based on an activity level and a match measure. In particular, they define the activity level of $y_S^{(k)}(\cdot)$ as a local energy measure:

$$a_S^{(k)}(\boldsymbol{n}|p) = \sum_{\Delta \boldsymbol{n} \in \mathcal{W}^{(k)}(p)} \left| y_S^{(k)}(\boldsymbol{n} + \Delta \boldsymbol{n}|p) \right|^2 , \tag{3.14}$$

and the match between $y_A^{(k)}(\cdot)$ and $y_B^{(k)}(\cdot)$ as

$$m_{AB}^{(k)}(\boldsymbol{n}|p) = \frac{2 \sum_{\Delta \boldsymbol{n} \in \mathcal{W}^{(k)}(p)} y_A^{(k)}(\boldsymbol{n} + \Delta \boldsymbol{n}|p) y_B^{(k)}(\boldsymbol{n} + \Delta \boldsymbol{n}|p)}{a_A^{(k)}(\boldsymbol{n}|p) + a_B^{(k)}(\boldsymbol{n}|p)} , \tag{3.15}$$

with $\mathcal{W}^{(k)}(p)$ being either a 1 by 1, 3 by 3, or 5 by 5 window centered on the origin. The combination process is the weighted average

$$y_F^{(k)}(\cdot) = w_A\big(d^{(k)}(\cdot)\big) y_A^{(k)}(\cdot) + w_B\big(d^{(k)}(\cdot)\big) y_B^{(k)}(\cdot) ,$$

where the weights are determined by the decision process for each level k, band p and position $\boldsymbol{n}$ as $w_A\big(d^{(k)}(\cdot)\big) = 1 - w_B\big(d^{(k)}(\cdot)\big) = d^{(k)}(\cdot)$, with

$$d^{(k)}(\cdot) = \begin{cases} 1 & \text{if } m_{AB}^{(k)}(\cdot) \leq T \text{ and } a_A^{(k)}(\cdot) > a_B^{(k)}(\cdot) \\ 0 & \text{if } m_{AB}^{(k)}(\cdot) \leq T \text{ and } a_A^{(k)}(\cdot) \leq a_B^{(k)}(\cdot) \\ \frac{1}{2} + \frac{1}{2}\big(\frac{1 - m_{AB}^{(k)}(\cdot)}{1 - T}\big) & \text{if } m_{AB}^{(k)}(\cdot) > T \text{ and } a_A^{(k)}(\cdot) > a_B^{(k)}(\cdot) \\ \frac{1}{2} - \frac{1}{2}\big(\frac{1 - m_{AB}^{(k)}(\cdot)}{1 - T}\big) & \text{if } m_{AB}^{(k)}(\cdot) > T \text{ and } a_A^{(k)}(\cdot) \leq a_B^{(k)}(\cdot) \end{cases} \tag{3.16}$$

for some threshold T. Observe that in case of a poor match (no similarity between the inputs), the source coefficient having the largest activity level will yield the composite value; while otherwise, a weighted sum of the sources coefficients will be used. Because of this, the fusion is said to be performed by combination of selection and averaging. The authors claim that this approach provides a partial solution to the problem of combining components that have opposite contrast, since such components are combined by selection. In addition, the use of area-based (vs. sampled-based) operations and the gradient pyramid provide greater stability in noise, compared to the Laplacian pyramid-based fusion.

Ranchin and Wald [74] presented one of the first wavelet-based fusion systems. This approach is also used by Li *et al.* in [51]. Their implementation considers the maximum absolute value within a window as the activity level measure associated with the sample centered in the window:

$$a_S^{(k)}(\boldsymbol{n}|p) = \max_{\Delta \boldsymbol{n} \in \mathcal{W}^{(k)}(p)} \left(\left| y_S^{(k)}(\boldsymbol{n} + \Delta \boldsymbol{n}|p) \right| \right) .$$

For each position in the transform domain, a simple maximum selection rule is used to determine which of the inputs is likely to contain the most useful information. This results in a preliminary decision map which indicates, at each position, which source should be used in the combination map. This decision map is then subject to a consistency verification. In particular, they apply a majority filter in order to remove possible wrong selection decisions caused by impulsive noise. The authors claim that their scheme performs better than the Laplacian pyramid-based fusion due to the compactness, directional selectivity and orthogonality of the wavelet transform.

Wilson *et al.* [109] used a DWT fusion method and a perceptual-based weighting based on the frequency response of the human visual system. Indeed, their activity level measure is computed as a weighted sum of the Fourier transform coefficients of the wavelet decomposition, with the weights determined by the contrast sensitivity[7]. They define a perceptual distance between the sources as

$$D_{AB}^{(k)}(\cdot) = \left| \frac{a_A^{(k)}(\cdot) - a_B^{(k)}(\cdot)}{a_A^{(k)}(\cdot) + a_B^{(k)}(\cdot)} \right| ,$$

---

[7]The contrast sensitivity is defined as the reciprocal of the threshold contrast required for a given spatial frequency to be perceived.

and use it together with the activity level to determine the weights of the wavelet coefficients from each source. Observe that this perceptual distance is directly related to the matching measure: the smaller the perceptual distance, the higher the matching measure. The final weighting is given by $w_A\big(d^{(k)}(\cdot)\big) = 1 - w_B\big(d^{(k)}(\cdot)\big) = d^{(k)}(\cdot)$, with:

$$
d^{(k)}(\cdot) = \begin{cases} 1 & \text{if } D_{AB}^{(k)}(\cdot) > T \text{ and } a_A^{(k)}(\cdot) > a_B^{(k)}(\cdot) \\[2mm] 0 & \text{if } D_{AB}^{(k)}(\cdot) > T \text{ and } a_A^{(k)}(\cdot) \le a_B^{(k)}(\cdot) \\[2mm] 1 - \frac{1}{2}\frac{a_B^{(k)}(\cdot)}{a_A^{(k)}(\cdot)} & \text{if } D_{AB}^{(k)}(\cdot) \le T, \end{cases}
$$

for some threshold T. In the experimental results presented by the authors, the fused images obtained with their method are visually better than the ones obtained by fusion techniques based on the gradient pyramid or the ratio-of-low-pass pyramid.

Koren *et al.* [44] used a steerable wavelet transform for the MR decomposition. They advocate their choice because of the shift-invariance and no-aliasing properties this transform offers. For each frequency band, the activity level is a local oriented energy. Only the components corresponding to the frequency band whose activity level is the largest are included for reconstruction (maximum selection rule). Liu *et al.* [53] take a completely different point of view. They also used a steerable pyramid but rather than using it to fuse the source images, they fuse the various bands of this decomposition by means of a Laplacian pyramid.

In [79], Rockinger considered an approach based on a shift-invariant extension of the DWT. The detail coefficients are combined by a maximum selection rule, while the coarse approximation coefficients are merged by averaging. Due to the shift-invariance representation, the proposed method is particularly useful for image sequence fusion, where a composite image sequence has to be built from various input image sequences. The author shows that the shift-invariant fusion method outperforms other MR fusion methods with respect to temporal stability[8] and consistency[9].

Zhang and Blum [114] compared different MR-based image fusion approaches within a generic MR fusion methodology. As we mentioned before, our framework has been inspired by their work.

Pu and Ni [73] proposed a contrast-based image fusion method using the wavelet transform. They measure the activity level as the absolute value of what they call directive contrast:

$$
a_S^{(k)}(\boldsymbol{n}|p) = \left| \frac{y_S^{(k)}(\boldsymbol{n}|p)}{y_S^{(k)}(\boldsymbol{n}|0)} \right| \qquad p = 1, \dots, P\,,
$$

and use a maximum selection rule as the combination method of the wavelet coefficients. They also proposed an alternative approach where the combination process is performed on the directive contrast itself. The fused image can be then reconstructed by reversing the computation of the directive contrast to obtain a composite MR decomposition, and finally applying the inverse wavelet transform.

Li and Wang [52] examined the application of discrete multiwavelet transforms to multisensor image fusion. The composite coefficients are obtained through a sample-based maximum selection rule, and the fused image is obtained by taking the corresponding discrete multiwavelet reconstruction of the composite coefficients. The authors showed experimental results where their fusion scheme performs better than those based on comparable scalar wavelet transforms.

Another MR technique is proposed by Scheunders in [82] where the fusion consists of retaining the modulus maxima [57] of the wavelet coefficients from the different bands and combining them. Noise reduction can be applied during the fusion process by removing noise-related modulus maxima. In the experiments presented, the proposed method outperforms other wavelet-based fusion techniques.

Mukhopadhyay and Chanda [66] presented a fusion scheme using multiscale morphology. In particular, they employ two MR top-hat transforms [83, 87] for extracting bright and dark details from the sources. For each source x_S, they derive two MR structures $y_{S,b}$ and $y_{S,d}$, by applying the bright and dark top-hat transforms respectively. A sample-based maximum selection rule on all $y_{S,b}$ yields a 'bright composite' $y_{F,b}$. Likewise, the same selection rule on all $y_{S,d}$ produces a 'dark composite'

---

[8]A fused image sequence is temporal stable if the graylevel changes in the fused sequence is only caused by the graylevel changes in the input sequences.

[9]A fused image sequence is temporal consistent if the graylevel changes occurring in the input sequences is present in the fused sequence without any delay or contrast change.

$y_{F,d}$. In both cases, the activity measure of each sample is taken to be its amplitude. Finally, the two composite MR representations are combined by subtracting the 'dark' details $y_{F,d}^{(k)}$ from the 'bright' ones $y_{F,b}^{(k)}$ and summing up all the entries. The 'dark' and 'bright' approximations are averaged and added to the detail to obtain an output fused image x_F.

3.4 Experimental results

Fig. 2-Fig. 5, in Section 1, are examples of fused imaged obtained by choosing different alternatives in the fusion blocks. Here, we give a few more examples. Some of them correspond to well-known fusion schemes which have also been discussed in Section 3.3, while others correspond to 'new' fusion schemes. In this latter case, the purpose is not to outperform the already existing approaches but to give an idea of the flexibility and freedom our framework offers.

In all cases, we have used the sources shown in Fig. 1. They correspond to visual (Fig. 1(a)) and infrared (Fig. 1(b)) image modalities. The images are each 360×270 pixels and for displaying purposes the gray values of the pixels have been scaled between 0 and 255 (histogram stretching). Unless otherwise stated, three levels of decomposition (i.e., $K = 3$) have been used for the MR decomposition of the sources.

3.4.1 Existing schemes

Fig. 26 shows some examples of fused images obtained by some of the fusion algorithms which exist in the literature.

The first row of Fig. 26 corresponds to the special case where $K = 0$ and thus, no MR decomposition is done. Fig. 26(a) is the result of a pixel-by-pixel average of the sources, while Fig. 26(b) is a weighted average where the weights have been determined by a principal component analysis (PCA).

In Fig. 26(c) we have used Burt's method [10]: a Laplacian pyramid decomposition and the combination algorithm specified in (3.12)- (3.13). The same combination algorithm but using a ratio-of-low pass pyramid for the decomposition (Toet's method [95]) yields the result in Fig. 26(d).

Fig. 26(e) illustrates the fusion algorithm proposed by Burt and Kolczynski in [14]. The activity level, match measure and weights are computed as in (3.15)-(3.16), with a 3×3 window centered on the origin and a threshold $T = 0.85$.

Fig. 26(f) shows the fused image obtained by the fusion scheme proposed by Li *et al.* in [51]. We have used a Daubechies orthogonal wavelet of order 2 to implement the DWT, and a 3×3 window for the consistency check.

3.4.2 Other schemes

We mainly concentrate on the different choices of MR decompositions and use conventional fusion rules such as the ones described in Section 3.3.

In Fig. 27(a) a steerable pyramid with $P = 4$ is used for the decomposition of the sources. The combination algorithm is the same as the one proposed by Burt and Kolczynski [14] and defined in (3.15)-(3.16), with a 3×3 window centered on the origin and a threshold $T = 0.85$. Fig. 27(b) has also been obtained with the same combination algorithm but performing, in addition, consistency check, and using a translation invariant Haar wavelet for the MR representation of the sources.

For the remaining examples, the simple combination algorithm specified in (3.12)-(3.13) is used. That is, the detail coefficients are sample-based selected by a maximum selection rule while the approximation coefficients are averaged.

Fig. 27(c)-Fig. 27(d) are examples of fused images where a median pyramid is employed for the MR decomposition. In Fig. 27(c), however, the ratio of the median filtered approximations instead of the standard difference has been used.

In Fig. 27(e)-Fig. 27(f), the lifting scheme is used to perform the MR decomposition on a quincunx lattice. Fig. 27(e) shows the fused image obtained with the max-lifting scheme[10] [38], while Fig. 27(f) depicts the result obtained with a lifting scheme where the prediction and update operators are Neville filters [45] of order 2.

---

[10]The max-lifting scheme is based on morphological operators. A major characteristic of this scheme is that it preserves local maxima.
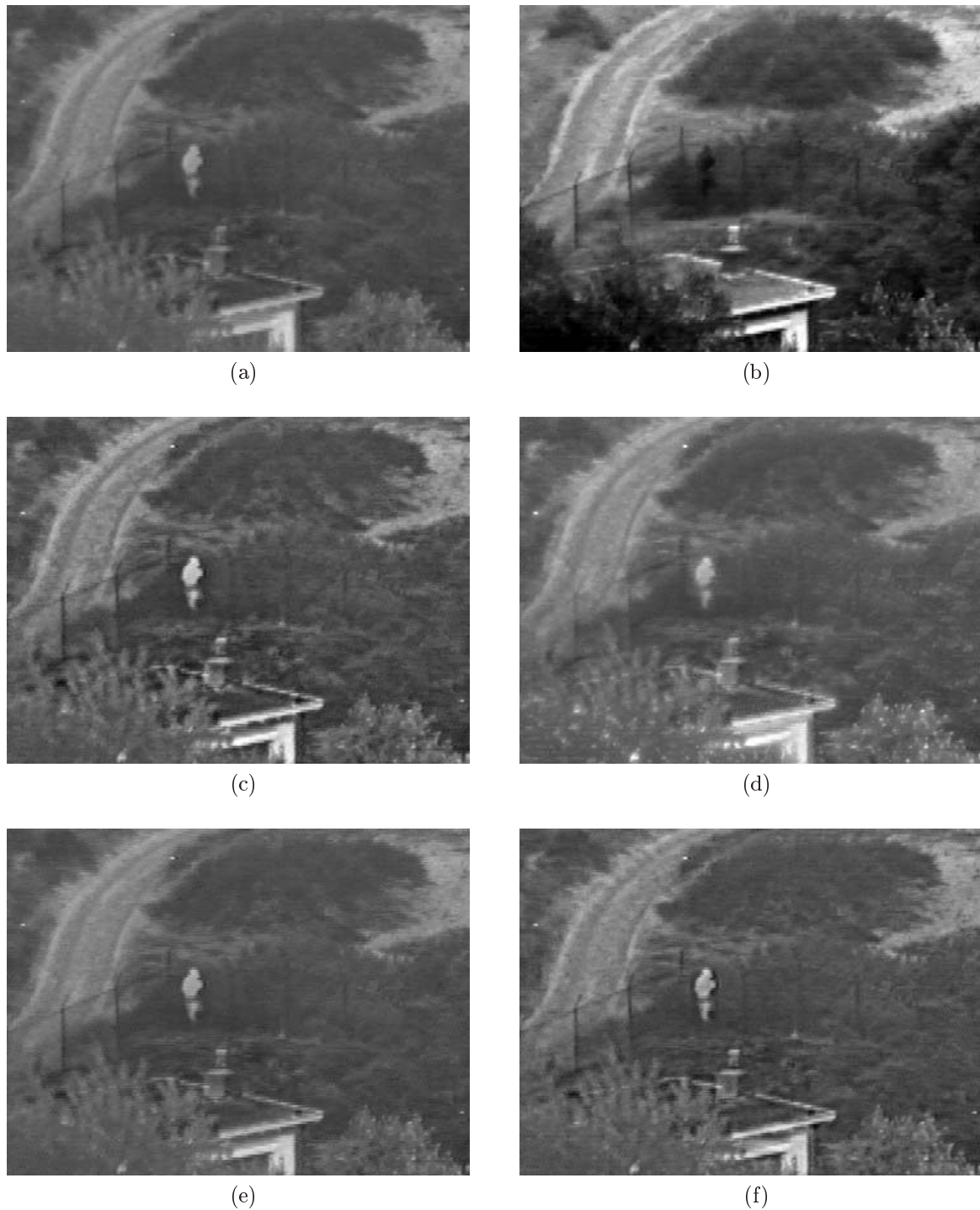
Figure 26: *Examples of fused images by some existing methods: (a) average; (b) weighted average by PCA; (c) Burt's method; (d) Toet's method; (e) Burt and Kolczynski's method; (f) Li et al.'s method.*

(a)   (b)

(c)   (d)

(e)   (f)
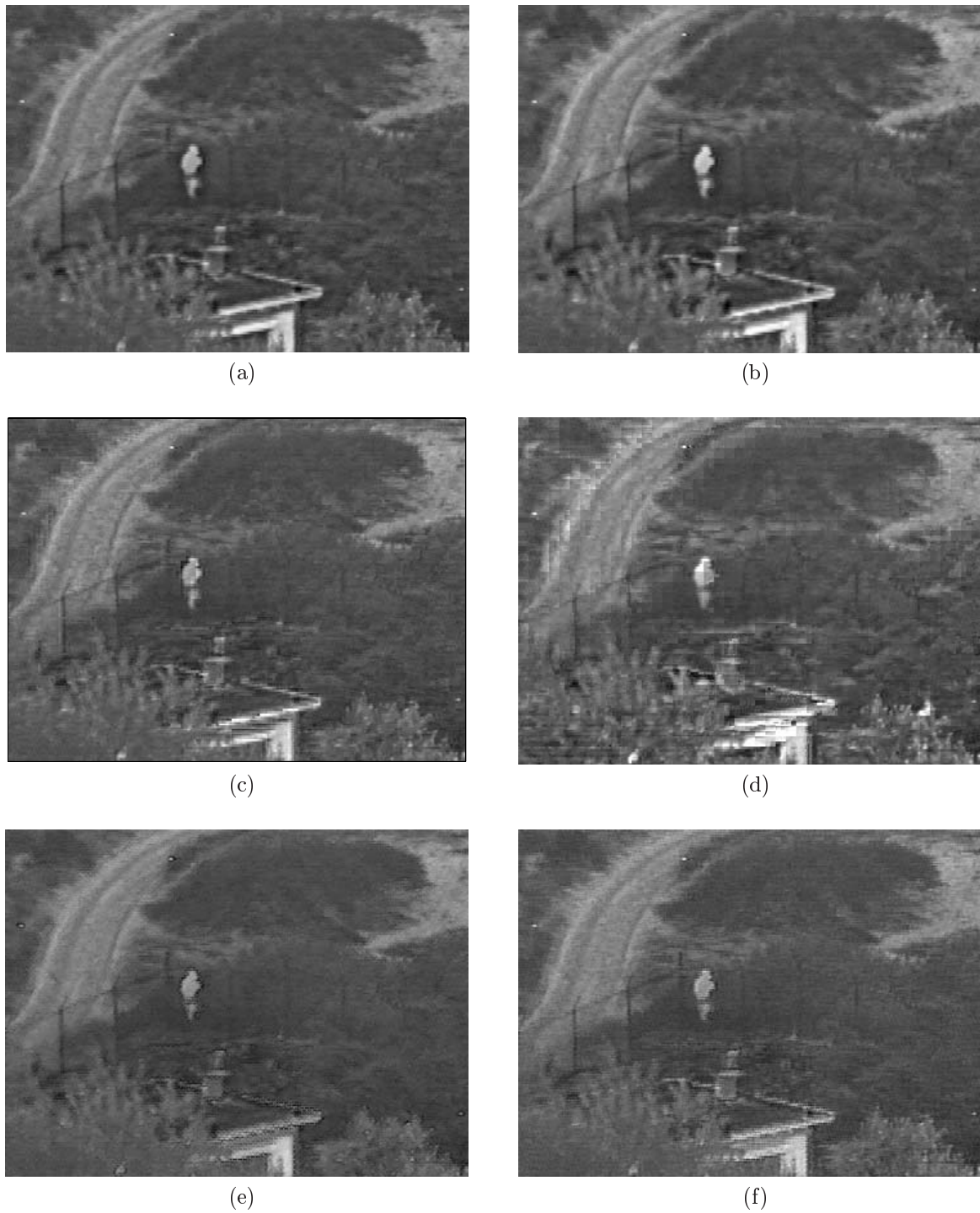
Figure 27: *Examples of fused images by 'new' methods: (a) steerable pyramid with Burt and Kolczyn-ski's combination algorithm; (b) undecimated DWT with Burt and Kolczynski's combination algorithm and consistency check. For (c)-(f), it has been used the combination algorithm in (3.12)-(3.13), and the following MR decompositions: (c) median pyramid; (d) ratio-of-median pyramid ; (e) max-lifting wavelet scheme (in quincunx lattice); (f) linear lifting with Neville filters wavelet scheme (in quincunx lattice).*

4 A region-based MR image fusion algorithm

In this section, we introduce a new region-based approach to MR fusion, which combines aspects of feature and pixel-level fusion. The basic idea is to make a segmentation based on all different source images and to use this segmentation to guide the combination process.

4.1 Motivation

The algorithms based on MR techniques that we have discussed in the previous sections are mainly pixel-based approaches where each individual coefficient of the MR decomposition (or possibly the coefficients in a small fixed window) is treated more or less independently. However, for most, if not all, image fusion applications, it seems more meaningful to combine objects rather than pixels. For example, assuming the input images depicted in Fig. 1, a fused image containing objects such as the house, the bushes, the hills, etc., as well as the person from the IR source and the fence from the visual source, would represent a rather accurate description of the underlying scene. Therefore, when fusing the images, it is reasonable to consider the pixels which constitute these objects as single entities instead of using the standard approach of combining the pixels without reference to the object they belong to. As an intermediate step from pixel-based toward object-based fusion schemes, one might consider region-based approaches. Such approaches have the additional advantage that the fusion process becomes more robust and avoids some of the well-known problems in pixel-level fusion such as blurring effects and high sensitivity to noise and misregistration.

4.2 The overall scheme

4.2.1 Introduction

Our region-based fusion scheme (see Fig. 28) extends the pixel-based fusion approach discussed in Section 3 (see Fig. 25). Indeed, it includes all the blocks described before. The major difference between the two schemes consists hereof that the region-based scheme also contains a segmentation module which uses all sources x_S as input and returns a single MR segmentation $\mathcal{R}$ (i.e., a partition of the underlying image domains into regions) as output. Thus, we use MR decompositions to represent the input images at different scales and, additionally, we introduce a multiresolution/multimodal[11] (MR/MM) segmentation to partition the image domain at these scales. The activity level and match measures are computed for every region in the decomposed input images. These measures may correspond to low-level as well as intermediate-level structures. Furthermore, the MR segmentation $\mathcal{R}$ allows us to impose data-dependent consistency constraints based on spatial as well as inter- and intra-scale dependencies. All this information, i.e, the measures and the consistency constraints, is integrated to yield a decision map d which governs the combination of the coefficients of the transformed sources. This combination results in a MR decomposition y_F, and by MR synthesis we obtain a fused image x_F.

The main functional blocks of this fusion strategy are depicted in Fig. 28. Since we already discussed most of them in Section 3, we concentrate on the segmentation module and its interaction with the other modules.

4.2.2 MR/MM Segmentation

This block uses the various source images as input and returns a single MR segmentation

$$\mathcal{R} = \{\mathcal{R}^{(1)}, \mathcal{R}^{(2)}, \ldots, \mathcal{R}^{(K)}\}$$

as output. Here $\mathcal{R}^{(k)}$ represents a segmentation at level k, i.e., a partitioning of the domain at level k.

Loosely speaking, $\mathcal{R}$ provides a MR representation of the various regions of the underlying scene. This representation will guide the other blocks of the fusion process; hence instead of working at pixel-level, they will take into consideration the regions inferred by the segmentation. From an intuitive point of view, we can regard these regions constituent parts of the objects in the overall scene.

---

[11]Here, the adjective multimodal means that there are more than one input image to be segmented.

$$x_A \qquad\qquad\qquad\qquad x_B$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

Segmentation

$\mathcal{R}$

$\Psi \qquad\qquad\qquad \Psi$

$y_A \qquad\qquad\qquad\qquad y_B$

Match

m_{AB}

Activity $\qquad$ Activity

$a_A \qquad\qquad\qquad a_B$
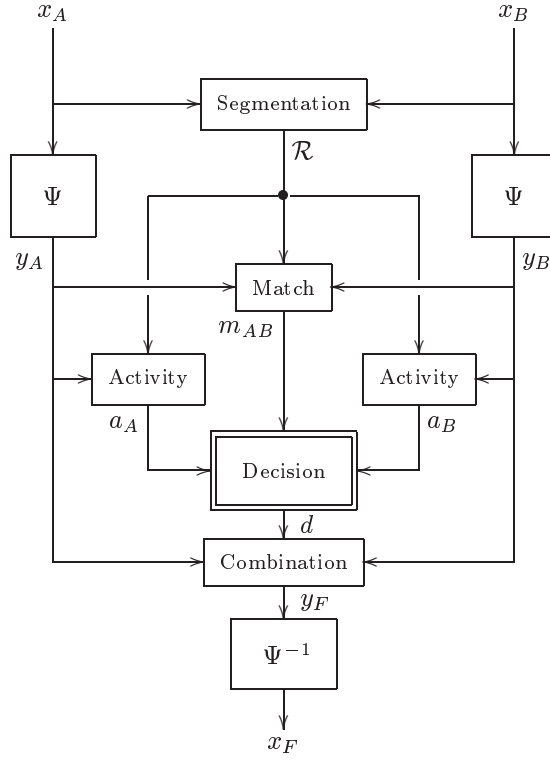
Decision

d

Combination

y_F

Ψ^{-1}

x_F

Figure 28: *Generic region-based MR fusion scheme with two input sources x_A and x_B, and one output fused image x_F.*

In our image fusion problem, the segmentation is merely a preparation step toward the actual fusion. In fact, we are not interested in the actual segmentation of the images, but rather in having a rough partition of the underlying scene. Therefore, the segmentation process does not need to be extremely accurate. For our purposes, we have developed a MR/MM segmentation algorithm based on the linked pyramid [13]. We describe our segmentation algorithm in Section 4.3. Obviously, other segmentation methods can be used. We require, however, that the sampling structure in $\mathcal{R}$ is the same as in y_S, so that each partition $\mathcal{R}^{(k)}$ corresponds to a partition of the detail $y_S^{(k)}(\cdot|p)$.

4.2.3 Combination algorithm

Since the building blocks of the combination algorithm in the region-based approach are essentially the same as in the pixel case, the combination algorithms discussed in Section 3 can be easily extended to the region-based approach. For example, we can define the activity level of each region $R \in \mathcal{R}^{(k)}$ in $y_S^{(k)}(\cdot|p)$ by

$$a_S^{(k)}(R|p) = \frac{1}{|R|} \sum_{\boldsymbol{n} \in R} a_S^{(k)}(\boldsymbol{n}|p)\,, \qquad\qquad (4.1)$$

where $|R|$ is the area of region R. Similarly, we can define the match measure of each region $R \in \mathcal{R}^{(k)}$ in the image bands $y_A^{(k)}(\cdot|p)$ and $y_B^{(k)}(\cdot|p)$ by

$$m_{AB}^{(k)}(R|p) = \frac{1}{|R|} \sum_{\boldsymbol{n} \in R} m_{AB}^{(k)}(\boldsymbol{n}|p)\,. \qquad\qquad (4.2)$$

Given these measures, the decision map can be constructed in several ways as discussed in Section 3.2.5, with the only difference that $a_S^{(k)}(R|p)$, $m_{AB}^{(k)}(R|p)$ are used instead of $a_S^{(k)}(\boldsymbol{n}|p)$, $m_{AB}^{(k)}(\boldsymbol{n}|p)$. For instance, a combination algorithm based on a maximum selection rule (see (3.10) for the pixel-based

case) would read:

$$y_F^{(k)}(\boldsymbol{n}|p) = \begin{cases} y_A^{(k)}(\boldsymbol{n}|p) & \text{if } a_A^{(k)}(R|p) > a_B^{(k)}(R|p) \\ y_B^{(k)}(\boldsymbol{n}|p) & \text{otherwise} \end{cases} \qquad \text{for all } \boldsymbol{n} \in R. \qquad (4.3)$$

As in the pixel-based scheme, once the decision map is constructed, the mapping performed by the combination process is determined for all coefficients, and the synthesis process yields the fused image x_F.

Note that for the particular case in which each region corresponds to a single point $\boldsymbol{n}$, the region-based approach reduces to a pixel-based approach. Thus, the region-based MR fusion scheme extends and generalizes the pixel-based approach and offers a general framework for MR-based image fusion which encompasses most of the existing MR fusion algorithms.

4.3 MR/MM segmentation based on pyramid linking

In this section, we present a MR/MM segmentation algorithm based on pyramid linking. We first review the basics of the conventional pyramid linking segmentation method. Then, we modify and extend this method for the segmentation of multimodal input images.

4.3.1 The linked pyramid

The linked pyramid structure was first described by Burt *et al.* [13] (subsequent related work can be found in [6, 17, 40, 68, 105]). It consists of a number of levels with the bottom level containing the full-resolution image and each successive higher level being a filtered/subsampled version derived from the level below it (see Section 2.2). The various levels of the pyramid are 'linked' by means of so-called child-parent relations (see Fig. 29) between their samples (pixels); such child-parent links are established during an iterative processing procedure to be described below.

A conventional linked pyramid is constructed as follows. First, an approximation pyramid is produced by low-pass filtering and sampling. Then, child-parent relations are established by linking each pixel in a level (called *child*) to one of the pixels in the next higher level (called *parent*) which is closest in gray value or in another pixel attribute. The attribute values of the parents are then updated using the values of their children. The process of linking and updating is repeated until convergence (which always occurs [13]). Finally (or during the linking process), some pixels are labeled as *roots*. In the simplest case, only the pixels in the top level of the pyramid are roots. A root and the pixels which (directly or via other links) are connected with the root induce a tree in the pyramid. The leaves of each tree correspond to pixels in the full-resolution image which define a *segment* or *region*. Thus, the linked pyramid provides a framework for an iterative process of image segmentation. For example, in Fig. 29, pixel T is a root which represents in the bottom level a segment composed of pixels a, b, c and d.
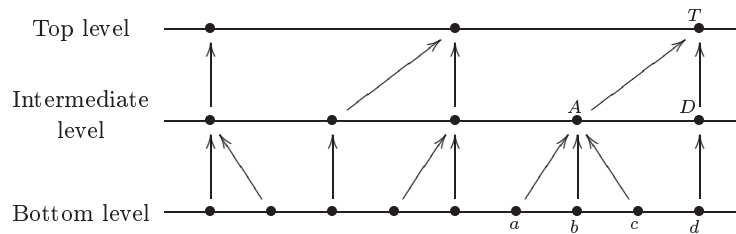

Figure 29: *A diagram illustrating linking relationships. E.g., pixel A is the parent of children a, b and c, and it is also the child of pixel T.*

There are a lot of variations on the scheme: in the way the initial pyramid is built, in the manner pixels are linked to each other, in determining when pixels should be declared as roots, in the size of the neighborhood in which children can look for a parent to link to, in the attribute that is being used

(e.g., gray value, edge, local texture), etc. As a result, a general pyramid linking method is hard to define, and most research has focused on specific problems or aspects.

Here, we want to address two major problems, namely the enforcement of connectivity in the segmented regions and the root labeling. The first problem arises from the fact that standard algorithms do not guarantee connectivity. Pixels which are adjacent in some higher level do not necessarily represent adjacent regions in the lower levels. This can cause the creation of disconnected regions at the bottom level. To avoid such anomalies, one can use the *connectivity preservation criteria* proposed by Nacken in [68]. The second problem concerns root characterization. In Burt's original approach, only the pixels at the top level are defined as roots, and therefore, the number of segments (which equals the number of roots) is fixed. Posterior approaches avoid such a prior choice and define the roots as those pixels which are not 'strongly' enough linked to a parent [40, 105]. Now, the problem is reduced to a definition of link strength and the determination of a root labeling threshold.

4.3.2 MR segmentation algorithm using linking

Our basic algorithm follows the classical "50% overlapping 4×4" structure [13]. This means that each parent is derived from the pixels in the 4×4 neighborhood immediately below it, and this neighborhood overlaps 50% of that of its 4 neighbors. Thus, each pixel has 16 candidate children and each child up to 4 candidate parents. The bottom of the pyramid corresponds to level zero and, for simplicity, is assumed to be of size $N \times N$ with N a power of 2. The maximum height of the pyramidal structure is considered to be $K_M = log_2 N - 1$.

At each level k, the pixels are indexed by the vector $\boldsymbol{n} = (n, m)$, where $n, m = 0, \ldots, \frac{N}{2^k} - 1$. We denote by $\mathcal{C}(\boldsymbol{n})$ the set of candidate children of pixel $\boldsymbol{n}$ at level $k > 0$; that is,

$$\mathcal{C}(\boldsymbol{n}) = \left\{ (n', m') \,|\, n' \in \{2n - 1, 2n, 2n + 1, 2n + 2\} , \, m' \in \{2m - 1, 2m, 2m + 1, 2m + 2\} \right\}.$$

Similarly, we denote by $\mathcal{P}(\boldsymbol{n})$ the set of candidate parents of pixel $\boldsymbol{n}$ at level $k < K_M$:

$$\mathcal{P}(\boldsymbol{n}) = \left\{ (n', m') \,|\, n' \in \{\lfloor \tfrac{1}{2}(n - 1) \rfloor, \lfloor \tfrac{1}{2}n \rfloor, \lfloor \tfrac{1}{2}(n + 1) \rfloor\} , \, m' \in \{\lfloor \tfrac{1}{2}(m - 1) \rfloor, \lfloor \tfrac{1}{2}m \rfloor, \lfloor \tfrac{1}{2}(m + 1) \rfloor\} \right\},$$

where $\lfloor \cdot \rfloor$ denotes the integer part of the enclosed value. The set of pixels to which the pixel $\boldsymbol{n}$ is connected at the bottom level is called *receptive field*.

To each pixel we associate one or more variables representing the attributes on which the segmentation will be based. In this study, we assign to each pixel $\boldsymbol{n}$ at level k its grayscale value $x^{(k)}(\boldsymbol{n})$, and the area $A^{(k)}(\boldsymbol{n})$ of its receptive field.

Consider an input image $x = x^{(0)}$. Our pyramid segmentation algorithm consists of three steps.

1. Initialization
We associate to each pixel $\boldsymbol{n}$ in level zero the gray value $x^{(0)}(\boldsymbol{n})$ of the original image, and to each pixel $\boldsymbol{n}$ in level $k > 0$ a gray value $x^{(k)}$ computed from the average of the gray values of its candidate children:

$$x^{(k)}(\boldsymbol{n}) = \frac{1}{16} \sum_{\boldsymbol{n}' \in \mathcal{C}(\boldsymbol{n})} x^{(k-1)}(\boldsymbol{n}').$$

2. Linking

(a) Pixel linking and root labeling.

For each child, a suitable parent is sought among the candidate parents: it is linked to its most 'similar' parent or it becomes a root (see below). Here, 'similarity' is based on geometrical and grayscale proximity. A distance measure between the child and each of its four candidate parents is computed. The link is established with the parent that minimizes that distance. It might occur that more than one candidate parent minimizes the distance measure. In this case we arbitrary choose among the minimal candidates. A simple choice for the distance measure is the difference in grayscale. Examples of other distances can be found in [68, 105].

In our approach, we perform the root labeling within the linking step. That is, when linking to a parent, if the distance measure is above some threshold, the link is not established and the pixel is labeled as a root (thus, it is not considered to be a child any more). We refer to [40, 105] for other alternatives.

(b) Updating of area $A^{(k)}$ and gray values $x^{(k)}$.

The attributes of each parent are recomputed using only the children that are linked to him:

$$A^{(k+1)}(\boldsymbol{n}) = \sum_{\boldsymbol{n}' \in \mathcal{C}(\boldsymbol{n})} A^{(k)}(\boldsymbol{n}')$$

$$x^{(k+1)}(\boldsymbol{n}) = \frac{\sum\limits_{\boldsymbol{n}' \in \mathcal{C}(\boldsymbol{n})} x^{(k)}(\boldsymbol{n}') A^{(k)}(\boldsymbol{n}')}{A^{(k+1)}(\boldsymbol{n})},$$

where $A^{(0)}(\boldsymbol{n}) = 1$ for all $\boldsymbol{n}$ in level zero.

(c) Iteration of (a) and (b) until convergence.

3. Segmentation

The actual segmentation is obtained by using the tree structure of the created links. In each level k, all pixels that are connected to a common root are classified as a single region segment R. In this way, at each level k, we obtain a segmented image $\mathcal{R}^{(k)}$ which contains the different regions R at this level.

4.3.3 MR/MM segmentation algorithm using linking

So far we have discussed how to obtain a MR segmentation from a single input. We now turn to consider a simultaneous MR and MM segmentation in which from various input images a unique segmented output is derived.

The segmentation method presented in the last subsection can be easily extended for the case where we have several input images x_S, $S \in \mathcal{S}$. In this case, the initialization step is performed as before for each image and, in the linking step, in order to measure the similarity between a child $\boldsymbol{n}$ and each candidate parent $\boldsymbol{n}' \in \mathcal{P}(\boldsymbol{n})$, we use the distance measure given by the expression

$$\left(\sum_{S \in \mathcal{S}} \left(x_S^{(k)}(\boldsymbol{n}) - x_S^{(k+1)}(\boldsymbol{n}') \right)^2 \right)^{1/2}. \tag{4.4}$$

As in the scalar case, the candidate $\boldsymbol{n}'$ which minimizes this distance is selected to become the parent unless the distance is above some threshold, in which case $\boldsymbol{n}$ is labeled as a root. Using the new links, the gray values are updated for each $S \in \mathcal{S}$, and the process of linking and updating is iterated until convergence. After the linking step, although the gray values of the various inputs $S \in \mathcal{S}$ will in general differ, the linking structure (child-parent relations) is the same. Thus, we have a single linked pyramid structure and we can apply the same segmentation step as before.

We summarize the basic steps of our MR/MM segmentation in the following algorithm.

Algorithm

1. For each input $S \in \mathcal{S}$

   – Construct an approximation pyramid $\{x_S^{(k)}\}$.

2. For each level $k < K_M$

   – While not convergent,
      * For each child $\boldsymbol{n}$ at level k, look for the parent $\boldsymbol{n}' \in \mathcal{P}(\boldsymbol{n})$ which minimizes (4.4). If this distance is above some threshold, $\boldsymbol{n}$ is set as a root, otherwise it is linked to $\boldsymbol{n}'$.
      * For each parent $\boldsymbol{n}$ at level $k + 1$, update $A^{(k+1)}(\boldsymbol{n})$ and $x_S^{(k+1)}(\boldsymbol{n})$ for all $S \in \mathcal{S}$.

3. For each level k

   – All pixels $\boldsymbol{n}$ at level k connected to a common root (or being themselves a root) are classified as a single region segment R (at level k).

The segmentation is based on the approximation pyramids (computed from the grayscale values of the pixels) of the different input sources x_S, which are all treated equally. Obviously this is a very naive approach since different sources may present different amplitude ranges and may not be equally reliable. Thus, prior to segmentation, one might pre-process the input images so that they are comparable in their attributes. As an alternative one can choose to modify the distance measure in (4.4) and use, for instance,

$$\left(\sum_{S \in \mathcal{S}} \mu_S \left(x_S^{(k)}(\boldsymbol{n}) - x_S^{(k+1)}(\boldsymbol{n}') \right)^2 \right)^{1/2} ,$$

where μ_s is a normalization factor which may depend on several factors such as the dynamic range, noise estimation, entropy, etc.

Additionally, the segmentation algorithm can be improved by the use of connectivity preservation criteria, adaptive windows and probabilistic linking [68, 105].

Note that, by construction, the MR segmentation $\mathcal{R}$ obtained with our algorithm has a pyramidal structure where the bottom level is at full resolution (same size as x_S) and each successive coarser level is $1/4$ of its predecessor. However, this might not be true for the MR decompositions y_S obtained with the MR analysis block. Note also that the levels from the above MR segmentation $\mathcal{R}$ range from zero to K_M, whereas the levels from the MR decompositions y_S go from one to K. In practice, K is smaller than K_M, so we assume henceforth that $K \leq K_M$. In addition, we also assume the same lattice structure. Since we require all decompositions to have the same sampling structure, the MR/MM segmentation module should associate to each level k and band p in y_S, the segmentation level k' such that the domain of $y^{(k)}(\cdot|p)$, i.e. $I_y^{(k)}(p)$, has the same dimensions as $\mathcal{R}^{(k')}$. For instance, if y_S corresponds to a Laplacian decomposition, then $k' = k - 1$, for $k = 1, \ldots, K$; while if y_S corresponds to a DWT, then $k' = k$ for $k = 1, \ldots, K$ and all $p = 1, \ldots, 3$.

We assume that these associations are performed inside the MR/MM segmentation module so that we get the output $\mathcal{R} = \{\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(K)}\}$ which has the same sampling structure as $y_S = \{y_S^{(1)}, \ldots, y_S^{(K)}\}$.

4.4 Case studies

In this section, we present some experimental results obtained with one of the simplest implementations of the region-based fusion approach.

We consider two input sources x_A and x_B. For their MR decomposition, we use a Laplacian pyramid (thus, we only have a single orientation band, i.e, $P = 1$). We employ the MR/MM segmentation algorithm discussed in Section 4.3. In the combination algorithm, we do not use a matching measure and define the activity level of each region $R \in \mathcal{R}^{(k)}$ as in (4.1), with $a_S^{(k)}(\boldsymbol{n}|p) = |y_S^{(k)}(\boldsymbol{n}|p)|$. The combination process is performed as in (3.6), with $w_A(\delta) = \delta$ and $w_B(\delta) = 1 - \delta$. In the decision process, each component of d is obtained by the following simple decision rules:

- For $p = 0$,

$$\delta = d^{(K)}(\boldsymbol{n}|0) = \frac{1}{2} , \text{ for all } \boldsymbol{n} .$$

- For $p = 1$,
  - for each level k and for each region $R \in \mathcal{R}^{(k)}$:

$$\delta = d^{(k)}(\boldsymbol{n}|1) = \begin{cases} 1 & \text{if } a_A^{(k)}(R|1) > a_B^{(k)}(R|1) \\ 0 & \text{otherwise} \end{cases} \qquad \text{for all } \boldsymbol{n} \in R.$$

Note that according to this algorithm, the composite approximation image $y_F^{(K)}(\cdot|0)$ is the pixel-wise average of the approximation images $y_S^{(K)}(\cdot|0)$ and therefore, the region information $\mathcal{R}^{(K)}$ is neglected. The composite detail images $y_F^{(k)}$, however, are constructed by a selective combination as in (4.3).

We have tested our algorithm on several pairs of images. Three examples are given here to illustrate the fusion process described above. In all cases, we have chosen $K = 3$ and, when displaying the images, the gray values of the pixels have been scaled between 0 and 255 (histogram stretching). The input

sources x_A and x_B are displayed, respectively, on the left and right top of the corresponding figure. For the decision maps, pixels with $\delta = 0, 1$ are displayed in black and white, respectively. Thus, according to our algorithm, coefficients corresponding to 'white zones' are selected from $y_A^{(k)}$, while coefficients corresponding to 'black zones' are selected from $y_B^{(k)}$.

Fig. 30 shows the fusion of a visible and an IR wavelength images. The first level of the resulting segmentation and decision map are shown in the middle row. The corresponding second levels are displayed on the bottom left. It is interesting to note that, according to $d^{(2)}$, although most of the background is selected from the visual image $y_A^{(2)}$, the region corresponding to the person is selected from the IR image $y_B^{(2)}$. The fused image is depicted at the bottom right of Fig. 30.

Fig. 31 shows the fusion of images with different focus points. The segmentation and decision for the first level are displayed in the second row of Fig. 31. Note that since the digit '8' is connected to a particular region located within the left clock, the binary decision map $d^{(1)}$ points out, wrongly, to take the '8' from $y_A^{(1)}$ instead from $y_B^{(1)}$. The same happens in level $k = 2$ (not displayed here). For this particular example, we also show the corresponding fused output we would have obtained using a pixel-based MR fusion algorithm with the same fusion rules as in the region-based algorithm. We also illustrate how we can improve the region-based fused image by filtering the decision map. Here, we have filtered both decision maps $d^{(1)}$, $d^{(2)}$ with a morphological alternating filter: an opening followed by a closing [83, 87]. The filtered $d^{(1)}$ is shown at the bottom left of Fig. 31. One can see that small white and black regions have been removed and that the boundaries of the non-removed regions have been smoothed. The fused image obtained with the filtered decision maps is shown at the bottom right of Fig. 31.

Fig. 32 shows the fusion of MRI and CT images. In this last example, we illustrate the combination of the approximation coefficients using an activity level based on a local variance (see below). More precisely, we perform the selective combination in (4.3) for both detail and approximation coefficients but using different activity level measures. For the details, $a_S^{(k)}(R|1)$ is defined as before, while for the approximation we consider

$$a_S^{(K)}(R|0) = \frac{1}{|R|} \sum_{\boldsymbol{n} \in R} \left(y_S^{(K)}(\boldsymbol{n}|0) - \bar{y}_S^{(K)}(R|0) \right)^2$$

where $\bar{y}_S^{(K)}(R|0) = \dfrac{1}{|R|} \displaystyle\sum_{\boldsymbol{n} \in R} y_S^{(K)}(\boldsymbol{n}|0)$.

We also show (Fig. 32, bottom left) the corresponding fused output we would have obtained using (i) the decision rules used in the previous experiments (i.e., the selective combination only for the details and pixel-wise average for the approximation); (ii) a pixel-based MR fusion algorithm with the same fusion rules as in (i).

4.5   Discussion

From the experiments presented, we can see that, despite the crudeness of the current implementation, the visual performance is surprisingly good. This suggests that the region-based approach proposed here can at least be competitive with (but more likely outperform) other MR fusion techniques.

Further investigations are necessary for the fine-tuning of parameters as well as for the proper selection of the different ingredients of the scheme. Toward this end, performance assessment criteria will be developed to evaluate and demonstrate the capacities of the new fusion technique, as well as to compare its performance with others MR fusion schemes. Test will be carried out based both on objective and subjective criteria.

5   Performance assessment

Performance measures are essential to determine the possible benefits of fusion as well as to compare results obtained with different algorithms. Furthermore, they are necessary in order to obtain an optimal setting of parameters for a specific fusion algorithm.

In most cases, image fusion is only a preparatory step to some specific task such as human monitoring, and thus the performance of the fusion algorithm has to be measured in terms of improvement of
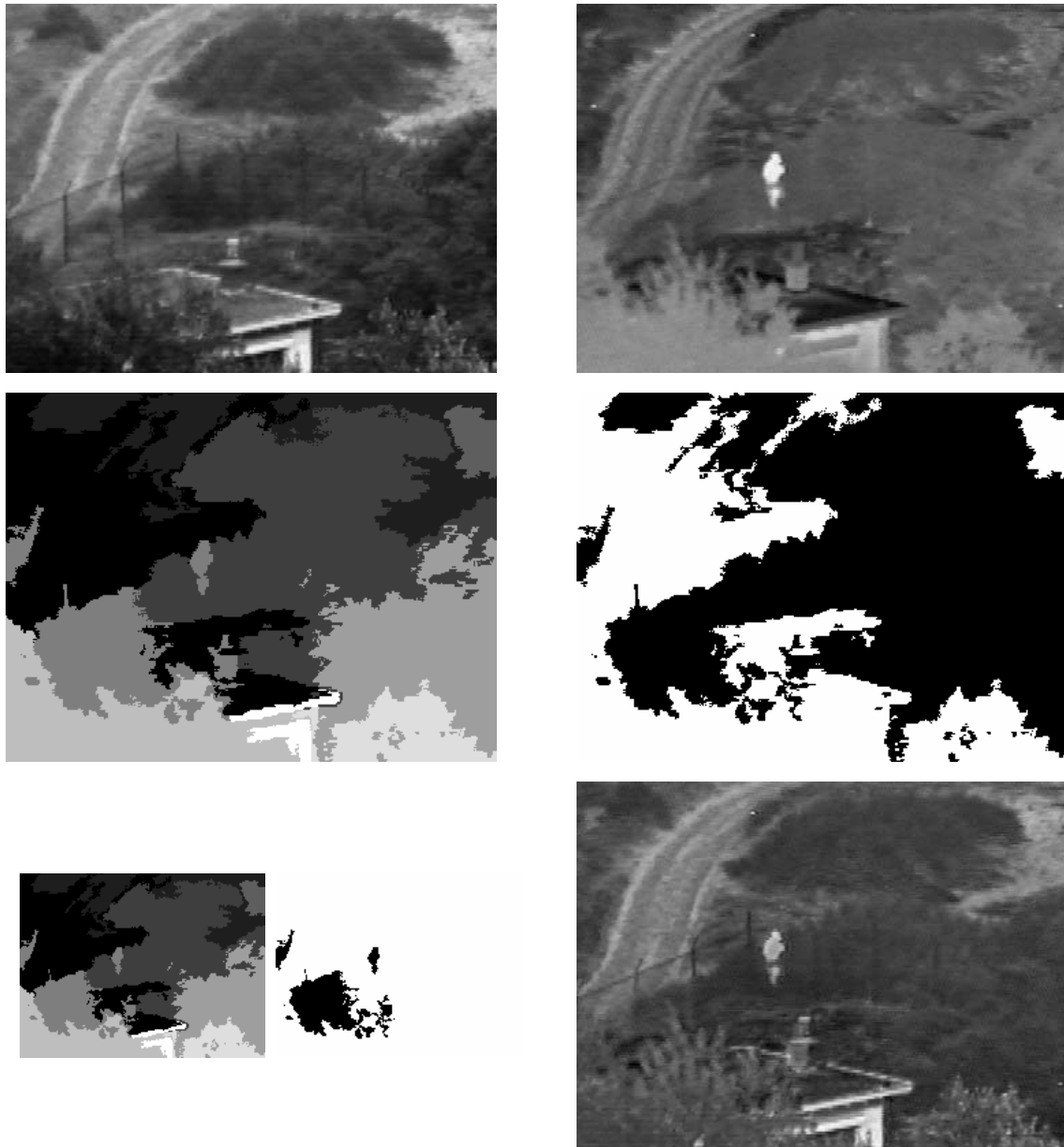
Figure 30: *Example 1. Top: visual (left) and IR (right) test images; middle: 1st level of segmentation (left) and decision map (right); bottom: 2nd level of segmentation (left) and decision map (central), and fused image (right).*
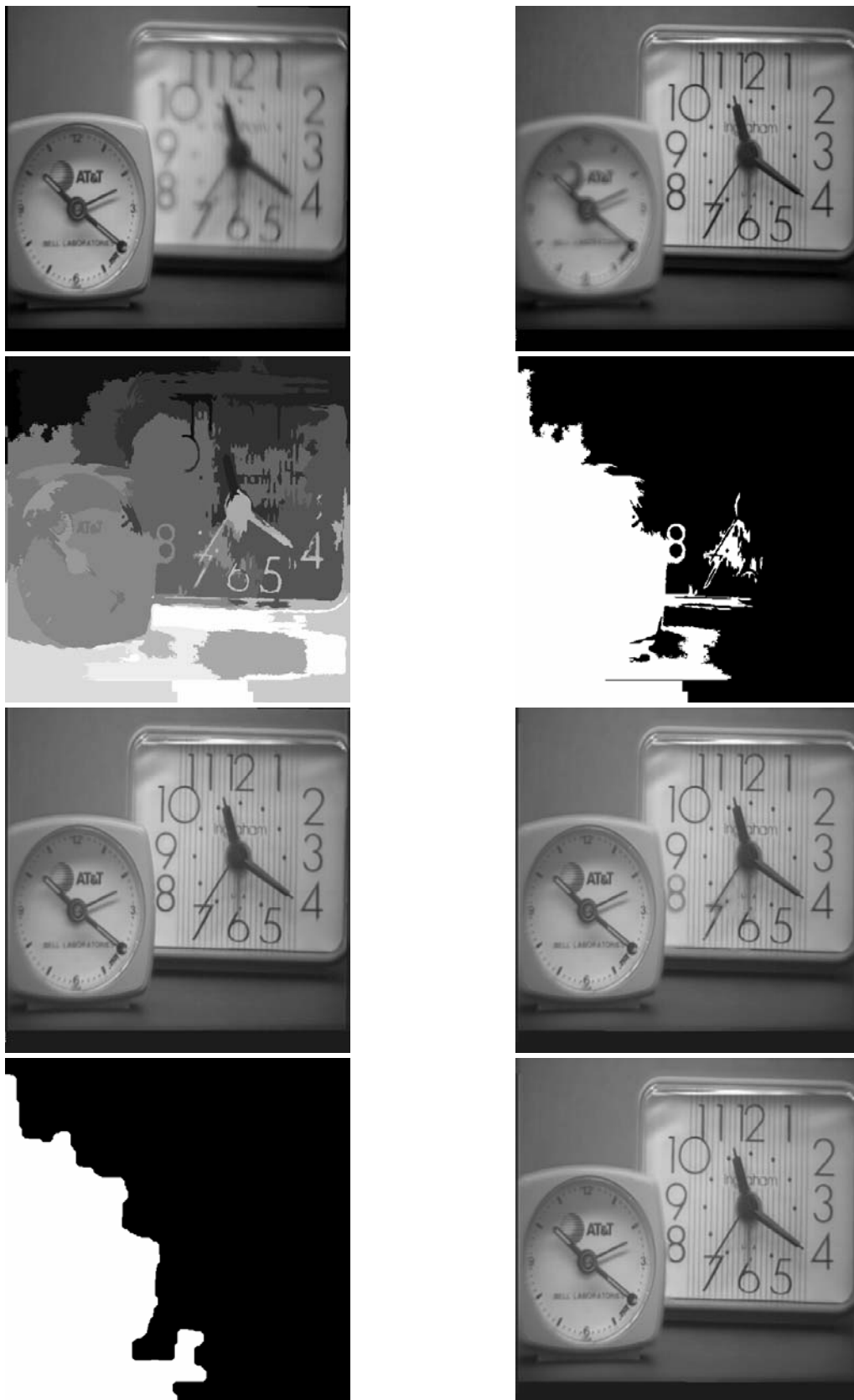
Figure 31: *Example 2. Top: multi-focus test images; second row: 1st level of segmentation (left) and decision map (right); third row: fused images with pixel-based (left) and region-based (right) approach; bottom: filtered decision map (left) and corresponding fused image.*
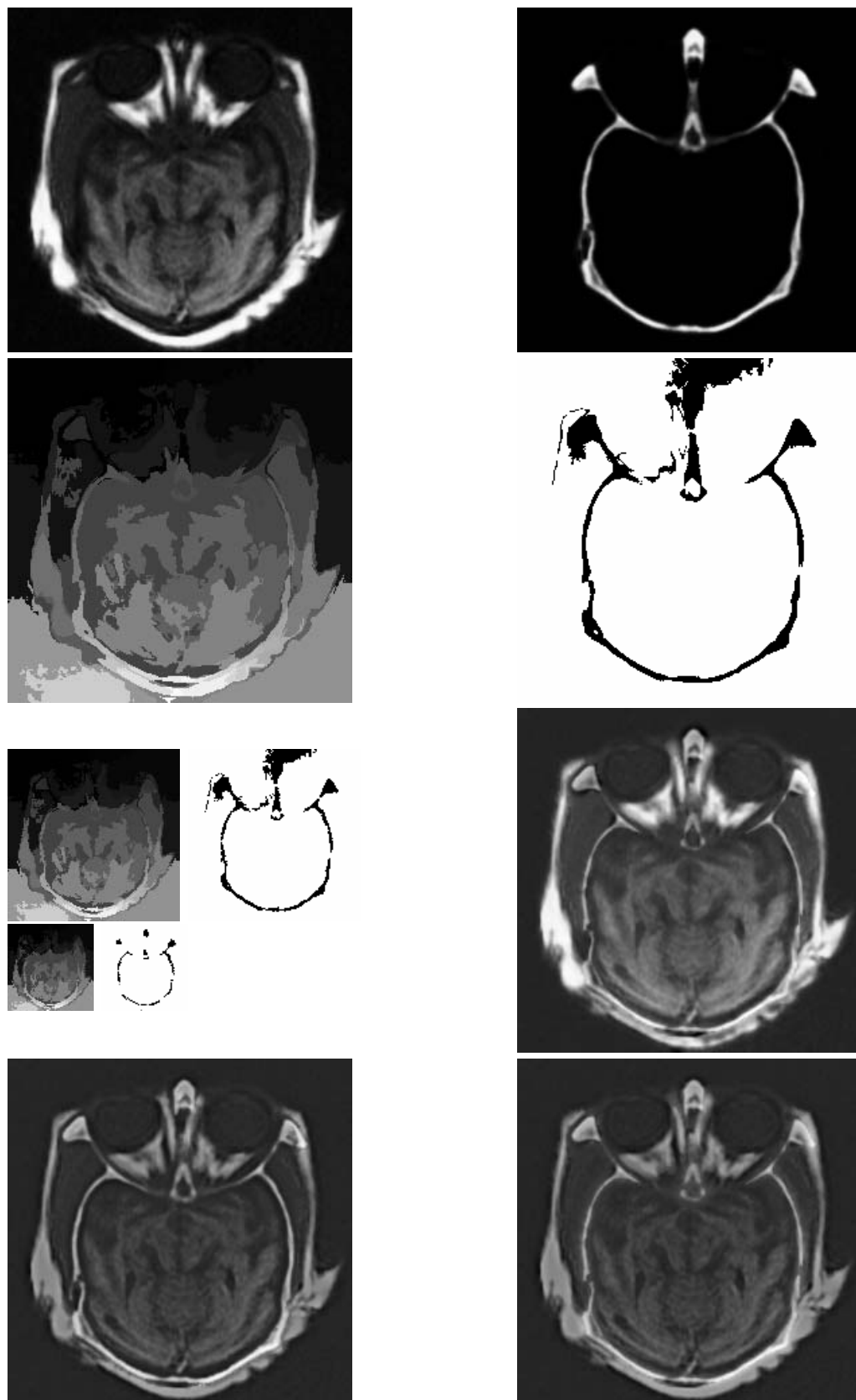
Figure 32: *Example 3. Top: MRI (left) and CT (right) test images; second row: 1st level of segmentation (left) and decision map (right); third row: 2nd and 3rd level of segmentation (left) and decision map (central), and fused image (right); bottom: pixel-based (left) and region-based (right) fused images with pixel-wise average combination for K = 3.*

the subsequent tasks. For example, in classification tasks, a common evaluation measure is the ratio of the correct number and the total number of classifications. This requires that the 'true' correct classifications are known. In experimental setups, however, the availability of a ground-truth is not guaranteed.

In this paper, we do not consider task-specific evaluation methods. Instead, we focus on general performance measures which can be applied independently of the task. In this sense, we consider an image fusion technique to be successful to the extent that it creates a composite that retains salient information from the sources while minimizing the number of artifacts or the amount of distortion that could interfere with interpretation.

In many applications, the ultimate user or interpreter of the fused image is a human. Consequently, the human perception of the fused image is of paramount importance and therefore, fusion results are mostly evaluated visually [80, 99, 100]. This involves human observers to judge the quality of the resulting fused images. Since the 'human quality measure' depends highly on phsycovisual factors, these subjective tests are difficult to reproduce and verify, as well as time consuming and expensive. Hence, although it cannot be denied that subjective tests are important in characterizing fusion performance, objective performance metrics appear as a valuable complementary method. But how can a subjective impression like image quality be quantified? This problem is usually solved by associating quality with the deviation of the experimental fused image from the 'ideal' fused image. Then, another problem arises, namely, how to define the 'ideal' fused image. A less usual approach is to design performance measures which, without assuming knowledge of a ground-truth, can be used for quality assessment of the fused image. These performances measures quantify the degree to which the fused image is 'related' to the input sources.

The work by Li *et al.* [51] is an example where out-of-focus image fusion is evaluated by comparison of the fused image with an 'ideal' composite created by a manual 'cut and paste' process. Indeed, various fusion algorithms presented in literature have been evaluated by constructing some kind of ideal fused image and using it as a reference for comparing with the experimental fused results. Mean squared error based metrics are widely used for these comparisons, and despite their well-known limitations, they can be helpful if used carefully. An example of such metrics is the *root mean square error* (RMSE):

$$\text{RMSE} = \left(\frac{1}{MN} \sum_{n=1}^{N} \sum_{m=1}^{M} \left(x_R(n,m) - x_F(n,m) \right)^2 \right)^{1/2} , \tag{5.1}$$

where x_R is the ideal reference, x_F the obtained fused image, and M, N are the dimensions of the images.

Information-theory related metrics [22] such as *mutual information* have also been proposed for fusion evaluation [106, 112]. Given two images x_F and x_R we define their mutual information as

$$I(x_R; x_F) = \sum_{u=1}^{L} \sum_{v=1}^{L} h_{R,F}(u,v) \log_2 \frac{h_{R,F}(u,v)}{h_R(u) h_F(v)} , \tag{5.2}$$

where h_R, h_F are the normalized graylevel histograms of x_R, x_F respectively, $h_{R,F}$ is the joint graylevel histogram of x_R and x_F, and L is the number of bins.

Let x_R, x_F correspond to the reference and fused images respectively; $I(x_R; x_F)$ indicates how much information the fused image x_F conveys about the reference x_R. Thus, the higher the mutual information between x_F and x_R, the more likely x_F resembles the ideal x_R.

Some authors [70, 93] have also proposed some objective quality measures based on perceptual criteria where the human visual system is taken into account. These measures are more elaborate and computationally intensive, but fit better with empirical subjective tests.

In [112], Zhang compares different MR image fusion approaches by standard evaluation measures such as RMSE and mutual information. He also proposes some new techniques for blindly (without a reference) estimating the quality of a noisy and blurred image, and suggests that such quality measures may be used to guide the fusion process. In those techniques, a mixture of probability densities functions is used to model the edge intensity histogram of an image. By studying the effects of the noise on the parameters of the mixture model, an estimation of the amount of noise and image quality is made.

More recently, Xydeas and Petrović [111] proposed an objective performance metric that models the accuracy with which visual information is transferred from the source images to the fused image. In their approach, important visual information is associated with edge information measured for each pixel. Correspondingly, by evaluating the relative amount of edge information that is transferred from the input images to the fused image, a measure of fusion performance is obtained.

From the current literature, we can conclude that the topic of objective performance assessment is an open problem which has received little attention. Besides, in most existing approaches, the quality of a fused image is evaluated only at very low level, and the methods used do not seem to be easily applicable to objects (high level). Much more research is required to provide valuable objective evaluation methods for fused images and, in particular, to object-oriented quality measures.

6   Conclusions

In this paper, we have introduced a general framework for MR image fusion. The proposed framework not only encompasses most of the existing MR image fusion schemes, but also allows the construction of new ones, either pixel or region-based approaches.

The region-based MR fusion scheme presented in this paper is an extension of the classical pixel-based schemes. The basic idea is to perform a MR/MM segmentation of the various input images in order to guide the fusion process. For this purpose, we developed a MR/MM segmentation method based on pyramid linking and suggested some combination algorithms which make use of the resulting segmentation. Experimental results have also been shown.

The implementation of our region-based fusion approach is still in a preliminary stage and in the experiments performed we did not attempt to optimize its performance. However, the results obtained so far suggest that our approach may be useful for several image fusion applications. We need to investigate this more thoroughly in the future. In particular, we plan to study the effect of the different parameters and functions in the scheme on the final fusion process. We also intend to design new combination algorithms and replace the MR/MM segmentation by pyramid linking by some other techniques, such as a hierarchical watershed from mathematical morphology.

A substantial part of our efforts will be devoted to the design of objective measures for fusion performance assessment. We intend to use such objective measures to evaluate and demonstrate the capacities of our region-based fusion approach, as well as to compare its performance with other MR fusion schemes. We also plan to study how these objective measures can be used to guide the fusion and improve the fusion performance.

Acknowledgements

References

[1] ABIDI, M. A., AND GONZALEZ, R. C., Eds. *Data Fusion in Robotics and Machine Intelligence.* Academic Press, San Diego, 1992.

[2] AKERMAN III, A. Pyramids techniques for multisensor fusion. In *Proceedings of SPIE* (1992), vol. 1828, pp. 124–131.

[3] BASTIERE, A. Methods for multisensor classification of airborne targets integrating evidence theory. *Aeroespace Science and Technology 2*, 6 (1998), 401–411.

[4] BECKERMAN, M., AND SWEENEY, F. J. Segmentation and cooperative fusion of laser radar image data. In *Proceedings of SPIE* (1994), vol. 2233, pp. 88–98.

[5] BÉTHUNE, S. D., MULLER, F., AND BINARD, M. Adaptive intensity matching filters: a new tool for multi-resolution data fusion, 29 September-2 October 1997. Paper presented at Symposium on Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar.

[6] BISTER, M., CORNELIS, J., AND ROSENFELD, A. A critical view of pyramid segmentation algorithms. *Pattern Recognition Letters 11* (1990), 605–617.

[7] BORGHYS, D., VERLINDE, P., PERNEEL, C., AND ACHEROY, M. Multilevel data fusion for the detection of targets using multispectral image sequences. *Optical Engineering 37*, 2 (1998), 477–484.

[8] BROUSSARD, R. P., ROGERS, S. K., OXLEY, M. E., AND TARR, G. L. Physiologically motivated image fusion for object detection using a pulse coupled neural network. *IEEE Transactions on Neural Networks 10*, 3 (1999), 554–563.

[9] BROWN, L. G. A survey of image registration techniques. *ACM Computing Survey 24*, 4 (December 1992), 325–376.

[10] BURT, P. J. The pyramid as a structure for efficient computation. In *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed. Springer-Verlag, Berlin, Germany, 1984, pp. 6–35.

[11] BURT, P. J. A gradient pyramid basis for pattern selective image fusion. In *Proceedings of the Society for Information Display Conference* (1992).

[12] BURT, P. J., AND ADELSON, E. H. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications 31* (1983), 532–540.

[13] BURT, P. J., HONG, T. H., AND ROSENFELD, A. Segmentation and estimation of image region properties through cooperative hierarchical computation. *IEEE Transactions on Systems, Man, and Cybernetics 11*, 12 (December 1981), 802–809.

[14] BURT, P. J., AND KOLCZYNSKI, R. J. Enhanced image capture through fusion. In *Proceedings of the 4th International Conference on Computer Vision* (Berlin, Germany, May 1993), pp. 173–182.

[15] CASTELLANOS, A., AND TARDOS, J. D. *Mobile Robot Localization and Map Building: A Multisensor Fusion Approach.* Kluwer Academic Publishers, Boston, MA, 2000.

[16] CHIPMAN, L. J., AND ORR, T. M. Wavelets and image fusion. In *Proceedings of the IEEE International Conference on Image Processing* (Washington D.C., October 1995), pp. 248–251.

[17] CIBULSKIS, J. M., AND DYER, C. R. An analysis of node linking in overlapped pyramids. *IEEE Transactions on Systems, Man, and Cybernetics 14* (May/June 1984), 424–436.

[18] CLAYPOOLE, R. L., DAVIS, G., SWELDENS, W., AND BARANIUK, R. D. Nonlinear wavelet transforms for image coding. In *Proceedings of the 31st Asilomar Conference on Signals, Systems, and Computers, Volume 1* (1997), pp. 662–667.

[19] CLÉMENT, V., GIRAUDON, G., HOUZELLE, S., AND SANDAKLY, F. Interpretation of remotely sensed images in a context of multisensor fusion using a multispecialist architecture. *IEEE Transactions on Geoscience and Remote Sensing 31*, 4 (1993), 779–791.

[20] COSTANTINI, M., FARINA, A., AND ZIRILLI, F. The fusion of different resolution SAR images. *Proceedings of the IEEE 81*, 1 (1997), 139–146.

[21] COULOIGNER, I., RANCHIN, T., VALTONEN, V. P., AND WALD, L. Benefit of the future SPOT-5 and of data fusion to urban roads mapping. *International Journal of Remote Sensing 19*, 8 (1998), 1519–1532.

[22] COVER, T. M., AND THOMAS, J. A. *Elements of information theory*. John Wiley and Sons, New York, 1991.

[23] DANIEL, M. M., AND WILLSKY, A. S. A multiresolution methodology for signal-level fusion and data assimilation with application to remote sensing. *Proceedings of the IEEE 85*, 1 (1997), 164–180.

[24] DASARATHY, B. V. *Decision Fusion*. IEEE Computer Sociaty Press, Los Alamitos, California, 1994.

[25] DASARATHY, B. V. Fuzzy evidential reasoning approach to target identity and state fusion in multisensor environments. *Optical Engineering 36*, 3 (1997), 699–683.

[26] DAUBECHIES, I. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

[27] DAUBECHIES, I., AND SWELDENS, W. Factoring wavelet transforms into lifting steps. *Journal of Fourier Analysis and Applications 4*, 3 (1998), 245–267.

[28] DE QUEIROZ, R. L., FLORÊNCIO, D. A. F., AND SCHAFER, R. W. Nonexpansive pyramid for image coding using a nonlinear filterbank. *IEEE Transactions on Image Processing 7* (1998), 246–252.

[29] DUBUISSON, M., AND JAIN, A. K. Contour extraction of moving objects in complex outdoor scenes. *International Journal of Computer Vision 14* (1995), 83–105.

[30] EGGER, O., LI, W., AND KUNT, M. High compression image coding using an adaptive morphological subband decomposition. *Proceedings of the IEEE 83* (1995), 272–287.

[31] FECHNER, T., AND GODLEWSKI, G. Optimal fusion of TV and infrared images using artificial neural networks. In *Proceedings of SPIE* (May 1995), vol. 2492, pp. 919–925.

[32] FLORÊNCIO, D. A. F., AND SCHAFER, R. W. A non-expansive pyramidal morphological image coder. In *Proceedings of the IEEE International Conference on Image Processing* (Austin, Texas, 1994), pp. 331–335.

[33] GEREK, O. N., AND ÇETIN, A. E. Image coding using adaptive subband decomposition. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Seattle, Washington, May 1998).

[34] GERONIMO, S., HARDIN, D. P., AND MASSOPUST, P. R. Fractal functions and wavelet expansions based on several functions. *Journal of Approximation Theory 78*, 3 (1994), 373–401.

[35] GOUTSIAS, J., AND HEIJMANS, H. J. A. M. Multiresolution signal decomposition schemes. Part I: Morphological pyramids. *IEEE Transactions on Image Processing 9*, 11 (2000).

[36] HAMPSON, F. J., AND PESQUET, J.-C. A nonlinear subband decomposition with perfect reconstruction. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Atlanta, Georgia, May 7-10, 1996), pp. 1523–1526.

[37] HAMPSON, F. J., AND PESQUET, J.-C. *M*-band nonlinear subband decompositions with perfect reconstruction. *IEEE Transactions on Image Processing 7* (1998), 1547–1560.

[38] HEIJMANS, H. J. A. M., AND GOUTSIAS, J. Nonlinear multiresolution signal decomposition schemes. Part II: morphological wavelets. *IEEE Transactions on Image Processing 9*, 11 (2000), 1897–1913.

[39] HILL, D., EDWARDS, P., AND HAWKES, D. Fusing medical images. *Image Processing 6*, 2 (1994), 22–24.

[40] HONG, T. H., AND ROSENFELD, A. Compact region extraction using weighted pixel linking in a pyramid. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*, 2 (1984), 222–229.

[41] JEON, B., AND LANDGREBE, D. A. Decision fusion approach for multitemporal classification. *IEEE Transactions on Geoscience and Remote Sensing 37*, 3 (1999), 1227–1233.

[42] KAM, M., ZHU, X., AND KALATA, P. Sensor fusion for mobile robot navigation. In *Proceedings of the IEEE* (January 1997), vol. 85, pp. 108–119.

[43] KOHONEN, T. *Self-organizing maps*. Springer-Verlag, 1995.

[44] KOREN, I., LAINE, A., AND TAYLOR, F. Image fusion using steerable dyadic wavelet transforms. In *Proceedings of the IEEE International Conference on Image Processing* (Washington D.C., October 1995), pp. 232–235.

[45] KOVACEVIĆ, J., AND SWELDENS, W. Wavelet families of increasing order in arbitrary dimensions. *IEEE Transactions on Image Processing 9*, 3 (2000), 480–496.

[46] KOVACEVIĆ, J., AND VETTERLI, M. Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathbb{R}^n$. *IEEE Transactions on Information Theory 38* (1992), 533–555.

[47] KRUSKAL, J. B. Non metric multidimensional scaling: a numerical method. *Psychometrika 29* (1964), 115–129.

[48] LALLIER, E. Real-Time Pixel-Level Image Fusion through Adaptive Weight Averaging. Technical report, Royal Military College of Canada, 1999.

[49] LECKIE, D. G. Synergism of SAR and visible/infrared data for forest type discrimination. *Photogrammetric Engineering and Remote Sensing 56* (1990), 1237–1246.

[50] LI, H., MANJUNATH, B. S., AND MITRA, S. K. A contour based approach to multisensor image registration. *IEEE Transactions on Image Processing 4*, 3 (March 1995), 320–334.

[51] LI, H., MANJUNATH, B. S., AND MITRA, S. K. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing 57*, 3 (May 1995), 235–245.

[52] LI, S. T., AND WANG, Y. N. Multisensor image fusion using discrete multiwavelet transform. In *Proceedings of the 3rd International Conference on Visual Computing* (Mexico city, Mexico, September 2000).

[53] LIU, Z., TSUKADA, K., HANASAKI, K., HO, Y. K., AND DAI, Y. P. Image fusion by using steerable pyramid. *Pattern Recognition Letters 22* (2001), 929–939.

[54] LOU, K. N., AND LIN, L. G. An intelligent sensor fusion system for tool monitoring on a machining centre. *International Journal of Advanced Manufacturing Technology, 13* (1997), 556–565.

[55] LUO, R. C., AND KAY, M. G. *Multisensor Integration and Fusion for Intelligent Machines and Systems*. Ablex Publishing Corporation, 1995.

[56] MALLAT, S. G. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, California, 1998.

[57] MALLAT, S. G., AND ZHONG, S. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence 14* (1992), 710–732.

[58] MANDUCA, A. Multispectral image visualization with nonlinear projections. *IEEE Transactions on Image Processing 5*, 10 (1996), 1486–1490.

[59] MANGOLINI, M. *Apport de la fusion d'images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation.* PhD thesis, University of Nice-Sophia Antipolis, France, November 1994.

[60] MARR, D. *Vision.* W. H. Freeman and Company, 1982.

[61] MATSOPOULOS, G. K., AND MARSHALL, S. Application of morphological pyramids: fusion of MR and CT phantoms. *Journal of Visual Communication and Image Representation 6*, 2 (1995), 196–207.

[62] MATSOPOULOS, G. K., MARSHALL, S., AND BRUNT, J. N. H. Multiresolution morphological fusion of MR and CT images of the human brain. In *Proceedings of the IEE on Vision, Image and Signal Processing* (June 1994), vol. 141, pp. 137–142.

[63] MCMICHAEL, D. W. Data fusion for vehicle-borne mine detection. In *EUREL Conference on Detection of Abandoned Land Mines* (1996), pp. 167–171.

[64] MILISAVLJEVIĆ, N., AND ACHEROY, M. An approach to the use of bayesian rule in decision level fusion for multisensor mine detection. In *Proceedings of Physics in Signal and Image Processing Conference* (Paris, France, 1999), pp. 216–266.

[65] MIRHOSSEINI, A. R., HONG, Y., KIN, M. L., AND TUAN, P. Human face image recognition: an evidence aggregation approach. *Computer Vision and Image Understanding 71*, 2 (1998), 213–230.

[66] MUKHOPADHYAY, S., AND CHANDA, B. Fusion of 2D grayscale images using multiscale morphology. *Pattern Recognition 34* (2001), 1939–1949.

[67] MURPHY, R. R. Sensor and information fusion improved vision-based vehicle guidance. *IEEE Intelligent Systems 13*, 6 (1999), 49–56.

[68] NACKEN, P. F. M. Image segmentation by connectivity preserving relinking in hierarchical graphs structures. *Pattern Recognition 9*, 6 (1995), 907–920.

[69] NEWMAN, E. A., AND HARTLINE, P. H. The infrared vision of snakes. *Scientific American 246*, 3 (1982), 116–127.

[70] PAPPAS, T. N., AND SAFRANEK, R. J. Perceptual criteria for image quality evaluation, 1999. Available at http://ftp.eecs.umich.edu/people/neuhoff/.

[71] PIELLA, G., AND HEIJMANS, H. J. A. M. Adaptive lifting schemes with perfect reconstruction. Research Report PNA-R0104, CWI, Amsterdam, February 2001. To appear in IEEE Transactions on Signal Processing.

[72] POHL, C., AND GENDEREN, J. L. Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing 19*, 5 (1998), 823–854.

[73] PU, T., AND NI, G. Contrast-based image fusion using the discrete wavelet transform. *Optical Engineering 39*, 8 (August 2000), 2075–2082.

[74] RANCHIN, T., AND WALD, L. The wavelet transform for the analysis of remotely sensed images. *International Journal of Remote Sensing 14* (1993), 615–619.

[75] RANCHIN, T., AND WALD, L. Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. *Photogrammetric Engineering and Remote Sensing 66* (2000), 49–61.

[76] RANCHIN, T., WALD, L., AND MANGOLINI, M. The ARSIS method: a general solution for improving spatial resolution of images by the means of sensor fusion. In *Proceedings of the EARSeL Conference on Fusion of Earth Data* (Cannes, France, February 1996).

[77] REED, J. M., AND HUTCHINSON, S. Image fusion and subpixel parameter estimation for auto-mated optical inspection of electronic components. *IEEE Transactions on Industrial Electronics 43*, 3 (1996), 346–354.

[78] RICHARDS, J. A. Thematic mapping from multitemporal image data using the principal com-ponent transformation. *Remote Sensing of Environment 16* (1984), 36–46.

[79] ROCKINGER, O. Pixel-level fusion of image sequences using wavelets frames. In *Proceedings of the 16th Leeds Applied Shape Research Workshop* (1996), Leeds University Press.

[80] RYAN, D., AND TINKLER, R. Night pilotage assessment of image fusion. In *Proceedings of Hel-met and Head-Mounted Displays and Symbology Design Requirements* (1995), vol. 2465, SPIE, pp. 50–67.

[81] SAMMON, J. W. A nonlinear mapping for data nalysis. *IEEE Transactions on Computers C-18* (1969), 401–409.

[82] SCHEUNDERS, P. Multiscale edge representation applied to image fusion. In *Proceedings of Wavelet Applications in Signal and Image Processing VIII* (San Diego, USA, 30 July-4 August 2000), SPIE.

[83] SERRA, J. *Image Analysis and Mathematical Morphology*, vol. 1. Academic Press, New York, 1982.

[84] SHARMA, R. K., AND PAVEL, M. Adaptive and statistical image fusion. *Society for Information Display Digest of Technical Papers 27* (1996), 969–972.

[85] SIMONCELLI, E. P., AND FREEMAN, W. T. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *Proceedings of the IEEE International Conference on Image Processing* (1995), pp. 444–447.

[86] SIMONCELLI, E. P., FREEMAN, W. T., ADELSON, E. H., AND HEEGER, D. J. Shiftable multi-scale transforms. *IEEE Transactions on Information Theory 38*, 2 (1992), 587–607.

[87] SOILLE, P. *Morphological Image Analysis*. Springer-Verlag, Berlin, 1999.

[88] STRELA, V., HELLER, N., AND STRANG, G. The applications of multiwavelets filter banks to signal and image processing. *IEEE Transactions on Image Processing 8*, 4 (1996), 548–563.

[89] STRELA, V., AND STRANG, G. Finite element multiwavelets. In *Proceedings NATO Conference* (Boston, MA, 1995), Kluwer.

[90] SWELDENS, W. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. In *Wavelet Applications in Signal and Image Processing III* (1995), A. F. Lain and M. Unser, Eds., Proceedings of SPIE, vol. 2569, pp. 68–79.

[91] SWELDENS, W. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis 3* (1996), 186–200.

[92] SWORDER, D. D., BOYD, J. E., AND CLAPP, G. A. Image fusion for tracking manoeuvring targets. *International Journal of Systems Science 28*, 1 (1997), 1–14.

[93] TEO, P. C., AND HEEGER, D. J. Perceptual image distortion. In *Proceedings of the IEEE International Conference on Image Processing* (1994), vol. 2, pp. 982–986.

[94] THÉVENAZ, P., AND UNSER, M. An efficient mutual information optimizer for multiresolution image registration. In *Proceedings of the IEEE International Conference on Image Processing* (1998), vol. 1, pp. 833–837.

[95] TOET, A. Image fusion by a ratio of low-pass pyramid. *Pattern Recognition 9* (1989), 245–253.

[96] TOET, A. A morphological pyramidal image decomposition. *Pattern Recognition Letters 9* (1989), 255–261.

[97] TOET, A. Hierarchical image fusion. *Machine Vision Application* (March 1990), 1–11.

[98] TOET, A. Multiscale contrast enhancement with application to image fusion. *Optical Engineering 31*, 5 (1992), 1026–1031.

[99] TOET, A., IJSPEERT, J. K., WAXMAN, A. M., AND AGUILAR, M. Fusion of visible and thermal imagery improves situational awareness. In *Proceedings of SPIE Conference on Enhanced and Synthetic Vision* (1997), vol. 3088, pp. 177–188.

[100] TOET, A., SCHOUMANS, N., AND IJSPEERT, J. K. Perceptual evaluation of different nighttime imaging modalities. In *Proceedings of the 3rd International Conference on Information Fusion* (Paris, France, July 2000), vol. 1, International Society of Information Fusion (ISIF).

[101] TOET, A., AND WALRAVEN, J. New false color mapping for image fusion. *Optical Engineering 35*, 3 (1996), 650–658.

[102] TOUTIN, T. SPOT and Landsat stero fusion for data extraction over mountainous areas. *Photogrammetric Engineering and Remote Sensing 64*, 2 (1998), 109–113.

[103] VAN ELSEN, P. A., POL, E. J. D., AND VIERGEVER, M. A. Medical image matching– A review with classification. *IEEE Transactions on Engineering Medical Biology 12* (March 1993), 26–39.

[104] VETTERLI, M., AND KOVAČEVIĆ, J. *Wavelets and Subband Coding*. Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[105] VINCKEN, K. L., KOSTER, A. S. E., AND VIERGER, M. A. Probabilistic multiscale image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 2 (1997), 109–120.

[106] WALD, L., RANCHIN, T., AND MANGOLINI, M. Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing 63*, 3 (1997), 691–699.

[107] WAXMAN, A. M., GOVE, A. N., FAY, D. A., RACAMATO, J. P., CARRICK, J. E., SEIBERT, M. C., AND SAVOYE, E. D. Color night vision: opponent processing in the fusion of visible and IR imagery. *Neural Networks 10*, 1 (1997), 1–6.

[108] WICKERHAUSER, V. M., AND COIFMAN, R. R. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory 38* (1992), 713–718.

[109] WILSON, T. A., ROGERS, S. K., AND MEYERS, L. R. Perceptual based hyperspectral image fusion using multiresolution analysis. *Optical Engineering 34*, 11 (1995), 3154–3164.

[110] WONG, S. T. C., KNOWLTON, R. C., HAWKINS, R. A., AND LAXER, K. D. Multimodal image fusion for noninvasive epilepsy surgery planning. *IEEE Transactions on Computer Graphics and Applications 16*, 1 (1996).

[111] XYDEAS, C., AND PETROVIĆ, V. Objective pixel-level image fusion performance measure. In *Proceedings of SPIE* (April 2000), vol. 4051, pp. 88–99.

[112] ZHANG, Z. *Investigations of image fusion*. PhD thesis, Lehigh University, April 1999.

[113] ZHANG, Z., AND BLUM, R. A region-based image fusion scheme for concealed weapon detection. In *Proceedings of the 31th Annual Conference on Information Sciences and Systems* (March 1997), pp. 168–173.

[114] ZHANG, Z., AND BLUM, R. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE 87*, 8 (1999), 1315–1326.