Adaptive wavelets and their applications to image fusion and compression

Piella, G.

Citation for published version (APA):
Chapter 7

Region-based multiresolution image fusion

The algorithms based on multiresolution (MR) techniques that we have discussed in the previous chapter 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, in the input images depicted in Fig. 6.1, a composite 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 these images, it is reasonable to consider the pixels which constitute these objects as entities instead 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 may consider region-based approaches. Such approaches have the additional advantage that the fusion process becomes more robust and may help to circumvent some of the well-known drawbacks of pixel-based techniques, such as blurring effects and high sensitivity to noise and misregistration.

In this chapter, we introduce a new region-based approach to MR fusion which combines aspects of feature and pixel fusion. The basic idea is to build a segmentation based on all different source images and to use this segmentation to guide the combination process. A major difference with other existing region-based approaches [105,174] is that the segmentation performed is: (i) multisource, in the sense that a single segmentation is obtained from all the input images, and (ii) multiresolution, in the sense that it is computed in an MR fashion (thus, we do not merely compute independent segmentations of images at different resolutions). For instance, in [174] the regions are obtained by segmenting (independently) each of the approximation images $x_n^S$ and by exploiting the tree-structure in the decomposition: every detail coefficient $y_n^S(\cdot)$ is assigned to a region in $x_n^S$. 

157
7.1 The overall scheme: from pixels to regions

7.1.1 Introduction

Our region-based fusion scheme depicted in Fig. 7.1 extends the pixel-based fusion approach discussed in Chapter 6 (see Fig. 6.7). 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 \( R \) (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/multisource (MR/MS) segmentation to partition the image domain at these scales into regions. The activity and match measures are now computed for every such region. These measures may correspond to low-level as well as intermediate-level structures. Furthermore, the MR segmentation \( R \) 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 sources. This combination results in an MR decomposition \( y_F \), and by MR synthesis we obtain a composite image \( x_F \).

The main functional blocks of this fusion strategy are depicted in Fig. 7.1. Since we already discussed most of them in Chapter 6, we concentrate on the segmentation module and its interaction with the other modules.

7.1.2 MR/MS segmentation

The MR/MS segmentation uses the various source images as input and returns a single MR segmentation

\[
R = \{R^1, R^2, \ldots, R^K\}
\]

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

Loosely speaking, \( R \) provides an 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 as the constituent parts of the objects in the overall scene.

In our image fusion problem, segmentation is merely a preparatory step toward actual fusion. In fact, we are not interested in the segmentation of the images \( \text{per se} \), but rather in a coarse partition of the underlying scene. Therefore, the segmentation process does not need to be extremely accurate. For our purposes, we have developed an MR/MS segmentation algorithm based on pyramid linking [20]. We describe our segmentation algorithm in Section 7.2. Obviously, other segmentation methods can be used, as long as they meet the constraint that the sampling structure in \( R \) is the same as in \( y_s \), so that each partition \( R^k \) corresponds to a partition of the detail image \( y^k_s(\cdot|p) \).
7.1. The overall scheme: from pixels to regions

![Diagram of region-based MR fusion scheme]

Figure 7.1: Generic region-based MR fusion scheme with two input sources $x_A$ and $x_B$, and one output composite image $x_F$.

7.1.3 Combination algorithm

Since the building blocks of the combination algorithm in the region-based approach are essentially the same as in the pixel-based case, the combination algorithms discussed in Chapter 6 can be easily extended to the region-based approach. For example, we can define the activity of each region $R \in \mathcal{R}^k$ in $y_S^k(\cdot|p)$ by

$$a_S^k(R|p) = \frac{1}{|R|} \sum_{n \in R} a_S^k(n|p). \quad (7.1)$$

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

$$m_{AB}^k(R|p) = \frac{1}{|R|} \sum_{n \in R} m_{AB}^k(n|p). \quad (7.2)$$

Given these measures, the decision map can be constructed in several ways as discussed in Section 6.2.7, with the only difference that $a_S^k(R|p), m_{AB}^k(R|p)$ are used instead of $a_S^k(n|p)$. 
\( m^{k}_{AB}(n|p) \). For instance, a combination algorithm based on a maximum selection rule (see (6.5) for the pixel-based case) would read:

\[
y^{k}_{F}(n|p) = \begin{cases} 
  y^{k}_{A}(n|p) & \text{if } a^{k}_{A}(R|p) > a^{k}_{B}(R|p) \\
  y^{k}_{B}(n|p) & \text{otherwise}
\end{cases} \quad \text{for all } n \in R. \tag{7.3}
\]

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 composite image \( x_{F} \).

Note that for the particular case in which each region corresponds to a single point \( n \), 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.

### 7.2 MR/MS segmentation based on pyramid linking

In this section we present an MR/MS 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 several input images.

#### 7.2.1 The linked pyramid

The linked pyramid structure was first described by Burt et al. [20] (related work can be found in [9, 28, 75, 113, 164]). It consists of an MR decomposition of an image with the bottom level containing the full-resolution image, and each successive higher level containing a filtered and subsampled image derived from the level below it. The various levels of the pyramid are 'linked' by means of so-called child-parent relations (see Fig. 7.2) 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 (see Section 2.3) is produced by low-pass filtering and sampling. Then, child-parent relations are established by linking each pixel at a given level (called \textit{child}) to one of the pixels in the next higher level (called \textit{parent}) which is closest in gray value (or in some other 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 [20]). At the end (or possibly during the linking process), some pixels are labeled as \textit{roots}. In the simplest case, only pixels at the top level of the pyramid are roots. Every root and the pixels which are connected to it induce a tree in the pyramid. The leaves of each tree correspond to pixels in the full-resolution image which define a \textit{segment} or \textit{region}. Thus, the linked pyramid provides a framework for an iterative process of image segmentation. For example, in Fig. 7.2, pixel T is a root which represents a segment at the bottom level composed of pixels a, b, c and d.

There exist many variations on the scheme above. This may concern the way the initial pyramid is built, the manner pixels are linked to each other, the choice when pixels should be
declared as roots, the size of the neighborhood in which children can look for a parent to link to, 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 been focused on specific problems or aspects. Two major problems are 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 of regions. Pixels which are adjacent at some higher level do not necessarily represent adjacent regions at 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 [113]. The second problem concerns root characterization. In Burt's original approach, only 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 roots as those pixels which are not linked 'strongly' enough to a parent [75,164]. Now, the problem of root characterization is reduced to the definition of link strength and the choice of a root labeling threshold.

### 7.2.2 MR segmentation algorithm using linking

Our basic algorithm follows the classical '50% overlapping 4 x 4' structure [20]. This means that each parent is derived from the pixels in the 4 x 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; see Fig. 7.3. 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 highest level is considered to be $K_M = \log_2 N - 1$.

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

$$C(\mathbf{n}) = \{ (m', n') \mid m' \in \{2m - 1, 2m, 2m + 1, 2m + 2 \}, n' \in \{2n - 1, 2n, 2n + 1, 2n + 2 \} \} .$$

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

$$P(\mathbf{n}) = \{ (m', n') \mid m' \in \{ \lfloor \frac{1}{2}(m - 1) \rfloor, \lfloor \frac{1}{2}m \rfloor, \lfloor \frac{1}{2}(m + 1) \rfloor \}, n' \in \{ \lfloor \frac{1}{2}(n - 1) \rfloor, \lfloor \frac{1}{2}n \rfloor, \lfloor \frac{1}{2}(n + 1) \rfloor \} \} ,$$

where $\lfloor . \rfloor$ denotes the integer part of the enclosed value. We define the receptive field of a pixel $\mathbf{n}$ as the set of pixels at level zero that belong to the same tree as $\mathbf{n}$. 

---

**Figure 7.2:** 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.
Chapter 7. Region-based multiresolution image fusion

Figure 7.3: Parent-child relations. The dark pixel at the lower level should choose a parent within the 4 candidate parents at the next finer level. Each of the candidate parents has 16 children. E.g., the bottom-left gray pixel at the finer level has as children the pixels shaded in gray in the lower level.

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 $n$ at level $k$ its grayscale value $x^k(n)$, and the area $A^k(n)$ 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 $n$ at level zero the gray value $x^0(n)$ of the original image, and to each pixel $n$ at level $k > 0$ a gray value $x^k(n)$ computed from the average of the gray values of its candidate children:
   
   $$x^k(n) = \frac{1}{16} \sum_{n' \in C(n)} x^{k-1}(n').$$

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 grayscale proximity. A distance measure between the child and each of its four candidate parents is computed. A link is established with the parent that minimizes that distance. It may occur that more than one candidate parent minimizes the distance measure. In this case we arbitrarily pick one of them. A simple choice for the distance measure is the absolute difference in grayscale. Examples of other distances can be found in [113,164].

   In our approach, we perform the root labeling within the linking step. That is, when trying to link to a parent, the link is not established if the minimal distance is above some threshold. In such a case the pixel is labeled as a root (thus, it is not considered to be a child any more). We refer to [75,164] for other alternatives.
7.2. MR/MS segmentation based on pyramid linking

An advantage of this method is its speed: a single operation will identify all roots. A disadvantage is that it is not clear beforehand how many roots (and therefore, how many segments) will be found. Defining a good root labeling threshold is not straightforward. When the threshold is too high, few pixels become roots, whereas many pixels are labeled as root if the threshold is too low. For simplicity, we use a threshold \( T = 0.25 \Delta x^0 \) where \( \Delta x^0 \) is the length of the dynamic range of the input image \( x^0 \).

(b) Updating area \( A^k \) and gray values \( x^k \).

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

\[
A^{k+1}(n) = \sum_{n' \in C(n)} A^k(n')
\]

\[
x^{k+1}(n) = \frac{\sum_{n' \in C(n)} x^k(n') A^k(n')}{A^{k+1}(n)}
\]

where \( A^0(n) = 1 \) for all \( n \) at level zero.

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

3. Segmentation

The actual segmentation is obtained by using the tree structures that have been created. At a given level \( k \), pixels that are connected to a common root are classified as a single region segment. In this way, we obtain a segmentation \( R^k \) of the \( k \)th-level approximation image.

7.2.3 MR/MS segmentation algorithm using linking

In the previous subsections we have discussed how to obtain an MR segmentation from a single input. Now we address the more difficult problem of how to compute a single MR segmentation based on multiple source images. We show that the segmentation method presented in the last subsection can be extended to 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, the distance between a child \( n \) and a candidate parent \( n' \in \mathcal{P}(n) \) is given by the expression

\[
\left( \sum_{S \in \mathcal{S}} (x^k_S(n) - x^{k+1}_S(n'))^2 \right)^{1/2}
\]

As in the scalar case, the candidate \( n' \) which minimizes this distance is selected to become the parent unless the distance is above some threshold, in which case \( n \) 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. In this way, we obtain a single linked pyramid structure and we can apply the same segmentation step as before.

We summarize the basic steps of our MR/MS segmentation in the following algorithm.
Chapter 7. Region-based multiresolution image fusion

Algorithm

1. For each input $S \in S$
   - Construct an approximation pyramid $\{x^k_S\}$.

2. For each level $k < K_M$
   - While no convergence,
     * For each child $n$ at level $k$, find parent $n' \in \mathcal{P}(n)$ which minimizes the distance given by (7.4). If this distance is above some threshold, $n$ is set as a root, otherwise it is linked to $n'$.
     * For each parent $n$ at level $k + 1$, update $A^{k+1}(n)$ and $x^{k+1}_S(n)$ for all $S \in S$.

3. For each level $k$
   - All pixels $n$ at level $k$ connected to a common root are classified as a single region segment in $\mathcal{R}^k$.

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 may pre-process (e.g., normalization of amplitudes, denoising, etc.) the input images so that their attributes become comparable. Alternatively, one can modify the distance measure in (7.4) and use, for instance,

$$\left( \frac{\sum_{S \in S} \mu_S (x^k_S(n) - x^{k+1}_S(n'))^2}{\mu_S} \right)^{1/2},$$

where $\mu_S$ is a nonnegative 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, other root criteria, adaptive windows and probabilistic linking [113, 141, 164].

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 0 to $K_M$, whereas the levels from the MR decompositions $y_S$ go from 1 to $K$. In practice, $K$ is smaller than $K_M$, so we assume henceforth that $K \leq K_M$. In addition, we assume that the MR/MS segmentation module associates to each image $y^k(\cdot|p)$, the partition $\mathcal{R}^{k'}$ such that they have the same dimensions and sampling structure. 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 discrete wavelet transform, then $k' = k$ for $k = 1, \ldots, K$ and all $p = 1, \ldots, 3$. 
7.3 Experimental results

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

7.3.1 Case studies

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/MS segmentation algorithm discussed in Section 7.2. In the combination algorithm, we do not use a matching measure and define the activity of each region $R \in \mathcal{R}^k$ as in (7.1), with $d^k_S(n|p) = |g^k_S(n|p)|$. The combination process is performed as in (6.3), 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 $p = 0$,
  \[ \delta = d^k(n|0) = \frac{1}{2}, \text{ for all } n. \]

- For $p = 1$,
  \[ \delta = d^k(n|1) = \begin{cases} 1 & \text{if } a^k_A(R|1) > a^k_B(R|1) \\ 0 & \text{otherwise} \end{cases}, \text{ for all } n \in R. \]

Note that according to this algorithm, the composite approximation image $y^k_F(\cdot|0)$ is the pixel-wise average of the approximation images $y^k_S(\cdot|0)$ and therefore, the region information $\mathcal{R}^k$ is neglected. The composite detail images $y^k_F$, however, are constructed by a selective combination as in (7.3).

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 first row of each figure shows the input sources $x_A$ and $x_B$. The second row depicts the segmentation and decision map at the first level of decomposition. For the decision maps, black and white pixels correspond to $\delta = 0$ and $\delta = 1$, respectively. Thus, according to our algorithm, coefficients corresponding to ‘white zones’ are selected from $y^k_A$, while coefficients corresponding to ‘black zones’ are selected from $y^k_B$.

Case 1: fusion of visible and IR wavelength images - Fig. 7.4

Fig. 7.4 shows the input images, their corresponding segmentation and decision maps at levels 1 and 2, and the resulting composite image. It is interesting to note that, according to $d^2$ (bottom-middle of Fig. 7.4), 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 $y_B^2$. Note that in the
composite image (bottom-right of Fig. 7.4) there is less contrast between the person and the background than in the IR image. This is due to the fact that the approximation images at the coarsest level are averaged, i.e., no region information has been used there.

Figure 7.4: Case 1. Top: visual (left) and IR (right) input images. Middle: segmentation (left) and decision map (right) at level 1; Bottom: segmentation (left) and decision map (middle) at level 2, and composite image (right).
7.3. Experimental results

Case 2: fusion of images with different focus points - Fig. 7.5
As before, the second row (in Fig. 7.5) shows the segmentation and decision map at level 1. 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 $y^1_L$ instead from $y^1_B$. The same happens at level $k = 2$ (not displayed here). The third row shows, from right to left, the resulting composite image and the composite image corresponding to a pixel-based MR fusion with the same fusion rules as in the region-based case. The bottom row illustrates how we can improve the region-based composite 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 [140]. The filtered $d^1$ is shown at the bottom left of Fig. 7.5. One can see that small details have been removed and that the boundaries have been smoothed. The composite image obtained with the filtered decision maps is shown at the bottom right.

Case 3: fusion of a magnetic resonance image (MRI) and a computer tomography (CT) image - Fig. 7.6
In this last example, we illustrate the combination of the approximation coefficients using an activity based on a local variance (see below). More precisely, we perform the selective combination in (7.3) for both detail and approximation coefficients but using different activity measures. For the details, $a^k_S(R|1)$ is defined as before, while for the approximation we consider

$$a^k_S(R|0) = \frac{1}{|R|} \sum_{n \in R} \left( y^k_S(n|0) - \tilde{y}^k_S(R|0) \right)^2,$$

where $\tilde{y}^k_S(R|0) = \frac{1}{|R|} \sum_{n \in R} y^k_S(n|0)$.

Fig. 7.6 shows the input images, their corresponding segmentation and decision maps at levels 1, 2 and 3, the resulting composite image and, for comparison, two other composite images obtained with different algorithms than the above. By visual inspection, we can see that the proposed region-based fusion (right image of the third row) preserves the soft tissue depicted in the MRI image better than the pixel-based fusion (left image of the bottom row). The right image of the bottom row is the composite image resulting from a fusion algorithm where the region-based approach is only used for the detail images (as we did in the previous experiments).

7.3.2 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.

The importance of the various parameters in our approach is a topic that needs much more investigation. In particular, our MR/MS segmentation is very sensitive to the root labeling criteria and the threshold proposed in Section 7.2.2 does not always give a satisfactory segmentation.
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 have been developed (see Chapter 8) to evaluate and demonstrate the capacities of the new fusion technique, as well as to compare its performance with other MR fusion schemes.
Figure 7.5: Case 2. Top: multi-focus input images. Second row: segmentation (left) and decision map (right) at level 1. Third row: composite images with pixel-based (left) and region-based (right) approach. Bottom: filtered decision map (left) and corresponding composite image (right).
Figure 7.6: Case 3. Top: MRI (left) and CT (right) input images. Second row: segmentation (left) and decision map (right) at level 1. Third row: segmentation (left) and decision map (middle) at levels 2 and 3, and composite image (right). Bottom: pixel-based (left) and region-based (right) composite images with pixel-wise average combination for the approximation images.