International Journal of Computer Vision
© 2006 Springer Science + Business Media, LLC. Manufactured in The Netherlands.
DOI: 10.1007/s11263-005-6299-0

Scale-Sets Image Analysis

LAURENT GUIGUES
Institut Géographique National, MATIS Lab., 2-4. av. Pasteur 94165 Saint-Mandé Cedex France
laurent.guigues@creatis.insa-lyon.fr

JEAN PIERRE COCQUEREZ
Université de Technologie de Compiègne, HEUDIASYC Lab., UMR CNRS 6599,
BP 20529 Compiègne Cedex France
尺度集 cocquerez@hds.utc.fr

HERVÉ LE MEN
Institut Géographique National, MATIS Lab., 2-4. av. Pasteur 94165 Saint-Mandé Cedex France
herve.le-men@ign.fr

Received September 13, 2004; Revised November 8, 2005; Accepted November 8, 2005

First online version published in April, 2006

Abstract. This paper introduces a multi-scale theory of piecewise image modelling, called the scale-sets theory, and which can be regarded as a region-oriented scale-space theory. The first part of the paper studies the general structure of a geometrically unbiased region-oriented multi-scale image description and introduces the scale-sets representation, a representation which allows to handle such a description exactly. The second part of the paper deals with the way scale-sets image analyses can be built according to an energy minimization principle. We consider a rather general formulation of the partitioning problem which involves minimizing a two-term-based energy, of the form $\lambda \mathsf{C} + \mathsf{D}$, where D is a goodness-of-fit term and C is a regularization term. We describe the way such energies arise from basic principles of approximate modelling and we relate them to operational rate/distorsion problems involved in lossy compression problems. We then show that an important subset of these energies constitutes a class of multi-scale energies in that the minimal cut of a hierarchy gets coarser and coarser as parameter λ increases. This allows us to devise a fast dynamic-programming procedure to find the complete scale-sets representation of this family of minimal cuts. Considering then the construction of the hierarchy from which the minimal cuts are extracted, we end up with an exact and parameter-free algorithm to build scale-sets image descriptions whose sections constitute a monotone sequence of upward global minima of a multi-scale energy, which is called the "scale climbing" algorithm. This algorithm can be viewed as a continuation method along the scale dimension or as a minimum pursuit along the operational rate/distorsion curve. Furthermore, the solution verifies a linear scale invariance property which allows to completely postpone the tuning of the scale parameter to a subsequent stage. For computational reasons, the scale climbing algorithm is approximated by a pair-wise region merging scheme: however the principal properties of the solutions are kept. Some results obtained with Mumford-Shah's piece-wise constant model and a variant are provided and different applications of the proposed multi-scale analyses are finally sketched.

Keywords: image segmentation, multi-scale modelling, energy minimization, hierarchy

1. Introduction

1.1. *The Need for Multi-Scale Low Level Image Descriptions*

Following David Marr's computational theory of vision (Marr, 1982), a number of image analysis systems are based on a bottom-up architecture, made up of two global stages. A *low level* analysis of the incoming signal is first performed, which builds a low level description of the data in terms of primitive structures (characteristic points, contours, regions, optical flow, disparity map. . .). Based on this low level description, *high level* vision tasks are then issued (object recognition, scene modelling, interpretation. . .). In its most general—and radical—form, Marr's theory assumes a complete independence between the low and the high level stages of vision. The low level processes are supposed to output a general-purpose description of the image that is a description independent from any specific high level task.

Now, the structures which can be useful to high level reasoning can be located at arbitrary positions in an image, can have any orientation and can be of any size. Also, useful structures can have arbitrary image contrasts, that is they can be very salient as well as very poorly contrasted. Hence, a low level image description should be *uncommitted* in terms of position, orientation, size and contrast (Lindeberg, 1994). Size refers to the characteristic spatial extension of a phenomenon while contrast refers to its characteristic extension in the image's values space. Size and contrast can thus be thought of as two different measures of the "scale" of a structure along the two basic dimensions of an image: the spatial dimension and the values dimension. Hence, low-level image analyses should be multi-scale analyses both in the spatial and in the value sense.

1.2. *From Scale-Space. . .*

Marr proposed to start image analyses with the elaboration of a low level description of the image in terms of *contours*—the *raw primal sketch*—and put forward the idea to look for transitions of different spatial extensions. Indeed, a progressive transition in an image content becomes more and more local as the distance of observation or as the size of an analyzing tool increases. "The critical observation" by Marr (1982) (p. 69) "is that, at their own scale, these things [the transitions in some image features] can all be thought of as spatially localized". The raw primal sketch was based on convolving the image with the Laplacian of Gaussian kernels of different spatial extensions (variances), detecting the zero crossings and looking for the coincidence of responses for different scales (2 or 3).

Gaussian filtering does nothing else than blurring the image: it deletes the finest structures first but it also progressively 'blurrs' the shapes, coarsening their *geometrical description*. To cope with this problem, Witkin (1983) proposed to consider the family of convolutions results with gaussians of different size as a whole, a multi-scale representation of the signal called the *scale space*. The original idea of Witkin was to detect global contours at large scales and to track them downward along the scale dimension in order to get a fine geometrical description. The scale-space theory then greatly developed to become a fully coherent theory of multi-scale low-level image processing.

Important milestones in this development include Koenderink's work (Koenderink, 1984), who formalized the Gaussian scale-space theory in $\mathbb{R}^n$, introducing a fundamental principle of multi-scale analysis, the *causality principle*, which formalizes a very commonsensical idea: details can only disappear by zooming out. To act as a multi-scale analysis, a sequence of descriptions—or models—of a datum must not create information when scale increases. In other words, the scale axis is fundamentally oriented. Koenderink also related Gaussian convolutions with the solutions of the heat equation and thus transposed the scale-space theory from the filtering domain to the domain of continuous evolution modelling by partial differential equations (PDE). Badaud et al. (1986) then showed that the gaussian scale-space is the unique linear scale-space. Subsequent scale-space models, such as Perona and Malik anisotropic diffusion scheme (Perona and Malik, 1990) or the affine morphological scale-space (Sapiro and Tannenbaum, 1993; Alvarez et al., 2001), involved non linear PDE. Alvarez et al. (2001) finally derived axiomatically a complete catalogue of all the possible PDE generated scale-spaces, the nature of the analysis depending on the set of the invariances obeyed. Another important family of scale-spaces are generated by morphological operators (Jackay and Deriche, 1987; Salembier and Serra, 1995; Park and Lee, 1996; Bosworth and Acton, 2003).

1.3. *. . . To Scale-Sets*

While low-level contour extraction and region delineation are very close problems, region-oriented image analysis, i.e. image segmentation, has followed a rather different development than edge detection. Following Leclerc, the image segmentation problem can be informally defined as the problem "to delineate regions that have, to a certain degree, coherent attributes in the image". Indeed, "it is an important problem as, on the whole, objects or coherent physical processes in the scene project into regions with coherent image attributes" (Leclerc, 1989).

Now in general, "coherent" regions can be found at different scales in an image: global structures which are perceived as "coherent" wholes at a large scale of analysis are made up of subparts which appear "coherent" at a lower scale of analysis, and which in turn decompose into coherent subparts... The existence of nested "coherent" structures in an image is obvious in the case of the texton/texture relationship; however it is much more general. For instance, a region which is "homogenous" to a certain degree, if it is not strictly constant, is then made up of more "homogenous" subparts. Hence, it can be argued that the general organization of the "coherent" regions of an image is a *hierarchical* organization, modelling the existence of nested meaningful structures in an image.

Despite this obvious remark, the image segmentation problem has been traditionally defined as a *partitioning* problem: decompose the image into *disjoint* regions (Haralick and Shapiro, 1991), that is produce a single scale, 'flat', analysis of the image. A reason for this might be that contrarily to edge detection, region extraction is not always clearly conceived as a low level process (Marr, 1982). Indeed, as semantic units of a scene get mapped into regional units in an image, there is a possible confusion between the low-level task of primitive structure extraction from images and the high-level task of object recognition and delineation. Furthermore, even a very classical segmentation algorithm, say a simple region growing method, is able to delineate structures which correspond to certain semantic units, *when its parameters are appropriately tuned*. Hence the temptation to use a low-level segmentation tool as an object recognition tool.

However, automatic parameter tuning is a hard problem which greatly affects the practical operationality of most segmentation algorithms. If the images a system has to analyze have a little variability, then it is dubious that a single parameter setting will allow to correctly segment all the images: the target objects will be over-segmented in some images, under-segmented in some other or even over-segmented in some parts of a given image and under-segmented in other parts of the same image. The reason for this is that the parameters of a segmentation algorithm are related to low-level criteria, such as homogeneity or geometrical regularity, which are not in general robust criteria in terms of object *discrimination*, except in very constrained imaging conditions (e.g. industrial inspection). Moreover, homogeneity or goodness-of-fit terms are dependent on the image dynamics; hence, if the image contrast changes, the parameters have to be adapted in order to keep a satisfactory result. This point is of course closely related to the fundamental role of invariance in vision and to the fact that vision algorithms should in general be independent of the viewing conditions (pose, distance, lighting...).

Like Morel and Solimini, we think that "the only basis parameter in computer vision is 'scale'" (Morel and Solimini, 1995), that is the fineness of analysis, and that in general, the useful scale of analysis for a given task cannot be chosen a priori, that is before viewing the image. The task of selecting the useful scale(s) for a given application is inherently a high level recognition task which cannot be robustly solved based on the low level criteria embedded in segmentation algorithms. Hence, a low level segmentation tool should remain scale uncommitted and output a multi-scale description of the image in which higher level processes can navigate freely in order to find their objects of interest. We claim that devising multi-scale low-level analyses is the only way to get robust computer vision systems, and in particular to solve the difficult problems of parameter tuning and of the respect of fundamental invariances.

Those different ideas have led us to develop a multiscale approach of low level region-oriented image analysis, which we have called the scale-sets theory because, as we shall see, it can be regarded as a region-oriented scale-space theory.

1.4. Organization of the Paper

The organization of the paper is as follows: In the spirit of the scale-space theory, the first part of the paper (Section 2) derives the general structure of geometrically unbiased multi-scale region-oriented image descriptions from a few basic axioms. It then introduces the scale-sets representation, an implicit representation which fully captures the structure of such descriptions. Its general properties are studied and our proposal is compared to related work.

The second part of the paper (Section 3) studies the way scale-sets image descriptions can be built according to an energy-minimization principle. In a much classical way now, the single-scale segmentation problem is put as the problem of finding a piece-wise defined model minimizing a two-term based energy, of the form $\lambda \mathsf{C} + \mathsf{D}$ where D is a goodness-of-fit term and C is a regularization term. We show that under broad assumptions, such energies induce monotone sequences of minimal cuts in a hierarchy when λ browses $\mathbb{R}^+$ and we design an efficient algorithm to compute the complete scale-sets representation of these sequences exactly. In a kind of feedback loop, this allows us to design a locally optimal algorithm to build the hierarchy from which the minimal cuts are extracted. The principle is to progressively regularize the energy by increasing λ and to follow the energy minimum along a monotone sequence of partitions. This strategy leads to a parameter-free region grouping criterion which we implement within a pair-wise region merging scheme. The scale-sets descriptions obtained are linearly scale invariant, which implies that the scale parameter λ is completely removed from the low-level segmentation stage. This property also allows to confer other invariance properties to the solution.

Before concluding, Section 4 presents some experimental results. Appendix A recalls some usual definitions on partitions and hierarchies and takes up the global notations of the paper. Appendices B and C provide the proofs applicable to Theorems 8 and 10.

2. Unbiased Multi-Scale Segmentations and their Scale-Sets Representation

2.1. Multi-Scale Segmentations

Let **P** be a partitioning algorithm depending on a positive real parameter λ. Given an image I defined over a domain $\mathcal{D}$ and a value of λ, **P** outputs a partition P_λ of $\mathcal{D}$. Intuitively, λ behaves as a scale parameter if increasing λ systematically coarsens the partition produced by **P**, in the sense of the fineness relationship between partitions (see Appendix A). Certainly because it is very intuitive, this definition has been adopted as a starting point for multi-scale segmentation by various authors, such as (Koepfler et al., 1994; Serra and Salembier, 1993). We show here that it can be derived from two fundamental principles of multi-scale analysis : causality and absence of boundary delocalization.

Let $P = (P_\lambda)_{\lambda\in\mathbb{R}^+}$ be the complete family of the partitions produced by **P** on I when λ browses $\mathbb{R}^+$. P can be thought of as a mapping from $\mathbb{R}^+$ to $\mathbb{P}(\mathcal{D})$ or equivalently as a point of $\mathbb{P}(\mathcal{D})^{\mathbb{R}^+}$. Assume that $\mathcal{D}$ is a part of a topological space and that the algorithm **P** only produces partitions into *connected regions*. A partition P_λ is then fully determined by the set of its *boundaries*, which we denote by δP_λ.

The **causality principle** is certainly the most fundamental principle of multi-scale analysis (Koenderink, 1984). From this principle, for any couple of scales $\lambda_2 > \lambda_1$, the "structures" found at scale λ_2 should find a "cause" at scale λ_1. Following Witkin's original idea (Witkin, 1983), we apply this principle to the *edges* produced by algorithm **P**. In this case, the parameter λ behaves as a *scale parameter* if and only if for all $\lambda_2 > \lambda_1$, the boundaries of partition P_{λ_2} are in a one-to-one mapping with a subset of the boundaries of P_{λ_1} (their "cause"). Now, to be topologically meaningful, the mapping which relates δP_{λ_2} to a subset of δP_{λ_1} must be continuous. We thus propose the following definition:

Definition 1 (causal structure). We say that a sequence $(P_\lambda)_{\lambda\in\mathbb{R}^+} \in \mathbb{P}(\mathcal{D})^{\mathbb{R}^+}$ has a **causal structure** if $\forall(\lambda_1, \lambda_2) \in \mathbb{R}^{+2}$ such that $\lambda_2 \geq \lambda_1$ there is a diffeomorphism ϕ of $\mathcal{D}$ such that $\phi(\delta P_{\lambda_2}) \subseteq \delta P_{\lambda_1}$.

In terms of partitions, P has a causal structure if for all $\lambda_2 > \lambda_1$, P_{λ_2} can be morphed to an under-partition of P_{λ_1}, that is can be obtained by first applying a continuous deformation to P_{λ_1} and then deleting some of its boundaries, that is merging some of its regions.

In general, the morphism ϕ is dependent on λ_1 and λ_2. We thus write it down $\phi(\lambda_1, \lambda_2)$. Now, in order to discard sudden displacements of boundaries when scale increases it is natural to require ϕ's continuity with respect to λ. Let Id denote the identical mapping on $\mathcal{D}$. As setting $\phi(\lambda, \lambda) = Id$ for all λ is the natural solution for $\phi(\lambda, \lambda)$, one is naturally led to the definition:

Definition 2 (continuous causal structure). We say that a sequence $(P_\lambda)_{\lambda\in\mathbb{R}^+} \in \mathbb{P}(\mathcal{D})^{\mathbb{R}^+}$ has a **continuous causal structure** if it has a causal structure and the family of diffeomorphisms ϕ verifies

$$\forall\lambda \in \mathbb{R}^+ \quad \lim_{\varepsilon\to 0} \phi(\lambda, \lambda + \varepsilon) = Id.$$

Finally, if one assumes that the best *localization* of a structure can be achieved at the finest scale of analysis, one should always keep the *geometrical information* gathered at the finest scale and thus discard any boundary deformation when scale increases. This can only be reached by setting $\phi(\lambda_1, \lambda_2) = Id$ for all scales. This additional condition then leads to a structure in which δP_{λ_2} is always a subset of δP_{λ_1}, and thus to a structure in which P_{λ_2} is always an under-partition of P_{λ_1}. Hence the definition:

Definition 3 (unbiased multi-scale segmentation). A sequence $(P_\lambda)_{\lambda\in\mathbb{R}^+} \in \mathbb{P}(\mathcal{D})^{\mathbb{R}^+}$ is called an **unbiased multi-scale segmentation** if and only if the application $\lambda \to P_\lambda$ is increasing, that is if and only if

$$\forall(\lambda_1, \lambda_2) \in \mathbb{R}^{+^2} \qquad \lambda_2 \geq \lambda_1 \Rightarrow P_{\lambda_2} \geq P_{\lambda_1}. \quad (1)$$

As a conclusion, if one considers an analysis into *connected regions*, then the intuitive structure of a multi-scale segmentation—a monotone mapping from $\mathbb{R}^+$ to $\mathbb{P}(\mathcal{D})$—follows from two basic principles applied to the *boundaries* of the partitions: (i) the causality principle, which is a *topological/structural* principle and (ii) a principle of *geometrical* accuracy. The scale parameter λ then controls a process which does topological/structural caricature of information but *no* geometrical caricature. These two conditions obviously meet the intuitive requirements for an ideal "multi-scale analyzer: an analyzer which delineates the *global* structures of an image together with a fine, or *local*, geometry. This property is called "strong causality" in Morel and Solimini (1995). Note that a partitioning algorithm **P** whose results on an image verify the relation 1 can also be regarded as a geometrically unbiased multi-scale edge detector which produces closed contours. We shall get back to the issue of region/contour duality in Section 2.3 below.

2.2. *The Scale-Sets Representation*

As was argued in introduction, a low-level segmentation algorithm should not focus on a specific scale but should output a multi-scale description of an image; and as we just saw, in a region-oriented approach, this description should be what we have called an unbiased multi-scale segmentation (Definition 3). However, this poses a representational problem: an unbiased multi-scale segmentation P is an increasing mapping from $\mathbb{R}^+$ to $\mathbb{P}(\mathcal{D})$, and as this mapping starts from the *continuous* real line, it cannot be directly represented on a computer.

Of course, one can always approximate the multi-scale structure by *sampling* the scale axis, i.e. by setting a series $(\lambda_1, \lambda_2 \ldots \lambda_k)$ of increasing scales and collecting the associated partitions, ending with a pyramid of segmentations $(P_{\lambda_1} \leq \cdots \leq P_{\lambda_k})$. However, the sampling points are necessarily arbitrary: they must be chosen *before* seeing the image. Such a strategy is thus *scale-committed*. Instead, we would like to get a complete representation of P which would allow to browse the multi-scale structure of an image freely, e.g. which would allow to set any value of λ *a posteriori*, and retrieve the corresponding partition P_λ. This can be achieved by what we call the scale-sets representation of P, as we explain now.

Rather than focusing on the partitions P_λ, consider the set of all the *regions* which compose them. Formally, consider $H = \bigcup_{\lambda\in\mathbb{R}^+} P_\lambda$. Obviously, as P is a sequence of monotone partitions, the regions which belong to H are either disjoint or nested. Thus, if P_0 is the absolute over-partition and if for a sufficiently large scale L, P_L is the absolute under-partition $\{\mathcal{D}\}$, then H is a hierarchy on $\mathcal{D}$ (see Appendix A). For a digital image defined on a domain $\mathcal{D}$ of N pixels, H is finite and contains at most $2N - 1$ elements, a bound which is reached when H is a binary hierarchy.

Also consider for each region x of H the set $\Lambda(x)$ of the scales for which x belongs to P_λ:

$$\Lambda(x) \triangleq \{\lambda \mid x \in P_\lambda\}.$$

One can easily verify that the relation 1 implies that $\Lambda(x)$ is an *interval*, which can be written

$$\Lambda(x) = [\lambda^+(x), \lambda^-(x)[.$$

Definition 4. $\Lambda(x)$ is called the **interval of persistence** of the set x. $\lambda^+(x)$ is its **scale of appearance** in P and $\lambda^-(x)$ is its **scale of disappearance** in the sequence P.

$\lambda^+(x)$ and $\lambda^-(x)$ are the equivalent of the so-called inner and outer scales of x in scale-space theory (Lindeberg, 1994). One then immediately verifies that a set $x \in H$ disappears when its father in H appears, that is

$$\forall x \in H \qquad \lambda^{-}(x) = \lambda^{+}(\mathcal{F}(x)). \tag{2}$$

Hence, the couple $S = (H, \lambda^{+})$, where H provides the sets which compose the partitions of the sequence P and λ^{+} gives access to the range of scales over which they "live", captures the complete structure of P. We thus propose the following definition :

Definition 5 (scale-sets representation). If $P = (P_\lambda)_{\lambda\in\mathbb{R}^+}$ is an unbiased multi-scale segmentation then the structure $\mathcal{S}(P) \triangleq (H, \lambda^{+})$ where

$$H \triangleq \bigcup_{\lambda\in\mathbb{R}^+} P_\lambda$$

$$\lambda^{+} \triangleq \begin{cases} H \mapsto \mathbb{R}^+ \\ x \to \min\{\lambda \in \mathbb{R}^+ \mid x \in P_\lambda\} \end{cases}$$

is called the **scale-sets representation** of P.

From Eq. (2), λ^{+} is an *increasing function* with respect to the set-inclusion partial order:

$$x \subset y \Rightarrow \lambda^{+}(x) < \lambda^{+}(y).$$

In the domain of hierarchical classification, such a function is called an *index* on H and the couple (H, λ^{+}) is called an *indexed hierarchy* or a *stratified hierarchy*. In our context, H represents a family of regions of the image's domain and λ^{+} maps them onto a 1D "scale" axis, hence the scale-sets terminology. The whole structure completely represents a volume of partitions, yet in an implicit manner. Any partition P_λ can then be immediately retrieved by *sectioning* the scale-sets structure:

Definition 6 (sections of a scale-sets). Let $\mathcal{S} = (H, \lambda^{+})$ be a scale-sets on $\mathcal{D}$. Let B_λ be the family of boolean predicates on H:

$$B_\lambda(x) \text{ is true} \Leftrightarrow \lambda^{+}(x) \leq \lambda$$

then the coarsest cut of H whose elements verify B_λ is called $\mathcal{S}$' **section** of scale λ, or $\mathcal{S}$' **λ-section**. We denote it by $S_\lambda(\mathcal{S})$.

$S_\lambda(\mathcal{S})$ is the partition made up of the largest regions of H which have a scale of appearance lower than λ. Obviously, thanks to the increasingness of λ^{+}, this partition exists and is unique for all $\lambda \geq 0$. Furthermore, if P is an unbiased multi-scale segmentation, then

$$\forall\lambda \in \mathbb{R}^+ \qquad P_\lambda = S_\lambda(\mathcal{S}(P)).$$

Hence the family $(P_\lambda)_{\lambda\in\mathbb{R}^+}$ is the family of the sections of $\mathcal{S}(P)$.

In practice, a scale-sets $\mathcal{S}$ can be stored in a computer as a real-valued tree. It can be graphically represented by a dendrogram in which the y axis represents the scale axis. Each *horizontal cut* in this diagram then corresponds to a section of $\mathcal{S}$ (see Fig. 1).

2.3. Discussion

2.3.1. Related Work. As regards the image segmentation problem, the idea of obtaining partitions as cuts in hierarchies of regions goes back to the famous split and merge algorithm by Horowitz and Pavlidis (1976). However, the authors considered regular hierarchies (quad-trees) and only looked for a single cut, using a homogeneity predicate-based formulation of the partitioning problem.

Later, different authors have proposed to return a stack of monotone partitions - fine to coarse - as a segmentation result, also called a *pyramid of segmentations*, either based on a structural/graph-based approach (Montanvert et al., 1991; Jolion and Montanvert, 1992), on an energy minimization-based approach (Koepfler et al., 1994; Ballester et al., 1994; Fuchs and Le Men, 1999) or on a morphological approach (Serra and Salembier, 1993; Salembier and Serra, 1995)

Other authors consider returning a tree of regions, or hierarchy, built by a classical region merging algorithm (Salembier and Garrido, 2000a) or by a recursive divisive approach based on binary Markov Random Fields (Poggi and Ragozini, 1999).

In all these works, different levels of details are proposed; however either the levels are not related to the values of a scale parameter, or the scale axis is sampled.

As we have seen, in order to obtain a complete representation of a sequence of monotone partitions with respect to a real parameter, one needs to consider an *implicit* representation of the partitions, as the family of the sections of a stratified hierarchy. A monotone mapping $\lambda \to P_\lambda$ is fully characterized by a set of critical events: namely the appearance of new regions at some specific scales, which are the unions of some regions existing at lower scales. By nature, these

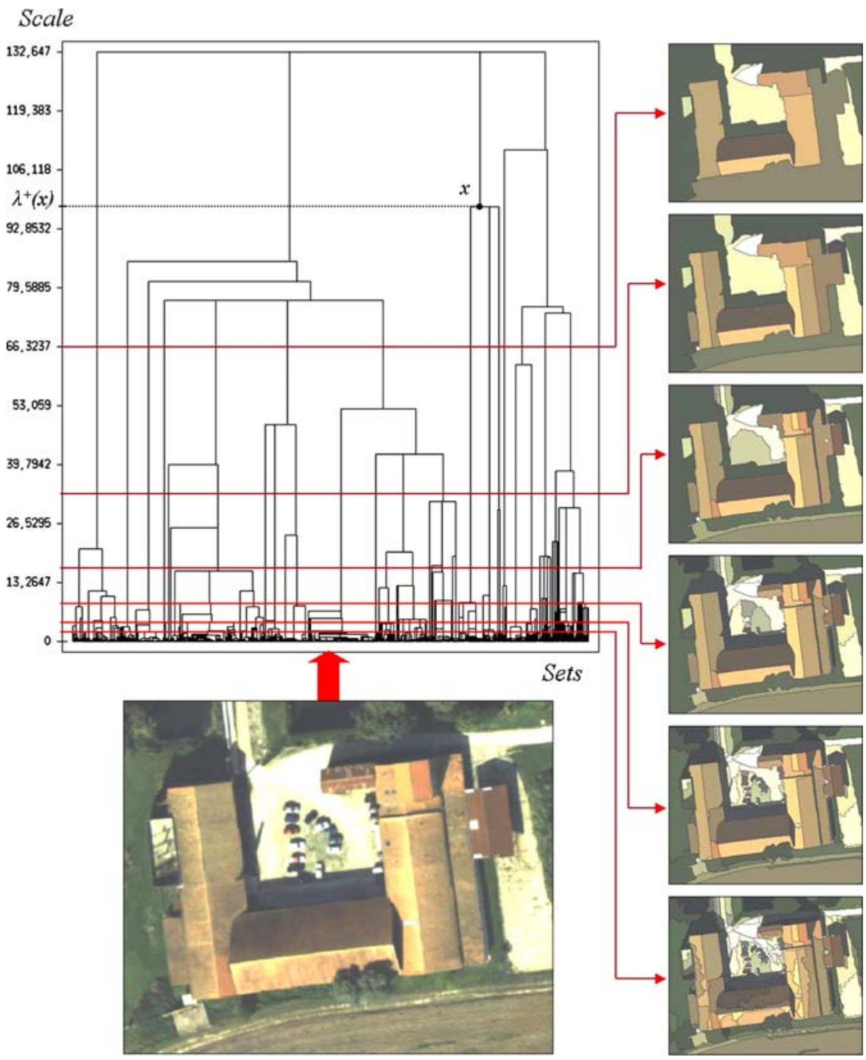


Figure 1. A scale-sets image description and some of its sections. This scale-sets was obtained by the scale climbing algorithm from a watershed over-segmentation of the image (see below).

merging events are discrete and in the case of an image defined on a discrete domain, they are in finite number. A complete representation can thus be obtained by making *explicit* these critical events and the scales at which they occur. It corresponds to a reverse point of view from the pyramidal approach which gives an approximate description of the mapping $\lambda \rightarrow P_\lambda$. The pyramidal approach amounts to setting some specific scales and asking: which sets are present at that scales? From the scale-sets point of view, the key-point is: at which scale does a specific set appear [1]? Answering this question will be a key-point in the energy minimization-based approach to build scale-sets descriptions presented below.

2.3.2. Ultrametric Distances and the Region/Contour Duality. It is well known that the datum of a stratified hierarchy on $\mathcal{D}$ is equivalent to that of an **ultrametric distance** δ on $\mathcal{D}$, that is of a distance which verifies

$$\forall (x, y, z) \in \mathcal{D}^3 \qquad \delta(x, z) \leq \max\{\delta(x, y), \delta(y, z)\}.$$

The closed balls of an ultrametric space $(\mathcal{D}, \delta)$ constitute a hierarchy H and the diameter of the balls is an index on H. Conversely, a couple (H, λ^+) induces an ultrametrics $\lambda^-_\star$ on $\mathcal{D}$: $\lambda^-_\star(x, y)$ is the smallest scale for which x and y are grouped, i.e. belong to a same

Figure 2. The contour disappearance map associated with the scale-sets of Fig. 1 (logarithmic scale).

region. Formally

$$\lambda_\star^-(x, y) \triangleq \lambda^+(\{x\} \vee \{y\})$$

where $\vee$ denotes the supremum operator in the sup-semi-lattice $(H, \subseteq)$. The elements of the section of scale λ of (H, λ^+) are the maximal balls of $(\mathcal{D}, \lambda_\star^-)$ which have a diameter lower than λ.

Now, if $\mathcal{D}$ is a domain of a topological space and if the regions of H are *connected* then the ultrametrics is fully determined by its values for the couples of *neighbouring* points in $\mathcal{D}$. $\lambda_\star^-(x, y)$ then represents the scale at which the *boundary* between x and y *disappears*. As a consequence, an unbiased multi-scale segmentation *into connected regions* can be equivalently represented by a **contour disappearance map** (see Fig. 2). This contour map has this particular property that thresholding it always leads to closed contours, hence to a partition of the image's domain.

For practical uses, this contour-oriented representation can be further simplified. Let $G = (\mathcal{D}, \mathcal{N})$ denote the region adjacency graph of P_0, i.e. of the base of H. Value its edges by $\lambda_\star^-$ and consider a *minimum spanning tree (MST)* T of $(G, \lambda_\star^-)$. Removing the k edges of T which have a lower scale than λ then splits the tree into $k + 1$ connected components which correspond to the regions of the section of scale λ of (H, λ^+).[2] The scale-sets representation, by an indexed inclusion tree, and the indexed spanning tree representation are two dual representations of an unbiased multi-scale segmentation, in the sense of the duality between connected region-based and closed contour-based descriptions.

Fernand Meyer's morphological multi-scale segmentation approach is based on the spanning tree representation (Meyer, 1999a,b, 2001). Meyer proposes to start from a watershed transform of the magnitude of a gradient of the image, build the region adjacency graph of the catchment basins and value its edges by a measure of dissimilarity. A multi-scale segmentation is then obtained by considering the MST of this graph. Meyer proposes different measures of dissimilarity which amount to simulating different synchronous flooding processes of the gradient magnitude. The scale parameter only depends on the contrast, the surface or the volume (contrast$\times$surface) of the regions.

In Guigues et al. (2003) we proposed another way to obtain multi-scale segmentations from a dissimilarity-based grouping approach. We defined a *cocoon* of a valued graph as a connected set of nodes whose maximal internal dissimilarity is lower than the minimal dissimilarity with the exterior nodes. We proved that the set of the cocoons of a graph is a hierarchy and released an associated ultrametrics. The hierarchy of the cocoons of a graph is related to complete-linkage clustering while the MST of a graph is related to single-linkage clustering, see (Lance and Williams, 1967; Guigues et al., 2003).

However, dissimilarity-based approaches do not allow to introduce *geometrical criteria* in the segmentation process. In contrast, the multi-scale analyses which we propose here are based on optimizing two-term-based energies, one term being a geometrical regularization term. We now turn to the energy-minimization part of our approach.

3. Scale-Sets Generation from a Variational Principle

The first part of the paper has discussed the general structure of unbiased multi-scale segmentations and the way they can be represented exactly, considering a scale-sets representation. We shall now show that one can build scale-sets whose sections approximate the solutions of a general energy minimization problem which embeds most optimization-based formulations of the segmentation problem (variational, Markovian, minimal encoding). The energies involved have two global terms whose relative weight is controlled by a real parameter which, as we shall see, generally behaves as a scale parameter.

3.1. Scale as an Inherent Parameter of Approximate Modelling Problems

In a very classical way now, we put the single-scale segmentation problem as an optimal piece-wise image modelling problem (Mumford and Shah, 1989; Geman and Geman, 1984). Let $\mathcal{I}$ be the set of the images one is interested in and $\mathcal{M}$ be a set of possible "models" of these images. In deterministic approaches of segmentation, a model is a piece-wise defined image: piece-wise constant or smooth (Mumford and Shah, 1989), polynomial (Leclerc, 1989; Kanungo et al., 1995)... In probabilistic approaches, a model is a piece-wise defined stochastic process whose realizations are images: piece-wise i.i.d. Gaussian process... In what follows, we focus on deterministic models; however a strictly parallel reasoning can be followed for probabilistic models.

Given an image $I \in \mathcal{I}$, the objective is to pick in $\mathcal{M}$ the "best model" of I. By a "comparison principle" (Koepfler et al., 1994), this can always be formalized as the problem of finding $M \in \mathcal{M}$ which minimizes a certain energy $\mathsf{E} : \mathcal{I} \times \mathcal{M} \mapsto \mathbb{R}^+$. Such an approach contains three aspects: the definition of $\mathcal{M}$, the definition of E, and the optimization itself.

In a deterministic framework, a model of an image is also an image. Hence, one can at once choose a distance D between models and images and define the "best model" M of I as the one which minimizes $\mathsf{D}(I, M)$. However, the solution for this type of problem is in general not satisfactory: most of the time the model set contains the image set, so that there is always a trivial solution $M = I$ which achieves $D(I, M) = 0$. For instance, if one looks for a piece-wise constant model of an image, then there is always an exact solution which consists in tessellating the image into as many regions as pixels. In practice, one looks for regions which correspond to coherent phenomena which are larger than a pixel. Hence, one does not look for the absolutely flat zones of an image but looks for regions over which the image is *approximately* constant. A *precision of approximation* then needs to be introduced.

If I is an image and ε is a positive number, let us say that a model $M \in \mathcal{M}$ such that $\mathsf{D}(I, M) \leq \varepsilon$ is an ε-approximation of I. For a given ε, there are a priori numerous ε-approximations of I. The way to choose one of them must then be principled, a task which can again be formalized as an energy-minimization problem: assume that one is capable of defining an energy $\mathsf{C} : \mathcal{M} \mapsto \mathbb{R}^+$ which fully captures his preferences *a priori* for the different models; the problem of the ε-approximation of an image I can then be expressed as a constrained optimization problem which we denote by $\mathcal{P}_\varepsilon(I)$:

$$\begin{aligned} &\text{Find } M \in \mathcal{M} \text{ which minimizes } \mathsf{C}(M) \\ &\text{subject to } \mathsf{D}(I, M) \leq \varepsilon. \end{aligned} \tag{3}$$

The energy C can be of different kinds. It can translate objective knowledge on the data or be based on relatively subjective criteria and aim to favour a specific interpretation. However, this constrained problem takes a particular meaning when C can be interpreted as a measure of the *complexity* of a model (objective or subjective). In this case, the problem $\mathcal{P}_\varepsilon(I)$ formalizes the idea that *between two models that equally fit the data, the simpler of the two should be preferred*. It thus can be viewed as an expression of Occam's Razor principle. Now, Occam's Razor principle has a dual formulation : *between two equally complex models, the closer to the data of the two should be preferred*, which corresponds to a dual minimization problem $\mathcal{P}^\star_\gamma(I)$:

$$\begin{aligned} &\text{Find } M \in \mathcal{M} \text{ which minimizes } \mathsf{D}(I, M) \\ &\text{subject to } \mathsf{C}(M) \leq \gamma. \end{aligned} \tag{4}$$

These two problems are well known in information theory where they formalize lossy compression issues (Shannon, 1959). In this context, the energy C represents the communication rate or the number of bits needed to encode the model and D measures the distortion of the data caused by the compression-decompression process. Given a certain quota of losses, the problem is to obtain the best compression ratio or dually, given a certain compression ratio, the problem is to minimize the losses. Classically, the minimization of a constrained problem is obtained by minimizing its associated Lagrangian. The Lagrangians associated to $\mathcal{P}_\varepsilon(I)$ and $\mathcal{P}^\star_\gamma(I)$ respectively read:

$$\mathsf{E}_\mu(I, M) = \mathsf{C}(M) + \mu \mathsf{D}(I, M) \tag{5}$$

$$\mathsf{E}^\star_\lambda(I, M) = \lambda \mathsf{C}(M) + \mathsf{D}(I, M) \tag{6}$$

where μ and λ are the Lagrange multipliers.

Under large assumptions, one proves that there is a bijection between the solutions of $\mathcal{P}_\varepsilon(I)$ (resp. $\mathcal{P}^\star_\gamma(I)$) and the minima of $\mathsf{E}_\mu(I, M)$ (resp. $\mathsf{E}^\star_\lambda(I, M)$). Every (single) value of ε (resp. γ) corresponds to a particular value of μ (resp. λ) for which the minimum of the Lagrangian is identical to the solution of the problem under constraint.

With the lagrangian expressions, one finds again the usual energies used in image analysis. $\mathsf{E}^{\star}_{\lambda}(I, M)$ corresponds for instance to the general expression of a regularized ill-posed variational problem. D is then usually called a "goodness-of-fit" term and C a "regularization" term. For segmentation purposes, the archetype of this kind of formulation is given by Mumford and Shah's functional (Mumford and Shah, 1989). Within the framework of Bayesian inference, $\mathsf{E}^{\star}_{\lambda}(I, M))$ can also be interpreted as the Gibbs potential associated to a conditional probability of the model given the data, or posterior probability (Geman and Geman, 1984). In this case, $\exp(-\mathsf{D}(I, M))$ corresponds to the probability of the data knowing the model, or likelihood, and $\exp(-\lambda\mathsf{C}(M))$ corresponds to the prior probability of the model. Minimizing $\mathsf{E}^{\star}_{\lambda}(I, M)$ is equivalent to maximizing $\exp(-\mathsf{E}^{\star}_{\lambda}(I, M))$ which is proportional to the posterior probability of M.

Now, viewing an energy $\mathsf{E}_{\mu}(I, M)$ as the lagrangian of a constrained minimization problem allows to interpret its minimization as the resolution of a problem of optimal ε-approximation of an image and thus puts in light the role of the parameter μ. μ controls the precision ε of the approximation found and thus can be thought of as a *scale parameter* along the image's values dimension: the distance between the optimal model and the image monotonously decreases as μ increases. Besides, the minima of $\mathsf{E}_{\mu}(I, M)$ and $\mathsf{E}^{\star}_{\lambda}(I, M)$ are identical when $\lambda = 1/\mu$. Consequently, the families of solutions for the two dual constrained optimization problems coincide. Every precision ε corresponds to a bound of "complexity" $\gamma(\varepsilon)$. Looking for an ε-approximation of minimal complexity is equivalent to looking for the closest approximation whose complexity is lower than $\gamma(\varepsilon)$. The precision and the complexity of an optimal model thus vary altogether and the Lagrangian parameter is the bridge between them. As we will see, when the complexity is a notion of spatial complexity and the problem is put in a hierarchical framework then the Lagrangian parameter λ is a scale parameter *both* in the spatial sense and in the values sense.

3.2. *Statement of the Problem and the Strategy Adopted*

We have described above two different notions of scale associated with piece-wise image description problems: from a structural/ensemblistic point of view, a "scale parameter" allows to browse a sequence of nested spatial partitions and thus controls the "structural" fineness of a description. From an optimal modelling point of view, a "scale parameter" controls both the approximation's fidelity and the approximation's complexity. One would naturally like to see both notions of scale coincide, i.e. to see the modelling/optimization multi-scale framework match the structural/ensemblistic multi-scale framework.

However, in general, the partitions minimizing an energy of the form 6 are not monotone, mainly because boundary displacements occur when λ increases. An example involving Mumford-Shah's functional for piecewise constant image approximation is provided in Guigues (2003). Furthermore, it is well known that even the simplest energy-minimization problems on discrete partitions are NP-hard (see e.g. (Boykov et al., 1999) for a proof of it involving Potts energy).

As the minima of usual two term-based energies on partitions are not nested, we propose to enforce this structure and put the multi-scale segmentation problem as the one to **find a scale-sets description of an image whose sections are as close as possible to the partitions minimizing a two term-based energy of the form 6**. As we shall see, imposing a hierarchical structure to the solution will lead us to a fast approximation scheme of the solutions.

The way we address this problem is driven by the structure of the scale-sets representation. Two distinct components are needed to fully determine a multi-scale segmentation: a hierarchy of regions H and a mapping λ^+ of these regions onto a one-dimensional axis, the 'scale' axis. H provides the sets but only induces a *partial order* on them through the inclusion relation. λ^+ then defines a *total order* on the structure, which implicitly defines a family of monotone cuts of H. We shall study the construction of these two components independently and in *reverse order*:

(1) In Section 3.3 we assume that we already know the structural part of the solution, i.e. the hierarchy H, and we prove that for an important class of two term-based energies, the family of the minimal cuts of H *is* a multi-scale analysis. Hence, for this class of energies, knowing H fully determines λ^+. Furthermore, we show that λ^+ can be efficiently computed by a dynamic programming-based algorithm.

(2) In Section 3.4 we get back to the construction of the hierarchy itself. The results of Section 3.3 lead us to a natural principle to build a scale-sets whose sec-

tions track the energy's minima, which is called the **scale climbing** principle. The method is parameter-free and leads to scale-sets which verify important properties.

3.3. *Multi-Scale Minimal Cuts of a Hierarchy*

This section develops the first part of the strategy described above. From there on, we shall restrict the type of energies considered to what we call **affine separable energies (ASE)**, which are the energies which can be written for a partition P

$$\mathsf{E}_\lambda(P) = \sum_{R\in P} \lambda \mathsf{C}(R) + \mathsf{D}(R) \tag{7}$$

Most energies usually considered for image segmentation can be separated. Examples of it are the Mumford-Shah functional (Mumford and Shah, 1989), Potts' priors for Markov Random Fields (Geman and Geman, 1984), and most description length criteria such as the one proposed by Leclerc (1989). From a probabilistic point of view, the separability of E_λ amounts to an independence assumption between the processes that generated the different regions of an image.

The datum of an ASE E_λ on the partitions of a domain $\mathcal{D}$ is equivalent to the datum of a couple (C, D) of energies on the *parts* of $\mathcal{D}$. We can thus also write $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$.

3.3.1. Hierarchical Minimization of Separable Energies. Forget for a while the scale parameter and simply consider a separable energy $\mathsf{E}(P) = \sum_{R\in P} \mathsf{E}(R)$. If H is a hierarchy then H's cut which minimizes E can be easily computed by dynamic programming (DP):

$\forall x \in H$, let $C^*(H(x))$ be the minimal cut of H's partial hierarchy rooted at x. Let $\mathsf{E}^*(H(x))$ be the energy of this cut. As the union of two cuts of two disjoint hierarchies is a cut of their union, and as the energy E is separable, then $\forall x \in H$ the following Bellman's dynamic programming equations hold:

$$\mathsf{E}^*(H(x)) = \min\left\{\mathsf{E}(x), \sum_{s\in\mathcal{S}(x)} \mathsf{E}^*(H(s))\right\} \tag{8}$$

$$C^*(H(x)) = \begin{cases} \{x\} & \text{if } \mathsf{E}(x) \leq \sum_{s\in\mathcal{S}(x)} \mathsf{E}^*(H(s)) \\ \bigcup_{s\in\mathcal{S}(x)} C^*(H(s)) & \text{otherwise} \end{cases} \tag{9}$$

One can thus optimize E starting from the leaves of H by applying Eqs. (8) and (9) successively to all the nodes of H in hierarchical order. $C^*(H)$ and $E^*(H)$ then provide the minimal cut of H and its energy. The overall procedure has a linear complexity with respect to H's size.

This procedure has been used for classification purposes in the CART algorithm (Breiman et al., 1984) and for wavelet bases construction in the Best-basis algorithm (Donoho, 1997). In these applications, the hierarchies involved are quad-trees. Salembier and Garrido also used this minimal cut algorithm to build optimal image thumbnails in a rate/distorsion sense (Salembier and Garrido, 2000a). The cut is extracted from a hierarchy of regions obtained by classical region-merging technique which has no particular affinity with the minimization problem. We also employed this method in an image segmentation algorithm based on minimizing a Minimum Description Length (MDL) criterion in a hierarchy of cocoons (Guigues et al., 2003, 2001).

The dynamic programming procedure allows to extract a single optimal partition from a hierarchy. Given an affine separable energy E_λ, we shall now study the behavior of H's minimal cut with respect to λ.

3.3.2. Multi-Scale Minimal Cuts Let H be a hierarchy and E_λ be an ASE. For each $\lambda \in \mathbb{R}^+$, we call λ*-cut* the cut of H which minimizes E_λ. We denote this cut by $C^*_\lambda(H)$. A region $x \in H$ which belongs to $C^*_\lambda(H)$ is called λ-optimal.

Examining the structure of the dynamic programming Eq. (8), one can easily verify that

Proposition 7. *A set $x \in H$ is λ-optimal if and only if the following two properties hold:*

(i) *x is* partially optimal, *i.e.* $\forall Y \in \mathbb{C}(H(x))\, \mathsf{E}_\lambda(\{x\}) \leq \mathsf{E}_\lambda(Y)$.
(ii) *x is* maximal *in H for the property (i), that is no $y \in H$ such that $x \subset y$ is also partially optimal.*

In what follows $P^*_\lambda(H)$ will denote the set of all *partially* λ-optimal nodes of H, that is the set of the nodes which verify the property (i) of Proposition 7. Our approach then rests on the following Theorem:

Theorem 8 (*Multi-scale minimal cuts*). *Let $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$ be an affine separable energy. If either term C or term D is monotone with respect to the fineness partial order relation on the partitions of $\mathcal{D}$, then for any hier-*

archy H on $\mathcal{D}$, the sequence $(C^*_\lambda(H))_{\lambda\in\mathbb{R}}$ of the minimal cuts of H for E_λ is a multi-scale segmentation.

More precisely, if C is decreasing in $\mathbb{P}(\mathcal{D})$, that is if

$$\forall (P,Q)\in\mathbb{P}^2(\mathcal{D}) \quad Q\geq P \Rightarrow \mathsf{C}(Q)\leq \mathsf{C}(P) \quad (10)$$

or if D is increasing in $\mathbb{P}(\mathcal{D})$ then $(C^*_\lambda(H))_{\lambda\in\mathbb{R}}$ is an increasing sequence of partitions. If C is increasing or D is decreasing then $(C^*_\lambda(H))_{\lambda\in\mathbb{R}}$ is a decreasing sequence.

The proof is provided in Appendix B.

From there on, we shall focus on the energies which lead to increasing sequences of solutions, fine to coarse when λ increases, hence to ASES built either on a decreasing regularizer C or on an increasing goodness-of-fit term D. Such energies shall be called **multi-scale ASES (MASEs)**.

Note that the decreasingness of a separable energy in the partition lattice is equivalent to the *subadditivity* of the corresponding energy on the set lattice. More precisely, one easily shows that a separable energy $\mathsf{C}(P)=\sum_{R\in P}\mathsf{C}(R)$ verifies 10 if and only if

$$\begin{aligned}&\forall (R,S)\in\mathcal{P}^2(\mathcal{D})\\ &R\cap S=\emptyset \quad \Rightarrow \mathsf{C}(R\cup S)\leq \mathsf{C}(R)+\mathsf{C}(S). \quad (11)\end{aligned}$$

Usual regularizing energies are decreasing. Quantifying the complexity of a segmentation by its number of regions gives a decreasing regularizer. Also, summing up any positive quantity (length, curvature, ...) along the boundaries of the partition leads to a decreasing regularizer: deleting a boundary by merging some adjacent regions systematically reduces the energy. Indeed, the decreasingness condition matches Occam's idea that entities should not be multiplied without necessity. Starting from this idea, the regions of a solution should only be multiplied if it increases the model's goodness-of-fit. Hence among two ordered solutions one should a priori prefer the coarser of the two.

On the contrary, usual goodness-of-fit energies are increasing. Assume for example that one models the image within a region by a certain parametric model (e.g. constant, polynomial) and measures the model-image distance in the L_p norm. Indeed, the overall energy (the sum of the residuals of an L_p regression) is always larger when one fits a single model to the union of two regions rather than two separate models to each region. In other words, in the context of fixed order parametric modelling in the L_p norm, increasing the number of model pieces always improves the model's fidelity.

As a conclusion, the energies which naturally arise in image segmentation lead to multi-scale solutions in a hierarchical framework. We shall now proceed with the consequences of the Theorem 8.

3.3.3. Appearance, Disappearance and Persistence of Regions. Indeed, if $(C^*_\lambda(H))_{\lambda\in\mathbb{R}}$ is a multi-scale segmentation, it can be represented by a scale-sets structure. Moreover, this scale-sets can be computed exactly and efficiently by generalizing the above dynamic programming procedure. We shall first explain the relationship between partial optimality and maximality in H and the scales of appearance and disappearance of H's sets.

For each region x of the hierarchy, let

$$\Lambda^*(x)\triangleq\{\lambda\,|\,x\in C^*_\lambda(H)\}$$

be the set of the scales for which x belongs to the λ-minimal cut of H. Also define

$$\Lambda^*_\uparrow(x)\triangleq\{\lambda\,|\,x\in P^*_\lambda(H)\}$$

which is the set of the scales for which x is partially λ-optimal.

When E_λ is a MASE, if a given region x is partially optimal for a scale λ then so is it for any larger scale $\lambda'>\lambda$. Thus, $\forall x\in H$, the set $\Lambda^*_\uparrow(x)$ is an interval of the form $[a,+\infty[$. Now, from the maximality condition (ii) in Proposition 7, x is globally optimal at scale λ if and only if no upper node on its branch to the top of the hierarchy is also partially λ-optimal. We thus define

$$\Lambda^*_\downarrow(x)\triangleq\bigcup_{y\in H,\,x\subset y}\Lambda^*_\uparrow(y)$$

which represents the set of the scales for which x *cannot* be maximal and thus cannot be globally optimal. As $\Lambda^*_\downarrow(x)$ is a union of intervals of the form $[a,+\infty[$, it is also an interval of this form.

Finally, the two conditions of global optimality of a node provided by Proposition 7 can be summarized by

$$\Lambda^*(x)=\Lambda^*_\uparrow(x)\backslash\Lambda^*_\downarrow(x).$$

We thus conclude that

Proposition 9. *If H is a hierarchy and E_λ is a MASE then $\forall x \in H$ the set $\Lambda^*(x)$ is an interval of the form $\Lambda^*(x) = [\lambda^+(x), \lambda^-(x)[$, where*

$$\lambda^+(x) = \inf \Lambda^*_\uparrow(x) \tag{12}$$

$$\lambda^-(x) = \inf \Lambda^*_\downarrow(x) = \min_{y \in H,\ x \subset y} \lambda^+(y) \tag{13}$$

The rightmost part of Eq. (13) shows that, knowing $\lambda^+(x)$ for each element x of H, λ^- can be easily computed by a top-down traversal of H. We shall now explain how λ^+ can be efficiently computed by generalizing the dynamic programming method described above.

3.3.4. Functional Dynamic Programming and Persistent Hierarchies. In the single-scale DP method, each node $x \in H$ is attributed by a single energy value. We now consider attributing each node $x \in H$ by a *function* of λ which represents its energetic behaviour with respect to scale.

$\forall x \in H$, $\mathsf{E}_\lambda(x)$ represents the energy of x at scale λ. Now rewrite it as a function of λ:

$$\mathsf{E}_x : \lambda \to \mathsf{E}_x(\lambda) = \lambda \mathsf{C}(x) + \mathsf{D}(x).$$

and call it the **self-energy** of x. E_x is an *affine* function of λ which has a positive slope and y-intercept.

For each partial hierarchy $H(x)$, also rewrite the energy of its minimal cuts as a function of λ:

$$\mathsf{E}^*_x : \lambda \to \mathsf{E}^*_x(\lambda) = \mathsf{E}^*_\lambda(H(x)).$$

and call it the **partial energy** of x.

Clearly, for any base node of H (whose partial hierarchy reduces to a single set) these two energy functions coincide, that is $\forall b \in \underline{H}\ :\ \mathsf{E}^*_b = \mathsf{E}_b$.

Now, as addition and infimum operations on functions are constructed from the point-wise operations, for any node $x \in H$, the dynamic programming Eq. (8) can be rewritten as a functional equation:

$$\mathsf{E}^*_x = \inf \left\{ \mathsf{E}_x, \sum_{s \in \mathcal{S}(x)} \mathsf{E}^*_s \right\} \tag{14}$$

We then have the

Proposition 10 (Functional dynamic programming). *Let H be a hierarchy and $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$ a multi-scale affine separable energy, then $\forall x \in H$*

(i) E^*_x *is a piece-wise affine, non decreasing, continuous and concave function.*

(ii) $\forall \lambda \in \mathbb{R}^+$

$$\mathsf{E}^*_x(\lambda) = \begin{cases} \sum_{s \in \mathcal{S}(x)} \mathsf{E}^*_s(\lambda) & \text{if} \quad \lambda < \lambda^+(x) \\ \mathsf{E}_x(\lambda) & \text{otherwise} \end{cases} \tag{15}$$

(iii) *If* C *is strictly decreasing or* D *is strictly increasing then $\lambda^+(x)$ is real and is the unique solution of*

$$\mathsf{E}_x(\lambda) = \sum_{s \in \mathcal{S}(x)} \mathsf{E}^*_s(\lambda).$$

The proof is provided in Appendix C.

The function E^*_x which provides the energies of the optimal cuts of $H(x)$ with respect to λ is thus a rather simple piece-wise affine function (see Fig. 3(b)). It can be explicitly stored for each node of H, typically by the list of the endpoints of its affine pieces. Then, for

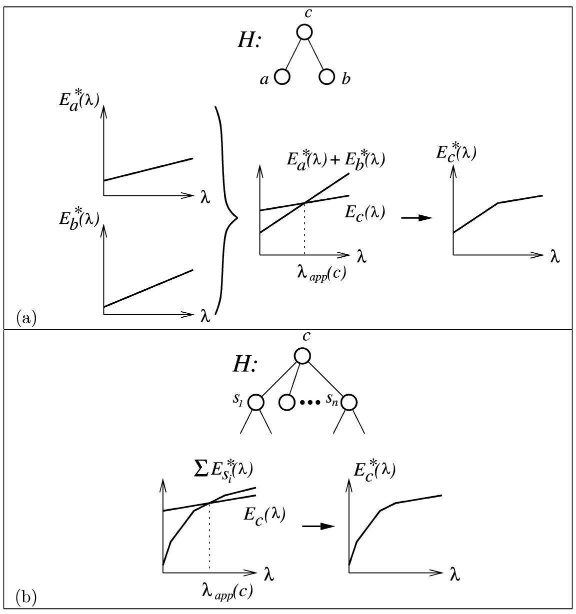


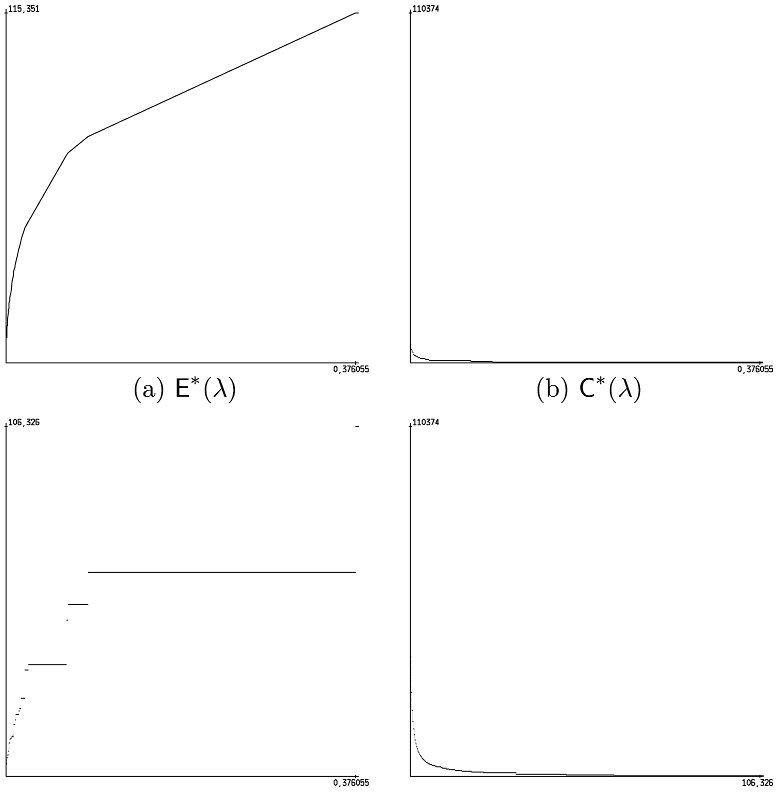
Figure 3. Computation of the partial energy function E^*_c and of the scale of appearance $\lambda^+(c)$ of a node c knowing E^* for its sons. (a) Case of a first level node having two sons. (b) General case.

each node x, computing E^*_x only requires two operations on piece-wise affine functions, a sum and an intersection with an affine function, which can be both computed exactly (see Fig. 3). If the driving energy (C or D) is strictly monotone, then by the proposition 10 (iii), the intersection between $\sum_{s\in\mathcal{S}(x)} E^*_s$ and E_x is unique and provides $\lambda^+(x)$. Note that when the driving energy is (even casually) additive, the intersection may exist and be unique, but it also may not exist, in which case $\lambda^+(x) = +\infty$, or span all over $\mathbb{R}$, in which case $\lambda^+(x) = -\infty$. At last, knowing $\lambda^+(x)$, E^*_x is then provided by Eq. (15). This functional dynamic programming (FDP) method thus allows to compute λ^+ and E^* for each region of H by a single bottom-up traversal of H. Afterwards, $\lambda^-(x)$ can be computed by a top-down traversal of H using Eq. (13).

After implementing two traversals of the hierarchy, bottom-up then top-down, one gets three quantities on each node: its partial energy function and its scales of appearance and disappearance. In particular, the partial energy $\mathsf{E}^*_{\hat{H}}$ of the top of the hierarchy gives the energy of each λ-minimal cut $C^*_\lambda(H)$ of H. We simply denote it by E^*. Let also $\mathsf{C}^*(\lambda) = \mathsf{C}(C^*_\lambda(H))$ denote the complexity of $C^*_\lambda(H)$, that is in a compression context the rate function, and $\mathsf{D}^*(\lambda) = \mathsf{D}(C^*_\lambda(H))$ its distance to the image, or distortion function. Knowing E^*, these functions are given by:

$$\begin{cases} \mathsf{C}^*(\lambda) = \dfrac{\partial \mathsf{E}^*}{\partial \lambda}(\lambda) \\ \mathsf{D}^*(\lambda) = \mathsf{E}^*(\lambda) - \lambda \dfrac{\partial \mathsf{E}^*}{\partial \lambda}(\lambda). \end{cases}$$

$\mathsf{C}^*(\lambda)$ is decreasing and $\mathsf{D}^*(\lambda)$ is increasing, hence the operational rate/distorsion function (C^* as a function of D^*) is well defined and is decreasing (see Fig. 4). This result, which holds in a particular operational context, matches the general rate/distorsion monotonicity result of Shannon (1959).



(a) $\mathsf{E}^*(\lambda)$ (b) $\mathsf{C}^*(\lambda)$

(c) $\mathsf{D}^*(\lambda)$ (d) $\mathsf{C}^*(\mathsf{D}^*)$

Figure 4. Typical energy curves associated to the minimal cuts of a hierarchy.

After FDP, some nodes might have an empty range of persistence, whenever $\lambda^+(x) \leq \lambda^-(x)$, which means that they never enter an optimal cut of H. These non persistent nodes do not belong to the scale-sets which represent H's minimal cuts for E_λ. They may thus be deleted from the initial hierarchy. Hence the definition:

Definition 11 (Persistent hierarchy and scale-sets associated to (H, E_λ)). Let H be a hierarchy and E_λ be a MASE, then the hierarchy $H^* \subset H$ defined by

$$H^* \triangleq H \setminus \{x \in H \mid \lambda^+(x) \leq \lambda^-(x)\}$$

is called the persistent hierarchy associated with H and E_λ. (H^*, λ^+) is the scale-sets representation of the minimal cuts of H for E_λ.

Figure 5 illustrates the different steps of the computation of this scale-sets and Fig. 6 shows the effect of

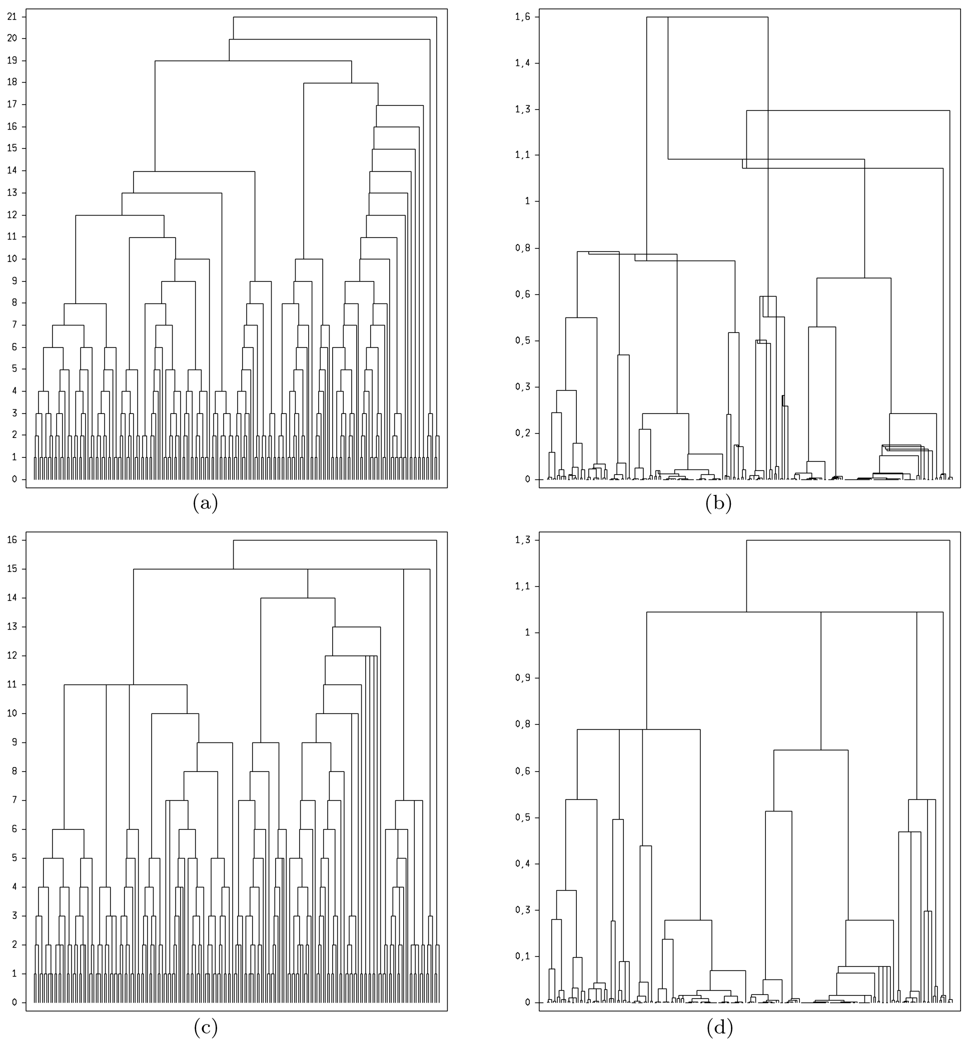


Figure 5. Multi-scale optimization in a hierarchy. (a) An initial binary hierarchy H drawn with a vertical axis corresponding to the levels of the regions in the hierarchy. After a bottom-up traversal of H, each region is attributed a scale of appearance in an optimal solution for the multi-scale energy considered. (b) represents the hierarchy H with a vertical axis corresponding to the scale axis. In this representation, the non persistent nodes, which never appear in an optimal solution, are upper than some of their ancestors. A top-down traversal of H then allows to compute the scales of disappearance of the regions, which reveal these non persistent nodes. (c) and (d) Represent the persistent hierarchy H^* obtained after removing the non persistent regions. Remark that deleting some intermediate stages leads to n-ary hierarchies. (d) illustrates that the scale of appearance is an increasing function in H^*.

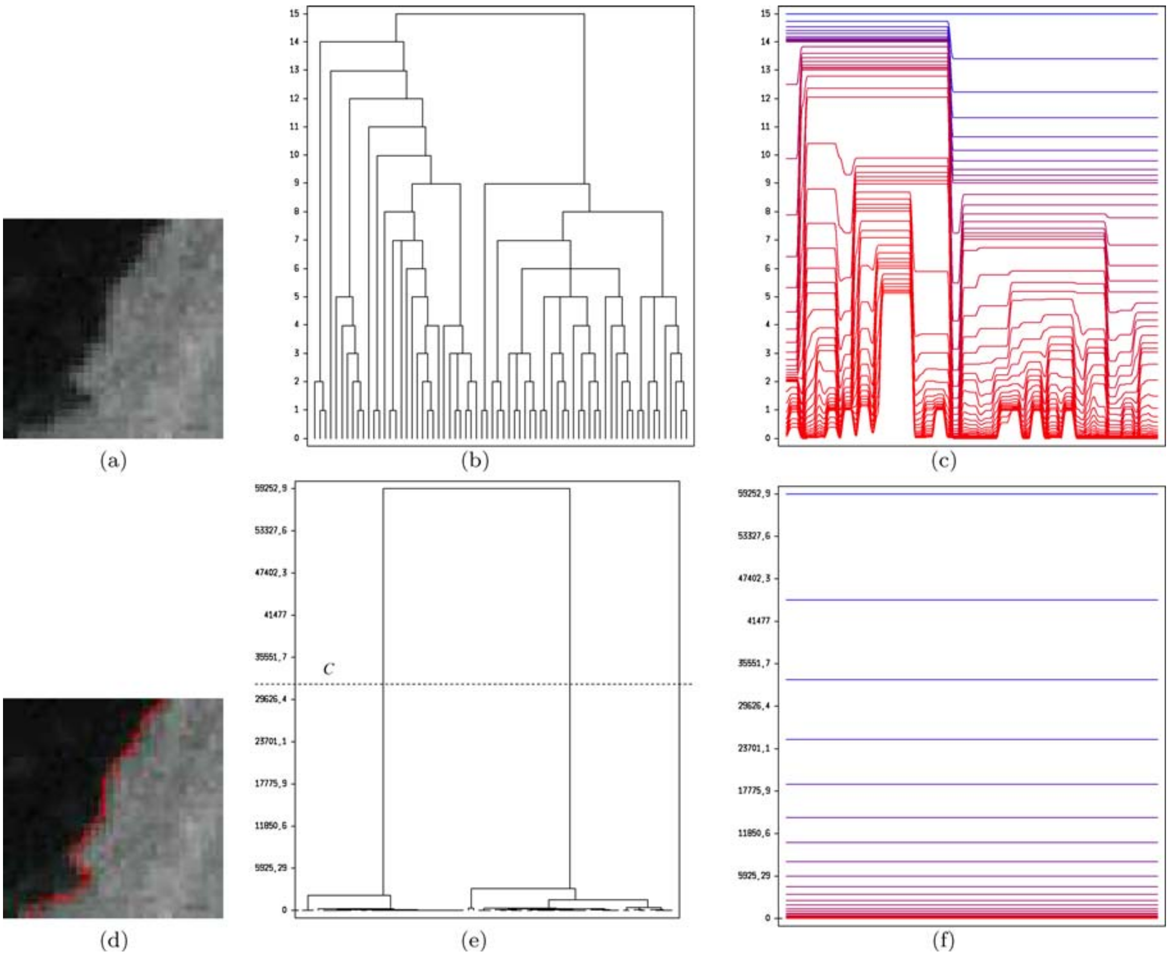


Figure 6. The scale-sets representation maps the family of the minimal cuts of a hierarchy into a family of horizontal cuts, or sections. (a) a 30 × 30 image. (b) The persistent hierarchy indexed by the level of its nodes. (c) The mininal cuts of the hierarchy (b) drawn in the same coordinate system than in (b). (d) The partition obtained by sectioning the scale-sets at half the scale of appearance of the top of the hierarchy. (e) The persistent hierarchy indexed by the scale of appearance of its nodes. C is the section corresponding to the partition in (d). (f) The minimal cuts of the hierarchy (b) drawn in the same coordinate system than in (e). The hierarchy was obtained by the scale-climbing algorithm using the Mumford-Shah functional (see below). The scale-sets exhibits two distinct clusters, revealing the global binary structure of the image.

the scale-sets representation on the family of minimal cuts of a hierarchy.

3.3.5. Discussion

Negative scales? As one might have noticed, nothing above constraints the scale parameter to be *positive*. In particular, all the base regions a priori have a scale of appearance of $-\infty$, and the top of the hierarchy has a scale of disappearance of $+\infty$. These two cases are side effects due to the relativity of the optimality criterion: the appearance depends on the energy of the sub-structures of a region and the disappearance depends on the energy of some super-structures which are undefined at the bounds of the sets lattice. The infinite ranges of persistence of both the base and the top are thus meaningless.

Apart from the extrema, it can be noted that, if the energy D also decreases (this may only happen casually), some regions which do not belong to the base might also get a negative scale of appearance. The idea of obtaining negative scales is somewhat unnatural: what does a negative regularization mean? If no regularization at all is needed to prefer a region described as a whole rather than described by its parts, it means that the "goodness-of-fit" energy D itself embeds a regularizing term. A negative scale of appearance can then be interpreted as an "anti-regularization" factor needed to compensate the internal regularization of D, in order to ... split the region ! Note that if D is always decreasing then the top of H appears at a negative scale: the optimal cut immediately jumps to the top and the whole image is merged for all positive scales.

Conversely, it can be shown that if D increases then all the scales of appearance are positive (this appears as a consequence of proposition 14 below). Thus, "true" opponent energy schemes—involving a decreasing prior and an increasing goodness-of-fit term—lead to intuitive solutions, belonging to the positive scales domain. Other reasons might be invoked to restrict the range of solutions to positive scales. For example, in minimum encoding frameworks negative scales correspond to negative code lengths, which are meaningless. In practice, we only used "true" or nearly "true" opponent energies and we restricted the final solution to the positive scales domain.

Multi-dimensional extension. One can easily check that the optimization technique can be extended to multi-dimensional regularizers by considering a vector-valued "scale-parameter" $\lambda \in \mathbb{R}^n$ and also a vector of regularizing energies $\mathsf{C} : \mathcal{P}(X) \rightarrow \mathbb{R}^n$. If $< \cdot, \cdot >$ denotes the inner product of $\mathbb{R}^n$, one can then consider an energy of the form $\mathsf{E}_\lambda(R) =< \lambda, \mathsf{C}(R) > +\mathsf{D}(R)$. If all the components of C are sub-additive then all the framework is still valid. The partial energy functions are then piece-wise affine hyperplanes of $\mathbb{R}^{n+1}$.

3.4. The Scale Climbing Principle

The previous section has explained how *for a given hierarchy of regions*, the full sequence of its optimal cuts for a multi-scale energy E_λ could be computed and represented exactly. The methodology is general-purpose: it can be used after any process which builds a hierarchy. One could for example use it on a quad-tree or on a hierarchy obtained by recursively splitting the image using Shi and Malik's normalized cut criterion (Shi and Malik, 2000). However, the results obtained so far put forward a natural way to build a hierarchy whose minimal cuts seek the global minima of the energy considered.

As we have seen, given a hierarchy H and a multi-scale energy E_λ, H's λ-cut climbs the hierarchy when λ increases. It jumps to a higher region x each time λ reaches a specific scale namely $\lambda^+(x)$. Increasing λ amounts to regularizing the solution by strengthening the constraint on the model's complexity or—by duality—by softening the constraint on the model's goodness-of-fit. Thus consider starting from the finest solution—the one which involves the slightest regularization ($\lambda = 0^+$)—and simulating a continuous increase of λ. For most classes of models, the 0^+-regularization solution is straightforward, e.g. for a piece-wise constant model it is made up of the flat zones of the image. Call this initial partition C_0. Now check among all the under-partitions of C_0 *which one would be reached first* during a continuous increase of λ. Assume it is reached at λ_1 and call it C_{λ_1}. Collect it in the scale-sets as its λ_1-section and repeat the process: increase λ until a different partition is reached, say at λ_2, collect it in the scale-sets, and so on. As we shall now explain, such a continuous simplification process can be handled exactly within our hierarchical framework.

3.4.1. "Pure" Scale Climbing.

We first define what we call the "pure" scale climbing strategy to build the hierarchy. Assume we want to build H by a greedy technique: starting from the empty hierarchy and adding sets one by one to the solution. Let H_0 be the minimal hierarchy on $\mathcal{D}$, that is the one which does not contain any other set than the singletons and $\mathcal{D}$ itself. Let H_k be the hierarchy obtained after k sets have been added to H_0. Let R_k be the set of the regions that can be added to H_k without breaking the hierarchical structure of the solution: $R_k = \{R \in \mathcal{P}(\mathcal{D}) | R \notin H_k \text{ and } H_k \cup \{R\} \text{ is a hierarchy}\}$. Indeed, at each step k, the scale of appearance of any $R \in R_k$ in a minimal cut of the hierarchy $H_k \cup \{R\}$ is well defined. The "pure" scale climbing algorithm is then

Definition 12 (Pure scale climbing). Starting from the empty hierarchy H_0, recursively add to H_k the region R of R_k which would have the smallest scale of appearance in $H_k \cup \{R\}$. Stop when $R_k = \emptyset$. The final hierarchy is called the **scale climbing hierarchy (SCH)** associated with E_λ.

As shows the next proposition, the pure scale climbing algorithm follows a sequence of *upward global minima* of E_λ when scale increases:

Theorem 13. *Let H_k be the hierarchy obtained after k steps of the scale climbing strategy for the multi-scale energy E_λ and let H^* be the hierarchy obtained* *after the scale climbing procedure is completed. Let C_k^u be the coarsest cut of H_k which is not the trivial cut $\{\mathcal{D}\}$. Let $R^* = argmin_{R \in R_k} \lambda^+(R)$ be the next region to be added to H_k by scale climbing and $S^* = argmin_{R \in R_k \setminus R^*} \lambda^+(R)$ be the region of R_k which has the second smallest scale of appearance after R^*. Then, for all $\lambda \in [\lambda^+(R^*), \lambda^+(S^*)[$, H^*'s λ-minimal cut minimizes E_λ in the set of all the under-partitions of C_k^u.*

Sketch of the Proof: Obviously, the set R_k of the regions which can be added to the solution gets smaller and smaller, namely whatever region is added at each step: $R_{k+1} \subset R_k$. Furthermore, one shows that the scales of appearance of the sets which remain in R_{k+1} after a completion step always *increase*, that is $\forall R \in R_{k+1}$: $\lambda^+(R)$ in $H_{k+1} \cup \{R\}$ is always greater than $\lambda^+(R)$ in $H_k \cup \{R\}$ (see Guigues, 2003). Hence, adding R^* to H_k ensures that it will persist in the final scale-sets at least over the range $[\lambda^+(R^*), \lambda^+(S^*)[$. Moreover, as adding sets of R_k to H_k allows to build any under-partition of C_k^u, the partition built is optimal among all the under-partitions of C_k^u. □

Hence, the pure scale climbing strategy tracks a global minimum among all the under-partitions of the current solution along the scale dimension.

The point is then that adding the sets in the order of their scale of appearance always leads to adding some *supersets* of the sets previously added. Indeed, as the scales of appearance of the completion hypotheses always increase after a completion step, no reversal of order can occur. As a consequence, all sets of a pure scale climbing hierarchy are persistent. This bottom-up direction of construction matches the direction of the dynamic programming method to compute λ^+, thus if the dynamic programming computation has already been made on the partial solution H_k, then the value of $\lambda^+(R)$ for any $R \in R_k$ can be obtained by a single dynamic programming step. The whole framework thus points out a privileged way to build the hierarchy, namely *the causal way*.

Please note that, whereas scale climbing is a discrete procedure, it properly simulates a *continuous* increase of λ. The point is that the procedure explicitly computes the scales at which a simplification of model would occur. It thus can be viewed as a *graduated non-convexity* (Blake and Zisserman, 1987) or a *continuation* minimization method (Leclerc, 1989) constrained to produce ordered partitions. Indeed, the "scale" parameter λ provides a natural embedding of the energy. The following proposition gives a rate/distortion interpretation of this continuation:

Proposition 14. *The scale climbing strategy amounts to choosing at each completion step the region $R \in R_k$ which minimizes*

$$-\frac{\Delta \mathsf{D}}{\Delta \mathsf{C}} = -\frac{\mathsf{D}(\{R\}) - \mathsf{D}(P)}{\mathsf{C}(\{R\}) - \mathsf{C}(P)}$$

where P is the coarsest minimal cut of the partial hierarchy $H_{k+1}(R)$ which is not $\{R\}$.

Proof: $\lambda^+(R)$ represents the scale for which the energy of $\{R\}$ becomes equal to the energy of the coarsest minimal cut P of $H_{k+1}(R)$ which is not $\{R\}$. Hence:

$$\begin{aligned} \lambda^+(R) = \lambda \ \text{ such that } & \\ & \lambda \mathsf{C}(\{R\}) + \mathsf{D}(\{R\}) = \lambda \mathsf{C}(P) + \mathsf{D}(P) \\ = -&\frac{\mathsf{D}(\{R\}) - \mathsf{D}(P)}{\mathsf{C}(\{R\}) - \mathsf{C}(P)}. \end{aligned}$$

□

The expression $\frac{\Delta \mathsf{D}}{\Delta \mathsf{C}}$ is the discrete form of the derivative of the distortion D with respect to the rate C. The scale climbing strategy can thus be interpreted as a steepest descent strategy along the operational rate/distortion curve of the problem.

Moreover, the scale climbing algorithm produces hierarchies which are robust to linear transforms of the energies:

Theorem 15 *(Linear invariance of SCH).* *If (H^*, λ^+) is the SCH associated with the energy $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$, then the SCH associated with $\mathsf{E}'_\lambda = (\mu \mathsf{C}, \nu \mathsf{D})$ is $(H^*, \frac{\nu}{\mu}\lambda^+)$.*

Sketch of the Proof: The proof can be made by induction on the dynamic programming step of the computation of λ^+, on the basis that if $f_{(a,b)}(x) = ax + b$ is an affine function and $I_x(f, g)$ denotes the intersection abscissa of two affine functions f and g then

$$I_x\big(f_{(\mu a, \nu b)}, f_{(\mu a', \nu b')}\big) = \frac{\nu}{\mu} I_x\big(f_{(a,b)}, f_{(a',b')}\big).$$

This relation remains valid for sum and infimum operations on piece-wise affine functions. Hence, as all scales of appearance are multiplied by ν/μ, the scale climbing order is unchanged by linear transforms of the energies. □

This invariance property has very important consequences.

The *structure* of a SCH is invariant by any linear transform on C or D. Adjusting the relative weight of the initial energies just stretches the scale axis of the final scale-sets proportionally. The energies used can thus be stated up to a multiplicative term and the right balance between complexity and goodness-of-fit of a model can always be set *a posteriori*. This allows us to achieve our initial goal, which was to extricate the scale parameter from the low level segmentation stage. It then becomes a real "potentiometer" allowing to explore the scale dimension.

This linear invariance also allows for the transfer of invariance properties from the energies to the final description. For instance, if the goodness-of-fit term is

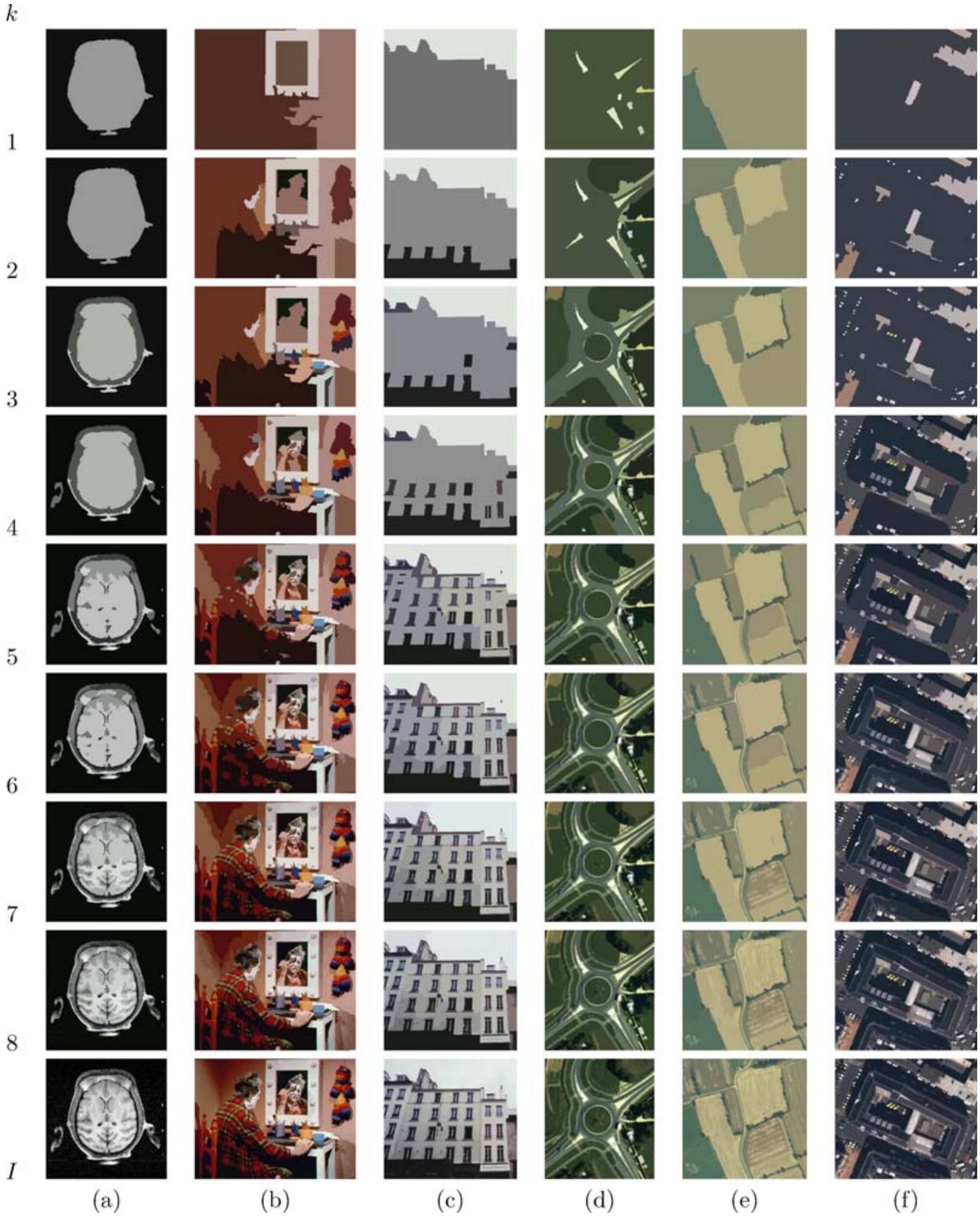


Figure 7. Dichotomies of some scale-sets obtained by scale climbing from the image pixels using Mumford-Shah's piece-wise constant model. k is the order of the dichotomy (see text). The bottom row represents the original images.

only scaled when the image values are scaled then so is the SCH. For example, one can easily verify that the solutions obtained with the Mumford and Shah's piecewise constant model are invariant to linear transforms of the image values and to homotheties of the image domain.

3.4.2. Binary Scale Climbing. Of course, the pure scale climbing strategy is computationally intractable: if $|\mathcal{D}| = N$ then the set R_0 of the initial possible completions has almost 2^N elements. In order to get a practical algorithm, we thus restrict the search space to a local search space. As before, we denote by C_k^u the coarsest cut of H_k which is not the whole domain. Instead of considering all the connected supersets of the regions of C_k^u, we only consider the supersets which can be obtained by merging pairs of adjacent regions of C_k^u. The algorithm obtained works as follows:

Definition 16 (Binary scale climbing). Let E_λ be a multi-scale energy and P_0 a fine segmentation of $\mathcal{D}$. Set $P \leftarrow P_0$ and $H \leftarrow P_0$. While $|P| \neq 1$ merge the pair $\{R, S\}$ of adjacent regions of P whose scale of appearance $\lambda^+(R \cup S)$ is minimal and add $R \cup S$ to H. The final hierarchy H is called the **Binary Scale Climbing Hierarchy (BSCH)** associated with E_λ and P_0.

The construction of a BSCH is parameter-free, given an over-segmentation and a couple of antagonist energies on the regions of an image. It is based on a classical priority queue region merging algorithm which can be efficiently implemented using a region adjacency graph and a heap structure to manage the queue of merging hypotheses (Kurita, 1995; Haris et al., 1998). The bottom-up dynamic programming optimization is realized simultaneously to the construction of the hierarchy. Each time a pair of regions is merged, a DP step allows to compute the scale of appearance of the union of the new region with its neighbors. After the hierarchy is completed, a top-down propagation finally provides the scales of disappearance of the regions which allow to prune the hierarchy in order to get the persistent hierarchy. Note that the initial binary hierarchy then becomes n-ary. We show in Guigues (2003) that the whole procedure can be implemented with a worst case complexity in $\mathcal{O}(N^2 \log N)$, where N is the size of the initial over-segmentation. For usual images, building a BSCH takes almost linear time.

While the search space is reduced, the major properties of the "pure" scale climbing strategy remain valid. The global minimum is only tracked locally, among the immediate under-partitions of the current solution. However, the local step is still chosen by steepest descent along the rate/distorsion curve and the final structure is also linearly invariant. Furthermore, the

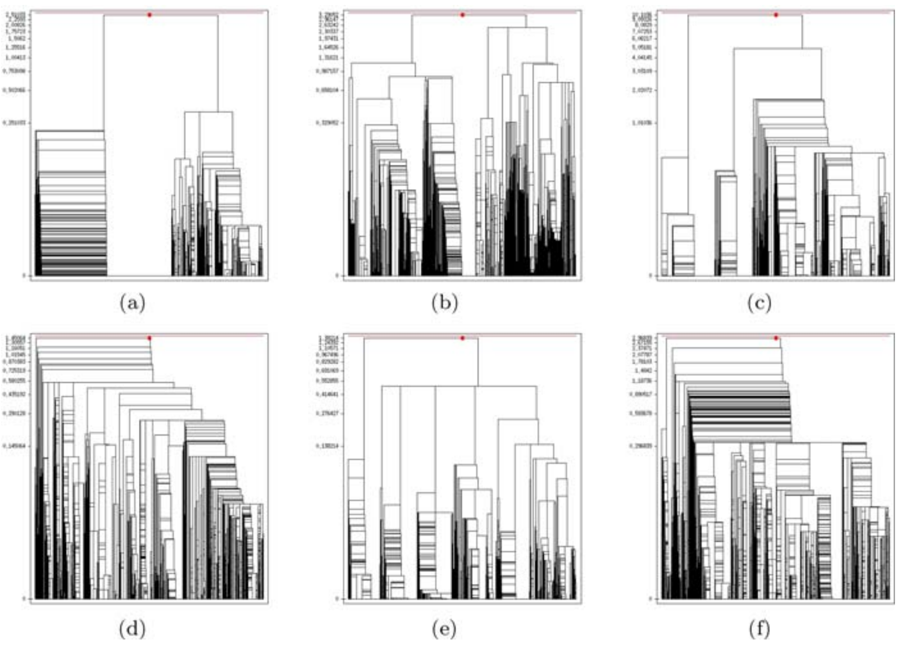


Figure 8. Scale-sets corresponding to the images of Fig. 7. The scale axis is logarithmic with a precision p of 256 (see Eq. (17)).

persistent hierarchies obtained are almost binary hierarchies (see below), which indicates that most of the time the "pure" scale climbing completion step would certainly have been chosen among the immediate supersets of the current upper sets, like in binary scale climbing.

4. Experimental Results

We shall now proceed with some experimental results.

We first consider Mumford and Shah's classical piece-wise constant approximation (Mumford and Shah, 1989). The discrete version of this model assigns the following energies to a region R:

$$\begin{aligned} \mathsf{C}(R) &= |\delta R| \\ \mathsf{D}(R) &= \sum_{x \in R} ||I(x) - \overline{I_R}||^2. \end{aligned} \tag{16}$$

$\mathsf{C}(R)$ is the length of the boundary of R, and D is the sum of the quadratic deviations from the mean value of the image within the region. This second term is valid for any number of channels in the image.

Remark that a particular scale arises after the computation of a scale-sets, namely the scale of appearance of the top of the hierarchy, $\lambda^+(\mathcal{D})$ which represents the minimum regularization strength needed to get the image modeled by a single region. This scale depends on the image content, contrast and size. However recall that with the Mumford-Shah functional, the scales

Figure 10. The same image segmented into 15 regions with a) a boundary length-based regularizer (Mumford-Shah), and b) a number-of-regions-based regularizer.

of appearance are linearly covariant with respect to the image dynamics and size. Hence sectioning a scale-sets at a scale which is defined *relatively* to $\lambda^+(\mathcal{D})$ leads to an analysis which is invariant to linear transforms of the image values and to homotheties of the image domain. In particular, in order to present a visual result, we consider the pyramid of N segmentations obtained by sectioning a scale-sets at scales $\lambda^+(\mathcal{D})/2^k, k = 1 \ldots N$. We call this pyramid the dichotomy of the scale-sets.

Figure 7 shows the dichotomies of the scale-sets obtained with Mumford and Shah's energy for six different images. Grouping started from the image pixels hence the algorithm is absolutely parameter-free. With our implementation, the average computation time is 42 seconds on a 1, 2 GHz computer for a 256×256 image. Of course, starting from a coarser segmentation

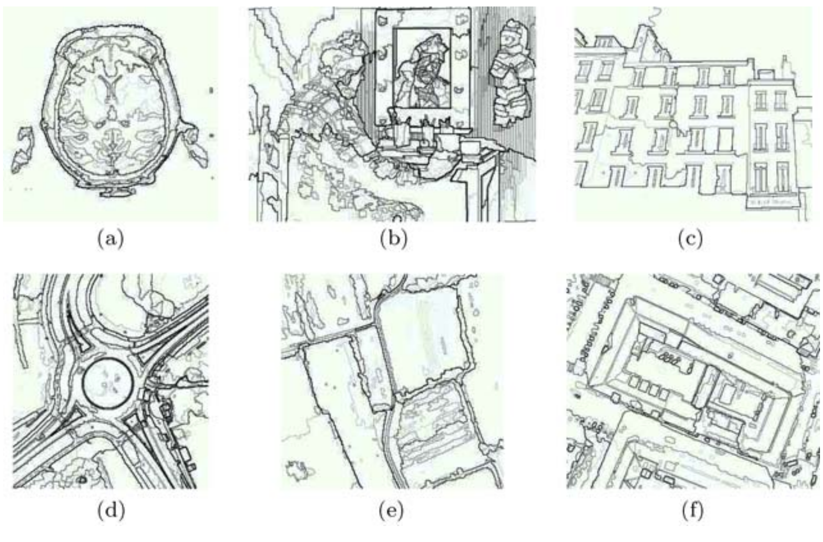


Figure 9. Contour disparition maps (ultrametrics) corresponding to the images in Fig. 7. Logarithmic scale axis of precision $p = 256$. Null log-scales are mapped to white and the log-scale of appearance of the whole domain is mapped to black.

than the pixel-wise tesselation greatly reduces the computation time. With the same energy, starting from a watershed segmentation (Beucher and Meyer, 1993) reduced the computation time to 2 to 3 seconds (the average size of the catchment basins ranged from 15 to 20 pixels).

Figure 8 shows the scale-sets corresponding to the images of Fig. 7. The scale axis is represented with a logarithmic scale, which is the natural representation for intensive units as it maps ratios into differences. As $\log(x) < 0$ when $x < 1$, an origin of the representation has to be chosen in order to get a representation

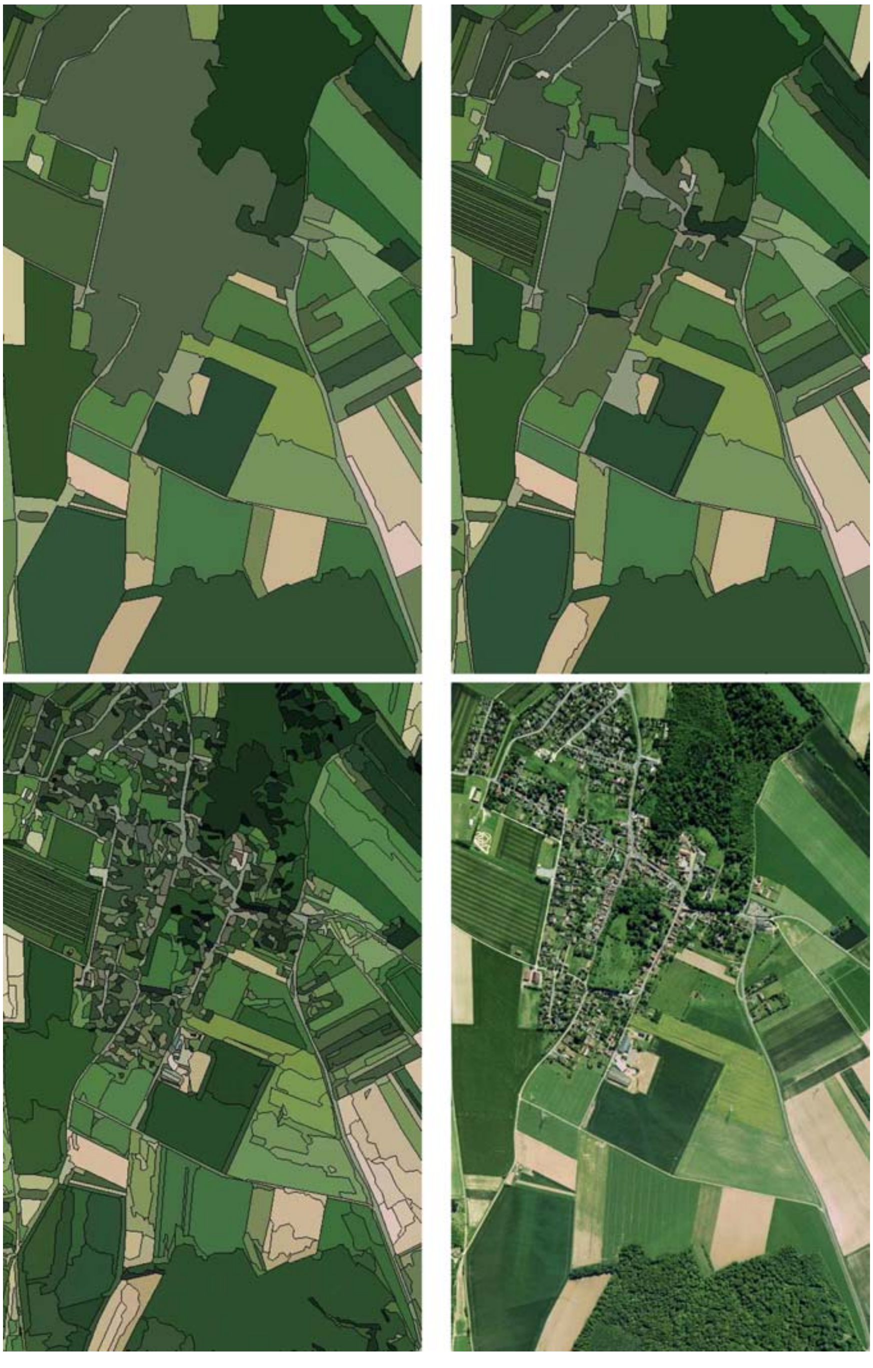


Figure 11. An aerial image.

belonging to $\mathbb{R}^+$. Again, a mapping relative to $\lambda^+(\mathcal{D})$ is natural:

$$\lambda \to \begin{cases} \log\left(\dfrac{\lambda p}{\lambda^+(\mathcal{D})}\right) & \text{if} \quad \lambda \geq \lambda^+(\mathcal{D})/p \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

where p is a parameter called the "precision" of the representation. In the examples, $p = 2^8$, hence matching the dichotomies in Fig. 7. The same mapping was used for the contour-oriented representation (ultrametrics) of Fig. 9.

Figure 7 clearly illustrates that meaningful objects appear at different scales in an image. The MR image of a head in Fig. 7(a) is a good example of it. Large scale values allow to separate the whole head from its background. Decreasing the scale parameter then progressively adds details to the description: the brain and the skull are first described as wholes and then internal details appear. Note that the scale-sets 8 (a) is clearly bimodal, revealing the object/background structure of the image. The right hand side of the hierarchy corresponds to the head while the left hand side corresponds to the background. The partial hierarchy which corresponds to the background is characteristic of an unstructured region: no large persistent region, no particular stage during the grouping. Also note that the contours corresponding to a scale-sets are based on global information and not on local intensity. Hence, the poorly contrasted contour of the top of the head in Fig. 7(a) gets the same "saliency" than the lower contour of the head which is very contrasted: both are equally meaningful because they participate in the boundary of the same globally coherent zone.

In order to illustrate the importance of geometrical regularization in segmentation we consider the segmentation of the same image with (i) the Mumford and Shah's energy and (ii) the Mumford and Shah's energy in which the energy C was replaced by $\mathsf{C}(R) = 1$. Hence in the second energy, the "complexity" of a partition is simply measured by the number of regions, independently of their geometry. Figure 10 shows the same image segmented into 15 regions with these two energies. The improvement of using a boundary length-based regularizer is clear.

Figures 11 and 12 show results obtained with a more sophisticated geometrical model than the simple length-based energy of Mumford and Shah.

Let $L = (m_1, m_2, \ldots m_k)$ be a sequence of k 2D points defining a closed 2D polygonal line. If $v_i = m_{i+1} - m_i$ denotes the displacement vector associated with the segment $[i, i+1]$, we say that the quantity

$$\mathsf{S}(L) = \sum_{i=1}^{k} |\widehat{v_i, v_{i+1}}|$$

where $\widehat{a, b}$ denotes the angle made up by vectors a and b, is the **concavity energy** of polyline L.

The name "concavity" comes from the fact that this energy has a value of 2π for all convex polygons and

Figure 12. Group photograph.

increases as the total curvature of the concavities of a polygon increases. To measure the concavity of a discrete region, we first perform a slight polygonal approximation of its boundary with a classical algorithm in order to suppress the pixelic artefacts.

In Examples 11 and 12, the initial over-partition was obtained by a watershed algorithm based on the magnitude of a Canny-Deriche edge detector tuned to $\alpha = 2$ (Deriche, 1987). As before, we considered piece-wise constant image modelling and employed the L_2 norm as a fidelity criterion (Eq. (16)). The difference with the Mumford-Shah functional is that the regularizing energy is the concavity energy.

We observed on several experiments that this energy produces more meaningful regions than Mumford-Shah's energy (the scale-sets of Fig. 1 was also computed with this energy). A point which might appear surprising for a piece-wise constant model is that this energy allows to find the right delineation for very textured units at large scales, such as the village in Fig. 11. The point is that energy minimization criteria are relative criteria: even if a very textured region is absolutely not constant, if it is different enough from its environment then for a low precision, the best piece-wise constant model is a model which separates the region from its background.

5. Conclusion and Future Work

A general methodology for multi-scale region-oriented image analysis has been presented. It can be viewed as a synthesis between structural—hierarchical grouping—approaches and energy minimization based approaches or as a region-oriented counterpart to the scale-space theory. The main result of the paper is that usual piece-wise modelling problems expressed in a variational way intrinsically contain a free parameter which behaves as a scale parameter. We then developed a methodology which provides an approximation of the full structure of the solutions with respect to scale, which we called a scale-sets description in reference to the scale-space theory. The scale-climbing algorithm is absolutely parameter-free given a multi-scale energy and a base segmentation which can be the pixel-wise tessellation of the image. The fundamental property of the structure obtained is its linear scale invariance which ensures that scale tuning is totally deferred to a subsequent stage.

The final hierarchical representations allow the dynamical browsing of the complete set of the multi-scale solutions. The useful scale of description for a specific application can thus be tuned *a posteriori*, either interactively or automatically (e.g. by actively searching for specific objects). Also, partitions made up of objects living at different scales may be composed, which correspond to non horizontal cuts in the hierarchy.

While we only present segmentation results on natural images, the multi-scale modelling framework is not dedicated to a specific kind of image, nor is it to the partitioning of image data. In Taillandier et al. (2003), the scale-climbing algorithm method is used to perform multi-scale piece-wise plane descriptions of range images obtained from a classical stereovision algorithm.

Different applications of the proposed framework have already been developed by different authors. Chehata et al. (2003) proposed a stereovision algorithm based on matching the regions of two scale-sets descriptions of a stereo pair. The robustness of the algorithm comes from the fact that regions of different levels of the hierarchies can be matched, hence avoiding the classical problem of differences in level of segmentation in region-oriented stereo-matching. We observed that, due to its robustness, the scale-climbing algorithm often segments in much the same way the matching regions of a stereo-pair, however not necessarily at the same scale. Roger and Marc (2004) also used the contour-based representation of a scale-sets (the ultrametrics) to register a cadaster graph onto an aerial image segmented with the scale-climbing algorithm.

Salembier and Garrido and co-workers have developed a number of image processing techniques based on hierarchies of regions. They have shown how such structures could be used for a large number of processing goals such as object detection and recognition, visual browsing and region-oriented image compression, filtering, image data-base indexing and similarity based retrieval. All these techniques can naturally rely on scale climbing hierarchies. We refer the reader to the authors' papers (Salembier et al., 1998; Garrido and Salembier, 1999; Salembier and Garrido, 2000a; Salembier and Garrido, 2000b) for detailed descriptions. Part of our future work will be centered on the development of such applications using SCHs.

Meyer (1999b); Marcotegui et al. (1999); Zanoguera et al. (1999) have proposed interactive segmentation systems based on the multi-scale approach of Meyer Meyer (1999a). Again, these interactive tools could be based on SCHs.

In the scale-space theory, the evolution diagram (when scale increases) of the features for which the analysis is causal (e.g. the image maxima) is known as the scale-space *fingerprint* of an image (Jackay et al., 1994). The scale-sets itself, that is the stratified hierarchy structure, is the fingerprint associated with a multi-scale segmentation. Like a scale-space fingerprint, it could for example be used in content-based image retrieval problems.

Obviously, the persistence of the regions of a SCH is a very meaningful information. The range of scales over which a region lives is a measure of its stability, in the sense of (Leclerc, 1989). It represents the efforts needed to merge it with its neighborhood and thus quantifies the region's saliency. Starting from a point in the image and climbing the associated branch of the hierarchy, a chain of nested regions is browsed. When the "growing" region reaches a maximal coherent part of an image, a peak of persistence can be observed, which generally means that the region delineated corresponds to a meaningful object of the scene. The issue of exploiting the persistences (and also energetical variations related to merging operations) in order to extract salient objects of an image will be part of our future work.

To conclude, we would like to get back to some methodological aspects of the approach.

First and back with David Marr, we'd like to stress the "duality" between the "representation and the processing of information" (Marr, 1982, first sentences, p. 3). The present paper clearly shows that the investigation of how the solutions to a problem can be represented can lead to important hints on the way to solve it. Also, our methodology relies on addressing an intractable problem into a subspace of the space of its solutions (the cuts of a hierarchy rather than the whole partitions lattice), in which it can be exactly solved. As we have seen, finding exact resolution methods in structured subspaces can then, in a kind of feedback loop, lead to privileged ways to build the subspaces and thus to fine approximation methods. Finally, we would like to go back over the functional dynamic programming principle. It amounts to handling the explicit form of a parametric energy towards its parameters and considering the inverse problem of finding the values of the parameters for which an element of the search space is optimal. When such an approach is successful, it turns a parameter into a real "potentiometer" which allows flexible algorithm tuning. It also allows to quantify the stability of a solution, a point which is certainly as important as obtaining one. This "inverse parametric" approach is certainly worth being investigated for other computer vision or pattern recognition problems.

Appendices

A. Definitions and Notations

This section recalls some general definitions and puts the global notations of the paper.

- Given a set $\mathcal{D}$, a subset $S \subset \mathcal{P}(\mathcal{D})$, where $\mathcal{P}(\mathcal{D})$ denotes the set of the parts of $\mathcal{D}$, is called a set system on $\mathcal{D}$.
- A set system made up of non overlapping sets covering $\mathcal{D}$ is a **partition** of $\mathcal{D}$. We denote by $\mathbb{P}(\mathcal{D})$ the set of the partitions of $\mathcal{D}$. Recall that a partition P is *finer* than a partition Q, $P \leq Q$, if and only if $\forall p \in P\ \exists q \in Q\ :\ p \subseteq q$. The inclusion relation, $\subseteq$, is a partial order on $\mathcal{P}(\mathcal{D})$ and the fineness relation, $\leq$, is a partial order on $\mathbb{P}(\mathcal{D})$. $\subseteq$ and $\leq$ induce complete lattice structures on $\mathcal{P}(\mathcal{D})$ and $\mathbb{P}(\mathcal{D})$.
- A set system H of $\mathcal{D}$ is a **hierarchy** on $\mathcal{D}$ if and only if

  (i) $\emptyset \notin H$
  (ii) $\mathcal{D} \in H$
  (iii) $\forall x \in \mathcal{D}\ :\quad \{x\} \in H$
  (iv) $\forall (X, Y) \in H^2\ :\quad X \cap Y \in \{\emptyset, X, Y\}$
  (X and Y are either disjoint or nested)

  The set $\mathcal{D}$ is called the top of H and we denote it by $\hat{H}$. The subset $\{\{x\}, x \in \mathcal{D}\}$ of H is called its base (or bottom) and we denote it by $\underline{H}$. For image segmentation problems, $\mathcal{D}$ is the domain of the image under scope, and H's base represents the finest partition considered: either the absolute over-partition of $\mathcal{D}$, into individual pixels, or a coarser one, such as its tessellation into flat zones (Serra and Salembier, 1993) or into the catchment basins of an image gradient (Beucher and Lantuejoul, 1979).

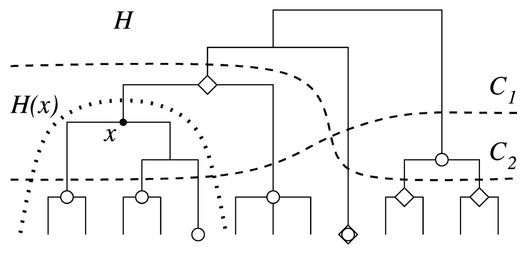


Figure 13. A hierarchy H represented as a tree. $H(x)$ is the partial hierarchy rooted at x. C_1 and C_2 are two cuts, resp. made up of circled and diamond-shaped nodes.

- A hierarchy H can be represented as a rooted tree $T = (H, S)$ (see Fig. 13). The nodes of T represent the regions of H (the root is the top of H), and the edges S of T represent the covering relation between the elements of H, i.e. $(x, y) \in S \Leftrightarrow x \subset y$ and $\nexists z \in H | x \subset z \subset y$. For any edge $(x, y) \in S$, y is called x's father and x is one of y's sons. We denote the father of a node x by $\mathcal{F}(x)$ and the set of the sons of a set y by $\mathcal{S}(y)$.
- $\forall x \in H$, the subset of H defined by $H(x) \triangleq \{y \in H, y \subseteq x\}$ is a hierarchy on x which we call H's **partial hierarchy** induced by x. The associated tree is the complete subtree rooted at x (see Fig. 13).
- A **cut** C of a hierarchy H is a subset of H which intersects any path from the base to the top of H exactly once. Equivalently, the cuts of a hierarchy H on a set $\mathcal{D}$ are the partitions of $\mathcal{D}$ that can be obtained by picking sets in H. A cut C can be graphically represented by a curve which divides the tree into two sets of nodes: the ones which remain connected to the top and the ones which remain connected to the bottom (see Fig. 13). The elements of the cut itself are the upper nodes remaining connected to the bottom. We denote by $\mathbb{C}(H)$ the set of all the possible cuts of a hierarchy H. Please note that:

i) Two cuts C_1 and C_2 of a hierarchy are in general *unordered* (C_1 is neither coarser nor finer than C_2). For example, the two cuts in Fig. 13 are unordered, which can be checked because their representative curves intersect.

ii) The number of possible cuts of a hierarchy is in general very large. One easily shows that if H is a balanced binary tree built on a set of size N then $|\mathbb{C}(H)| > \sqrt{2}^N$ (we provide a precise approximation of this cardinal in Guigues (2003)). It is interesting to compare this bound to the number of partitions of a planar domain in *connected regions*, which is lower than 4^N (Wang, 1998).

B. Proof of Theorem 8

We focus on the case of a decreasing energy C. The increasing case is symmetrical and the case of a monotony condition on D is obtained by the fact that the minimum of $\lambda\mathsf{C} + \mathsf{D}$ is the same as the minimum of $\mathsf{C} + \frac{1}{\lambda}\mathsf{D}$. The inversion in λ explains the change of direction of monotony in the solutions.

Let $\lambda_1 \in \mathbb{R}$. Expanding the partial λ_1-optimality property (proposition 7-i) for all x in $P^*_{\lambda_1}(H)$ then gives:

$$\forall x \in P^*_{\lambda_1}(H) \quad \forall Y \in \mathbb{C}(H(x)):$$

$$\lambda_1\mathsf{C}(x) + \mathsf{D}(x) \leq \sum_{y\in Y} \lambda_1\mathsf{C}(y) + \mathsf{D}(y)$$

$$\Leftrightarrow \quad \lambda_1\left(\mathsf{C}(x) - \sum_{y\in Y}\mathsf{C}(y)\right) \leq \sum_{y\in Y}\mathsf{D}(y) - \mathsf{D}(x)$$

As C is decreasing, and by definition Y is a partition of x, we get

$$\Delta\mathsf{C} = \mathsf{C}(x) - \sum_{y\in Y}\mathsf{C}(y)$$
$$= \mathsf{C}\left(\bigcup_{y\in Y} y\right) - \sum_{y\in Y}\mathsf{C}(y).$$

Hence, if C is a sub-additive function on sets then $\Delta\mathsf{C} \leq 0$, which implies that $\forall\lambda_2 \in \mathbb{R}$

$$\lambda_2 \geq \lambda_1 \Rightarrow \lambda_2\Delta\mathsf{C} \leq \lambda_1\Delta\mathsf{C} \leq \sum_{y\in Y}\mathsf{D}(y) - \mathsf{D}(x)$$

which expresses the partial λ_2-optimality of x. We thus obtain a "partial causality" property:

$$\forall(\lambda_1, \lambda_2) \in \mathbb{R}^2 \quad \lambda_2 \geq \lambda_1 \Rightarrow (\forall x \in H \quad x \in P^*_{\lambda_1}(H) \Rightarrow x \in P^*_{\lambda_2}(H)) \tag{18}$$

Let's now examine the conditions of global optimality of a node x. The condition i) of Proposition 7 can be rewritten

$$\forall\lambda_1 \in \mathbb{R}\left(\forall x \in H \quad x \in C^*_{\lambda_1}(H) \Rightarrow x \in P^*_{\lambda_1}(H)\right).$$

So, using the "partial causality" Eq. (18) we obtain $\forall(\lambda_1, \lambda_2) \in \mathbb{R}^2$

$$\lambda_2 \geq \lambda_1 \Rightarrow \left(\forall x \in H \quad x \in C^*_{\lambda_1}(H) \Rightarrow x \in P^*_{\lambda_2}(H)\right).$$

Now, if $x \in P^*_{\lambda_2}(H)$ then no partially λ_2-optimal node y such that $y \subset x$ may be maximal (verify Proposition 7ii)). As y is λ_2-optimal if and only if it is a maximal partially λ_2-optimal node, we conclude that $\forall(\lambda_1, \lambda_2) \in \mathbb{R}^2$

$$\lambda_2 \geq \lambda_1 \Rightarrow \forall x \in C^*_{\lambda_1}(H)$$
$$\forall y \in C^*_{\lambda_2}(H)\ x \cap y \neq \emptyset \Rightarrow x \subseteq y$$
$$\Leftrightarrow \quad \lambda_2 \geq \lambda_1 \Rightarrow C^*_{\lambda_2}(H) \geq C^*_{\lambda_1}(H).$$

C. Proof of Proposition 10

(i) The proof is made by induction on the dynamic programming step in H.
Indeed, $\forall x \in H$, its self-energy $E_x = \lambda \mathsf{C}(x) + \mathsf{D}(x)$ is affine with $\mathsf{C}(x) \geq 0$ and is thus a piece-wise affine, non decreasing, continuous and concave function (NDCC). In particular, as $\forall b \in \underline{H}$ $\mathsf{E}^*_b = \mathsf{E}_b$, the partial energies of the base nodes are NDCC. Now take $x \in H$ and assume that $\forall s \in \mathcal{S}(x)$, E^*_s is NDCC. As finite sums and infima of NDCC functions are also NDCC, then according to Eq. (14), E^*_x is also NDCC.

(ii) Let $\mathsf{E}^*_{\mathcal{S}(x)} = \sum_{s \in \mathcal{S}(x)} \mathsf{E}^*_s$. By the dynamic programming principle the range $\Lambda^*_{\uparrow}(x)$ of partial λ-optimality of a node x reads

$$\Lambda^*_{\uparrow}(x) = \{\lambda | \mathsf{E}_x(\lambda) \leq \mathsf{E}^*_{\mathcal{S}(x)}(\lambda)\}.$$

According to proposition 9, this range is given by

$$\Lambda^*_{\uparrow}(x) = [\lambda^+(x), +\infty[.$$

Thus,

$$\mathsf{E}_x(\lambda) \leq \mathsf{E}^*_{\mathcal{S}(x)}(\lambda) \quad \Leftrightarrow \quad \lambda \geq \lambda^+(x)$$

which means that E^*_x has the piece-wise form given by Eq. (15).

(iii) As all the functions involved are continuous

$$\lambda^+(x) = \min\{\lambda | \mathsf{E}_x(\lambda) = \mathsf{E}^*_{\mathcal{S}(x)}(\lambda)\}.$$

Thus consider $\Delta = \mathsf{E}_x - \mathsf{E}^*_{\mathcal{S}(x)}$ and look for its zero crossings. $\forall \lambda \in \mathbb{R}$, $\mathsf{E}^*_{\mathcal{S}(x)}(\lambda)$ represents the energy of a cut $P(\lambda)$ of $H(x)$ which is a strict over-partition of $\{x\}$, hence

$$\mathsf{E}^*_{\mathcal{S}(x)}(\lambda) = \lambda \sum_{R \in P(\lambda)} \mathsf{C}(R) + \sum_{R \in P(\lambda)} \mathsf{D}(R)$$

and $\forall \lambda \in \mathbb{R}$

$$\frac{\partial \mathsf{E}^*_{\mathcal{S}(x)}}{\partial \lambda}(\lambda) = \sum_{R \in P(\lambda)} \mathsf{C}(R).$$

We thus get

$$\begin{aligned} \frac{\partial \Delta}{\partial \lambda}(\lambda) &= \frac{\partial \mathsf{E}_x}{\partial \lambda}(\lambda) - \frac{\partial \mathsf{E}^*_{\mathcal{S}(x)}}{\partial \lambda}(\lambda) \\ &= \mathsf{C}(x) - \sum_{R \in P(\lambda)} \mathsf{C}(R). \end{aligned}$$

So, if C is strictly subadditive, then $\forall \lambda \in \mathbb{R}$: $\frac{\partial \Delta}{\partial \lambda}(\lambda) < 0$. Δ is thus strictly decreasing and crosses zero for a unique $\lambda \in \mathbb{R}$ which is $\lambda^+(x)$.

Notes

1. This reversal of point of view can be made mathematically precise: if $\mathbf{P}$ is viewed as a multivalued function of λ returning *sets*, then $\Lambda(x)$ can be written $\Lambda(x) = \mathbf{P}^{-1}(x)$.
2. This comes from the fact that the MST of a graph (G, d), where d is a dissimilarity function, is a graph-based representation of the maximal ultrametrics δ bounded above by d. When d is already an ultrametrics then $\delta = d$.

References

Alvarez, L., Guichard, F., Lions, P.-L. and Morel, J.-M. 1993. Axioms and fundamental equations of image processing. *Arch. Rational Mech. Anal.*, 123:199–257.

Badaud, J., Witkin, A.P., Baudin, M., and Duda, R.O. Uniqueness of the gaussian kernel for scale-space filtering. 1986. *IEEE trans. Pattern Anal. Mach. Intell.*, 8:26–33.

Ballester, C., Caselles, V., and González, M. 1994. Affine Invariant Segmentation by Variational Method. In *Proc. of 9th RFIA*, Paris, pp. 379–390.

Beucher, S. and Lantuejoul, C. 1979. Use of watersheds in contour detection. In *Proceedings of Int. Workshop on Image Processing, Real-time Edge and Motion Detection/Estimation, Rennes*, (France),

Beucher, S. and Meyer, F. 1993. The morphological approach to segmentation : the watershed transformation. In Dougherty, (ed.), *Mathematical Morphology in Image Marcel Dekker, Processing*, chapter 12, pp. 433–481.

Blake, A. and Zisserman, A. 1987. *Visual Reconstruction*. MIT Press, Cambridge, MA.

Bosworth, J. H. and Acton, S.T. 2003. Morphological scale-space in image processing. *Digital Signal Processing*, 13(3).

Boykov, Y., Veksler, O., and Zabih, R. 1999. Fast approximate energy minimization via graph cuts: *ICCV, Seventh International Conference on Computer Vision (ICCV'99)*, Vol. 1, p. 377.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. *Classification and Regression Trees*. Wadsworth and Brooks.

Chehata, N., Jung F., Pierrot-Deseilligny, M., and Stamon, G. 2003. A region based approach for 3D-roof reconstruction form HR satellite stereo pairs. In *DICTA 03, Vol II*, Sydney, Australia, pp. 889–898,

Deriche, R. 1987. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187 .

Donoho, D. 1997. CART and best-ortho-basis selection : A connection. *Annals of statistics*, 25:1870–1911.

Fuchs, F. and Le Men, H. 1999. Detecting planar patches in urban scenes. In *Proceedings of SPIE, Visual Information Processing VIII, Orlando, USA*, volume 3716, pp. 167–176, avril .

Garrido, L., and Salembier, P. 1999. Representing and retrieving regions using binary partition trees. In *IEEE Int. Conference on Image Processing ICIP'99*, Kobe, Japan, pp. 25–28.

Geman S., and Geman, D. 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

Guigues, L. 2003. *Modèles multi-échelles pour la segmentation d'images (Multi-scale models for image segmentation)*. PhD thesis, Univ. de Cergy-Pontoise, France.

Guigues, L., Le Men, H., and Cocquerez, J.-P. 2001. Segmentation d'image par minimisation d'un critère mdl dans une pyramide de segmentations. In *actes du XVIII^{ème} colloque GRETSI sur le Traitement des Images et du Signal*, Toulouse, France.

Guigues L., Men Hervé Le, and Cocquerez Jean-Pierre. 2003. The hierarchy of the cocoons of a graph and its application to image segmentation. *Pattern Recognition Letters*, 24(3):1024–1066.

Haralick and Shapiro 1991. Glossary of computer vision terms. *Pattern Recognition*, 24(1):69–93 .

Haris, K., Estradiadis, S.N., Maglaveras, N., and Katsaggelos, A.K. 1998. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans. on Image Processing*, 7(12):1684–1699.

Horowitz, S.L. and Pavlidis, Theodosios. 1976. Picture Segmentation by a Tree Traversal Algorithm. *J. of the Assoc. for Comp. Mach.*, 23(2):368–388.

Jackay, P.T., Boles, W.W., and Deriche, M. 1994. Morphological fingerprints and their use in object recognition in range images. In *Proc. of IEEE Conf. on Acoustics, Speech and Signal Processing*, Adelaide, Australia, pp. V5–V8.

Jackay, P. T. and Deriche, M. 1987. Scale-space properties of the multiscale morphological dilation-erosion. *IEEE trans. Pattern Anal. Mach. Intell.*, 9:532–550.

Jolion, J.-M. and Montanvert, A. 1992. The adaptive pyramid, a framework for 2d image analysis. *CVGIP: Image Understanding*, 55(3):339–348.

Kanungo, T., Dom, B., Niblack, W., Steele, D., and Sheinvald, J. 1995. MDL-Based Multi-Band Image Segmentation Using a Fast Region Merging Scheme. Technical Report RJ 9960 (87919), IBM Research Division.

Koenderink, J.J. 1984. The structure of images. *Biol. Cybern.*, 50:363–370.

Koepfler, G., Lopez, C. and Morel, J.M. 1994 A Multiscale Algorithm for Image Segmentation by Variational Method. *SIAM Journal on Numerical Analysis*, 31(1):282–299.

Kurita, T. 1995. A Efficient Clustering Algorithm for Region Merging. *IEICE Trans. on Information and Systems*, E78-D(12):1546–1551.

Lance and Williams 1967. A general theory of classificatory sorting strategies. *The Computer Journal*, 9:373–380.

Leclerc, Y.G. 1989. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3(1):73–102.

Lindeberg, T. 1994. *Scale-space Theory in Computer Vision*. Kluwer Academic.

Marcotegui, B., Zanoguera, F., Correia, P., Rosa, R., Marqués, F., Mech, R., and Wollborn, M. 1999. A video object generation tool allowing friendly user interaction. In *Proc. of ICIP'99*, Kobe (Japan).

Marr, D. 1982. *Vision: A computational Investigation Into the Human Representation and Processing of Visual Information*. W.H. Freeman and Company, NY 29–61.

Meyer, F. 1999a. Graph based morphological segmentation. In *Proc. 2nd IAPR TC-15 workshop on Graph based Representations*, Haindorf, Austria, pp. 51–60.

Meyer, F. 1999b. Morphological multiscale and interactive segmentation. In *Proc. of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Antalya, Turkey.

Meyer, F. 2001. An overview of morphological segmentation. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 15(7):1089–1118.

Montanvert, A., Meer, P., and Rosenfeld, A. 1991. Hierarchical image analysis using irregular tesselations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):307–316.

Morel, J.-M. and Solimini, S. 1995. *Variational methods in image segmentation*. Progress in Nonlinear Differential Equations and Their Applications, vol. 14, Birkhaüser, Boston.

Mumford, D., and Shah, J. 1989. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. on Pure and Applied Math.*, 17(4):577–685.

Park, K.-R. and Lee, C.-N. 1996. Scale-space using mathematical morphology. *IEEE trans. Pattern Anal. Mach. Intell.*, 8:1121–1126.

Perona, P. and Malik, J. 1990. Scale space and edge detection using anisotropic diffusion. *IEEE trans. on PAMI*, 12(7):629–639.

Poggi, G. and Ragozini, A.R.P. 1999. Image segmentation by a tree-structured markov random field. *IEEE Signal Processing Letters*, 6:155–157.

Salembier, P. and Garrido, L. 2000a. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Trans. on Image Processing*, 9(4):561–576.

Salembier, P. and Garrido, L. 2000b. Visual segments tree creation for MPEG-7 description schemes. In *IEEE Int. Conference on Multimedia and Exposition ICME'2000, NY city*, NY USA.

Salembier, P., Oliveras, Albert., and Garrido, Luis. 1998. Anti-extensive Connected Operators for Image and Sequence Processing. *IEEE Trans. on Image Processing*, 7(4):555–570.

Salembier, P. and Serra, J. 1995. Flat zones filtering, connected operators, and filters by reconstruction. *IEEE trans. on Image Processing*, 4(8):1153–1160.

Sapiro, G. and Tannenbaum, A. 1993. Affine invariant scale space. *Int. J. of Computer Vision*, 11:24–44.

Serra J. and Salembier, P. 1993. Connected operators and pyramids. In *SPIE conf. on Image Algebra and Mathematical Morphology, San Diego, CA*, volume 2030, pp. 65–76.

Shannon, C. 1959. Coding theorems for a discrete source with a fidelity criterion. *IRE Natl. Conv. Record,*, Part 4:142–163.

Shi, J. and Malik, J. 2000. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

Taillandier, F., Guigues, L., and Deriche, R. 2003. A framework for constrained multi-scale range image segmentation. In *Proc. of IEEE Int. Conf. on Image Processing (ICIP'03)*, Barcelona, Spain.

Roger, T.-S. and Marc, P.-D. 2004. An edge-based method for registering a graph onto an image with application to cadastre registration. In *Proc. of the 2004 International Conference on Pattern Recognition (ICPR 2004)*, Cambridge, UK.

Wang, J.-P. 1998. Stochastic relaxation on partitions with connected components and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6):619–636.

Witkin, A.W. 1983. Scale space filtering. In *Proc. of 8th Int. Joint Conf. on Artificial Intell., Karlsruhe, West Germany*, pp. 1019–1021.

Zanoguera, F., Marcotegui, B., and Meyer, F. 1999. A toolbox for interactive segmentation based on nested partitions. In *proc. of ICIP'99*, Kobe (Japan).