MULTISCALE MODEL. SIMUL.
Vol. 0, No. 0, pp. 000–000

LOCAL SCALE MEASURE FROM THE TOPOGRAPHIC MAP AND APPLICATION TO REMOTE SENSING IMAGES*

BIN LUO†, JEAN-FRANÇOIS AUJOL‡, AND YANN GOUSSEAU§

Abstract. In this paper, we propose a spatially accurate definition of scale for images. The proposed definition of local scale relies on the hierarchical structure of the topographic map, the set of all level lines of an image. Namely, a connected component of a level line is associated with each point of the image and the scale at this point is then computed from the corresponding shape. Level lines are selected under the assumption that the blur is uniform over the image, a realistic assumption for satellite images. First, we discuss the links between the proposed approach and recent scale definitions making use of the total variation flow or the Rudin–Osher–Fatemi model. This comparison sheds some light on the relationships between morphological and variational approaches to scale definitions. Then we perform several experiments on synthetic and satellite images and numerically compare our local scale definition with results using the linear scale space, variational, or morphological approaches. These experiments suggest that the proposed method enables a robust and spatially accurate computation of local scales, without the need for complex parameter tuning.

Key words. local scale, topographic map, remote sensing images

AMS subject classifications. AUTHOR: PLEASE PROVIDE

DOI. 10.1137/080730627

1. Introduction.

The notion of scale is of primary importance in digital image analysis. On the one hand, it is crucial information to tune the spatial extent of analysis tools. In remote sensing imaging, the best observation scale is closely related to the concept of *analysis scale* in geography [39]. The NATO standard STANAG 3769 [40] gives some examples of the best scales for interpreting certain objects in remote sensing images. In computer vision, most object recognition or detection methods include a scale computation while or before extracting significant features. As a side product, this also permits one to achieve scale invariance, which is necessary considering the arbitrariness of object distances to the camera.

On the other hand, scale itself is also useful for the classification of images. In particular, remote sensing images are often made of several regions showing some texture-like homogeneity. These regions usually exhibit different size distributions of their constituents. It is tempting to characterize each pixel of such images by the size of the object(s) (in a sense to be defined) it belongs to. The purpose of this paper is to show that this idea is feasible and yields spatially accurate features for describing remote sensing images.

A classical approach to estimate the scale of structures in an image has been introduced by [25, 26]. Local scales are computed by studying extrema of differential operators in the linear scale space. This method has been widely applied in computer

*Received by the editors July 18, 2008; accepted for publication (in revised form) January 20, 2009; published electronically DATE. This work was supported by the CNES-DLR-ENST Competence Centre.

http://www.siam.org/journals/mms/x-x/73062.html

†TELECOM ParisTech, CNES-DLR-ENST Competence Centre, Paris, France, and DIS Gipsa-Lab, Grenoble, Saint Martin d'Heres 38402, France (Bin.luo@gipsa-lab.inpg.fr). This author's research was supported by the COC and TELECOM ParisTech.

‡CMLA, ENS Cachan, CNRS, UniverSud, Cachan 94235, France (Jean-Francois.Aujol@cmla.ens-cachan.fr).

§TELECOM ParisTech, LTCI CNRS, Paris 75014, France (gousseau@enst.fr).

vision to select the optimal analysis scale of structures; see, e.g., [28, 35]. Similar to this approach, the methods presented in [20, 46, 47, 52] propose to estimate the salient scale of an image by considering extrema of various information theoretic operators in the linear scale space. For remote sensing images, it is also proposed in [30] to rely on a linear scale space to estimate a resolution invariant characteristic scale. The invariance is achieved by studying the effect of the image acquisition process on the linear scale space decomposition. Indeed, the linear scale space is very convenient since it enables one to easily model scale (see, e.g., [35]) or resolution [30] changes. However, using the linear scale space prevents such methods from achieving spatial accuracy. It is well known that the positions of structures that are detected by such methods are relatively erratic and very sensitive to contrast changes, especially for large structures. Therefore, one can expect some difficulty in building up spatially accurate indexing of images from these methods. Moreover, these methods disregard structures without significant gradients, for which no scale is computed.

In order to obtain spatially accurate scale measures, it is quite natural to look toward nonlinear approaches. In [21, 22] it is proposed to measure the significant scales of structures by computing the entropy of the joint distribution of pixels in a fixed neighborhood. It has been shown (see [22]) that such an approach yields better repeatability and is less sensitive to image perturbation than methods making use of Gaussian filtering. Other methods enabling the computation of spatially accurate scales are mainly based on nonlinear scale spaces, either in a morphological or variational framework. Recently, several scale measures based on nonlinear partial differential equations have been proposed. In [51, 50], a local definition of scale based on total variation regularization is introduced. In [8], it is proposed to estimate the local scales of structures from their evolution speed under the total variation flow. In both works, the idea is that the evolution time of structures is related to their scale. Independently, the mathematical morphology school has long ago proposed to characterize materials by looking at the size distribution of grains, through the use of the so-called granulometry [18, 43]. Following this idea, it is proposed in [31] to use the *pattern spectrum* of images (roughly, the derivative of granulometry) to index grayscale images. In the framework of remote sensing imaging, [5, 13] have proposed, in view of the classification of satellite images, to compute size distributions (called *derivative morphological profiles*) at each pixel. A closely related approach to the extraction of urban area was previously proposed in [27]. In these works, the proposed local features contain some information about scale. This information is well localized thanks to the use of connected morphological filtering of features.

In this paper, our purpose is twofold. First, we introduce a method to compute a local scale measure following the general approach of granulometries. Contrarily to previous morphological approaches, the proposed method is autodual (is invariant to contrast reversal) and does not necessitate any structuring element. Second, we establish theoretical links between the variational and morphological approaches to scale computation.

The first contribution is a method to compute a local scale measure (defining a characteristic scale at each pixel of a digital image) by using the topographic map [11] of the image. The main idea is that, for each pixel, we associate the scale of the most significant structure containing it. The definition of this structure relies on the topographic map, which is made of the connected components of the boundaries of level sets of the image. More precisely, we make use of the digital image transform presented in [38], an efficient tool to compute the topographic map, representing an image by a hierarchical structure (an inclusion tree) of shapes. From this tree we

search, for a given pixel, the most contrasted shape containing it and associate the scale of this shape with the pixel. In order to deal with optical blur that is inherent in satellite images, a new definition of shape is proposed. The underlying idea of this definition is to group level lines corresponding to the same structure in the image. This grouping is made possible under the assumption that the optical blur is constant over the image, an assumption that makes sense for remote sensing images but would be wrong for natural, everyday life images. Experimental results suggest that the scale measure is a pertinent feature for remote sensing image classification. In particular, the use of the topographic map yields very satisfactory spatial accuracy of the scale measure, especially when compared with linear methods (see section 6). Our approach is also compared with variational and morphological methods to estimate local scales.

The second contribution of the paper is to shed light on the relationship between morphological and variational approaches to the computation of local scales. In particular, it is shown that, under some hypotheses, approaches relying on the total variation flow or on the Rudin–Osher–Fatemi model amount to defining scale as a weighted average of the size of shapes containing each pixel. In this framework, the term "size" refers to the ratio of the area to the perimeter. This result is also investigated numerically on synthetic and real images.

The paper is organized as follows. In section 2, we present the topographic map of an image, the fundamental tool on which our work is based. Then, in section 3, we explain how an accurate scale definition can be given by using the topographic map. In section 4, alternative variational definitions of scale are given, and the link between these approaches and ours is investigated. In section 5, we illustrate the method with numerical examples, both on synthetic and real images. In section 6, we provide experimental comparisons between the proposed local scale measure, linear approaches, variational methods, a possible use of the so-called maximally stable extremal regions, and methods relying on the pattern spectrum. In section 7, we propose some extensions to the proposed definition of scale.

2. Topographic map. In this section, we introduce the main tool to be used in this paper, the topographic map of an image as introduced and studied in [11, 12, 37, 38]. It is made of the connected components of the topological boundaries of level sets, thereafter called *level lines*. It gives a complete representation of the image, invariant to local contrast changes. The topographic map has a tree structure, inherited from inclusion properties of level sets.

For an image $u : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$, its upper and lower level sets are defined, respectively, as

$$\Psi_\lambda = \{x \in \Omega,\, u(x) \geq \lambda\} \quad \text{and} \quad \Psi^\lambda = \{x \in \Omega,\, u(x) \leq \lambda\}$$

for $\lambda \in \mathbb{R}$. Observe that u can be reconstructed using the upper level sets by the following formula:

$$u(x) = \sup\{\lambda \in \mathbb{R} : x \in \Psi_\lambda\}.$$

Lower level sets also permit us to reconstruct u. Moreover, these sets are globally invariant with respect to contrast changes. Each of these family, upper sets on the one hand and lower sets on the other hand has a tree structure with respect to inclusion. Several authors [42, 11, 12, 19] have proposed the connected components of level sets as an efficient way to represent images. This representation enjoys the same contrast invariant and hierarchical structure as level sets representations while being local,

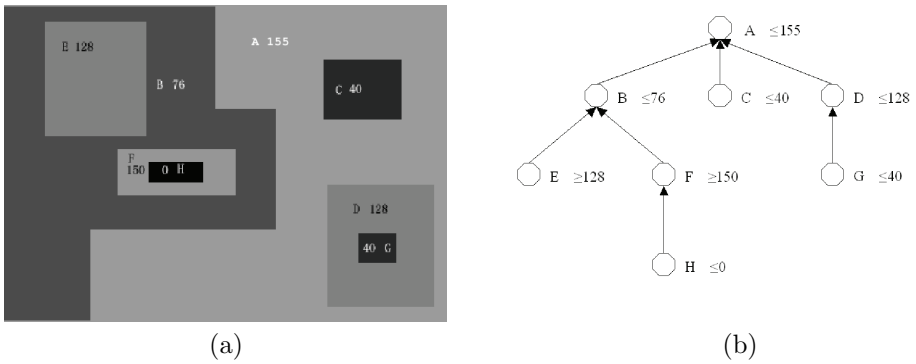


FIG. 1. *Example of FLST:* (a) *Synthetic image;* (b) *inclusion tree obtained with FLST.*

thus in particular robust to occlusions or local contrast changes. Again, connected components of upper sets have a tree structure, as well as connected components of lower sets. In order to merge both trees, Monasse and Guichard [37, 38] introduced the concept of tree of shapes. It relies on the concept of *shape*. A shape is defined as the union of a connected component of an upper or lower set and its holes. The holes of a set A are defined as the connected components of the complementary set of A which do not intersect with the boundary of Ω. It can be shown that these shapes have a tree structure. Under some regularity conditions, this tree of shapes is the same as the tree made of the connected components of topological boundaries of upper (or lower) sets of the image, the topographic map. In what follows, we use the term *topographic map* for this tree. An important property of this representation is its autoduality, that is, its invariance with respect to the operation $u \to -u$ (or, equivalently, to $u \to C - u$, with C a constant). This implies in practice that light and dark objects are treated in the same way. As we will see in the following, it is important for the purpose of this paper to deal with a single tree instead of two separated ones, since this will enable us to associate a unique contrasted shape with each pixel.

In order to apply the preceding definition to a discrete image I, it is convenient to consider a continuous image (defined at each point of its domain) associated with I. Two straightforward choices are piecewise constant interpolation (each pixel corresponds to a constant value) and bilinear interpolation. Of course, bilinear interpolation is preferable in order to obtain regular lines, free of pixelization effects. However, this also implies longer computation time. Moreover, we observed that for the purpose of this work, namely local scale computation from the topographic map, bilinear interpolation was unnecessarily complex. In the remainder of the paper, level lines are computed on piecewise constant interpolations of discrete images. Thanks to its tree structure, the topographic map of an image can be efficiently computed using the fast level set transform (FLST) algorithm presented in [38]. Figure 1 shows the result obtained with this algorithm on a synthetic image. Notice that another implementation has also been proposed in [45].

We end this section by giving some notation for the attributes of shapes that are used in what follows. For a pixel x of an image u, we denote by $\{f_i(x)\}_{i=1,\dots,N(x)}$ the set of shapes that contain x, $N(x)$ being the number of shapes containing x and the shapes being ordered so that $f_i(x) \subset f_{i+1}(x)$. For the sake of clarity, we will omit the x dependency when it is not necessary. For each shape f_i, we define $|f_i|$ its area,

$P(f_i)$ its perimeter, and $I(f_i)$ its gray level value, that is, the gray level of the upper or lower level set defining f_i. The contrast of the shape f_i is then defined as the absolute value of the difference between the gray level values associated, respectively, with f_i and f_{i+1}:

$$C(f_i) = |I(f_{i+1}) - I(f_i)|. \tag{2.1}$$

3. Scales of an image. The basic idea of this paper is to associate with each pixel a shape (i.e., a node in the FLST) from which its scale can be computed. Such shapes are obtained by filtering the topographic map. Basically, shapes are recursively grouped in order to account for structures present in the image and the most contrasted groups are kept, as detailed in section 3.1. Shape grouping is defined by taking advantage of the particular structure of satellite images, for which the blur is constant over the image and depends only on the (usually known) point spread function of the acquisition device. This fact is investigated in section 3.2. Section 3.3 then defines the local scale associated with each pixel.

3.1. Most contrasted shape. The aim of this section is to propose a method to extract the most contrasted shapes from the topographic map. A notion of meaningful shape was introduced in [14] and consists roughly in picking the shapes whose perimeters and contrasts are very unlikely in a noise image, the decision being made using an *a contrario* method. In our case, the objective is different. First, we intend to associate a shape with each pixel. Second, the selection criterion should not rely on the size of shapes because such a process would bias the following definition of characteristic scale. In this section, shapes are selected using solely some contrast information. In section 7.1, we briefly investigate the use of both contrast and regularity in the shape selection process.

Using the definition of contrast given by (2.1), the simplest definition of the most contrasted shape at a pixel would be to define

$$\hat{f}(x) = f_{\arg\max_{i=0,\dots,N(x)} C(f_i)}. \tag{3.1}$$

For instance, in Figure 1 the most contrasted shape associated with pixels of shape C is C, and the most contrasted shape associated with pixels of shape E is B. However, this definition is not satisfactory for real images. Indeed, the definition of contrast by formula (2.1) corresponds to the contrast of a given binary structure under the assumption that only one line is associated with the boundary of this structure. This is, for instance, the case for the synthetic image of Figure 1. Now, in a natural image, the contours of objects are always blurred by the acquisition process. As a consequence, a discrete image being quantified, a contour is in fact associated with a set of level lines. In practice, the contrast of each line is often equal to one. Therefore, the choice of the most contrasted shape using (3.1) can be ambiguous at best or even completely meaningless in the presence of strong noise, as will be illustrated in the experimental section.

A solution to this problem would be to compute the contrast of a line in a neighborhood and to select the most meaningful line along a monotone branch of the tree (see [14]). In the present work, we choose to group lines corresponding to a single structure by using a simple model of blur. To do so, we recursively sum up the contrasts of shapes f_i and f_{i+1} such that

$$|f_{i+1}| - |f_i| < \lambda P(f_i), \tag{3.2}$$

where λ is a constant.

Before explaining in more detail how this idea is implemented, let us observe that this criterion relies on the hypothesis that the level lines corresponding to a blurred contour are regularly separated by a distance λ. Let us remark that in the one-dimensional case, this hypothesis boils down to assuming that the blur is due to a kernel equal to a characteristic function. This fact will be detailed in Appendix A. If the kernel has a more complicated shape (e.g., Gaussian), then the distance between two lines depends on the contrast of the contour, the amount of blurring, and the quantification of the image. We will see in Appendix A that by assuming a minimum contrast of the contours and by knowing the width of the blurring kernel, it is possible to deduce an upper bound for the value of λ. Let us remark that the hypothesis of a constant blurring kernel for the whole image is realistic in the case of satellite images but is less consistent in the case of natural images for which the blur associated with an object depends on its position (see [23]).

We now explain how the most contrasted shape of a pixel is computed. For a given pixel x, we consider the shapes $(f_i(x))_{i=0,\dots,N(x)}$ containing it. The contrast of $f_0(x)$ is simply $C(f_0(x))$, as defined by (2.1). Then, for $i \geq 1$, we recursively define the cumulated contrast as

$$\bar{C}_x(f_i(x)) = \begin{cases} \bar{C}_x(f_{i-1}(x)) + C(f_i(x)) & \text{if } |f_i(x)| - |f_{i-1}(x)| < \lambda P(f_{i-1}(x)), \\ C(f_i(x)) & \text{otherwise.} \end{cases}$$

In other words, for a given pixel x, the cumulated contrast of the shape $f_i(x)$ (for $i > 0$) is defined as

$$\bar{C}_x(f_i(x)) = \sum_{k=a(i)}^{i} C(f_k(x)), \tag{3.3}$$

where, for all $i \leq N(x)$,

$$a(i) = \min\{j \geq 0 : j < i \text{ and } \forall k = j+1, \dots, i,\ |f_k(x)| - |f_{k-1}(x)| \leq \lambda P(f_{k-1}(x))\}.$$

If $a(i)$ is not defined (that is, if (3.2) is not satisfied), then $\bar{C}_x(f_i(x)) = C(f_i(x))$. The cumulated contrast of $f_i(x)$ is therefore obtained by adding the contrasts of close enough level lines. These lines usually correspond to the same contour in the image, as investigated in the next section. The most contrasted shape associated with x is then defined as

$$\bar{f}(x) = f_{\arg\max_i \bar{C}_x(f_i(x))}. \tag{3.4}$$

In the case when the maximum is reached at more than one index, then the smaller one is chosen. We conclude this section by noticing that a method to group level lines relying on criteria similar to formula (3.2) (but using no perimeter information) was proposed in [36] as an efficient alternative to shock filters, in the framework of image restoration.

3.2. Level lines, edges, and blur. In this section, we investigate the validity of the use of formula (3.2) for grouping lines corresponding to a single edge. More precisely, we are interested in finding values of λ for which this formula holds true for consecutive level lines along an edge.

Let f_i and f_{i+1} be two consecutive shapes corresponding to the same object, an object being defined as a constant times the indicator function of some set smoothed

by the acquisition kernel. Writing q for the quantization step and neglecting sampling, we have, for some gray level l, $f_i = \partial\Psi_l = \{x \in \Omega \ / \ u(x) = l\}$ and $f_{i+1} = \partial\Psi_{l+q}$ (where ∂ stands for the topological boundary). Now, as noticed in [16], if $x(s)$ is a parameterization of $\partial\Psi_l$, then $\partial\Psi_{l+q}$ can be approximated, for small q, by

$$\tilde{x}(s) = x(s) + q\frac{\nabla u}{|\nabla u|^2}. \tag{3.5}$$

If we now assume that $|\nabla u| \geq C$ for some $C > 0$, then $f_{i+1} \subset f_i \oplus D(qC^{-1})$, where $\oplus$ stands for the Minkowski[1] addition and $D(r)$ is a disk of radius r centered at the origin. On the other hand, assuming that f_i is a convex set, the area of $f_i \oplus D(qC^{-1})$ is (Steiner formula [49])

$$\left|f_i \oplus D\left(\frac{q}{C}\right)\right| = |f_i| + \pi\left(\frac{q}{C}\right)^2 + \frac{q}{C}P(f_i).$$

This suggests that formula (3.2) enables one to group level lines corresponding to the same edge as soon as $\lambda > qC^{-1}$. The values of C ensuring that $|\nabla u| \geq C$ depend on both the contrast of the object and the blur kernel. On the other hand, the value of λ should not be too large in order not to group level lines corresponding to different edges. For a blur kernel of width σ and an object of contrast G, the minimum value of C that can be chosen is approximately equal to $G\sigma^{-1}$, so that λ can roughly be chosen to be $q\sigma G^{-1}$. In Appendix A, this result is investigated further for two blur kernels (indicator function and Gaussian kernel) in the one-dimensional case. For typical values of q, G, and σ, we can use a value of $\lambda = 1$ (in pixels) to group lines corresponding to the same object.

3.3. Scale definition. Recall that the most contrasted shape at each pixel is defined by formula (3.4). In order to define the scale at each pixel, we choose to consider as the final shape associated with x the shape $\bar{f}(x)$ minus the most contrasted shapes embedded inside itself. Observe that these shapes provide a partition of the image domain. The main reason for this choice is as follows. Let us recall indeed that a shape is a connected component of a level set whose holes have been filled in. In Figure 1, the shape F contains the pixels of the shape H. Intuitively, this shape corresponds to the objects in the image minus the objects that occlude them. Now in satellite images occlusion is not preponderant, and contrasted shapes containing other contrasted shapes often correspond to road or river networks. To accurately represent such structures, we eventually decide to define the most contrasted shape associated with a pixel x as

$$\tilde{f}(x) = \bar{f}(x) \setminus \bigcup_{\bar{f}(y) \subsetneq \bar{f}(x)} \bar{f}(y), \tag{3.6}$$

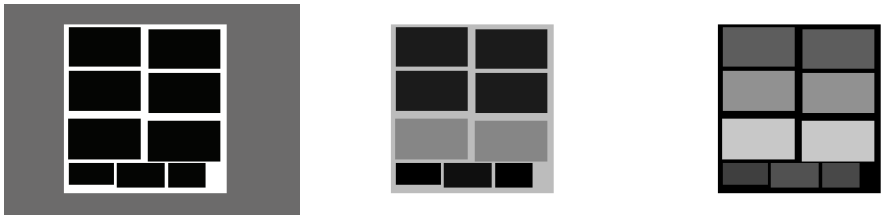
i.e., the shape $\bar{f}(x)$ minus the most contrasted shapes strictly embedded in it. Other choices would be possible in the framework of other applications.

The last step consists in associating a scale $E(x)$ with the pixel x. We choose to define the scale as

$$E(x) = |\tilde{f}(x)|/P(\tilde{f}(x)) \tag{3.7}$$

so that the geometry of $\tilde{f}(x)$ is taken into account. In particular, long and thin shapes (e.g., the roads) correspond to relatively small scales, even though their area can be

[1]Let A and B be two sets; $A \oplus B = \{x + y,\ x \in A,\ y \in B\}$.



(a) (b) (c)

FIG. 2. (a) *Synthetic image;* (b) *image of scale using the area;* (c) *image of scale using* (3.7). *Notice that in* (b) *or* (c)*, the darker the gray level, the smaller the shape.*

quite large. Figure 2 shows a synthetic example of scale computation illustrating this fact. Let us remark that if scale is defined as the area, then the scale of the thin structure around the rectangles is much larger than the ones of the rectangles. In the next section, this definition of scale is compared to related notions that have been proposed in a variational framework for image analysis. Then, in section 5, we validate definition (3.7) with various numerical experiments.

4. Link with variational definitions of scales. Two definitions of local scale related to the total variation of images have been recently proposed in the literature and are presented in section 4.1. In section 4.2 a geometrical interpretation of these definitions is given and the link between them is clarified. These are actually closely connected with the scale definition of this paper, as investigated in section 4.3. Roughly speaking, methods relying on the total variation define the scale at each pixel as a weighted average over many shapes of the ratio *area/perimeter*, whereas the approach proposed in this paper defines scale using the same ratio but relies on only one shape per pixel.

4.1. Variational definitions of scale.

4.1.1. Definition based on total variation regularization. Strong and Chan have proposed in [51] to define the scales in an image by using the Rudin–Osher–Fatemi (ROF) model [41]. Recall that the ROF model (or total variation regularization) consists, given an image f, in finding the solution u of

$$\inf_u \left(\int |Du| + \frac{1}{2T} \|f-u\|_{L^2}^2 \right). \tag{4.1}$$

It is shown in [51] that if the scale of a set E is defined as $\frac{|E|}{P(E)}$ (i.e., its area, divided by its perimeter, as done in (3.7)) and if f is a binary image of a disk, then the intensity change between u and f inside this disk is inversely proportional to its scale, i.e.,

$$\delta = \frac{T}{\text{scale}}. \tag{4.2}$$

Therefore, the idea in [51] to define scales in an image is to use the gray level difference at each pixel between u and f. The scale at each pixel x is defined as

$$\text{scale}(x) = T.|u(x) - f(x)|^{-1}. \tag{4.3}$$

Observe that, in general, this definition of scale depends on the parameter T.

It was later noticed in [50] that in fact this notion of *scale* is related to the polar seminorm associated with the total variation, the so-called G norm introduced by Meyer in [34]. Indeed, a possible definition of the G norm of an image f with zero mean is [50]

$$\|f\|_G = \sup_{E \subset \Omega} \frac{\int_E f}{P(E, \Omega)}, \tag{4.4}$$

where $P(E)$ stands for the perimeter of E, as defined, for instance, in [2]. The G norm of an image can therefore be seen as an area divided by a perimeter. These relations have been used in [50] to propose a variant of the ROF model where the user gives as input of the algorithm the minimal scale in the original image f which is to be preserved (smaller scales are considered as noise and therefore wiped out).

4.1.2. Definition based on the total variation diffusion. The properties of total variation diffusion have also been used in [7, 8] to introduce a definition of scale in images. Let us recall that the solution u of the total variation diffusion satisfies

$$\begin{cases} u(.,0) = f, \\ \frac{\partial u}{\partial t} = \operatorname{div}\left(\frac{Du}{|Du|}\right). \end{cases} \tag{4.5}$$

In [48], the authors have proved the equivalence for one-dimensional signals of total variation regularization (ROF model) and total variation diffusion. They have derived the same type of results as in [51] (where the considered functions were two-dimensional radially symmetric signals). In particular, when using the total variation diffusion on an image, a constant region evolves with speed $2/m$, where m is the number of pixels in the considered region.

Therefore in [8] the authors have proposed to define the scale m of a region (in dimension 1) as

$$\frac{1}{m} = \frac{1}{2} \frac{\int_0^T |\partial_t u| \, dt}{T}, \tag{4.6}$$

where T is the evolution time of the total variation diffusion.

In the same paper, the following definition of scale m is then proposed for two-dimensional images:

$$\frac{1}{m} = \frac{\int_0^T |\partial_t u| \, dt}{T}. \tag{4.7}$$

See [8] for practical details about this definition. Combining this scale feature with other orientations and magnitudes features, good segmentation results of textured images are reported in [8].

4.2. Equivalence and geometrical interpretation. As already said, results from [50] provide a geometrical interpretation of the scale definition given by formula (4.3). On the other hand, equivalence results between total variation regularization (formula (4.1)) and total variation flow (formula (4.5)) are provided in [48].

We now summarize some recent mathematical results in order to further investigate the definitions of scale given by formulas (4.3) and (4.7), as well as to clarify the link with the definition of scale given in the present paper, formula (3.7). These results have been proved by Caselles and his collaborators in a series of papers [3, 4, 10, 1]. In particular, it is shown that if an image f is the characteristic function of a convex set C, i.e., $f = \mathbf{1}_C$, then total variation regularization is equivalent to total variation

flow. In both cases, the evolution speed of a convex body C is $\frac{P(E)}{|E|}$, where E is the Cheeger set of C (see [1]); that is, E is a solution of $\min_{K\subset C}\frac{P(K)}{|K|}$. The set C is said to be a Cheeger set in itself if it itself is a solution to this minimization problem. In dimension 2, a necessary and sufficient condition for C to be a Cheeger set in itself is that C is convex and the curvature of C is smaller than $\frac{P(C)}{|C|}$. A disk is thus a Cheeger set in itself.

Assume that $f=\mathbf{1}_C$, with C a Cheeger set in itself. Then it is shown in [4] that the solution of the total variation flow, (4.5), or equivalently of the total variation regularization, (4.1), is given by

$$u(x,T)=\max\left(\left(1-T\frac{P(C)}{|C|}\right),0\right)\mathbf{1}_C. \tag{4.8}$$

The evolution speed of C is thus $\frac{P(C)}{|C|}$ (and in the case when C is a disk, this is what was proved by Chan and Strong in [51]).

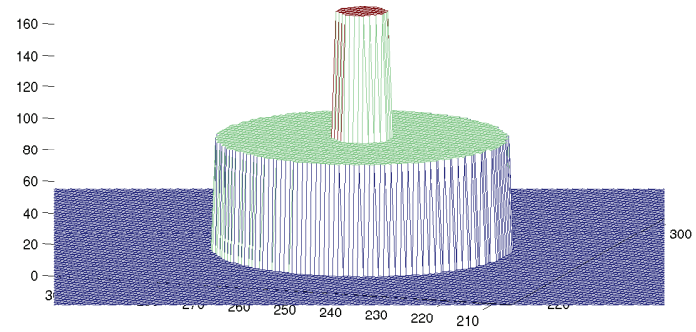
As a consequence, in the case when the considered image f is the characteristic function of a Cheeger set, then both definitions of scale (4.2) and (4.7) are equivalent. Notice that in this particular case these two definitions of scale are also equivalent to the one proposed in this paper (3.7). With all three definitions (4.2), (4.7), and (3.7), the *scale* of C is equal to $\frac{|C|}{P(C)}$. From now on this will be our definition of the scale of a Cheeger set. Of course, in the case of more complicated images, the equivalence no longer holds. However, in the case of *towers* of Cheeger sets (that is, a sequence of embedded Cheeger sets), provided that some geometric conditions are satisfied [3], then formula (4.8) can be generalized (both for total variation flow and regularization), and thus definitions (4.2) and (4.7) are still equivalent. In particular, this is the case for a *tower* of circles [3] (with no further assumptions).

4.3. Relationship with our definition of scale. We have already seen that in the case when the image is made of a single object, $f=\mathbf{1}_C$, with C a Cheeger set in itself, then the two variational scale definitions (4.2) and (4.7) are equivalent to the scale definition introduced in this paper, formula (3.7). In these cases, the *scale* of C is equal to $\frac{|C|}{P(C)}$.

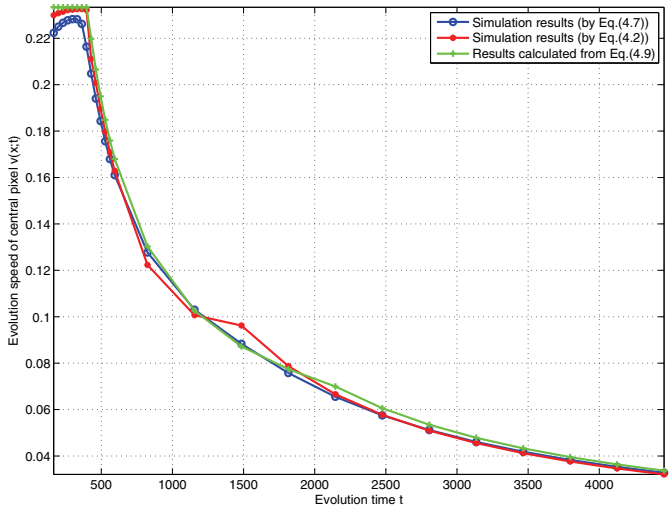
We now consider an image f with two objects. For the sake of clarity, we assume that f is made of two circles $\mathcal{C}_1$ and $\mathcal{C}_2$ with radiuses $r_1<r_2$ and with the same center (see Figure 3(a)). That is, the image has constant gray level values G_1, G_2, and G_3, respectively, on $\mathcal{C}_1$, $\mathcal{C}_2\backslash\mathcal{C}_1$, and the background. We also assume that $G_1\geq G_2\geq G_3$. We denote the contrast of objects by $\delta_1=G_1-G_2$ and $\delta_2=G_2-G_3$. We finally assume that $\delta_2\geq\frac{r_1}{r_2-r_1}\delta_1$.

Notice that from what we explained in the previous subsection, definitions (4.2) and (4.7) are equivalent in this case (*tower* of circles). We therefore consider only (4.7) to compute the scale. In order to simplify the computations, we assume that the image size is sufficiently large, so that we can ignore the evolution of the background. As soon as we start evolving (4.5), $\mathcal{C}_2$ goes down with speed $1/r_2$ and $\mathcal{C}_1$ with speed $1/r_1+1/r_2$. At $t_1=r_1\delta_1$, the small circle $\mathcal{C}_1$ disappears into $\mathcal{C}_2$, and $\mathcal{C}_2$ keeps going down at speed $1/r_2$ until $t_2=r_2\delta_2$, when it also disappears. Thus, the average evolution speed $v(x;t)$ of a pixel x in $\mathcal{C}_1$ is the following:

$$v(x;t)=\begin{cases}\frac{1}{r_1}+\frac{1}{r_2}, & t\leq t_1,\\ \frac{\delta_1+t/r_2}{t}=\frac{1}{r_2}+\frac{t_1}{t}\frac{1}{r_1}, & t_1\leq t\leq t_2,\\ \frac{1}{t}\left(\delta_1+\delta_2\right)=\frac{t_1}{t}\frac{1}{r_1}+\frac{t_2}{t}\frac{1}{r_2}, & t\geq t_2.\end{cases} \tag{4.9}$$



(a)

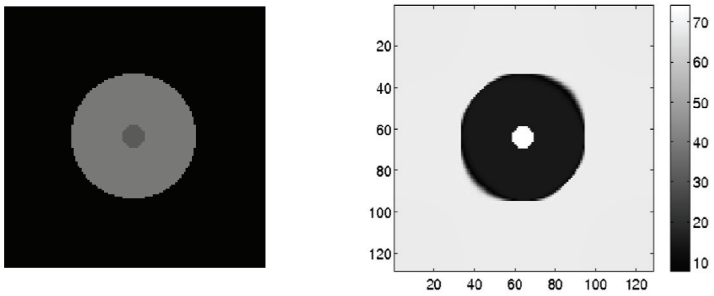


(b)

FIG. 3. (a) *Synthetic image (size* 512×512*) with two concentric circles (radiuses* $r_1 = 5$*,* $r_2 = 30$*,* $\delta_1 = 80$*,* $\delta_2 = 70$*);* (b) *graph of the average evolution speed* $v(x;t)$ *of the center pixel as a function of total evolution time* t*. The two other curves are the ones of the inverse of the scale computed with* (4.2) [51] *and* (4.7) [8]*. The three curves are almost equal: definitions* (4.2) *and* (4.7) *are equivalent in this case.*

Notice that in this case, the average speed $v(x;t)$ given by formula (4.9) is exactly the inverse of the scale at point x both with definitions (4.2) and (4.7). In Figure 3(b), we display the scale computed at the center pixel on the image using definitions (4.2) and (4.7) with finite difference schemes, as well as the scale obtained with formula (4.9). The equivalence is numerically confirmed in this case: the three graphs are almost equal.

Now, observe that the scale given by (4.9), which is the inverse of the average speed v, is a weighted average of the scales of the two shapes (disks in this case) containing the center pixel, r_1 and r_2. Therefore, in this simple case, both variational definitions of scale given by formula (4.2) or (4.7) yield a scale that is a weighted average of the shapes containing a pixel. Observe that different hypotheses on the gray level values of the image would also lead to some weighted averages similar to formula (4.9).



(a) (b)

FIG. 4. (a) *Synthetic image with two concentric circles (contrarily to Figure* 3(a)*, the gray level value of the small circle is smaller than the one of the large one);* (b) *corresponding scale map computed with definition* (4.2) [51] *(with* $t = 545$*). We can observe that since the evolution of the gray values of the pixels in the small circle is not monotonous the scale values computed on these pixels are not correct.*

In contrast, with the definition of scale introduced in this paper (formula (3.7)), the scale at each pixel is the scale of *only one shape*, the most contrasted shape containing this pixel. In this simple case, the scale of the central pixel is r_1 or r_2, depending on whether or not $\delta_1 > \delta_2$. In the example of Figure 3(a), the scale of a pixel of $\mathcal{C}_1$ is r_1.

Two consequences can be drawn from this example. First, and anticipating the experimental section, this explains why the definition of scale introduced in this paper is spatially more accurate than variational definitions from [8] or [51]. Moreover, this also explains an observation from [8], where it was noticed that the definition given by (4.7) works better when using an integration time T smaller than $\tilde{T}$, the time needed to obtain a constant image under total variation flow. The authors of [8] explain that they observe a bias towards large scales when using $\tilde{T}$: this bias is due to the averaging of scales over all shapes that contain the pixels. By using a smaller T, one avoids averaging over very large shapes. As one can notice in the graph of Figure 3(b), the scale value obtained with definitions (4.2) or (4.7) is far from being constant: the choice of the parameter T in (4.2) or (4.7) has a strong impact on the resulting scale.

4.4. Difference between definitions (4.2) and (4.7). We have just seen examples when definitions (4.2) and (4.7) are equivalent. However, this equivalence is not true in general. The equivalence is valid as long as the evolution of the gray level of the pixels is monotone. For instance, in Figure 3, the gray value of the central pixel is always decreasing during the evolution process of (4.1) and (4.5). But when the evolution is no longer monotone, definitions (4.2) and (4.7) are no longer equivalent, and (4.2) fails to give a correct scale.

Consider, for instance, the case of Figure 4(a) (contrarily to Figure 3(a), the gray level value of the small circle is smaller than the one of the large circle). During the evolution process, the gray values of the pixels in the small circle first increase, until the small circle merges with the large one. Afterwards, the gray values of these pixels decrease, until they reach the average gray level value of the original image. We can therefore choose a time t ($t \approx 545$ for the case of Figure 4(a)) for which the gray level value of the small circle is almost the same as its original value (before evolution): with definition (4.2), this implies a large scale. In Figure 4(b), we show the scale map

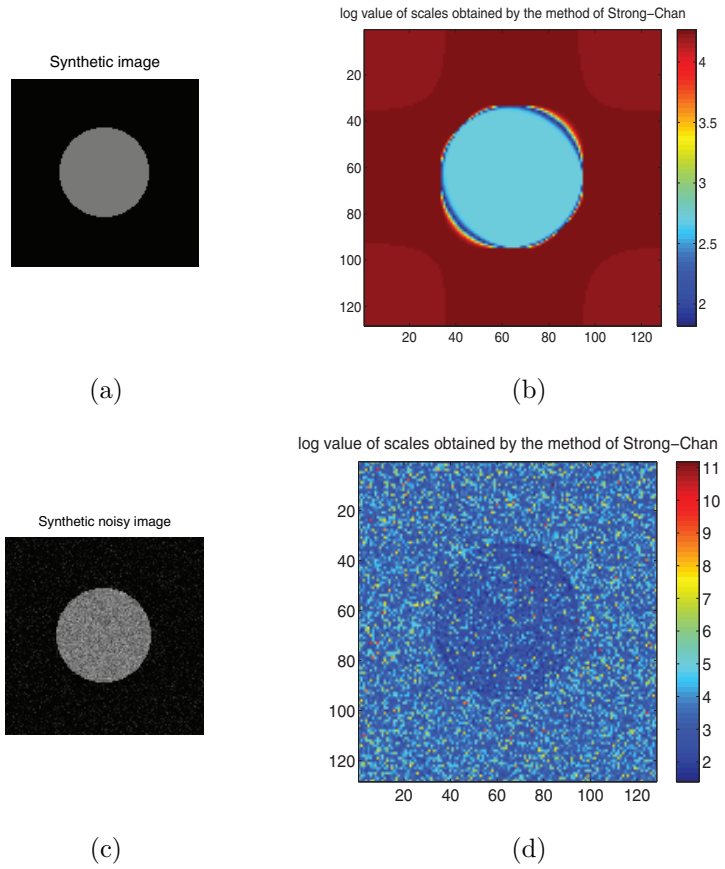


FIG. 5. (a) *Image of a circle without noise;* (b) *scale map corresponding to* (a) *computed with* (4.2). *For better visualization, the logarithmic values of the scales are presented.* (c) *Image of a circle with Gaussian noise of standard deviation* 20*;* (d) *scale map corresponding to* (c) *computed with* (4.2). *For the visualization issue, the logarithmic values of the scales are presented. This illustrates the sensitivity to noise of definition* (4.2) [51].

corresponding to Figure 4(a), computed using formula (4.2). It can be noticed that the larger scale is the one of the small circle (it is even larger than the one of the background). Thanks to the absolute value used in definition (4.7), (4.7) does not suffer from this problem when the evolution is not monotonous. Notice that with our definition of scale, formula (3.7), this type of problem does not occur either.

The case of Figure 4(a) is not that unusual in an image: indeed, as soon as an image is corrupted by some noise, then such cases occur. In Figure 5, we show the scale map computed with (4.2) on a noisy synthetic image. We can observe that even if the noise level is rather low, the scale map is not correct: this is due to the fact that many pixels have a nonmonotone behavior.

This is the reason why in section 6 we compare our method with the one relying on formula (4.7) (and not formula (4.2)).

5. Experiments. In this section, we perform experiments on both synthetic and real images. We first apply the method described in section 3.1, aiming at extracting the most contrasted shape at each pixel, to an image made of a single blurred and noisy shape. We then compute scale maps on various satellite images with high or

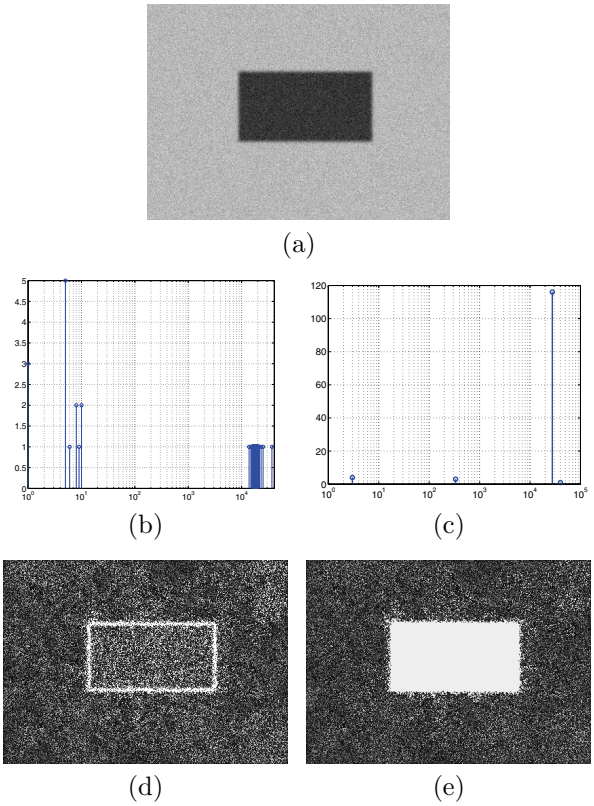


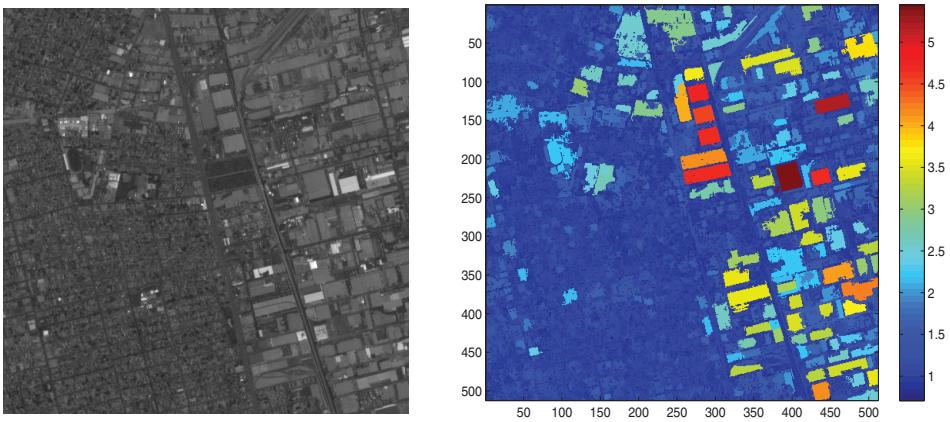
FIG. 6. (a) *Synthetic image of a rectangle degraded by blur and additive noise;* (b) *and* (c) *values of the contrasts (*(b): *plain contrast,* (c): *cumulated contrast) of the shapes containing the central pixel as a function of the area of the shape;* (d) *and* (e) *corresponding area map (*(d): *using plain contrast,* (e): *using cumulated contrast) (for visualization, the gray level values correspond to the logarithm of the area).*

very high resolutions. Our implementations make use of the FLST available in the free image processing library *Megawave*2.[2] Computation time is image dependent, but a typical value is of the order of 1 to 5 seconds for a 1000×1000 image on a Pentium 4 at 3.2 GHz.

5.1. Synthetic images. In this section, we perform an experiment on a simple synthetic image to demonstrate the utility of the cumulated contrast of shapes, as defined by formula (3.3), to extract the most contrasted shape at each pixel. We consider a rectangle blurred by a Gaussian kernel with standard deviation 3 and corrupted with a Gaussian noise with standard deviation 30 ($SNR \approx 17$); see Figure 6(a).

We first apply (3.1) to extract the most contrasted shapes as displayed in Figure 6(d). For visualization, the gray level values in this figure correspond to the areas of the shapes. Figure 6(b) shows the contrasts (as defined by formula (2.1)) of the shapes containing the center of the rectangle. The horizontal axis corresponds to the areas of the shapes and the vertical axis corresponds to the contrast values. We then perform the same experiment using formula (3.3) to define the (cumulated) contrast of shapes. Corresponding results are displayed in Figures 6(c) and (e). It can be ob-

[2]http://www.cmla.ens-cachan.fr/Cmla/Megawave/



(a) (b)

FIG. 7. (a) *Image of Los Angeles, SPOT5 (*5m*) ©CNES;* (b) *corresponding scale map. Notice the spatial accuracy of the method.*

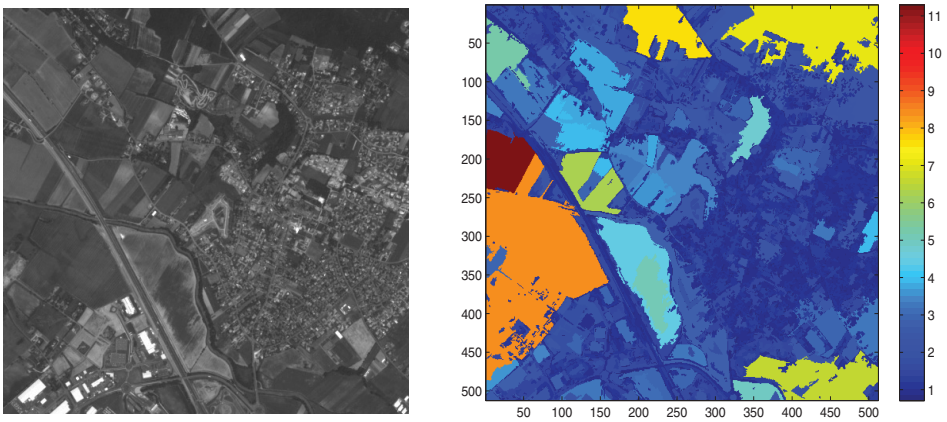
served that many contrasted shapes computed using formula (3.1) correspond to noisy structures, since the contrast of the rectangle is then spread over many shapes. In comparison, using the definition of cumulated contrast following formula (3.3) enables us to correctly retrieve the blurred rectangle.

5.2. Satellite images. In this section, we compute scale maps on real satellite images using the approach presented in section 3. We first consider SPOT5 HMA images that have a resolution of 5m. Most contrasted shapes are extracted using formula (3.4). We choose to use a value of $\lambda = 1$. As explained in section 3.2, this is reasonable for typical values of blur and contrast. Appendix A makes this observation more precise in the case of a one-dimensional Gaussian kernel, which constitutes an approximation of the point spread function (PSF) of satellite captors such as those of SPOT5 HMA; see [29].

In Figure 7 a SPOT5 HMA image of Los Angeles (mainly urban area) together with its computed scale map is displayed. It can be observed that the computed scales are spatially very accurate (e.g., at the edges of buildings and warehouses). Moreover, these scales are in good qualitative agreement with the sizes of structures of the image (large for warehouses and smaller for individual houses on the left). One also observes that computed scales are largely constant inside objects. Notice also that the road network is attributed relatively small scales, in agreement with formula (3.7). The same observations can be made on the SPOT5 HMA image of an urban area near Toulouse displayed in Figure 8(b). These experiments also suggest that the computed local scales are pertinent features for segmenting images. This fact will be further investigated in section 7.2.

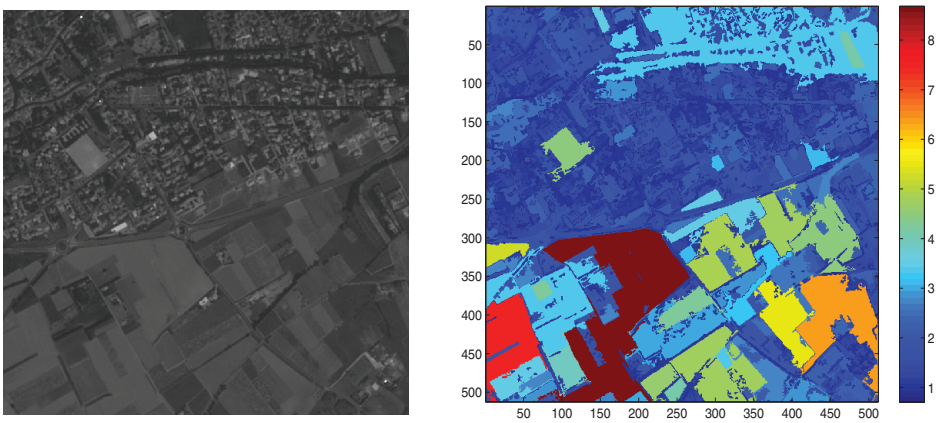
Let us also notice that the resolution of the images being known and the computed scale corresponding to the size of a geometrical element, this scale is directly related to the true size of the structures in the image. Such an accurate result cannot be obtained when using a definition of scale relying on a linear scale space approach, even though some approximations in this direction are possible; see [30].

Figure 9 shows the scale map obtained from a SPOT5 THR image [24] of Mar-



(a) (b)

FIG. 8. (a) *Image of Toulouse, SPOT*5 *(*5m*)* ©*CNES;* (b) *corresponding scale map.*

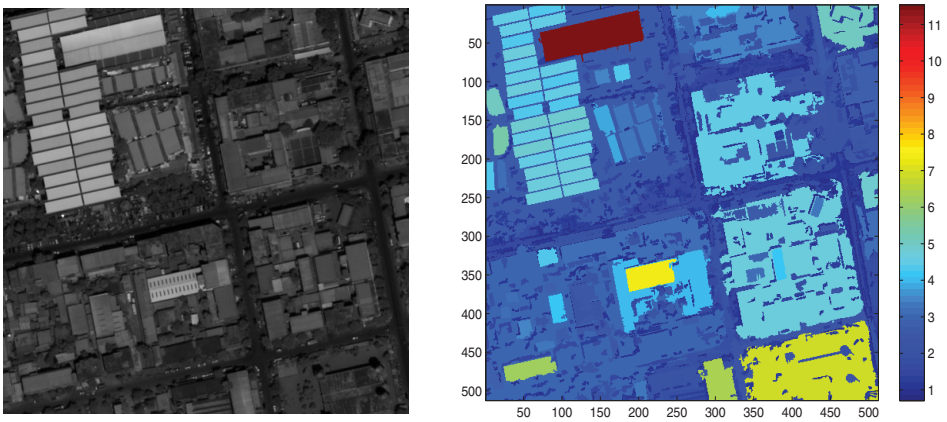


(a) (b)

FIG. 9. (a) *Image of Marseille, SPOT*5 *(*2.5m, 512×512*)* ©*CNES;* (b) *corresponding scale map.*

seille with resolution 2.5m. In this clever imaging system, two images captured by two different CCD line arrays are interpolated to generate the high resolution image. The PSFs of SPOT5 THR images are much more complicated than those of HMA images and cannot be modeled in a simple way, for instance using a Gaussian kernel. Therefore, the computations of Appendix A cannot be applied directly. However, the slope of these PSFs is sharp enough so that a value of $\lambda = 1$ still allows us to group shapes using the scheme presented in section 3.1, as explained in section 3.2.

Finally, we present the scale map for a QuickBird Panchromatic image with a resolution of 0.6m, taken at Ouagadougou. In [29], is has been shown that the PSFs of QuickBird satellites can also be approximated by a Gaussian kernel. Again we use a value of $\lambda = 1$. Considering the high resolution of QuickBird images,



(a) (b)

FIG. 10. (a) *Image of Ouagadougou, QuickBird (*0.6m, 512×512*);* (b) *corresponding scale map.*

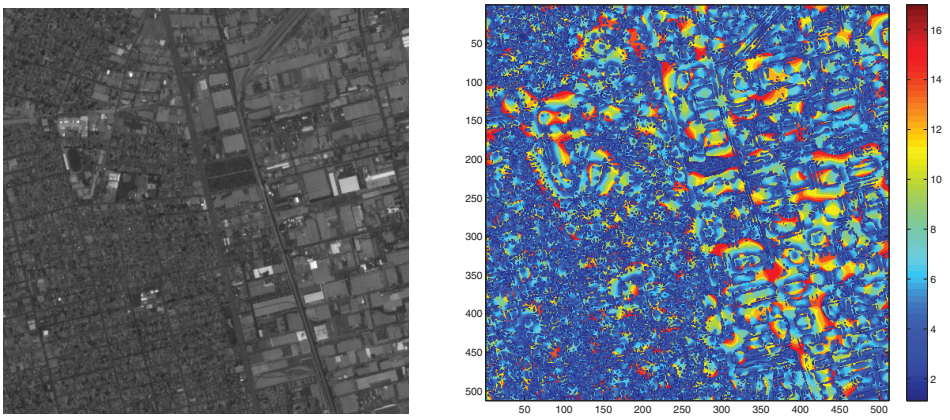
shapes smaller than 16 pixels are not taken into consideration. This is equivalent to the application of a grain filter of size 16 before the computation of the scale map; see [32]. The scale map is shown in Figure 10. Again it can be observed that for most structures, such as the big buildings on the top left, computed scales are spatially accurate. However, for the city block in the middle of the image, the scale of the shape corresponding to the whole block has been associated with all pixels of this shape. This shows one of the limitations of the method: each pixel is associated with the scale of exactly one structure. A natural extension of the method would be to compute a scale profile at each pixel, in a way similar to [13].

6. Comparisons with other methods.

6.1. Methods based on the linear scale space. Lindeberg [26] has proposed to use normalized derivatives in the linear scale space in order to select characteristic scales in an image. The main idea of this approach is to detect local extrema over scales of these normalized derivatives, the most popular of which is the normalized Laplacian. In order to get a scale map, that is, to define a scale at each pixel, we compute at each point the scale (if it exists) at which the normalized Laplacian reaches its extremum. In Figure 11, scales are computed for the previously used image of Los Angeles (SPOT5). Recall that this image has a resolution of 5m. It can be observed that the scale map computed by the method of Lindeberg is quite noisy. The linear scale space approach suffers from the following drawbacks:

- The normalized Laplacian does not always have a maximum at each pixel. Thus, scale is not defined at some pixels, in particular along contours.
- Even within the same physical object, computed scales are quite variable.
- Computed scales are not spatially accurate as a consequence of the use of a linear scale space.

A variant has been proposed in [30] to define a global scale for an image. It is based on extrema of the total variation. As shown in this paper, localizing this approach using a sliding window does not yield results that are spatially very accurate, as can be expected when using the linear scale space as an analysis tool.



(a) (b)

FIG. 11. (a) *Image of Los Angeles, SPOT*5 *(*5m*)* ©*CNES;* (b) *scale map obtained by plotting the scale (if any) at which the normalized Laplacian reaches its maximum in the linear scale space, following the general framework of* [26]. *Notice that the use of the linear scale space yields poor spatial accuracy.*

6.2. Methods based on non-linear scale-space.

6.2.1. Comparison with variational methods. In section 4, we have compared our scale definition with the scale definitions proposed by [8] and [51], which are based, respectively, on the total variation flow and the ROF model. We have also explained why these two approaches ([8] and [51]) are equivalent under some regularity assumptions. Therefore, in this subsection, we use only scale maps computed by the method of [8] for experimental comparisons.

We use two different numerical approaches. The first one is a classical finite difference scheme, while the second implementation is based on the FLST, as proposed in [15]. Using this second implementation, the evolution speed of a given pixel is the perimeter of the shape f_0 containing this pixel, divided by its area. That is, the gray value $u(x)$ of the pixel at location x evolves with speed $P(f_0(x))|f_0(x)|^{-1}$. We can therefore define the scale of a pixel using the evolution speed at this pixel in a way similar to (4.7). Notice that a consequence of (4.8) is that the total variation minimization based on the FLST (proposed in [15]) is therefore equivalent to the classical total variation minimization (4.5) in the case when the image is the indicator function of a set that is a Cheeger set in itself, as explained in section 4.2.

In Figure 12 the scale maps of the previously shown images of Toulouse, Los Angeles, Marseille, and Ouagadougou are displayed. In Figures 12(a)–(d), we show the results obtained with classical finite differences, using (4.5) and (4.7). In Figures 12(e)–(h), the results with the FLST implementation of (4.5) and (4.7) are shown. In both cases, the total evolution time is $T = 60$.

The first observation is that all these methods yield much better spatial accuracy than approaches based on the linear scale space (see Figure 11). Second, one can observe that the scale maps displayed in Figures 12(a)–(d) and (e)–(h) are more noisy than the ones obtained using the method presented in this paper; see Figures 7–10. Third, regions of the original images with homogeneous scales are more clearly identified in Figures 7–10 than with both methods using the total variation flow to

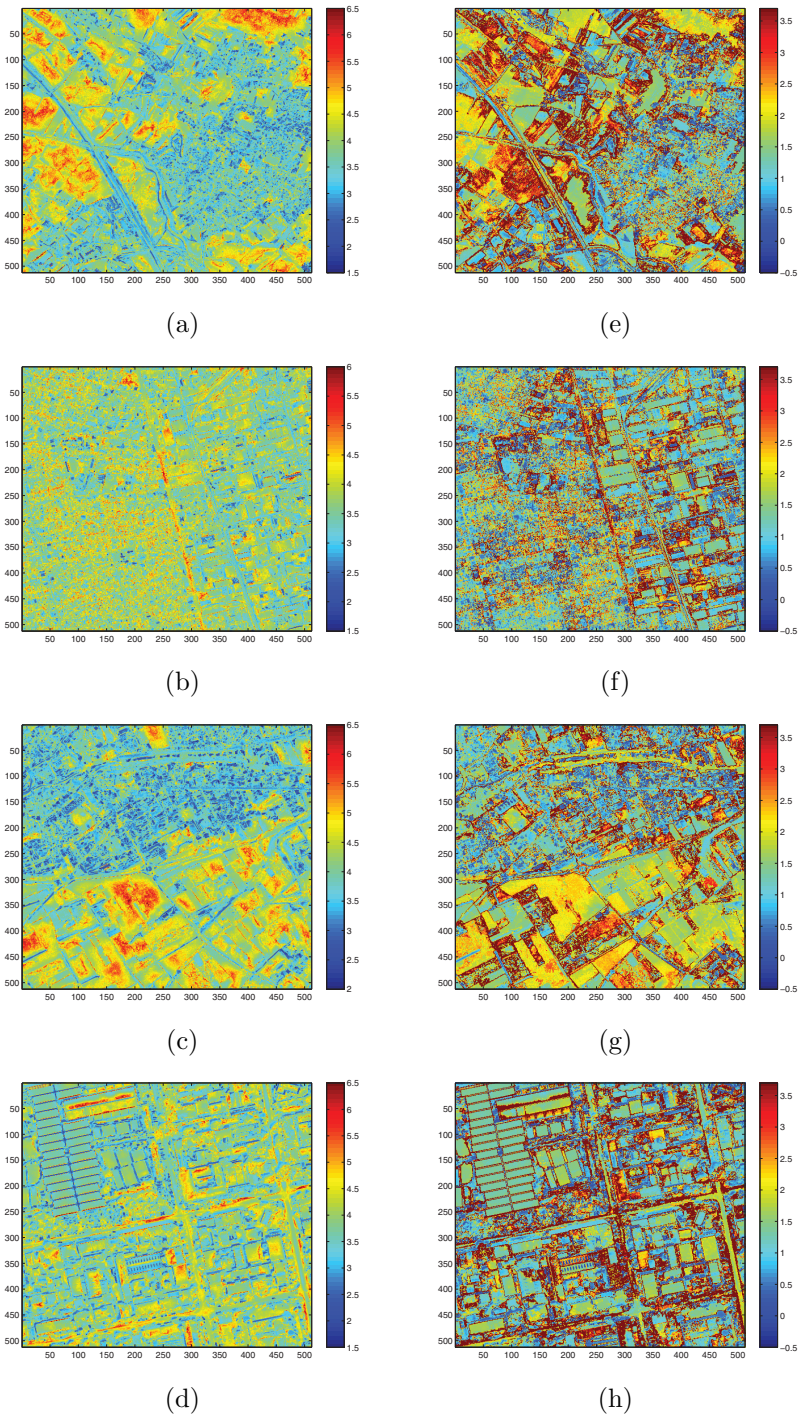


FIG. 12. (a)–(d) *Scale maps obtained by using the total variation diffusion; see* (4.7). (e)–(g) *Scale map obtained by using the total variation diffusion but this time using a numerical scheme relying on the FLST.*

define scales. Fourth, approximating the total variation using the FLST, Figures 12(e)–(h) yield results that seem more usable for classification tasks than the results obtained when approximating the flow with finite differences. But it appears quite clear in these four examples that the approach presented in this paper yields sharper results. This is probably due to the definition of scale given by formula (3.7): only one contrasted shape is selected for each pixel, in contrast with the averaging effects as described by formula (4.9) and previously illustrated in Figure 3(b).

6.2.2. Comparison with methods relying on connected filters. As mentioned in the introduction, the mathematical morphology school has proposed several ways to extract scale information from digital images, within the general framework of granulometry [18, 43]. These methods are clearly related to our approach, since the topographic map is computed from level sets of the image and in this sense is a morphological representation tool. More precisely, the topographic map is made of connected components of level sets and is therefore closely related to connected filters [42]. In [27], it is proposed to use such connected filters to compute the size distribution of structures in the image, in view of the classification of urban area. Similarly [44, 13] have proposed to use the *differential morphological profile* (DMP) to classify satellite images. Starting from an image I, a series of images I_r (respectively, I^r) are first obtained by applying opening (respectively, closing) by reconstruction with structuring elements of size r. The DMPs are then defined at each pixel as the derivative of gray level values as a function of r. The DMP obtained using I_r corresponds to bright structures, whereas the DMP obtained with I^r corresponds to dark structures. It is shown in [13] that such features are very efficient for remote sensing image classification.

In order to compare such approaches with the notion of scale introduced in this paper, we can define a scale at each pixel, obtained as the value of r for which the maximum value over the two DMPs is reached. We call this local scale the DMP scale. We use disks of radius r as structuring elements. In Figures 13(a)–(d) DMP scales for the same images as in the previous section are displayed, when the values of r used to compute the DMP vary from 1 to 10 pixels. In Figures 13(e)–(h) the results when r varies from 1 to 30 pixels are displayed. Several observations can be made from these experiments. First, DMP scales are spatially very accurate, similarly to the local scales introduced in this paper. This is a consequence of the use of connected filters. These results are also less noisy than the results shown in the previous section, using the total variation flow; see Figures 12(a)–(h).

However, the approach based on connected filters suffers several drawbacks. First, connected filters are usually not autodual; i.e., we have to use two operators (opening/closing by reconstruction in the case of Figure 13) in order to characterize bright and dark objects, respectively. Therefore, the use of connected filters makes it necessary to merge information originating from two distinct inclusion trees, preventing us from having simple local scale definitions. In contrast, using the topographic map to define scale enables us to deal with bright and dark objects in the same way and at the same time. Second, the computation of DMP necessitates the choice of a structuring element. In contrast, the topographic map permits us to handle shapes directly and to filter or select them depending on geometric attributes. Last, the results strongly depend on the values of r that are considered for the computation of the DMP, that is, on the choice of filter sizes, as shown in Figure 13. In particular, if one defines the scale as the value for which the DMP reaches its maximum, results will clearly be very dependent on the quantization of the DMP. When using the topographic map to

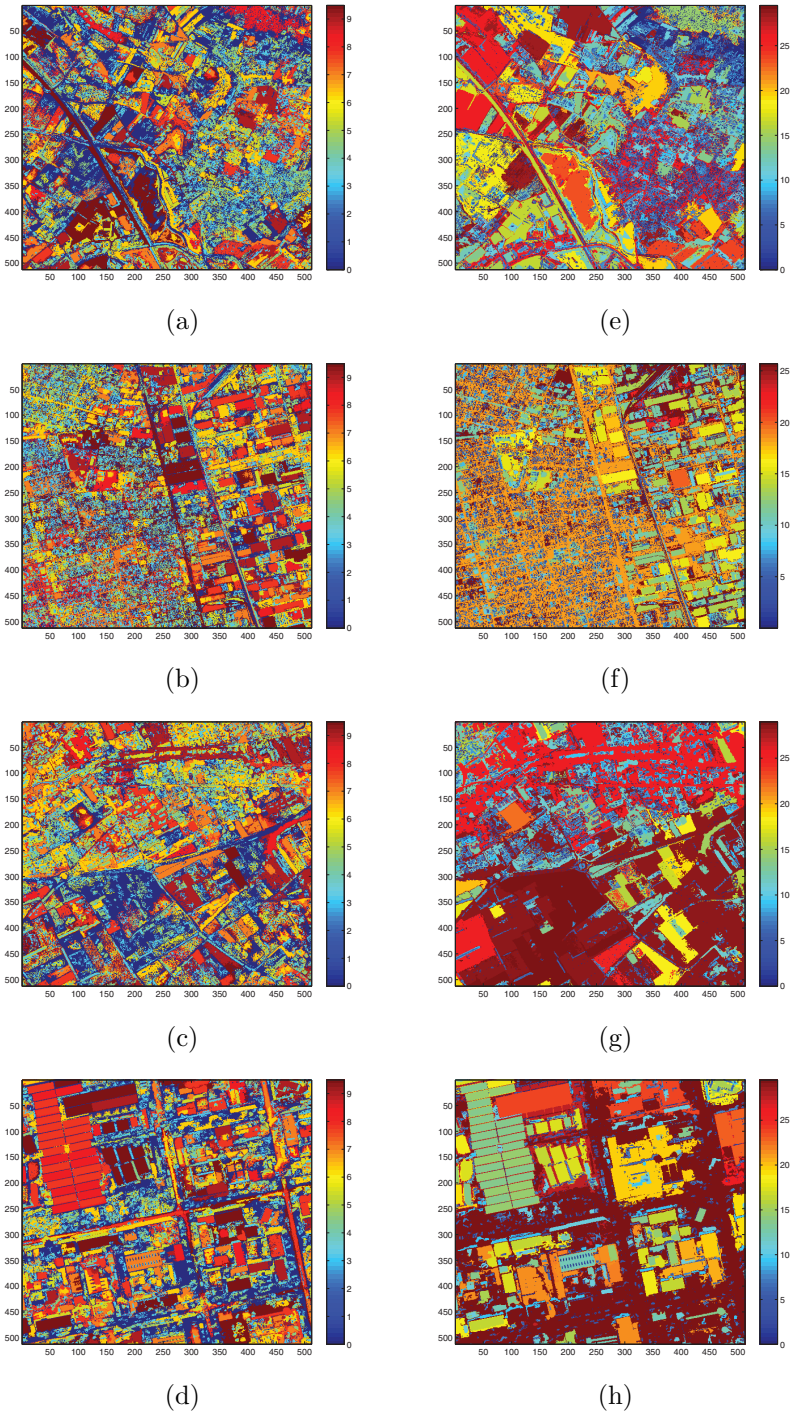


FIG. 13. (a)–(d) *Scale maps obtained by the method derived from* [13] *with connected filters of sizes varying from* 1 *to* 10 *pixels.* (e)–(g) *Scale maps obtained by the method derived from* [13] *with connected filters of sizes varying from* 1 *to* 30 *pixels.*

compute local scales, this issue is solved thanks to the notion of cumulated contrast introduced in section 3.1.

6.3. Comparison with the use of maximally stable extremal regions. We end this experimental comparative section by comparing the local scale proposed in this paper with the use of maximally stable extremal regions (MSERs) [33]. These image features are popular in the computer vision community and have proven efficient for, e.g., the wide baseline stereo problem. As mentioned in [9], these regions can be obtained from the topographic map and, using the notation introduced in section 2, correspond to shapes $f_i(x)$ for which the quantity $(|f_{i+1}| - |f_{i-1}|)/|f_i|$ reaches a local minimum. Therefore, MSERs can be seen as an alternative way to select shapes from the topographic map and these shapes could be useful to compute local scales. Observe that there may be several MSERs containing a given pixel, so that a further selection of regions must be performed in order to compute a scale measure as defined in this paper. In Figure 14, we compute scale maps obtained from MSERs on the two aerial images of Toulouse and Los Angeles shown in Figures 8(a) and 7(a), respectively. Two strategies are adopted to compute the scale measure. In the first one, we keep for each pixel x the shape $f_i(x)$ for which the quantity $(|f_{i+1}(x)| - |f_{i-1}(x)|)/|f_i(x)|$ reaches a global minimum (over all shapes $f_i(x)$, that is, shapes containing x). That is, we keep at each pixel the "best" (most stable) region. In the second strategy, the first shape for which $(|f_{i+1}(x)| - |f_{i-1}(x)|)/|f_i(x)|$ reaches a local minimum is kept. In both cases, the scale of the pixel x is defined as the ratio area/perimeter of the retained shape. The results are displayed in Figure 14. As can be expected, the resulting scale measure has sharp boundaries inherited from the regions used to define scales. However, the results strongly differ depending on the chosen strategy. In the first case (global minimum, Figures 14(a) and (b)) shapes tend to be too large to account for the scale of structures at hand, while in the second case (first local minimum, Figures 14(c) and (d)) shapes tend to be small. The main drawback of these two simple applications of MSERs to define local scales is that the chosen selection method imposes a strong bias in the definition of local scale.

7. Extensions. In this section, two extensions of the method introduced in section 3.3 to compute local scales are introduced. We first present an alternative choice for selecting the most significant shape at each pixel relying on both contrast and regularity instead of only contrast. We then propose an image segmentation method relying only on scale information.

7.1. Alternative criteria for extracting most significant shapes. Up to now, the criterion used to select the most significant shape containing a pixel is the contrast (see (3.4) and (3.6)). Nevertheless, depending on the application, this choice may be changed. For instance, it may be interesting to add a regularity criterion when selecting the most significant shape. We can define the most significant shape $\bar{f}(x)$ at a pixel x as

$$\bar{f}(x) = f_{\arg\max_{i \in A(x)} T(f_i)}, \tag{7.1}$$

where $T(f_i) = \bar{C}(f_i) \times \left(\frac{|f_i|}{P(f_i)^2}\right)^\gamma$ and γ is a weighting parameter controlling the regularity of the shape (the larger the γ, the more regular the shape)). Notice that this regularity criterion is scale invariant. The corresponding most significant connected component is then defined as

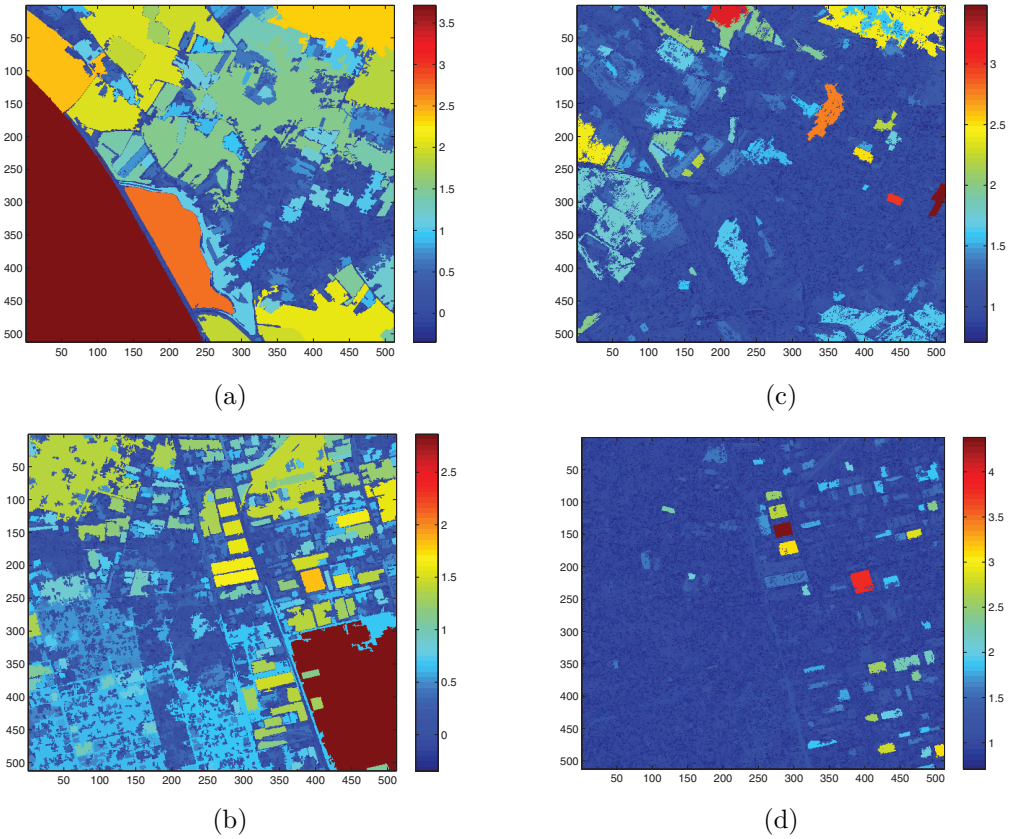
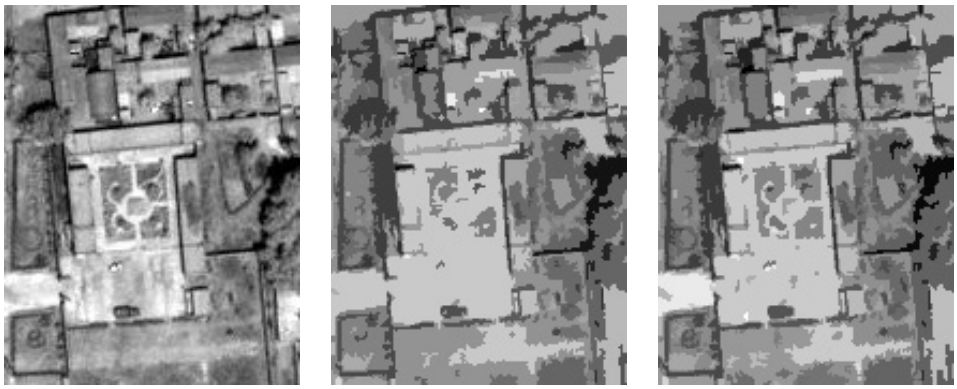


FIG. 14. *Scale maps obtained using maximally stable regions in the images of Toulouse (top) and Los Angeles (bottom). To obtain images* (a)–(b)*, the shape for which* $(|f_{i+1}(x)|-|f_{i-1}(x))/|f_i(x)|$ *is minimum is kept for each pixel. To obtain images* (c)–(d)*, the first shape for which a local minimum is reached is kept. In both cases, the scale is defined as the ratio area/perimeter of the kept region. For visualization, in* (a)–(b)*, the logarithmic values of the scales are shown.*

$$\tilde{f}(x) = \bar{f}(x) \setminus \bigcup_{\bar{f}(y)\subset\bar{f}(x)} \bar{f}(y). \tag{7.2}$$

This criterion favors regular and compact shapes (instead of elongated shapes or shapes having irregular perimeters). Remark that when $\gamma = 0$, (7.1) and (7.2) reduce to (3.4) and (3.6).

We illustrate the role of this alternative criterion in Figure 15. An image of Beijing (QuickBird (0.6m)) is displayed in Figure 15(a). In Figure 15(b), we show the scale obtained by using (3.4) and (3.6) (that is, by using only the contrast as a selection criterion). In Figure 15(c), we show the scale obtained by using (7.1) and (7.2) (that is, by using both contrast and regularity as selection criteria). We use $\gamma = 0.5$ in (7.1). Notice that by using this alternative criterion, we get more scales corresponding to the geometrical shapes (that have relatively low contrast) in the garden (in the center of the image) than when using only contrast. Let us finally emphasize that the geometric nature of the proposed algorithm (using significant shapes to define scales) enables large flexibility to assign a scale at each pixel.



(a) (b) (c)

FIG. 15. (a) *Image of a garden in Beijing (QuickBird (*0.6m*));* (b) *Most significant shapes are extracted from* (a) *by using* (3.4) *and* (3.6) *(using only contrast).* (c) *Most significant shapes are extracted from* (a) *by using* (7.1) *and* (7.2) *(using contrast and regularity) with* $\gamma = 0.5$. *The gray level value of each pixel* x *is the mean value of the pixels of the most significant shape at* x.

7.2. Image segmentation based on scale map. In this section, we briefly investigate how scale can be used as a feature for image classification or segmentation tasks. We propose to segment remote sensing images by using only the scale information. Notice that the scale attribute could, of course, be combined with other more classical features (e.g., wavelet features) to derive more sophisticated segmentation algorithms.

The segmentation scheme we use here is based on the following steps. We first compute a scale map of the image using (3.7). A presegmentation is then performed by a k-means [17] method using the local scale of each pixel, resulting in labels $l_x \in \{1, \dots, k\}$ at each pixel x. The initial cluster centers $(c_0(l))_{l=1,\dots,k}$ are chosen as $c_0(l) = l \times \frac{\sup_{x\in\Omega}(E(x))}{k}$.

After the presegmentation, a Markov random field is used to regularize the contour of the segmentation by minimizing a potential energy defined as follows:

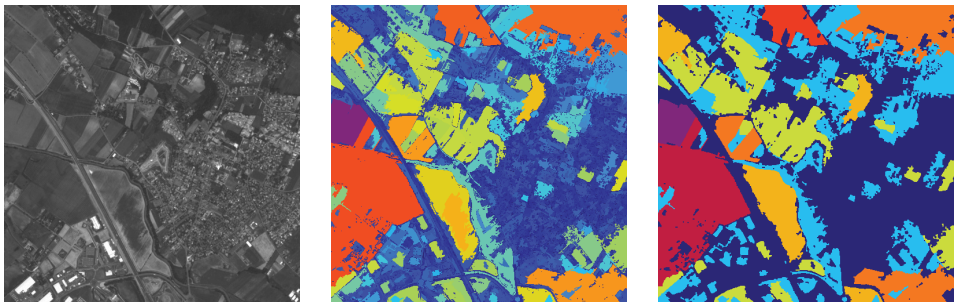
$$U(l_x|\hat{c}(l_x)) = \beta P(x; l_x) + \sum_{y\in N_x} \delta(l_x, l_y), \tag{7.3}$$

where $\hat{c}(l_x)$ is the cluster center after k-means, $P(x; l_x) = |E(x) - \hat{c}(l_x)|$, β is a nonnegative weighting parameter, N_x is the 4-neighborhood of x, and

$$\delta(l_x, l_y) = \begin{cases} -1, & l_x = l_y, \\ 1 & \text{otherwise.} \end{cases} \tag{7.4}$$

The iterative conditional mode (ICM) [6] algorithm is finally used to minimize the energy.

In Figure 16 a segmentation result obtained with this approach is shown. The original image of Toulouse is displayed in Figure 16(a), and the scale map obtained from this image is displayed in Figure 16(b). Notice that the scale map is already quite regular and not that far from a usable segmentation result. Figure 16(c) presents the segmentation result obtained from this scale map after k-means and MRF modelling. A total of eight labels have been used, and the parameter of (7.3) has been set to $\beta = 1$. Ten iterations are performed for the ICM algorithm. The first observation



(a) (b) (c)

FIG. 16. (a) *Image of Toulouse ©CNES;* (b) *scale map computed from* (b)*;* (c) *segmentation result.*

concerns the high spatial accuracy of the segmentation. Moreover, notice that using only one feature (the scale) a classification of the image between urban areas and nonurban areas is obtained: indeed, small scales occur in the urban parts of the image, while large scales characterize the nonurban parts.

8. Conclusions and perspectives. In this paper, we have proposed a method to define a local scale at each pixel of an image by using the topographic map. The scale of a pixel is defined as the scale of the most contrasted shape containing this pixel. Our approach has been validated by carrying out a large number of experiments on various satellite images. These experiments indicate that the method gives robust and spatially very accurate results. These results do not involve any complex parameter tuning. Notice also that this approach is devoted to remote sensing images, since, in order to define the contrast of level lines, it is assumed that blur is uniform over the image. Nevertheless, the method could be applied to natural images for which the blur is uniform (e.g., focus at infinity) or extended to the case of nonuniform blur (e.g., with an adaptive choice of λ in (3.2)).

Another contribution of this paper is the study of the links between the proposed method and previous variational definitions of scale. This somehow bridges a gap between morphological and variational methods to compute scales in an image.

The proposed scale measure appears to be an efficient feature for image classification or segmentation. Moreover, this feature can be expected to be complementary to more traditional features, such as wavelet features, which often rely on some linear filtering of the image. Combined with the fact that our scale measure is spatially very accurate, this could lead to interesting applications in remote sensing image indexing.

Finally, we point out that the proposed approach is relatively generic and that, as briefly presented in section 7.1, other selection criteria can be used to extract the most significant shape for each pixel. In a different direction, it seems quite promising to keep more than one scale information at each point and to investigate the hierarchical structure of shapes at each point. This could lead to interesting image analysis tools for satellite or texture images.

Appendix A. One-dimensional discontinuity and blur. In this appendix we investigate further, in the one-dimensional case, the choice of parameter λ in (3.2). Recall that this value should be chosen not too small in order to group level lines corresponding to the same edges and not too big to avoid grouping lines corresponding to different structures. In section 3.2, we use the approximation given by formula (3.5)

to explain why choosing a value of $\lambda = 1$ is reasonable for the usual values of contrast, blur, and quantization step. We now investigate more precisely the spacing between consecutive "lines" along a quantized step function in dimension one. In dimension two, computations become very tricky and were not carried out.

Let $H(x)$ be the Heaviside function

$$H(x) = \begin{cases} 1, & x \leq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

A contour in dimension 1 can be written as $CH(x)$, where C is the contrast of the contour. The smoothed contour is then $g(x) = CH(x) * h$, where h is the convolution kernel. We will consider two cases for the blurring kernel: rectangular function and Gaussian kernel. We denote by $g_q(x)$ the signal obtained by quantifying $g(x)$ with a quantification step of q and by $g_{q,r}$ the quantified and sampled signal with a sampling step of r. In the following computations, we always consider $r = 1$.

Rectangular kernel. If the convolution kernel is a rectangular function, i.e., $h = \frac{1}{\sigma}(H(x-\sigma) - H(x))$, we have

$$\begin{aligned} g(x) &= \frac{1}{\sigma} CH(x) * (H(x-\sigma) - H(x)) \\ &= \begin{cases} C, & x \leq -\sigma, \\ \frac{C}{2}(1 - \frac{x}{\sigma}), & x \in (-\sigma, \sigma), \\ 0, & x \geq \sigma. \end{cases} \end{aligned}$$

The value of $\tilde{\lambda}$ is the width between two consecutive jumps of $g_q(x)$, i.e.,

$$|g(\tilde{\lambda} + x) - g(x)| = q.$$

When $x \in (-\sigma, \sigma)$, we thus have

$$\tilde{\lambda} = \frac{2\sigma q}{C}. \tag{A.2}$$

We thus have $\tilde{\lambda} < 1$ (enabling one to use a value $\lambda = 1$ in formula (3.2)) as soon as σ is smaller than half the relative contrast of the edge, Cq^{-1}, which seems reasonable.

Gaussian kernel. If the convolution kernel is Gaussian, i.e., $h = k_\sigma(x)$, we have

$$g(x) = CH(x) * k_\sigma = \frac{C}{2}\left(1 - erf\left(\frac{x}{\sqrt{2}\sigma}\right)\right).$$

Contrarily to the previous case, the width of the jumps of the function $g_q(x)$ are different. We define $\tilde{\lambda}$ as the largest width between two consecutive jumps of $g_q(x)$. That is,

$$\tilde{\lambda} = \sup_l \{ \quad l \quad | \quad g_q(l + x) = g_q(x),\, l \in R,\, 0 < g_q(x) < C \}.$$

Let us remark that since $\partial_x g(x)$ is symmetric (the blurring kernel is symmetric), we can consider only $x < 0$ or $x > 0$. Moreover, since when $x < 0$, $|\partial_x g(x)|$ is nondecreasing, the largest value of l is indeed the width between the first and second jumps of $g_q(x)$. The first jump of $g_q(x)$ is computed as

$$\hat{x} = \sup\{ \quad x \quad | \quad |C - g(x)| \leq q \},$$

and therefore

$$\tilde{\lambda} = \sup_l \{ \quad l \quad | \quad |g(\hat{x}+l) - g(\hat{x})| \leq q \}.$$

We thus have

$$\text{(A.3)} \qquad \begin{cases} erf(\frac{\hat{x}}{\sqrt{2}\sigma}) = \frac{2q}{C} - 1, \\ erf(\frac{1}{\sqrt{2}\sigma}(\hat{x}+\tilde{\lambda})) = \frac{4q}{C} - 1. \end{cases}$$

As in the rectangular function case, we can see that $\tilde{\lambda}$ is nonincreasing with respect to the contrast of the contour, nondecreasing with respect to the width of the kernel, and nondecreasing with respect to the quantification step. In practice, for reasonable contrast and blur ($C > 15$ and $\sigma < 3$), (A.3) yields $\tilde{\lambda} < 1$, in which case the value $\lambda = 1$ can be used to apply formula (3.2).

Acknowledgments. We thank the CNES for providing us with the images shown in this paper. We thank H. Maitre and A. Giros for their comments and advices.

REFERENCES

[1] F. Alter and V. Caselles, *Uniqueness of the Cheeger set of a convex body*, Nonlinear Anal., 70 (2009), pp. 32–44.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variations and Free Discontinuity Problems*, Oxford Math. Monogr., The Clarendon Press, Oxford University Press, New York, 2000.

[3] G. Bellettini, V. Caselles, and M. Novaga, *The total variation flow in R^n*, J. Differential Equations, 184 (2002), pp. 475–525.

[4] G. Bellettini, V. Caselles, and M. Novaga, *Explicit solutions of the eigenvalue problem $-\mathrm{div}(\frac{Du}{|Du|}) = u$ in R^2*, SIAM J. Math. Anal., 36 (2005), pp. 1095–1129.

[5] J. A. Benediktsson, M. Pesaresi, and K. Arnason, *Classification and feature extraction for remote sensing images from urban areas based on morphological transformations*, IEEE Trans. Geosci. Remote Sens., 41 (2003), pp. 1940–1949.

[6] J. Besag, *On the statistical analysis of dirty pictures*, J. Roy. Statist. Soc. Ser. B, 48 (1986), pp. 259–302.

[7] T. Brox and J. Weickert, *A TV flow based local scale measure for texture discrimination*, in Computer Vision—ECCV 2004, Vol. 2, Springer, Berlin, 2004, pp. 578–590.

[8] T. Brox and J. Weickert, *A TV flow based local scale estimate and its application to texture discrimination*, J. Visual Communication and Image Representation, 17 (2006), pp. 1053–1073.

[9] F. Cao, J.-L. Lisani, J.-M. Morel, P. Musé, and F. Sur, *A Theory of Shape Identification*, Lecture Notes in Math. 1948, Springer, Berlin, 2008.

[10] V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Model. Simul., 6 (2007), pp. 879–894.

[11] V. Caselles, B. Coll, and J. M. Morel, *Scale space versus topographic map for natural images*, in Proceedings of the First International Conference on Scale-Space Theory in Computer Vision, Springer, London, 1997, pp. 29–49.

[12] V. Caselles, B. Coll, and J.-M. Morel, *Topographic maps and local contrast changes in natural images*, Int. J. Comput. Vision, 33 (1999), pp. 5–27.

[13] J. Chanussot, J. A. Benediktsson, and M. Fauvel, *Classification of remote sensing images from urban areas using a fuzzy possibilistic model*, IEEE Geosci. Remote Sens. Lett., 3 (2006), pp. 40–44.

[14] A. Desolneux, L. Moisan, and J. M. Morel, *Edge detection by Helmholtz principle*, J. Math. Imaging Vision, 14 (2001), pp. 271–284.

[15] F. Dibos and G. Koepfler, *Total variation minimization by the fast level sets transform*, in Proceedings of the IEEE Workshop on Variational and Level Set Methods in Computer Vision, 2001, pp. 179–185.

[16] F. Dibos, G. Koepfler, and P. Monasse, *Total variation minimization for scalar/vector regularization*, in Geometric Level Set Methods in Imaging, Vision, and Graphics, S. Osher and N. Paragios, eds., Springer, New York, 2003, pp. 121–140.
[17] E. W. Forgy, *Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications*, Biometrics, 21 (1965), pp. 768–769.
[18] A. Haas, G. Matheron, and J. Serra, *Morphologie mathématique et granulométries en place*, Annales des Mines, 11 (1967), pp. 736–753.
[19] H. J. A. M. Heijmans, *Connected morphological operators for binary images*, Comput. Vis. Image Underst., 73 (1999), pp. 99–120.
[20] M. Jägerstand, *Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach*, in Proceedings of the 5th International IEEE Conference on Computer Vision, Cambridge, MA, 1995, pp. 195–202.
[21] T. Kadir and M. Brady, *Scale, saliency and image description*, Int. J. Comput. Vision, 45 (2001), pp. 83–105.
[22] T. Kadir, A. Zisserman, and M. Brady, *An affine invariant salient region detector*, in Proceedings of the European Conference on Computer Vision, 2004.
[23] S. Ladjal, *Flou et quantification dans les images numériques*, Ph.D. thesis, ENS de Cachan, Cachan, France, 2005.
[24] C. Latry and B. Rouge, *SPOT5 THR mode*, in Proceedings of the SPIE Earth Observing Systems III, Vol. 3493, 1998, pp. 480–491.
[25] T. Lindeberg, *Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention*, Int. J. Comput. Vision, 11 (1993), pp. 283–318.
[26] T. Lindeberg, *Feature detection with automatic scale selection*, Int. J. Comput. Vision, 30 (1998), pp. 79–116.
[27] A. Lorette, *Analyse de texture par méthodes markoviennes et par morphologie mathématique : application à l'analyse des zones urbaines sur des images satellitales*, Ph.D. thesis, Université de Nice Sophia Antipolis, Nice, France, 1999.
[28] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, Int. J. Comput. Vision, 60 (2004), pp. 91–110.
[29] B. Luo, *Echelle et résolution en imagerie de télédétection*, Ph.D. thesis, Télécom ParisTech, Paris, France, 2007.
[30] B. Luo, J.-F. Aujol, Y. Gousseau, S. Ladjal, and H. Maître, *Resolution independent characteristic scale dedicated to satellite images*, IEEE Trans. Image Process., 16 (2007), pp. 2503–2514.
[31] P. Maragos, *Pattern spectrum and multiscale shape representation*, IEEE Trans. Pattern Anal. Mach. Intell., 11 (1989), pp. 701–716.
[32] S. Masnou and J.-M. Morel, *Image restoration involving connectedness*, in Proceedings of the 6th SPIE International Workshop on Digital Image Processing and Computer Graphics, 1997, pp. 84–95.
[33] J. Matas, O. Chum, U. Martin, and T. Pajdla, *Robust wide baseline stereo from maximally stable extremal regions*, in Proceedings of the British Machine Vision Conference, Vol. 1, British Machine Vision Association, Malvern, Worcestershire, UK, 2002, pp. 384–393.
[34] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures*, AMS, Providence, RI, 2001.
[35] K. Mikolajczyk and C. Schmid, *Scale and affine invariant interest point detectors*, Int. J. Comput. Vision, 60 (2004), pp. 63–86.
[36] P. Monasse, *Morphological Representation of Digital Images and Application to Registration*, Ph.D. thesis, University Paris IX, Paris, France, 2000.
[37] P. Monasse and F. Guichard, *Scale-space from a level lines tree*, in Proceedings of the Second International Conference on Scale-Space Theories in Computer Vision, Springer, Berlin, 1999, pp. 175–186.
[38] P. Monasse and F. Guichard, *Fast computation of a contrast-invariant image representation*, IEEE Trans. Image Process., 9 (2000), pp. 860–872.
[39] D. R. Montello, *Scale in Geography*, Pergamon Press, Oxford, UK, 2001.
[40] Nato, *STANAG* 3769*: Minimum Resolved Object Sizes and Scales for Imagery Interpretation*, 1998.
[41] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[42] P. Salembier and J. Serra, *Flat zones filtering, connected operators, and filters by reconstruction*, IEEE Trans. Image Process., 4 (1995), pp. 1153–1160.
[43] J. Serra, *Image Analysis and Mathematical Morphology*, Vol. 2, Academic Press, London, 1988.

[44] P. Soille and M. Pesaresi, *Advances in mathematical morphology applied to geoscience and remote sensing*, IEEE Trans. Geosci. Remote Sens., 40 (2002), pp. 2042–2055.
[45] Y. Song, *A topdown algorithm for computation of level line trees*, IEEE Trans. Image Process., 16 (2007), pp. 2107–2116.
[46] J. Sporring, *The entropy of scale-space*, in Proceedings of the 13th International IEEE Conference on Pattern Recognition, Vol. A, 1996, pp. 900–904.
[47] J. Sporring and J. Weickert, *On generalized entropies and scale-space*, in Proceedings of the First International Conference on Scale-Space Theories in Computer Vision, Springer, Berlin, 1997, pp. 53–64.
[48] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs*, SIAM J. Numer. Anal., 42 (2004), pp. 686–713.
[49] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 2nd ed., Wiley, Chichester, 1995.
[50] D. M. Strong, J.-F. Aujol, and T. F. Chan, *Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization*, Multiscale Model. Simul., 5 (2006), pp. 273–303.
[51] D. Strong and T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse Problems, 19 (2003), pp. 165–187.
[52] A. Winter, H. Maître, N. Cambou, and E. Legrand, *An original multi-sensor approach to scale-based image analysis for aerial and satellite images*, in Proceedings of the 1997 International IEEE Conference on Image Processing, Vol. II, Santa Barbara, CA, 1997, pp. 234–237.