Resolution independent characteristic scale with application to satellite images

Echelle caractéristique indépendante de la résolution et application aux images satellitaires

Bin Luo
Jean-François Aujol
Yann Gousseau
Saïd Ladjal
Henri Maître

2006C003

juillet 2006

Département Traitement du Signal et de l'Image
Groupe Traitement et Interprétation des Images

Dépôt légal : 2006 – 3ème trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 0751-1337 ENST C (Paris) (France 1983-9999)

Ecole Nationale Supérieure des Télécommunications

Groupe des Ecoles des Télécommunications - membre de ParisTech

46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.enst.fr

Département TSI

Echelle caractéristique indépendante de la résolution et application aux images satellitaires

Résumé : Nous nous intéressons dans cet article à la définition de l'échelle caractéristique d'une image satellitaire. Nous imposons à cette caractéristique une invariance par changement de résolution. Notre approche est fondée sur l'utilisation d'une espace échelle linéaire et de la variation totale. L'échelle caractéristique est définie comme l'échelle à laquelle la variation totale normalisée de l'image atteint son maximum.

Resolution independent characteristic scale with application to satellite images

Bin Luo [1] & Jean-François Aujol [2] & Yann Gousseau [1] & Saïd Ladjal [1] & Henri Maître [1]
[1] GET/Télécom Paris, CNRS UMR 5141 CNES-DLR-ENST Competence Center
[2] CMLA, CNRS UMR 8536

***Abstract*— We study the problem of finding the characteristic scale of a given satellite image. We want to define this feature so that it does not depend on the spatial resolution of the image. Our approach is based on the use of a linear scale space and the total variation. The critical scale is defined as the one at which the normalized total variation reaches its maximum.**

I. Introduction

Scale is usually regarded as one of the most significant features for image characterization. A wide body of literature has been devoted to the examination of images at different scales, giving birth to the popular scale-space theory. Several mathematical tools have concurrently been used to perform such an analysis : mathematical morphology, wavelet decompositions, differential equations, pyramid decompositions, etc.

While scale has a clear definition in several domains of engineering (architecture, cartography, etc.) and also in analog imagery (there it stands for the ratio of an object size in the image to the actual size of this object in the real world), it has a much fuzzier meaning in digital image processing. There, as in Physics, it reflects to some extent the level of refinement of the representation of the observed world [1]. In this rationale, a scale space representation offers a series of images where details are progressively filtered, from the thinnest to the coarsest ones, each level providing an image where no detail smaller than a given size is left.

If absolute scale remains a rather uncertain concept, characteristic scale receives a more widely accepted meaning. It is attached to a structure (object, group of objects or texture) and denotes this precise scale, in a scale space representation, where this structure is the most easily perceived. For thinner scales than the characteristic scale, fine details may interfere with the structure making it less salient ; for coarser scales, the contrast of the structure is blurred by low pass filtering or the structure may even have disappeared. Lindeberg strongly defended this approach [2] and, for an operational implementation, proposed an efficient definition by relating the characteristic scale to the scale where a suitable combination of derivatives assumes a local maximum [3].

The problem we address in this paper is issued from the world of remote sensing applied to Earth Observation, but similar problems exist in microscopic imaging or robot vision. Every time a same scene may be observed with different sensors having different resolutions, the question arises to measure an identical characteristic scale for a given structure, independent of the sensor resolution. Although some works report experimental results showing that a linear scale space applied to two sensors may provide convenient results (see for instance Figure 1 of [4]), it is our experience that without taking into consideration the impact of the sensor resolution, the derived characteristic scale is biased. Therefore, we propose a solution which explicitly incorporates the sensor impulse response in the characteristic scale estimation. Some preliminary results were presented in [5].

Many definitions of characteristic scales for images have been proposed in the literature. The most popular one is probably the aforementioned definition relying on linear scale space [2], [6]. Many alternatives also relying on the use of the linear scale space have been proposed in the field of Computer Vision, see e.g. [4]. Definitions relying on extrema of wavelet decompositions, see e.g. [7], can be put in the same category, as we will briefly see in Appendix D. Recently, it has been proposed to use non-linear scale spaces in a similar way, [8]. Several alternative approaches rely on information theory : in [9] the maximum entropy between consecutive wavelet subbands, in [10] the maximum Kullback divergence after increasing filtering by diffusion equations, in [11] the maximum change of entropy, in [12] the maximum change of generalized entropy, and in [13] the maximum entropy of grey level differences in the Gaussian scale space are used as definitions. A third kind of approach, popular in remote sensing imaging, relies on the use of the variogram of images, see [14]. However, most methods relying on the use of second order statistics assume that images follow some specific model, such as various point processes [15] or periodic functions [16] and are not suited to complex images for which such assumptions are not realistic. In this paper, we choose to follow the approach proposed by Lindeberg because the use of a linear scale space naturally allows us to take the acquisition process of the image into account when computing a characteristic scale.

The plan of the paper is the following. In Section II, is given a first definition of the characteristic scale based on the definition in [3], but differing by the mathematical norm used. In Section III, the main contribution of this paper is presented : we adapt the definition of the characteristic scale by taking into account the acquisition process in order to achieve resolution invariance. In Section IV, the behavior of the proposed characteristic scale definition is studied on various synthetic images. In Section V we test our approach on real data provided by the French space agency (CNES).

II. Basic tools and scale definition

In this section, we recall the models and mathematical tools to be used in this work, and give a definition of the characteristic scale of an image. Namely, we define the simplified acquisition process assumed for images, we introduce the classical linear scale space to be used for scale characterization and we define the total variation of images. We then define the characteristic scale as the maximizer of the total variation in the linear scale space.

a) Simplified sampling scheme: We assume that the scene under study is represented by a continuous function f, and that the digital image f_r at resolution r is obtained by convolution and sampling . Moreover, it is assumed that the convolution kernel is Gaussian, with a standard deviation $\sigma = r/\alpha$ proportional to the resolution. This can conveniently be modeled as :

$$f_r = \Pi_r . \left(f * k_\sigma\right), \tag{1}$$

where :

$$k_\sigma(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right), \tag{2}$$

and Π_r is the Dirac comb on $\mathbb{Z}^2$, that is,

$$\Pi_r = \sum_{i,j\in\mathbb{Z}} \delta_{(ir,jr)}.$$

In this context, our goal is to extract from f_r a characteristic scale related to f. Equation (1) is a rough approximation of the real acquisition process, neglecting some important aspects such as noise or quantization and assuming a simple form for the modulation transfer function of the imaging device. However, it will be shown in Section V that this model is sufficient for our purpose.

b) Linear scale space: As previously explained, the basic idea to extract characteristic scales is to track structural changes in scale spaces. In order to deal with images at various resolution (as expressed by (1)) we are naturally led to use a linear scale space [17]. For an image $f : \mathbb{R}^2 \mapsto \mathbb{R}$, its linear scale space is a function $L : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ defined as :

$$L(x,y;t) = k_t * f, \tag{3}$$

where k_t is defined by Formula (2). It is easily seen that $L(.,.,\sqrt{2t})$ is a solution of the heat equation $\partial_t L = \Delta L$, with initial condition $L(.,.;0) = f$ and that, under reasonable hypotheses, it is the only solution. For this reason, $L(.,.,\sqrt{2t})$ is the classical definition of linear scale-space. However, we prefer the definition given by Formula (3) that simplifies forthcoming computations and allows to directly define a scale that is homogeneous to a distance.

Various non-linear scale-spaces could also be considered, [18], [19], but we restrict ourselves to the linear one to be able to deal with resolution changes, as it will become clear soon.

c) Total variation: The structural changes to be quantified in the linear scale space are due to the objects present in the scene. These objects disappear as the scale increases. The basic idea of the proposed approach is to quantify the evolution of geometric structures of the image in the linear scale space. Therefore, we consider the total variation (TV) [20] of images, defined (when the image is regular enough) as $TV(f) = \int |\nabla f|$. Indeed, the semi-norm TV is related to the geometry of the image through the coarea formula, which implies that for a binary image $TV(f)$ is equal to the perimeter of the objects.

d) Scale definition: Following the general approach of [2], we define the characteristic scale of an image as the maximizer of a suitably normalized differentiable operator. To deal with the geometric contents of the image, we choose to use a normalized total variation, $NTV(t) = \phi(t)\, TV\left(k_t * f\right)$. The main idea is that the normalization term must compensate the decrease of the total variation caused by Gaussian smoothing. We denote by $t_{\max}$ the maximizer of the normalized TV over t. A natural requirement on $t_{\max}$ is that $t_{\max}(f) = s t_{\max}(f^s)$, where $f^s(\mathbf{x}) = f(s\mathbf{x})$. In Appendix B, we show that $\phi(t) = t$ is a good choice. That is, we define

$$NTV(t) = t\, TV\left(k_t * f\right) = t \int |\nabla k_t * f|, \tag{4}$$

and

$$t_{\max} = \text{argmax}_{\mathbb{R}_+^*} NTV(t). \tag{5}$$

This is in fact a special case of the normalization proposed by Lindeberg [2] for differential operators. Observe also that such a characteristic scale definition is invariant under linear contrast changes.

Recall now that we are interested in discrete images obtained from f through Equation (1). In the next section, we show how to adapt the definition of characteristic scale in this context.

III. Resolution invariance

The purpose of this section is to derive a method to ensure that the computed characteristic scale does not depend upon the resolution of the image. Recall that f is a continuous function corresponding to a given scene and since we assume that the acquisition system performs a convolution by a Gaussian kernel k_σ followed by a sampling at rate $r = \alpha\sigma$, we write :

$$f_r = \Pi_r.(f * k_\sigma),$$

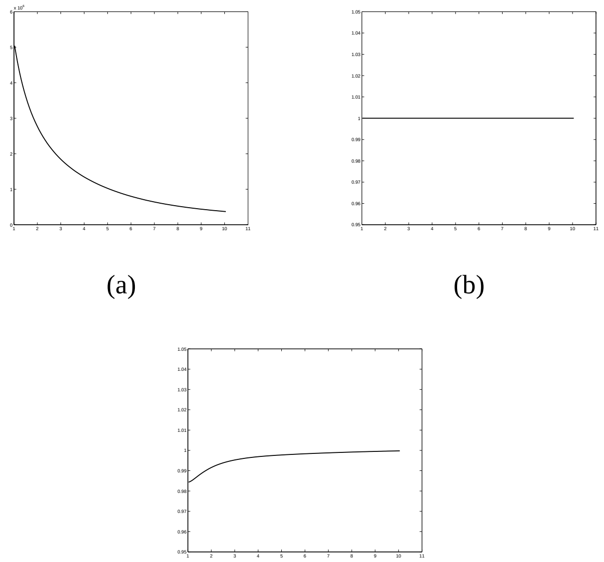
where f_r is the sampled version of f at resolution r. The parameter α is a characteristic of the acquisition process (the larger α, the more aliased the image). In the numerical experiments presented in the paper, we use $\alpha = 1$.

Denoting by $\tilde{k}_t$ the discrete version of the Gaussian kernel with standard deviation t (t expressed in pixels), we have $\tilde{k}_t \approx k_{rt}$ (up to some normalization constant which can be dropped). Let us define the discrete scale space as :

$$f_{r,t} = \tilde{k}_t \tilde{*} f_r = \tilde{k}_t \tilde{*} \left(\Pi_r . \left(k_\sigma * f\right)\right) \approx \Pi_r . \left(k_{rt} * \left(k_\sigma * f\right)\right). \tag{6}$$

where $\tilde{*}$ is the discrete convolution. The last approximation means that inverting convolution and sampling is possible, at least for non-aliased images such as $k_\sigma * f$. In Figure 1, we test the validity of this assumption on a real image. The result fully supports the hypotheses. In addition we can assume (for well-sampled images) that the total variation of the continuous and discrete versions are the same up to a normalization due to the zooming of factor r (this will be confirmed by the numerical experiments in the following sections). This leads to :

$$TV(f_{r,t}) \approx \frac{1}{r} TV(k_{rt} * k_\sigma * f) = \frac{1}{r} TV\left(k_{\sqrt{r^2t^2+\sigma^2}} * f\right). \tag{7}$$



(a) (b) (c)

Fig. 1. Validation test of Equations (6) and (7). Figure (a) shows the total variations of the last two terms of Equality (6) as functions of σ, where f is the image (b) of Figure 9. One sees that both curves are superimposed. Figure (b) shows the ratio between the two total variations displayed in (a). It is equal to 1 with precision 10^{-7}. In figure (c) is displayed the ratio between the two first terms of Equality 7 ; this ratio varies between 0.98 and 1. This experiment validates the assumption of Equation (7).

A normalization of the discrete total variation is now needed in order to relate it to the continuous normalized total variation NTV (defined in Equation (4)). Let us define :

$$G_r(t) = h(t)\, TV\left(f_{r,t}\right), \tag{8}$$

where the normalization factor $h(t)$ is to be chosen. Using Equation (7) :

$$\begin{aligned} G_r(t) &\approx \frac{1}{r} h(t)\, TV\left(k_{\sqrt{r^2t^2+\sigma^2}} * f\right) \\ &= \frac{1}{r} \frac{h(t)}{\sqrt{r^2t^2+\sigma^2}} NTV\left(f; \sqrt{r^2t^2+\sigma^2}\right). \end{aligned}$$

Hence :

$$G_r(t) \approx \frac{1}{r^2} \frac{h(t)}{\sqrt{t^2+\frac{1}{\alpha^2}}} NTV\left(f; \sqrt{r^2t^2+\sigma^2}\right). \tag{9}$$

If we choose :

$$h(t) = \sqrt{t^2 + \frac{1}{\alpha^2}}, \tag{10}$$

we then obtain :

$$G_r(t) \approx \frac{1}{r^2} NTV\left(f; \sqrt{r^2t^2+\sigma^2}\right) \tag{11}$$

and, since $t_{\max} = \text{argmax}_{\mathbb{R}_+^*}(NTV(t))$, we may define :

$$t_{\max_r} = \text{argmax}_{\mathbb{R}_+^*} G_r(t), \tag{12}$$

which provides the following relation :

$$t_{\max} = \sqrt{r^2 t_{\max_r}^2 + \sigma^2} = r\sqrt{t_{\max_r}^2 + \frac{1}{\alpha^2}}. \tag{13}$$

For a discrete image at resolution r we measure $t_{\max_r}$ and derive the value of $t_{\max}$ using Equation (13). Notice that it is impossible to find a characteristic scale $t_{\max}$ smaller than σ (which is comparable to r). More generally, when the resolution of the image is larger than the actual characteristic scale $t_{\max}$ the computation becomes unreliable. Experiments show that $t_{\max}$ is retrievable as long as $r < t_{\max}$.

From now on, the values of $t_{\max}$ will be deduced from Equation (13).

Remark about the normalization. In view of Equation (4), the intuitive normalization would not take into account the filtering process due to the change of resolution and, therefore, involve a factor t instead of $h(t)$:

$$A_r(t) = t \times TV\left(f_{r,t}\right) \tag{14}$$

If, according to this intuition, we set $t_{\max_r}$ as $\text{argmax}_{\mathbb{R}_+^*}(A_r(t))$ and deduce $t_{\max} = r \times t_{\max_r}$, then we can check numerically that the obtained value of $t_{\max}$ will depend much more on the resolution than with the definition from Equation (13). This fact will be precised in Section V, see Figure 14.

Notice also that when $t \gg 1$, then the definitions from Equations (8) and (14) are equivalent. The choice of the correct normalization given by Equation (10) is important when r approaches $t_{\max}$ (that is $t_{\max_r}$ approaches 1).

IV. Relating $t_{\max}$ to the geometry of the image

In this section, we investigate the link between the characteristic scale $t_{\max}$, as defined in Section II for a continuous image, and the geometric contents of the image. For this purpose, following the example in [2], we first consider various simple one-dimensional functions, for which we perform computations and numerical approximations. Then, we tackle the two-dimensional case by performing numerical simulations on discrete synthetic images.

A. Continuous one-dimensional examples

In order to consider cases with tractable computations, we define $t_{\max}$ for a one-dimensional function f as in Formula (5). For 1D signals the gradient is replaced by the derivative and k_t by a one-dimensional Gaussian in the computation of $NTV(t)$.

a) Sinus function: Assuming that f is a sinus of period D, restricted to $[-T, T] \subset \mathbb{R}$, it may be shown that if $T/D \to \infty$ (so that boundary effects can be neglected) then $t_{\max} \to D/2\pi \sim 0.15D$, as already mentioned in [21], [2].

b) Sum of Gaussians: Assume that f is a function defined on $[-T, T] \subset \mathbb{R}$ as :

$$f(x) = \sum_{k=-K}^{K-1} \frac{1}{\sqrt{2\pi v^2}} e^{\frac{-(x-(2k+1)D/2)^2}{2v^2}}, \tag{15}$$

ie., f is the restriction to $[-T, T]$ of a sum of Gaussians, the spatial period of this sum being D. Assuming that $K \gg 1$ (or $T \gg D$) in order to neglect boundary effects, we obtain :

$$NTV(t) \approx \frac{2t}{Dq\sqrt{2\pi}} \left(1 + e^{-\frac{K^2D^2}{4q^2}} + 2\sum_{k=1}^{2K-1} (-1)^k e^{-\frac{k^2D^2}{4q^2}}\right), \tag{16}$$

where $q = \sqrt{v^2 + t^2}$. This result is obtained by noticing that the total variation may be computed on each monotonous part. The graph of $NTV(t)$ as a function of t is shown on Figure 2. On Figure 3 (a), the graph of $t_{\max}$ is displayed as a function of D, v being constant. One observes that $t_{\max} \approx 0.15D$, a result very similar to the one for the sinusoidal case. Figure 3 (b) shows $t_{\max}/D$ as a function of v, D being constant. One can check that $t_{\max}/D \approx 0.15$. In this case, $t_{\max}$ is related to the period of the signal but not to the width of each Gaussian.

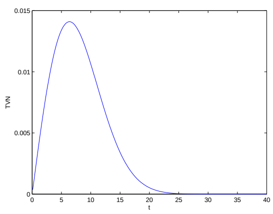
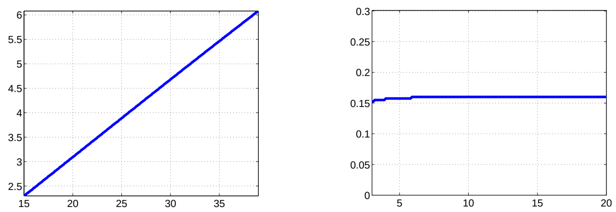


Fig. 2. Plot of $NTV(t)$ as a function of t, when f is a sum of Gaussians as in Equation (15), $t \in [0.1, 40]$ and $K = 10$, $D = 40$, $v = 10$. $NTV(t)$ reaches its maximum for $t_{\max} = 6.4$



(a) $t_{\max}$ as function of D (b) $t_{\max}/D$ as a function of v

Fig. 3. Plot of $t_{\max}$ for a sum of Gaussian functions (see Equation (15)) ; (a) :$t_{\max}$ as a function of D, with $v = 5$ (we check numerically that $t_{\max} \approx 0.15D$, D being the spatial period) ; (b) :$t_{\max}$ as a function of v, with $D = 40$.

c) Sum of Heaviside functions: In order to investigate the sensitivity of $t_{\max}$ to the shape of "objects", we consider the following example, still in 1D to yield tractable computations : f is defined on $[-T, T]$ by

$$f(x) = \sum_{i=-K}^{K} H(x - iD) \tag{17}$$

where

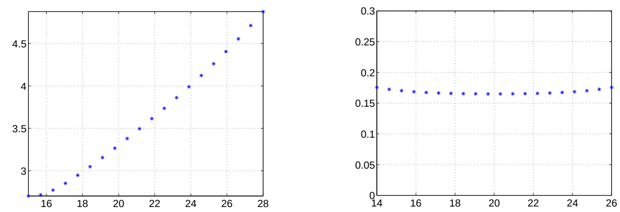
$$H(x) = \begin{cases} 1, & x \in [0, v] \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

with $v \in (0, D)$. Assuming that $t << D$ and $v \approx D/2$, then it may be shown that :

$$NTV(t) \approx Ct\left(2\text{erf}\left(\frac{D-v}{2\sqrt{2}t}\right) - \text{erf}\left(\frac{D-v}{\sqrt{2}t}\right)\right) + \left(2\text{erf}\left(\frac{v}{2\sqrt{2}t}\right) - \text{erf}\left(\frac{v}{\sqrt{2}t}\right)\right) \tag{19}$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}$ and C is a constant (see Appendix E).

Figure 4 (a) shows numerical computations of $t_{\max}$, taken as the zero of the derivative of Formula (19) (cf. Equation (35) in Appendix E) as a function of D and Figure 4 (b) shows the plot of $t_{\max}/D$ as a function of v. Here again we obtain $t_{\max} \approx 0.15D$ and observe that $t_{\max}/D$ depends very little on v.



(a) $t_{\max}$ as function of D (b) $t_{\max}/D$ as a function of v

Fig. 4. $t_{\max}$ for a sum of Heaviside functions (see Equation (17)). We check numerically that $t_{\max} \approx 0.15D$ where D is the spatial period. In Figure (a) $v = 10$ and in Figure (b) $D = 40$.

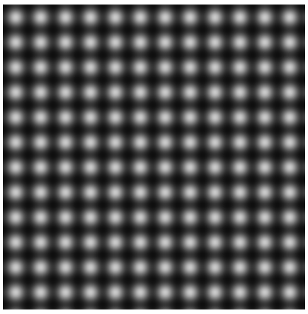
To summarize, in cases a), b) and c), it may be computed or observed that $t_{\max} \approx 0.15D$, which indicates that neither the shape nor the size of the pattern seem to influence much $t_{\max}$ in the cases of 1-D functions.

B. Discrete synthetic images

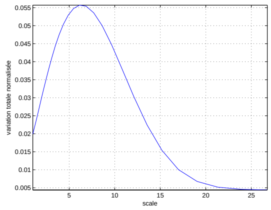
In order to confirm the linear relation between $t_{\max}$ and the spatial period of signals (D in the preceding examples) in the case of images, we simulated periodic images using various patterns. Two instances of such images are displayed in Figure 5 (sum of Gaussians with standard deviation v) and 6 (sum of squares with width v), together with the associated graphs of NTV as functions of t. Figure 7 (a) shows the graph of $t_{\max}$ as a function of D and Figure 7 (b) shows the graph of $t_{\max}/D$ as a function of v for sums of Gaussians. Figures 8 (a) and (b) show the same quantity for sums of squares. Comparing these two figures respectively with Figure 3 and Figure 4, we conclude that the shape of patterns as well as their size have little influence on the measure. Moreover, we see that $t_{\max} \approx 0.15D$ still holds in dimension 2.

V. Application to Satellite Images

In this Section several experiments are presented to demonstrate the invariance of $t_{\max}$ with respect to resolution on real images in the domain of remote sensing.

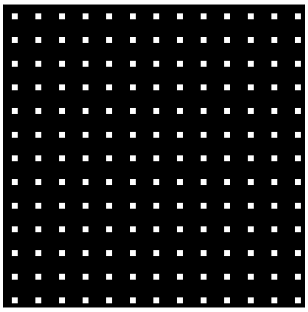


(a) Periodic sum of Gaussians

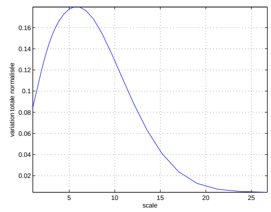


(b) Graph of NTV

Fig. 5. A periodic Gaussian function with $D = 40$ and $v = 10$ (standard deviation of each Gaussian) and the graph of the corresponding normalized total variation. The maximum is reached for $t = 6.1$.



(a) Periodic sum of squares

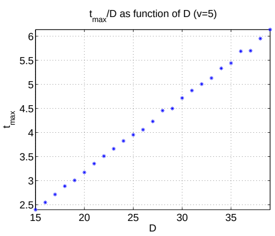


(b) Graph of NTV

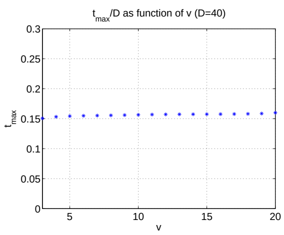
Fig. 6. An image composed of squares with spatial period $D = 40$, and the side of each square equal to 10 pixels, and the graph of the corresponding normalized total variation. The maximum is reached for $t = 6.1$.

At first, starting from several images with resolution equal to 25 cm, we create a series of lower resolution images using Formula (1), i.e. using a down-sampling scheme in which filtering is made using a Gaussian impulse response. We call it the ideal down-sampling.

Then, we make use of series of images provided by the CNES which precisely simulate the images which would be obtained with different sensors operating with various resolutions from a satellite, i.e. by taking into account the different effects of sampling, integration, acquisition time, etc. and therefore providing the actual impulse response of sensors.

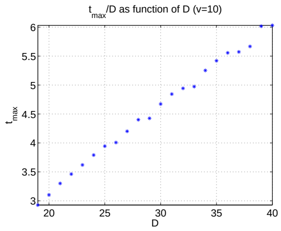


(a) t_{max} as a function of D

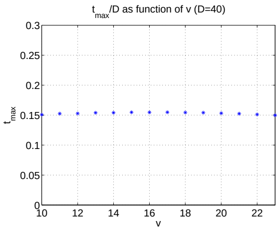


(b) t_{max}/D as a function of v

Fig. 7. (a) Graph of $t_{\max}$ as a function of D (with $v = 5$) for sums of Gaussians, we obtain $t_{\max} \approx 0.15D$; (b) Graph of $t_{\max}/D$ as a function of v (with $D = 40$) for sums of Gaussians, we obtain $t_{\max}/D \approx 0.15$.



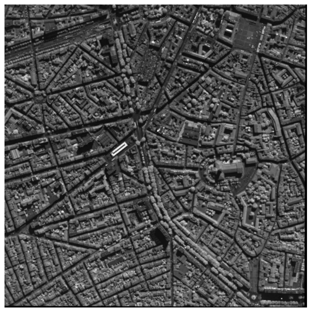
(a) t_{max} as a function of D



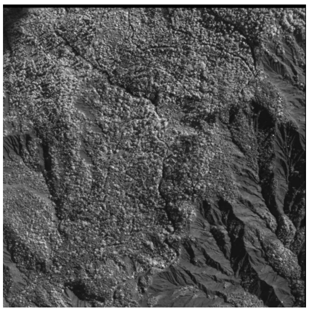
(a) t_{max} as a function of D

Fig. 8. (a) Graph of $t_{\max}$ as a function of D (with $v = 10$) for sums of squares, showing that $t_{\max} \approx 0.15D$; (b) Graph of $t_{\max}/D$ as a function of v (with $D = 40$) for sums of squares, showing that $t_{\max}/D \approx 0.15$.

(a) Marseille



(b) Toulouse

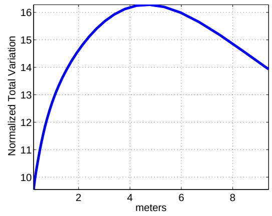


(c) Didrai

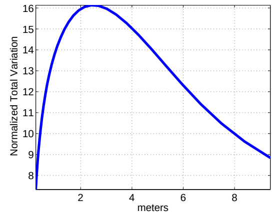
(d) Roujan

Fig. 9. Aerial images with 25 cm resolution ©CNES : (a) and (b) 2 cities with different urban tissues, (c) a forest, (d) agricultural fields.

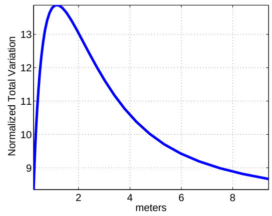
a) Computation of $t_{\max}$ using Formula (13): On Figure 10, we display the graphs of the normalized total variation for the 4 images shown in Figure 9 (at resolution $r = 25cm$). There is at least one local maximum in each case. In the case of cities (Marseille or Toulouse), the characteristic scale is related to the size of the buildings and streets. In the case of the Didrai image (forest), the scale is related to the vegetation. Note that in the case of the Roujan image (fields), there are two local maxima, the narrow one (zoomed in Figure 10 (e)) is related to the vineyards, and the large one to the fields. Figure 11 displays a zoomed area of Figure 9 composed of vineyards. The spatial period of the vineyards D may be computed from the characteristic scale $t_{\max}$, using the relation $t_{\max} \approx 0.15D$. We find that the distance between two vineyard rows is roughly 2.7m, a result which we were able to check on the image for such a regular and periodic structure.



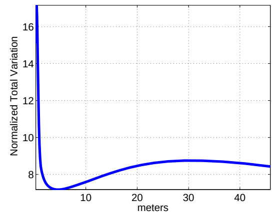
(a) Marseille



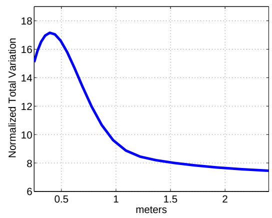
(b) Toulouse



(c) Didrai

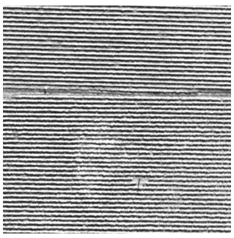


(d) Roujan

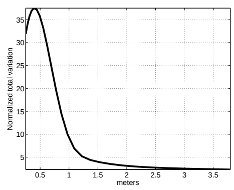


(e) Roujan(zoomed)

Fig. 10. NTV as a function of $\sqrt{r^2t^2+\sigma^2}$ computed for the 4 images of Figure 9 with resolution $r = 0.25m$. (a) Marseille : $t_{max} = 4.8m$; (b) Toulouse : $t_{max} = 2.4m$; (c) Didrai : $t_{max} = 1.2m$; (d) Roujan : there are two local maxima. The first one at position $0.4m$, the second one at position $30m$; (e) Zoom around the first local maximum at $0.4m$ shown in (d).



(a) Image of vineyards



(b) NTV calculated on the vineyards

Fig. 11. Zoom on vineyards issued from the Roujan image displayed on Figure 9 (d). The characteristic scale appears at resolution 0.4m. Using the relation $t_{\max} \approx 0.15D$, we see that the distance D between 2 rows of vineyards is roughly 2.7m.

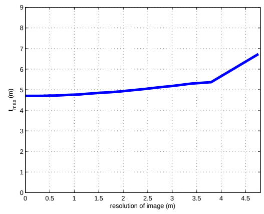
b) Resolution invariance: In order to confirm that the characteristic scale extracted from the images is independent from the resolution of the sensor ($t_{\max}$ does not depend on r), we made the following experiments. For a given scene, an image g_r at resolution r is generated (using Formula (1) with $\alpha = 1$), and the maximizer $t_{\max}$ is computed. Figure 12 shows the graph of $t_{\max}$ as a function of r. As expected, it shows that $t_{\max}$ is almost constant (as long as $r < t_{\max}$).

Remark that in the case of Roujan, where two different characteristic scales are present, the plot of $t_{\max}$ is coherent with the result shown in Figure 10 (d). When the resolution r is fine enough, $t_{\max}$ is the characteristic scale corresponding to the vineyards. But when r gets larger, then the vineyards disappear (one no longer sees them in the images), and $t_{\max}$ is then related to the size of the fields.

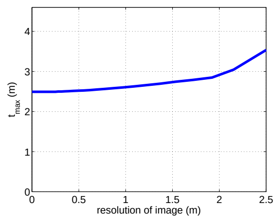
TABLE I

Available resolutions (meters)

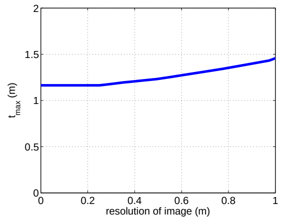
0.250	0.281	0.315	0.354	0.397	0.445	0.500
0.561	0.630	0.707	0.794	0.891	1.00	1.12
1.26	1.41	1.59	1.78	2.00	2.25	2.52
2.83	3.17	3.56	4.00	4.49	5.04	5.66
6.35	7.13	8.00	8.98	10.08		



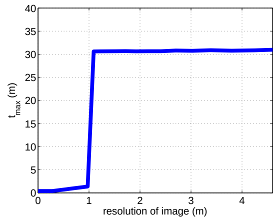
(a) Marseille



(b) Toulouse



(c) Didrai



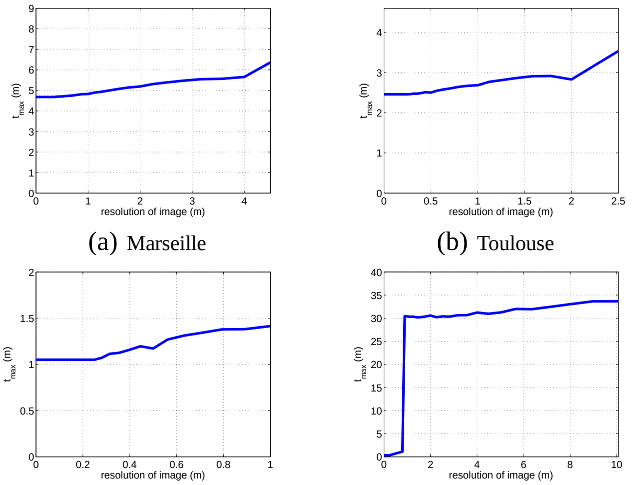
(d) Roujan

Fig. 12. Characteristic scales t_{max} as a function of the resolution, for the 4 scenes shown in Figure 9. The images at different resolutions are obtained by down-sampling the 25 cm images using the ideal acquisition model presented in Section III (with $\alpha = 1$), i.e. with a Gaussian convolution kernel. Notice that the characteristic scale is almost independent from the resolution.

c) Validation of the approximation of the acquisition model: In order to examine the case where different sensors with different resolutions and different impulse responses are used, we take advantage of a series of images provided by the CNES, including the four images of Figure 9. For each scene, 33 images are available at resolutions ranging from 25 cm to 10.08 m (see Table I), each one taken with the exact impulse response of a real sensor. These images have been obtained by numerical simulations performed by the CNES, using aerial images and a realistic model of data acquisition. The impulse response is resolution dependant, isotropic, and highly non-Gaussian. The use of a non-Gaussian impulse response in place of a Gaussian one makes the derivation of a relation similar to (13) difficult. However, we will see below that approximating the impulse response with a Gaussian kernel leads to good numerical results.

Figure 13 shows the graph of $t_{\max}$ as a function of the resolution. Results are very similar those of Figure 12. We observe that $t_{\max}$ is almost constant (as long as the resolution $r < t_{\max}$). We conclude that even though the kernel is not Gaussian, the approximations made in section III are still valid.

If instead of using the original relation (13) introduced in this paper, we make use of the intuitive normalization of Equation (14), we obtain the plots of Figure 14. As expected, in this case, the estimated $t_{\max}$ is much more sensitive to the resolution.



(a) Marseille (b) Toulouse

(c) Didrai (d) Roujan

Fig. 13. Characteristic scale t_{max} as a function of the resolution, on the 4 scenes shown in Figure 9. The images at different resolutions are issued from the series of images provided by CNES (therefore, the convolution kernel is no longer Gaussian). The characteristic scales are almost invariant when the resolutions changes.

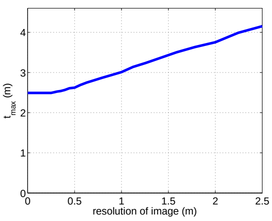


Fig. 14. Characteristic scale $t_{\max}$ as a function of the resolution, for the Toulouse image. The scale $t_{\max}$ is computed with the naive normalization given by Equation (14). Notice that the result is less invariant to resolution changes than in the case of Figure 13(b). In the range $[0.25, 2m]$, the variation of the value is 18% with the proposed method (Figure 13 (b)) and 40% with the naive normalization.

VI. Conclusion and future prospects

A new method to compute a characteristic scale for a given image has been proposed, which does not depend on the resolution (as long as the objects are larger than one pixel). This method explicitly takes into account the role of filtering in the down-sampling. It has been shown to be robust and stable on different images issued from the remote sensing domain. We have also shown on various examples that the position of the maximizer of the normalized TV does not depend on the object shapes, but merely on the distances between structures.

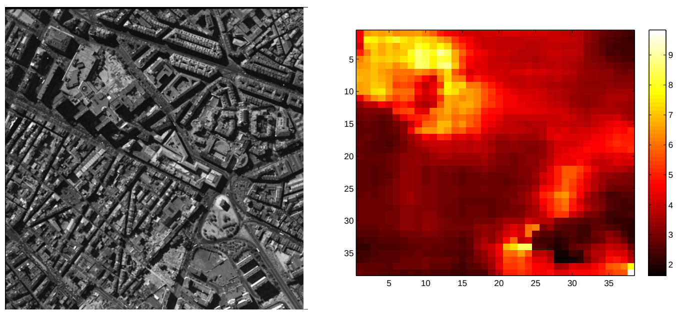
This approach is foreseen to find applications for the problem of satellite image indexing. In this case, it is indeed a major asset that features does not depend on the resolution [22]. Moreover, we expect to find the texture/geometry behavior of a scene [23], which indeed depends on the resolution and can be related to the characteristic scale. This could be useful for features selection. We also need to understand more deeply the effect of sampling on total variation, especially when the resolution is close to $t_{\max}$.

Appendix

A. Localization issue

The scale measurement we have introduced can be localized using a sliding window. The scale of a single pixel is then computed as the scale on the window centered around this pixel. To illustrate this approach, we have processed the Marseille image (see Figure 15 a). We use the image at resolution $0.707m$, with size 1440×1440. The analysis is made using a window with size 256×256, and the window is moved by 32 pixels at each step. On Figure 15 (b), we show the computed values of $t_{\max}$.

Notice in particular that $t_{\max}$ is larger in the top left corner of the image . Looking at Figure 15 (a), one sees that this corresponds to larger buildings and structures in the original image.



(a) (b)

Fig. 15. (a) Image of Marseille (resolution $0.707m$, size 1440×1440) ; (b) Image of the corresponding values of $t_{\max}$ (the larger $t_{\max}$, the whiter the gray level value in (b)).

B. Normalization issue revisited

The characteristic scale of an image f has been defined as :

$$t_{\max} = \text{argmax}_{\mathbb{R}_+^*} \phi(t) \int |\nabla(k_t * f)|,$$

with $\phi(t) = t$. In this section, we show why it is reasonable to choose $\phi(t) = t^B$ while the next section explains why $B = 1$ has been be chosen.

Since we want $t_{\max}$ to be related to the size of objects in the image f, we naturally assume that :

$$t_{\max}(f) = s t_{\max}(f^s), \tag{20}$$

where $f^s(\mathbf{x}) = f(s\mathbf{x})$. For any $t_0 > 0$ and $s > 0$, let us define

$$F_s(t_0) = \partial_t \log \left(\phi(t) \int |\nabla f^s * k_t| \right)(t_0).$$

Equation (20) implies that

$$F_1(t_0) = 0 \Rightarrow F_s\left(\frac{t_0}{s}\right) = 0. \tag{21}$$

Now,

$$
\begin{aligned}
F_s\left(\frac{t_0}{s}\right) &= s\partial_t \log\left(\phi(t/s)\int |\nabla f_s * k_{t/s}|\right)(t_0) \\
&= s\partial_t \log\left(\phi(t/s)s^{-1}\int |\nabla f * k_t|\right)(t_0) \\
&= \frac{\phi'}{\phi}(t_0/s) + s\partial_t \log \int |\nabla f * k_t|(t_0) \\
&= \frac{\phi'}{\phi}(t_0/s) + s\left(F_1(t_0) - \frac{\phi'}{\phi}(t_0)\right).
\end{aligned}
$$

Then Equation (21) implies that

$$\frac{\phi'}{\phi}(t_0/s) = s\frac{\phi'}{\phi}(t_0),$$

and therefore $\phi(t) = At^B$ for two constants A and B to be chosen.

C. *Power of normalization factor : why we set $B = 1$*

The constant A does not affect $t_{\max}$. The reason why we chose $B = 1$ is essentially of a numerical nature. If B is too small, then NTV decreases very fast, implying a very small value of $t_{\max}$. This becomes a severe drawback when computing the scale of low resolution images. On the other hand, B cannot be too large. Indeed, we have checked experimentally on the images provided by the CNES that in this case the graph of NTV becomes flat and the relative error for the numerical value of $t_{\max}$ gets larger. In such a case, the localization of the extremum is not reliable. We found experimentally that setting $B = 1$ is a good compromise between these two drawbacks. Moreover, this choice is coherent with the one in [2]. As an example, Figure 16 displays the graph of NTV with $B = 1.3$ in the case of Didrai image. We may see that this value already makes it difficult to compute t_{max}, whereas it is easier from Figure 10 (c).

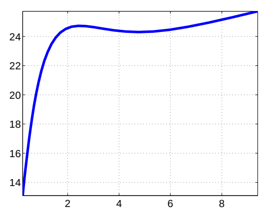


Fig. 16. NTV calculated on the image of Didrai with normalization factor $\phi(t) = t^{1.3}$.

D. *Relations with wavelet-based approaches*

For the sake of clarity, we only deal with the 1-Dimensional case in this appendix. For a detailed presentation of the wavelet theory, we refer the interested reader to [24], [25]. Let us consider a signal f with spatial period $2T$. We recall that we denote

$$f^{1/s}(t) = f\left(\frac{t}{s}\right) \tag{22}$$

Let ψ a wavelet. We define the wavelet coefficients of f at position y and scale t by :

$$W_f(y,s) = c_s \int_{\mathbb{R}} f(t)\psi^{1/s}(t-y)\,dt \tag{23}$$

c_s being a normalization coefficient which we will fix later. Let us now look at the following quantity :

$$h_f(s) = \frac{c_s}{2T}\int_{-T}^{T} |W_f(y,s)|\;dy \tag{24}$$

A straightforward computation (change of variables in the integral) leads to :

$$h_{f^{1/a}}(s) = \frac{c_s}{2Ta}\int_{-T/a}^{T/a}\left|\int_{\mathbb{R}} f(u)\psi\left(\frac{z-u}{as}\right)du\right| dz \tag{25}$$

Hence, by definition of $h_f(as)$, we get : $\frac{h_{f_{1/a}}(s)}{h_f(as)} = \frac{1}{a}\frac{c_s}{c_{as}}$. Now, remembering formula (22), one sees that in order to achieve scale invariance, the following equality must hold : $h_{f^{1/a}}(s) = h_f(as)$. We thus conclude that :

$$\frac{c_s}{c_{as}} = a \tag{26}$$

The exact form of c_s is given by the next standard lemma which we state without proof :

Lemma : Let us assume that c_s verifies (26) and that the function $s \mapsto \frac{1}{c_s}$ is continuous in 0, then there exists a constant $A > 0$ such that :

$$c_s = \frac{A}{s} \tag{27}$$

Let us choose $\psi = \phi'$, with ϕ the 1-Dimensional gaussian defined by : $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right)$. Noticing that $\left(\phi^{1/s}\right)' = \frac{1}{s}\psi^{1/s}$, we get from (26) that :

$$h_f(s) = \frac{As}{2T}\int_{-T}^{T}\left|\left(\frac{1}{2\pi s}\phi^{1/s} * f\right)'(y)\right| dy \tag{28}$$

Up to some positive multiplicacite constant, we therefore get the same expression as formula (4) (notice that the normalization by the standard deviation is already included in k_t (defined by (2)) in (4), contrary to $\phi^{1/s}$). This computation also confirms our choice of $B = 1$ in Appendix C.

E. *Computation of $t_{\max}$ for a sum of Heaviside functions*

We detail here the computation of $t_{\max}$ when f is defined by Formula (17). To simplify notations, we define $T_1 = v$ and $T_2 = D - v$.

$$
\begin{aligned}
TVN(f;t) &= t\int |\partial_x (f * k_t(x))|\,dx && (29)\\
&= t\int |(\partial_x f(x)) * k_t(x)|\,dx && (30)
\end{aligned}
$$

Recall that, in the sense of distributions, we have $\partial_x d(x) = \delta(0) - \delta(T_2)$ where δ is the Dirac distribution [26]. We therefore have :

$$\begin{aligned} & NTV(t) \quad (31) \\ = \quad & \int t \left| \left(\sum_{k=-K}^{K} \delta(kD) - \delta(kD+T_2) \right) * k_t \right| dx \\ = \quad & t \int \left| \sum_{k=-K}^{K} k_t(x-kD) - k_t(x-kD-T_2) \right| dx \\ = \quad & \frac{1}{\sqrt{2\pi}} \int \left| \sum_{k=-K}^{K} e^{-\frac{(x-kD)^2}{2t^2}} - e^{-\frac{(x-kD-T_2)^2}{2t^2}} \right| dx. \end{aligned}$$

We make the assumption that $K \gg 1$. In the above integration, we can thus restrict our attention to the interval $[-T_1, T_2]$. Besides, we also assume that $D \gg t$. We therefore have to take into account three Gaussians only, ($k_t(x+T_1)$, $k_t(x)$ and $k_t(x-T_2)$), and neglect the influences of the other ones in the interval $[-T_1, T_2]$. We then split the interval $[-T_1, T_2]$ into two intervals $[-T_1, 0]$ and $[0, T_2]$, and we suppose that v is close to $D/2$, so that we do not take into account $k_t(x-T_2)$ when we calculate the integral in the interval $[-T_1, 0]$ (and in the same way, we will neglect the influence of $k_t(x+T_1)$ in the interval $[0, T_2]$). We denote the integral in the interval $[-T_1, T_2]$ by $TVN_D(f;t)$. Remark that the interval $[-T_1, -T_1/2]$ and the interval $[-T_1/2, 0]$ are symmetric, as well as the interval $[0, T_2/2]$ and $[T_2/2, T_2]$. We therefore get :

$$\begin{aligned} TVN_D(f;t) \quad & \approx \frac{1}{\sqrt{2\pi}} \int_{-T_1}^{0} \left| -e^{-\frac{(x+T_1)^2}{2t^2}} + e^{-\frac{x^2}{2t^2}} \right| \\ & + \frac{1}{\sqrt{2\pi}} \int_{0}^{T_2} \left| -e^{-\frac{(x-T_2)^2}{2t^2}} + e^{-\frac{x^2}{2t^2}} \right| \\ & \approx \frac{1}{\sqrt{2\pi}} 2 \int_{-T_1/2}^{0} \left(-e^{-\frac{(x+T_1)^2}{2t^2}} + e^{-\frac{x^2}{2t^2}} \right) \\ & + 2 \frac{1}{\sqrt{2\pi}} \int_{0}^{T_2/2} \left(-e^{-\frac{(x-T_2)^2}{2t^2}} + e^{-\frac{x^2}{2t^2}} \right) \\ & \approx \frac{1}{\sqrt{2\pi}} \left(4 \int_0^{T_1/2} e^{-\frac{x^2}{2t^2}} - 2 \int_0^{T_1} e^{-\frac{x^2}{2t^2}} \right. \\ & \left. + 4 \int_0^{T_2/2} e^{-\frac{x^2}{2t^2}} - 2 \int_0^{T_2} e^{-\frac{x^2}{2t^2}} \right) \\ & \approx \frac{1}{\sqrt{2\pi}} \left(2\sqrt{2} t \left(2\mathrm{erf}\left(\frac{T_2}{2\sqrt{2}t}\right) - \mathrm{erf}\left(\frac{T_2}{\sqrt{2}t}\right) \right) \right. \\ & \left. + 2\sqrt{2} t \left(2\mathrm{erf}\left(\frac{T_1}{2\sqrt{2}t}\right) - \mathrm{erf}\left(\frac{T_1}{\sqrt{2}t}\right) \right) \right). \end{aligned} \quad (32)$$

where we recall that $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}$. We then denote by $z = \frac{1}{2\sqrt{2}t}$, i.e. $t = \frac{1}{2\sqrt{2}z}$. Then we have

$$\begin{aligned} TVN_D(f;z) \quad & \approx \frac{1}{2\sqrt{2}} \frac{1}{z} \left(2\mathrm{erf}(zT_1) - \mathrm{erf}(2zT_1) \right) \\ & + \frac{1}{2\sqrt{2}} \frac{1}{z} \left(2\mathrm{erf}(zT_2) - \mathrm{erf}(2zT_2) \right). \end{aligned} \quad (33)$$

Recall that $\partial_x \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}$. We then have :

$$\partial_t TVN_D(f;t) = \partial_z TVN_D(f;z) \partial_t z(t). \quad (34)$$

Recall that $t_{\max}$ is such that $\partial_t TVN_D(f;t) = 0$. Since $\partial_t z(t)$ cannot be zero, we just need to solve $\partial_z TVN_D(f;z) = 0$. Therefore we have to solve the following equation :

$$\begin{aligned} \partial_z TVN_D(f;z) \quad & \approx -\frac{1}{2\sqrt{2}} \frac{1}{z^2} \left(2\mathrm{erf}(zT_1) - \mathrm{erf}(2zT_1) \right) \\ & - \frac{1}{z^2} \frac{1}{2\sqrt{2}} \left(2\mathrm{erf}(zT_2) - \mathrm{erf}(2zT_2) \right) \\ & + \frac{1}{\sqrt{2}} \frac{T_1}{z\sqrt{\pi}} \left(e^{-z^2 T_1^2} - e^{-4z^2 T_1^2} \right) \\ & + \frac{1}{\sqrt{2}} \frac{T_2}{z\sqrt{\pi}} \left(e^{-z^2 T_2^2} - e^{-4z^2 T_2^2} \right) \\ & = 0. \end{aligned} \quad (35)$$

Hence we deduce Formula (19).

Acknowledgment We thank the CNES for providing us with the images, and Alain Giros (CNES) and Mihaï Datcu (DLR) for fruitful discussions.

References

[1] R. B. Fisher, K. Dawson-Howe, A. Fitzgibbon, C. Robertson, and E. Trucco, *Dictionnary of Computer Vision and Image Processing*. Wiley, 2005.

[2] T. Lindeberg, "Feature detection with automatic scale selection," *IJCV*, vol. 30, no. 2, pp. 79–116, August 1998.

[3] ——, "Feature detection with automatic scale selection," *Int. J. of Computer Vision*, vol. 30, pp. 79–116, 1998.

[4] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *IJCV*, vol. 60, no. 1, pp. 63–86, 2004.

[5] B. Luo, J.-F. Aujol, Y. Gousseau, S. Ladjal, and H. Maître, "Characteristic scale in remote sensing images," in *ICASSP*, vol. II, Toulouse FRANCE, 2006, pp. 809–812.

[6] T. Lindeberg, *Scale-Space Theory in Computer Vision*, ser. Norwell, MA, USA. Kluwer Academic Publishers, 1994.

[7] P. Abry, P. Gonçalvès, and P. Flandrin, *Wavelets, Spectrum Analysis and 1/f Processes*, ser. Lecture Notes in Statistics. Springer, 1995, no. 103, pp. 15–29.

[8] T. Brox and J. Weickert, "A TV flow based local scale measure for texture discrimination," in *ECCV 04*, vol. 2, May 2004, pp. 578–590.

[9] 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 *IEEE-ICIP-97*, vol. II, Santa Barbara, CA, USA, 1997, pp. 234–237.

[10] M. Jägerstand, "Saliency maps and attention selection in scale and spatial coordinates : An information theoretic approach," in *Proc. 5th Int. Conf. on Computer Vision*, 1995, pp. 195–202.

[11] J. Sporring, "The entropy of scale-space," in *ICPR*, vol. A, 1996, pp. 900–904.

[12] J. Sporring and J. Weickert, "On generalized entropies and scale-space," in *Scale-Space Theories in Computer Vision*, 1997, pp. 53–64.

[13] T. Kadir and M. Brady, "Scale, saliency and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[14] C. Woodcock, A. Strahler, and D. Jupp, "The use of variogram in remote sensing II : real digital images," *Remote Sensing and Environment*, vol. 25, pp. 349–379, 1988.

[15] A. Baddeley, *Stochastic Geometry : Likelihood and Computation*. Chapman & Hall / CRC, 1999, ch. Spatial Sampling and Censoring.

[16] M. Lavielle and C. Lévy-Leduc, "Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry signals," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2306–2315, 2005.

[17] J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.

[18] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*, ser. Applied Mathematical Sciences. Springer-Verlag, 2002, vol. 147.

[19] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel, "Axioms and fundamental equations of image processing," *Arch. for Rational Mechanics*, vol. 123, no. 3, pp. 199–257, 1993.

[20] L. Evans and R. Gariepy, *Measure Theory and Fine Properties of Functions*, ser. Studies in Advanced Mathematics. CRC Press, 1992, vol. 19.

[21] S. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 617–643, 1992.

[22] B. Luo, J.-F. Aujol, Y. Gousseau, and S. Ladjal, "Interpolation of wavelet features for satellite images with different resolutions," in *IGARSS*, Denver USA, 2006.

[23] Y. Meyer, "Oscillating patterns in image processing and in some nonlinear evolution equations," March 2001, the Fifteenth Dean Jacquelines B. Lewis Memorial Lectures.

[24] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[25] A. Cohen, *Numerical analysis of wavelet methods*. Elsevier, 2003.

[26] L. Evans, *Partial Differential Equations*, ser. Graduate Studies in Mathematics. American Mathematical Society, 1991, vol. 19.