KCS—New Kernel Family with Compact Support in Scale Space: Formulation and Impact

Lakhdar Remaki and Mohamed Cheriet, *Member, IEEE*

***Abstract*—Multiscale representation is a methodology that is being used more and more when describing real-world structures. Scale-space representation is one formulation of multiscale representation that has received considerable interest in the literature because of its efficiency in several practical applications and the distinct properties of the Gaussian kernel that generate the scale space. Together, some of these properties make the Gaussian unique. Unfortunately, the Gaussian kernel has two practical limitations: information loss caused by the unavoidable Gaussian truncation and the prohibitive processing time due to the mask size. In this paper, we propose a new kernel family derived from the Gaussian with compact supports that are able to recover the information loss while drastically reducing processing time. This family preserves a great part of the useful Gaussian properties without contradicting the uniqueness of the Gaussian kernel. The construction and analysis of the properties of the proposed kernels are presented in this paper. To assess the developed theory, an application of extracting handwritten data from noisy document images is presented, including a qualitative comparison between the results obtained by the Gaussian and the proposed kernels.**

***Index Terms*—Compact support, functional space, handwritten data, handwritten data extraction, image segmentation, kernels, multiscale representation, scale-space representation.**

I. Introduction

An inherent property of objects in the world and details in images is that they only exist as meaningful entities over certain ranges of scale [21]. This shows that the concept of scale is crucial when describing the structure of the world, as well as the structure of images, as in our application. The basic idea is to embed the original image in a one-parameter family of gradually-smoothed signals in which the fine-scale details are successively suppressed. Over the last few decades, mathematical tools have been developed for handling the scale concept in a coherent manner. Various multiscale representations have been proposed, such as quad-trees by Klinger [18], pyramid representation by Burt [4] and Crowley [9], and scale-space representation by Witkin [34] and Koenderink [19]. Almost all these approaches can be considered as particular cases of the multiresolution scheme of wavelet decomposition, which consists of a projection of a signal on a base of functions obtained by the translation and dilation of a single function called a wavelet. This allows decomposing the input signal onto frequency components according to a scale-adaptive resolution. For greater detail on this theory, refer to [11] and [33].

Manuscript received July 15, 1998; revised November 12, 1999. This work was supported under grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fondation du Conseil et Aide à la Recherche du Québec (FCAR). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lina J. Karam.

The authors are with the Imagery, Vision, and Artificial Intelligence Laboratory, École de Technologie Supérieure, Montreal, P.Q., Canada H3C 1K3 (e-mail: remaki@livia.etsmtl.ca; cheriet@gpa.etsmtl.ca).

Publisher Item Identifier S 1057-7149(00)04510-3.

A. Scale-Space Representation

Scale-space representation consists in using a convolution with a chosen kernel in the smoothing operation. Several authors show that under some posed hypotheses, the choice of the Gaussian kernel is unique and offers many beneficial properties, such as the semi-group property, which allows computing a representation at a given scale from any previous scale; a linear change of the scale parameter variable guarantees an invariant property under rescaling of the spatial domain; the strong regularization property; and the fact that the convolution of the Gaussian with any bounded initial condition gives the unique solution of the diffusion equation [21]. Hence, a physical interpretation of the scale-space representation is established. Moreover, when the scale parameter approaches zero, the original signal is recovered. One important property worth mentioning is the zero-crossing diminishing property (or the noncreation of artificial local extrema due to the smoothing operation itself that do not correspond to any important region in the finer-scale representation. This property is stronger than the zero-crossing diminishing property, and it has been shown that the Gaussian is the only kernel that adheres to it). Note that a scale space can be considered as a special case of continuous wavelet representation: simply take the first derivative of the Gaussian as the basic function that defined the translated and rescaled function family in wavelet theory (refer to [10] and [25] for more details on this multiscale representation theory). For additional Gaussian properties, see [21].

B. Motivation and Our Contribution

In scale-space representation, the Gaussian kernel has two practical limitations: information loss due to diminished accuracy when the Gaussian is cut off to compute the convolution product (kernel truncation), and the prohibitive processing time due to the mask's width (which is increased to minimize the loss of accuracy). To avoid this kind of failure, a number of solutions have been proposed in the literature, e.g., approximating the Gaussian using recursive filters [15]. This approach does reduce processing time; however, the information loss remains. In this paper, we propose a new kernel to replace the Gaussian. This kernel must respond to a number of criteria. The first criterion (the principal motivation for this paper) is to recover the information loss and improve processing time. Gaussian truncation is at the root of both these problems. To avoid this truncation, the kernel we propose is built using the compact support analytical

property (i.e., the kernel itself vanishes outside a given compact set). Consequently, we do not need to cut off the kernel when computing the convolution product. Furthermore, the mask is the support of the kernel, and its size is thereby considerably reduced (e.g., from 11.31σ [2], 8σ [20], or 6σ [6] to only 2σ). The second criterion is that the new kernel must retain the most important properties of the Gaussian kernel. It obviously cannot include all the properties that make the Gaussian unique. However, although the Gaussian boasts a vast array of properties, we do not necessarily need all of them to perform image segmentation properly. In this regard, the most important ones are

1) recovering the initial signal (image) when the scale parameter tends toward zero;
2) continuity with respect to the scale parameter;
3) strong regularization property;
4) zero-crossing diminishing property.

As shown in Sections III and IV, our new kernel, Kernel with Compact Support (KCS), retains properties 1)–3) by means of its building process. In addition, tests on a great number of perturbed functions show that property 4) is conserved overall, but we do not have more details. A similar, yet weaker property (the total-variation diminishing property) is proved, however. This new kernel has a compact support, and is derived from the Gaussian kernel by transforming the $\Re^2$ space into a unit ball through a change of variables. This transformation makes it possible to pack all the information in the unit ball. With the new variables, the Gaussian is defined on the unit ball and vanishes on the unit sphere. We then extend it over all $\Re^2$ by taking zero values outside the unit ball to make the convolution product possible. We note that the obtained kernel is still C^∞ (for the space of functions that have derivatives of any order, see [3]).

In Section II, we review the scale-space formulation. In Section III, we detail construction of the KCS, and in Section IV, we investigate its properties. In Section V, an application to extract handwritten data from gray-level images using the KCS and the Gaussian is described, and a series of comparative results is presented.

II. Scale-Space Formulation

A. Overview of the Scale-Space Concept

Different multiscale representations have been developed in the past few decades. The scale-space representation proposed by Witkin [34] and Koenderink [19] is widely used in practice for several applications because it offers accurate interpretation and very practical handling of the multiscale concept. The general idea of this formulation is to use a convolution with a given one-parameter family of kernels in the smoothing operation. Some questions naturally arise:

1) What relates these different spaces to the original signal?
2) How are the various generated spaces related to one another? This can be understood as a continuity with respect to the scale parameter.
3) What about the derivatives of the smoothed images? Can we differentiate them in any order we need to get a precise study of their local behavior?
4) How do the zero-crossings evolve when increasing the scale parameter?

The Gaussian kernel offers adequate responses to these questions, the fourth one in particular. This subject is widely discussed by several authors because it is an important property and its formulation is not obvious. Also, several authors under some posed hypothesis have proven the uniqueness of the Gaussian in scale-space representation in various manners. Koenderink [19] proved the uniqueness by combining causality, homogeneity and isotropy principles; Lindeberg [22]–[24] has combined avoiding the introduction of new extrema with increasing scale using the semi-group property; Florak *et al.* [12] combined the semi-group property with uniform scaling over scales. Other ways have been shown in Yuille and Poggio [35], Babaud *et al.* [1], and Hummel [16]. The Gaussian kernel possesses several other interesting properties widely detailed in [21].

B. Practical Limitations of the Gaussian Kernel and Our Solution

As mentioned above, in a scale-space formulation a convolution product with a Gaussian is computed at each scale. Although the integral defining the convolution must be computed over all the $\Re^2$ space, practically speaking, it is approached by its computation over a bounded set of $\Re^2$, commonly known as a mask. The accuracy of the computation depends on the mask size. Wider masks provide more precise computations, but increase the cost of processing time; smaller mask sizes decrease the processing time, but the accuracy is sometimes severely diminished, which induces information loss. As such, two fundamental practical limitations of the Gaussian kernel can be raised: information loss and prohibitive processing time. Some solutions to overcome these problems are proposed in the literature, such as the approximation of the Gaussian by recursive filters or using truncated exponential functions instead of the Gaussian [21]. The alternatives based on approximating the Gaussian reduce processing time, but the information loss remains, and is sometimes increased. These two problems constitute the main motivations for this paper. In order to recover the information loss without increasing the mask size considerably, we propose a kernel with a compact support. Thus, there is no need to cut off the kernel while the processing time is controlled because the mask is the support of the kernel itself. From this standpoint, we built our new KCS kernel. This kernel is derived from the Gaussian and preserves the properties 1)– 3) mentioned in the introduction. Furthermore, tests on a great number of perturbed functions show that property 4) is conserved overall, but we do not have more details. A similar and weaker property (the total-variation diminishing property) is proved, however. In addition, the KCS guarantees the most important properties required to perform image segmentation properly without contradicting the uniqueness of the Gaussian for such a purpose. In the following section, details on the construction of the KCS and investigation of some of its properties are given.

III. Building the KCS

The property of recovering the original signal when the scale parameter becomes close to zero is necessary for defining a

scale-space representation. In general, this property is not guaranteed for any kernel. For the Gaussian, the associated heat equation proves this property; however, this argument is not valid for the KCS. A fundamental convergence theorem in the functional analysis field guarantees this property under some hypotheses on the basic function defining the kernel family. The most important of these hypotheses is the compactness of the support, which is precisely the main purpose behind our kernel construction. We will consequently build our kernel in light of this theorem. Before we begin, let us recall the convergence theorem and some of the mathematical fundamentals necessary for understanding it.

Definition 1: We call functional space $L^p(\Re^2)$ with $1 \leq p \leq \infty$ all spaces

$$L^p(\Re^2) = \left\{ f : \Re^2 \to \Re \Big/ \int_{\Re^2} |(f)|^p \, dx < \infty \right\}.$$

This space with

$$\left(\int_{\Re^2} |(f)|^p \, dx \right)^{1/p}$$

norm is a Banach space.

Definition 2: We define the space of test functions (smooth functions with compact support) as

$$D(\Re^2) = \{\Phi \in C^\infty \text{ so that}, \exists \text{ a compact } K / \text{support}\, \Phi \subset K\}.$$

Definition 3: Let ρ be a nonnegative function in $D(\Re^2)$ satisfying the following conditions:
(A)

$$\int_{IR^2} \rho(x,y)\, dx\, dy = 1 \tag{1}$$

(B)

$$\text{support}\, \rho = \{(x,y) \in IR^2; x^2 + y^2 \leq 1\} = B(0,1). \tag{2}$$

For each $\sigma > 0$, we define

$$\rho_\sigma(x,y) = \frac{1}{\sigma^2} \rho\left(\frac{x}{\sigma}, \frac{y}{\sigma}\right) \tag{3}$$

$\rho_\sigma \in D(IR^2)$, it is nonnegative and satisfies
(A)

$$\int_{IR^2} \rho_\sigma(x)\, dx = 1 \tag{4}$$

(B)

$$\text{support}\, \rho_\sigma = \{(x,y) \in IR^2; x^2 + y^2 \leq \sigma^2\} = B(0,\sigma). \tag{5}$$

These functions ρ_σ are called *mollifiers*.

Theorem 1 (See [3]): Let $f \in L^p(\Re^2)$, for a given p with $1 \leq p \leq \infty$ then

$$\begin{aligned} &\rho_\sigma * f \in L^p(\Re^2) \\ &\rho_\sigma * f \underset{\sigma \to 0}{\to} f \quad \text{in } L^p(\Re^2) \end{aligned} \tag{6}$$

i.e.,

$$\lim_{\sigma \to 0} \int_{\Re^2} |(\rho_\sigma * f - f)|^p \, dx\, dy = 0. \tag{7}$$

Note that if f is an image, and $\tilde{f}$ is its extension by 0 for all $\Re^2$, then $\tilde{f} \in L^p(\Re^2)$ for all $1 \leq p < \infty$. Therefore, the previous theorem holds for every value of p, and we obtain the original image for $\sigma = 0$ in the sense given by the theorem. Now let us give the technical details of building the KCS.

A. Kernel Construction

To clarify the construction of the KCS, we proceed in two steps. The first is concerned with the topological deformation of the plane, and then the deduction of the appropriate change of variables that will be applied to the Gaussian.

Let $g_\sigma(x,y)$ be a two-dimensional (2-D) one-parameter family of normalized symmetrical Gaussian. We observe that this family can be obtained in the following manner.

Let $g(x,y)$ be the function

$$g(x,y) = \frac{1}{2\pi} e^{-[\frac{x^2+y^2}{2}]} \tag{8}$$

then

$$g_\sigma(x,y) = \frac{1}{\sigma^2} g\left(\frac{x}{\sigma}, \frac{y}{\sigma}\right) = \frac{1}{2\pi\sigma^2} e^{-[\frac{x^2+y^2}{2\sigma^2}]}. \tag{9}$$

In order to build the KCS kernel, we apply a change of variables to g_σ and define the one-parameter family of KCS kernels as in (9). Let $\Gamma(r,\theta)$ be the function that defines the polar change of coordinates

$$\Gamma(r,\theta) = (r\cos\theta, r\sin\theta). \tag{10}$$

Now define the function ω as follows:

$$\begin{aligned} &[0,1[\to \Re^+ \\ &t \to \omega(r) = \sqrt{\frac{1}{1-r^2} - 1} \quad \text{for } 1 > r \geq 0 \end{aligned} \tag{11}$$

where ω is a C^1 diffeomorphism; its inverse transforms $\Re^+$ into $[0,1[$ in a continuous manner. Thus, $\Gamma(\omega(r),\theta)$ transforms the unit ball $B(0,1)$ into an $\Re^2$ plane. Note that the choice of ω is not unique. However, within the framework of reducing the computational cost of the transform function as much as possible, and assuming that the desired function only depends on the square of its argument (as justified below), ω [as defined in (11)] is optimal. This claim can be justified as follows.

Note that the Gaussian is a function of $|\vec{x}|^2$ (the Euclidean norm of $\vec{x}$ in $\Re^n$), and that ω will replace the $\vec{x}$ norm in the expression of the Gaussian, i.e., $|\vec{x}| = \omega(r)$. ω should be a function of r^2 in order to obtain an expression comparable to that of the Gaussian, and to facilitate the change of variables from polar to Cartesian coordinates and vice-versa. If we set $r^2 = z$, the function $f(z) = (1/1-z)$ transforms the interval $[0,1[$ into $[1,+\infty[$. The f function values are only obtained after two operations. By only performing a single operation, we could not carry out the last topological transform. Lastly, by translating the function f to reach all $\Re^+$, we obtain the expression of $\omega^2(r)$. In the Gaussian expression, the standard deviation σ controls the

Gaussian's bandwidth, we therefore shall exploit this characteristic to derive the KCS, we first set $\gamma = 1/2\sigma^2$, and then apply the change of variables $(x, y) = \Gamma(\omega(r), \theta)$ in the expression of $(1/2\pi)\exp[-\gamma[x^2 + y^2]]$ and extend it by zero outside the unit ball. After normalization, we obtain the desired KCS kernel

$$\rho_{\sigma,\gamma}(r,\theta) = \begin{cases} \dfrac{1}{C_\gamma} e^{\frac{1}{2}(\frac{\gamma}{r^2-1}+\gamma)}, & \text{if } r^2 < 1 \\ 0, & \text{elsewhere.} \end{cases} \tag{12}$$

Or, with the variables (x, y)

$$\rho_{\sigma,\gamma}(x,y) = \begin{cases} \dfrac{1}{C_\gamma} e^{\frac{1}{2}(\frac{\gamma}{x^2+y^2-1}+\gamma)}, & \text{if } x^2 + y^2 < 1 \\ 0, & \text{elsewhere} \end{cases} \tag{13}$$

where

$$C_\gamma = \int_{-1}^{1}\int_{-1}^{1} e^{\frac{1}{2}(\frac{\gamma}{x^2+y^2-1}+\gamma)} 1_{B(0,1)}(x,y)\,dx\,dy$$

a normalization constant and

$$1_{B(0,1)}(x,y) = \begin{cases} 1, & \text{if } x^2 + y^2 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

This function satisfies

$$\rho_{\sigma,\gamma} \in C^\infty, \quad \rho_{\sigma,\gamma} \in D(IR^2) \quad \text{and} \quad \int_{IR^2} \rho_{\sigma,\gamma}(x,y) = 1. \tag{14}$$

We can now define the one-parameter family of KCS kernels as

$$\rho_{\sigma,\gamma}(x,y) = \frac{1}{\sigma^2}\rho_{\sigma,\gamma}\left(\frac{x}{\sigma}, \frac{y}{\sigma}\right)$$

which gives

$$\rho_{\sigma,\gamma}(x,y) = \begin{cases} \dfrac{1}{C_\gamma \sigma^2} e^{(\frac{\gamma\sigma^2}{x^2+y^2-\sigma^2}+\gamma)}, & \text{if } x^2 + y^2 < \sigma^2 \\ 0, & \text{elsewhere} \end{cases}$$

According to Definition 3, $(\rho_{\sigma,\gamma})$ is a family of mollifiers, and so Theorem 1 holds. The support of $\rho_{\sigma,\gamma}$ is $B(0,\sigma)$. It is therefore unnecessary to take mask sizes greater than 2σ. Fig. 1(a) depicts the curve of ρ_σ for a given σ. Practically speaking, the constant C_γ can be computed with a high level of accuracy using numerical integration, since the integration domain is bounded. More specifically

$$C_\gamma = \Delta x \Delta y e^{\frac{\gamma}{2}} \sum_{j=0}^{M}\sum_{i=0}^{N} e^{\frac{1}{2}(\frac{\gamma}{x_i+y_j})} 1_{B(0,1)}(x_i, y_j)$$

where

$$x_i = -1 + i\Delta x, \quad y_{ji} = -1 + j\Delta x \quad \Delta x = \frac{2}{N},$$
$$\Delta y = \frac{2}{M}$$

where N and M are the number of points in the discretization process. This double integral can be simplified to a single integral using polar coordinates

$$C_\gamma = 2\pi \int_0^1 e^{\frac{1}{2}(\frac{\gamma}{r^2-1}+\gamma)} r\,dr = \pi \int_0^1 e^{\frac{1}{2}(\frac{\gamma}{r-1}+\gamma)}\,dr$$
$$\approx \Delta r \pi \sum_0^N e^{\frac{1}{2}(\frac{\gamma}{r_i-1}+\gamma)}$$

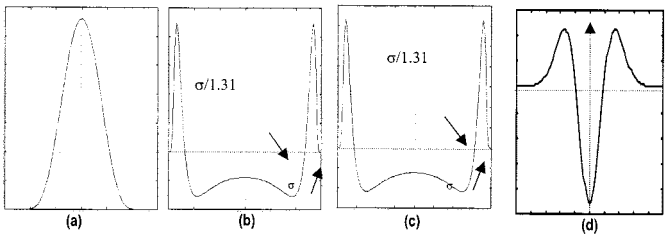

Fig. 1. KCS and the Laplacian of the KCS and Gaussian curves ($\gamma < 2$).

where $r_i = i\Delta r$, $\Delta r = 1/N$. Usually, the parameter γ is kept constant and only σ varies, as is the case for our application. As such, the constant C_γ is computed once at most during the process. The computational cost of the process is consequently unaffected. We do not even need the value of this constant in Section V because we are only concerned by the sign of the KCS' Laplacian. However, for other applications where the parameter γ is likely to vary during the process, we propose a high-speed algorithm that computes C_γ for every γ from C_{γ_0} and C_{γ_1} for the given values γ_0, γ_1. We consider the case of P-dimensions. (Note that in P-dimensions, the KCS formula is obtained by replacing the Euclidean norm $\Re^2$ in by the Euclidean norm in $\Re^p$. i.e. $x^2 + y^2$ by $\|\vec{x}\|^2 = \sum_{i=1}^{P} x_i^2$.)

Proposition 1: If $h(\gamma)$ is the solution of the following second-order differential equation:

$$h''(\gamma) - \left(\frac{1}{4} + \frac{P}{2\gamma}\right) h = 0, \quad \gamma \in [\gamma_0, +\infty[$$

with $\gamma_0 > 0$ on $[\gamma_0, +\infty[$ with $\gamma_0 > 0$, then $C_\gamma = h(\gamma)e^{\gamma/2}$ (for an appropriate initial condition).

(Proof: see the Appendix.)

Now we discretize the equation of the proposition using the second-order Taylor expansion, and obtain the following numerical scheme:

$$\begin{cases} h_{i+1} = \left(\Delta\gamma^2\left(\dfrac{1}{4} + \dfrac{P}{2\gamma_i}\right) + 2\right) h_i + h_{i-1} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{then } C_{\gamma_i} = h_i e^{\frac{1}{\gamma_i}} \\ h_0 = C_{\gamma_0}, \quad h_1 = C_{\gamma_1} \end{cases}$$

where $\Delta\gamma$ is the meshsize (step) of the discretization and $\gamma_i = \gamma_0 + i\Delta\gamma$. As we can see, we only compute C_γ for two values using discrete integration; all other values are obtained with a few operations with the numerical scheme above. Also note that this scheme is given for any dimension P, hence it is very useful for higher dimensions. Now let us give accuracy results for the recursive computation algorithm of h.

Proposition 2: For each γ_i $(i \ge 0) \in [\gamma_0, +\infty[$, we have $|h(\gamma_i) - h_i| = O((\Delta\gamma)^2)$. (Proof: see the Appendix.)

$\rho_{\sigma,\gamma}$ curves exhibit precisely the same behavior as the Gaussian [Fig. 1(a)]. However, does the same apply to their successive derivatives? Fig. 1(b) and (c), which depict the KCS Laplacian curves for $\sigma = 15$, $\gamma = 1$ and $\sigma = 20$, $\gamma = 1$, respectively, show that a new maximum appears at the origin when compared to the Gaussian Laplacian [Fig. 1(d)]. Since one of our concerns when building the KCS was to preserve a large number of Gaussian properties, and since in practice the first and second derivatives are used most frequently, we would like to guarantee that at least the first and second KCS

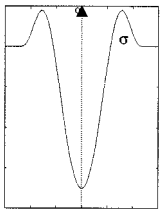

Fig. 2. Behavior of the KCS Laplacian (LoKCS) ($\gamma > 2$).

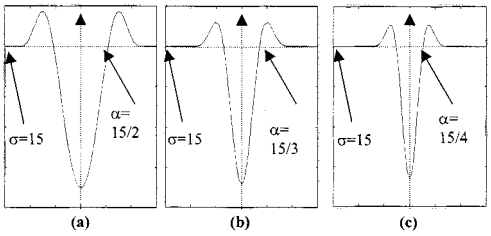

Fig. 3. The influence of parameter γ on the width of the peak in the KCS Laplacian, for $\sigma = 15$.

derivatives have the same behavior as the Gaussian's. We give in the following a lemma that proves that for a fixed chosen value of γ, the desired characteristics of the KCS derivatives are guaranteed (see Figs. 2 and 3). We will see again that the parameter γ plays another role that appears to be important in certain applications.

Lemma 1: If $\gamma > 2$, the first and second KCS derivatives (in one or two dimensions) exhibit the same behavior as the Gaussian's. (Proof: see the Appendix.)

Let α be the distance from the zero-crossing of the KCS Laplacian to the origin of the axes (see Fig. 2). As we will show in a practical case, it might be of interest if this distance could be controlled. We will show its effectiveness on the quality of output data in Section 5. The parameter γ allows us to control the value of α, as stated in the following lemma.

Lemma 2: If

$$\begin{cases} \gamma = \left(\frac{1}{2}n^2 - \frac{3}{2}\frac{1}{n^2} + 1\right) & \text{in one dimension} \\ \text{or} \\ \gamma = \left(n^2 - \frac{1}{n^2}\right) & \text{in two dimensions} \end{cases}$$

then for all σ, $\alpha = \sigma/n$, where n is an integer or real number. (Proof: see the Appendix.)

It could be concluded from this lemma that for a fixed n, γ is fixed and does not depend on the value of σ. This constitutes an important property, since in scale space we have several representations of the original signal corresponding to different values of the standard deviation σ. Fig. 3(a)–(c) show the influence of γ (for $n = 2, 3, 4$) on the width of the peak in the KCS Laplacian for a fixed σ value.

IV. Investigation of Selected Properties of the KCS

A. Morphological–Analytical Properties

Figs. 1(a), 2, and 3 depict the curve of $\rho_{\sigma,\gamma}$ (for $\gamma \geq 2$) and its Laplacian. These curves show the same behavior as the Gaussian. In the following, we discuss selected other properties of the KCS.

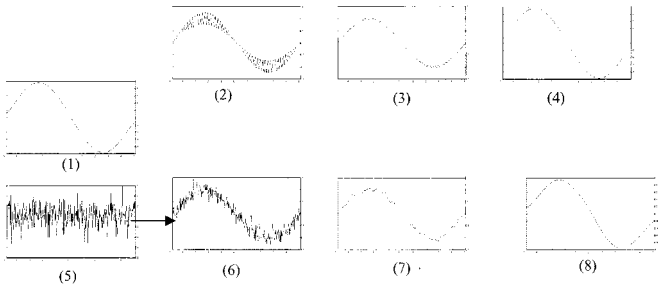

Fig. 4. Sinusoidal and white noise elimination using the KCS kernel.

1) Recovering the original signal when the scale parameter is close to zero

In Section III, we showed that the KCS satisfies a convergence theorem that ensures the recovering of the original signal when the scale parameter becomes close to zero. This property is a necessary condition for defining a scale-space representation, and in general is not guaranteed by any kernel.

2) Continuity (in the topological sense)

According to the well-known Lebesgue Theorem (see [3], for example), and since the KCS is bounded uniformly over $[\beta, +\infty[$ for all $\beta > 0$, the scale-space representation is continuous with respect to the scale parameter σ, in $]0, +\infty[$. This means that there will be no jump between successive representation spaces of the original signal.

3) Strong regularization

Thanks to the exponential function in the KCS expression, spatial derivatives of any order of $\rho_{\sigma,\gamma}$ are in $D(\Re^2)$. Hence, the convoluted signal with KCS is C^∞, according to the spatial coordinates. This means that the smoothed signal (image) can be derived in any order.

4) Regularization and zero-crossings over scales

Convolution with the KCS has been applied to a variety of perturbed signal in order to observe how the KCS operates with respect to smoothing, in particular, the evolution of the number of zero-crossings when increasing the scale parameter. We present here a sample of a one-dimensional (1-D) perturbed sinusoidal function and its convolution with the KCS for different σ values. Two kinds of noise are considered: sinusoidal noise and white noise generated by Matlab. In Fig. 4, we show the original signal, the perturbed signals, and two steps of the smoothing process. We observe that, for an appropriate value of σ ($\sigma = 1.5$ for the sinusoidal noise and $\sigma = 2$ for the white noise), we can perfectly restore the original signal. Another important observation is the diminishing of the number of extremas (or zero-crossing diminishing). This property was observed over all the perturbed signals we tested; we can consider that it is satisfied overall, but we do not have more details. In this section, we prove an interesting property nevertheless: instead of the zero-crossing diminishing property, we consider the total-variation diminishing property. This can be understood as follows: globally, the distance between two successive extremas decreases, which can be considered as a weak formulation of the zero-crossing diminishing or noncreation of artificial local extremas property. A definition and the theorem proving the total-variation diminishing property are given as follows.

Definition 4: We denote by $W^{1,1}(IR^2)$, the familiar Sobolev space

$$W^{1,1}(\Re^2) = \left\{ f \in L^1(\Re^2); \partial^\beta f \in L^1(\Re^2), \quad \forall \beta \in IN^2 | \beta | \leq 2 \right\}$$

and by $L^1_{loc}(IR^2)$, the functional space $L^1_{loc}(\Re^2) = \{f;$ for each compact set $K, f \in L^1(K)\}$.

Definition 5: Let g be a given function; we define the total variation of g over $\Re^2$ by

$$\mathrm{TV}_{IR^2} = \mathrm{Sup}\left\{ \sum_j |g(x_j) - g(x_{j-1})|, \quad (x_k)_k \text{ is a sequence in } \Re^2 \right\}.$$

Definition 6: If $g \in L^1_{loc}(IR^2)$, $\mathrm{TV}_{IR^2}(g)$ can be defined as

$$\mathrm{TV}_{IR^2}(g) = \mathrm{Sup}\left\{ \int_{IR^2} g \cdot \operatorname{div} \theta \, dx \, dy, \theta \in D(IR^2)^2, \quad \|\theta\|_{L^\infty(IR^2)} \leq 1 \right\}$$

where

$$\|\theta\|_{L^\infty(IR^2)} = \max_{i=1,2} \left(\sup_{x \in IR^2} |\theta_i(x)| \right).$$

Since the quantity TV_{IR^2} is generally not finite, we then define the useful space of bounded variation functions (BV-functions) as follows (for more details, see [13]).

Definition 7: A function $g \in L^1_{loc}(IR^2)$ is said to have a bounded variation in $\Re^2$ if $\mathrm{TV}_{IR^2}(g) < +\infty$. We set

$$BV(\Re^2) = \left\{ g \in L^1_{loc}(\Re^2); \mathrm{TV}_{IR^2}(g) < +\infty \right\}.$$

Remark 1: If g belongs in $W^{1,1}(\Re^2)$, we have

$\mathrm{TV}_{IR^2}(g) = \int_{IR^2} |\operatorname{grad} g| \, dx \, dy$. Thus, $W^{1,1}(IR^2) \subset BV(IR^2)$.

Let us now formulate the theorem of total-variation diminishing.

Theorem 2: Let I be a given signal belonging to $BV(IR^2)$, and $\rho_{\sigma,\gamma}$ a normalized KCS kernel. Suppose that the support of I is bounded. If we set $g_{\sigma,\gamma} = \rho_{\sigma,\gamma} * I$ then $\forall \sigma > 0, \gamma > 0$, we have

$$TV_{IR^2}(g_{\sigma,\gamma}) \leq TV_{IR^2}(I).$$

We say that $g_{\sigma,\gamma}$ has a total-variation diminishing property. (Proof: see the Appendix.)

5) Stability of the KCS family under the convolution operation

The convolution of two KCS ρ_{σ_1,γ_1} and ρ_{σ_2,γ_2} is a function with compact support equal to the sum of the supports of the original functions, and since the KCS is a symmetric function, this support is exactly equal to the ball $B(o, \sigma_1, +\sigma_2)$. Consequently, if we suppose that $\rho_{\sigma_1,\gamma_1} * \rho_{\sigma_2,\gamma_2}$ is a KCS (stability under the convolution operation), it will be written as $\rho_{\sigma_1,\sigma_2,\gamma}$ for an adequate γ. Using a discrete integration to compute the convolution product, we have observed that such γ exist. This is obviously not sufficient, but is highly interesting and encourages us to seek a thorough theoretical proof in a future research.

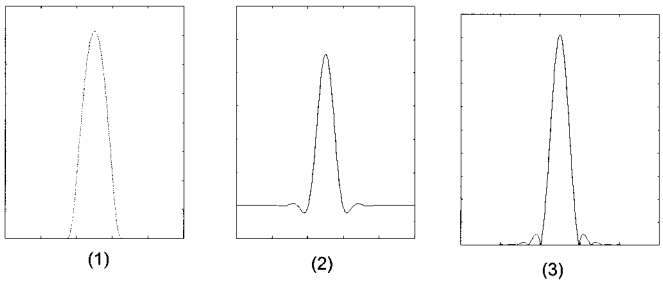

Fig. 5. KCS and its Fourier transform.

B. Spatial-Frequency Properties

1) Fourier Transform of the KCS: In this section, we will investigate the spatial-frequency properties of the 1-D KCS. Let $\rho_{\sigma,\lambda}$ be the Fourier transform of the KCS, given by

$$\tilde{\rho}_{\sigma,\lambda}(\xi) = \int_{-\infty}^{+\infty} \rho_{\sigma,\gamma}(x) e^{-2\pi i x \xi} \, dx = \int_{-\sigma}^{+\sigma} \rho_{\sigma,\gamma}(x) e^{-i2\pi x \xi} \, dx.$$

Since the support of the KCS is compact, the integral is computed over a bounded interval $]-\sigma, +\sigma[$. This way, a numerical approximation can be used with a very high level of accuracy. Let Δx be a given meshsize, we then have the approximation

$$\tilde{\rho}_{\sigma,\lambda}(\xi) = \Delta x \sum_0^N \rho_{\sigma,\gamma}(x_i) e^{-2\pi i x_i \xi}$$

$$\text{where } x_i = -\sigma + i\Delta x, \quad \Delta x = \frac{2\sigma}{N}.$$

Fig. 5 shows the curves of a KCS (with $\sigma = 3.5, \gamma = 2.5$), its Fourier transform and the absolute value of the Fourier transform. Note that for this γ value, the bandwidth in the frequency domain is in the same order as the duration in the spatial domain; in other words, there is no information dispersion between the two.

2) Heisenberg Uncertainty Principle: The Heisenberg uncertainty principle, also known as the duration bandwidth uncertainty principle, is an interesting concept in communication theory. Given any 1-D function V with Fourier transform $\tilde{V}$, define the normalized second-order moments $\langle V \rangle$ and $\langle \tilde{V} \rangle$ in the spatial and Fourier domains respectively by

$$\langle V \rangle^2 = \frac{\int_{-\infty}^{+\infty} x^2 V^2(x) \, dx}{\int_{-\infty}^{+\infty} V^2(x) \, dx}, \quad \langle \tilde{V} \rangle^2 = \frac{\int_{-\infty}^{+\infty} x^2 \tilde{V}^2(x) \, dx}{\int_{-\infty}^{+\infty} \tilde{V}^2(x) \, dx}.$$

Then, the uncertainty relationship states that (see [29])

$$\langle V \rangle \langle \tilde{V} \rangle \geq \frac{1}{4\pi}.$$

This relationship means that a waveform cannot simultaneously have an arbitrarily small bandwidth and an arbitrarily small duration. Let us first see what happens when the parameter γ is close to zero. The response is given by the following lemma.

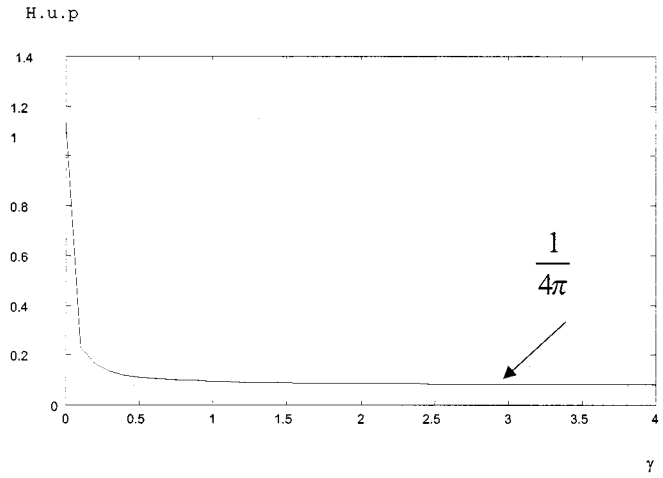

Fig. 6. Curve of the Heisenberg uncertainty product according to parameter γ.

Lemma 3: For every $1 \leq p < \infty$ we have $\rho_{\sigma,\gamma} \rightarrow_{\gamma \rightarrow 0} h_\sigma$ in $L^p(IR)$ where h_σ is the gate function defined by

$$h_\sigma(x) = \begin{cases} \dfrac{C_\lambda}{2\pi\sigma}, & \text{if } x^2 < \sigma^2 \\ 0, & \text{elsewhere.} \end{cases}$$

(Proof: see the Appendix.)

The Fourier transform of h_σ is the Sinus cardinal function given by

$$h_\sigma(\tilde{\xi}) = \frac{C_\lambda}{2\pi^2\sigma} \frac{\sin(\sigma\xi)}{\xi} \xi \in \Re.$$

The second moment of this function is not bounded, therefore the Heisenberg uncertainty product (H.u.p.) is not bounded either. More precisely, we have shown that the H.u.p. of the KCS explodes when the parameter tends toward zero. Fortunately, however, and according to Lemma 1, we are not concerned by the small values of γ. Let us now investigate this feature for different γ values. The H.u.p. is numerically computed with a precision of 10^{-6}. Note that the H.u.p. does not depend on the scale parameter; so in Fig. 6, σ is set to 1, and γ varies from 10^{-9} to 50 with step 0.2. The curve obtained shows that when γ increases, the H.u.p. decreases very quickly and becomes close to $1/4\pi \approx 0.079\,617$, which is the optimal value reached by the Gaussian kernel.

V. Application to Extracting Handwritten Data

Automatic segmentation, which has received considerable attention in the literature [5]–[8], [30]–[32], is the most difficult task in image processing. In this section, we investigate the practical aspects of the KCS and its impact on extracting handwritten data from degraded and noisy images as a target application. A comparison is also established between results obtained by using the Gaussian and the KCS. Before we begin, however, we shall briefly recall the methodology used in this application.

To obtain multiscale representations of a discrete signal I, it is possible to define a set of discrete functions I_i ($\mathsf{i} \in [0 \ldots p]$) as in [21], where I_0 represents the original image and each new level is calculated by convolution from the previous one as follows:

$$i = 1, \ldots I_i = I_{i-1} * K_i \quad \text{where } K_i \text{ is a given kernel.} \tag{15}$$

The purpose of this application is to extract data from handwritten documents and business forms. We focus on segmenting noisy data with strong intensity variation, such as for postal mail envelopes [17], [28]. In the literature, the Gaussian kernel is used in edge detection [2], [8], [26]. In [6] and [7], we presented a generalization of using the LoG operator for full-shape data segmentation (for more details concerning the LoG operator, refer to [20]). The methodology is briefly described in this section. We refer by LoKCS to the Laplacian of KCS. The operator is defined by convoluting the image with LoG

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2},$$

$$\text{then LoG} = \nabla^2 g(x,y) = \left(\frac{x^2+y^2}{\sigma^4} - \frac{1}{\sigma^2} \right) e^{-[\frac{x^2+y^2}{2\sigma^2}]}$$

and LoKCS as shown in the formula at the bottom of the page. The decision criterion in segmenting data [6] is the detection of the convex parts of the smoothed image I_I at each scale level of resolution (ith step); it is determined by the sign of $\text{LoKCS} * I_{i-1}$ ($\text{LoG} * I_{i-1}$). Since most of the information for the Gaussian is contained within the range [6σ [20], 11.32σ [2]], where σ is the standard deviation of the Gaussian, the mask size will be in this range in order to recover the maximum information when using the Gaussian. For the KCS, we have shown that all the information is contained in the interval $]-\sigma, \sigma[$, so the mask size is equal to only 2σ. We have used the algorithm proposed in [6] to perform the experiments, and replaced the LoG operator by the LoKCS operator. In this algorithm, (15) is replaced by $I_i = I_0 * K_i$, which means that the corresponding image is generated from the original at each scale. For the Gaussian, the semi-group property ensures that these two formulas are equivalent. For the KCS, this could be justified by the stability property numerically observed (Section IV-A). At the very least, this last formula can be used even if we do not have theoretical proof of the stability. In this case, we speak of multiscale representation instead of scale space.

VI. Experimental Results and Discussion

In this section, a practical comparison is made between the LoG and LoKCS operators. Some examples are illustrated in

$$\text{LoKCS} = \nabla^2 \rho_{\sigma,\gamma}(x,y) = \begin{cases} \dfrac{\gamma}{C_\gamma} \left[\dfrac{(x^2+y^2)^2 + \gamma\sigma^2(x^2+y^2) - \sigma^4}{(x^2+y^2-\sigma^2)4} \right] e^{(\frac{\gamma\sigma^2}{x^2+y^2-\sigma^2}+\gamma)}, & \text{if } x^2+y^2 < \sigma^2 \\ 0, & \text{elsewhere} \end{cases}$$

order to show the influence of the γ parameter on the shape of the segmented images.

A. Practical Comparison Between LoG and LoKCS

To compare their performance, our tests focused exclusively on qualitative comparison. The methodology we followed to perform this comparison is based on the capacity of each kernel to resist to progressively degraded input images without adapting their parameters to the degradation. In other words, this methodology measures the influence of information loss on the results, as well as processing time. To perform this task, we have set appropriate parameters with respect to a certain number of images for both the LoKCS and LoG operators. We then used these parameters to process a large image database. We center the discussion on four sets of images.

1) In the first set (a) [see Fig. 7(a)], both operators yield respectable results with high visual quality. However, the processing time required for the LoKCS operator is drastically less than that required by the LoG operator. Indeed, processing time is akin to the mask size, which is equal to 2σ for the LoKCS. For the LoG, we have chosen the mask size as suggested in [2], i.e., mask size $= 11.32\sigma$. In our tests, for the multiscale representation we decreased the σ value from 4 to 2 with a step of 0.5 for the LoKCS and from 3 to 1 with the same step of 0.5 for the LoG. In terms of mask dimension, from 8×8 to 4×4 for the LoKCS and from 33.96×33.96 to 11.32×11.32 for the LoG. We can appreciate the reduction of the mask dimension and hence the overall processing time.
2) In the second set (b) [see Fig. 7(b)], we can see that the LoKCS always yields satisfactory results (with the same parameters), but a certain degradation appears on the images segmented by the LoG. We can explain this phenomenon by the recovered information that is lost when using the Gaussian.
3) The third set (c) [see Fig. 7(c)] represents samples of degraded images. With the same parameters used in (a) and (b), we observe that the information loss when using the Gaussian kernel becomes too sensitive, and we appreciate the capability of the KCS kernel to recover it, not to mention the processing time gained.
4) In set (d) [Fig. 7(d)], the results show the limitations of both kernels when dealing with extremely degraded images.

We can conclude that to minimize the information loss when using the Gaussian and improve the results obtained for sets (b) and (c), we must adapt the Gaussian's parameters to the images of these sets by sufficiently increasing the mask size. This naturally brings about a drastic amplification of the processing time, which is already too long when compared to using the KCS. On the other hand, the parameter settings for the KCS remain unchanged and robust for the different sets of data, leading us to make the assumption of "easy generalization," depending on the applications.

We were curious to see what would happen when we set the mask dimension of the Gaussian equal to that of the KCS, which means that the processing time is the same for both kernels. To

Fig. 7. Comparison between LoKCS and LoG results.

vary the LoG mask dimension from 8×8 to 4×4 (as for the LoKCS), we keep the same interval variation for σ as above and set mask size $=2.6\sigma$, while the KCS parameters are left unchanged. Fig. 7(e) shows the results obtained on a contrasted image for both kernels. We can see the drastic information loss when the LoG's processing time is forced to be the same as that of the LoKCS.

B. Influence of γ Value on the Experimental Results

To observe the practical influence of the γ value, we have segmented the image in Fig. 8 using the same σ parameter range used in Section VI-A for different γ values. This test shows how the parameter γ influences the shape of the characters. In

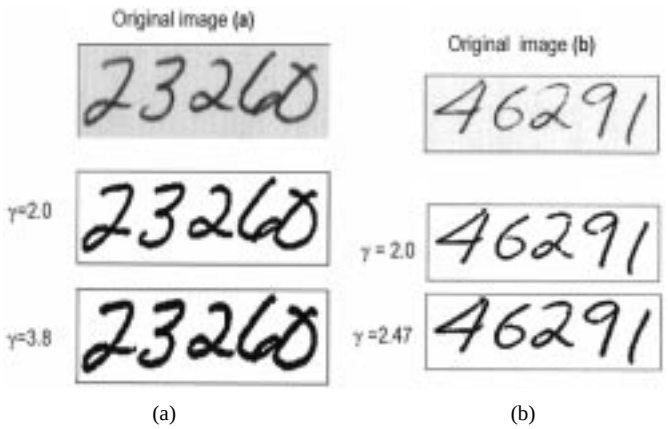

Fig. 8. Influence of the γ value (a) on the shape width of the characters (b) and its role in improving the image segmentation.

Fig. 8(a), we observe that the thickness of the characters diminishes when decreasing the γ value. Hence, an appropriate choice of γ may also improve the segmented image quality as in Fig. 8(b). By virtue of Lemma 1, γ should be made greater than 2. More experiments are currently underway on a large database from SUNY Buffalo (CEDAR) in order to more precisely evaluate the KCS algorithm settings.

VII. Conclusion

In this paper, we have presented a new family of kernels (KCS) to generate scale-space in multiscale representation, where the Gaussian kernel is usually employed. The new kernels are derived from the Gaussian by deforming the plane ($\Re^N$ space in the general case) into a unit ball. We therefore obtain kernels with compact support. We first have shown that the transform function used is optimal in terms of computational cost, and that the obtained kernels preserve the most important properties of the Gaussian kernel in order to perform image segmentation efficiently. In addition, a thorough study of complementary analytical, morphological and statistical properties are provided by giving theoretical proofs whenever possible, and numerical proofs in the other cases. Thanks to their compact support property, the proposed kernels give an effective response to both practical limitation problems when using the Gaussian kernel, namely the information loss caused by the truncation of the Gaussian kernel and the prohibitive processing time due to the wide mask size. We have presented an application to extracting handwritten data, and have made a qualitative comparison of the results obtained by both the Gaussian and the KCS kernels. For this purpose, we designed a methodology able to measure the influence of the information loss and the gain in processing time. These results confirm the theoretical virtues of the KCS we have shown.

Appendix

Proof of Proposition 1: In P-dimensions the KCS formula is given by

$$\rho_{\sigma,\gamma}(x_1,x_2,\ldots,x_p)=\begin{cases}\dfrac{1}{C_\gamma\sigma^2}e^{\left(\frac{\gamma\sigma^2}{||\vec{X}||^2-\sigma^2}+\gamma\right)}, & \text{if } ||\vec{X}||^2<\sigma^2\\ 0, & \text{elsewhere}\end{cases}$$

where $||\vec{X}||^2=\sum_{i=1}^{P}x_i^2$. Let $\varphi_{\sigma,\gamma}$ be the none normalized KCS, i.e.,

$$\varphi_{\sigma,\gamma}(x_1,x_2,\ldots,x_p)=\begin{cases}\dfrac{1}{\sigma^2}e^{\left(\frac{\gamma\sigma^2}{||\vec{X}||^2-\sigma^2}\right)}, & \text{if } ||\vec{X}||^2<\sigma^2\\ 0, & \text{elsewhere}\end{cases}$$

then $C_\gamma=e^\gamma\int_{\Re^N}\varphi_{1,\gamma}\,dX$. Note that this integral is computed over the support of $\varphi_{1,\gamma}$, which is the unit ball $B(0,1)$. Now, let us compute the first and second derivatives of $\varphi_{1,\gamma}$ according to the variable γ, we obtain

$$\varphi'_{1,\gamma}=\frac{1}{||\vec{X}||^2-1}e^{\left(\frac{\gamma}{||\vec{X}||^2-1}\right)}$$
$$\varphi''_{1,\gamma}=\frac{1}{(||\vec{X}||^2-1)^2}e^{\left(\frac{\gamma}{||\vec{X}||^2-1}\right)}$$

we can easily see that $|\varphi'_{\sigma,\gamma}|\le\varphi'_{\sigma,\gamma_0}$ and $|\varphi''_{\sigma,\gamma}|\le\varphi''_{\sigma,\gamma_0}$ for every $\gamma\in[\gamma_0,+\infty[$. Using the fact that the integral in C_γ is over a bounded set, and the Lebesgue convergence theorem (see [3]) we deduce that: if we set $V(\gamma)=\int_{\Re^N}\varphi_{1,\gamma}\,dX$, then $V'(\gamma)=\int_{\Re^N}\varphi'_{1,\gamma}\,dX$ and $V''(\gamma)=\int_{\Re^N}\varphi''_{1,\gamma}\,dX$. Now, by integration by part in the expression of $V(\gamma)$ for each direction $x_i, 1\le i\le N$, we obtain

$$V(\gamma)=\int_{B(0,1)}\frac{2\gamma x_i^2}{(||\vec{X}||^2-1)^2}e^{\left(\frac{\gamma}{||\vec{X}||^2-1}\right)}dX.$$

Now, summing over index i, we obtain

$$\begin{aligned}V(\gamma)&=\frac{1}{P}\int_{B(0,1)}\frac{2\gamma||\vec{X}||^2}{(||\vec{X}||^2-1)^2}e^{\left(\frac{\gamma}{||\vec{X}||^2-1}\right)}dX\\&=\frac{1}{P}\int_{B(0,1)}\frac{2\gamma}{(||\vec{X}||^2-1)}e^{\left(\frac{\gamma}{||\vec{X}||^2-1}\right)}dX\\&\quad+\frac{1}{P}\int_{B(0,1)}\frac{2\gamma}{(||\vec{X}||^2-1)^2}e^{\left(\frac{\gamma}{||\vec{X}||^2-1}\right)}dX.\end{aligned}$$

The last two quantities are exactly $(2\gamma/P)V'(\gamma)$ and $(2\gamma/P)V''_\gamma$. Thus, the function $V(\gamma)$ is a solution of the second-order differential equation

$$\frac{2\gamma}{P}(V''(\gamma)+V'(\gamma))-V(\gamma)=0,\quad\gamma\in[\gamma_0,+\infty[\gamma_0>0.$$

Now set $V(\gamma)=h(\gamma)\exp[(-1/2)\gamma]$ (which implies that $C_\gamma=h(\gamma)\exp[(1/2)\gamma]$). By placing it in the above equation, we have

$$h''(\gamma)-\left(\frac{1}{4}+\frac{P}{2\gamma}\right)h=0,\quad\gamma\in[\gamma_0,+\infty[\quad\text{with }\gamma_0>0.$$

This completes the proof. ■

Proof of Proposition 2: To obtain the proposed scheme, the above equation is discretized using the following approximation:

$$h''(\gamma)\approx\frac{h(\gamma_{i+1})-2h(\gamma_i)+h(\gamma_{i-1})}{(\Delta\gamma)^2},$$
$$\text{where }\gamma_i=\gamma_0+i\Delta\gamma=\gamma_{i-1}+\Delta\gamma$$

using the Taylor expansion we obtain

$$\begin{aligned}h''(\gamma_i)=\frac{h(\gamma_{i+1})-2h(\gamma_i)+h(\gamma_{i-1})}{(\Delta\gamma)^2}-\frac{(\Delta\gamma)^4}{4!}\\\times(h''''(\gamma_i+\theta\Delta\gamma)+h''''(\gamma_i+\beta\Delta\gamma)).\end{aligned}$$

Using the differential equation verified by $h(\gamma)$ and the last equality, we obtain

$$|h(\gamma_i) - h_i| = \left(\frac{1}{4} + \frac{P}{2\gamma_i}\right)^{-1} \frac{(\Delta\gamma)^4}{4!} |h''''(\gamma_i + \theta\Delta\gamma) + h''''(\gamma_i + \beta\Delta\gamma)|.$$

The quantity $((1/4)+(P/2\gamma))^{-1}$ is bounded on $[\gamma_0, +\infty[$, and from the integral definition of $h(\gamma)$, it is easy to see that $h''''(\gamma)$ is also bounded on $[\gamma_0, +\infty[$. We deduce that $|h(\gamma_i) - h_i| = O((\Delta\gamma)^2)$, which completes the proof. ∎

We will prove both lemmas in the case of two dimensions; the 1-D case proofs are similar. Since some computations of Lemma 2 will be used in proof of Lemma 1, we begin by giving the proof of Lemma 2.

Proof of Lemma 2: Recall the Laplacian of KCS ($\Delta\rho_{\sigma,\gamma}$) formula shown in the first equation at the bottom of the page. Let us set $x^2 + y^2 = z$; then the zero-crossings of $\Delta\rho_{\sigma,\gamma}$ in $]-\sigma, \sigma[$ are given by those of the second order polynomial $p(z) = z^2 + \gamma\sigma^2 z - \sigma^4$. The nonnegative zero of this polynomial function is given by

$$z_1 = \frac{(-\gamma + \sqrt{\gamma^2 + 4})\sigma^2}{2}. \qquad (\star)$$

First note that this zero has a nonnegative value for every value of γ. Hence, for a given integer (or nonnegative real) n, we can set $z_1 = (\sigma^2/n^2)$, thus (*) gives: $\gamma = n^2 - (1/n^2)$. The lemma is therefore proved. ∎

Proof of Lemma 1: To prove that first and second derivatives exhibit the same behavior as those of the Gaussian, it suffices to prove the result for the second derivative, or the Laplacian of the KCS. To do so, we need to look for the KCS Laplacian gradient zeros. For each direction ($y = \delta x, \delta \in \Re$), we should have exactly three zeros, one at the origin and two in the $]-\sigma, \sigma[$ interval, with the latter two zeros symmetric. We are only interested in the nonnegative one (i.e., in $]0, \sigma[$). Let us now compute the KCS Laplacian gradient

$$\operatorname{grad} \Delta\rho_{\sigma,\gamma}(x,y) = \nabla^2 \Delta\rho_{\sigma,\gamma}(x,y) = \begin{pmatrix} \partial_x \Delta\rho_{\sigma,\gamma}(x,y) \\ \partial_y \Delta\rho_{\sigma,\gamma}(x,y) \end{pmatrix}.$$

Note that $\Delta\rho_{\sigma,\gamma}$ could be presented as a function of the variable $z = x^2 + y^2$; then the gradient could be computed as follows:

$$\begin{pmatrix} \partial_x \Delta\rho_{\sigma,\gamma}(x,y) \\ \partial_y \Delta\rho_{\sigma,\gamma}(x,y) \end{pmatrix} = \begin{pmatrix} \partial_x \Delta\rho_{\sigma,\gamma}(x,y)\partial_x z \\ \partial_z \Delta\rho_{\sigma,\gamma}(x,y)\partial_y z \end{pmatrix} = 2\partial_z \Delta\rho_{\sigma,\gamma}(x,y) \begin{pmatrix} x \\ y \end{pmatrix}.$$

From the last expression of the gradient, we deduce that the origin (0, 0) is an obvious zero. However, the expression $\partial_z \Delta\rho_{\sigma,\gamma}(x,y)$ in terms of z is given by (to avoid making the expressions overly heavy, we use the same function symbol for z and (x, y) variables) the second equation at the bottom of the page. The zeros of $\partial_z \Delta\rho_{\sigma,\gamma}(z)$ in interval $]0, \sigma[$ are given by those of the polynomial function

$$Q(z) = -2z^3 + (2 - 4\gamma)\sigma^2 z^2 + (-\gamma^2 + 2\gamma + 6)\sigma^4 z + (2\gamma - 4)\sigma^6.$$

In order to discuss the zeros of $Q(z)$, we compute its first derivative

$$Q'(z) = -6z^2 + (4 - 8\gamma)\sigma^2 z + (-\gamma^2 + 2\gamma + 6)\sigma^4.$$

The $Q'(z)$ zeros are

$$\xi_1 = \frac{[(4 - 8\gamma) + \sqrt{40\gamma^2 - 16\gamma + 160}]\sigma^2}{12},$$
$$\xi_2 = \frac{[(4 - 8\gamma) - \sqrt{40\gamma^2 - 16\gamma + 160}]\sigma^2}{12}.$$

We now establish the hypothesis $\gamma > 2$; and we deduce: $Q(0) > 0$ and $\xi_2 < 0$. The fact that $Q(0) > 0$, and $Q(+\infty) = -\infty$ implies that $Q(z)$ has at least one zero in the nonnegative part of the real axis, i.e., $]0, +\infty[$. We should now prove that $Q(z)$ really has just one zero β in $]0, +\infty[$ that belongs to $]0, \sigma[$. Suppose that there are two zeros in $]0, +\infty[$, combining with $Q(0) > 0$ and $Q(+\infty) = -\infty$, we deduce that $Q(z)$ has two extremas that are reached in $]0, +\infty[$. These extremas are necessarily ξ_1 and ξ_2, which contradicts the fact that $\xi_2 < 0$. It remains to be proved that β is in $]0, \sigma[$. From Lemma 2 given above, we show that a nonnegative zero-crossing of the KCS Laplacian is given by $\sqrt{z_1}$ (where $z_1 = (-\gamma + \sqrt{\gamma^2 + 4})\sigma^2/2$). Since $\gamma < \sqrt{\gamma^2 + 4} < \gamma + 2$, we deduce that $0 < z_1 < \sigma^2$. However,

$$\Delta\rho_{\sigma,\gamma} = \begin{cases} \dfrac{\gamma}{C_\gamma} \left[\dfrac{(x^2+y^2)^2 + \gamma\sigma^2(x^2+y^2) - \sigma^4}{(x^2+y^2-\sigma^2)4}\right] e^{\left(\frac{\gamma\sigma^2}{x^2+y^2-\sigma^2} + \gamma\right)}, & \text{if } x^2 + y^2 < \sigma^2 \\ 0, & \text{elsewhere} \end{cases}$$

$$\begin{aligned} &\partial_z \Delta\rho_{\sigma,\gamma}(x,y) \\ &\quad = \partial_z \Delta\rho_{\sigma,\gamma}(z) \\ &\quad = \begin{cases} \dfrac{\gamma}{C_\gamma} \left[\dfrac{-2z^3 + (2-4\gamma)\sigma^2 z^2 + (-\gamma^2 + 2\gamma + 6)\sigma^4 z + (2\gamma - 4)\sigma^6}{(z - \sigma^2)^6}\right] e^{\left(\frac{\gamma\sigma^2}{(z-\sigma^2)} + \gamma\right)}, & \text{if } z < \sigma \\ 0, & \text{elsewhere} \end{cases} \end{aligned}$$

σ is also a zero-crossing for the KCS Laplacian; this implies that the gradient of the KCS Laplacian vanishes at some value of the interval $]\sqrt{z_1}, \sigma[$, which proves that α is in $]0, \sigma[$. The extrema at β of the KCS Laplacian is a local maximum, otherwise we would have more then one extrema, since $\Delta\rho_{\sigma,\gamma}(0) < 0$ and $\Delta\rho_{\sigma,\gamma}(\sqrt{z_1}) = \Delta\rho_{\sigma,\gamma}(\sigma) = 0$. This completes our proof. ■

Proof of Lemma 3: We must show that the limit of $\int_{-\sigma}^{+\sigma}(\rho_{\sigma,\gamma}(x) - h_\sigma(x))^p\,dx$ tends to zero when γ tends to zero, for every $1 \le p < \infty$. The function $(\rho_{\sigma,\gamma} - h_\sigma)^p$ is uniformly bounded on $[-\sigma, \sigma]$ ($[-\sigma, \sigma]^N$ in N-Dimension), and the limit of this function is zero except in sets with a Lebesgue measuring zero (the set $\{-\sigma, +\sigma\}$ for one dimension, the sphere defined by $x_1^2 + x_2^2 + \cdots + x_N^2 = \sigma^2$ for N dimension). According to the Lebesgue theorem (see [3]), the limit on γ can be commuted with the integral symbol, and thus the Lemma is proved. ■

Proof of Theorem 2: Since the support of I is bounded, and $\rho_{\sigma,\gamma}$ support is compact, there exists an open bounded subset C of IR^2 which contains the support of $\mathsf{g}_{\sigma,\gamma}$ and I^ε. We then have the following proposition.

Proposition: If we set $\mathsf{I}^\varepsilon = \mathsf{I} * \mathsf{g}_{\varepsilon,\gamma}$ and $\mathsf{g}^\varepsilon_{\sigma,\gamma} = \mathsf{g}_{\sigma,\gamma} * \mathsf{g}_{\varepsilon,\gamma} = \rho_{\sigma,\gamma} * \mathsf{I} * \mathsf{g}_{\varepsilon,\gamma} = \rho_{\alpha,\gamma} * \mathsf{I}^\varepsilon$.

We have: $\mathrm{TV}_C(I) = \lim_{\varepsilon\to 0} \mathrm{TV}_C(I^\varepsilon)$ and $\mathrm{TV}_C(g_{\sigma,\gamma}) = \lim_{\varepsilon\to 0} \mathrm{TV}_C(g^\varepsilon_{\sigma,\gamma})$.

A proof of this proposition in a general case can be found in [14].

Let us now prove the theorem.

Since $\mathsf{g}^\varepsilon_{\sigma,\gamma}$ is a regular function, it belongs in $\mathsf{W}^{1,1}$. According to Remark 1, we have

$$\mathrm{TV}_{IR^2}(g^\varepsilon_{\sigma,\gamma}) = \int_{IR^2} \left|\mathrm{grad}\, g^\varepsilon_{\sigma,\gamma}\right| \,dx\,dy.$$

Remark 2: $\mathrm{TV}_{IR^2}(g^\varepsilon_{\sigma,\gamma}) = \mathrm{TV}_C(g^\varepsilon_{\sigma,\gamma})$, and so for I^ε, since C contains the support of $g^\varepsilon_{\sigma,\gamma}$ and I^ε.

Now choose C large enough to have the following property:

$$\forall\,(\mathsf{x},\mathsf{y}) \in G = \mathrm{Support}\, g^\varepsilon_{\sigma,\gamma}, \cup \mathrm{Support}\, \mathsf{I}^\varepsilon$$
$$\forall\,(\mathsf{h}_1,\mathsf{h}_2) \in \mathsf{B}(0,\sigma), (\mathsf{x}-\mathsf{h}_1, \mathsf{y}-\mathsf{h}_2) \in C. \qquad \text{(P)}$$

We now compute the total variation of $g^\varepsilon_{\sigma,\gamma}$

$$\begin{aligned}
&\mathrm{TV}_C\left(g^\varepsilon_{\sigma,\gamma}\right)\\
&\quad= \int_C \left|\mathrm{grad}\, g^\varepsilon_{\sigma,\gamma}\right| \,dx\,dy = \int_C |\mathrm{grad}\,\rho_{\sigma,\gamma} * I^\varepsilon|\,dx\,dy\\
&\quad= \int_C |\rho_{\sigma,\gamma} * \mathrm{grad}\, I^\varepsilon|\,dx\,dy\\
&\quad\le \int_C \left[\int_{B(0,\sigma)} \rho_{\sigma,\gamma}(h_1,h_2)\right.\\
&\qquad\left.\times\ |\mathrm{grad}\, I^\varepsilon(x-h_1, y-h_2)|\,dh_1\,dh_2\right]\,dx\,dy\\
&\quad= \int_{B(0,\sigma)} \left[\int_C \rho_{\sigma,\gamma}(h_1,h_2)\right.\\
&\qquad\left.\times\ |\mathrm{grad}\, I^\varepsilon(x-h_1, y-h_2)|\,dx\,dy\right]\,dh_1\,dh_2\\
&\quad= \int_{B(0,\sigma)} \rho_{\sigma,\gamma}(h_1,h_2)\\
&\qquad\times\left[\int_C |\mathrm{grad}\, I^\varepsilon(x-h_1, y-h_2)|\,dx\,dy\right]\,dh_1\,dh_2.
\end{aligned}$$

Using the property (P), we have

$$\begin{aligned}
&\int_C |\mathrm{grad}\, I^\varepsilon(x-h_1, y-h_2)|\,dx\,dy\\
&\qquad= \int_C |\mathrm{grad}\, I^\varepsilon(z_1,z_2)|\,dz_1\,dz_2 = \mathrm{TV}_C(I^\varepsilon).
\end{aligned}$$

Thus

$$\mathrm{TV}_C\left(g^\varepsilon_{\sigma,\gamma}\right) \le \mathrm{TV}_C(I^\varepsilon) \int_{B(0,\sigma)} \rho_{\sigma,\gamma}(h_1,h_2)\,dh_1\,dh_2.$$

Since $\rho_{\sigma,\gamma}$ is normalized, we obtain

$$\mathrm{TV}_C\left(g^\varepsilon_{\sigma,\gamma}\right) \le \mathrm{TV}_C(I^\varepsilon)$$

by tending ε to zero, and using the proposition above we have

$$\mathrm{TV}_C(g_{\sigma,\gamma}) \le \mathrm{TV}_C(I).$$

According to Remark 2, the theorem is proved. ■

Remark 3: We have proved Theorem 2 for functions in $\mathsf{BV}(\mathsf{IR}^2)$, for the coherence of this paper. However, it is still verified for higher dimensions (for functions in $\mathsf{BV}(\Re^n)$, for any $n \ge 1$).

ACKNOWLEDGMENT

The authors would also like to thank Dr. C. Suen from Concordia University and Dr. Amar Mitiche from INRS-Telecom for their fruitful remarks about this work. They are also indebted to the reviewers for their thoughtful comments and careful correction of the paper.

REFERENCES

[1] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 26–33, Jan. 1986.

[2] F. Bergholm, "Edges focusing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 726–741, June 1987.

[3] H. Brezis, *Analyze Fonctionnelle et Application*. Paris, France: Masson, 1986.

[4] P. J. Burt, "Fast filter transform for image processing," *Comput. Vis., Graph., Image Processing*, vol. 16, pp. 20–51, 1981.

[5] M. Cheriet, J. N. Said, and C. Y. Suen, "A recursive thresholding technique for image segmentation," *IEEE Trans. Image Processing*, vol. 7, pp. 918–922, June 1998.

[6] M. Cheriet, "Extraction of handwritten data from noisy gray-level images using a multi-scale approach," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 13, no. 5, pp. 665–685, 1999.

[7] M. Cheriet, R. Thibault, and R. Sabourin, "A multi-resolution based approach for handwriting segmentation in gray-level images," in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, 1994, pp. 159–168.

[8] J. J. Clark, "Singularity theory and phantom edges in scale-space," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 720–727, May 1988.

[9] J. L. Crowley, "A Representation for Visual Information," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, 1981.

[10] I. Daubechies, "Orthogonal bases of compactly-supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, Nov. 1988.

[11] ——, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[12] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "Scale and the differential structure of images," *Image Vis. Comput.*, vol. 10, pp. 376–388, 1992.

[13] E. Giusti, "Minimal surfaces and functions of bounded variation," in *Monographs in Mathematics*. Cambridge, MA: Birkhäuser, 1984, vol. 80.

[14] E. Godlewski and P. A. Raviart, "Hyperbolic systems of conservation laws," in *Mathématiques & Applications*. Paris, France: Ellipses-Edition Marketing, 1991.

[15] R. Horaud and O. Monga, *Vision par Ordinateur, outils fondamentaux*, Paris, France: Hermes, 1995, p. 425.

[16] R. A. Hummel, "The scale-space formulation of pyramid data structures," in *Parallel Computer Vision*, L. Uhr, Ed. New York: Academic, 1987, pp. 187–223.

[17] R. Kasturi and L. O'Gorman, "Special issue: Document image analysis," *Mach. Vis. Applicat.*, vol. 6, no. 2–3, pp. 67–180, 1993.
[18] A. Klinger, "Pattern and search statistics," in *Optimizing Methods in Statistics*, J. S. Rustagi, Ed. New York: Academic, 1971.
[19] J. J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 53, pp. 363–370, 1984.
[20] R. Lewis, *Practical Digital Image Processing*, London, U.K.: Ellis Horwood, 1990, p. 253.
[21] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Norwell, MA: Kluwer, 1994, p. 423.
[22] ——, "Scale-space behavior of local extrema and blobs," *J. Math. Imag. Vis.*, vol. 1, pp. 65–99, Mar. 1992.
[23] ——, "Scale-space for discrete signals," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 234–254, 1990.
[24] T. Lindeberg and J.-O. Eklundh, "Scale detection and region extraction from a scale-space primal sketch," in *Proc. 3rd Int. Conf. Computer Vision*, Osaka, Japan, Dec. 1990, pp. 416–426.
[25] S. Mallat, "A theory for multi-resolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.
[26] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. London*, ser. B, vol. 207, pp. 187–217, 1980.
[27] T. Pavlidis and G. Wolberg, "An algorithm for the segmentation of bi-level images," in *Proc. IEEE Computer Vision Pattern Recognition Conf.*, Miami Beach, FL, 1986, pp. 570–575.
[28] *Proc. 1st Int. Conf. Document Analysis and Recognition*, Saint Mălo, France, 1991.
[29] H. Reinhard, *Eléments de Mathématiques du Signal*, Paris, France: Dunod, 1995.
[30] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, "A survey of thresholding techniques," *Comput. Vis., Graph., Image Process.*, vol. 41, pp. 233–260, 1980.
[31] O. D. Trier and T. Taxt, "Evaluation of binarization methods for document images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 312–315, Mar. 1995.
[32] O. D. Trier and A. K. Jain, "Goal-directed evaluation of binarization methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 1191–1201, Dec. 1995.
[33] M. Vetterli and C. Herley, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[34] A. P. Witkin, "Scale-space filtering," in *Proc. 8th Int. Joint Conf. Artificial Intelligence*, Karlsruhe, Germany, Aug. 1983, pp. 17–45.
[35] A. L. Yuille and T. A. Poggio, "Scaling theorems for zero-crossings," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 15–25, 1986.

Lakhdar Remaki received the B.S. degree in mathematics from the Université des Sciences et Technologie d'Alger, Algiers, Algeria, in 1991, and the M.Sc. and Ph.D. degrees in applied mathematics from the Université Claude Bernard de Lyon, France, in 1992 and 1997, respectively.

Since 1998, he has been a Postdoctoral Fellow with the École de Technologie Supérieure, Université du Québec, Montréal, P.Q., Canada. His research interests include mathematical modeling for image process, partial differential equations, and numerical analysis.

Mohamed Cheriet (M'95) received the B.Eng. degree in computer science from the Université des Sciences et Technologie d'Alger, Algiers, Algeria, in 1984, and the M.Sc. and Ph.D. degrees in computer science from the University of Pierre et Marie Curie (Paris VI) in 1985 and 1988, respectively.

From 1988 to 1990, he was a Research Associate with LAFORIA/CNRS Laboratory. He then joined CENPARMI, Concordia University, Montreal, P.Q., Canada, where he was a Postdoctoral Fellow for two years. He was appointed Assistant Professor in 1992, Associate Professor in 1996, and Full Professor in 1998, with the Department of Automation Engineering, Ecole de Technologie Supérieure, University of Quebec, Montréal. His research focuses on image processing, pattern recognition, character recognition, text processing, handwritten documents analysis and recognition, and perception. He has published more than 60 technical papers in the field. He was a guest co-editor of the *International Journal of Pattern Recognition and Artificial Intelligence* and the machine, perception, and artificial intelligence series books published by World Scientific.

Dr. Cheriet is an active member of LIVIA and CENPARMI.