Building New Kernel Family with Compact Support, in Scale-Space.

L. Remaki and M. Cheriet[1]

Laboratoire d'Imagerie, de Vision et d'Intelligence Artificielle

École de Technologie Supérieure de l'Université du Québec

1100 rue Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

remaki@livia.etsmtl,ca; cheriet@gpa.etsmtl.ca

ABSTRACT

Scale-space representation is one formulation of the multi-scale representation which has received considerable interest in the literature, because of its efficiency in several practical applications, and the distinct properties of the Gaussian kernel which generates the Scale-space. However, in practice, we note some undesirable limitations when using the Gaussian kernel: information loss caused by the unavoidable Gaussian truncation and the prohibitive processing time due to the mask size. To give a solution to both problems, we propose in this paper a new kernel family with compact support derived from the Gaussian, which are able to recover the information loss while reducing drastically the processing time. This family preserves a great part of the useful Gaussian properties, without contradicting the uniqueness of the Gaussian kernel. The construction and some properties of the proposed kernels are developed in this paper. Furthermore, in order to asses our contribution, an application of extracting handwritten data from noisy document images is presented, including a qualitative and quantitative comparison between results obtained by the Gaussian and the proposed kernels.

Key Words Multi-scale representation, scale-space representation, compact support, kernels, functional space, image segmentation, handwritten data, handwritten data extraction.

1. INTRODUCTION

During the last few decades, mathematical tools have been developed for handling the scale concept in a coherent manner. Different multi-scale representations have been proposed, such as quad-trees in Klinger [11], pyramid representation in Burt [4] and Crowley [8], and scale-space representation in Witkin [17] and Koenderink [12]. The latter formulation consists of using a convolution with a chosen kernel in the smoothing operation. Several authors, show that, under some posed hypothesis, the choice of the Gaussian kernel is unique, and it has many beneficial properties.

In his paper Cheriet [6] has shown how to derive a new multi-scale paradigm which conjugate data quality to noise immunity with low computational cost. Indeed, although it is an intuitive based approach, choosing mask size as scale parameter leads to the desired objectives of efficient segmentation method and it allows us to reduce sensitively the maximum mask size from 30×30 to 17×17 in the case of the variation coefficient VC>0.49 (see [6]). In this paper we propose to assess these empirical results with a strong theoretical foundations. Since taking mask size as scale parameter reduces the processing time and improves the image quality (minimize the information loss). This intuition has pointed us in a new direction for deriving a well founded new formalism able to make a rigorous and explicit relation between the "classical" scale parameter, namely the standard deviation and the mask size of the kernel. To pursue this line we propose a new family of kernels derived from the Gaussian with compact supports (KCS) (i.e. the kernel itself vanishes, outside some compact set). Thus we don't need to cutoff the kernel when computing the convolution product which is the principal cause of the information loss Furthermore, the mask is precisely the support of the kernel, and its relation (as we will show it in this paper) with the standard deviation is simple and exact, concretely : mask size = 2σ (σ is the standard deviation). After doing this, we can appreciate how much the mask size (and then, the processing time) is reduced, (e.g. from 11.31σ [2,15] or 6σ [6,13] to only 2σ), and the information loss due to the truncation of the Gaussian kernel is recovered. The diagram shown in Figure1, summarizes the functional and the topological transforms which allow the derivation of the new kernel from the Gaussian one, and their impact on the kernel mask sizes.

2. SCALE-SPACE GENERATION, PRACTICAL LIMITATIONS OF THE GAUSSIAN KERNEL & OUR CONTRIBUTION

Because we are concerned with a two dimension signal (image) in the application given in section 5, we recall the definition of scale-space in two dimensions, although this definition is generalized to any higher dimension without loss of the properties mentioned above. Let $f : \Re^2 \to \Re$ be a given signal. The scale space representation $L : \Re^2 \times \Re \to \Re$ is defined such that : at zero scale it gives the original signal

$$L(.;0) = f, \qquad (1)$$

and at every scale σ, the coarser signal representation is given by the convolution

$$L(.;\sigma) = g(.;\sigma) * f. \qquad (2)$$

or

$$L(x,y,\sigma) = \int_{\Re^2} g(\mu;\lambda;\sigma)\, f(x-\mu, y-\lambda)\, d\mu d\lambda \qquad (3)$$

where $g(x,y,\sigma)$, is the 2-dimensional Gaussian kernel :

$$g(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (4)$$

This presents a scale-space formulation using the Gaussian kernel where the convolution product is computed at each scale. The integral defining the convolution must be computed over all space $\Re^2$, however, it is approached practically 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 give more precise computation, but they increase the cost of processing time; lower mask sizes decrease the processing time, but the accuracy is sometimes severely diminished, which induces information loss. Hence, two fundamental practical limitations of the Gaussian kernel can be raised: information loss and prohibitive processing time. Some solutions are proposed in the literature to overcome these problems, such as the approximation of the Gaussian by recursive filters, or using truncated exponential functions instead of the Gaussian [14]. All these alternatives based on approximating the Gaussian, reduce the processing time, but the information loss still persists, and sometimes is increased. These two problems constitute the main motivations of 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 cutoff the kernel while the processing time is controlled, since the mask is the support of the kernel itself.

3. BUILDING NEW KERNELS WITH COMPACT SUPPORT (KCS)

3.1 Recall of some Mathematical Fundamentals

In order to make the KCS construction explicit, we recall some definitions:

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

$$L^p(\Re^2) = \left\{ \begin{array}{l} f : \Re^2 \to \Re \ \ such \ \ that \\ \int_{\Re^2} |(f)|^p \ dx < \infty \end{array} \right\}$$

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:

$$D(\Re^2) = \left\{ \begin{array}{l} \Phi \in C^\infty \ so \ that \ there \ exists \\ a \ compactset \ K \ such \ that \\ \text{support}\, \Phi \subset K \end{array} \right\}$$

Definition 3: let ρ be a non-negative function in $D(\Re^2)$ satisfying the following conditions:

$$(A) \ \int_{\Re^2} \rho(x,y) dxdy = 1 \qquad (5)$$

$$(B) \ \ \text{support} \ \ \rho = \{(x,y) \in \Re^2 ; x^2 + y^2 \le 1\} = B(0,1). \qquad (6)$$

Now, for each $\sigma > 0$, we define

$$\rho_\sigma(x,y) = \frac{1}{\sigma^2} \rho(\frac{x}{\sigma}, \frac{y}{\sigma}) \qquad (7)$$

Then, $\rho_\sigma \in D(\Re^2)$ is non-negative, and satisfies :

$$(A) \ \int_{\Re^2} \rho_\sigma(x) dx = 1 \qquad (8)$$

$$(B) \ \ \text{support}\rho_\sigma = \{(x,y) \in \Re^2 ; x^2 + y^2 \le \sigma^2\} = B(0,\sigma). \qquad (9)$$

These functions ρ_σ are called mollifiers.

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

$$\rho_\sigma * f \in L^p(\Re^2) \qquad (10)$$

$$\rho_\sigma * f \xrightarrow[\sigma \to 0]{} f \ \ in \ \ L^p(\Re^2) \qquad \text{i.e.:}$$

$$\lim_{\sigma \to 0} \int_{\Re^2} |(\rho_\sigma * f - f)|^p \ dxdy = 0 \qquad (11)$$

We notice 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 $1 \le p < \infty$. Thus the previous theorem holds for each value of *p*, and we obtain the original image for $\sigma = 0$ in the sense given by the theorem.

3.2 Kernel Construction

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

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

$$g(x,y) = \frac{1}{2\pi} e^{-\left[\frac{x^2+y^2}{2}\right]} \qquad (12)$$

then

$$g_\sigma(x,y) = \frac{1}{\sigma^2} g\left(\frac{x}{\sigma}, \frac{y}{\sigma}\right) = \frac{1}{2\pi\sigma^2} e^{-\left[\frac{x^2+y^2}{2\sigma^2}\right]} \qquad (13)$$

In order to build the KCS kernel, we apply a change of variable to g instead of g_σ and define the one-parameter family of KCS kernels as above. First, recall the polar coordinates which change the variables (x,y) into (r,θ); when (x,y) belongs to IR^2, (r,θ) belongs to $\Re^+ \times [0, 2\pi]$. Let $\Gamma(r,\theta)$ be the function which defines this change of coordinates:

$$\Gamma(r,\theta) = (r\cos\theta, r\sin\theta) \qquad (14)$$

Now define the function ω as follows:

$$[0, 1[\;\rightarrow\; \Re^{+}$$

$$r \;\rightarrow\; \omega(r) \quad \text{with}$$

$$\omega(r) = \sqrt{\frac{1}{1-r^2} - 1} \qquad (15)$$

$\Gamma(\omega(r),\theta)$ transforms the unit ball $B(0,1)$ into a $\Re^2$ plane. Now Make the change of variables $(x, y) = \Gamma(\omega(r),\theta)$ in the expression of the Gaussian $g(x, y)$, and extend it by zero outside the unit ball; we obtain then the desired KCS kernel :

$$\rho(r,\theta) = \begin{cases} \frac{1}{2\pi} e^{\frac{1}{2}\left(\frac{1}{r^2-1}+1\right)} & if \;\; r^2 < 1 \\ 0 & elsewhere \end{cases} \qquad (16)$$

or, with the variables (x, y) :

$$\rho(x,y) = \begin{cases} \frac{1}{2\pi} e^{\frac{1}{2}\left(\frac{1}{x^2+y^2-1}+1\right)} & if \;\; x^2+y^2 < 1 \\ 0 & elsewhere \end{cases} \qquad (17)$$

This function, well known in mathematics, satisfies (up to a normalizing constant):

$$\rho \in C^{\infty}, \;\; \rho \in D(IR^2) \;\; and \;\; \int_{\Re^2} \rho(x,y) = 1 \qquad (18)$$

Now, we can define the one-parameter family of KCS kernels, as follows:

$$\rho_{\sigma}(x,y) = \frac{1}{\sigma^2}\,\rho\left(\frac{x}{\sigma},\frac{y}{\sigma}\right) \quad which \;\; gives:$$

$$\rho_{\sigma}(x,y) = \begin{cases} \frac{1}{2\pi\sigma^2} e^{\frac{1}{2}\left(\frac{\sigma^2}{x^2+y^2-\sigma^2}+1\right)} & if \;\; x^2+y^2 < \sigma^2 \\ 0 & elsewhere \end{cases}$$

The KCS formula

According to definition 3, (ρ_{σ}) is a family of mollifiers, and theorem 1 holds. The support of ρ_{σ} is B(0,σ). Then it is unnecessary to take mask's sizes greater than 2σ . Figure 2 depicts the curve of ρ_{σ} for a given σ.

4. DISCUSSION & APPLICATION

From the theoretical aspect, we have built in the previous section the new kernel family KCS, and given its formula. This kernel is derived from the Gaussian and keeps the most important properties (relative to the image segmentation process) of the Gaussian kernel. Obviously, it cannot have all of the properties which make the Gaussian unique. However, even the Gaussian is very rich in properties as seen above, we do not necessarily need all of them to perform image segmentation properly**.** In this sense, the most important ones are: i) continuity with respect to scale parameter; ii) recovering the initial signal (image) when the scale parameter tends to zero; iii) the strong regularization property; and iv) the zero crossing diminishing property (i.e. a number of zeros of a given signal decreases when it is convoluted with a Gaussian kernel). Our new kernel (KCS: kernel with compact support), conserves the properties i) to iii). A thorough and deep discussion and proofs of theses properties can be found in [16].
For the iv)th property, tests on a great number of perturbed functions show that this property is conserved in general (a sample is given in Figures 3 to 13), but we cannot say more. However, the following theorem, proves a similar but weaker property (The total-variation diminishing) :

Theorem 2 (total-variation diminishing property): Let I be a given signal belonging to BV($\Re^2$), and $\rho_{\sigma,\gamma}$ a normalized KCS kernel. Suppose that the support of I is bounded. Then, if we put $g_{\sigma} = \rho_{\sigma} * I$ than $\forall\sigma > 0$, we have:

$$TV_{\Re^2}(g_{\sigma}) \leq TV_{\Re^2}(I)$$

We say that g_{σ} has a total-variation diminishing property. For the definition of the bounded variation functions space BV($\Re^2$) and the total variation $TV_{\Re^2}()$, please see [9-10].

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 non-creation of artificial local extremas property which is proper to the Gaussian (see [1]). For the proof of this theorem please see [16].

For the practical aspect, in order to investigate the KCS impact on extracting handwritten data from degraded and noisy images, we have chosen postal mail envelops as a target application. A comparison is also established between results obtained by using the Gaussian and KCS kernels. First, we recall briefly the methodology being used in this application.

To obtain multi-scale representations of a discrete signal I, one can define a set of discrete functions I_i ($i \in$ [0 .. n]) as in [14], where I_0 represents the original image and each new level is calculated by convolution from the previous one as follows :

i=1,… $I_i = I_{i-1} * K_i$ where K_i is a given kernel.

In the literature, the Gaussian kernel is used in edge detection (Marr and Hildreth [15]). Other studies related to edge detection include [2,7]. We have presented in [5] a generalization of using the LOG operator for full shape data segmentation. This methodology is briefly described in this section. For instance, we describe the method by giving the KCS Laplacian (LOKCS) and recalling the Laplacian of Gaussian (LOG) operators; for more details concerning the LOG operator please see [2,5]. The operator is defined by convoluting the image with LOKCS (LOG).

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \quad \text{then} \quad LOG =$$

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

and

$$LOKCS = \nabla^2 \rho_\sigma (x,y)$$

$$= \begin{cases} \frac{1}{\pi^4} \left[\frac{(x^2+y^2)^2 + \sigma^2 (x^2+y^2) - \sigma^4}{(x^2+y^2-\sigma^2)4} \right] \times e^{\left(\frac{\sigma^2}{x^2+y^2-\sigma^2} + 1 \right)} & \text{if } x^2+y^2 < \sigma^2 \\ 0 & elsewhere \end{cases}$$

The decision criterion in segmenting data [5] is the detection of the convex parts of the smoothed image I_i, at each scale level of resolution (i^{th} step); it is determined by the sign of $LOKCS*I_{i-1}$ ($LOG*I_{i-1}$). Since most of the information for the Gaussian is contained within the range [6σ[6,13], 11.32σ[15,2]], 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]-σ, σ [, then the mask size is equal to 2σ only.

In order to compare their performance, we have fixed appropriate parameters with respect to a few images for both the LOKCS and the LOG operators. Then we use these parameters to process a large image data base. We focus the discussion on two sets of images:

1) In the first set (a) (see figure14-a), both operators give good results with high visual quality. However the required processing time for the LOKCS operator is drastically less than that required by the LOG operator. Indeed, the processing time depends on the mask size, which is equal to 2σ for the LOKCS. For the LOG, we have chosen the mask size as suggested in [15,2], i.e., mask size = 11.32σ. In our tests, for the multi-scale representation we have 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 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) The second set (b) (see figure 14-b) represents some degraded images. With the same parameters used in (a), one can see that the information loss when using the Gaussian kernel becomes too sensitive, and we can appreciate the capability of the KCS kernel to recover it, without forgetting the processing time gained.

4. CONCLUSION

In this paper, we have presented a new family of kernels (KCS) to generate scale-space in Multi-scale representation where the Gaussian kernel is usually used. The new kernels are derived from the Gaussian, by deforming the plane ($\Re^n$ space in the general case) into a unit ball. Hence, we obtain kernels with compact support. The obtained kernels preserve the most important properties of the Gaussian kernel to perform image segmentation efficiently. 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 and quantitative comparisons of the results obtained by both the Gaussian and the KCS kernels, which confirm the KCS theoretical virtues we have shown. More experimental results are being undertaken on a large and real data base in order to give the new kernel family a thorough and a complete discussion with respect to the analytical and practical issues.

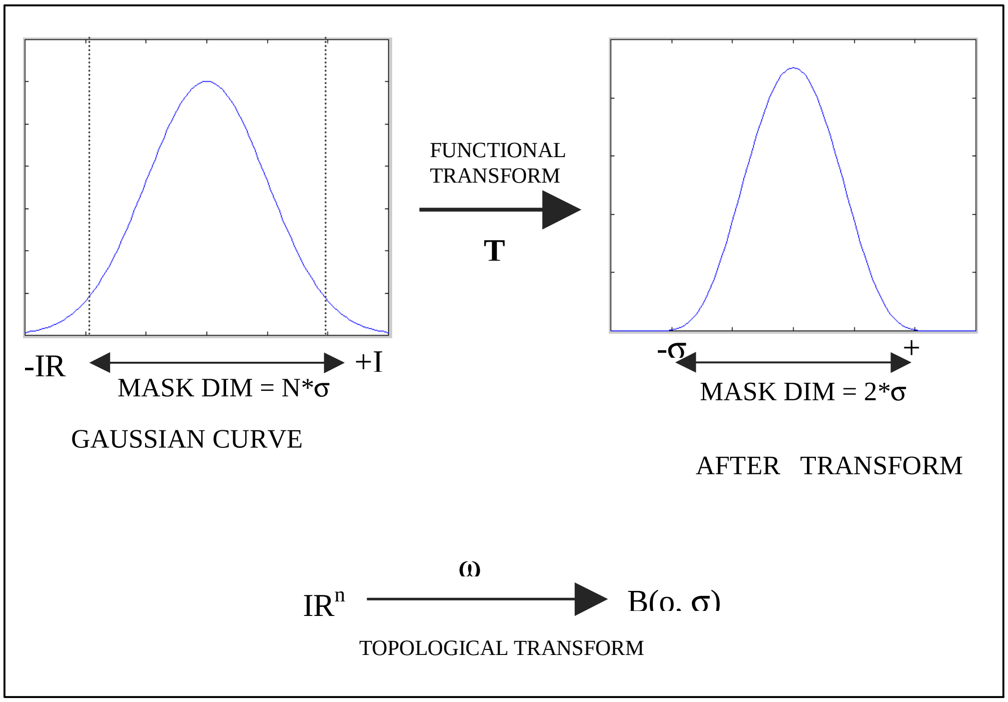


Figure 1 : Derivation of the KCS kernel from the Gaussian kernel: the topological transform that transforms the Gaussian kernel to a compact support kernel.

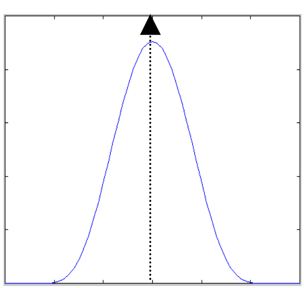


Figure2: The curve behavior

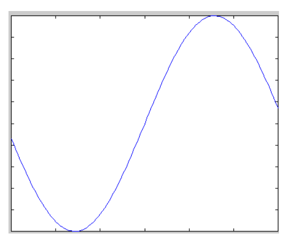


Figure 3: Curve of sinusoidal $S_p(x)=S(x)$ +

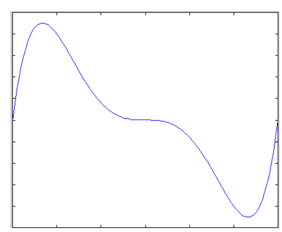


Figure 4: Curve of polynomial $P_p(x)=P(x)$ +

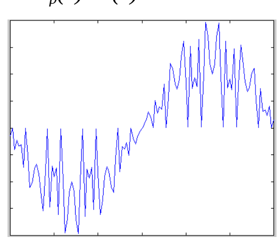


Figure 5: Curve of perturbed

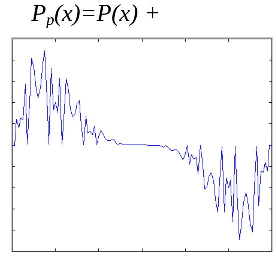


Figure 6: Curve of perturbed

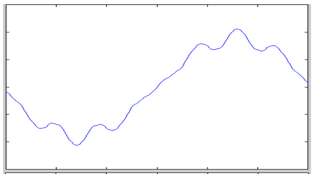
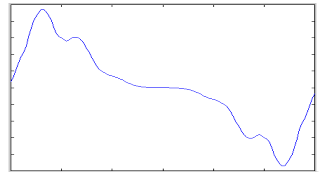


Figure 7: Regularization with σ=3



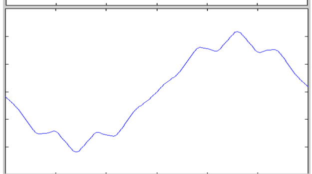
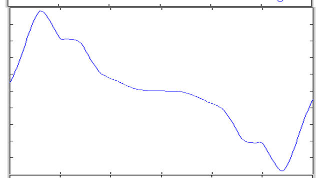


Figure 8: Regularization with σ=4



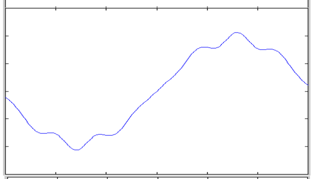
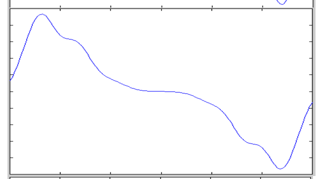


Figure 9: Regularization with σ=5



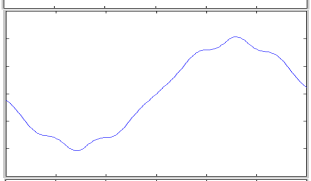
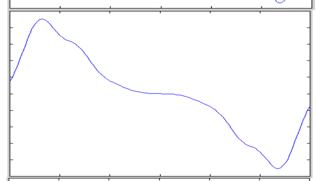


Figure 10: Regularization with σ=6



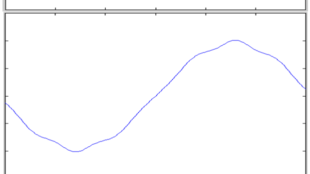
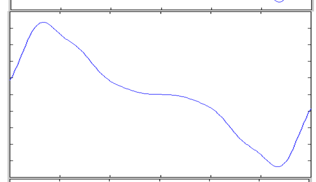


Figure 11: Regularizatio n with σ=7



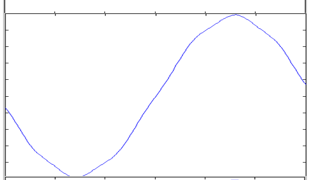
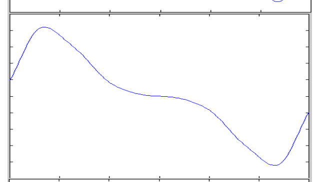


Figure 12: Regularization with σ=8



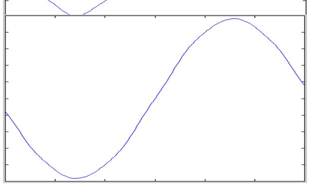
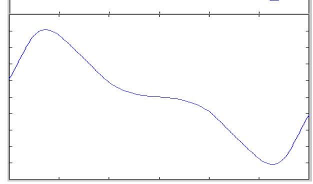
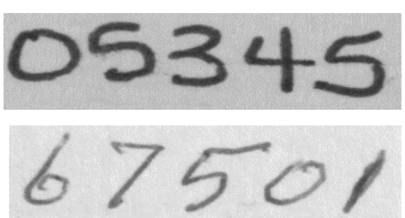
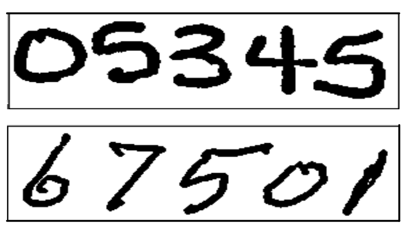


Figure 13: Regularization with σ=9

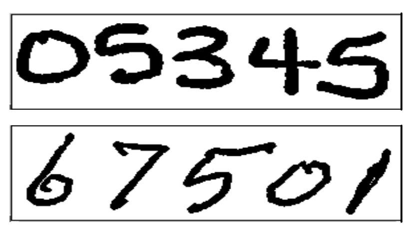




Original images of set **(a)**

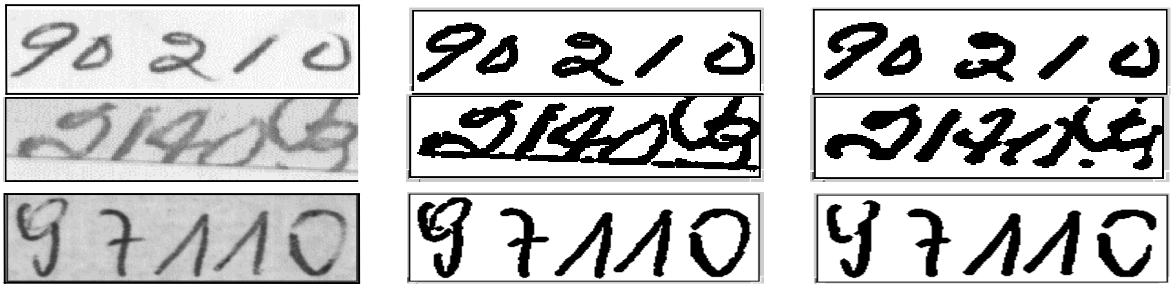


Segmentation using the KCS kernel. Maxdimask=8×8



Segmentation using the Gaussian kernel. Maxdimask=33.96×33.96

Figure14-a: segmented_images of set (a) shows that both operators LOKCS and LOG give good results. However the maximum mask size is reduced from 33.96×33.96 for the LOG to 8×8 for the LOKCS.



Original images of set **(b)**

Segmentation using the KCS kernel.
Maxdimask = 8×8

Segmentation using the Gaussian kernel.
Maxdimask = 33.96×33.96

Figure 14-b- set (b) represent some degraded images. With the same parameter used in (a) , one can see the sensitive loss of information when using the LOG, and the capability of LOKCS to recover it, when the processing time is drastically reduced as in (a) .

ACKNOWLEDGEMENT

This work was supported by research grants received from IRIS, the National Network of Centres of Excellence research program of Canada, NSERC, the Natural Sciences and Engineering Research Council of Canada, and FCAR, the Fondation du Conseil et Aide à la Recherche du Québec. We would also like to thank Dr. Ching Suen from Concordia University and Dr. Amar Mitiche from INRS for their fruitful remarks about this work.

REFERENCES :

[1] J. Babaud, A. P. Withkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtring, " IEEE Trans. PAMI, Vol. 8, no. 1, pp. 26-33, 1986.

[2] F. Bergholm, "Edges focusing," IEEE Trans. PAMI, vol. 9, no. 6, pp.726-741, 1987.

[3] H. Brezis, "Analyse fonctionnelle et application". Masson, Paris 1986.

[4] P. J. Burt, "Fast filter transform for image processing," Computer Vision, Graphics, and Image Processing, Vol. 16, pp. 20-51, 1981.

[5] M. Cheriet, R. Thibault and R. Sabourin," A multi-resolution based approach for handwriting segmentation in gray level images," In IEEE Int. Conf. on IP, Austin, pp. 159-168, 1994.

[6] M. Cheriet, “Extraction of Handwritten Data from Noisy Gray-level Images Using a Multi-scale Approach,” To appear in The Int Journal of Pattern Recognition and Artificial Intelligence, June 1999

[7] J.J. Clark, "Singularity theory and phantom edges in scale-space," IEEE Trans. PAMI, vol. 10, no. 5, pp.720-727, 1988.

[8] J. L. Crowley, "A representation for visual information". PhD thesis, Carnegie-Mellon University, Robotics Institute, Pittsburgh, Pennsylvania, 1981.

[9] E. Giusti, "Minimal surfaces and functions of bounded variation," Monograph in Mathematics, Vol 80, Birkhäuser, 1984.

[10] E. Godlewski and P. A. Raviart, "Hyperbolic Systems of Conservation laws," Mathématiques & Applications, Ellipses-Edition Marketing, 1991.

[11] A. Klinger, "Pattern and search statistics," in Optimizing Methods in Statistics(J.S. Rustagi, ed.), (New York), Academic Press, 1971.

[12] J.J. Koenderink, "The structure of images," Biological Cybernetics, vol. 53, pp. 363-370, 1984.

[13] R. Lewis, "Practical Digital Image Processing", Ellis Horwood, 253 pages, England 1990.

[14] T. Lindeberg, "Scale-space theory in computer vision," Kluwer Academic Pub., 423 pages, 1994.

[15] D. Marr and E. Hildreth, "Theory of edge detection," In Proc. of the Royal Society of London, Series B, vol. 207, pp. 187-217, 1980.

[16] L. Remaki and M. Cheriet, “KCS - New Kernel Family with Compact Support in Scale Space: Formulation & Impact,” Submitted to IEEE Trans. On Image Processing (August 1998).

[17] A.P. Witkin, "Scale-space filtering," In Proc. 8th Int. Joint Conf. Art. Intell., Karlsruhe, Germany, pp. 17-45, Aug. 1983.