Binomial Convolutions and Derivatives Estimation from Noisy Discretizations

Rémy Malgouyres[1], Florent Brunet[2], and Sébastien Fourey[3]

[1] Univ. Clermont 1, LAIC, IUT Dépt Informatique, BP 86, F-63172 Aubière, France
remy.malgouyres@laic.u-clermont1.fr
[2] Univ. Clermont 1, LAIC, IUT Dépt Informatique, BP 86, F-63172 Aubière, France
brunet@laic.u-clermont1.fr
[3] GREYC Image – ENSICAEN, 6 bd marchal Juin, 14050 Caen CEDEX, France
sebastien.fourey@greyc.ensicaen.fr

Abstract. We present a new method to estimate derivatives of digitized functions. Even with noisy data, this approach is convergent and can be computed by using only the arithmetic operations. Moreover, higher order derivatives can also be estimated. To deal with parametrized curves, we introduce a new notion which solves the problem of correspondence between the parametrization of a continuous curve and the pixels numbering of a discrete object.

1 Introduction

In the framework of image and signal processing, as well as shape analysis, a common problem is to estimate derivatives of functions, when only some (possibly noisy) sampling of the function is available from acquisition. This problem has been investigated through finite difference methods ([1]), scale-space ([4,8]) and discrete geometry ([6,7,5]).

We present a new approach to estimate derivatives from discretized data. As in scale-space, our method relies on simple computations of convolutions. However, our approach is oriented toward integer-only models and algorithms and is based on a discrete point of view of analysis. Unlike estimators proposed in [6], our method still works on noisy data. Moreover, we are able to estimate higher order derivatives.

Regarding the order of convergence, we have proved that our approach is as good as the one proposed in [6], that is, in $O(h^{2/3})$ for first order derivatives. Besides, this order of convergence is uniform. To the best of our knowledge, there is no uniform convergence results for estimation of derivatives from discretizations using scale-space.

The asymptotic computational complexity is worst case $O(\sqrt{-\ln(h)}/h)$ for first order derivatives which is similar to [6].

To deal with parametrized curves of $\mathbb{Z}^2$, we introduce a new notion, the *pixel length parametrization*, which solves the problem of correspondence between the pixels numbering of a discrete object and the parametrization of a continuous curve.

Through this paper, we also present some experimental results showing the behavior of our derivative estimator.

2 Derivatives Estimation for Real Functions

First, we consider the case of the real functions. We call *real functions* functions for which input and output sets are of dimension 1 without doing any hypothesis on the nature of these sets. We call *discrete function* a function for which the input set is such that the cardinal of all bounded subset is finite (this is the case, for example, with $\mathbb{Z}$ and $\mathbb{N}$). For simplification purposes, we consider in the sequel that a discrete function is a function from $\mathbb{Z}$ to $\mathbb{Z}$.

2.1 Principle

As said in the introduction, the principle of our derivative estimator relies on discrete convolution products. So, we are going to define what a discrete convolution product is and, then, how we are using it to construct a digital derivative estimator.

Discrete Convolution Product

Definition 1. *Let $F : \mathbb{Z} \longrightarrow \mathbb{Z}$ and $K : \mathbb{Z} \longrightarrow \mathbb{Z}$ be two discrete functions. The* discrete convolution product *of the function F and the kernel K, denoted $F * K$, is the discrete function defined by the following formula:*

$$F * K : \mathbb{Z} \longrightarrow \mathbb{Z}$$
$$a \longmapsto \sum_{i \in \mathbb{Z}} F(a - i) K(i) \tag{1}$$

For practical purpose, we consider that the kernel K has a finite support (i.e. K is zero out of some bounded subset I of $\mathbb{Z}$). By doing so, a computer can handle the calculation of the convolution product. Then, the computation of $F * K(a)$ can be seen as a balanced sum of values of F for abscissas close to a with coefficients given by the kernel K.

Definition 2. *Let $F : \mathbb{Z} \longrightarrow \mathbb{Z}$ be a discrete function and $K : \mathbb{Z} \longrightarrow \mathbb{Z}$ a convolution kernel. We define the operator Ψ_K aimed to modify the function F by convolution with the kernel K:*

$$\Psi_K F : \mathbb{Z} \longrightarrow \mathbb{Z}$$
$$a \longmapsto F * K(a) \tag{2}$$

We will now construct the kernel at the basis of our derivative estimator.

Finite Differences. Finite differences are a widely used method for estimating derivatives of continuous functions and solving partial differential equations.

Definition 3. *Let $F : \mathbb{Z} \longrightarrow \mathbb{Z}$ be a discrete function and let $a \in \mathbb{Z}$. The backward finite difference of F at the point of abscissa a is defined by:*

$$F(a) - F(a-1) \tag{3}$$

Considering that the derivative of F at a is the slope of the function tangent at this point, backward finite differences are a method to estimate the derivative of a discrete function.

Proposition 1. *For a discrete function F, the backward finite difference can be computed with $\Psi_\delta F$ where δ is the kernel defined as follow:*

$$\delta(a) = \begin{cases} 1 & \text{if } a = 0 \\ -1 & \text{if } a = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Smoothing Kernel.

Definition 4. *For $n \in \mathbb{N}$, we define the* smoothing kernel, *denoted H_n, using binomial coefficients, as follow:*

$$H_n(a) = \begin{cases} \binom{n}{a+\frac{n}{2}} & \text{if } n \text{ is even and } a \in \left\{-\frac{n}{2}, \ldots, \frac{n}{2}\right\} \\ \binom{n}{a+\frac{n+1}{2}} & \text{if } n \text{ is odd and } a \in \left\{-\frac{n+1}{2}, \ldots, \frac{n-1}{2}\right\} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

H_n can be seen as the n-th line of the Pascal's triangle recentered on 0. The expression $\frac{1}{2^n}\Psi_{H_n}F(a)$ is a weighted mean of a set of values of F near the point of abscissa a giving a stronger weight for values near a and, on the contrary, a weaker weight for farther values. The division by 2^n comes from the fact that the sum of the n-th line coefficients of the Pascal's triangle is always equal to 2^n.

Derivative Kernel. Since preliminaries are set up, we can define the core kernel of our derivative estimator.

Definition 5. *For $n \in \mathbb{N}$, we define the* derivative kernel, *denoted D_n, by:*

$$D_n = \delta * H_n \tag{6}$$

So, our derivative estimator for a discrete function F is defined by:

$$\frac{1}{2^n}\Psi_{D_n}F \tag{7}$$

Composed of two parts, H_n and δ, the derivative kernel D_n has two effects. Thanks to H_n, the discrete function F is smoothed and so, the possible noise is reduced. Then, the kernel δ evaluates the derivative but on a cleansed function.

2.2 Convergence

At this point, we have defined our digital derivative estimator. Now, we want to estimate the quality of this estimator. To do so, given a continuous function and its discretization, we want to evaluate the error made by our estimator with respect to the theoretical derivative of the original continuous function.

The first thing to do is to establish a correspondence between a theoretical continuous function and its discretization. Let $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous function and let $\Gamma : \mathbb{Z} \longrightarrow \mathbb{Z}$ be a discrete function. Let h be the discretization step (i.e. the size of a pixel). Let $K > 0$ and $\alpha \in]0, 1]$ be two additional parameters. We say that Γ is a (possibly noisy) discretization of ϕ if for all $a \in \mathbb{Z}$ we have that:

$$|h\Gamma(a) - \phi(ha)| \leq Kh^\alpha \tag{8}$$

In other words, the point $(a, \Gamma(a))$ is close to the point $(a, \frac{\phi(ha)}{h})$ for all a of the domain. The deviation between these two points is bounded by the quantity Kh^α where K is a general constant and α is a parameter which aims to represent the amount of noise (considering a small discretization step, $h < 1$, the allowed noise is more important with α close to 0 than with α close to 1).

Theorem 1. *With the following hypothesis:*

- $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ *is a C^3 function*
- $\phi^{(3)}$ *is bounded*
- $\alpha \in]0, 1]$, $K \in \mathbb{R}_+^*$ *and $h \in \mathbb{R}_+^*$*
- $\Gamma : \mathbb{Z} \longrightarrow \mathbb{Z}$ *is such that $|h\Gamma(a) - \phi(ha)| \leq Kh^\alpha$*
- $n = \lfloor h^{2(\alpha-3)/3} \rfloor$

there exists a function $\sigma_{\phi,\alpha} : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ with $\sigma_{\phi,\alpha} \in O(h^{(2/3)\alpha})$ such that:

$$\left| \frac{1}{2^n} \Psi_{D_n} \Gamma(a) - \phi'(ha) \right| \leq \sigma_{\phi,\alpha}(h) \tag{9}$$

In other words, we provide a derivative estimation converging at rate $h^{(2/3)\alpha}$ of functions known through their discretizations with step h and an arbitrary noise bounded by Kh^α.

We don't report here the proof of Theorem 1 for lack of space. We can just point out that our proof relies on the use of the Floater's Theorem ([2]) which gives the order of convergence of the k-th derivative of the Bernstein approximation[4] of a function and the k-th derivative of the function itself. The hypothesis made in Theorem 1 about the fact that ϕ must be C^3 with $\phi^{(3)}$ bounded come from the hypothesis required by the Floater's Theorem.

---

[4] The Bernstein approximation of a function $f : [0, 1] \longrightarrow \mathbb{R}$ is the Bézier curve with control points $\left(\frac{i}{n}, f\left(\frac{i}{n} \right) \right)_{i \in \{0,...,n\}}$.

2.3 Reducing the Complexity

Considering the maximum noise allowed by the Theorem 1 (that is, $\alpha = 0$) the computational complexity of our derivative estimator is quadratic ($O(h^{-2})$, where h is the discretization step). This complexity is not as good as the one induced by tangent estimators from discrete geometry which is in $O(h)$ for the best ([6]). However, it is possible to reduce the complexity of our estimator by considering that values at extremities of the smoothing kernel H_n are negligible.

The following theorem comes from the Hoeffding inequality ([3]).

Theorem 2. *Let* $\beta \in \mathbb{N}^*$ *and* $n \in \mathbb{N}$*. If* $k = \frac{n}{2} - \sqrt{\frac{\beta n \ln(n)}{2}}$ *then:*

$$\frac{1}{2^n} \sum_{j=0}^{k} \binom{n}{j} \leq \frac{1}{n^\beta} \tag{10}$$

Theorem 2 means that the sum of the k first and the k last coefficients of the smoothing kernel are negligible with respect to the whole kernel. The parameter β enables to define what negligible mean. Typically, we chose $\beta = 2$. Using the result of Theorem 2, it is possible to reduce the size of the derivative kernel D_n recommended by Theorem 1. As a consequence, the computational complexity of our method can be improved without spoiling the quality of the derivative estimation. For example, with $\beta = 2$, the computational complexity is reduced to $O\left(\frac{\sqrt{-\ln(h)}}{h}\right)$ (with the hypothesis that the noise is maximal).

2.4 Higher Order Derivatives

One of the most interesting point about our method is that higher order derivatives can also be estimated.

Definition 6. *Let* $n \in \mathbb{N}$ *be the size of the kernel and* $k \in \mathbb{N}^*$ *the derivation order. We define the* kernel for higher order derivatives, *denoted* D_n^k*, as follow:*

$$D_n^k = \underbrace{\delta * \ldots * \delta}_{k\ times} * D_n \tag{11}$$

As for first order derivatives, we have proved the convergence of the estimator for higher order derivatives.

Theorem 3. *With the following hypothesis:*

- $k \in \mathbb{N}^*$
- $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ *is a* C^{k+2} *function*
- $\phi^{(k+2)}$ *is bounded*
- $\alpha \in]0,1]$*,* $K \in \mathbb{R}_+^*$ *and* $h \in \mathbb{R}_+^*$
- $(\alpha_r)_{r \in \mathbb{N}}$ *a sequence such that* $\alpha_r = (2/3)^r \alpha$
- $\Gamma : \mathbb{Z} \longrightarrow \mathbb{Z}$ *is such that* $|h\Gamma(a) - \phi(ha)| \leq Kh^\alpha$

$$- \; n = \sum_{l=0}^{k-1} \left\lfloor h^{2(\alpha_l - 3)/3} \right\rfloor$$

there exists a function $\sigma_{\phi,\alpha,k} : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ *with* $\sigma_{\phi,\alpha,k} \in O(h^{(2/3)^k \alpha})$ *such that:*

$$\left| \frac{h^{-k+1}}{2^n} \Psi_{D_n^k} \Gamma(a) - \phi^{(k)}(ha) \right| \leq \sigma_{\phi,\alpha,k}(h) \tag{12}$$

The proof of Theorem 3 is obtained by iterating the result of the Theorem 1. As a remark, we can point out that the Theorem 3 is identical to Theorem 1 if we consider the case $k = 1$.

2.5 Experimental Results

As an example, Fig. 1 shows the derivative estimation (in gray) of a noisy sine (in black). On this first example, we have used a kernel size $n = 25$.
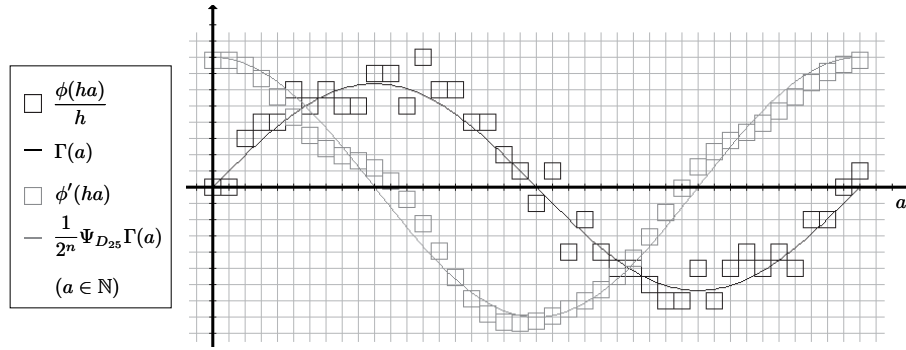

Fig. 1. Example of derivative estimation of a noisy sine wave.

In Fig. 2, we start from a simple continuous sine wave. This original function is then discretized. After that, we compute the derivative from the discretized data using our estimator. Finally, the result is compared with the theoretical continuous derivative we should get. To do so, we use the cumulative error, that is, the sum of absolute differences between the theoretical values we should get and the estimated ones. These measures are repeated for several discretization steps (h) and derivative kernel widths (n). The grayed part in Fig. 2 is the place where the cumulative error happens to be minimum. This area corresponds to the optimal derivative kernel width for a given discretization step. We can see that this zone is compatible with the result of Theorem 1.

3 Derivatives Estimation for Parametrized Curves

A parametrized function is a function $\phi : \mathbb{R} \longrightarrow \mathbb{R}^2$ made up of two coordinates $\phi_x : \mathbb{R} \longrightarrow \mathbb{R}$ and $\phi_y : \mathbb{R} \longrightarrow \mathbb{R}$. In the same manner, a discrete parametrized
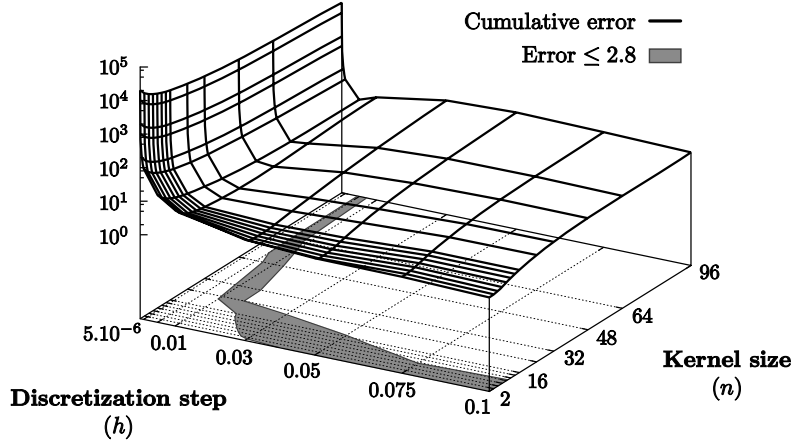
Fig. 2. Cumulative error.

function is a function $\Gamma : \mathbb{Z} \longrightarrow \mathbb{Z}^2$ made up of two discrete coordinates Γ_x and Γ_y. In order to estimate the derivative of a discrete parametric function, we naturally want to apply our estimator to its coordinates. After that, we want to evaluate the quality of such an estimator. So, we build a correspondence between the parametrization of the continuous function ϕ and the pixels numbering of the discrete function Γ in a manner similar to the case of real functions (see Equ. 8).

3.1 Pixel Length Parametrization

As we have seen for the case of real functions, we need to establish a correspondence between a theoretical continuous function and its discretization in order to evaluate the quality of the derivative estimation. So, given a continuous parametric function ϕ and its discretization Γ, we would like an expression of the form:

$$\|h\Gamma(a) - \phi(ha)\| \leq \varepsilon \tag{13}$$

The problem is that Equation 13 is not necessarily satisfied if we consider ε small enough to be meaningful. This is due to the non-uniform parametrization of ϕ and the non-isotropic character of $\mathbb{Z}^2$. Therefore, we build a reparametrization $\overline{\phi}$ of the function ϕ such as:

$$\left\|h\Gamma(a) - \overline{\phi}(ha)\right\| \leq \varepsilon \tag{14}$$

Remark that the quality estimation of the derivative estimator will consider the reparametrization of the initial curve instead of the initial curve itself. As a consequence, the norm of the tangent estimation will not be meaningfull but its orientation will.

The problem of correspondence between parametrizations is solved by introducing a new notion, the pixel length parametrization. This notion is similar to the parametrization with respect to arc length but relies on the 1-norm (the taxicab norm) despite of the euclidean norm.

Definition 7. *Let $\phi : I \subset \mathbb{R} \longrightarrow \mathbb{R}^2$ be a parametric function. We define the length function, denoted L, by:*

$$L : I \longrightarrow \mathbb{R}_+$$
$$u \longmapsto \int_{\min(I)}^{u} \|\phi'(t)\|_1 \, dt \tag{15}$$

The function L gives the arc length according to the 1-norm of the curve ϕ between the origin of the parametrization and the parameter u.

Definition 8. *Let $\phi : I \subset \mathbb{R} \longrightarrow \mathbb{R}^2$ a continuous parametric function. Similarly to arc length, we define the pixel length reparametrization of the function ϕ, denoted $\overline{\phi}$, by:*

$$\overline{\phi} : L(I) \longrightarrow \mathbb{R}^2$$
$$x \longmapsto \phi\big(L^{-1}(x)\big) \tag{16}$$

We have proved that there is a correspondence between the continuous parametrization of $\overline{\phi}$ and the pixels numbering of an 8-connected discretization Γ of ϕ such that Equ. 14 is true.

Since we are able to connect a parametric continuous function and its discretization, we want to evaluate the quality of our derivative estimator. Unfortunately, there is still a problem. Indeed, the pixel length reparametrization is only piecewise C^3 and C^1 where the curve admits a vertical or horizontal tangent. This comes from the use of the 1-norm which involves absolute values. As a consequence, the hypothesis of Theorem 1 are not satisfied and it cannot be directly applied to prove the convergence of our derivative estimator for parametrized curves. At the moment, we have an another theorem for the parametric case proving the convergence of our derivative estimator which gives a convergence rate of $O(h^{2/3})$ almost everywhere on the curve and $O(h^{1/3})$ in the neighborhood of vertical and horizontal tangents. Though, from experimental results, it seems that the convergence rate is uniform $O(h^{2/3})$.

3.2 Experimental Results

Fig. 3 is an example of derivative estimation for parametrized curves. The tangents for all points of the discrete curve (black squares) are represented with black segments. Note that in this figure, the gray grid is a representation of $\mathbb{Z}^2$. From this illustration, it seems that derivative estimation is not worse for horizontal and vertical tangents than elsewhere.

Fig. 4 is a more precise experimentation intended to illustrate the behavior of our derivative estimator for vertical and oblique tangents. In this test, we start from a parametric continuous circle (centered on 0 and with radius 1). Then we
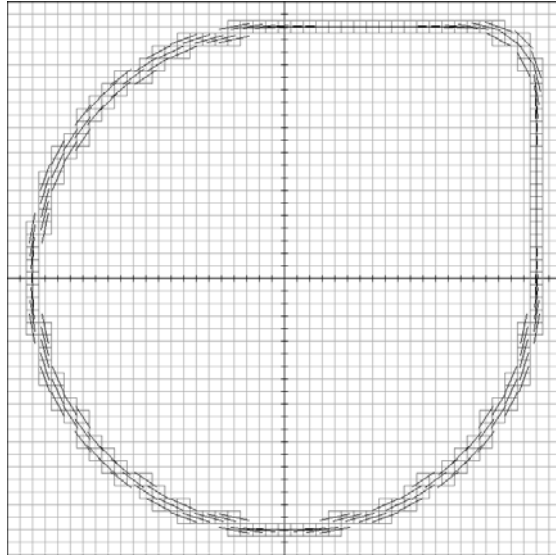
Fig. 3. Example of tangent estimation for a parametrized curve.

discretize this base function for several discretization steps using the pixel-length reparametrization. After that, we compute the tangent with our derivative estimator for two locations: the first one is a vertical tangent and the second one is an oblique tangent. The results are then compared with the theoretical tangents we should get (computed analytically from the continuous original function). The error is defined as the angle between the estimated tangent and the theoretical one (in radians). From Fig. 4, we can see that our estimator is not worse for the vertical tangent than for the oblique one. Besides, we can also notice than the estimator gives better results for small discretization steps.

4   Conclusion

In this paper, we have defined a new approach based on discrete convolution products to estimate derivatives of digitized functions. We showed that under some conditions this new method has a good order of convergence even on noisy data. Given a discretization step and a noise strength, it is possible to automatically select the size of our derivative kernel to achieve the best convergence rate. As a result, the user won't have to set too many parameters.

An another benefit of this approach is that computations involved in the derivative estimation can be done using integer only arithmetic. Moreover, higher order derivatives can be estimated.

By reducing the size of the derivative kernel in such a way that the lost of quality is negligible, our computational complexity is comparable to [6].
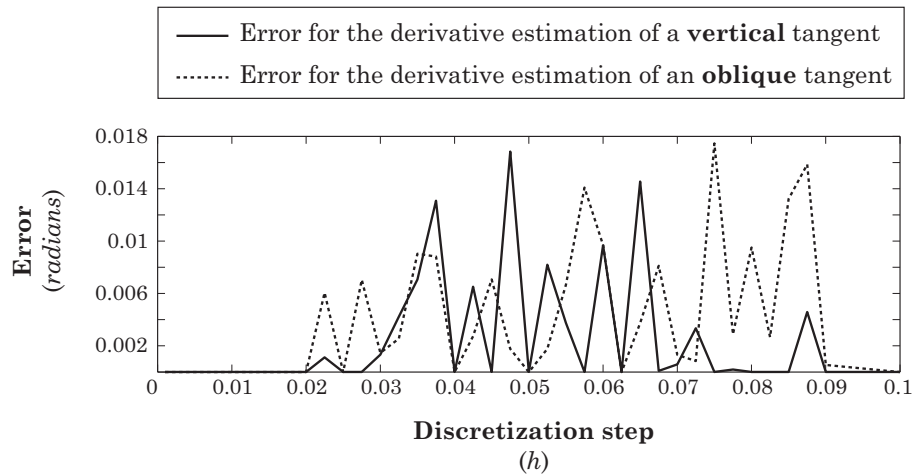
Fig. 4. Comparison of derivative estimator in the parametric case for vertical and oblique tangents.

Even if there is some technical problems for proving the exact convergence rate, experimental results showed that our derivative estimator still works for parametrized curves. Moreover, we have set up the new notion of pixel length parametrization which solves the problem of correspondence between the pixels numbering of a discrete curve and the parametrization of a continuous function.

References

1. S.H. Chen. *Finite Difference Method.* High-Field Physics and Ultrafast Technology Laboratory, Taipei, Tawan, 2006.
2. Michael S. Floater. On the convergence of derivatives of Bernstein approximation. *J. Approx. Theory*, 134(1):130–135, 2005.
3. Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, march 1963.
4. Ron Kimmel, Nir A. Sochen, and Joachim Weickert, editors. *Scale Space and PDE Methods in Computer Vision, 5th International Conference, Scale-Space 2005, Hofgeismar, Germany, April 7-9, 2005, Proceedings*, volume 3459 of *Lecture Notes in Computer Science*. Springer, 2005.
5. Jacques-Olivier Lachaud and François de Vieilleville. Convex shapes and convergence speed of discrete tangent estimators. In *ISVC (2)*, pages 688–697, 2006.
6. Jacques-Olivier Lachaud, Anne Vialard, and François de Vieilleville. Analysis and comparative evaluation of discrete tangent estimators. In *DGCI*, pages 240–251, 2005.
7. Jacques-Olivier Lachaud and François De Vieilleville. Fast accurate and convergent tangent estimation on digital contours. *Image and Vision Computing*, 2006. In press.

8. Tony Lindeberg. Scale-space for discrete signals. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(3):234–254, 1990.