International Journal of Computer Vision 69(1), 7–25, 2006

DOI: 10.1007/s11263-006-6849-5

Stochastic Motion and the Level Set Method in Computer Vision: *Stochastic Active Contours*

OLIVIER JUAN, RENAUD KERIVEN* AND GHEORGHE POSTELNICU
The Odyssee Lab (ENPC CERTIS/ENS Paris/INRIA Sophia)

Received April 7, 2004; Revised November 17, 2004; Accepted November 23, 2004

First online version published in April, 2006

Abstract. Based on recent work on Stochastic Partial Differential Equations (SPDEs), this paper presents a simple and well-founded method to implement the stochastic evolution of a curve. First, we explain why great care should be taken when considering such an evolution in a Level Set framework. To guarantee the well-posedness of the evolution and to make it independent of the implicit representation of the initial curve, a *Stratonovich* differential has to be introduced. To implement this differential, a standard *Ito plus drift* approximation is proposed to turn an implicit scheme into an explicit one. Subsequently, we consider shape optimization techniques, which are a common framework to address various applications in Computer Vision, like segmentation, tracking, stereo vision etc. The objective of our approach is to improve these methods through the introduction of stochastic motion principles. The extension we propose can deal with local minima and with complex cases where the gradient of the objective function with respect to the shape is impossible to derive exactly. Finally, as an application, we focus on image segmentation methods, leading to what we call *Stochastic Active Contours*.

Keywords: level sets, stochastic partial differential equations, shape optimization, stochastic optimization, segmentation, active contours

1. Introduction

1.1. Why Adding Noise?

Shape optimization techniques are a common framework to address various applications in Computer Vision, like segmentation, tracking, stereo vision etc. The objective of our approach is to improve these methods through the introduction of stochastic motion principles. These problems are most of the time stated as the minimization with respect to some hyper-surface Γ of $\mathbb{R}^N$ of some objective function $E(\Gamma)$. This is usually achieved using a gradient-descent method. Yet, in complex cases, E does not have any computable gradient with respect to Γ (see Section 5.2). In other cases, the minimization process gets stuck into some local minimum, while no multi-resolution approach can be invoked. To deal with those two frequent problems, one can naturally turn to a stochastic optimization approach. Even a simple Simulated Annealing method might be powerful enough to escape from local minima and to cope with an approximation of the shape gradient. Indeed, adding noise to the motion of a curve is a prerequisite to developing this idea.

*Corresponding author. CERTIS, ENPC, 77455 Marne la Vallee, France. keriven@certis.enpc.fr

1.2. Context

We are interested in letting $\Gamma(t)$ evolve according to the equation

$$\frac{\partial \Gamma}{\partial t}(t, p) = \beta(t, p)\mathbf{n}(t, p) = \dot{\eta}(t, p)\mathbf{n}(t, p) \quad (1)$$

where p is some parameterization of Γ, $\mathbf{n}$ the normal to $\Gamma(t)$ at point $\Gamma(t, p)$ and where the normal velocity β depends on some stochastic perturbation $\dot{\eta}$—here, the notation $\dot{\eta}$, standing for the "derivative" of the noise η w.r.t. time, will become clear further. The mean curvature motion $\beta = \kappa$ as well as many other problem oriented choices of β and their implementation with the Level Sets method (Osher and Sethian, 1988) are well known. The novelty in our work is the implementation of the recently proposed stochastic flow (1) (see (Walsh, 1994)) and its application to Computer Vision.

Stochastic dynamics of interfaces have been widely discussed in later years in the physics literature. The work in fields like front propagation or front transition is aimed at modeling and studying the properties of a moving frontier between two media that is subject to macroscopic constraints and random perturbations (which are due to the bulk). The natural translation of this dynamic in mathematical language is done through Stochastic Partial Differential Equations (SPDEs). These equations were introduced by Walsh in Walsh (1994) and their mathematical properties were studied using mostly partial differential equations tools. Nevertheless, the problems researchers have to deal with are various and there is more than one way to add a stochastic perturbation to a PDE. An up to date survey of the existing models on stochastic motions by mean curvature can be found in Yip (2002). It was only in recent years that the notion of viscosity solution for a SPDE was developed by Lions and Souganidis in a series of articles (Lions and Souganidis, 1998a, b, 2000a, b). Their notion of weak viscosity solution is very attractive for numerical applications, since they define the solution as a limit in a convenient space for a set of approximations. Since their pioneering work, related work has been done by Yip (1998) and by Katsoulakis et al. (2001). Another independent approach concerned with viscosity solutions of stochastic partial differential equations is due to Buckdahn and Ma (2001). Their approach is not well suited for Level Sets evolutions, though, since they do not allow certain functional dependencies that are common to all Level Sets evolutions.

1.3. What Should not be Done!

Mistake #1: Considering Eq. (1), one should be tempted to make the perturbation $\dot{\eta}$ depend on the parameter p, or, to make it intrinsic,[1] on σ, the arc length parameter of the curve. Such an evolution, namely

$$\frac{\partial \Gamma}{\partial t}(t, p) = \dot{\eta}(t, \sigma(t, p))\mathbf{n}(t, p)$$

is actually unstable, even if η has some regularity with respect to σ and even if σ is normalized with respect to the total length of the curve. Suppose the curve develops a kink at some point. Then, its length will increase at that point and that will lead to adding more noise around the point where the kink formed itself. Consequently, that might lead to an unstable character of the evolution. That is why we will consider the stochastic perturbation as a function of space and the corresponding evolution:

$$\frac{\partial \Gamma}{\partial t}(t, p) = \dot{\eta}(t, \Gamma(t, p))\mathbf{n}(t, p)$$

Our noise will be regular in space and white in time. Indeed, there is no reason one would expect the random perturbations be correlated from one time step to the other, but considering noise that is white in space may have disastrous effects w.r.t. the regularity of the contours.

Mistake #2: A first simple choice for η is to suppose it constant in space and to consider that each of the increments $\dot{\eta}(t)$ is an independent Gaussian random variable. Therefore, $\eta(t)$ is a Brownian motion W:

$$\frac{\partial \Gamma}{\partial t} = \dot{W}(t)\mathbf{n}(t, p) \tag{2}$$

An error would be to believe that it should be implemented with the explicit scheme:

$$\Gamma(t + \Delta t, p) = \Gamma(t, p) + \Delta t\, \mathcal{N}_{(0,1)}(t)\, \mathbf{n}(t, p) \tag{3}$$

where $\mathcal{N}_{(0,1)}(t)$ denotes a standard Gaussian random variable. This would be incorrect, since the statistical properties of the curve would then depend upon the discretization of the time grid. To see this, consider independent variables $x_i \sim \mathcal{N}_{(0,1)}$ and notice that the previous evolution at time T would amount to $\sum_{i=1}^{n} x_i \Delta t = \sum_{i=1}^{n} \frac{T}{n} x_i$ where $\Delta t = T/n$ is the discretization step. Given the independence of the x_i, the previous sum is a Gaussian variable $\mathcal{N}_{(0, \frac{T^2}{n})}$, thus depending upon the discretization of the time interval $[0, T]$. We will see that the correct scheme involves $\sqrt{\Delta t}$ instead of Δt (see Eq. (13)):

$$\Gamma(t + \Delta t, p) = \Gamma(t, p) + \sqrt{\Delta t}\, \mathcal{N}_{(0,1)}(t)\, \mathbf{n}(t, p) \tag{4}$$

The previous argument is actually a trivial example that shows the difference between stochastic integration and Stieltjes integration. This is to say that our modelization will have to obey rules of Stochastic Calculus relying on intrinsic properties of Brownian motion. But we must also rely on the theory of viscosity solutions, since it is a necessary ingredient in the Level Set framework—allowing for rarefactions and shocks in evolutions and providing stable numerical schemes. The theory we need must then fill the gap between Stochastic Calculus and viscosity solutions. This is where the recent theory developed by Lions and Souganidis comes into place.

Mistake #3: Let us now try to implement (2) in a Level Set framework, where $\Gamma(t)$ is the zero level set of some implicit function $u(t, x)$ driven by the evolution $\frac{\partial u}{\partial t} = |Du|\dot{W}(t)$. Actually, a more correct way to write this SPDE is:

$$du = |Du|dW(t) \tag{5}$$

As pointed out in Lions and Souganidis (1999), this equation is not reasonable and suffers from:

- *Non invariance*: Let $\alpha(\cdot)$ be some smooth increasing function with $\alpha(0) = 0$. If $u(t, x)$ is solution of (5) given some initial condition $u_0(x)$, then $\alpha(u(t, x))$ is not solution of (5) with initial condition $\alpha(u_0)$. Moreover, the solution of (5) with initial condition $\alpha(u_0)$ has not the same zero level set than $u(t, x)$: the evolution depends on the implicit representation of the initial curve (see Section 2.2)!
- *Ill-posedness*: Let us take $N = 1$ for the sake of simplicity. The equation $du = u_x dW + \lambda u_{xx} dt$, obtained by adding the curvature term λu_{xx} to (5), reveals to be an inverse heat equation for $0 < \lambda < 1/2$

It turns out that the differential used in (5) is the *Ito differential*, and that those two difficulties are overcome introducing the *Stratonovich differential*:

$$du = |Du| \circ dW(t) \tag{6}$$

In the sequel, we briefly introduce the essential notions needed to understand the difference between the Ito and the Stratonovich cases. Citing (Lions and Souganidis, 1998a), we will see that the notion of viscosity solution can be extended to the SPDE case. We propose an effective implementation of (6), that we extend to the case when the noise term depends on the space variable as well; then we investigate some geometrical properties of the evolution that could guide the user toward correct noise parameters. Subsequently, we explain how the stochastic motion can improve the shape optimization based methods in Computer Vision. Finally, as an application, we focus on image segmentation, leading to what we call *Stochastic Active Contours*.

2. Mathematics

2.1. Some Notions of Stochastic Calculus

This subsection is meant to offer the reader an intuition of the notion of Stochastic Calculus and of the supplementary challenges it poses. Focusing on the definition of the integral itself, we suppose the reader is familiar with the Brownian motion. We shall equally use the idea of a standard probability space, martingale, quadratic variation. Rigorous and complete introductions of Stochastic Calculus can be found in Karatzas and Shreve (1991), Gard (1988) or Kunita (1990). Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a standard probability space. We will consider that $W = (W_1(t), \ldots, W_m(t))_{t \geq 0}$ is a standard m-dimensional Brownian motion issued from 0. We are interested in finding an appropriate way of introducing the notion of stochastic differential with respect to the process W. To better understand the difficulty here, it is worth while mentioning that the paths of the Brownian motion are only $\frac{1}{2}$-Holder continuous, so they are nowhere differentiable. Hence, in order to give a meaning to what the term $dW(t)$ might mean, one can first define an integral with respect to W—the stochastic integral. Once this integral is defined, the differential is obtained using the integral defined. To keep the presentation as clear as possible, we suppose that $m = 1$ and all our processes are real-valued. The considered approach for the construction of such an integral is to define it as an isometry in the appropriate functional space. Indeed, consider a square integrable process $\Phi = (\Phi(t, \omega))_{t \geq 0}$. Trying to define the stochastic integral $\int_0^T \Phi(t, \omega) dW(t)$, one would start with Riemann approximations

$$I_\Delta(\Phi)(T) = \sum_{i=1}^{n-1} \Phi(t_i, \omega)\,(W(t_{i+1}) - W(t_i)) \tag{7}$$

with $\Delta = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ and hope to find a suitable space where the above sum

would converge when $|\Delta| \to 0$. By proving some completeness results, one can show that every square integrable continuous process is integrable w.r.t. the Brownian motion and one can obtain the **Ito stochastic integral** as the unique, square integrable martingale $I(\Phi) = (I(\Phi)(t))_{t\geq 0}$ which is the limit of $(I_\Delta(\Phi)(t))_{t\geq 0}$ when $|\Delta| \to 0$. The convergence happens in the pseudo-metric $\|X\| = \sum_{n=1}^{\infty} \frac{\min(1, \sqrt{\mathbb{E}(X_n^2)})}{2^n}$.

The price to pay for the convergence of the Riemann sums is that it happens in a process space, hence the limit, which is denoted as the stochastic integral $\int_0^T \Phi(s)dW(s)$, does not hold a meaning path-wise, but only as a process. This means that the set of events $\omega \in \Omega$ where the Riemann sums mentioned above do not converge to the above introduced integral is of measure 0. In the sequel, the convergence suggested above can be extended to an arbitrary dimension. Moreover, the extension can be taken with respect to local continuous martingales. For details, the reader is referred to one of the references above. Once this integral is defined, it is imperative one is given some chain rule formula. This is where the Ito lemma comes in. Consider a process $X = (X_t)_{t\geq 0}$ and $\alpha : \mathbb{R} \to \mathbb{R}$ a function of class $\mathcal{C}^2$. Then the Ito formula states that $Y = (\alpha(X(t)))_{t\geq 0}$ verifies the dynamics

$$dY(t) = \alpha'(X(t))dt + \frac{1}{2}\alpha'' d\langle X, X\rangle(t) \tag{8}$$

The main difference when compared to the regular chain rule is the appearance of an extra term, also called Ito term, **or drift**, which involves the second derivative of α and the quadratic form $\langle X, X\rangle$, also known as the quadratic variation of the process X. The quadratic variation is not zero when X has some dependence upon a stochastic process and can be computed in the following manner. Suppose that $X_1(t) = \int_0^t f_1(s)dW(s) + A_1(t)$, where f_1 is some continuous square integrable function and $A_1 : \mathbb{R} \to \mathbb{R}$ is continuous and increasing. Suppose similarly that $X_2(t) = \int_0^t f_2(s)dW(s) + A_2(t)$. Then

$$\langle X_1, X_2\rangle(t) = \int_0^t f_1(s) f_2(s)\, ds$$

Note that $\langle X, X\rangle$, depends solely on the stochastic part of the dynamics and is independent of the bounded variation part. When $f \equiv 0$, then the classical chain rule is obtained in (8).

Now, using the Ito formula, a variation of the stochastic integral introduced above can be obtained so that the classical chain rule is satisfied. Consider a process $X(t) = x + M(t) + B(t)$ where M is a local continuous martingale[2] and B an increasing process. Consider equally another continuous process $Y(t) = y + N(t) + C(t)$ where N is a local continuous martingale and C an increasing process. The **Stratonovich integral** of Y with respect to X is then given by the formula

$$\int_0^t Y(s) \circ dX(s) = \int_0^t Y(s)dX(s) + \frac{1}{2}\langle M, N\rangle(t) \tag{9}$$

Suppose now α is of class $\mathcal{C}^3$ and apply the Ito formula (8) to $\alpha'(X)$. Then

$$d\alpha'(X(t)) = \alpha''(X(t))dW(t) + \frac{1}{2}\alpha^{(3)}(X(t))dt$$

and consequently $d\langle \alpha'(X), X\rangle(t) = \alpha''(X(t))dt$. Hence,

$$\alpha(X_t) = \alpha(X(0)) + \int_0^t \alpha'(X(s)) \circ dX(s) \tag{10}$$

by noticing that the quadratic variation term that we obtained is equal to the Ito term in Eq. (8).

We conclude this section by the approximation of the stochastic integral mentioned above. For the Ito integral, we saw that, if $\Delta = \{0 = t_0 < t_1 < \cdots < t_n = T\}$, then

$$\lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} Y(t_i)(X(t_{i+1}) - X(t_i)) = \int_0^T Y(s)\, dX(s)$$

For the Stratonovich case, it can be proved that we have the two equally useful limits:

$$\lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} \frac{Y(t_i) + Y(t_{i+1})}{2}(X(t_{i+1}) - X(t_i)) = \int_0^T Y(s) \circ dX(s) \tag{11}$$

$$\lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} Y\left(\frac{t_i + t_{i+1}}{2}\right)(X(t_{i+1}) - X(t_i)) = \int_0^T Y(s) \circ dX(s) \tag{12}$$

As an example that is meant to emphasize the difference between the two integrals previously introduced, consider as before a Brownian motion W and

$\lambda \in [0, 1]$. Let $\Delta = \{a = t_0 < t_1 < \cdots < t_n = b\}$, $\Delta W_i = W(t_{i+1}) - W(t_i)$ and consider $\phi^\lambda(t_i) = (1 - \lambda)W(t_i) + \lambda W(t_{i+1})$. It can be proved that

$$\begin{aligned}\int_a^b W(s) \circ dW(s) &= \lim_{|\Delta|\to 0} \phi^{\frac{1}{2}}(t_i)\Delta W_i \\ &= \frac{1}{2}[W^2(b) - W^2(a)] \\ \int_a^b W(s)dW(s) &= \lim_{|\Delta|\to 0} \phi^0(t_i)\Delta W_i \\ &= \frac{1}{2}[W^2(b) - W^2(a)] - \frac{1}{2}(b - a)\end{aligned}$$

Note also that one should be careful when simulating ΔW_i (see Section 1.3):

$$\Delta W_i \sim \mathcal{N}_{(0,t_{i+1}-t_i)} \sim \sqrt{t_{i+1} - t_i}\mathcal{N}_{(0,1)} \tag{13}$$

2.2. *Proposed Model for the Stochastic Curve Evolution*

In applications, the stochastic term will add to a deterministic force $F(D^2u, Du, x, t)$ (one of the simplest examples, analyzed in detail in Yip and Souganidis (2004), is concerned with the coupled evolution $du = \kappa|Du|dt + \dot{W}|Du|dt$). Hence, a naive way to write down the coupled evolution is

$$du = F|Du|dt + \dot{W}|Du|dt$$

The above equation will have a meaning if written as

$$du = F|Du|dt + |Du|dW(t) \tag{14}$$

Concentrating on the stochastic part again, we remark that we made an implicit choice by considering the Ito integral in the above formula, but we could have decided to go for the Stratonovich integral. So what is the difference between the two integrals from a Level Sets point of view? Let us consider the following invariance property that is required when working within a Level Sets framework: consider just a random evolution of the type

$$du = |Du|dW(t) \text{ with } u(0, \cdot) = u_0 \tag{15}$$

where W is a one-dimensional Brownian motion. Then this evolution codes for the corresponding contour evolution

$$\frac{\partial \Gamma}{\partial t} = \dot{W}\mathbf{n} \text{ with } \Gamma(0) = \Gamma_0 \tag{16}$$

where $\dot{W}$ is Gaussian white noise and Γ_0 is the zero-level of u_0. The idea behind the Level Sets evolution framework is to have all the level sets of the implicit function given by (15) evolve according to the same dynamics (16). A smooth change of scale of a function satisfying (15) that leaves the zero-level unchanged should not influence the dynamics of the level sets contour—since the corresponding contour evolution (16) is not affected by this change of scale. Consider then a function $\alpha : \mathbb{R} \to \mathbb{R}$ such that $\alpha' > 0$ and $\alpha(0) = 0$ and the initial value problem

$$du = |Du|dW(t) \quad \text{with} \quad u(0, \cdot) = u_0(\cdot) \tag{17}$$

If we consider u the solution of (17), then $v = \alpha(u)$ should verify the same dynamics, but with a different initial condition

$$dv = |Dv|dW(t) \quad \text{with} \quad v(0, \cdot) = \alpha(u_0(\cdot))$$

as is the case in the deterministic framework. Nevertheless, one can apply the Ito rule to the dynamics (15) and see that

$$\begin{aligned}dv = d\alpha(u) = \alpha'(u)du = |Dv|dW(t) \\ +\frac{1}{2}\alpha''(u)|Du|^2dt\end{aligned}$$

and the assertion is not verified due to the additional Ito term. Hence, the problem (15) is ill-posed from a Level Sets point of view: **for a given initial curve $\Gamma(0)$, the choice of the initial implicit function u_0 modifies the solution of the equation!**. However, as observed by Lions and Souganidis, this invariance condition is verified if one replaces the Ito integral with the Stratonovich integral, since the latter does not include any additional term anymore. Hence, the *right way* to insert stochastic evolutions in the Level Sets framework is through the Stratonovich integral. We rewrite (15) accordingly

$$du = |Du| \circ dW(t) \quad \text{with} \quad u(0, \cdot) = u_0(\cdot) \tag{18}$$

Then, if we consider $v = \alpha(u)$ (hence the corresponding initial condition will be $\alpha(u_0)$) the dynamics

verified by v are

$$\begin{aligned} dv &= \alpha'(u) \circ du = \alpha'(u)|Du| \circ dW(t) \\ &= |Dv| \circ dW(t) \end{aligned}$$

and the invariance property is verified this time. Now, given the previous ingredient, the proposed random curve evolution model is given by

$$du = Fdt + |Du| \circ dW(t) \tag{19}$$

Here, we used the Stratonovich integral, as opposed to (14).

A second example that suggests Stratonovich integration should be used when working with stochastic partial differential equations is concerned with the 1-dimensional perturbed heat equation $du = u_x dW(t) + \lambda u_{xx} dt$. It can be shown that this equation reduces to a backward heat equation when considering Ito integration for $\lambda \in (0, \frac{1}{2})$–hence ill-posed. Once again, the extra term Stratonovich integration solves this problem (for more details, see (Lions and Souganidis, 1999)).

What is the difference between the evolution (19) and a classical Level Sets evolution such as $du = Fdt$? Suppose the initial condition function is a signed distance function. Since the stochastic term only depends upon $|Du|$ (which equals 1 in this case) and the time parameter, all the points of the contour will have an extra random force which will be the same on the entire contour at each time step. This type of perturbation is indeed very important from a theoretical point of view, but we would like something more flexible in our applications. Typically, we would be interested in having white noise in both the time and spatial parameters. Nevertheless, white noise in space appears to add a lot of technical difficulties to the problem and the return on investment is quite small, since most of our models will evolve on discrete grid spaces. That is why we have opted for colored spatial noise, that is typically given by

$$W(t, x) = \sum_{i=1}^{m} \phi_i(x) W_i(t)$$

where $\phi_i : \mathbb{R}^N \to \mathbb{R}$ are smooth functions with compact support. Note that other choices of colored spatial noise are possible. The final evolution model we propose is thus

$$du = F|Du|dt + |Du| \sum_{i=1}^{m} \phi_i(x) \circ dW_i(t) \tag{20}$$

As a simplification, in practice we choose the functions with the same profile, but centered around a number of points x_i, that we call *noise sources*. Thus, our typical choice is

$$\phi_i(x) = \phi(x - x_i)$$

where ϕ is some convenient regular function.

2.3. *Stochastic Viscosity Solutions*

The theory developed earlier needs some sort of convergence results. As mentioned before, the proper type of solutions need to be used, so that the previous results from the Level Sets theory apply here. The notion of stochastic viscosity solution for fully nonlinear, second-order, possibly degenerate, stochastic partial differential equations such as the ones considered previously is put forward in a series of articles: Lions and Souganidis (1998a, b, 2000a, b). Their theory is meant to apply precisely to equations such as (20), with $F = F(D^2u, Du, x, t)$. So far, a limit of their theory, which stands even today as an open question, is that they do not treat equations where the noise depends upon the space parameter (they only treat the case $\phi_i \equiv 1$, with our previous notation). However, experimental data suggests that their theory applies in cases like ours as well (see Section 3). Precisely, consider the equations

$$\begin{aligned} du &= F(D^2u, Du, x, t)dt + \epsilon|Du| \circ dW(t) \\ &\quad \text{with} \quad u(\cdot, 0) = u_0(\cdot) \end{aligned} \tag{21}$$

$$\begin{aligned} du &= F(D^2u, Du, x, t)dt + |Du|\dot{\xi}_\alpha(t) \\ &\quad \text{with} \quad u(\cdot, 0) = u_\alpha(\cdot) \end{aligned} \tag{22}$$

where $\epsilon \geq 0$ and ξ_α is a family of smooth functions $\xi_\alpha : \mathbb{R}_+ \to \mathbb{R}$. Then we can cite the following theorem, summarizing their results:

Theorem 1. *The following hold a.s. in ω:*

1 *There exists a unique solution to (21).*
2 *Let $\{\xi_\alpha(t)\}_{t\geq 0}$ and $\{\eta_\beta(t)\}_{\beta>0}$ be two families of smooth functions such that as α and $\beta \to 0$, ξ_α and η_β converge to W uniformly on any compact in t and a.s. in ω. Let $\{u_\alpha\}_{\alpha>0}$ and $\{v_\beta\}_{\beta>0}$ in $BUC(\mathbb{R}_+ \times \mathbb{R}^N)$[3] be the solutions of (22). If $\lim_{\alpha,\beta\to 0} \|u_\alpha(\cdot, 0) - v_\beta(\cdot, 0)\|_{C(\mathbb{R}^N)} = 0$, then, for all $T > 0$, $\lim_{\alpha,\beta\to 0} \|u_\alpha - v_\beta\|_{C([0,T]\times\mathbb{R}^N)} = 0$. In*

particular, any smooth approximations of W produce solutions converging to the unique function stochastic viscosity solution of (21).

3 *As* $\epsilon \to 0$*, the solution* u^ϵ *of (21) converges in* $\mathcal{C}(\mathbb{R}_+ \times \mathbb{R}^N)$ *to the solution of (21) with* $\epsilon = 0$*.*

Consequently, their result **allows us to simulate the solutions of such equations and be sure that the result of our computer simulation is what we expect it to be**. Furthermore, we mention that according to Lions, the convergence takes place in $C(\mathbb{R}_+ \times \mathbb{R}^N)$, which means that the numerical solutions we develop will be continuous and that they will be converge uniformly almost surely in $\omega \in \Omega$.

We end this theoretical part with an example by Souganidis on the explicit solution of the equation

$$du = |Du|\dot{\eta}dt \quad \text{with} \quad u(0,x) = |x| \qquad (23)$$

where $\eta : \mathbb{R}_+ \to \mathbb{R}$ is a function of class $\mathcal{C}^1$ such that $\eta(0) = 0$. The explicit viscosity solution of this equation is given by

$$u(t,x) = \max\left[(|x| + \eta(t))_+ \,,\, \max_{s\in[0,t]}(\eta(s))_+\right]$$

where $(x)_+ = \max(0,x)$ (for a simple proof of the statement above, consider the case when $\eta(t) \equiv 1$ and compute the viscosity solution of the equation in that case); then, one can see that uniform convergence of $\eta \to W$ is sufficient to obtain the solution of the associated SPDE. Moreover, this simple case allows one to see that the random path η has a different effect on the solution that depends mainly on its sign. Indeed, as it can be observed from formula (23), there is a qualitative difference between the behavior of the solution depending on whether $\dot{\eta} > 0$ or $\dot{\eta} < 0$. This can be better understood watching a sample evolution in Fig. 1. Moreover, numerical artifacts will develop due to very frequent changes of sign of $\dot{\eta}$, since the use of η is only for heuristic purpose (the Brownian motion is nowhere differentiable). As a result, the regular reinitialization of the implicit function—a standard technique of the Level Set framework—is indispensable in the stochastic case.

2.4. *Numerical Scheme*

The main problem when implementing Stratonovich evolutions is that they amount often to implementing implicit numerical schemes. Consider again the

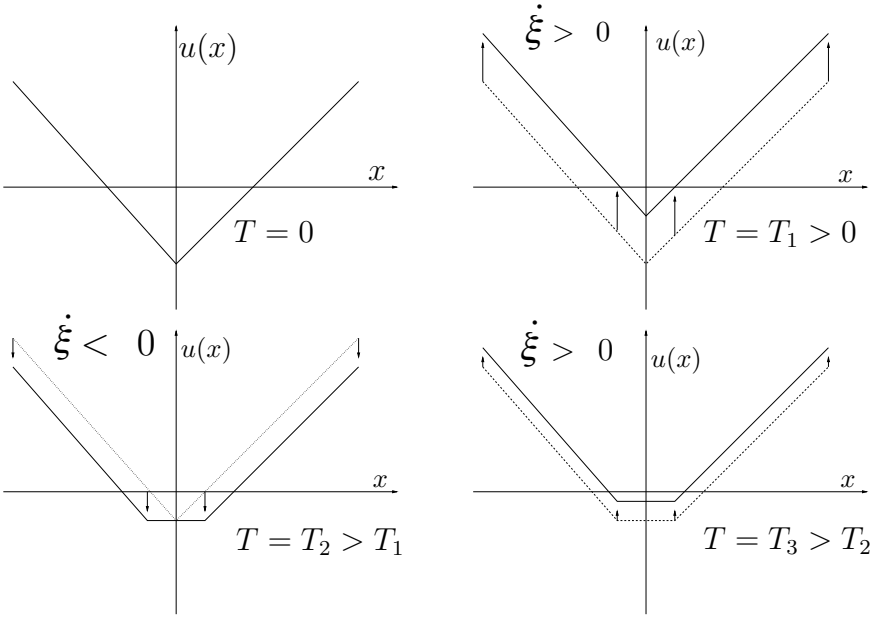

Figure 1. Examples of a typical evolution following the dynamics of Eq. (23). The extremely frequent changes of sign of the increments will alter the profile of the implicit function which does not remain a distance function. Hence, from an implementation point of view, some regular *reinitialization* of u is advisable.

simple evolution $du = |Du| \circ dW(t)$. According to the approximating scheme (11), the direct way of simulating such a process is through the following implicit scheme:

$$u_{i+1} = u_i + \frac{1}{2}(|Du_i| + |Du_{i+1}|)\Delta W_i$$

To avoid working with an implicit scheme, notice that the schema presented for the simulation of the Ito integral is an explicit one and use the fact that the Stratonovich integral is equal to the Ito integral plus an additional drift. Consider the evolution

$$du = H(Du, x) \circ dW(t) \tag{24}$$

where we have compacted the notation used previously. Here $H(p, x)$ is a function from $\mathbb{R}^N \times \mathbb{R}$ with real values. The typical example is $H(p, x) = |p|\phi(x)$, where, ϕ is some convenient regular function which is smooth enough. Such an evolution is equivalent, according to the definition of the Stratonovich integral, with the Ito evolution given by

$$du = H(Du, x)dW(t) + \frac{1}{2}d\langle H(Du, x),\ W\rangle(t) \tag{25}$$

To compute the drift, we start by rewriting the above dynamics in an integral form

$$u(t, x) = u_0(x) + \int_0^t H(Du(s, x),\ x) \circ dW(s)$$

We can then take the derivative with respect to the spatial parameter x and obtain

$$\begin{aligned} Du(t, x) = {} & Du_0(x) \\ & + \int_0^t \big[D^2u(s, x)D_pH(Du(s, x), x) \\ & + D_xH(Du(s, x), x)\big] \circ dW(s) \end{aligned}$$

where D_pH (resp. D_xH) denotes the gradients of H w.r.t. p (resp. x) and D^2u denotes the spatial Hessian. Then, applying the Ito rule, we have

$$\begin{aligned} H(Du(t, x), x) = {} & H(Du_0(x), x) \\ & + \int_0^t [D_pH \cdot (D^2uD_pH) \\ & + D_pH \cdot D_xH] \circ dW(s) \end{aligned}$$

Finally, if we consider the simplifying notation $A[u] = A(u, u)$, when A is some quadratic form, then the drift from Eq. (25) can be written as

$$\begin{aligned} & \frac{1}{2}\langle H(Du, x),\ W\rangle(t) \\ & \quad = \frac{1}{2}\int_0^t (D^2u(s, x)[D_pH(Du(s, x), x)] \\ & \qquad + D_pH(Du(s, x), x) \cdot D_xH(Du(s, x), x))\, ds \end{aligned}$$

When $H = |p|\phi(x)$, the previous formula becomes

$$\begin{aligned} & \langle H(Du, x),\ W\rangle(t) \\ & \quad = \int_0^t \left[\phi^2(x)D^2u(s, x)\left[\frac{Du(s, x)}{|Du(s, x)|}\right]\right. \\ & \qquad \left. + \phi(x)\, D\phi(x) \cdot Du(s, x)\right] ds \end{aligned}$$

We can remark that the second order term in the above formula is a smoothing term. It can also be written

$$\begin{aligned} D^2u\left[\frac{Du}{|Du|}\right] &= \Delta u - |Du|\mathrm{div}\left(\frac{Du}{|Du|}\right) \\ &= \Delta u - |Du|\kappa \end{aligned}$$

where κ denotes the mean curvature of the level set at point x. One can be alarmed by the presence of $-|Du|\kappa$. Nevertheless, the overall term is positive, since D^2u is a positive semi-definite matrix.

The above calculation remains valid if the dynamics depends on more than one Brownian motion. In conclusion, to simulate an evolution of the type

$$du = F|Du|\, dt + |Du| \sum_{i=1}^m \phi_i(x) \circ dW_i(t) \tag{26}$$

we use

$$\begin{aligned} du = {} & F|Du|\, dt + |Du| \sum_{i=1}^m \phi_i(x) dW_i(t) \\ & + \frac{1}{2}\left(\left(\sum_{i=1}^m \phi_i^2(x)\right) D^2u\left[\frac{Du}{|Du|}\right]\right. \\ & \left. + \left(\sum_{i=1}^m \phi_i(x)D\phi_i(x)\right) \cdot Du\right)\mathrm{dt} \end{aligned} \tag{27}$$

or, in the general case when the stochastic Hamiltonian is given by $H(p, x)$:

$$\mathrm{d}u = F|Du|\,\mathrm{d}t + H(Du, x)\mathrm{d}W(t) + \frac{1}{2}(D^2u[D_pH] + D_pH \cdot D_xH)\,\mathrm{d}t$$

3. Validation

In this section, we test our scheme and investigate some simple geometrical properties the evolution that could guide the user toward a correct choice of noise.

3.1. One Gaussian Noise

Let us begin with the simple case of a Gaussian noise constant in space. We thus consider $du = |Du| \circ dW(t)$ and implement:

$$du = |Du|dW(t) + \frac{1}{2}D^2u(t,x)\left[\frac{Du(t,x)}{|Du(t,x)|}\right]$$

We use a standard WENO3 scheme (Jiang and Peng, 2000) in space with step Δx and a first order explicit scheme in time with step Δt and verify the convergence of the approximation when the space step and/or the time step tend to zero. Again, please recall the use of $\sqrt{\Delta t}$:

$$u(t+\Delta t, x) = u(t,x) + |Du(t,x)|\sqrt{\Delta t}\mathcal{N}_{(0,1)}(t) + \frac{1}{2}D^2u(t,x)\left[\frac{Du(t,x)}{|Du(t,x)|}\right]$$

Because of the stochastic character of the evolution, one can only compare the different approximations through some statistical quantity[4]. For a given initial condition and a given final time T, *the variance of the area of the interior of the final curve* provides a simple and meaningful way to compare two approximations. The left part of Fig. 2 shows, for different values of Δx, the convergence of this variance when $1/\Delta t$ increases. As a reminder to avoid a naive mistake, we also implemented the evolution with Δt instead of $\sqrt{\Delta t}$ and verify that the variance of the area tends to zero!

As a test of the invariance of the Stratonovich differential, we compare, for a given initial curve $\Gamma(0)$, the mean of the area of the curve at a given final time T for different choices of the initial implicit function $u(0)$ (namely the signed distance function $d_{\Gamma(0)}$ to $\Gamma(0)$ and $\alpha(d_{\Gamma(0)})$ with $\alpha(x) = e^x - 1$). The right part of Fig. 2 shows, for different values of T and different initial curves, the relative difference between the means of the final area for the initial conditions $d_{\Gamma(0)}$ and for $\alpha(d_{\Gamma(0)})$ in both the Ito and the Stratonovich cases . Note how the Stratonovich scheme is much more insensitive with respect to the choice of $u(0)$.

It could be proved (Juan et al., 2004) that, for a given initial curve, the variance of the area of the curve at time T is a polynomial of degree N in T where N is the space dimension. In practice, for reasonable values of T, the relation between T and this variance is linear. As a final test, the left part of Fig. 3 shows, for $N = 1$ and $N = 2$, how this relation is respected by our scheme.

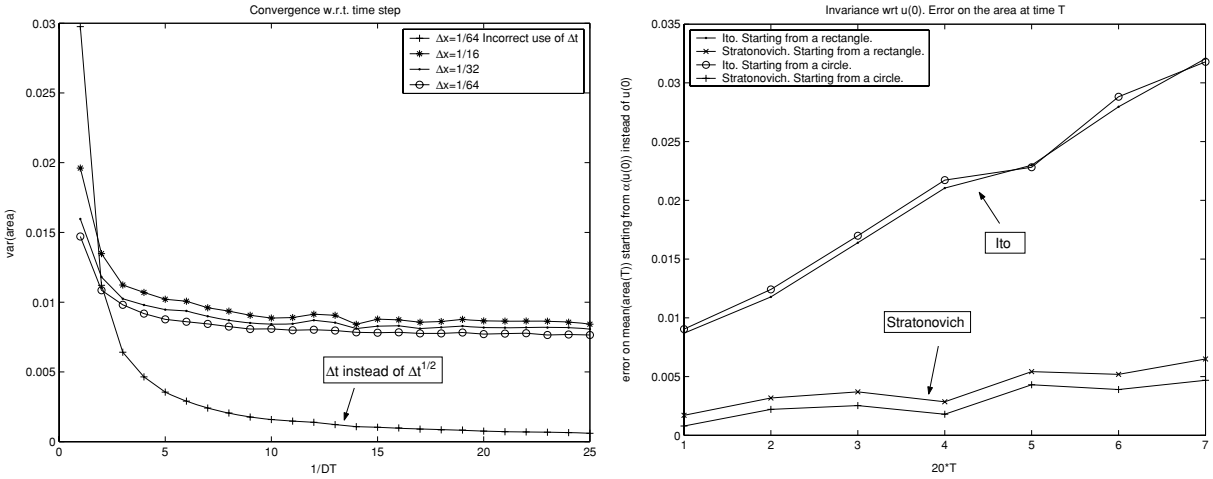

Figure 2. One Gaussian source. Left: convergence of the variance of the area at a given time T when Δt tends to 0 (plus the erroneous case when using (3)). Right: invariance of Stratonovich w.r.t. the choice of the initial implicit function.

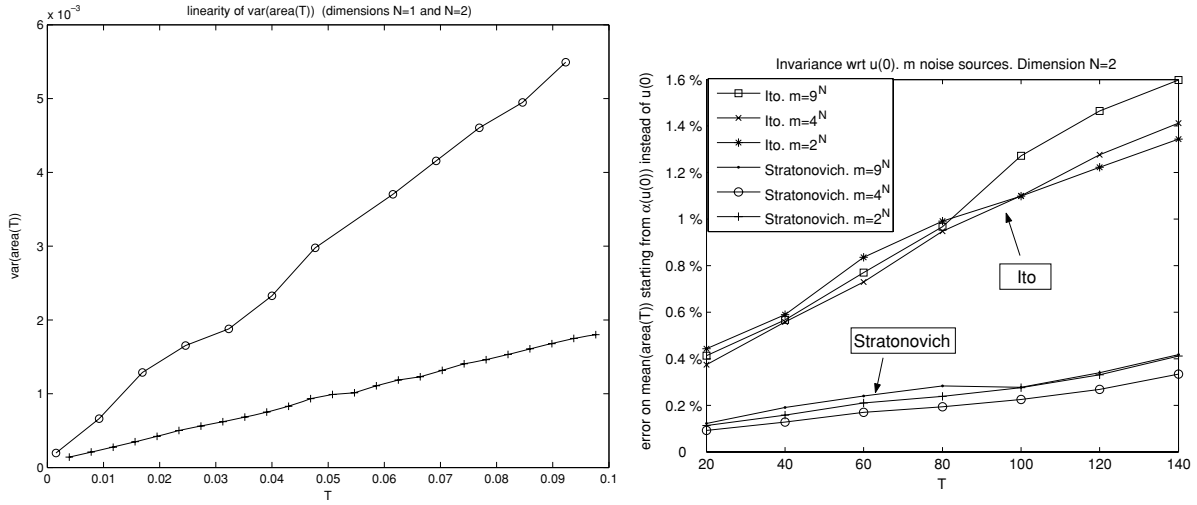

Figure 3. Left: Linear dependency between the final time and the variance of the area (one Gaussian). Right: invariance of Stratonovich wrt to the choice of the implicit function (several Gaussian sources).

3.2. Several Gaussian Noise Sources

Having the whole curve shrink or grow at the same time is not very useful. We will use a spatially dependent noise although the viscosity solution result is still an open question in this case. For a given number m of random sources, we implement the evolution (26) with $F = 0$ using the scheme (27). The m sources are equally distributed on a grid $\{x_i\}$ and $\phi_i(x) = \phi(x - x_i)$ where ϕ is such that $\phi_i(x_j) = \delta_{ij}$ and ϕ_i decreases smoothly from x_i to its neighbors. In practice, although not derivable in x_i, the classical multi-linear interpolation functions are sufficient. Note also that $\sum_{i=1}^{m} \phi_i(x) dW_i(t)$ is no longer of variance 1 for all x, so that the stochastic motion would be weaker between two sources. Using $\phi_i(x)/(\sum_{j=1}^{m} \phi_j^2(x))^{1/2}$ instead of $\phi_i(x)$ recovers a constant variance 1.

The drift will have a spatial derivative term (see (27)). Like Fig. 2 for one noise, the right part of Fig. 3 shows, for different values of m, how the Stratonovich scheme makes the evolution invariant with respect to the choice of $u(0)$.

With more than one source of noise, the points of the curve do not move at the same speed anymore, leading to the desired stochastic global deformation. As one should expect, with a large number of sources, the deformation is very noisy but the contributions of the sources tend to annihilate one each other. Thus, the curve does not move very far from its initial position. On the contrary, with a medium number of sources, the deformation is smoother but with ampler motions (see Fig. 4). Depending on his/her own application, the user might want to choose the optimal number of sources. As a first attempt to quantify the phenomenon, we measure how long it takes to the curve to move away from its initial position. For a given distance δ, we call the *expected exit time* the quantity $T(\delta) = \mathbb{E}(\inf\{t : \exists x \in \Gamma(t), d(x, \Gamma(0)) \geq \delta\})$ where $\mathbb{E}$ denotes the expectation. For a Brownian motion, the expected exit time from a ball is a quadratic function of the radius of the ball. In our case, such a result would be certainly hard to prove. Yet, our experiments show a similar relationship $T(\delta) \approx \xi(m)\delta^2$: (see the left part of Fig. 5). This useful relationship indicates clearly how long the user has to wait to see his/her curve getting away from its initial position. The right part of Fig. 5 plots ξ as a function of the number of sources. As expected, a large number of sources m induces a larger exit time, thus a larger ξ. Surprisingly, the smallest values of m give also a large ξ. We do not have any satisfactory explanation for this phenomenon... Anyway, these are only some very first steps toward the understanding of the geometric properties of this kind of stochastic motion and many other quantities would be of great interest: the variations of the curvature, the time to get the curve split, etc.

4. Applications to Computer Vision

Many Computer Vision problems consist in recovering a certain surface or region through a shape optimization

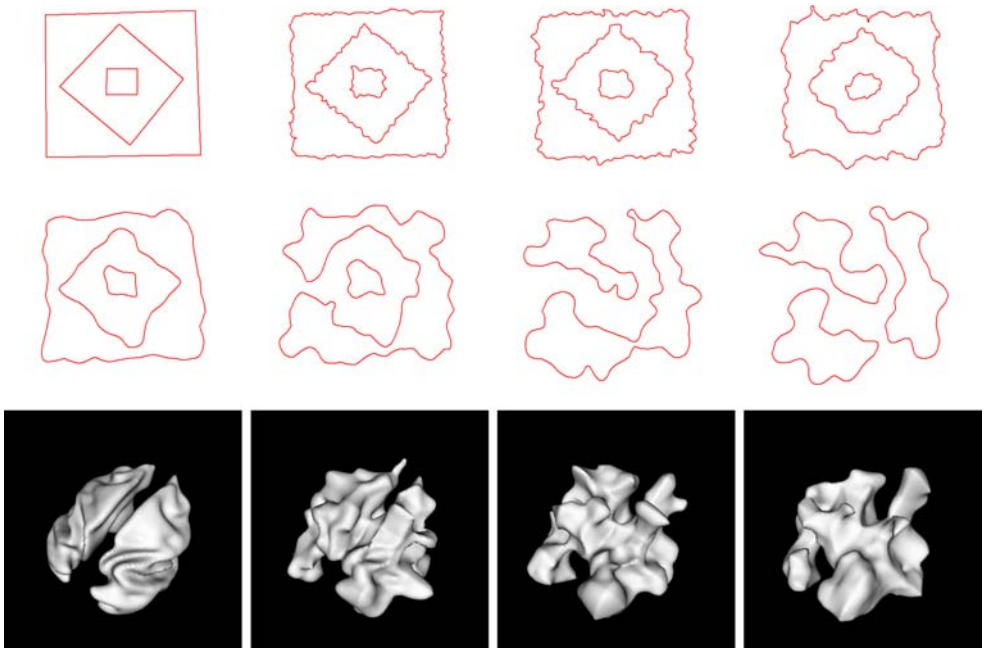

Figure 4. Different number of Gaussian noise sources. Top row: starting from the initial curve (top left), three time steps of the evolution with a large number of Gaussian sources. Middle row: from the same initial curve, four time steps of the evolution with a spatially smoother noise (small number of sources). Bottom row: a 3D example starting from the cortex of a monkey.

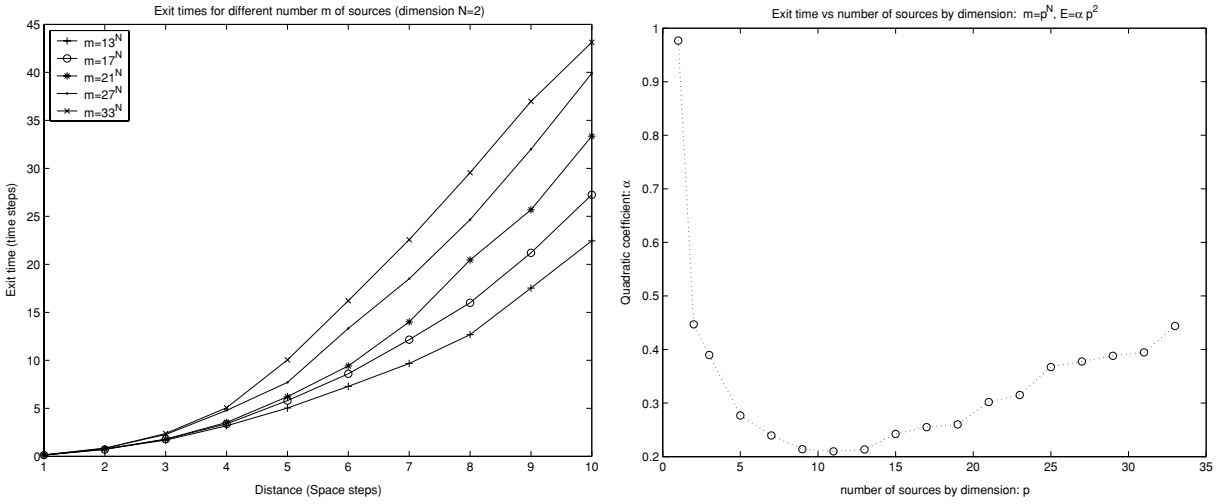

Figure 5. Several Gaussian sources. Left: Quadratic relation between the distance δ and the expected exit time from the band of thickness δ. Right: variation of the exit time w.r.t. the number of sources.

framework (Caselles et al., 1997; Faugeras and Keriven, 1998; Vese and Chan, 2002; Paragios and Deriche, 2002). These methods suffer from being sometimes stuck in local minima. The dynamics presented earlier, coupled with a decision mechanism, can be used to overcome this problem. As a first step toward more sophisticated genetic methods, we turned our attention to the Simulated Annealing algorithm.

Based on the work of Metropolis et al. (1953), Simulated Annealing was first mentioned by Kirkpatrick in Kirkpatrick et al. (1983) as a nice application of statistical physics to optimization problems. Its purpose is to introduce a probabilistic decision mechanism for finding global minima in higher dimension. First, we would like to briefly comment upon the previous works oriented toward the use of genetic programming in Computer Vision.

4.1. Comparison with Previous Work in Computer Vision

In a lot of cases, the stochastic theory is used to help researchers develop an intuition of the macroscopic dynamics at a microscopic level. This is the case in Ben-Arous et al. (2002), Unal et al. (2003), where an algorithm for stochastic approximations to a curve shortening flow is built. Another example is given by Unal et al. (2002), where the authors develop a model of anisotropic diffusion using the information gained by analyzing the stochastic differential equation associated to a linearized version of the geometric heat equation. Note that one should not confuse these stochastic simulations of a deterministic motion with our use of stochastic motion. In other cases, stochastics are actively used in selection algorithms meant to overcome some classical dynamics difficulties. In Storvik (1994) Storvik used Simulated Annealing combined with a Bayesian dynamics and developed applications in medical imagery. He used a node-oriented representation technique for the contour representation. Thus, his algorithm can only detect simply connected domains in an image. Moreover, self-intersections are not allowed, due to the complications they would involve. More recently, Ballerini et al developed in Ballerini (1999) an interesting application to medical image segmentation using a genetic algorithm, *genetic snakes*. They used a model that they fit using a number of control points. Their application cannot, therefore, be extended to a more general framework.

Please note that the main ingredient of our work is not the Simulated Annealing part, but rather the implementation of the stochastic motion and its use in shape optimization problems. It is obvious that the stochastic approach adds to the power and flexibility of the Level Sets technique into a very powerful tool. We can thus use this mechanism through skillfully applied controls, while continuing to allow for topological changes and weak regularity assumptions. Simulated Annealing is used in our experiments. In the future, more evolved genetic programming selection techniques might be considered, but it is encouraging that such simple ingredients added to the Level Sets framework provide good practical results.

4.2. Principle

Given some Computer Vision problem in a variational framework where we have to find the region Γ that minimizes an energy $E(\Gamma) = E(u)$, we use the following simple Simulated Annealing decision scheme:

1. Start from some initial guess u_0
2. Compute u_{n+1} from u_n using some dynamics, e.g. $\mathrm{d}u = |Du| \sum_{i=1}^{m} \phi_i(x) \circ \mathrm{d}W_i(t)$
3. Compute the energy $E(u_{n+1})$
4. Accept u_{n+1}:
   - if $E(u_{n+1}) < E(u_n)$
   - otherwise, accept u_{n+1} with probability $\exp\left(-\frac{E(u_{n+1})-E(u_n)}{T(n)}\right)$
5. Loop back to step 2, until some stopping condition is fulfilled

Here, $T(n)$ is a time-dependent function that plays the same role as a decreasing temperature. Its choice is not obvious. If the temperature decreases too fast, the process may get stuck in a local minimum; on the contrary, decreasing too slowly, it may postpone convergence. A classical choice is $T(n) = T_0/\sqrt{n}$. The classical way to solve the previous minimization problem is to use a gradient descent method. The Euler-Lagrange equation is computed, leading to some evolution $\partial\Gamma/\partial t = \beta_c \mathbf{n}$, or equivalently, in the Level Set framework, to $\partial u/\partial t = \beta_c |Du|$. We will actually use the classical motion as heuristics that drive the evolution faster toward a minimum, and replace the dynamics of step 2, by

$$du = \beta_c |Du| dt + |Du| \sum_{i=1}^{m} \phi_i(x) \circ dW_i(t) \quad (28)$$

As often with genetic algorithms, the proof of the convergence of this algorithm toward a global minimum is still an open problem. However, we use numerical experiments to show how this algorithm performs better by avoiding some local minima that are problematic in the deterministic case. This is our main motivation, since local minima are the major problem of

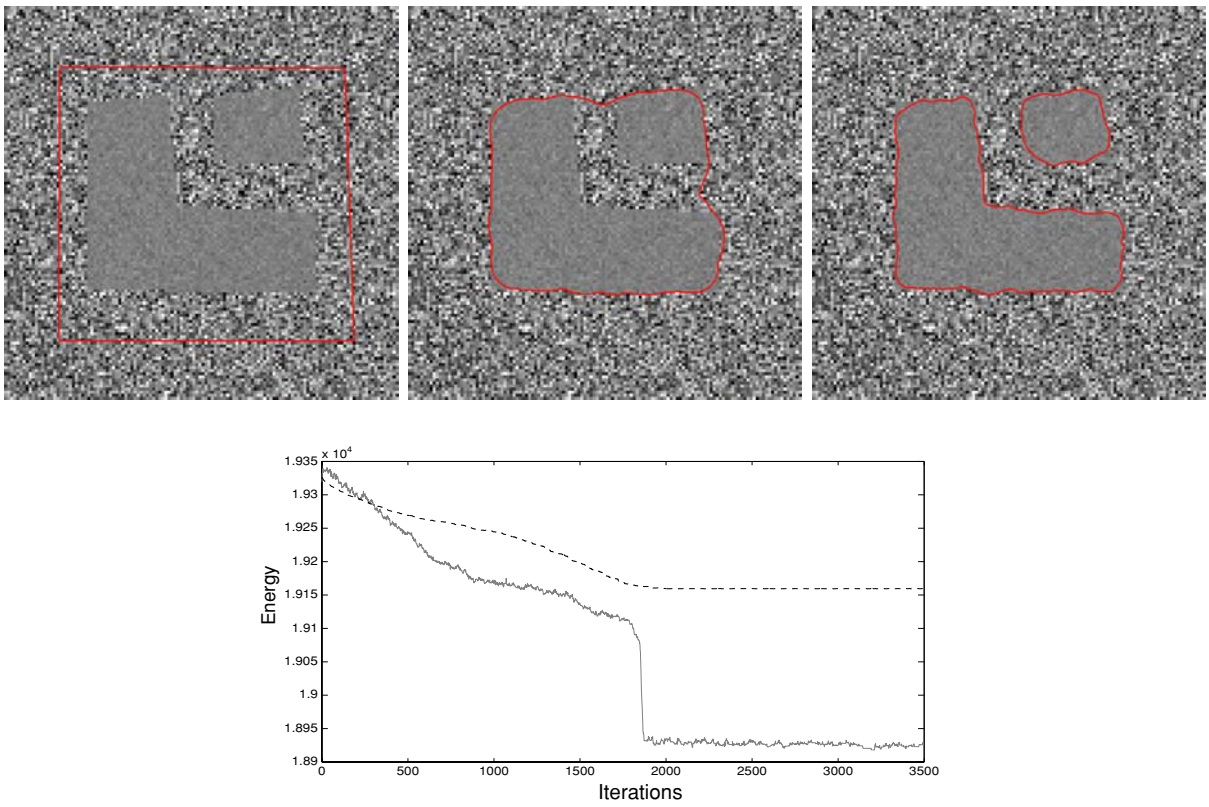

Figure 6. Segmentation of two regions modeled by two unknown Gaussian distributions (same mean, different variances). Top row: the initial curve, the final state of the classical approach stuck in a local minimum, and the final state of our method. Bottom row: evolution of the energy (dashed: deterministic method, solid: our method). The two results were obtained with identical common parameters; moreover, the result is not too sensitive wrt. the choice of the stochastic parameters.

classical approaches. Note also, as already mentioned, that our framework can be used in cases when the shape gradient is too complex from a mathematical or computational point of view, or even impossible to compute.

5. Stochastic Active Contours

Our scheme could be used in the Geodesic Active Contours framework (Caselles et al., 1997) where segmentation is based upon gradient intensity variations. Yet, a multiscale approach is often used successfully in that context to overcome the local minimum problem. Other segmentation schemes (Paragios and Deriche, 2002) use a region model (eg. texture, statistics) that is less adapted to multiscale. We will first focus on one such case, namely the single Gaussian statistics model in Rousson and Deriche (2002).

5.1. Single Gaussian Model

In their unsupervised segmentation framework (Rousson and Deriche, 2002), the authors model each region of a gray-valued or color image I by a single Gaussian distribution of unknown mean μ_i and variance Σ_i. The case of two regions segmentation turns into minimizing the following energy:

$$E(\Gamma, \mu_1, \Sigma_1, \mu_2, \Sigma_2) = \int_{\Omega_1} e_1(x) + \int_{\Omega_2} e_2(x) + \nu \text{length}(\Gamma)$$

where Ω_1 is the region inside Γ, Γ_2 the outside, and $e_i(x) = -\log p_{\mu_i \Sigma_i}(I(x))$ with $p_{\mu_i \Sigma_i}(I(x)) = C|\Sigma_i|^{-1/2} e^{-(I(x)-\mu_i)^T \Sigma_i^{-1}(I(x)-\mu_i)/2}$ being the conditional probability density function of a given value $I(x)$ with respect to the hypothesis (μ_i, Σ_i). The

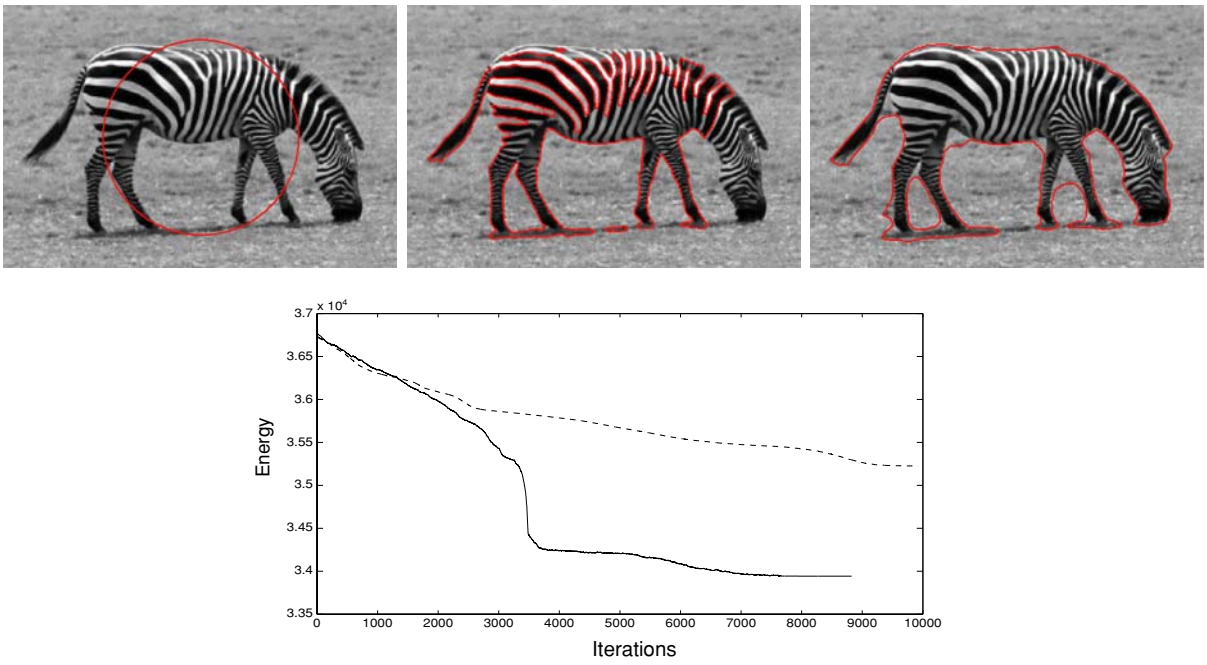

Figure 7. Segmentation of two regions modeled by two unknown Gaussian distributions. Top row: the initial curve, the final time step of the classical method, again stuck in a local minimum and the final step of our method. Bottom row: evolution of the energy (dashed: deterministic method, solid: our method).

parameters (μ_i, Σ_i), estimated from the pixel actually inside and outside Γ, are functions of Γ. Thus, the energy is a function of Γ only: $E(\Gamma, \mu_1, \Sigma_1, \mu_2, \Sigma_2) = E(\Gamma)$. Its Euler-Lagrange equation is not obvious, but finally simplifies into the minimization dynamics

$$\beta_c = e_2(x) - e_1(x) + \nu \mathrm{div}\left(\frac{\mathrm{D}u}{|\mathrm{D}u|}\right)$$

The authors successfully segment two regions, even when they have the same mean but only different variances. However, the evolution could easily be stuck into some local minimum while a multiscale approach might modify the statistics so that no segmentation would be possible anymore. As demonstrated Fig. 6, a simple Simulated Annealing scheme with dynamics (28) overcomes this problem. Figure 7 shows the same phenomenon on a real image. Note that this image was successfully segmented by the authors of Paragios and Deriche (2002). Yet, they used an adapted model of texture. Here, the Stochastic Active Contours framework succeeds in making a simple unsupervised single Gaussian model recover the correct regions.

5.2. *Gaussian Mixtures*

As an illustration of the case when the Euler-Lagrange equation cannot be computed, we extend the previous method to region statistics modeled by a mixture of Gaussian distributions of parameters $\Theta_i = (\pi_i^1, \mu_i^1, \Sigma_i^1, \ldots, \pi_i^{n_i}, \mu_i^{n_i}, \Sigma_i^{n_i})$. with $\sum_j \pi_i^j = 1$. The conditional probability density function of a given value $I(x)$ becomes:

$$p_{\Theta_i}(I(x)) = \sum_{j=1}^{n_i} \pi_j p_{\mu_i^j \Sigma_i^j}(I(x))$$

The number of Gaussian distributions can be given, estimated at the initial time step, or dynamically evaluated using a Minimum Description Length criterion (Rissanen, 1989) or the Minimal Message Length method (Wallace and Boulton, 1968). A large literature is dedicated to the problem of estimating Θ_i from input samples. We have used the original *K*-Means algorithm pioneered by MacQueen (1967), although we have tested extensions like the Fuzzy-K-Means (Bezdek, 1981; deGruijter and McBratney, 1988), the K-Harmonic-Means (Zhang and Hsu, 1999), and the Expectation-

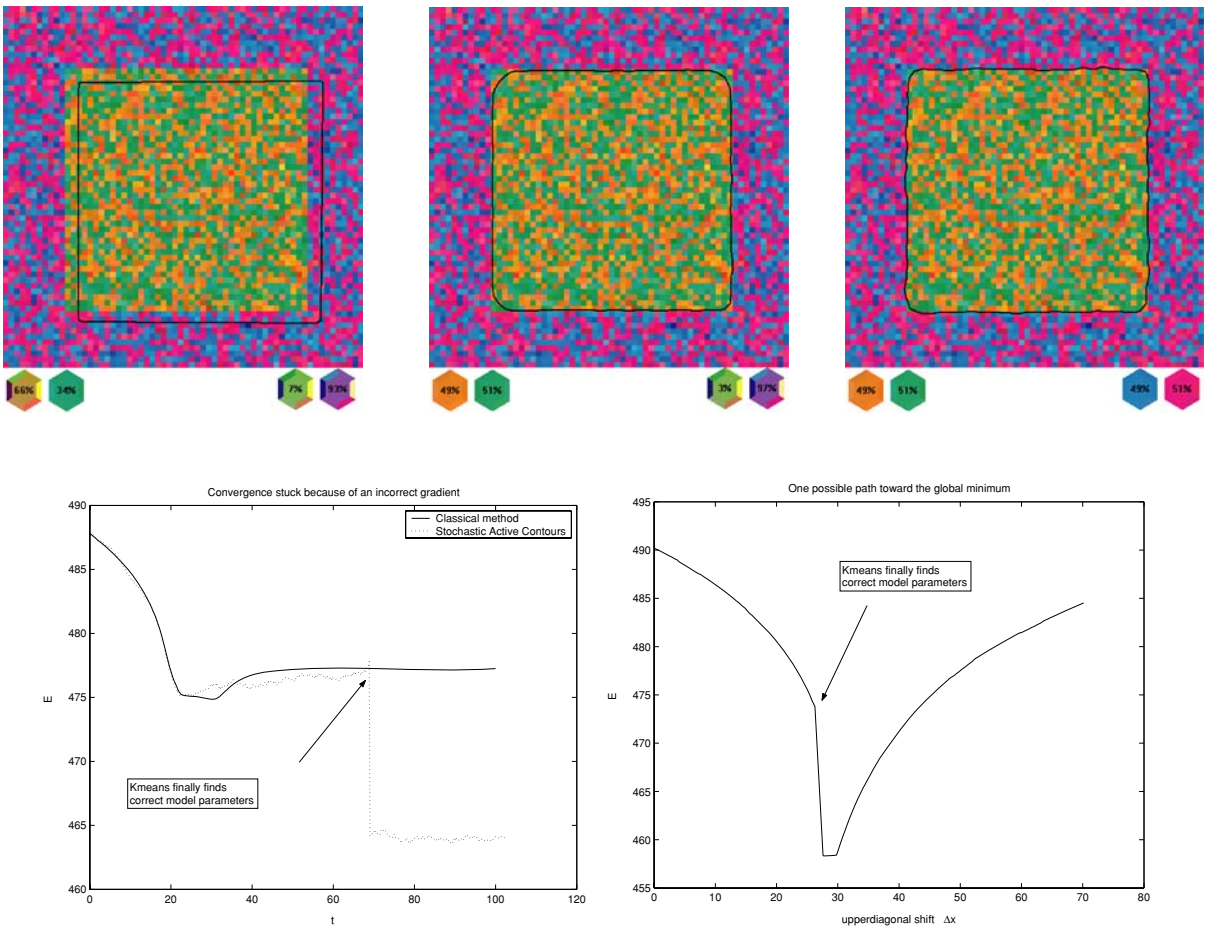

Figure 8. A case where the gradient is not correct (see text). Top row, from left to right: initial position, final position with the classical method (the model is not correctly recovered - see percentages in the hexagons), leading to rounded corners), final position with our method (the model is correctly recovered). Bottom left: evolution of the energy in both cases. Bottom right: energy for a translation of the curve that goes through the correct segmentation.

Maximization algorithm (EM), first proposed in Dempster et al. (1977). The latter solves iteratively $\widehat{\Theta}_i = \text{argmax}_{\Theta_i} \int_{x\in\Omega_i} \log p_{\Theta_i}(I(x))\, dx$ (Please refer to (McLachlan and Krishnan, 1997; McLachlan and David, 2000) for details and to (Ueda et al., 2000) for extensions).

Our segmentation problem still consists in minimizing the same energy, with now $e_i(x) = -\log p_{\Theta_i}(I(x))$. Unfortunately, we now have to deal with a complex dependency of Θ_i with respect to Γ. In fact, the learning algorithm acts as a "black box" implementing $\Gamma \rightarrow \Theta_i(\Gamma)$. As a consequence, the Euler-Lagrange equation of the energy $E(\Gamma, \Theta_1(\Gamma), \Theta_2(\Gamma)) = E(\Gamma)$ cannot be computed. A deterministic contour evolution driven by $\beta_c = e_2 - e_1 + \nu\kappa$ may get stuck just because $\beta_c \mathbf{n}$ is not the exact gradient. Yet, the Stochastic Active Contours can still be used, with β_c as heuristics. As a simple illustration of this, let us consider the synthetic example of Fig. 8. The region to segment is a square. The square and the background are each modeled by a mixture of two equally weighted Gaussian distributions: $\Theta_i = (\frac{1}{2}, \mu_i^1, \Sigma, \frac{1}{2}, \mu_i^2, \Sigma)$. As an initial guess, we shift the square toward the bottom-right corner. Although a bit of the background is in Ω_1, Θ_1 is correctly estimated. Yet, for some reason, the K-Means algorithm estimates Θ_2 approximatively by $(1-\epsilon, \frac{\mu_2^1+\mu_2^2}{2}, \Sigma', \epsilon, \frac{\mu_1^1+\mu_1^2}{2}, \Sigma'')$. During its convergence, the deterministic method keeps such an incorrect Θ_2 and finally gets stuck at a roughly correct place but with an incorrect model, leaving some

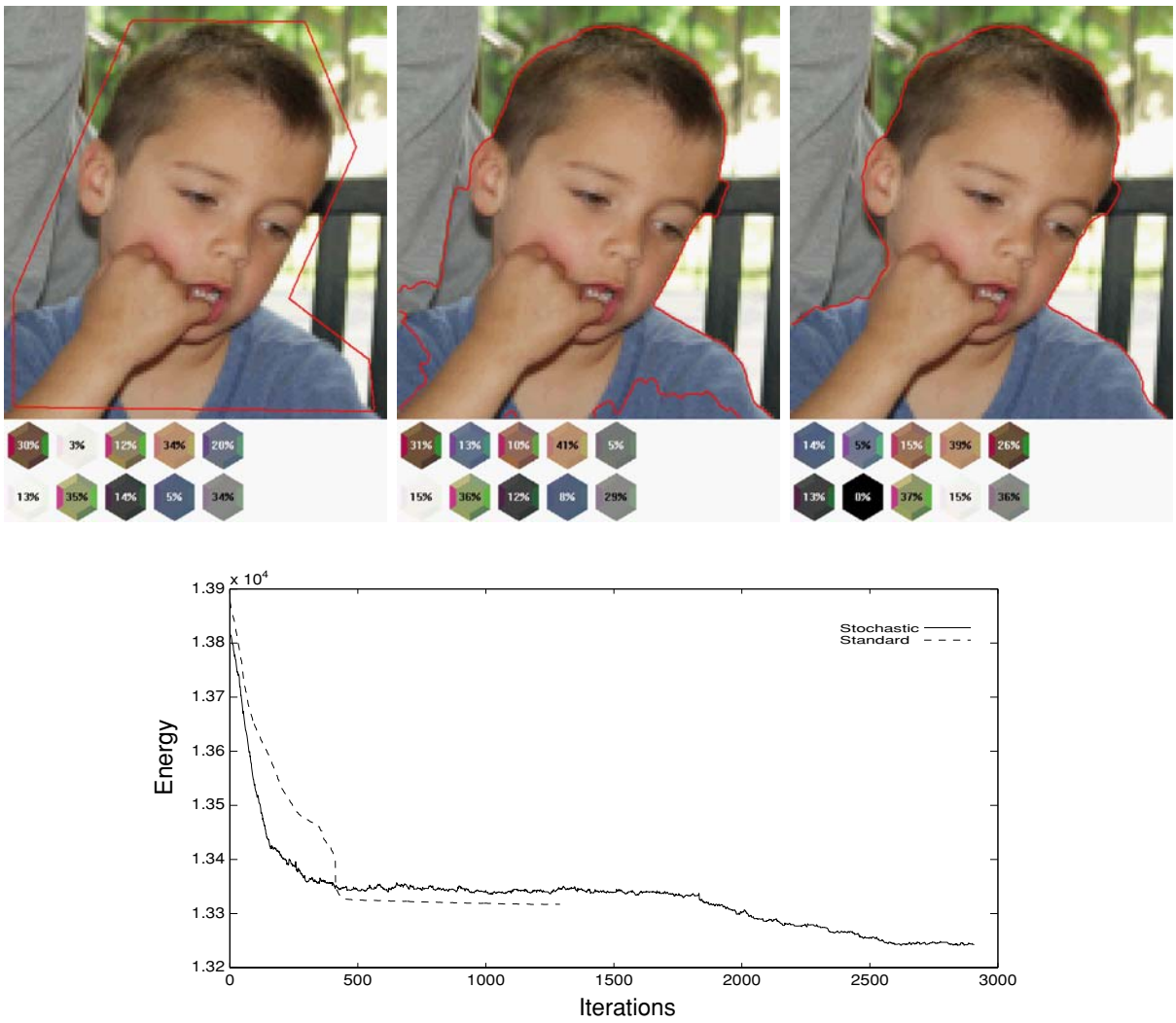

Figure 9. Segmentation of two regions modeled by two unknown Gaussian mixtures. Top row: the initial curve, the final state of the deterministic method, stuck in a local minimum and the final state of our method. Bottom row: evolution of the energy.

interior pixels outside (especially in the corners, because of the smoothing term of the energy). The colored hexagons below the images indicate the means and variances of the mixtures components and their respective weights. See also how the energy increases in the end! On the contrary, our method does not rely completely on the incorrect gradient only and finally *"discovers"* the correct model, leading to a somehow better fit. Notice the energy level drop-down when the K-Means algorithm ejects the interior pixels as negligible and shifts to the correct model. The last graph of Fig. 8 is a plot of the energy when the initial square is *manually* translated from the bottom-right corner to the upper-left one, going through the correct position. into some local minimum. It clearly shows that the heuristic gradient by itself gets stuck in a local minimum , whereas our method comes much closer to the desired minimum.

6. Results

Even when the deterministic scheme converge more or less, our method shows a better ability to overcome local minima: Fig. 9 shows how Γ can be stuck leading to a dramatic evolution toward completely false regions.

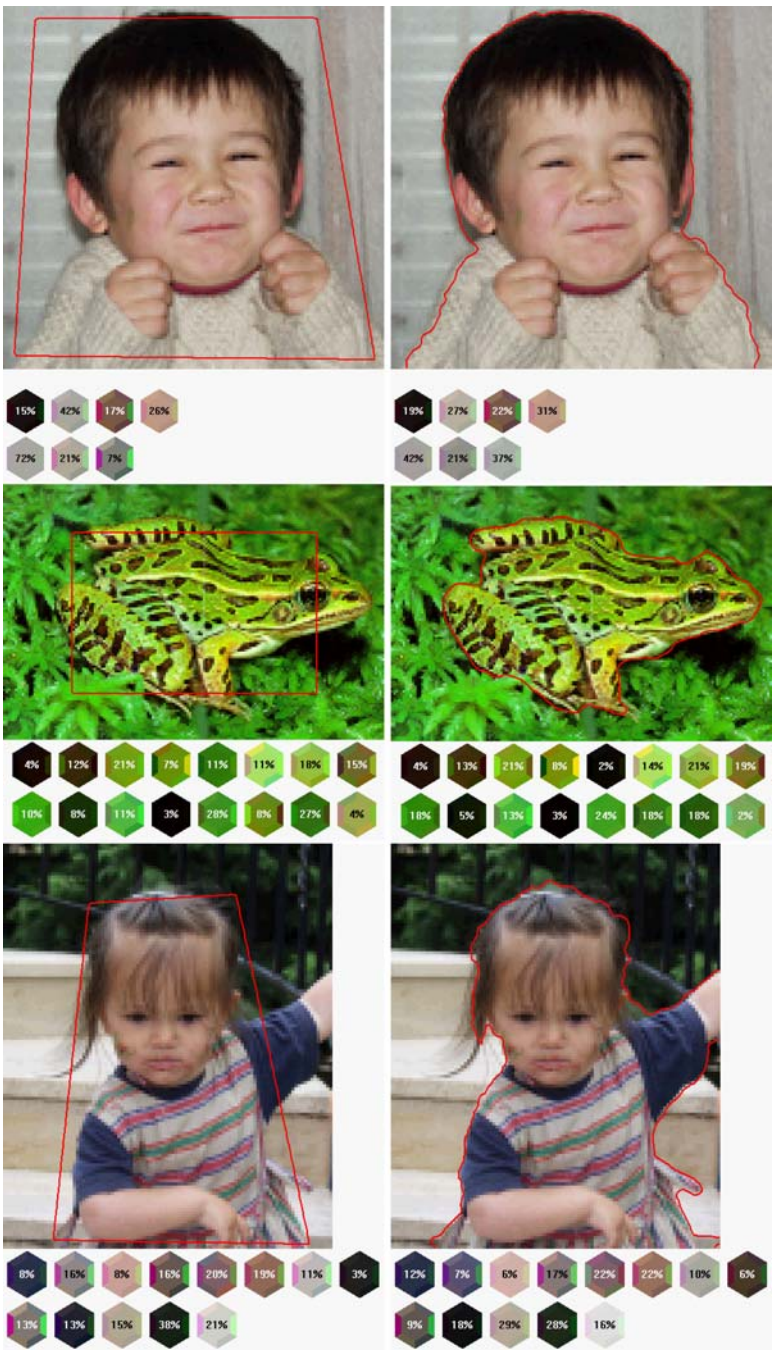

Figure 10. Segmentation of two regions modeled by two unknown Gaussian mixtures. Left column: the initial states. Right column: the corresponding final states of our method.

In all the cases, the common parameters were identical in the standard and stochastic cases. Moreover, the method is not too sensitive wrt. the choice of the stochastic parameters. Finally, Fig. 10 shows some more examples on other real images. Animations corresponding to all the presented examples can be downloaded at `http://cermics.enpc.fr/~juan/IJCV/`.

7. Conclusion

Based on recent work on Stochastic Partial Differential Equations by Lions and Souganidis, we have presented a simple and well-founded method to implement the stochastic motion of a surface in a Level Set framework. This method is used as the key point of a stochastic extension to standard shape optimization methods in Computer Vision. In the particular case of segmentation, we introduced the *Stochastic Active Contours*, a stochastic extension of the well-known active contours. Our method is more likely to overcome local minima and can also be used when the Euler-Lagrange equation of the energy is out of reach. This extension is not time consuming: the only computational effort is computing the energy. Convincing results are presented with the segmentation of regions modeled by unknown statistics, namely single Gaussian distributions or mixtures of Gaussian distributions. The way is now open for applying our principle to other Computer Vision problems but also in different fields where shape optimization problems arise, like in theoretical chemistry (Cances et al., 2004).

Notes

1. That is to say, depending on the curve itself but not on the choice of the parameterization p.
2. Typically, this would be some Ito stochastic integral $M(t) = \int_0^t \Phi dW$.
3. The space of **Bounded Uniformly Continuous** functions.
4. Actually, for a given time step Δt, we might fix the event ω and compare the approximations for different Δx but with the same Brownian. We also successfully used such a path-wise comparison when testing the invariance property of our scheme.

References

Ballerini, L. 1999. Genetic snakes for image segmentation. In *Evolutionary Image Analysis, Signal Processing and Telecommunication*, vol. 1596 of *Lecture Notes in Computer Science*, Springer, pp. 59–73.

Ben-Arous, G., Tannenbaum, A., and Zeitouni, O. 2002. Stochastic approximations to curve shortening flows via particle systems. Technical report, Technion Institute.

Bezdek, J.C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers.

Buckdahn, R. and Ma, J. 2001. Stochastic viscosity solutions for nonlinear stochastic partial differential equations. *Stochastic Processes and their Applications*, 93:181–228.

Cances, E., Keriven, R., Lodier, F., and Savin, A. 2004. How electrons guard the space: Shape optimization with proba bility distribution criteria. *Theoret. Chem. Acc.*

Caselles, V., Kimmel, R., and Sapiro, G. 1997. Geodesic active contours. *The International Journal of Computer Vision*, 22(1): 61–79.

deGruijter, J. and McBratney, A. 1988. A modified fuzzy k means for predictive classification. In Bock, H.H.(ed), *Classification and Related Methods of Data Analysis*, Elsevier Science, Amsterdam., pp. 97–104.

Dempster, A.P., Laird, N.M., and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39:1–38.

Faugeras, O. and Keriven, R. 1998. Variational principles, Surface Evolution, PDE's, level set methods and the Stereo Problem. *IEEE Transactions on Image Processing*, 7(3):336–344.

Gard, T. 1988. *Introduction to Stochastic Differential Equations*. Marcel Dekker, New York and Basel.

Jiang, G.S. and Peng, D. 2000. Weighted ENO schemes for hamilton-jacobi equations. *SIAM Journal of Scientific Computing*, 21(6):2126–2143.

Juan, O., Keriven, R., and Postelnicu, G. 2004. Stochastic motion and the level set method in computer vision: Stochastic active contours. Technical Report 04-01, CERTIS.

Karatzas, I. and Shreve, S. 1991. *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

Katsoulakis, M. and Kho, A. 2001. Stochastic curvature flows: Asymptotic derivation, level set formulation and numerical experiments. *Journal of Interfaces and Free Boundaries*, 3:265–290.

Kirkpatrick, S., Gelatt, C.J., and Vecchi, M. 1983. Optimization by Simulated Annealing. *Science*, 220(4598).

Kunita, H. 1990. *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press.

Lions, P. and Souganidis, P. 1998a. Fully nonlinear stochastic partial differential equations. *C.R. Acad. Sci. Paris Ser. I Math*, 326:1085–1092.

Lions, P. and Souganidis, P. 1998b. Fully nonlinear stochastic partial differential equations: Nonsmooth equations and applications. *C.R. Acad. Sci. Paris Ser. I Math*, 327:735–741.

Lions, P. and Souganidis, P. 1999. Equations aux derivees partielles stochastiques non lineaires et solutions de viscosite. Seminaire X-EDP (Unpublished Lecture).

Lions, P. and Souganidis, P. 2000a. Fully nonlinear stochastic partial differential equations with semilinear stochastic dependence. *C.R. Acad. Sci. Paris Ser. I Math*, 331:617–624.

Lions, P. and Souganidis, P. 2000b. Uniqueness of weak solutions of fully nonlinear stochastic partial differential equations. *C.R. Acad. Sci. Paris Ser. I Math*, 331:783–790.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.

McLachlan, G. and David, P. 2000. *Finite Mixture Models*, Wiley, New York.

McLachlan, G. and Krishnan, T. 1997. *The EM Algorithm and Extensions*, Wiley, New York.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and T.E. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092.

Osher, S. and Sethian, J. 1988. Fronts propagating with curvature dependent speed: Algorithms based on the hamilton–jacobi formulation. *Journal of Computational Physics*, 79:12–49.

Paragios, N. and Deriche, R. 2002. Geodesic active regions and level set methods for supervised texture segmentation. *The International Journal of Computer Vision*, 46(3):223.

Rissanen, J. 1989. *Stochastic Complexity in Statistical Inquiry*, World Scientific Publishing Company.

Rousson, M. and Deriche, R. 2002. A variational framework for active and adaptative segmentation of vector valued images. In *Proc. IEEE Workshop on Motion and Video Computing*, Orlando, Florida, pp. 56–62.

Storvik, G. 1994. A bayesian approach to dynamic contours through stochastic sampling and simulated annealing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(10):976–986.

Ueda, N., Nakano, R., Ghahramani, Z., and Hinton, G.E. 2000. SMEM algorithm for mixture models. *Neural Computation*, 12(9):2109–2128.

Unal, G., Krim, H., and Yezzi, A. 2002. Stochastic differential equations and geometric flows. *IEEE Transactions in Image Processing*, 11(12):1405–1416.

Unal, G., Nain, D., Ben-Arous, G., Shimkin, N., Tannenbaum, A., and Zeitouni, O. 2003. Algorithms for stochastic approximations of curvature flows. In *IEEE International Conference on Image Processing*.

Vese, L. and Chan, T. 2002. A multiphase level set framework for image segmentation using the mumford and shah model. *International Journal of Computer Vision*, 50(3):271–293.

Wallace, C. and Boulton, D. 1968. An information measure for classification. *Computer Journal*, 11(2):185–194.

Walsh, J. 1994. An introduction to stochastic partial differential equations. In *Ecole d'Été de Probabilités de Saint-Flour*, vol. XIV-1180 of *Lecture Notes in Math.* Springer.

Yip, N. 1998. Stochastic motion by mean curvature. *Arch. Rational Mech. Anal.*, 144:331–355.

Yip, N. 2002. Stochastic curvature driven flows. In G. Da Prato and L. Tubaro (eds.), *Stochastic Partial Differential Equations and Applications*, vol. 227 of *Lecture Notes in Pure and Applied Mathematics*. Springer, pp. 443–460.

Yip, N. and Souganidis, P. 2004. Uniqueness of motion by mean curvature perturbed by stochastic noise. *Annales de l'Institut Henri Poincare - analyse non lineaire*, (21):1–23.

Zhang, B. and Hsu, M. 1999. K-Harmonic Means—A Data Clustering Algorithm.