Exploiting information geometry to improve the convergence of nonparametric active contours

Marcelo Pereyra, Hadj Batatia and Steve McLaughlin

***Abstract*—In this paper we seek to exploit information geometry in order to define the Riemannian metric of the manifold associated with nonparametric active contour models from the exponential family. This Riemannian metric is obtained through a relationship between the contour's energy functional and the likelihood of the categorical latent variables of a mixture model. Accordingly contours form a statistical manifold equipped with a natural metric which is determined by the model's Fisher information matrix. Mathematical developments show that this matrix has a closed-form analytic expression and is diagonal. Based on this, we subsequently develop a Riemannian steepest descent algorithm for the active contour, with application to image segmentation. Because the proposed method performs optimisation on the parameter's natural manifold it attains dramatically faster convergence rates than the Euclidean gradient descent algorithm commonly used in the literature. A segmentation experiment on an ultrasound image is presented and confirms that the proposed natural gradient algorithm converges extremely fast and delivers accurate segmentation results in few iterations.**

***Index Terms*—active contours, level sets, variational methods on Riemannian manifolds, information geometry.**

I. Introduction

Active contour (AC) models are a powerful framework for estimating the boundaries of an object within a given image. In this framework, contours are represented as curves that evolve subject to certain constraints to minimize an energy functional. This paper considers nonparametric ACs, a particularly useful class of segmentation methods where curves are represented implicitly as the zero level set of a surface that evolves with a fictitious time t [1]. In particular, we focus on region-based ACs that evolve according to the statistical characteristics of the object of interest and the background, as opposed to evolving according to image gradients or edges.

Region-based nonparametric ACs where first postulated in the seminal work of Chan and Vese [2], which defined an active contour for images composed by a foreground and background with Gaussian statistics. That work was subsequently generalized to images with other specific statistics, such as Rayleigh [3], gamma [4], Weibull [5] and Laplace [6]. A unified framework for AC models for distribution from the exponential family was finally proposed in [7].

Inherent in active contour segmentation problem is the solution of the Euler-Lagrange differential equations that guide the contour's evolution. In most applications these are solved using standard first-order Euler methods which are relatively simple to derive and implement. However, it is well known that Euler's method can take a large number of iterations to converge. This drawback has recently motivated numerous papers in the literature that study more advanced algorithms to solve active contour (a survey of the state-of-the-art has been presented in [8]). However, these fast algorithms have been designed for the Chan-Vese active contour and cannot be directly applied to other contours from the exponential family.

This paper presents a general Riemannian optimisation method for nonparametric ACs from the exponential family and is structured as follows. In the next section we introduce the underlying mathematical concepts in region-based active contour segmentation and briefly highlight some of the difficulties that our approach seeks to address. In Section III we develop an information geometry framework for active contours from the exponential family and propose a Riemannian steepest descent method to compute the contour's evolution. More precisely, we derive a smooth natural gradient descent algorithm [9] for nonparametric active contour, with application to image segmentation. Following on from this in Section IV we present some results illustrating the power of the approach and contrast the performance on alternative existing approaches. Finally we draw some conclusions and discuss potential extensions of the approach.

II. Region-Based Level Set Segmentation

Let Ω be a bounded subset of $\mathbb{R}^D$ and $I : \Omega \to \mathbb{R}^p$ a D-dimensional image composed by p channels (i.e., pixels take their values in $\mathbb{R}^p$). Moreover, I is assumed to be constituted by a foreground Ω_F and a background Ω_B, each characterised its own statistical distribution. Precisely, it is assumed that image pixels are distributed according to the following statistical model

$$\begin{array}{ll} I(\boldsymbol{x}) \sim f(\boldsymbol{\theta}_F) & \text{if } \boldsymbol{x} \in \Omega_F \\ I(\boldsymbol{x}) \sim f(\boldsymbol{\theta}_B) & \text{if } \boldsymbol{x} \in \Omega_B \end{array} \tag{1}$$

where $\boldsymbol{\theta}_F$ and $\boldsymbol{\theta}_B$ are the statistical parameters associated with the foreground and background respectively and where f is an arbitrary distribution from the natural exponential family, i.e,

$$f(I(\boldsymbol{x}); \boldsymbol{\theta}) = h(I(\boldsymbol{x})) \exp\left[\boldsymbol{\theta}^T S(I(\boldsymbol{x})) - A(\boldsymbol{\theta})\right] \tag{2}$$

with sufficient statistic $S(I(\boldsymbol{x}))$ and log-normalizer $A(\boldsymbol{\theta})$. Note that (2) comprises most distributions used in signal and image processing such as the normal, exponential, gamma, Poisson, Rayleigh, binomial, categorical, log-normal and Dirichlet.

Following an active contour approach, the segmentation of I is addressed by finding a curve $\hat{C} \subset \Omega$ that minimizes the following energy functional [7]:

$$\begin{aligned}\hat{C} = \underset{C}{\operatorname{argmin}} &-\int_{inside(C)} \log f\left(I(\mathbf{x}); \boldsymbol{\theta}_1\right) d\mathbf{x} \\ &-\int_{outside(C)} \log f\left(I(\mathbf{x}); \boldsymbol{\theta}_2\right) d\mathbf{x}.\end{aligned} \tag{3}$$

For simplicity we assume that $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are known a-priori. If this is not the case then equation (3) is typically minimized by updating iteratively C, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Note that for a fixed C the exact minimization of (3) with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ is straightforward by using maximum likelihood estimators.

This paper considers nonparametric contours where the curve $C \in \Omega$ is defined implicitly as the zero level set of a Lipschitz function $\phi : \Omega \to \mathbb{R}$, such that

$$C = \{(\boldsymbol{x}) : \phi(\boldsymbol{x}) = 0\} \tag{4}$$

$$inside(C) = \{(\boldsymbol{x}) : \phi(\boldsymbol{x}) > 0\} \tag{5}$$

$$outside(C) = \{(\boldsymbol{x}) : \phi(\boldsymbol{x}) < 0\} \tag{6}$$

and the energy minimization problem (3) is restated as follows

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} -F(I; \phi) \tag{7}$$

where

$$\begin{aligned}F(I;\phi) \triangleq &-\int_{\Omega} \log f\left(I(\mathbf{x}); \boldsymbol{\theta}_1\right) H(\phi(\boldsymbol{x})) d\boldsymbol{x} \\ &+\int_{\Omega} \log f\left(I(\mathbf{x}); \boldsymbol{\theta}_2\right) H(-\phi(\boldsymbol{x})) d\boldsymbol{x}\end{aligned} \tag{8}$$

and where $H(\cdot)$ denotes the Heaviside function. The functional optimisation problem (7) can be solved by introducing a fictitious time t and solving the associated Euler-Lagrange differential equations $\partial_t \phi = \partial_\phi F$, which lead to the following flow for ϕ

$$\partial_t \phi(\boldsymbol{x}) = -\delta\left(\phi(\boldsymbol{x})\right) \left[\log f\left(I(\mathbf{x}); \boldsymbol{\theta}_1\right) - \log f\left(I(\mathbf{x}); \boldsymbol{\theta}_2\right)\right] \tag{9}$$

where the $\delta(\cdot)$ is the Dirac delta function and where $\partial_\phi F$ is the 1st variation of F with respect to ϕ.

In practice, equations (8) and (9) must be computed over a discrete space-time grid and using sampled functions $\boldsymbol{I} = (\boldsymbol{I}_1, \ldots, \boldsymbol{I}_N)$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$. Equation (8) becomes

$$F(\mathbf{I}; \boldsymbol{\phi}) = \sum_{i=1}^{N} \left(\log f\left(I_i; \boldsymbol{\theta}_1\right) H(\phi_i) + \log f\left(I_i; \boldsymbol{\theta}_2\right) H(-\phi_i)\right). \tag{10}$$

Similarly, equation (9) is now a discrete flow

$$\boldsymbol{\phi}^{t+1} = \boldsymbol{\phi}^t + \eta \nabla_\phi F(\mathbf{I}; \boldsymbol{\phi}^t) \tag{11}$$

where η is the time step, which is bounded by the Courant-Friedrich-Levy (CFL) stability condition, and the gradient $\nabla_\phi F(\mathbf{I}; \boldsymbol{\phi}^t)$ approximates $\partial_\phi F$. This approximation leads to the following iterative algorithm

$$\phi_i^{t+1} = \phi_i^t + \eta \nabla_\phi F_\varepsilon(\mathbf{I}; \boldsymbol{\phi}^t) \tag{12}$$

with

$$\nabla_\phi F_\varepsilon(\mathbf{I}; \boldsymbol{\phi}^t) = \delta_\varepsilon(\phi_i) \left(\log f\left(I_i; \boldsymbol{\theta}_1\right) - \log f\left(I_i; \boldsymbol{\theta}_2\right)\right) \tag{13}$$

which is widely used within the level set community (for numerical stability, AC algorithms use a smooth approximation $\delta_\epsilon(\cdot)$ of Dirac's delta function $\delta(\cdot)$ [2]).

As explained previously, this iterative algorithm is known to have several shortcomings. In particular, it can take a large number of iterations to converge, especially in cases where the gradient $\nabla_\phi F(\mathbf{I}; \boldsymbol{\phi}^t)$ is strongly anisotropic (depends on the direction) (see [8] for more details). In such cases, preconditioning, which changes the geometry of the parameter space, can improve convergence dramatically [8], [9].

The limitations of (12) have motivated numerous papers in the literature that study alternative methods to solve ACs. Most recent papers interpret (12) as a steepest descent problem in the Euclidean space $\mathbb{R}^N$ and investigate alternatives based on more sophisticated optimisation algorithms. Indeed, most state-of-the-art algorithms are steepest descent methods on alternative spaces or manifolds whose inner-products induce favorable properties on gradient flows. For instance, some papers propose to solve (12) in a Sobolev space because its inner product acts as a smoothing operator inducing favorable regularity properties on the contour [10], [11]. Alternatively, Bar *et al.* presented a generalized Newton method that combines an anisotropy reducing inner-product derived from the energy functional's Hessian with a smoothing inner product obtained from a non-canonical space [12]. Bar *et al.* went further by highlighting two important open problems "*the selection of the most appropriate inner product associated with a particular functional*" and "*to incorporate non flat manifolds instead of Euclidean spaces*". These problems were recently addressed in [8], where the authors used information geometry to derive the statistical manifold associated with the Chan-Vese active contour and subsequently proposed a Riemannian steepest descent methods based on the intrinsic metric tensor of that manifold. Experiments reported in [8] show that the resulting algorithm converges extremely fast and produces accurate segmentation results in only a few iterations. Unfortunately these state-of-the-art optimisation methods [8], [11], [12] have been developed for the Chan-Vese active contour and cannot be directly applied to active contours of the exponential family.

III. Proposed optimisation method

This section presents a general Riemannian steepest descent method that can be applied to any active contour of the exponential family. In a manner akin to [8], the method is derived by first using information geometry to obtain the statistical manifold associated with the active contour and then proposing a Riemannian steepest descent method based on the intrinsic metric tensor of that manifold. Note that the method presented here includes as a particular case the Riemannian steepest descent proposed in [8] for the Chan-Vese AC.

We begin by noting that algorithms for minimizing the energy functional (10) can also be interpreted as methods for maximizing the following likelihood

$$f(\boldsymbol{I}; \boldsymbol{\phi}) = \prod_{\{i:\phi_i >= 0\}} f(I_i; \boldsymbol{\theta}_1) \prod_{\{i:\phi_i < 0\}} f(I_i; \boldsymbol{\theta}_2) \tag{14}$$

where $f(I_i;\boldsymbol{\theta})$ has been defined in (2). Based on this, we propose to derive the Riemmanian metric of the statistical manifold $\mathcal{M}$ whose points are the distributions $f(\boldsymbol{I};\boldsymbol{\phi})$, parametrized by $\boldsymbol{\phi}$. This allows us to compute gradient flows on (the tangent spaces of) $\mathcal{M}$ and to define a steepest descent on this manifold. This algorithm, also known as natural gradient descent, generally exhibits very good convergence properties [9] and has been successfully applied to the Chan-Vese AC in [8].

According to information geometry, the *natural* or *intrinsic* metric of $\mathcal{M}$ is the collection of inner products given by the Fisher information matrix (FIM) [13]

$$(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} \triangleq -E\left(\frac{\delta^2}{\delta\phi_i\delta\phi_j}\log\left[f(\mathbf{I};\boldsymbol{\phi})\right]\middle|\boldsymbol{\phi}\right) \tag{15}$$

where $E(\dots|\boldsymbol{\phi})$ denotes the expectation with respect to the distribution $f(\mathbf{I};\boldsymbol{\phi})$. Precisely, at any given point $\boldsymbol{\phi}$, the manifold $\mathcal{M}$ can be approximated locally by an Euclidean tangent space $\mathcal{T}_\phi$ endowed with the inner product $\langle\boldsymbol{\phi}',\boldsymbol{G}(\boldsymbol{\phi})\,\boldsymbol{\phi}\rangle$.

Once $\boldsymbol{G}(\boldsymbol{\phi})$ is known, a steepest descent on $\mathcal{M}$ to optimise F is defined as

$$\phi_i^{t+1} = \phi_i^t + \eta\boldsymbol{G}(\boldsymbol{\phi})^{-1}\nabla_\phi F_\varepsilon(\mathbf{I};\boldsymbol{\phi}^t) \tag{16}$$

where $\boldsymbol{G}(\boldsymbol{\phi})^{-1}\nabla_\phi F(\mathbf{I};\boldsymbol{\phi}^t)$ is the gradient of F on $\mathcal{T}_\phi$ (see [8] for more details).

Unfortunately, deriving the FIM associated with the likelihood (14) is not possible because $H(\cdot)$ is not differentiable. We propose to define an alternative FIM by using the following smooth approximation of $H(\cdot)$ introduced in [2]

$$H_\varepsilon(s) \;=\; \frac{1}{2}\left(1 \;+\; \frac{2}{\pi}\arctan\frac{s}{\varepsilon}\right) \tag{17}$$

which is widely accepted within the level set community. Thus,

$$\begin{aligned}(\boldsymbol{G}_\varepsilon(\boldsymbol{\phi}))_{(i,j)} = \\ &- E\left(\frac{\delta^2}{\delta\phi_j\delta\phi_k}\left[\sum_{i=1}^{N}\log f(I_i,\boldsymbol{\theta}_1)H_\varepsilon(\phi_i)\right]\middle|\boldsymbol{\phi}\right)\\ &- E\left(\frac{\delta^2}{\delta\phi_j\delta\phi_k}\left[\sum_{i=1}^{N}\log f(I_i,\boldsymbol{\theta}_2)H_\varepsilon(-\phi_i)\right]\middle|\boldsymbol{\phi}\right).\end{aligned} \tag{18}$$

By developing the derivatives in (18) we observe that $\mathbf{G}_\varepsilon$ is diagonal and is

$$\begin{aligned}(\boldsymbol{G}_\varepsilon(\boldsymbol{\phi}))_{(i,j)} = &- E\left[\log f(I_i;\boldsymbol{\theta}_1)\delta'_\varepsilon(\phi_i)\middle|\phi_i\right]\\ &- E\left[\log f(I_i;\boldsymbol{\theta}_2)\delta'_\varepsilon(-\phi_i)\middle|\phi_i\right]\end{aligned} \tag{19}$$

if $i=j$ and

$$(\boldsymbol{G}_\varepsilon(\boldsymbol{\phi}))_{(i,j)} = 0$$

otherwise, where $E(\dots|\phi_i)$ denotes the expectation with respect to the marginal likelihood

$$f(I_i;\phi) = \begin{cases} f(I_i;\boldsymbol{\theta}_1) & \text{if } \phi_i \geq 0\\ f(I_i;\boldsymbol{\theta}_2) & \text{if } \phi_i < 0\end{cases} \tag{20}$$

and where

$$\delta'_\varepsilon(x) = \frac{-2\epsilon}{\pi}\frac{x}{(\epsilon^2+x^2)^2}.$$

Given that $\delta'_\varepsilon(-x) = -\delta'_\varepsilon(x)$, if $i=j$ then

$$(\boldsymbol{G}_\varepsilon(\boldsymbol{\phi}))_{(i,j)} = -\delta'_\varepsilon(\phi_i)E\left[\log\left(\frac{f(I_i,\boldsymbol{\theta}_1)}{f(I_i,\boldsymbol{\theta}_2)}\right)\middle|\phi_i\right]. \tag{21}$$

Equation (21) can be further simplified by noting that the expectations in (21) can be expressed in terms of Kullback-Leibler divergences [14], i.e.,

$$(\boldsymbol{G}_\varepsilon(\boldsymbol{\phi}))_{(i,j)} = \begin{cases} -|\delta'_\varepsilon(\phi_i)|\,\mathrm{D}_{\mathrm{KL}}(f(\boldsymbol{\theta}_1)||f(\boldsymbol{\theta}_2)) & \text{if } \phi_i \geq 0\\ -|\delta'_\varepsilon(\phi_i)|\,\mathrm{D}_{\mathrm{KL}}(f(\boldsymbol{\theta}_2)||f(\boldsymbol{\theta}_1)) & \text{if } \phi_i < 0.\end{cases} \tag{22}$$

Since $f(\boldsymbol{\theta})$ belongs to the exponential family, equation (22) can also be expressed in terms of Bregman divergences that admit closed-form expressions [14]

$$(\boldsymbol{G}_\varepsilon(\boldsymbol{\phi}))_{(i,j)} = \begin{cases} -|\delta'_\varepsilon(\phi_i)|\,\mathrm{B}_f(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1) & \text{if } \phi_i \geq 0\\ -|\delta'_\varepsilon(\phi_i)|\,\mathrm{B}_f(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)) & \text{if } \phi_i < 0\end{cases} \tag{23}$$

where

$$\mathrm{B}_f(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1) \triangleq A(\boldsymbol{\theta}_2) - A(\boldsymbol{\theta}_1) - \langle\boldsymbol{\theta}_2 - \boldsymbol{\theta}_2, \nabla A(\boldsymbol{\theta}_1)\rangle \tag{24}$$

and where it is recalled that $A(\boldsymbol{\theta})$ is the logarithm of the normalizing constant of $f(\boldsymbol{\theta})$ defined in (2). Thus, for energy functionals of the form of (10), associated with distributions from the exponential family (2), $\boldsymbol{G}_\varepsilon(\boldsymbol{\phi})$ is diagonal, has a closed-form expression and is trivial to compute and invert.

Moreover, using (23) in equation (16) leads to the following Riemannian steepest descent iteration

$$\phi_i^{t+1} = \phi_i^t + \eta\boldsymbol{G}_\varepsilon^{-1}\left(\boldsymbol{\phi}^t\right)\nabla_\phi F_\varepsilon(\mathbf{I};\boldsymbol{\phi}^t) \tag{25}$$

where $\nabla_\phi F_\varepsilon(\mathbf{I};\boldsymbol{\phi}^t)$ has been defined in (13). Note that this natural gradient algorithm preconditions the Euclidean gradient with $\mathbf{G}_\varepsilon^{-1}\left(\boldsymbol{\phi}^t\right)$ to ensure isotropic convergence (see [8], [9] for more details).

Finally, we now need to modify (25) so that it will promote smooth solutions. This regularization of $\boldsymbol{\phi}$ improves significantly segmentation results by introducing prior knowledge about image structure, i.e., pixels belonging to the foreground and background are organized in spatial groups (as opposed to being randomly distributed across the image). In a manner akin to [11], [12] and [8], we propose to regularize $\boldsymbol{\phi}$ by smoothing gradient flows. Precisely, we define the following iterative algorithm based on a smooth natural gradient [8]

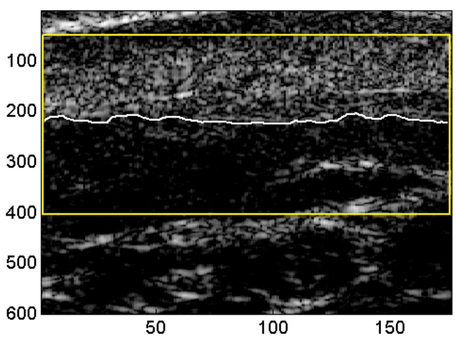
$$\boldsymbol{\phi}^{t+1} = \boldsymbol{\phi}^t + \eta\mathbf{H}_\sigma\mathbf{G}_\varepsilon^{-1}\left(\boldsymbol{\phi}^t\right)\nabla_\phi F_\varepsilon(\mathbf{I};\boldsymbol{\phi}^t) \tag{26}$$

where $\mathbf{H}_\sigma$ is a smoothing operator. In this work we set $\mathbf{H}_\sigma = \text{Toeplitz}(h_\sigma)$, where $h_\sigma(s,u) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{s^2+u^2}{2\sigma^2}\right]$ is a Gaussian kernel of scale σ. The parameter σ determines the width of the Gaussian kernel and therefore the amount of smoothness enforced by $\mathbf{H}_\sigma$. More details regarding the motivation for choosing this particular smoothing operator and about the selection of σ can be found in [8].

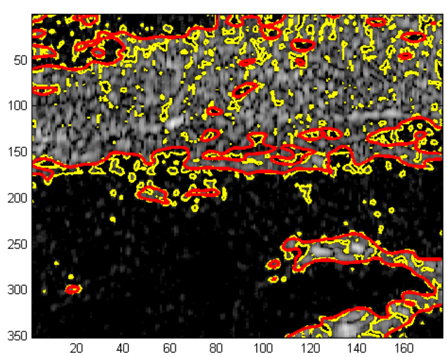
IV. Experimental Results and Observations

Here we demonstrate the proposed methodology on a challenging image segmentation problem for which Euler's method converges very slowly. Specifically, we consider a Rayleigh AC model with application to ultrasound image segmentation [3] (additional experiments using the Chan-Vese AC can be found in [8]). To the best of our knowledge the Euler method proposed in [3] is the only algorithm that has been applied to the Rayleigh AC. In order to apply (26) to this problem we have evaluated (23) using the log-normalising constant of the Rayleigh distribution. To guarantee that the comparisons are fair both algorithms use the same initialization and step size (we used $\eta = 0.1$ as recommended in [3]).

Fig. 1(a) shows a B-mode ultrasound image of in-vivo human dermis the dermis-hypodermis junction has been annotated approximately by an expert (coarse white line). The region of interest used in the experiments is depicted in yellow. Fig. 1(b) shows in coarse red the segmentation results obtained with our method and in yellow those obtained with [3].



(a)



(b)

Fig. 1. Comparison of our method (red) with Euler's method (yellow) on a Rayleigh AC with application to ultrasound image segmentation.

We observe the proposed method is robust to speckle noise and clearly estimates the dermis-hypodermis junction. More importantly, the proposed method converged in only 18 iterations, which took 0.34 seconds, whereas Euler's method required 9860 iterations and 269 seconds to produce a stable solution (experiments computed on an Intel Core 2 Duo @2.1 GHz processor workstation running MATLAB R2010b). This dramatic difference in speed results from the fact that proposed method takes steps along the steepest directly on the parameter's intrinsic manifold, as opposed to using the default Euclidean gradient. Finally, note that this extremely fast convergence is in agreement with the experiments reported in [8].

V. Conclusion

This paper has shown how information geometry can be used to define the Riemannian metric of the statistical manifold associated with nonparametric active contour models. Through appropriate mathematical developments we have shown that the Fisher Information matrix, which determines the natural metric of the statistical manifold is diagnonal and invertible. This insight enables a fast converging segmentation methodology which we have demonstrated on an ultrasound image segmentation problem where a 2 orders of magnitude increase in segmentation speed is achieved.

Acknowledgments

We are grateful to the Hospital of Toulouse and Pierre Fabre Laboratories for the ultrasound image used in IV.

References

[1] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations," *J. Comput. Physics*, vol. 79, no. 1, pp. 12–49, 1988.

[2] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, feb 2001.

[3] A. Sarti, C. Corsi, E. Mazzini, and C. Lamberti, "Maximum likelihood segmentation of ultrasound images with rayleigh distribution," *IEEE Trans. Ultrason. Ferroelect. Freq. Contr.*, vol. 52, no. 6, pp. 947–960, june 2005.

[4] I. Ben Ayed, A. Mitiche, and Z. Belhadj, "Multiregion level-set partitioning of synthetic aperture radar images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 27, no. 5, pp. 793–800, 2005.

[5] I. Ayed, N. Hennane, and A. Mitiche, "Unsupervised variational image segmentation/classification using a Weibull observation model," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3431–3439, 2006.

[6] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *J. Math. Imaging Vis.*, vol. 28, no. 2, pp. 151–167, June 2007.

[7] F. Lecellier, J. Fadili, S. Jehan-Besson, G. Aubert, M. Revenu, and E. Saloux, "Region-based active contours with exponential family observations," *J. Math. Imaging Vis.*, vol. 36, pp. 28–45, January 2010.

[8] M. Pereyra, H. Batatia, and S. McLaughlin, "Exploiting information geometry to improve the convergence properties of variational active contours," *IEEE J. Sel. Topics Signal Processing*, vol. 7, no. 4, pp. 1–8, Aug. 2013.

[9] S. Amari and S. Douglas, "Why natural gradient?" in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, vol. 2, Seattle, USA, May 1998, pp. 1213 –1216.

[10] G. Sundaramoorthi, A. Yezzi, A. Mennucci, and S. G., "New possibilities with sobolev active contours," *International Journal of Computer Vision*, vol. 84, no. 2, pp. 113–129, May 2009.

[11] R. J. Renka, "Image segmentation with a sobolev gradient method," *J. Nonlinear Analysis: Theory, Methods and Applications*, vol. 71, no. 12, pp. e774–e780, Dec. 2009.

[12] L. Bar and G. Sapiro, "Generalized newton-type methods for energy formulations in image processing," *SIAM J. Imaging Sciences*, vol. 2, no. 2, pp. 508–531, 2009.

[13] S.-I. Amari, *Differential-Geometrical Methods in Statistics (Lecture Notes in Statistics 28)*. Springer, 1990.

[14] F. Nielsen and R. Nock, "Entropies and cross-entropies of exponential families," 2010, pp. 3621–3624.