Region-based active contours with exponential family observations

F. Lecellier[1], J. Fadili[1], S. Jehan-Besson[2], G. Aubert[3], M. Revenu[1], E. Saloux[4]
1: GREYC UMR CNRS 6072, 6 bd MARECHAL JUIN, 14050 CAEN CEDEX, FRANCE
2: LIMOS UMR CNRS 6158, 63173 AUBIERE CEDEX, FRANCE
3: Laboratoire J.A. Dieudonné UMR CNRS.6621, 06108 NICE CEDEX, FRANCE
4: CHU de CAEN, 14000 CAEN, FRANCE

2008-01-24

Abstract

In this paper, we focus on statistical region-based active contour models where image features (e.g. intensity) are random variables whose distribution belongs to some parametric family (e.g. exponential) rather than confining ourselves to the special Gaussian case. In the framework developed in this paper, we consider the general case of region-based terms involving functions of parametric probability densities, for which the anti-log-likelihood function is a special case. Using shape derivative tools, our effort focuses on constructing a general expression for the derivative of the energy (with respect to a domain), and on deriving the corresponding evolution speed. More precisely, we first show by a counterexample that the estimator of the distribution parameters is crucial for the derived speed expression. On the one hand, when using the maximum likelihood (ML) estimator for these parameters, the evolution speed has a closed-form expression that depends simply on the probability density function. On the other hand, complicating additive terms appear when using other estimators, e.g. method of moments. We then proceed by stating a general result within the framework of multi-parameter exponential family. This result is specialized to the case of the anti-log-likelihood score with the ML estimator and to the case of the relative entropy. Experimental results on simulated data confirm our expectations that using the appropriate noise model leads to the best segmentation performance. We also report preliminary experiments on real life Synthetic Aperture Radar (SAR) images to demonstrate the potential applicability of our approach.

keywords Segmentation, Region-based active contours, Exponential family, Shape derivation, Maximum likelihood, Relative Entropy.

1 Introduction

State of affairs

In image segmentation, the main issue is to extract one or several regions according to a given criterion. Since the seminal work of [1, 2], active contours have proven their efficiency for such a task. The general idea behind active contours model is to apply partial differential equations (PDEs) to deform a curve towards the boundaries of the objects of interest. Snakes [1], balloons [3] and geodesic active contours [2] were pioneering works on active contour model. In these methods, the contour is driven towards image edges. More recently, region-based active contours (i.e. RBAC) were proposed [4, 5, 6, 7, 8]. In these approaches, region-based terms can be advantageously combined with boundary-based ones. The evolution equation is generally deduced from a general criterion to minimize that includes both region integrals and boundary integrals. The combination of those two terms in the energy functional allows the use of photometric image properties, such as texture [9, 10, 11, 12] and noise [13, 14, 15], as well as geometric properties such as the prior shape of the object to be segmented [16, 17, 18, 19, 20, 21], see also the review in [22]. RBACs have proven their efficiency for a wide range of applications such as medical image segmentation [23, 24], video object segmentation or tracking [25]. The main issues when dealing with RBACs models are the definition of an appropriate criterion for the given segmentation problem and the derivation of the PDEs from this criterion.

Concerning the latter issue, we first point out that the PDE derivation is not trivial when the energy criterion involves region functions. This is mostly due to the fact that the set of image regions does not have the structure of a vector space, preventing us from using in a straightforward way gradient descent methods. To circumvent this problem, we propose to take benefit of the framework proposed in [26, 27, 28], based on shape derivation principles developed in [29, 30]. This framework is particularly well-adapted when dealing with global information of the region such as statistical image features (e.g. mean, variance, entropy, histogram) viewed as empirical probability density function (pdf)-based estimates. In this case, one must pay attention to the fact that these features are globally attached to the region and must then be taken into account in the shape derivation framework [26, 27, 28].

As far as the definition of the criterion is concerned, we focus on region-based terms that allow to take benefit of statistical image properties. And we pay a particular attention to the random part that contaminates the image coming during its acquisition process, i.e. the noise model. In many papers [13, 14, 15], the authors proposed region-based terms that involve functions of the pdf of some image attributes within the region. The minimization of the anti-log-likelihood is classically used for the segmentation of homogeneous regions [6, 7, 31, 32, 33]. However, many works implicitly consider that the intensity inside homogeneous regions follows a Gaussian law. This is the case in [34] when minimizing the mean squared error between the pixel intensity and the mean of the region, or in [7, 11] where they use the two parameters of Gaussian distributions (μ, σ^2). However such an assumption is not always justified. For example, it is well known that under appropriate conditions (large number of randomly located scatters), the Rayleigh distribution is well suited to model the noise in echographic data [35]. Some recent papers consider a more general case where the pdf belongs to other families [36, 33, 37, 38]. The pdf is then characterized by one or more parameters describing the underlying statistical model. These parameters depend on the region and must be estimated at each evolution step of the active contour. This necessitates to estimate these region-dependent parameters, which can be achieved using various estimators, typically method of moments, maximum likelihood (ML). Nonetheless, to the best of our knowledge, the influence of the estimation method on the computation of the evolution equation has never been investigated.

Contributions

In the framework developed hereafter, we consider the general case of region-based terms involving functions of parametric pdfs without restricting ourselves solely to the anti-log-likelihood function. Using shape derivative tools [30, 26, 27, 28], our effort focuses on constructing a general expression for the derivative of the energy (with respect to a domain) and on deriving the corresponding evolution speed term. Shape derivation tools lead us to obtain very general results and to rigorously take into account the

way the distribution parameters were estimated. As far as the derivation is concerned, we particularly pay attention to the fact that distribution parameters depend on the region. More precisely, we show by a counterexample that the estimator of the distribution parameters is crucial for the derived speed expression. On the one hand, when using ML estimators for the parameters, the evolution speed has a closed-form expression that depends simply on the pdf. On the other hand, complicating additive terms appear when using other estimators such as the method of moments. We then provide a general result for the evolution equation within the framework of multi-parameter exponential family. The rationale behind using the exponential family is that it includes, among others, Gaussian, Rayleigh, Poisson and Bernoulli distributions that have proven to be useful to model the noise structure in many real image acquisition devices (e.g. Poisson for photon counting devices such as X-ray or CCD cameras, Rayleigh for ultrasound images, etc). Our general framework is also specialized to some particular cases, such as the anti-log-likelihood score or the relative entropy. These particular cases are chosen because they allow to formulate a wide range of region criteria in image segmentation, e.g. anti-log-likelihood as region homogeneity measure, relative entropy for region competition. Interestingly, when particularizing our general expression to the anti-log-likelihood associated to the ML estimator for the parameters, we recover the same expressions as those previously published in the literature, e.g. [6, 13]. The evolution using relative entropy for the whole exponential family has, to the best of our knowledge, never been investigated. Some works propose to use the relative entropy for medical image segmentation [] but they restric their study to the Gaussian model. Beside theoretical arguments, we also provide experimental results to support our claims and show the influence of the noise model on the final segmentation.

This paper is organized as follows. In Section 2 we briefly remind some key concepts on shape derivation tools. Then, in Section 3 we present the statistical framework for our noise model, we review some properties of the exponential family and we state a first result within this framework. In Section 4, we put all previous ingredients together and give our main theoretical results on the whole exponential family. In Section 5 we detail our segmentation algorithm and our experimental results. In Section 6 we conclude and present some perspectives. Proofs are deferred to the appendix awaiting inspection by the interested reader.

Notations

Notation	Signification
Ω_I	The image domain
Ω	A region of the image
$\partial\Omega$	The boundary of the region Ω
$\|\Omega\|$	Region size $\int_\Omega d\mathbf{x}$
$I(\mathbf{x})$	The intensity of the pixel at the location $\mathbf{x}$
$k(\mathbf{x},\Omega)$	The region descriptor of Ω
$k_b(\mathbf{x})$	The boundary descriptor of $\partial\Omega$
Ω_{in}	Inside region
Ω_{out}	Outside region
$< J'(\Omega), \mathbf{V} >$	The Eulerian derivative of domain energy criterion $J(\Omega)$
$p(y,\boldsymbol{\eta})$	Probability density function (pdf) of the random variable Y
$\boldsymbol{\eta}$	Hyper-parameter vector of the pdf
$\mathbb{E}[Y]$	Expectation of the random variable Y
$k'(\mathbf{x},\Omega,\mathbf{V})$	The domain derivative of the function k
$\mathbf{speed}(\mathbf{x},\Omega)$	Evolution speed of the active contour

2 Shape derivation theory

In this part we introduce the general criterion to minimize and we briefly remind some necessary details on shape derivative tools and on the associated evolution equation.

2.1 Introduction of a general criterion

Let $\mathcal{U}$ be a class of domains (open, regular bounded sets, i.e. C^2) of $\mathbb{R}^d$ ($d = 2$ in two-dimensions), and Ω an element of $\mathcal{U}$. The boundary $\partial\Omega$ of Ω is sometimes denoted by Γ. The segmentation problem consists in extracting one or more regions Ω from the complete image Ω_I. We search for the domain or the partition of the image that minimizes a general functional which is composed of a combination of region-based terms and boundary based terms. The minimization is performed using active contours that are driven by an evolution equation directly deduced from the functional.

On the one hand, the region-based term corresponds to a global information on the region of interest. It can for example describe the homogeneity of a region. The region-based term is usually expressed as a domain integral of a function k called descriptor of the region or by abuse of terminology, homogeneity criterion:

$$J_r(\Omega) = \int_\Omega k(\mathbf{x}, \Omega) d\mathbf{x} \ . \tag{1}$$

In the general case the function k may depend on the domain. For example when minimizing the difference between the intensity and the mean value inside the region, we take k equal to $(I(\mathbf{x}) - \mu(\Omega))^2$ and $\mu(\Omega)$ obviously depends on the domain Ω.

On the other hand, the boundary-based term corresponds to a local information on the boundary of the region of interest. It can be used as a regularization term. The boundary-based term is expressed as a boundary integral of a function k_b coined boundary descriptor of the region or by abuse of terminology, regularity criterion:

$$J_b(\delta\Omega) = \int_{\delta\Omega} k_b(\mathbf{x}) d\mathbf{a}(\mathbf{x}) \ , \tag{2}$$

where $d\mathbf{a}(\mathbf{x})$ is the area element. An example of boundary-based term can be the curve length $J_b(\partial\Omega) = \int_s ds = L(\partial\Omega)$.

For the sake of clarity, and without loss of generality, we consider in the rest of our paper the segmentation into two regions Ω_{in} and Ω_{out}, Γ represents the interface between the two domains. We then look for the partition $\{\Omega_{in}, \Omega_{out}, \Gamma\}$ of the image that minimizes the following functional:

$$J(\Omega_{in}, \Omega_{out}, \Gamma) = \int_{\Omega_{in}} k_{in}(\mathbf{x}, \Omega_{in}) d\mathbf{x} + \int_{\Omega_{out}} k_{out}(\mathbf{x}, \Omega_{out}) d\mathbf{x} + \int_\Gamma k_b(\mathbf{x}) d\mathbf{a} \ . \tag{3}$$

Once the functional has been expressed, the problem now is to deduce the evolution equation that will drive the RBAC towards a minimum of the functional. This is achieved in our case through the derivation of the criterion using shape derivative tools.

2.2 Shape derivation tools

In order to be comprehensive, we here give a brief summary of the shape derivation theory. The interested reader may refer to [30, 28] for further details.

As stated in the introduction, the set of image regions does not have the structure of a vector space. Consequently, the derivation of the region-term is performed using domain derivation tools. We remind some definitions [30] that we state in full for the reader convenience. We can consider that Ω evolves in a velocity vector field, $\mathbf{V}$ and calculate the variations of $J(\Omega(t))$ in the direction $\mathbf{V}$:

Definition 1 *The Eulerian derivative of $J_r(\Omega)$ in the direction $\mathbf{V}$ is defined as follows:*

$$< J_r'(\Omega), \mathbf{V} >= \lim_{\tau \to 0} \frac{J_r(\Omega(\tau)) - J_r(\Omega)}{\tau} \ . \tag{4}$$

Definition 2 *The domain derivative of k in the direction $\mathbf{V}$ denoted $k'(\mathbf{x}, \Omega, \mathbf{V})$ is defined as follows:*

$$k'(\mathbf{x}, \Omega, \mathbf{V}) = \lim_{\tau \to 0} \frac{k(\mathbf{x}, \Omega(\tau)) - k(\mathbf{x}, \Omega)}{\tau} \ . \tag{5}$$

Theorem 1 *There is a relation between the Eulerian derivative of $J_r(\Omega)$ in the direction $\mathbf{V}$, and the domain derivative of k denoted $k'(\mathbf{x}, \Omega, \mathbf{V})$:*

$$< J_r'(\Omega), \mathbf{V} >= \int_\Omega k'(\mathbf{x}, \Omega, \mathbf{V}) d\mathbf{x} - \int_{\partial\Omega} k(\mathbf{x}, \Omega) \, (\mathbf{V} \cdot \mathbf{N}) d\mathbf{a}(\mathbf{x}) \ , \tag{6}$$

where $\mathbf{N}$ is the unit inward normal to $\partial\Omega$.

A proof can be found in [29, 30] and an elementary one in [28]. The first integral comes from the dependence of the descriptor $k(\mathbf{x}, \Omega)$ with the region, while the second term comes from the evolution of the region itself.

2.3 Evolution equation

From the shape derivative, we can derive the evolution equation that will drive the active contour towards a (local) minimum of the criterion.

Let us suppose that the shape derivative of the region Ω may be written as follows:

$$< J_r'(\Omega), \mathbf{V} >= - \int_{\partial\Omega} \mathbf{speed}(\mathbf{x}, \Omega)(\mathbf{V}(\mathbf{x}) \cdot \mathbf{N}(\mathbf{x})) d\mathbf{a}(\mathbf{x}) \ . \tag{7}$$

We can then deduce [28] the following evolution equation:

$$\frac{\partial \Gamma}{\partial \tau} = \mathbf{speed}(\mathbf{x}, \Omega) \, \mathbf{N}(\mathbf{x}) \ , \tag{8}$$

with $\Gamma(\tau = 0) = \Gamma_0$.

Let us consider the classical homogeneity descriptor proposed by [34], i.e. $k(\mathbf{x}, \Omega) = (I(\mathbf{x}) - \mu(\Omega))^2$, where $\mu(\Omega)$ is the mean over region Ω. Using Theorem 1, the shape derivative becomes [28]:

$$< J_r'(\Omega), \mathbf{V} >= - \int_{\partial\Omega} (I - \mu(\Omega))^2 (\mathbf{V}(\mathbf{x}) \cdot \mathbf{N}(\mathbf{x})) d\mathbf{a}(\mathbf{x}) \ , \tag{9}$$

because the term $\int_\Omega k'(\mathbf{x}, \Omega, \mathbf{V}) d\mathbf{x}$ vanishes in this case. It is immediate to see that the evolution speed is:

$$\mathbf{speed}(\mathbf{x}, \Omega) = (I(\mathbf{x}) - \mu(\Omega))^2 \ , \tag{10}$$

which is precisely the classical result found by [34]. Note however, that in the case of a derivation using shape derivation tools, we get rid of the regularizing Dirac distribution δ_ϵ that appears when minimizing the criterion with the method described by [34].

3 Statistical Framework for variational segmentation

Statistical decision and estimation theory are theoretical tools that allow to handle rigorously the problem of information extraction from an image under random fluctuations. In this paper, we focus on the segmentation of a region of interest in images. We want to take benefit of a more complete modeling of the noise formation process in images so as to incorporate this feature in a variational segmentation framework. We begin by giving arguments supporting the validity of the exponential family in real image segmentation problems.

3.1 Statistical Image Modeling and Segmentation

When observing an image, one does not observe the true gray level of pixels, but rather random fluctuations of those levels. The statistical framework appears as a rigorous and flexible candidate to describe this randomness. In statistical image modeling, an image is generally considered either as the observation of a random variable at different locations, or as a random field. We here adopt the former model. The source of random fluctuations will be imputed to noise, which is introduced by the imaging system and has its own pdf that might be known or not. Fortunately, in many cases of interest, the noise pdf can be modeled from a known parametric family. For example, The assumptions of the classical Synthetic Aperture Radar (SAR) image generation model lead to a Rayleigh noise model. Ultrasound noise in images is also approximately Rayleigh-distributed. Modulus of Magnetic Resonance Images (MRI) is known to follow a Rician distribution, etc. In fact, it turns out that the classical Gaussian model is an exception rather than a rule.

In our work we want to incorporate a noise model in RBACs. The exponential family is here chosen because it includes a wide range of models such as Bernoulli, Binomial, Poisson, Gamma, Beta, Rayleigh, etc. One may have noticed that the few imaging modalities just cited involve noise models whose pdfs are special instances of this general and comprehensive family.

The introduction of the noise model in RBACs goes through the definition of an appropriate region descriptor k. Let us consider $p(y(\mathbf{x}), \eta)$ the parametric pdf of the observations $y(\mathbf{x})$. Here, we consider $y(\mathbf{x})$ as an image feature (e.g. the intensity) at location $\mathbf{x}$. The pdf often depends on some parameters denoted η. In this paper, we focus on the fact that these parameters evolve with the domain Ω.

Let us now introduce some useful definitions that will allow to bridge the gap between the statistical framework and the minimization of a criterion as involved in RBACs.

Definition 3 *The likelihood function or score L for a given observed data y is:*

$$L_y(\boldsymbol{\eta}) = p(y, \boldsymbol{\eta}), \quad \boldsymbol{\eta} \in \mathcal{E} \subseteq \mathbb{R}^{\kappa}. \tag{11}$$

Thus, when we think of $L_y(\boldsymbol{\eta})$ as a function of $\boldsymbol{\eta}$, it gives, for some observed y the "likelihood" or "plausibility" of various $\boldsymbol{\eta}$. More clearly, we can think of $L_y(\eta)$ as a measure of how likely $\boldsymbol{\eta}$ is to have produced the observation y. If the domain of the image characteristic $y(\mathbf{x})$ were discrete (e.g. gray level), then for each $\boldsymbol{\eta}$, $L_y(\boldsymbol{\eta})$ would correspond to the probability of observing $y(\mathbf{x})$. We can easily extend this definition to the RBACs. Hence, the likelihood function becomes:

$$L_y(\boldsymbol{\eta}, \Omega) = p(y(\mathbf{x}) : \mathbf{x} \in \Omega, \boldsymbol{\eta}), \boldsymbol{\eta} \in \mathcal{E} \ . \tag{12}$$

In general, we use the log-likelihood score:

$$l_y(\boldsymbol{\eta}, \Omega) = \log(L_y(\boldsymbol{\eta}, \Omega)) = \log(p(y(\mathbf{x}) : \mathbf{x} \in \Omega, \boldsymbol{\eta})) \ . \tag{13}$$

$\log(p)$ turns to be the best monotone function of p to be considered for many reasons [39]. A typical one is that if the $y(\mathbf{x}_i)$ are independent and identically distributed (iid) with a density or frequency function $p(y(\mathbf{x}_i), \boldsymbol{\eta})$ for $i = 1, ..., n$, then, with $y = (y(\mathbf{x}_1), ..., y(\mathbf{x}_n))$

$$l_y(\boldsymbol{\eta}, \Omega) = \log(p(y, \boldsymbol{\eta}, \Omega)) = \log \prod_{i=1}^{n} p(y(\mathbf{x}_i) : \mathbf{x}_i \in \Omega, \boldsymbol{\eta}) = \sum_{i=1}^{n} \log p(y(\mathbf{x}_i) : \mathbf{x}_i \in \Omega, \boldsymbol{\eta}) \ . \tag{14}$$

This justifies the anti-log-likelihood criterion that is classically used:

$$J_r(\Omega) = -\int_{\Omega} \log(p(y(\mathbf{x}), \boldsymbol{\eta}) \, d\mathbf{x} \ . \tag{15}$$

This criterion was proposed in active contours by [6]. In this paper, we consider a more general setting replacing the $-\log$ by any proper function Φ, see in Section 4.

3.2 Exponential family

As argued in Section 3.1, the exponential family covers most noise models commonly encountered in image acquisition systems. This section provides the necessary material on the exponential family.

3.2.1 Definition

The multi-parameter exponential family is naturally indexed by a κ-dimensional real parameter vector and a κ-dimensional natural statistic vector $\mathbf{T}(Y)$. A simple example is the normal family when both the location and the scale parameters are unknown ($\kappa = 2$). The normal, Poisson and Rayleigh distributions exhibit the interesting feature that there is a natural sufficient statistic whose dimension as random vector is independent of the sample size. The class of families of distributions that we introduce in this section was first discovered in statistics by [40] through investigations of this property. Subsequently, many other properties of these families were discovered and they have became an important class of the modern theory of statistics.

Definition 4 *The family of distributions of a Random Variable (RV) Y $\{\mathcal{P}_{\boldsymbol{\eta}} : \boldsymbol{\eta} \in \mathcal{E} \subseteq \mathbb{R}^\kappa\}$, is said a κ-parameter* **canonical** *exponential family, if there exist real-valued functions $\eta_1, \ldots, \eta_\kappa : \Theta \mapsto \mathbb{R}$ and $A(\boldsymbol{\eta})$ on $\mathcal{E}$, and real-valued functions $h, T_1, \ldots, T_\kappa : \mathbb{R}^\kappa \mapsto \mathbb{R}$, such the pdf $p(y, \boldsymbol{\eta})$ of the $\mathcal{P}_{\boldsymbol{\eta}}$ may be written:*

$$p(y, \boldsymbol{\eta}) = h(y) \exp[\langle \boldsymbol{\eta}, \mathbf{T}(y) \rangle - A(\boldsymbol{\eta})], y \in \chi \subset \mathbb{R} \ . \tag{16}$$

where $\mathbf{T} = (T_1, \ldots, T_\kappa)^T$ is the natural sufficient statistic, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_\kappa)^T$ and $\mathcal{E}$ are the natural parameter vector and space, $\langle \boldsymbol{\eta}, \mathbf{T} \rangle$ denotes the scalar product, the natural parameter space is defined as $\mathcal{E} = \{\boldsymbol{\eta} \in \mathbb{R}^\kappa; -\infty < A(\boldsymbol{\eta}) < +\infty\}$.

We draw the reader's attention to the fact than $\boldsymbol{\eta}$ is a function of $\theta \in \Theta$ which is the parameter of interest in most applications.

Table 1 provides a synthetic description of some common distributions of the exponential family, with the associated parameters, functions (see Definition 4) and sufficient statistics.

Distribution	$\boldsymbol{\theta}^T$	$\boldsymbol{\eta}(\boldsymbol{\theta})^T$	$\mathbf{T}(y)^T$	$A(\boldsymbol{\eta})$	$\mathcal{E}$
Normal	(μ, σ^2)	$(\frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2})$	(y, y^2)	$\frac{1}{2}\left(-\frac{\eta_1^2}{2\eta_2} - \log\frac{-\eta_2}{\pi}\right)$	$\mathbb{R} \times \mathbb{R}^{-*}$
Gamma	(λ, p)	$(-\lambda, p-1)$	$(y, \log y)$	$-(\eta_2+1)\log -\eta_1 + \log\Gamma(\eta_2+1)$	$\mathbb{R}^{-*}\times]-1, +\infty[$
Beta	(r, s)	$(r-1, s-1)$	$(\log y, \log(1-y))$	$-\log B(\eta_1+1, \eta_2+1)$	$\mathbb{R}^+ \times \mathbb{R}^+$
Poisson	μ	$\log\mu$	y	e^η	$\mathbb{R}$
Exponential	λ	$-\lambda$	y	$-\log -\eta$	$\mathbb{R}^{-*}$
Rayleigh	θ^2	$-1/2\theta^2$	y^2	$-\log -2\eta$	$\mathbb{R}^{-*}$

Table 1: Some common canonical exponential families. $B(\alpha, \beta)$ is the Euler Beta function.

In order to illustrate this table, let us develop the form of the natural parameters for the Normal law:

$$\begin{aligned} p(y(\mathbf{x}); \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y(\mathbf{x})-\mu)^2}{2\sigma^2}} &= \exp\left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} + \frac{\mu y}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right) \\ &= \exp\left(\langle \boldsymbol{\eta}, \mathbf{T}(y) \rangle - \frac{1}{2}\log(2\pi\sigma^2) - \frac{\mu^2}{2\sigma^2}\right) . \end{aligned}$$

It follows that

$$h(y) = 1, \quad \mathbf{T}(y) = \begin{bmatrix} y \\ y^2 \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix}, \quad A(\boldsymbol{\eta}) = \frac{1}{2}\left(\log(2\pi\sigma^2) + \frac{\mu^2}{2\sigma^2}\right) = -\frac{1}{2}\left(\frac{\eta_1^2}{2\eta_2} + \log(-\frac{\eta_2}{\pi})\right) \ .$$

3.2.2 Properties

The following results will be useful for our RBAC scheme based on the exponential family. Their proofs may be found in [39].

Theorem 2 *Let $\mathcal{P}$ a κ-parameter canonical exponential family with natural sufficient statistic $\mathbf{T}(Y)$ and open natural parameter space $\mathcal{E}$ then :*

(i) $\mathcal{E}$ *is convex.*

(ii) $A : \mathcal{E} \mapsto \mathcal{S} \subseteq \mathbb{R}$ *is convex where* $\mathcal{S} = A(\mathcal{E})$.

(iii) $\mathbb{E}[\mathbf{T}(Y)] = \nabla A(\boldsymbol{\eta})$.

(iv) $\mathrm{Cov}[T(Y)] = \ddot{A}(\boldsymbol{\eta})$.

where $\nabla A = (\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}, .., \frac{\partial A}{\partial \eta_\kappa})^T$ *represents the gradient of A, and* $\ddot{A}$ *is the Hessian matrix of A with* $\ddot{A}_{ij} = \frac{\partial^2 A}{\partial \eta_i \partial \eta_j}$.

The following theorem establishes the conditions of strict convexity of A, and then those for $\dot{A}$ to be 1-1 on $\mathcal{E}$. This is a very useful result for optimization (derivation) purposes:

Theorem 3 *Let $\mathcal{P}$ a full rank (i.e. $\mathrm{Cov}[T(Y)]$ is a positive-definite matrix) κ-parameter canonical exponential family with natural sufficient statistic $\mathbf{T}(Y)$ and open natural parameter space $\mathcal{E}$ [39].*

(i) $\nabla A : \mathcal{E} \mapsto \mathcal{S}$ *is 1-1.*

(ii) The family may be uniquely parameterized by $\boldsymbol{\mu}(\boldsymbol{\eta}) \equiv \mathbb{E}[\mathbf{T}(Y)] = \nabla A(\boldsymbol{\eta})$.

(iii) The anti-log-likelihood function is a strictly convex function of $\boldsymbol{\eta}$ *on* $\mathcal{E}$.

These results establish a 1-1 correspondence between $\boldsymbol{\eta}$ and $\mathbb{E}[\mathbf{T}(Y)]$ such that:

$$\mathcal{S} \ni \boldsymbol{\mu} = \nabla A(\boldsymbol{\eta}) = \mathbb{E}[(\mathbf{T}(Y)] \Leftrightarrow \mathcal{E} \ni \boldsymbol{\eta} = \psi\left(\mathbb{E}[\mathbf{T}(Y)]\right) \tag{17}$$

holds uniquely with ∇A and ψ continuous. At this stage, it is interesting to mention that an alternative solution to establish this bijection, is to use the Legendre conjugate (convex analysis) in the same vein as in the work of [41] which used it to prove the bijection between exponential families and Bregman divergences.

It is also interesting to point out that in (17), when the expectation $\mathbb{E}[\mathbf{T}(Y)]$ is replaced with the empirical estimate of the mean $\overline{\mathbf{T}(Y)}$, the obtained estimate coincides with the ML estimator (MLE) of $\boldsymbol{\eta}$. For example, when dealing with the Rayleigh distribution, we have:

$$\eta = \frac{-1}{2\theta^2} \quad , \quad A(\eta) = -\log(-2\eta) \quad \text{and} \quad T(y) = y^2 \ . \tag{18}$$

By computing $A'(\eta) = \overline{\mathbf{T}(Y)}$, we find that:

$$-\frac{1}{\eta} = \frac{1}{|\Omega|} \int_{\Omega} y(\mathbf{x})^2 d\mathbf{x} \ , \tag{19}$$

which corresponds to the MLE of the parameter θ^2 given by:

$$\widehat{\theta^2}_{\mathrm{ML}} = \frac{1}{2|\Omega|} \int_{\Omega} y(\mathbf{x})^2 d\mathbf{x} \ \hat{A}0. \tag{20}$$

3.3 A primer case for the statistical region-based active contours

This section is devoted to illustrate the influence of the parameters estimator on the expression of the RBAC evolution speed. We consider here the minimization of the anti-log-likelihood function when the image feature is Rayleigh-distributed, and we compute the shape derivative of this region-based term. For such a computation, we need to evaluate the parameter of the law using some estimation method. The MLE, which is an alternative to the minimum variance unbiased estimator (MVUE), can then be computed. The MLE, in contrast to the MVUE, does not necessarily satisfy any optimality criterion, but it can almost always be computed, either through exact formulas or numerical techniques. For this reason, the MLE is one of the most common estimation procedure used in practice. There is another classical estimation method, namely the method of moments. This alternative is a technique for constructing estimators of the parameters that is based on matching the sample moments with the corresponding distribution moments. We propose here to compare the evolution equation of the active contour obtained using the MLE and the method of moments when minimizing the anti-log-likelihood criterion for the Rayleigh law.

3.3.1 Prologue

The functional to minimize is then:

$$J_r(\Omega) = -\int_\Omega \log(p(y(\mathbf{x}),\theta)\, d\mathbf{x} \ . \tag{21}$$

with p the Rayleigh pdf of scalar parameter θ, i.e $p(y(\mathbf{x}),\theta) = \frac{y(\mathbf{x})}{\theta^2}\exp\left(\frac{-y(\mathbf{x})^2)}{2\theta^2}\right)$.
Classically, one can compute an estimate of the parameter θ using the method of moments. In this case the estimator is given by the sample mean :

$$\widehat{\theta}_{\mathrm{MO}} = \sqrt{\frac{2}{\pi}}\ \frac{1}{|\Omega|}\int_\Omega y(\mathbf{x})d\mathbf{x} = \sqrt{\frac{2}{\pi}}\ \overline{y(\mathbf{x})}\ , \tag{22}$$

where $\overline{y(\mathbf{x})} = \frac{1}{|\Omega|}\int_\Omega y(\mathbf{x})d\mathbf{x}$ denotes the sample mean inside the region Ω.
Alternatively, one can also compute an estimate through the MLE as given by:

$$\widehat{\theta}_{\mathrm{ML}} = \sqrt{\frac{1}{2|\Omega|}\int_\Omega y(\mathbf{x})^2 d\mathbf{x}}\ . \tag{23}$$

Shape derivative with the moment estimator

Theorem 4 *The Gâteaux derivative, in the direction of* $\mathbf{V}$*, of the functional* $J_r(\Omega) = -\int_\Omega \log(p(y(\mathbf{x}),\widehat{\theta}_{\mathrm{MO}})d\mathbf{x}$ *with* p *a Rayleigh pdf, is the following:*

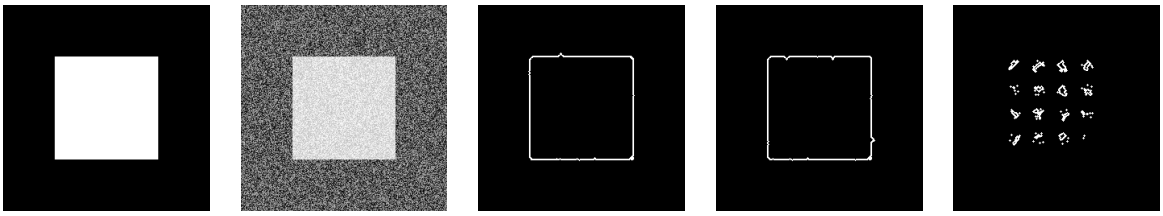
$$< J_r'(\Omega), \mathbf{V} > \ = \ \int_{\partial\Omega}\Big(\log\Big(p(y(\mathbf{x}),\widehat{\theta}_{\mathrm{MO}})\Big) + Add(y(\mathbf{x}),\Omega)\Big)\,(\mathbf{V}\cdot\boldsymbol{N})d\mathbf{a}(\mathbf{x})\ , \tag{24}$$

$$Add(y(x),\Omega) = \Big(2 - \frac{\pi}{4}\frac{\overline{y^2(\mathbf{x})}}{\overline{y(\mathbf{x})}^2}\Big)\Big(1 - \frac{y(\mathbf{x})}{\overline{y(\mathbf{x})}}\Big)\ .$$

The proof can be found in Appendix A.

In this case, it follows that the evolution equation is :

$$\frac{\partial\Gamma(p,\tau)}{\partial\tau} = -\Big[\log\Big(p(y(\mathbf{x}),\widehat{\theta}_{\mathrm{MO}})\Big) + Add(y(\mathbf{x}),\Omega)\Big]\,\mathbf{N}(\mathbf{x})\ . \tag{25}$$



(a) (b) (c) (d) (e)

Figure 1: (a) Original square image, (b) Image corrupted by Rayleigh noise, (c) Final boundary segmentation using MLE, (d) Final boundary segmentation using moment estimator and additive term, (e) Final boundary segmentation using moment estimator with no additive term.

Shape derivative with the MLE

Theorem 5 *The Gâteaux derivative, in the direction of* $\mathbf{V}$*, of the functional* $J_r(\Omega) = -\int_\Omega \log(p(y(x), \widehat{\theta}_{\mathrm{ML}})d\mathbf{x}$ *with* p *a Rayleigh pdf, is the following:*

$$< J_r'(\Omega), \mathbf{V} >= \int_{\partial\Omega} \log\Big(p(y(\mathbf{x}), \widehat{\theta}_{\mathrm{ML}})\Big)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) \ . \tag{26}$$

This theorem is a particular case of Corollary 1, see Section 4.2.

The evolution equation becomes here :

$$\frac{\partial\Gamma(p,\tau)}{\partial\tau} = -\log\Big(p(y(\mathbf{x}), \widehat{\theta}_{\mathrm{ML}})\Big)\mathbf{N}(\mathbf{x}) \ . \tag{27}$$

3.3.2 Epilogue

One can see that the parameter estimator has a clear impact on the evolution speed expression, as the additive term $Add(y(x), \Omega)$ appears in the shape derivative when the moment estimator is used. These theoretical results are confirmed by experimental tests on a synthetic image as depicted in Fig. 1. These tests were carried out using MLE and moments-based theoretical evolution speeds given in (27) and (25), and also using the moments-based speed where the additive term was deliberately dropped.

The segmentation results on this test image are visually similar using either ML or moment estimators. However when the moment additive term is neglected, the contour does not converge to the desired minimum. This confirms the importance of the estimation of the pdf parameters. Of course, in the exponential family, the MLE always exists (except in some pathological cases such as the segmentation of a one pixel region). But, unfortunately, closed-form analytical expressions of the MLE are not always accessible, and one may have to resort to alternative estimators such as the method of moments or cumulants.

4 General Segmentation Setting

We now turn to the general case and consider the following functional.:

$$J_r(\Omega) = \int_\Omega \Phi(p(y(\mathbf{x}), \boldsymbol{\eta}(\Omega)))\, d\mathbf{x}\ \hat{\mathcal{A}}0, \tag{28}$$

where Φ is a continuously differentiable and integrable function, $p(.)$ belongs to the multi-parameter exponential family and is expressed by (16) and $\boldsymbol{\eta}$ the natural parameter vector.

4.1 A general result for the exponential family

In the sequel, for the sake of simplicity, we will invariably denote $\boldsymbol{\eta}$ for the natural parameter and its finite sample estimate over the domain (without a slight abuse of notation, this should be $\hat{\boldsymbol{\eta}}$). We are now ready to state our main result:

Theorem 6 *The Gâteaux derivative, in the direction of* $\mathbf{V}$*, of the functional* (28)*, is:*

$$< J_r'(\Omega), \mathbf{V} >= -\int_{\partial\Omega} \Phi(p(y(\mathbf{x}), \boldsymbol{\eta}(\Omega)))(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x})$$
$$+\int_{\Omega} p(y(\mathbf{x}), \boldsymbol{\eta}(\Omega))\Phi'(p(y(\mathbf{x}), \boldsymbol{\eta}(\Omega)))\langle\nabla_{\mathbf{V}}\boldsymbol{\eta}, \mathbf{T}(y(\mathbf{x})) - \nabla A(\boldsymbol{\eta})\rangle d\mathbf{x}\ , \tag{29}$$

with $\nabla_{\mathbf{V}}\boldsymbol{\eta} =< \boldsymbol{\eta}'(\Omega), \mathbf{V} >$ *the Gâteaux derivative of* $\boldsymbol{\eta}$ *in the direction of* $\mathbf{V}$*, and* $\langle .,. \rangle$ *is the usual scalar product of two vectors.*

The proof is detailed in Appendix B.

This result can be divided into two parts. On the one hand, the term of the evolution of the region, on the other hand, an additive term coming from the dependence of the parameters $\boldsymbol{\eta}$ to the region.

4.2 Specializing to ML

In a finite sample setting, when using the MLE, we can replace $\nabla A(\boldsymbol{\eta})$ by $\overline{\mathbf{T}(y)}$, the 1st order sample moment of $\mathbf{T}(y)$. This is a consequence of Theorem 2. Thus, when using the anti-log-likelihood function, the second term becomes equal to $\int_\Omega \langle \nabla_{\mathbf{V}}\boldsymbol{\eta}, \mathbf{T}\left(y(\mathbf{x})\right) - \overline{\mathbf{T}(y)}\rangle d\mathbf{x}$, and hence vanishes. The following corollary is straightforward:

Corollary 1 *The Gâteaux derivative, in the direction of* $\mathbf{V}$*, of the functional* $J_r(\Omega) = -\int_\Omega \log(p(y(\mathbf{x}), \hat{\boldsymbol{\eta}}_{\mathrm{ML}}(\Omega))d\mathbf{x}$ *where* $\hat{\boldsymbol{\eta}}_{\mathrm{ML}}$ *is the ML estimate, is the following:*

$$< J_r'(\Omega), \mathbf{V} >= \int_{\partial\Omega} (\log\left(p(y(\mathbf{x}), \hat{\boldsymbol{\eta}}_{\mathrm{ML}}(\Omega))\right)(\mathbf{V}\cdot\boldsymbol{N})d\mathbf{a}(\mathbf{x})\ . \tag{30}$$

This provides an alternative and elegant proof to the result of [6, 13]. Nonetheless, we here point out that, in the work of [13], the role of the parameters estimator was not elucidated. Indeed, the evolution speed expression of Corollary 1 is only valid when the MLE is used for $\boldsymbol{\eta}$, otherwise complicated additive terms may appear with others estimators.

A straightforward exercise is to apply (30) to some common members of the exponential family. This is the goal of Table 2 which summarizes the computed speed expressions of some common distributions, when minimizing the anti-log-likelihood and using the MLE for the parameters.

Law	Parameter $\boldsymbol{\eta}$	Speed Expression
Normal	(μ, σ)	$\log(\sigma_{in}) + \frac{(I-\mu_{in})^2}{2\sigma_{in}^2} - \log(\sigma_{out}) - \frac{(I-\mu_{out})^2}{2\sigma_{out}^2}$
Exponential	λ	$-\log(\lambda_{in}) + \lambda_{in}I + \log(\lambda_{out}) - \lambda_{out}I$
Poisson	μ	$-I\log(\mu_{in}) + \mu_{in} + I\log(\mu_{out}) - \mu_{out}$
Rayleigh	θ^2	$\frac{I^2}{2\theta_{in}^2} - \log(\frac{I}{\theta_{in}^2}) - \frac{I^2}{2\theta_{out}^2} + \log(\frac{I}{\theta_{out}^2})$
Gamma	(λ, p)	$\log(\Gamma(p_{in})\lambda_{in}^{p_{in}}) + \frac{I}{\lambda_{in}} - p_{in}\log(I) - \log(\Gamma(p_{out})\lambda_{out}^{p_{out}}) - \frac{I}{\lambda_{out}} + p_{out}\log(I)$

Table 2: Speed expressions of some common members of the exponential family, when minimizing the anti-log-likelihood associated to the MLE for the parameters.

4.3 Beyond ML : Relative Entropy

As far as the segmentation of homogeneous regions is concerned, we may also focus our attention on energy criteria based on information measures [42] such as the relative entropy. The latter is also known as the Kullback-Leibler divergence (KLD). Note that in this case the integrals are over the real line and no more over the domain. In this section, we first give a general result concerning the derivation of the expectation of a function of a pdf. This general result is then applied to obtain the derivative of the KLD.

4.3.1 Derivation of the Expectation of a function of a pdf

Let us introduce the following quantity which designs the expectation over a pdf $p_{in}(y, \boldsymbol{\eta}(\Omega_{in}))$ of a function of the pdf $p_{out}(y, \boldsymbol{\eta}(\Omega_{out}))$:

$$H_Y(\boldsymbol{\eta}_{in}, \boldsymbol{\eta}_{out}) = \mathbb{E}_{in}[\Phi(p_{out}(Y, \boldsymbol{\eta}_{out}))] = \int_{\chi} p_{in}(y, \boldsymbol{\eta}_{in}) \Phi(p_{out}(y, \boldsymbol{\eta}_{out})) dy \ , \tag{31}$$

where χ is the domain of the random variable Y and Φ is a continuously differentiable and integrable function When we consider the pdf of a feature y within a region Ω, the pdf parameters vectors are again region-dependent, that is $H_Y(\boldsymbol{\eta}_{in}, \boldsymbol{\eta}_{out}) = H_Y(\boldsymbol{\eta}_{in}(\Omega_{in}), \boldsymbol{\eta}_{out}(\Omega_{out}))$. Such a quantity appears in many interesting information measures such as the cross entropy (with $\Phi(p) = -log(p)$), the differential entropy (with $in = out$ and $\Phi(p) = -log(p)$), or the KLD which is detailed thereafter. It is then interesting to compute the Gâteaux derivative with respect to the domain of this quantity:

Theorem 7 *The Gâteaux derivative, in the direction $\mathbf{V}$, of the functional $H_Y(\boldsymbol{\eta}_{in}, \boldsymbol{\eta}_{out})$ defined in equation (31) with $p_{in}(y, \boldsymbol{\eta}_{in})$ and $p_{out}(y, \boldsymbol{\eta}_{out})$ two members of the exponential family that belong to the same parametric law, is as follows:*

$$\begin{aligned} < H_Y(\boldsymbol{\eta}_{in}, \boldsymbol{\eta}_{out}), \mathbf{V} > &= < \nabla_{\mathbf{V}} \boldsymbol{\eta}_{out}, \mathbb{E}_{in}\left[p_{out}\, \Phi'(p_{out})(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{out}))\right] > \\ &+ < \nabla_{\mathbf{V}} \boldsymbol{\eta}_{in}, \mathbb{E}_{in}\left[\Phi(p_{out})(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{in}))\right] > \end{aligned}$$

The proof can be found in Appendix C.

4.3.2 Region competition using the KLD

When considering the segmentation of an image into two regions Ω_{in} and Ω_{out}, we propose here to consider the maximization of the relative entropy or KLD defined as follows:

Definition 5 *The relative entropy between the pdf $p_{in}(y, \boldsymbol{\eta}_{in})$ of the feature y within the region Ω_{in} and the pdf $p_{out}(y, \boldsymbol{\eta}_{out})$ of the feature y within the region Ω_{out} is defined as follows:*

$$D(p_{in} \| p_{out}) = \int_{\chi} p_{in}(y, \boldsymbol{\eta}_{in}) \log \frac{p_{in}(y, \boldsymbol{\eta}_{in})}{p_{out}(y, \boldsymbol{\eta}_{out})} dy \ . \tag{32}$$

This quantity is always positive, convex but non-symmetric. It can be expressed using the expectation under the pdf p_{in}, denoted by $\mathbb{E}_{in}$, as follows:

$$D(p_{in} \| p_{out}) = \mathbb{E}_{in}[\log(p_{in}(Y, \boldsymbol{\eta}_{in}))] - \mathbb{E}_{in}[\log(p_{out}(Y, \boldsymbol{\eta}_{out}))] \tag{33}$$

To get the gist of using KLD as a criterion in an RBAC functional, consider the data $y_i = \{y(\mathbf{x}) | \mathbf{x} \in \Omega_i\}$ as an iid sequence from the statistical model $p_i(y, \boldsymbol{\eta}_i)$. Maximizing the KLD between the two pdfs p_{in} and p_{out} can be seen as equivalent to minimizing the log-likelihood score [42]:

$$\frac{1}{|\Omega_{in}|} \int_{\Omega_{in}} \log(p(y(\mathbf{x}), \boldsymbol{\eta}_{out})\, d\mathbf{x} \ ,$$

which tends to $\mathbb{E}_{in}[\log(p_{out}(y,\boldsymbol{\eta}_{out}))]$ using the weak law of large number for a very large domain Ω_{in}. In other words, the KLD-maximization based segmentation criterion will look for the configuration that maximizes the log-likelihood of the data y_i under their actual model p_{in}, while minimizing the plausibility of the same data under p_{out}. Thus, translating this into a segmentation setting, the KLD acts as a region competition criterion.

When the two pdfs belong to the exponential family, we can compute the Gâteaux derivative of $D(p_{in}\|p_{out})$.

Theorem 8 *Suppose that $p_{in}(y,\boldsymbol{\eta}_{in})$ and $p_{out}(y,\boldsymbol{\eta}_{out})$ are two members of the exponential family that follow the same distribution with distinct parameters. The Gâteaux derivative, in the direction $\mathbf{V}$, of the functional $D(p_{in}\|p_{out})$ is:*

$$< D'(p_{in}\|p_{out}), \mathbf{V} >= \langle \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \ddot{A}(\boldsymbol{\eta}_{in})(\boldsymbol{\eta}_{in}-\boldsymbol{\eta}_{out})\rangle + \langle \nabla_{\mathbf{V}}\boldsymbol{\eta}_{out}, \nabla A(\boldsymbol{\eta}_{out}) - \nabla A(\boldsymbol{\eta}_{in})\rangle \ . \tag{34}$$

The proof is given in Appendix D.
We can specialize this result when the parameter $\boldsymbol{\eta}$ is estimated using the ML method for the full rank exponential family and using two complementary domains Ω_{in} and Ω_{out}. The two domains then share the same boundary with normals pointing to opposite direction.

Corollary 2 *The Gâteaux derivative, in the direction $\mathbf{V}$, of the functional $D(p_{in}\|p_{out})$ with $p_{in}(y,\boldsymbol{\eta}_{in})$ and $p_{out}(y,\boldsymbol{\eta}_{out})$ two members of the exponential family with $\boldsymbol{\eta}_{in}$ and $\boldsymbol{\eta}_{out}$ the parameters estimated using the MLE, is as follows:*

$$\begin{aligned} < D'(p_{in}\|p_{out}), \mathbf{V} >= \ & - \ \langle \nabla_{\mathbf{V}}\overline{\mathbf{T}(y)}_{in}, \boldsymbol{\eta}_{in}-\boldsymbol{\eta}_{out}\rangle \\ & + \ \langle \nabla_{\mathbf{V}}\overline{\mathbf{T}(y)}_{out}, [\ddot{A}(\boldsymbol{\eta}_{out})]^{-1}[\nabla A(\boldsymbol{\eta}_{out}) - \nabla A(\boldsymbol{\eta}_{in})]\rangle \ , \end{aligned} \tag{35}$$

where

$$\nabla_{\mathbf{V}}\overline{\mathbf{T}(y)}_{in} = \frac{1}{|\Omega_{in}|}\int_{\partial\Omega_{in}} \left(\overline{\mathbf{T}(y)}_{in} - \mathbf{T}(y(\mathbf{x}))\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) \ , \tag{36}$$

and

$$\nabla_{\mathbf{V}}\overline{\mathbf{T}(y)}_{out} = -\frac{1}{|\Omega_{out}|}\int_{\partial\Omega_{out}} \left(\overline{\mathbf{T}(y)}_{out} - \mathbf{T}(y(\mathbf{x}))\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) \ . \tag{37}$$

The proof is given in Appendix E.

Again, for the reader's convenience, let us take the example of the Rayleigh pdf where:

$$T(y)=y^2, \quad \eta_i = \frac{-1}{\overline{T(y)}} = \frac{-1}{2\overline{\theta_i^2}}, \quad A'(\eta_i) = \frac{-1}{\eta_i}, \quad A''(\eta_i) = \frac{1}{\eta_i^2}, \quad A''(\eta_i)^{-1} = \eta_i^2 \ , \tag{38}$$

where $i = in$ or out. After appropriate substitutions in (35) and some rearrangements, we obtain:

$$\begin{aligned} < D'(p_{in}\|p_{out}), \mathbf{V} > \ = \ & \left(1-\frac{\theta_{in}^2}{\theta_{out}^2}\right)\left(\frac{1}{|\Omega_{in}|}\int_{\partial\Omega_{in}}\left(1-\frac{y^2(\mathbf{x})}{2\theta_{in}^2}\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x})\right) \\ + \ & \left(1-\frac{\theta_{in}^2}{\theta_{out}^2}\right)\left(\frac{1}{|\Omega_{out}|}\int_{\partial\Omega_{out}}\left(1-\frac{y^2(\mathbf{x})}{2\theta_{out}^2}\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x})\right) \end{aligned} \tag{39}$$

In order to maximize the relative entropy between the two pdfs, we can make an active contour evolve using the following velocity:

$$\mathbf{speed}(\mathbf{x},\Omega) = \left(1-\frac{\theta_{in}^2}{\theta_{out}^2}\right)\left(\frac{1}{|\Omega_{in}|}\left(1-\frac{y^2(\mathbf{x})}{2\theta_{in}^2}\right)+\frac{1}{|\Omega_{out}|}\left(1-\frac{y^2(\mathbf{x})}{2\theta_{out}^2}\right)\right) \tag{40}$$

Table 3 summarizes the evolution speed expressions of some common distributions belonging to the exponential family, when maximizing the relative entropy and using the MLE for the parameters.

Law	$\boldsymbol{\eta}$	Speed Expression
Normal	(μ, σ)	$-\frac{1}{\lvert\Omega_{in}\rvert}\left(\left(\frac{1}{2\sigma_{in}^2}-\frac{1}{2\sigma_{out}^2}\right)\left(\frac{\mu_{in}^2}{2\sigma_{in}^2}-\sigma_{in}^2-y^2\right)-\left(\frac{\mu_{in}}{\sigma_{in}^2}-\frac{\mu_{out}}{\sigma_{out}^2}\right)(y+\mu_{in})\right)$ $+\frac{\mu_{in}-\mu_{out}}{2\lvert\Omega_{out}\rvert\sigma_{out}^4}\left(\sigma_{in}^2-\sigma_{out}^2+\mu_{out}-2\mu_{in}\right)\left(\frac{\mu_{out}^2}{2\sigma_{out}^2}-\sigma_{out}^2-y^2\right)$ $-\frac{\mu_{in}-\mu_{out}}{\lvert\Omega_{out}\rvert\sigma_{out}^4}\left(\mu_{out}^2-\sigma_{out}^2\right)\left(\sigma_{in}^2+\sigma_{out}^2+\mu_{out}-\mu_{in}\right)(y+\mu_{out})$
Exponential	λ	$-\frac{1}{\lvert\Omega_{in}\rvert}\left((\frac{1}{\lambda_{in}}-y(\mathbf{x}))(\lambda_{out}-\lambda_{in})\right)+\frac{1}{\lvert\Omega_{out}\rvert}\left((\frac{1}{\lambda_{out}}-y(\mathbf{x}))(\lambda_{out}-\frac{\lambda_{in}}{\lambda_{out}})\right)$
Poisson	μ	$-\frac{1}{\lvert\Omega_{in}\rvert}\left((\mu_{in}-y(\mathbf{x}))(\log(\frac{\mu_{out}}{\mu_{in}}))\right)+\frac{1}{\lvert\Omega_{out}\rvert}\left((\mu_{out}-y(\mathbf{x}))(1-\frac{\mu_{in}}{\mu_{out}})\right)$
Rayleigh	θ^2	$\left(1-\frac{\theta_{in}^2}{\theta_{out}^2}\right)\left(\frac{1}{\lvert\Omega_{in}\rvert}\left(1-\frac{y^2(\mathbf{x})}{2\theta_{in}^2}\right)+\frac{1}{\lvert\Omega_{out}\rvert}\left(1-\frac{y^2(\mathbf{x})}{2\theta_{out}^2}\right)\right)$
Gamma	(λ, p)	$-\frac{1}{\lvert\Omega_{in}\rvert}\left(\left(\frac{p_{in}}{\lambda_{in}}-y\right)(\lambda_{in}-\lambda_{out})-(\psi_0(p_{in})-\log(\lambda_{in}y))(p_{in}-p_{out})\right)$ $+A\left(\frac{p_{in}}{\lambda_{in}}-\frac{p_{out}}{\lambda_{out}}\right)\left(\lambda_{out}^2\psi_1(p_{out})(\frac{p_{out}}{\lambda_{out}}-y)-\lambda_{out}(\psi_0(p_{out})-\log(\lambda_{out}y))\right)$ $+A\left(\psi_0(p_{in})-\psi_0(p_{out})+\log(\frac{\lambda_{in}}{\lambda_{out}})\right)\left(p_{out}(\psi_0(p_{out})-\log(\lambda_{out}y))-\lambda_{out}(\frac{p_{out}}{\lambda_{out}}-y)\right)$ where $A=\frac{1}{\lvert\Omega_{out}\rvert(p_{out}\psi_1(p_{out})-1)}$

Table 3: Speed expressions of some common distributions of the exponential family when maximizing the relative entropy and using the MLE for the parameters.

4.4 Summary

So far, we have introduced mainly three general region descriptors taking into account the statistical properties of the noise, and we have given expressions of the associated evolution speeds. Let us now summarize how these descriptors will be incorporated in our RBAC-based segmentation algorithm. Hereafter, the feature y of the image is chosen to be the intensity $I(\mathbf{x})$ at pixel $\mathbf{x}$.

- The first type of RBAC functional to be minimized can be finally written as:

$$J(\Omega_1,..,\Omega_n,\Gamma)=\sum_{i=1}^{m}\int_{\Omega_i}\Phi(p_i(I(\mathbf{x}),\boldsymbol{\eta}_i))\,d\mathbf{x}+\beta E_b(\Gamma)\ , \tag{41}$$

  where $\Gamma=\bigcup_{i=1}^{m}\partial\Omega_i$, and as in (2), E_b is a regularization term (e.g. curve length) balanced with a positive real parameter β. Here the RBAC segmentation problem aims at finding a partition of the image into $m\geq 2$ regions $\{\Omega_1,..,\Omega_m\}$. This functional is typically chosen with the anti-log-likelihood descriptor where one can assign a specific noise model to each region Ω_i all possibly different. The evolution speed associated to this energy is directly obtained from the results of Section 4.1 or 4.2.

- As explained in 4.3.2, the KLD acts as a region competition criterion, and makes sense only when segmenting in two regions. The criterion to maximize becomes:

$$J(\Omega_{in},\Omega_{out},\Gamma)=\mathbb{E}_{in}[\log(p_{in}(I,\boldsymbol{\eta}_{in}))]-\mathbb{E}_{in}[\log(p_{out}(I,\boldsymbol{\eta}_{out}))]+\beta\,E_b(\Gamma), \tag{42}$$

  and the associated evolution speed is obtained from the results given Section 4.3.2.

5 Experimental results

5.1 Synthetic data

This section presents some experimental results on noisy images. The synthetic noise-free test image is shown in Fig.2.

The first experiment that we report here sheds the light on the role of the noise model on the actual performance of the RBAC segmentation algorithm. For four different Battacharya distances (BD),



Figure 2: Synthetic noise-free test image.

we have systematically corrupted the test image with two types of noise: Poisson and Rayleigh. The Battacharya distance is used as a measure of "contrast" between the objects and the background. It is defined as :

$$\mathcal{D}(p_b(y), p_o(y)) = -\log \int_{\mathcal{X}} \sqrt{p_b(y)p_o(y)}dy \ . \tag{43}$$

with p_b and p_o are respectively the background and the object pdfs.

For each combination of BD value and noise type, 50 noisy images were generated. Each noisy image was then segmented using five different energy functions, namely Chan-Vese [34], and our functional (41) with the anti-log-likelihood score and MLE with four assumed noise models: Gaussian, Rayleigh, Exponential and Poisson. The evolution speeds associated to our noise models were picked up from Table 2. For each segmented image with each method at each BD value, the average false positive fraction (FPF) and true positive fraction (TPF), over the 50 simulations were computed. The bottom-line of the experiment is to show that using the appropriate noise model will lead to the best performance in terms of compromise between specificity (over-segmentation as revealed by the FPF) and sensitivity (under-segmentation as revealed by the TPF).

Fig.4 depicts the average TPF (left) and FPF (right) as a function of the BD for Rayleigh ((a)-(b)) and Poisson ((c)-(d)) noises. As expected, the FPF exhibits a decreasing tendency as the BD increases, while the TPF increases with BD, which is intuitively acceptable. More interestingly, the best performance in terms of compromise between FPF and TPF is reached when the contaminating noise and the noise model in the functional (Table 2) are the same. This behavior is more salient at low BD values, i.e. high noise levels. These quantitative results are confirmed by visual results as portrayed in Fig.3(e)-(h).

We also compared the two RBAC functionals (41) and (42). In this experiment, the noise-free image was corrupted with a Rayleigh noise and 50 noisy images were generated. Each noisy image was segmented using each of the three energy functionals, where $p_{in,out}$ are Rayleigh pdfs and MLE was used for their parameters. The associated speeds are respectively obtained from Table 2 and 3.

The results are reported in Fig.3 and Fig.5. In Fig.3(b), the noisy image looks badly contrasted because of noise predominance. For the reader's convenience and for better visual rendering, the contrast was artificially enhanced as shown in Fig.3(c). The segmentation results are also superimposed over the contrast-enhanced image. By visual inspection of Fig.3(h)-(i), the segmentation results provided by the two functionals are quite comparable. This visual impression was more deeply investigated by a careful quantitative study of the fractions FPF and TPF depicted in Fig.5. Again, TPF and FPF respect the intuitive monotonic tendencies (FPF decreases while TPF increases). The anti-log-likelihood functional seems to be the best followed by the KLD.

5.2 SAR imaging data

Our RBAC segmentation algorithm was applied to a real SAR image shown on Fig.6 (a).

The segmentation results of the original image are depicted in Fig.7. The initial contour (Fig.6 (b)) is composed of many bubbles uniformly spead over the image. We see clearly that the five assumed noise

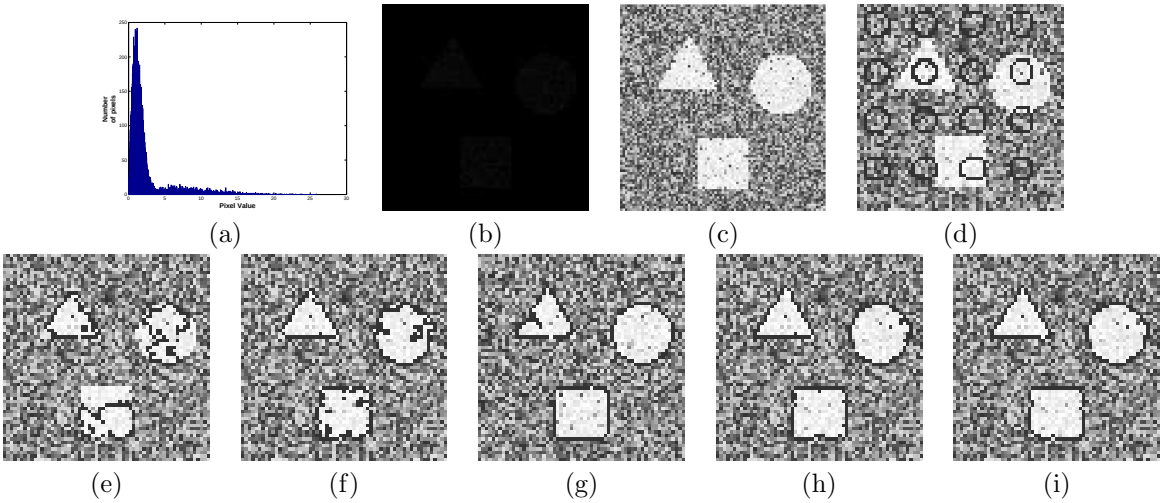


Figure 3: (a) Histogram of the noisy image, (b) Rayleigh-noisy image with original contrast (Battacharya distance is 1.0), (c) Noisy image after contrast enhancement, (d) Initial contour, (e) Final contour with Chan-Vese criterion, (f) Final contour with Poisson criterion and anti-log-likelihood score,(g) Final contour with Exponential criterion and anti-log-likelihood score, (h) Final contour with Rayleigh criterion and anti-log-likelihood score, (i) Final contour with KLD and Rayleigh pdf.

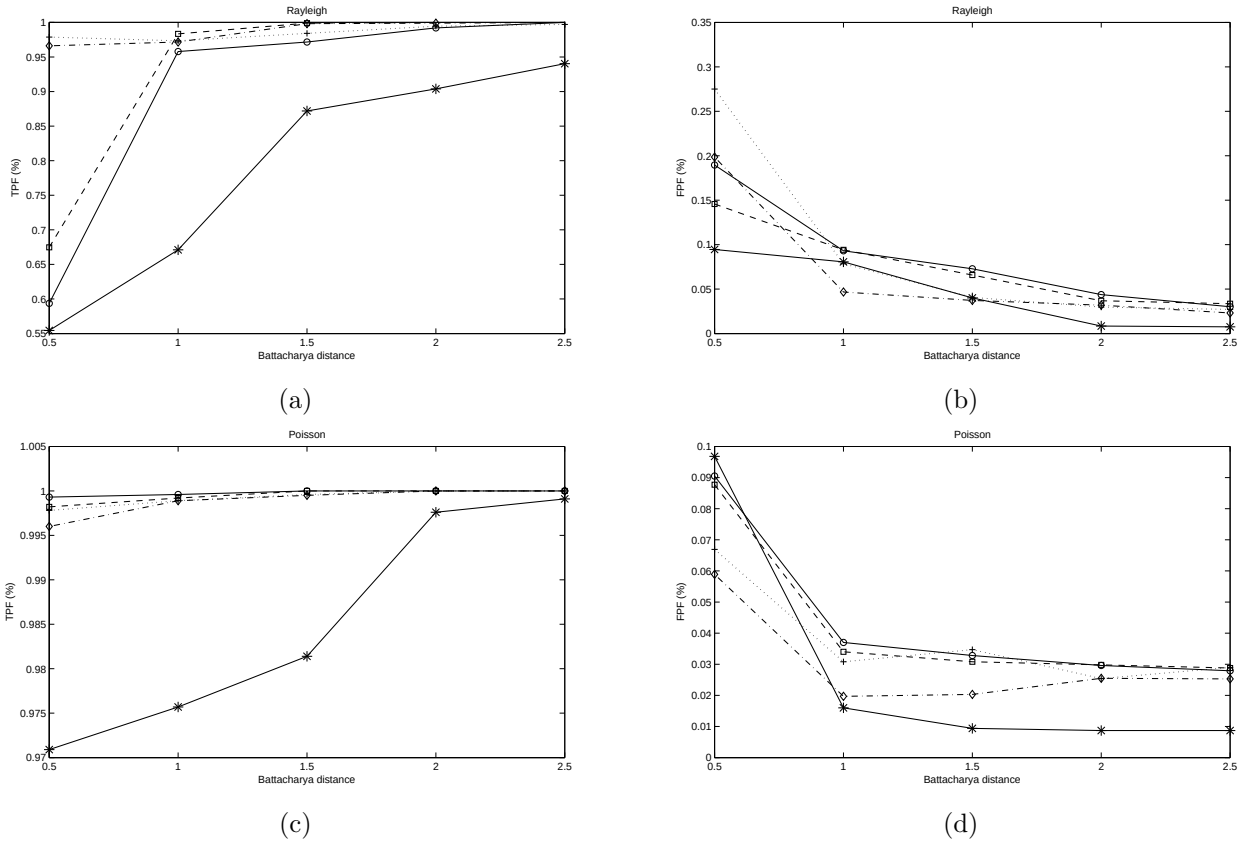
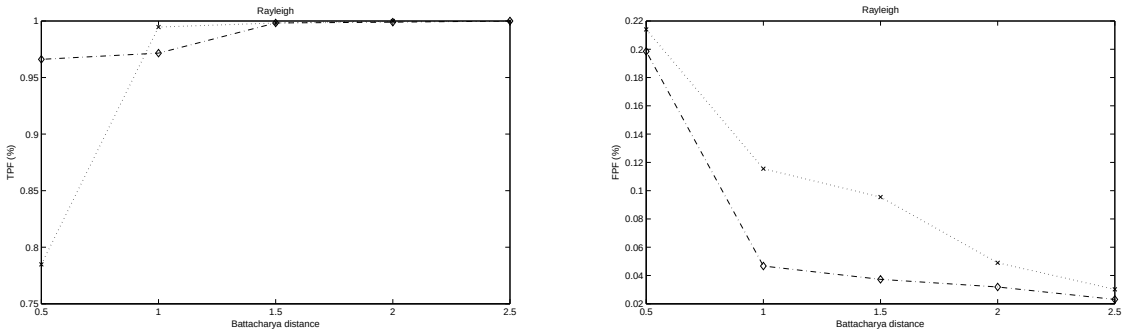
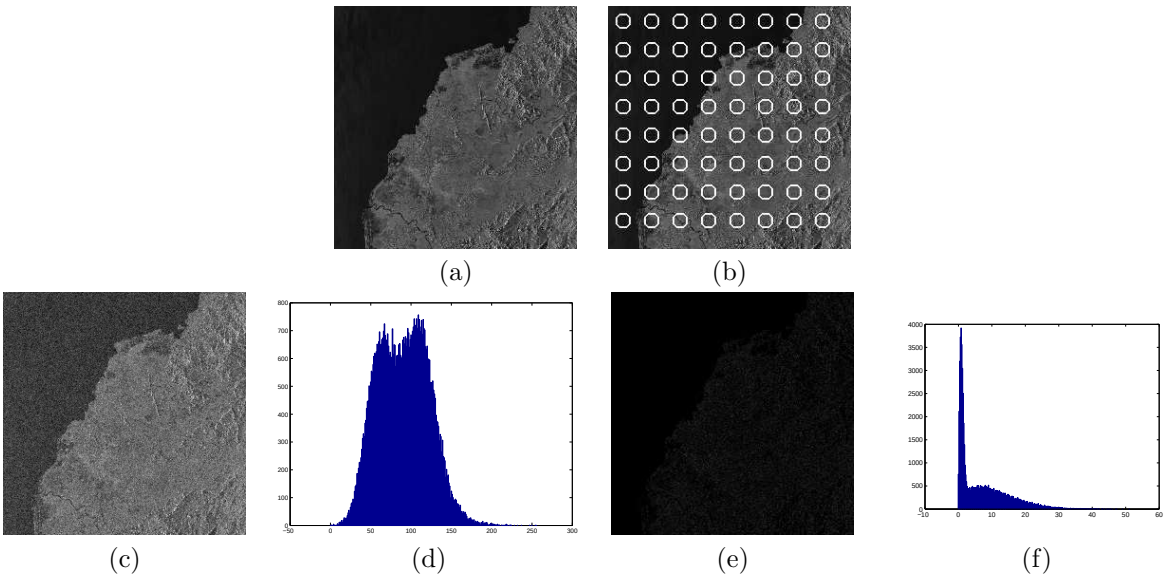


Figure 4: TPF and FPF as a function of BD for the anti-log-likelihood function using five models: Chan-Vese (dotted line with crosses), Gaussian (dashed line with squares), Poisson (solid line with circles), Rayleigh (dash-dot line with diamonds), Exponential (solid line with stars). (a) TPF for Rayleigh noise, (b) FPF for Rayleigh noise, (c) TPF for Poisson noise, (d) FPF for Poisson noise.



(a) (b)

Figure 5: Comparison of FPF and TPF as a function of BD for the anti-log-likelihood (dash-dot line with diamonds) and KLD (dotted line with cross). The noise is Rayleigh-distributed. (a) TPF, (b) FPF.



(a) (b)

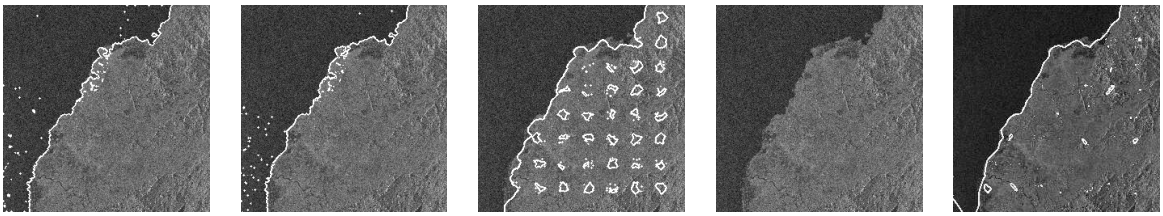
(c) (d) (e) (f)

Figure 6: SAR Images. (a) Original image, (b) Initial contour on original image, (c) Noisy image corrupted by a Gaussian noise, (d) Histogram of the Gaussian noisy image, (e) Noisy image corrupted by a Rayleigh noise (f) Histogram of the Rayleigh noisy image



(a) (b) (c) (d) (e)

Figure 7: Segmentation results of the original image (a) Chan-Vese model, (b) Gaussian model, (c) Poisson model, (d) Rayleigh model, (e) Exponential model. Initialization Fig.6 (b).



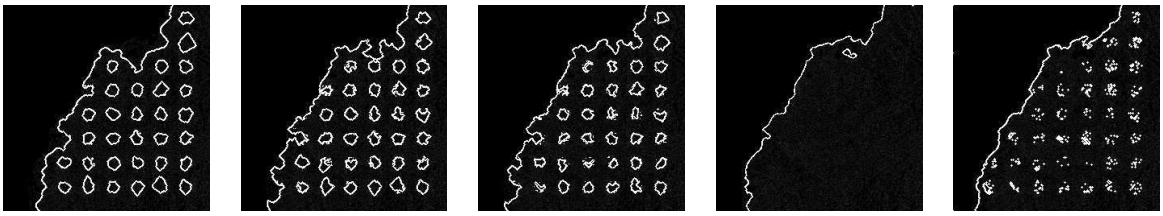
(a) (b) (c) (d) (e)

Figure 8: Segmentation results of the noisy image corrupted by a Gaussian noise (a) Chan-Vese model, (b) Gaussian model, (c) Poisson model, (d) Rayleigh model, (e) Exponential model. Initialization as in Fig.6 (b).

models (Chan-Vese, Gaussian, Poisson, Exponential and Rayleigh) provide rather good segmentation results of the image. This shows the applicability of our method to noise-free images and the stability results of all the different models presented before. Moreover, we see clearly that the segmentation of noise-free images, or equivalently images with a large Battacharya distance, is virtually the same whatever the assumed noise model. This confirms the findings of the synthetic image in the previous section when we quantified the performance using FPF and TPF measures. When the Battacharya distance became high, the segmentation errors were negligible. Of course, here we consider a real image where no ground-truth is available. But, visually, the results look satisfactory.

We performed the same tests the SAR image corrupted by a Gaussian noise with 2.5 as Battacharya distance, see Fig.6 (c). The histogram of this Gaussian noisy image is given in Fig.6 (d). Visually, the image seems to be relatively simple to segment. However, at this noise level, the two Gaussians have a large overlap, and a simple thresholding would be definitely awkward. The results on Fig.8 show the final segmentation using the five models cited just before. Differences are now visible: on the one hand, the two models based on a Gaussian noise (i.e. Chan-Vese Fig.8 (a) and Gaussian (b)) give good and comparable results. On the other hand, the other noise models perform badly except perhaps the Exponential model. For the Gaussian model, this was predictable, because it is precisely the type of noise that contaminates the image. The Chan-Vese model is implicitely a Gaussian noise model where the standard deviation is assumed to be known and constant over the image. As far as the Poisson model is concerned Fig.8 (c), the darkest (i.e. left) part of the image is rather properly segmented while the initial contour (bubbles) did not evolve that much on the brightest (i.e. right) part. One may think that this is an initialization issue. We carried out other tests using many other (random) initializations, and the segmentation was always poor on the brightest part. Our interpretation of this behaviour is that a Poisson random variable tends to be Gaussian when its mean gets very large (i.e. very bright image). But the mean and variance at this asymptotic regime are the same. That is, on the brightest part of the image, the noise would be almost Gaussian but with a highly signal-dependent variance. Finally, using a Rayleigh model Fig.8 (d), the contour does not even converge whatever the initialization is.

We also carried out the same experiment when a Rayleigh noise with Battacharya distance of 2 contaminates the SAR image Fig.6 (e). In this case, manual segmentation is challenging and difficult. The histogram of this Rayleigh noisy image is portrayed in Fig.6 (f). Now, the four models converge (Fig.9) but the Rayleigh model (Fig.9 (d)) is the only one to provide a segmentation result that is consistent with visual expectations.



(a) (b) (c) (d) (e)

Figure 9: Segmentation results of the noisy image corrupted by a Rayleigh noise (a) Chan-Vese model, (b) Gaussian model, (c) Poisson model, (d) Rayleigh model, (e) Exponential model. Initialization as in Fig.6 (b).

6 Conclusion

In this work, we proposed a novel statistical region-based active contours method, where the region descriptor is written as a function Φ of some pdf belonging to the exponential family. The particular case of anti-log-likelihood score is developed and obtained as a special case of our general setting. We also shed light on the influence of parameters' estimation. That is to say, while using likelihood score and MLE (i.e. the more classical case) our results are strictly the same as the literature. But, using another estimation method, (i.e. Moments method), complicating additive terms appear. We also presented other functions Φ that may be interesting candidates in segmentation problems as for example the relative entropy for region competition.

Some experimental results prove the applicability of our method. First, we have illustrated the influence of additive terms using the moments-based estimator and the necessity to take care of the dependence of the pdf parameters to the domain. Secondly, the Monte-Carlo simulations demonstrated that using the appropriate noise model that matches the noise corrupting the images leads to the best segmentation. prove that our model is flexible enough to incorporate other photometric or geometric priors [].

Our ongoing work is directed towards extending the implementation of our approach to $m > 2$ regions by adapting the multiphase method [34]. We also want to add to our noise model a texture model to segment textured images.

A Appendix A : Proof of Theorem 4

Proof : In the case of Rayleigh law, the pdf is:

$$p(y(\mathbf{x}),\theta) = \frac{y(\mathbf{x})}{\theta^2}\exp\left(\frac{-y(\mathbf{x})^2)}{2\theta^2}\right) , \tag{44}$$

and the region criterion becomes:

$$J_r(\Omega) = -\int_\Omega \log y(\mathbf{x})d\mathbf{x} + \frac{1}{2\widehat{\theta}^2_{\rm MO}}\int_\Omega y(\mathbf{x})^2 d\mathbf{x} + |\Omega|\log(\widehat{\theta}^2_{\rm MO}(\Omega)) . \tag{45}$$

We first compute the Gâteaux derivative in the direction of $\mathbf{V}$, of the moment estimator $\widehat{\theta}_{\rm MO}$. We obtain:

$$< \widehat{\theta}'_{\rm MO}(\Omega), \mathbf{V} >= \sqrt{\frac{2}{\pi}}\frac{1}{|\Omega|}\int_{\partial\Omega}(\overline{y} - y(\mathbf{x}))(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) . \tag{46}$$

Using Theorem 1, the Gâteaux derivative of $J_r(\Omega)$ is:

$$\begin{aligned} < J'_r(\Omega), \mathbf{V} > &= \int_{\partial\Omega}\log y(\mathbf{x})(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) - \frac{< \widehat{\theta}_{\rm MO}(\Omega), \mathbf{V} >}{\widehat{\theta}^3_{\rm MO}}\int_\Omega y(\mathbf{x})^2 d\mathbf{x} \\ &- \frac{1}{2\widehat{\theta}^2_{\rm MO}}\int_{\partial\Omega} y(\mathbf{x})^2(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) - \log(\widehat{\theta}^2_{\rm MO})\int_{\partial\Omega}(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) \\ &+ 2|\Omega|\frac{< \widehat{\theta}'_{\rm MO}(\Omega), \mathbf{V} >}{\widehat{\theta}_{\rm MO}} , \end{aligned}$$

which gives:

$$\begin{aligned} < J'_r(\Omega), \mathbf{V} > &= \int_{\partial\Omega}\left(\log\left(p(y(\mathbf{x},\widehat{\theta}_{\rm MO}))\right)\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) \\ &+ \frac{< \widehat{\theta}_{\rm MO}(\Omega), \mathbf{V} >}{\widehat{\theta}^3_{\rm MO}}|\Omega|(2\widehat{\theta}^2_{\rm MO} - \overline{y^2}) , \end{aligned} \tag{47}$$

where $\overline{y^2} = \frac{1}{|\Omega|}\int_\Omega y^2(\mathbf{x})d\mathbf{x}$. We replace $\widehat{\theta}_{\rm MO}$ and its derivative by their expressions which leads to:

$$\begin{aligned} < J'_r(\Omega), \mathbf{V} > &= \int_{\partial\Omega}\left(\log\left(p(y(\mathbf{x}),\widehat{\theta}_{\rm MO}))\right)\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) \\ &+ \frac{\pi}{2}\frac{((4/\pi)\overline{y}^2 - \overline{y^2})}{\overline{y}^3}\int_{\partial\Omega}(\overline{y} - y(\mathbf{x}))(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) . \end{aligned} \tag{48}$$

Thus,

$$\begin{aligned} < J'_r(\Omega), \mathbf{V} > &= \int_{\partial\Omega}\left(\log\left(p(y(\mathbf{x}),\widehat{\theta}_{\rm MO}))\right)\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) \\ &2\left(1 - \frac{\pi}{4}\frac{\overline{y^2}}{\overline{y}^2}\right)\int_{\partial\Omega}\left(1 - \frac{y(\mathbf{x})}{\overline{y}}\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) , \end{aligned} \tag{49}$$

and the result follows.

■

B Appendix B : Proof of Theorem 6

Proof : To compute $< J'_r(\Omega), \mathbf{V} >$, we must first get the derivative of $p(y(\mathbf{x}), \boldsymbol{\eta})$ with respect to the domain, and apply the chain rule to $\Phi(p(y(\mathbf{x}), \boldsymbol{\eta}))$. To simplify the notation we write the Eulerian derivative of $\boldsymbol{\eta}$ as $< \boldsymbol{\eta}'(\Omega), \mathbf{V} >= \nabla_{\mathbf{V}}\boldsymbol{\eta}$.

Using the definition of $p(y,\boldsymbol{\eta})$ given in (16) and the chain rule applied to $A\left(\boldsymbol{\eta}(\Omega)\right)$, we obtain:

$$\begin{aligned} < p'(y,\boldsymbol{\eta}), \mathbf{V} > &= h(y)\left(\langle \nabla_{\mathbf{V}}\boldsymbol{\eta}, \mathbf{T}(y)\rangle - \langle \nabla_{\mathbf{V}}\boldsymbol{\eta}, \nabla A(\boldsymbol{\eta})\rangle\right) e^{\langle \boldsymbol{\eta}(\Omega), \mathbf{T}(y)\rangle - A(\boldsymbol{\eta}(\Omega))} \\ &= p(y,\boldsymbol{\eta})\langle \nabla_{\mathbf{V}}\boldsymbol{\eta}, \mathbf{T}(y) - \nabla A(\boldsymbol{\eta})\rangle \ . \end{aligned} \tag{50}$$

By the chain rule applied to $\Phi(p(y(\mathbf{x}),\boldsymbol{\eta}))$, we get:

$$< \Phi'(p(y,\boldsymbol{\eta})), \mathbf{V} >=< p'(y,\boldsymbol{\eta}), \mathbf{V} > \ \Phi'(p) \ , \tag{51}$$

where $\Phi'(r)$ is the derivative of the function $\Phi(r)$ according to r, which completes the proof.

■

C Appendix C : Proof of Theorem 7

Proof : Using the derivative of the pdf $p(y,\boldsymbol{\eta})$ in the direction $\mathbf{V}$ given in (50), we find :

$$\begin{aligned} < H_Y(\boldsymbol{\eta}_{in},\boldsymbol{\eta}_{out}), \mathbf{V} > &= \mathbb{E}_{in}\left[p_{out}\, \Phi'(p_{out}) < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{out}, \mathbf{T}_{out}(Y) - \nabla A_{out}(\boldsymbol{\eta}_{out}) >\right] \\ &+ \mathbb{E}_{in}\left[\Phi(p_{out}) < \langle \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \mathbf{T}_{in}(Y) - \nabla A_{in}(\boldsymbol{\eta}_{in}) >\right], \end{aligned}$$

If the two pdfs belong to the same parametric law, we have $\mathbf{T}_{in}(y) = \mathbf{T}_{out}(y) = \mathbf{T}(y)$, $A_{in}(\boldsymbol{\eta}) = A_{out}(\boldsymbol{\eta}) = A(\boldsymbol{\eta})$ and $h_{in}(y) = h_{out}(y)) = h(y)$. The derivative then reduces to:

$$\begin{aligned} < H_Y(\boldsymbol{\eta}_{in},\boldsymbol{\eta}_{out}), \mathbf{V} > &= < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{out}, \mathbb{E}_{in}\left[p_{out}\, \Phi'(p_{out})(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{out}))\right] > \\ &+ < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \mathbb{E}_{in}\left[\Phi(p_{out})(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{in}))\right] > \end{aligned}$$

■

D Appendix D : Proof of Theorem 8

Proof : We can develop the relative entropy as follows:

$$D(p_{in}(y,\boldsymbol{\eta}_{in}) \| p_{out}(y,\boldsymbol{\eta}_{out})) = \mathbb{E}_{in}[log(p_{in}(y,\boldsymbol{\eta}_{in})] - \mathbb{E}_{in}[log(p_{out}(y,\boldsymbol{\eta}_{out})] = H_Y(\boldsymbol{\eta}_{in},\boldsymbol{\eta}_{in}) - H_Y(\boldsymbol{\eta}_{in},\boldsymbol{\eta}_{out}) \ .$$

If the two pdfs belong to the same parametric law, we have $\mathbf{T}_{in}(y) = \mathbf{T}_{out}(y) = \mathbf{T}(y)$, $A_{in}(\boldsymbol{\eta}) = A_{out}(\boldsymbol{\eta}) = A(\boldsymbol{\eta})$ and $h_{in}(y) = h_{out}(y)) = h(y)$. Using Theorem 7, we can deduce the following derivative for KLD :

$$\begin{aligned} < D'(p_{in}\|p_{out}), \mathbf{V} > &= < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \mathbb{E}_{in}\left[(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{in}))\right] > \\ &+ < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \mathbb{E}_{in}\left[\log(p_{in})(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{in}))\right] > \\ &- < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{out}, \mathbb{E}_{in}\left[(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{out}))\right] > \\ &- < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \mathbb{E}_{in}\left[\log(p_{out})(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{in}))\right] > \end{aligned}$$

Using Theorem 2, we can replace $\mathbb{E}_{in}[\mathbf{T}(y)]$ by $\nabla A(\boldsymbol{\eta}_{in})$, which yields:

$$\begin{aligned} < D'(p_{in}\|p_{out}), \mathbf{V} > &= < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \mathbb{E}_{in}\left[(\log(p_{in}) - \log(p_{out}))(\mathbf{T}(Y) - \nabla A(\boldsymbol{\eta}_{in}))\right] > \\ &+ < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{out}, \nabla A(\boldsymbol{\eta}_{out}) - \nabla A(\boldsymbol{\eta}_{in})) > \end{aligned}$$

Since p_{in} and p_{out} belong to the same parametric law within the exponential family, we can replace $log(p_{in}) - log(p_{out})$ by $< \boldsymbol{\eta}_{in} - \boldsymbol{\eta}_{out}, \mathbf{T}(y) > -A(\boldsymbol{\eta}_{in}) + A(\boldsymbol{\eta}_{out})$ and then rearranging the terms we find:

$$\begin{aligned} < D'(p_{in}\|p_{out}), \mathbf{V} > &= < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \mathbb{E}_{in}\left[< \boldsymbol{\eta}_{in} - \boldsymbol{\eta}_{out}, \mathbf{T}(Y) > (\mathbf{T}(Y) - \mathbb{E}_{in}[\mathbf{T}(y)])\right] > \\ &+ < \nabla_{\mathbf{V}}\boldsymbol{\eta}_{out}, \nabla A(\boldsymbol{\eta}_{out}) - \nabla A(\boldsymbol{\eta}_{in})) > \end{aligned}$$

Using the fact that $\ddot{A}(\boldsymbol{\eta}) = Cov[\mathbf{T}(Y)]$ and then that $\ddot{A}(\boldsymbol{\eta})_{ij} = \mathbb{E}[T_i(Y)T_j(Y)] - \mathbb{E}[T_i(Y)]\mathbb{E}[T_j(Y)]$, we find:

$$< D'(p_{in}\|p_{out}), \mathbf{V} >= \langle \nabla_{\mathbf{V}}\boldsymbol{\eta}_{in}, \ddot{A}(\boldsymbol{\eta}_{in})(\boldsymbol{\eta}_{in} - \boldsymbol{\eta}_{out})\rangle + \langle \nabla_{\mathbf{V}}\boldsymbol{\eta}_{out}, \nabla A(\boldsymbol{\eta}_{out}) - \nabla A(\boldsymbol{\eta}_{in})\rangle \ , \tag{52}$$

which is the desired result.

■

E Appendix E : Proof of Corollary 2

Proof : As discussed after Theorem 3, the parameter $\boldsymbol{\eta}$ for the exponential family can be uniquely parameterized as a function of $\psi\left(\mathbb{E}[\mathbf{T}(Y)]\right)$. When using the MLE, the term $\mathbb{E}[\mathbf{T}(Y)]$ can be empirically estimated with $\overline{\mathbf{T}(y)}$ and so derived easily with respect to the domain Ω. We propose to directly derive the expression $\nabla A(\boldsymbol{\eta}) = \overline{\mathbf{T}(y)}$. This expression can be written as:

$$\frac{\partial A}{\partial \eta_i}(\boldsymbol{\eta}) = \overline{\mathbf{T}_i(y)} \quad \forall i \in [1, \kappa] \ . \tag{53}$$

We can then compute the shape derivative of this expression, which gives:

$$\sum_{j=1}^{\kappa} \langle \eta_j', \mathbf{V} \rangle \frac{\partial^2 A}{\partial \eta_i \partial \eta_j}(\boldsymbol{\eta}) = \langle \overline{\mathbf{T}_i(y)}', \mathbf{V} \rangle \quad \forall i \in [1, \kappa] \ , \tag{54}$$

which can be written in the compact form:

$$\nabla_{\mathbf{V}}(\overline{\mathbf{T}}) = \ddot{A}(\boldsymbol{\eta}) \nabla_{\mathbf{V}} \boldsymbol{\eta} \ , \tag{55}$$

where $\nabla_{\mathbf{V}} \overline{\mathbf{T}(y)} = \Big(< \overline{\mathbf{T}_1(y)}', \mathbf{V} >, < \overline{\mathbf{T}_2(y)}', \mathbf{V} >, .. < \overline{\mathbf{T}_\kappa(y)}', \mathbf{V} > \Big)^T$.
Restricting our study to the full rank exponential family, where $\ddot{A}(\boldsymbol{\eta})$ is a symmetric positive-definite, hence invertible, matrix (Theorem 3), the domain derivative of the parameters $\boldsymbol{\eta}$ is uniquely determined by:

$$\ddot{A}(\boldsymbol{\eta})^{-1} \nabla_{\mathbf{V}}(\overline{\mathbf{T}}) = \nabla_{\mathbf{V}} \boldsymbol{\eta} \ . \tag{56}$$

This equation holds for $\boldsymbol{\eta}_{in}$ and $\boldsymbol{\eta}_{out}$ and thus, the domain derivative of $D(p_{in} \| p_{out})$ can be expressed as:

$$\begin{aligned} < D'(p_{in} \| p_{out}), \mathbf{V} > &= \langle \ddot{A}(\boldsymbol{\eta}_{in})^{-1} \nabla_{\mathbf{V}}(\overline{\mathbf{T}}_{in}), \ddot{A}(\boldsymbol{\eta}_{in})(\boldsymbol{\eta}_{in} - \boldsymbol{\eta}_{out}) \rangle \\ &+ \langle \ddot{A}(\boldsymbol{\eta}_{out})^{-1} \nabla_{\mathbf{V}}(\overline{\mathbf{T}}_{out}), \nabla A(\boldsymbol{\eta}_{out}) - \nabla A(\boldsymbol{\eta}_{in}) \rangle \ . \end{aligned} \tag{57}$$

The corollary follows after simplification using the fact that the matrix $\ddot{A}(\boldsymbol{\eta}_i)^{-1}$ is symmetric.
The domain derivative $\nabla_{\mathbf{V}}(\overline{\mathbf{T}}_{in})$ is computed using Theorem 1 and is given by

$$\nabla_{\mathbf{V}}(\overline{\mathbf{T}}_{in}) = \frac{1}{|\Omega_{in}|} \int_{\partial\Omega} \Big(\overline{\mathbf{T}(y)}_{in} - \mathbf{T}(y(\mathbf{a})) \Big) (\mathbf{V} \cdot \mathbf{N}) d\mathbf{a}(\mathbf{x}) \ , \tag{58}$$

The domain Ω_{out} is the complement of the domain Ω_{in}. They share the same boundary with normals pointing to opposite direction. Similar results can then be found when considering the domain derivative of $\overline{\mathbf{T}}_{out}$:

$$\nabla_{\mathbf{V}}(\overline{\mathbf{T}}_{out}) = -\frac{1}{|\Omega_{out}|} \int_{\partial\Omega_{out}} \Big(\overline{\mathbf{T}(y)}_{out} - \mathbf{T}(y(\mathbf{a})) \Big) (\mathbf{V} \cdot \mathbf{N}) d\mathbf{a}(\mathbf{x}) \ . \tag{59}$$

■

References

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes : Active contour models," International Journal of Computer Vision, vol. 1, pp. 321–332, 1988.

[2] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," International Journal of Computer Vision, vol. 22, no. 1, pp. 61–79, 1997.

[3] L. D. Cohen, "On active contour models and balloons," Computer Vision, Graphics, and Image Processing. Image Understanding, vol. 53, no. 2, pp. 211–218, 1991.

[4] L. Cohen, E. Bardinet, and N. Ayache, "Surface reconstruction using active contour models," in SPIE Conference on Geometric Methods in Computer Vision, San Diego, 1993.

[5] R. Ronfard, "Region-based strategies for active contour models," International Journal of Computer Vision, vol. 13, no. 2, pp. 229–251, 1994.

[6] S. Zhu, T.S. Lee, and A. Yuille, "Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation," in International Conference on Computer Vision, 1995, pp. 416–423.

[7] S. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, pp. 884–900, September 1996.

[8] A. Chakraborty, L. Staib, and J. Duncan, "Deformable boundary finding in medical images by integrating gradient and region information," IEEE Transactions on Medical Imaging, vol. 15, pp. 859–870, December 1996.

[9] N. Paragios and R. Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," International Journal of Computer Vision, vol. 46, no. 3, pp. 223, 2002.

[10] J.-F. Aujol, G. Aubert, and L. Blanc-Féraud, "Wavelet-based level set evolution for classification of textured images," IEEE Transactions on Image Processing, vol. 12, no. 12, pp. 1634–1641, 2003.

[11] M. Rousson, C. Lenglet, and R. Deriche, "Level set and region based surface propagation for diffusion tensor MRI segmentation," in Computer Vision Approaches to Medical Image Analysis (CVAMIA) and Mathematical Methods in Biomedical Image Analysis (MMBIA) Workshop, 2004.

[12] I. Karoui, R. Fablet, J. M. Boucher, and J. M. Augustin, "Region-based image segmentation using texture statistics and level-set methods," in ICASSP, 2006, vol. 2, pp. 693–696.

[13] P. Martin, P. Réfrégier, F. Goudail, and F. Guérault, "Influence of the noise model on level set active contour segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, pp. 799–803, 2004.

[14] F. Lecellier, S. Jehan-Besson, J. Fadili, G. Aubert, and M. Revenu, "Statistical region-based active contours with exponential family observations," in ICASSP, 2006, vol. 2, pp. 113–116.

[15] F. Galland, N. Bertaux, and P. Réfrégier, "Multi-component image segmentation in homogeneous regions based on description length minimization: Application to speckle, Poisson and Bernoulli noise," Pattern Recognition, vol. 38, pp. 1926–1936, 2005.

[16] M. Leventon, Statistical Models for Medical Image Analysis, Ph.D. thesis, MIT, 2000.

[17] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr, "Diffusion snakes : Introducing statistical shape knowledge into the Mumford-Shah functional," IJCV, vol. 50, pp. 295–313, 2002.

[18] A. Tsai, A. Yezzi, and W. Wells, "A shape-based approach to the segmentation of medical imagery using level sets," IEEE Transactions on Medical Imaging, vol. 22, pp. 137–154, 2003.

[19] M. Gastaud, M. Barlaud, and G. Aubert, "Tracking video objects using active contours and geometric priors," in IEEE 4th E.W.I.A.M.I.S., April 2003, pp. 170–175.

[20] A. Foulonneau, P. Charbonnier, and F. Heitz, "Geometric shape priors for region-based active contours.," in ICIP, 2003.

[21] F. Lecellier, S. Jehan-Besson, J. Fadili, G. Aubert, M. Revenu, and E. Saloux, "Region-based active contours with noise and shape priors," in ICIP, 2006, vol. 1, pp. 1649–1652.

[22] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape," International Journal of Computer Vision, vol. 72, no. 2, pp. 195–215, April 2007.

[23] P. Y. Lau and S. Ozawa, "A region-based approach combining marker-controlled active contour model and morphological operator for image segmentation," in IEEE EMBS, 2004, pp. 165–1655.

[24] L. Cheng, J. Yang, and X. Fan, "A new region-based active contour for object extraction using level set method," Pattern Recognition and Image Analysis, pp. 285–291, 2005.

[25] F. Precioso, M. Barlaud, T. Blu, and M. Unser, "Robust real-time segmentation of images and videos using a smooth-spline snake-based algorithm," IEEE Transactions on Image Processing, vol. 14, pp. 910–924, 2005.

[26] S. Jehan-Besson, M. Barlaud, and G. Aubert, "Video object segmentation using eulerian region-based active contours," in International Conference on Computer Vision, Vancouver, Canada, 2001.

[27] S. Jehan-Besson, M. Barlaud, and G. Aubert, "DREAM^{2}S: Deformable regions driven by an eulerian accurate minimization method for image and video segmentation," International Journal of Computer Vision, , no. 53, pp. 45–70, 2003.

[28] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson, "Image segmentation using active contours: Calculus of variations or shape gradients ?," SIAM Applied Mathematics, vol. 63, no. 6, pp. 2128–2154, 2003.

[29] J. Sokolowski and J. P. Zolésio, Introduction to shape optimization, vol. 16 of Springer series in computational mathematics, Springer-Verlag, 1992.

[30] M. C. Delfour and J. P. Zolésio, Shape and geometries, Advances in Design and Control SIAM, 2001.

[31] N. Paragios and R. Deriche, "Coupled geodesic active regions for image segmentation: A level set approach," in European Conference in Computer Vision, Dublin, Ireland, June 2000.

[32] N. Paragios and R. Deriche, "Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision," Journal of Visual Communication and Image Representation, vol. 13, pp. 249–268, 2002.

[33] P. Martin, P. Réfrégier, F. Goudail, and F. Guérault, "Influence of the noise model on level set active contour segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 6, pp. 799–803, june 2004.

[34] L. Vese and T. Chan, "A multiphase level set framework for image segmentation using the mumford and shah model," International J. of Comp. Vision, vol. 50, no. 1, pp. 271–293, 2002.

[35] J. W. Goodman, "Some fundamental properties of speckle," J. of Optical Society of America, vol. 66, pp. 1145–1150, 1976.

[36] I. Dydenko, D. Friboulet, and I. Magnin, "A variational framework for affine registration and segmentation with shape prior : Application in echocardiographic imaging," IEEE Workshop on V.G.L.S.M. in Computer Vision, vol. 1, pp. 201–208, 2003.

[37] G. Slabaugh, G. Unal, T. Fang, and M. Wels, "Ultrasound-specific segmentation via decorrelation and statistical region-based active contours," in International Conference on Computer Vision and Pattern Recognition, 2006, vol. 1, pp. 45–53.

[38] C. Chesnaud, P. Réfrégier, and V. Boulet, "Statistical region snake-based segmentation adapted to different physical noise models.," IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 11, pp. 1145–1157, 1999.

[39] P. J. Bickel and K. A. Docksum, Mathematical statistics: basic ideas and selected topics, vol. I, Prentice-Hall, London, 2nd edition, 2001, ISBN 013850363-X(v. 1).

[40] P. O. Koopman, "On distributions admitting a sufficient statistic," Trans. Am. Math. Soc., vol. 39, pp. 399–409, 1936.

[41] A. Banerjee, I. Dhillon, J. Ghosh, and S. Merugu, "An information theoretic analysis of maximum likelihood mixture estimation for exponential families," in ICML, 2004, pp. 57–64.

[42] T. M. Cover and J. A. Thomas, Elements of Information Theory, New York: Wiley, 1991.