# ALPHA-DIVERGENCE MAXIMIZATION FOR STATISTICAL REGION-BASED ACTIVE CONTOUR SEGMENTATION WITH NON-PARAMETRIC PDF ESTIMATIONS

Leila Meziou, Aymeric Histace

ETIS - UMR 8051 - CNRS ENSEA - Cergy-Pontoise University, France {leila.meziou,aymeric.histace}@ensea.fr

## ABSTRACT

In this article, a complete original framework for unsupervised statistical region-based active contour segmentation is proposed. More precisely, the method is based on the maximization of alphadivergences between non-paramterically estimated probability density functions (PDFs) of the inner and outer regions defined by the evolving curve. We define the variational context associated to distance maximization in the particular case of alpha-divergences and provide the complete derivation of the partial differential equation leading the segmentation. Results on synthetic data, corrupted with a high level of Gaussian and Poisson noises, but also on clinical X-ray images show that the proposed unsupervised method improves standard approaches of that kind.

*Index Terms*— Image segmentation, active contours, distance maximization, probability density function, alpha-divergences.

## 1. INTRODUCTION

Originally proposed in [1], the basic idea of active contour segmentation is to iteratively evolve an initial curve towards the boundaries of target objects. The evolution equation of the curve is generally derived from a variational principle in which the energy functional is optimized thanks to the combination of external forces, induced from the image (intensity, texture...) and internal forces determined by the geometry of the evolving curve (shape prior of the target object). In the particular framework of statistical region-based active contour segmentation, external forces involve optimization of distance (or divergence) between probability density functions (PDFs) characterizing the inner  $(\Omega_{in})$  and the outer  $(\Omega_{out})$  regions delimited by the boundaries of the active curve. Depending on the formulation of the problem, minimization or maximization, the segmentation process completely differs: the minimization formulation leads to a supervised segmentation problem since the objective is to minimize the distance between the current PDF of active curve inner (resp. outer) region and an inner (resp. outer) reference PDF; the maximization formulation does not require any reference since the objective is to maximize at each iteration the distance between current inner  $(p_{in})$ and outer  $(p_{out})$  PDFs.

Whatever the optimization formulation is, main key points of statistical region-based active contour segmentation approaches are: the distance used for comparing PDFs, and the way the different PDFs are estimated (parametrically or not). Considering the minimization approach, standard proposed distances of the literature are the  $\chi_2$  distance, the Kullback-Leibler divergence (KL) or the Hellinger distance [2, 3, 4, 5]. Recently, we have proposed [6] to minimize alpha-divergence criterion with non-parametric estima-

Frédéric Precioso

I3S - UMR 6070 - CNRS Nice Sophia Antipolis University, France precioso@polytech.unice.fr

tion of the PDFs using Parzen window technic [7]. We showed that our criterion outperforms existing distance criteria. In the context of maximization formulation, only few works have dealt with these problems: In [8],  $p_{in}$  and  $p_{out}$  are parametrically estimated and compared using a maximization of the Bhattacharaya distance (which is directly linked to Hellinger distance), and in [5], maximization of the Kullback-Leibler divergence is considered between parametrically estimated  $p_{in}$  and  $p_{out}$  PDFs. In this article, we propose to consider the maximization problem of the alpha-divergence criterion between non-parametrically estimated PDFs ( $p_{in}$  and  $p_{out}$ ) of the inner and outer regions defined by the active curve in order to extend our previous work [6] to unsupervised context.

The remaining of this paper is organized as follows: section 2 is focused on the theoretical part of this work: First the distance maximization mathematical framework is remained and considering a non-parametric estimation of  $p_{in}$  and  $p_{out}$ , the complete original calculations that lead to the corresponding partial differential equation (PDE) are exposed. Second, the specific case of alpha-divergences is considered. In section 3, evaluation of the method is proposed. Segmentation results obtained on synthetic noisy images are first compared with performances of standard distances. Finally experimental results obtained with segmentation of X-ray images using alpha-divergence maximization are compared to the standard KL divergence and Hellinger distance.

# 2. ALPHA-DIVERGENCE MAXIMIZATION BETWEEN NON PARAMETRIC PDF FOR IMAGE SEGMENTATION

#### 2.1. Derivation of the general PDE

As said in the previous section, the functional from which is derived the PDE driving the active curve evolution is based on the optimization of a distance criterion between the PDFs of the inner region  $p_{in}$ and the outer region  $p_{out}$  delimited by the active curve. From the maximization point of view, this distance is defined as follows:

$$D(p_{in} || p_{out}, \Omega) = \int_{\Re^m} \varphi(p_{in}, p_{out}, \lambda) \, d\lambda \,, \tag{1}$$

where  $\varphi$  is a cost function related to the maximized distance Dand  $\Omega$  the image domain.  $p_{in}$  and  $p_{out}$  are normalized histograms such as  $p_i(\lambda): \Re^m \to [0, 1]$  represents probability distribution of pixel intensity  $\lambda$  in the image. In this article, m = 1 since we will only consider grayscale images and PDFs are *non-parametrically* estimated at each iteration of the segmentation process using Parzen Window:

$$p_i(\lambda) = \frac{1}{|\Omega_i|} \int_{\Omega_i} g_\sigma(\mathbf{I}(\mathbf{x}) - \lambda) \, d\mathbf{x}, \text{ where } i = \{in, out\}, \quad (2)$$

with  $g_{\sigma}$  is the Gaussian kernel of standard deviation  $\sigma$  (see [4] for optimal choice of  $\sigma$ ) used in the Parzen window estimation,  $\mathbf{I}(\mathbf{x})$  is the intensity function of the segmented image at a given pixel  $\mathbf{x}$  of the image,  $\Omega_{in}$  is the region inside the active curve  $\Gamma$  and  $\Omega_{out}$  the region outside.

Considering the maximization of Eq. (1), the corresponding general PDE is usually deduced from the Euler Derivative of D given by :

$$< D'(\Omega), \mathbf{V} > = -\int_{\partial\Omega} \mathbf{v}(\mathbf{x}, \Omega) < \mathbf{V} \cdot \mathbf{N} > da(\mathbf{x}),$$
 (3)

where  $\partial\Omega$  is the boundary of the region  $\Omega$  and da an area element of  $\Omega$ , with **N** the inner normal vector of the curve and **v** the velocity of the curve.

In our case, Eq. (3) then becomes:

$$< D'(\Omega), \mathbf{V} > = dD(p_{in} || p_{out}, \Omega, \mathbf{V})$$

$$= \int_{\mathfrak{B}} d\varphi(p_{in}, p_{out}, \lambda, \mathbf{V}) d\lambda.$$

$$(4)$$

The problem is now then shifted to the calculation of the Euler derivative of the  $\varphi$  function. To achieve this, let us introduce f the function such as:

$$\varphi(p_{in}, p_{out}, \lambda) = \varphi\left(\frac{G_{1,in}}{G_{2,in}}, \frac{G_{1,out}}{G_{2,out}}, \lambda\right)$$
$$= f(G_{1,in}, G_{2,in}, G_{1,out}, G_{2,out}, \lambda), \quad (5)$$

with 
$$G_{1,i}(\lambda, \Omega_i) = \int_{\Omega_i} g_\sigma(\mathbf{I}(\mathbf{x}) - \lambda) d\mathbf{x}$$
  
and  $G_{2,i}(\Omega_i) = |\Omega_i| = \int_{\Omega_i} d\mathbf{x}.$  (6)

From Eq. (5) and Eq. (6), we can then deduce that:

$$d\varphi(p_{in}, p_{out}, \lambda, \mathbf{V}) = df(p_{in}, p_{out}, \lambda, \mathbf{V})$$
  
$$= \sum_{i=\{in, out\}} \frac{\partial f}{\partial G_{1,i}} dG_{1,i}(\lambda, \Omega_i, \mathbf{V})$$
  
$$+ \sum_{i=\{in, out\}} \frac{\partial f}{\partial G_{2,i}} dG_{2,i}(\Omega_i, \mathbf{V}).$$

Since the function  $g_{\sigma}(\mathbf{I}(\mathbf{x}) - \lambda)$  does not depend on the region  $\Omega_i$ , we have:

$$dG_{1,i}(\lambda,\Omega_i,\mathbf{V}) = -\int_{\partial\Omega_i} g_{\sigma}(\mathbf{I}(\mathbf{x}) - \lambda) < \mathbf{V} \cdot \mathbf{N} > da(\mathbf{x}),$$
  
$$dG_{2,i}(\Omega_i,\mathbf{V}) = -\int_{\partial\Omega_i} < \mathbf{V} \cdot \mathbf{N} > da(\mathbf{x}),$$

and partial derivatives of f are given by:

$$\frac{\partial f}{\partial G_{1,i}} = \frac{1}{|\Omega_i|} \partial_k \varphi(p_{in}, p_{out}, \lambda)$$
$$\frac{\partial f}{\partial G_{2,i}} = -\frac{p_i}{|\Omega_i|} \partial_k \varphi(p_{in}, p_{out}, \lambda),$$

where  $\{i, k\} = \{\{in, 1\}, \{out, 2\}\}.$ 

where  $\partial_1 \varphi$  and  $\partial_2 \varphi$  are the derivatives of  $\varphi$  with respect to the first  $(p_{in})$  and the second  $(p_{out})$  variables.

Merging all those intermediate calculations and noticing that by convention, the curve  $\Gamma = \partial \Omega_{in} = -\partial \Omega_{out}$ , the Euler derivative of the maximized functional D becomes :

$$dD(p_{in} || p_{out}, \Omega, \mathbf{V}) = \int_{\Gamma} \left( \frac{-1}{|\Omega_{in}|} (A_1 - C_1) \right)$$
(7)  
+ 
$$\frac{1}{|\Omega_{out}|} (A_2 - C_2) < \mathbf{V} \cdot \mathbf{N} > da(\mathbf{x}),$$
  
with 
$$A_k = \partial_k \varphi(p_{in}, p_{out}, \lambda) * g_{\sigma}(\mathbf{I}(\mathbf{x}))$$

$$C_k = \int_{\Re} \partial_k \varphi(p_{in}, p_{out}, \lambda) p_i d\lambda,$$
  
where  $\{i, k\} = \{\{in, 1\}, \{out, 2\}\}.$ 

Finally, the PDE corresponding to the maximization of a distance D between two non-parametrically estimated PDFs is obtained thanks to the Gateaux derivative gradient flow:

$$\frac{\partial \Gamma}{\partial t} = \left[\frac{1}{|\Omega_{in}|}(A_1 - C_1) - \frac{1}{|\Omega_{out}|}(A_2 - C_2)\right]\mathbf{N}.$$
 (8)

## 2.2. Alpha-divergence distance

In [6], we introduced alpha-divergence measure [9, 10] as a distance criterion for statistical region-based active contour segmentation. More precisely, from a minimization perspective (supervised approach), we showed the adaptability of this parameter to some very different contexts of noise. In the case of grayscale images, the energy functional  $D_{\alpha}$  related to this particular divergence could be defined using Eq. (1) with the cost function given by:

$$\varphi_{\alpha}(p_{in}, p_{out}, \lambda) = \frac{1}{\alpha(1-\alpha)} \Big( \alpha p_{in}(\lambda) + (1-\alpha) p_{out}(\lambda) \\ - [p_{in}(\lambda)]^{\alpha} [p_{out}(\lambda)]^{1-\alpha} \Big),$$
(9)

where  $\alpha \in \Re$ .

If a complete study about the mathematical properties of alphadivergences can be found in [11], let us highlight that for specific values of  $\alpha$ , some aforementioned standard distances can be connected to alpha-divergences. For instance:  $D_2(\Omega) = \frac{1}{2}D_{\chi^2}(\Omega)$ ,  $D_{\frac{1}{2}}(\Omega) = 2D_{Hellinger}(\Omega), D_{KL}(\Omega) = \lim_{\alpha \to 1} D_{\alpha}(\Omega)$ . This makes alpha-divergence a generic distance estimation, with multiple tuning possibilities via  $\alpha$  parameter and as a consequence, a very flexible measure.

In the context of maximization, in order to properly define the corresponding PDE of Eq. (9) for unsupervised segmentation, we calculate first and second derivatives of corresponding  $\varphi_{\alpha}$  function with respect to  $p_{in}$  and  $p_{out}$ :

$$\partial_{1}\varphi_{\alpha}(p_{in}, p_{out}, \lambda) = \frac{1}{1 - \alpha} \left( 1 - \left[ \frac{p_{out}}{p_{in}}(\lambda) \right]^{1 - \alpha} \right)$$
  
$$\partial_{2}\varphi_{\alpha}(p_{in}, p_{out}, \lambda) = \frac{1}{\alpha} \left( 1 - \left[ \frac{p_{in}}{p_{out}}(\lambda) \right]^{\alpha} \right),$$
  
(10)

which completely defines the iterative process of segmentation.

#### **3. EXPERIMENTS**

In order to be able to segment images presenting more than one target object, we propose to embed the alpha-divergence maximization within the now usual level-set framework [12, 13]. In this framework, considering the standard level-set embedding function  $\phi : \Re^2 \times \Re^+ \to \Re$  and preliminary calculations given by Eq. (8), the following evolution PDE is obtained:

$$\frac{\partial \phi}{\partial t} = \delta \phi \left( \beta \nabla \cdot \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - \xi \left( \frac{1}{|\Omega_{in}|} (A_1 - C_1) + \frac{1}{|\Omega_{out}|} (A_2 - C_2) \right) \right),$$
(11)

where  $A_1, A_2, C_1$  and  $C_2$  are taken from Eq. (8) and Eq. (10),  $\beta$  and  $\xi$  positive weighting parameters and  $\nabla$  the gradient operator. The first term of Eq. (11) consists in a regularization constrain on the total length of the final segmentation and second and third terms are related to the iterative maximization of the alpha-divergences between  $p_{in}$  and  $p_{out}$  (see Eq. (8) for corresponding general PDE). Practically speaking, the implementation of Eq. (11) is achieved with a semi-implicit version of the Additive Operator Splitting scheme first introduced in [14] and successfully used in [15].

## 3.1. Segmentation of noisy synthetic images

In order to evaluate the performance of the proposed method based on maximization of alpha-divergences, we first propose to achieve the segmentation of synthetic images corrupted by various types of noises (see Fig. 1 for illustation). Mainly, corrupting noises considered here are zero-mean Gaussian and Poisson ones: The Gaussian noise is a standard one in the majority of acquisition systems, and the Poisson distribution will model the corrupting process of X-Ray imaging system that will be studied in next section. Moreover, in order to highlight the benefit from the level-set implementation of Eq. (11), the synthetic image presents two objects to segment. Finally, the initialization of the active curve is a set of little circles regularly dispatched on the whole image which allows not to consider a too specific initialization process (too close to the boundaries of the objects to segment for example). Some results of segmentation are shown in Fig. 1. The first row shows results obtained with the Gaussian noise and the second row with the Poisson distribution. In both cases, we purposely chose to highly corrupt the original image  $(PSNR = 10 \ dB)$  and to set the regularization parameter  $\beta$  of Eq. (11) to 10 whereas the weigthing parameter for distance maximization is fixed to  $\xi = 0.01$ . As one can notice, regarding the value of  $\alpha$  parameter (restricted to  $[0 \cdot \cdot 1]$  in this study), the segmentation results is very different: considering the Gaussian noise, best results are obtained with non-standard values of  $\alpha$  parameter like  $\alpha = 0.4$ (Fig. 1.a). Usual distances like Hellinger and KL do not lead to satisfying segmentations: In the first case, the main object is not finally well-segmented (Fig. 1.b) and in the second, the segmentation process does not even really starts owing to unsuficient generated forces in terms of magnitude by the alpha-divergence measure. This can not be balanced by a more important regularization: In this case, the active contour can not even stop to the boundaries of the two objects. For Poisson noise, same kinds of results are obtained: best segmentation is achieved thanks to 0.3-divergence (which remains a non-standard value), whereas Hellinger and KL do not lead to proper segmentations (Fig. 1.e and 1.f).



Fig. 1. Some results of segmentation using distance maximization between PDF of inner and outer regions of the synthetic peanut corrupted by Gaussian (a, b, c) and Poisson (d, e, f) noises of  $PSNR = 10 \ dB$ .

#### 3.2. Segmentation of X-ray images



Fig. 2. In green, typical structures of the bones related to osteoporosis pathology. In red, classical segmentation result using a parametric Chan et Vese like method. X-Ray imaging remains of primary interest for diagnosis and follow-up of pathologies related to bones. More precisely, segmentations of some bone structures are required to quantify gold standard parameters (as density, curvature, spacement...) that lead clinicians to a precise diagnosis and follow-up of the considered pathology. Segmentation of that kind of images is challenging for two main reasons: First, these acquisitions are corrupted by a strong Poisson noise that makes its segmentation not always that easy with standard approaches like Chan and Vese [13] one (which is known to be unadapted to clinical image analysis); Second, bones area are characterized by a trabecular texture that can

not be easily parametrically-estimated.

In this context, the first application we propose is the nonsupervised segmentation of X-ray images of hip bone in the framework of osteoporosis diagnosis. Fig. 2 shows the particular structure to highlight (see green circles) for the achievement of a quantification of the severity of the pathology. Moreover, we also show on Fig. 2 a classic result of segmentation (in red) obtained thanks to standard active contour segmentation based on the minimization of the mean and the variance of the inner and the outer regions of the curve. As one can notice, the segmentation results are not satisfying since the important structures of the bone are not preserved due to the presence of some areas of less density. Calculations of quantitative parameters like curvature of the bone are then biased. Fig. 3 shows now segmentation obtained with the proposed approach of this article and for different distances. As one can notice on Fig. 3, usual distances do not make possible a satisfying segmentations: The Hellinger distance provides a segmentation result (Fig. 3.b and 3.e) too smooth that leads to an oversegmentation of the whole bone, and the KL divergence definitely do not fit to this segmentation task (Fig. 3.c and 3.f). Finally, this is a non-standard value of  $\alpha$ 



**Fig. 3.** Hip segmentations from X-ray acquisition for different  $\alpha$  value (to each row corresponds a different acquisition): (a, d)  $\alpha = 0.75$ , (b, e)  $\alpha = 0.5$  (Hellinger/Bhattacharaya distance), (c, f)  $\alpha \rightarrow 1$  (Kullback-Leibler divergence)



Fig. 4. Segmentation of vertebrae structure with different  $\alpha$  value. (a)  $\alpha = 0.75$ , (b)  $\alpha \rightarrow 1$  (Kullback Leibler divergence)

(0.75) that leads to the best segmentation results (Fig. 3.a and 3.d).

Another clinical application proposed for illustration is X-ray vertebrae segmentation. In the framework of this particular application, clinicians are interested in the quantification of the distance between the different vertebrae of the spine in order to characterize some abnormalities of the main structure (see [16] for a clinical description of the problem). Results of segmentation using the proposed approach are shown on Fig. 4. Once again, one can notice that better segmentation (in terms of global shape extraction) is possible with a non-standard value of  $\alpha$  parameter (0.75) compared to usual distances (here only KL divergence result is shown since for Hellinger criterion, results where not exploitable).

### 4. CONCLUSION

In this paper, we have proposed an unsupervised statistical regionbased active contour method integrating maximization of alphadivergences between PDF of inner and outer regions of the active curve. The proposed approach is a generalization of the standard methods of that kind, based on KL or Hellinger divergences mainly, for which we also proposed the original complete derivation of the PDE considering non-parametric estimations of  $p_{in}$  and  $p_{out}$ . The preliminary studies made first on synthetic images - corrupted by high Gaussian and Poisson noises - and, second, on X-ray images charaterized by high Poisson noise and complex trabecular texture, have shown the flexibility of the alpha-divergence parameter in the framework of unsupervised segmentation. The main improvement we are currently working on is to automatically and locally adapt  $\alpha$  parameter to the segmentation context.

## 5. REFERENCES

- M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [2] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson, "Image segmentation using active contours: Calculus of variations or shape gradients?," *SIAM J. Appl. Math.*, vol. 63, pp. 2128–2154, 2003.
- [3] S. Jehan-Besson, Modèles de contours actifs basés régions pour la segmentation d'images et de vidéos, Ph.D. thesis, Université de Nice-Sophia Antipolis, 2003.
- [4] A. Herbulot, S. Jehan-Besson, S. Duffner, M. Barlaud, and G. Aubert, "Segmentation of vectorial image features using shape gradients and information measures," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 3, pp. 365–386, 2006.
- [5] F. Lecellier, S. Jehan-Besson, J. Fadili, G. Aubert, and M. Revenu, "Optimization of divergences within the exponential family for image segmentation," in *SSVM '09*.
- [6] L. Meziou, A. Histace, F. Precioso, B. Matuszewski, and M. Murphy, "Confocal Microscopy Segmentation Using Active Contour Based on Alpha-Divergence," in *Proceedings of ICIP*, September 2011, pp. 3138–3141.
- [7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, New York, 2. edition, 2001.
- [8] O.V. Michailovich, Y. Rathi, and A. Tannenbaum, "Image segmentation using active contours driven by the bhattacharyya gradient flow," vol. 16, no. 11, pp. 2787–2801, 11 2007.
- [9] H. Zhu and R. Rohwer, "Information geometric measurements of generalisation," Tech. Rep. NCRG/4350, Aston University, 1995.
- [10] A.O Hero, B. Ma, O. Michel, and J.D. Gorman, "Alphadivergence for classification, indexing and retrieval (revised)," Tech. Rep. CSPL-328, The University of Michigan, June 2002.
- [11] A. Beirami, V. Cevher, B. Bower, and K. Tsianos, "Proofs of alpha divergence properties," Tech. Rep. STAT 631 / ELEC 639, Rice University, September 2008.
- [12] S. Osher and J. A. Sethian, "Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.
- [13] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. on IP*, vol. 10, no. 2, pp. 266–277, 2001.
- [14] J. Weickert, B. M. Ter Haar Romeny, and M. A. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Transactions on Image Processing*, vol. 7, pp. 398–410, Mar. 1998.
- [15] Y. Zhang, B. Matuszewski, A. Histace, and F. Precioso, "Statistical Shape Model of Legendre Moments with Active Contour Evolution for Shape Detection and Segmentation," in *Proceedings of the 14th CAIP Conference*, Springer, Ed., Sevilla, Espagne, Aug. 2011, vol. 6854 of *LNCS*, pp. 51–58.
- [16] F. Lecron, M. Benjelloun, and S. Mahmoudi, "Vertebra segmentation for the computer-aided cervical mobility evaluation on radiographs," vol. 6, no. 1, pp. 46–52, june 2011.