

VARIATIONAL LEVEL SET METHODS: FROM CONTINUOUS TO DISCRETE SETTING, APPLICATIONS IN VIDEO SEGMENTATION AND TRACKING

Françoise Dibos, Florent Ranchin

CEREMADE (UMR CNRS 7534), Université Paris Dauphine,
Place du Maréchal De Lattre De Tassigny,
75775 Paris cedex 16, France

Abstract

In this paper, we investigate some recent active contour models used in image and video segmentation and we transpose them into a discrete form to apply the Iterated Conditional Modes (ICM) algorithm. This work can be seen as an extension of the recent work of T. Chan and B. Song for other functionals than the Mumford-Shah/Chan-Vese one. We investigate it for video segmentation and tracking applications.

1. INTRODUCTION

For many years, variational level set methods are often used to tackle segmentation problems (static/moving objects detection in image/video [4], image segmentation or texture discrimination ([9])). Solving a segmentation problem is usually achieved by minimizing a given energy functional of the general form

$$J(\Omega) = \int_{\Omega} k^{\text{in}}(x, \Omega) dx + \int_{\Omega^c} k^{\text{out}}(x, \Omega) dx + \int_{\partial\Omega} k^{\text{b}}(x, \Omega) d\mathcal{H}^1(x).$$

Thus the aim is to minimize J with respect to an opened subset Ω of the image domain D .

Optimization is performed either with the shape sensitivity analysis techniques introduced in [10] or with the heaviside function techniques relying on an implicit representation of Ω by a function u positive on Ω and negative on its complementary domain, following the Osher-Sethian level set method [8]. This leads to a new functional which depends on u

$$J(u) = \int_D k^{\text{in}}(x, u) H(u(x)) dx + \int_D k^{\text{out}}(x, u) (1 - H(u(x))) dx + \int_D k^{\text{b}}(x, u) |\nabla H(u)| dx.$$

In order to compute the Gâteaux derivative of J , the heaviside distribution is approximated by a differentiable

function H_ϵ . Whatever the method that is used, we obtain a partial differential equation discretized with classical numerical schemes.

The aim of this paper is to transpose continuous variational level set methods in a discrete setting. The function u is taken as a binary image taking the values $+1$ and -1 . The discrete functional that corresponds to J is given by

$$J_d(u) = \sum_{(i,j)} k^{\text{in}}(i, j, u) H(u(i, j)) + \sum_{i,j} k^{\text{out}}(i, j, u) (1 - H(u(i, j))) + L_{k^{\text{b}}}(u).$$

The last term $L_{k^{\text{b}}}$ is the discretization of $\int_D k^{\text{b}}(x, u) |\nabla H(u)| dx$. Then it is given by

$$\sum_{i,j} k^{\text{b}}(i, j, u) \left((H(u(i+1, j)) - H(u(i, j)))^2 + (H(u(i, j+1)) - H(u(i, j)))^2 \right)^{1/2}.$$

This way of discretizing functionals has been used for the first time in [11] for the minimization of the discrete version of the piecewise constant Mumford and Shah functional [7]. In this work, we will suppose $k^{\text{b}} = \mu > 0$, then we want to find a region with a boundary smooth enough.

Our purpose in this work is to apply the Iterated Conditional Modes (ICM) algorithm for the minimization of discrete functionals arising in segmentation problems and to use the length discretization we have introduced above. First, we will recall the principles of this algorithm. Then, we apply it on video segmentation and tracking models in which k^{in} and k^{out} do not depend on u . In the last section, we apply the ICM to a slightly generalized version of the Mumford-Shah/Chan-Vese functional [4] for a video segmentation application and we discuss about the application to other functionals involving mean and variance of the image.

2. BAYESIAN FORMULATION, ICM ALGORITHM

In this section, for notation simplicity, we denote a pixel $s = (i, j)$ and the set of pixels S .

In an interpretation from a bayesian point of view, the functional J_d can be seen as an *a posteriori energy* conditionally to the image data I if $k^{\text{in}} = -\log(\mathbb{P}(I|u = 1))$ and $k^{\text{out}} = -\log(\mathbb{P}(I|u = -1))$.

We assume u being a sampling of a Gibbs random variable \mathcal{U} , that is to say

$$\mathbb{P}(u) \triangleq \mathbb{P}(\mathcal{U} = u) = \frac{1}{Z} \exp(-\mu L(u))$$

where $Z = \sum_{u \in \mathcal{X}} \exp(-\mu L(u))$ where \mathcal{X} is the set of all possible configurations u , in our case $\mathcal{X} = \{-1, +1\}^S$.

I is assumed to be a sampling of a random variable \mathcal{I} . $\mathcal{I}|\mathcal{U} = +1$ (*resp.* $\mathcal{I}|\mathcal{U} = -1$) is assumed to be a Gibbs random variable of energy k^{in} (*resp.* k^{out}).

Then the energy $E(u|I) = J_d(u)$ can be decomposed in the following way

$$\begin{aligned} E(u|I) &= -\sum_{s, u(s)=+1} \log(\mathbb{P}(I|u = +1)) \\ &\quad -\sum_{s, u(s)=-1} \log(\mathbb{P}(I|u = -1)) + \mu L(u) \\ &= -\sum_s \log(\mathbb{P}(I(s)|u(s)) - \log(\mathbb{P}(u))) + cte \\ &= -\log(\prod_s \mathbb{P}(I(s)|u(s))\mathbb{P}(u)) + cte \\ &= -\log(\mathbb{P}(I|u)\mathbb{P}(u)) + cte. \end{aligned}$$

Thus by the minimization of J_d , we achieve the classical Maximum A Posteriori (MAP) estimation since it consists to maximize the probability *a posteriori* $\mathbb{P}(u|I) \propto \mathbb{P}(I|u)\mathbb{P}(u)$ with respect to u . For MAP estimation, Geman and Geman [6] proposed to use the Simulated Annealing algorithm. Unfortunately, the convergence of this algorithm is very slow and it is often used with a geometrical temperature decreasing instead of the logarithmic theoretical one. Here we use the ICM algorithm introduced by Besag [3], this is a suboptimal procedure, in the sense that it converges only towards a local minimum but it is known for being very fast.

The ICM algorithm consists in minimizing $E(u(s)|I, u(s'), s' \neq s)$. By the relations above and since u is a sampling of a markovian random variables ($L(u)$ is the sum of an expression that only involves the values of u in three neighboring sites), we obtain $E(u(s)|I, u(s'), s' \neq s) = E(I(s)|u(s)) + E(u(s)|u(s'), s' \in \mathcal{V}(s))$ (\mathcal{V} denotes a *neighborhood system* [2-connectivity, 3-connectivity, ...]) up to an additive constant. Then the ICM is performed by sweeping over all the pixels then by minimizing $E(I(s)|v) + E(v|u(s'), s' \in \mathcal{V}(s))$ with respect to v , given the values of u at other pixels. We iterate this procedure until convergence.

3. APPLICATIONS OF THE ICM

3.1. Region independent descriptors

The three functions k^{in} , k^{out} and k^{b} are called region and boundary “descriptors” by Barlaud *et al* in [2]. In this section, we consider region independent descriptors and a constant boundary descriptor: $k^{\text{in}} = k^{\text{in}}(\mathbf{x})$, $k^{\text{out}} = k^{\text{out}}(\mathbf{x})$ and $k^{\text{b}} = \mu$. Such a model arise in video segmentation and tracking:

- For a video segmentation purpose ([2]), Aubert, Barlaud and Jehan-Besson consider a video sequence taken by a static camera and an extracted background image B and they use $k^{\text{in}} = \alpha$ and $k^{\text{out}} = |B - I_t|$ for each image I_t in the video. Thus the aim is to have B and I_t well matched outside the moving regions that are searched, α representing a minimal contrast between B and I_t over the moving regions.
- For tracking purpose, Aron, Mansouri and Mitiche [1] consider the following problem: given an object associated to a reference region Ω_{ref} in image I_t , we want to find its location in the image I_{t+1} . They use the two matching errors

$$\xi_1(x) = \inf_{|v| \leq \alpha, x+v \in \Omega_{ref}} |I(x) - I_{ref}(x+v)|^2$$

and

$$\xi_2(x) = \inf_{|v| \leq \alpha, x+v \in \Omega_{ref}^c} |I(x) - I_{ref}(x+v)|^2$$

which should be minimal respectively on Ω and its complementary. Then it is natural to take $k^{\text{in}} = \xi_1$ and $k^{\text{out}} = \xi_2$.

One step of the ICM is equivalent to compute the energy variation when we change the value of u in s from the current value $u(s)$ to the opposite one. Then, in our case of region independent descriptors, this can be very easily computed:

$$\begin{aligned} \Delta E &= (k^{\text{in}} - k^{\text{out}})(1 - 2H(u(s))) \\ &\quad + \Delta E(u(s)|u(s'), s' \in \mathcal{V}(s)). \end{aligned}$$

The variation of the *a priori* conditional energy $E(u(s)|u(s'), s' \in \mathcal{V}(s))$ can also be computed since it involves pixels which are neighbors of s in the sense of \mathcal{V} . So the implementation of the ICM is quite straightforward here.

We give some experimental results on the Aubert–Barlaud–Jehan-Besson model in figure 2 for $\alpha = 25$. For the initialization choice, we have taken $\{|B - I| \leq \alpha_i\}$ with α_i in a set of values. $\{|B - I| \leq 16\}$ gives the lowest energy after application of the ICM. Choosing different



Fig. 1. One image of the sequence and the background computed by time median filter.



Fig. 2. Aubert, Barlaud and Jehan-Besson video segmentation model by comparison of the images and a given background. The initialization (the first image) is chosen as $\{|B - I| \leq 16\}$. The result with the ICM superimposed to the original image.

initialisations allow to avoid irrelevant local minima. For the tracking model of Mansouri *et al.*, some results are given in figure 3 (the initialization on the first image is performed with a classical snake).

3.2. Region dependent descriptors

In the case of region dependent descriptors, the energy variation expression is more complicated than in the region independent descriptors case. But the particular case of the Mumford-Shah/Chan-Vese functional has been studied by Chan and Song in [11]. We have then $k^{in}(x) = (I(x) - \mu_1)^2$ and $k^{out}(x) = (I(x) - \mu_2)^2$ where μ_1 and μ_2 are the mean values of I over $\{u = +1\}$ and $\{u = -1\}$. When changing $u(s)$ from $+1$ to -1 (for the change from -1 to $+1$, the indices are inverted), we get ([11]) the energy variation (n_1 denotes the number of pixels in the region and

μ_1 the mean value of I over this region)

$$(I(s) - \mu_2)^2 \frac{n_2}{n_2 + 1} - (I(s) - \mu_1)^2 \frac{n_1}{n_1 - 1} + \Delta E(u(s)|u(s'), s' \in \mathcal{V}(s)).$$

A slightly generalized version of the Mumford-Shah/Chan-Vese functional lies on the descriptors $k^{in}(x) = \log(\sigma_1^2) + (\frac{I(x) - \mu_1}{\sigma_1})^2$ and $k^{out}(x) = \log(\sigma_2^2) + (\frac{I(x) - \mu_2}{\sigma_2})^2$. It has been used by Deriche and Rousson in [9] for texture segmentation and tracking. From a probability point of view, this model relies on the hypothesis that the image is a mixture of two gaussians with different means and variances. Actually, we can throw away the terms $(\frac{I(x) - \mu_i}{\sigma_i})^2$, indeed

$$\sum_{s, u(s)=+1} (\frac{I(s) - \mu_1}{\sigma_1})^2 = \frac{\sigma_1^2}{\sigma_1^2} n_1 = n_1,$$

and

$$\sum_{s, u(s)=-1} (\frac{I(s) - \mu_2}{\sigma_2})^2 = \frac{\sigma_2^2}{\sigma_2^2} n_2 = n_2,$$

and thus the summation of these two terms gives $n_1 + n_2 = n_{tot}$ the total number of pixels in the image which is constant, thus the energy can be expressed very easily by

$$n_1 \log(\sigma_1^2) + n_2 \log(\sigma_2^2) + E(u(s)|u(s'), s' \in \mathcal{V}(s)).$$

In comparison, the Chan-Vese energy is given by

$$n_1 \sigma_1^2 + n_2 \sigma_2^2 + E(u(s)|u(s'), s' \in \mathcal{V}(s)),$$

and thus the expression of the energy variation for the Rousson-Deriche model is given by

$$\begin{aligned} \log\left(\frac{\sigma_1^2}{\sigma_1^2}\right) &+ \log\left(\left(\frac{n_1}{n_1-1}\right)^{n_1-1} \left(\frac{n_2}{n_2+1}\right)^{n_2+1}\right) \\ &+ (n_1 - 1) \log\left(1 - \left(\frac{I - \mu_1}{\sigma_1}\right)^2 \frac{1}{n_1 - 1}\right) \\ &+ (n_2 + 1) \log\left(1 + \left(\frac{I - \mu_2}{\sigma_2}\right)^2 \frac{1}{n_2 + 1}\right). \end{aligned}$$

As an application, video segmentation with a given background image is still possible (replace I by $|B - I|$ the difference between the image and the background in the computations above, see [5]). The advantage over the model of Aubert *et al* is that it does not depend anymore on a parameter α . We display the result on figure 4 for $\mu = 50$.

4. CONCLUSION

In this paper, we have proved that the Chan and Song algorithm can be seen as a particular case of the well known ICM algorithm. Here we have applied it to other functionals involving either region-dependent or region-independent descriptors. We have shown applications in video segmentation and tracking, but texture segmentation is also possible, since the Deriche-Rousson model was also designed for this purpose.

5. ACKNOWLEDGEMENTS

The second author would like to thank Jean-Michel Marin for his comments and corrections, and for having introduced us the ICM algorithm.

© 2005 by the authors, submitted to International Conference on Image Processing. This paper may be reproduced, in its entirety, for non-commercial purposes.

6. REFERENCES

- [1] M. Aron, A.-R. Mansouri, and A. Mitiche. Pde based tracking without motion computation by joint space-time segmentation. In *IEEE International Conference on Image Processing*, volume 3, pages 113–116, 2003.
- [2] G. Aubert, M. Barlaud, and S. Jehan-Besson. Video object segmentation using eulerian region-based active contours. In *International Conference in Computer Vision Proceedings*, Vancouver, Canada, 2001.
- [3] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48:259–302, 1986.
- [4] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.
- [5] R. Deriche and N. Paragios. Geodesic active regions for motion estimation and tracking. *Proceedings of the Int. Conf. in Computer Vision*, pages 224–240, 1999.
- [6] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1993.
- [7] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.
- [8] S. Osher and J. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on the hamilton-jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1990.
- [9] M. Rousson and R. Deriche. Adaptive segmentation of vector valued images. *Geometric Level Set Methods in Imaging, Vision and Graphics*, pages 195–205, 2003.
- [10] J. Sokolowski and J.-P. Zolésio. *Introduction to Shape Optimization. Shape sensitivity analysis*. Springer Ser. Comput. Math. Springer-Verlag, 1992.
- [11] B. Song and T. F. Chan. A fast algorithm for level set based optimization. Technical Report 02-68, UCLA, 2002.



Fig. 3. Results of the tracking method of Aron, Mansouri and Mitiche with the ICM from the second to the fourth image of the speaker sequence movie ($\mu = 10$ and $\alpha = 3$).

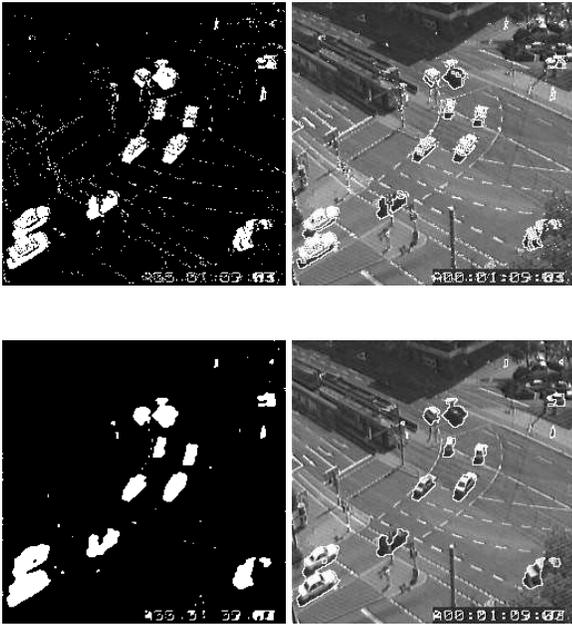


Fig. 4. Application of the ICM with the Rousson-Deriche model on the difference image $|B - I|$ for $\mu = 50$. On the top of the figure, the initialization of the algorithm: we threshold $|B - I|$ at the level 20; the left image is the mask and the right image is the mask superimposed upon the image.