Minimal Stochastic Complexity Image partitioning with nonparametric statistical model

G. Delyon, Ph. Réfrégier and F. Galland

Fresnel Institute, UMR CNRS 6133, EGIM, Domaine Universitaire de Saint-Jérôme, 13397 Marseille Cedex 20, France.

Abstract. We propose to analyse a general statistical image partitioning into homogeneous regions. This method is based on a polygonal grid which can have an arbitrary topology and whose number of region and regularity of its boundaries are obtained by minimizing of the stochastic complexity of the image. It thus leads to optimize a criterion that can be expressed as the sum of two terms: a likelihood term and a regularization term. This criterion, whose two terms are elements of a global entropy and are expressed in the same unit, does not contain any parameter to be tuned by the user. We analyse the performance of this approach in comparison with a more standard formulation for which there is a weighting parameter between the likelihood and the regularization terms analogous to the one used in snake-based techniques.

Keywords: Image partitioning, Stochastic Complexity, active contour, Density Estimation
PACS:

INTRODUCTION

Image segmentation is a topic of growing interest in computer vision and image processing and has numerous applications. The sensors used to obtain these images can thus be of various types and can lead to images corrupted with different noise models: Gaussian, Gamma, Poisson, ... It is therefore important to take into account the physical nature of the images in statistical techniques for image segmentation.

Snake-based segmentation techniques generally depend on the minimization of a criterion that is the sum of two energies: an external energy, that is based on the grey levels of the image and on a statistical model, and an internal energy, that allows one to regularize the boundaries of the regions of the image. Most of the developed statistical segmentation techniques are based on a Bayesian formulation of the statistical model which generally leads to the optimization of a criterion that has at least one parameter to be tuned by the user. It is the case of the Markov Random Fields (MRF) approaches [1], [2], [3], [4] that introduce regularization parameters that cannot be easily determined automatically and which can leads to difficult optimization problems. It is also the case of variational methods [5], [6], [7], [8]. With deformable models, the desirable properties such as continuity and smoothness of the contours are enforced by introducing regularization terms in the functional to optimize. The situation is thus analogous to the one obtained with MRF approaches since parameters are in general present in the criterion to optimize.

Reducing the number of free parameters in the criterion to optimize thus appears as one of the key problems in image segmentation. The Stochastic Complexity (SC) minimization principle, introduced by Rissanen [9], [10] has been early used to adress the issue of parameter estimation. Based on information theory, this principle allows one to estimate the number of needed parameters for parametric description of observed data. In the context of image segmentation, the estimation of the number of needed parameters for parametric description of observed data can be an interesting alternative to the introduction of regularization terms.

Leclerc [11] first proposed to apply this approach to optical image segmentation. Then, Zhu and Yuille [12] proposed a region competition algorithm deduced from a MDL criterion. Their approach allows one to segment complex images. However this technique is based on a Bayesian framework and contains free parameters in the energy criterion. Figueiredo *et al.* [13] demonstrated that the SC principle can be successfully applied to B-spline snake-based models with Gaussian or Rayleigh noise. These results have been generalized to polygonal snake [14] and to level set implementation [15]. Finally, a new approach of SAR image partitioning into homogeneous regions has been proposed by Galland *et al.* [16]. This algorithm is based on a polygonal grid whose number of region, number of nodes and position of nodes are automatically estimated by minimizing the SC of the image. This technique has been then generalized to other PDF of the exponential family [17].

Although these techniques provide interesting results, they have some limitations. Indeed, they need *a priori* knowledge of the family of the PDF of the image grey levels. In this paper, we analyse a generalization of the approach introduced in [16] to images with unknown noise model for which a parametric model of grey level's fluctuations is not available. The homogeneous partitioning is based on a deformable partition defined by a polygonal grid which can have an arbitrary topology and is obtained by minimizing the SC complexity of the image which is a criterion without parameter to be tuned by the user. We analyse quantitatively on a synthetic image and qualitatively on a real image the relevance of this approach compared to Bayesian techniques which generally introduce regularization parameters to balance the weight of the different terms that compose the criterion.

STOCHASTIC COMPLEXITY BASED ALGORITHM

Our goal is to estimate the partition of the image which contains R regions. In each region of the image, we assume that the pixel intensities are quantized on M levels (for example $M = 256$) and are realizations of independent and identically distributed random variables with PDF P_r. The SC Δ of the image, which is an approximation of the number of bits needed to describe the image, given the partition and the statistical model, can be expressed as the sum of two terms [10]: a likelihood term Δ_L which is equivalent to the external energy of the snake-based approach and a regularization term Δ_M which is equivalent to the internal energy. One thus obtains:

$$\Delta = \Delta_L + \Delta_M. \quad (1)$$

Given a partition and the PDF P_r in each region of the image, Δ_L is the number of bits needed to describe the grey levels of the image. For that purpose, the PDF are estimated with irregular step functions of order Q as shown on figure 1. For a fixed value of Q, the parameters of a step function are the width b_j of the bin number j and the probability $p_r(j)$ of observing a grey level in the bin number j, for all $j \in [1,Q]$. The values of Q and of the b_j are chosen to be the same for all the regions of the image. The expression of Δ_L is provided in [18].

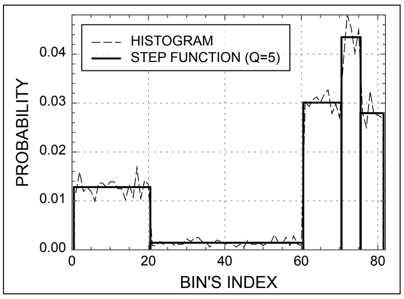


FIGURE 1. Illustation of the estimation of a grey level distribution defined on $M = 81$ grey levels (dashed line) with a step function of order $Q = 5$ and with $b_1 = 20$, $b_2 = 40$, $b_3 = 10$, $b_4 = 5$ and $b_5 = 6$ (solid line).

Δ_M is the sum of the code lenght needed to describe the PDF of each region and the code length needed to describe the partition of the image. The expression of the length of the encoding of the PDF is given in [18]. The partition is specified with a polygonal active grid (i.e. a set of nodes linked by segments). This grid allows us to get a partition in an arbitrary number of regions and with arbitrary topologies. The length of its encoding is given in [16].

OPTIMIZATION PROCEDURE

The optimal partition Π^{opt}, which is the parameter of interest and which depends on the number of steps Q and the widths of the steps b_j that are nuisance parameters, is obtained by minimizing Δ. The proposed optimization algorithm, that allows one to estimate the partition and the nuisance parameters, is based on three stages as shown on figure 2.

Stage 1: first estimation of the partition

For that purpose, we perform the convergence procedure of the grid described in [16] using step functions of order $Q_0 = 10$ and with bins of equal length and starting with an initial regular grid Π_0. The results obtained at the end of this stage are shown on figure 3.

Stage 2: estimation of Q and the b_j

For that purpose we initialize the steps functions with $Q = M$ and $b_j = 1$ for all $j \in [1,Q]$. Given the estimated partition of stage 1, we then evaluate the SC Δ obtained

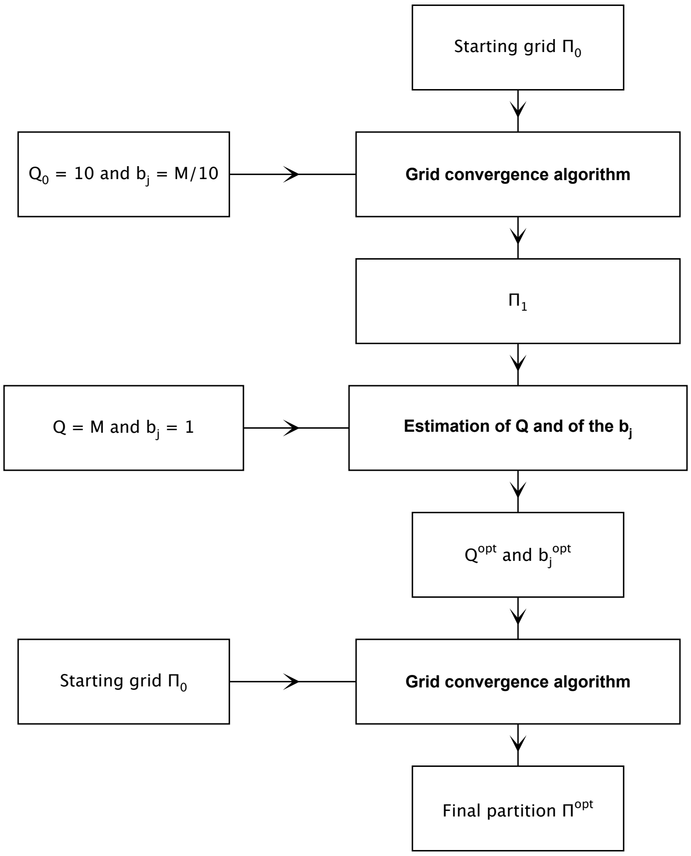
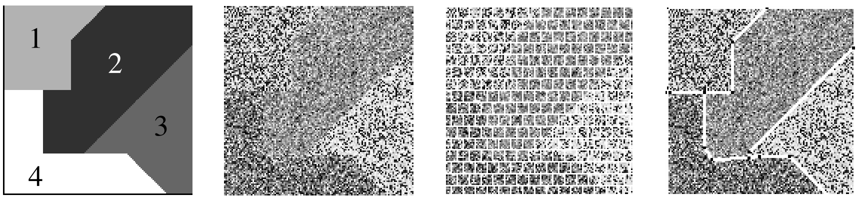


FIGURE 2. Description of the three stage algorithm. The three stages are written in bold font symbol.



(*a*) (*b*) (*c*) (*d*)

FIGURE 3. **Stage 1** of the three stages algorithm. (*a*) synthetic image with the numeration of its regions - (*b*) noisy version quantized on $M = 81$ grey levels with random grey levels distributed with step functions of order $Q = 5$ with $b_1 = 20$, $b_2 = 40$, $b_3 = 10$, $b_4 = 5$ and $b_5 = 6$ - (*c*) starting grid - (*d*) partition obtained at the end of stage 1 with $Q_0 = 10$.

after each fusion of two adjacent bins j and $j+1$ and we choose the fusion which leads to the greatest decrease of Δ. This fusion provides a new bin of width $b_j + b_{j+1}$ so that the probability of observing a grey level in this bin is $p_r(j) + p_r(j+1)$. One iterates this procedure so that the order Q decreases from $Q = M$ to $Q = 2$. The order Q^{opt} and the corresponding b_j^{opt} which minimize Δ are finally selected (figure 4).

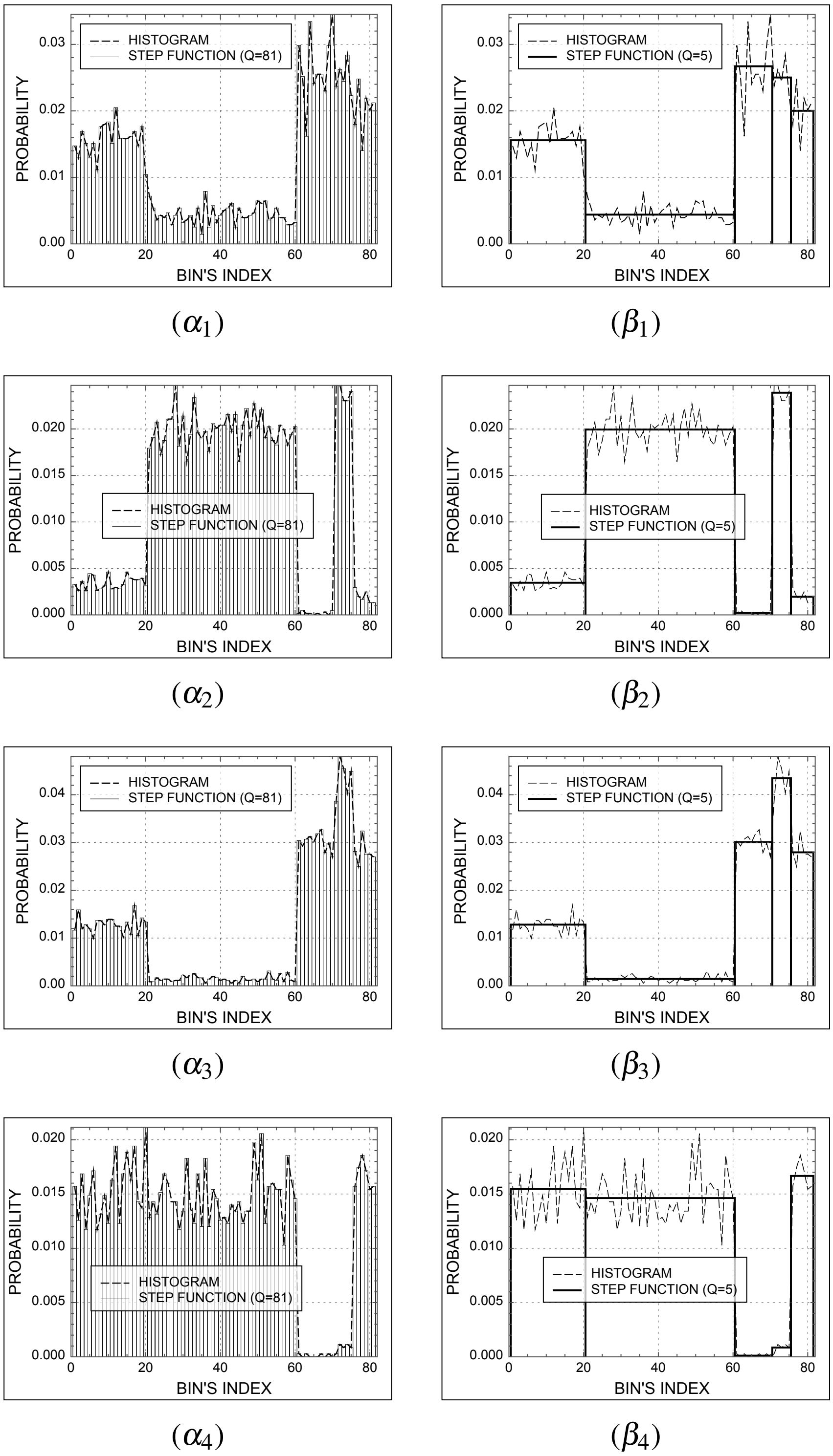
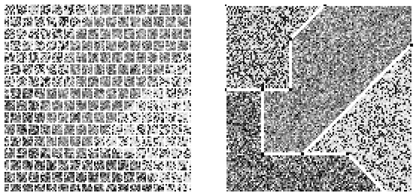


FIGURE 4. Stage 2: estimation of Q and the b_j. (α_i) initialization of the step functions on region number i with $Q = M$ and $b_j = 1$ - (β_i) result on region number i obtained after the fusion procedure. One obtains $Q^{\text{opt}} = 5$ with $b_1^{\text{opt}} = 20$, $b_2^{\text{opt}} = 40$, $b_3^{\text{opt}} = 10$, $b_4^{\text{opt}} = 5$ and $b_5^{\text{opt}} = 6$.

Stage 3: final estimation of the partition

With the estimated parameters Q^{opt} and the b_j^{opt}, a final convergence procedure of the grid is applied (figure 5). Since it is easy to be trapped in a local minimum, the initial grid of this stage is the same as in the first stage. One can see on the last figure the importance of the second convergence which allows us to correct some details of the boundaries between regions (the boundary between region 2 and 4 for example) that were not correctly detected with the first convergence. Indeed, the number of misclassified pixels (NMP) [19], which is a measure for evaluating the quality of the partition is equal to 29 at the end of stage 1 and is equal to 0 at the end of stage 3.



(a) (b)

FIGURE 5. Stage 3: final estimation of the partition. (a) starting grid - (b) partition obtained after the second convergence of the polygonal active grid and using step functions whose parameters have been estimated at stage 2.

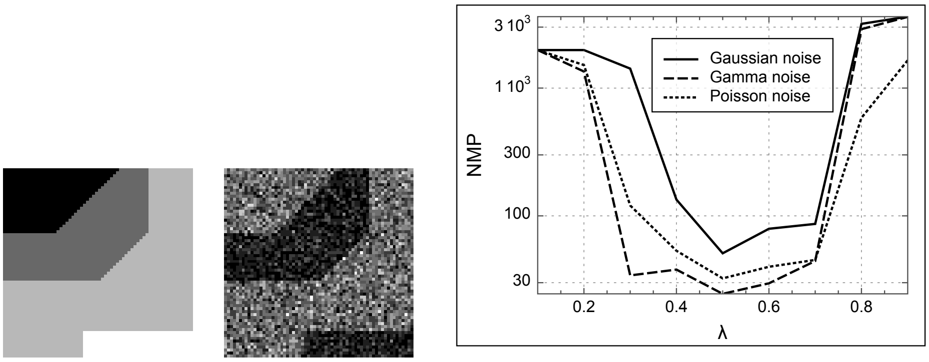
RELEVANCE OF THE STOCHASTIC COMPLEXITY BASED CRITERION

The specificity of the SC is to provide a criterion without parameter to be tuned, contrary to Bayesian approachs. Indeed, in Bayesian techniques, like those proposed by Mumford and Shah [6] or Zhu and Yuille [12], regularization parameters are introduced in the functional to optimize in order to enforce continuity and smoothness of the contours. In the criterion based on the SC, the two terms (the likelihood term and the regularization term) used to describe the image are elements of a global entropy and are expressed in a common unit, namely *the bit*. Consequently, these terms do not need any parameter to balance their weight.

In order to analyse the efficiency of the proposed approach, we have added a regularization parameter λ in the proposed criterion which leads to the following new criterion:

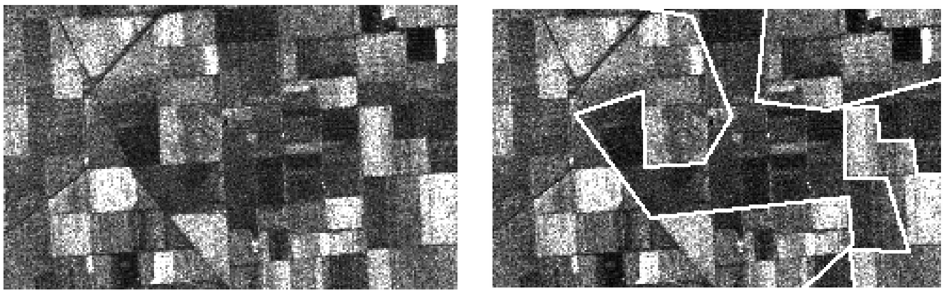
$$\Delta(\lambda) = \lambda \Delta_L + (1-\lambda)\Delta_M. \tag{2}$$

We have then applied the three stage algorithm on the synthetic image of figure 6 by optimizing $\Delta(\lambda)$ for different values of λ. It is thus equivalent to put more ($\lambda > 0.5$) ou less ($\lambda < 0.5$) weight in the likelihood term in comparison with the regularization term. We show on this figure the evolution of the NMP as a function of λ. The performances are clearly optimal for $\lambda = 0.5$, which correspond to optimizing the SC of the image. These results show that if one wants to improve the quality of the partition obtained with the SC based criterion by introducing a regularization parameter λ, the estimation of this parameter, i.e. the value of λ which minimize the NMP, leads to the SC based criterion.

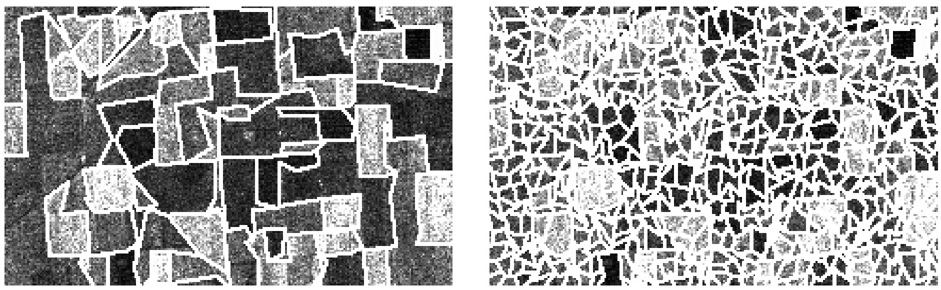


(a) (b) (c)

FIGURE 6. Number of misclassified pixels *NMP* (c), estimated on 50 realizations of gaussian noise, gamma noise and poisson noise (b) of synthetic image (a), obtained after the three stage algorithm but using the criterion $\Delta(\lambda)$, as a function of λ.



$\lambda = 0.1$



$\lambda = 0.5$ $\lambda = 0.9$

FIGURE 7. Partitioning results of a SAR agricultural image from Ukraine obtained by the ERS-1 satellite (provided by the CNES and distributed by the ESA). The results have been obtained by minimizing the criterion $\Delta(\lambda)$ for three different values of λ : 0.1, 0.5 and 0.9.

This result thus emphasizes the fact that minimizing the SC based criterion corresponds to minimizing a global entropy which does not need any parameter to be tuned by the user.

We illustrate this point on the real SAR agricultural image of figure 7. We show partitioning results obtained by optimizing $\Delta(\lambda)$ for three different values of λ : 0.1, 0.5 and 0.9. We can see on this figure that the more satisfactory result is obtained with

$\lambda = 0.5$, i.e. with the SC based criterion.

CONCLUSION

We have analysed a general statistical technique for image partitioning into homogeneous regions based on the SC minimization principle. This principe allows us to obtain a method based on the optimization of a criterion that does not need *ad hoc* parameters to be tuned by the user. We have analysed the performance of this approach in comparison with a Bayesian formulation for which a weighting parameter is generally introduced between the likelihood and the regularization terms analogous to the one used in snake-based techniques.

ACKNOWLEDGMENTS

The authors are grateful to the European and French Space Agencies (ESA and CNES) for supplying the SAR image and also thank the Phyti team for fruitful discussions.

REFERENCES

1. P. Kelly, H. Derin, and K. Hartt, *IEEE Trans. Acoustics, Speech and Signal Processing* **36**, 1628–1641 (1988).
2. P. C. Smits, and S. G. Dellepiane, *IEEE Trans. Geoscience and Remote Sensing* **35**, 844–857 (1997).
3. S. A. Barker, and P. J. W. Rayner, *Pattern Recognition* **33**, 587–602 (2000).
4. Z. Tu, and S.-C. Zhu, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24**, 657–673 (2002).
5. M. Kass, A. Witkin, and D. Terzopoulos, *International Journal of Computer Vision* **1**, 321–331 (1988).
6. D. Mumford, and J. Shah, *Communications on Pure and Applied Mathematics* **42**, 577–684 (1989).
7. L. D. Cohen, *CVGIP: Image Understanding* **53**, 211–218 (1991).
8. V. Caselles, R. Kimmel, and G. Sapiro, *International Journal of Computer Vision* **22**, 61–79 (1997).
9. J. Rissanen, *Automatica* **14**, 465–471 (1978).
10. J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore, 1989.
11. Y. Leclerc, *Computer Vision* **3**, 73–102 (1989).
12. S. C. Zhu, and A. Yuille, *IEEE Trans. Pattern Anal. and Machine Intell.* **18**, 884–900 (1996).
13. M. Figueiredo, J. Leitão, and A. K. Jain, *IEEE Trans. Image Processing* **9**, 1075–1087 (2000).
14. O. Ruch, and P. Réfrégier, *Opt. Lett.* **26**, 977–979 (2001).
15. P. Martin, P. Réfrégier, F. Goudail, and F. Guérault, *IEEE Trans. Pattern Anal. and Machine Intell.* **26**, 799–803 (2004).
16. F. Galland, N. Bertaux, and P. Réfrégier, *IEEE Transactions on Image Processing* **12**, 995–1006 (2003).
17. F. Galland, N. Bertaux, and P. Réfrégier, *Pattern Recognition* **38**, 1926–1936 (2005).
18. G. Delyon, F. Galland, and P. Réfrégier, *to appear in IEEE Trans. on Image Processing* **15** (2006).
19. G. Delyon, and P. Réfrégier, *to appear in IEEE Trans. on Geoscience and Remote Sensing* **44** (2006).
20. P. Ferrand, M. Egen, R. Zentel, J. Seekamp, S. G. Romanov, and C. M. Sotomayor Torres, *Applied Physics Letters* **83**, 5289–5291 (2003).