

A comparison of non-parametric segmentation methods

Bruno SCIOLLA¹ Paola CECCATO¹, Thibaut DAMBRY², Benoît GUIBERT² Philippe DELACHARTRE¹

¹CREATIS, INSA Lyon, France

²Atys Medical, Soucieu-en-Jarrest, France

bruno.sciolla@creatis.insa-lyon.fr

Résumé – Pour la segmentation d’images, différentes méthodes ont été proposées pour segmenter une image à partir d’estimateurs de Parzen des distributions d’intensités, par exemple la distance de Bhattacharyya, de Kullback-Leibler, ou la log-vraisemblance. Nous comparons plusieurs méthodes couramment utilisées et montrons que les méthodes basées sur la log-vraisemblance sont les plus robustes, et en particulier sont exemptes de problèmes de bords rencontrés dans toutes les autres méthodes testées. Ces résultats donnent des indications claires sur quelles méthodes doivent être préférées et nous avançons quelques arguments théoriques dans cette direction.

Abstract – In image segmentation, level-set methods discriminating regions with Parzen estimates of their intensity distributions have proven useful in a broad variety of contexts. A number of area cost terms have been proposed to achieve this goal, such as log-likelihood, Bhattacharyya coefficient, Kullback-Leibler divergence and several others. In this work we compare the performance of the most widespread criteria and show that log-likelihood and assimilated methods have a clear advantage in terms of robustness. In particular, the other methods tested suffer from a boundary instability due to small region/small initialization/hard to distinguish regions. We also give some theoretical arguments supporting our experimental results on synthetic and real images.

1 Introduction

Several methods have been proposed to segment image via estimates of their intensity distribution. Some methods maximize the likelihood that the contour divides the image into two distinct areas [1]. Other methods tend to make the distributions in the two areas as different as possible, according to some measure of a distance between distributions. The measure may be the Hellinger distance [2, 3, 4], the Kullback-Leibler divergence [2, 5], the L2-norm on cumulated distribution functions [6]. However, little information is available about the efficiency, robustness and specificity of these various methods. In the following we compare the robustness and accuracy of a pool of methods on images where the target and background areas are difficult to distinguish, a practical problem which arises for example in medical imaging.

2 Notations and problem

We consider the problem of segmenting an image into two areas Ω_1 and Ω_2 . In the following we consider only standard images in two dimensions but the conclusions are valid in any dimensions. The level-set method is a type of active contour method [7] where the region Ω_1 is delimited by the zeros of a level-set function $\phi(\mathbf{x})$, such that $\Omega_1 = \{\mathbf{x} | \phi(\mathbf{x}) \geq 0\}$ and $\Omega_2 = \{\mathbf{x} | \phi(\mathbf{x}) < 0\}$ where \mathbf{x} spans the image coordinates. The segmentation problem is then

formulated as an optimization problem, assuming that the target segmentation minimizes:

$$E[\phi] = \alpha \int d\mathbf{x} \delta\phi(\mathbf{x}) \|\nabla\phi\| + \beta E^{\text{DA}}[\phi] \quad (1)$$

where the integration over \mathbf{x} covers the entire image. The energy contains a regularization term with coefficient α ensuring smoothness of the contour and a data attachment term, with coefficient β . In this article we focus on the methods where the data attachment term E^{DA} tends to split the image into Ω_1 , Ω_2 where the *intensity distributions differ as much as possible*. Then, the contour is obtained as the stationary solution of a gradient descent $\partial_t \phi(\mathbf{x}) = -\delta E[\phi] / \delta \phi(\mathbf{x})$ for some appropriate initial condition.

In all the non-parametric methods we consider here, the distributions of intensities in region Ω_1 and Ω_2 are supposed to be unknown during the segmentation process and have to be estimated at each time using Parzen estimates [8]:

$$\hat{P}_1(I) = \int_{\Omega_1} d\mathbf{x} K_\sigma(I(\mathbf{x}) - I) / A_1 \quad (2)$$

where K_σ is a Gaussian distribution of fixed width σ and A_1 is the area of Ω_1 . A similar estimate is given for $\hat{P}_2(I)$. Given these distributions, a number of data attachment terms E^{DA} can be proposed, listed in table 1.

The measures we consider fall into two distinct categories: the first one $C^I = \{\text{(LL)}, \text{(EE)}\}$ rely on a statistical framework. The (LL) method quantifies the likelihood

of a given contour, based on a maximum a posteriori estimate. This method is also related to the mutual information between the contour and the image [1] and is also related to the extensive entropy (EE) [9]. The second category $C^{II} = \{(BC) \dots (L2)\}$ of methods consider the respective distributions in the two domains, $\hat{P}_1(I)$ and $\hat{P}_2(I)$, and tends to make them differ as much as possible. Various quantities have been proposed to achieve this, and we only list a few.

Max. a posteriori Log-likelihood (LL) [1]:
$-\sum_{x \in R_1} \log(P_1(I_x)) - \sum_{x \in R_2} \log(P_2(I_x))$
Extensive entropy (EE):
$-A_1 \sum_I P_1(I) \log(P_1(I)) - A_2 \sum_I P_2(I) \log(P_2(I))$
Bhattacharyya coefficient (BC) [8]:
$\sum_I \sqrt{P_1(I)P_2(I)}$
Hellinger distance (HD) [2]:
$\sqrt{1 - B}$
Kullback-Leibler divergence (KL) [2, 5]:
$\frac{1}{2} \sum_I P_1(I) \log \frac{P_1(I)}{P_2(I)} + \frac{1}{2} \sum_I P_2(I) \log \frac{P_2(I)}{P_1(I)}$
Kolmogorov-Smirnov statistic (KS):
$\sup_I C_1(I) - C_2(I) $
L^2 difference (L2):
$\sum_I (P_1(I) - P_2(I))^2$

TAB. 1: Data attachment energy E^{DA} for the different methods compared here. Type C^I/C^{II} methods are in light gray/dark gray respectively. Notice that all criterion (except for BC) must be maximized, so we take E^{DA} as minus those quantities (plus for BC). $C_1(I)$, $C_2(I)$ are the cumulative distribution functions associated with $P_1(I)$, $P_2(I)$.

3 Comparison of the different data terms

3.1 Experiment layout

We first compare the various data terms in a controlled setting. We consider synthetic images formed of i. i. d. random variables with two different distributions in two target regions T_1 and T_2 , as shown in figure 1. In order to understand precisely the properties of the data terms, we

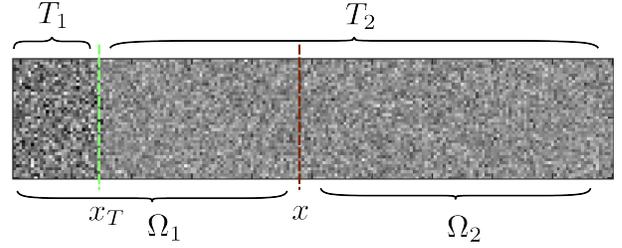


FIG. 1: Random image with target areas T_1 and T_2 with target boundary at coordinate x_T . A trial contour is shown at position x .

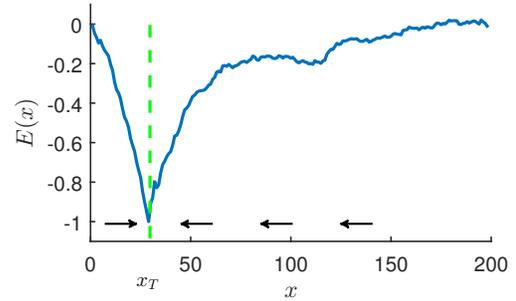


FIG. 2: Energy $E^{\text{DA}}(x)$ for the log-likelihood criterion (LL) as a function of contour position x . The level-set algorithm is a gradient descent in $E^{\text{DA}}(x)$. The direction of the flow is indicated with arrows.

consider a simplified numerical experiment in which the smoothing term is excluded. We suppose that T_1 , T_2 are vertical bands and we put the hard constraint that the contour is a vertical line parametrized only by its position x .

For one realization of the random image, one can compute the data term $E^{\text{DA}}(x)$ as a function of the boundary position x as depicted in figure 2. The segmentation is performed, finding the global minima of $E^{\text{DA}}(x)$ for all x . The level-set method does a gradient descent of the data term $E^{\text{DA}}(x)$.

3.2 Non-monotonicity of the gradient flow

Since the energy is dependent on the image realization, we show the average of $E^{\text{DA}}(x)$ for different methods in figure 3. Unlike in the case shown previously, the data term $E^{\text{DA}}(x)$ is not a monotonous function of x for (BC) and (L2) criteria. As a result, if the initial contour x_i is too close to the boundaries, the gradient descent drives the contour in the wrong direction and the gradient descent diverges. We find that this phenomenon occurs for all distances in the class C^{II} , and not for C^I methods. A qualitative explanation is given in section 5. This effect is stronger when the area of one of the regions T_1 , T_2 is small, or when the two distributions are more difficult to distinguish.

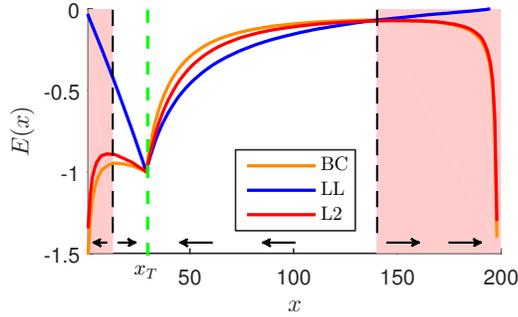


FIG. 3: Average energy $E^{\text{DA}}(x)$ (over many random image realizations) as a function of contour position x for three criterion (BC), (LL) and (L2). $E^{\text{DA}}(x)$ is shifted and rescaled for readability. $E^{\text{DA}}(x)$ reaches a local minimum for $x = x_T$ for all considered methods. However, there are additional minima at the boundaries for (BC) and (L2) methods. The direction of the gradient descent flow is indicated with arrows for method (BC). One can see that the method does not converge to the correct result for initial points in the red areas.

4 Comparison on synthetic images

In this section, we consider the full gradient descent of energy (1) with a level-set formalism. The smoothing parameter is set to its optimal value for the problem at hand, $\alpha = \{2, 3\}$.

We find that the different methods yield good results for simple cases with very well-distinguished regions, thus we focus on the more difficult cases. For definiteness we focus on the two representative methods of C^I and C^{II} , the (LL) and (HD) methods. We use a synthetic image again with a square target area T_1 in a larger background region T_2 with Gaussian intensity distributions of average and standard deviation $\mu_1 = \mu_2 = 6, \sigma_1 = 2.2, \sigma_2 = 1.5$.

4.1 Final contour

For initial contours close to the target contours, in the case where all methods converge, we find that the (LL) method provides a significantly better segmentation than the (HD) method, as shown in figure 4.

Over a large number (100) of random image realizations, the medium pixel misclassification ratio $R = A(\Omega_1 \cup T_1 - \Omega_1 \cap T_1)/A(T_1)$ is of 17.8% for the (LL) method and 41% for the (HD) method. Moreover, in our test the error with (LL) is lower than that with (HD) for each single run. The histogram of R is shown in figure 4, showing that the outcome is much more fluctuating for (HD) and that divergence occasionally occurs.

4.2 Convergence/divergence

According to our results in the previous sections, with the (HD) distance the gradient flow should not always

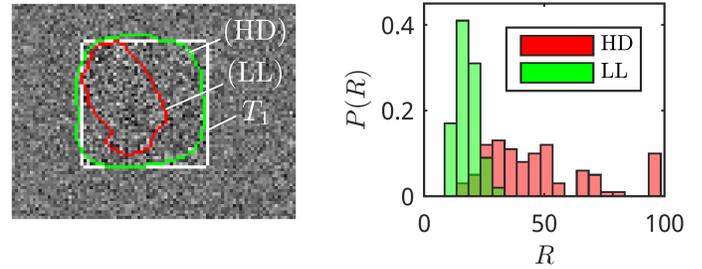


FIG. 4: Left panel: Results of the (HD), (LL) methods compared with the target contour T_1 (white line). Right panel: Histograms of the misclassification R (in %) for the (HD) and (LL) methods over 100 images. 100% error means divergent outcome.

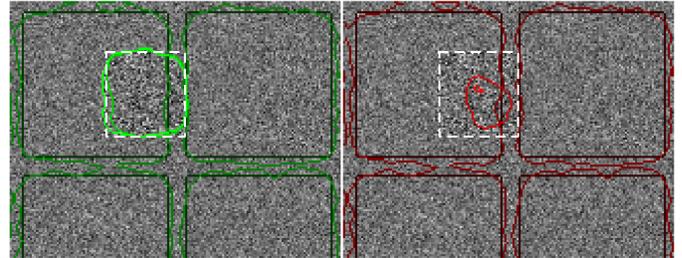


FIG. 5: Four contours with increasing time (black to light) for the (LL) method (left) and the (HD) method (right). The initial contour is on a grid of large squares. For (LL) the contour converges towards the target area T_1 (white dashed line). For the (HD) method the contour shrinks down to a point (divergence).

be monotonous and can drive the contour in the opposite direction than the expected one. This leads to instabilities where one of the areas shrinks down to zero. As shown previously in figure 3, the gradient descent is divergent if the initial contour lies too far from the target boundaries.

We have observed a similar behavior in images, where a simple initial contour encircling the target area converges like in figure 4. On the other hand, a divergent behavior may occur for different initial contours, like a square grid shown in figure 5. For the (LL) method, we find that the gradient descent converges for a much broader range of initializations and provides consistent results.

5 Rationale

We now give a qualitative explanation of the divergence in C^{II} methods. The divergence occurs whenever one of the two segmented regions (say Ω_1) is relatively small, and thus the statistical estimate $\hat{P}_1(I)$ is not representative of the true distribution in the target area T_1 . In this case, shrinking the area Ω_1 causes larger fluctuations in the distribution $\hat{P}_1(I)$. As illustrated in figure 6, the (L2) distance between the two regions Ω_1 and Ω_2 is increased. The decrease of (L2) distance yields an incorrect behavior,

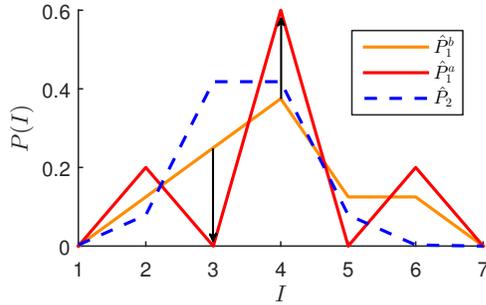


FIG. 6: Illustration of the divergence. Histograms on the two regions Ω_1, Ω_2 before ($\hat{P}_1(I)^b$) and after ($\hat{P}_1(I)^a$) the area Ω_1 is shrunk. We assume that $A_2 \gg A_1$ and thus $\hat{P}_2(I)$ is weakly changed. The (L2) distance is increased from 0.05 to 0.26 by the change of boundary. In this example, due to the low number of samples, some histograms ($I = 3, 5$) are emptied by the change of boundary, increasing the (L2) distance. Furthermore, the other intensities ($I = 4$) may over-represented and further increase the distance. We show histograms for the sake of clarity, but the same holds for Parzen estimates of the distributions.

since the region Ω_1 should be expanded, not decreased further. This argument explains the gradient flow towards the boundaries of the image, shown in figure 3, and the collapse of the (HD) method in figure 5. Any method of type C^{II} is subject to this problem. On the other hand, the C^I method rely on a statistical method which is aware of the respective sizes of the two regions (for example, the entropy criterion (EE) is extensive, i. e. scales like the area) and therefore of the statistical relevance of the estimates $\hat{P}_1(I), \hat{P}_2(I)$.

6 Conclusions

We have compared different methods of segmentation based on intensity distributions and found that the methods based on likelihood of the contour are stable whereas methods involving a distance between distributions may diverge. Further study is necessary to extend these results beyond Gaussian distributions, and to investigate the performance as a function of the size of the region of interest. Nevertheless, it appears that the log-likelihood methods rely on a safer theoretical background and have better efficiency, thus we believe that, should the above results be confirmed in a broader variety of contexts, the log-likelihood method should be preferred in practical applications for best robustness.

Acknowledgements. This work was funded by the ANR-14-LAB3-0006-01 LabCom AtysCrea and was supported by the LABEX CeLyA (ANR-10-LABX-0060) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

References

- [1] J. Kim, J.W. Fisher, A. Yezzi, M. Cetin, and A.S. Willsky, "A nonparametric statistical method for image segmentation using information theory and curve evolution," *Image Processing, IEEE Transactions on*, vol. 14, no. 10, pp. 1486–1502, Oct 2005.
- [2] L. Meziou, A. Histace, and F. Precioso, "Alpha-divergence maximization for statistical region-based active contour segmentation with non-parametric pdf estimations," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 861–864.
- [3] Gilles Aubert, Michel Barlaud, Olivier Faugeras, and Stéphanie Jehan-Besson, "Image segmentation using active contours: Calculus of variations or shape gradients?," *SIAM Journal on Applied Mathematics*, vol. 63, no. 6, pp. 2128–2154, 2003.
- [4] S. Jehan-Besson, M. Barlaud, G. Aubert, and O. Faugeras, "Shape gradients for histogram segmentation using active contours," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, Oct 2003, pp. 408–415 vol.1.
- [5] F. Lecellier, S. Jehan-Besson, and J. Fadili, "Statistical region-based active contours for segmentation: An overview," *IRBM*, vol. 35, no. 1, pp. 3 – 10, 2014.
- [6] D. Freedman, R.J. Radke, Tao Zhang, Yongwon Jeong, D.M. Lovelock, and G.T.Y. Chen, "Model-based segmentation of medical imagery by matching distributions," *Medical Imaging, IEEE Transactions on*, vol. 24, no. 3, pp. 281–292, March 2005.
- [7] R. Delgado-Gonzalo, V. Uhlmann, D. Schmitter, and M. Unser, "Snakes on a plane: A perfect snap for bioimage analysis," *Signal Processing Magazine, IEEE*, vol. 32, no. 1, pp. 41–48, Jan 2015.
- [8] O. Michailovich, Y. Rathi, and A. Tannenbaum, "Image segmentation using active contours driven by the bhattacharyya gradient flow," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2787–2801, Nov 2007.
- [9] A. Herbulot, S. Jehan-Besson, M. Barlaud, and G. Aubert, "Shape gradient for image segmentation using information theory," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, May 2004, vol. 3, pp. iii–21–4 vol.3.