A PROBABILISTIC INTERPRETATION OF GEOMETRIC ACTIVE CONTOUR SEGMENTATION

Jonas De Vylder, Dirk Van Haerenborgh, Jan Aelterman and Wilfried Philips

Department of Telecommunications and Information Processing,
iMinds - Image Processing and Interpretation,
Ghent University, St-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

ABSTRACT

Active contours or snakes are widely used for segmentation and tracking. These techniques require the minimization of an energy function, which is typically a linear combination of a data-fit term and regularization terms. This energy function can be tailored to the intrinsic object and image features. This can be done by either modifying the actual terms or by changing the weighting parameters of the terms. There is, however, no sure way to set these terms and weighting parameters optimally for a given application. Although heuristic techniques exist for parameter estimation, often trial and error is used. In this paper, we propose a probabilistic interpretation to segmentation. This approach results in a generalization of state of the art active contour segmentation. In the proposed framework all parameters have a statistical interpretation, thus avoiding ad hoc parameter settings.

Index Terms— Active contours, segmentation, convex optimization, statistical estimator

1. INTRODUCTION

Since Kass et al. [1] introduced the active contours, the framework has become a constant recurring topic in segmentation and tracking literature [2, 3, 4, 5, 6]. In the active contour framework, an initial contour is moved and deformed in order to minimize a specific energy function. This energy function should be minimal when the contour is delineating the object of interest. Two main groups can be distinguished in the active contour framework: one group representing the active contour explicitly as a parametrized curve, a second group of techniques represent the contour implicitly, e.g. using level-sets. In the first group, also called snakes, the contour commonly converges towards edges in the image [1, 5]. The second group generally has an energy function based on region properties, such as the variance of intensity of the enclosed segment [3, 7]. These level-set approaches have gained a lot of interest since they have some benefits over snakes. For example, they don't need any parametrization and can easily change their topology, e.g. splitting a segment into multiple unconnected segments.

Level-set based techniques have some benefits over parametrized active contours, but they still have one major drawback in common: the energy function is generally non-convex, thus the end result depends on the initialization. This can be overcome by choosing a different representation than level-sets to represent the contour. By representing the contour using a characteristic function, i.e. a binary function where one represents a foreground pixel and a zero corresponds with a background pixel, a convex energy function can be found which minimizes the popular active contours without edges (ACWE) [8, 3]. This convexity allows for generally faster and global optimization of the energy function [4, 6]. The benefit is that their results no longer depend on the initialization.

While the active contour framework shows good results for numerous applications, it is not always straightforward to tune the parameters in order to get an optimal result. Some work has interpreted the active contour framework using probability theory [9, 2, 10], giving a probabilistic meaning to the parameters. These methods show good results, avoiding ad hoc parameter settings, but these approaches are either limited to parametric active contours or restricted to specific shape priors. The parametric active contours restrict the methods to segmentation of objects with fixed topology and good initialization. Geometric active contours with a probabilistic approach either limit the usability by imposing strong but restrictive shape priors [9], or start from a probabilistic interpretation for the data-fit but but use generic shape energies without any probabilistic interpretation, thus returning to ad hoc parameter tuning [11]. In this work we propose a probabilistic interpretation for geometric global optimal active contours. In contrast to many methods in literature, both the data-fidelity and the shape prior are statistically modelled within our framework. The proposed method corresponds to the minimization of a convex energy term where all parameters have a statistical interpretation. This allows to calculate the optimal parameters from a training data set.

This paper is arranged as follows: the next section provides a brief description of convex energy active contours. In section 2 our proposed algorithm is presented, while the section 3 elaborates on the optimization of our proposed method. Section 4 shows some results of our technique. Section 5 recapitulates and concludes.

2. PROBABILISTIC ACTIVE CONTOURS

While the active contour framework has been proven useful for segmentation [12, 13, 14], it is not straightforward to define optimal parameters. Furthermore many alternative regularization terms and data-fit energy terms have been proposed. This makes it difficult to define a good energy function for a given segmentation problem. We look at a segmentation problem as an inverse problem, which we solve using estimation theory. Assume that an image is the result of a function, $g(.)$, which maps pixels belonging to a segment to a specific intensity, while mapping background pixels to different intensities. This model is not perfect, e.g. there can be noise in the image or blur or other objects might be visible in the image, which are not of interest for the end user, etc. We include these discrepancies in the model as noise, $n(.)$. The image $f(.)$ can be defined as:

$$f(\vec{x}) = g(u(\vec{x})) + n(u(\vec{x})) \tag{1}$$

with $u(.)$ a characteristic function, i.e. a binary function which has results one for pixels belonging to the segment, and zero for all other pixels. Note that we consider the noise to be dependent on the segmentation. Using eq. (1), an optimal segmentation result can then be calculated in a Maximum A Posteriori (MAP) sense as:

$$\hat{u}(.) = \arg\max_{u(.)\in S} \log \mathrm{p}_{U}\left(u(.)\right) + \log \mathrm{p}_{F|U}\left(f(.) \mid u(.)\right) \tag{2}$$

Where S represents the set of characteristic functions. The quality of the segmentation resulting from this MAP estimator strongly depends both on the knowledge of the mapping function, $g(.)$, on the noise of the model, $n(.)$, and on the prior knowledge that can be used on $u(.)$. We will discuss different priors and image and noise models in the next two subsections.

2.1. Likelihood

Assume a simple piecewise constant data-fit model, i.e. that an image has a constant value for the segment pixels and a different constant value for all background pixels. This results in the following image model:

$$g(a) = \begin{cases} \mu_f & \text{if } a = 1 \\ \mu_b & \text{if } a = 0 \end{cases} \tag{3}$$

We assume that the noise statistics on this model can differ between segment pixels and background. If we model the noise for both segment and background using a Normal distribution, then the likelihood of $f(.)$ corresponds to:

$$\mathrm{p}_{F|U}\left(f(\vec{x}) \mid u(\vec{x})\right) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_f^2}} e^{\frac{-(f(\vec{x})-\mu_f)^2}{2\sigma_f^2}} & \text{if } u(\vec{x}) = 1 \\ \frac{1}{\sqrt{2\pi\sigma_b^2}} e^{\frac{-(f(\vec{x})-\mu_b)^2}{2\sigma_b^2}} & \text{if } u(\vec{x}) = 0 \end{cases} \tag{4}$$

Since we assume that $u(.)$ represents a characteristic function, i.e. $u(\vec{x}) \in \{0, 1\}$, we can rewrite the log likelihood as follows

$$\begin{aligned} & \log \mathrm{p}_{F|U}\left(f(\vec{x}) \mid u(\vec{x})\right) \\ = \ & u(\vec{x})\left(\frac{-(f(\vec{x})-\mu_f)^2}{2\sigma_f^2} + \log\frac{1}{\sqrt{2\pi\sigma_f^2}}\right) \\ & +(1-u(\vec{x}))\left(\frac{-f(\vec{x})-\mu_b)^2}{2\sigma_b^2} + \log\frac{1}{\sqrt{2\pi\sigma_b^2}}\right) \\ = \ & u(\vec{x})\left(\frac{1}{2\sigma_b^2}(f(\vec{x})-\mu_b)^2 - \frac{1}{2\sigma_f^2}(f(\vec{x})-\mu_f)^2\right) \\ & +u(\vec{x})\left(\log\frac{1}{\sqrt{2\pi\sigma_f^2}} - \log\frac{1}{\sqrt{2\pi\sigma_b^2}}\right) + c \end{aligned} \tag{5}$$

with c a constant independent of $u(\vec{x})$, thus we can ignore this constant with regard to the MAP estimator. Note that for $\sigma_f = \sigma_b$ that this comes down to the same data-fit term used in the piecewise constant Mumford-Shah model [8]. This also gives a statistical interpretation for the weighting parameter in the ACWE model [3, 4]. This probabilistic interpretation is more general than the model used in the ACWE model, i.e. the model allows different variances for segment and background noise. If the variance of the segment noise is bigger than the background noise, i.e. $\sigma_f > \sigma_b$, then a positive bias is added for all foreground pixels, thus favoring foreground pixels in the MAP estimator. If this model is used in a maximum likelihood estimator (ML) instead of a MAP estimator, i.e. without any prior knowledge on $u(.)$, this actually comes down to hard thresholding.

2.2. Segmentation prior

The quality of the MAP estimator will depend on the prior used. While many shape priors exist for specific applications, we will restrict our priors to simple but generic priors. Common generic shape priors use the area, m_1, and perimeter, m_2, of the detected objects [8, 3]. For a continuous characteristic function, i.e. a binary function, these measurements can respectively be expressed as $m_1(u(.)) = |u(.)|$ and $m_2(u(.)) = |\nabla u(.)|$ (see [15] for more details). In this work we model the probability of the area/perimeter of a segment using a generalized Laplacian, i.e.

$$\mathrm{p}_{M_i}(m_i(u(.))) = \frac{\beta_i}{2\alpha_i\Gamma(\frac{1}{\beta_i})} e^{-\left(\frac{|m_i(u(.))-\mu_i|}{\alpha_i}\right)^{\beta_i}} \quad (6)$$

where $\Gamma(x)$ represents the Gamma function, i.e. $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. For simplicity we assume that area and perimeter are independent of each other. This results in the following segmentation prior:

$$\begin{aligned} &\log \mathrm{p}_U(u(.)) \\ = &-\left(\frac{||u(.)|-\mu_1|}{\alpha_1}\right)^{\beta_1} + \log\left(\frac{\beta_1}{2\alpha_1\Gamma(\frac{1}{\beta_1})}\right) \\ &-\left(\frac{||\nabla u(.)|-\mu_2|}{\alpha_2}\right)^{\beta_2} + \log\left(\frac{\beta_2}{2\alpha_2\Gamma(\frac{1}{\beta_2})}\right) \quad (7) \end{aligned}$$

Note that since $u(\vec{x})$ is positive for all $\vec{x}$, that the term related to $m_1(u(.))$ is convex. However the term related to $m_2(u(.))$ is not necessarily convex. In practice the ML estimator of $u(.)$ generally results in a segment with a bigger perimeter than the real object of interest. This is due to noise, clutter, blur, etc. By assuming that the perimeter of $u(.)$ is bigger than μ_2, we can simplify ex. (7) to the following convex prior:

$$\begin{aligned} \log \mathrm{p}_U(u(.)) &= -\left(\frac{||u(.)|-\mu_1|}{\alpha_1}\right)^{\beta_1} \quad (8) \\ &-\left(\frac{|\nabla u(.)|-\mu_2}{\alpha_2}\right)^{\beta_2} + c \end{aligned}$$

with c a constant independent of $u(.)$, thus irrelevant with regard to the MAP estimator.

3. OPTIMIZATION

Combining the likelihood of eq. (5) and the prior in eq. (8) results in the following MAP estimator:

$$\begin{aligned} \hat{u}(.) &= \arg\min_{u(.)\in S} \left(\frac{||u(.)|-\mu_1|}{\alpha_1}\right)^{\beta_1} + \left(\frac{|\nabla u(.)|-\mu_2}{\alpha_2}\right)^{\beta_2} \\ &+\langle u(.), r(.)\rangle \quad (9) \end{aligned}$$

with $\langle .,. \rangle$ the inner product and

$$\begin{aligned} r(\vec{x}) &= \left(\frac{1}{2\sigma_f^2}(f(\vec{x})-\mu_f)^2 - \frac{1}{2\sigma_b^2}(f(\vec{x})-\mu_b)^2\right) \\ &+\left(\log\frac{1}{\sqrt{2\pi\sigma_b^2}} - \log\frac{1}{\sqrt{2\pi\sigma_f^2}}\right) \quad (10) \end{aligned}$$

Since all parameters have a statistical interpretation, they can easily be estimated based on a small training data set. Note that S, the set of binary functions, is a non convex set. In [4, 16] this problem was circumvented by relaxing the binary constrained to functions with co-domain equal to $[0, 1]$, thus optimizing over the convex set S'. After finding the optimal $\hat{u}(.)$, the function is binarised by thresholding the function. This binarised version is also a global optimizer of eq. (9) under specific conditions (This is proven in [17]), but even when these conditions were not met, the binarised version showed good segmentation results. Although the optimization strategy is generic for all β_i, the actual optimization scheme, i.e. the explicit formulas depend on β_i. As an example, we will consider the situation where $\beta_i = 1$, i.e. where the prior is modeled using Laplacian distributions. Then the MAP estimator in (9) can be rewritten using the matrix-vector formulation:

$$\hat{\vec{u}} = \arg\min_{\vec{u}\in S'} \left(\frac{||\vec{u}|-\vec{\mu}_1|}{\alpha_1}\right) + \frac{|\nabla\vec{u}|}{\alpha_2} + \langle\vec{u}, \vec{r}\rangle \quad (11)$$

Where $\vec{\mu}_1$ represents a vector of the same length as $\vec{u}$, with all elements equal to μ_1. Applying the Legendre-Frenchel transform [18] results in the following primal dual formulation:

$$\hat{u}(.) = \arg\min_{u(.)\in S'} \max_{y\in Y} \langle\vec{u}, K^H\vec{y}\rangle + \langle\vec{u}, \vec{r}\rangle + \delta_Y(\vec{y}) \quad (12)$$

with $\vec{y} \in Y$ the dual variable, $.^H$ denotes the Hermitian transpose, and $K = \begin{bmatrix} \nabla \\ I \end{bmatrix}$. The function $\delta_Y(.)$ denotes the indicator function of the convex set Y:

$$\delta_Y(\vec{y}) = \begin{cases} 0 & \text{if } \vec{y} \in Y \\ \infty & \text{if } \vec{y} \notin Y \end{cases} \quad (13)$$

Where the convex set Y is given by:

$$Y = \{\vec{y} \in Y : \|\vec{y} - \vec{b}\|_\infty \leq 1\}$$

with $\vec{b} = \begin{bmatrix} \vec{0} \\ \vec{\mu}_1 \end{bmatrix}$. The primal dual formulation in eq. (12) can be optimized using the following iterative steps:

$$\begin{aligned} \vec{y}_{n+1} &= \mathrm{prox}_\delta(\vec{y}_n - K\vec{z}_n) \quad (14) \\ \vec{x}_{n+1} &= \mathrm{prox}_{<.,\vec{r}>}(\vec{x}_n - K^H\vec{y}_{n+1}) \quad (15) \\ \vec{z}_{n+1} &= \vec{x}_{n+1} + \theta(\vec{x}_{n+1} - \vec{x}_n) \quad (16) \end{aligned}$$

Since $\delta_Y(.)$ is an indicator function of a convex set, the proximal operator reduces to pointwise Euclidean projectors on L^2 balls [18]:

$$\mathrm{prox}_\delta(\vec{p}) = \vec{b} + \frac{\vec{p}-\vec{b}}{\max(1, |\vec{p}-\vec{b}|)} \quad (17)$$

The proximal operator corresponding with the inner product can be calculated using the following closed formula:

$$\text{prox}_{<.,\vec{r}>}(\vec{p}) = \vec{p} - \vec{r} \tag{18}$$

4. EXPERIMENTAL RESULTS

The proposed framework is in the first place a statistical justification of the ACWE model, which already has proven its merit for numerous applications [3, 4, 6]. However the proposed work is more general than the ACWE model, which assumes that the intensity variance of segments should be equal to the intensity variance of the background. The proposed work is not hampered by this constraint. An example of such an application is shown in Fig. 1(a). The figure shows a fluorescence microscopy image of pathological nuclei. Due to certain pathologies, the cell nuclei does not have a sharp boundary, but appears frayed. This results in higher variance in intensity for segment pixels. The background however only has low variance in the background. The segmentation result using the convex energy ACWE model [4] is shown in Fig. 1.(b). The expected intensity for segment and background pixels are estimated from a manually annotated training dataset. The weighting parameters were empirically chosen to get the best result. The result of the proposed method is shown in Fig 1.(c). The segmentation is more accurate, which can be seen at the thin section of the lower nucleus at the left, or at the fraying part of the two right nuclei.

For a qualitative validation of the segmentation results, we use the Dice coefficient as a quality metric. If S is the resulting segment from the active contour and GT the ground truth segment, then the Dice coefficient between S and GT is defined as: $d(S, GT) = \frac{2\ \text{Area}(S \wedge GT)}{\text{Area(S)}+\text{Area}(GT)}$ where $S \wedge GT$ consist of all pixels which both belong to the detected segment as well as to the ground truth segment. If S and GT are equal, the Dice coefficient is equal to one. The Dice coefficient will approach zero if the regions hardly overlap. We analyzed a larger data-set containing 28 nuclei. These nuclei where manually delineated in order to get ground truth. The proposed method resulted in an average Dice coefficient of 0.973, compared to an average Dice coefficient of 0.957 using the classical ACWE model.

5. CONCLUSION

This paper proposes a statistical interpretation to the famous geometric active contour framework. The work describes the segmentation problem as an inverse problem. This approach results in a generalization of the ACWE model, where all parameters have a statistical interpretation, thus avoiding ad hoc parameter tuning. The segmentation result can be calculated using an efficient optimizer based on primal-dual optimization. The method described is the result of assuming specific distributions for the noise on the model. However the same reasoning can be applied for other noise distributions. This allows to calculate specific active contour models, based on the statistics of both the objects of interest and on the statistics of the image. Furthermore we would like to emphasise that this is a generic framework. We used a Gaussian model for intensity modeling and simple shape priors such as area and perimeter. However other shape priors and other probability distributions could be used as well. Another possible extension is not to model the global intensity but local intensity or local features such as texture could be incorporated in our statistical framework.

6. ACKNOWLEDGEMENTS

The authors would like to thank dr. Winnok De Vos (Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University) for sharing the fluorescent microscopic images and for manually generating ground truth.

7. REFERENCES

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes - active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1987.

[2] M. Isard and A. Blake, *Active contours*, Springer, 1998.

[3] T. Chan and L. Vese, "An active contour model without edges," *Scale-Space Theories in Computer Vision*, vol. 1682, pp. 141–151, 1999, Bs60h Times Cited:50 Cited References Count:21 Lecture Notes in Computer Science.

[4] T. F. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *Siam Journal on Applied Mathematics*, vol. 66, no. 5, pp. 1632–1648, 2006.

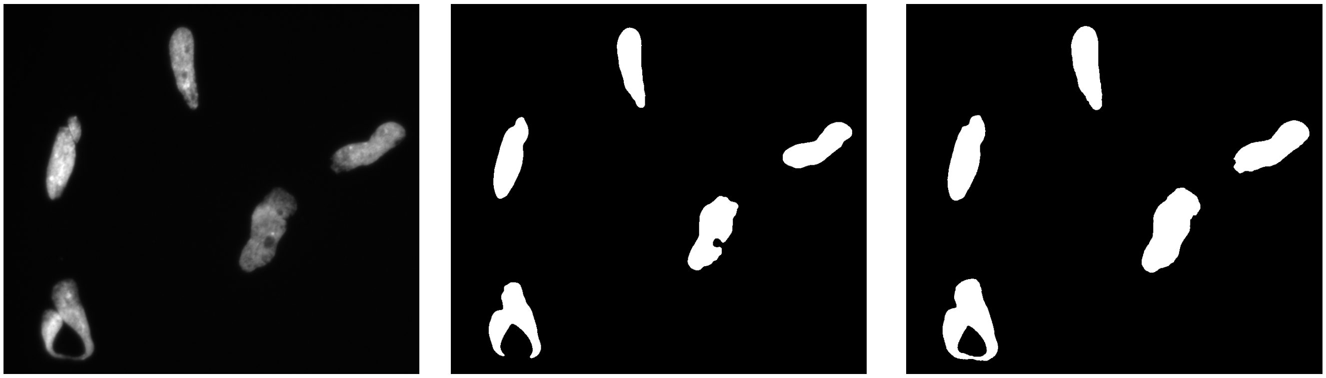
[5] M. A. Charmi, S. Derrode, and F. Ghorbel, "Fourier-based geometric shape prior for snakes," *Pattern Recognition Letters*, vol. 29, no. 7, pp. 897–904, May 1 2008.

[6] T. Goldstein, X. Bresson, and S. Osher, "Geometric applications of the split bregman method: Segmentation and surface reconstruction," *Journal of Scientific Computing*, vol. 45, no. 1-3, pp. 272–293, Oct 2010.

[7] R. Goldenberg, R. Kimmel, E. Rivlin, and M. Rudzsky, "Fast geodesic active contours," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1467–1475, Oct 2001.

[8] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989.

[9] D. Cremers, F. R. Schmidt, and F. Barthel, "Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions," *2008 IEEE Conference on Computer Vision and Pattern Recognition, Vols 1-12*, pp. 810–815, 2008.



(a) original image (b) segmentation using ACWE (c) segmentation using proposed method

Fig. 1. An example for biomedical segmentation. (a) shows cell nuclei images using fluorescence microscopy. (b-c) the segmentation results.

[10] J. De Vylder, D. Ochoa, L. Chaerle, D. Van Der Straeten, and W. Philips, "Leaf segmentation and tracking using probabilistic parametric active contours," in *Proc. of Mirage 2011*. October 2011, Lecture Notes in Computer Science, Springer Verlag.

[11] Fuhua Chen and Yunmei Chen, "A stochastic variational model for multi-phase soft segmentation with bias correction," *Advanced Modeling and Optimization. v12 i3*, pp. 339–345, 2010.

[12] J. A. Sethian, "A review of level set and fast marching methods for image processing," *Modern Methods in Scientific Computing and Applications*, vol. 75, pp. 365–396, 2002.

[13] L. X. Gong, L. Ng, S. D. Pathak, I. Tutar, P. S. Cho, D. R. Haynor, and Y. Kim, "Prostate ultrasound image segmentation using level set-based region flow with shape guidance," *Medical Imaging 2005: Image Processing, Pt 1-3*, vol. 5747, pp. 1648–1657 2200, 2005, Part 1-3 Bcn01 Times Cited:3 Cited References Count:43 Proceedings of the Society of Photo-Optical Instrumentation Engineers (Spie).

[14] J. Mille, "Narrow band region-based active contours and surfaces for 2d and 3d segmentation," *Computer Vision and Image Understanding*, vol. 113, no. 9, pp. 946–965, Sep 2009, 474SB Times Cited:0 Cited References Count:51.

[15] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, Studies in advanced mathematics. CRC Press, Boca Raton, 1992.

[16] X. Bresson, S. Esedoglu, P. Vandergheynst, J. P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 2, pp. 151–167, Jun 2007.

[17] X. Bresson and T. F. Chan, "Active contours based on chambolle's mean curvature motion," *2007 IEEE International Conference on Image Processing, Vols 1-7*, pp. 33–36 3371, 2007.

[18] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.