IMAGE SEGMENTATION WITH BACKGROUND CORRECTION USING A MULTIPLICATIVE SMOOTHING-SPLINE MODEL

Ramtin Madani, Aurélien Bourquard, and Michael Unser

Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), Switzerland

ABSTRACT

This paper presents an image-segmentation method which compensates multiplicative distortions based on smooth regularity assumptions. In this work, we generalize the original Chan-Vese functional to handle a continuous multiplicative bias. In the derivation of our model, we show that the optimal correction function is necessarily a spline, which we express in terms of discrete coefficients. Following an iterative technique, we propose to find the solution by an alternate optimization of this map and of the segmented domains. In order to maximize the overall efficiency, graph cuts are combined with a specifically designed multigrid algorithm. Our experiments demonstrate the relevance of our approach for biomedical data.

***Index Terms*—** Chan-Vese model, graph cuts, image segmentation, intensity inhomogeneity, multigrid methods, nonuniform illumination, B-splines, variational methods.

1. INTRODUCTION

The acquisition of biomedical images often involves some sort of bias, which can be attributed to several factors among which nonuniform illumination, static-field inhomogeneity, or reception-coil-sensitivity inhomogeneity in the case of MRI. These distortions render the analysis or post-processing of the data intricate [1]. In particular, most classical segmentation methods suffer from such biases.

There is an important dependency between intensity-inhomogeneity correction and segmentation. Indeed, knowing the solution of either of the two problems causes the other one to have a simpler form [2]. As a consequence, approaches for joint segmentation and bias correction have been proposed by several authors. In order to yield a satisfactory result, these techniques are iterative and rely on specific models for the bias [3].

In the presence of noise or similar visual artifacts, segmentation models typically entail the minimization of an energy functional according to some shape prior. The Chan-Vese model, for instance, regularizes the length and area of the solution contours, while being based on the expected intensity values of the background and object of interest [4]. In this framework, Wang *et al.* have provided a modified energy functional which compensates local multiplicative inhomogeneities in the image [5]. Meanwhile, other researchers have proposed to compensate a global multiplicative bias based on a log-type formulation [6].

In this paper, we propose a variational method based on a generalized Chan-Vese functional for the joint segmentation and multiplicative-bias compensation of an image. This technique allows for precise segmentation as well as accurate global correction of image intensities. The key aspect of our strategy lies in the use of a correction term which models the bias as a multiplicative smoothing-spline function. In order to efficiently retrieve this correction term along with the segmented image, we devise an iterative minimization strategy consisting in a multigrid algorithm combined with graph cuts. We perform several experiments on MRI and autofluorescence-microscopy data which illustrate the applicability of our approach.

2. SEGMENTATION MODEL

Let $v : \Omega \subset \mathbb{R}^2 \longmapsto \mathbb{R}$ be the continuous-domain intensities of a given image. Considering the two-zone segmentation problem, we are looking for two open sets $\Omega_1 \subset \Omega$ and $\Omega_2 = \Omega \setminus \overline{\Omega}_1$ which correspond to the region of interest and the background of the image, respectively. The spatial distribution of Ω_i is represented by the characteristic function $\chi : \Omega \longmapsto \{0, 1\}$ as

$$\chi(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \Omega_1 \\ 0, & \mathbf{x} \in \overline{\Omega}_2. \end{cases} \tag{1}$$

In this work, according to [7] we model the intensity inhomogeneity of v as a slowly varying multiplicative term $s : \Omega \longmapsto \mathbb{R}$ such that

$$v(\mathbf{x}) = s(\mathbf{x})u(\mathbf{x}) + n(\mathbf{x}), \tag{2}$$

where u and n are the inhomogeneity-free intensity function and some additive Gaussian noise, respectively.

According to the segmentation model of [4], a piecewise-constant approximation of the corrected image can be inferred

This work was funded (in part) by the Swiss National Science Foundation under Grant 200020-121763.

from χ as

$$\begin{aligned} u(\mathbf{x}) &\approx v(\mathbf{x})/s(\mathbf{x}) \\ &\approx \chi(\mathbf{x})q_1 + (1-\chi(\mathbf{x}))q_2, \end{aligned} \tag{3}$$

where the scalars q_i are the expected intensities of Ω_i.

As long as the effect of s is negligible as it approaches unity, the Chen-Vese formulation [4] can be used to find q_1, q_2, and χ. This amounts to minimizing the energy

$$\begin{aligned} F(q_1,q_2,\chi) &= \mu V(\chi,\Omega) + \nu\|\chi\|_{L_1(\Omega)} \\ &\quad + \gamma_1\|\chi(u-q_1)\|^2_{L_2(\Omega)} \\ &\quad + \gamma_2\|(1-\chi)(u-q_2)\|^2_{L_2(\Omega)}, \end{aligned} \tag{4}$$

where V is the total-variation operator (*i.e.*, $V(\chi,\Omega)$ represents the length of the boundary $\partial\Omega_1$ of Ω_1), $\gamma_1,\gamma_2>0$, and $\mu,\nu\geq 0$ are regularization constants. This yields the optimal values $q_1 = \frac{1}{\|\chi\|}\int_{\Omega_1} u(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and $q_2 = \frac{1}{\|1-\chi\|}\int_{\Omega_2} u(\mathbf{x})\,\mathrm{d}\mathbf{x}$.

In our case, s can significantly depart from unity, due to the type of data we consider. In order to properly handle this non-ideality, our contribution is to propose a generalized version of the Chen-Vese functional F. For the sake of accuracy, we also account for the sampled nature of the available image. Accordingly, we define our new functional J as

$$\begin{aligned} J(q_1,q_2,s,\chi) &= \mu\mathrm{D}(\chi^1,v^1) \\ &\quad + \|\chi^1(v^1 - s^1 q_1)\|^2_{\ell_2(\Omega)} \\ &\quad + \|(1-\chi^1)(v^1 - s^1 q_2)\|^2_{\ell_2(\Omega)} \\ &\quad + \lambda\|\mathrm{L}s\|^2_{L_2(\Omega)}, \end{aligned} \tag{5}$$

where $s^1 : \Omega\cap\mathbb{Z}^2 \longmapsto \mathbb{R}$ is the sampled sequence defined as $s^1[\mathbf{k}] = s(\mathbf{x})|_{\mathbf{x}=\mathbf{k}}$. The same definitions apply for χ^1 and v^1 relative to χ and v. In this compound energy, D denotes a *graph-representable* regularizer [8]. Compared to (4), the essential difference is the presence of the multiplicative map s; the latter is regularized by the smooth term $\|\mathrm{L}s\|^2_{L_2(\Omega)}$ weighted by $\lambda\geq 0$, where L is a spline-admissible differential operator [9].

Lemma 1. *Assuming that the function s belongs to the generalized Sobolev space $\mathcal{W}_2^{\mathrm{L}}$ associated with L, its optimal form minimizing (5) is an $\mathrm{L}^*\mathrm{L}$ spline.*

Proof. The property $s\in\mathcal{W}_2^{\mathrm{L}}$ implies that $\|\mathrm{L}s\|^2_{L_2}$ is a spline energy which decomposes into

$$\|\mathrm{L}s\|^2_{L_2} = \|\mathrm{L}s_{\text{int}}\|^2_{L_2} + \|\mathrm{L}(s-s_{\text{int}})\|^2_{L_2},$$

where s_{int} is the unique $\mathrm{L}^*\mathrm{L}$ spline interpolating s at integer positions. Since all terms of (5) except $\|\mathrm{L}s\|^2_{L_2}$ are a mere function of the sampled solution s^1, and given the above decomposition, (5) is optimally minimized iff $s = s_{\text{int}}$. In other words, the optimum is itself an $\mathrm{L}^*\mathrm{L}$ spline, which concludes the proof. □

Using Lemma 1, we express the continuous solution s without any loss of generality as

$$s(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^2\cap\Omega} c[\mathbf{k}]\varphi(\mathbf{x}-\mathbf{k}), \tag{6}$$

where φ is the corresponding $\mathrm{L}^*\mathrm{L}$-spline basis function, and where c are the spline coefficients. From (6), we rewrite the cost to be minimized as a function of c. Namely,

$$\begin{aligned} J(q_1,q_2,s,\chi) &= \mu\mathrm{D}(\chi^1,v^1) \\ &\quad + \|\chi^1(v^1 - (\phi\star c)q_1)\|^2_{\ell_2(\Omega)} \\ &\quad + \|(1-\chi^1)(v^1 - (\phi\star c)q_2)\|^2_{\ell_2(\Omega)} \\ &\quad + \lambda\|r\star c\|^2_{\ell_2(\Omega)}, \end{aligned} \tag{7}$$

where $\star$ denotes a discrete convolution, and where the discrete filters ϕ and r are defined as

$$\phi[\mathbf{k}] = \varphi(\mathbf{x})|_{\mathbf{x}=\mathbf{k}} \tag{8}$$

$$r[\mathbf{k}] = \mathrm{L}\varphi(\mathbf{x})|_{\mathbf{x}=\mathbf{k}}, \tag{9}$$

respectively. The problem thus amounts to finding the discrete coefficients c of the continuous map s, given the samples v^1 and the constants q_i.

3. OPTIMIZATION

In order to obtain the segmented zones along with the multiplicative bias s, we have to minimize (7) with respect to q_i, χ, and s. Our strategy is to optimize these quantities sequentially one at the time, keeping the others fixed. Accordingly, the optimal q_i corresponds to the update rule

$$q_i = \frac{\int_{\Omega_i} v(\mathbf{x})/s(\mathbf{x})\,\mathrm{d}\mathbf{x}}{\int_{\Omega_i} 1/s(\mathbf{x})\,\mathrm{d}\mathbf{x}}. \tag{10}$$

Regarding the optimization of χ, we follow a standard graph-cut approach. The details of the graph construction in the general case and the fast minimization algorithm are presented in [8] and [10], respectively. Finally, the minimization of (7) with respect to the coefficients of the correction map s corresponds to the resolution of a quadratic problem which we describe below. Our overall algorithm first initiates the values of χ and q_i. Starting from these initial conditions, it then updates s, χ, and q_i through M successive cycles.

4. MULTIGRID ALGORITHM

The minimum of each functional with respect to the spline coefficients c corresponds to the solution of the linear system

$$\mathbf{A}\mathbf{c} = \mathbf{f} \tag{11}$$

in matrix notation, where

$$\mathbf{A} = \mathbf{B}^T\mathbf{W}\mathbf{B} + \lambda\mathbf{R} \tag{12}$$

$$\mathbf{f} = \mathbf{B}^T\left[(q_1\boldsymbol{\chi} + q_2(\mathbf{I}-\boldsymbol{\chi}))\,\mathbf{v}\right]. \tag{13}$$

The vectors $\mathbf{v}$ and $\mathbf{c}$ contain the corresponding sequences v^1 and c in lexicographic ordering, respectively. The diagonal matrices $\mathbf{W}$ and $\boldsymbol{\chi}$ implement a pointwise multiplication with the sequences $(q_1^2 - q_2^2)\chi^1 + q_2^2$ and χ^1, respectively, and $\mathbf{B}$ and $\mathbf{R}$ are convolution matrices associated to the discrete kernel ϕ and to the finite-difference filter $r \star r[-\cdot]$, respectively.

In practice, the dimensions of $\mathbf{A}$ in (11) are such that this matrix cannot be inverted directly. To find an approximate solution, we resort to an iterative approach to reduce the norm of the error $\boldsymbol{\epsilon} = \mathbf{A}^{-1}\mathbf{f} - \mathbf{c}$, or, equivalently, of the *residual* $\boldsymbol{\rho} = \mathbf{f} - \mathbf{A}\mathbf{c}$. Since it involves no matrix inversion, $\boldsymbol{\rho}$ can be computed exactly. For instance, the gradient descent updates the solution as $\mathbf{c} \leftarrow \mathbf{c} + \omega\boldsymbol{\rho}$ with step ω.

Given the particular properties of the multiplicative bias, most components of the solution $\mathbf{s}$ tend to be low-frequency. In this case, mathematical analysis demonstrates that iterating at additional coarse scales accelerates the convergence [11].

Let us first reformulate Problem (11) at nominal discretization as $\mathbf{A}^1\mathbf{c}^1 = \mathbf{f}^1$, where the superscript h denotes the sampling step in each dimension associated to the grid Ω^h of the domain Ω. Following N_1 iterations at that scale, the remaining error $\boldsymbol{\epsilon}^1$ tends to be mostly low-frequency, as these components are less efficiently reduced asymptotically [11].

Our strategy to improve the convergence rate without increasing N_1 is to transfer the corresponding residual $\boldsymbol{\rho}^1 = \mathbf{A}\boldsymbol{\epsilon}^1$ onto a coarser grid Ω^2. This transfer operation is performed as $\boldsymbol{\rho}^2 = \mathbf{I}_h^{2h}\boldsymbol{\rho}^1$, where $\mathbf{I}_h^{2h}$ is a *restriction operator* that resamples the data from Ω^1 to Ω^2. At that resolution, the remaining low-frequency components of $\boldsymbol{\epsilon}^1$ can then be identified efficiently through $\boldsymbol{\epsilon}^2$, iterating on the coarse problem $\mathbf{A}^2\boldsymbol{\epsilon}^2 = \boldsymbol{\rho}^2$, where $\mathbf{A}^2$ is a low-resolution counterpart of $\mathbf{A}^1$. The components of $\boldsymbol{\epsilon}^1$ become essentially high-frequency in $\boldsymbol{\epsilon}^2$, which is associated to higher convergence rates.

The obtained $\boldsymbol{\epsilon}^2$ can then be used to correct $\mathbf{c}^1$ after resampling it back to Ω^1. Using the *prolongation operator* $\mathbf{I}_{2h}^h$ as the dual of $\mathbf{I}_h^{2h}$, we write these operations as $\mathbf{c}^1 \leftarrow \mathbf{c}^1 + \mathbf{I}_{2h}^h\boldsymbol{\epsilon}^2$. In order to compensate for the approximate grid transfers, the solution is again iterated N_2 times on Ω^1.

In our implementation, we follow a recursive version of this approach called full-multigrid cycles [11], which involves N_0 update cycles through V distinct grids. At each grid, we use successive over-relaxation as an iterative method. The definition of the transfer operators depends on the basis φ. For instance, using tensor-product B-splines of order n yields

$$\mathbf{I}_h^{2h} = \mathbf{D}_2\mathbf{H}_2^T, \tag{14}$$

where $\mathbf{D}_2$ is a downsampling-by-two matrix, and where $\mathbf{H}_2$ is a circulant matrix that corresponds to the scaling filter h_2 of order n defined in [12]. Regarding the coarser-scale expressions of $\mathbf{A}^1$, we constrain their general form to maintain

$$\mathbf{A}^h = \mathbf{B}^T\mathbf{W}^h\mathbf{B} + \lambda\mathbf{R}^h. \tag{15}$$

According to the *variational properties* [11], an ideal coarse-grid matrix $\mathbf{A}^{2h}$ should be defined as $\mathbf{A}^{2h} = \mathbf{I}_h^{2h}\mathbf{A}^h\mathbf{I}_{2h}^h$. Applying this principle to our convolutive terms $\mathbf{R}^h$, we obtain the spatial-domain relation

$$r^{2h} = \{h_2 \star r^h \star {h_2}^T\}_{\downarrow 2}, \tag{16}$$

where $_{\downarrow 2}$ denotes a downsampling by 2 of a sequence in each dimension, and where $r^1 = r$. Regarding the terms $\mathbf{B}^T\mathbf{W}^h\mathbf{B}$, we rather resort to a pyramid of weights because the variational properties do not comply with a simple matrix structure. Defining $w^1 = w$, the diagonal matrix $\mathbf{W}^h$ is found through the spatial-domain relation

$$\tilde{w}^{2h}[\mathbf{k}] = \{{h_2}^T \star w^h\}_{\downarrow 2}[\mathbf{k}]. \tag{17}$$

5. EXPERIMENTS

In this section, we apply our algorithm to biomedical data. In our implementation, we specify L as the Laplacian operator. Its nullspace is adapted to the segmentation of images with linear background variations. This choice corresponds to a fourth-order *polyharmonic spline* which is spanned by radial basis functions [13]. For computational reasons, we select the tensor-product B-spline basis functions (*i.e.*, cubic splines) first mentioned in Section 4. In our graph-cut implementation, we adapt the 8-neighbor-system regularizer D presented in [14] to the characteristic function χ.

In the experiments described below, we have obtained our results using less than 4 iterations in each case. The code has been implemented in Java.

We show in Figure 1 a result of our segmentation method applied to the image of a *Caenorhabditis elegans* acquired with a point confocal microscope, where the intensities are nonuniform due to some parts that are in defocus. We note that the correction of the background allows one to properly identify the segmented areas of the biological structure. As expected, the multiplicative background correction takes higher values when the uncorrected image has a low intensity in the corresponding locations. We also note the high smoothness of the correction map due to regularization, and its relative independence with respect to the structure of the object. The estimated inhomogeneity factor in (d) lies approximately in an order-of-magnitude range.

In Figure 2, we show an MRI scan of a brain where the presence of system coils has resulted in a bias field. In this second experiment, we compare this image to its background-corrected version using our method. Taking the weighted average of twelve distinct coil acquisitions as a reference, the observed SNR improvement reaches about 6 dB.

6. CONCLUSIONS

We have devised a segmentation method. It is based on a smoothing-spline model and can be used to compensate for multiplicative distortions of image intensities. The results obtained on real biomedical data are encouraging, and we have

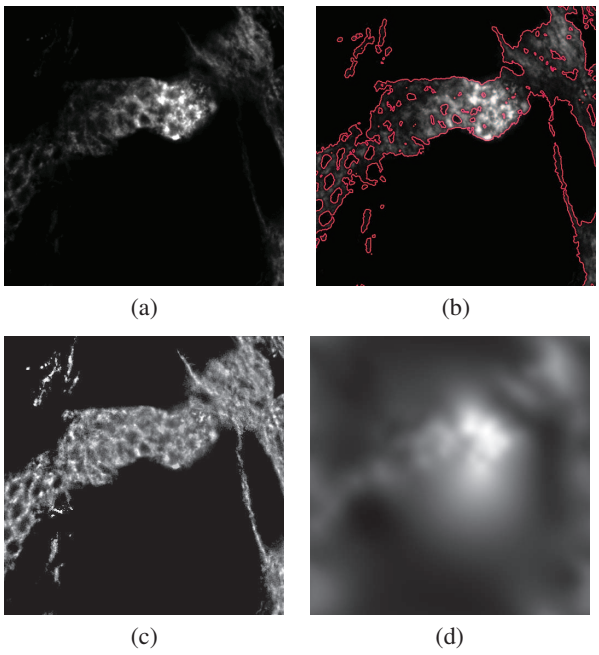

Fig. 1. Simultaneous segmentation and intensity correction of a microscopic image: (a) original image, (b) segmented data, (c) background-corrected image, (d) estimated inhomogeneity.

been able to perform a satisfying segmentation despite high intensity variations in the available data. While the method is inherently iterative, it tends to stabilize after relatively few updates of the solution.

7. ACKNOWLEDGMENTS

The images used in the experimental section are courtesy of François Lazeyras, CIBM, University Hospitals of Geneva, Switzerland, and of Olivier Burri, Swiss Federal Institute of Technology, Lausanne, Switzerland.

8. REFERENCES

[1] T. Tasdizen, E. Jurrus, and R.T. Whitaker, "Non-uniform illumination correction in transmission electron microscopy," in *MICCAI Workshop on Microscopic Image Analysis with Applications in Biology*, New York City, USA, September 5-6, 2008.

[2] U. Vovk, F. Pernuš, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *IEEE Transactions on Medical Imaging*, vol. 26, no. 3, pp. 405–421, 2007.

[3] W.M. Wells, III, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 429–442, 1996.

[4] T.F. Chan and L.A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

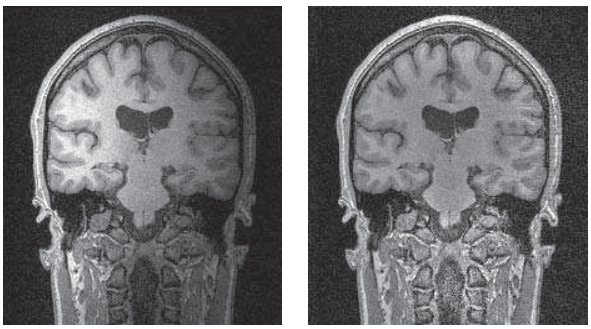

Fig. 2. Background correction of an MRI scan of a brain: (a) original data acquired with one of twelve coils, (b) background-corrected result.

[5] X.F. Wang, D.S. Huang, and H. Xu, "An efficient local Chan-Vese model for image segmentation," *Pattern Recognition*, vol. 43, no. 3, pp. 603–618, 2010.

[6] A.W.C. Liew and H. Yan, "An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation," *IEEE Transactions on Medical Imaging*, vol. 22, no. 9, pp. 1063–1075, 2003.

[7] J.G. Sled, A.P. Zijdenbos, and A.C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Transactions on Medical Imaging,*, vol. 17, no. 1, pp. 87–97, 1998.

[8] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[9] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the Wiener filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2146–2159, 2005.

[10] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[11] W.L. Briggs, V.E. Henson, and S.F. McCormick, *A Multigrid Tutorial*, vol. 72, Society for Industrial Mathematics, 2000.

[12] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.

[13] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Sobolev spaces," *Constructive Theory of Functions of Several Variables*, pp. 85–100, 1977.

[14] N. El Zehiry, S. Xu, P. Sahoo, and A. Elmaghraby, "Graph cut optimization for the Mumford-Shah model," in *The Seventh IASTED International Conference on Visualization, Imaging and Image Processing*. ACTA Press, Palma de Mallorca, Spain, August 29-31, 2007, pp. 182–187.