Local Region Based Medical Image Segmentation Using J-Divergence Measures

Wanlin Zhu, Tianzi Jiang and Xiaobo Li
Medical Imaging and Computing, National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences

Abstract—In this paper, we propose a novel variational formulation. The originality of our formulation is on the use of J-divergence (symmetrized Kullback-Leibler divergence) for the dissimilarity measure between local and global regions. The intensity of a local region is assumed to follow Gaussian distribution. Thus, two features - mean and variance of the distribution of every voxel are introduced to ensure the robustness of the algorithm when noise appeared. Then, J-divergence is used to measure the “distance” between two distributions. The proposed method is verified on synthetic and real medical images. The experimental results are very encouraging for medical image segmentation.

I. INTRODUCTION

Image segmentation plays a major role in medical image analysis since the segmentation results often influence further image analysis for diagnoses and therapy. However, medical images suffer from digital noise. In order to accurately delineate objects in images, robust algorithms that are insensitive to noise and poor image contrast are expected.

There are many segmentation algorithms in literature. Among them, deformable models are very effective and popular. As in [1], deformable models can be classified into two categories. One is parametric and another geometric. Parametric deformable models represent shapes explicitly in their parametric form. This model has difficulty to handle topology changes and hard to be implemented in 3D image applications. Another drawback is the numerical instability when two points are very close. Geometric deformable models use implicit shape representations and evolve curves with a higher dimension function. It based on the level set theory [2], which was first developed in fluid dynamics. With the long development of computational fluid dynamics, numerous numerical techniques have been proposed to solve the level set equation. [3] first applied the level set method to solve segmentation problem. where curve evolved directly using level set equation. At the same time, the authors of [4] deduce curve evolution equation represented with level set function from a variational approach.

Early geometric deformable models only consider boundary information in curve evolution. The segmentation results based on these models are very sensitive to noise and highly dependent on initialization. Region information is then introduced in variational segmentation frameworks [5]-[8]. The authors of [5] propose an energy functional which is a piecewise constant Mumford-Shah functional. This also equals to represent a region by its mean. A voxel is assigned to a region if its intensity is close to the region’s mean. Without curve length constraint, the method reduces to K means clustering [13]. Another general variational segmentation model is obtained by Maximum a posterior from observed data in [6]-[8]. According to the assumptions that regions are independent, voxels follow identical independent distribution in a region and the prior probability is uniform. It becomes the Maximum Likelihood of observed data. In addition, boundary information is incorporated. The models in [5]-[8] behave very well when the noise level is not too high. However, when image is corrupted with high noise, the results are not satisfactory. In [8], the author suggests a multi-scale preprocessing to filter noise and accelerate the evolution. It is equivalent to use a voxel’s neighborhood weighted mean instead of its own intensity. This method improves the results. However it is ineffective if regions have similar means but different variances. To solve this problem, more information should be considered, such as shape prior. However, shape prior is not guaranteed to be easily available in all cases. Incorporation of variance information is another natural way to deal with this issue [14]-[16]. [14] uses average probability in the neighborhood of a voxel so that second order statistics are considered. [16] proposes a general region-based variational segmentation framework, where any information from region and boundary can be generalized as descriptor. As an example, region descriptor using mean and variance is performed.

In this paper, we propose an energy functional incorporated voxels neighbor information. The variational framework can be deduced from a MDL criterion as in [14] or is a special case of [16]. The major contribution is the energy measured distribution between local and global regions and using information divergence. Thus mean and variance are naturally integrated in the proposed functional.

It is worth noticing the work in [12], this approach is based on region competition and implemented within a level set framework. It deduces that the average probability distribution function (PDF) is equivalent to Kullback-Leibler (KL) divergence between PDF of local and global region. Very recently, [17][18] proposed a variational segmentation framework incorporating mean and variance in the same way as ours. However, for the reason that their application is to extract objects from natural image, the distribution of regions is assumed to be Generalized Laplacian. We make assumption that every region follows a Gaussian distribution.
In addition, the measure in their work is by KL divergence which is directional but by J- divergence in ours, therefore we derive completely different evolution equation from theirs. Last their study is to segment 2D natural images, whereas our work mainly focused on 3D medical image applications.

The reminder of the paper is organized as follows. In section 2 we briefly introduce the variational segmentation model using J- divergence and corresponding evolution equation. Experimental results are demonstrated in section 3. We conclude our work in Section 4.

II. METHODS

In this section, we first introduce the proposed variational functional, as well as its representation with level set. After that the evolution equation coupled with level set is deduced. Last, numerical implementation is briefly presented.

A. Variational segmentation model

At the beginning, region homogeneity is defined. In this study, we confine our approach to intensity image. Vector image can be inferred from the same principle. We adopt the definition in [14], “A region R is considered to be homogeneous if its intensity values are consistent with having been generated by one of family of prespecified probability distributions p(I|θ), where θ are the parameters of the distribution”.

Next, let I : Ω ∈ R^p → R^q be an input image, where Ω is an open and bounded image domain (p = 2 or p = 3) in Euclidean space, q is the dimension of observed data. For example q = 1 is an intensity image, q = 3 is a color image. Image I is composed of nonoverlapping regions {R_i} and their boundaries {Γ_i}. They satisfy R_i ∩ R_j = ∅ when i ≠ j, ∪_{i=1}^{m} R_i = Ω, and Γ_i = ∂R_i.

Based on the aforementioned definition, we make some assumptions: 1. The intensity in a region is homogeneous. This indicates that the intensities in local and global regions follow the same distribution. In this study, for computational efficiency, we assume that every region follows a Gaussian distribution with different parameters. 2. The local region statistics are most similar to the global region to which the voxel belongs. The assumptions are degenerated in the vicinity of boundary. However in our experiments, this inaccuracy of boundary localization is tolerable. Then we can write the energy functional as follows:

\[ E(\Gamma, \theta) = \sum_{i=1}^{n} \left[ \int_{R_i} D(p(N(x) \parallel p(R_i | \theta_i)) dx \right] + \frac{\mu}{\nu} \text{length}(\Gamma) \]  

Where p(N(x)) is PDF of local region around voxel I(x), p(R_i | θ_i) is PDF of i_{th} region with parameters θ_i. D(·) is the dissimilarity measure between two distributions. The energy in Equation (1) measures the dissimilarity between global region and every voxel’s local region in it. Without length of curve constraint, it intuitively can be treated as a weighted volume. When the optimal contour is achieved, the minimum weighted volume is obtained. For the sake of simplicity, we only consider bimodal segmentation.

B. Representation with Level sets

Our curve/surface evolution is implemented using level set equation because of its topology flexibility and numerical stability. As in [2], we defined a Lipschitz continuous function. Commonly we choose the signed distance function and define it as follows:

\[
\begin{align*}
\phi(x) &= -D(x, \Gamma) & \text{if } x \in R_1 \\
\phi(x) &= 0 & \text{if } x \in \Gamma \\
\phi(x) &= D(x, \Gamma) & \text{if } x \in R_2 
\end{align*}
\]

Where \( R_1 \cup R_2 = \Omega \), \( \Gamma \) is zero level set, \( D(x, \Gamma) \) is the Euclidean distance between \( x \) and \( \Gamma \). To represent our energy functional with the level set function and compute the associated Euler Lagrange equation, we also need a regularized version of the Heaviside and Dirac function as defined in [5].

\[ H_\epsilon(z) = \frac{1}{2} \left( 1 + \frac{2}{\pi} \arctan \left( \frac{z}{\epsilon} \right) \right) \]

Now we can rewrite Equation (1) as follows:

\[
E(\phi, \theta_1, \theta_2) = \\
\lambda \cdot \int_{\Omega} D(p(N(x)) \parallel p(R_1 | \theta_1)) \cdot H_\epsilon(-\phi) dx + \\
\lambda \cdot \int_{\Omega} D(p(N(x)) \parallel p(R_2 | \theta_2)) \cdot H_\epsilon(\phi) dx + \\
\nu \cdot \int_{\Omega} | \nabla H_\epsilon(\phi) | dx 
\]

C. Evolutionary equation

To minimize the energy functional, a two-step algorithm is performed. First fix region parameters and compute the Euler-Lagrange equation of the proposed energy functional for level set function. Then using the gradient descent flow, we can get the curve evolution equation:

\[
\frac{\partial \phi}{\partial t} = \delta(\phi) \cdot \left( \nu \cdot \nabla \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right) - \lambda \cdot D(p(N(x)) \parallel p(R_1 | \theta_1)) + \lambda \cdot D(p(N(x)) \parallel p(R_2 | \theta_2)) 
\]

Second, update region parameters with fixed level set function. It is to find the optimal parameters which minimize the energy function with constant \( \phi \). As in [10], we can update these parameters of the corresponding region simply using their sample means and variances:

\[
\mu_i = \frac{\int_{R_i} I(x) dx}{V_i}, \quad \sigma_i^2 = \frac{\int_{R_i} (I(x) - \mu_i)^2 dx}{V_i} 
\]

Where \( V_i = \int dx \) is the volume of i_{th} region. After that we need an appropriate dissimilarity measure between two distributions. The J-divergence is applied here for the reason that it is symmetric and can be analytically expressed for Gaussian distribution. Thus the evolution equation is simplified to a function of sample means and variances. It is
very efficient for computation. The J-divergence is defined as follows:

$$D(p(x) \parallel q(x)) = \frac{1}{2} \int (p(x) \log \frac{p(x)}{q(x)} + q(x) \log \frac{q(x)}{p(x)}) dx$$ (6)

Finally, the curve evolution equation can be rewritten as follows:

$$\frac{\partial \phi}{\partial t} = \frac{1}{\delta} \left( \mu \cdot \nabla \left( \frac{\nabla \phi}{\sqrt{\delta}} \right) + \frac{\sigma_1^2 + (\mu_2 - \mu)^2}{4\delta \sigma_2^2} - \frac{\sigma_2^2 + (\mu_2 - \mu)^2}{4\delta \sigma_1^2} \right)$$ (7)

Equation (7) can reduce to Tony Chan’s [5] equation if $\sigma_1 = \sigma_2$. Replaced with KL divergence, Equation (7) becomes Paragios’s [6]-[8] curve evolution equation.

### D. Numerical implementation

The numerical scheme we used is the semi-implicit finite difference scheme introduced in [5]. When image intensity varied drastically, the velocity discrepancy is large. Thus convergence rate is slow. We use normalized velocity to obtain a smooth and slowly varied velocity as follows:

$$g(z) = \frac{2}{\pi} \left( \arctan(\frac{z}{\delta}) \right)$$

Where $\delta$ is a parameter to control the range of smooth velocity.

![Normalized velocity distribution](image1)

**Fig. 1.** Normalized velocity distribution

### III. EXPERIMENTAL RESULTS

In this section, Segmentation experiments are conducted on both synthetic and real data to illustrate the effectiveness of our algorithm. Comparisons with Paragios’s method [6]-[8] for synthetical data are also presented. Before real data experiments, Skull-stripping had been applied to all real data. All experiments are running on a PC (PentiumIV 2.4GHz, 512M).

#### A. Synthetic Data

To validate the proposed algorithm that is not only effective for Gaussian noise but also for other type noise. We performed our method and Paragios’s for Gaussian and pepper noise degenerated square pattern image. The visual results from Fig. 2 demonstrate the stability and robustness of our algorithms.

![Segmentation results](image2)

**Fig. 2.** Segmentation of square pattern from background. The first row shows results of proposed algorithm, Second row shows results of algorithm in [8], all experiments with the same initialization. (a) and (d), Gaussian noise, background mean/variance = 10/100, square mean/variance = 10225. (b) and (e) Gaussian noise, background mean/variance = 0/0.1, square mean/variance = 0/1 (in (e), the curve will not stop until expand out image domain). (c) and (f), salt and pepper noise, background and square mean $= 1$ noise density $= 0.5$.

#### B. Real Data

In this subsection, we use two real data sets to assess our algorithm. The first data are MRI data obtained on a 1.5T GE Signa Twin Speed scanner with a 3D Spoiled Gradient-Recalled (SPGR). The imaging parameters are as follows: $TR = 11.3 ms, TE = 4.2 ms, and Flip Angle = 15^\circ$. In order to delineate lateral ventricle, a necessary preprocessing that manually crop the third ventricle was performed to satisfied bimodal condition. The local window size is $(3*3*3)$.

![MRI brain data](image3)

**Fig. 3.** Segmentation of ventricle on MRI brain data.

The second data are DT-MRI data acquired on a 1.5 T SIEMENS Sonata with high speed gradients. The diffusion sensitizing gradients are applied along 6 non-collinear directions with b value $= 900 s/mm^2$. Imaging parameters are: $TR = 8000 ms, TE = 106 ms, number of excitations = 3$ and voxel size is $0.9 * 0.9 * 4.0 mm^3$.

Diffusion tensor matrix is calculated according to Stejksal and Tanner equation [19]. Our algorithm is performed on
fractional anisotropy (FA) image computed from diffusion tensor image. In order to extract the corpus callosum, we cropped FA image around the region of interest so as to assure the bimodel condition. The initial surface is a sphere centered in the image with radius equal to 20 voxels. Fig. 4 shows the evolution process and the final results.

![Fig. 4. Segmentation of corpus callosum from 3D DTI data.](image)

From the figures, we see that our method can successfully segment the ventricle and the corpus callosum in 3D real images. It demonstrates the potential of proposed algorithm for medical image segmentation.

IV. CONCLUSIONS AND FUTURE RESEARCH

We proposed a variational segmentation algorithm which measured the dissimilarity between local and global regions. Region boundary was modeled as a regular surface implicitly so that evolution equation represented with level set function was derived. The mean and variance in a neighborhood of every voxel were both considered in the approach. Thus our algorithm is insensitive to noise. Experimental results from synthetic and real data demonstrated this.

Our further work will focus on taking into account the influence of local window size and shape on boundary localization in the approach. Especially when the interested region is tiny in the image, the error will be intolerable compared with the size of the region. Second we will also consider nonparametric PDF replaced the Gaussian distribution assumption in our algorithm. Extension to color images is another interesting work, with which multi-variate distribution has to be dealt.

ACKNOWLEDGEMENT

Authors want to thank NDI Inc. for their support to participate the conference.

REFERENCES


