Shape-Appearance Guided Level-Set Deformable Model for Image Segmentation

Fahmi Khalifa, Ayman El-Baz  
Bioengineering Department, University of Louisville, Louisville, Kentucky, USA  
{fakhal01,aselba01}@louisville.edu

Georgy Gimel’farb*  
Department of Computer Science, University of Auckland, New Zealand  
g.gimelfarb@auckland.ac.nz

Rosemary Ouseph  
Department of Medicine, University of Louisville, Louisville, Kentucky, USA  
r0ouse01@louisville.edu

Mohamed Abu El-Ghar  
Urology and Nephrology Department, University of Mansoura, Egypt  
maboelghar@mans.edu.eg

Abstract

A new speed function to guide evolution of a level-set based active contour is proposed for segmenting an object from its background in a given image. The guidance accounts for a learned spatially variant statistical shape prior, 1st-order visual appearance descriptors of the contour interior and exterior (associated with the object and background, respectively), and a spatially invariant 2nd-order homogeneity descriptor. The shape prior is learned from a subset of co-aligned training images. The visual appearances are described with marginal gray level distributions obtained by separating their mixture over the image. The evolving contour interior is modeled by a 2nd-order translation and rotation invariant Markov-Gibbs random field of object / background labels with analytically estimated potentials. Experiments with kidney CT images confirm robustness and accuracy of the proposed approach.

1. Introduction

Accurate segmentation of anatomical structures in medical images is very important in practice but still remains a challenging problem due to large variations of goal shapes, image noise and inhomogeneities, discontinuous boundaries due to occlusions, similar visual appearance of adjacent parts of different structures, etc. [1]. Geometric level-set based deformable models are popular and powerful tools in segmenting medical images because of flexibility and independence of parameterizations of an evolving contour on the x-y plane [2]. The object-background boundary at each moment $t$ is represented by a zero level $\phi_t(x,y) = 0$ of an implicit level-set function – a distance map $\phi_t(x,y)$ signaling the signed minimum Euclidean distances from every point $(x,y)$ to the boundary (negative for interior and positive for exterior points). The distance map is evolved iteratively as [2]:

$$
\phi_{n+1}(x,y) = \phi_n(x,y) - \tau V_n(x,y) |\nabla \phi_n(x,y)| \quad (1)
$$

where $n$ denotes the time instants $t = n\tau$ with step $\tau > 0$; $V_n(x,y)$ is a speed function guiding the evolution, and $\nabla \phi_n = (\frac{\partial \phi_n}{\partial x}, \frac{\partial \phi_n}{\partial y})$ is the gradient of $\phi_n(x,y)$.

Conventional speed functions $V_n(x,y)$ accounting for image intensities (e.g., [3]), object edges (e.g., [4]), an intensity gradient vector flow (e.g., [5]), etc., are unsuccessful on very noisy images with low object-background intensity gradients. Additional prior shape knowledge results in more effective segmentation: see e.g., [6–8]. Leventon et al. [8] augmented speed functions with terms attracting the level set front to more likely shapes specified by analyzing principal components of a training set of objects. Tsai et al. [6] approximated an object shape with a linear combination of distance maps for a set of mutually aligned images of goal objects. But due to a very large dimensionality
of the distance maps’ space, only a few top-rank principal components can be included in the linear combination to make the model feasible. To overcome over-segmentation, Cheng et al. [7] incorporated a prior shape knowledge into the improved Chan-Vese’s level-set model [9] using a variational approach.

But speed functions based solely on the shape priors lead to large segmentation errors under discontinuous object boundaries and large image noise and inhomogeneities (see e.g., [10]). Also, the approximation of a distance map \( \phi_n(x, y) \) with the closest linear combination of the maps for a set of co-registered training images (see e.g., [6, 7]) has a serious inherent drawback: because the space of distance maps is not closed with respect to linear operations, the zero level of a linear combination of the training maps may produce an invalid or even physically impossible boundary.

To overcome these limitations, this paper introduces a new speed function combining a mean curvature of an evolving contour with a spatially variant joint Markov-Gibbs random field (MGRF) model of an image and its object-background prior independent random field of region labels. As shown in Fig. 1, a joint MGRF model of pairs \((g, m)\) is described with probability distribution \( P(g, m) = P(g|m)P(m) \) combining a simple 2\textsuperscript{nd}-order MGRF \( P(m) \) of region maps and a conditionally independent random field of image intensities \( P(g|m) = \prod_{(x,y)\in\mathbb{R}} p(g_{x,y}|m_{x,y}) \) given the map.

The map model \( P(m) = P_\phi(m)P_\lambda(m) \) has two independent parts: a shape prior being a spatially variant independent random field of region labels \( P_\phi(m) = \prod_{(x,y)\in\mathbb{R}} p_{\phi(x,y)}(m_{x,y}) \), for a set of co-aligned training maps, and a 2\textsuperscript{nd}-order MGRF model \( P_\lambda(m) \) of a spatially homogeneous evolving map \( m \) for the image \( g \) aligned to the training set. The pixel-wise priors \( p_{\phi(x,y)}(1) \) and \( p_{\phi(x,y)}(0) = 1 - p_{\phi(x,y)}(1) \) are the empirical object-background probabilities for the manually segmented training images co-aligned by a rigid 2D transformation. Figure 2 illustrates the steps for building the shape prior: kidneys in the training set are (i) manually segmented (Fig. 2.b) and (ii) co-aligned by a rigid 2D registration maximizing their mutual information [11]; and (iii) the pixel-wise object probabilities \( p_{\phi(x,y)}(1) \) are estimated (Fig. 3) by counting how many times the pixel \((x, y)\) was segmented as kidney.

To build an initial region map \( m \), a given image \( g \) is aligned to one of the training images. The empirical mixture \( P_{\text{mix}}(g) = \prod_{(x,y)\in\mathbb{R}} p_{\text{mix}}(g_{x,y}) \), is separated into the object and background marginal components, \( (p|q_\lambda : q \in \mathbb{Q}; \lambda \in \mathbb{L}) \), just as in [12], by close approximation with a linear combination of discrete Gaussians (LCDG). The LCDG restores transitions between dominant distribution modes in a mixture more accurately than conventional mixtures of only positive Gaussians [3], thus yielding a better initial map, \( m \).

By symmetry considerations and for simplicity, the homogeneity of the initial map is modeled with a popular Potts model with the nearest pixel 8-neighborhood and a bi-valued Gibbs potential depending only on whether the nearest pairs of labels are equal or not. An approximate analytical potential estimate in line with [13]: \( V_{\text{eq}} \equiv -V_{\text{ise}} \equiv 4f_{\text{eq}}(m) - 2 \) allows for re-computing pixel-wise conditional probabilities \( p_{\phi(x,y)}(m_{x,y}) \) of labels, \( m_{x,y} \equiv \lambda \); \( \lambda \in \mathbb{L} \), given their 8-neighbors, at each step of the contour evolution. Here, \( f_{\text{eq}}(m) \) is the relative frequency of equal labels in all the neighboring pixel pairs \((x, y), (x + \xi, y + \eta)\) \( \in \mathbb{R}^2 \);
\((\xi, \eta) \in \{(1,0), (1,1), (0,1), (-1,1)\}\).

The speed function in Eq. (1) is \(V(x, y) = \partial(x, y) \kappa\) where \(\kappa\) is the mean contour curvature and \(\partial(x, y)\) specifies the evolution magnitude and direction:

\[
\partial(x, y) = \begin{cases} 
-P_{1:x,y} & \text{if } P_{1:x,y} > P_{0:x,y} \\
-P_{1:x,y} & \text{otherwise}
\end{cases}
\]

where

\[
P_{1:x,y} = \frac{W_{1:x,y}}{W_{1:x,y} + W_{0:x,y}}, P_{0:x,y} = 1 - P_{1:x,y}
\]

\[
W_{1:x,y} = p(g(x,y) | \Phi_{h:x,y}(1) \Phi_{s:x,y}(1))
\]

\[
W_{0:x,y} = p(g(x,y) | 1 - \Phi_{h:x,y}(1)) (1 - \Phi_{s:x,y}(1))
\]

and \(p_{h:s}(x,y)(1)\) is the probability of transition to the object label in the Potts model \(P(m)\) at \((x, y)\) at the current evolution step of the region map \(m\). The initial distance map \(\phi_0(x, y)\) is computed with respect to a circular contour initialized inside the object on the initial region map and then incremented or decremented at every pixel in line with Eq. (2). If a pixel \((x, y)\) is classified as the object, the level set function is incremented and the contour expands; otherwise it is incremented and the boundary contracts.

For a more smooth evolution and thus more accurate segmentation results, the level set function \(\phi(x, y)\) has to be reinitialized to a signed distance function (i.e., \(|\nabla \phi| \approx 1\)). All the experimental results presented below were obtained with a level set function having been reinitialized every 10 iterations in line with [14].

3. Experimental Results and Conclusions

The proposed approach has been tested on 2D CT kidney images whose marginal intensity distributions are mixtures with the three dominant modes: two modes for darker image parts and mid-gray abdominal tissues (relating both to the background) and one mode for a brighter kidney object. The LCDG to separate the modes including numbers of its positive and negative components is learned with the Expectation-Maximization-based algorithm introduced in [12]. Basic density estimation steps using the LCDG model are illustrated in Fig. 4. One of the images and its step-wise segmentation results with the final error of 0.62% w.r.t. the ground truth are shown in Fig. 5. Additional results are shown in Fig. 6.

For comparison, the same images have been segmented using two other speed functions that rely on either the image intensities only [12] or both the intensities and spatial information [15]. The results (see Fig. 7) are too inaccurate – the errors are about 31%. Table 1 compares results for the 21 test CT kidney images with the known ground truth (manual segmentation by an expert) that were not used for training. Differences between the mean errors for the proposed approach, the shape based level-set approach of Tsai et al. [6], intensity-based segmentation [12], and stochastic segmentation [15] are statistically significant by the unpaired \(t\)-test (the two-tailed \(P\)-values are less than 0.0001). To evaluate the sensitivity of our approach to selection of manually segmented training images, we divide the data into three groups and performed the 3-fold cross validation. Table 2 shows...
that such a sensitivity is statistically insignificant.

In total, embedding the probabilistic shape and appearance models into the speed function notably improves the level set evolution and results in precise segmentation of complex images like the kidney CT images. The initial rigid alignment of an object to be segmented to the probabilistic shape prior is a basic step making the proposed approach suitable for noisy and inhomogeneous goal objects such as anatomical structures. The proposed segmentation is robust and accurate, and due to the simple models involved it is invariant to object translations, rotations, and scaling. Therefore, it might be useful not only for medical imaging but also for other computer vision applications.

Table 1. Comparative segmentation accuracy.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Our</th>
<th>[6]</th>
<th>[12]</th>
<th>[15]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min error, %</td>
<td>0.6</td>
<td>4.5</td>
<td>16.7</td>
<td>9.1</td>
</tr>
<tr>
<td>Max error, %</td>
<td>5.6</td>
<td>11.8</td>
<td>41.6</td>
<td>48.1</td>
</tr>
<tr>
<td>Mean error, %</td>
<td>2.3</td>
<td>4.9</td>
<td>31.4</td>
<td>31.6</td>
</tr>
<tr>
<td>St.deviation, %</td>
<td>1.4</td>
<td>2.2</td>
<td>6.9</td>
<td>9.3</td>
</tr>
</tbody>
</table>

References


