MULTIPLE-ATLAS-BASED AUTOMATIC SEMENTATION OF HIPPOCAMPUS FOR LATERALIZATION IN TEMPORAL LOBE EPILEPSY

Alireza Akhondi-Asl,1,2,3, Kourosh Jafari-Khouzani3, Hamid Soltanian-Zadeh,1,2,3

1Control and Intelligent Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
2School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics (IPM), Tehran, Iran
3Image Analysis Laboratory, Department of Diagnostic Radiology, Henry Ford Health System, Detroit, Michigan, USA

ABSTRACT
We introduce a 3D segmentation framework which uses principal shapes. The probabilistic energy function of the method is defined based on intensity, tissue type, and location information of the structures using a multiple atlas method. For intensity information, nonparametric probability density function is used which considers intensity relation of different structures. To find a local minimum of the energy function, a two-step optimization strategy is used. In the first step, shape parameters are optimized based on the analytic first and second derivatives of the energy function. In the second step, shapes of the structures are fine-tuned using a level set method. The proposed method is shown to be superior to some popular methods in the literature using a dataset of 64 patients with mesial temporal lobe epilepsy. In addition, the method can be used for lateralization with accuracy close to that of manual segmentation.

Index Terms—Image Segmentation, Atlas-Based, Hippocampus, Epilepsy

1. INTRODUCTION
Medical image segmentation is one of the most important issues in computer aided diagnosis and also one of the most challenging topics in image processing [1]. Many segmentation methods have an energy function, a shape model, and an optimization algorithm. Each of these parts plays an important role in the design of an accurate segmentation algorithm. Prior knowledge is incorporated in a variety of fashions in order to design more accurate segmentation algorithms [2-5]. Prior knowledge can be used in the energy function, the shape models, or the optimization algorithm. For example, we have used prior knowledge to extract constraints about the energy function parameters in [6]. Another category of segmentation methods use relation between different structures [7-9].

In all of these three methods, the authors used principal component analysis (PCA) to extract shape relation between different structures in the training datasets. In this framework, each structure can be defined based on the weights of the principal shapes. In this paper, we introduce a 3D segmentation framework based on the PCA which considers shape relation between different structures. It uses principal shapes and their weights to define shapes of different structures. The probabilistic energy function of the method is defined based on intensity, tissue type, and location information of the structures. To extract shape models, tissue type, and location information, a multiple atlas method is employed. For intensity information, the nonparametric probability density function (pdf) is used which considers intensity relation of different structures. To find the local minimum of the energy function, a two-step optimization strategy is used. In the first step, shape parameters are optimized based on the analytic first and second derivatives of the energy function with respect to the parameters. In the second step, shapes of the structures are fine-tuned using a level set method.

2. ENERGY FUNCTION
Suppose that we want to simultaneously segment m different structures in a skull stripped image I(x) and we have n skull stripped training intensity datasets \(I_{m}^{i}(x)\), \(i \in \{1,...,n\}\) and their corresponding training label datasets \(L_{m}^{i}(x)\), \(i \in \{1,...,n\}\) in which all of the m structures are segmented and labeled with labels \(l \in \{0,...,m\}\). In each label image, label 0 is assigned to region outside all of these m structures (outside region).

For each voxel in the image we define the probability of a pixel belonging to the region \(\Omega_{k}\) with
\[
p(x|x \in \Omega_{k}) = p_{T}\end{ab} \times p_{\alpha}(x)\times p_{\alpha}(x) \quad \text{where} \quad k \in \{1,...,m+1\}
\]
with \(m+1\) indicating the outside region. In this definition, \(p_{T}\end{ab} (x)\) explains the information about the tissue type (the intensity ranges), \(p_{\alpha}(x)\) considers the location information, and \(p_{\alpha}(x)\) demonstrates the intensity uniformity of the kth structure. To compute \(p_{\alpha}(x)\) and \(p_{\alpha}(x)\) in the first step, we use multiple atlas strategy to register \(I_{m}^{i}(x)\) \(i \in \{1,...,n\}\) to \(L(x)\) with a non-rigid registration method. The corresponding transformation is named \(T\). There are many methods in the literature that use multiple atlases for segmentation [10-12]. We employ a different method in which local information is used for atlas construction. In the previous methods in the literature, local information is applied for each voxel independently, but here we use a local weight for each structure which is not global in the form of [11] and not local.
in the form of \([10, 12]\). We apply \(T^i\) to the label volumes \(L^i_n(x)\) \(i \in \{1, \ldots, n\}\) to get \(T^i L^i_n(x)\). In the next step, for each structure we find a region of interest (ROI) based on the registered datasets and name it \(R_k\), which shows the ROI of the \(k\)th structure. In addition, we define \((I(x))_{k_{iS}}\) to show the part of the image \(I(x)\) selected by \(R_k\). Then, we use the following equation to find the similarity between structures of the training datasets and the test image:

\[
M^i_k = \text{MI}\left(\left(T^i L^i_n(x)\right)_{k_{iS}}, (I(x))_{k_{iS}}\right)
\]

(1)

where \(\text{MI}(A,B)\) show the mutual information between the images \(A\) and \(B\).

If \(M^i_k > M^i_{k'}\) for \(i = 1 \ldots m, i \neq i'\) based on \(R_k\) for structure \(k\), it means that the \(k\)th structure of the dataset \(i\) has highest similarity with the \(k\)th structure of the test image \(I(x)\) based on the mutual information metric. Next, for each label we construct the following image:

\[
\bar{\phi}^i_k(x) = \sum_k M^i_k \times S\left(\delta_i \left(T^i L^i_n(x)\right)\right)
\]

(2)

where \(\delta_i(u(x)) = \begin{cases} 1 & \text{if } u(x) = k \\ 0 & \text{if } u(x) \neq k \end{cases}\) and \(S\) is the sign distance function (SDF) of the binary. This image has the property that a point with a lower value has a higher probability of falling inside the structure of interest. Thus, using \(\bar{\phi}^i_k(x)\) we can define \(p_s(x)\) for the \(k\)th structure as:

\[
p_s(x) = \begin{cases} \left(1 - e^{-\phi^i_k(x)}\right) \left(\text{sign}(\phi^i_k(x)) + e^{\phi^i_k(x)}\right) & \phi^i_k(x) > -\frac{1}{\varepsilon} \\ \varepsilon & \text{otherwise} \end{cases}
\]

(3)

This function is 1 when \(\bar{\phi}^i_k(x)\) takes its minimum and 0.5 when \(\bar{\phi}^i_k(x) = 0\). The parameter \(\varepsilon\) is used to resolve the numerical problems associated with using \(\text{MI}(p_s(x))\) in the next steps. To calculate \(p_s(x)\), we use a probability mass function (pmf) with a new strategy explained next. We apply clustering twice with 3 and 10 clusters. The 3-class segmentation results capture the global intensity information of the tissues in the brain while the 10-class segmentation captures their local or fine intensity information. To use the results, we model the pmf of the tissue type of each structure, i.e., the probabilities of each of the 10-class clusters being present in the structure, by the model shown in Figure 1 whose parameters are estimated using:

\[
L_1 = \arg \min_{\sum_{\leq 0}} \left| C_{i}(2) - C_{i}(0) \right| \quad L_2 = \arg \min_{\sum_{\leq 0}} \left| C_{i}(3) - C_{i}(0) \right| - 2
\]

(4)

In this equation, \(C_{i}(k)\) shows the intensity of the center of the \(k\)-th class (cluster) in the \(n\)-class clustering. For the region outside all of the \(m\) structures (region \(n+1\)), if all of the other structures have the same pmf, we use \(1-p\) but if there are different pmf's, we use \(p = 0.5\), as described in:

\[
\begin{align*}
p_m(x) &= \frac{1}{2} \left[ 1 - p_{m+1}(x) \right] \\
p_m(x) &= \frac{1}{2} \left[ 1 - p_{m+1}(x) \right] + \frac{1}{2} \left[ 1 - p_{m+1}(x) \right]
\end{align*}
\]

(5)

To calculate \(p_m(x)\), we use the Parzen window estimator that is defined in the form of \(p_m(x) = 1 \frac{1}{|\Omega|} \int_{\Omega} K(I(x) - I(x)) dx\) [13]. In this equation, \(K(t)\) is a Gaussian kernel with variance \(\sigma\) which is estimated by the method of [14]. In addition, \(|\Omega|\) is used for the cardinality that is the number of members in the set or the region \(\Omega\). Thus, we can estimate pmfs for each \(m+1\) using the following equations:

\[
\begin{align*}
p_k(x) &= \frac{1}{2} \sum_{i=1}^{m+1} \int_{\Omega} K(I(x) - I(x)) dx \\
p_m(x) &= \frac{1}{2} \sum_{i=1}^{m+1} \int_{\Omega} K(I(x) - I(x)) dx
\end{align*}
\]

(6)

Finally, we write the energy function as:

\[
\begin{align*}
\mathcal{J}(\Omega_1, \ldots, \Omega_m) &= \sum_{i=1}^{m} \int_{\Omega_i} F(p(x) | x \in \Omega_i) dx \\
&= \sum_{i=1}^{m} \int_{\Omega_i} F(p(x) | x \in \Omega_i) dx \times p_s(x) dx
\end{align*}
\]

(7)

This function depends on the \(m\) regions and the probability of a voxel being within each region. To find the best segmentation, we should find the regions \((\Omega_1, \ldots, \Omega_m)\). It should be noticed \(\Omega_{m+1}\) is a function of the other regions and is not an independent region. When all of the regions are as uniform as possible, same type regions have more similar intensity distributions, all of the regions are in the correct tissue as much as possible and all of them are in the proper locations, the energy function takes its optimum. If decreasing \(F(p)\) such as \(-\ln(p)\) is used, the optimization will be a minimization problem. To optimize this function, first a shape (region) representation method is needed. In the next section, we use training datasets to find a shape model which considers shape relation of different structures.

3. SHAPE REPRESENTATION

Sign distance function is one of the most effective shape representation methods in the literature [15] which is used in our framework. There are essential information about the shapes and relation of the structures that can be used in segmentation. In [9], we used these relations and showed their effectiveness. Here, we use the same idea. Suppose that we define a sign distance function (SDF) \(\bar{\phi}(x)\) which is negative inside region \(k \in \{1, 2, \ldots, m\}\). We need to find this SDF's for a correct segmentation. We know that there are relations between shapes of different structures and want to use these relationships for segmentation. We have \(n\) training label
maps \( L_r^+ (x) \) with \( m \) non-zero labels \( k \in \{1, \ldots, m\} \) representing \( m \) structures. To extract shape relations between different structures, for each label we register binary images of all training images to the one with highest similarity to \( I(x) \) to find the \( A^-_k \)'s. Each \( A^-_k \) is an affine transform with 12 parameters. It should be noticed that for each structure \( k \) we have 12-\( n(l) \) parameters that should be optimized because for the fixed one (for the structure \( k \) it is \( p_k \)), \( A^-_k = I \). We use the location differences of the center of masses of the fix and the remaining images for initialization.

In the next step, we compute the SDF every \( \partial_i^q(L_r^+ (x)) \) and construct \( \varphi = [\varphi_1^1, \ldots, \varphi_m^k]^T \) where \( \varphi = \sum \varphi_i \) and \( X^T \) is the transpose of \( X \). The function \( \varphi(f) \) changes the 3D image \( f \) with the size of \( d_1 \times d_2 \times d_3 \) to a vector of size \( d_1d_2d_3 \). In our approach, the shape priors are modeled using PCA as some basis shapes \( \{\varphi_i\} \), where deformation model of multiple structures is considered to be the combination of these basis priors \( \varphi = \varphi_0 + \sum w_i \varphi_i \) where \( \varphi = \frac{1}{n} \sum \varphi_i \). In this model, we are seeking proper values of \( w_i \) that minimize our energy function which depend on functions \( \varphi_i \) which are part of \( \varphi = [\varphi_1, \ldots, \varphi_m]^T \). Thus, for each set of \( w_i \)'s we can find \( \{\varphi_1, \ldots, \varphi_m\} \) to compute the energy function. In the next section, we summarize the energy function and design the optimization strategy.

4. OPTIMIZATION

To consider the pose variances in the segmentation process, for each structure an individual affine transform is used. These transforms give flexibility to the regions and can be used for the local alignment. In this case, for each region, we add 12 parameters to the problem. Thus, we have \( l + 12 \times m \) parameters which we put in the vector \( P \).

\[
\min_j \int P(x) - \sum_{j=1}^{m} \int_{\varphi_j (x)}^{\varphi_j (x)} F(\varphi_j (x)) \sum_{\varphi_j (x) \times \varphi_j (x)} F(\varphi_j (x)) \mathrm{d}x
\]

In this equation, \( \varphi \) is Heaviside function. Any optimization problem needs an initialization. For principal shapes, we use \( \varphi = \varphi_0 \) and for the Affine transforms, we set the translation part of the matrices such that the center of mass of \( \varphi_i (x) < 0 \) (with \( \varphi = \varphi_0 \)) be the same as the center of mass of \( \varphi_i (x) < 0 \). Other parts of the affine matrices are same as the Identity matrix.

To optimize the energy function, many methods use first order derivatives and Quasi-Newton or steepest descent optimization strategies [7, 9]. To use this algorithm, we need to compute first order derivatives of the energy function for the parameters in \( P \). For the energy function, it can be shown that the derivative of the function \( J \) with respect to a parameter \( \varphi \) is:

\[
\frac{\partial}{\partial \varphi} J = \frac{\partial}{\partial \varphi} \int \varphi(x) \int F(\varphi_j (x)) \sum_{\varphi_j (x) \times \varphi_j (x)} F(\varphi_j (x)) \mathrm{d}x
\]

In the next step and after optimization of the parameters of the energy function using constraint optimization algorithm, to capture details of the structures that cannot be extracted from the principal shapes, we remove shape dependency due to the principal shapes and define the following function:

\[
\min_j \int \varphi(x) \int F(\varphi_j (x)) \sum_{\varphi_j (x) \times \varphi_j (x)} F(\varphi_j (x)) \mathrm{d}x
\]

In this function, there are \( m \) level set functions that should be optimized in the same constraint as before. Thus, we can use the following energy function where the second term is for the smoothness of the shapes.

\[
\mathcal{F} = \sum_{j=1}^{m} \int \varphi(x) \int F(\varphi_j (x)) \sum_{\varphi_j (x) \times \varphi_j (x)} F(\varphi_j (x)) \mathrm{d}x
\]

Using the following equation, we can update each shape iteratively where \( t \) shows time.

\[
\frac{\partial}{\partial t} \varphi_j (x) = \frac{\varphi_j (x) - \varphi_j (x)}{\varphi_j (x) \times \varphi_j (x)} F(\varphi_j (x)) \sum_{\varphi_j (x) \times \varphi_j (x)} F(\varphi_j (x)) \mathrm{d}x
\]

5. RESULT

To test and evaluate the proposed method, we have applied it to the MRI of 64 patients with mesial temporal lobe epilepsy (mTLE) acquired on a General Electric 1.5T Sigma System (GE Medical Systems, Milwaukee, WI) [16]. The hippocampi were manually segmented by an expert. Comparison of the hippocampus segmentation results generated using our method and the expert segmentation is shown in Figure 2. It is known that the mTLE may cause changes in the size and shape of the hippocampus. Thus, correct segmentation of hippocampus in MRI of mTLE patients is a complicated task. The method is also compared with other previous methods (Table 1) which shows the superiority of our method. In addition, we compared lateralization output using right/left volume ratio of automatic methods and the manual segmentation in Table 2. It can be seen that except HAMMER other methods have almost same standard lateralization percentage.
### Table 2. Comparison of lateralization accuracy of mTLE patients extracted using 4 segmentation methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Dice</th>
<th>F95</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Method</td>
<td>Mean±Std</td>
<td>0.71±0.08</td>
<td>0.06±0.35</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>0.52</td>
<td>-0.46</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.81</td>
<td>1.39</td>
</tr>
<tr>
<td>HAMMER [17]</td>
<td>Mean±Std</td>
<td>0.65±0.08</td>
<td>-0.13±0.19</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>0.39</td>
<td>-0.45</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.76</td>
<td>0.74</td>
</tr>
<tr>
<td>Free Surfer  [18]</td>
<td>Mean±Std</td>
<td>0.63±0.13</td>
<td>-0.28±0.30</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>0.00</td>
<td>-0.62</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.79</td>
<td>1.41</td>
</tr>
</tbody>
</table>

### 6. CONCLUSION

We have presented a new method for segmentation of the brain subcortical structures using their shape relationships and tissue type and location information using multiple atlas strategy. The energy function used for segmentation is based on the entropy of different structures. Tissue type information is used to improve robustness and accuracy of the segmentation process. In addition, location information is used to further improve the segmentation process. With a powerful automatic initialization of the structures which considers tissue type and location information and the use of the quasi-Newton algorithm, a local minimum of the energy function is found. In the final step, a level-set based optimization is applied. To achieve accurate results, the intensity pdf's are calculated in each iteration of the algorithm and gradients are computed analytically. Experimental results have illustrated superiority of the proposed framework to the other methods in the literature for the segmentation of hippocampus in MRI of mTLE patients. In addition, our proposed method can be used for lateralization with accuracy close to that of the manual segmentation.

### 7. REFERENCES