A NOVEL, FAST, AND COMPLETE 3D SEGMENTATION OF VERTEBRAL BONES

Melih S. Aslan, Asem Ali, Ham Rara, Ben Arnold∗, Rachid Fahmi, Aly A. Farag, and Ping Xiang∗

University of Louisville
Computer Vision and Image Processing Laboratory
Louisville, KY, 40299 USA

Image Analysis, Inc.
1380 Burkesville St.
Columbia, KY, 42728, USA

ABSTRACT

Bone mineral density (BMD) measurements and fracture analysis of the spine bones are restricted to the vertebral bodies (VBs), especially the trabecular bones (TBs). In this paper, we propose a novel, fast, and robust 3D framework to segment VBs and trabecular bones in clinical computed tomography (CT) images without any user intervention. The Matched filter is employed to detect the VB region automatically. To segment the whole VB, the graph cuts method which integrates a linear combination of Gaussians (LCG) and Markov Gibbs Random Field (MGRF) is used. Then, the cortical and trabecular bones are segmented using local volume growing methods. Validity was analyzed using ground truths of data sets (expert segmentation) and the European Spine Phantom (ESP) as a known reference. Experiments on the data sets show that the proposed segmentation approach is more accurate than other known alternatives.

Index Terms—Spine Bone, Vertebral Body (VB), trabecular bone, graph cuts segmentation.

1. INTRODUCTION

The spine bone consists of the VB and spinal processes. In this paper, we are primarily interested in volumetric computed tomography (CT) images of the vertebral bone of spine column with a particular focus on the lumbar spine. The primary goal of the proposed work is in the field of spine densitometry where bone mineral density (BMD) measurements are restricted to the vertebral bodies, especially trabecular bones (see Fig. 1 for regions of spine bone).

Various approaches have been introduced to tackle the segmentation of skeletal structures in general and of vertebral bodies in particular for the anatomical definition of a VB. For instance, Kang et al. [1] proposed a 3D segmentation method for skeletal structures from CT data. Their method is a multi-step method that starts with a three dimensional region growing step using local adaptive thresholds followed by a closing of boundary discontinuities and then an anatomically-oriented boundary adjustment. Applications of this method to various anatomical bony structures are presented and the segmentation accuracy was determined using the European Spine Phantom (ESP) [2]. Later, Mastmeyer et al. [3] presented a hierarchical segmentation approach for the lumbar spine in order to measure bone mineral density. This approach starts with separating the vertebrae from each other. Then, a two step segmentation using a deformable mesh followed by adaptive volume growing operations are employed in the segmentation. The authors conducted a performance analysis using two phantoms: a digital phantom based on an expert manual segmentation and the ESP. They also reported that their algorithm can be used to analyze three vertebrae in less than 10min. This timing is far from the real time required for clinical applications but it is a huge improvement compared to the timing of 1 – 2h reported in [5]. Recently, in the context of evaluating the Ankylosing Spondylitis, Tan et al. [6, 7] presented a technique to segment whole vertebrae with their syndesmophytes using a 3D multi-scale cascade of successive level sets. The seed placement was done manually and results were validated using synthetic and real data. Other techniques have been developed to segment skeletal structures and can be found for instance in [8, 9] and the references therein.

The VB consists of trabecular and cortical bones. The main objective of our algorithm is to segment the VB, and then the trabecular bone. In this paper, we propose a novel automatic VB segmentation approach that uses subsequently: i) the Matched filter which is used in automatic determination of the VB region, ii) the LCG method to approximate the gray...
level distribution of the VB (object) and surrounding organs (background), and iii) the graph cuts to obtain the optimal segmentation. First, we use the Matched filter to determine the VB region in CT slice. In this method, no user interaction is needed. Also, this method helps the LCG method to initialize the gray level distributions more accurately. After the LCG method initializes the labels, graph cuts segmentation method is employed in the segmentation. Because the VB and surrounding organs have very close gray level information and there are no strong edges in some CT images, we depend on both the volume gray level information and spatial relationships of voxels in order to overcome any region inhomogeneity existing in CT images as shown in the Fig. 2. In this study, the interpolation and level set methods using various post-processing steps are tested and compared with the proposed algorithm. After we segment the VB, cortex and trabecular bones are extracted from each other using the local adaptive region growing algorithm.

Section 2 discusses the background of Matched filter, graph cuts method, and local adaptive region growing methods. Section 3 describes the alternative methods, explain the experiments, and compare the results.

2. PROPOSED FRAMEWORK

2.1. Matched Filter

In the first step, the Matched filter [10] is employed to detect the VB automatically. This procedure eliminates the user interaction and improves the segmentation accuracy. Let \( f(x, y) \) and \( g(x, y) \) be the reference and test images, respectively. To compare the two images for various possible shifts \( \tau_x \) and \( \tau_y \), one can compute the cross-correlation \( c(\tau_x, \tau_y) \) as

\[
c(\tau_x, \tau_y) = \int \int g(x, y) f(x - \tau_x, y - \tau_y) dx dy.
\]

(1)

where the limits of integration are dependent on \( g(x, y) \). Equation 1 can also be written as

\[
c(\tau_x, \tau_y) = FT^{-1}(G(f_x, f_y)F^*(f_x, f_y))
\]

(2)

\[
eq \int \int G(f_x, f_y)F^*(f_x, f_y) e^{j2\pi(f_x\tau_x + f_y\tau_y)} df_x df_y.
\]

where \( G(f_x, f_y) \) and \( F(f_x, f_y) \) are the 2-D FTs of \( g(x, y) \) and \( f(x, y) \), respectively with \( f_x \) and \( f_y \) denoting the spatial frequencies. The test image \( g(x, y) \) is filtered by \( H(f_x, f_y) = F^*(f_x, f_y) \) to produce the output \( c(\tau_x, \tau_y) \). Hence, \( H(f_x, f_y) \) is the correlation filter which is complex conjugate of the 2-D FT of the reference image \( f(x, y) \).

2.2. Graph Cuts Segmentation Framework

In the graph cuts method, a VB (object) and surrounding organs (background) are represented using a gray level distribution models which are approximated by a linear combination of Gaussians (LCG) to better specify region borders between two classes (object and background). Initial segmentation based on the LCG models is then iteratively refined by using MGRF with analytically estimated potentials. In this step, the graph cuts is used as a global optimization algorithm to find the segmented data that minimize a certain energy function, which integrates the LCG model and the MGRF model.

To segment a VB, the volume is initially labeled based on its gray level probabilistic model. Then we create a weighted undirected graph with vertices corresponding to the set of volume voxels \( \mathcal{P} \), and a set of edges connecting these vertices. Each edge is assigned a nonnegative weight. The graph also contains two special terminal vertices \( s \) (source) “VB”, and \( t \) (sink) “background”. Consider a neighborhood system in \( \mathcal{P} \), which is represented by a set \( \mathcal{N} \) of all unordered pairs \( \{p, q\} \) of neighboring voxels in \( \mathcal{P} \). Let \( \mathcal{L} \) the set of labels \( \{0^\prime, 1^\prime\} \), correspond to VB and background regions respectively. Labeling is a mapping from \( \mathcal{P} \) to \( \mathcal{L} \), and we denote the set of labeling by \( f = \{f_1, \ldots, f_p, \ldots, f_{|\mathcal{P}|}\} \). In other words, the label \( f_p \), which is assigned to the voxel \( p \in \mathcal{P} \), segments it into VB or background region. Now our goal is to find the optimal segmentation, best labeling \( f \), by minimizing the following energy function:

\[
E(f) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{\{p, q\} \in \mathcal{N}} V(f_p, f_q),
\]

(3)

where \( D_p(f_p) \), measures how much assigning a label \( f_p \) to voxel \( p \) disagrees with the voxel intensity, \( I_p \). \( D_p(f_p) = -\ln P(I_p|f_p) \) is formulated to represent the regional properties of segments. The second term is the pairwise interaction model which represents the penalty for the discontinuity between voxels \( p \) and \( q \).

To initially label the VB volume and to compute the data penalty term \( D_p(f_p) \), we use the modified EM [11] to approximate the gray level marginal density of each class \( f_p \), VB and background region, using a LCG with \( C^+_{f_p} \) positive and \( C^-_{f_p} \) negative components as follows:

\[
P(I_p|f_p) = \sum_{r=1}^{C^+_{f_p}} w^+_{f_p,r} \varphi(I_p|\theta^+_{f_p,r}) - \sum_{l=1}^{C^-_{f_p}} w^-_{f_p,l} \varphi(I_p|\theta^-_{f_p,l}),
\]

(4)

where \( \varphi(.,|\theta) \) is a Gaussian density with parameter \( \theta = (\mu, \sigma^2) \) with mean \( \mu \) and variance \( \sigma^2 \). \( w^+_{f_p,r} \) means the \( r^{th} \) positive component frequency.
positive weight in class $f_p$ and $w_{f_p,l}^-$ means the $l^{th}$ negative weight in class $f_p$. These weights have a restriction $\sum_{r=1}^{C_f} w_{f_p,r}^+ - \sum_{l=1}^{C_f} w_{f_p,l}^- = 1$.

The simplest model of spatial interaction is the Markov Gibbs random field (MGRF) with the nearest 6-neighborhood. Therefore, for this specific model the Gibbs potential, $\gamma$, can be obtained analytically using our maximum likelihood estimator (MLE) for a generic MGRF in [12, 13]. So, the resulting approximate MLE of $\gamma$ is:

$$\gamma^* = \left( K - \frac{K^2}{K - 1} f_{\text{neq}}(f) \right) .$$

where $K = 2$ is the number of classes in the volume and $f_{\text{neq}}(f)$ denotes the relative frequency of the not equal labels in the voxel pairs. To segment a VB volume, we use a 3D graph where each vertex in this graph represents a voxel in the VB volume. Then we define the weight of each edge as shown in table 2.2. After that, we get the optimal segmentation surface between the VB and its background by finding the minimum cost cut on this graph. The minimum cost cut is computed exactly in polynomial time for two terminal graph cuts with positive edges weights via s/t Min-Cut/Max-Flow algorithm [14].

$$\begin{align*}
\{p, q\} & : \gamma^* \quad f_p \neq f_q \\
\{s, p\} & : -\ln[P(f_p \mid "1")], \quad p \in \mathcal{P} \\
\{p, t\} & : -\ln[P(f_p \mid "0")], \quad p \in \mathcal{P}
\end{align*}$$

### 2.3. Local Adaptive Volume Growing Method

Starting from the segmented VB, check every voxel on its outer surface. If the intensity value or Hounsfield units (HU) of this voxel is greater than a local threshold then it will be used to initiate a local volume growing. This volume growing classification is based on the mean intensity value, $\mu$, and its standard deviation, $\sigma$, in the 26-neighborhood of the considered voxel, $v$, as follows:

$$\begin{align*}
\text{if } I(v) \geq \mu - \alpha \sigma, \text{ label } v \text{ as cortical}, \\
\text{if } I(v) < \mu - \alpha \sigma, \text{ label } v \text{ as trabecular},
\end{align*}$$

with $\alpha$ being a small positive real number. In our experiment, we accept that $\alpha = 1$.

### 3. EXPERIMENTS AND DISCUSSION

To assess the accuracy and robustness of our proposed framework, we tested it using clinical data sets, as well as, the phantom (ESP), which is an accepted standard for quality control [2] in bone densitometry. The real data sets were scanned at 120kV and 2.5mm slice thickness. The ESP was scanned at 120kV and 0.75mm slice thickness. All algorithms are run on a PC 3Ghz AMD Athlon 64 X2 Dual, and 3GB RAM. All implementations are in C++.

To compare the proposed method with other alternatives, VBs are subsequently segmented using the spline interpolation and level sets method including some post-processing steps. Finally, segmentation accuracy is measured for each method using the ground truths (expert segmentation). M1 represents the proposed algorithm. The alternative methods used in the experiments are represented as M2 (for spline-based interpolation), M3 (for level sets with morphological closing post-process), M4 (for level sets without any post-process), and M5 (for level sets with interpolation post-process).

To evaluate the results we calculate the percentage segmentation error as follows:

$$\text{error}\% = \frac{100 \ast \text{Number of misclassified voxels}}{\text{Total number of VB voxels}} .$$

Preliminary results are very encouraging and the test results was achieved for 10 data sets and the ESP. The statistical analysis of our method is shown in the Table 1. In this table the results of the proposed segmentation method and other four alternatives are shown. The average error of the VB segmentation on 10 clinical 3D image sets is 5.6% for the proposed method. The average error of in the trabecular bone segmentation is 2.14%. It is worth mentioning that the segmentation step is extremely fast thanks to automatically
detection of the VOI step using the Matched Filter. The segmentation time is much faster than that reported in [3, 5] and other alternatives tested in our experiment. The spline based interpolation method, represented as M2, has the closest segmentation accuracy for the clinical data set as shown in the Table 1. An example that shows 3D segmentation results of all tested methods for a clinical data set is shown in Fig 5. In this figure, the red color represents the misclassified voxels. The result of M1 has less misclassified voxels than other methods. Some 3D results of the proposed method are shown in the Fig. 6.

The Figure 7 shows some CT images of the ESP used in our experiment. Because clinical CT images have gray level inhomogeneity, noise, and weak edges in some slices, the ESP was scanned with the same problems to validate the robustness of the method. The VB segmentation error on the ESP is 3.0% for the proposed method. The level set method without any post-processing has the closest (but not less) segmentation error which is 9.9%. The Fig. 8 shows 3D segmentation results for the ESP using M1 (proposed method) and M4. Because the proposed algorithm uses both gray level information and spatial interaction between the voxels, it is superior than other alternatives.

4. CONCLUSION

In this paper, we have presented a novel, fast, and robust 3D segmentation framework for VBs and TBs in clinical CT images. User interaction is completely eliminated using the Matched filter which detects the VB region automatically. This step also improves the segmentation accuracy of the graph Cuts method. Validity was analyzed using ground truths of data sets and the European Spine Phantom (ESP) as a known reference. Experiments on the data sets show that the proposed segmentation approach is more accurate and robust than other known alternatives.

5. REFERENCES


Table 1. Accuracy and time performance of our VB segmentation on 10 data sets. Average volume 512x512x14.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. error, %</td>
<td>2.1</td>
<td>3.5</td>
<td>7.3</td>
<td>8.2</td>
<td>7.2</td>
</tr>
<tr>
<td>Max. error, %</td>
<td>12.6</td>
<td>8.6</td>
<td>34.3</td>
<td>41.4</td>
<td>37.2</td>
</tr>
<tr>
<td>Mean error, %</td>
<td>5.6</td>
<td>6.3</td>
<td>13.7</td>
<td>15.5</td>
<td>14.5</td>
</tr>
<tr>
<td>Stand. dev., %</td>
<td>4.3</td>
<td>2.4</td>
<td>11.5</td>
<td>14.5</td>
<td>12.8</td>
</tr>
<tr>
<td>Average time, sec</td>
<td>7.5</td>
<td>34.5</td>
<td>12.5</td>
<td>6.9</td>
<td>43.6</td>
</tr>
</tbody>
</table>