Bone Segmentation on MRI Scans

Todor Markov, William McCloskey
Advised by Samir Menon
Stanford University

I. INTRODUCTION

Diagnostic imaging is a principal method of analysis in modern medicine. Doctors and other medical professionals gain invaluable information about their patients through medical scans such as X-Rays, computed tomography (CT) cans, and magnetic resonance imaging (MRI). With this visual information, medical practitioners diagnose and monitor the health of their patients, and medical researchers gain insight into anatomical features of the subjects of the images. It is therefore desirable to improve the process as much as possible, and one way to do this is to automatically segment medical images.

Despite the bulk of research done on medical image segmentation, however, the successes are only partial. A multitude of algorithms have produced varying degrees of success, though no algorithm has performed well enough or universally enough to become the canonical choice for diagnostic image segmentation (Elnakib et al. 2011). Segmentation of medical images therefore remains a relatively open problem, and successful segmentation occurs only on a situational basis.

This project focused on finding an effective algorithm for bone segmentation in MRI scans. We wanted particularly to improve the process of musculoskeletal modeling, where most bone segmentation is done completely by hand. In detail, we input an MRI scan with four dimensions - three dimensions for the location of a voxel, and one dimension for the voxel's greyscale value - and we aimed to return an MRI scan with "bone" voxels highlighted differently from the rest of the tissue in the scan. The main feature that characterizes our problem is the prominence of bone among the rest of the MRI contents: bone mass generally appears much whiter than the surrounding tissue and has a sharp boundary around most but not all of its edges. (see Figure 1).

Our approach was as follows. For preprocessing, we ran anisotropic diffusion smoothing in order to mitigate noise in the images. We then ran variations of bias field correction in order to learn the bias field of the image and then to correct intensity inhomogeneities. Two models were used to segment the bones: K-means clustering, and region growing. Though several different variations of K-means were implemented, the resulting segmentations were seldom better than mediocre. Region growing, on the other hand, performed exceedingly well. Where a musculoskeletal researcher would spend several hours to segment a single bone, region growing performed the segmentation to a reasonable degree of accuracy in under fifteen minutes.

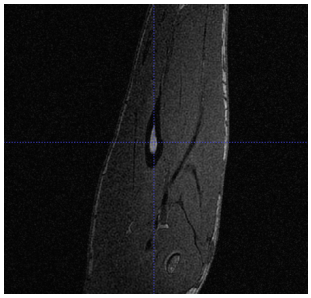

FIG. 1. A representative MRI bone scan of a forearm. The bone voxels are distinguishable from the other tissue by their white color.

II. DATA

Our data was a set of unlabeled MRI scans provided by the Stanford Artificial Intelligence Laboratory. For the purposes of musculoskeletal modeling, it was sufficient to find an algorithm that resulted in a fairly accurate - though perhaps not perfect - segmentation. Additionally, a quick automatic segmentation that segments the "obvious" bone voxels can be followed up by a careful hand segmentation, still saving a significant amount of time. Since relative accuracy can be judged by the naked eye, and since it would take multiple hours per bone to hand-label our data, we opted not to use labeled data.

III. PREPROCESSING

A. Anisotropic Diffusion

Anisotropic diffusion, also called Perona-Malik diffusion, is a technique aimed at reducing image noise without blurring edges or removing other significant parts of the image content. We opted to use it because it allows us to achieve a more homogenous intensity distribution in bone voxels while keeping the quality of the bone edges.

As the technique is not related to the topics we covered in class (and machine learning in general), we will only explain the general idea behind it. The way ordinary diffusion works is by taking the intensity of each voxel. Then we apply a discretized version of the heat equation to the image (we set the value of each voxel to a weighted mean of its 8-neighborhood or 26-neighborhood) for a number of iterations. However, this way, the blurring of the image affects the edges as well. To compensate for that, we change the conductivity factor in the heat equation from a constant to an appropriate function g of the image which promotes

diffusion within regions and inhibits it between regions. We tested both functions proposed in the Perona-Malik paper:

$$g(I) = \exp(-(\frac{||\nabla I||}{K})^2)$$

$$g(I) = \frac{1}{1 + (\frac{||\nabla I||}{K})^2}$$

Where I is the original image, ∇I is the gradient image, and K is a constant which controls sensitivity to edges. Both of the functions seemed to enhance the segmentation quality by a small amount (by reducing the number of false negatives), but neither was obviously better than the other.

B. Bias Field Correction

A bias field is a distortion of an MRI image. This field varies smoothly over the image, making entire regions of an MRI darker than other regions (see Figure 2). Bias fields are a problem for bone segmentation because non-bone voxels in light regions can easily become lighter than bone voxels in darker regions. Due to decreased visibility, bias fields also obstruct manual segmentation. All of our data had a noticeable bias field, which needed to be corrected. We tested two different approaches to bias field correction, obtained respectively from Juntu et al. and Mohamed et al., outlined below.

FIG. 2. An example bias field. It might be difficult to see, but there are actually two bones in this picture, which will be clear in the results section.

1. Parametric Surface Fitting

In our first approach, we extract a background that represents the lighting conditions from the original image, and then divide the original image by the background image in order to get the bias field corrected image. The background image is obtained by first applying a low-pass filter to the original image (we used Gaussian blurring with a large bandwidth). Then, several datapoints are selected from the blurred image (taking their coordinates and greyscale values), and a 3D topological manifold (we used a smooth polynomial surface) is fitted to this data. Finally, the fitted equation is used to generate the final background image representing the bias field.

We did the surface fitting by using the damped least squares method (also known as the Levenberg-Marquardt algorithm). Given a polynomial $f(x, y, z; a_1, a_2, \ldots, a_m)$ with parameters $a_1, a_2, \ldots, a_m$, we substitute the data points in it to get a set of equations:

$$g_1 = f(x_1, y_1, z_1; a_1, a_2, \ldots, a_m)$$

$$\vdots$$

$$g_n = f(x_n, y_n, z_n; a_1, a_2, \ldots, a_m)$$

We obtain the solution $a = (a_1, a_2, \ldots, a_n)$ to this set of equations by minimizing the sum of squares of the differences

$$Q(a) = \frac{1}{2}\sum_{i=1}^{n}(g_i - f(x_i, y_i, z_i; a_1, a_2, \ldots, a_m))^2$$

This is done iteratively by selecting an initial estimate of the parameters a_0 and then using a step similar to the Newton-Rhapson method:

$$a_{i+1} = a_i - (H + \lambda\, diag|H|)^{-1}\nabla Q(a_i)$$

Where H and ∇ are the Hessian and the gradient, and $diag|H|$ is the diagonal elements of the Hessian, and λ is a variable adjusted at each iteration which controls the speed of reduction.

2. Fuzzy C-Means

Our second algorithm was a Fuzzy C-Means bias field correction. As is common in the literature, we modeled the bias field additively:

$$y_k = x_k + \beta_k. \tag{3.1}$$

The observed intensity y_k is a "true" intensity x_k plus a bias term β_k that varies smoothly over the three-dimensional image.

In order to learn the bias field β_k, we minimized the following cost function of Mohamed et al., which uses Fuzzy C-means:

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^p ||y_k - \beta_k - v_i||^2 + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^p \left(\sum_{y_r \in N_k} ||y_r - \beta_r - v_i||^2\right).$$

Here, $\{v_i\}_{i=1}^c$ are the prototypes of the clusters, and the array $[u_{ik}] = U$ represents the partition matrix; N is the image volume (total number of voxels), N_k stands for the set of neighbors that exist in a window around voxel x_k, and N_R is the cardinality of N_k. This is done by taking the derivatives with respect to u_{ik}, v_i, and β_k and setting them to zero. The relevant equations we obtain

after we set the derivatives to 0 are below.
For the partition matrix:

$$u_i k = \frac{1}{\sum_{j=1}^{c} (\frac{D_{ik} + \frac{\alpha}{N_R}\gamma_i}{D_{jk} + \frac{\alpha}{N_R}\gamma_j})^{\frac{1}{p-1}}}$$

Where $D_{ik} = ||y_k - \beta_k - v_i||^2$ and $\gamma_i = \sum_{y_r \in N_k} ||y_r - \beta_r - v_i||^2$;
For the clusters:

$$v_i = \frac{\sum_{k=1}^{N} u_{ik}^p \Big((y_k - \beta_k) + \frac{\alpha}{N_R} \sum_{y_r \in N_k} (y_r - \beta_r) \Big)}{(1+\alpha) \sum_{k=1}^{N} u_{ik}^p};$$

For the bias field:

$$\beta_k = y_k - \frac{\sum_{i=1}^{c} u_{ik}^p v_i}{\sum_{i=1}^{c} u_{ik}^p}.$$

The algorithm works by selecting initial cluster prototypes and initial bias field estimate, and then iteratively updating the partition matrix, the cluster values and the bias field using the rules above until the clusters converge (or for a set number of iterations). Once the bias field is learned, the "true" intensities can be solved for in equation 3.1 to obtain a corrected image.

IV. MODELS

A. Region Growing

The region growing algorithm uses local spatial data to grow a homogeneous region. The algorithm takes in human-specified seed voxel, which is initialized to be the region. Then points are added to the region if they neighbor a voxel of the region and if their intensity falls within a certain threshold. Once this process terminates, the result is a region of voxels of homogeneous intensity connected spatially to the seed point.

We used two different ways to determine the intensity threshold required to add a voxel to the region. In the first implementation, we allow the user to supply an upper and lower bound. In the second implementation, we compute the mean and standard deviation of the intensity of all the voxels classified as bone, and use those values to set the threshold (upper and lower bounds are the mean plus a constant times the standard deviation.) Once a voxel is added to the region, the mean and standard deviation are recomputed before adding the next voxel o the region.

Since bones usually show up brighter than the rest of the scan and have prominent edges, region growing is an natural way to grow a bone voxel into a bone with minimal risk of leakage.

B. K-means Clustering

Our implementation of K-means is as follows. We reshaped the MRI from a three-dimensional image into a one-dimensional image. Then K-means clustering was performed on the one-dimensional image to separate the points into K clusters, where each cluster contained the points in a certain range of grayscale values. Since the "true" intensities of the bone voxels are brighter than the "true" intensities of the none-bone voxels, and since we correct for the "true" intensities with bias field correction, we expected the bone voxels to constitute the brightest clusters. To segment the image, we iterated through each the resulting clusters, setting the grayscale values of the points in the cluster to the grayscale value of the centroid representing the cluster. Then we mapped the points back to their original position in the three-dimensional MRI and displayed the result. In the final image, the whitest color clusters represented bone, and the darker colors represent other features of the scan.

1. Bone Brightening

Our main method for augmenting K-means was a technique we called bone brightening. We observed that the majority of a bone is surrounded by a sheath of black voxels. This sheath can be seen in the figures included throughout the paper. The white part of the scan, the part that shows up as bone, corresponds to bone marrow; the black sheath surrounding the bone corresponds to the hard bone around the marrow, which is prominent everywhere except at the joint and cartilage areas. Bone brightening attempts to use this black sheath to brighten the bone voxels more than the rest of the voxels. The algorithm is in some sense a bias field correction for bones: if the bones were brightened to a level much higher than the rest of the tissues, then K-means could successfully cluster the bones out of the rest of the image.

In our implementation, we first marked all of the voxels in the black sheath using region growing. Then, for each marked voxel, we brightened every non-marked voxel in a circle of a certain radius by a constant factor.

The geometry of the sheath and the bone ensures that the constant increments pile up on the bone voxels more than the surrounding region. To see this, consider the two-dimensional case: the black sheath corresponds to an annulus, the center of the annulus corresponds to bone, and the outside of the annulus corresponds to non-bone voxels. Assume we are brightening around a circle of radius r. A voxel v is brightened proportional to the area of the intersection of the annulus and the circle of radius r centered at v (because these are the marked sheath voxels that will brighten the bone). As long as the annulus is not too thin, an appropriate choice of r will cause the circles in bone voxels to intersect with the annulus at a large area (the area of the circle increases proportional to the square of the radius, so the area of the annulus is usually quite large.) In addition, all voxels outside of the annulus's outer circle intersect with the annulus at at most half their area, because the outer circle is convex. Hence the bone voxels are brightened more than the non-bone voxels.

V. RESULTS

Region growing resulted in quick and successful segmentation of a large majority of images. We visually estimated that region growing classifies at least 80 percent of the "obvious" bone voxels. The following scans are representative of region growing segmentation on our data set.

FIG. 3. Original wrist scan.

FIG. 4. Wrist after region growing.

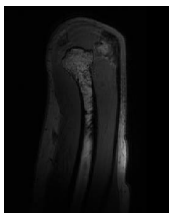

FIG. 5. Original Humerus scan.

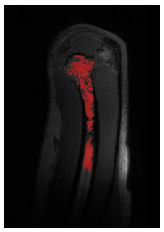

FIG. 6. Humerus after region growing.

On average, region growing takes around ten minutes per bone, where a manual segmentation might take multiple hours. The segmentations classify a majority of the bone, and, more importantly, rarely miscategorize a non-bone voxel as bone due to the sharp edges around most of the bone. Region growing struggles most near joints and in regions of severe intensity inhomogeneity where edges become less clear, though we found that both of these obstacles can usually be combated with appropriate choice of seed point and threshold.

Bias field correction successfully reduced intensity inhomogeneities, brightening the images and making them much more navigable to the human eye, as seen in Figure 7. While bias field correction dramatically improves

FIG. 7. Bias field correction on Figure 2 makes the second bone visible.

the process of manual segmentation, we found that bias field correction had little effect on the quality of region growing segmentation. In fact, we decided not to use bias field correction in our region growing segmentations because it smooths the edges to some degree in the original image, increasing the risk of leakage.

In addition, bias field correction was not effective enough to allow K-means to perform at a satisfactory level. K-means clusters solely based on voxel intensity, and hence is extremely sensitive to global intensity inhomogeneities. Thus, while K-means successfully clusters the brightest bone voxels, it often conflates dark bone voxels and light non-bone voxels:

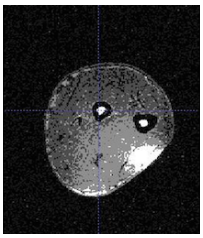

FIG. 8. K-means on a bias field corrected image

In a completely corrected image, K-means should perform very well; nevertheless, the true image could be only partially recovered.

Our other attempt at combating the bias field was bone brightening, which tries to brighten the bones relative to the surrounding tissue to a level where they could be clustered.

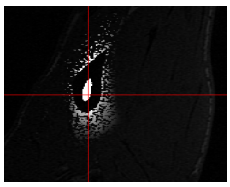

FIG. 9. Bone brightening successful locally.

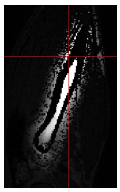

FIG. 10. Bone brightening unsuccessful globally.

Bone brightening performed exactly as desired on a local level, but it created its own type of bias field on a global level. The black sheath becomes narrower and wider at different areas along the bone; since brightening occurs proportional to the area of the circle intersecting the local black sheath, areas with a thick sheath become much brighter than areas with a thin sheath. We were unable to overcome this problem in bone brightening, though we are optimistic about the potential of some modification of the algorithm.

VI. ANALYSIS

The initial struggle of K-means surprised us, but in retrospect there is little mystery in the failings of the algorithm. K-means makes what turns out to be a strong assumption that voxel intensity is spatially invariant. We expected bias field correction to supplement K-means,

but both bias field corrections that we implemented were not strong enough to separate the bone from the rest of the tissue.

For the same reason, region growing's success is also unmysterious. Bone voxels have relatively homogeneous intensity, especially after edge preserving smoothing. Moreover, the majority (though not all) of the bone is surrounded by a black sheath, so the potential for region growing to leak into non-bone areas is present but minimal. Region growing takes into account only local inhomogeneities, so it still functions in a fairly efficient manner even in the presence of a bias field.

The relative inefficiency of bias field correction, on the other hand, is difficult to explain. While it could be explained by failure to converge for the Fuzzy C-means approach, and a poor choice of a surface polynomial and / or datapoints for the surface fitting approach, we still expected it to have a more noticeable effect on segmentation quality, especially given the fact that the bias field corrected images were much easier to do manual segmentation on.

VII. CONCLUSION AND FUTURE WORK

This paper implemented machine learning and segmentation algorithms with the aim of segmenting bones from MRI scans. Bias field correction performed at a level adequate for human segmentation, though not enough to aid the segmentation algorithms algorithms. K-means clustering gave subpar results despite modifications. Region growing performed at a reasonable level of accuracy, and gave rapid, quality segmentations. The region growing segmentation was at the level the project set out to achieve in its segmentation.

We identify a number of areas in need of future work. The most central of these is a satisfactory clustering algorithm. Although it is more difficult to implement clustering than region growing in the face of intensity in homogeneities, we think that the ceiling of accuracy for a clustering algorithm is much higher than that for region growing - because, in a "true" image, clustering should work excellently. One step towards this goal would be to cluster based on features that take into account on spatial data. We did try some simple features that took into account neighboring voxel intensities to no avail. But perhaps some more complex features might result in more accurate segmentation.

Another step toward better clustering algorithm is a better method of overcoming the bias field. We tried two methods of finding the bias field that are common in the literature, but neither was adequate. Perhaps different bias fields can be chosen to improve clustering. We are more optimistic about the prospects of bone brightening, which was a novel algorithm developed in this paper. K-means clustering on two-dimensional bone-brightened slices results in a quality clustering, so an extension of bone brightening to the global level should enhance K-means dramatically. The extension of bone brightening to the global level might appear to be as difficult as correcting bias field inhomogeneities. However, we deem the extension of bone brightening to be much more tractable than bias field correction because, rather than attempting to calculate a low level smooth distortion over an arbitrary image, bone brightening needs to investigate the properties of the black sheath, much of which is known beforehand.

Finally, the region growing algorithm can be improved upon and fine-tuned for better bone segmentation. This could be done by finding a better threshold, or by running region growing in tandem with other segmentation algorithms, such as a hypothetically improved clustering algorithm.

[1] P. Perona and J. Malik. "Scale-space and edge detection using anisotropic diffusion." Pattern Analysis and Machine Intelligence, IEEE Transactions on 12.7 (1990): 629- 639.

[2] M. Ahmed et al. "A modified fuzzy c-means algorithm for estimation and segmentation of MRI data." Medical Imaging, IEEE Transactions on 21.3 (2002): 193- 199.

[3] R. Adams and L. Bischof, Seeded Region Growing, IEEE Transactions on Pattevn Analysis and Machine Intelligence, vol. 16, no. 6, June 1994.

[4] A. Elnakib, G. Gimelfarb, J.S. Suri, A. El-Baz. Medical image segmentation: A Brief Survey. Multi modality state-of-the-art medical image segmentation and registration methodologies. Springer, New York (2011).

[5] Juntu, J. Sijbers, D.V. Dyck, J. Gielen. Bias field correction for MRI images. M. Kurzynski, E. Puchala, M. Wozniak, A. Zolnierek (Eds.), Proceedings of the 4th international conference on computer recognition systems, CORES 05; May 2225, 2005, Rydzyna Castle, Poland. Advances in soft computing, vol. 30. Springer (2005), pp. 543 - 551.