SEGMENTATION OF NEWBORN BRAIN MRI

N. I. Weisenfeld†, A. U. J. Mewes†, S. K. Warfield†

warfield@crl.med.harvard.edu
†Computational Radiology Laboratory
Children’s Hospital and Brigham and Women’s Hospital
Harvard Medical School, Boston, MA, USA
and
⋄Department of Cognitive and Neural Systems
Boston University, Boston, MA, USA

ABSTRACT
Segmenting neonatal brain images is challenging due to reduced MR signal-intensity contrast between tissues of interest compared to older subjects. In this paper we discuss a new strategy that we are currently employing in our neonatal brain studies and present the segmentation problem within a Bayesian framework. We show results validating this method and a previously published method from our group against a group of manually drawn segmentations. We achieve good to excellent results, on particularly challenging data, with greater speed and ease of use.

1. INTRODUCTION
Our group collaborates with researchers interested in studying the health, development, and the impact of various interventions on infants born premature. We have an institutional review board approved protocol for scanning newborn infants as soon as they are medically stable and at various times after an initial scan. We are interested in quantifying the development of grey matter, the maturation (myelination) of white matter, and any structural abnormalities that are present initially or develop over time. In order to facilitate this work, we seek to perform accurate classification of tissue types in magnetic resonance (MR) images. In this paper we present a new analysis strategy and validation results.

2. MATERIALS AND METHODS
2.1. Data Acquisition
Data were acquired on a 1.5T clinical MR scanner (General Electric Medical Systems, Wakeusa, WI, USA) using a head-surface RF receiver coil. The present study utilizes two MR scans: a T1-weighted SPGR (spoiled gradient recall) sequence and a T2-weighted fast spin echo scan. Our regular scan protocol includes other sequences not utilized in this study. The scans were all coronally-acquired and are 256x256 voxels in-plane. Slice thickness was 1.5mm for the SPGR and 3.0mm for the T2W scans and each voxel represents approximately 0.7mm x 0.7mm x 1.5mm,3.0mm of tissue.

2.2. Segmentation Process
2.2.1. Intensity Correction
Several factors contribute to intensity artifacts in MR images. Artifacts resulting from RF coil sensitivity differences across the imaging volume are well characterized by a smoothly varying, multiplicative field. These artifacts serve to confound statistical tissue classification by causing identical tissue samples to lead to differing MR intensities in different parts of the image volume. The correction of such artifacts has been studied in detail, and post-acquisition correction techniques [1, 2] in which the artifact is estimated from the imaging data, have been found to be the most suitable, due to their ability to recover subject-specific RF coil loading characteristics. Such correction is not possible with phantom-based methods since such phantoms exhibit different intensity artifacts than the brains under consideration here.

We have implemented and refined a method proposed in [3] to perform this correction. This method is practically parameter-free and hence very easy to initialize and robust in practice. The underlying model is that different tissue classes should be characterized by narrow intensity profiles, but that MR intensity artifacts cause this signal to vary across a wider range of values. The result is a spreading of the peaks in an intensity histogram, characterize this spreading of intensity histograms. Our implementation uses the information-theoretic principle of minimizing entropy, as proposed by Viola, together with a rapid simulated annealing algorithm [4],
as proposed by [3], to derive a correction field for the image volume that minimizes entropy in the image histogram while preserving the local intensity features of the original image.

2.2.2. Edge-Preserving Noise Smoothing

Typical magnetic resonance acquisitions, especially those acquired with small voxel dimensions, exhibit signal intensity fluctuations due to thermal noise. An effective increase in signal-to-noise ratio can be achieved by smoothing these random variations, but it is important to preserve the location and magnitude of edges between different structures. A number of anisotropic diffusion processes have been proposed, and have been demonstrated to successfully reduce the appearance of noise while preserving edges [5]. Most recently we have adopted and used successfully the curvature-based anisotropic diffusion edge preserving noise smoothing strategy implemented in the open source Insight toolkit (www.itk.org) [6].

2.2.3. Tissue Classification

The goal of our segmentation process is to generate a tissue classification by identifying with each voxel, on the basis of the MRI signal intensity and prior information, the most likely type of tissue present. With our studies of newborns, we seek to identify and distinguish cortical gray matter (cGM), myelinated and unmyelinated white matter (mWM and uWM), cerebrospinal fluid (CSF), basal ganglia structures (BG), and extracerebral tissue.

The signal intensity distribution for each of these tissue classes may overlap with each other. In such circumstances, identification of different types of tissue on the basis of signal intensity alone is not possible. It is possible, however, to reduce the amount of feature space overlap between different classes by considering the signal intensities of images with different contrast mechanisms simultaneously. We construct nonparametric estimates of the underlying class conditional probability density functions in order to maximize our capacity to differentiate between different structures. We can combine signal intensity information from any number of images (limited only by workstation memory size) and we have commonly used high spatial resolution MPRAGE or SPGR acquisitions alone, dual channel (for example, T1-weighted SPGR and T2 weighted MRI) acquisitions, and three channels formed by aligning a T1-weighted SPGR with proton density and T2 weighted images.

In our earlier work we have demonstrated that supervised statistical classification (the use of training data to learn how to classify different tissue types) has higher accuracy and robustness than unsupervised statistical classification (approaches which try to learn to identify different tissue types solely from algorithmic principles, without regard to training data or positive and negative examples) [7]. The MRI characteristics of the healthy and pathological tissue may include subtle contrast differences between different structures. We describe below the approach that we have developed for accurately estimating the class conditional probability density functions needed for our segmentation.

We estimate tissue class properties after intra-subject alignment of the enhanced, noise smoothed, MRI [8]. We then form a training data set by interactively labeling selected voxels with the correct tissue classification \( w_i; i = 1 \ldots W \) where \( W \) is the total number of tissue classes under consideration. Each such selected voxel is considered a sample from an \( n \)-dimensional feature space formed by the \( n \) types of images (e.g. MR channels). For each of the tissue classes we wish to segment, we identify a collection of labeled training samples.

Consider the problem of estimating an unknown probability density function (pdf), simply from a collection of samples drawn from the pdf. This process is described in detail in [9].

The probability that a sample \( x \) will fall in a region of the feature space \( R \) is given by \( P = \int_R p(x') dx' \). The probability \( P \) is therefore a smoothed (locally averaged) version of the density function \( p(x) \). We can therefore estimate the smoothed version of \( p(x) \) by estimating \( P \).

Consider a set consisting of \( n \) samples drawn independently from \( p(x) \). The probability that \( k \) of these \( n \) samples fall in the region \( R \) is given by the binomial law (since each sample either falls inside or outside the region \( R \)): \( P_k = \binom{n}{k} P^k (1 - P)^{n-k} \), where \( \binom{n}{k} = \frac{n!}{k!(n-k)!} \) is the number of ways of choosing \( k \) unordered samples from a set of \( n \) samples, and the expected value for \( k \) is \( E[k] = np \). This binomial distribution for \( k \) peaks sharply around the mean, so we can expect that a good estimator for \( P \) is \( E[k] / n \). If we also assume \( p(x) \) is continuous and that \( p(x) \) doesn’t vary appreciably inside the region \( R \) (say because \( R \) is small) then \( P = \int_R p(x') dx' \approx p(x) V \) where \( V \) is the volume of the feature space enclosed by the region \( R \). Hence, an estimator for the locally averaged probability density function is \( p(x) = k / n \). Furthermore, if \( k_i \) of our \( k \) samples that fall inside region \( R \) belong to class \( w_i \), then an estimator for the joint probability is \( p(x, w_i) = k_i / n \). The posterior probability of any tissue class can then be estimated by [9]:

\[
p(w_i | x) = \frac{p(x, w_i)}{\sum_{j=1}^W p(x, w_j)} = \frac{k_i}{k}
\]

This also gives us a robust and effective estimator for the likelihood \( p(x | w_i) = p(x, w_i) / p(w_i) \). We use this relationship to carry out maximum likelihood estimation of the tissue class for each voxel in the image.

The prior in the previous equation, \( p(w_i) \), is considered stationary across the image. Since the likelihood of finding a given class of tissue is not, in fact, the same everywhere in the brain, it may be beneficial to incorporate a statistical atlas (spatially varying prior) for each tissue class. Indeed, certain tissue classes could not be distinguished solely based on their MR signal properties alone, but are highly localized.
In order to construct a statistical atlas, one needs a method for representing the anatomy of anatomically similar individual subjects in the same coordinate system. A number of schemes have been proposed for this non-rigid, intrasubject registration of images. We employed a method proposed by Zöllei in her doctoral dissertation that achieves simultaneous registration of a group of subjects. This allows construction of an unbiased atlas and the performance of this algorithm is explored in [10].

For our atlas, twenty preterm infants were segmented by an expert reviewer using a previously published supervised classification system [11] and those segmentations were used to generate a statistical atlas. At a particular voxel coordinate, the prior probability of a tissue class \( p(w_i) \) is approximated by the number of times that tissue class appeared in that location, divided by the number of subjects. If \( b_{ij} \) is a binary labeling for tissue class \( w_i \) in subject \( j \), then \( p(w_i) \approx \frac{1}{N} \sum_{j=1}^{N} b_{ij} \), where \( N \) is the number of subjects.

The scheme presented thus far enables voxelwise independent tissue classification, based on the signal intensities at a particular voxel and, potentially, an atlas of voxelwise prior probabilities. A useful extension is to capture the spatial homogeneity of tissue classes, and a commonly used model is that of a Markov Random Field (MRF). An extensive investigation of this has been reported by Elfadel and it has been used to impose a spatial homogeneity constraint for image segmentation [12]. The mean field approximation has been used for constrained surface reconstruction, and its use has been motivated by the fact that it is the minimum variance Bayes estimator of the true field [13]. The mean field, \( \bar{P}_{ij} \), of the probability that class \( i \) is present at voxel \( j \) can be found by iterating the following relation to convergence:

\[
\bar{P}_{ij}^{(t+1)} \leftarrow \frac{1}{Z} p(x|w_i) p(w_i) \exp \left( \sum_m \sum_n J_{in} \bar{P}_{nm}^{(t)} \right)
\]

Here \( Z \) is a normalizing constant, \( n \) is an index over tissue classes, and \( m \) is an index over spatial locations in the neighborhood of voxel \( j \). \( J_{in} \) is the tissue class compatibility matrix, which acts to enhance spatial labelings that have compatible neighborhood labelings and to reduce the probability of spatial labelings with incompatible neighborhoods. \( J_{in} \) is often estimated by computing the joint occurrence statistics of label \( i \) and label \( n \) in training data. It can also be extended to incorporate spatial anisotropy in the tissue compatibility. Further discussion of this parameter is in, for example, [14].

The iteration is initialized by setting the initial estimates of the MRF to zero. In this way, the first iteration proceeds as in the voxelwise independence case, with the prior probabilities of each tissue class \( p(w_i) \), which may be estimated from previously segmented training data, and by estimating the class conditional probabilities \( p(x|w_i) \) for each voxel as described above. In practice, we carry out this iteration until the estimate \( \bar{P}_{ij}^{(t+1)} \) at each voxel converges as measured by changes of only a small amount from the previous iteration. Usually three to six iterations are sufficient.

### 3. RESULTS

Validation was performed by comparing the segmentation to ground truth estimates in each of 5 subjects. An estimate of ground truth was provided by having an expert ratemanually assign tissue labels to each voxel corresponding to a single MRI slice in each of our test subjects. None of these subjects was included in the atlas construction. The atlas was developed from 20 healthy preterm children, each born after 28 weeks gestational age (GA) and scanned at approximately 42 weeks. The test set comprises 5 subjects, each born prior to or at 28 weeks, two without white-matter injury, two with moderate white-matter injury, and a fifth with severe white-matter injury. These data provide a challenging test for our methods.

The two automatic segmentations were compared to the manual segmentation and an example is shown in Figure 1. Dice similarity coefficients (see [15]) are shown in Table 1 for each subject and each tissue class. Zijdenbos notes that greater than 0.7 is considered excellent agreement [15]. The previously reported semi-automatic method (“old”) [11] is compared against the present method both with the atlas of priors (“new”) and without (“no atlas”). The current method equals or outperforms the previously validated method and achieves excellent or near-excellent results. The “no atlas” results are shown for completeness, but they are not surprising. Without a spatially varying prior, it is difficult to distinguish cortical grey matter from myelin and there is a negative impact on the results for these two classes. Additionally, subcortical grey matter is difficult to distinguish from the partial-volume averaging that occurs at the grey-matter white-matter boundary. In this data, the Markov random field processing had little impact on the final results, so only the results with MRF were presented.
Table 1. Comparison showing higher accuracy and robustness of the new method versus old (with and without statistical prior “atlas”) when compared with single-slice manual segmentation using Dice similarity coefficient (see text).

<table>
<thead>
<tr>
<th>subj</th>
<th>alg</th>
<th>cGM</th>
<th>CSF</th>
<th>mWM</th>
<th>uWM</th>
<th>BG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>old</td>
<td>0.63</td>
<td>0.76</td>
<td>0.67</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>no atlas</td>
<td>0.23</td>
<td>0.77</td>
<td>0.48</td>
<td>0.76</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>0.76</td>
<td>0.66</td>
<td>0.67</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>2</td>
<td>old</td>
<td>0.67</td>
<td>0.67</td>
<td>0.77</td>
<td>0.61</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>no atlas</td>
<td>0.32</td>
<td>0.68</td>
<td>0.39</td>
<td>0.71</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>0.68</td>
<td>0.66</td>
<td>0.52</td>
<td>0.71</td>
<td>0.72</td>
</tr>
<tr>
<td>3</td>
<td>old</td>
<td>0.65</td>
<td>0.69</td>
<td>0.76</td>
<td>0.70</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>no atlas</td>
<td>0.30</td>
<td>0.72</td>
<td>0.33</td>
<td>0.73</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>0.77</td>
<td>0.71</td>
<td>0.77</td>
<td>0.77</td>
<td>0.73</td>
</tr>
<tr>
<td>4</td>
<td>old</td>
<td>0.70</td>
<td>0.43</td>
<td>0.56</td>
<td>0.62</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>no atlas</td>
<td>0.56</td>
<td>0.54</td>
<td>0.30</td>
<td>0.71</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>0.73</td>
<td>0.51</td>
<td>0.71</td>
<td>0.71</td>
<td>0.64</td>
</tr>
<tr>
<td>5</td>
<td>old</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>no atlas</td>
<td>0.48</td>
<td>0.69</td>
<td>0.39</td>
<td>0.72</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>0.81</td>
<td>0.70</td>
<td>0.75</td>
<td>0.72</td>
<td>0.76</td>
</tr>
</tbody>
</table>

mean

<table>
<thead>
<tr>
<th></th>
<th>old</th>
<th>0.67</th>
<th>0.65</th>
<th>0.69</th>
<th>0.65</th>
<th>0.65</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no atlas</td>
<td>0.03</td>
<td>0.13</td>
<td>0.09</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>0.38</td>
<td>0.69</td>
<td>0.39</td>
<td>0.72</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>0.14</td>
<td>0.09</td>
<td>0.08</td>
<td>0.02</td>
<td>0.04</td>
</tr>
</tbody>
</table>

mean

<table>
<thead>
<tr>
<th></th>
<th>old</th>
<th>0.75</th>
<th>0.65</th>
<th>0.69</th>
<th>0.74</th>
<th>0.72</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no atlas</td>
<td>0.05</td>
<td>0.08</td>
<td>0.1</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>0.05</td>
<td>0.08</td>
<td>0.1</td>
<td>0.03</td>
<td>0.05</td>
</tr>
</tbody>
</table>

4. DISCUSSION AND CONCLUSION

Our previous segmentation method has been shown to predict 2-year outcomes from scans taken at 40 weeks gestational age. Here we have shown the current work to equal or exceed the previous method in accuracy, on average, for five challenging segmentations. However, the previous methodology took a trained reviewer several iterations to achieve adequate results and similar, and often better, results were achieved here in far less time. The clinical significance of studies done with this technique, combined with the current speed and ease of use, make it possible that this technique will find routine use in the clinical evaluation of premature infants at our institution. Finally, the current formulation provides an intuitive and consistent conceptual framework with verifiable results at each stage of processing. Future work should include testing on a larger sample of children and further analysis of the effect of each stage of the processing pipeline.

5. REFERENCES


