Available online at www.sciencedirect.com
SCIENCE DIRECT®
Medical Image Analysis 9 (2005) 457–466
MEDICAL IMAGE ANALYSIS
www.elsevier.com/locate/media

Automatic segmentation of MR images of the developing newborn brain ☆

Marcel Prastawa [a,*], John H. Gilmore [b], Weili Lin [c], Guido Gerig [a,b]

[a] *Department of Computer Science, University of North Carolina, CB #3175 Sitterson Hall, Chapel Hill, NC 27599, USA*
[b] *Department of Psychiatry, University of North Carolina, Chapel Hill, NC 27599, USA*
[c] *Department of Radiology, University of North Carolina, Chapel Hill, NC 27599, USA*

Available online 12 July 2005

Abstract

This paper describes an automatic tissue segmentation method for newborn brains from magnetic resonance images (MRI). The analysis and study of newborn brain MRI is of great interest due to its potential for studying early growth patterns and morphological changes in neurodevelopmental disorders. Automatic segmentation of newborn MRI is a challenging task mainly due to the low intensity contrast and the growth process of the white matter tissue. Newborn white matter tissue undergoes a rapid myelination process, where the nerves are covered in myelin sheathes. It is necessary to identify the white matter tissue as myelinated or non-myelinated regions. The degree of myelination is a fractional voxel property that represents regional changes of white matter as a function of age. Our method makes use of a registered probabilistic brain atlas. The method first uses robust graph clustering and parameter estimation to find the initial intensity distributions. The distribution estimates are then used together with the spatial priors to perform bias correction. Finally, the method refines the segmentation using training sample pruning and non-parametric kernel density estimation. Our results demonstrate that the method is able to segment the brain tissue and identify myelinated and non-myelinated white matter regions.

Keywords: Automatic brain MRI classification; Automatic brain MRI segmentation; Early brain development; Kernel density estimation; Neonatal MRI; Robust estimation

1. Introduction

The segmentation of newborn brain structures from magnetic resonance images (MRI) is crucial for the study of normal development and comparison to neurodevelopmental disorders at early stages. The development of new segmentation methods for this age group is driven by the increasing use of MRI to study newborns, for example our ongoing study of early brain development in normal and high risk children (Zhai et al., 2003; Gilmore et al., 2004) and the lack of appropriate segmentation methodology. Manual segmentation of newborn brains is tedious, time consuming, and lacks reproducibility. Therefore, it is necessary to use automatic segmentation methods for clinical studies with multiple subjects. This task is considerably more challenging compared to automatic segmentation of adult brain MRI due to the early development process; Rutherford (2002) provides an excellent description of newborn MRI and the dynamic changes seen over the early development period.

In newborn infant brains, the white matter structure undergoes myelination, where the fibers are being

☆ This research is supported by the UNC Schizophrenia Research Center, an NIMH Conte Center MH064065 (PI J. A. Lieberman and J.H. Gilmore) and the UNC Neurodevelopmental Disorders Research Center HD 03110 (PI J. Piven). Marcel Prastawa is supported by NIH-NIBIB R01 EB000219 (PI E. Bullitt).
* Corresponding author. Tel.: +1 919 962 1836.
E-mail address: prastawa@cs.unc.edu (M. Prastawa).

1361-8415/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2005.05.007

covered in myelin sheathes. At birth, the white matter of the brain stem and the posterior limbs of the internal capsule are myelinated and have white to gray matter contrast similar to that of adults (white matter is brighter than gray matter in the T1w image). Other regions of white matter such as the centrum semi-ovale are not myelinated and the white to gray matter contrast is inverted (white matter is darker than gray matter in the T1w image). As the child ages from birth to one year, myelination progresses through the anterior limbs of the internal capsule, the occipital radiations, and then to the frontal white matter. As this happens, the MR relaxation times of these regions change with the new myelinated fibers consequently changing the MRI signal. By 1.5 years, the MR image contrast is almost adult-like. Fig. 1 shows an example of a newborn MR image with the myelinated and non-myelinated white matter regions.

Several methods have been developed for automatically segmenting healthy adult brain MRI, mostly variations of multi-variate statistical classification techniques. Wells et al. (1996) proposed an Expectation-Maximization scheme that interleaves segmentation and intensity bias correction. This method was extended by Van Leemput et al. (1999b) through the use of a probabilistic brain atlas. Warfield et al. (2000) described a k-nearest neighbor classification algorithm that is combined with template matching. Cocosco et al. (2003) combines atlas sample selection and robust sample pruning using minimum spanning trees (MST) for intensity-based classification. A different class of segmentation techniques uses deformable templates that transfer the labeling of a template to each new subject. For example, the work by Collins et al. (1999) which combines neural network classification with nonlinear image matching.

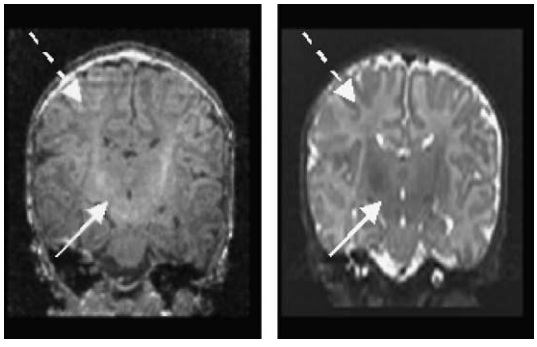

Fig. 1. MR images of a newborn brain (subject 0096, coronal view). Left: T1w image, right: T2w image. The arrows show the white matter structure. The arrow with the solid line indicates myelinated white matter, the arrow with the dashed line indicates non-myelinated white matter. Early myelination in white matter is shown as bright regions in the T1w image and dark regions in the T2w image.

Automatic segmentation methods for healthy adult brain MRI typically fail in segmenting all the different structures apparent in newborn brain MRI, particularly the myelinated white matter regions. Methods that use probabilistic brain atlases (Van Leemput et al., 1999b; Collins et al., 1999) or templates (Warfield et al., 2000) cannot be directly applied to newborn brain MRI since the spatial prior information for rapidly changing myelination property would be very difficult to define. Warfield et al. (2000) uses a specific template for newborn brains with predefined classifications for myelinated and non-myelinated white matter. Methods that are driven by image intensities (Wells et al., 1996; Cocosco et al., 2003) would have difficulties in the initialization phase. The MR image intensities for newborn brains are significantly affected by both low contrast and RF inhomogeneity, which can be difficult to overcome without spatial prior information.

Matsuzawa et al. (2001) presented a segmentation method for infant brain MRI, as part of a study of early brain development. Their method does not identify myelinated white matter and non-myelinated white matter separately. The results show that their method has difficulties dealing with tissue separation. Hüppi et al. (1998) and Inder et al. (2005) showed segmentation results of newborn infants, using the method of Warfield et al. (2000). They study both prematurely born infants and normal infants. The prematurely born infants tend to have simpler cortical folding compared to normal newborns. The segmentation method identifies non-myelinated and myelinated white matter. Boardman et al. (2003) used image deformation for detecting regions of major development.

Automatic segmentation of newborn brain MRI is significantly more challenging than the segmentation of adult brain MRI. This is mainly due to the biology and the rapid growth process. The specific challenges are:

(1) The white matter and gray matter contrast to noise ratio (CNR) for newborn MRI can be as low as half of the one for adult brain MRI. Factors that reduce CNR are the small size of the infant brains and the short scanning period. The small head size requires them to be scanned at higher resolution, which leads to higher noise levels. The infants need to be scanned in very short time since they are not sedated or constrained. The low CNR causes difficulty in segmenting the partial volume regions.
(2) Typically, newborn brain MRI exhibits motion artifacts even with very short scan sequences. The infants may not stay motionless during the scan period. This problem can be difficult to solve since the infants are not mentally aware, and healthy infants cannot be sedated or restrained due to ethical reasons.

(3) The process of myelination separates white matter tissue into two types: myelinated and not myelinated. We treat myelination as a fractional property because the MR image intensities reflect the degree of myelination and partial voluming. The dividing boundaries between regions that are fully myelinated and non-myelinated are generally ambiguous (Rutherford, 2002). The myelinated white matter regions are mostly distributed near the spine (central posterior) and parts of the internal capsule. We also observed the presence of myelinated white matter around the regions associated with the sensory and motor cortex.
(4) Each tissue type in newborn brain MRI exhibits significant levels of intensity inhomogeneity and variability, which may be due to a combination of RF inhomogeneity and biological properties of the developing tissue (Kandel et al., 2000).
(5) The different tissues have large overlaps in their intensity characteristics, as shown in Fig. 2. The decision boundaries for intensity-based classification are typically ambiguous and complex.

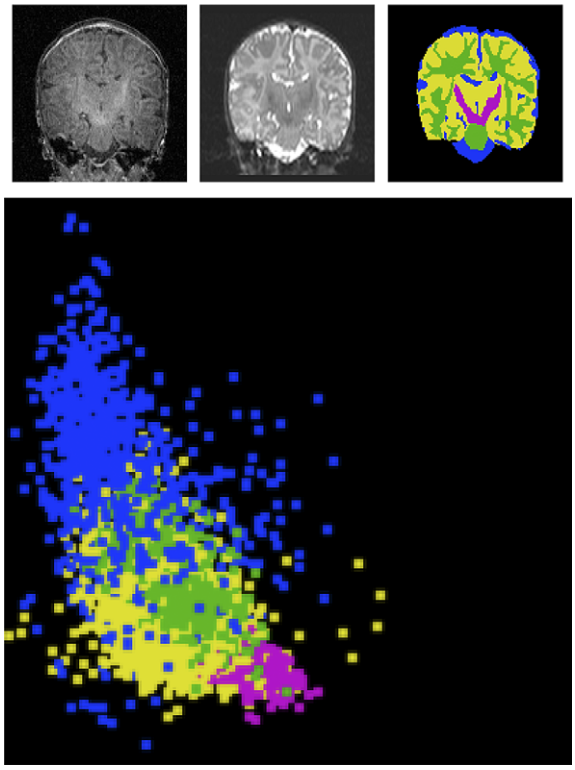

Fig. 2. Intensity characteristics of one coronal slice of a newborn brain MRI dataset (subject 0096). Top, from left to right: T1w image, T2w image, and the manually assigned labels. Purple is myelinated white matter, green is non-myelinated white matter, yellow is gray matter, and blue is cerebrospinal fluid. Bottom: the scatterplot of the tissue intensities, the horizontal axis represents T1w intensities and the vertical axis represents T2w intensities. There is significant overlap between the intensities of different tissues, and there are ambiguities in the decision boundaries.

We developed an atlas based segmentation algorithm for newborn brain MRI that addresses the challenges listed above. The method incorporates the robust clustering method proposed by Cocosco et al. (2003) and the robust parameter estimation method presented by Rousseeuw and Van Driessen (1999) to deal with noisy data. It uses the intensity inhomogeneity estimation scheme from spatial classification proposed by Van Leemput et al. (1999a). The complex decision boundaries are modeled using non-parametric kernel density estimates, using the efficient method of Girolami and He (2003). The probabilistic atlas is used as a spatial prior in the classification process as proposed by Van Leemput et al. (1999a).

2. Method

Due to the large overlap in the tissue intensity distributions, it becomes necessary to use spatial priors for the segmentation. The spatial priors that we use is part of a probabilistic brain atlas of newborn MRI, shown in Fig. 3. The atlas provides voxel prior probabilities for white matter, gray matter, and cerebrospinal fluid (csf). Myelinated white matter and non-myelinated white matter are combined as one white matter class in the atlas. This is necessary because it is difficult to model the different dynamic growth patterns across subjects given the significant changes during early brain development. With the combined white matter prior, the discrimination between the two different white matter classes is primarily driven by the image intensities. The atlas was created by averaging three semi-automatic segmentations registered using affine transformation. Each segmentation was done by a human rater that selects samples for each tissue types for k-nearest neighbor segmentation. The outputs of the k-nearest neighbor classification are then edited by manual outlining. The number of subjects is insufficient to create prior

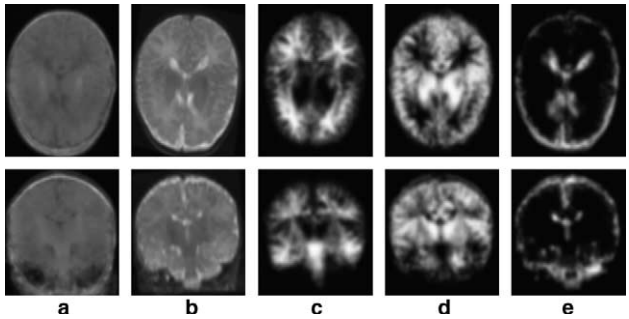

Fig. 3. The probabilistic brain atlas of a newborn brain. From left to right: (a) the T1w average image, (b) the T2w average image, and the spatial prior probability values for (c) white matter (either myelinated or non-myelinated), (d) gray matter, and (e) csf. Top: axial view. Bottom: coronal view.

probabilities that reflect the variability in the population. At this point, we are limited by the size of our datasets and the amount of time for manual processing. To compensate for the lack of available data, an additional blurring is applied to the average segmentations. The blurred spatial probabilities simulate an atlas with a higher level of population variability.

Our segmentation framework is composed of three major steps, as shown in Fig. 4. First, it obtains rough estimates of the class intensity clusters. It then iteratively performs inhomogeneity correction and parametric classification. Finally, it refines the segmentation using non-parametric kernel density estimates.

Before segmentation, we register the atlas to the subject using affine transformation and the mutual information image match metric (Maes et al., 1997). In this study, we use T1w and T2w MR images, where the T2w images are registered to the T1w images also by affine transformation and mutual information metric. The lower resolution T2w images are up-sampled with spline interpolation. The registered subject MR images are filtered using anisotropic diffusion (Gerig et al., 1992) to reduce noise and motion artifacts.

2.1. Robust intensity distribution estimation

The segmentation of newborn brain MRI involves classifying each voxel into different categories C, where C is commonly defined to be {myelinated white matter, non-myelinated white matter, gray matter, and cerebrospinal fluid}. The first step in the segmentation process is to determine the rough estimates of the class intensity distributions. We obtain samples for class C_i at location $\vec{x}$ with high atlas prior probability values, for example $\Pr(\vec{x}|C_i) > 0.9$ as presented by Cocosco et al. (2003).

The white matter samples are constrained to have low image gradient magnitude values to avoid choosing samples near the transition regions between myelinated and non-myelinated white matter and at white/gray matter boundaries. The value we use for the gradient magnitude of our 3-D images is the 2-norm of the vector of individual gradient magnitudes, where

$$G(\vec{x}) = \sqrt{|\nabla I_1(\vec{x})|^2 + \cdots + |\nabla I_n(\vec{x})|^2}. \quad (1)$$

We only retain samples for the white matter class with $G(\vec{x})$ lower than the average of $G(\vec{x})$ over the white matter prior, $\gamma = [\sum_{\vec{x}} \Pr(\vec{x}|\text{white-matter})G(\vec{x})]/[\sum_{\vec{x}}\Pr(\vec{x}|\text{white-matter})]$. The 2-norm gradient magnitude metric is more sensitive to noise compared to the vector field gradient magnitude metric described in Lee and Cok (1991). This is a desired property since we want to avoid sampling noisy regions.

We then process the obtained intensity samples to remove outliers and false positives. We use the minimum covariance determinants (MCD) estimator (Rousseeuw and Van Driessen, 1999) to generate the robust mean and covariance estimates of the unimodal distributions (gray matter and csf). The MCD estimator computes the robust mean and covariance that have the smallest determinant of covariance and covers at least half of the data. For the bi-modal white matter distribution, we use a robust graph based clustering method, similar to the one described in (Cocosco et al., 2003). The clustering method creates the minimum spanning tree (MST) graph (Cormen et al., 2001) from the sample points and breaks long edges to form the clusters (Duda et al., 2001). The minimum spanning tree is the graph where all the points are connected such that the total edge lengths are minimized. The MST graph does not have any closed loops (cycles). The removal of samples with high image gradient helps in the MST clustering process, as shown in Fig. 5.

The algorithm searches for myelinated white matter and non-myelinated white matter intensity clusters by iteratively breaking long edges of the MST. At each iteration, we break an undirected edge $e(v,w)$ that connects vertices v and w if it is longer than $A(v) \times T$ or $A(w) \times T$. $A(v)$ is the average length of edges incident on vertex v, $A(v) = \frac{1}{n_s}\sum_s |e(v,s)|$, while T is a distance multiplier. The edge breaking results in subtrees where each subtree forms an intensity cluster. For each detected cluster, an intensity location estimate is computed. The cluster intensity location estimate provides an approximation of where most points in the cluster is distributed in the intensity space. The iterative algorithm terminates when

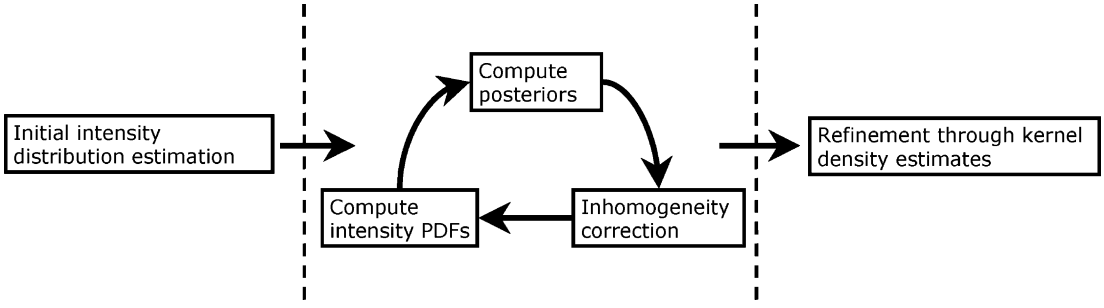

Fig. 4. The segmentation framework.

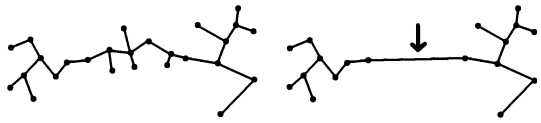

Fig. 5. Illustrations of the minimum spanning trees for white matter obtained using different sampling strategies. Left: Samples with high probability values. Right: Samples with high probability values and low gradient magnitude. Choosing only samples with low gradient magnitude helps to remove samples from the transition regions between myelinated white matter and non-myelinated white matter and gray/white boundary voxels. This is crucial for clustering based on edge breaking. As seen on the right picture, breaking the longest edge marked by the arrow would give two well separated clusters.

two clusters are found with intensity location estimates that are in the proper order. For example, the order of intensities for the classes in T2w from darkest to brightest is myelinated white matter, gray matter, non-myelinated white matter, and csf.

The intensity location estimates for the two white matter classes are also computed using the MCD estimator. We use the robust MCD mean values, as opposed to the standard location estimates such as the mean or median, to make sure that we obtain reasonable sample clusters. The standard location estimates such as mean or median may not always be optimal for the noisy newborn MRI data. The mean value can be skewed by a single outlier sample, while the median value only uses one sample point and ignores contributions of other samples. The initial intensity distributions for non-myelinated white matter and myelinated white matter are computed as the MCD mean and covariance estimates of the largest detected clusters. The MCD estimator therefore serves to estimate the initial intensity distributions. The initial gray matter and csf distributions are the MCD estimates of the atlas sampled data. The initial white matter distributions are the MCD estimates of the atlas samples that are clustered and pruned using MST. The steps involved in the intensity distribution estimation are listed in Algorithm 1.

Algorithm 1. Initial intensity distribution estimation

1: Obtain samples by thresholding atlas prior probabilities
2: Remove white matter samples with gradient magnitude higher than γ
3: Compute robust mean intensity values for gray matter and csf (μ_{gm} and μ_{csf}) using the MCD estimator
4: Construct Minimum Spanning Tree from white matter samples
5: $T \leftarrow 2$
6: **repeat**
7: Break edges longer than $T \times A$, where A is the average length of connected neighbor edges
8: Find largest myelinated white matter cluster, where $\mu_{myelinated} < \mu_{gm}$ in T2w
9: Find largest non-myelinated white matter cluster, where $\mu_{gm} < \mu_{non\text{-}myelinated} < \mu_{csf}$ in T2w
10: $T \leftarrow T - 0.01$
11: **until** both white matter clusters are found or $T \leqslant 1$
12: **if** $T < 1$
13: Algorithm fails
14: **end if**
15: Compute white matter Gaussian distribution parameters from detected clusters

2.2. Inhomogeneity correction

Newborn brain MRI exhibit higher intensity variability for each tissue and low intensity contrast compared to adult brain MRI. These two factors severely hamper the estimation of intensity inhomogeneity. Histogram based intensity inhomogeneity estimation methods, such as the ones proposed by Sled et al. (1998) and Styner et al. (2000), are likely to have difficulties in obtaining the optimal solution. The histogram of a newborn brain MR image is generally smooth with weak maximas.

In the case of inhomogeneity correction of newborn brain MRI, the spatial information is useful to deal with the low intensity contrast. We have chosen to use the method developed by Van Leemput et al. (1999a). The scheme uses the spatial posterior probabilities to estimate the intensity inhomogeneity, which helps to overcome problems with low contrast and high variability. The inhomogeneity estimation method is an iterative generalized expectation maximization algorithm. It interleaves classification with inhomogeneity estimation at each iteration. The Gaussian distributions obtained from the previous segmentation step are used as initial parameters for the iterative inhomogeneity estimation algorithm.

The intensity likelihoods are modeled using parametric Gaussian functions, and the inhomogeneity is modeled using polynomials:

$$p(\vec{I}(\vec{x})|C_i) = \phi_{\Sigma_i}(\vec{I}(\vec{x}) - \mu_i - \sum_k \beta_k q_k(\vec{x})), \quad (2)$$

where ϕ is the Gaussian function, with mean μ_i and covariance Σ_i, the intensity inhomogeneity is the linear combination of the coefficients β_k and the basis polynomials q_k. The intensity inhomogeneity is estimated by least squares fitting of the polynomial coefficients to the log difference of the original image and the reconstructed image. The reconstructed image is the homogeneous image computed with the mean values μ_i and the class posterior probabilities. The class posterior probabilities are computed using the image intensity likelihood probabilities and the atlas prior probabilities:

$$p(C_i|\vec{I}(\vec{x})) = \frac{\sum_i p(\vec{I}(\vec{x})|C_i)\Pr(\vec{x}, C_i)}{\sum_i \Pr(\vec{x}, C_i)}, \quad (3)$$

where $\Pr(C_i, \vec{x})$ is a combination of the atlas prior probabilities ($\Pr(\vec{x}|C_i)$) and the global class prior probabilities ($\Pr(C_i)$): $\Pr(\vec{x}, C_i) = \Pr(\vec{x}|C_i)\Pr(C_i)$. The myelinated and non-myelinated white matter shares the same atlas prior, $\Pr(\vec{x}|C_1) = \Pr(\vec{x}|C_2)$. The global class prior probabilities can be tuned based on the age of the newborns to be segmented. For the results presented here, we set the global class priors such that white matter is more likely to be not myelinated: $\Pr(C_1) = 0.2$, $\Pr(C_2) = 0.8$, and $\Pr(C_i) = 1$ for the other classes. The use of the atlas spatial prior probabilities $\Pr(\vec{x}|C_i)$ helps resolve ambiguities that are caused by the low image contrast, following the formulation used in Van Leemput et al. (1999b).

2.3. Segmentation refinement using kernel density estimation

The class intensity likelihoods are modeled as Gaussian probability density functions in the segmentation and inhomogeneity correction to obtain an optimal parametric solution. The use of the parametric Gaussian distribution eases the computation of the maximum likelihood estimate. However, Gaussian distributions can have significant overlap and therefore result in degenerate decision boundaries. In order to capture the complex and ambiguous intensity characteristics of newborn brain MRI, we switch from the parametric Gaussian distribution to a non-parametric distribution estimate. We refine the classification by sampling the inhomogeneity corrected images, pruning the outliers and false positives from the intensity samples, and then estimating the intensity distribution using kernel density functions (Duda et al., 2001; Hastie et al., 2001).

The non-parametric intensity probability density function for each class is estimated as follows:

$$\hat{p}(\vec{I}(\vec{x})|C_i) = \sum_{j=1}^{N_i} w_{ij} K_h(\vec{I}(\vec{x}) - T_{ij}), \quad (4)$$

where K_h is the Gaussian kernel with standard deviation h, N_i is the number of training samples for class C_i, and T_{ij} is the jth training sample for the ith class. Each training sample has an associated weight w_{ij}, where for each class C_i, $\sum_{j=1}^{N_i} w_{ij} = 1$. The kernel density estimates are used to produce the final classification results, which are the class posterior probabilities:

$$\hat{p}(C_i|\vec{I}(\vec{x})) = \frac{\sum_i \hat{p}(\vec{I}(\vec{x})|C_i)\Pr(C_i, \vec{x})}{\sum_i \Pr(C_i, \vec{x})}. \quad (5)$$

The atlas spatial prior probabilities are also used at this stage. The spatial priors are combined with the non-parametric kernel densities to provide class posterior probabilities that are capable of capturing more complex intensity characteristics.

The set of training samples T for the kernel density estimates are obtained by sampling the MR images using the previously obtained posterior probabilities. Each sample T_{ij} is obtained by selecting features at location $\vec{x}$ where

$$\arg\max_{C_k} p(C_k|\vec{I}(\vec{x})) = C_i. \quad (6)$$

The samples are pruned and clustered using the robust MST-based method proposed by Cocosco et al. (2003). This step removes the false positives and outliers in the intensity data resulting from using Gaussian distribution estimates in the previous step.

The method proposed by Girolami and He (2003) is applied to efficiently estimate the kernel density function. This method speeds up the density estimation process by reducing the size of the training set. The weights w_{ij} are chosen to minimize the integrated squared error between the true density function and the estimated kernel density function. Redundant training features are assigned lower weight values compared to characteristic training features. This minimization process for the sample weight assignment is similar to the quadratic optimization process for Support Vector Machines, for which an efficient solution exists (Schölkopf et al., 2001). The samples with zero weights are removed from the training set, which effectively eliminates the redundant features in the training set. Compared to other fast density estimation techniques such as pre-binning (Scott and Sheather, 1985) and multi-scale selection using hyperdiscs (Mitra et al., 2002), this method has the advantage of having only one user specified parameter: the kernel width or the standard deviation of the Gaussian kernels.

3. Results

We show the application of our new segmentation method to four different subjects. Fig. 6 shows the coronal view of the MR images along with two sets of manual segmentation slices done by different raters. Fig. 7 shows the coronal view of the automatic segmentation results. The 3D volumes for the automatically segmented structures are listed in Table 1. The four cases are samples from a large neonatal study at UNC Chapel Hill to assess early brain development in normal and high risk children (Zhai et al., 2003; Gilmore et al., 2004). We currently have over 50 datasets of neonatal MRI and will collect a total of 125, with some of them followed-up at the age of one year. As part of the study, we plan to measure the cortical folding and the cortical thickness of the newborn brains. Fig. 8 illustrates the 3D view of the relevant structures for one of the subjects. Visual inspection of the results show that the myelinated

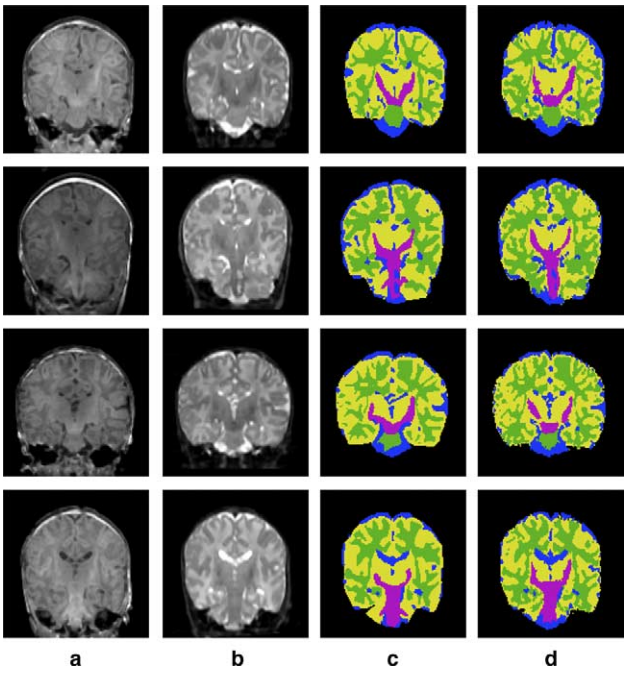

Fig. 6. The MR images along with the manually segmented labels. From left to right: (a) T1w image, (b) T2w image, (c) color image showing the segmentation done by the first human rater, and (d) color image showing the segmentation done by the second human rater. Purple is myelinated white matter, green is non-myelinated white matter, yellow is gray matter, and blue is csf. From top to bottom: subject 0096, 0117, 0118, and 0123.

white matter regions are mostly distributed near the spine (central posterior) and internal capsule. We have also observed the presence of small regions of myelinated white matter around the regions associated with the sensory and motor cortex.

Images were acquired on a Siemens head-only 3T scanner (Allegra, Siemens Medical System, Erlangen, Germany). Two structural imaging sequences were used: a magnetization prepared rapid gradient echo (MPRAGE) T1-weighted and a turbo spin echo (TSE), dual-echo (proton density and T2 weighted). Total scan time for structural scans was approximately 10 min. The imaging parameters for the MP-RAGE sequence were: repeat time TR = 11.1 ms, echo time TE = 4.3 ms, inversion time TI = 400 ms, slice thickness TH = 1 mm, in-plane resolution = 0.898×0.898 mm^2. A total of 128 sagittal images were acquired to cover the entire brain. The imaging parameters for the TSE sequence were: TR = 7 s, TE = 15 and 90 ms, TH = 1.95 mm, in-plane resolution 1.25×1.25 mm^2, and 56 slices.

Validation of the automatic segmentation results are difficult because a gold standard does not exist. The common standard, manual segmentations, is difficult to obtain since highly convoluted structures in low-contrast, noisy data are very hard to trace. In addition to that, the myelinated white matter and the non-myelinated white matter have ambiguous boundaries, which

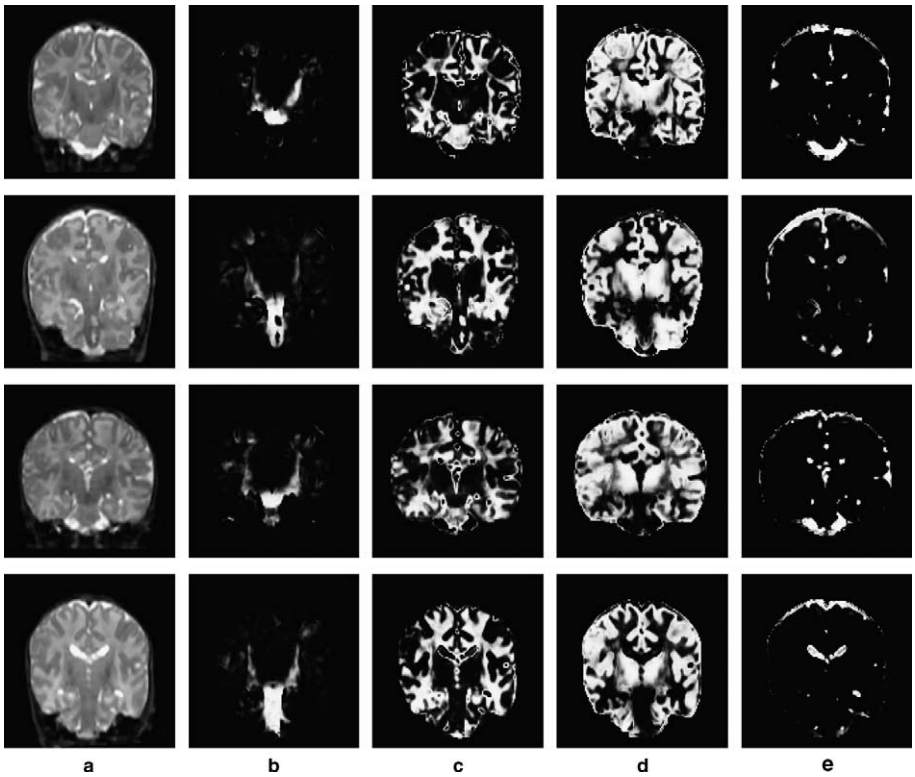

Fig. 7. Coronal view of the 3D automatic segmentation results. From left to right: (a) the T2w image and the class posterior probabilities for (b) myelinated white matter, (c) non-myelinated white matter, (d) gray matter, and (e) cerebrospinal fluid. From top to bottom: subject 0096, 0117, 0118, and 0123.

would make manual segmentation results highly variable and difficult to reproduce. We have done a limited validation of the segmentation results by restricting the validation to only a 2D coronal slice of each dataset. The human raters assign discrete labels to each voxel in the slice. The degree of myelination is not specified because it is extremely difficult for the raters to consistently assign a continuous weight for myelination. The posterior probabilities that are generated by the automatic segmentation method are discretized following Eq. (6).

We use Cohen's κ (Cohen, 1960) to measure the segmentation variability. The κ value given two observations and N classes is defined as

$$\kappa = \frac{\sum_{i=1}^{N} \text{agree}(C_i) - \sum_{i=1}^{N} \text{ef}(C_i)}{N - \sum_{i=1}^{N} \text{ef}(C_i)}, \tag{7}$$

where $\text{agree}(C_i)$ is the number of agreements between the two observers for class C_i and $\text{ef}(C_i)$ is the expected frequency of agreement by chance for class C_i. Given the number of observations of class C_i for the first and second observer, a_i and b_i, the expected frequency of agreement is $\text{ef}(C_i) = \frac{1}{N} a_i b_i$. This reliability measure places equal weight on the samples of each class, so classes with larger number of observations will have more influence on the final result. The κ value is normalized, 0 indicates independence and 1 indicates complete agreement. κ values above 0.7 is generally interpreted to reflect a satisfactory level of reliability. The κ values comparing the two manual segmentations and the manual segmentation against the automatic segmentation is shown in Table 2.

We also measure the overlap of the segmentations of each class using the Dice similarity coefficient (Dice, 1945). For two segmentations A and B, the Dice similarity coefficient is defined as $2|A \cap B|/(|A| + |B|)$. This overlap measure is normalized, where 0 indicates complete dissimilarity and 1 indicates complete agreement. The overlap values reflecting inter-rater variability is shown in Table 3. The overlap comparison between the manual raters and the automatic segmentation method is shown in Tables 4 and 5. Since the validation is only done on 2D slices presenting complex folding structures, the number of samples is low and consequently leads to low overlap values.

The κ values show that there is insufficient level of reliability for the two manual segmentations. The non-myelinated white matter and gray matter classes have higher number of observations compared to the other classes and therefore dominate the κ measurements. The κ values are low because the segmentations of the brain tissue classes tend to be ambiguous. The overlap measures show that the automatic segmentation method has similar level of variability to the two manual segmentations. The overlap values for csf for the automatic method are generally lower due to misclassifications in the partial volume regions.

Table 1
The volumes of the segmented structures for the four subjects

Subject	ICV	Myelinated WM	Non-myelinated WM	Gray matter	CSF
0096	504,724	15,353	157,160	289,133	43,078
0117	527,885	12,678	234,706	250,161	30,340
0118	514,760	11,480	193,307	255,849	54,124
0123	499,775	28,487	170,227	252,056	49,005

These include the intra cranial volume (ICV) and the volumes of the individual structures (myelinated white matter, non-myelinated white matter, gray matter, and cerebrospinal fluid). All volumes are measured in cubic millimeters.

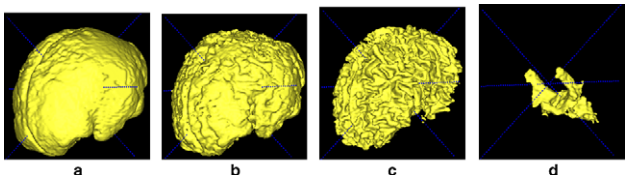

Fig. 8. Surface renderings of the segmented structures of subject 0123. From left to right: (a) intra cranial volume, (b) gray matter, (c) non-myelinated white matter, and (d) myelinated white matter.

Table 2
The κ coefficients that measure the level of agreement between manual raters, first manual rater against the automatic method, and second manual rater against the automatic method

Subject	Rater 1 vs rater 2	Rater 1 vs automatic	Rater 2 vs automatic
0096	0.658	0.604	0.558
0117	0.627	0.577	0.587
0118	0.603	0.561	0.500
0123	0.625	0.626	0.542

Table 3
The Dice similarity values that measure the overlap between the two manual segmentations

Subject	Myelinated white matter	Non-myelinated white matter	Gray matter	CSF
0096	0.715	0.767	0.777	0.738
0117	0.760	0.771	0.741	0.662
0118	0.683	0.738	0.752	0.696
0123	0.787	0.757	0.750	0.639

Table 4
The Dice similarity values that measure the overlap between the segmentation results of the first human rater and the automatic method

Subject	Myelinated white matter	Non-myelinated white matter	Gray matter	CSF
0096	0.634	0.676	0.809	0.681
0117	0.637	0.725	0.776	0.491
0118	0.681	0.661	0.782	0.598
0123	0.777	0.724	0.790	0.569

Table 5
The Dice similarity values that measure the overlap between the segmentation results of the second human rater and the automatic method

Subject	Myelinated white matter	Non-myelinated white matter	Gray matter	CSF
0096	0.774	0.649	0.777	0.598
0117	0.651	0.739	0.760	0.577
0118	0.606	0.630	0.742	0.601
0123	0.719	0.694	0.732	0.478

4. Discussion

We have presented an atlas-based automatic segmentation method for multi-channel newborn brain MRI. The method uses graph clustering and robust estimation to obtain initial distribution estimates from the noisy data. These estimates are then used to generate spatial posterior probabilities for correcting the intensity inhomogeneity inherent in the image. The segmentation is finally refined through the use of non-parametric kernel density estimates.

The use of a probabilistic atlas or a template such as the one used in Warfield et al. (2000) is crucial to overcome the intensity contrast limitations. A probabilistic brain atlas that captures the variability of the large population is essential for the proposed method, such as the one described in Evans et al. (1993) for adult brain MRI. The creation of a true newborn brain atlas requires the segmentations of a large set of representative data. Also, with the high level of brain shape variability in infants, it is likely that a non-linear registration will be required for a reliable atlas formation. These factors make the creation of a newborn brain atlas highly challenging. The current atlas that is built from a small set of data seems sufficient for our data. However, we are working on an improved atlas that shows better population variability.

Visual inspection of the results shows that the major structures are segmented consistently. The segmentations of regions largely affected by partial voluming is still insufficient and is an inherent problem with voxel-based classification. Illustrations of segmentations of the four cases demonstrate that the new method can cope with variable brain shapes. Also, location and shape of the early myelination structures across the subjects seem quite similar. The new segmentation technique is currently applied to the whole database of over 50 neonates (age range is 42.7 ± 1.8 weeks of gestational age) to study volume and structure of brain tissue at this early age. The reproducibility of the results is optimal since the method is fully automatic.

Due to the lack of a gold standard, we have performed only a limited validation of our results. The κ coefficient values and the volume overlap measures show that our segmentation results have similar level of variability to the inter-rater variability for manual segmentations. A significant challenge in newborn brain MRI segmentation is the construction of a gold standard for validation. To our knowledge, there is no standard dataset available to the community to measure and compare the performance of segmentation methods for neonatal MRI. This problem is solved for adult brain MRI by using web-based archives with simulated datasets (Cocosco et al., 1997; Collins et al., 1998) and a large collection of manually segmented real datasets (MGH, 2004).

Acknowledgments

We acknowledge Koen van Leemput for providing the MATLAB code that helped with the development of the bias correction software and the Insight Toolkit community (Insight Consortium, 2004) for providing the software framework for the segmentation algorithm. We wish to thank the anonymous reviewers for their constructive suggestions which benefited the revision of this paper.

References

Boardman, J.P., Bhatia, K., Counsell, S., Allsop, J., Kapellou, O., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2003. An evaluation of deformation-based morphometry in the developing human brain and detection of volume changes associated with preterm birth. In: Proceedings of the MICCAI.

Cocosco, C.A., Kollokian, V., Kwan, R.K.-S, Evans, A.C., 1997. BrainWeb: online interface to a 3D MRI simulated brain database. NeuroImage 5 (4).

Cocosco, C.A., Zijdenbos, A.P., Evans, A.C., 2003. A fully automatic and robust brain MRI tissue classification method. Medical Image Analysis 7, 513–527.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurements 20, 37–46.

Collins, D.L., Zijdenbos, A.P., Baaré, W.F.C., Evans, A.C., 1999. ANIMAL+INSECT: Improved cortical structure segmentation. In: Kuba, A., Sámal, M., Todd-Pokropek, A. (Eds.), Information Processing in Medical Imaging, vol. 1613 of Lecture Notes in Computer Science. Springer, Heidelberg, Germany, pp. 210–223.

Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C., 1998. Design and construction of a realistic digital brain phantom. IEEE Transactions of the Medical Imaging 17 (3), 463–468.

Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C., 2001. Introduction to Algorithms, second ed. MIT Press, Cambridge, MA.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302.

Duda, R.O., Hart, P.E., Stork, D., 2001. Pattern Classification, second ed. Wiley.

Evans, A.C., Collins, D.L., Mills, S.R., Brown, E.D., Kelly, R.L., Peters, T.M., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. In: Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 1813–1817.

Gerig, G., Kübler, O., Kikinis, R., Jolesz, F., 1992. Nonlinear anisotropic filtering of MRI data. IEEE TMI 11, 221–232.

Gilmore, J., Zhai, G., Wilber, K., Smith, J., Lin, W., Gerig, G., 2004. 3T magnetic resonance imaging of the brain in newborns. Psychiatry Research Neuroimaging 132, 81–85.

Girolami, M., He, C., 2003. Probability density estimation from optimally condensed data samples. IEEE Transactions of the Pattern Analysis and Machine Intelligence 25 (10), 1253–1264.

Hastie, T., Tibshirani, R., Friedman, J.H., 2001. The Elements of Statistical Learning. Springer-Verlag.

Hüppi, P., Warfield, S., Kikinis, R., Barnes, P., Zientara, G., Jolesz, F., Tsuji, M., Volpe, J., 1998. Quantitative magnetic resonance imaging of brain development in premature and normal newborns. Annals of Neurology 43, 224–235.

Inder, T.E., Warfield, S.K., Wang, H., Hüppi, P.S., Volpe, J.J., 2005. Abnormal cerebral structure is present at term in premature infants. Pediatrics 115 (2), 286–294.

Insight Consortium, 2004. Insight Toolkit. Available from: <http://www.itk.org>.

Kandel, E.R., Schwartz, J.H., Jessel, T.M., 2000. Principles of Neural Science, fourth ed. McGraw Hill.

Lee, H.C., Cok, D.R., 1991. Detecting boundaries in a vector field. IEEE Transactions of the Signal Processing 39 (5), 1181–1194.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE TMI 16 (2), 187–198.

Matsuzawa, J., Matsui, M., Konishi, T., Noguchi, K., Gur, R., Bilder, W., Miyawaki, T., 2001. Age-related volumetric changes of brain gray and white matter in healthy infants and children. Cerebral Cortex 11 (4), 335–342.

MGH, 2004. Internet Brain Segmentation Repository. Available from: <http://www.cma.mgh.harvard.edu/ibsr>.

Mitra, P., Murthy, C.A., Pal, S.K., 2002. Density based multiscale data condensation. IEEE Transactions of the Pattern Analysis and Machine Intelligence 24 (6).

Rousseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics 41 (3), 212–223.

Rutherford, M. (Ed.), 2002. MRI of the Neonatal Brain. W.B. Saunders.

Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., Williamson, R., 2001. Estimating the support of a high-dimensional distribution. Neural Computation 13, 1443–1471.

Scott, D.W., Sheather, S.J., 1985. Kernel density estimation with binned data. Communications and Statistics – Theory and Methods 14, 1353–1359.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of tissue nonuniformity in MRI data. IEEE Transactions of the Medical Imaging 17 (1), 87–97.

Styner, M., Brechbuhler, C., Szekely, G., Gerig, G., 2000. Parametric estimate of intensity inhomogeneities applied to MRI. IEEE TMI 19 (3), 153–165.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999a. Automated model-based bias field correction of MR images of the brain. IEEE Transactions of the Medical Imaging 18, 885–896.

Van Leemput, K, Maes, F., Vandermeulen, D., Suetens, P., 1999b. Automated model-based tissue classification of MR images of the brain. IEEE Transactions of the Medical Imaging 18, 897–908.

Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R., 2000. Adaptive, template moderated, spatially varying statistical classification. Medical Image Analysis 4 (1), 43–55.

Wells, W.M., Kikinis, R., Grimson, W.E.L., Jolesz, F., 1996. Adaptive segmentation of MRI data. IEEE Transactions of the Medical Imaging 15, 429–442.

Zhai, G., Lin, W., Wilber, K., Gerig, G., Gilmore, J.H., 2003. Comparison of regional white matter diffusion in healthy neonates and adults using a 3T head-only MR scanner. Radiology 229, 673–681.