Automatic tissue and structural segmentation of neonatal brain MRI using Expectation-Maximization

Antonios Makropoulos[1], Christian Ledig[2], Paul Aljabar[3], Ahmed Serag[2], Joseph V. Hajnal[3], A. David Edwards[3], Serena J. Counsell[3], and Daniel Rueckert[2]

[1]Centre for the Developing Brain, Imperial College, London, United Kingdom, [2]Computing, Imperial College, London, United Kingdom and [3]Division of Imaging Sciences and Biomedical Engineering, King's College, London, United Kingdom

Abstract. Accurate automated image segmentation in neonates is challenging due to the lower contrast-to-noise ratio compared to adult scans, the partial volume effect and large anatomical variation. In this paper, we present a technique for brain segmentation into different tissues and structures of interest. Atlas priors and subject-specific tissue priors are used to initialize an Expectation-Maximization (EM) scheme. The proposed implementation incorporates a Markov Random Field (MRF) regularization to account for the spatial dependency of labels, prior relaxation to adapt the priors according to the individual brain appearance and partial volume (PV) correction. The algorithm is evaluated against manually segmented data from the NeoBrainS12 MICCAI challenge.

1 Introduction

Magnetic Resonance (MR) imaging is increasingly being used to assess brain development in neonates. Manual segmentation is extremely time consuming and so an accurate automatic segmentation technique is required.

Segmentation of neonatal MR images is considerably more challenging than the segmentation of adult brain MR images due to the lower contrast-to-noise ratio (CNR), the partial volume (PV) effect as a result of the inverted white matter signal intensity and the large changes in appearance of the brain from the early preterm period to term-equivalent age. A number of studies have been proposed for the segmentation of tissues in the neonatal images [1–5]. Most of the methods use priors in form of templates that are non-rigidly registered to the subject image in order to propagate prior anatomical knowledge to the subject space. The priors are often combined with a model of the image intensities to refine the labels and can be adapted according to the subject image [5, 6].

In this paper, we propose an Expectation-Maximization (EM) framework for the segmentation of neonatal brains into 7 regions: cerebrospinal fluid (CSF), grey matter (GM), unmyelinated white matter (WM), brainstem, cerebellum, basal ganglia and thalami and ventricles. Priors of these structures are propagated with the use of a

probabilistic atlas and are combined with priors obtained with the use of a clustering technique to provide a good initialization of the EM algorithm. Our implementation includes a Markov Random Field (MRF) scheme to penalize the proximity of anatomically distant regions, adaptation of the priors and PV correction similar to previous studies [2, 5].

2 Methods

2.1 Subjects

The methodology described in this paper has been applied to the T2 images of the subjects provided from the NeoBrainS12 MICCAI challenge [7]. Three different sets of images were provided: axial scans acquired at 40 weeks corrected age (set 1) with 2 training and 3 test data, coronal scans acquired at 30 weeks corrected age (set 2) with 2 training and 3 test data, and coronal scans acquired 40 weeks corrected age (set 3) with 3 test data. All of the images were skull stripped using the Brain Extraction Tool (BET) [8] and corrected for field inhomogeneity using the N4 algorithm [9].

2.2 Atlas Priors

In this study we used the spatiotemporal non-rigid atlas developed by [10] to propagate the probabilistic spatial prior of each structure according to age. For each subject the corresponding atlas template according to age was rigidly, affinely and non-rigidly registered to the subject space. The non-rigid registration applied within this study uses normalized mutual information (NMI) as the similarity measure with free-form deformations and control point grid spacings of 20mm, 10mm, 5mm and 2.5mm [11]. The spatial prior of each structure is then transformed into the subject's native space.

Since some structures of the template had different definitions than the delineations provided from the challenge, the spatial priors were manually modified to comply to the protocol of the challenge. The CSF prior was divided into left and right ventricle and the extra-cerebral CSF. Parts of the deep gray matter map were extracted and merged with the cortical gray matter map and the brainstem mask was altered to result in a better agreement to the challenge structure definitions.

2.3 Subject – specific tissue priors

Subject-specific priors that reflect the tissue proportions of the individual images were obtained with the use of k-means clustering [12]. The image intensities of each subject were segmented into four classes that represent the three tissue memberships (CSF, GM, WM) and the background. The intensities that belong to the higher intensity class were subsequently divided into two parts with another k-means scheme. The lower part is mainly attributed to the low CSF intensities and the high WM intensities often found around the term equivalent age in the frontal and occipital lobe. The

higher intensity class corresponds to the 'pure' CSF intensities. A Gaussian distribution is initialized with each of the five resulting k-means classes' centroids and variances.

The normalized likelihoods of the Gaussians represent the tissue prior for the 5 classes (extra-cranial space, GM, WM, high intensity WM/low intensity CSF, CSF). To avoid local minima with the use of k-means, the tissue priors were blurred with a Gaussian kernel.

The atlas priors of each subject propagated from the atlas are combined with the k-means tissue priors to result in refined priors for the structures in the native space of the subject. The combined priors are used to initialize the mixing coefficients of the EM algorithm, providing a better initialization than the atlas priors alone. The resulting classes introduced into the EM model are: CSF, low intensity CSF, GM, high intensity WM, WM, basal ganglia and thalami, brainstem, cerebellum, left ventricle, right ventricle and outside space.

2.4 Expectation-Maximization formulation

In this study we adopt a methodology similar to [2, 5, 13] for the segmentation of the neonatal images. The image is approximated with a Gaussian mixture of the K = 11 structures described in the previous section.

The label $k \in 1,..,K$ of each voxel $i \in 1,2,..,N$ is represented with the variable $z_i = e_k$, where e_k is a unit vector with the k-th component equal to 1. The prior distribution of the z_i's, i.e. $P(z_i = e_k) = \pi_{ik}$, is provided by the subject-specific structure priors π_{ik} for each of the structures, as described in the previous section.

Assuming that the observed intensity y_i of the voxel *i* is independent from the rest of the voxels in the image, the segmentation problem can be formalised as the Maximum a Posteriori (MAP) estimation of the means μ_j and standard deviations σ_j of the Gaussians of the K structures, $\Phi_y = \{\mu_1, \mu_2, .., \mu_K, \sigma_1, \sigma_2, .., \sigma_K\}$ [13]. The parameters $\hat{\Phi}_y$ are estimated with the EM algorithm, at each iteration m as:

Expectation step:

$$p_{ik}^{(m+1)} = P\left(z_i = e_k \mid y_i, \Phi_y^{(m)}\right) = \frac{P\left(y_i \mid z_i = e_k, \Phi_y^{(m)}\right) P\left(z_i = e_k\right)}{\sum_{j=1}^{K} P\left(y_i \mid z_i = e_j, \Phi_y^{(m)}\right) P\left(z_i = e_j\right)} \quad (1)$$

where $P\left(y_i \mid z_i = e_k, \Phi_y^{(m)}\right) = G\left(y_i \mid \mu_k, \sigma_k\right)$ the likelihood of the Gaussian distribution G with parameters μ_k, σ_k .

Maximization step:

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^{N} p_{ik}^{(m+1)} y_i}{\sum_{i=1}^{N} p_{ik}^{(m+1)}}$$
$$\left(\sigma_k^{(m+1)}\right)^2 = \frac{\sum_{i=1}^{N} p_{ik}^{(m+1)} \left(y_i - \mu_k^{(m+1)}\right)^2}{\sum_{i=1}^{N} p_{ik}^{(m+1)}} \quad (2)$$

However, since each voxel's intensity is dependent on the surrounding voxels, we include a spatial regularization term that enforces a smooth labelling among neighbouring voxels [13]. The regularization term is implemented with a MRF that is a non-binary extension of a multiclass Potts model as defined in [5],

$$P(z_i = e_k) = P\left(z_i = e_k \mid p_{N_i}, \Phi_z, \pi_{ik}\right) = \frac{\pi_{ik} e^{-U_{MRF}\left(e_k, \mid p_{N_i}, \Phi_z\right)}}{\sum_{j=1}^{K} \pi_{ij} e^{-U_{MRF}\left(e_j, \mid p_{N_i}, \Phi_z\right)}} \quad (3)$$

where N_i are the first-order neighbours of voxel i. U_{MRF} is the energy function :

$$U_{MRF}\left(e_k, \mid p_{N_i}, \Phi_z\right) = \sum_{j=1}^{k} G_{kj} \left(\sum_{l \in N_i^x} s_x p_{lj} + \sum_{l \in N_i^y} s_y p_{lj} + \sum_{l \in N_i^z} s_z p_{lj} \right) \quad (4)$$

Here $s = \{s_x, s_y, s_z\}$, accounts for the anisotropic spacing in each direction of the image. The connectivity strength matrix G_{kj} is defined a priori according to the anatomical proximity of the structures.

Due to anatomical variability, we assume that the proportions π_{ik} are not known a priori. Instead, we consider π_{ik} as a sample drawn from a distribution derived from the statistical atlas, i.e., we consider them as a posterior of a Dirichlet distribution.

Here, π_{ik} is updated by

$$\pi_{ik} = (1-\alpha)\pi_{ik} + \alpha(G_k * p_{ik}) \quad (5)$$

where G_k is a Gaussian kernel. The amount of relaxation is dependent on the parameter α, $0 \le \alpha \le 1$, controlling the amount of adaptation of the priors. In our experiments the relaxation factor is set to α=0.5.

The penalty introduced by the MRF allows a smooth labeling by removing isolated voxels. However, in regions that misclassified voxels are neighboring each other, the MRF energy will not be sufficient to remove them since they will favor each other through the MRF term. To account for this problem we adopted a knowledge-based driven approach based on morphological operations similar to [2]. The priors of the detected misclassified structure m, and the appropriate structures c_j are adapted as

$$\pi_{ic_j} = \pi_{ic_j} + w_{ic_j}(1-\lambda)\pi_{im} \quad (6)$$

$$\pi_{im} = \lambda\pi_{im} \quad (7)$$

where $w_{ic_j} = \dfrac{\pi_{ic_j}}{\sum_n \pi_{ic_n}}$ the appropriate structures' adaptation weight according to the prior probability. λ is set to 0.5 in all the experiments. Additionally, a separate PV class as described in [5] is defined between GM and WM for the early neonatal scans (30 weeks). The PV class is merged after the EM convergence to the wm class in order to improve the boundary of the resulting WM. The low intensity CSF class is appended to the 'pure' CSF class and the high intensity WM class to the 'pure' WM class. Voxels that belong to small, connected components of the segmented brainstem and cerebellum are reclassified according to the maximum posterior probability of the rest of the classes and holes inside the ventricles are filled in.

3 Results

The performance of the proposed algorithm was assessed on the test data provided by the challenge. Since the myelinated wm class was not segmented with the proposed technique, the results are presented for the rest of the structures. The average Dice coefficients and 95th-percentile Hausdorff distances with respect to the manual segmentation for the different structures are presented in Table 1 and 2 respectively.

	Set 1	**Set 2**	**Set 3**	**Over all sets**
gm	0.837	0.700	0.747	0.761
dgm	0.907	0.820	0.863	0.863
wm	0.89	0.907	0.857	0.885
brainstem	0.837	0.747	0.730	0.771

cerebellum	0.913	0.873	0.920	0.902
ventricles	0.797	0.860	0.823	0.827
csf	0.76	0.837	0.753	0.783
average	**0.849**	**0.821**	**0.813**	**0.827**

Table 1. Average Dice coefficients of the segmented structures

	Set 1	**Set 2**	**Set 3**	**Over all sets**
gm	0.73	1.00	0.86	0.86
dgm	0.80	2.01	1.15	1.32
wm	0.63	0.50	0.95	0.69
brainstem	1.04	2.40	1.65	1.70
cerebellum	0.70	0.81	0.52	0.68
ventricles	11.74	1.43	12.22	8.46
csf	1.68	1.32	2.24	1.75
average	**2.47**	**1.35**	**2.80**	**2.21**

Table 2. Average 95^{th}-percentile Hausdorff distance of the segmented structures

The average runtime of the proposed automatic segmentation technique is 95 minutes per subject. The segmentations of all the subjects were run in parallel on a server with 24 cores and 64 GB RAM.

4 Conclusion

We have presented an algorithm for the segmentation of neonatal T2 images into 7 structures: CSF, GM, unmyelinated WM, brainstem, cerebellum, basal ganglia and thalami and ventricles. Priors of each structure are propagated from a spatiotemporal probabilistic atlas and are combined with subject-specific tissue priors obtained with k-means. The resulting structures' priors are used to initialize an EM optimization. Our implementation includes a MRF term, PV correction and relaxation of the labels to refine the structures. The quality of the algorithm is assessed according to the manually segmented neonatal brains provided by the MICCAI challenge.

References

1. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic segmentation of MR images of the developing newborn brain. Med Image Anal. 9, 457–466 (2005).
2. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., Edwards, A.D., Rueckert, D., Hajnal, J.V.: Automatic segmentation and reconstruction of the cortex from neonatal MRI. Neuroimage. 38, 461–477 (2007).

3. Song, Z., Awate, S.P., Licht, D.J., Gee, J.C.: Clinical neonatal brain MRI segmentation using adaptive nonparametric data models and intensity-based Markov priors. Med Image Comput Comput Assist Interv. 10, 883–890 (2007).
4. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain MRI. Neuroimage. 47, 564–572 (2009).
5. Cardoso, M.J., Melbourne, A., Kendall, G.S., Modat, M., Hagmann, C.F., Robertson, N.J., Marlow, N., Ourselin, S.: Adaptive neonate brain segmentation. Med Image Comput Comput Assist Interv. 14, 378–386 (2011).
6. Shiee, N., Bazin, P.-L., Cuzzocreo, J.L., Blitz, A., Pham, D.L.: Segmentation of brain images using adaptive atlases with application to ventriculomegaly. Inf Process Med Imaging. 22, 1–12 (2011).
7. http://neobrains12.isi.uu.nl.
8. Smith, S.M.: Fast robust automated brain extraction. Hum Brain Mapp. 17, 143–155 (2002).
9. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 29, 1310–1320 (2010).
10. Serag, A., Aljabar, P., Ball, G., Counsell, S.J., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. Neuroimage. 59, 2255–2265 (2012).
11. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imaging. 18, 712–721 (1999).
12. Macqueen, J.: Some methods of classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. pp. 281–297 (1967).
13. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Trans Med Imaging. 18, 897–908 (1999).