Prior Knowledge Driven Multiscale Segmentation of Brain MRI

Ayelet Akselrod-Ballin[1], Meirav Galun[1], John Moshe Gomori[2], Achi Brandt[1], and Ronen Basri[1,⋆]

[1,⋆⋆]Dept. of Computer Science and Applied Math, Weizmann Institute of Science, Rehovot

[2]Dept. of Radiology, Hadassah University Hospital, Jerusalem, Israel

Abstract. We present a novel automatic multiscale algorithm applied to segmentation of anatomical structures in brain MRI. The algorithm which is derived from algebraic multigrid, uses a graph representation of the image and performs a coarsening process that produces a full hierarchy of segments. Our main contribution is the incorporation of prior knowledge information into the multiscale framework through a Bayesian formulation. The probabilistic information is based on an atlas prior and on a likelihood function estimated from a manually labeled training set. The significance of our new approach is that the constructed pyramid, reflects the prior knowledge formulated. This leads to an accurate and efficient methodology for detection of various anatomical structures simultaneously. Quantitative validation results on gold standard MRI show the benefit of our approach.

1 Introduction

Segmentation of anatomical structures in brain magnetic resonance images (MRI) is crucial for medical image analysis. It includes a wide range of applications such as therapy evaluation, image guided surgery and neuroimaging studies [1,2,3]. The challenge in brain MRI segmentation is due to issues such as noise, intensity non-uniformity (INU), partial volume effect, shape complexity and natural tissue intensity variations. Under such conditions, incorporation of a priori medical knowledge, commonly represented in anatomical brain atlases by state-of-the-art studies [1,2,3,4] is essential for robust and accurate automatic segmentation.

Automatic segmentation of brain structures in MRI has been extensively studied in scientific literature (see [2,3]). A popular approach is to utilize deformable models in a variational formulation. In [5] the templates were initialized by nonlinear registration of an MRI atlas to the input and then modified to minimize an energy based on expected textural and shape properties. Alternatively,

⋆ Research was supported in part by the Israel Institute of Technology. Research was supported in part by the Binational Science foundation, Grant No. 2002/254.

⋆⋆ Research was supported in part by the European Commission Project IST-2002-506766 Aim Shape. Research was conducted at the Moross Laboratory for Vision and Motor Control at the Weizmann Institute of Science.

N. Ayache, S. Ourselin, A. Maeder (Eds.): MICCAI 2007, Part II, LNCS 4792, pp. 118–126, 2007.
© Springer-Verlag Berlin Heidelberg 2007

[6] performed segmentation of several anatomical structures using a level set formulation. Numerous classification approaches have been proposed, including supervised techniques such as artificial neural networks [1], k-nearest neighbors (kNN) and unsupervised clustering techniques such as k-means and fuzzy c-means. An adaptive fuzzy c-means (AFCM) algorithm was developed in [7] using a multigrid algorithm. Statistical-based methods have been widely used. These approaches typically model the intensity distribution of brain tissues by a Gaussian mixture model and classify voxels according to the intensity distribution of the data. Given the distribution, the optimal segmentation can be estimated by a maximum a posteriori (MAP) or a maximum likelihood (ML) formulation. The expectation maximization (EM) is a popular algorithm for solving the estimation problem. It was pioneered by [8] to simultaneously perform brain segmentation and estimate the INU correction and was further extended by many others to incorporate spatial considerations. The authors in [9] applied the EM approach to a hierarchical segmentation model. Their idea of combining a hierarchical graph pyramid with atlas information in a statistical model, is closely related to our work. A Bayesian approach that uses manually labeled data for brain segmentation was presented by [10]. The core idea, of using a manual training set for incorporation of prior statistics and class conditional densities resembles our approach. Yet, instead of using Markov random fields (MRFs) our novel multiscale approach allows us to capture both the local and global geometric relations between the structures. A recent multiscale segmentation and classification approach (ISCA) [11] integrated segmentation with a second classification stage and applied it to tissue classification. Our novel **probabilistic multiscale** segmentation avoids the second stage by incorporating prior information into the framework. The work reported here differs from past work in several aspect: First, to the best of our knowledge this is the first application of a **multiscale algorithm** derived from a fast multi-level solver called algebraic multi-grid (AMG) for segmentation of deep brain structures. The benefits of using a multiscale approach include, the reduction in time computation, the development of a framework that can adopt a rich set of multiscale measurements and the hierarchical representation which as noted by [9] is indeed a powerful and flexible representation useful for many applications. Second, the incorporation of prior information, represented by a **probabilistic atlas** and a likelihood function (based on a manually labeled training set), into the multiscale segmentation using **Bayes formulation**. Finally, in contrast with other approaches that are typically tuned to a particular set of structures or tasks, we propose a general framework for segmenting the three main **brain tissue classes and their substructures** simultaneously, by using the same parameter values for all structures. Our approach can be generalized to other tasks and modalities, because any probabilistic atlas and registration scheme can be used and due to the ability to define a likelihood function based on any training set. The restriction is that the probability information of the training data (atlas and likelihood function) represent the specific population.

The remainder of the paper is organized as follows. In section 2 our probabilistic multiscale approach is introduced. Section 3 presents the experiments

and compares our results to other approaches on a gold standard data base. Section 4 summarizes with conclusions.

2 Methods

Our aim is to perform segmentation of anatomical structures, by incorporating prior anatomical information into a multiscale segmentation framework, through a Bayesian formulation. We utilize a probabilistic atlas where each region of image space is assigned a prior probability of belonging to a variety of anatomical structures. This prior probability is aggregated and used throughout the segmentation process. The likelihood in the model is computed based on an automatic learning process derived from labeled training data. Consequently, the posterior probability (the chance) that two neighboring aggregates reside in the same segment is estimated by integrating the Bayesian formulation into the multiscale segmentation algorithm. This section describes the segmentation framework together with the probability-based criteria computed for the aggregates.

2.1 Segmentation Methodology

Our method is based on the Segmentation by Weighted Aggregation (SWA) [12], which was extended to handle 3D multi-channel and anisotropic data [11]. The algorithm uses a graph representation of the images and constructs a "pyramid", i.e., a sequence of progressively smaller ("coarser") graphs ("levels" or "scales"), which adaptively represents progressively larger aggregates of voxels of similar properties. The nodes and the couplings (edge values) of the initial graph are the voxels of the given images and similarity measures between neighboring voxels, respectively. The algorithm recursively coarsens the graph, level after level, by softly aggregating several similar nodes of a finer level into a single node of the next coarser level. The couplings of the coarser graph are based on tunable statistical measures, called **aggregative features** which are scale dependent properties computed along with the segmentation process. Features obtained for small aggregates at a fine level affect the aggregation formation of larger aggregates at coarser levels, according to features similarity. This work employs as aggregative features both the **average intensity** of voxels at an aggregate i, denoted by $\bar{I}(i)$ and the **average atlas prior probabilities** of an atlas structure at an aggregate i, denoted for example by $\overline{P}(i \in WM)$ for finding the white matter (WM) in an aggregate i. The scheme provides a recursive mechanism for calculating the aggregative features (see [12]).

2.2 Incorporating the Probabilistic Model into the Segmentation

Let $L = \{l_1, \ldots, l_\nu, \ldots, l_K\}$ be a collection of K anatomical structures. In our experiment we use twelve structure classes ($K = 12$) referring to white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), Caudate(CN), Putamen(Pu), Thalamus(Th), Pallidum(GP), Brainstem(Bs), Ventral Diencephalon

(VDC), Hippocampus(H), Amygdala(Am) and "Other". The **prior probability** $\overline{P}(i \in l_\nu)^{[s=0]}$ of a voxel i at the finest graph level ($s = 0$) is defined by the spatial distribution of the probability atlas aligned with the test data set where the atlas construction is performed based on the training set (see Sec. 3.1). At coarser levels of the segmentation pyramid the prior probability $\overline{P}(i \in l_\nu)^{[s]}$ of an aggregate i is accumulated as an aggregative feature, so that at level s it is modelled by the average prior probabilities of its sub-aggregates obtained at level $s-1$. The **posterior probability** of a structure being present at an aggregate i can be obtained using Bayes formula as follows (Z is a normalization factor):

$$\overline{P}(i \in l_\nu | \bar{I}(i))^{[s]} = \frac{1}{Z}[\overline{P}(i \in l_\nu)^{[s]} \overline{P}(\bar{I}(i) | i \in l_\nu)^{[s]}], \tag{1}$$

The **likelihood** $\overline{P}(\bar{I}(i)|i \in l_\nu)^{[s]}$ represented by the second term in the right hand side of Eq. 1 reflects the conditional intensity probability for an aggregate i given that the structure class is l_ν. It is computed for each structure l_ν, based on the voxel intensity and the ground-truth structure, as determined by the manual labeling of the training data. The histograms for each structure, are then averaged over all training subjects and used as the likelihood for the test set.We decided to model the voxel intensity probability with a nonparametric distribution, since the real distribution can differ from the commonly used Gaussian distribution model. In addition, the ability to define a likelihood function based on the training set allows generalization to other tasks and clinical population.The **average intensity** of voxels in aggregate i, is accumulated as an aggregative feature. Thus, at every level of the pyramid, we find the histogram bin corresponding to $\bar{I}(i)$ and determine the likelihood value for each aggregate and structure accordingly.

The role of the posterior probabilities (Eq. 1) in the segmentation is to determine if two neighboring aggregates at level $(s-1)$ reside in the same aggregate at the level s. The algorithms goal is to accurately segment a set of anatomical structures which consist of both clearly and weakly defined boundaries. Subcortical structures (e.g., the thalamus) are commonly defined by weakly visible boundaries, since their intensity pattern is often similar to their neighbors. In such cases, the atlas prior is critical. However, in the case of the main tissue classes, the borders are more visible and the smooth atlas prior can impede the delineation accuracy. The multiscale framework allows us to adjust the influence of the aggregative features on the coarse graph couplings across scale during the pyramid construction. Once reaching an intermediate scale the aggregates have gathered sufficient statistics, therefore the probability criteria can fully control the segmentation process. Yet, at finer scales ($s \leq 4$), we have experienced that for the three main tissues, the probability criteria needs to be regulated by intensity. Accordingly, the coupling weights of aggregates i, j are modified based on their aggregative feature similarity denoted by $\exp(-\zeta a_{ij})$ and defined as follows:

$$a_{ij} = \begin{cases} \lambda(\Delta\overline{P}_{ij})^{\frac{1}{2}} + (1-\lambda)|\bar{I}(i) - \bar{I}(j)| & \text{if } s \leq 4 \\ (\Delta\overline{P}_{ij})^{\frac{1}{2}} & \text{otherwise} \end{cases}, \tag{2}$$

where $\lambda = (\Delta\overline{P}_{ij})^{\frac{1}{\eta}}$ ($\eta = 10$,$\zeta = 7$ in our implementation), and

$$\Delta\overline{P}_{ij} = \sum_{\nu=4}^{K-1} (\overline{P}(i \in l_\nu | \bar{I}(i)) - \overline{P}(j \in l_\nu | \bar{I}(j)))^2. \quad (3)$$

The expression for the finer scales ($s \leq 4$) combines an intensity term and posterior probability term based on the internal structures. λ is derived from the posterior probability itself and controls the relative weight of the two terms. When the posterior difference between two aggregates is high, then the first term controls the expression, otherwise the average intensity controls the process.

The algorithm's **solution in terms of voxels** is computed as follows. First, voxel occupancy of the aggregates is determined by projection of all aggregates onto the data voxels using the interpolation matrix (see [12]). A voxel is associated with the aggregate for which it has the maximal interpolation weight. Then, each aggregate in the coarsest scale of the pyramid is matched to one structure based on its maximal a-posteriori probability.

Computational cost is linear in the size of the data. Our implementation on a standard Xeon 1.7GHz PC takes 5 min for the segmentation of all 12 structures on a $150 \times 150 \times 60$ region of interest. This does not include the atlas construction which takes about 8min using 5 brains and likelihood histogram computation which is done in advance.The efficiency of our approach is superior to previously reported results, for instance [13] requires 5min for segmentation of the caudate structure on a pentium 4 2Ghz and [5] takes 6min for four structures on a pentium 3, 1GHz (the training phase took about 20 hours).

3 Experiments

The methods performance was assessed on a gold standard database. The MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/. The data set contains 18 real T1-weighted normal MR brain scans and the manual segmentation of 43 structures, performed by a trained expert. The selection of the 12 structures that commonly appear in literature, was motivated by memory considerations. The MR scans are $256 \times 256 \times 128$ volumes acquired with 1.5_{mm} coronal slice thickness resolution and pixel dimension going from 0.84_{mm} to 1_{mm} on each slice. The approach was tested on the same $150 \times 150 \times 60$ region of interest in all data sets, which contains the entire volume of all internal structures. Since we are not interested in the skull or the scalp, the images were skull stripped using the automatic BET procedure [14], and present validation results for the structures.

3.1 Atlas Construction

Similarly to other MRI brain segmentation methods [1,2,3,4] we employ probabilistic anatomical atlases to determine the prior information. The atlas is constructed based on affine co-registrations of the manually labeled training set to the test subject, implemented using the publicly available AIR5.0 [15] registration algorithm with 12-parameter affine transformation. First, the training sets

are aligned with the test set and then the atlas is created by voxel-wise averaging of the neuroanatomical structures over the manually labeled training data sets. The atlas prior is formed by the frequency that each structure occurred at a voxel across the training sets. Thus, the atlas represents the prior probability of each voxel in the test set to belong to a particular structure.

3.2 Results

We applied the algorithm to the data set and performed a "leave one out" learning strategy, leading to eighteen separate experiments. In each experiment one subject is removed from the training set and considered as the test set. The probabilistic atlas is constructed from five data sets randomly selected from training sets with $K = 12$ labels. Table 1 presents our segmentation results on all structures in comparison to other approaches. Only the four first rows in the table report on the same data set used here. The first and second rows present our novel approach with a baseline comparison based on the atlas priors. The third row presents the ISCA [11] applied on the same data set, whereas the lower rows are based on published results. Denote by S, the set of voxels automatically segmented as a specific structure, and by R the set of voxels labeled as the same structure in the 'ground truth' reference. The similarity between S and R is evaluated using the following validation measures:

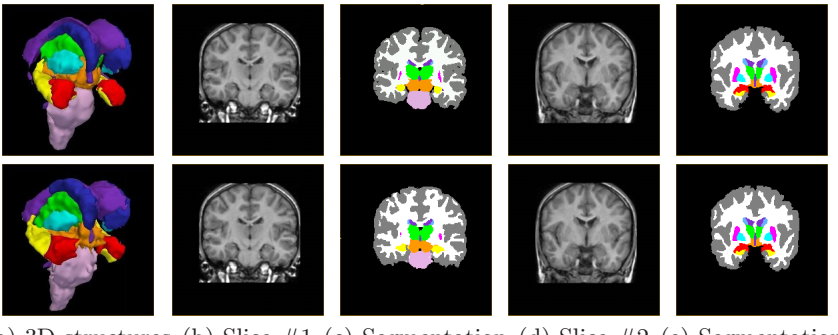
- **Dice similarity statistics** κ**:** $2|S \cap R|/(|S| + |R|)$
- **Hausdorff distance** H_f**:** The maximum of the minimum Eucledian distance between S and its closest voxel on R. $H(S,R) = max_{v \in S}(min_{w \in R} d_{Euclidean}(v,w))$ The distance is symmetrized by taking the maximum of both symmetric measures. Due to outliers H_{95} measures the $f = 95\%$ percentile of the Hausdorff distance.
- **Mean distance** M**:** the mean distance between the boundarys of S and R.

Quantitative analysis shows that the algorithm can be considered as a promising platform for segmentation of anatomical structures. Figure 1 demonstrates our results comparing the manual and automatic segmentation in 2D and 3D views. It is important to note that the same parameter values were used on all structures. Our results compare favorably with results reported *on the same gold standard data set*. They are better than [6] on all structures and validation measures, except for the M measure on Pu. The significant difference ($p \leq 0.05$) between [11] and our novel approach on all structures except Pu, shows that the latter which is faster and less complex, is also more accurate. A significant difference to the Naive approach was found on 8 of the 11 structures. The results reported in [10] are better. However, the analysis in Fischl et al. is based on 7 subjects whereas our work presents results on 18 subjects of a gold-standard database. Comparison to other approaches *applied on different data sets*, shows that our H_{95} results for the Hippocampus and caudate(CN) are inferior compared to [5], but our M is better, indicating that the error is due to a small percent of outliers. The results for CN, in terms of κ are lower than [9,10,16], yet either M or H_{95} are better than [5,6,13]. This may be explained by the

Table 1. Segmentation scores for brain structures by various algorithms

Method:	CN	Pu	Th	GP	H	Am	Bs	VDC	WM	GM	CSF
Multiscale Bayesian:-κ:	0.80	0.79	0.84	0.74	0.69	0.63	0.84	0.76	0.87	0.86	0.83
-M:	1.44	1.6	1.44	1.43	1.88	1.67	1.62	1.43	-	-	-
-H_{95}:	3.07	3.36	2.9	2.75	4.57	3.38	3.42	2.84	-	-	-
Naive Prior:-κ:	0.65	0.77	0.83	0.72	0.62	0.65	0.81	0.77	0.69	0.72	0.53
ISCA [11] **-κ:**	0.74	0.78	0.80	0.70	0.64	0.58	0.62	0.66	0.84	0.84	0.79
-M:	1.84	1.72	1.59	1.55	1.91	1.78	2.18	1.7	-	-	-
-H_{95}:	4.46	3.89	3.41	3.2	4.44	3.89	5.26	3.75	-	-	-
C. Ciofolo et al. [6] -κ:	0.65	0.70	0.77	0.62	-	-	-	-	-	-	-
-M:	1.71	1.46	1.70	1.51	-	-	-	-	-	-	-
B. Fischl et al. [10]-κ :	0.88	0.86	0.87	0.78	0.80	0.67	0.89	-	-	-	-
K. Pohl et al. [9]-κ :	0.866	-	0.894	-	-	-	-	-	0.87	0.9	0.7
A. Pitiot et al. [5] -M:	1.6	-	-	-	2.1	-	-	-	-	-	-
-H_{95} :	2	-	-	-	3	-	-	-	-	-	-
B.M. Dawant et al. [16]-κ :	0.86	-	-	-	-	-	-	-	-	-	-
D. Nain et al. [13]-H_{95} :	3.16	-	-	-	-	-	-	-	-	-	-

small volume of the CN, since in such volumes, small differences in placement of boundaries between S and R can have a large effect on κ. In sum, comparing the results obtained on several deep cortical structures to other approaches, we conclude that the results are not as high as the results reported by [9,10], but comparable or superior to results reported by other algorithms. We believe that the results are lower than the results reported by [9,10], which are the only study reporting results on both the tissues and their substructures, mainly due to the need to include additional features in the framework.



(a) 3D structures (b) Slice #1 (c) Segmentation (d) Slice #2 (e) Segmentation

Fig. 1. Manual and automatic segmentation (upper, lower row). Presented in a 3D view (a) and super imposed on two coronal 2D slices (c,e) corresponding to their original input (b,d) respectively.

4 Discussion

An automatic multiscale probabilistic method for segmentation of anatomical structures in MRI is introduced. The inclusion of prior information is a critical feature of the algorithm. Additional imaging contrasts protocols, and multiscale features can be easily incorporated into the framework. Future work will extend this work by enforcing shape, texture, spatial and neighborhood relations between the structures. Such constraint can be modelled implicitly and explicitly during the pyramid construction [9,5,10]. Also we plan to further improve our approach by using more sophisticated registration techniques. In sum, the method's strength is demonstrated on gold standard real MRI, by performing accurate and efficient segmentation of many structures simultaneously, including both subcortical structures and brain tissues.

References

1. Zijdenbos, A., Forghani, R., Evans, A.: Automatic pipeline analysis of 3D MRI data for clinical trials: application to MS. IEEE TMI 21(10), 1280–1291 (2002)
2. Pham, D., Xu, C., Prince, J.: Current methods in medical image segmentation. Annual Review of Biomedical Engineering 2, 315–337 (2000)
3. Sonka, M.M., Fitzpatrick, J.M. (eds.): Handbook of Medical Imaging. SPIE (2000)
4. Van-Leemput, K.: Probabilistic brain atlas encoding using bayesian inference. MICCAI 1, 704–711 (2006)
5. Pitiot, A., Delingette, H., Thompson, P.M., Ayache, N.: Expert knowledge guided segmentation system for brain MRI. NeuroImage 23(1), S85–S96 (2004)
6. Ciofolo, C., Barillot, C.: Brain segmentation with competitive level sets and fuzzy control. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 333–344. Springer, Heidelberg (2005)
7. Pham, D., Prince, J.: Adaptive fuzzy segmentation of magnetic resonance images. IEEE TMI 18, 737–752 (1999)
8. Wells, W.M., Grimson, W., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. IEEE TMI 15, 429–442 (1996)
9. Pohl, K., Bouix, S., Kikinis, R., Grimson, W.: Anatomical guided segmentation with non-stationary tissue class distributions in an expectation-maximization framework. IBSI, 564–572 (2004)
10. Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S.: Whole Brain SegmentationAutomated Labeling of Neuroanatomical Structures in the Human Brain. Neuron 33(3), 341–355 (2002)
11. Akselrod-Ballin, A., Galun, M., Gomori, J.M., Basri, R., Brandt, A.: Atlas guided identification of brain structures by combining 3D segmentation and SVM classification. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, Springer, Heidelberg (2006)
12. Galun, M., Sharon, E., Basri, R., Brandt, A.: Texture segmentation by multiscale aggregation of filter responses and shape elements. In: ICCV, pp. 716–723 (2003)
13. Nain, D., Haker, S., Bobick, A., Tannenbaum, A.: Shape-driven 3D segmentation using spherical wavelets. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, Springer, Heidelberg (2006)

14. Smith, S.: Fast robust automated brain extraction. Human Brain Mapping 17(3), 143–155 (2002)
15. Woods, R., Grafton, S., Holmes, C., Cherry, S., Mazziotta, J.: Automated image registration: I. general methods and intrasubject, intramodality validation. Journal of Computer Assisted Tomography 22, 139–152 (1998)
16. Dawant, B.M., Hartmann, S.L., Thirion, J., Maes, F., Vandermeulen, D., Demaerel, P.: Automatic 3D segmentation of internal structures of the head in MRI using a combination of similarity and free-form transformations. IEEE TMI 18 (1999)