Towards A Parts-Based Approach to Sub-Cortical Brain Structure Parsing

Digvijay Gagneja[a,b], Caiming Xiong[a] and Jason J. Corso[a]

[a]Computer Science and Engineering, SUNY at Buffalo, USA
[b]Electrical Engineering, Indian Institute of Technology, Kharagpur, India

ABSTRACT

The automatic localization and segmentation, or parsing, of neuroanatomical brain structures is a key step in many neuroscience tasks. However, the inherent variability in these brain structures and their appearance continues to challenge medical image processing methods. The state of the art primarily relies upon local voxel-based morphometry, Markov random field, and probabilistic atlas based approaches, which limits the ability to explicitly capture the parts-based structure inherent in the brain. We propose a method that defines a principled parts-based representation of the sub-cortical brain structures. Our method is based on the pictorial structures model and jointly models the appearance of each part as well as the layout of the parts as a whole. Inference is cast as a maximum a posteriori problem and solved in a steepest-descent manner. Experimental results on a 28-case data set demonstrate high accuracy of our method and substantiate our claim that there is significant promise in a parts-based approach to modeling medical imaging structures.

Keywords: neuroimaging, sub-cortical structures, segmentation, pictorial structures, parts-based representation

1. INTRODUCTION

Parsing neuroanatomical brain structures, such as the caudate, putamen and hippocampus that we target in this paper, is a core problem in neuroimaging. Accurate parsing—detection, localization, and segmentation—facilitates rich study into the brain's structure and function. Although the human brain has a predictable structure, e.g., most brains have symmetric left-right lobe sub-cortical structures, there is significant variation in the shape and local appearance of the various brain parts, which has challenged automatic parsing methods. Automatic brain parsing hence remains an active area of study.[1–5]

Most of the past work in automatic brain structure parsing operates under the assumption that local modeling is sufficient. A seminal work along these lines is Fischl et al.[4] who model the appearance of a collection of neuroanatomical parts as a mixture of Gaussians with one component linked to each part. The authors note the significant overlap in T1 appearance between various parts. More recently, researchers such as Tu et al.[5] have begun incorporating local filter responses in a fixed cubic window around each voxel to capture its appearance. They have shown marked improvement over the voxel only methods, but remain local and can blur part boundaries due to a fixed, global scale of the feature response. Random fields have been used[4,5] to enforce some amount of smoothness in the result, but still remain local. Part location and shape are typically modeled by a probabilistic atlas[6] which basically estimates a class-conditional density for the probability that a given voxel is a member of a particular type as learned by a population of *registered* human brains. The difficulties in adapting such priors to various age groups, sexes, or sick brain is quite clear.

In this paper, we introduce parts-based models to the brain neuroanatomical parsing community. Parts-based models, such as the pictorial structures models we study here,[7,8] have the potential to overcome many of the drawbacks of these existing local-only methods. The parts-based models explicitly account for each neuroanotomical part and have a part-independent way of capturing appearance while also establishing a principled distribution over the structural layout of the parts. We apply the pictorial structures model to the problem

---

of localizing six sub-cortical structures in T1 MRI imagery. We are able to localize all parts with a maximum Hausdorff distance of 8 voxels (about 3 millimeters in our data) on a data set of 28 images. This parts-based method for localizing the various sub-cortical structures is a first step towards a parts-based comprehensive brain structure parsing model, which is our ultimate goal.

2. METHODS

We describe the modeling and inferential approach in the next few section, beginning with a high-level view of the parts-based approach to sub-cortical structure localization (Section 2.1) and then proceeding to the specific choices for modeling the appearance of each part.

2.1 The Pictorial Structures Model

A pictorial structure model is described as an undirected graph $G = (V, E)$ where the vertices of the graph $V = \{v_1, \ldots, v_n\}$ represent the n-parts (6 sub-cortical structures in our case) and connected vertices are given by the set of edges E, i.e., for every connected pair v_i and v_j, $(v_i, v_j) \in V$. Conceptually, each node in the model represents a single sub-cortical structure, i.e., a part of the brain, and the overall model specifies a joint distribution on the location and appearance of each part as well as the relative positions of each pair of parts. We now make these ideas more precise.

A specific instance of the model, called a *configuration*, is specified by $L = \{l_1, \ldots, l_n\}$ where each l_i represents the location of vertex v_i. Let θ specify the set of (appearance and geometrical) parameters for a particular pictorial structures model. The probability of a particular configuration L is given by

$$P(L|I, \theta) \propto P(I|L, \theta) P(L|\theta) \ , \tag{1}$$

where we denote the input image as I and we have made use of Bayes Theorem. The first term $P(I|L, \theta)$ is the appearance model, described in Section 2.2.

The second term $P(L|\theta)$ defines the distribution over the spatial layout of the parts in L. Following the standard pictorial structures literature,[7,8] we think of each pair of vertices (v_i, v_j) in the model as being connected by a spring, which allows us to establish an ideal (i.e., prior) relative location of pair. This pairwise relative term is assumed to be Gaussian distributed, defined by distance s_{ij} and covariance Σ_{ij}, corresponding to the stiffness of the spring connected the two parts. The operant assumption here is the spring model, and hence the Gaussian is the appropriate model. Then, the full structure distribution is given by

$$P(L|\theta) = \prod_{(v_i, v_j) \in E} P(l_i, l_j | s_{ij}, \Sigma_{ij}) \tag{2}$$

$$\propto \prod_{(v_i, v_j) \in E} \exp\left[-\frac{1}{2} \left((l_i - l_j) - s_{ij} \right)^\mathsf{T} \Sigma_{ij}^{-1} \left((l_i - l_j) - s_{ij} \right) \right] \tag{3}$$

The actual set of connections specified in the model is defined by a minimum spanning tree-based search through the undirected graph, during the training process.

The ideal configuration is sought as the maximum a posteriori solution $L^* = \mathrm{argmax}_L P(L|I, \theta)$. The MAP formulation allows various statistical inference methods to be used, and we have developed a direct gradient-based search in our work, initialized by the spatial priors of each part, $P(l_i)$, which are estimated from training data.

2.2 Part Appearance Models

The local appearance around each part v_i is described by a vector of 3D Gaussian derivative filters. The filterbank enumerates a selection of orientations centered around the part's current location (voxel) l_i. We use 16 filters sampled at $\frac{\pi}{4}$ radians intervals along both coordinates axes, but not the off-axes. Let $\alpha(l_i; I)$ define the filterbank response at location l_i in image I. The filter bank response is modeled by a Gaussian distribution with parameters

Figure 1. Views of a Gaussian Derivative filter along 3 axes.

(μ_i, Σ_i), for simplicity, but with a diagonal covariance for tractability. Finally, the appearance term in the full model (1) is given by

$$P(I|L,\theta) = \prod_{i=1}^{n} P(I|l_i, \mu_i, \Sigma_i) \propto \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\alpha(l_i, I)^\mathsf{T}\Sigma_i^{-1}\alpha(l_i, I)\right] \; . \tag{4}$$

2.3 Learning Model Parameters

2.3.1 Learning Appearance Parameters

Suppose we have a set of labelled images $(I^1, \ldots, I^n)$ with configurations $(L^1, \ldots, L^n)$. We first compute the filterbank response at each image defined by $(\alpha(L^1, I^1), \ldots, \alpha(L^n, I^n))$. Assume there are m parts in the pictorial structures model. Thus, we can write:

$$\alpha(L^n, I^n) = \alpha((l_1^n, \ldots, l_m^n), I^n),$$

We estimate the filterbank response parameters (μ_i, Σ_i) for each part l_i as described in Section (2.2) as follows:

$$\mu_i = \frac{\sum_{j=1}^{n} \alpha(l_i, I^j)}{n}; \forall i \in [1, m], \tag{5}$$

$$\Sigma_i = \sqrt{\frac{\sum_{j=1}^{n} (\mu_i - \alpha(l_i, I^j))^2}{n}}; \forall i \in [1, m], \tag{6}$$

2.3.2 Learning Pictorial Structures Model

Let s_{ij}^n denote the distance between i and j parts in the image I^n. We estimate the parameters (s_{ij}, Σ_{ij}) as follows:

$$s_{ij} = \frac{\sum_{k=1}^{n} s_{ij}^k}{n}; \forall i, j \in [1, m], \tag{7}$$

$$\Sigma_{ij} = \frac{\sqrt{\sum_{k=1}^{n} \left(s_{ij} - s_{ij}^k\right)^2}}{n}; \forall i, j \in [1, m], \tag{8}$$

The set of connections defining the tree in the undirected graph are determined by a minimum spanning tree-based search where the metric is the energy between two vertices v_i and v_j given by:

$$E_{ij} = -log\left(P(l_i, l_j | s_{ij}, \Sigma_{ij})\right) \tag{9}$$

$$= \frac{1}{2}\left[(l_i - l_j - s_{ij})^\mathsf{T}\Sigma_{ij}^{-1}\left((l_i - l_j) - s_{ij}\right]\right. \tag{10}$$

The MST based search picks up only those set of connections that are more consistent in their relative locations over all example images. Thus, it defines our initial pictorial structures model.

2.4 MAP Estimate

Our goal is to find a configuration $L^* = \{l_1^*, \ldots, l_n^*\}$ of our pictorial model that maximizes the a-Posteriori estimate $P(L|I, \theta)$, i.e. $L^* = \text{argmax}_L P(L|I, \theta)$. We can write the full model (1) as

$$P(L|I, \theta) \propto P(I|L, \theta)P(L|\theta) \propto \prod_{i=1}^{n} P(I|l_i, \mu_i, \Sigma_i) \prod_{(v_i, v_j) \in E} P(l_i, l_j | s_{ij}, \Sigma_{ij}) \tag{11}$$

Let $m_i(l_i) = -\log P(I|l_i, \mu_i, \Sigma_i)$ be the match cost of part v_i when placed at location l_i on the image i.e. lower match cost defines better appearance match. Similarly, let $d_{ij}(l_i, l_j) = -\log P(l_i, l_j | s_{ij}, \Sigma_{ij})$ define the deformation cost when parts v_i and v_j are placed at locations l_i and l_j respectively i.e. higher deviation of the relative locations of these parts from the mean distance s_{ij} (Section 2.1) corresponds to higher deformation cost. Thus, the configuration L^* can be defined as

$$L^* = \underset{L}{\text{argmax}}\, P(L|I, \theta) = \underset{L}{\text{argmin}} \left(\sum_{i=1}^{n} m_i(l_i) + \sum_{v_i, v_j \in V} d_{ij}(l_i, l_j) \right) \tag{12}$$

We perform this task using a recursive algorithm. Given tree $G = (V, E)$, let the root vertex $v_r \in V$ s.t. v_r has maximum children (in order to reduce computational complexity). For any vertex v_i of depth d_i in the tree G, let C_i define the children of v_i, i.e. all vertices (if any) of depth $(d_i + 1)$. For any vertex v_j with no children in the tree, the quality of the best location l_j^* for v_j can be defined in terms of its parent v_i as:

$$Q_j(l_i) = \min_{l_j} \left(m_j(l_j) + d_{ij}(l_i, l_j) \right), \tag{13}$$

i.e., the best location l_j^* gives minimum cost as a sum of its match cost and deformation cost w.r.t. its parent. Similarly, for a vertex v_j that has children vertices defined by $v_c \in C_j$ and a parent vertex v_i, the quality of the best location can be defined as

$$Q_j(l_i) = \min_{l_j} \left(m_j(l_j) + d_{ij}(l_i, l_j) + \sum_{v_c \in C_j} Q_c(l_j) \right), \tag{14}$$

where, $Q_c(l_j)$ is the quality defined for a child v_c with parent l_j. This equation thus forms a recursive function with the boundary condition given by equation(13). The best location for the root vertex v_r can thus be defined as

$$l_r^* = \underset{l_j}{\text{argmin}} \left(m_r(l_r) + \sum_{v_c \in C_r} Q_c(l_r) \right). \tag{15}$$

The energy minimization in equation (12) can be achieved by first computing a list of candidate locations for each vertex and then applying equation (15) recursively. List of candidate locations for a vertex v_i are obtained by applying a threshold on $P(I|l_i, \mu_i, \Sigma_i)$ defined by equation (4) in section (2.2).

3. EXPERIMENTS AND RESULTS

We work with a dataset[3] of 28 high-resolution 3D SPGR T1-weighted MR images that was acquired on a GE Signa 1.5T scanner as series of 124 contiguous 1.5 mm coronal slices (256x256 matrix; 20cm FOV). Brain volumes were roughly aligned and linearly scaled to perform 9 parameter registration. Four control points were manually given to perform this global registration. The sub-cortical structures (caudate, putamen, and hippocampus) were manually labeled by expert neuroanatomists. We split the dataset into 14 training images and 14 testing images and learn each of the model parameters once, offline. Then we compute localization by maximizing (1).
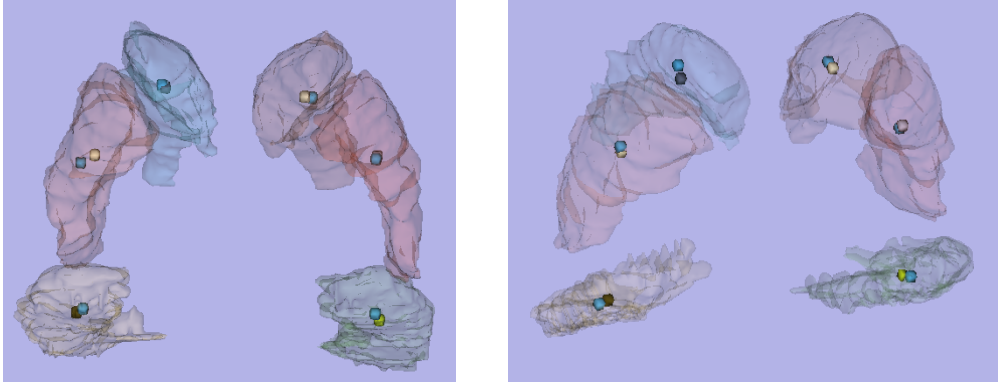
Figure 2. Two test results of our parts-based localization method for the sub-cortical brain structures. The blue dots depict the computed location of each respective part, the goal dot depicts the actual location.

Part	Average Error (in voxels)	Median Error (in voxels)	Hausdorff Error (in voxels)
Right Hippocampus	4.65	4.47	7.48
Left Hippocampus	5.01	5.54	8.33
Right Putamen	3.61	3.37	6.71
Left Putamen	5.50	5.69	8.31
Right Caudate	5.62	5.83	8.18
Left Caudate	4.83	4.89	7.61

Table 1. Localization Errors for Each Sub-Cortical Structure on Testing Images.

We are able to localize each of the six parts in all ten images with a worst-case Hausdorff distance of 8 voxels. For scale purposes, we estimate the average size of a given part bounding box to be 40 by 40 by 90 voxels. Table 1 summarizes the average errors, Median Errors and Hausdorff Distances of the localized parts in testing images. Table 2 summarizes the testing results on training images. Figure 2 shows visual results on two separate testing cases.

3.1 Implementation

Note that in our implementation, all model parameters were automatically estimated and required no manual inputs or "tweaks." The learning stage can be split into two stages: learning the pictorial model requires 2 to 5 seconds on a PC (Core i5 2.27 GHz, 4GB RAM). learning the appearance model requires approximately 20 seconds to train on 14 images. Testing time for each image varies from 10-20 minutes. If the tree obtained as a result of Minimum Spanning Tree based search as described in Section 2.3.2 is a straight chain, the testing takes more time in the recursive algorithm (Section 2.4) since there are more layers in the tree for each candidate location of the root. It takes less time if the root of the tree obtained has more children.

Part	Average Error (in voxels)	Median Error (in voxels)	Hausdorff Error (in voxels)
Right Hippocampus	3.57	3.67	8.16
Left Hippocampus	4.93	5.04	8.27
Right Putamen	2.82	2.45	5
Left Putamen	5.11	5.14	7.07
Right Caudate	4.25	4.18	7.60
Left Caudate	2.72	2.45	4.47

Table 2. Localization Errors for Each Sub-Cortical Structure on Training Images.

4. DISCUSSION

We have presented a parts-based approach to localizing sub-cortical structures in the brain. The method builds upon the pictorial structures literature, which has been thoroughly developed in the 2D natural imaging domain, but not in the medical imaging domain. Parts-based models have the potential to overcome many of the limitations inherent in the common voxel-based morphometry and atlas-based approaches. We hence believe this paper is first step towards a comprehensive, principled parts-based strategy for neuroanatomical structure parsing. Our accurate results on a 28-case data set substantiate these claims.

To the best of our knowledge, this is the first paper in the medical imaging literature on neuroanatomical structure parsing that takes an explicit and principled parts-based approach to the modeling task. We have demonstrated high accuracy in our method. In addition to the neuroanatomical parsing problem we've discussed here, many other anatomical and pathological problems in the medical imaging community could benefit from a parts-based methodology. We hence belief this paper could have a major impact on the medical image processing community.

However, there is a key limitation in the current formulation that may limit the direct applicability of the proposed ideas. Although the parts-based method has shown potential to localize the centroid of each part robustly, it does not present an immediate way to convert this localization of the part centroid to a delineation of the part in its entirety. Our current work directly investigates methods to overcome this challenge. We are investigating the potential to extend the framework to a multilayer idea (as we have, for example, in the lumbar imaging case[9]) as well as directly coupling the part representation with a segmentation process.

REFERENCES

[1] Held, K., Kops, E. R., Krause, B. J., Wells, W. M., I., Kikinis, R., and Muller-Gartner, H. W., "Markov random field segmentation of brain MR images," *Medical Imaging, IEEE Transactions on* **16**(6), 878–886 (1997).

[2] Zhang, Y., Brady, M., and Smith, S., "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging* **20**, 45–57 (January 2001).

[3] Corso, J. J., Tu, Z., Yuille, A., and Toga, A. W., "Segmentation of Sub-Cortical Structures by the Graph-Shifts Algorithm," in [*Proceedings of Information Processing in Medical Imaging*], Karssemeijer, N. and Lelieveldt, B., eds., 183–197 (2007).

[4] Fischl, B., Salat, D. H., Busa, E., Albert, M., Deiterich, M., Haselgrove, C., Kouwe, A. v. d., Killiany, R., Kennedy, D., Klaveness, S., Monttillo, A., Makris, N., Rosen, B., and Dale, A. M., "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron* **33**, 341–355 (January 2002).

[5] Tu, Z., Narr, K. L., Dinov, I., Dollar, P., Thompson, P. M., and Toga, A. W., "Brain anatomical structure segmentation by hybrid discriminative/generative models," *IEEE Transactions on Medical Imaging* **27**(4), 495–508 (2007).

[6] Mazziotta, J. C., Toga, A. W., Evans, A. C., Fox, P. T., and Lancaster, J., "A probabilistic atlas of the human brain: Theory and rationale for its development," *Neuroimage* **2**, 89–101 (1995).

[7] Fischler, M. A. and Elschlager, R. A., "The representation and matching of pictorial structures," *IEEE Transactions on Computers* **C-22**(1), 67–92 (1973).

[8] Felzenszwalb, P. F. and Huttenlocher, D. P., "Pictorial structures for object recognition," *International Journal of Computer Vision* **61**(1), 55–79 (2005).

[9] Alomari, R. S., Corso, J. J., and Chaudhary, V., "Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model," *IEEE Transactions on Medical Imaging* **30**(1), 1–10 (2011).