Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification

Jason J. Corso, *Member, IEEE,* Eitan Sharon, Shishir Dube, Suzie El-Saden, Usha Sinha, and Alan Yuille, *Member, IEEE*

***Abstract*— We present a new method for automatic segmentation of heterogeneous image data that takes a step toward bridging the gap between bottom-up affinity-based segmentation methods and top-down generative model based approaches. The main contribution of the paper is a Bayesian formulation for incorporating soft model assignments into the calculation of affinities, which are conventionally model free. We integrate the resulting model-aware affinities into the multilevel *segmentation by weighted aggregation* algorithm, and apply the technique to the task of detecting and segmenting brain tumor and edema in multichannel MR volumes. The computationally efficient method runs orders of magnitude faster than current state-of-the-art techniques giving comparable or improved results. Our quantitative results indicate the benefit of incorporating model-aware affinities into the segmentation process for the difficult case of glioblastoma multiforme brain tumor.**

***Index Terms*— Multilevel segmentation, normalized cuts, Bayesian affinity, brain tumor, glioblastoma multiforme.**

I. Introduction

Medical image analysis typically involves heterogeneous data that has been sampled from different underlying anatomic and pathologic physical processes. In the case of *glioblastoma multiforme* brain tumor (GBM), for example, the heterogeneous processes in study are the tumor itself, comprising a necrotic (dead) part and an active part, the edema or swelling in the nearby brain, and the brain tissue itself. To complicate matters, not all GBM tumors have a clear boundary between necrotic and active parts, and some may not have any necrotic parts. In Figure 1, we show a 2D slice of an MR image in the T1 weighted and T2 weighted channels presenting an enhancing GBM brain tumor. On the right, we outline the different heterogeneous regions of the brain tumor and label them as edema, active, or necrotic.

J.J. Corso is with the Department of Computer Science and Engineering, University at Buffalo SUNY. This work was completed while he was with the Department of Radiological Sciences, University of California, Los Angeles. He is the corresponding author: jcorso@cse.buffalo.edu.

E. Sharon is with the Department of Electrical Engineering, Technion - Israel Institute of Technology.

S. Dube is with the Department of Biomedical Engineering, University of California, Los Angeles.

S. El-Saden is with the Department of Radiological Sciences, University of California, Los Angeles.

U. Sinha is with the Department of Radiological Sciences, University of California, Los Angeles.

A. Yuille is with the Departments of Statistics and Psychology, University of California, Los Angeles.

Corso and Dube are funded by NLM Grant # LM07356. Yuille and Sharon are funded by the Keck Foundation.

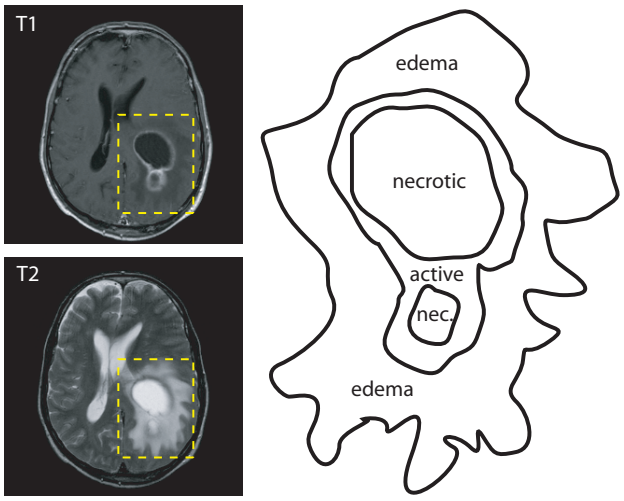

Fig. 1. Labeled example of a brain tumor illustrating the importance of the different modalities (T1 with contrast and T2).

It is assumed that a distinct statistical distribution of imaging features exists for each heterogeneous process, and that each distribution can be estimated from training data. In the constrained medical imaging domain, it is plausible to capture such feature distributions with relatively low-dimensional models that generalize to an entire population. This plausibility in medical imaging comes in contrast to the natural imaging domain in which the feature distribution can be extremely complex due to external phenomena like lighting and occlusion.

A key problem in medical imaging is automatically segmenting an image into its constituent heterogeneous processes. Automatic segmentation has the potential to positively impact clinical medicine by freeing physicians from the burden of manual labeling and by providing robust, quantitative measurements to aid in diagnosis and disease modeling. One such problem in clinical medicine is the automatic segmentation and quantification of brain tumors. We consider the GBM tumor because it is the most common primary tumor of the central nervous system, accounting for approximately 40% of brain tumors across patients of all ages [1], and the median postoperative survival time is extremely short (8 months) with a 5-year recurrence-free survival rate of nearly zero [2].

Quantifying the volume of a brain tumor is the key indicator of tumor progression [3]. However, like most segmentation problems, automatic detection and quantification of a brain

tumor is very difficult. In general, it is impossible to segment a GBM tumor by simple thresholding techniques [4]. Brain tumors are highly varying in size, have a variety of shape and appearance properties, and often deform other nearby structures in the brain [2]. In the current clinic, the tumor volume is approximated by the area of the maximal cross-section, which is often further approximated to an ellipse. Such a rough approximation is used because the time cost to compute a more accurate manual volume estimate is too high. Liu et al. [3] present an interactive system for computing the volume that reduces the cost of manual annotation and shows promise in volume estimates on a small number of cases.

However, no completely automatic segmentation algorithm has yet been adopted in the clinic. In Table I we present a concise review of the prior art in automatic tumor segmentation. Both GBM and non-GBM methods are given in the table for completeness. Fuzzy clustering methods (voxel-based) across all tumor types appear to be the most popular approach. Philips et al. [5] give an early proof-of-concept fuzzy clustering for brain tumor by operating on the raw multi-sequence data. They visually demonstrated that even with multi-sequence data the intensity distributions for tumor and normal tissue overlap. This led future researchers to incorporate additional knowledge into the feature vectors being clustered. Clark et al. [6] integrate knowledge-based techniques and multi-spectral histogram analysis to segment GBM tumors in a multichannel feature space. Fletcher-Heath et al. [7] take a knowledge-based fuzzy clustering approach to the segmentation followed by 3D connected components to build the tumor shape. Prastawa et al. [4] also present a knowledge-based detection/segmentation algorithm based on learning voxel-intensity distributions for normal brain matter and detecting outlier voxels, which are considered tumor. The distributions are learned with kernel-based density estimation methods, and the initial outlier detection is followed by a region competition algorithm.

Voxel-based statistical classification methods include [9], [10]. Kaus et al. [10] use the adaptive template-moderated classification algorithm [11] to segment the MR image into five different tissue classes: background, skin, brain, ventricles, and tumor. Their technique proceeds as an iterative sequence of spatially varying classification and non-linear registration. Prastawa et al. [9] define a parametric distribution across multiple channels of tumor as a mixture of Gamma and Gaussian components. They use the Expectation-Maximization algorithm [18] to perform segmentation and iteratively adapt the model parameters to the case at hand.

These two sets of methods are limited by their extreme degree of locality, i.e., they are voxel-based and do not take local or global context into account. While they have had some success in segmenting low-grade gliomas and meningiomas (relatively homogeneous) on a good-sized data set [10], their success is limited in the more relevant GBM (heterogeneous) segmentation examples. Furthermore, it is not clear this limited success will scale to the more difficult inevitable cases arising in larger data-sets (like the one used in this paper). There have been few attempts at solving this problem of local ambiguity. One method of note is the recent work of Lee et al. [14] that uses the context-sensitive discriminative random fields model [19], [20]. They use a set of knowledge-based features [21] coupled with support vector machines to perform the segmentation and classification. The use of energy and shape models (e.g., level-sets [13] and active contours [16]) is promising but is generally iterative in nature and therefore sensitive to initialization, which, unless interactive, is nearly as difficult as the entire segmentation for brain tumor.

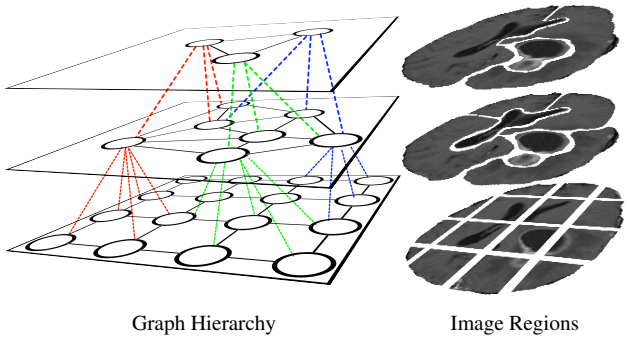

Fig. 2. The SWA algorithm gives a graph hierarchy of potential voxel segments at different scales. This figure shows an explanatory 2D graph hierarchy and the corresponding image region of each lattice element. Only a few interlevel connections are drawn; note how one node can have multiple parents. In practice, the individual voxels form the lowest graph layer.

In this paper[1], we present a new method for automatic segmentation of heterogeneous image data that is applicable in any case for which distinct feature distributions can be learned for the heterogeneous regions. To demonstrate such an application, we experiment with the task of detecting and segmenting brain tumors but note the method is generally applicable. Our method combines two of the most effective approaches to segmentation. The first approach, exemplified by the work of Tu et al. [23], [24], uses class models to explicitly represent the different heterogeneous processes. The tasks of segmentation and classification are solved jointly by computing solutions that maximize a posterior distribution that has been learned from training data. To make the optimization tractable, the posterior is often represented as a product distribution over generative models on sets of pixels, or segments. Hence, we call these methods *model-based*. This type of approach is very powerful as the solutions are guaranteed to be from a statistical distribution that has been learned from training data, but the algorithms for obtaining these estimates are comparatively slow and model choice is difficult. Some techniques have been studied to improve the efficiency of the inference, e.g. Swendsen-Wang sampling [25], but these methods still remain comparatively inefficient.

The second approach is based on the concept of graph cuts [26]. In these affinity-based methods, the input data induces a sparse graph, and each edge in the graph is given an *affinity* measurement that characterizes the similarity of

[1]This paper is an extended version of [22]. Here, we present a more complete technical discussion, full qualitative comparison to the literature, complete results and failure mode analysis, and a new mathematical formulation for learning the parameters of the model-aware affinity functions from labeled training data.

Authors	Description	Type	# Cases	Accuracy	Time
Liu et al. [3]	Fuzzy clustering (semi-automatic)	GBM	5	99%	16 min.
Phillips et al. [5]	Fuzzy clustering	GBM	1	N/A	N/A
Clark et al. [6]	Knowledge-based fuzzy clustering	GBM	7	70%	N/A
Fletcher-Heath et al. [7]	Knowledge-based fuzzy clustering	NE	6	53%-90%	N/A
Karayiannis and Pin [8]	Fuzzy clustering (VQ)	MG	1	N/A	N/A
Prastawa et al. [4]	Knowledge-based/outlier detection	GBM	4	68%-80%	90 min.
Prastawa et al. [9]	Statistical classification via EM	GBM	5	49%-71%	100 min.
Kaus et al. [10], [11]	Statistical classification with atlas prior	LGG, MG	20	99%	10 min.
Vinitski et al. [12]	k-Nearest neighbor	N/A	9	N/A	2 min.
Ho et al. [13]	3D level sets	GBM	3	85%-93%	N/A
Lee et al. [14]	Discriminative Random Fields and SVM	GBM, AST	7	40%-89%	N/A
Peck et al. [15]	Eigenimage analysis	N/A	10	N/A	N/A
Zhu and Yan [16]	Hopfield neural network and active contours	N/A	2	N/A	N/A
Zhang et al. [17]	Support vector machines	N/A	9	60%-87%	N/A
Our Method	Multilevel Bayesian segmentation	GBM	20	27%-88%	7 min.

TABLE I

SUMMARY OF RELATED METHODS IN AUTOMATIC BRAIN TUMOR SEGMENTATION. THE TYPE ABBREVIATIONS ARE GBM: GLIOBLASTOMA MULTIFORME, AST: ASTROCYTOMA, NE: NON-ENHANCING, LGG: LOW-GRADE GLIOMA, MG: MENINGIOMA. N/A IS USED WHENEVER THE INFORMATION IS NOT GIVEN IN THE PAPER. ACCURACIES ARE COMPUTED AS VOLUME OVERLAP, WHICH IS ALSO CALLED THE JACCARD SCORE. WE DISCUSS OUR RESULTS IN DETAIL IN SECTION VI-E AND INCLUDE FAILURE MODE ANALYSIS FOR THE FEW CASES AT THE LOWER END OF THE RANGE (THE MAJORITY OF OUR CASES SCORE GREATER THAN 70%).

the two neighboring nodes in some predefined feature space. Cuts are sets of edges that separate the graph into two subsets, which are typically computed by analyzing the eigenspectrum [26], [27] or pairwise-predicate measures [28]. These methods have led to the hierarchical segmentation by weighted aggregation (SWA) algorithm due to Sharon et al. [29]–[31]. SWA was first extended to the 3D image domain by Akselrod-Ballin et al. [32] for the problem of multiple sclerosis segmentation.

SWA operates by recursively coarsening the initial graph using an adapted algebraic multigrid algorithm [33]; it is shown to approximate the normalized cut measure [26]. The SWA algorithm produces a multilevel segmentation of the data with each node in the hierarchy representing a potential segment (see Figure 2 for a simple example). The hierarchy can capture interesting multiscale properties like, for example, the necrotic and active parts of the tumor as initially separate segments to be joined at a higher level in the hierarchy as a single segment. However, the original algorithm does not give a method for selecting individual segments to produce a final classification of the data. SWA is rapid and effective, but does not explicitly take advantage of the class models used in [23].

The main contribution of this paper is the model-aware affinity, which is step toward unifying these two disparate segmentation approaches by incorporating models into the calculation of the affinities on the graph and then using the models to extract a final classification from the hierarchy. Both the model parameters and the model-aware affinity parameters are learned from labeled training data. Our method incorporates information from multiple scales and thus has greater potential to avoid the local ambiguity problem that affects the prior voxel-based classification and clustering methods. Furthermore, our algorithm defines a feed-forward process that requires no initialization and is capable of doing classification during this process. We demonstrate encouraging results and cross-validate them on a comparatively large GBM dataset.

The organization of the paper is as follows: first, we discuss the necessary background in generative models and the notation that will be used in the paper (Section II). Next, we describe (Section III) how we incorporate Bayesian model classification into the calculation of affinities. The proposed model-aware affinity leads to improved cuts by allowing the use of affinity functions tailored to the specific models in use. We extend the SWA algorithm to include the model-aware affinity in Section IV. In Section V, we describe a method to extract the segmentation from the SWA hierarchy that makes explicit use of the model probabilities from the new affinity function. Finally, in Section VI we discuss the application of our method to the problem of brain tumor segmentation in multichannel MR. We describe the specific class models and probability functions used in the experimental results. We conclude with a failure mode analysis of the proposed method and discuss potential improvements.

II. MATHEMATICAL BACKGROUND

In this section, we first make the definitions and describe the notation necessary for the technical discussion. Then, we introduce the necessary background concepts.

A. Notation

The input data induces a graph, $G = (\mathcal{V}, \mathcal{E})$, on which all of the analysis occurs. Associated with each node in the graph, $u, v \in \mathcal{V}$, are properties, or statistics, denoted $s_u \in \mathcal{S}$, where $\mathcal{S}$ is the space of properties (e.g., $\mathbb{R}^3$ for red-green-blue image data). Edges in the graph, $e_{uv} \in \mathcal{E}$, are created based on connectivity relations in the input data. Define a *cluster* to be a connected set of nodes $C \subset \mathcal{V}$ in the graph such that $C_k \cap C_l = \varnothing$ when $k \neq l$ and $\bigcup C_k = \mathcal{V}$.

Associated with each node is a random variable, m_u, called the model variable that takes values from a discrete set of process models $\mathcal{M}$ that is problem specific; in the brain

tumor example this set would be {brain, tumor, edema}. Additionally, associated with each edge is a binary random variable, X_{uv}, called the edge activation variable, and the set of these over $\mathcal{E}$ is denoted $\mathcal{X}$. An edge activation variable takes value 1 if u and v are in the same cluster and value 0 if the two nodes are not in the same cluster. Thus, an instance of $\mathcal{X}$, an *activation set*, defines a segmentation of the data into clusters.

For a given image, there may be multiple plausible activation sets. These multiple interpretations often arise from the inherent scale ambiguity in biomedical images: for example, at one scale, a tumor is composed of separate necrotic and active segments, but at a higher scale, the two subparts of the tumor are joined giving a single tumor segment. We thus note that the clusters are not deterministically defined by assignments to the model variables: both sets of variables are stochastic, there is an interdependent relationship between the two, and nodes with different model variables can reside in the same cluster.

B. Generative Models

The model based methods define a likelihood function $P(\{s_u\}|\{m_u\})$ for the probability of the observed statistics $\{s_u\}$ conditioned on the model variables $\{m_u\}$ of the pixels $\{u\}$. The methods also put prior probabilities $P(\{m_u\})$ on the model variables defining what is termed a *generative model* [34]. Intuitively, the term generative means that by explicitly modeling the likelihood and prior, the creative, or generative, process has been captured. Thus, one can generate random samples from the model that *resemble* the real images from which the model was trained. Examples of such generative models include simple point processes like those used in this paper, maximum entropy models of texture [35], and stochastic grammars [36].

Computing estimates of the class labels that maximize the posterior probability

$$P(\{m_u\}|\{s_u\}) \propto P(\{s_u\}|\{m_u\})P(\{m_u\}) \tag{1}$$

is the modus operandi of the model-based segmentation methods. However, such distributions are typically very high-dimensional and require very sophisticated modeling and inference algorithms.

C. Affinity Methods

In contrast, the affinity-based methods define a comparatively efficient bottom-up strategy for computing the segmentation of the input data. In the induced graph, each edge is annotated with a weight that represents the affinity of the two nodes. The affinity is denoted by w_{uv} for connected nodes u and $v \in \mathcal{V}$. Conventionally, the affinity function is of the form

$$w_{uv} = \exp\left(-D(s_u, s_v; \theta)\right) \tag{2}$$

where D is a non-negative *distance* measure and θ are predetermined parameters. To promote efficient calculation, the affinity function is typically defined on a simple feature space like intensity or texture. For example, on intensities a common function is $\theta\,|s_u - s_v|_1$. The parameters θ are fixed and predetermined through some heuristic techniques or learned from training data [37].

The goal of affinity methods is to detect salient clusters defined as those clusters giving small values of the following function

$$\Gamma(C) = \frac{\sum_{u\in C, v\notin C} w_{uv}}{\sum_{u,v\in C} w_{uv}} \ . \tag{3}$$

Such clusters have low affinity across their boundaries and high affinity within their interior. This is the so-called normalized cut criterion [26]. Eigenvector techniques [27] were originally used to compute the clusters, but, more recently, an efficient multiscale algorithm for doing this was proposed [29]–[31] and is described in Section IV.

III. Integrating Models and Affinities

In this paper, we restrict ourselves to the simple generative model where $P(s_u|m_u)$ is the conditional probability of the statistics s_u at a node u with model m_u, and $P(m_u, m_v)$ is the prior probability of model labels m_u and m_v at nodes u and v. We assume the edge activation variables are conditionally independent given the properties at its nodes.

We use probabilities to combine the generative model methods with the affinities. The affinity between nodes $u, v \in \mathcal{V}$ is defined to be the probability of the binary event X_{uv} that the two nodes lie in the same cluster. This probability is calculated by treating the class labels as hidden variables that are summed out:

$$\begin{aligned}
&P(X_{uv}|s_u, s_v) = \\
&\sum_{\substack{m_u \\ m_v}} P(X_{uv}|s_u, s_v, m_u, m_v) P(m_u, m_v|s_u, s_v) \ , \\
\propto &\sum_{\substack{m_u \\ m_v}} P(X_{uv}|s_u, s_v, m_u, m_v) P(s_u, s_v|m_u, m_v) P(m_u, m_v) \ , \\
= &\sum_{\substack{m_u \\ m_v}} P(X_{uv}|s_u, s_v, m_u, m_v) P(s_u|m_u) P(s_v|m_v) P(m_u, m_v) \ ,
\end{aligned} \tag{4}$$

where the third line follows from the assumption that the nodes are conditionally independent given class assignments. This Bayesian *model-aware affinity* avoids making premature hard assignments of nodes to models by integrating over all possible models and weighting by the class evidence and prior. The formulation also makes it plausible to define a custom affinity function for each model pair. The first term in the sum of (4) is a model specific affinity:

$$P(X_{uv}|s_u, s_v, m_u, m_v) = \exp\Big(-D\left(s_u, s_v; \theta[m_u, m_v]\right)\Big) \ . \tag{5}$$

Note that the property of belonging to the same region is not uniquely determined by the model variables m_u, m_v. Pixels with the same model may lie in different regions and pixels with different model labels might lie in the same region.

This definition of affinity is suitable for heterogeneous data since the affinities are explicitly weighted by the evidence $P(s_u|m_u)$ for class membership at each pixel u, and so can

adapt to different classes. This differs from the conventional affinity function $w_{uv} = \exp(-D(s_u, s_v; \theta))$, which does not model class membership explicitly. The difference becomes most apparent when the nodes are aggregated to form clusters as we move up the pyramid, see the multilevel algorithm description in Section IV. Individual nodes, at the bottom of the pyramid, will typically only have weak evidence for class membership (i.e., $P(s_u|m_u)$ is roughly constant as a function of m_u). But as we proceed up the pyramid, clusters of nodes will usually have far stronger evidence for class membership, and their affinities will be modified accordingly.

The formulation presented here is general; in this paper, we integrate these ideas into the SWA multilevel segmentation framework (Section IV). In Section VI, we discuss the specific forms of these probabilities used in our experiments.

IV. Segmentation By Weighted Aggregation

We now review the segmentation by weighted aggregation (SWA) algorithm of Sharon et al. [29]–[31], and describe our extension to integrate model-aware affinities. As earlier, define a graph $G^t = (\mathcal{V}^t, \mathcal{E}^t)$ with the additional superscript indicating the level in a pyramid of graphs $\mathcal{G} = \{G^t : t = 0, \ldots, T\}$. Denote the multichannel intensity vector at voxel i as $I(i) \in \mathbb{R}^C$, with C being the number of channels.

A. Original SWA Algorithm

The finest layer in the graph $G^0 = (\mathcal{V}^0, \mathcal{E}^0)$ is induced by the voxel lattice: each voxel i becomes a node $v \in \mathcal{V}$ with 6-neighbor connectivity, and node properties set according to the image, $s_v = I(i)$. The affinities, w_{uv}, are initialized as in Section III using $D(s_u, s_v; \theta) \doteq \theta |s_u - s_v|_1$. SWA proceeds by iteratively coarsening the graph according to the following algorithm:

1) $t \leftarrow 0$, and initialize G^0 as described above.
2) Choose a set of representative nodes $\mathcal{R}^t \subset \mathcal{V}^t$ such that $\forall u \in \mathcal{V}^t$ and $0 < \beta < 1$
$$\sum_{v \in \mathcal{R}^t} w_{uv} \geq \beta \sum_{v \in \mathcal{V}^t} w_{uv} \quad . \tag{6}$$
3) Define graph $G^{t+1} = (\mathcal{V}^{t+1}, \mathcal{E}^{t+1})$:
   a) $\mathcal{V}^{t+1} \leftarrow \mathcal{R}^t$, and edges will be defined in step 3f.
   b) Compute interpolation weights
   $$p_{uU} = \frac{w_{uU}}{\sum_{V \in \mathcal{V}^{t+1}} w_{uV}} \quad , \tag{7}$$
   with $u \in \mathcal{V}^t$ and $U \in \mathcal{V}^{t+1}$.
   c) Accumulate statistics to coarse level:
   $$s_U = \sum_{u \in \mathcal{V}^t} \frac{p_{uU} s_u}{\sum_{v \in \mathcal{V}^t} p_{vU}} \quad . \tag{8}$$
   d) Interpolate affinity from the finer level:
   $$\hat{w}_{UV} = \sum_{(u \neq v) \in \mathcal{V}^t} p_{uU} w_{uv} p_{vV} \quad . \tag{9}$$
   e) Use coarse affinity to modulate the interpolated affinity:
   $$W_{UV} = \hat{w}_{UV} \exp(-D(s_U, s_V; \theta)) \quad . \tag{10}$$
   f) Create an edge in $\mathcal{E}^{t+1}$ between $U \neq V \in \mathcal{V}^{t+1}$ when $W_{UV} \neq 0$.
4) $t \leftarrow t + 1$.
5) Repeat steps 2 $\rightarrow$ 4 until $|\mathcal{V}^t| = 1$ or $|\mathcal{E}^t| = 0$.

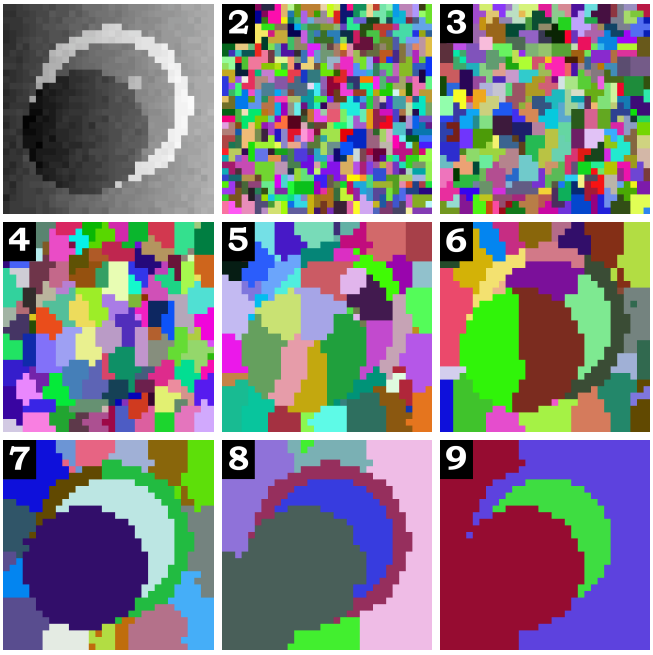

Fig. 3. Example SWA hierarchy on a synthetic grayscale image. The numbers indicate the level in the hierarchy (0 would be the pixels).

The parameter β in step 2 governs the amount of coarsening that occurs at each layer in the graph (we set $\beta = 0.2$ in this work). There is no explicit constraint to select a minimum set of representative nodes that satisfy (6). However, the set $\mathcal{R}^t$ should be programmatically chosen to be a minimum set or the height of the resulting graph hierarchy is potentially unbounded. [30] shows that this algorithm preserves the saliency function (3).

In Figure 3, we show the hierarchy resulting from running the SWA coarsening algorithm on a synthetic grayscale image. The input image is drawn in the top-left corner of the figure; it consists of a bright annulus, a dot, a dark circle and a noise process in the background. The levels of the pyramid depict the segments (drawn with arbitrary colors) outputted by the iterative coarsening process. Until we encounter some of the objects of interest, the coarsening follows an isotropic growth. At levels 3 and 4, the small dot is segmented well. At level 7 the dark circle is detected, and at level 8 the annulus is detected. Eventually, all of the segments merge into a single segment (on level 10, not shown).

In the example, we see that each of the salient foreground objects in the image is correctly segmented at some level in the hierarchy. However, being objects of different scale, they are not detected at the same level. Sharon et al. [30], [31] suggest thresholding the saliency function (3) to detect the salient objects at their *intrinsic* scale. In our experimentation, we found this method to be inadequate for medical imaging data because the objects of interest are often not the only salient

objects and seldom the most salient objects in the imaging volume resulting in many false positives. In Section V, we propose a new method for extracting segments from the hierarchy that incorporates the model likelihood information, and in Section VI, we show the model-aware approach performs significantly better than the saliency based approach.

B. Incorporating Model-Aware Affinities

The two terms in (10) convey different affinity cues: the first affinity $\hat{w}_{UV}$ is comprised of finer level (scale) affinities interpolated to the coarse level, and the second affinity is computed from the coarse level statistics. For all types of regions, the same function is being used. However, at coarser levels in the graph, evidence for regions of known types (e.g., tumor) starts appearing making it sensible to compute a model-specific affinity (step 3e below). Furthermore, the model-aware affinities compute the model likelihood distribution, $P(s_U|m_U)$, and we can also associate a most likely model m_U^* with each node (step 3f below). The final algorithm follows:

1) $t \leftarrow 0$, and initialize G^0 as earlier.
2) Choose a set of representative nodes $\mathcal{R}^t$ satisfying (6).
3) Define graph $G^{t+1} = (\mathcal{V}^{t+1}, \mathcal{E}^{t+1})$:
   a) $\mathcal{V}^{t+1} \leftarrow \mathcal{R}^t$, and edges will be defined in step 3g.
   b) Compute interpolation weights according to (7).
   c) Accumulate statistics according to (8).
   d) Interpolate affinity according to (9).
   e) Apply the model-aware affinity as a modulation factor:
   $$W_{UV} = \hat{w}_{UV} P(X_{UV}|s_U, s_V) \ , \qquad (11)$$
   where $P(X_{UV}|s_U, s_V)$ is evaluated as in (4).
   f) Associate a class label with each node:
   $$m_U^* = \arg\max_{m \in \mathcal{M}} P(s_U|m) \ . \qquad (12)$$
   g) Create an edge in $\mathcal{E}^{t+1}$ between $U \neq V \in \mathcal{V}^{t+1}$ when $W_{UV} \neq 0$.
4) $t \leftarrow t + 1$.
5) Repeat steps $2 \rightarrow 4$ until $|\mathcal{V}^t| = 1$ or $|\mathcal{E}^t| = 0$.

We demonstrate a quantitative improvement from integrating models into the affinity calculation in the results presented in Section VI-E.

V. Extracting Segments from the Hierarchy

Both the original and the modified SWA algorithms produce a graph hierarchy during the iterative coarsening of the input image. As briefly discussed through the example in Figure 3, extracting the segments corresponding to objects of interest from the hierarchy is non-trivial. In this section, we propose two extraction algorithms. First, we discuss an approach that uses saliency (3) and is derivative of the original SWA papers [30], [31]. Second, we present a new extraction algorithm that is based on tracing a voxel's model *signature* up the hierarchy. The second method relies exclusively on the generative models that have been learned from data, and in the results (Section VI), we show it outperforms the original saliency-based approach.

A. Saliency-Based Extraction

This method associates each voxel with the most salient segment of which it is a part in the graph hierarchy. The routine proceeds for each voxel independently; neighborhood information for the voxels is implicitly incorporated due to the agglomerative nature of the graph hierarchy. First, associate every voxel with a node at each level using the Gauss-Seidel relaxation sweeps algorithm [30]. For voxel i, denote the node v at level t with which it is associated by v_i^t. Then, traverse the hierarchy to find the level at which the associated node is most salient:

$$t^* = \arg\min_{t=\{1\ldots T\}} \Gamma\left(v_i^t\right) \ . \qquad (13)$$

Finally, label the voxel with the class associated with this most salient node: $m_i \leftarrow m_{v_i^{t^*}}$.

B. Model-Based Extraction

We focus on the model likelihood function that is computed during the Bayesian model-aware affinity calculation (4). In this extraction method, we conserve the soft clustering nature of the SWA algorithm in contrast to the saliency based method that makes a hard assignment of a voxel to a node at each level. Again, we proceed independently for each voxel letting the neighborhood information be captured by the multiscale nature of the graph hierarchy.

For each voxel i with corresponding node v, create a variable m_v^0 to store the most likely model as in (12).

$$m_v^0 = \arg\max_{m \in \mathcal{M}} P(s_v|m) \qquad (14)$$

Then, recursively proceed up the hierarchy creating such a model variable for the voxel at each level in the hierarchy. Explicitly use the interlevel interpolation weights (7) to incorporate the soft node coarsening from SWA. For example, at level one, the function is easily written:

$$m_v^1 = \arg\max_{m \in \mathcal{M}} \sum_{V \in \mathcal{V}^1} p_{vV} P(s_V|m) \qquad (15)$$

From the $T+1$ resulting model variables, associate the voxel with the model that occurs most frequently. As discussed earlier, the model likelihood distribution will be roughly constant at the fine (initial) levels in the hierarchy but will tend to quickly peak for one model. In most cases, the likelihood will remain peaked until the node gets joined to some other larger node of a different class at which time, the distribution will shift to prefer that class.

VI. Application to Brain Tumor Segmentation

In this section, we discuss the application to automatic segmentation and volume estimation of GBM brain tumors. As discussed in the introduction, brain tumors are highly varying in size, have a variety of shape and appearance properties, and often deform nearby structures in the brain [2]. Quantifying the volume of the tumor is a key indicator of tumor progression [3]. An accurate volume measurement can be used to analyze the effectiveness of new treatments. However, no automatic approach yet exists for automatically and accurately computing the volume.

A. Data Processing

We work with a dataset of 20 expert-annotated GBM studies. Using FSL tools [38], we pre-process the data through the following pipeline: (1) spatial registration, (2) noise removal, (3) skull removal, and (4) intensity standardization. The intensity standardization is intended to align the gray and white matter peaks in the intensity histogram; this procedure can be corrupted by the presence of a large tumor. We have taken no extra step in the standardization to make it robust to such corruption. We use the T1 weighted pre- and post-contrast (T1CE), FLAIR and the T2 weighted MR sequences. The 3D data is 1×1 mm^2 resolution in the axial plane but it varies (even for a single subject) in the slice resolution. For example, the typical slice resolution used in the T1CE channel is near isotropic ($1 \times 1 \times 1$ mm^3), but the slice resolution in the T2 channel is highly anisotropic (e.g., $1 \times 1 \times 10$ mm^3). Since the proposed method assumes all image data lies on the same lattice, we subsample all channels to match the resolution of the lowest channel (since subsampling is generally a more reliable task than extrapolating). While accurate volume measurements are difficult under such anisotropy, the data reflects current clinical practices in diagnostic radiology. The resulting 3D data is 256×256 with an average of 24 slices (the range is 22 to 26 slices). To facilitate training and testing, we split the dataset into two sets and denote them as (A##) and (B##). For most of the results, we use set A for training and set B for testing. We do include two-fold cross-validation results (train on B and test on A).

B. Class Models and Feature Statistics

We model four classes of data: non-data (outside of head), brain matter, tumor, and edema. The tumor class includes the necrotic (dead) part of the GBM tumor, the enhancing (active) part of the tumor, the ambiguous tissue in between necrotic and enhancing, as well as possible tumor infiltration and non-enhancing tumor. The edema represents non-tumor, healthy tissue that has a swelling response to the tumor. We assume the underlying statistics of the tumor and edema intensities conform to a mixture model with Gaussian components (GMM). The mixture model is plausible because of its relative simplicity and generality: it can be shown that given enough components, a mixture model can fit any finite set of empirical data. The single channel intensity histograms for the tumor and edema data are given in Figure 4. We can see the enhancing tissue as a heavy tail in the T1CE channel. The mixture model can capture the different subparts of the tumor model as separate components.

Denote the parameters of a Gaussian component $\psi_i = \{\phi_i, \mu_i, \Sigma_i\}$, where μ_i is the mean vector and Σ_i is the covariance matrix. The ϕ_i parameter is called the mixing coefficient and describes the relative weight of component i in the complete model. For the complete model, write $\Psi = \{k, \psi_i, \ldots \psi_k\}$, where k is the number of components in the data. A mixture model on d-dimensional data x is written

$$P(x;\Psi) = \sum_{i=1}^{k} \phi_i P\left(x;\mu_i,\Sigma_i\right)$$

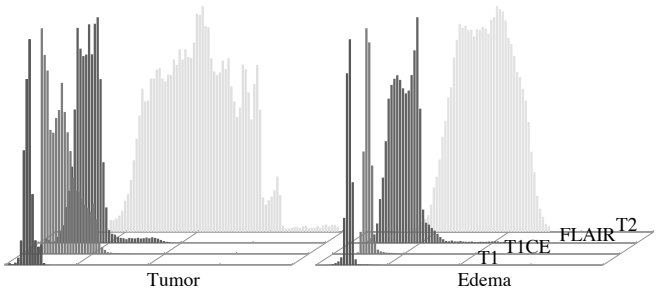

Fig. 4. Class intensity histograms in each channel independently for the whole population.

$$= \sum_{i=1}^{k} \phi_i \frac{\exp\left(-\frac{1}{2}\left(x-\mu_i\right)^{\mathsf{T}} \Sigma_i^{-1}\left(x-\mu_i\right)\right)}{(2\pi)^{\frac{d}{2}}\left|\Sigma_i\right|} \quad . \quad (16)$$

The standard Expectation-Maximization algorithm [18] is used to estimate the parameters of each class mixture model in a maximum likelihood formulation. Training these parameters takes about two minutes per class on a standard Linux PC.

The node-class likelihoods $P(s_u|m_u)$ are computed directly against this mixture model. The class prior term, $P(m_u, m_v)$, encodes the obvious hard constraints (i.e. tumor cannot be adjacent to non-data), and sets the remaining unconstrained terms to be uniform according to the maximum entropy principle.

C. Model-Aware Affinity Definition and Learning

For the model-aware affinity term (5), we use a class dependent weighted distance:

$$P(X_{uv}|s_u, s_v, m_u, m_v) = \exp\left(-\sum_c \theta^c_{m_u m_v}\left|s^c_u - s^c_v\right|\right) \quad , \quad (17)$$

where superscript ^{c} indicates vector $\underline{\theta}$ element at index c (the channel c). The class dependent coefficients are learned from the labeled training data by constrained optimization. Intuitively, the affinity between two nodes of the same (different) class should be near 1 (0). Thus, we learn the coefficients by optimizing the following function under the constraint that the coefficients sum to one for each class pair, $\sum_c \theta^c_{m_u,m_v} = 1, \quad \forall\{m_u, m_v\}$.

$$\underline{\theta}^*_{m^*_u,m^*_v} = \begin{cases} \arg\max\limits_{\underline{\theta}} \sum\limits_{\substack{u:m_u=m^*_u \\ v:m_v=m^*_v}} \exp\left(-\sum\limits_c \theta^c\left|s^c_u - s^c_v\right|\right) & \text{if } m^*_u = m^*_v \\ \arg\min\limits_{\underline{\theta}} \sum\limits_{\substack{u:m_u=m^*_u \\ v:m_v=m^*_v}} \exp\left(-\sum\limits_c \theta^c\left|s^c_u - s^c_v\right|\right) & \text{otherwise.} \end{cases} \quad (18)$$

To optimize the coefficients for each class-pair, we perform an initial stochastic search for the best parameters followed by a steepest coordinate-descent procedure. The gradient of the function is estimated numerically at each iteration and the single coordinate that optimally modifies the affinities is

adjusted. The procedure is terminated when no adjustment will improve the affinity over the training data. Because this is an offline procedure in a non-linear space with many local minima, we emphasize the utility of the initial stochastic search to locate a good starting point for the coordinate-descent step. Typically, the training is complete in about one minute (on a standard Linux PC) for ten thousand iterations of stochastic search and the ensuing descent.

We present the learned coefficients in Table II (we abbreviate non-data, ND). We have included only those pairs not excluded by the hard constraints discussed above. The learned coefficients reflect the general intuition about what channels are best utilized to analyze certain region types. For example, the affinity between brain and edema region is measured solely in the FLAIR and T2 channels while the affinity between tumor and edema incorporates the T1CE, FLAIR, and T2 channels. Note that the coefficients are symmetric (i.e., equal for Brain, Tumor and Tumor, Brain).

m_u, m_v \ c	T1	T1CE	FLAIR	T2
ND, Brain	0.00	0.01	0.03	0.96
Brain, Brain	0.41	0.59	0.00	0.00
Brain, Tumor	0.00	0.15	0.36	0.49
Tumor, Tumor	1.00	0.00	0.00	0.00
Brain, Edema	0.00	0.00	0.54	0.46
Tumor, Edema	0.00	0.31	0.20	0.49
Edema, Edema	0.45	0.55	0.00	0.00

TABLE II

MODEL-AWARE COEFFICIENTS THAT WERE LEARNED FROM THE TRAINING DATA. ROWS ARE (SYMMETRIC) CLASS PAIRS AND COLUMNS ARE CHANNELS.

The sole feature statistic that we accumulate for each node in the graph (SWA step 3c) is the average intensity; i.e., the intensity of a super node is the weighted average of its children's intensities. The feature statistics, model choice and model-aware affinity form is specific to our problem; many other choices could also be made in this and other domains for these functions.

D. Implementation and Computational Efficiency

The multilevel approach based on the segmentation by weighted aggregation algorithm is approximately linear in the number of input voxels ($\left|\mathcal{V}^0\right|$) [30]. With the addition of the Bayesian model-aware affinity calculation a multiplicative factor in the number of models squared is imposed. This number is typically small, four in this case, and in our experience has not greatly affected the computational efficiency.

However, the memory requirement of the multilevel algorithm is high. The burden is not in the graph nodes, which are $O(2\left|\mathcal{V}\right|)$ since each coarsening procedure cuts the number of nodes in half, roughly. Instead, the cost of maintaining the adaptive neighborhood structure and soft assignment during the agglomeration is large, even in the case of a sparse initial graph. For the pixel layer, it is a linear term (each node is connected to a fixed number of neighbors). However, while the number of nodes decreases at each coarser layer in the hierarchy, the neighborhood structure grows. We observe a roughly constant total number of edges in a graph layer, which gives a memory requirement of $O(|\mathcal{V}|\log|\mathcal{V}|)$. The explicit representation of the soft interlevel weights requires a second $O(|\mathcal{V}|\log|\mathcal{V}|)$ order memory term. Sharon et al. [31] give suggestions for dealing with such memory cost. In our implementation, we rely on an out-of-core memory buffer to store the node relationships at each layer in the graph hierarchy.

The algorithm is implemented in pure Java (v1.5) with no native bindings. On a typical image volume of size $256\times256\times24$, the entire volume is completely segmented and classified in less than 1 minute using a 3GHz P4 Linux machine with a heap size of 1.5GB and less than 2 minutes using a 1.67GHz PowerPC Mac OS X laptop with a heap size of 1.5GB. Including the cost of preprocessing, which is about 5 minutes, these times are orders of magnitude faster than the current state of the art in medical image segmentation, specifically brain tumor segmentation as summarized in Table I. For example, the execution time given in [4] is about 90 minutes on a 2GHz Xeon machine. We note that the apparent speedup observed in our approach may be caused by some intrinsic characteristic of the data itself (e.g., the highly anisotropic voxels) in comparison with other methods in the literature. However, a direct comparison is not possible since each paper works with a different dataset and often prior works have not given the voxel resolution. But, as discussed above, the proposed method scales linearly with the size of the input image, and has been observed to perform at similar rapid rates in other situations with higher resolution data.

E. Results

In this section, we show some results, both qualitative and quantitative, from the experiments. For space reasons, in most cases we show a single, indicative slice from the volume (all processing is in 3D). In the classification figures, we use green to represent the tumor label and red to represent edema. The colors used to depict different segments in the hierarchy are arbitrary.

1) Hierarchy Example: Figure 5 shows an example segmentation hierarchy. In this typical example, we can see that even at finer levels (5 and 6), the agglomeration process begins to capture the subregions of the enhancing and necrotic tumor tissues. At level 8 the entire necrotic subregion is segmented while the active region and the edema region are split into parts. The edema is never grouped into a single region before joining with part of the tumor, which is partly necrotic and partly enhancing (due to 3D processing, on a different slice). Finally, at level 10, the tumor and edema regions are completely absorbed by the brain region.

2) Quantitative Results: In Table III, we show the volume overlap (Jaccard) results taking a weighted averaged over the set. Let T be the true positive, F_p be the false positive, and F_n be the false negative. The Jaccard score is $T/(T+F_p+F_n)$. The algorithm we propose in this paper is labeled (in bold) "Model-Based SWA." The single voxel classifier uses the same learned models and applies a Bayes classification rule

Image	Tumor Accuracy (%)			Tumor Volume (vx^3)		Tumor Surface Error (mm)			Edema Accuracy (%)		
	Jaccard	Prec.	Recall	Auto	Manual	Mean	Median	Hausdorff	Jaccard	Prec.	Recall
A1	88	96	92	18720	19415	0.3	0	3	71	80	86
A2	52	72	65	2180	2432	5.17	0	50.14	32	50	47
A3	30	80	32	4712	11722	1.94	1	19.13	55	57	94
A4	31	32	92	476	166	1.72	1	15.3	55	62	83
A5	47	48	99	11014	5286	2.52	1	17.29	61	82	71
A6	47	52	83	2537	1590	2.42	1	44.4	20	28	43
A7	70	81	84	6909	6621	0.4	0	26.1	50	66	68
A8	69	69	99	7273	5116	1.54	0	24.39	72	77	92
A9	75	75	99	15014	11444	4.66	0	52.36	72	80	88
A10	83	87	95	25963	23812	0.39	0	11.36	28	82	30
B1	58	64	87	4868	3552	19.98	0	127.46	73	92	78
B2	73	84	85	11520	11306	0.67	0	39.74	65	81	77
B3	80	93	85	8885	9704	0.62	0	21.12	73	79	89
B4	78	86	90	2302	2184	0.58	0	15.39	83	92	89
B5	72	72	100	11483	8251	1.01	1	14.7	66	85	74
B6	37	90	38	2068	4867	1.25	1	7.55	13	26	21
B7	27	47	39	6412	7824	2.39	1	19.52	33	54	46
B8	85	92	92	10837	10867	0.32	0	14.14	73	83	86
B9	72	74	96	4417	3433	2.53	0	44.73	75	80	93
B10	82	85	97	22572	19849	0.74	0	66.72	66	85	75

TABLE IV

QUANTITATIVE SCORES FOR ACCURACY, VOLUME, AND SURFACE DISTANCE OF THE AUTOMATICALLY DETECTED TUMOR FOR EACH CASE IN THE DATASET. SET A WAS USED FOR TRAINING AND B FOR TESTING TO COMPUTE THESE RESULTS.

Algorithm	Tumor		Edema	
	Train	Test	Train	Test
Single Voxel Classifier	42%	49%	43%	56%
Saliency-Based Extraction	44%	48%	47%	56%
Conventional Affinity	58%	63%	54%	59%
Model-Based SWA	**69%**	**69%**	**63%**	**62%**
Cross-Validation	68%	55%	65%	54%

TABLE III

SUMMARY (JACCARD OVER ENTIRE SET) VOLUME OVERLAP RESULTS AND TWO-FOLD CROSS-VALIDATION RESULTS (USING MODEL-BASED SWA ON FLIPPED TRAINING AND TESTING SETS).

to each voxel independently. It is clear that even with the strong mixture models, the independent voxel classification rule is not robust to the variations in an individual image. The single voxel classifier is outperformed by the other methods, each incorporating some multilevel information during inference. Essentially, this single voxel classifier approximates the voxel-based statistical classification methods [9]–[11] and (less so) the fuzzy clustering methods [4]–[8]. These results demonstrate the benefit of incorporating multilevel information during inference rather than voxel level information alone.

Intuitively, the capability to use refined affinity functions depending on the model classes should result in a more accurate segmentation with difficult regions being extracted when they would otherwise be missed. To quantitatively demonstrate this intuition, we compare the proposed model-aware affinity method (row labeled "Model-Based SWA") against the conventional affinity and show about 9% improvement in the training sets and 6% in the testing sets. We also show a near 20% improvement when comparing the model-based extraction against the saliency-based extraction. We show a visual comparison of these three methods in Figure 6 for 12 slices of case A1.

Finally, from the scores in Table III, it is also evident that the generative models generalize to the testing set. Two-fold cross-validation also confirm this generalization albeit with slightly worse testing scores. However, as explained next in Table IV and in the next section on failure modes, this comparatively large dataset of GBM tumor contains some peculiarities not entirely captured by our models.

In Table IV, we show a complete set of volumetric and surface accuracy results. The precision is $T/(T+F_n)$, and the recall is $T/(T+F_p)$. We include additional error measurements (volume and surface distance) to facilitate comparison, but note that the relevance of these measurements is questionable in the presence of the gross anisotropic voxels in our dataset, which are typical of diagnostic radiology. The "mean" column contains the average distance (Euclidean, in metric space) from the voxels on the extracted surface to the nearest voxel on the manually labeled surface. Likewise, the "median" column contains the median distance, and the Hausdorff distance is a conservative maxi-min distance.

In most cases, the median distance is 0 indicating that the majority of the voxels on the automatically extracted surface exactly lie on the manually labeled surface. However, it's clear from the scores in the mean and Hausdorff columns that there are some examples with spurious false positives. Some of these cases are elucidated in the next section on failure modes.

3) Failure Modes: Figure 7 shows some examples of the failure modes in our current system. The left two depict the case where a spurious false positive region is classified as tumor. However, the two examples here demonstrate an important detection (B1) and an erroneous one (B10). The detected distant region in B1 is a malignant tumor, but it is not a GBM tumor. Even though our models were not designed to detect this non-GBM tumor, they did. However, in B10, the erroneous detection occurs in the small region of enhancement inside the ventricles. Similar errors occur in A2, A9, and B9

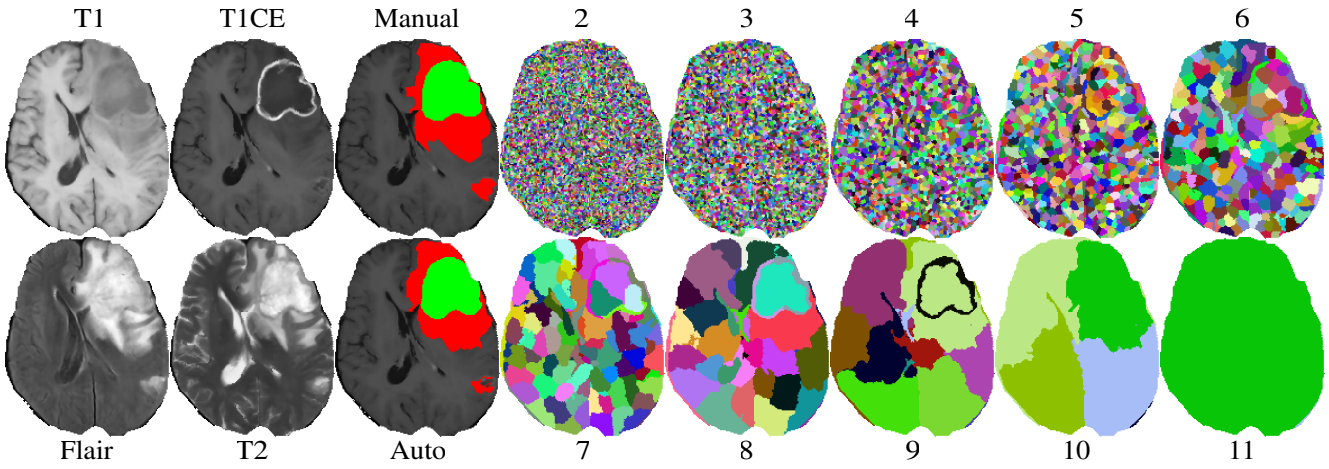

Fig. 5. Example hierarchy on image B8 (from test-set). Levels of the hierarchy (2-11 shown) demonstrate increasingly salient tumor regions being segmented.

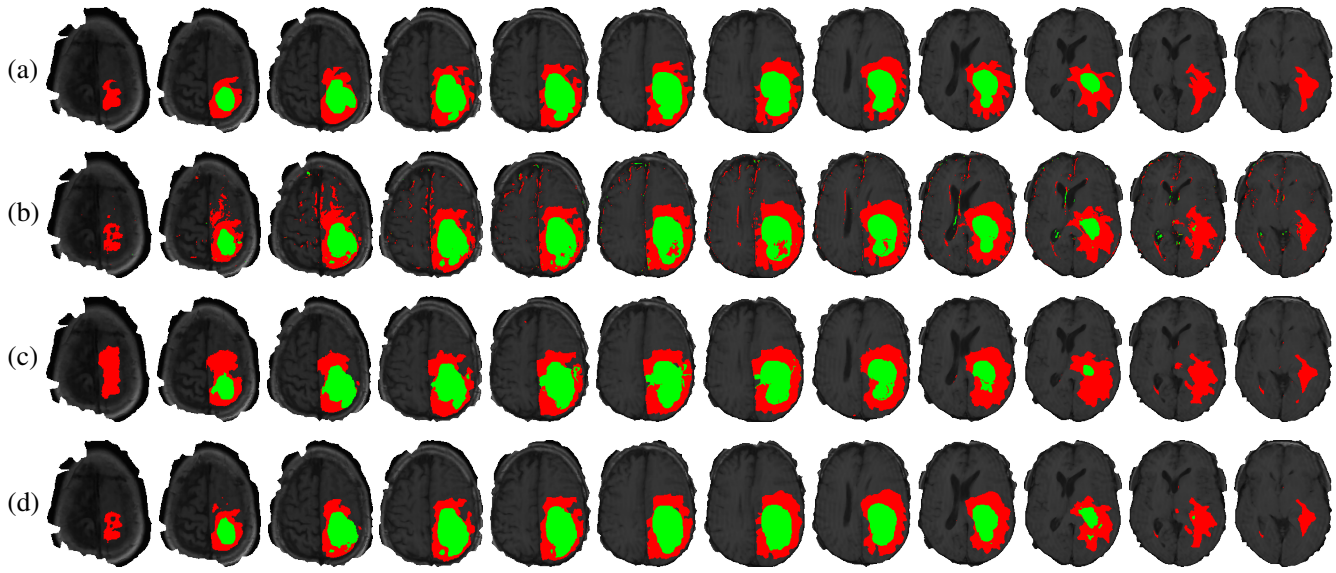

Fig. 6. Classification example on case A1. Each column is a sequential axial slice and each row depicts a different algorithm: (a) manual labeling, (b) single pixel Bayes classifier, (c) saliency based SWA method, and (d) our approach (model based SWA).

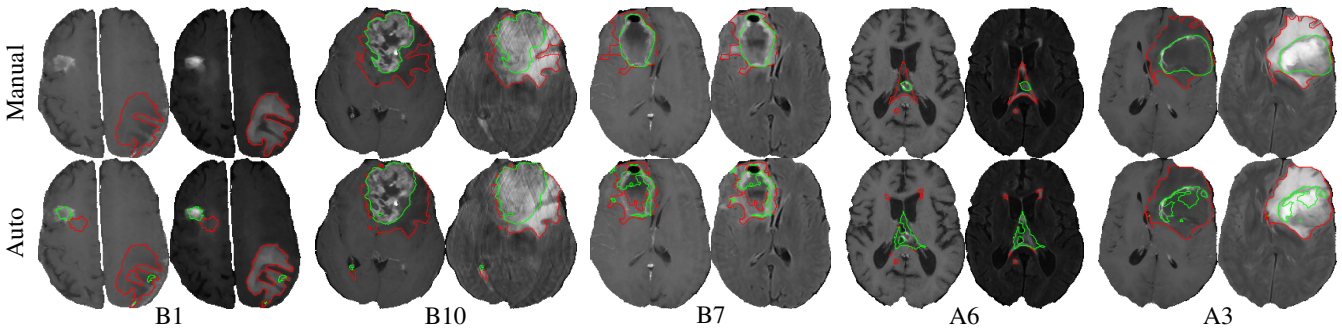

Fig. 7. These five cases represent the different failure modes in our current system. In each group, the left column shows the T1CE channel and the right column shows the FLAIR. Complete discussion in text.

explaining their large Hausdorff distances. Our models classify each region based only on the local multilevel statistics, and do not incorporate any global contextual information or knowledge. To remove errors of this type, we can improve the extraction stage to both incorporate prior knowledge similar to [4], [6], [7] and enforce the global context model.

The next two columns, B7 and A6, represent particularly difficult tumor regions. B7 is the sole case in our set that had a biopsy prior to being imaged and an air-pocket is visibly present in the slice. The disturbed tissue has an anomalous intensity signature. The tumor in A6 is located near the middle of the slice between the two ventricles. The resulting intensity character is quite ambiguous with the nearby ventricles. Similar phenomena occur in A4 and A5. To resolve the ambiguity a context model that incorporates normal anatomy (cortical and sub-cortical structures) could again help in both the classification and extraction stages of the system.

The last column, A3, is the single case in our dataset where the GBM tumor contains a non-contrast-enhancing component. It, thus, classifies a large part of the tumor region as edema. With more examples of this rare phenomenon, our current approach would be able to handle it.

VII. Conclusion

We have made three technical contributions in this paper. The main contribution is the mathematical formulation presented in Section III for bridging graph-based affinities and generative model-based techniques. Second, we extend the SWA algorithm to integrate model-based terms into the affinities during the coarsening. The model-aware affinities integrate classification without making premature hard, class assignments. Using model-specific affinity functions has clear advantages over conventional static affinity methods, both intuitively and justified in the experimental results. The third contribution is a mathematical formulation for learning the parameters of the model-specific affinity functions directly from training data. Furthermore, the algorithm is computationally efficient, running orders of magnitude faster than current state of the art methods.

We apply these techniques to the difficult problem of segmenting and classifying GBM brain tumor in multichannel MR volumes. Our approach improves upon the current state-of-the-art in GBM brain tumor segmentation by incorporating information at multiple scales. The results show good segmentation and classification on a comparatively large dataset. We note that the technical contributions in this paper are general and can be applied to other problems with the proper application-specific models.

We thoroughly analyze the failure modes of our algorithm. While the majority of the cases are segmented with accuracies near 70%, the failure modes will need to be addressed before the method is ready for the clinic, which is the goal. To that end, we have suggested possible solutions to fixing them, and we are developing a global context model of normal brain anatomy (cortical and sub-cortical structures [39]) and brain tumor that will help disambiguate the complex phenomena exhibited in some of the more difficult cases. We are currently investigating stochastic methods to solve the extraction problem by treating the graph hierarchy as a set of model proposals as in Swendsen-Wang sampling [25]. In future work, we will include more complex classification models involving additional feature information (e.g. shape and texture) and models for the appearance of GBM tumor.

Acknowledgments

The authors would like to thank Dr. Cloughesy (Henry E. Singleton Brain Cancer Research Program, University of California, Los Angeles, CA, USA) for the image studies, and the anonymous reviewers for their constructive comments.

References

[1] J. G. Smirniotopoulos, "The new WHO classificaiton of brain tumors." *Neuroimaging Clinics of North America*, vol. 9, no. 4, pp. 595–613, Nov 1999.

[2] M. R. Patel and V. Tse, "Diagnosis and staging of brain tumors," *Seminars in Roentgenology*, vol. 39, no. 3, pp. 347–360, 2004.

[3] J. Liu, J. Udupa, D. Odhner, D. Hackney, and G. Moonis, "A system for brain tumor volume estimation via mr imaging and fuzzy connectedness," *Computerized Medical Imaging and Graphics*, vol. 29, no. 1, pp. 21–34, 2005.

[4] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A brain tumor segmentation framework based on outlier detection," *Medical Image Analysis Journal, Special issue on MICCAI*, vol. 8, no. 3, pp. 275–283, Sep. 2004.

[5] W. E. Phillips, R. P. Velthuizen, S. Phupanich, L. O. Hall, L. P. Clarke, and M. L. Silbiger, "Applications of Fuzzy C-Means Segmentation Technique for Tissue Differentiation in MR Images of a Hemorrhagic Glioblastoma Multiforme," *Journal of Magnetic Resonance Imaging*, vol. 13, no. 2, pp. 277–290, 1995.

[6] M. C. Clark, L. O. Hall, D. B. Goldgof, R. Velthuizen, R. Murtagh, and M. S. Silbiger, "Automatic tumor segmentation using knowledge-based techniques," *IEEE Transactions on Medical Imaging*, vol. 17, no. 2, pp. 187–201, 1998.

[7] L. M. Fletcher-Heath, L. O. Hall, D. B. Goldgof, and F. Reed Murtagh, "Automatic segmentation of non-enhancing brain tumors in magnetic resonance images," *Artificial Intelligence in Medicine*, vol. 21, pp. 43–63, 2001.

[8] N. B. Karayiannis and P. Pai, "Segmentation of Magnetic Resonance Images Using Fuzzy Algorithms for Learning Vector Quantization," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 172–180, 1999.

[9] M. Prastawa, E. Bullitt, N. Moon, K. V. Leemput, and G. Gerig, "Automatic brain tumor segmentation by subject specific modification of atlas priors," *Academic Radiology*, vol. 10, pp. 1341–1348, 2003.

[10] M. Kaus, S. Warfield, A. Nabavi, P. M. Black, F. A. Jolesz, and R. Kikinis, "Automated segmentation of mr images of brain tumors," *Radiology*, vol. 218, pp. 586–591, 2001.

[11] S. K. Warfield, M. R. Kaus, F. A. Jolesz, and R. Kikinis, "Adaptive template moderated spatially varying statistical classification." in *Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*, W. H. Wells, A. Colchester, and S. Delp, Eds. Springer-Verlag, 1998, pp. 431–438.

[12] S. Vinitski, C. F. Gonzalez, R. Knobler, D. Andrews, T. Iwanaga, and M. Curtis, "Fast tissue segmentation based on a 4d feature map in characterization of intracranial lesions Fast Tissue Segmentation Based on a 4D Feature Map in Characterization of Intracranial Lesions," *Journal of Magnetic Resonance Imaging*, vol. 9, no. 6, pp. 768–776, 1999.

[13] S. Ho, E. Bullitt, and G. Gerig, "Level set evolution with region competition: Automatic 3-d segmentation of brain tumors," in *Proceedings of International Conference on Pattern Recognition*, vol. I, 2002, pp. 532–535.

[14] C. H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, "Segmenting brain tumor with conditional random fields and support vector machines," in *Proceedings of Workshop on Computer Vision for Biomedical Image Applications at International Conference on Computer Vision*, 2005.

[15] D. J. Peck, J. P. Windham, L. L. Emery, H. Soltanian-Zadeh, D. O. Hearshen, and T. Mikkelsen, "Cerebral Tumor Volume Calculations Using Planimetric and Eigenimage Analysis," *Medical Physics*, vol. 23, no. 12, pp. 2035–2042, 1996.

[16] Y. Zhu and H. Yan, "Computerized Tumor Boundary Detection Using a Hopfield Neural Network," *IEEE Transactions on Medical Imaging*, vol. 16, no. 1, pp. 55–67, 1997.

[17] J. Zhang, K. Ma, M. H. Er, and V. Chong, "Tumor Segmentation from Magnetic Resonance Imaging By Learning Via One-Class Support Vector Machine," in *International Workshop on Advanced Image Technology*, 2004, pp. 207–211.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society – Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[19] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in *Proceedings of International Conference on Machine Learning*, 2001.

[20] S. Kumar and M. Hebert, "Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification," in *International Conference on Computer Vision*, 2003.

[21] M. Schmidt, I. Levner, R. Greiner, A. Murtha, and A. Bistritz, "Segmenting brain tumors using alignment-based features," in *Proceedings of International Conference on Machine Learning and Applications*, 2005.

[22] J. J. Corso, E. Sharon, and A. Yuille, "Multilevel Segmentation and Integrated Bayesian Model Classification with an Application to Brain Tumor Segmentation," in *Medical Image Computing and Computer Assisted Intervention*, vol. 2, 2006, pp. 790–798.

[23] Z. Tu and S. C. Zhu, "Image Segmentation by Data-Driven Markov Chain Monte Carlo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 657–673, 2002.

[24] Z. Tu, X. R. Chen, A. L. Yuille, and S. C. Zhu, "Image Parsing: Unifying Segmentation, Detection and Recognition," *International Journal of Computer Vision*, 2005.

[25] A. Barbu and S. C. Zhu, "Generalizing Swendsen-Wang to Sampling Arbitrary Posterior Probabilities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1239–1253, 2005.

[26] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[27] Y. Weiss, "Segmentation Using Eigenvectors: A Unifying View," *International Conference on Computer Vision*, vol. 2, pp. 975–982, 1999.

[28] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[29] E. Sharon, M. Galun, D. Sharon, R. Basri, and A. Brandt, "Hierarchy and adaptivity in segmenting visual scenes," *Nature*, vol. 442, no. 7104, pp. 810–813, 2006.

[30] E. Sharon, A. Brandt, and R. Basri, "Fast Multiscale Image Segmentation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. I, 2000, pp. 70–77.

[31] ——, "Segmentation and Boundary Detection Using Multiscale Intensity Measurements," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. I, 2001, pp. 469–476.

[32] A. Akselrod-Ballin, M. Galun, M. J. Gomori, M. Filippi, P. Valsasina, R. Basri, and A. Brandt, "Integrated Segmentation and Classification Approach Applied to Multiple Sclerosis Analysis," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[33] A. Brandt, S. McCormick, and J. Ruge, "Algebraic multigrid (AMG) for automatic multigrid solution with application to geodetic computations," in *Inst. for Computational Studies, POB 1852, Fort Collins, Colorado*, 1982.

[34] S. C. Zhu, "Statistical Modeling and Conceptualization of Visual Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 691–712, 2003.

[35] S. C. Zhu, Y. N. Wu, and D. B. Mumford, "FRAME: Filters, Random field And Maximum Entropy: — Towards a Unified Theory for Texture Modeling," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 1–20, 1998.

[36] F. Han and S. C. Zhu, "Bottom-Up/Top-Down Image Parsing by Attribute Graph Grammar," in *Proceedings of International Conference on Computer Vision*, 2005.

[37] C. Fowlkes, D. Martin, and J. Malik, "Learning affinity functions for image segmentation: combining patch-based and gradient-based approaches," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 54–61.

[38] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. D. Luca, I. Drobnjak, D. E. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. D. Stefano, J. M. Brady, and P. M. Matthews, "Advances in Functional and Structural MR Image Analysis and Implementation as FSL." *NeuroImage*, vol. 23, no. S1, pp. 208–219, 2004.

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