Title
Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification

Authors
Corso, Jason J.
Sharon, Eitan
Dube, Shishir
et al.

Publication Date
2008-02-20

Peer reviewed

Manuscript submitted to IEEE Transactions on Medical Imaging. Do not distribute.

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 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. We consider the GBM tumor because it is the most common primary tumor of the central nervous system, accounting for approximately 40% of brain tumor 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]. In Figure 1, we show a 2D slice of a MR image in the T1 weighted and T2 weighted channels presenting an enhancing *glioblastoma multiforme* brain tumor. On the right, we outline the different heterogeneous regions of the brain tumor and label them as edema, active, or necrotic.

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 where the feature distribution can be extremely complex due to external phenomena like lighting and occlusion.

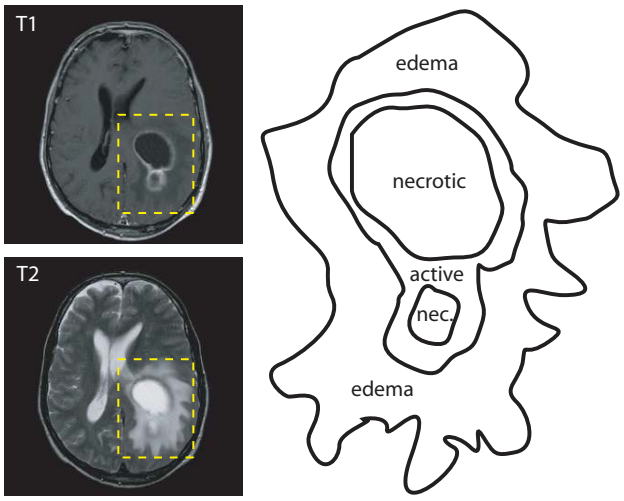

Fig. 1. Labeled example of a brain tumor illustrating the importance of the different modalities (T1 with contrast and T2).

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 tumor.

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 brain tumor is very difficult. In general, it is impossible to segment 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.

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.

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] gave 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 features 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, which 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], they're success is limited in the more relevant GBM (heterogeneous) segmentation examples. Furthermore, it's 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 uses of energy and shape models (e.g., level-sets [13] and active contours [16]) have promise but are generally iterative in nature and therefore sensitive to initialization, which, unless interactive, is nearly as difficult as the entire segmentation.

In this paper, we present a new method for automatic segmentation of heterogeneous image data that is applicable in any case that 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 more generally applicable. Our method combines two of the most effective approaches to segmentation. The first approach, exemplified by the work of Tu et al. [22], [23], 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 [24], but these methods still remain comparatively inefficient.

The second approach is based on the concept of graph cuts [25]. In these affinity-based methods, the input data induces a sparse graph, and each edge in the graph is given

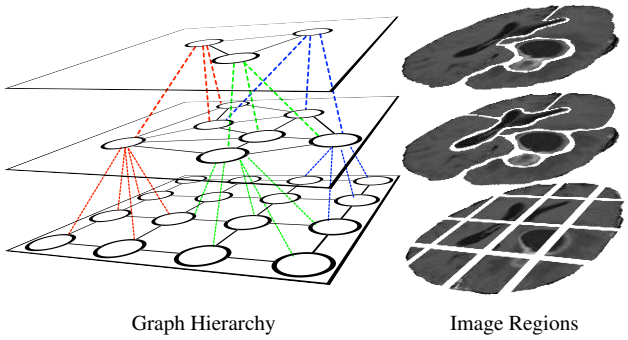

Fig. 2. The SWA algorithm gives a graph hierarchy of potential voxel segments at different scales. This figure shows an explanatory 2D graph hierarchy and the corresponding image region of each lattice element. Only a few interlevel connections are drawn; note how one node can have multiple parents.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

an *affinity* measurement that characterizes the similarity of the two neighboring nodes in some predefined feature space. Cuts are sets of edges that separate the graph into two subsets, typically by analyzing the eigen-spectrum [25], [26] or pairwise-predicate measures [27]. These methods have led to the hierarchical segmentation by weighted aggregation (SWA) algorithm due to Sharon et al. [28]–[30]. SWA was first extended to the 3D image domain by Akselrod-Ballin et al. [31] for the problem of multiple sclerosis segmentation.

SWA operates by recursively coarsening the initial graph using an adapted algebraic multigrid algorithm [32]; it is shown to approximate the normalized cut measure [25]. 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 extremely rapid and effective, but does not explicitly take advantage of the class models used in [22].

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 require 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 **??** we discuss the application of our method to the problem of segmenting brain tumor in multichannel MR. We describe the specific class models and probability functions used in the experimental results.

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 [33]. 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 [34], and stochastic grammars [35].

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 intensty 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 [36].

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 [25]. Eigenvector techniques [26] were originally used to compute the clusters, but, more recently, an efficient multiscale algorithm for doing this was proposed [29], [30] 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 sume 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 $w_{uv} = \exp\left(-D(s_u, s_v; \theta)\right)$, 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 algorithmic 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 **??**, 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. [28]–[30], 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 $\mathcal{G} = \{G^t \colon t = 0, \ldots, T\}$. Denote the multichannel intensity vector at voxel i as $I(i) \in \mathbb{R}^C$, with C being the number of channels.

A. Original 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 $D(s_u, s_v; \theta) \doteq \theta \left| s_u - s_v \right|_1$. 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:

$$W_{UV} = \hat{w}_{UV} \exp\left(-D(s_U, s_V; \theta)\right) \quad . \tag{10}$$

   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 $|\mathcal{E}^t| = 0$.

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. [29] 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.

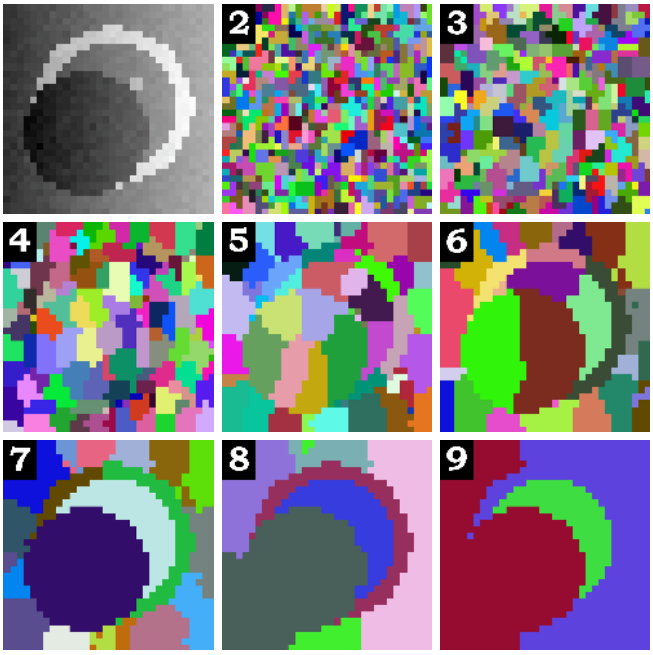

Fig. 3. Example SWA hierarchy on a synthetic grayscale image. The numbers indicate the level in the hierarchy (0 would be the pixels).

In the example, we see that each of the salient foreground objects in the image are correctly segmented at some level in the hierarchy. However, being objects of different scale, they are not detected at the same level. This is another example of the important phenomena we mentioned in the introduction and in Figure 2. Sharon et al. [29], [30] 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 **??**, 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 give an example in Section **??** (Figure **??**) showing the added power of integrating models into the affinity calculation in the case of heterogeneous data like tumor and edema.

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 dicussed 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 [29], [30]. Second, we present a novel 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 **??**), 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 [29]. 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. An extraction algorithm of this sort is especially suited to the i.i.d. generative models we use in this paper.

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] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

[30] ——, "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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.