

Multilevel Segmentation and Integrated Bayesian Model Classification with an Application to Brain Tumor Segmentation

Jason J. Corso¹, Eitan Sharon², and Alan Yuille²

¹ Medical Imaging Informatics, University of California, Los Angeles, CA, USA,
jcorso@mii.ucla.edu

² Department of Statistics, University of California, Los Angeles, CA, USA

Abstract. We present a new method for automatic segmentation of heterogeneous image data, which is very common in medical image analysis. The main contribution of the paper is a mathematical formulation for incorporating soft model assignments into the calculation of affinities, which are traditionally model free. We integrate the resulting model-aware affinities into the multilevel *segmentation by weighted aggregation* algorithm. We apply the technique to the task of detecting and segmenting brain tumor and edema in multimodal MR volumes. Our results indicate the benefit of incorporating model-aware affinities into the segmentation process for the difficult case of brain tumor.

1 Introduction

Medical image analysis typically involves image data that has been generated from heterogeneous underlying physical processes. Segmenting this data into coherent regions corresponding to different anatomical structures is a core problem with many practical applications. For example, it is important to automatically detect and classify brain tumors and measure properties such as tumor volume, which is a key indicator of tumor progression [1]. But automatic segmentation of brain tumors is very difficult. 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 general, it is impossible to segment tumor by simple thresholding techniques [3].

In this paper, we present a new method for automatic segmentation of heterogeneous image data and apply it to the task of detecting and describing brain tumors. The input is multi-modal data since different modes give different cues for the presence of tumors, edema (swelling), and other relevant structure. For example, the T2 weighted MR modality contains more pertinent information for segmenting the edema (swelling) than the T1 weighted modality.

Our method combines two of the most effective approaches to segmentation. The first approach, exemplified by the work of Tu and Zhu [4], uses class models to explicitly represent the different heterogeneous processes. The tasks of segmentation and class membership estimation are solved jointly. This type of approach is very powerful but the algorithms for obtaining these estimates are comparatively slow. The second approach is based on the concept of normalized cuts [5] and has led to the segmentation by weighted aggregation (SWA)

algorithm due to Sharon et al. [6, 7]. SWA was first extended to the 3D image domain by Akselrod-Ballin et al. [8]. In their work, multiple modalities are used during the segmentation to extract multiscale segments that are then classified in a decision tree algorithm; it is applied to multiple sclerosis analysis. SWA is extremely rapid and effective, but does not explicitly take advantage of the class models used in [4].

In this paper, we first describe (§2) how we unify these two disparate approaches to segmentation by incorporating Bayesian model classification into the calculation of affinities in a principled manner. This leads (§3) to a modification of the SWA algorithm using class-based probabilities. In §4 we discuss the specific class models and probability functions used in the experimental results.

2 Integrating Models and Affinities

In this section we discuss the main contribution of the paper: a natural way of integrating model classes into affinity measurements without making premature class assignments to the nodes.

Let $G = (\mathcal{V}, \mathcal{E})$ be a graph with nodes $u, v \in \mathcal{V}$. Each node in the graph is augmented with properties, or statistics, denoted $s_v \in \mathcal{S}$, where \mathcal{S} is the space of properties (e.g., \mathbb{R}^3 for standard red-green-blue data). Each node also has an (unknown) class label $c_u \in \mathcal{C}$, where \mathcal{C} is the space of class labels. The edges are sparse with each node being connected to its nearest neighbors. 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 $w_{uv} = \exp(-D(s_u, s_v; \theta))$ where D is a non-negative *distance* measure and θ are predetermined parameters.

The goal of affinity methods is to detect regions $R \subset \mathcal{V}$ with low saliency defined by

$$\Gamma(R) = \frac{\sum_{u \in R, v \notin R} w_{uv}}{\sum_{u, v \in R} w_{uv}} . \quad (1)$$

Such regions have low affinity across their boundaries and high affinity within their interior. An efficient multiscale algorithm for doing this is described in §3.

By contrast, the model based methods define a likelihood function $P(\{s_u\}|\{c_u\})$ for the probability of the observed statistics/measurements $\{s_u\}$ conditioned on the class labels $\{c_u\}$ of the pixels $\{u\}$, and put prior probabilities $P(\{c_u\})$ for the class labels of pixels. Their goal is to seek estimates of the class labels by maximizing the posterior probability $P(\{c_u\}|\{s_u\}) \propto P(\{s_u\}|\{c_u\})P(\{c_u\})$. In this paper, we restrict ourselves to a simple model where $P(s_u|c_u)$ is the conditional probability of the statistics s_u at a node u with class c_u , and $P(c_u, c_v)$ is the prior probability of class labels c_u and c_v at nodes u and v .

We combine the two approaches by defining class dependent affinities in terms of probabilities. The affinity between nodes $u, v \in \mathcal{V}$ is defined to be the probability $\hat{P}(X_{uv}|s_u, s_v) = w_{uv}$ of the binary event X_{uv} that the two nodes lie in the same region. This probability is calculated by treating the class labels as

hidden variables which are summed out:

$$\begin{aligned}
P(X_{uv}|s_u, s_v) &= \sum_{c_u} \sum_{c_v} P(X_{uv}|s_u, s_v, c_u, c_v)P(c_u, c_v|s_u, s_v) , \\
&\propto \sum_{c_u} \sum_{c_v} P(X_{uv}|s_u, s_v, c_u, c_v)P(s_u, s_v|c_u, c_v)P(c_u, c_v) , \\
&= \sum_{c_u} \sum_{c_v} P(X_{uv}|s_u, s_v, c_u, c_v)P(s_u|c_u)P(s_v|c_v)P(c_u, c_v) , \quad (2)
\end{aligned}$$

where the third line follows from the assumption that the nodes are conditionally independent given class assignments. The first term of (2) is the model-aware affinity calculation:

$$P(X_{uv}|s_u, s_v, c_u, c_v) = \exp\left(-D(s_u, s_v; \theta[c_u, c_v])\right) . \quad (3)$$

Note that the property of belonging to the same region X_{uv} is not uniquely determined by the class variables c_u, c_v . Pixels with the same class may lie in different regions and pixels with different class labels might lie in the same region.

This definition of affinity is suitable for heterogeneous data since the affinities are explicitly modified by the evidence $P(s_u|c_u)$ 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 algorithmic description in §3. Individual nodes, at the bottom of the pyramid, will typically only have weak evidence for class membership (i.e., $p(c_u|s_u)$ is roughly constant). 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 (§3). In §4, we discuss the specific forms of these probabilities used in our experiments.

3 Segmentation By Weighted Aggregation

We now review the segmentation by weighted aggregation (SWA) algorithm of Sharon et al. [6, 7], and describe our extension to integrate class-aware affinities. As earlier, define a graph $G^t = (\mathcal{V}, \mathcal{E})$ with the additional superscript indicating the level in a pyramid of graphs $\mathcal{G} = \{G^t: t = 0, \dots, T\}$. Denote the multimodal intensity vector at voxel i as $I(i) \in \mathbb{R}^M$, with M being the number of modalities.

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 §2 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 $\forall u \in \mathcal{V}^t$
 $\sum_{v \in \mathcal{R}^t} w_{uv} \geq \beta \sum_{v \in \mathcal{V}^t} w_{uv}$.
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 $p_{uU} = \frac{w_{uU}}{\sum_{V \in \mathcal{V}^{t+1}} w_{uV}}$, with $u \in \mathcal{V}^t$ and $U \in \mathcal{V}^{t+1}$.
 - (c) Accumulate statistics to coarse level: $s_U = \sum_{u \in \mathcal{V}^t} \frac{p_{uU} s_u}{\sum_{v \in \mathcal{V}^t} p_{vU}}$.
 - (d) Interpolate affinity from the finer level: $\hat{w}_{UV} = \sum_{(u \neq v) \in \mathcal{V}^t} p_{uU} w_{uv} p_{vU}$.
 - (e) Use coarse affinity to modulate the interpolated affinity:

$$W_{UV} = \hat{w}_{UV} \exp(-D(s_U, s_V; \theta)) . \quad (4)$$

- (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). [6] shows that this algorithm preserves the saliency function (1).

Incorporating Class-Aware Affinities. The two terms in (4) 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 resulting in more accurate, class-aware affinities:

$$W_{UV} = \hat{w}_{UV} P(X_{UV} | s_U, s_V) , \quad (5)$$

where $P(X_{UV} | s_U, s_V)$ is evaluated as in (2). We give an example in §4 (Figure 4) showing the added power of integrating class knowledge into the affinity calculation in the case of heterogeneous data like tumor and edema. Furthermore, the class-aware affinities compute model likelihood terms, $P(s_U | c_U)$, as a by-product. Thus, we can also associate a most likely class with each node in the graph: $c_U^* = \arg \max_{c \in \mathcal{C}} P(s_U | c)$.

4 Application to Brain Tumor Segmentation

Related Work in Brain Tumor Segmentation. Segmentation of brain tumor has been widely studied in medical imaging. Here, we review some major approaches to the problem; [1, 3] contain more complete literature reviews. Fletcher-Heath et al. [9] take a fuzzy clustering approach to the segmentation followed by 3D connected components to build the tumor shape. Liu et al. [1] take a similar fuzzy clustering approach in an interactive segmentation system that is shown to give an accurate estimate of tumor volume. Kaus et al. [10] use the *adaptive template-moderated classification* algorithm to segment the MR image into five different tissue classes: background, skin, brain, ventricles, and

tumor. Prastawa et al. [3] present a detection/segmentation algorithm based on learning voxel-intensity distributions for normal brain matter and detecting outlier voxels, which are considered tumor. Lee et al. [11] use the recent discriminative random fields coupled with support vector machines to perform the segmentation.

Data Processing. We work with a dataset of 30 manually annotated glioblastoma multiforme (GBM) [2] brain tumor studies. Using FSL tools [12], we pre-processed the data through the following pipeline: (1) spatial registration, (2) noise removal, (3) skull removal, and (4) intensity standardization. The 3D data is 256×256 with 24 slices. We use the T1 weighted pre and post-contrast modality, and the FLAIR modality in our experimentation.

Class Models and Feature Statistics. We model four classes: non-data (outside of head), brain matter, tumor, and edema. Each class is modeled by a Gaussian distribution with full-covariance giving 9 parameters, which are learned from the annotated data. The node-class likelihoods $P(s_u|c_u)$ are computed directly against this Gaussian model. Manual inspection of the class histograms indicates there are some non-Gaussian characteristics in the data, which will be modeled in future work.

The class prior term, $P(c_u, c_v)$, 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. For the model-aware affinity term (3), we use a class dependent weighted distance:

$$P(X_{uv}|s_u, s_v, c_u, c_v) = \exp\left(-\sum_{m=1}^M \theta_{c_u c_v}^m |s_u^m - s_v^m|\right), \quad (6)$$

where superscript m indicates vector element at index m . The class dependent coefficients are set by hand using expert domain knowledge. They are presented in Table 1 (we abbreviate non-data, ND). For a given class-pair, the modalities that do not present any relevant phenomena are set to zero and the remaining are set uniformly according to the principle of maximum entropy. For example, edema is only detectable in the FLAIR channel, and thus at the Brain-Edema boundary, only the FLAIR channel is used. We will experiment with techniques to automatically compute the coefficients in future work.

m \ c_u, c_v	ND, ND	ND, Brain	Brain, Tumor	Brain, Edema	Tumor, Edema
T1	$\frac{1}{3}$	0	0	0	0
T1 with Contrast	$\frac{1}{3}$	$\frac{1}{2}$	1	0	$\frac{1}{2}$
FLAIR	$\frac{1}{3}$	$\frac{1}{2}$	0	1	$\frac{1}{2}$

Table 1. Coefficients used in model-aware affinity calculation. Rows are modality and columns are (symmetric) class pairs.

Note that the coefficients are symmetric (i.e., equal for Brain, Tumor and Tumor, Brain), and we have included only those pairs not excluded by the hard constraints discussed above. The sole feature statistic that we accumulate for each node in the graph (SWA step 3c) is the average intensity. 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.

Extracting Tumor and Edema from the Pyramid. To extract tumor and edema from the pyramid, we use the class likelihood function that is computed during the Bayesian model-aware affinity calculation (2). For each voxel, we compute its most likely class (i.e., c_u^* from §3) at each level in the pyramid. Node memberships for coarser levels up the pyramid are calculated by the SWA interpolation weights (p_{uU}). Finally, the voxel is labeled as the class for which it is a member at the greatest number of levels in the pyramid. We also experimented with using the saliency function (1) as recommended in [6], but we found the class likelihood method proposed here to give better results.

Results. We computed the Jaccard score (ratio of true pos. to true pos. plus false pos. and neg.) for the segmentation and classification (S+C) on a subset of the data. The subset was chosen based on the absence of necrosis; the Gaussian class models cannot account for the bimodality of the statistics with necrosis present. The Jaccard score for tumor is 0.85 ± 0.04 and for edema is 0.75 ± 0.08 . From manual inspection of the graph pyramid, we find that both the tumor and edema are segmented and classified correctly in most cases with some false positives resulting from the simple class models used.

On page 8, we show three examples of the S+C to exemplify different aspects of the algorithm. For space reasons, we show a single, indicative slice from each volume. Each example has a pair of images to compare the ground truth S+C against the automatic S+C; green represents tumor and red represents edema. Figure 2 shows the pyramid overlaid on the T1 with contrast image. Figure 3 demonstrates the importance of using multiple modalities in the problem of brain tumor segmentation. On the top row, the pyramid is overlaid on the T1 with contrast modality and on the bottom row, it is overlaid on the FLAIR modality. We easily see (from the left most column) that the tumor and edema phenomena present quite different signals in these two modalities. Yet, the two phenomena are accurately segmented; in Figure 1, we show the class labels for the layer in the third column. E indicates edema, T indicates tumor, and any non-labeled segment is classified a brain. Figure 4 shows a comparison between using (rows 1 and 2) and not using (row 3) class-aware affinities. We see that with our extension, the segmentation algorithm is able to more accurately delineate the complex edema structure.

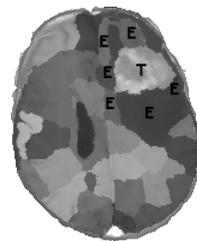


Fig. 1. Class labels overlaid on a slice layer.

5 Conclusion

We have made two contributions in this paper. The main contribution is the mathematical formulation presented in §2 for bridging graph-based affinities and model-based techniques. In addition, we extended 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. We applied these techniques to the difficult problem of segmenting and classifying brain tumor in multimodal MR volumes. The results showed good segmentation and classification and demonstrated the benefit of including the model information during segmentation.

The model-aware affinities are a principled approach to incorporate model information into rapid bottom-up algorithms. In future work, we will extend this work by including more complex statistical models (as in [4]) involving additional feature information (e.g. shape) and models for the appearance of GBM tumor. We will further extend the work to automatically learn the model-aware affinity coefficients instead of manually choosing them from domain knowledge.

Acknowledgements

Corso is funded by NLM Grant # LM07356. 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, Shishir Dube and Usha Sinha.

References

1. Liu, J., Udupa, J., Odhner, D., Hackney, D., Moonis, G.: A system for brain tumor volume estimation via mr imaging and fuzzy connectedness. *Computerized Medical Imaging and Graphics* **29**(1) (2005) 21–34
2. Patel, M.R., Tse, V.: Diagnosis and staging of brain tumors. *Seminars in Roentgenology* **39**(3) (2004) 347–360
3. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis Journal, Special issue on MICCAI* **8**(3) (2004) 275–283
4. Tu, Z., Zhu, S.C.: Image segmentation by data-driven markov chain monte carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5) (2002) 657–673
5. Shi, J., Malik, J.: Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8) (2000) 888–905
6. Sharon, E., Brandt, A., Basri, R.: Fast multiscale image segmentation. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Volume I.* (2000) 70–77
7. Sharon, E., Brandt, A., Basri, R.: Segmentation and boundary detection using multiscale intensity measurements. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Volume I.* (2001) 469–476
8. Akselrod-Ballin, A., Galun, M., Gomori, M.J., Filippi, M., Valsasina, P., Basri, R., Brandt, A.: Integrated segmentation and classification approach applied to multiple sclerosis analysis. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.* (2006)
9. Fletcher-Heath, L.M., Hall, L.O., Goldgof, D.B., Reed Murtagh, F.: Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial Intelligence in Medicine* **21** (2001) 43–63
10. Kaus, M., Warfield, S., Nabavi, A., Black, P.M., Jolesz, F.A., Kikinis, R.: Automated segmentation of mr images of brain tumors. *Radiology* **218** (2001) 586–591
11. Lee, C.H., Schmidt, M., Murtha, A., Bistriz, A., Sander, J., Greiner, R.: 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)
12. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., Luca, M.D., Drobnjak, I., Flitney, D.E., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J.M., Matthews, P.M.: Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage* **23**(S1) (2004) 208–219

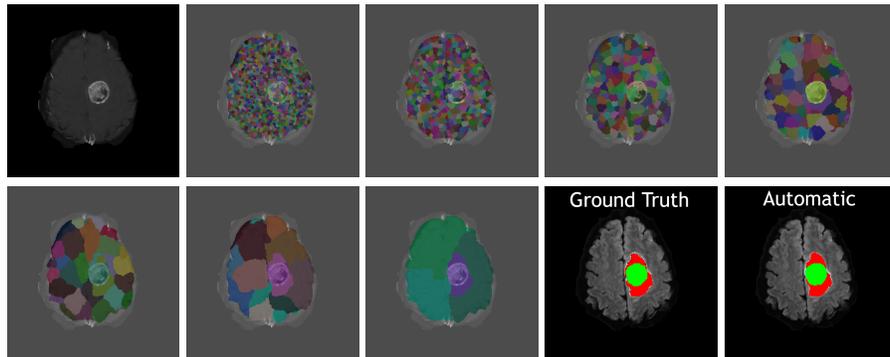


Fig. 2. Example of tumor detection and segmentation in post-contrast T1 weighted.

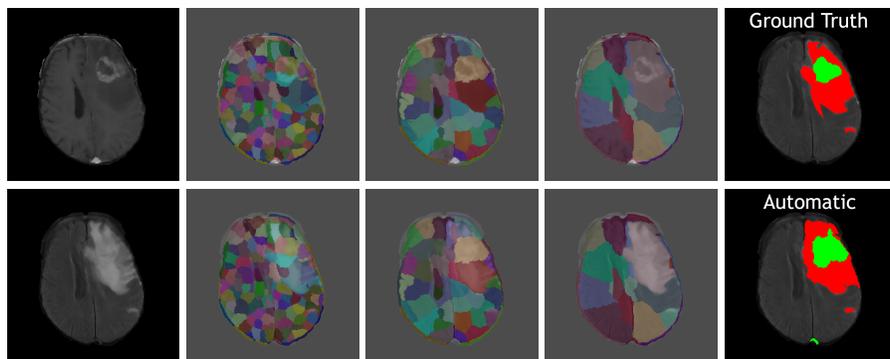


Fig. 3. Example segmentation demonstrating importance of multiple modalities.

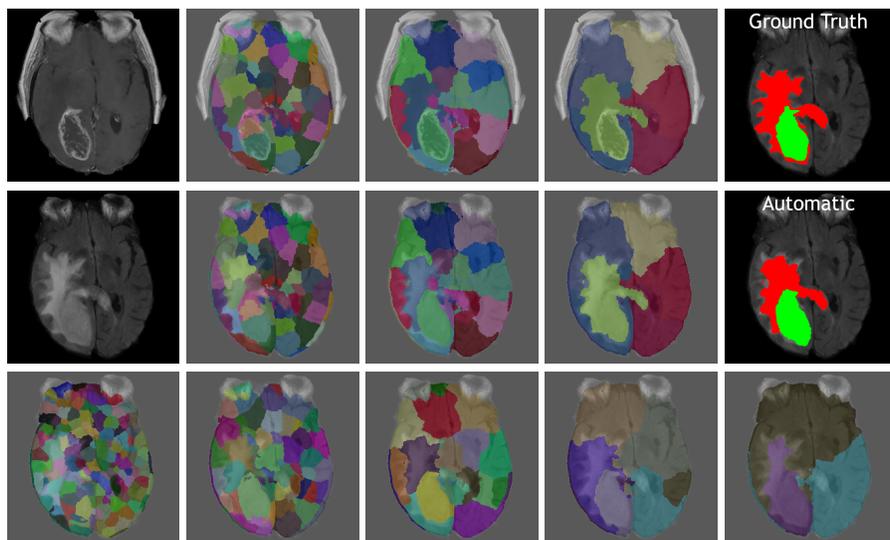


Fig. 4. Example segmentation comparing the integration of class information into the affinity calculation (rows 1 and 2, T1 post-contrast and FLAIR, resp.) versus no class information (row 3). See §4 for a full explanation of the figures on this page.