

4D ACTIVE CUT: AN INTERACTIVE TOOL FOR PATHOLOGICAL ANATOMY MODELING

Bo Wang^{†,‡}, Wei Liu^{†,‡}, Marcel Prastawa^{†,‡}, Andrei Irimia[§],
Paul M. Vespa[¶], John D. van Horn[§], P. Thomas Fletcher^{†,‡}, Guido Gerig^{†,‡,*}

[†] Scientific Computing and Imaging Institute,
[‡] School of Computing, University of Utah

[§] Institute for Neuroimaging and Informatics, USC
[¶] Brain Injury Research Center, UCLA

ABSTRACT

4D pathological anatomy modeling is key to understanding complex pathological brain images. It is a challenging problem due to the difficulties in detecting multiple appearing and disappearing lesions across time points and estimating dynamic changes and deformations between them. We propose a novel semi-supervised method, called *4D active cut* for lesion recognition and deformation estimation. Existing interactive segmentation methods passively waits for user to refine the segmentations which is a difficult task in 3D images that change over time. *4D active cut* instead actively selects candidate regions for querying the user, and therefore obtains the most informative user feedback. A user simply answers ‘yes’ or ‘no’ to a candidate object without having to refine the segmentation slice by slice. Compared to single-object detection of GrabCut, our method also detects multiple lesions with spatial coherence using Markov random fields constraints. Results show improvement on the lesion detection, which subsequently improves deformation estimation.

Index Terms— Active learning, graph cuts, longitudinal MRI, Markov Random Fields, semi-supervised learning,

1. INTRODUCTION

Quantitative studies in longitudinal pathology such as traumatic brain injury (TBI), autism, and Huntington’s disease are important for measuring treatment efficacy or making predictions. The modeling of 4D pathological anatomy is essential to understand the complex dynamics of pathologies and enables other analysis such as structural pathology and brain connectivity [1]. This is a challenging task because of the difficulties in localizing multiple lesions at specific time points and estimating deformations between time points. In effect, this involves solving interdependent segmentation and registration problems. In the analysis of magnetic resonance (MR) images of severe TBI patients, more difficulties arise such as multiple lesions with complex shapes, multi-modal information, and lack of prior knowledge of lesion location.

Graph cuts based user-interactive image segmentation methods have been applied to 2D and 3D imaging data [2, 3]. In particular, the GrabCut method [3] integrates human experts’ knowledge and the algorithm’s computation power. In GrabCut, one typically gives an initial guess of the region of interest (foreground), for example a bounding box. The algorithm estimates globally optimal boundary between the foreground and background regions. The user then inspects the segmentation results, and adjusts the boundary by drawing a region as a hard constraint. Such process works well on 2D images, as users can quickly find incorrectly classified regions. However, such interaction is a huge burden to users once applied to 3D volume data, since one has to visually inspect each slice and correct it. The algorithm is in a *passive learning* state, and its performance entirely depends on user’s active inspection and correction. This passive learning process is the bottleneck of the GrabCut algorithm when applied to 3D data.

We propose *4D active cut*, a method that uses active learning for guiding interactive segmentation of 4D pathological images, specifically longitudinal TBI data. We adopt a minimalist approach such that our algorithm needs the least amount of user involvement. Using active learning, the algorithm queries the user only on the most important regions. Accordingly, the user’s response to such query will be the most informative, and the number of user interaction is minimized. Our algorithm learns the model from a simple user initialization, finds the candidate objects and submits them to a user for inspection. The user is now in a passive state, and are only required to answer a ‘yes’ or ‘no’ question for the queried candidates. The algorithm does the remaining work including refining the candidate objects, learning model parameters and running graph cut algorithm to find a globally optimal map. In addition, with sufficient confidence, the algorithm automatically identifies additional lesion objects without user interaction by self-training. Self-training further decreases user involvement, without losing segmentation accuracy. Our algorithm can detect multiple lesions, while standard graph cut algorithm [3] only detects single objects. Moreover, we introduce spatial context constraints using Markov random fields (MRF) so that the estimated label maps will become piecewise constant. We also define

* Supported by grants: National Alliance for Medical Image Computing (NA-MIC) U54 EB005149 (GG) and the Utah Science Technology and Research (USTAR) initiative at the University of Utah.

the MRF prior on the candidate objects submitted to the user, so the candidates are spatially coherent objects instead of individual voxels.

Several researchers have applied active learning to 3D image segmentation. Top et al. [4] proposed to query users the most uncertain 2D slices. Iglesias et al. [5] used active learning for constructing manually labeled dataset for training supervised classifiers. Veeraraghavan et al. [6] used grow cut for segmentation and estimated the uncertainty using a support vector machine classifier. Our method is different from them in that 1) it solves both segmentation and registration in a unified 4D framework, and 2) instead of learning 2D slices or individual voxels, it learns new 3D objects that better fit human visual perception due to their spatial coherence.

In the remaining part of the paper, we discuss the model and interaction framework in section 2, give the 4D modeling results in section 3 and conclude in section 4.

2. METHOD

An overview of our algorithm is shown in Fig. 1. At each time point, the algorithm takes a bounding box as the initial user input, learns the model parameters, estimates graph cuts for detecting the lesions, and queries a user by submitting one or more candidate objects. The 3D label information from all time points are then integrated by fitting a 4D model that describes the changes across time as deformations.

2.1. Extension to GrabCut

We extend GrabCut [2, 3] to include hierarchical smoothness constraints. We build two GMM models for the normal brain regions and the lesions, and use the expectation maximization (EM) method to estimate both the class labels and the model parameters θ . In the graph cut step, the algorithm takes model parameters of both GMMs as input, and estimates a hard label $\alpha \in \{0, 1\}$ representing if each voxel n is normal ($\alpha_n = 0$) or lesions ($\alpha_n = 1$). In addition to smoothness constraint on the lesion boundary, we also apply MRF constraints on the label maps within normal and lesion regions. These soft constraints guarantee that the estimated labels are spatially coherent.

Within-Class MRF and Variational Inference: Define the class label map $\mathbf{z} = \{z_1, \dots, z_N\}$, where z_n is a K dimensional binary random variable for voxel n such that $\sum_k z_{nk} = 1$. An element $z_{nk} = 1$ indicates voxel n is in Gaussian component k . The prior distribution of \mathbf{z} takes the MRF constraints into account as well as the atlas, and is defined as

$$p(\mathbf{z}) = \frac{1}{C} \exp\left(\sum_{n=1}^N \sum_{k=1}^K z_{nk} \log \pi_{nk} + \beta \sum_{(n,m)} \langle z_n, z_m \rangle\right),$$

where π_{nk} is the affine-registered atlas prior probability of n being in component k , (n, m) is a pair of neighboring voxels, and $\langle \cdot, \cdot \rangle$ is the inner product of two vectors, representing our preference of a piecewise constant label map, with

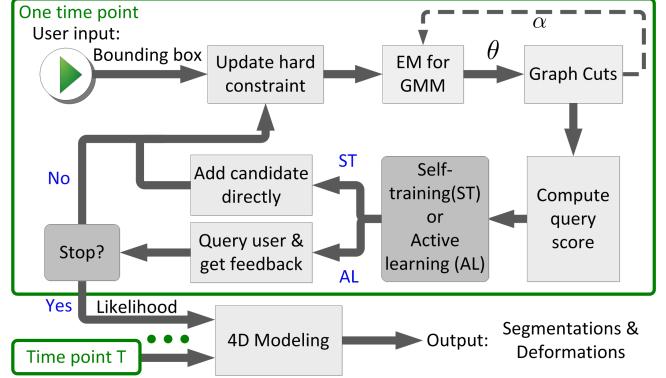


Fig. 1. Flowchart of the proposed algorithm.

the constraint strength given by β . Given z_n , the likelihood function is defined as a multivariate Gaussian $p(x_n | z_n) = \mathcal{N}(x_n; \theta(z_n))$, with x_n being the observed data vector of the multi-modality images at n . In EM, one needs to evaluate $\mathbb{E}_{P(\mathbf{z}|\mathbf{x})}[\log p(\mathbf{x}, \mathbf{z})]$, which is intractable due to the MRF prior on \mathbf{z} . Here we use the variational inference method that iteratively updates $p(z_n | z_{\mathcal{N}(n)}, \mathbf{x}_n)$ given the expected value of z_n 's neighbors $\mathcal{N}(n)$. The update equation takes the form

$$\log p(z_{nk}) = z_{nk} \pi_{nk} + \sum_{m \in \mathcal{N}(n)} \langle \bar{z}_m, z_n \rangle + \log \mathcal{N}(x_n; \theta(z_n)),$$

where \bar{z}_m is the expected value at neighbor m . We compute $\log p(z_{nk})$ for all k and compute $p(z_n)$ by taking the exponential and normalize. \bar{z}_n is just $p(z_n)$ for binary variables and is used for updating z_n 's neighbors. In the M step, we use \mathbf{z} to estimate $\theta = \{\mu, \Sigma\}$ for all components. The variational procedure stops when there are no changes of z_n . Given α , EM runs on both GMMs separately, with a uniform atlas map on lesion's GMM.

2.2. Guided User Interaction Using Active Learning

The algorithm conducts active learning by taking α and θ as input, and computing the probability of each voxel belonging to the lesions. It then builds a connected component map, since a high-level object-based representation is convenient for user interaction. We sort the multiple objects in a descending order of the probability of being lesions, and submit the top ranked objects for queries to the user. When the algorithm detects that the top ranked object should be lesion with a confidence above a user-given threshold, it adds the object to lesion in a self-training process without querying users, further reducing user involvement.

Query Score: The log-odds of α_n is defined as

$$a_n = \log p(\alpha_n = 1) + \mathbb{E}_{p(z_n | x_n)}[\log p(x_n, z_n; \theta(\alpha_n = 1))] - \log p(\alpha_n = 0) + \mathbb{E}_{p(z_n | x_n)}[\log p(x_n, z_n; \theta(\alpha_n = 0))],$$

where \mathbb{E} is the expectation, α is a MRF in the form of $p(\alpha) = (1/C_\alpha) \exp(\eta \sum_{(\alpha_m, \alpha_n)} \psi(\alpha_m, \alpha_n))$ to model its spatial coherence, with $\psi = 1$ if $\alpha_m = \alpha_n$, and $\psi = 0$ otherwise. The predictive probability of a voxel being in lesion is computed by the standard logistic function $p(\alpha_n = 1 | \mathbf{x}_n) = 1/(1 + \exp(-a_n))$ once a is estimated by the variational inference. We obtain a binary map w by thresholding the predictive map at 0.5, and identify a series of objects R_i by running a connected component detection on w . To further select the most salient objects, we sort the objects in descending order of the following score:

$$q(R_i) = \left(\sum_{n \in R_i} p(\alpha = 1 | \mathbf{x}_n) \right) / |\{n : n \in \mathcal{B}(R_i)\}|. \quad (1)$$

$\mathcal{B}(R_i)$ is the set of voxels on R_i 's boundary, and the denominator denotes the number of voxels on the boundary of R_i .

The above query score prefers objects with larger volumes of posterior probability. The score also prefers blob-like objects since such an object has large volume-surface ratio. Such criteria reflects our prior knowledge on the lesion object's shape.

4D Modeling Pathological Anatomy: We integrate the 3D information across time by taking \mathbf{z} and θ for both normal and lesion classes as input, and estimate the atlas prior π by computing the deformation from the healthy template to the images at each time point. We follow the 4D pathological anatomy modeling framework of Wang et al. [7], and define $\pi_{k,t} = A_k \circ \phi_t + Q_{k,t}$, where A is the tissue class probability that is initially associated with the healthy template, ϕ_t is the diffeomorphic deformation from time t to the atlas, and Q_t is the non-diffeomorphic probabilistic change. We use alternating gradient descent to estimate A , ϕ_t and Q_t by optimizing $\mathcal{F}(A, \phi_t, Q_t) = -\sum_{t=1}^T \mathbb{E}_{p(\mathbf{z}|\mathbf{x})} [\log p(\mathbf{z}, \mathbf{x} | \theta, \pi_t)]$.

Interaction Framework: With the above framework, the learning process is described in Fig. 1. Given the user defined bounding box, the hard-constrain map is set to 'normal' outside the box, and 'uncertain' inside. Then, we initialize the α map such that $\alpha = 1$ (lesion) in the 'uncertain' region of the hard-constrain, and $\alpha = 0$ elsewhere. The EM learns θ of both GMMs given α , then a graph-cut step [3] updates only the α for voxels within the 'uncertain' region. This update strategy of graph-cuts prevents false-positive detections. Upon EM & graph cuts convergence, there may be some false-negative voxels in 'normal' regions. We group these voxels into spatially coherent objects and find the one with highest score computed from (1). The algorithm either automatically adds the candidate to lesions or queries user for answer, depending on its confidence on the candidate. We then update the hard-constraint map to reflect the knowledge learned from the new object, and a new EM & graph cuts iteration starts. Objects already accepted or rejected in previous steps are recorded so they will be excluded from future queries. This learning process repeats until user stops the algorithm. Finally, we integrate information in the longitudinal datasets in all time points by fitting a 4D pathological anatomy model.

3. RESULTS

We applied our method to a dataset of longitudinal multi-modal MR images of four severe TBI patients. Each patient was scanned twice: an acute scan at about 5 days and a chronic scan at about 6 months post injury. Each subject's data includes T1, T2, FLAIR, and GRE. In lesions' GMM model, we chose $K = 2$ for bleeding and edema component, and in normal regions we chose $K = 3$ for gray matter, white matter and CSF. We took a 6 neighborhood system and set the smoothness parameters $\gamma = \beta = \eta = 1$. The image intensity of each modality was normalized to the range [0, 255] before the learning process. We used the K-Means clustering together with the atlas prior to initialize the GMM. The threshold used to decide self-training or active-learning was set to 3.0.

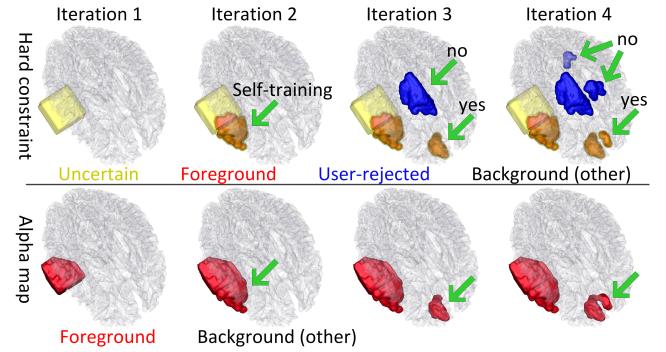


Fig. 2. Illustration of the iterative process of 4D active cut on subject I. Iteration 1: User-initialized bounding box. Iteration 2: Self training. Iteration 3 and 4: Active learning. The user stopped the process at 5th iteration. Arrows point to the changes between iterations. The white matter surface is visualized in all figures as reference.

Fig. 2 shows the dynamic learning process of subject I. In iteration 1, given the user initialized bounding box, 4D active cut successfully detected the lesion object, as shown in the bottom α map. In iteration 2, the candidate object has a large query score, so the algorithm decided to do self training and added the candidate to the lesion class. This object is indeed part of the lesion in the previous bounding box. That shows when the user leaves a portion of the lesion outside the box, the algorithm still detects the missing portion. This result shows the robustness of the algorithm given inaccurate user initialization. In iteration 3 and 4, the user rejected some candidate objects and accepted some others. The accepted objects are not connected to the major lesion but our algorithm still captured them. This result shows that our method is able to detect multiple objects. We also note that by using object volume, predictive probability and shape information, most of the top-ranked candidates are indeed true positive ones; showing the effectiveness of the proposed query score.

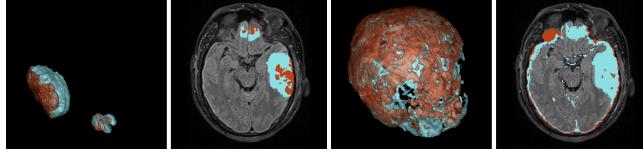


Fig. 3. Comparison between GrabCut and the proposed method. Left: our method in 3D space and axial view. Right: GrabCut. Light blue: edema. Brown: bleeding. Note the large false-positive detection of GrabCut in the CSF region.

Subject	Baseline		4D active cut		
	NHL	HL	NHL	HL	UI
I	0.2503	0.0613	0.6349	0.5698	5
II	0.3269	-	0.6910	-	4
III	0.1311	0.2288	0.4512	0.4840	6
IV	0.0918	0.0998	0.3503	0.1153	5

Table 1. Dice values comparing *4D active cut* and GrabCut to ground truth. HL and NHL are acute hemorrhagic and non-hemorrhagic lesions. UI denotes the number of interactions a user performed using *4D active cut*. Subject II has no ground truth for HL due to the lack of GRE modality.

In order to show the efficacy of the proposed method, we used GrabCut without user interaction as a baseline method and also allowed α outside of the bounding box to switch labels in order to detect multiple lesion objects. Fig. 3 shows the qualitative comparison of the baseline method and *4D active cut*. Without user interaction, the baseline algorithm detected a large number of false-positive voxels due to the ambiguity between lesion and CSF. Table 1 shows the quantitative comparison of both methods. The proposed method is able to significantly improve the segmentation.

As part of *4D active cut*, we updated the atlas prior using a 4D modeling method [7]. Fig. 4 shows the parcellation labels mapped to the space of each time point by using the estimated deformation field. The result shows the obtained parcellation maps match the data well, even in the presence of large deformations at different time points. Therefore, our integrated framework has the potential of becoming an important processing step for connectivity analysis of pathological brain with longitudinal data [1], where the mapping of parcellation labels to individual time points in the presence of large pathologies presents the biggest obstacle and currently requires tedious manual corrections and masking.

4. CONCLUSIONS

We presented *4D active cut* for quantitative modeling of pathological anatomy. The new algorithm can detect multiple lesion objects with minimal user input. The MRF prior ensures the spatially coherent label maps within class and within candidates. In the future, we will explore integration

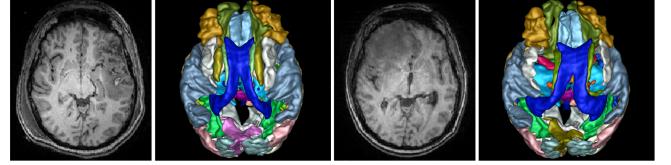


Fig. 4. Result of mapping parcellation labels associated with the healthy template to each time point. The left two images are the T1 reference TBI image at acute stage and mapped parcellations, and the right two images are the same at chronic stage.

of active learning and 4D modeling, as well as the validation and verification on other image data presenting pathologies.

5. REFERENCES

- [1] A. Irimia, B. Wang, S.R. Aylward, M.W. Prastawa, D.F. Pace, G. Gerig, D.A. Hovda, R. Kikinis, P.M. Vespa, and J.D. Van Horn, “Neuroimaging of structural pathology and connectomics in traumatic brain injury: Toward personalized outcome prediction,” *NeuroImage: Clinical*, vol. 1, no. 1, pp. 1–17, 2012.
- [2] Yuri Boykov, Olga Veksler, and Ramin Zabih, “Fast approximate energy minimization via graph cuts,” *IEEE PAMI*, vol. 23, no. 11, pp. 1222–1239, 2001.
- [3] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” in *ACM TOG*, 2004, vol. 23, pp. 309–314.
- [4] Andrew Top, Ghassan Hamarneh, and Rafeef Abugharbieh, “Active learning for interactive 3D image segmentation,” in *MICCAI 2011*, pp. 603–610. Springer, 2011.
- [5] Juan Eugenio Iglesias, Ender Konukoglu, Albert Montillo, Zhuowen Tu, and Antonio Criminisi, “Combining generative and discriminative models for semantic segmentation of CT scans via active learning,” in *IPMI*. Springer, 2011, pp. 25–36.
- [6] Harini Veeraraghavan and James V Miller, “Active learning guided interactions for consistent image segmentation with reduced user interactions,” in *IEEE ISBI*, 2011, pp. 1645–1648.
- [7] Bo Wang, Marcel Prastawa, Avishek Saha, Suyash P Awate, Andrei Irimia, Micah C Chambers, Paul M Vespa, John D Van Horn, Valerio Pascucci, and Guido Gerig, “Modeling 4D changes in pathological anatomy using domain adaptation: Analysis of TBI imaging using a tumor database,” in *MBIA*. 2013, pp. 31–39, Springer.