

White Matter Lesion Segmentation Using Machine Learning and Weakly Labeled MR Images

Yuchen Xie[†] Xiaodong Tao^{‡*}

[†] Dept. of CISE, University of Florida, Gainesville, Florida 32611 USA

Email: yxie @ cise.ufl.edu

[‡] GE Global Research Center, One Research Circle, Niskayuna, NY 12309 USA

Email: taox @ ge.com

ABSTRACT

We propose a fast, learning-based algorithm for segmenting white matter (WM) lesions for magnetic resonance (MR) brain images. The inputs to the algorithm are T1, T2, and FLAIR images. Unlike most of the previously reported learning-based algorithms, which treat expert labeled lesion map as ground truth in the training step, the proposed algorithm only requires the user to provide a few regions of interest (ROI's) containing lesions. An unsupervised clustering algorithm is applied to segment these ROI's into areas. Based on the assumption that lesion voxels have higher intensity on FLAIR image, areas corresponding to lesions are identified and their probability distributions in T1, T2, and FLAIR images are computed. The lesion segmentation in 3D is done by using the probability distributions to generate a confidence map of lesion and applying a graph based segmentation algorithm to label lesion voxels. The initial lesion label is used to further refine the probability distribution estimation for the final lesion segmentation. The advantages of the proposed algorithm are: 1. By using the weak labels, we reduced the dependency of the segmentation performance on the expert discrimination of lesion voxels in the training samples; 2. The training can be done using labels generated by users with only general knowledge of brain anatomy and image characteristics of WM lesion, instead of these carefully labeled by experienced radiologists; 3. The algorithm is fast enough to make interactive segmentation possible. We test the algorithm on nine ACCORD-MIND MRI datasets. Experimental results show that our algorithm agrees well with expert labels and outperforms a support vector machine based WM lesion segmentation algorithm.

Keywords: Segmentation, Machine Learning, White Matter Lesion, Region of Interest, Weak Label

1. INTRODUCTION

White matter lesions (WMLs) are commonly found in healthy elders and in patients with multiple sclerosis (MS), cerebrovascular disease (CVD), stroke, and other neurological disorders. It is believed that the total volume of the lesions (lesion load) and their progression relate to the aging

^{*}This work is partially supported by Grant R01 EB006733 from the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, through the University of North Carolina, Chapel Hill. All findings, opinions, and recommendations expressed in this paper are these of the authors and do not constitute these of the University of North Carolina, Chapel Hill or the National Institutes of Health.

This work was done while the first author was a summer intern at GE Global Research Center.

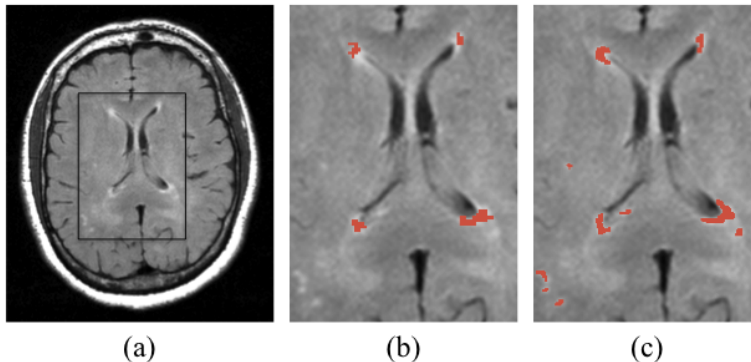


Figure 1. Inter-rater variability. (a) One axial slice of a FLAIR image. (b) and (c) Manual lesion segmentation by two raters.

process as well as disease process. Therefore, segmentation and quantification of white matter lesions is very important in understanding the aging process and diagnosis and assessment of these diseases.^{1,2}

Manual segmentation of WM lesions, which is still of routine use in clinic practices,³ is not only time consuming but also shows high inconsistency among human raters. In recent years, a number of computer-assisted lesion segmentation methods have been developed using segmentation techniques and machine learning algorithms. One approach developed by Leemput¹ uses an unsupervised tissue clustering algorithm with anatomical model of the brain to segment multi-spectral MR brain images into different tissue types. The outliers of this model are considered lesions. Some popular supervised classifiers such as SVM and Adaboost have also been used to segment WM lesions.² Those supervised tissue classification methods, however, suffer from two problems. First, because of the large variability in image appearance between different datasets, the classifiers need to be re-trained from each data source to achieve good performances. Second, this type of algorithms rely on manually labeled training datasets to compute the multi-spectral intensity distribution of the lesions. High intra- and inter-rater variability makes the manual labeled training dataset unreliable. For example, one typical training image from MICCAI 2008 MS Lesion Segmentation Grand Challenge³ is shown in Figure 1. The lesions labeled by two raters are very different and the DICE similarity between these two segmentation results is only 0.28.

Inspired by interactive segmentation algorithms in computer vision, we propose an efficient, semi-automatic WM lesion segmentation method with simple user input—a few rectangular ROI’s loosely around lesion regions (Figure 2 (a)). In order to distinguish these rectangular ROI’s from lesion ROI’s manually labeled by human raters, where the precise boundary of the lesions are delineated, we call these rectangular ROI’s *weak labels*. In our experiments, four such ROI’s are chosen on one axial slice for each of the datasets. Lesions in the entire 3D volume are segmented using statistics learned from these weak labels. The described methodology is validated in multi-channel MRI sequences against expert labels and state-of-art SVM based WM lesion segmentation algorithm.²

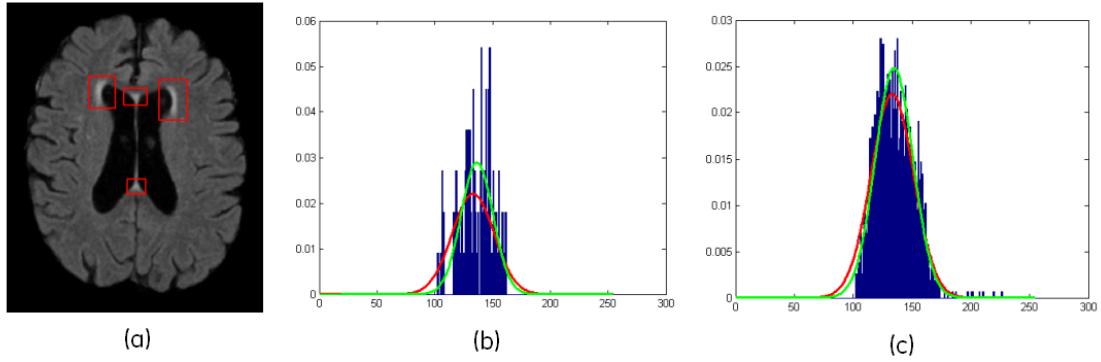


Figure 2. (a) ROI's provided by user. (b) Histograms for lesion voxels inside the 2D ROI's. (c) Refined estimation of histogram of lesion voxels in 3D. The green curves show the fitted normal distribution and the red curves show the distribution computed from manually labeled voxels.

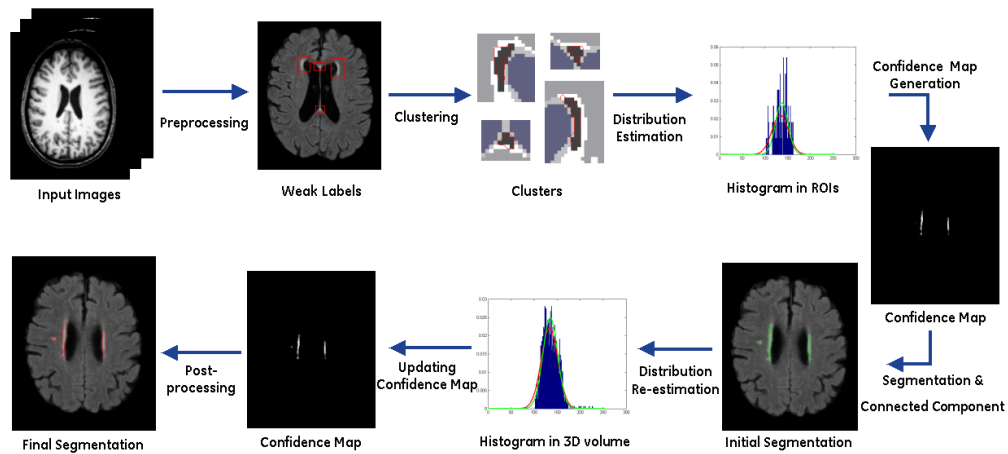


Figure 3. Overview of our white matter lesion segmentation pipeline.

2. METHODS

Figure 3 summarizes the pipeline of our WM lesion segmentation process. The inputs to the algorithm include T1, T2 and FLAIR images. During the pre-processing steps, images with different contrast are registered into a common space of T1 weighted images. The registration is limited to rigid transform with mutual information as image similarity measures. Skull stripping is performed on T1 weighted image.⁴ The resulting brain mask is applied to remove extracranial tissue on the other images. Tissue classification with inhomogeneity correction is also applied to the T1 weighted image to generate probability map for white matter.⁵ These pre-processing steps are common to most of other WM lesion segmentation algorithms.^{1,2}

After the user provides weak labels including lesion voxels, we first perform a k-means clustering in these rectangular ROIs using all image channels. Since lesions have high intensity in FLAIR image, the cluster with highest FLAIR intensity is considered as lesions. Because the k-means algorithm may only find a local optimum and cannot guarantee the approximation accuracy, we use the initialization approach proposed in k-means++,⁶ which is $\Theta(\log k)$ -competitive with the optimal clustering. We then compute the histograms for each channel in lesion cluster and approximate them using normal distribution. Figure 2 (b) shows the histogram (FLAIR) of lesion cluster in the rectangular ROIs as blue bars, estimated lesion distribution from lesion cluster (green curve), and the distribution computed from manually labeled lesion in ROIs (red curve). The estimated distribution from lesion cluster is used to generate a confidence map by computing a p-value for each voxel. Since FLAIR intensity is the most discriminative feature in lesion segmentation,⁷ we mainly consider the estimated distribution in FLAIR to compute p-value and use distributions in other channels for cleanup. If the confidence value is bigger than a threshold, the voxel is classified as lesion. This threshold is automatically decided as follows. Suppose the intensity distribution in ROIs could be modeled by mixture of Gaussians:

$$p(x) = \sum_{i=1}^k w_i N(x|\mu_i, \sigma_i), \quad (1)$$

where $0 \leq w_i \leq 1$ and $\sum_i w_i = 1$. $p(x)$ could also be considered as the mixture of lesion distribution $p_l(x)$ and background distribution $p_b(x)$: $p(x) = w_l p_l(x) + w_b p_b(x)$, where $0 \leq w_l, w_b \leq 1$ and $w_l + w_b = 1$. Since we already model the lesion distribution as Gaussian, the background distribution will be a mixture of Gaussians. Thus it is reasonable to take the p-value at the intersection of lesion and background curves as the segmentation threshold (Figure 4).

So far, we only consider intensity values in the segmentation process. Next, we apply the spatial context around each detected lesion mass to detect false positives. If the average WM probability of a lesion region’s neighbors is lower than some threshold (we use 0.3 in our experiments), we consider it as a false positive and remove it from the lesion map. One example is illustrated in Figure 5. Some false detections that have similar intensity values as lesion regions are successfully cleaned using the spatial information of white matter.

The next step is to refine lesion segmentation using grow cut image segmentation algorithm.⁸ In essence, any other segmentation algorithms such as Graph Cut⁹ and random walks¹⁰ can be used. We chose grow cut algorithm because of its ease of implementation. The lesion voxels are set as foreground and the voxels with confidence value smaller than $q = 0.001$ are set as background. The result of this step is a lesion segmentation in the entire 3D volume. However, since the bounding boxes drawn by the user are not enough to represent the entire 3D volume, some lesion regions may

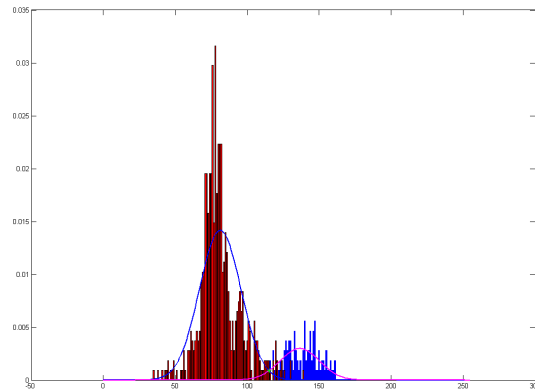


Figure 4. Intersection of lesion and background distributions in ROIs. Blue bars and red bars represent lesion and background histograms. The blue curve represents the estimated background distribution and the magenta curve represents the estimated lesion distribution. Their intersection is labeled using green star.

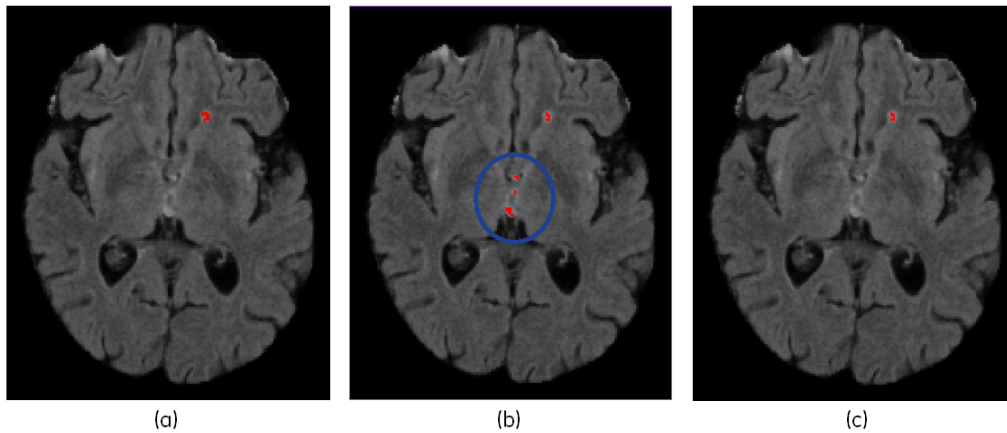


Figure 5. Cleanup using white matter probability map. (a) Manual label. (b) Segmentation result before cleanup. Some false detections are shown in blue circle. (c) Segmentation result after cleanup.

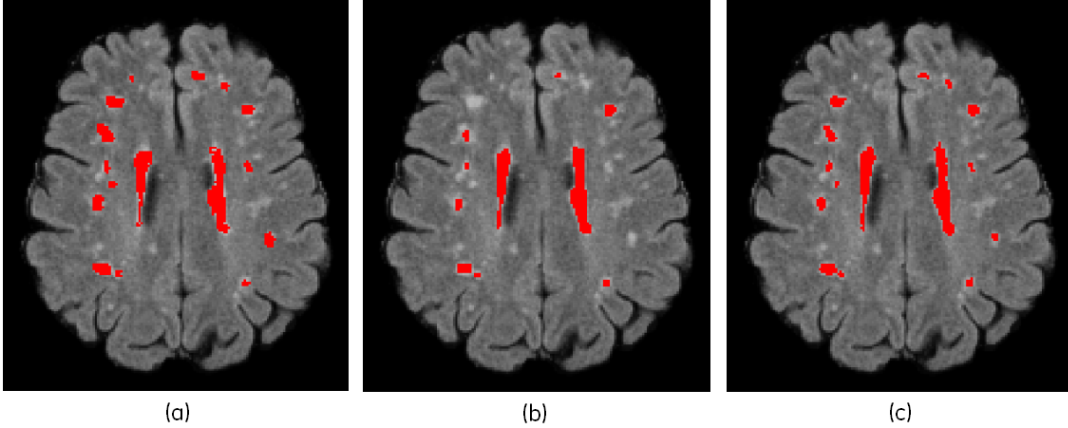


Figure 6. (a) Manual label. (b) Segmentation result using lesion distribution estimated from 2D ROIs. (c) Final segmentation result using the refined estimation of intensity distribution of the lesion voxels from the entire 3D volume.

be missed in the initial segmentation as shown in Figure 6 (b). Therefore, based on this initial lesion segmentation in 3D, we recompute the histograms and re-estimate the probability distribution of lesion voxels for each channel. Figure 2 (c) shows the refined distribution (magenta curve), which agrees better with the lesion distribution computed from manually labelled lesion voxels (red curve). The final lesion segmentation result is obtained by a confidence map computed from the refined distribution. In Figure 6 (c), we could see that the final segmentation contains most lesion regions missed in the initial segmentation and agrees well with the manual labels.

3. RESULTS

We applied the proposed WM lesion segmentation algorithm to 9 subjects from ACCORD-MIND MRI dataset (<http://www.accordtrial.org/>). For all cases, we manually drawn four weak labels on one axial slice. Two typical results of manually segmented lesions and the result of our algorithm are shown in Figure 7. We also compare the performance of our method with an SVM based WM algorithm² using an implementation available as a 3DSlicer extension^{11,12}. This method employs a two-fold cross validation for ACCORD dataset. The DICE scores ($DSC = \frac{2|Seg \cap Ref|}{|Seg| + |Ref|}$), True Positive ($TP = \frac{|Seg \cap Ref|}{|Ref|}$) and False Positive ($FP = \frac{|Seg| - |Seg \cap Ref|}{|Seg|}$) of segmentation results are tabulated in Table 1.

Overall, the DICE score of the proposed algorithm is markedly better than the SVM based algorithm, indicating a better agreement between the computer-assisted lesion segmentation and manual segmentation by human raters. In the proposed algorithm, although the true positive is sacrificed (from an average of 0.82 in SVM based algorithm to an average of 0.70), the overall consistency across subject, as measured by the standard deviations are significantly improved. These results demonstrated the reliability of the proposed algorithm.

For all the cases, the algorithm takes about 10 seconds on an average PC using non-optimized C++ code. We also implement our algorithm as a module in 3DSlicer and make it feasible as an

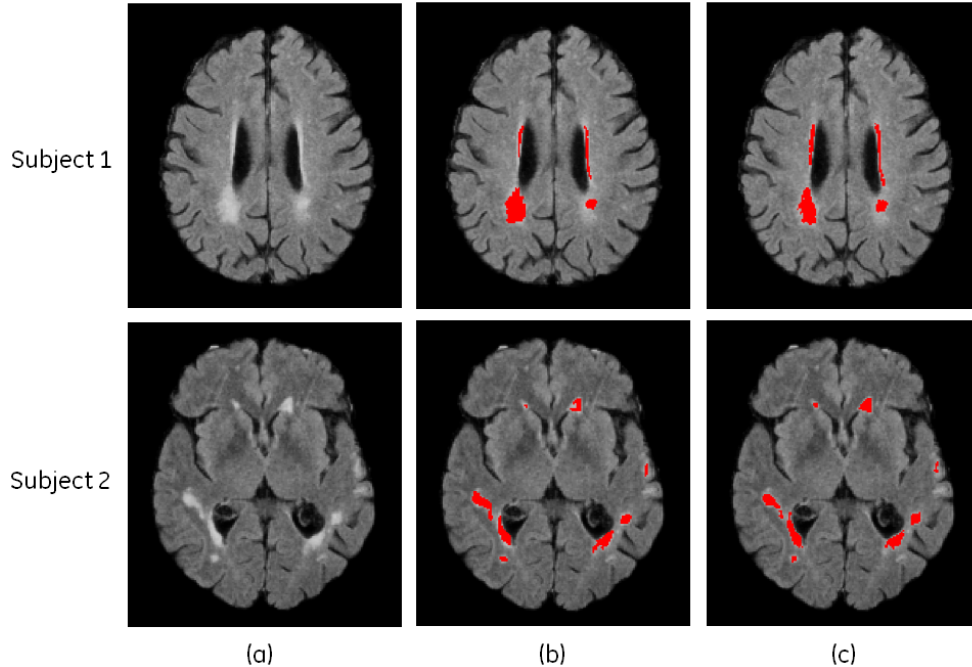


Figure 7. Comparison of manually segmented lesions and the results of our method for two subjects. (a) One axial slice of the FLAIR image; (b) Manual segmentation; (c) Results of the proposed algorithm.

interactive segmentation tool.

4. CONCLUSION

We proposed a new learning-based algorithm for segmenting WM lesions in MR images. In addition to the input T1, T2, and FLAIR image, the user provides a few rectangular ROIs around lesion regions. Our method effectively learns lesion’s appearance information and segments lesions in the entire 3D volume. Our method is demonstrated on ACCORD-MIND MRI dataset with an average DICE score of 0.705, which is higher than the average DSC of the state-of-art SVM based lesion segmentation method.

We note that the proposed algorithm is semi-automated. It requires as input a few weak labels containing lesion voxels. Although the full automation is sacrificed, we gained significantly more consistency in the lesion segmentation result because the learning is conducted on the same patient images presumably acquired on the same scanner during the same imaging session. In the proposed algorithm, the generation of the training examples is very simple comparing with these required for most other learning based lesion segmentation algorithms, where the precise boundaries of lesion need to be delineated by highly skilled and experienced experts.

In the future, we would like to perform more thorough tests with larger database from different data sources. Understanding sensitivity of the algorithm to the placement of weak labels is also important and will be investigated in the future.

Table 1. Comparison of our method with SVM based lesion segmentation algorithm.

Patients	SVM based method			Our method		
	DSC	TP	FP	DSC	TP	FP
ACCORD01	0.47	0.91	0.68	0.69	0.63	0.24
ACCORD02	0.63	0.93	0.52	0.75	0.79	0.28
ACCORD03	0.71	0.75	0.33	0.76	0.79	0.27
ACCORD04	0.61	0.58	0.34	0.68	0.74	0.36
ACCORD05	0.41	0.92	0.73	0.63	0.63	0.36
ACCORD06	0.60	0.74	0.50	0.74	0.81	0.31
ACCORD07	0.64	0.69	0.40	0.70	0.69	0.30
ACCORD08	0.34	0.95	0.80	0.76	0.83	0.31
ACCORD09	0.32	0.88	0.80	0.64	0.77	0.45
Average	0.53	0.82	0.57	0.705	0.741	0.32
StdDev	0.14	0.13	0.19	0.05	0.08	0.06
Max	0.71	0.95	0.80	0.76	0.83	0.45
Min	0.32	0.58	0.33	0.63	0.63	0.24

ACKNOWLEDGEMENT

The authors would like to thank Drs. Dinggang Shen, Jim Miller, and Minjeong Kim for many helpful discussions.

REFERENCES

- [1] Leemput, K. V., Maes, F., Vandermeulen, D., Colchester, A., and Suetens, P., “Automated segmentation of multiple sclerosis lesions by model outlier detection,” *IEEE Trans. on Medical Imaging* **20**, 677–688 (2001).
- [2] Lao, Z., Shen, D., Liu, D., Jawad, A., Melhem, E., Launer, L., Bryan, R., and Davatzikos, C., “Computer-assisted segmentation of white matter lesions in 3d mr images using support vector machine,” *Academic Radiology* **15**, 300–313 (2008).
- [3] Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., and Warfield, S., “3d segmentation in the clinic: A grand challenge ii: Ms lesion segmentation,” *MIDAS Journal*, 1–5 (2008).
- [4] Tao, X. and Chang, M.-C., “A skull stripping method using deformable surface and tissue classification,” in [*Medical Imaging: Image Processing*], *Proc. SPIE* **7623** (2010).
- [5] Chang, M.-C. and Tao, X., “Subvoxel segmentation and representation of brain cortex using fuzzy clustering and gradient vector diffusion,” in [*Medical Imaging: Image Processing*], *Proc. SPIE* **7623** (2010).
- [6] Arthur, D. and Vassilvitskii, S., “k-means++: The advantages of careful seeding,” in [*SODA*], (2007).
- [7] Geremia, E., Menze, B., Clatz, O., Konukoglu, E., Criminisi, A., and Ayache, N., “Spatial decision forests for ms lesion segmentation in multi-channel mr images,” in [*MICCAI*], (2010).
- [8] Vezhnevets, V. and Konouchine, V., “Grow-cut - interactive multi-label n-d image segmentation,” in [*Proc. Graphicon*], 150–156 (2005).

- [9] Boykov, Y. and Jolly, M.-P., “Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images,” in [*ICCV*], 105–112 (2001).
- [10] Grady, L., “Random walks for image segmentation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence* **28**, 1768–1783 (2006).
- [11] Pieper, S., Lorensen, B., Schroeder, W., and Kikinis, R., “The na-mic kit: Itk, vtk, pipelines, grids and 3d slicer as an open platform for the medical image computing community,” in [*ISBI 2006*], 698–701 (2006).
- [12] Shen, D., Wu, G., Miller, J., Kim, M., and Tao, X., “Hammer and wml modules for 3d slicer.” <http://www.nitrc.org/projects/hammerwml/>.