3D Segmentation in CT Imagery with Conditional Random Fields and Histograms of Oriented Gradients

Chetan Bhole[1], Nicholas Morsillo[1], and Christopher Pal[2]

[1] University of Rochester
[2] École Polytechnique de Montréal

Abstract. In this paper we focus on the problem of 3D segmention in volumetric computed tomography imagery to identify organs in the abdomen. We propose and evaluate different models and modeling strategies for 3D segmentation based on traditional Markov Random Fields (MRFs) and their discriminative counterparts known as Conditional Random Fields (CRFs). We also evaluate the utility of features based on histograms of oriented gradients or HOG features. CRFs and HOG features have independently produced state of the art performance in many other problem domains and we believe our work is the first to combine them and use them for medical image segmentation. We construct 3D lattice MRFs and CRFs, use variational message passing (VMP) for learning and max-product (MP) inference for prediction in the models. These inference and learning approaches allow us to learn pairwise terms in random fields that are non-submodular and are thus very flexible. We focus our experiments on abdominal organ and region segmentation, but our general approach should be useful in other settings. We evaluate our approach on a larger set of anatomical structures found within a publicly available liver database and we provide these labels for the dataset to the community for future research.

Keywords: MRF, CRF, generative, discriminative, 3D segmentation, HOG

1 Introduction

Accurate detection and segmentation of organs, vessels and other regions of interest is a key problem in medical imaging. Markov random fields (MRFs) provide an attractive framework for image segmentation. Recent insights into modeling techniques have given rise to a new distinction between traditional MRFs which define a joint distribution over both segmentation classes and features and conditional random fields (CRFs) which model the conditional distributions of the segmentation field directly. CRFs have had a major impact in machine learning in recent years and our work here explores their application to 3D image segmentation in detail. Image descriptors based on histograms of oriented image gradients or HOG features have received a lot of attention in computer vision recently. For example, both the widely used SIFT descriptors of Lowe [10] and the person detector of Dalal et al. [4] are based on different forms of HOG. Depending on the size and spatial extent of the HOG, such features are capable of capturing the entire shape of small organs or contour segments of larger organs. In our work here, we propose and explore a HOG based feature descriptor specifically designed for detecting

anatomical structures and contours in CT imagery of the abdomen. To visualize some key elements of our approach, we show an example of a preprocessed axial image slice, a manual segmentation into organs of interest and some pre-processing steps in Fig. 1. We model image volumes using the 3D graphical models as shown in Fig. 2. Each 2D lattice corresponds to an axial slice of the medical image.

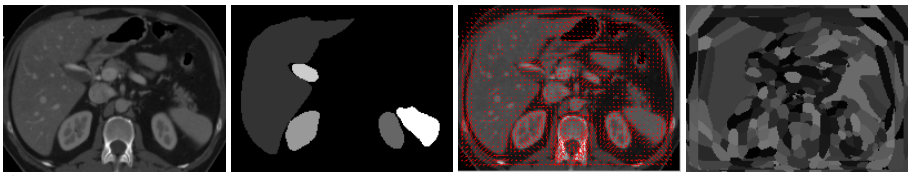

Fig. 1: Left to right: A coarsely registered axial data slice and its manual segmentation. Each shade denotes a different class label. The third image is a visualization of the HOG features (zoom for clarity). Arrows indicate gradient orientations and magnitude. Only prominent gradient bins are shown to reduce clutter. The image on the right is the classification using HOG codewords.

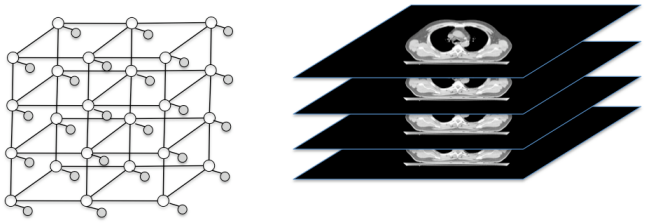

Fig. 2: The figure shows a graphical model modeling the image volume. The shaded nodes denote the observed feature node variables associated with a pixel of an image. The unshaded nodes are the segmentation variables representing the anatomy classes.

Markov Random Fields have a long history in image processing and computer vision; however, they started to receive more intense attention in the medical image analysis around the time of Zhang, Brady and Smith's influential work [16] on hidden Markov Random Fields for segmenting Magentic Resonance (MR) imagery of the brain. They created a traditional MRF in the sense that in that they created a likelihood term for each pixel and a spatially coupled prior in the form of a random field. One of their contributions was to propose and explore the use of Gaussian mixture models for the local likelihood terms, thus introducing hidden variables and producing a hidden MRF. They then used the Expectation Maximization (EM) algorithm to perform unsupervised learning in the hidden MRF for segmentation. While exploring the problem of general interactive image segmentation Blake et al. [1] used a similar Gaussian mixture model based random field (GMMRF) approach along with a small amount of training data. This allowed one to learn MRF models in a supervised manner; however, the model still had the form of a traditional MRF defining the joint probability of features and segmentation class labels. They however do not learn pairwise term parameters. Lafferty, McCallum and Pereira's landmark work [7] on conditional random fields (CRFs) presented a new MRF approach based on directly modeling the conditional distribution of a field of labels. Their work focused on the use of chain structured CRFs and applications in natural language processing. Soon after Kumar and Hebert [6] explored the use of two dimensional lattice structures for general image segmentation.

While there has been an explosion of activity in the general computer vision community using CRFs, there has been comparatively less exploration of CRFs in the medical image segmentation literature. Some notable previous work includes Tsechpenakis et al. [13] coupled CRFs with deformable models for 3D eye segmentation. The deformable model captures shape information and is fed to the CRF model as observations. Lee et al. [8] used pseudo-CRFs for segmenting brain tumors. As global inference in a true CRF can be expensive during learning, they broke the problem up into two components, one of them not depending on spatial interactions and another term that accounted for spatial interactions which they view as a regularizer.

While modeling choices are a key part of solving the problem of image segmentation, the choice of image features is equally critical. Image features based on HOG features [4] have generated considerable interest in computer vision for tasks such as human detection in photographs. While HOG techniques have received an explosion of exploration in computer vision there has been comparatively little work in medical image analyis. There has been some recent work using HOG like techniques such as [11] which used a HOG like feature in their work on segmenting brain structures from Diffusion Tensor (DT-MR) images. Graf et al. [5] recently explored the use of a pyramidal HOG for detecting vertebrae in 2D CT imagery. We are motivated to use HOG based features here as oriented gradients provide a way to capture shapes and partial curves through gradient profiles which can be more robust to variations across patients and different spatial contexts. To the best of our knowledge, we provide the first exploration of HOG features for multi-organ 3D CT image segmentation. Our experimental evaluation confirm their utility in our setting here.

Other exemplary work on organ segmentation such as Siefert et al. [12] used landmark detectors consisting of 3D Haar features for image volumes. They target full body segmentation and use 6 organs for segmentation. The landmarks are used in an MRF framework to obtain organ centers. After obtaining organ centers, marginal space learning classifiers which are a sequence of Probabilistic Boosting Trees are used to perform organ segmentation. They achieve full body organ localization and segmentation using a heirarchical and contextual approach. In contrast, Ling et al. [9] used a hierarchical approach, marginal space learning and steerable features to segment a liver and improve upon previous methods of shape initializations. Other recent work [3] has used regression forests to detect and localize abdominal organs. Regression forests could be integrated into the MRF approaches we explore here. Varshney et al. [14] provides a survey of the different approaches that have been used for segmentation of abdominal organs. Neural networks, level set methods, model fitting and rule based methods have received particular attention.

We make a number of contributions in this paper. First, we demonstrate how discriminative parameter learning in CRFs indeed leads to better segmentation performance compared to than their generative MRF counterparts. Second, we demonstrate how HOG features are indeed well suited to the anatomical segmentation problem. Third, we model pairwise terms using non-submodular functions and perform full 3D inference using VMP and MP.

Funding for this research was provided by Kodak research and Carestream

2 Our Approach

We use 22 3D volumes of patients to perform our experiments. 19 of these volumes are taken from the liver data set (http://www.sliver07.org/) and we provide the additional labels for the dataset to the community for future research. We use the cases that had segmented liver results provided. Our goal is to segment five organs from the background: the liver, two kidneys, spleen and gall bladder. In the beginning, we very coarsely align the volumes using 6 manually selected landmark key points and a global 3D affine transformation. This is done because there is a lot of variation in the images regarding the relative positions of these organs. 2 points capture the axial extent (height) of the abdominal region of interest. We then select 4 landmark points on an image slice that has the first appearance of the right kidney as we look at the slices from top to bottom. Two of these points are marked on the two sides (left and right) of the sternum and the other 2 points are marked on the top of the rib cage and bottom of the vertebra. We use linear interpolation so that the resultant scan has 150 slices and an image size of 300(width) x 400(height). We scale the image set down by a half on all dimensions in our experiments for computational reasons. We use a higher range of intensity values (-180 to 1200HU) so as to be able to accomodate any contrast enhancement or gall bladder or kidney stones. We quantize intensity values into bins and model intensity as a histogram or discrete distribution. The gradient between two neighboring pixels is also modeled as a histogram of gradient values. The gradient is the difference of pixel intensities of the neighboring pixels. We model the spatial location of organs using a mixture of Gaussians and the choice of the number of Gaussians in the mixture is determined by cross validation set likelihood and we note that the results tend to depend on the organ's size and compactness. Appearance features are modeled as a 5x5 patch around each pixel and clustering is used. HOG features are also clustered to create code books. To compute the HOG features we use non overlapping blocks where each block is made of 3x3 cells and each cell of 8x8 pixels has 10 orientation bins.

The feature variables for the pixel p are x_p^{int}, x_p^{app}, x_p^{hog} and x_p^{loc} to denote features of intensity, appearance, HOG and location respectively. The gradient feature variable for neighboring pixels p and q is given by x_{pq}^{grad}. We depict a discriminative component with an edge and a dark node on the edge and a generative component with a directed edge so that the arrow points to the variable that is generated.

We look at two different probabilistic models for image segmentation. The models are either fully discriminative CRFs or contain both generative and discriminative components (CRF-MRF). Fig 3 shows the model structures. A conditional random field

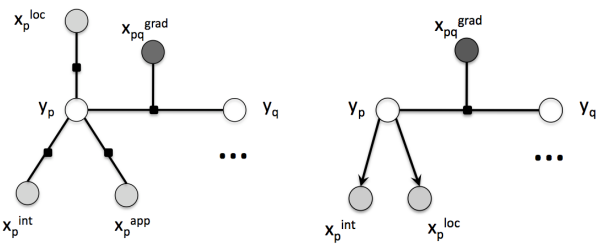

Fig. 3: The crf model and the crf-mrf graphical models explained in the text.

[7] can be expressed as an undirected graph or random field which has associated with it a conditional distribution $p(\mathbf{y}|\mathbf{x})$ where $\mathbf{y}$ denotes the output variables (e.g. image segmentation classes) and $\mathbf{x}$ denotes the input variables (such as features of an image). Let $\mathbf{x} \equiv [\mathbf{x^{int}}, \mathbf{x^{app}}, \mathbf{x^{loc}}, \mathbf{x^{grad}}]$, then our CRF can be written as:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_p \exp\Big(-\sum_{c=1}^{C}\sum_{b=1}^{B} \lambda_{c,b}\rho_\lambda(x_p^{int}, y_p) - \sum_{c=1}^{C}\sum_{l=1}^{L_c} \alpha_{c,l}\rho_\alpha(x_p^{loc}, y_p)\Big) \tag{1}$$

$$\prod_p \exp\Big(-\sum_{c=1}^{C}\sum_{t=1}^{T} \tau_{c,t}\rho_\tau(x_p^{app}, y_p)\Big) \prod_{p,q\in N} \exp\Big(-\sum_{c1,c2}\sum_{g=0}^{G} \gamma_{g,c1,c2}\rho_\gamma(y_p, y_q, x_{pq}^{grad})\Big)$$

where C indicates total number of class labels, B indicates number of bins per class (and we assume that each class has the same number of bins), T indicates number of appearance clusters per class, L_c indicates number of Gaussian distributions used to model the class c, G indicates the number of gradient bins and $Z(\mathbf{x})$ is the normalization constant. We note that we use HOG features in place of appearance features in the later experiments and the product of HOG feature potentials that is included in the equation (with $\mathbf{x} \equiv [\mathbf{x^{int}}, \mathbf{x^{hog}}, \mathbf{x^{loc}}, \mathbf{x^{grad}}]$) is as follows.

$$\prod_p \exp\Big(-\sum_{c=1}^{C}\sum_{h=1}^{H} \beta_{c,h}\rho_\beta(x_p^{hog}, y_p)\Big) \tag{2}$$

The parameters (in equations 1 and 2) of the data cost term are $\lambda_{c,b}$ one for each bin of each class label c, $\tau_{c,t}$ one for each appearance texton of each class label c, $\beta_{c,h}$ one for each HOG feature cluster for each class label c, $\alpha_{c,b}$ one for each location Gaussian for each class label c and the interaction parameters are $\gamma_{g,c1,c2}$, one for each gradient bin for each pair of classes. The feature functions have the following definitions.

$$\begin{aligned}
\rho_\lambda(x_p^{int}, y_p) &= \{1 \text{ if } x_p^{int} \in \text{ bin } b \text{ and } y_p = c,\ 0 \text{ otherwise }\} \\
\rho_\tau(x_p^{app}, y_p) &= \{1 \text{ if } y_p = c \text{ and } x_p^{app} = t,\ 0 \text{ otherwise }\} \\
\rho_\beta(x_p^{hog}, y_p) &= \{1 \text{ if } y_p = c \text{ and } x_p^{hog} = h,\ 0 \text{ otherwise }\} \\
\rho_\alpha(x_p^{loc}, y_p) &= \{1 \text{ if } y_p = c \text{ and } x_p^{loc} = l,\ 0 \text{ otherwise }\} \\
\rho_\gamma(y_p, y_q, x_{pq}^{grad}) &= \begin{cases} 1 & \text{when } y_p = c1 \text{ and } y_q = c2 \text{ and the gradient is in bin } g \\ 0 & \text{otherwise ,} \end{cases}
\end{aligned}$$

where $x_p^{app} = t$ indicates that the appearance feature at pixel p is closest to the texton codeword t. $x_p^{loc} = l$ indicates that the location feature at pixel p has highest probability of belonging to the Gaussian distribution $\{c, l\}$.

For our CRF-MRF model (as shown in Fig. 3) the local (datacost) features are modeled generatively while the interactive terms are modeled discriminatively. The equations of the model are given as follows:

$$p(\mathbf{x^{loc}}, \mathbf{x^{int}}, \mathbf{y}|\mathbf{x^{grad}}) = p(\mathbf{x^{loc}}|\mathbf{y})p(\mathbf{x^{int}}|\mathbf{y})p(\mathbf{y}|\mathbf{x^{grad}}) \tag{3}$$

$$= \prod_{p\in V} \mathcal{N}(x_p^{loc}|\mu_{y_p}, \Sigma_{y_p}) \prod_{p\in V} p(x_p^{int}|y_p) \frac{1}{Z(\mathbf{x^{grad}})} \prod_{p,q\in N} \exp\big(-\sum_{c1,c2}\sum_{g=0}^{G} \gamma_g\rho_\gamma(y_p, y_q, x_{pq}^{grad})\big),$$

where $p(\mathbf{x^{loc}}|\mathbf{y})$ for location is modeled as a Gaussian mixture model and $p(\mathbf{x^{int}}|\mathbf{y})$ is modeled as a discrete distribution. Learning here is easier as the learning of the interaction parameters is not affected by the intensity, location or appearance features.

We perform maximum likelihood (ML) or conditional ML learning using standard gradient descent. In contrast to [8] and other pseudo-likelihood or autoregression approaches such as [1] we use a fully globally defined CRF. However, for learning with gradient descent we need model expectations involving intractable marginal distributions during learning due to the cyclic nature of lattices. We therefore use a form of approximate global inference known as variational message passing (VMP) [15] to obtain approximate marginals for learning and loopy max-product for final predictions and segmentations. Interestingly, when learning pairwise potentials in both MRFs and CRFs we have often found that pairwise functions can become non-submodular and thus popular alternatives for inference such as those based on graph-cuts [2] cannot be used due to their modularity constraints.

3 Experiments and Results

We identify the background, liver, right kidney, left kidney, gall bladder and spleen as **C1**, **C2**, **C3**, **C4**, **C5** and **C6** respectively. These organs were selected because they pose a challenge to the learning algorithms. For example, the intensity profiles of the spleen and liver are very similar as well as to the stomach and intestines that are included in the background class. We perform 3 different experiments.

The first two experiments are in an interactive setup where the training slices and test slices come from the same patient. The goal is to accelerate the laborious task of labeling volumes in which a few slices are labeled with some time for an offline learning phase. The user can correct the results in a subsequent session. For both of these, we use 17 patient volumes. 8 of these volumes are used to compute the location information and 9 volumes to train and test on. We use 50 bins per class for intensity and 45 bins per pair of classes for gradient pairwise bins. Location features are modeled as mixture of Gaussians and the choice of the number of Gaussians in the mixture is determined by the organ's size and compactness. We use 11 Gaussians for the background, 6 for the liver, 3 for the right kidney, 3 for the left kidney, 3 for the gall bladder and 5 for the spleen. Initial clusters are obtained from K-means clustering on sampled points (taken from a completely different set of patient volumes and not from the training or testing slices), then running a Expectation Maximization (EM) based Gaussian mixture model with diagonal covariance matrices. We use 180 appearance code words and 120 HOG codewords. We choose a pair of 2 consecutive slices (mostly so we can include all the classes) for training and so that we can extract gradients in the 3rd (z-axis) dimension during training. For testing, we use 4 pairs of 4 consecutive slices. We can use larger number of consecutive slices at the cost of increased memory requirements.

In our first experiment, we compare the simple logistic regression with CRF and compare different features. We compare use of intensity, location (IL) with the addition of appearance (ILA) or the addition of HOG (ILH) or addition of both appearance and HOG (ILAH). Though the appearance patches 5x5 are not identical to the HOG dimensions (cell size of 8x8), we use a larger number of clusters for appearance. We observed that when we used appearance patches of 9x9, the segmentation results were blockier and hence worse. We initialize the datacost CRF parameters using the corresponding logistic regression parameters. The results are shown in Table. 1(a). In the second experiment,

	logreg				CRF			
	IL	**ILA**	**ILH**	**ILAH**	**IL**	**ILA**	**ILH**	**ILAH**
C1	96.2	95.7	97.5	95.6	98.2	97.3	98.4	96.4
C2	90.8	78.2	91.3	81.2	90.5	79.3	93.6	81.6
C3	67.4	55.6	75.0	51.4	72.9	64.0	80.5	57.7
C4	68.7	65.9	75.0	56.6	75.4	69.3	78.6	61.7
C5	19.2	7.5	40.3	8.0	19.7	12.7	42.4	21.8
C6	85.5	56.2	86.5	60.9	89.3	62.2	90.3	71.6
ACA	71.3	59.8	77.6	59.0	74.4	64.1	80.6	65.1
PA	92.9	88.9	94.5	89.1	94.9	90.9	95.9	90.7

(a) Experiment 1

	logreg			CRF		
	IL	**IH**	**ILH**	**IL**	**IH**	**ILH**
C1	82.5	83.8	83.1	83.2	83.8	85.7
C2	67.8	78.0	75.7	68.6	78.0	74.9
C3	63.3	39.1	66.9	63.3	39.1	64.9
C4	82.8	42.5	85.5	82.8	42.5	83.8
C5	97.9	50.2	89.7	97.7	50.2	91.9
C6	84.6	67.3	83.8	84.6	67.3	78.2
ACA	79.8	60.1	80.7	80.0	60.1	80.0
PA	79.9	81.0	81.6	80.6	81.0	83.4

(b) Experiment 3

Table 1: This table compares logistic regression with CRFs and also compares different features. I, L, A and H stand for intensity, location, appearance and HOG respectively. ACA and PA stands for average class accuracy and pixel accuracy respectively. All values are percentage accuracy.

we compare the CRF to the CRF-MRF. We use only intensity and location to compare the performance of these two approaches. Results are shown in Table. 2.

In the third experiment, we use all 22 patient volumes in a typical training/testing framework. We leave out 3 volumes for testing. Of the remaining 19 volumes, we use 15 volumes for training and 4 for cross-validation to select different modeling and parameter settings. We use the full 3D volume of patient for training and inference. The logistic regression is modified to handle imbalance of classes (Appendix A) [3]. The datacost parameters learnt here are used in the CRF and kept fixed. The smoothness term parameters are learnt in the CRF. We use 100 bins per class for intensity and 45 bins per pair of classes for gradient pairwise bins. Location features are modeled as mixture of Gaussians and the choice of the number of Gaussians in the mixture is determined by validation set likelihood and we note that the results tend to depend on the organ's size and compactness. The best set of Gaussians were 30 for the background, 20 for the liver, 5 for the right kidney, 3 for the left kidney, 2 for the gall bladder and 7 for the spleen. A weighted form of Expectation Maximization (EM) (Appendix B) based Gaussian mixture model with full covariance matrices was used to allow using more data points. 800 HOG codewords were generated using K-means. More number of codewords belonged to larger classes. The results are shown in Table 1(b).

4 Discussion and Conclusions

The experiments show an increase in pixel and class accuracy when HOG is used in most of the cases. The interactive setup experiment shows the use of a discriminative structured model like the CRF doing better than the simpler logistic regression. The appearance features do not tend to do well and it is possible that more appearance clusters or code words are required. In the 3rd experiment, we note that use of the modified logistic regression improved average class accuracy without which classes like gall bladder and kidney have very large error rates. We also note that in our setting, CRFs improve the pixel accuracy to some extent while the class accuracy remains almost the same. Using intensity and HOG (IH) which is less reliant on the coarse registration step shows

[3] http://www.cs.rochester.edu/~bhole/medicalseg

promise. The class accuracy increases from 30% (only intensity) to 60% (IH) (table not shown). The location feature however is more dependent on the manual registration step. The second experiment shows improved performance when using the discriminative CRF model compared to the MRF-CRF even with simpler features. CRF-MRF model completely misses some organs like the spleen (C6). In general, the goal is to reduce the labeling time of the user and so training can be done offline in this setup so the user can make corrections in subsequent interactions. Future work will involve use of shape models along with discriminative models to improve segmentations further.

Predicted (CRF)

Actual	C1	C2	C3	C4	C5	C6
C1	98.2	0.9	0.2	0.2	0.1	0.2
C2	9.0	90.5	0	0.2	0.2	0.1
C3	22.0	0.0	72.9			5.1
C4	15.9	8.5	0.2	75.4		0
C5	69.9	10.4			19.7	0
C6	8.1	1.1	1.4	0.0	0	89.3

ACA = 74.36 , PA = 94.91

Predicted (CRF-MRF)

Actual	C1	C2	C3	C4	C5	C6
C1	98.1	1.1		0.2	0.6	
C2	44.2	52.4		2.3	1.1	
C3	100.0					
C4	93.7	0.4		6.0	0.0	
C5	54.8	6.2		0.2	38.8	
C6	99.9	0.1				

ACA = 32.54 , PA = 81.8

Table 2: The tables above are the confusion matrices obtained using CRF (left) and CRF-MRF (right). The zeros are left out of the confusion matrices for clarity.

References

1. Blake, A., Rother, C., Brown, M., Perez, P., Torr, P.: Interactive image segmentation using an adaptive GMMRF model. In: ECCV. pp. 428–441 (2004)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001)
3. Criminisi, A., Shotton, J., Robertson, D.P., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: MCV. pp. 106–117 (2010)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE CVPR. pp. 886–893. IEEE Computer Society, Washington, DC, USA (2005)
5. Graf, F., Kriegel, H.P., Schubert, M., Strukelj, M., Cavallaro, A.: Fully automatic detection of the vertebrae in 2d ct images. In: SPIE Medical Imaging. vol. 7962 (2011)
6. Kumar, S., Hebert, M.: Discriminative random fields: A discriminative framework for contextual interaction in classification. In: ICCV. pp. 1150–1157 (2003)
7. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data (2001)
8. Lee, C.H., Wang, S., Murtha, A., Brown, M.R.G., Greiner, R.: Segmenting brain tumors using pseudo-conditional random fields. In: MICCAI. pp. 359–366 (2008)
9. Ling, H., et al: Hierarchical, learning-based automatic liver segmentation. IEEE CVPR (2008)
10. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2) (2004)
11. Motwani, K., Adluru, N., Hinrichs, C., Alexander, A.L., Singh, V.: Epitome driven 3-d diffusion tensor image segmentation: on extracting specific structures. In: NIPS (2010)
12. Seifert, S., et al: Hier. parsing and semantic nav. of full body CT data. In: Proc. SPIE (2009)
13. Tsechpenakis, G., Wang, J., Mayer, B., Metaxas, D.: Coupling CRFs and deformable models for 3D medical image segmentation. pp. 1–8 (2007)
14. Varshney, L.: Abdominal organ segmentation in ct scan images: A survey (2002)
15. Winn, J., Bishop, C.M.: Variational message passing. J. Mach. Learn. Res. 6, 661–694 (2005)
16. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden markov random field model and the EM algorithm. IEEE Trans. Med. Imaging 20(1), 45–57 (2001)