Tumor Segmentation with Multi-Modality Image in Conditional Random Field Framework with Logistic Regression Models

Yu-chi Hu[1], Michael Grossberg[2], and Gig Mageras[3]

***Abstract*— We have developed a semi-automatic method for multi-modality image segmentation aimed at reducing the manual process time via machine learning while preserving human guidance. Rather than reliance on heuristics, human oversight and expert training from images is incorporated into logistic regression models. The latter serve to estimate the probability of tissue class assignment for each voxel as well as the probability of tissue boundary occurring between neighboring voxels given the multi-modal image intensities. The regression models provide parameters for a Conditional Random Field (CRF) framework that defines an energy function with the regional and boundary probabilistic terms. Using this CRF, a max-flow/min-cut algorithm is used to segment other slices in the 3D image set automatically with options of addition user input. We apply this approach to segment visible tumors in multi-modal medical volumetric images.**

I. Introduction

A key requirement in high-precision radiotherapy is the accurate spatial delineation of the tumor and the critical normal organs abutting it. For this purpose, it is crucial to use images from multiple modalities. Computed tomography (CT) images provide high resolution images of both soft tissues and bony structures. Relative to CT, magnetic resonance imaging (MRI) provides improved soft tissue contrast in many anatomical sites. Positron emission tomography (PET) reveals functional data, though with much poorer resolution than CT or MRI. Figure 1(a) shows images from these modalities for the same patient. The complementary information provided by these images is heavily relied upon by the radiation oncologists in defining the tumor volume and designing an optimized treatment plan.

However, there are challenges in how to combine these images in treatment planning. For example, it is well known that large variability exists in target-delineation by different doctors using a single modality image [4]. In addition, it has been reported [22] that significantly different tumor volumes can be delineated by the same observer on images from different modalities. Given the above, it would be desirable to develop a computer-aided multiple modality image segmentation tool that incorporates both expert-user-input and machine-learning capabilities. We hypothesize that such an approach will reduce both inter-modality and inter-observer variability.

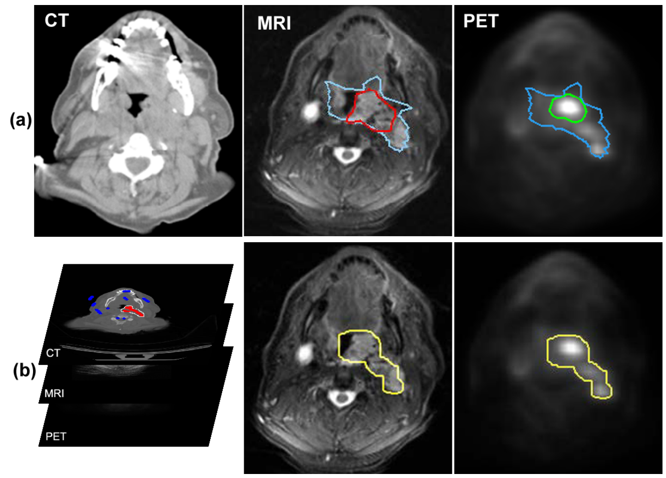

Fig. 1. (a) An image slice from co-registration of 3 modalities: CT, MRI and PET, showing greatly different image characteristics. CT virtually provides no distinguishable information on the tumor. The red contour was manually drawn by an expert on MRI only. The green contour was manually drawn by an expert on PET only. The blue contour was drawn by an export probably referring to both MRI and PET. The variation of the tumor contours is considerable. (b) Our method obtained training on one of the image slice in the same volume with all 3 modalities(left). The trained models were used to segment the same image slice shown on row (a). The resulting segmentation is the yellow contour on MRI and PET to show how our method utilizes information across the modalities by learning from the user's inputs.

We have developed a semi-automatic statistical framework for multi-modality image segmentation. Our work is based on Conditional Random Fields (CRF) [14] and energy minimization with a max flow/min cut algorithm. We define purely probabilistic regional and boundary terms in an energy function. The terms are based on logistic regression models statistically estimated. The parameters of the models are learned online from a training multi-modal image with the inputs of the expert user. The other image slices in the same data set then are segmented using the same parameters of the CRF framework without the need of user interactions on every image slice. Optional user input on further slices allows for correction of the segmentation as well as refinement of the CRF framework parameters. An overview of our segmentation method is shown in Figure 2. The novelty of the present work is exploring the use of CRF in simultaneously segmenting images from multi-modalities, and the use of regression models, trained with human supervision, for both the regional and boundary properties in context of tumor

[1]Y.-C. Hu is with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA and the Department of Computer Science, The Graduate Center, City University of New York, 365 Fifth Avenue, New York, NY 10016, USA huj at mskcc.org

[2]M. Grossberg is with the Department of Computer Science, City College of New York, 160 Convent Avenue, New York NY 10031, USA grossberg at cs.ccny.cuny.edu

[1]G.. Mageras is with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA magerasg at mskcc.org

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

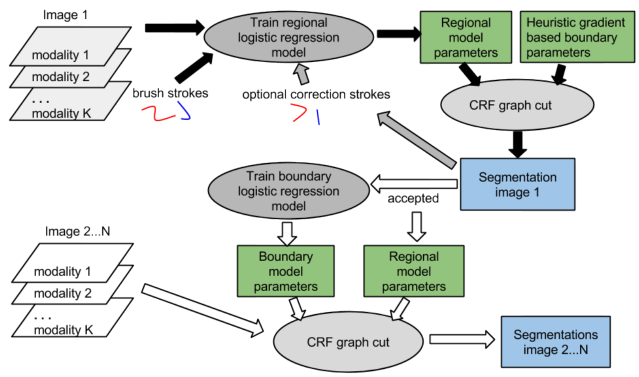

Fig. 2. The overview of the segmentation work flow. In a training phase, brush strokes are drawn on a training multi-modal image slice (Image 1). Parameters of the logistic regression model for the regions are calculated and basic graph cut method with contrast based boundary energy term is used to get a segmentation. Additional brush strokes are optional for correction to re-estimate the parameters and obtain new segmentation. Once the segmentation on the training image is accepted, the parameters of the logistic regression model for the boundary are estimated from the accepted segmentation. Both regional and boundary regression models are used in our energy function defined in CRF to segment the other images. Further training can be added at any point later to improve the segmentation.

image-segmentation.

A. *Previous work*

Active contour and level-set methods [10][15][20] are commonly available in tools for medical image segmentation. These methods evolve an initial contour iteratively with a speed function usually defined from gradient based edge features which attract the contour toward the boundary or with an external force constraint that maintains the regional homogeneities inside the contour [16]. Due to the shape of the target anatomical structure, the parameters controlling how the contour evolves need to be manually adjusted and usually are not intuitive to medical specialists. State-of-art commercial software, such as Smart Segmentation [9] from Varian (Palo Alto, CA) and Smart Probabilistic Image Contouring Engine (SPICE) [17] from Philips (Eindhoven, Netherlands), now employs deformable models and/or atlas-based methods for automatic segmentation. Recent studies [6][9][21], however, reported that reliability of advanced model-based methods is generally still inferior to interactive methods due to the large variation of the shapes in some anatomical structures. Furthermore, these methods are designed to work with normal organs, thus may not suitable for tumor segmentation since representing irregularly shaped tumors with a mutual shape model or an atlas is improbable.

Markov Random Fields (MRF) with graph cut based energy minimization, originally for binary image denoising [5], was extended for image segmentation [2][13][11][18] to include a contrast based boundary term in the energy function. The contrast based boundary term, like other boundary based methods such as level set, usually results in a segmentation that favors strong edges in the images. For medical images, however, the target structure often may not have a strong boundary due to the tissue type or due to the nature of image modalities, such as PET.

We have previously [7][8] introduced a simple non-parametric histogram-based boundary model trained from an accepted segmentation. With single image modality, the model showed superior segmentations for normal organs than segmentations using contrast based boundary models. By extending this work, we have developed a framework for multi-modality images to utilize information from different image modalities. Histogram based models, however, are too expensive in multi-dimensions when the number of the modalities grows. We choose logistic regression models for both regional and boundary image features to get segmentations closer to what physicians may draw in the application of tumor segmentation. There are prior approaches [3][19] that learn a segmentation from a training image and applies the trained models to segment the other related images In contrast, our approach integrates the option of user interaction for training and correction within single work flow. In addition we use purely statistical regional and boundary energy terms in the random field framework.

II. METHODS

A. *Conditional Random Fields*

Let $G = (V, E)$ be the graph representing an image with N voxels or nodes indexed by i, where $V = \{i|1 \leq i \leq N\}$, $E = \{\{i, j\}|i \in V, j \in N_i\}$ and N_i are neighbors of voxel i in a neighborhood system such as 4-connected or 8-connected system in 2D. We treat the segmentation problem as a classification problem in a stochastic process, that is, assigning a class label for each node in the graph based on the observed image features. Let $X = (X_i)_{i \in V}$ be the multivariate random variable of such assignments. x is an assignment instance and x_i is the class assignment for node i. Let $Y = (Y_i)_{i \in V}$ be the multivariate random variable of images and y_i be the extracted image feature vector at node i. For example, $y_i = (CT_i, MRI_i, PET_i)$ is the image intensity of CT, MRI and PET at voxel i. The segmentation problem can be simply described as finding an assignment x such that the conditional probability $P(X = x|Y = y)$ or $P(x|y)$ is maximum, that is, obtaining a *Maximum-A-Posteriori* (MAP) estimate of x.

A CRF obeys the Markov property as does MRF. By definition, the factorization of the conditional probability characterized by a Gibbs distribution has the form

$$P(x|y) \propto \exp \sum_i [r_i(x_i, y) + \lambda \sum_{j \in N_i} u_{ij}(x_i, x_j, y)] \quad (1)$$

For image segmentation, r_i represents a regional term that describes the relationship between the image y and label assignment x_i. u_{ij} represents a boundary term that describes the relationship between the image y and label assignments for the pair of neighboring x_i and x_j. λ is a constant weight to penalize discontinuity. It should be pointed out that global y is in both r_i and u_{ij}. That is, CRF allows the use of not only the image feature limited at pixel i but also, without breaking the MAP inference of random fields, more remote image features such as contextual local pattern [3] and image patches [19]. In this work, we simply used the observed image intensity values from multiple modalities.

B. Logistic regression models

To keep the CRF framework purely statistical without restrictive assumptions, a probabilistic function is a sensible choice for both regional and boundary terms. We define

$$r_i(x_i, y) = -\ln p(x_i|y_i) \tag{2}$$

and

$$u_{ij}(x_i, x_j, y) = \begin{cases} 0, & \text{if } x_i = x_j \\ -\ln p(x_i \neq x_j|y_i, y_j), & \text{if } x_i \neq x_j \end{cases} \tag{3}$$

To estimate the probabilities, we assume a logistic regression model of probabilities. Let $\beta = (\beta_0, \beta_1, ..., \beta_K)$, where K is number of modalities, be the model parameters for the regional term and $y'_i = (1, y_i)$. Thus the probability that pixel i is target class is :

$$p_i = p(x_i = 0|y_i) = \frac{1}{1 + \exp(-\beta \cdot y'_i)}, \tag{4}$$

and the probability for non-target class $p(x_i = 1|y_i)$ is simply:

$$p(x_i = 1|y_i) = 1 - p_i. \tag{5}$$

As for the boundary term, we define a new feature vector $z_{ij} = (z_{ij1}, z_{ij2}, ..., z_{ijK})$, $j \in N_i$, where $z_{ijk} = |y_{ik} - y_{jk}|$. Let the model parameters be $\gamma = (\gamma_0, \gamma_1, ..., \gamma_K)$ and $z'_{ij} = (1, z_{ij})$ then the probability that there is a boundary between neighboring pixel i and j is:

$$p_{ij} = p(x_i \neq x_j|y_i, y_j) = \frac{1}{1 + \exp(-\gamma \cdot z'_{ij})}, \tag{6}$$

C. Training

Instead of building a training database, we collect samples on-line for both regional and boundary logistic regression models on a training image slice. To incorporate human guidance interactively, we use a similar approach as Boykov et al. [1][2] that enables the user to use paint brushes to identify locations of target and background regions. The brush strokes provide a collection of target and non-target samples for training of the regional model. The segmentation on the first training image is done using a traditional contrast based boundary term. The user can apply additional brush strokes for correction of the segmentation. Once the user accepts the segmentation on the training image slice, the set of pairs of neighboring pixels straddling the boundary of the accepted segmentation provides the samples for boundary training. Pairs of neighboring pixels randomly selected outside and inside the accepted segmentation serve as non-boundary samples. To estimate the parameters of logistic regression models we use stochastic gradient descent. The trained regional and boundary models are used to segment other image slices in the same data set.

III. Experiments

A. Tumor segmentation

We randomly selected 3 clinical head-and-neck patient cases consisting of image studies in all 3 image modalities: CT, MRI and PET, from the patient archive in the Department of Radiation Oncology at MSKCC (Memorial Sloan-Kettering Cancer Center, New York, USA) These image studies were acquired at different times for different purposes such that the patient geometry may be significantly different between images. For example, the patient head in each image modality may have a different inclination. In each case, the 3 image modalities were manually registered using in-house rigid registration software. Since the tumor is relatively small and the slice spacing is 4mm to 5mm, we used small regions of interest (ROIs) for registrations in order to obtain accurate registrations around the tumor locations. The ROIs contain 6-8 image slices per case. After registration, MRI and PET were re-sampled under the same image space of CT. The dimension of each image slice is 512x512. For comparison, we use Seg3D, a volume segmentation and processing tool developed by the NIH Center for Integrative Biomedical Computing (CIBC) at University of Utah (Salt Lake City, UT), which implemented the level set segmentation using Insight toolkit(ITK.) The level set method works for single modality only and we chose MRI. The number of iterations is set to 200 and all other parameters for the level set method are set to default.

λ in Eq 1 is 2 for all the cases in our method. Figure 3 shows the results of tumor segmentation from our multi-modality method (in yellow) compared to the manual contours (in red) and level set segmentations (in cyan.) Manual contours drawn by physicians served as the reference. Physicians also make use of their professional knowledge of tumors which cannot be learned merely from the images alone. Inter-observer variation should also be noted but the manual contours still represent an acceptable ground truth. An alternative brush stroke set #2, shown in Figure 4, was used for comparison of results using different brush strokes.

In this study, tumors are virtually indistinguishable from the normal tissue on CT. On the other hand, PET has more discriminating information for determining the tumor segmentation. Our model parameters, obtained from the training image in Figure 3(a), reflect such observations. For example, the parameters of regional model in Case 1 using brush stroke set #1 are (-0.549, 1.516, 8.495) for CT, MRI and PET respectively and the boundary model parameters are (-1.092, 1.759, 3.082). PET, however, has no clear boundary and the segmentations from our method utilized MRI for the boundary as shown in Figure 3. This study demonstrates that our method captures the available information across modalities and how the probabilistic boundary term helps in these cases.

We use the Dice coefficient to measure the similarity between segmentations. On average, using manual contours as the reference, our method using brush stroke sets #1 and #2 achieve 0.751 and 0.768 respectively while Seg3D level

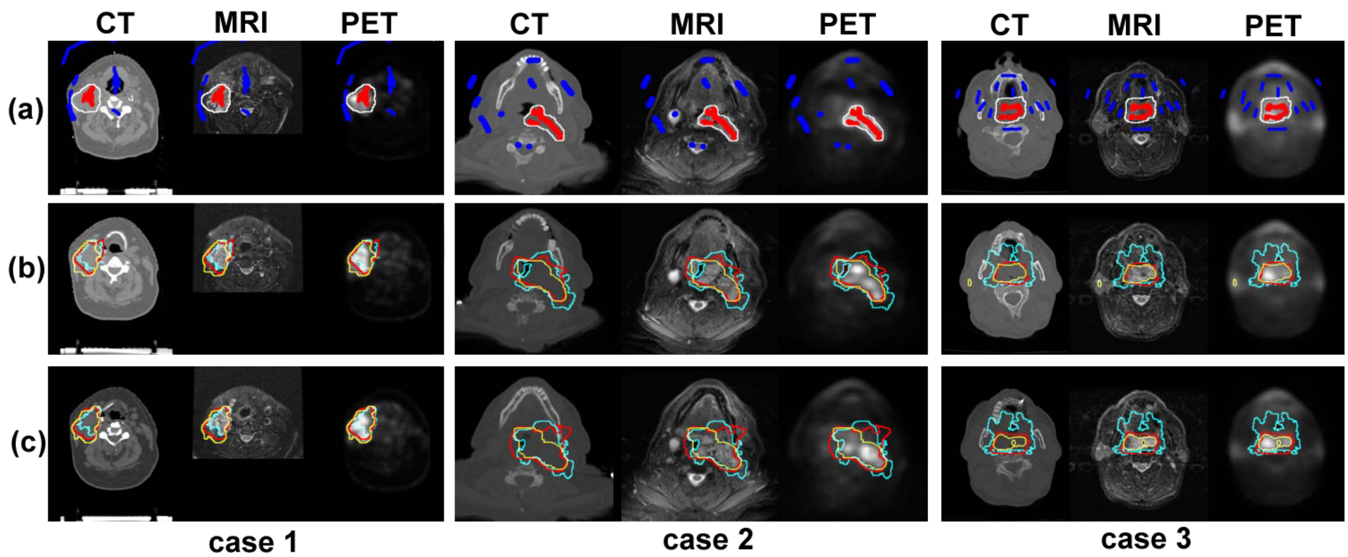

Fig. 3. Multi-modality tumor segmentations of 3 cases. For each case, 1st column is CT, 2nd column is MRI and 3rd column is PET. The 3 modalities are co-registered in each case. Row (a) shows the training images for each case in which the user drew brush strokes and accepted the segmentations (white contours) to train our method. Row (b) and (c) show two other image slices of the same image set in each case for comparison. Red contours are manual contours drawn by physicians, yellow contours are segmentations from our method and cyan contours are segmentations from the Seg3D level set method. On these images, our segmentations were done automatically without additional brush strokes. Probabilistic regional and boundary terms are estimated from logistic regression models trained from row (a). Our method utilizes all information across all modalities for the segmentation. For example, in case 1 row (c), the boundary of our segmentation follows the boundary of PET on the lower side but also the boundary of MRI on the upper side.

case	1	2	3	Average
with brush stroke set #1	0.810	0.735	0.708	0.751
with brush stroke set #2	0.815	0.773	0.716	0.768
Seg3D	0.740	0.772	0.577	0.696

TABLE I

DICE COEFFICIENT RESULTS.

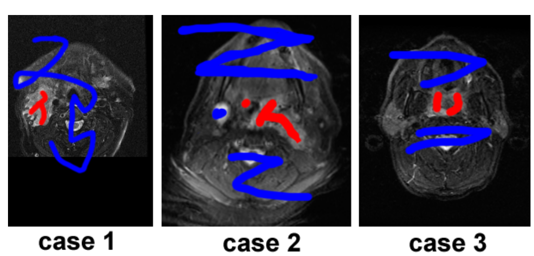

Fig. 4. For comparing the results of using different brush strokes in our method, an alternative brush stroke set (#2) is shown here on the training MRI images of the 3 testing cases.

set is 0.696. The results for each case are shown in Table I.

IV. CONCLUSIONS AND FUTURE WORKS

We have explored multi-modality medical image segmentation using CRF. In the energy function, we have introduced purely probabilistic regional and boundary terms that are estimated from logistic regression models. The models are directly trained from expert user inputs without ad-hoc assumptions that favor strong edges. We showed that tumor segmentations from our method on multi-modality images are more similar to physicians' manual segmentations, compared to other semi-automatic techniques such as level set.

Future work will study the following using the same purposed framework:

- Because CRF relaxes strong independence assumptions and retains the MAP inference, we will investigate using contextual regional properties such as texture or patch profile in the energy function.
- Logistic regression models are good for binary classification. For multi-class segmentation, for example, multiple normal organs on multi-modality images, we may use other discriminative models or non-parametric methods such as k-NN.
- We will continue to explore applications in other domains such as color video segmentation and time series of multi-spectrum remote sensing images.

REFERENCES

[1] Y. Boykov and G. Funka-Lea. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, 70:109–131, 2006.

[2] Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *Proc. International Conference of Computer Vision*, pages 105–112, 2001.

[3] J. Cui, Q. Yang, F. Wen, Q. Wu, C. Zhang, L. V. Gool, and X. Tang. Transductive object cutout. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.

[4] J. L. Fox, R. Rengan, W. OMeara, E. Yorke, Y. Erdi, S. Nehmeh, S. A. Leibel, and K. E. Rosenzweig. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for nonsmall-cell lung cancer? *International Journal of Radiation Oncology Biology Physics*, 62(1):70 – 75, 2005.

[5] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistics Society*, 51:271–279, 1989.

[6] T. Heimann, B. van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, F. Bello, G. Binnig, H. Bischo, A. Bornik, P. Cashman, Y. Chi, A. Cordova, B. Dawant, M. Fidrich, J. Furst, D. Furukawa, L. Grenacher, J. Hornegger, D. Kainmueller, R. Kitney, H. Kobatake, H. Lamecker, T. Lange, J. Lee, B. Lennon, R. Li, S. Li, H.-P. Meinzer, G. Nemeth, D. Raicu, A.-M. Rau, E. van Rikxoort, M. Rousson, L. Rusko, K. Saddi, G. Schmidt, D. Seghers, A. Shimizu, P. Slagmolen, E. Sorantin, G. Soza, R. Susomboon, J. Waite, A. Wimmer, and I. Wolf. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans. Med. Imaging*, 28:1251–1265, 2009.

[7] Y.-C. Hu, M. Grossberg, and G. Mageras. Semi-automatic medical image segmentation with adaptive local statistics in conditional random fields framework. In *EMBS 2008. 30th Annual International Conference of the IEEE*, pages 3099–3102, 2008.

[8] Y.-C. Hu, M. D. Grossberg, A. Wu, N. Riaz, C. Perez, and G. S. Mageras. Interactive semiautomatic contour delineation using statistical conditional. *Medical Physics*, 39:4547–4558, 2012.

[9] D. P. Huyskens, P. Maingon, L. Vanuytsel, V. Remouchamps, T. Roques, B. Dubra, B. Haas, P. Kunz, T. Coradi, R. Bhlman, R. Reddic, A. V. Esch, and E. Salamon. A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer. *Radiotherapy and Oncology*, 90:337–345, 2009.

[10] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1:321–331, 1988.

[11] P. Kohli and P. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 922–929. IEEE, 2005.

[12] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

[13] M. Kumar, P. Torr, and A. Zisserman. Obj cut. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 18–25, 2005.

[14] J. Lafferty, A. McCallum, and A. F. Pereira. Conditional random fields:probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conference on Machine Learning*, pages 282–289, 2001.

[15] R. Malladi, J. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, 1995.

[16] I. E. Naqa, D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, J. D. Bradley, P. Grigsby, and J. Deasy. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. *Medical Physics*, 34:4738–49, 2007.

[17] A. Qazi, V. Pekar, J. Kim, J. Xie, S. Breen, and D. Jaffray. Autosegmentation of normal and target structures in head and neck ct images: A feature-driven model-based approach. *Medical Physics*, 38:6160–6170, 2011.

[18] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 309–314, New York, NY, USA, 2004. ACM.

[19] Y. Schnitman, Y. Caspi, D. Cohen-Or, and D. Lischinski. Inducing semantic segmentation from an example. In P. Narayanan, S. Nayar, and H.-Y. Shum, editors, *Computer Vision ACCV 2006*, volume 3852 of *Lecture Notes in Computer Science*, pages 373–384. Springer Berlin Heidelberg, 2006.

[20] J. A. Sethian. *Level set methods and fast marching methods*. Cambridge University Press, 1999.

[21] N. Sheth, S. Murphy, Y. Chen, and C. Yang. Comparison of model-based segmentation systems for contouring of male pelvic structures. *Medical Physics*, 37:3125, 2010.

[22] R. J. Steenbakkers, J. C. Duppen, I. Fitton, K. E. Deurloo, L. J. Zijp, E. F. Comans, A. L. Uitterhoeve, P. T. Rodrigus, G. W. Kramer, J. Bussink, K. D. Jaeger, J. S. Belderbos, P. J. Nowak, M. van Herk, and C. R. Rasch. Reduction of observer variation using matched CT-PET for lung cancer delineation: A three-dimensional analysis. *International Journal of Radiation Oncology Biology Physics*, 64(2):435 – 448, 2006.