Probabilistic Model for 3D Interactive Segmentation

Ohad Shitrit[1], Tsachi Hershkovich[1][*], Tamar Shalmon[2], Ilan Shelef[2] and
Tammy Riklin Raviv[1]

[1]Department of Electrical and Computer Engineering, Ben-Gurion University of the
Negev, Beer-Sheva, Israel
[2]Department of Radiology, Soroka Medical Center, Ben-Gurion University of the
Negev, Beer-Sheva, Israel

Abstract. Fully-automated segmentation algorithms offer fast, objective, and reproducible results for large data collections. However these techniques cannot handle tasks that require contextual knowledge not readily available in the images alone. Thus, the expertise of an experienced physician is necessary.

We present a generative approach to image segmentation, which supports an intuitive and convenient user interaction subject to the bottom-up constraints introduced by the image intensities. The user "dialogue" with the segmentation algorithm, via several mouse clicks in regions of disagreement, is formulated as an additional, spatial term in a global cost functional for 3D segmentation. The method is exemplified for the segmentation of cerebral hemorrhages (CH) in human brain CT scans.

1 Introduction

Being fundamental to medical imaging analysis, image segmentation is actively studied, and numerous approaches exist. Recent trends focus on fully automatic segmentation frameworks, being much faster than manual annotation, less biased and repeatable. Usually, the required workload for processing and analyzing large datasets is far behind the ability of a human rater. Moreover, the computational advancements of the machine in cases that require modality fusion or 3D visualization cannot be competed even by an expert. Nevertheless, as the outcome of the image analysis process might have critical implications on the patient recuperation prospects the expertise of a clinician must be considered.

Interactive segmentation (IS) approaches can be classified based on the form and the type of input provided by the user as well as the underlying segmentation framework, see [19] and references therein. The pioneering IS work, which led to the development of the *live wire* technique or *intelligent scissors*, independently suggested by [5,11] is based on the image edge map. The shortest paths between the user's mouse clicks, calculated by the Dijkstra algorithm form the contour of the region of interest (ROI). Here, as well as in the *united snakes* [9], which relies

---

[*] Ohad Shitrit and Tsachi Hershkovitch equally contributed to the manuscript

on a classical active contour framework known as *snakes* [8], the user 'plants' anchors or seed points along the desired boundary, providing guidance for the segmentation.

Mouse scribbles seem to be the most common form of user interaction. The marked regions provide information about the ROI and the background intensity distributions. A well known IS approach is the *GrabCut* technique [17] which is based on the *graph-cut* [1]. Representing the image pixels by nodes in a graph, the graph-cut addresses a foreground-background image segmentation by solving a min-cut, max-flow problem. The user's annotated regions are assigned to either the source or the sink of the graph. In a recent paper by [12] marked user regions, via mouse scribbles were used for gathering spatially varying color statistics for multi-label segmentation. Level-set based segmentation framework with UI which are designed for medical images were suggested in [6,7]. Other recent IS techniques include [15,18,14,10].

We present a user interactive framework for the segmentation of volumetric ROIs. We chose to use the level-set framework as it is parametrization-free, allows automatic topological changes and enables a straightforward generalization to high-dimension and multi-label segmentation. Segmentation is obtained by solving a maximum a posteriori (MAP) problem. Using calculus of variation we search for the optimum of a cost functional which is based on a generative probabilistic model. The initial, fully automatic segmentation process is dominated by the gray level distributions of the ROI and the background partitions. The user input, which is provided by a few mouse clicks, adds spatial constraints to the segmentation problem. These constraints are soft and express, what we term here the *user's certainty disagreement map*. The user's disagreement map is convolved with a Gaussian kernel which defines the local extent of the user influence and controls the level of disagreement (or user's confidence) . This contribution allows to establish a user-machine dialogue. The user by no way edits the segmentation. Instead, our model provides an elegant framework to refine segmentation by resolving voxel annotation ambiguities, through *"negotiation"*.

The suggested method is exemplified on the segmentation of cerebral hemorrhages (CH) in human brain CT scans. CH volume estimates provided by the segmentation are critical to the determination of the therapy procedure which may involve surgery in addition to medicine intake. It is important to note that CHs have well defined intensity range, expressed in Hounsfield units (CT numbers). Nevertheless, similar gray-level values as a result of calcification or proximity to the skull bone can fail a fully automatic segmentation process and requires the insight of a physician. Our user-interactive segmentation tool was tested by two clinical experts and compared with a commercial toolbox. The advantage of the suggested method was obvious, considering both operating convenience (user interaction) and accuracy (user satisfaction). Both initialization and the first stage of the segmentation process were fully automatic. Usually, not more than a single or a couple of user interaction steps were needed to complete the segmentation. High Dice scores were obtained in a comparison with independent, fully manual annotations.

2 Methods

2.1 Problem definition and formulation

Let I be a gray-level image defined on a 3D image domain Ω. Our goal is to find the ROI/background image partitions denoted by ω and $\Omega \setminus \omega$, respectively. Let $S \colon \Omega \to \{0, 1\}$ denote the unknown binary voxel annotation of I corresponding to this partition. We assume that the observed image I is generated i.i.d. by the respective segmentation S with probability $p(I|S; \psi)$ where ψ are Gaussian mixture model (GMM) parameters of image intensities. A fully automatic segmentation can be stated as a MAP problem:

$$p(S|I; \psi) \propto p(I|S; \psi)p(S), \tag{1}$$

where the prior $p(S)$ is used for regularization as we shall see in the following. Incorporating the user input, the proposed segmentation framework is carried out in a step-by-stap manner, where at each step k the user provides a set of spatial parameters through the interaction, after observing the previous segmentation estimate, i.e. S^{k-1}. We denote this set by η^k and assume that S^k is generated with probability $p(S^k|S^{k-1}; \eta^k)$. Note, the similarity of this conditional probability to a first order Markov chain where η^k can be considered as the transition probability parameters. Given S^{k-1}, the intensity distribution parameters of the observed image, denoted by ψ and the user input, i.e. η^k, we can define the posterior probability of S^k, using the Bayes theorem and the chain rule, as follows:

$$p(S^k|I, S^{k-1}; \psi, \eta^k) \propto p(I|S^k; \psi)p(S^k|S^{k-1}; \eta^k). \tag{2}$$

We will look for $\hat{S}^k = \operatorname{argmax}_{S^k} p(S^k|I, S^{k-1}; \psi, \eta^k)$ by minimizing the following cost functional:

$$\mathcal{E} = -\log p(I|S^k; \psi) - \log p(S^k|S^{k-1}; \eta^k), \tag{3}$$

using the proportion $\mathcal{E} \propto -\log p$. In the following, an explicit formulation of the image likelihood term $p(I|S^k; \psi)$ and the user input term $p(S^k|S^{k-1}; \eta)$ will be presented. We assume i.i.d. distribution of the image voxels. Therefore the probability over the entire image domain will be presented by the product of probabilities of each voxel. A continuous form of $\mathcal{E}$, based on a level-set framework will be introduced next, followed by a gradient descent optimization process to estimate S^k.

2.2 Image likelihood term

We use mixtures of Gaussians to model the intensity distribution within and outside the ROI. Let $\psi_{in} = \{\mu_i^{in}, \sigma_i^{in}, w_i^{in}\}_{i=1}^{N^{in}}$ and let $\psi_{out} = \{\mu_i^{out}, \sigma_i^{out}, w_i^{out}\}_{i=1}^{N^{out}}$, where N^{in} and N^{out} are the respective numbers of Gaussians and w_i are the weights. For each image voxel $\mathbf{x}$ with intensity $I(\mathbf{x})$ we define

$$P_{in}(\mathbf{x}) \propto \sum_{i=1}^{N^{in}} w_i^{in} \exp\left(-\frac{(I(\mathbf{x}) - \mu_i^{in})^2}{(\sigma_i^{in})^2}\right) \text{ and } P_{out}(\mathbf{x}) \propto \sum_{i=1}^{N^{out}} w_i^{out} \exp\left(-\frac{(I(\mathbf{x}) - \mu_i^{out})^2}{(\sigma_i^{out})^2}\right).$$

The image likelihood cost, for a given ROI, could be defined as follows:

$$\mathcal{E}_{IL} = -\sum_{\mathbf{x}\in\omega^k} \log P_{in}(I(\mathbf{x}),\psi_{in}) - \sum_{\mathbf{x}\in\Omega\setminus\omega^k} \log P_{out}(I(\mathbf{x}),\psi_{out})$$

$$= -\sum_{\mathbf{x}\in\Omega} \left[S^k(\mathbf{x}) \log P_{in}(I(\mathbf{x}),\psi_{in}) + (1 - S^k(\mathbf{x})) \log P_{out}(I(\mathbf{x}),\psi_{out}) \right] \quad (4)$$

Note, that both the ROI (or equivalently, the segmentation S^k) and the intensity parameters ψ are unknown, and are therefore estimated via an alternating minimization scheme which will be detailed in the following.

2.3 User input term

Using mouse clicks, the user indicates disagreement with the current segmentation. The collection of user clicks at a time step k can be considered as a *forest* of impulse responses $U^k = \sum_{m=1}^{M^k} \delta(\mathbf{x}_m^k)$ defined on the 3D image domain, where $\mathbf{x}_m^k = (x_m^k, y_m^k, z_m^k)$ are the coordinates of a marked voxel and M^K is the number of clicks at time step k. Let $\hat{\omega}^{k-1}$ be the ROI estimate at step $k-1$. We then define $p(S^{k-1}(\mathbf{x})) \triangleq p(\mathbf{x} \in \hat{\omega}^{k-1})$. To the first approximation, the transition probability, given the user input takes the following form:

$$p(S^k|S^{k-1};U^k) = \begin{cases} p(S^{k-1}) & \text{if } U^k = 0 \\ 1 - p(S^{k-1}) & \text{if } U^k = 1 \end{cases} . \quad (5)$$

or, alternatively,

$$p(S^k|S^{k-1};U^k) = p(S^{k-1})^{(1-U^k)}(1 - p(S^{k-1}))^{U^k} \quad (6)$$

In practice, we smooth the binary user input by a convolution with a ,Gaussian kernel: $\eta^k \triangleq U^k * G_{\sigma^U}$. The transition probability can now be rephrased as follows:

$$p(S^k|S^{k-1};\eta^k) = p(S^{k-1})^{(1-\eta^k)}(1 - p(S^{k-1}))^{\eta^k}. \quad (7)$$

Note, that the voxels background/foreground assignments are not altered due to the user input. Instead, the assignment probability of a voxel at $\mathbf{x}$ is *flipped* with probability $\eta^k(\mathbf{x})$, where η^k represents the user's confidence or *certainty disagreement* map at step k. The extent of the user's influence, i.e. σ^u can be either determined by the user or set to a default value. To avoid cases in which a user click within the ROI affects the background or vice versa we define ROI-background user maps as follows: $\eta_{in}^k = S^{k-1}\eta^k$ and $\eta_{out}^k = (1 - S^{k-1})\eta^k$. We now define a user interaction energy term as a sum of two components referring to the user's clicks within the ROI and outside it:

$$\mathcal{E}_{UI} = -\sum_{\mathbf{x}\in\omega^k} \log_{in} p(S^{k-1}(\mathbf{x}),\eta_{in}^k(\mathbf{x})) - \sum_{\mathbf{x}\in\Omega\setminus\omega^k} \log p_{out}(S^{k-1}(\mathbf{x}),\eta_{out}^k(\mathbf{x})) \quad (8)$$

$$= -\sum_{\mathbf{x}\in\Omega} \left[S^k(\mathbf{x}) \log p_{in}(S^{k-1}(\mathbf{x}),\eta_{in}^k) + (1 - S^k(\mathbf{x})) \log p_{out}(S^{k-1}(\mathbf{x});\eta_{out}^k) \right],$$

where, $\log p_{in}(S^{k-1},\eta_{in}^k) \triangleq \log p(S^k|S^{k-1};\eta_{in}^k)$, and $\log p_{out}(S^{k-1},\eta_{out}^k) \triangleq \log p(S^k|S^{k-1};\eta_{out}^k)$, are defined by Eq. (7).

3 Level-set framework for interactive segmentation

3.1 Probabilistic view

We use a level-set framework to formulate the UI-segmentation problem [13]. Let ϕ^k define a level-set function, such that $\partial\omega^k = \{\mathbf{x}|\phi^k(\mathbf{x}) = 0\}$ denotes the estimated ROI boundaries in I at step k, and let ω_k denote the corresponding ROI domain. As in [2], the binary segmentation S^k can be represented by applying the Heaviside function to ϕ^k, i.e. $H(\phi^k)$. Adopting the probabilistic approach in [16], we use the logistic regression sigmoid, which can be used as a regularized form of the Heaviside function, to represent the soft segmentation $p(S^k)$:

$$H_\epsilon(\phi^k) = \frac{1}{2}\left(1 + \tanh\left(\frac{\phi^k}{2\epsilon}\right)\right) = \frac{1}{1 + e^{-\phi^k/\epsilon}}, \tag{9}$$

where ϵ can be use to determine the fuzziness of the estimated ROI boundaries [16]. As is [16], we now define the level-set function ϕ^k, as follows:

$$\phi^k(\mathbf{x}) \triangleq \epsilon \, \text{logit}(p) = \epsilon \, \log\frac{p(\mathbf{x} \in \omega^k)}{1 - p(\mathbf{x} \in \omega^k)} = \epsilon \, \log\frac{p(\mathbf{x} \in \omega^k)}{p(\mathbf{x} \in \Omega \setminus \omega^k)}. \tag{10}$$

It can be shown by substitution of Eq. (10) into Eq. (9) that $H_\epsilon(\phi^k)$ is equivalent to $p(\mathbf{x} \in \omega^k)$. This relation is fundamental in the proposed probabilistic level-set framework. It is important to note that ϕ^k is generated in an iterative manner, using a gradient descent framework. However, k is not an iteration index (which will be denoted by τ) but the final form of a level-set function after a first step of a fully automatic segmentation followed by $k - 1$ steps of user interaction. Usually the number of steps is not higher than two or three. We next, use the equivalence between $p(S^k)$ and $H_\epsilon(\phi^k)$, and the continuous forms of equations above to resolve the segmentation problem via a level-set based gradient descent optimization.

3.2 Cost functional

The proposed cost functional for a level-set based segmentation includes an image likelihood term, $\mathcal{E}_{IL}$, a regularization term, $\mathcal{E}_{REG}$, and a user interaction term $\mathcal{E}_{UI}$:

$$\mathcal{E}(\phi^k|I, \phi^{k-1}, \psi, \eta^k) = \mathcal{E}_{IL}(\phi^k|I, \psi) + \mathcal{E}_{UI}(\phi^k|\phi^{k-1}, \eta^k) + \mathcal{E}_{REG}(\phi^k). \tag{11}$$

Adopting a continuous formulation and a soft definition of the ROI, the sum over the image voxels, i.e. $\sum_{\mathbf{x}} \in \Omega$ is replaced with an integration: $\int_\Omega d\mathbf{x}$, and the binary segmentation $S^k(\mathbf{x})$ is replaced by the probability that a voxel $\mathbf{x}$ belongs to the ROI: $p(S^k(\mathbf{x}))$ or, using the level-set formulation, by $H_\epsilon(\phi^k(\mathbf{x}))$.

The explicit form of the energy functional in Eq. (11) is as follows:

$$\mathcal{E}(\phi^k) = \int_\Omega W_{UI} \left[H_\epsilon(\phi^k(\mathbf{x})) \log p_{in}(\eta_{in}^k) + H_\epsilon(-\phi^k(\mathbf{x})) \log p_{out}(\eta_{out}^k) \right] \quad (12)$$

$$+ W_{IL} \left[H_\epsilon(\phi^k(\mathbf{x})) \log p_{in}(I; \psi_{in}) + H_\epsilon(-\phi^k(\mathbf{x})) \log p_{out}(I; \psi_{out}) \right]$$

$$+ W_{LEN} |\nabla H_\epsilon(\phi^k(\mathbf{x}))| d\mathbf{x}.$$

$$(13)$$

The last term in Eq. (12), i.e. $|\nabla H_\epsilon(\phi^k(\mathbf{x}))|$, is known, in the level-set literature as the smoothness or regularization term [2]. The relation to $-\log p(S^k)$ where $p(S^k)$ is the prior in Eq. (1) is shown in [16]. The weights W_{UI}, W_{IL}, W_{LEN}, are positive scalars that balance the contribution of each term. In the first, fully automatic segmentation phase W_{UI} is set to zero. Similarly, W_{LEN} can be set to a lower value (or zero) in the presence of user interaction.

We look for the segmentation ϕ^k that optimizes the energy functional Eq. (12) via a gradient descent process: $\phi_{\tau+\Delta\tau}^k = \phi_\tau^k + \Delta\tau \frac{\partial \phi^k}{\partial \tau}$, where, ϕ_τ^k is the level-set estimate at iteration τ in step k. The gradient descent step $\frac{\partial \phi^k}{\partial \tau}$ is derived from the first variation of the functional above and determines the evolution of the segmentation:

$$\frac{\partial \phi^k}{\partial \tau} = \delta_\epsilon(\phi^k)\{W_{UI} \left[\log p_{in}(\eta_{in}^k) - \log p_{out}(\eta_{out}^k) \right] \quad (14)$$

$$+ W_{IL} \left[\log p_{in}(I; \psi_{in}) - \log p_{out}(I; \psi_{out}) \right]$$

$$+ W_{LEN} \operatorname{div}(\frac{\nabla \phi^k}{|\nabla \phi^k|}) \},$$

where, div is the divergence operator and $\delta_\epsilon(\phi^k)$ is the derivative of $H_\epsilon(\phi^k)$ with respect to $\phi^k : \delta_\epsilon(\phi) = \frac{dH_\epsilon(\phi)}{d\phi} = \frac{1}{4\epsilon}\operatorname{sech}^2(\frac{\phi}{2\epsilon})$. The scalar maps $p_{in}(\eta_{in}^k), p_{out}(\eta_{out}^k)$ are updated once after each segmentation step, if a user input is provided. The GMM of the image intensities ψ_{in} and ψ_{out} are re-estimated in correspondence to ϕ_τ^k, which determines the ROI boundaries, using expectation maximization (EM) [3]. Therefore, $p_{in}(I; \psi_{in}), p_{out}(I; \psi_{out})$ are calculated at every iteration, based on the updated intensity distribution parameters.

4   Experimental Results

We exemplify our method on cerebral hemorrhages (CH) segmentation of brain CT scans. Segmentation is necessary for an accurate estimate the hemorrhage volume for further medical treatment. Scans where acquired with Philips Brilliance CT 64 system without radiocontrast agents injection. The data resolution is $512 \times 512 \times [90 - 100]$ with voxel size of 0.48mm $\times$ 0.48mm $\times$ 3mm, with 1.5 $[mm]$ overlap in the axial direction. We tested 15 cases of CH seizures. Qualitative results of some of the cases are shown in Fig. 1. Quantitative results,

which were obtained with respect to manual annotation of clinical experts are shown in Table 1. This includes the Dice scores [4], sensitivity, specificity, accuracy of the segmentation results obtained by the proposed method. The reader is encouraged to view our demo, located at http://youtu.be/aRn7TulrWLY which demonstrates the UI segmentation using the GUI.

Table 1. Dice, Sensitivity, Specificity and Accuracy. Averages of 15 different slices from different patients.

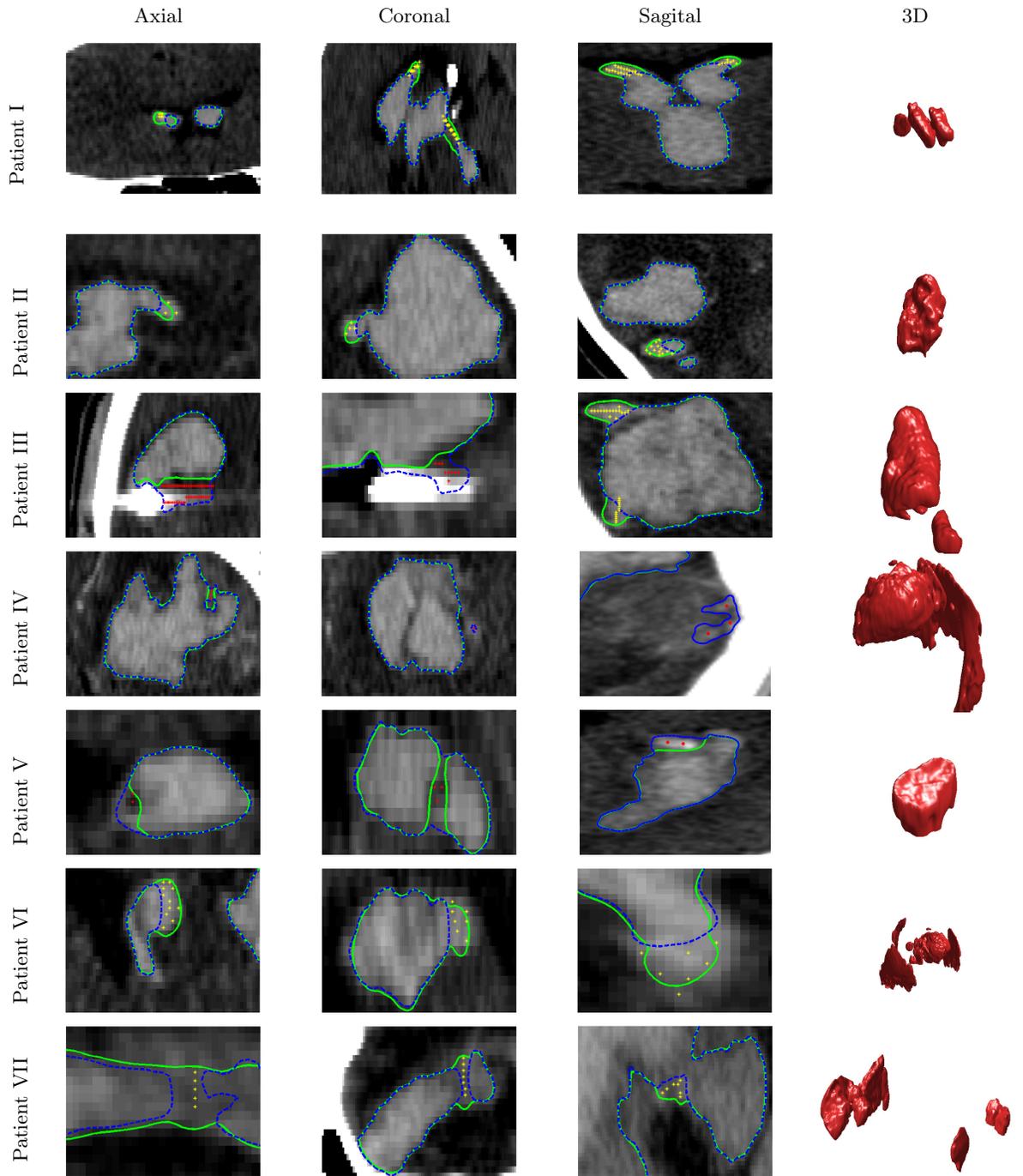| Phase vs Score	Dice	Sensitivity	Specificity	Accuracy
Automatic	0.874 ± 0.034	0.864 ± 0.073	0.99 ± 0.0019	0.996 ± 0.0033
With user interaction	0.905 ± 0.027	0.87 ± 0.063	0.999 ± 0.0003	0.997 ± 0.0022

5   Conclusions

Image segmentation in necessary for measuring ROI features, such as pose, size, location, texture, etc., which may be critical for diagnosis, treatment planning and image guided therapy. User interaction is therefore essential for resolving classification ambiguities and errors due to imaging artifacts, poor contrast and noise. We proposed a novel probabilistic model for a semi-automatic segmentation in which the user interacts with the segmentation algorithm providing spatial information. Accurate segmentation results, with a minimal user input are obtained for the segmentation of CH in brain CT images. According to radiologists evaluation, our method provides an intuitive interface and a more accurate segmentation results, when compared to another, commercially distributed, toolbox. Moreover, the overall number of interaction steps needed was not higher than two. As no prior information is assumed and the image intensity distribution is learned throughout the segmentation, the proposed method can be used for a wider range of segmentation applications and imaging modalities such as MRI or ultrasound. A toolbox with a user friendly GUI was developed as a part of this research project and is demonstrated in http://youtu.be/aRn7TulrWLY.

References

1. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.
2. T.F. Chan and L.A. Vese. Active contours without edges. *TIP*, 10(2):266–277, 2001.
3. A. Dempster, N. Laird, and D. Rubin. Maximal likelihood form incomplete data via the EM algorithm. *Proceedings of the Royal Statistical Society*, 39:1–38, 1977.
4. L. Dice. Measure of the amount of ecological association between species. *Ecology*, 26(3):297–302, 1945.
5. A. Falcao *et. al.* User-steered image segmentation paradigms - live wire and live lane. *Graph. Models and Image Proc.*, 60(4):233–260, 1998.

6. Y. Gao, R. Kikinis, S. Bouix, M.E. Shenton, and A. Tannenbaum. A 3D interactive multi-object segmentation tool using local robust statistics driven active contours. *Medical image analysis*, 16(6):1216–1227, 2012.

7. P. Karasev, I. Kolesov, K. Fritscher, P. Vela, P. Mitchell, and A. Tannenbaum. Interactive medical image segmentation using PDE control of active contours. 2013.

8. M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, Jan. 1988.

9. J. Liang, T. McInerney, and D. Terzopoulos. United snakes. 10(2):215–233, 2006.

10. K. McGuinness and N. E. OĊonnor. Toward automated evaluation of interactive segmentation. *CVIU*, 115(6):868–884, 2011.

11. E. N. Mortensen and W. A. Barrett. Interactive segmentation with intelligent scissors. *Graph. Models and Image Proc.*, 60(5):349–384, 1998.

12. C. Nieuwenhuis and D. Cremers. Spatially varying color distributions for interactive multilabel segmentation. *PAMI*, 35(5):1234–1247, 2013.

13. S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

14. L. Paulhac, J-Y. Ramel, and T. Renard. Interactive segmentation of 3D images using a region adjacency graph representation. In *Image Analysis and Recognition*, pages 354–364. 2011.

15. J. S. Prassni, T. Ropinski, and K Hinrichs. Uncertainty-aware guided volume segmentation. *Visualization and Computer Graphics*, 16(6):1358–1365, 2010.

16. T. Riklin Raviv, K. Van Leemput, B.H. Menze, W.M. Wells, and P. Golland. Segmentation of image ensembles via latent atlases. *Medical Image Analysis*, 14(5):654–665, 2010.

17. C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *SIGGRAPH*, 2004.

18. W. Yang, J. Cai, J. Zheng, and J. Luo. User-friendly interactive image segmentation through unified combinatorial user inputs. *TMI*, 19(9):2470–2479, 2010.

19. F. Zhao, , and X. Xie. An overview of interactive medical image segmentation. *Annals of the BMVA*, 7:1–22, 2013.

Fig. 1. Segmentations of the CH of different patients. Fully automatic segmentation in blue; Final segmentation with user-interaction in green. User clicks are in red (false positive) and yellow (false negative).