Interactive Multi-organ Segmentation Based on Multiple Template Deformation

Romane Gauriau[1,2], David Lesage[1], Mélanie Chiaradia[3], Baptiste Morel[4], and Isabelle Bloch[2]

[1] Philips Research MediSys, Paris, France
[2] Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, Paris, France
[3] H. Mondor Hospital APHP, Medical Imaging Department, Créteil, France
[4] A. Trousseau Hospital APHP, Radiology Department, Paris, France

Abstract. We present a new method for the segmentation of multiple organs (2D or 3D) which enables user inputs for smart contour editing. By extending the work of [1] with user-provided hard constraints that can be optimized globally or locally, we propose an efficient and user-friendly solution that ensures consistent feedback to the user interactions. We demonstrate the potential of our approach through a user study with 10 medical imaging experts, aiming at the correction of 4 organ segmentations in 10 CT volumes. We provide quantitative and qualitative analysis of the users' feedback.

1 Medical Motivation and Overview

Despite constant improvements of fully automatic segmentation approaches, perfect accuracy remains unreachable in many image processing scenarios, especially when inter- and intra-patient variabilities are important. In a clinical context, the possibility of incorporating user corrections is particularly valuable. From a user point of view, the interactions should be: (i) simple, easy to perform, (ii) fast (ideally with real-time feedback), (iii) intuitive (well-behaved algorithm feedback). Designing efficient and user-friendly algorithms meeting these criteria is particularly difficult.

Many works on interactive segmentation can be found in the literature. For instance, the live wire technique [2] is a highly interactive approach, close to fully manual 2D delineation. This approach can be extended to 3D and performed in real-time [3], but it remains very time-consuming for the end user. Various methods aim at optimizing globally an energy taking into account image information and user-provided initializations (e.g. through strokes). This problem is often tackled within discrete optimization frameworks [4,5,6,7,8]. The different formulations found in the literature propose different properties in terms of robustness, speed and sensitivity to initialization. Image partitioning from user inputs can also be formulated as a continuous variational problem [9] and include more global and contextual information. Globally optimizing an energy that changes with each new user input can have counter-intuitive effects, as the algorithm *forgets* previous results while not putting particular emphasis on the

N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 55–62, 2015.
DOI: 10.1007/978-3-319-24574-4_7

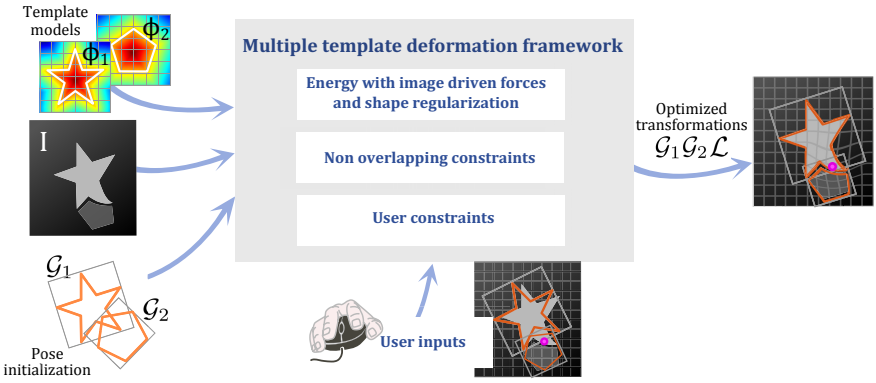

Fig. 1. Illustration of the framework principle on a toy example with two objects.

latest inputs. The non-convexity of some segmentation formulations can be exploited to derive sequential approaches [10,11]. After each user interaction, the contour evolves towards a local minimum of the new energy, starting from its previous location. With these methods, the impact of new inputs remains global. Resulting contours may change at locations distant from the latest inputs. Few methods were proposed for more local corrections [12,13] but they generally do not guarantee shape consistency. Finally, very few works are dedicated to the simultaneous segmentation of multiple objects [14], even less so in 3D [15].

In this work, we propose a framework for multi-organ interactive segmentation with: (i) simple interactions (point-wise mouse clicks), (ii) fast and *on the fly* user interactions, (iii) intuitive results (good tradeoff between user input and image information). We rely on the multiple template deformation framework of [1] for its efficiency and robustness, with shape priors and non-overlapping constraints. We extend it with user inputs expressed as hard optimization constraints (Sec.2). As an important refinement, we show how to handle user interactions in a spatially local fashion. Our approach is evaluated through a user study with 10 medical imaging experts, aiming at the correction of 4 organ segmentations in 10 CT volumes (Sec.3). The qualitative and quantitative feedback highlights the user-friendliness of our framework, which, we believe, is a decisive criterion towards its suitability to clinical workflows.

2 Methodology

This work is based on the multiple template deformation framework of [1]. We extend it with user constraints and propose a fast numerical optimization making possible real-time user interactions. The approach is illustrated in Fig.1.

2.1 Multiple Implicit Template Deformation with User Constraints

We denote an image $I : \Omega \to \mathbb{R}$ where $\Omega \in \mathbb{R}^d$ is the image domain ($d = 3$ in this work). For each object indexed by $n \in [\![1, N]\!]$ we associate an implicit shape

template $\phi_n : \Omega_n \to \mathbb{R}$ where Ω_n are the template referentials (ϕ_n is positive inside and negative outside the contour). In general the implicit shape template is a distance function whose zero level corresponds to the contour. For each object we define the transformations $\psi_n : \Omega \to \Omega_n$ that map back the image domain to the template domains. Each of these transformations is advantageously decomposed as $\psi_n = \mathcal{G}_n \circ \mathcal{L}$, where $\mathcal{G}_n : \Omega \to \Omega_n$ corresponds to the pose of object n (e.g. a similarity transformation) and $\mathcal{L} : \Omega \to \Omega$ corresponds to the local deformation common to the set of templates in the domain Ω.

The approach aims at finding the transformations $\mathcal{G}_n$ and $\mathcal{L}$ that best fit the template onto the image while following image-driven forces (specific to each object and defined by $f_n : \Omega \to \mathbb{R}$) and not deviating too much from the original shape. To prevent the objects from overlapping, specific penalizations are added on template pairs intersections (Eq.2) after pose transformation $\mathcal{G}_n$ [1]. The corresponding energy equation is given below (Eq.1), where H is the Heaviside function, λ is a constant balancing the shape prior and U is a reproducing kernel Hilbert space defined by a Gaussian kernel. We use the same forces f_n as in [1] integrating both region intensities and edge information.

$$\min_{\mathcal{G}_1,..\mathcal{G}_N,\mathcal{L}} \left\{ E(\mathcal{G}_1,..,\mathcal{G}_N,\mathcal{L}) = \sum_{n=1}^{N} \left(\int_\Omega H(\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(x)).f_n(x)dx \right) + \frac{\lambda}{2}\|\mathcal{L} - Id\|_U^2 \right\} \tag{1}$$

subject to

$$\forall (i,j) \in [\![1,N]\!]^2, i<j,\ C_{i,j} = \int_\Omega H(\phi_i \circ \mathcal{G}_i(x)) H(\phi_j \circ \mathcal{G}_j(x))dx = 0, \tag{2}$$

$$\forall n \in [\![1,N]\!], \forall q_n \in [\![1,K_n]\!],\ \gamma_{q_n}\ \phi_n \circ \mathcal{G}_n \circ \mathcal{L}(x_{q_n}) \geq 0,\ \gamma_{q_n} \in \{-1,1\} \tag{3}$$

We integrate user inputs into this framework as point-wise hard constraints similarly to [16]. These correspond to very simple interactions (mouse clicks). To modify the contour of object n, the user can add a point x_{q_n} *outside* (denoted as $\gamma_{q_n} = 1$) or *inside* ($\gamma_{q_n} = -1$) the object. Adding a point outside (respectively inside) the object indicates that the template should be deformed to include (respectively exclude) the point. These constraints can be expressed with regards to the sign of the deformed implicit template ϕ_n, as formulated in Eq. 3. For instance, for an outside point ($\gamma_{q_n} = 1$), the implicit template is constrained to become positive, i.e. to include point x_{q_n}: $\gamma_{q_n}\phi_n \circ \mathcal{G}_n \circ \mathcal{L}(\mathbf{x}_{q_n}) \geq 0$.

2.2 Numerical Optimization

To optimize the constrained problem of Eq. 1-3 we do not use the penalty method as in [1], as it may suffer from instability due to ill-conditioning. Instead, we use the augmented Lagrangian scheme presented in [17] to turn the problem into a series of unconstrained minimizations:

$$\min_{\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L}} \left\{ \hat{E}_k = E + \sum_{\substack{1\leq i\leq N \\ i<j\leq N}} h(C(\mathcal{G}_i,\mathcal{G}_j),\alpha_{i,j}^k,\mu_k) + \sum_{\substack{1\leq n\leq N \\ 1\leq q_n\leq K_n}} h(\gamma_{q_n}\phi_n\circ\mathcal{G}_n\circ\mathcal{L}(x_{q_n}),\alpha'_{q_n},\mu'_k) \right\} \tag{4}$$

where we denote $E = E(\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L})$ and $\hat{E}_k = \hat{E}_k(\mathcal{G}_1,...,\mathcal{G}_N,\mathcal{L})$, $\alpha_{i,j}^k$ and α'_{q_n} are the Lagrange multipliers, μ_k and μ'_k are the penalty parameters of the constraints, and h is the function defined by:

$$h(c,\alpha;\mu) = \begin{cases} -\alpha c + \frac{\mu}{2}c^2 & \text{if } c - \frac{\alpha}{\mu} \geq 0, \\ -\frac{\alpha^2}{2\mu} & \text{otherwise.} \end{cases} \tag{5}$$

The unconstrained energy of Eq.4 can then be optimized following a gradient descent. At each optimization step k, the Lagrange multipliers are fixed, the energy is optimized and new Lagrange multipliers estimates can be obtained for the optimization step $k+1$. As in [1], the parameters of the transformations $\mathcal{G}_n$ and $\mathcal{L}$ are updated jointly and iteratively.

Efficient Implementation. Note that the gradients of the energy can be efficiently computed as: (i) integrals over the volume can be turned into integrals over surfaces, (ii) many terms are only needed near the zero level of the implicit functions, (iii) a collision detection step can be added to prevent the systematic computation of the non-overlapping constraints, (iv) if a constraint gets verified then it is not optimized. For instance the automatic segmentation of 6 organs in a typical abdominal CT takes about 30 seconds to converge in the absence of user corrections (including localization). New user inputs add relatively localized constraints taking minimal effort to satisfy, in general within a few seconds.

Convergence. The optimization procedure ensures the convergence towards a local minimum of the (non-convex) energy, which is not guaranteed to be the global minimum. In our applicative scope, this turns into an advantage. In complex medical imaging tasks, the global minimum rarely corresponds to the exact desired result. Intuitively, user constraints will quickly drive the segmentation to the desired local minimum. Note also that contradictory constraints can be easily detected and mitigated in practice.

2.3 Enhancing the Framework for Local Contours Editing

When correcting the contours of pre-segmented objects, a user may expect the impact of his inputs to remain spatially local. A proper algorithm behavior would take into account the user inputs while relying on the image information in a ROI around the user input location. In such a case we suppose that the objects are already correctly positioned in the image and that only local deformations occur. Hence we propose a new formulation of the energy E of Eq.1:

$$E(\mathcal{L}) = \sum_{n=1}^{N} \left\{ \int_\Omega \left(K_\sigma * \sum_{q_n=1}^{K_n} \delta_{q_n}(x) \right) H(\phi_n\circ\mathcal{L}(x)).f_n(x)dx \right\} + \frac{\lambda}{2}\|\mathcal{L} - Id\|_U^2 \tag{6}$$

where K_σ is a Gaussian kernel with fixed width (in practice 2-3cm) and $\delta_{q_n} = \delta(x - x_{q_n})$ (δ is the Dirac distribution). With this new energy, the image-driven forces act in the neighborhood of the user inputs only. Note that the shape contours remain consistent.

The numerical optimization with this new energy equation is similar to the one presented in Sec.2.2, except that the pose transformations are not optimized and the non-overlapping constraint term is reduced to the empty set.

2.4 Flexibility of the Framework

The method proposed is very flexible and can be adapted to different usages. Any type of image-driven forces can be implemented. The algorithm can work in automatic mode (given an initialization of the models, e.g. with regression localization as in [1]) or with user inputs. The user constraints can be added while the algorithm is running, which allows for live interactions[1].

3 A Study for the Evaluation of the User Interactions

The idea behind our experiments is to *reproduce* a clinical context where the clinicians could use automatic segmentation results (given from the original framework [1], possibly run off-line) and correct them with our method with local corrections (energy of Eq.6) if needed.

Material. Our database is composed of 156 3D CT images coming from 130 patients with diverse medical conditions and from different clinical sites (which implies different fields of view, resolution etc.). Slice and inter-slice resolutions range from 0.5 to 1 mm and from 0.5 to 3 mm, respectively. The organs of interest have been manually segmented on all the database. The database has been split randomly into 50 and 106 volumes for training (localization part) and testing, respectively. Our method was implemented in C++.

A Simple Interface. The interface is made as simple as possible. There is one button to activate the corrections and one button to remove the last correction. Once the correction button is pressed, the algorithm runs continuously and the user can add point constraints in any orthogonal view and at any moment without waiting. The right mouse button is used to select the organ to correct and the left one is used to add a point constraint in the volume. A left click oustide the selected object will attract the contour (*inside* constraint) while a left click inside the object will move away the contour (*outside* constraint).

Protocol. To ensure clinical-like conditions we propose to use our framework without user constraints to segment 4 abdominal organs (liver, kidneys and spleen) on our database of 106 CT volumes. The volumes are sorted with regards

[1] Demonstration video: `http://perso.telecom-paristech.fr/~gauriau/`

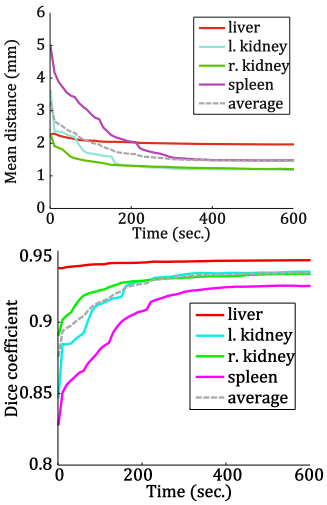

Fig. 2. Average mean distance and dice results per image in function of the correction time.

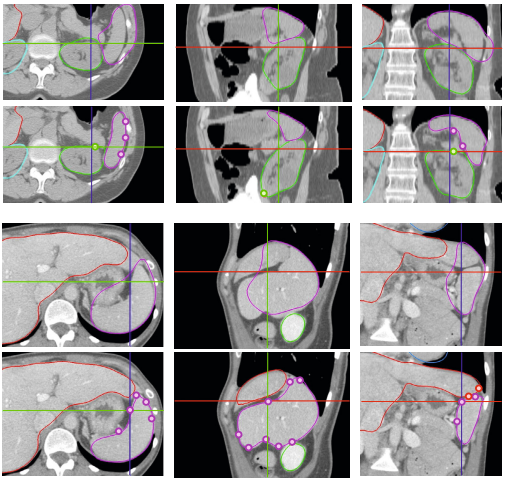

Fig. 3. Examples of results in two different volumes, after automatic segmentation (1st and 3rd lines) and after user constraints (2nd and 4th lines) in the three orthogonal views.

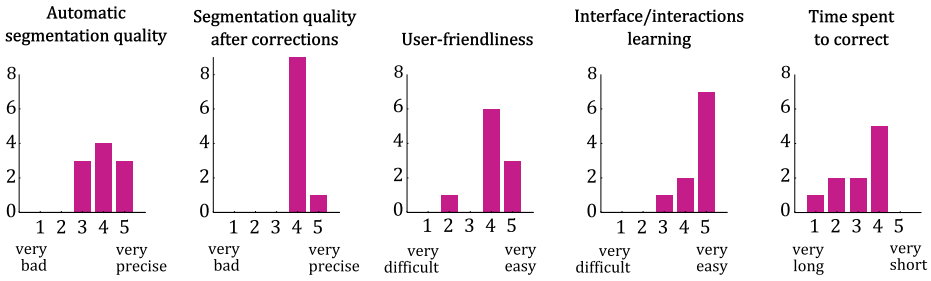

Fig. 4. Feedback form results from the 10 experts.

to their Dice coefficient to the ground truth and we select 10 volumes uniformly spread on this basis. We have then a sample of CT volumes representing the variability of the automatic segmentation results that we could find in a clinical context with an automatic method. Then 10 experts of the medical imaging domain (among them two radiologists) were asked to correct the results of the automatic segmentation in these volumes. Note that they were not asked for extreme precision. None of them knew the interface and the algorithm before using it. They only had few minutes to understand the tool before starting the experiments. Note that the experiments were performed on different computers with various configurations. During the experiments, each click was recorded and intermediate segmentation results were saved. At the end of the experiments, each user was asked to fill a form and give their feedback.

Results of the User Study. On average the users spent 345 seconds per volume (median: 228s). Figure 2 gives, for each organ, the average mean distance and dice coefficient according to the ground truth in function of the correction time. Note that the accuracy converges rather rapidly. After about 300 seconds an average distance of 1.5mm is reached. Considering that there are 4 organs to look at and correct, this remains reasonable in terms of time. The liver reaches a mean distance of 2mm explained mainly by the user variability and tolerance (e.g. sometimes the user includes or not part of the aorta and the inferior vena cava). Figure 3 shows examples of results before and after corrections. We observe that small corrections as well as important deformations can be handled by the algorithm. Finally, Fig. 4 shows the results of the feedback survey. The users seem satisfied with the final segmentation results. They confirm that the tool is very easy to learn. They suggest that some effort should be spent on the reactivity of our prototype, which is expected to be much improved with further code optimization. They also highlighted one limit of this approach: as large deformations are penalized (with the regularization term), large errors are more difficult to correct.

4 Conclusion

In this article we presented a fast and robust multi-organ segmentation method integrating user inputs in an intuitive manner. While benefiting from the original template deformation framework of [1], the efficient numerical optimization scheme with augmented Lagrangian results in a fast and stable algorithm allowing live user interactions. We also proposed a new formulation of the energy to take into account user inputs in a spatially local fashion. This extended framework can be used to build a complete and coherent tool chain: organs can be automatically segmented off-line (in about 30 sec.) and the clinicians can correct these results if needed. Thanks to our study with 10 users, we showed that this tool is easy to learn and results in fast, coherent and accurate corrections. Our experiments gave us precious insight for possible improvements. First, our local correction scheme could be made more adaptive, e.g. by adapting the width of the kernel K_σ to the distance between the user input and the object contour. Second, we are working on improving the performance of our software through code optimization and parallelization, as reactivity is a crucial aspect of such clinical tools. Finally, we saw that our approach may have difficulties with large errors, requiring large deformations to be corrected. We are currently exploring refinements better suited to such use cases.

Acknowledgments. This work is supported in part by an ANRT grant (008512012). We are very thankful to Vincent Auvray, Maxim Fradkin, Hélène Langet, Thierry Lefèvre, Paolo Piro and Jean-Michel Rouet for their participation in the study.

References

1. Gauriau, R., Ardon, R., Lesage, D., Bloch, I.: Multiple template deformation. application to abdominal organ segmentation. In: ISBI, pp. 359–362 (2015)
2. Mortensen, E., Morse, B., Barrett, W., Udupa, J.: Adaptive boundary detection using live-wire two-dimensional dynamic programming. In: IEEE Computers in Cardiology, pp. 635–638 (1992)
3. Falcao, A.X., Udupa, J.K., Miyazawa, F.K.: An ultra-fast user-steered image segmentation paradigm: live wire on the fly. IEEE TMI 19(1), 55–62 (2000)
4. Boykov, Y., Jolly, M.-P.: Interactive organ segmentation using graph cuts. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 276–286. Springer, Heidelberg (2000)
5. Grady, L.: Random walks for image segmentation. IEEE PAMI 28(11), 1768–1783 (2006)
6. Bai, X., Sapiro, G.: A geodesic framework for fast interactive image and video segmentation and matting. In: IEEE ICCV, pp. 1–8 (2007)
7. Criminisi, A., Sharp, T., Blake, A.: GeoS: Geodesic image segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 99–112. Springer, Heidelberg (2008)
8. Zhang, J., Zheng, J., Cai, J.: A diffusion approach to seeded image segmentation. In: IEEE CVPR, pp. 2125–2132 (2010)
9. Zhao, Y., Zhu, S.C., Luo, S.: Co3 for ultra-fast and accurate interactive segmentation. In: International Conference on Multimedia, pp. 93–102. ACM (2010)
10. Cremers, D., Fluck, O., Rousson, M., Aharon, S.: A probabilistic level set formulation for interactive organ segmentation. In: SPIE, vol. 6512 (2007)
11. Mory, B., Ardon, R., Yezzi, A.J., Thiran, J.: Non-euclidean image-adaptive radial basis functions for 3D interactive segmentation. In: IEEE International Conference on Computer Vision, pp. 787–794 (2009)
12. Grady, L., Funka-Lea, G.: An energy minimization approach to the data driven editing of presegmented images/volumes. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 888–895. Springer, Heidelberg (2006)
13. Harrison, A.P., Birkbeck, N., Sofka, M.: IntellEditS: Intelligent learning-based editor of segmentations. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 235–242. Springer, Heidelberg (2013)
14. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: IEEE ICCV, vol. 1, pp. 105–112 (2001)
15. Fleureau, J., Garreau, M., Boulmier, D., Leclercq, C., Hernandez, A.: 3D multi-object segmentation of cardiac MSCT imaging by using a multi-agent approach. In: IEEE Annual International Conference, pp. 6003–6006. EMBS (2009)
16. Mory, B., Somphone, O., Prevost, R., Ardon, R.: Real-time 3D image segmentation by user-constrained template deformation. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 561–568. Springer, Heidelberg (2012)
17. Nocedal, J., Wright, S.J.: Numerical optimization. Springer, New York (2006)