

Collaborative Multi Organ Segmentation by Integrating Deformable and Graphical Models^{*}

Mustafa Gökhan Uzunbaş¹, Chao Chen¹, Shaoting Zhang¹, Kilian M. Pohl²,
Kang Li³, and Dimitris Metaxas¹

¹ CBIM, Rutgers University, Piscataway, NJ, USA

² University of Pennsylvania, Philadelphia, PA, USA

³ Dept. of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

Abstract. Organ segmentation is a challenging problem on which significant progress has been made. Deformable models (DM) and graphical models (GM) are two important categories of optimization based image segmentation methods. Efforts have been made on integrating two types of models into one framework. However, previous methods are not designed for segmenting multiple organs simultaneously and accurately. In this paper, we propose a hybrid *multi organ* segmentation approach by integrating DM and GM in a coupled optimization framework. Specifically, we show that region-based deformable models can be integrated with Markov Random Fields (MRF), such that multiple models' evolutions are driven by a maximum a posteriori (MAP) inference. It brings global and local deformation constraints into a unified framework for simultaneous segmentation of multiple objects in an image. We validate this proposed method on two challenging problems of multi organ segmentation, and the results are promising.

1 Introduction

Segmenting anatomical regions from medical images has been studied extensively and it is a critical process in many medical applications. Two important categories of optimization based image segmentation methods are Graphical Model (GM) based [13,1,11] and Deformable Model (DM) based [9,3,14,12,6,10] methods. Both categories are able to achieve satisfactory results by combining the confidence based on local evidence such as colors and textures and the confidence based on global evidence such as the smoothness of the boundary. However, each model has its own strengths and weaknesses.

GM methods can reach the global optimum because of efficient algorithms inspired by the max-flow min-cut theorem. Although in most cases, the used algorithm such as α -expansion [2] is still a local search algorithm, one can claim the solution is approximately optimal due to the extremely large local search neighborhood. The unique solution of GM methods could achieve reasonably good segmentation quality, but sometimes misses the fine details, which could be critical in medical context. On the other

^{*} This grant was partially supported based on funding from the following grants NIH-R01-HL086578, NIH-R21-HL088354 and NSF-MRI-1229628.

hand, DM methods, given a reasonably good initialization, would have the flexibility to deform to nearby local minimum of the energy which could achieve segmentation with nice details. However, due to the complicated energy model, DM methods use gradient descent method, and thus could be trapped in local minima that are very far from the ground truth.

Based on these observations, several works combining the two models have been introduced in the computer vision and the medical imaging literature. The general idea is to use graphical model at the gradient computation step of the deformable model, so that the deforming contour is less likely to be trapped at local minima that is far from the ground truth. Chen et al. [4] deform the contour by iteratively solving a graphical model optimization problem. At each iteration, the graphical model enforces that the deformation should respect priors trained offline and also should not be too far from the current contour. Huang et al. [7] formulated the segmentation task as a joint inference problem of contour and pixel labeling so that the two models are tightly coupled. At each step of the iteration, a graphical model is constructed. Instead of the MAP, the marginals of the graphical model are used as part of the gradient in deformable model.

However, these methods only use binary labeling when constructing the graphical model, which cannot be used directly for segmenting multi organs. One could extend these methods to segment multi organs straightforwardly by constructing a binary graphical model for each deforming contour. In this setting, the foreground and background labels correspond to the current concerned organ and the union of the rest, respectively. Unfortunately, this method would have difficulty to distinguish the two neighboring regions if they only have slightly different intensity distributions but are very different from the remaining regions (see an example in Figure 1).

In this paper, we propose a hybrid model to naturally cope with the *multi organ* segmentation problem by integrating GM and DM methods. Our main idea is to formulate a multi label

graphical model problem at each iteration, and use the MAP inference result as part of the gradient (i.e., external forces) of deformable model. Such inference task could be solved efficiently using α -expansion algorithm [2]. The main advantage of our model is that in the multiple labeling graphical model, regions of different labels bring high level constraint naturally and implicitly to the external forces of deformable models. Therefore, two similar neighboring regions can still be separated easily, and the resulting regions are more accurate as per the ground truth. See Figure 1 for an illustration and Section 3 for more comparisons. Besides the accuracy and robustness for multi organ segmentation, this proposed model does not need any offline learning (in contrast to [4]), has few parameter, and contains the advantages of both GM and DM

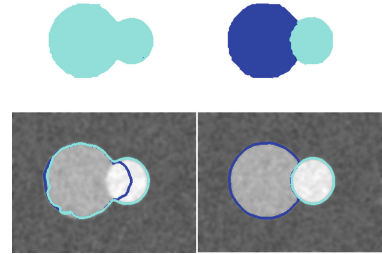


Fig. 1. Top: binary graphical model (left) would produce wrong regions for either labels, but multi-label graphical model (right) is correct; Bottom: comparison of combining DM with binary GM (left) and multi-label GM (right).

methods. We have applied this proposed method to two challenging clinical applications, i.e., knee joint bone segmentation and cardiac segmentation, and achieved promising results.

2 Methodology

The goal of our segmentation method is to find multiple regions with smooth and closed boundaries. We start with the overall energy functional for m models forming a set $\mathbf{C}$ as:

$$E(\mathbf{C}) = E_{int}(\mathbf{C}) + \sum_{i=1}^m E_{ext}^i(\mathcal{C}_i) \quad (1)$$

Here, the first term E_{int} is the smoothness term (see [9] and [3]) and E_{ext}^i is the data term for contour i . In contrast to the classical choice such as the difference from a constant value [12] or the negative log likelihood of a given distribution [14], we assume a region of interest (ROI) for contour i is given (R_i), and define the i -th external energy as

$$E_{ext}^i(\mathcal{C}_i) = \frac{1}{Vol} \iiint_{\Omega} (\Phi_{\mathcal{C}_i}(x) - \Phi_{\partial R_i}(x))^2 dx \quad (2)$$

Here Ω is the image domain and Vol is its volume. We denote $\Phi_{\mathcal{C}_i}$ and $\Phi_{\partial R_i}$ as the signed distance functions of the contour and the boundary of the region of interest, respectively. Intuitively, minimizing this term would pull the contour, $\mathcal{C}_i$, towards the boundary of ROI, ∂R_i . Our algorithm minimizes $E(\mathbf{C})$ using gradient descent method. At each iteration, we compute the gradient $\partial E / \partial \mathbf{C}$ and evolve the contours accordingly.

At each iteration, we define the ROIs by constructing a multi-label graphical model depending on the current contours $\mathbf{C}$. The MAP of the graphical model gives us the set of ROIs. We denote L as a labeling, which assigns to each pixel/voxel a label belonging to the label set $\mathcal{L} = \{1, \dots, m, m+1\}$, corresponding to regions inside the m contours and the background ($m+1$). We compute the labeling optimizing the conditional probability

$$L^* = \arg \max_L \{P(L \mid I, \mathbf{C})\} \quad (3)$$

The ROI R_i is then the set of voxels with label l^i in L^* . Assuming that the image data I and the deformable models $\mathbf{C}$ are independent and also conditionally independent given the labeling L , we define the posterior probability of labeling L :

$$P(L \mid I, \mathbf{C}) = \frac{P(I, \mathbf{C} \mid L)P(L)}{P(I, \mathbf{C})} = \frac{1}{P(L)} \cdot \frac{P(I \mid L)P(L)}{P(I)} \cdot \frac{P(\mathbf{C} \mid L)P(L)}{P(\mathbf{C})} \quad (4)$$

With the assumption that each labeling has equal prior probability we have

$$P(L \mid I, \mathbf{C}) \propto P(L \mid I) \cdot P(L \mid \mathbf{C}) \quad (5)$$

Here the negative log likelihood of $P(L \mid I)$ is the same as the energy in conventional multi-label graphical model, and $P(L \mid \mathbf{C})$ is the model shape prior, defined as $P(L \mid \mathbf{C}) = \prod_{j \in V} P(L_j \mid \mathbf{C})$, where V is the set of all voxels. The shape prior of each

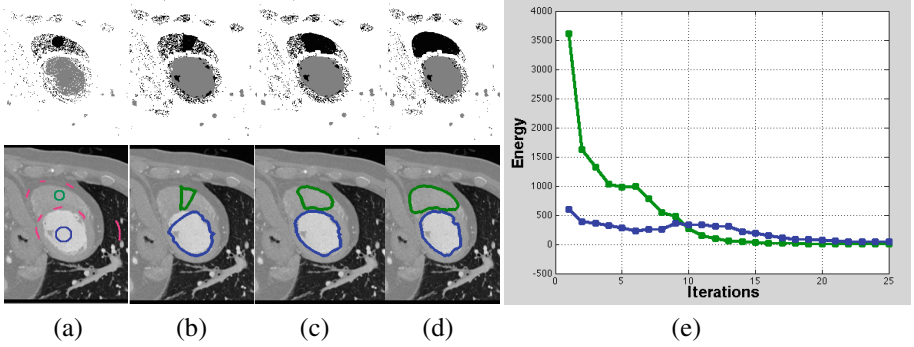


Fig. 2. Iterative, model constrained estimation of labels. (a) User provided model initializations and background cues (strokes in magenta). (b-d) multi-phase graph cut results and states of the deformable models at consecutive iterations. In top row, gray color represents the label for blue model (left ventricle) and black color represents the label for green model (right ventricle), white label represents the background. In bottom row, deformable models at consecutive iterations are shown. (e) Plot of E_{ext} energy values computed at each iterations.

individual voxel $P(L_j | \mathbf{C})$ is inversely proportional to the distance from the model. Let $i = L_j$, we have

$$P(L_j | \mathbf{C}) = P(L_j | \mathcal{C}_i) = \begin{cases} 1 & \text{if } \Phi_{\mathcal{C}_i}(j) \geq 0, \\ 1 - \frac{\|\Phi_{\mathcal{C}_i}(x_i)\|}{\|(\Phi_{\mathcal{C}_i})\|_\infty} & \text{otherwise} \end{cases}$$

According to the above definition, voxels which are closer to a model are more likely to belong to the label of that particular model. This leads to the final energy for graphical model as

$$-\log(P(L | I, \mathbf{C})) \propto E(L) = \sum_j^N (u_j(L_j) - \log(P(L_j | \mathbf{C}))) + \sum_{\substack{p, q \in \mathcal{N} \\ L_p \neq L_q}} b_{pq} \quad (6)$$

In the above equation, $u_j(L_j)$ is the cost of assigning label L_j to j^{th} voxel and are computed as $\|I_j - \mu_i\|^2 / \sigma_i$ where $i = L_j$, μ_i and σ_i are mean and standard deviation of the intensities inside the region enclosed by $\mathcal{C}_i$. b_{pq} is the typical binary term of MRF and is defined as $(|I_p - I_q|^2 / \sigma) \times \text{dist}(p, q)^{-1}$. For more details about construction of the graph and inference method we refer the reader to [11] and [2]. We solve the MAP inference problem given in Eq. 3 using α -expansion algorithm [2].

Integrating Deformable Models and Graph Cuts: According to the energy function defined in Eq. 1, the models are deformed under smoothness constraints and the attraction force coming from the ROIs. The minimization problem can be achieved by using an alternating minimization scheme where we do coordinate descent and split it into two problems: 1) fix multi phase labeling conditioned on the given models and image data; 2) locally deform the models minimizing the other energy terms such as smoothness

and image gradient etc. A single deformable model $\mathcal{C}^i$ is represented explicitly in terms of splines (2D) or meshes (3D) as in [8], [5] and deformed according to the deformable model dynamics explained there. Our segmentation process starts with initialization of the models $\{\mathcal{C}_1, \dots, \mathcal{C}_m\}$ for the foreground objects and providing markings (seeds) for the background. Before starting the iterative process, a graph is constructed only once with the desired connectivity (can be 8 in 2D or 26 in 3D) using the initial models and the seeds. The unary and compatibility potentials along with the model shape constraints are computed for the graph cut. Then we calculate the labeling according to the minimization of Eq. 6. Once we obtain the labels for each pixel, we select the ROIs that intersect the models, and for each model, we compute driving forces using the associated ROI (see Eq. 2). We continue this alternating process iteratively and at each iteration after the models are deformed we update the parameters μ , σ and $P(L_j | \mathbf{C})$. We continue deformation in the same scheme until convergence. In Fig. 2, we demonstrate this process for two foreground object segmentation. Fig. 2(a) shows the user initialized models for the foregrounds (blue and green) and strokes for the background (in magenta). Fig. 2(b-d) top row show the states of MRF labeling at consecutive steps. One can observe that, as the iterations continue not only the smoothness of the labels are enhanced but also the accuracy of the labeling gets better and better. Thus, E_{ext} provides more accurate driving force as the iterations continue. Fig. 2(e) shows the energy minimization process at each iteration which is calculated from Eq. 2. According to the plot, it is clear that the energy computed for each model is monotonically minimized and converges.

Implementation Detail: Multi phase graph cut method is not guaranteed to always return smooth and distinct region segments for each label, especially when target regions present similar/identical intensity properties. Due to this similarity, the calculated unaries could be in-distinctive and resulting labeling would not return structured segments. In such a case, the selected ROI per deformable model would not be accurate enough to drive the model correctly. To tackle this problem we develop an online label unification method which adaptively identifies the labels of the models to be merged. We do this unification operation according to the Kullback-Leibler divergence (KL) between the kernel density estimated intensity distributions underneath the models. In Fig. 3(a), the estimated distributions for each model region is shown. In this scenario, according to KL score, the algorithm decides to unify the yellow model label with the red model label, and the light blue label with the dark blue label. With the unification of the labels target label size becomes 3 including the background. In (b), we compare the resulting labels of multi phase graph cut w/o using the unification. As seen in the bottom image, the labeled regions are more smooth relative to the top image. Note that the label set is automatically shrunk to 2 foreground labels but this does not mean that number of models are changed. The models always select the binary ROI that intersect with itself. Then, the distance map of the ROI is used in the E_{ext} term as shown in Eq. 2. In Fig. 3(c) and (d), we show the effect of our unification mechanism to the deformation force. As seen in the bottom image the distance map computed for right ventricle and the left atrium are more accurate. Also in Fig. 3(e) we compare resulting segmentations at convergence to demonstrate the effect of our unification process. With this alternating minimization scheme, we take advantage of the strengths of these two methods. At each

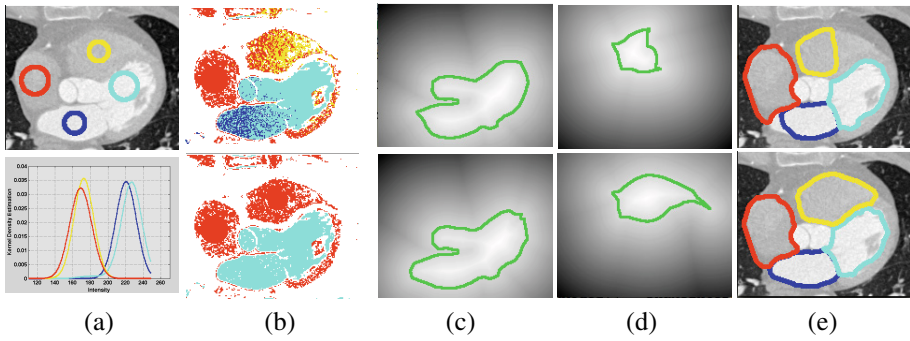


Fig. 3. Label unification and its effect to the segmentation: (a) Example kernel density estimate of the regions underneath the models. (b) Top row MAP-MRF labeling without unification (4 foreground labels). Bottom row, MAP-MRF labeling with unification (2 foreground labels). (c)-(d) Signed distance maps for two different models with (bottom) and without (top) unification. (e) Resulting segmentations obtained w/o unification.

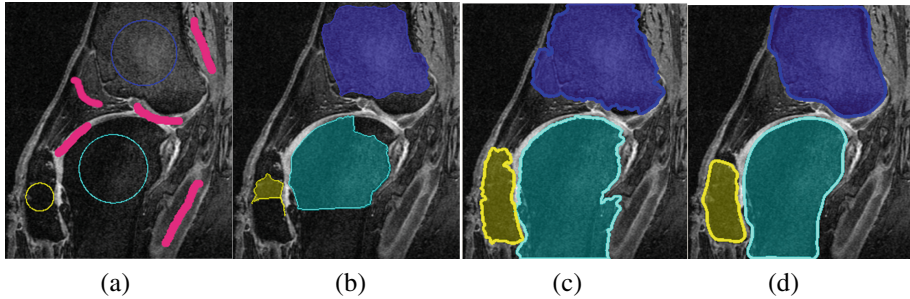


Fig. 4. Experiment on knee joint MR image. (a) Three model initializations and background cues (strokes) for segmentation of tibia, femur and patella. (b) [8]. (c) [7]. (d) Our method .

iteration the models are updated locally with globally computed forces and the global parameters of the MAP-MRF are updated locally. Moreover as the models start getting close to the actual target object shape, the system reaches convergence very precisely.

3 Experiments

We validate this proposed method on two multi organ segmentation applications. Our method is compared with two relevant approaches, 1) Metamorphs [8] which integrates texture information into deformable models, and 2) a graphical model coupling MRFs and deformable models [7]. They are evaluated on both MR and CT data sets whose ground truths were manually annotated by clinical experts. Our algorithm was implemented in MATLAB with C programming extensions. We tested the algorithm on a quad core (3.4 GHz) computer with 8Gb of memory.

Knee Joint Bone Segmentation: We segment knee joint bones femur, tibia and patella from 23 MR scans. The data scan protocol consists of 3D DESS scan with water excitation having $0.36 \times 0.36 \times 0.70$ mm voxel size. The qualitative comparisons are shown in Fig. 4. Due to the intensity heterogeneity inside the bone structures (particularly in regions close to the cartilage), [8] does not perform well and stuck in local minima as it uses the local online intensity modeling. Starting from the same initializations, our method converges to the final state for all 3 organs within 30 iterations (~ 6.5 sec/iter) while [8] stops at the final stage after 18, 23 and 38 iterations (~ 8 sec/iter) for patella, tibia and femur, respectively. For fair comparison, we tune deformation parameters (i.e. smoothness, image gradient, balloon) for method [8] to achieve the best performance and keep them exactly same for our method.

To compare with [7], we also carefully selected its parameters in order to obtain the best results. We observed that the Expectation Maximization approach in [7] performs better than [8] when updating the attraction force for the deformable models. However, it takes more iterations due to its narrow band limitation also running time per iteration takes ~ 20 sec which is 3 times the running time of our approach. In addition, our method takes advantage of multi phase MRF labeling and the ROIs per models are estimated more accurately. Thus, our hybrid approach gets out of local minima and also avoid possible leakages towards muscle regions.

Cardiac Segmentation: We evaluated our algorithm on segmenting the cardiac structures such as Right Atrium (RA), Right Ventricle (RV), Left Ventricle (LV), Left Atrium (LA) from a set of 15 CT volumes. The data scan protocol consisted of 3D CT scan with $1.0 \times 1.0 \times 1.0$ mm voxel size. The figure is shown in the supplementary materials due to page limitation. Compared to [8], our method performs slightly better in terms of avoiding leakages towards the heart muscle. In addition, the myocardium between LV and RV is identified better. For the LV case, the papillary muscles are nicely included into the segmentation owing to the smoothness factor of the graphical model and the parameter update scheme of the deformation. With our method, all models converge to the final state within 50 iterations (~ 4.9 sec/iter). Compared to [7], our method achieves better average accuracy of all organs since the background is identified well within the multi-region labeling scheme. Most significant accuracy differences between [7] and our method are observed for the RV and LV cases, due to local minima problems.

Table 1. Volumetric and surface errors for 23 MRI and 15 CT scans

mean $\pm$ std	Overlap Err. [%]	Avg. Surf. Dist. voxel
femur	7.34 ± 2.75	4.2 ± 2.42
tibia	6.27 ± 2.22	3.3 ± 0.57
patella	3.9 ± 1.37	1.4 ± 0.32
LA	7.12 ± 2.21	1.9 ± 1.45
LV	8.12 ± 1.35	3.1 ± 1.52
RA	5.88 ± 2.33	1.92 ± 0.32
RV	9.12 ± 2.9	3.1 ± 0.42

Table 1 shows quantitative results of our method in both applications. We reported the mean and standard deviation of voxel distances between segmented surfaces and ground truth, and volume overlap errors in proportions. Due to page limitations, we provide more comparisons in the supplementary materials. In general, our method achieves more accurate results than the other two hybrid approaches, and is also more efficient.

4 Conclusions and Future Work

In this paper we proposed a new hybrid multi object segmentation approach, in which deformable models and multi label graphical models are integrated into an alternating optimization framework. We integrate multi phase graph cut labeling into deformable model framework so that it provides the desired speed term for each deformable model to converge to the true boundary. We provide solutions for potential drawbacks of the two methods by combining the benefits of them to segment multiple objects efficiently and simultaneously using global and local constraints. We validated our method on medical images (MR, CT) and real-world images. As a future direction, we are currently working on speeding up the running time of our algorithm in 3D. We are planing to use a supervoxel approach which could reduce the graph cut processing time drastically. We also consider using conditional random fields model for a learning based multi object segmentation that might possibly use coupled prior information as well.

References

1. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation. *Int. J. Computer Vision* 70(2), 109–131 (2006)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. on PAMI* 23(11), 1222–1239 (2001)
3. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *Int. J. Computer Vision* 22(1), 61–79 (1997)
4. Chen, X., Udupa, J.K., Bagci, U., Zhuge, Y., Yao, J.: Medical image segmentation by combining graph cuts and oriented active appearance models. *IEEE Trans. Image Processing* 21(4), 2035–2046 (2012)
5. Cohen, L.D., Cohen, I.: Finite element methods for active contour models and balloons for 2d and 3d images. *IEEE Trans. on PAMI* 15, 1131–1147 (1991)
6. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Trans. PAMI* 23(6), 681–685 (2001)
7. Huang, R., Pavlovic, V., Metaxas, D.: A graphical model framework for coupling MRFs and deformable models. In: *CVPR*, vol. 2, pp. II-739–II-746 (2004)
8. Huang, X., Metaxas, D., Chen, T.: Metamorphs: Deformable shape and texture models. In: *CVPR*, vol. 1, pp. 496–503 (2004)
9. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *Int. J. Computer Vision* 1, 321–331 (1988)
10. Kohlberger, T., Uzunbaş, M.G., Alvino, C., Kadir, T., Slosman, D.O., Funka-Lea, G.: Organ segmentation with level sets using local shape and appearance priors. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II. LNCS*, vol. 5762, pp. 34–42. Springer, Heidelberg (2009)
11. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? *IEEE Trans. on PAMI* 26(2), 147–159 (2004)
12. Vese, L.A., Chan, T.F.: A multiphase level set framework for image segmentation using the mumford and shah model. *Int. J. of Computer Vision* 50, 271–293 (2001)
13. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *TMI* 20(1), 45–57 (2001)
14. Zhu, S.C., Yuille, A.: Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Trans. on PAMI* 18(9), 884–900 (1996)