

# Optimal Weights for Convex Functionals in Medical Image Segmentation

Chris McIntosh and Ghassan Hamarneh

Medical Image Analysis Lab  
School of Computing Science, Simon Fraser University, Canada  
{cmcintos,hamarneh}@cs.sfu.ca  
<http://mial.cs.sfu.ca>

**Abstract.** Energy functional minimization is a popular technique for medical image segmentation. The segmentation must be initialized, weights for competing terms of an energy functional must be tuned, and the functional minimized. There is a substantial amount of guesswork involved. We reduce this guesswork by analytically determining the optimal weights and minimizing a convex energy functional independent of the initialization. We demonstrate improved results over state of the art on a set of 470 clinical examples.

## 1 Introduction

Image segmentation is a key task in visual computing. For medical image analysis, segmentation is important for quantifying the progression of diseases and quantifying anatomical variation. In such applications, a high degree of accuracy is sought, as the anatomical variation itself can be small and thus consumed by segmentation error. For this and other reasons, the automatic segmentation of medical images remains a daunting task. Many segmentation approaches rely on the minimization of objective functions, including several landmark papers: from the seminal paper of Snakes for 2D segmentation [1] and other explicit models [2] to implicit models [3,4], graph approaches [5,6], and variants thereof.

Objective function-based methods are commonly built using five essential building blocks: (*i*) an objective function whose minima provide good segmentations; (*ii*) an appropriate shape representation; (*iii*) a set of parameters including weights to balance the competing terms of the energy functional; (*iv*) an initialization; and (*v*) a method for minimization, whether it be local or global, continuous or combinatorial.

Each of these common blocks is known to have certain challenges. In particular, the parameter setting, initialization, and minimization phases are well known to be problematic. Often there are unanswered questions: what if a different initialization was used, what about a different minimizer, what if different weights were used between competing terms of the energy functional? Hence, erroneous segmentations cannot be directly attributed to the energy functional or one of the many unknowns of the segmentation process. The goal is to reduce and ultimately remove these points of uncertainty.

Uncertainty with initializations and minimizers can be addressed by formulating problems as convex energy functionals over convex domains, or submodular in the case of combinatorial approaches. Papers dealing with this issue are now common ground for both continuous [7,8] and discrete [5,6] optimization, but they all have uncertainty stemming from the free weights in their energy functionals.

Uncertainty with energy functional weights can be addressed by determining the optimal weights for each image to be segmented, else another set of weights may exist that provides better results. Recently, we developed an analytical expression describing the optimal functional weights [9]. Our method solves for the optimal functional weights for a training set of image-segmentation pairs, and then infers the optimal parameters for a novel image via geodesic interpolation over the training set. Our results demonstrated the importance of not only using the optimal weights for a functional, but how those weights vary from image to image. However, the method was not without its drawbacks.

We build upon and extend our earlier work by addressing two key issues. First, our previous work was done using non-convex functionals. There is uncertainty that a different initialization or optimization process may have yielded improved results. Instead, in this work, we focus on convex functionals, ensuring global optima and thus removing uncertainty related to local minima and initializations. Second, our analytical expression for optimal parameters included an implicit weighting between its two competing terms; a weighting which we seek to address here. We remove the implicit weighting using a convex quadratic formulation under a linear constraint, and thus remove the uncertainty implied by it. Our latest results show significantly improved accuracy.

Though we are focusing on continuous functionals, a related field of approaches has come up in the study of combinatorial problems. The first set is based on recent advances in maximum margin estimation, wherein the parameters of the objective function are sought such that the highest scoring structures (in our case segmentations) are as close as possible to the ground truth [10,11,12,13]. However, in addition to being limited to combinatorial objective functions rather than continuous ones, these methods propose a fixed set of parameters for novel samples (in our case images), whereas we follow the direction of [9] using geodesic interpolation to infer the optimal parameters on a per-image basis. In other words, these works assume that a single set of parameters works for an entire test set. As shown in [9], this is often not the case in image segmentation, and greatly improved results can be obtained by adapting the parameters to the individual images (as we do). The second related direction, was introduced recently in [14]. Though this work is also restricted to combinatorial objective functions, an optimal parameter is indeed sought on a per-image basis. Given a parameter range, the method simultaneously solves the objective function for a set of parameters that bound how the parameters influence the solution. Each solution is then treated as a potential segmentation. They propose a number of heuristics, including user intervention, to select the best segmentation from a set of potential ones.

In what follows we describe the theory behind our method (Sec. 2), how we build convex functionals (Sec. refsec:energy), and how to analytically describe the optimal weights of a functional (Sec. 2.2). We then detail how to apply our method to novel images (Sec. refsec:methods) and validate our improvements over [9] (Sec. 4). Finally, we discuss our results and future work (Sec. 4).

## 2 Theory: Notations and Uncertainty in Segmentation

In order to more formally explain where the uncertainty lies, and how it needs to be addressed, we first give a more detailed view on the energy minimization based segmentation process. We define a gray-level image  $I$ , and its corresponding segmentation  $S$ . Then  $\mathbf{I} = \{I_1, I_2, \dots, I_N\}$  and  $\mathbf{S} = \{S_1, S_2, \dots, S_N\}$  are training sets of images and their corresponding, correct segmentations.

The first step is the identification of the form of the energy functional. It may be convex or non-convex, as can the shape space over which it is minimized. A common general form is  $E(S, I, \mathbf{w}) = w_1 \times \text{internal}(S) + w_2 \times \text{external}(S, I)$ . Notice the free parameter  $\mathbf{w} = [w_1, w_2]$ . Depending on its value, minima of  $E$  favor the internal energy, or the external energy.

The segmentation problem is to solve  $S^* = \arg \min_S E(S|I, \mathbf{w})$ , which involves choosing a  $\mathbf{w}$  and, depending on the nature of the energy functional, may also require training appearance and/or shape priors, and setting an initialization. A gradient descent-based solver is typically used but combinatorial approaches have also been explored for discretized versions of the problem [5]. Here we focus on continuous problems, and thus assume a gradient descent solver.

When using gradient descent, non-convexity can be quite problematic. There is no guarantee that another solution does not exist which better minimizes the energy, and thus is potentially a better segmentation. Ideally both functional and shape space are convex; guaranteeing globally optimal solutions.

Simply obtaining a global optima does not, however, guarantee a correct segmentation in the general case. If not appropriately set, the weights  $\mathbf{w}$  can cause significant error. Optimizing the weights has been shown to have dramatic effects; reducing error in large data sets by as much as 30% [9]. However, optimizing the weights by hand for even a single image can be a long and tedious task, with no real guarantee of obtaining the correct segmentation.

Instead of *guessing* the optimal weights, suppose we write a function  $\gamma(\mathbf{w}|I_j, S_j)$  evaluating how well weight  $\mathbf{w}$  works for a given image-segmentation pair  $(I_j, S_j)$ ; such that a parameter is deemed better when it causes  $S^*$  to approach  $S_j$ , i.e. the minimum of  $E$  to be the correct segmentation. Given  $S_j$ , we could then calculate the ideal weights for a particular image  $I_j$  by solving  $\mathbf{w}^* = \arg \min_{\mathbf{w}} \gamma(\mathbf{w}|I_j, S_j)$ . It is important that  $\gamma$  itself be convex or globally solvable in  $\mathbf{w}$ . If  $\gamma$  was not globally solvable, uncertainty would remain in that another  $\mathbf{w}^*$  may better minimize  $\gamma$ , and thus better segment the image. Similarly,  $\gamma$  can not contain free parameters, else those parameters would themselves introduce uncertainty; as was the case in [9].

### 2.1 Convex Energy Functionals

We make use of recent research into convex functionals for image segmentation, specifically that of Cremers et al. where a convex energy functional  $E$  is minimized over a convex shape space represented as probability maps, i.e.  $S(x) \in [0, 1]$  for all points  $x$  in the image domain  $\Omega$ , to yield a convex segmentation problem [15]. A shape model is then constructed via principal component analysis (PCA) on a set of training shapes forming a  $k$ -dimensional approximation to the shape space, with  $\alpha_1, \dots, \alpha_k$  eigen coefficients, a mean shape  $\bar{S}$ , and eigenvectors  $\psi_1, \dots, \psi_k$ . Shapes can now be reconstructed as  $S = \bar{S} + \sum_{i=1}^k \alpha_i \psi_i$ .

Writing  $S$  in terms of the vector of shape parameters  $\alpha = \{\alpha_i\}_{i=1}^k$ , convex  $E$  can be written as a sum of convex energy terms:

$$E(\alpha|I = I_j, \mathbf{w} = \hat{\mathbf{w}}) = \int_{\Omega} \hat{w}_1 J_1(\alpha|I_j) + \dots + \hat{w}_n J_n(\alpha|I_j) dx \tag{1}$$

for a fixed image  $I_j$  and arbitrary, fixed weights  $\hat{\mathbf{w}}$ , where  $J_i$  is a convex energy term and  $\mathbf{w} = [w_1, \dots, w_n]$  with  $w_i \in [0, 1]$  are weights. Consequently,  $E$  is a convex functional since the positively weighted sum of a set of convex terms is itself convex. For proofs of convexity and more details see [15].

Minimizing  $E$  optimally can then be performed via gradient descent on  $\alpha$  using derivative:  $E_{\alpha}(\alpha|I = I_j, \mathbf{w} = \hat{\mathbf{w}}) = \hat{w}_1 T_1(\alpha|I_j) + \dots + \hat{w}_n T_n(\alpha|I_j)$  where  $E_{\alpha}$  denotes the derivative of  $E$  with respect to  $\alpha$ , and  $T_i$  is the derivative of  $J_i$ th term. However, since  $\hat{\mathbf{w}}$  is arbitrary nothing can be said about its optimality for the particular image  $I_j$ .

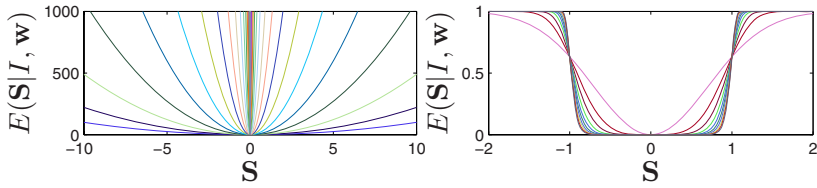
### 2.2 Optimal Energy Functional Weights

For each  $(I_j, S_j)$ ,  $I_j \in \mathbf{I}$  and  $S_j \in \mathbf{S}$ , the task is to find the optimal values for the free weights  $\mathbf{w}(I_j)$ . This section explores the notion of ‘optimal’.

One computationally intractable approach for finding  $\mathbf{w}^*$  is to try all possible weight combinations and run the segmentation method then select the weights with the least segmentation error. A better approach, as outlined in [9], is to find the weights  $\mathbf{w}^*$  that minimize the magnitude of the derivative, in our case  $E_{\alpha}$ , of the energy functional at the correct segmentation  $\alpha^j$  (i.e.  $\alpha^j = [\psi_1 \psi_2 \dots \psi_k]^+ (S_j - \bar{S})$ )<sup>1</sup>. Doing so encourages  $\alpha^j$  to be a minimum of  $E$  (i.e.  $E_{\alpha}(\alpha^j|I_j, \mathbf{w}^*) = 0$ ). Since  $E_{\alpha}$  is in our case a vector of length  $k$  and  $w_i(I_j)$  a scalar function, we measure its magnitude as  $|E_{\alpha}(\mathbf{w}|\alpha^j, I_j)|^2$ . McIntosh and Hamarneh go further to minimize  $|E_{\alpha}(\mathbf{w}|\alpha, I_j)|^2$  for  $\alpha = \alpha^j$  while maximizing it for all other possible shapes (in a direction toward the optimal solution). Adopting their approach, for the time being, but with the new convex setup, we proceed as follows.

For a given shape  $\alpha^i$ , a vector  $(\alpha^i - \alpha^j)$  in  $\mathbb{R}^k$  represents the direction towards  $\alpha^j$ . Since  $E_{\alpha}(\mathbf{w}|\alpha^i, I_j)$  is the vector in  $\mathbb{R}^k$  dictating in what direction, and

<sup>1</sup> We assume the chosen eigenvectors explain 99% of the variance and thus the error incurred by representing  $S$  as  $\alpha$  is negligible.



**Fig. 1.** Varying the shape of energy functionals. Left: Various functionals that show an increasing extent to which the gradient at neighboring shapes points towards the correct segmentation, represented by  $S = 0$ . Right: Various functionals that show how a set of neighboring shapes can also become minima, a degenerate case.

by what amount, the solution will change at the point  $\alpha^i$ , a normalized dot-product (projection-like approach) will measure how much in the right direction  $E_{\alpha}(\mathbf{w}|\alpha^i, I_j)$  points.

So for an energy functional with a form like those in (1), and following [9], for now, we define  $\gamma(\mathbf{w}|I_j, S_j)$  as

$$\gamma(\mathbf{w}) = \left( \left| E_{\alpha}(\mathbf{w}|\alpha^j, I_j) \right|^2 - \lambda \sum_{i \in \mathcal{N}_{\mathbf{S}}} F_{\mathcal{N}_{\mathbf{S}}}(\alpha^j, \alpha^i) \frac{E_{\alpha}(\mathbf{w}|\alpha^i, I_j) \cdot (\alpha^i - \alpha^j)}{|\alpha^i - \alpha^j|} \right) \quad (2)$$

where  $\mathcal{N}_{\mathbf{S}}$  denotes a set of nearby (or similar) shapes in the domain of  $E$ , and  $F_{\mathcal{N}_{\mathbf{S}}}$  is used to weight closer segmentations according to their proximity. The neighborhood  $\mathcal{N}_{\mathbf{S}}$  is used instead of the entire shape space to reduce computational complexity. The second term, dubbed the neighborhood term, is negative and  $|E_{\alpha}(\mathbf{w}|\alpha^i, I_j)|$  is omitted from the normalized dot-product to reward large steps in the correct direction. Solving  $\mathbf{w}^*(I_j) = \underset{\mathbf{w}}{\operatorname{arg\,min}} \gamma(\mathbf{w}|I_j, S_j)$ , yields the optimal weights for image  $I_j$ .

However, notice that there is a weighting  $\lambda$  between competing terms of (2), which was implicit in [9], i.e. was assumed equal to unity and not addressed. With two competing terms, a balance must be struck between: (i) the degree by which  $\alpha^j$  is a minimum of  $E$ ; and (ii) the degree by which the derivative at neighboring points in the shape space points towards  $\alpha^j$  (Fig. 1-left). Make  $\lambda$  too small and  $\alpha^j$  might be a minimum, but so might the entire neighborhood (Fig. 1-right). Make  $\lambda$  too large and the neighborhood will point in the right direction, but  $\alpha^j$  might no longer be a minimum.

To rectify this problem, we make the following observation: when the energy functional  $E$  is convex our only concern is making  $\alpha^j$  as much a minimum as possible, while avoiding the degenerate case that the neighboring points are minima (Fig. 1-right). As a result we can replace the neighborhood term by a *constraint rather than a cost term* since the degree to which the neighbors point towards  $\alpha^j$  does not change whether or not  $E_{\alpha}(\mathbf{w}|\alpha^j, I_j) = 0$  (i.e. we must avoid forcing the gradient in the neighborhood to point at  $\alpha^j$  at the cost of making  $E_{\alpha}(\mathbf{w}|\alpha^j, I_j) \neq 0$ ). Thus instead of (2) we re-define  $\mathbf{w}^*(I_j)$  as

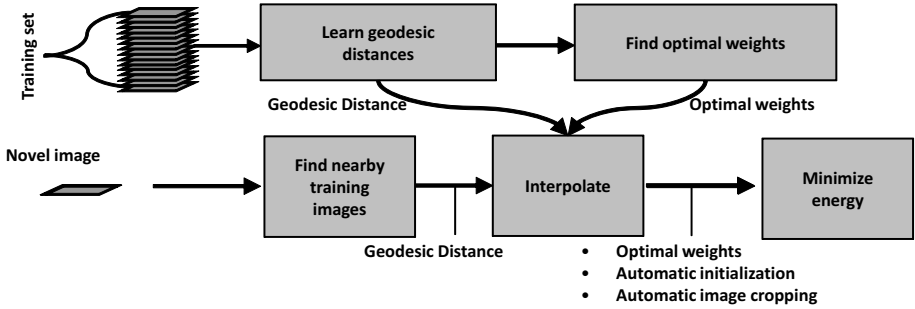


Fig. 2. Overview of our proposed method

$$\begin{aligned}
 \mathbf{w}^* &= \arg \min_{\mathbf{w}} \gamma(\mathbf{w}) = \arg \min_{\mathbf{w}} |E_{\alpha}(\mathbf{w}|\alpha^j, I_j)|^2 \\
 \text{s.t. } &\sum_{i \in \mathcal{N}_{\mathbf{S}}} F_{\mathcal{N}_{\mathbf{S}}}(\alpha^j, \alpha^i) E_{\alpha}(\mathbf{w}|\alpha^i, I_j) \cdot \frac{(\alpha^j - \alpha^i)}{|\alpha^j - \alpha^i|} \geq 0
 \end{aligned} \tag{3}$$

The result is a convex function in  $\mathbf{w}$  under a linear constraint, since the first term in (2) has been shown to be a convex quadratic [9] and the second term linear. A convex function under a linear constraint can be solved via convex optimization, and thus the optimal  $\mathbf{w}(I_j)$  is guaranteed.

### 3 Method: Segmenting Novel Images

Given our set of training image-segmentation pairs we will have  $N$  samples of  $\mathbf{w}^*(I)$ , from which we can interpolate to find values at new points (i.e. novel images). In order to interpolate, we need a metric for measuring distances between images<sup>2</sup>. The set of images with the shortest distances constitutes the neighboring images,  $\mathcal{N}_{\mathbf{I}}$ , and  $\mathcal{N}_{\mathbf{S}}$  are their corresponding correct segmentations.

We assume  $\mathbf{I}$  is smooth over its domain, the space of a particular class of images (e.g. MRI brain scans of normal adults), and that the mapping from images to segmentations is smooth. In other words, we assume that similar images have similar parameters, and similar segmentations. As such, we use a normalized Gaussian kernel, defined over the image distances, to interpolate both the parameters and initializations. For shape and appearance priors, we limit the training data to  $\mathcal{N}_{\mathbf{I}}$  and  $\mathcal{N}_{\mathbf{S}}$ , since we are more confident that the correct shape and appearance information is similar to the training data lying in those neighborhoods. The process is summarized in figure 2.

Manifold learning methods are a special class of nonlinear dimensionality reduction techniques that enable the calculation of geodesic distances between data points. We make use of these techniques to calculate distances between both images and segmentations. Distances between neighboring segmentations allow us to define

<sup>2</sup> The choice of metric is beyond the scope of this paper (Sec. 5).

$$F_{\mathcal{N}_S}(S_j, S_i) = 1 - \frac{g(S_i, S_j)}{\sum_{t \in \mathcal{N}_S} g(S_t, S_j)} \quad (4)$$

where  $g(S_i, S_j)$  is the learned geodesic distance between shape  $S_i$  and shape  $S_j$ . The function  $F_{\mathcal{N}_S}(S_j, S_i)$  then acts as a weighting for the given neighborhood  $\mathcal{N}_S$ , normalized to sum to one (i.e. the weight decreases as a function of distance from the center of the neighborhood). Here we use the geodesic distance between shapes, as opposed to their linear distance in the PCA subspace, for greater accuracy. The linear shape space is well suited for our shape representation because it forms the basis for a convex optimization problem (as previously noted). In essence, we assume an underlying non-linear shape space exists, but use a higher dimensional linear space to represent it. This of course, allows non-valid shapes to be represented, but it brings the benefit a convex energy functional, and with a good energy functional those non-valid shapes will not be minima anyway (as our results show).

Though numerous points of uncertainty have been addressed by our method, there are a few remaining free parameters: (i)  $k$ , the dimensionality of the PCA shape space used for our shape representation; (ii) the dimensionality of the shape manifold; (iii) the dimensionality of the image manifold; (iv) the input image-distance metric used as input to the manifold learning algorithm; (v) the manifold learning algorithm to be used; and finally (vi) the interpolation function used to determine parameters for novel images, as a function of their distance to similar images. As these choices are somewhat application dependent, we include a specification of their values in the experiment section. We also include a discussion of the implications of these parameters in section 5.

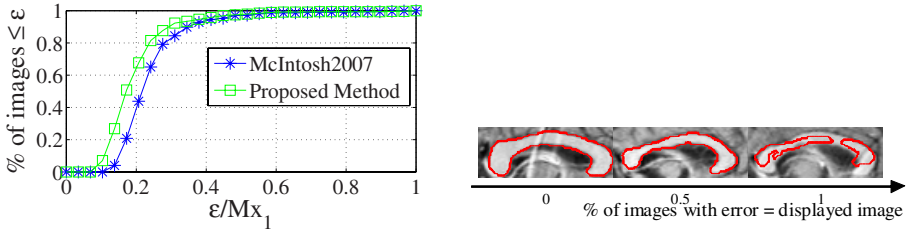
## 4 Experiments

We validate our method on a set of 470  $256 \times 256$  affine registered mid-sagittal MR images, with corresponding expert-segmented corpora callosa (CC). Our energy functional takes the form:

$$E(\alpha) = \int_{\Omega} (w_1(I)f(x)S(x) + w_2(I)g(x)(1 - S(x)) + w_3(I)h(x)|\nabla S(x)|) dx + w_4(I)\alpha^T \Sigma^{-1}\alpha \quad (5)$$

where  $f = -\log(P_{obj}(I))$ ,  $g = -\log(P_{bk}(I))$ , for object and background histograms  $P_{obj}, P_{bk}$ ,  $h = \frac{1}{1+|\nabla I|}$ , and  $\Sigma^{-1}$  characterizes the allowable shape distribution (see [15] for details).

To learn the distances, we used a MATLAB implementation of  $K$ -ISOMAP [16] from <http://isomap.stanford.edu/>, with Euclidean distance between images as the input distance matrix. As this paper is about the application of optimal parameters to segmentation, issues related to learning the manifold will not be addressed in this work. For  $K$ -ISOMAP we set  $K = 10$ , and reduce the image space to a 5-manifold; chosen as the elbow of the scree plot. One caveat with using ISOMAP is that it does not directly extend to novel samples. Though



**Fig. 3.** CC segmentation results. (Left) Error plot where  $Mx_1$  is the maximum measured value for  $\epsilon$ . (Right) Segmentations demonstrating the full range of error.

out-of-sample extensions for ISOMAP have been published, for simplicity we choose to simply re-run ISOMAP to include the novel image, as this only takes a few seconds. For PCA on the shape space we set  $k = 10$ ; the elbow of the corresponding scree plot.

Firstly, to show that eq. (3) can balance an ideal energy functional (one for which the minimum can be positioned exactly at the correct segmentation using the optimal weights) we calculate the error of energy functional (5) plus a fifth term whose unique global minimum is always the correct segmentation for the given image (i.e. we gave the functional weights  $\mathbf{w}$  the power to achieve 100% accuracy by adding a rigged, strictly convex term). We measure the error using a modified Dice metric:  $\epsilon = Area(A \cup G - A \cap G) / Area(G)$ , where  $A$  and  $G$  are the binary automatic segmentation and the ground truth, respectively. Using this rigged term, we obtained  $\epsilon = 0$  for all 470 images, validating that our method can achieve the full potential of a given functional.

To compare to the original weight optimization equation (eq. (2)), as originally presented in [9], we performed validation on the set of 470 images using energy functional (5). For each image, the optimal parameters are learned directly, using the ground truth segmentation, rather than interpolation from the manifold. Doing so isolates the error induced by the weights, not the interpolation method or the ability to locate the position of novel images on the manifold. Using eq. (2) we obtained an error of 0.1201 vs 0.1099 with eq. (3), a clear improvement.

To validate our segmentation method as it pertains to novel data, we perform leave-one-out validation on the set of 470 images using energy functional (5). Under error metric  $\epsilon$ , we found the average error to be 0.13, improving over the average error of 0.16 reported in [9]. Our results are summarized in Fig. 3. The figure shows the percentage of images with  $\epsilon \leq$  the specified value on the x-axis. So, for example, with our proposed method approximately 65% of the images have  $\epsilon \leq 0.2$  as opposed to only 44% using the method proposed in [9]. If an error of 0.2 was the cut off for the segmentation method to be clinically useful, our method would have successfully segmented an additional 21% of the data, or about 100 images more than [9].



## 5 Discussion

Our results demonstrate our method's ability to optimize the weights for convex energy functionals. In doing so, we have addressed a few key areas of uncertainty typically found in objective function based segmentation methods. Specifically, our method does not suffer from uncertainty with local minima, initializations, or hand-tuned parameters. However, new questions remain: Was the optimal manifold learned? Is this the best way to describe the optimal weights? And is this the best interpolation function for novel images? As already shown in [9], even with the inherit manifold uncertainty, this technique of analytically describing the optimal weights is better than the alternative (hand-tuning the weights, and/or fixing the weights as constant values over the set of application images).

The difference  $\delta\varepsilon = 0.02$  between what our method achieved in practice ( $\varepsilon = 0.13$ ) and the error using weights calculated directly from the ground truth segmentations, ( $\varepsilon = 0.1099$ ), is due to our localization of novel images on the manifold and the interpolation over  $\mathbf{w}^*(I)$ . An important area of future work is thus to lower this difference by using better manifold learning techniques. Improved image distance metrics [17] may also work, as they can simplify the learning problem.

Finally, we have provided two improvements over [9]: we removed the uncertainty with local minima and initializations; and we improved their weight optimization equation by removing its implicit weight. We demonstrated how our proposed weight optimization equation yields improved weights, and that our method has a significantly lower error overall. In the end, we were able to segment an additional 100 images under a reasonable cutoff, which is an important improvement in a clinical setting.

## References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *IJCV* 1(4), 321–331 (1987)
2. Lobregt, S., Viergever, M.: A discrete dynamic contour model. *IEEE TMI* 14, 12–24 (1995)
3. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *IJCV* 22(1), 61–79 (1997)
4. Chan, T., Vese, L.: Active contours without edges. *IEEE TIP*, 266–277 (2001)
5. Boykov, Y., Kolmogorov, V.: Computing geodesics and minimal surfaces via graph cuts. In: *ICCV* (2003)
6. Grady, L.: Random walks for image segmentation. *IEEE TPAMI* 28(11), 1768–1783 (2006)
7. Nikolova, M., Esedoglu, S., Chan, T.F.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics* 66(5), 1632–1648 (2006)
8. Bresson, X., Esedoglu, S., Vanderheynt, P., Thiran, J.P., Osher, S.: Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision* 28(2), 151–167 (2007)

9. McIntosh, C., Hamarneh, G.: Is a single energy functional sufficient? adaptive energy functionals and automatic initialization. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 503–510. Springer, Heidelberg (2007)
10. Taskar, B., Chatalbashev, V., Koller, D., Guestrin, C.: Learning structured prediction models: a large margin approach. In: ICML 2005: Proceedings of the 22nd international conference on Machine learning, pp. 896–903. ACM, New York (2005)
11. Anguelov, D., Taskar, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G., Ng, A.: Discriminative learning of markov random fields for segmentation of 3d scan data. In: CVPR 2005: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), Washington, DC, USA, vol. 2, pp. 169–176. IEEE Computer Society, Los Alamitos (2005)
12. Szummer, M., Kohli, P., Hoiem, D.: Learning crfs using graph cuts. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 582–595. Springer, Heidelberg (2008)
13. Finley, T., Joachims, T.: Training structural svms when exact inference is intractable. In: ICML 2008: Proceedings of the 25th international conference on Machine learning, pp. 304–311. ACM, New York (2008)
14. Kolmogorov, V., Boykov, Y., Rother, C.: Applications of parametric maxflow in computer vision. In: IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, October 2007, pp. 1–8 (2007)
15. Cremers, D., Schmidt, F.R., Barthel, F.: Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. In: IEEE CVPR (2008)
16. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323 (2000)
17. Souvenir, R., Pless, R.: Image distance functions for manifold learning. *Image and Vision Computing* 25(3), 365–373 (2007); Articulated and Non-rigid motion