

# Convex Multi-Region Probabilistic Segmentation with Shape Prior in the Isometric Log-Ratio Transformation Space

Shawn Andrews, Chris McIntosh, Ghassan Hamarneh  
Medical Image Analysis Lab, Simon Fraser University  
8888 University Drive, Burnaby, BC, Canada, V5A 1S6  
{sda56, cmcintos, hamarneh}@sfu.ca

## Abstract

*Image segmentation is often performed via the minimization of an energy function over a domain of possible segmentations. The effectiveness and applicability of such methods depends greatly on the properties of the energy function and its domain, and on what information can be encoded by it. Here we propose an energy function that achieves several important goals. Specifically, our energy function is convex and incorporates shape prior information while simultaneously generating a probabilistic segmentation for multiple regions. Our energy function represents multi-region probabilistic segmentations as elements of a vector space using the isometric log-ratio (ILR) transformation. To our knowledge, these four goals (convex, with shape priors, multi-region, and probabilistic) do not exist together in any other method, and this is the first time ILR is used in an image segmentation method. We provide examples demonstrating the usefulness of these features.*

## 1. Introduction

Image segmentation is a fundamental task in computer vision, and many segmentation algorithms can be formulated as energy minimization problems. These energy functions encode a wide variety of information designed to increase the accuracy of the segmentation.

In this paper, we will focus on constructing an energy function that achieves four important goals. We will show how our choice of segmentation representation allows us to achieve these goals simultaneously. The four goals are:

1. To make our energy function convex. Even if a non-convex energy function can provide good results in practice, it can still give poor results with certain initializations, rendering the method difficult to evaluate and less robust. Convex energy functions are not as sensitive to initialization and can often be minimized more quickly than nonconvex energies.

2. To incorporate a shape prior into our energy function. Energy functions with shape priors allow for more accurate segmentation of regions that are not readily described using image information alone, and allow the segmentation of images that would otherwise be too noisy. Statistical analysis of training data, for example principal component analysis (PCA), is often performed on training data to obtain a prior distribution of the shape space. Enforcing the shape prior can be done by restricting segmentations to the span of the modes of greatest variance in the training data [9].
3. To enable our energy function to encode multi-region segmentations. An energy function that explicitly handles multi-region segmentation is important, as many energy functions designed for binary segmentation do not extend well to multi-region segmentation due to the added complexity. Explicitly multi-region algorithms capture interactions between regions easily [20, 4].
4. To enable our energy function to encode probabilistic segmentations. Probabilistic segmentations contain more information than crisp (non-probabilistic) segmentations, and by analyzing both the absolute probabilities and their relative values, we gain a richer statistical understanding of the variability and uncertainty in region boundaries. Probabilistic segmentations should have thought out interpretations [12, 26, 22], something which relaxed 0 – 1 segmentations often lack as they are not designed with probabilistic results in mind.

We note that pose estimation is another important goal in image segmentation, but it is not included in the above goals due to the fact that including pose parameters into an energy function while maintaining convexity is an unsolved problem. Thus, to maintain convexity, our energy function does not minimize pose parameters explicitly. However, we do discuss a method for pose estimation, if convexity is not enforced, in Sec. 2.3.

Table 1: A comparison of certain features of popular algorithms and our proposed algorithm.

Methods	Features			
	Convex Energy	Shape Prior	Multi-Region	Probabilistic
[5]	✓	✗	✗	✓*
[14]	✓	✗	✓†	✗
[24]	✓	✓‡	✗	✗
[25]	✗	✓	✓	✗
[20, 4, 10]	✓	✗	✓	✗
[16, 19, 27]	✓	✗	✓	✓*
[23]	✓	✓‡	✓	✗
[21]	✗	✓	✓	✓
[12]	✓	✗	✓	✓
[9]	✓	✓	✗	✓
Our Method	✓	✓	✓	✓

\* Relaxed 0-1 segmentations could be informally treated as probabilities.

† Allows only limited regional interaction terms.

‡ Only applicable to restricted classes of shapes.

The work required for minimization varies greatly between energy functions. How successfully information can be encoded in an energy function and how optimally it can be minimized depends on how segmentations are represented, as this determines the domain of the energy function. A key step in developing our method is the use of a segmentation representation encoding a multi-region probabilistic segmentation as an element of a vector space. To provide context, we will discuss some existing segmentation representations and how the methods that use them relate to the goals (convex, with shape prior, multi-region, and probabilistic). All of the methods surveyed below include some but not all of these goals (Table 1).

A segmentation’s boundary can be specified explicitly using parametric representations [15, 7]. There are difficulties in extending this type of representation to handle multiple regions, topological adaptability, or higher dimensions. These drawbacks have sparked research on level-set representations [18, 17], a popular example of which is signed distance maps (SDMs). SDMs have two main drawbacks: SDMs are not closed under linear combinations, which creates difficulties when trying to perform statistical analysis [21]; and energy functionals over SDMs are generally not convex mainly because the constraints enforcing unity gradient magnitude are nonconvex [8].

A characteristic function is an alternative segmentation representation which assigns every region an integer from 0 to  $(D - 1)$ , where  $D$  is the number of regions, and assigns one of those numbers to each pixel. These often lead to non-convex discrete constraints that make global optimization

difficult [5]. Techniques to globally optimize these representations via relaxation have been introduced [5, 4]; however, these methods do not include shape priors. Further, the resulting relaxed segmentations do not have the statistical basis to be considered probabilities.

The success of constraint relaxation suggests using continuous representations, e.g. assigning a vector of regional probabilities to each point in the image. In [12], Grady’s random walker algorithm (RW) uses a probabilistic segmentation of this nature in a convex, multi-region method, but no ways of enforcing a shape prior for RW while maintaining its convexity are known. In [21], Pohl et al. transform SDMs to probability maps, thereby allowing the construction of a statistical shape prior. The resulting energy’s convexity is not guaranteed, however. In [9], Cremers et al. incorporate a statistical shape prior into a convex energy functional that generates a probabilistic segmentation; however, no extension of this scheme to multi-region segmentation has been presented.

In this paper, we achieve the four goals (convexity, shape prior, multi-region, and probabilistic) by using a new representation of a segmentation based on performing the isometric log-ratio (ILR) [11, 6] transformation on the multi-regional probabilities, bijectively mapping a probabilistic segmentation to a point in a vector space of real numbers. This mapping removes the need for constraints to guarantee a valid segmentation [9] (with the probabilities at each pixel summing to unity). We will train multi-region probabilistic shape priors in the vector space provided by ILR and, using those, we will define a convex energy function that captures image and shape prior information. As summarized above and in Table 1, we do not know of any other methods that achieve these four goals without significant restrictions on the type of shape prior.

## 2. Method

We start in Sec. 2.1 with a review of the method from Cremers et al. [9], as we will later extend this convex method to multi-region segmentation and remove the need for constraints in the optimization process. To do this, we use an alternate representation of a segmentation involving ILR, described in Sec. 2.2. We then describe our proposed method in Sec. 2.3.

### 2.1. Convex binary segmentation with shape prior

Following [9], we will define a convex energy functional as the sum of convex terms :

$$E(q) = E_{Edge}(q) + E_{Region}(q) + E_{Shape}(q) . \quad (1)$$

A binary segmentation is represented by a function  $q : \Omega \rightarrow [0, 1]$  that maps points on the image domain  $\Omega$  to probabilities. This probabilistic representation makes it so that con-

vex combinations of segmentations are still functions from  $\Omega$  to  $[0, 1]$ , and thus correspond to valid segmentations.

To define a shape energy term, we need to construct a shape prior. Still following [9], given a set of training segmentations  $\{q_1, \dots, q_m\}$ , a shape prior is constructed by finding the mean of the training data,  $q_0$ , and performing PCA to find the data's eigenmodes. Segmentations are represented by introducing a new dependent variable,  $\alpha$ , to  $q$  such that  $q(x, \alpha) = q_0(x) + \Psi(x)\alpha$ , where  $x \in \Omega$ ,  $\Psi$  is a matrix with the  $K$  eigenmodes of greatest variance as its columns, and  $\alpha \in \mathbb{R}^K$  weights the eigenmodes.

Writing  $E$  as a function of  $\alpha$ , it will still be convex since  $q(x, \alpha)$  is linear in  $\alpha$ . Many works focus on developing convex  $E_{Edge}$  and  $E_{Region}$  terms, either directly over probabilities or as relaxations of crisp terms. As developing novel energy terms is not the goal of this work, we refer the reader to some such works [16, 19, 27] for example terms. We use an  $E_{Shape}$  of the form:

$$E_{Shape}(\alpha) = \alpha^T \Lambda^{-1} \alpha, \quad (2)$$

where  $\Lambda$  is chosen to provide a Mahalanobis type energy term, penalizing the eigenmode weights  $\alpha$  according to the corresponding variances. Since all terms are convex, with  $E_{Shape}(\alpha)$  being strictly convex,  $E(\alpha)$  will have a unique global minimum when it is optimized, i.e.

$$\alpha_{min} = \min_{\alpha} E(\alpha), \quad (3)$$

then the corresponding segmentation is expressed as

$$q_{min} = q_0 + \Psi \alpha_{min}. \quad (4)$$

However,  $q_{min}$  must be restricted to the range  $[0, 1]$ , and thus (3) requires constraints. Specifically,  $\alpha$  is constrained to the convex domain  $\{\alpha | q(x, \alpha) \in [0, 1] \text{ for all } x \in \Omega\}$ .

If we can represent probabilistic segmentations by elements of a vector space and then define energies over this vector space, then we will not need explicit constraints on our energy minimization, as any element of the vector space will correspond to a valid segmentation. We will do this by mapping probabilities in the range  $[0, 1]$  to real numbers using the ILR transform [11, 6]. We note ILR is defined for probability vectors of any length, allowing us to easily handle multiple regions. Thus our main improvements to the above method are the extension to multi-region segmentation and the removal of constraints from the optimization.

## 2.2. Multi-region probabilistic segmentation representation using ILR

In this section, we will describe a method for representing an arbitrary  $D$  region probabilistic segmentation as an element of a vector space. In the next section we will discuss the application of this representation to convex multi-region probabilistic segmentation with shape prior.

We represent the probabilistic segmentation of an image into  $D$  regions as a vector-valued function over  $\Omega$  using the ILR transform [11, 6]. We use the ILR transform because it has several desirable properties: the transformed probabilities form a vector space; it is an isomorphism between the probability simplex and vectors of real numbers; and it is symmetric between regions. We note that while the LogOdds transform [21] maps probabilities to a vector space, it is not an isomorphism. We will see that using an isomorphic transform like ILR allows us to construct energy functionals in the transformed space whose meanings we can relate intuitively back to properties of the probabilities.

We define  $\mathbb{S}^D$  to be the simplex of all  $D$ -length probability vectors.  $\mathbb{S}^D$  is a  $D - 1$  dimensional inner product space [11, 1] with an inner product defined as

$$\langle p, q \rangle_a = \sum_{d=1}^D \log \left( \frac{p_d}{\mu_g(p)} \right) \log \left( \frac{q_d}{\mu_g(q)} \right), \quad (5)$$

where  $p, q \in \mathbb{S}^D$  and  $\mu_g(v)$  is the component-wise geometric mean of a vector  $v$ . This inner product, known as the Aitchison inner product [1], can also be used to define the Aitchison distance metric on  $\mathbb{S}^D$ :

$$d_a(\mathbf{p}, \mathbf{q}) = \left[ \sum_{d=1}^D \left( \log \frac{p_d}{\mu_g(\mathbf{p})} - \log \frac{q_d}{\mu_g(\mathbf{q})} \right)^2 \right]^{\frac{1}{2}}. \quad (6)$$

We represent the ILR transformation (including the background) by a function  $\phi_{D-1} : \mathbb{S}^D \rightarrow \mathbb{R}^{D-1}$ . Using the Aitchison inner product and following Egozcue et al. [11], we define an orthonormal basis,  $B_E$ , over the simplex  $\mathbb{S}^D$ , where  $B_E = \{e_1, \dots, e_{D-1}\}$  with  $e_i \in \mathbb{S}^D \forall i$ , such that  $\langle e_i, e_j \rangle_a = 1$  if  $i = j$  and 0 otherwise. We use

$$e_{i,j} = \frac{1}{Z} \exp \left( \begin{cases} \sqrt{\frac{1}{i(i+1)}} & \text{if } j \leq i \\ -\sqrt{\frac{i}{i+1}} & \text{if } j = i+1 \\ 0 & \text{o.w.} \end{cases} \right) \quad (7)$$

where  $Z$  ensures  $e_i$  sums to 1.

$\phi_{D-1}$  is the Aitchison-based projection onto  $B_E$ :

$$\phi_{D-1}(p) = [\langle p, e_1 \rangle_a, \dots, \langle p, e_{D-1} \rangle_a] \in \mathbb{R}^{D-1}. \quad (8)$$

It is shown in [11] that the Aitchison distance between  $p_1$  and  $p_2$  equals the Euclidean distance between  $\phi_{D-1}(p_1)$  and  $\phi_{D-1}(p_2)$ , i.e. the transformation is isometric.

We use the convention that if  $q : \Omega \rightarrow \mathbb{S}^D$  is a vector-valued function representing a  $D$ -region probabilistic segmentation then  $\eta = \phi_{D-1}(q) : \Omega \rightarrow \mathbb{R}^{D-1}$ . We note that since  $\phi_{D-1}$  is bijective,  $\eta$  is a vector-valued function that corresponds to a unique multi-region probabilistic segmentation. We will refer to the space of all segmentations represented this way as  $\eta$ -space.

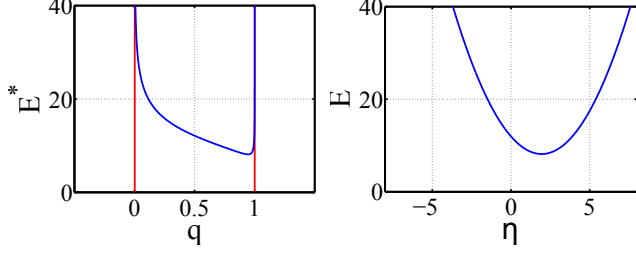


Figure 1: An example of how if  $\eta = \phi_1(q)$  and  $E(\eta)$  is convex in  $\eta$ , then the domain of  $E^*(q) = E(\phi_1(q))$  is restricted to functions taking values in  $[0, 1]$ .  $E^*$  is quasi-convex in  $q$ , thus  $E^*$  can be globally minimized with no constraints required.

### 2.3. Convex energies with shape prior in ILR space

We define an energy as a functional of  $\eta$  similar to (1):

$$E(\eta) = E_{Edge}(\eta) + E_{Region}(\eta) + E_{Shape}(\eta). \quad (9)$$

Unlike (1), this energy supports multi-region segmentation. We will assume  $E$  and all of its components are convex.

As our final probabilistic segmentation will be  $q = \phi_{D-1}^{-1}(\eta)$ , it is important to examine how  $q$  is affected by the minimization of  $E(\eta)$ . We gain insight into the relationship between  $q$  and  $\eta$  by looking at a convex energy functional  $E(\eta)$  as a functional  $E^*(q)$ :

$$E^*(q) = E(\phi_{D-1}(q)) = E(\eta) \quad (10)$$

where  $E^*$  is the pullback of  $E$  by  $\phi_{D-1}$ . As the range of  $\phi_{D-1}$  is  $\mathbb{S}^D$ , the functional  $E^*$  is only defined for functions that range over  $\mathbb{S}^D$ . Theorems 1 and 2 below show that  $E^*$  is quasi-convex in  $q$ . Thus the global minima of  $E^*(q)$  could be found directly, but we will instead take the approach of minimizing the convex  $E(\eta)$  and then using  $\phi_{D-1}^{-1}$  to find a global minima of  $E^*(q)$ . The following theorems demonstrate that finding a global minima of  $E(\eta)$  is equivalent to finding a global minima of  $E^*(q)$ , and thus minimizing  $E(\eta)$  gives an optimal segmentation.

**Theorem 1.** *If  $E^*(q)$  has a critical point at  $q_1$ , then  $E(\eta)$  has a critical point at  $\eta_1 = \phi_1(q_1)$ .*

*Proof.* Follows from differentiating (10) (See [3]).  $\square$

**Theorem 2.** *Let  $H_{min}$  be the set of global minima of  $E(\eta)$ ,*

$$H_{min} = \{\eta_1 | E(\eta_1) = \min_{\eta} E(\eta)\}. \quad (11)$$

*Then  $G_{min} = \phi_1^{-1}(H_{min})$  satisfies*

$$G_{min} \subseteq \{q_1 | E^*(q_1) = \min_q E^*(q)\}, \quad (12)$$

*that is,  $G_{min}$  is a subset of all global minima of  $E^*(q)$ .*

*Proof.* Follows from the bijectivity of  $\phi_{D-1}$  (See [3]).  $\square$

Theorems 1 and 2 imply that wherever  $E^*(q)$  has a critical point, it must be a global minimum, and by finding a global minimum of  $E(\eta)$  and applying  $\phi_{D-1}^{-1}$ , we find a global minimum of  $E^*(q)$ . However,  $E^*(q)$  is not guaranteed to have positive second derivative everywhere, and thus it is only guaranteed to be quasi-convex. Fig. 1 shows how a specific example of  $E$  and  $E^*$  relate.

As in Sec. 2.1, to define our shape energy term, we will construct a shape prior using PCA. Given a set of probabilistic training segmentations, we will create segmentations in  $\eta$ -space,  $\{\eta_1, \dots, \eta_m\}$ , by applying  $\phi_{D-1}$ . Since  $\phi_{D-1}$  is not defined for probability vectors that include 0's or 1's, 0's and 1's are replaced in the multi-region training data with  $\epsilon$  and  $(1 - (D - 1)\epsilon)$ , respectively. We then perform PCA on this training data. As we are performing PCA on all of the regions simultaneously, PCA will capture not only the variations of individual regions but the interactions between regions (i.e. the joint probabilities).

PCA yields the mean of  $\{\eta_1, \dots, \eta_m\}$ ,  $\eta_0$ , and a matrix  $\Gamma$  whose columns are their  $K$  eigenmodes of greatest variance. As with  $q(x, \alpha)$  in Sec. 2.1, we will enforce the shape prior by introducing a new dependent variable,  $\gamma$ , to  $\eta$  such that  $\eta(x, \gamma) = \eta_0(x) + \Gamma(x)\gamma$ , where  $x \in \Omega$  and  $\gamma$  weights the eigenmodes. For any fixed  $\gamma$ ,  $\eta(x, \gamma)$  represents a probabilistic segmentation. We redefine  $E$  as a (convex) function of  $\gamma$ .

Examples of convex energy terms in  $\eta$  space are easily constructed due to the isomorphic nature of the ILR transform.  $E_{Edge}$  terms usually involve the difference in  $q$  between neighboring pixels, so replacing  $q$  with  $\eta$  should give a useful  $E_{Edge}$  based on Aitchison instead of Euclidean distance. We do not explore edge terms in depth, however, as we found  $E_{Shape}$  provided adequate regularization in our results. As a convex  $E_{Region}(\gamma)$ , we use:

$$E_{Region} = \int_{x \in \Omega} d_a(p(x), q(x))^2 dx, \quad (13)$$

where  $p : \Omega \rightarrow [0, 1]$  are the regional intensity prior probabilities. The Mahalanobis type energy term used in (2) can be used here as a strictly convex example of  $E_{Shape}(\gamma)$ .

We minimize the energy function:

$$\gamma_{min} = \min_{\gamma} E(\gamma), \quad (14)$$

and take as our probabilistic segmentation

$$q_{min} = \phi_{D-1}^{-1}(\eta_0 + \Gamma\gamma_{min}). \quad (15)$$

In this method, we are performing PCA on functions ranging over  $\mathbb{R}^{D-1}$ , so the minimum of our energy function is guaranteed to correspond to a valid segmentation

and our energy function can be optimized without constraints. As our  $E_{Region}$  and  $E_{Shape}$  are derived from training data, both of these terms can be viewed as negative log-likelihoods of Gaussian priors over  $\eta$  derived from the shape analysis and regional intensity profiles, respectively.  $E(\gamma)$  is a strictly convex energy function that includes a shape prior and can be globally minimized without constraints to give a multi-region probabilistic segmentation; therefore, this segmentation approach is the first technique to realize the four stated goals.

If we wish to estimate pose, we note that since  $\eta$  is a function from  $\Omega \subset \mathbb{R}^d$  to  $\mathbb{R}^{D-1}$ , we can express a rigid body transformation as  $\eta'(x) = \eta(Rx + t)$ , where  $x \in \Omega$ ,  $R$  is a rotation matrix, and  $t$  is a translation vector. If  $u$  is a pose parameter (translation distance or rotation angle), we can update it by calculating the derivative

$$\frac{\partial E}{\partial u} = (\nabla_{\eta} E) \cdot \left( \frac{\partial \eta}{\partial u} \right), \quad (16)$$

and then performing gradient descent. Thus pose estimation can be included by iteratively solving (14) and then updating the pose parameters for the resulting segmentation. To do this, we would have to sacrifice the overall convexity of our method, so we often rely on affine pre-registration of novel images to provide adequate alignment.

### 3. Results

When testing our algorithm, we focus on two things: showing it provides accurate segmentations of images, and showing how our four goals make our energy robust and useful. We compare the results achieved by our method, which satisfies the four aforementioned goals, to other methods that fail to satisfy one or more of these four goals (Table 1). We show how turning off the shape prior greatly reduces the accuracy of the segmentation. We show the benefits of multi-region segmentation over a binary method like [9]. We compare to LogOdds [21] to demonstrate the benefits of having a convex method while retaining the remaining 3 goals. We also demonstrate the benefits of a probabilistic segmentations.

While a PCA-based shape space won't contain the exact ground truth, we can project the ground truth onto the shape space to find the best possible result. This lets us differentiate error in the energy from error in the shape space. Specifically, if  $q_{gt}$  is the ground truth and  $q_{proj}$  is its projection onto the shape space, referred to as the projected ground truth, then

$$q_{proj} = \phi_{D-1}^{-1}(\eta_0 + \Gamma \Gamma^T (\phi_{D-1}(q_{gt}) - \eta_0)). \quad (17)$$

In the following, we demonstrate our method using both synthetic and real examples. We begin with the synthetic images seen in Fig. 2. A shape model is trained using

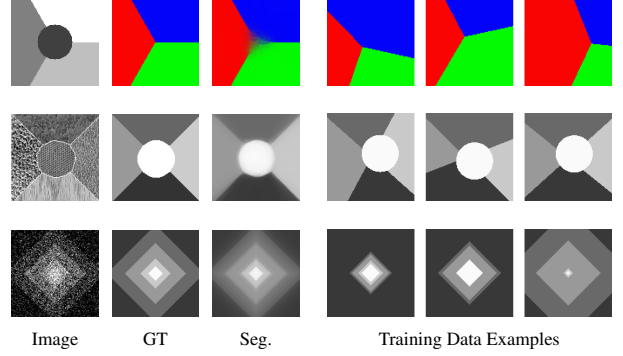


Figure 2: (Color Figure) Segmentation of synthetic images. For each of the three rows we see from left to right: the original image, its ground truth segmentation, the segmentation our method generated, and three examples of the training data. These figures show the shape prior overcoming occlusion, noise, and regions with similar intensity profiles. We note that blurred areas of the probabilistic segmentations (near region boundaries or occluded areas) correspond to areas of greater uncertainty (cf. Figure 6).

1000 training segmentations from similar images and finding their  $K = 500$  largest variance eigenmodes. Some examples of the training segmentations are shown. These synthetic images show how our shape prior can overcome occlusion, noise, and regions with similar intensity profiles.

We demonstrate the usefulness of our goals when segmenting real data. We took 20 magnetic resonance imaging (MRI) slices of size  $60 \times 160$  of different brains, along with crisp expert segmentations of the images into  $D = 4$  regions: white matter, gray matter, cerebrospinal fluid (CSF), and background. These slices were affinely registered to create 20 aligned 2D images and their corresponding ground truth segmentations. Following [13], we address the problem of limited training ground truths by deforming each of these images 200 times to varying degrees to create 4000 training samples. This will increase the size of the shape space, allowing more images to be accurately segmented. When an image was segmented, it and all 200 images derived by deforming it were removed from the training data (i.e. cross validation). The  $K = 500$  largest variance eigenmodes of the training samples were used to create a shape model. Since the training data is crisp, we convert it to  $\eta$ -space by creating probabilistic data using the method described in Sec. 2.3 with  $\epsilon = \frac{1}{100}$ .

We now demonstrate the power of shape priors in multi-region segmentations. The fourth column in Fig. 3 shows the resulting multi-region ( $D = 4$ ) segmentations of 3 of the 20 brain slices without shape prior, e.g. only satisfying the 3 other goals, like RW [12]. For measuring accuracy, we use the Dice similarity coefficient (DSC) between a segmentation (thresholded to crisp, and averaged across regions) and its ground truth (GT). The average DSC of all 20 of the seg-



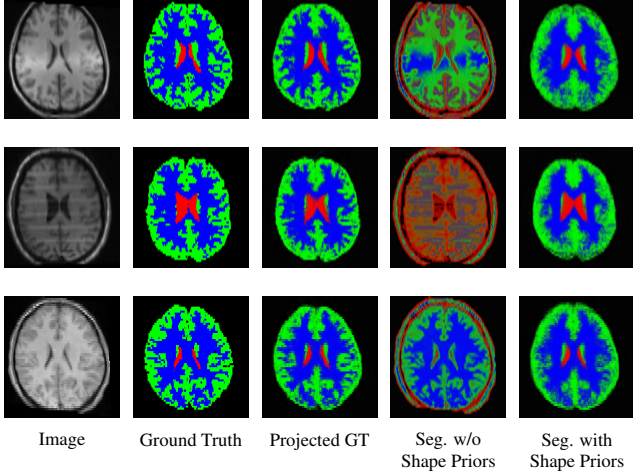


Figure 3: (Color Figure) Examples of our segmentation algorithm applied to axial brain slice images. We see the shape prior greatly improves the segmentation. In most color figures of brain images seen here, red corresponds to CSF, green to gray matter, blue to white matter, and black to background. In gray scale images, lighter shades correspond to higher intensities/probabilities/entropies.

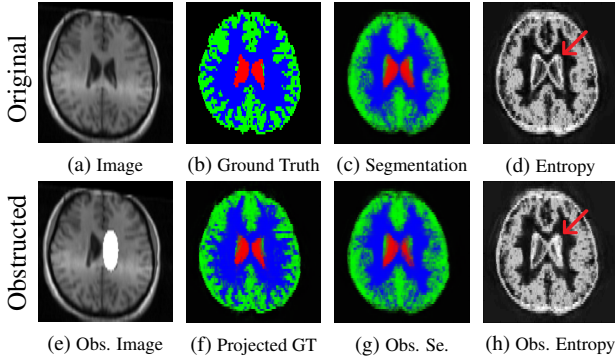


Figure 4: (Color Figure) The segmentation of both an image and the same image partially obstructed, demonstrating the usefulness of our shape prior. The DSC of the segmentation without and with the obstruction is 0.8114 and 0.7995, respectively. Thus, despite the obstruction, the DSC only drops 1.5%. In 4d and 4h we note the increase in entropy under the obstruction (red arrow), corresponding to uncertainty in our probabilistic segmentation. Note that while such a glaring occlusion does not typically appear in MRI images, small artifacts and noise do. This example is designed to make the effect of the shape prior obvious.

mentations without a shape prior was  $0.5226 \pm 0.1411$ . The DSC was greatly improved by the inclusion of the shape prior, to  $0.8273 \pm 0.0345$ . This is close to the DSC of the optimal segmentation in the shape space (the projected ground truth):  $0.9029 \pm 0.0136$ .

Handling occlusion is a natural way to demonstrate the power of a shape prior. Because of our shape prior, our

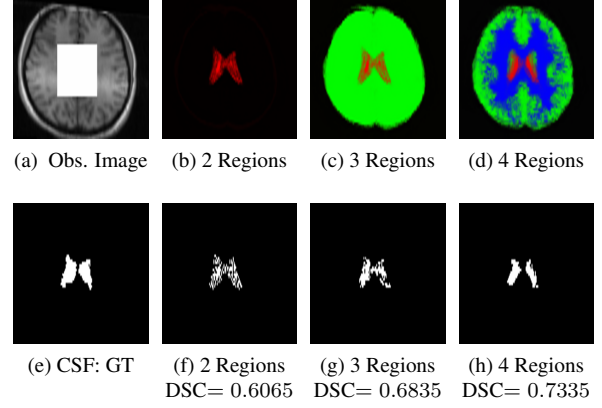


Figure 5: (Color Figure) An obstructed image is segmented into 2, 3, and 4 regions, showing how multi-region interactions improve the segmentation of individual regions. The bottom row shows the ground truth for the CSF and how the CSF’s segmentation improves as more regions are included.

segmentation algorithm is able to segment the ventricles in Fig. 4 despite the fact that they are about partially occluded. Our algorithm approximates the shape of the left ventricle based on the shape of the right ventricle due to the lateral symmetry encoded in the shape prior.

Extending this concept to multi-region segmentations, we show how the shape prior encodes region interactions, which is not possible with binary segmentation methods like [9]. To demonstrate how this improves the segmentation, we take the extreme example of completely obstructing the ventricles, and then segmenting the image into 2, 3, and 4 regions (Fig. 5). The segmentation of the obstructed ventricles becomes more accurate as more regions are considered because the shapes of the other regions correlate to the shape of the hidden region, improving the segmentation.

To demonstrate the usefulness of a probabilistic segmentation, we return to Fig. 4 and compare the segmentations obtained for the image in Fig. 4a and its obstructed counterpart, the image in Fig. 4e. We see that in Fig. 4h the obstructed area has a much higher entropy (red arrows) than in Fig. 4d, where no obstruction existed.

To further demonstrate the usefulness of a probabilistic segmentation over a crisp segmentation, we show how uncertainty in a probabilistic segmentation is related to uncertainty in the intensity and shape priors. In Fig. 6, we have a close up of the red-colored ventricles in the center of one of the brain slices. In Fig. 6h, we see that the pixels along the bottom and inside boundaries of the ventricles have high entropy. This is caused by the large gradient in the image intensity there (Fig. 6c) creating uncertainty in the intensity prior. In Fig. 6h we see that the pixels corresponding to the green regions on the outside of either of the ventricles also have high entropy. As the segmentation in this

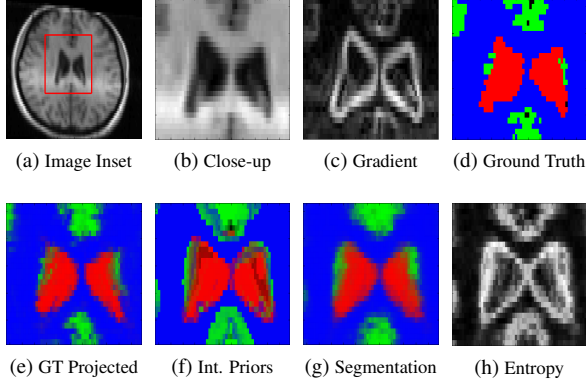


Figure 6: (Color Figure) Analysis of segmentation uncertainty. A close up of the CSF region in the center of an image for the purposes of understanding the sources of the uncertainty in the segmentation, shown by its entropy in the bottom right figure. We can infer the origin of the entropy of the segmentation (6h) from the intensity priors (6f) and the projected ground truth (6e). Areas of high intensity gradient (6c) create uncertainty in the intensity prior, and thus in the segmentation. Areas where the ground truth (6d) and the projected ground truth differ indicate features not representable by the shape prior, and again this affects the segmentation’s certainty.

area differs between the ground truth (Fig. 6d) and the projected ground truth (Fig. 6e), we deduce that this feature is not easily represented in the shape space, which creates greater uncertainty in the shape prior. Crisp segmentations produced by [20, 4, 10, 23] are not able to highlight these areas of uncertainty.

As our method is strictly convex, no initialization is required. To demonstrate the advantages of the convexity of our energy, we show how different initializations affect the results of a non-convex method that achieves our other goals, the LogOdds EM-based method from [21]. In Fig. 7, we segment the image from the second row of Fig. 3 using a shape space constructed from the same training data used for our method, again with  $K = 500$  eigenmodes. The projected GT in this LogOdds shape space has DSC with the GT of 0.8741, less than the 0.9046 of the projected GT in our shape space. This may be because converting the GT to an SDM and then back to probabilities using LogOdds can create inaccuracies.

The LogOdds algorithm takes an element of the shape space and a set of intensity inhomogeneity parameters per pixel as initialization. In the top row of Fig. 7, we initialize with the mean shape and zeros for the inhomogeneity parameters, resulting in a poor segmentation with a DSC of 0.6766 with the GT. However, if we use the GT to find the optimal initialization, we see in the bottom row of Fig. 7 that the resulting segmentation is quite good, with a DSC of 0.8337 with the GT. It is possible for LogOdds to perform

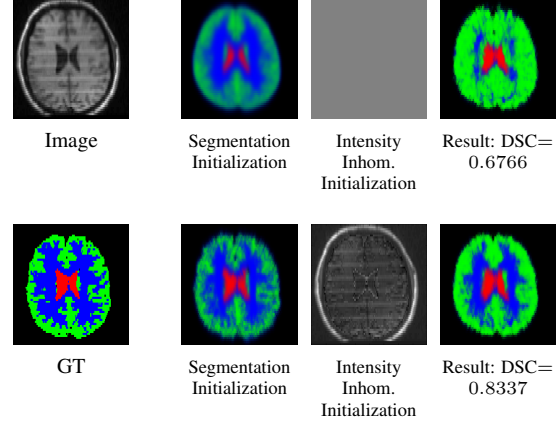


Figure 7: (Color Figure) The results of segmenting a brain slice using the LogOdds method from [21] with different initializations. The initialization consists of an initial probabilistic segmentation in the shape space and an initial intensity inhomogeneity parameter for each pixel. When given the optimal initialization (bottom row), a good segmentation is found, having a DSC of 0.8337. However, when given an initialization of zeros (top row), the results are much worse, with the resulting segmentation having a DSC of 0.6766. The minimization gets stuck in a local minimum, as can often happen with non-convex energies. See the second row of Fig. 3 for our method’s performance on this image (requiring no initialization).

well here with certain initializations, but given a standard initialization, the method gets stuck in local minima corresponding to poor segmentations. What initializations are “good” cannot necessarily be determined beforehand, and thus local minima cause problems for non-convex methods. Such problems are avoided with a strictly convex method, with results that are independent of initialization.

Finally, in Fig. 8, we segment 3 of the vertebrae in an image of a spine. We found  $K = 200$  eigenmodes from 248 training segmentations. This is another practical demonstration for the use of a shape prior, as the regional intensity priors and edge terms are very weak due to the low contrast and similar intensities between regions. This is made clear when the segmentation performed without shape priors is very different from the ground truth and the DSC drops from 0.8273 with shape priors to 0.1336 without.

## 4. Conclusion

We have provided a unique and innovative method for image segmentation that, to our knowledge, is the first to simultaneously achieve four important goals: convex, with shape prior, multi-region, and probabilistic. We define a convex energy function that captures desirable features, like shape and regional priors, to ensure an accurate segmentation. Our algorithm returns a meaningful probabilistic segmentation with uncertainties that reflect the relative uncer-

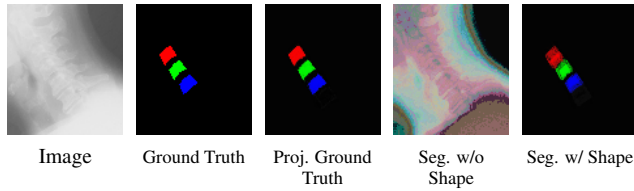


Figure 8: (Color Figure) The segmentations of the vertebrae in the image of a spine. Note that the regional intensities priors and edge terms are weak, so when the shape priors are eliminated the segmentation is very poor. The images came from the U.S. National Library of Medicine website at <http://archive.nlm.nih.gov/proj/ftp/ftp.php> and were accompanied by expert segmentations used as the ground truths.

tainties inherent in the information used to build the energy function. We extended our method to multi-region segmentation while maintaining all of our previous qualities, thus achieving the stated goals. A remaining challenge is to incorporate pose estimates into our method [9, 25] while maintaining the four goals.

We have recently explored specific energy terms in ILR space for thigh muscle segmentation [2], with good results. In the future we will look at other specific energy terms for different tasks.

## References

- [1] J. Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 139–177, 1982.
- [2] S. Andrews, G. Hamarneh, A. Yazdanpanah, B. HajGhanbari, and W. D. Reid. Probabilistic multi-shape segmentation of knee extensor and flexor muscles. In *MICCAI*, 2011.
- [3] S. Andrews, C. McIntosh, and G. Hamarneh. Convex multi-region probabilistic segmentation with shape prior in the isometric logratio transformation space. Technical Report TR 2011-03, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada, July 2011.
- [4] E. Brown, T. Chan, and X. Bresson. Convex formulation and exact global solutions for multi-phase piecewise constant Mumford-Shah image segmentation. Technical report, UCLA CAM Report, July 2009.
- [5] T. Chan, S. Esedoglu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *Algorithms*, 66(5):1632–1648, 2004.
- [6] N. Changizi and G. Hamarneh. Probabilistic multi-shape representation using an isometric log-ratio mapping. In *MICCAI*, pages 563–570. Springer, 2010.
- [7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *CVIU*, 61(1):38–59, 1995.
- [8] D. Cremers. Dynamical statistical shape priors for level set-based tracking. *TPAMI*, 28(8):1262–1273, 2006.
- [9] D. Cremers, F. Schmidt, and F. Barthel. Shape priors in variational image segmentation: convexity, Lipschitz continuity and globally optimal solutions. In *CVPR*, pages 1–6, 2008.
- [10] A. Delong and Y. Boykov. Globally optimal segmentation of multi-region objects. In *ICCV*, pages 285–292, 2009.
- [11] J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal. Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3):279–300, 2003.
- [12] L. Grady. Multilabel random walker image segmentation using prior models. In *CVPR*, pages 763–770, 2005.
- [13] G. Hamarneh, P. Jassi, and L. Tang. Simulation of ground-truth validation data via physically-and statistically-based warps. In *MICCAI*, pages 459–467. Springer, 2008.
- [14] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *TPAMI*, 25(10):1333–1336, 2003.
- [15] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *IJCV*, 1(4):321–331, 1988.
- [16] J. Lellmann, F. Becker, and C. Schnorr. Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In *ICCV*, pages 646–653, 2009.
- [17] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *CVPR*, pages 316–323, 2002.
- [18] S. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.
- [19] T. Pock, A. Chambolle, D. Cremers, and H. Bischof. A convex relaxation approach for computing minimal partitions. In *CVPR*, pages 810–817, 2009.
- [20] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *ECCV*, pages 792–805, 2008.
- [21] K. Pohl, J. Fisher, S. Bouix, M. Shenton, R. McCarley, W. Grimson, R. Kikinis, and W. Wells. Using the logarithm of odds to define a vector space on probabilistic atlases. *MIA*, 11(6):465–477, 2007.
- [22] A. Saad, G. Hamarneh, and T. Moeller. Exploration and visualization of segmentation uncertainty using shape and appearance prior information. *TVCG*, 16(6):1366–1375, 2010.
- [23] Q. Song, X. Wu, Y. Liu, M. Garvin, and M. Sonka. Simultaneous searching of globally optimal interacting surfaces with shape priors. In *CVPR*, pages 2879–2886, 2010.
- [24] O. Veksler. Star shape prior for graph-cut image segmentation. In *ECCV*, pages 454–467. Springer, 2008.
- [25] N. Vu and B. Manjunath. Shape prior segmentation of multiple objects with graph cuts. In *CVPR*, pages 1–8, 2008.
- [26] S. Warfield, K. Zou, and W. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *TMI*, 23(7):903, 2004.
- [27] C. Zach, D. Gallup, J. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. *Vision, modeling, and visualization*, page 243, 2008.