

Image Segmentation Using Disjunctive Normal Bayesian Shape and Appearance Models

Fitsum Mesadi, *Student Member, IEEE*, Ertunc Erdil, Mujdat Cetin, *Member, IEEE*, and Tolga Tasdizen, *Senior Member, IEEE*

Abstract—The use of appearance and shape priors in image segmentation is known to improve accuracy; however, existing techniques have several drawbacks. For instance, most active shape and appearance models require landmark points and assume unimodal shape and appearance distributions, and the level set representation does not support construction of local priors. In this paper, we present novel appearance and shape models for image segmentation based on a differentiable implicit parametric shape representation called disjunctive normal shape model (DNSM). The DNSM is formed by disjunction of polytopes which themselves are formed by conjunctions of half-spaces. The DNSM’s parametric nature allows the use of powerful local prior statistics, and its implicit nature removes the need to use landmarks and easily handles topological changes. In a Bayesian inference framework, we model arbitrary shape and appearance distributions using nonparametric density estimations, at any local scale. The proposed local shape prior results in accurate segmentation even when very few training shapes are available, because the method generates a rich set of shape variations by locally combining training samples. We demonstrate the performance of the framework by applying it to both two and three dimensional datasets with emphasis on biomedical image segmentation applications.

Index Terms—Segmentation, disjunctive normal forms, shape priors, appearance models, Bayesian, variational methods.

I. INTRODUCTION

THE use of prior information about shape and appearance is critical in many image segmentation problems. These include scenarios where the object of interest is poorly differentiated from surrounding structures due to low image contrast, noise, missing or diffuse boundaries, and occlusion. For example, in medical image segmentation, tissue boundaries can be missing due to a low signal-to-noise ratio of the acquisition device, or there may be no clear boundary due to similarities of the surrounding tissues [1]. In such cases, segmentation of object(s) of interest without use of the prior knowledge of the object’s shape and appearance cannot give satisfactory results. In this paper, we propose a general segmentation framework that integrates shape and appearance priors in a way that overcomes some of the major limitations of the techniques currently available in the literature.

F. Mesadi is with the Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT, USA e-mail:fitsum.mesadi@utah.edu

E. Erdil, is with the Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey

M. Cetin, is with the Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey; and he is also with Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

T. Tasdizen is with the Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT, USA

A. Related Work

One of the earliest attempts to include a shape prior in image segmentation is the active contour model, also called Snakes, by Kass *et al.* [2]. In the Snakes, a general regularity term is used as the shape prior, where the smoothness and length of the curve serve as a penalty, which is based on the assumption that smoother and shorter curves are more likely [3]. However, in many applications a more informative object-type specific shape prior can be learned from training samples. In this regard, active shape models (ASM) by Cootes *et al.* [4] are powerful techniques for segmentation using shape priors. Variants of the ASM and their applications to different image segmentation problems can be found in [5]–[7], and a review in [8]. In the original ASM, a training set of shapes represented by landmarks is used to construct allowable shape variations via principal component analysis (PCA). The use of linear analysis tools such as PCA in ASMs limits the domain of applicability of these techniques to shape priors that can only be modeled as unimodal densities. That is, the original ASMs assume that the training shapes are distributed according to a Gaussian-like distribution; hence, the technique cannot model more complex (multimodal) shape distributions.

Several methods have been proposed to handle multimodal distributions of shapes by extending ASMs [9]–[11]. These approaches include the use of mixture of Gaussians [9] and manifold learning techniques [10]. However, these approaches use parametric probability distributions, which may not model very complex shape variations [12]. In addition, the explicit shape representation used in ASM models has two major shortcomings. First, annotating landmark points with correct correspondences across all example shapes can be difficult and time consuming. Second, the extensions of the technique to handle topological changes and segment multiply connected objects are not straightforward. Active appearance models (AAM), an extension of the ASM, are also proposed by Cootes *et al.* [13] to model the possible appearance variations in the training samples. However, like the ASM, the AAM also assumes a Gaussian-like distribution of the appearances and models the appearance variations using PCA.

To overcome some of the limitations of ASMs, level set based shape priors were proposed [1], [3], [14]–[17]. Because of their implicit nature, level set methods do not need landmarks and can easily handle topological changes. In [14] and [1], shape variability is captured using PCA on signed distance functions of level sets. However, the space of signed distance functions is a nonlinear manifold; hence, linear combinations

of shapes no longer correspond to valid shapes. In order to solve this challenge, Kim *et al.* [3] and Cremers *et al.* [15] proposed nonparametric density estimation based shape priors within a level set framework. Their priors do not assume the space of signed distance functions to be linear and also allow a good approximation of multimodal shape distributions. However, the use of level set based shape and appearance priors for image segmentation has several drawbacks; some of the major ones are discussed next.

First, the similarities between two shapes are compared only on a global scale in level set based approaches. Therefore, during segmentation using level set based shape priors, the candidate shape is forced to move towards the globally similar training shapes without consideration for any local shape similarities [3], [15]. For example, in Fig. 1 the top half of shape (a) is similar to the top half of shape (b), and the bottom half of shape (a) is similar to the bottom half of shape (c). Therefore, if the shapes in Fig. 1 (b) and (c) are our training samples, and the shape in (a) is our test sample, then during segmentation, looking at local shape similarities can greatly improve the segmentation accuracy. However, level set based shape prior segmentation frameworks can compute only how close the shape in Fig. 1 (a) is to shapes in (b) and (c), and not to their local regions. Because of their inability to locally combine training shapes in their segmentation framework, level set based shape priors need large training samples with all the possible shape variations as an independent training sample [15]. Since medical datasets usually don't have many ground truth segmentation shapes (as it is labor intensive to obtain), methods that can develop shape models from a limited number of training samples are advantageous.

Second, level set based shape priors methods use region-based shape dissimilarity metrics [3], [15]. Region-based shape dissimilarities such as the template metric or L2 distance do not always correspond to the true shape dissimilarities observed by humans [3], [15].

Furthermore, the appearance models used in the level set framework are limited to mainly encoding the global object properties such as intensity (color) and texture. Because of lack of locality in the level set shape representation, it is difficult to use more powerful localized appearance models [18]. For instance, modeling the appearance of a prostate region (which has a central gland and peripheral regions that usually have different intensity levels) with a single density function, as is done in global appearance models, may not capture sufficient information for successful segmentation. Therefore, learning the appearance distributions at small local scales results in a better expressive power of the model, and hence can improve the segmentation performance. We refer the reader to a survey by Cremers *et al.* [19] for an overview of the use of priors - including shape and appearance - in level set based image segmentation

In this paper, we compare our algorithm with the level set, ASM, and AAM frameworks, because of their popularity in the medical image segmentation applications and their similarity to the proposed method. However, there are many other priors-based image segmentation methods available in the literature. For instance, Zhang *et al.* [20], [21] proposed

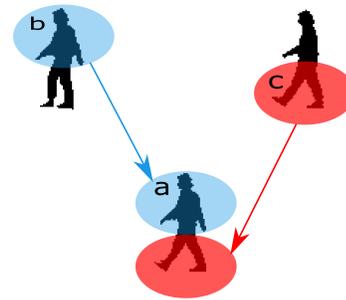


Fig. 1. Similarities of shapes at a local level. The test shape (a) is similar to the training shapes (b) and (c) at different regions.

a sparse shape composition and applied it to various medical image segmentation problems. In [22], watershed segmentation using prior shape and appearance knowledge is presented. In [23], an adaptive shape prior is proposed using a graph cut image segmentation framework. Patenaude *et al.* [24] developed a Bayesian model of shape and appearance for subcortical brain segmentation. In [25], Nguyen *et al.* proposed a classifier-based implicit shape representation using support vector machines. Recently, deep learning methods for shape priors using Boltzmann machines have been presented in [26], [27]. However, most of these techniques require a large number of manually segmented training samples, which are time consuming and labor intensive to obtain [28].

B. Contributions

In order to solve some of the limitations of the currently available techniques in the literature, we propose a novel appearance and shape priors based segmentation in a Bayesian framework. We use an implicit and parametric shape model called Disjunctive Normal Shape Models (DNSM) [29]–[31] to construct our shape and appearance models. The DNSM approximates the characteristic function of a shape as a union of convex polytopes which themselves are represented as intersections of half-spaces. Figure 2(a) shows the intersection of five half-spaces that form a polytope. If we use more half-spaces, we can represent smoother convex polytopes as in Fig. 2(b). The DNSM uses the union of convex polytopes to represent complex shapes, as shown in Fig. 2(c). DNSM is

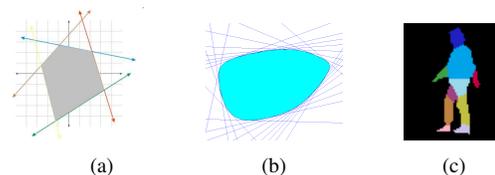


Fig. 2. Illustration of the DNSM shape representation. In (a), the intersection of five discriminant half-spaces is used to create a polygon. In (b), a larger number of half-spaces are used to create a smoother polytope. In (c), the union of ten polytopes are used to represent a walking person shape (each color represents a polytope).

a differentiable shape representation which allows the use of variational approaches. The DNSM's parametric nature allows the use of powerful local prior statistics, and its implicit nature

removes the need to use landmark points and easily handles topological changes.

The major contributions of this paper include:

- *Shape Priors*: We propose DNSM-based global and local shape priors. The DNSM shape representation allows the estimation of nonparametric shape densities at any local scale. The main observation here is that a rich set of candidate shapes can be generated by locally combining training shapes. For instance, in Fig. 1 our model can easily learn the local shape distributions from the two training samples (b and c), and hence can correctly segment the noisy or occluded version of the test shape in (a). Therefore, previously unseen shapes in the testing dataset can be segmented by combining the local regions that are closest to the candidate shape from the training shapes. The advantage of our local shape prior model is particularly significant when there are limited training shapes, as is usually the case in most medical image segmentation problems. In addition, we show that the traditional level set based shape prior methods are a special case of our general shape prior framework.
- *Appearance Models*: Similar to the shape priors, we also propose local and global appearance models using the DNSM. The main observation here is that an object of interest can have varying appearances at different locations, and hence, learning appearance statistics at small corresponding local regions from the training samples can have a better expressive power, compared to a single global appearance distribution. Since the DNSM shape representation has a concept of locality, for instance the different discriminants or the polytopes shown with different colors in Fig. 2, local nonparametric appearance models can easily be constructed at any spatial scale.
- *Segmentation Framework*: We propose a general segmentation framework that incorporates appearance and shape priors in a Bayesian framework.

The rest of the paper is organized as follows. In section II, we review the DNSM shape representation and introduce how the model can be used in image segmentation in a Bayesian framework. In section III, we present novel DNSM global and local shape priors together with a new shape dissimilarity measure. In section IV, we construct global and local appearance models using the DNSM. In section V, we provide qualitative and quantitative analysis of our segmentation framework by applying it to five different datasets: walking person, dendritic spine, brain tumor, central gland, and full prostate. Finally, section VI provides the conclusions. Preliminary results of this work were presented at the MICCAI conference [32]. Compared to [32], this journal version: provides detailed explanations and derivations of the algorithms, formulates the proposed algorithm in a Bayesian framework and tie together the shape and appearance models, gives a generic appearance model that can be used with different features, provides the formulations of the global and local priors for both shape and appearance models, and gives a significantly expanded experimental section. The DNSM has also recently been formulated as a parametric level set framework for image

segmentation in [33], [34].

II. DISJUNCTIVE NORMAL SHAPE MODEL

A. DNSM Shape Representation

Consider the characteristic function $f : \mathbf{R}^D \rightarrow \mathbf{B}$ where $\mathbf{B} = \{0, 1\}$. Let $\Omega^+ = \{\mathbf{x} \in \mathbf{R}^D : f(\mathbf{x}) = 1\}$. Let us approximate Ω^+ as the union of N convex polytopes $\tilde{\Omega}^+ = \cup_{i=1}^N P_i$, where the i 'th polytope is defined as the intersection of $P_i = \cap_{j=1}^M \mathcal{H}_{ij}$ of M half-spaces. \mathcal{H}_{ij} is defined in terms of its indicator function

$$h_{ij}(\mathbf{x}) = \begin{cases} 1, & \sum_{k=0}^D w_{ijk}x_k + b_{ij} \geq 0 \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

where w_{ijk} and b_{ij} are the weights and the bias term, and D is the dimension. Since any Boolean function can be written in disjunctive normal form [35], we can construct

$$\tilde{f}(\mathbf{x}) = \bigvee_{i=1}^N \underbrace{\left(\bigwedge_{j=1}^M h_{ij}(\mathbf{x}) \right)}_{B_i(\mathbf{x})}, \quad (2)$$

such that $\tilde{\Omega}^+ = \{\mathbf{x} \in \mathbf{R}^D : \tilde{f}(\mathbf{x}) = 1\}$. Since $\tilde{\Omega}^+$ is an approximation to Ω^+ , it follows that \tilde{f} is an approximation to f . Our next step is to provide a differentiable approximation to \tilde{f} , which is important because it allows us to use variational approaches; in other words, it allows us to formulate various energy functions and to minimize them with respect to the parameters of the model. First, the conjunction of binary variables $\bigwedge_{j=1}^M h_{ij}(\mathbf{x})$ can be replaced by the product $\prod_{j=1}^M h_{ij}(\mathbf{x})$. Then, using De Morgan's laws [35] we replace the disjunction of the binary variables $\bigvee_{i=1}^N B_i(\mathbf{x})$ with $\neg \bigwedge_{i=1}^N \neg B_i(\mathbf{x})$, which in turn can be replaced by the expression $1 - \prod_{i=1}^N (1 - B_i(\mathbf{x}))$. Finally, we approximate $h_{ij}(\mathbf{x})$ with logistic sigmoid functions $\sigma_{ij}(\mathbf{x}) = \frac{1}{1 + e^{-\sum_{k=0}^D w_{ijk}x_k + b_{ij}}}$ to get the differentiable approximation of the characteristic function $\hat{f}(\mathbf{x})$

$$\hat{f}(\mathbf{x}; \mathbf{W}) = 1 - \prod_{i=1}^N \left(1 - \prod_{j=1}^M \frac{1}{\underbrace{1 + e^{\sum_{k=0}^D w_{ijk}x_k + b_{ij}}}_{g_i(\mathbf{x})}} \right), \quad (3)$$

where $\mathbf{x} = \{x, y, 1\}$ for two-dimensional (2D) shapes and $\mathbf{x} = \{x, y, z, 1\}$ for three-dimensional (3D) shapes, and $\mathbf{W} = \{w_{ijk}\}$. By appending 1 to the pixel coordinates \mathbf{x} , we use w_{ijk} to represent both the weights and the biases in the rest of the paper.

The only adaptive parameters, w_{ijk} , are the weights and biases of the first layer of logistic sigmoid functions $\sigma_{ij}(\mathbf{x})$ which define the orientations and positions of the linear discriminants that form the shape boundary. The conjunctions of $\sigma_{ij}(\mathbf{x})$ form the convex polytopes $g_i(\mathbf{x})$. Fig. 2(a) and (b) show convex polytope examples with different choices for M . The disjunction of $g_i(\mathbf{x})$ forms the shape model $\hat{f}(\mathbf{x}; \mathbf{W})$ (Fig. 2 (c)). Note that, $f(\mathbf{x}) : \mathbf{R}^D \rightarrow [0, 1]$, and the level set $f(x) = 0.5$ represents the interface between the foreground

$f(\mathbf{x}) > 0.5$ (inside the shape) and background $f(\mathbf{x}) < 0.5$ (outside the shape) regions.

Next, we discuss how to initialize the parameters of the DNSM. We use the domain knowledge to decide on the small number of polytopes needed N . For example, in order to represent a walking person shape with DNSM, we can use around $N=10$ polytopes to capture the limbs, head, torso, etc as shown in Fig. 2. Similarly, we can easily pick the right number of polytopes for medical organ (or tissue). DNSM shape representation is less sensitive to the number of discriminants per polytope, M . We use $M = 16$ for 2D images and $M = 32$ for 3D segmentation. The initialization polytopes are approximated as discs for 2D (and spheres for 3D) of a fixed radius. The set of N seed points, $C(i)$, $i = 1$ to N , are chosen such that they are well distributed in the region of interest. For 2D image segmentation, the initial disc approximations are obtained by choosing the parameters, w_{ijk} , as

$$w_{ijk} = \begin{cases} \cos\left(\frac{2\pi j}{M}\right), & k = 0 \\ \sin\left(\frac{2\pi j}{M}\right), & k = 1 \\ -\left(r + \left(C_x(i) \times \cos\left(\frac{2\pi(j-1)}{M}\right)\right)\right) \\ + \left(C_y(i) \times \sin\left(\frac{2\pi(j-1)}{M}\right)\right), & k = 2 \end{cases} \quad (4)$$

where r is the radius of the initial disc. $C_x(i)$ and $C_y(i)$ are the center point coordinates for disc polytope i . $j = [0, \dots, M-1]$. For 3D images segmentation, we can use similar approach to obtain the parameters, w_{ijk} , that approximate sphere polytopes.

B. Image Segmentation Using DNSM

The goal of DNSM-based segmentation can be formulated as the estimation of the optimal DNSM parameters, $\mathbf{W} = \{w_{ijk}\}$, given an image $I : \Omega \rightarrow R$. In the Bayesian framework, this can be computed by maximizing the posterior distribution

$$\mathcal{P}(\mathbf{W}/I) \propto \mathcal{P}(I/\mathbf{W})\mathcal{P}(\mathbf{W}). \quad (5)$$

The most probable segmentation of a given image is then obtained by maximizing the posterior probability in equation (5), which is equivalent to minimizing its negative logarithm. Therefore the energy to be minimized is given as

$$E(\mathbf{W}) = \underbrace{-\log \mathcal{P}(I/\mathbf{W})}_{E_{\text{Appr}}} - \underbrace{\log \mathcal{P}(\mathbf{W})}_{E_{\text{Shape}}} \quad (6)$$

where the first term is an image-based term which depends on the observation I that can consist of appearance features such as intensity, color and texture. The second term is based on geometry, and it can be used to introduce prior shape knowledge [19]. Therefore, the first term in (6) can be seen as energy from appearance, E_{Appr} , and the second term as energy from shape, E_{Shape} . The segmentation is achieved by minimizing the weighted average of the shape and appearance

energy terms. The energy minimization implies computing the derivatives of both the shape and appearance terms with respect to each discriminant parameters, w_{ijk} . By applying gradient descent to the combined energy, the update to the discriminants w_{ijk} is then given as

$$w_{ijk} \leftarrow w_{ijk} - \alpha \frac{\partial E_{\text{Shape}}}{\partial w_{ijk}} - \gamma \frac{\partial E_{\text{Appr}}}{\partial w_{ijk}} \quad (7)$$

In the next two sections, we present the details on how we formulate the E_{Shape} and E_{Appr} for DNSM-based image segmentation.

III. DNSM SHAPE PRIORS

In this section, we describe how a DNSM shape prior can be constructed from a set of training shapes and used in the segmentation of new images. The set of parameters $\mathbf{W} = \{w_{ijk}\}$ of the DNSM is used to represent shapes; hence, shape statistics will be constructed in this parameter space. Therefore, in this section, we first present how to represent the training shapes using $\mathbf{W} = \{w_{ijk}\}$ parameters of the DNSM. Then, we detail the construction of global and local shape prior energy terms.

A. Representation of Training Shapes with DNSM Parameters

In order to obtain pure shape statistics, it is important to first remove pose variations (scale, translation, and rotation) from the training samples [8]. Procrustes analysis [36] or other shape alignment (matching) techniques can be used to first align the training samples before representing them with the DNSM parameters. Alternatively, the DNSM can also be used for binary image registration in the discriminant parameter domain. Shape transformation for alignment using the DNSM has a significant computational advantage during segmentation, and it is given in the Appendix.

The DNSM discriminant parameters, \mathbf{W}^t , that represent a given training shape, t , can be obtained by choosing the weights that minimize the energy

$$E(\mathbf{W}^t) = \int_{\Omega} (f(\mathbf{x}) - q_t(\mathbf{x}))^2 d\mathbf{x} + \eta \sum_i \sum_{r \neq i} \int_{\Omega} g_i(\mathbf{x}) g_r(\mathbf{x}) d\mathbf{x} \quad (8)$$

where $g_i(\mathbf{x})$ represents the individual polytopes of $f(x)$ as given in equation (3). $q_t(\mathbf{x})$ is the ground truth segmentation (1 for object and 0 for background) of the t^{th} training shape, and η is a constant. The first term in (8) fits the model to the training shape by minimizing the mean square error between the level set value f and the ground truth q . The energy from the first term in (8) is minimum when $f = 1$ inside the object shape (where the intensity value of the ground truth is $q = 1$), and $f = 0$ outside the object shape (where the intensity value of the ground truth is $q = 0$). The second term in (8) minimizes the overlap between the different polytopes. The motivation for including such an overlap constraint is to obtain a representation which can provide better correspondence across training samples, in addition to helping us develop local appearance model for each region represented by a unique polytope in Section IV. An η value of 0.1 is experimentally

found to be sufficient to avoid the overlapping of the polytopes. Figure 3 shows the effect of the second term in (8). There is almost no overlap of the polytopes in Fig. 3(c) compared to (b).

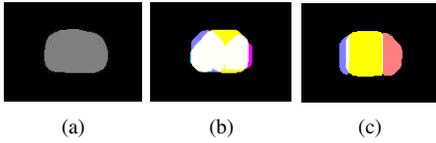


Fig. 3. Demonstration of the effect of the overlap constraint term of equation (8). (a) One slice of a prostate ground truth to be represented by the DNSM. (b) and (c) show the DNSM representation with 3 polytopes. In (b), no overlap constraint is used, $\eta = 0$. In (c), the overlap constraint of $\eta = 0.05$ is used. The different colors show the individual polytopes, and the white part in (b) and (c) shows the area where at least two of the polytopes overlap.

We minimize (8) using gradient descent to obtain \mathbf{W}^t which represents the t^{th} training shape. The derivative of (8) with respect to the discriminant parameters, w_{ijk} , is given as

$$\frac{\partial E}{\partial w_{ijk}} = 2(f(\mathbf{x}) - g_t(\mathbf{x})) \frac{\partial f(\mathbf{x})}{\partial w_{ijk}} + \eta \sum_{r \neq i} g_r(\mathbf{x}) \frac{\partial g_i(\mathbf{x})}{\partial w_{ijk}} \quad (9)$$

where $\frac{\partial f(\mathbf{x})}{\partial w_{ijk}} = \left(\prod_{r \neq i} (1 - g_r(\mathbf{x})) \right) \frac{\partial g_i(\mathbf{x})}{\partial w_{ijk}}$, and $\frac{\partial g_i(\mathbf{x})}{\partial w_{ijk}} = -g_i(\mathbf{x})(1 - \sigma_{ij}(\mathbf{x}))x_k$. Therefore, during the evolution, the discriminant parameters are updated on each iteration as $w_{ijk} \leftarrow w_{ijk} - \gamma \frac{\partial E}{\partial w_{ijk}}$, where γ is the step-size.

We have found that a common initialization (e.g. the mean shape) for all the registered training shapes together with the second term in (8) is sufficient to keep the correspondence among the discriminants and polytopes across the training shapes for most applications. Figure 4(d-f) shows the correspondence achieved between the polytopes across the shapes in (a-c). This is an advantage over ASMs, which can require manually placed landmark points to ensure correspondence. We use walking person silhouettes to demonstrate our shape model, since person silhouettes have shapes that are easily recognizable by any reader and they also have larger shape deformations (and variations) compared to most biomedical organs.

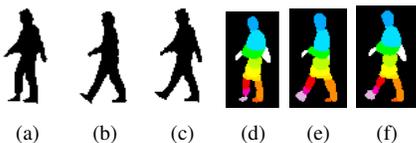


Fig. 4. (a)-(c) are shapes from the walking silhouettes dataset [15]. (d)-(f) show the non overlapping polytopes for shapes in (a)-(c) respectively, using the DNSM. The fifteen polytopes ($N=15$) are shown with different colors.

B. DNSM Global Shape Prior

Our goal is to estimate how likely a candidate shape represented by parameters \mathbf{W} is, given T training shapes $\mathbf{W}^1, \dots, \mathbf{W}^t, \dots, \mathbf{W}^T$. Following an approach similar to [3], we assume that the T aligned shapes are i.i.d. We can write

the Parzen probability density function, $\mathcal{P}(\mathbf{W})$, given the discriminant parameters of the training shapes, \mathbf{W}^t , as

$$\mathcal{P}(\mathbf{W}) = \frac{1}{T} \sum_{t=1}^T \mathcal{K}(d(\mathbf{W}, \mathbf{W}^t), \sigma) \quad (10)$$

where \mathcal{K} is a one-dimensional Gaussian kernel of standard deviation σ , and $d(\mathbf{W}, \mathbf{W}^t)$ is the shape distance between the candidate shape and t^{th} training sample. We can define the distance between two shapes as the sum of the absolute difference between their corresponding normalized discriminant parameters, which is given as

$$d(\mathbf{W}, \mathbf{W}^t) = \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^{D+1} \left(\left| \frac{w_{ijk}}{w_{kA}} - \frac{w_{ijk}^t}{w_{kA}^t} \right| \right) \quad (11)$$

where w_{kA} is the average of the k^{th} weight across all discriminants. Notice that we are dividing the weight of each discriminant by its average value to normalize its contribution to the overall distance summation. This normalization is necessary because the bias weights are typically much larger than the other weights. The shape dissimilarity distance in (11) has some advantages compared to the approaches used in level set based shape prior techniques. In [15] and [3], region overlap based techniques such as template and L2 metrics are used to evaluate the shape dissimilarity distance. However, these region-based metrics do not always provide a good shape dissimilarity measure. For instance, a rectangle with a small protruding part is visually very different from a normal rectangle, but the region-based similarity metrics still show small shape dissimilarity (or large shape similarity) between them. On the other hand, using the proposed method in equation (11), in order for two shapes to have good similarity, the corresponding discriminant planes (or lines in 2D) of the two shapes should have similar angles and locations (biases). Hence, our shape dissimilarity is based on the geometric representation of the shapes and has the potential to correspond closer to true shape dissimilarity observed by humans.

As presented in equation (6), the shape energy is defined as the negative logarithm of (10). The energy minimization implies computing the derivatives of the shape energy terms with respect to each discriminant parameters, w_{ijk} .

C. DNSM Local Shape Prior

One of the major limitations of existing level set based shape priors is that the similarity between the candidate and training shapes is computed only globally. Since no local shape similarity is considered, these approaches cannot generate shape variations by locally combining training shapes. As discussed in Section I-A using Fig. 1, the capability to measure and use shape similarity at local level helps to segment shapes that are not seen in the training. That is, we do not need to have a specific shape in the training sample as long as there are other training samples that can be locally combined to create the candidate shape.

Unlike the traditional level set formulations, our DNSM shape representation allows shape density estimation at any local scale. Such local shape priors are of particular interest

in applications in which different object parts can have independent (uncoupled) statistical shape variability.

Next, we discuss how the global shape probability density in (10) can be modified to obtain a local shape density. Let a given shape be represented by N polytopes and M discriminants per polytope using the DNSM. Let us also assume that each local region is represented by a single polytope (see Fig. 4(d-f)). We make this assumption for ease of explanation. It can easily be relaxed so that the local region can be of any size. We study shape density estimation for each region (polytope) independently by decoupling the entire shape into N local regions (polytopes). We can write the probability density function of the candidate's i^{th} polytope shape, represented by the weight \mathbf{W}_i , given the discriminant parameters of the training shapes for the corresponding polytope \mathbf{W}_i^t as

$$\mathcal{P}(\mathbf{W}_i) = \frac{1}{T} \sum_{t=1}^T \mathcal{K}(d(\mathbf{W}_i, \mathbf{W}_i^t), \sigma_i) \quad (12)$$

where \mathcal{K} is a Gaussian kernel of standard deviation σ_i , and $d(\mathbf{W}_i, \mathbf{W}_i^t)$ is the i^{th} polytope (local) shape distance between the candidate shape and t^{th} training sample. We define the distance between two polytopes as the sum of the absolute difference between their corresponding normalized discriminant parameters (similar to equation (11), but limited to the two polytopes under consideration).

The shape energy for the i^{th} polytope is then defined as the negative logarithm of (12). During segmentation, the update to the discriminant weights $\left(w_{ijk} \leftarrow w_{ijk} - \alpha \frac{\partial E_{\text{Shape},i}}{\partial w_{ijk}} \right)$ for the i^{th} polytope from the shape term is obtained by minimizing the energy using gradient descent as

$$\frac{\partial E_{\text{Shape},i}}{\partial w_{ijk}} = \frac{1}{p(\mathbf{W}_i)n\sigma_i^2} \sum_{t=1}^T \mathcal{K}(d(\mathbf{W}_i, \mathbf{W}_i^t), \sigma_i) (w_{ijk} - w_{ijk}^t). \quad (13)$$

Equation (13) shows that at local minima, the candidate polytope shape is a weighted average of the corresponding polytope training shapes, where the weight depends on the similarity ($d(\mathbf{W}_i, \mathbf{W}_i^t)$) between the polytope of the candidate shape and that of the given training sample. Therefore, in the local region represented by a given polytope, the shape prior term forces that part of the segmented image to move towards the locally closest plausible shapes. Although we have only shown when the local region size is defined to be a single polytope, the local size can easily be changed (to multiple of polytopes) by modifying equations (12 - 13) accordingly. If necessary, the combination of shape priors computed at varying local scales can be used during segmentation.

IV. DNSM APPEARANCE MODELS

Appearance models (priors), learned from the training datasets, are frequently used in image segmentation. In this section, we present how to construct global and local appearance models. Since the main application of this paper is medical image segmentation, we only use intensity and texture to demonstrate the model. However, the proposed appearance model framework can be used with color and any

other advanced appearance features (such as HOG [37] and CNN [38]) depending on the application area.

The $\mathcal{P}(I/\mathbf{W})$ appearance factor in equation (5) is approximated by an appearance distributions at the foreground and background regions, where the regions are formed by \mathbf{W} . Maximization of the posterior appearance pdf is the same as minimizing its negative logarithm. Since $f(\mathbf{x})$ is close to 1 inside the object and close to 0 outside, it acts as a Heaviside function. Therefore, the energy to be minimized is given as

$$E_{\text{GAppr}}(\mathbf{W}) = - \int_{\Omega} (f(\mathbf{x}) \log \mathcal{P}_f + (1 - f(\mathbf{x})) \log \mathcal{P}_b) d\mathbf{x} \quad (14)$$

by assuming independence of appearances at the foreground and background regions, where \mathcal{P}_f and \mathcal{P}_b are the pdfs in the foreground and background regions, respectively. We use \mathcal{P}_f instead of $\mathcal{P}_f(I(\mathbf{x}), \mathbf{x})$ to simplify the notation. We also use f or $f(\mathbf{x})$ instead of $f(\mathbf{x}; \mathbf{W})$ for a similar reason.

Given a training image dataset the intensity probability density functions for the foreground, \mathcal{P}_f , and for the background, \mathcal{P}_b , regions can be constructed using nonparametric density estimation [39], [40]. Smoothed normalized intensity histograms can be used to estimate the nonparametric Parzen density [41] pdfs from the training images. Therefore, the energy minimization implies computing the derivatives of the appearance energy with respect to each discriminant parameters, w_{ijk} . During segmentation, the update to the discriminant weights from the appearance term is obtained by minimizing the energy using gradient descent.

In our DNSM model, in addition to intensity, we use texture features to obtain appearance priors. Grey-level co-occurrence matrix (GLCM) texture measurements [42] are used to obtain eight features (energy, entropy, correlation, difference moment, inertia, cluster shade, cluster prominence, and Haralick's correlation) that summarize the texture T of a given image. Similar to the intensity pdf, the texture pdf at the foreground \mathcal{P}_{Tf} and at the background \mathcal{P}_{Tb} are estimated using the nonparametric Parzen density [41] from the training images. Assuming the texture and intensity pdfs are independent, the energy from the appearance model given in (14) can now be modified as

$$E_{\text{GAppr}} = - \int_{\Omega} (\log \mathcal{P}_f(I/\mathbf{W}) + \log \mathcal{P}_{Tf}(T/\mathbf{W})) f(\mathbf{x}) + (\log \mathcal{P}_b(I/\mathbf{W}) + \log \mathcal{P}_{Tb}(T/\mathbf{W})) (1 - f(\mathbf{x})) d\mathbf{x} \quad (15)$$

The above probability densities and the corresponding energy from appearance in (15) are global functions. That is, the probability densities are estimated from the foreground and background training pixels, without any consideration for the local relative positions of the pixels. However, images and objects usually have spatially varying appearance distributions. For example, the prostate in Fig. 5(a) is surrounded by different structures (tissues) from different sides. In addition, the prostate itself can have a spatially varying appearance distribution. Hence, studying appearances at smaller local regions has the potential to improve the expressive power of our statistical framework, thereby increasing image segmentation accuracy. Our DNSM allows the estimation and use of appearance models at any scale. The local scale can be

the neighborhood of a discriminate plane (see Fig. 2(a)), or a polytope (see Fig. 4(d-f)), or even a union of polytopes. For our next discussion, let us assume the local scale to be the neighborhood of a given discriminant plane. It can easily be extended to the cases when the local scales are a polytope or combination of polytopes.

The first step in local appearance pdf estimation is representing the training shapes with the DNSM using (8). Then, each pixel in the region of interest is assigned to its closest discriminant plane using point-to-plane distance (point-to-line distance in 2D images). In order to have a proper local appearance prior, the different polytopes should cover non-overlapping regions, which is achieved by the second term of equation (8), as seen in Fig. 4(d-f) and Fig. 3. In the previous global appearance prior case, we simply estimated the global intensity and texture pdfs for foreground and background regions. However, in the current local appearance prior case, each discriminant plane h_{ij} has its own set of pixels (or voxels) that are closest to it. Let us denote the h_{ij} discriminant with just h to simplify the notations.

We use a similar approach discussed above (for global energy), except that we compute the different pdf for each discriminant h using its local region. For the h^{th} discriminant, the intensity pdf for the foreground $\mathcal{P}_{h,f}$ and for the background $\mathcal{P}_{h,b}$, and the texture pdf for the foreground $\mathcal{P}_{h,Tf}$ and for the background $\mathcal{P}_{h,Tb}$, are obtained (as discussed before) using a smoothed normalized histograms for a nonparametric density estimation. The global appearance model in equation (15) can now be modified to obtain a local appearance energy given as

$$E_{\text{Appr}} = - \int_{\Omega} (\log \mathcal{P}_{h,f}(I/\mathbf{W}) + \log \mathcal{P}_{h,Tf}(T/\mathbf{W})) f(\mathbf{x}) \\ + (\log \mathcal{P}_{h,b}(I/\mathbf{W}) + \log \mathcal{P}_{h,Tb}(T/\mathbf{W})) (1 - f(\mathbf{x})) d\mathbf{x} \quad (16)$$

See Fig. 5 for an example of intensity, texture, and combined (intensity and texture) appearance probability map obtained using local appearance density estimation. The figure shows the probability in the region-of-interest. The region-of-interest (ROI) used is a sphere with a radius large enough to encompass all the prostate sizes. The radius of the sphere is set to twice the size of the prostate with the largest size in the training dataset. The ROI is placed automatically with its center set to the mean location obtained from the center of masses of all the prostates in the training. The accuracy of our segmentation algorithm is not sensitive to the ROI, since it is set to include all the voxels (regions) that can plausibly be part of a prostate. The only purpose of using the ROI is to avoid unnecessary computation on regions that are known to not contain any prostate. Since prostate cannot be found outside that region of interest, we automatically assign the areas outside the ROI to zero.

Note that, the appearance energy given in (14) can easily be reformulated to represent the popular Chan and Vese (CV) [43] region-based method that approximates the image into piecewise constant regions. In this case, since the image is assumed to be made up of two piecewise constant regions, a Gaussian probability density with fixed standard deviation and constant mean can be used for the \mathcal{P}_f and \mathcal{P}_b in equation

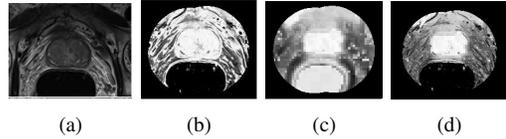


Fig. 5. a) Prostate MRI section. Appearance map using intensity in (b), texture in (c), and combination of both intensity and texture in (d). The brightness corresponds to the probability of the point to be inside the prostate. Note that, we automatically assign zero confidence to areas outside the region-of-interest in the appearance confidence maps.

(14). Hence, $\mathcal{P}_f \propto \exp^{-\eta(I(\mathbf{x})-c_1)^2}$, and $\log \mathcal{P}_f(I(\mathbf{x}), \mathbf{x}) \propto -(I(\mathbf{x}) - c_1)^2$. Therefore, the CV equivalent of the proposed DNSM two-phase piecewise constant region-based variational energy to be minimized is given as

$$E_{\text{Appr}} = \int_{\Omega} (I(\mathbf{x}) - c_1)^2 f(\mathbf{x}) + (I(\mathbf{x}) - c_2)^2 (1 - f(\mathbf{x})) d\mathbf{x} \quad (17)$$

where c_1 and c_2 are the mean intensities in the foreground and background regions, respectively.

V. EXPERIMENTS

In this section, we present experiments conducted on five different datasets. In Section V-A, we use the 2D walking person data [15] to compare the performances of the global and local shape priors. In Section V-B, we analyze the performances of using both the shape and appearance priors on the 2D dendritic spine dataset [44]. In this section, we also evaluate the different combinations of no-priors, global, or local priors, for both the shape and appearance. In Section V-C, we use our appearance model to segment 3D brain tumors using the MICCAI BRATS dataset [45]. No shape priors are used here since brain tumors do not have well-defined shapes. We segment the prostate central gland using 3D data from the ISBI challenge [46] in Section V-D, and full prostate using the MICCAI challenge 3D dataset [47] in Section V-E. In all the experiments, we compare our results with results using a level set framework, and (or) with results from the challenge participants. Since the nonparametric density estimation based level set methods are already shown to perform better than ASM and AAM on multimodal shape datasets [3], [15], we give no explicit comparison of our algorithm with ASM and AAM (except for when they are used by the challenge participants).

A. Walking Person Segmentation

Example images from the walking person dataset, obtained from [15], are shown in Fig. 4. Cremers et al. [15] used the same dataset to evaluate their level set based shape prior segmentation. In order to compare and evaluate the segmentation performances of the global and local shape priors, we add Gaussian noise to the walking person images. We use this section to present the results of two experiments.

First, we show the segmentation result for the walking person example (Fig. 1) introduced at the beginning of this paper to demonstrate our shape prior concept (in section III).

We use the two training samples shown in Fig. 1(b) and (c). Fig. 6(b) is our test image, which is a noise-corrupted version of (a). From the figure, it is clear that the segmentation using the proposed DNSM local shape prior, Fig. 6(f), is closer in shape to the ground truth (a), because the top half of the test image has moved closer to the training shape in Fig. 1(b), whereas the bottom half of the test image has moved closer in shape to the training sample in Fig. 1(c). On the other hand, both the DNSM global shape prior and the level set based shape prior methods move the shape of the test image to one of the two training shapes in Fig. 1(b or c).

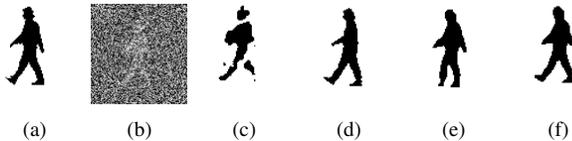


Fig. 6. Comparison of segmentation results with no shape prior, with global, and with local shape priors for noisy walking person. (a) is the test image, (b) is noisy version of (a) to be used for testing. The rest are the segmentation results using: without any shape prior (c), with level set shape prior [3] (d), with the DNSM global shape prior (e), with the DNSM local shape prior (f).

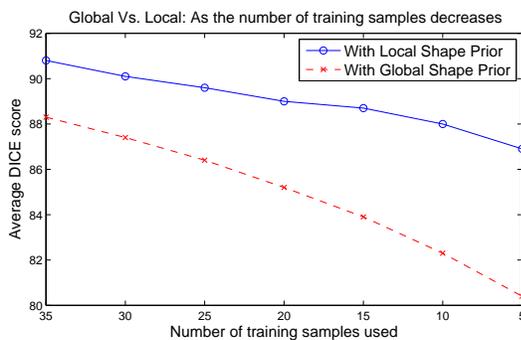


Fig. 7. Segmentation DICE score as a function of the number of training samples.

Next, we show the effect of the number of training shapes on the segmentation result. Figure 7 shows the average DICE score as a function of the number of training samples: using the global shape prior (dashed line), and using the local shape prior (solid line). The graph shows that the performance of global-based shape prior technique fall drastically as the number of training shapes decreases. Whereas, our proposed DNSM local shape prior performs reasonably well even with very small training shapes because the local shape prior method generates a rich set of shape variations by locally combining the few training shapes available.

B. Dendritic Spine Segmentation

In this section, we demonstrate the effect of both the shape and appearance priors using experiments conducted on the segmentation of dendritic spines in 2-photon microscopy images. Dendritic spines are the post-synaptic partners of a synapse. It has been shown that the structure of dendritic spines and their function are highly dependent [44]. Studying such structures

requires accurate segmentation of spines. Segmentation of spines is challenging because spine heads and the dendrite regions usually have similar appearances. In addition, the spine neck region generally has very low contrast and is blurry in 2-photon microscopy images [44]. In this section, we segment 40 spines (with leave-one-out) using images obtained from [44]. Figure 8 shows some of the spines to be segmented and their corresponding ground truth segmentation. From the figure it is clear that segmentation of spines can benefit from both shape and appearance priors. A coarse level classification of spine shapes reveals two spine classes: spines that have a neck, and those that do not. Therefore, this example also helps to demonstrate the performance of the proposed shape priors when the shape distribution involves two shape classes, leading to a multimodal shape density. We use the spines

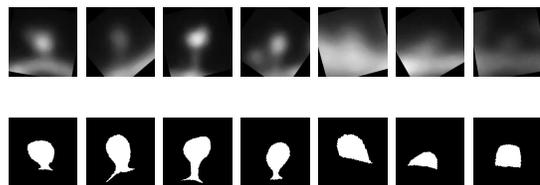


Fig. 8. On the first row are some example spine images; and on the second row the corresponding ground truth segmentation (by a domain expert) binary images.

TABLE I
SPINE SEGMENTATION QUANTITATIVE RESULTS USING DICE COEFFICIENTS

Prior Combinations	DNSM	Level set
No Shape + No Appearance Prior	32.1	31.0
No Shape + Global Appearance Prior	60.3	-
No Shape + Local Appearance Prior	64.8	-
Global Shape + No Appearance Prior	66.7	63.3
Global Shape + Global Appearance Priors	77.2	75.8
Global Shape + Local Appearance Priors	79.4	-
Local Shape + No Appearance Priors	70.5	-
Local Shape + Global Appearance Priors	79.3	-
Local Shape + Local Appearance Priors	82.6	-

in this subsection to compare the performances of the global and local (shape and appearance) priors. Table I gives the average DICE score obtained using the different shape and appearance prior combinations. It can be seen from the table that the result improves as we go from no priors to global priors, and then to local priors (in terms of both the shape and appearance priors). Table I also compares segmentation using the DNSM priors with the level set based priors of [3], [40]. For the level set based segmentation, only the global shape and global appearance prior results are given in the table, since existing level set based frameworks are limited to global models. The DICE score using our DNSM global shape and global appearance priors is slightly better than the corresponding level set based shape and appearance priors. We argue that this is because the proposed DNSM shape distance is better than the template metric used in the level set segmentation. However, the greatest improvement comes

from the ability of our DNSM to model local shape and local appearance priors. In Table I, it should be noted that in cases (rows) when no appearance prior is used, the region-based data term of equation (17) is employed for the DNSM, and the Chan-Vese method [43] is applied for the level set. Therefore, in the Table, no shape plus no appearance means using only equation (17) for DNSM, and Chan-Vese [43] for level set. In this experiment, we used $N = 8$ polytopes and $M = 16$ discriminants per polytope to represent a spine shape with DNSM.

Figure 9 gives an example qualitative segmentation comparison of local DNSM, level set based prior of [3], and the joint classification and segmentation approach of [44]. The DICE scores for the example in Fig. 9 are 83.1 using [3] (c), 83.9 using [44] (d), and 90.2 using our local DNSM approach (e).

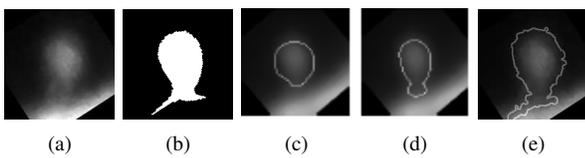


Fig. 9. Spine segmentation: (a) the original intensity spine image to be segmented, (b) the ground truth segmentation image, (c) the segmentation result using the level set based shape prior [3], (d) the segmentation result using [44], and (e) the segmentation result using our DNSM local shape and local appearance model.

One way to get additional insight into the significance of the different variants of our algorithm is to look at standard deviation of the segmentation results on the image samples. Table II shows the minimum, maximum and standard deviation of the variations of our algorithm for the spine segmentation. In the table, min score is the DICE segmentation value for the sample that is segmented poorly, out of the total 40 samples. Similarly, max score is for the sample that is best segmented by the algorithm in terms of the DICE score. From the table, we can see that local appearance + local shape has the smallest standard deviation, since it results in good segmentation on almost all the image samples. Whereas, no appearance + no shape prior has large performance difference for different image samples, resulting in large standard deviation.

Using the mean and standard deviations provided in Tables I and II, we can show if the results obtained with the different variants of our algorithm is statistically significant. By first computing the variance between two different variants of our algorithm and then using t-score, we can obtain the p-value of the hypothesis. An experiment is statistically significant when the p-value is less than the significance level α such as 0.05. First, by using the result of the algorithm that does not use any shape or appearance priors (first row in Table I) as a null hypothesis and the variant that uses local shape + local appearance priors (last row in Table I) as our testing hypothesis, we get a t-score of 18.598 and the corresponding p-value is less than 0.00001. Second, by using the result of the algorithm that does not use any shape or appearance priors as a null hypothesis and the variant that uses global shape + no appearance priors (see Tables I and II) as our testing hypothesis, we get a t-score of 11.147 and the corresponding

p-value is less than 0.00001. Similarly, using global shape + global appearance priors as our null hypothesis and the variant that uses local shape + local appearance as our testing hypothesis we get a t-score of 6.904 and the corresponding p-value is less than 0.00001. Hence, all of these comparisons are deemed statistically significant. Note that, we used one-tailed hypothesis since the improvement in the segmentation result is from one side of the distribution, and the degree of freedom is based on 40 training samples.

TABLE II
SPINE SEGMENTATION STATISTICAL ANALYSIS TO SHOW THE DIFFERENCES BETWEEN THE VARIATIONS OF THE PROPOSED ALGORITHM

Prior Combinations	Min score	Max Score	STD
No Shape + No Appearance	18.4	61.8	15.3
No Shape + Global Appearance	36.8	78.6	12.9
No Shape + Local Appearance	38.5	80.4	12.5
Global Shape + No Appearance	39.7	81.6	12.3
Global Shape + Global Appearance	51.4	85.1	10.6
Global Shape + Local Appearance	54.6	87.2	10.4
Local Shape + No Appearance	45.5	82.6	11.3
Local Shape + Global Appearance	62.7	89.9	8.9
Local Shape + Local Appearance	68.9	92.3	7.8

C. Brain Tumor Segmentation

In this section, we apply our method to segment tumors in multimodal brain images using the MICCAI BRATS-2012 dataset [45]. Segmentation of brain tumors is an important but difficult task in many clinical applications. In this experiment, we use 20 high-grade (HG) and 10 low-grade (LG) images available in the BRATS dataset to segment the active tumor (core) regions. Since tumors do not have a well-defined shape, no shape prior is used in this experiment. Instead, the appearance model (prior) energy of equation (15) is used for the automatic segmentation (with leave-one-out) of the active tumors.

Quantitative results for comparing our segmentation with the top performing automatic methods from the challenge participants [45] are given in Table III. It should be noted that the methods from the challenge [45] are specifically designed for brain tumor segmentation, but we made no special modifications to our general segmentation framework. Recently, deep convolutional neural network is applied to segmentation of brain tumors in [48] using MICCAI BRATS-2013 dataset (which has larger training samples compared to BRATS-2012 we are using for this experiment).

TABLE III
QUANTITATIVE COMPARISON IN % DICE COEFFICIENT FOR SEGMENTATION OF THE BRATS-2012 DATASET

Method	HG	LG
Zikic <i>et al.</i> [49]	71	62
Bauer <i>et al.</i> (with std. dev.) [50]	62 ± 0.27	49 ± 0.26
Geremia <i>et al.</i> [51]	58	20
DNSM	66	67

Figure 10 and 11 show examples of segmentation results for the low-grade and high-grade images, respectively. Figure 11 also shows the potential of our DNSM shape representation for handling topological changes. The tumor locations are first detected using the appearance energy followed by the gradient descent step. The gradient descent of our segmentation starts from the detected bounding box (in green) in Fig. 11(b), and after several iterations the box splits gracefully to capture the multiple small tumor components (see Fig. 11(c)).

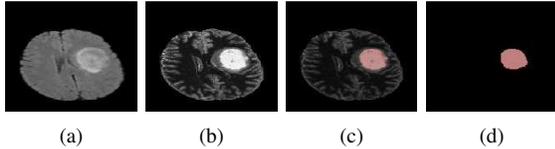


Fig. 10. Segmentation of brain tumors from low-grade multimodal MR images: a) FLAIR image b) the appearance probability map, c) the segmentation (in red) overlaid on the appearance map image, d) the segmentation (in red) overlaid on the ground truth segmentation.

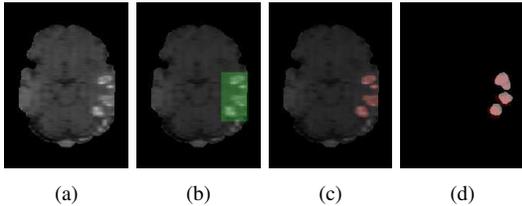


Fig. 11. Segmentation of brain tumors from high-grade multimodal MR images: a) T1 image, b) bounding box (in green) detected using our appearance energy, c) segmentation (in red) overlaid on the T1 image, d) the final segmented tumors (in red) overlaid on the manual segmentation.

D. Prostate Central Gland Segmentation

We use the NCI-ISBI 2013 Challenge - Automated Segmentation of Prostate Structures [46]- MRI dataset to evaluate the performance of our shape and appearance priors on 3D image segmentation. The training data has 60 samples with half obtained at 1.5 T and the other half at 3T. The test data has 10 samples. Since the two acquisition setups (1.5T and 3T) have different appearances, we decided to construct two training appearance models one for each setup. During segmentation (testing), we first detect the class of the image (whether it is 1.5T or 3T), and then use the training data of that particular class. We used $N = 8$ polytopes and $M = 32$ discriminants to represent a central gland shape with DNSM. We first apply histogram normalization to the MRI images, and then we use the appearance and shape prior energy terms of equation (7) to segment the central gland. Automatic segmentation of the central gland in MRI is challenging due to its variability in size, shape, location, and its similarity in appearance to the surrounding structures. Figure 12 shows one slice of MR image segmentation using DNSM local and global shape priors. The local appearance model is used in this experiment.

Table IV compares our algorithm with the top performing results from the NCI-ISBI challenge. From the table it is clear

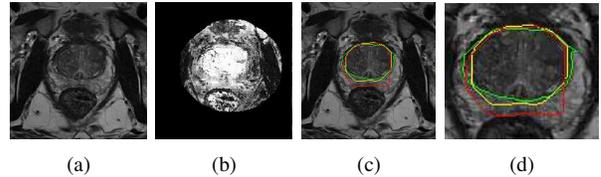


Fig. 12. Central gland segmentation: a) MRI section and b) local appearance map image (the brightness corresponds to the probability of the point to be a central gland). c) Segmentation results with and without the shape prior in yellow and red, respectively; green is the ground truth. d) Result in (c) zoomed in.

that DNSM segmentation with local shape and local appearance models performs better than with the global models. Our algorithm produces better results compared to the top methods in the challenge, both in terms of the DICE coefficient and the mean distance (MD) between the segmented and the ground truth surfaces, for the central gland segmentation part of the challenge. For MD, the smaller the better. Rusu *et al.* in [46] is an atlas-based segmentation, whereas the RUNMC in [46] used an interactive segmentation. Table IV also compares our result with the level set based shape and appearance priors method of [40].

TABLE IV
CENTRAL GLAND SEGMENTATION QUANTITATIVE RESULTS USING AVERAGE DICE AND MEAN DISTANCE (MD).

Method	DICE	MD
DNSM: Local Appearance + No Shape	75.2	2.13
DNSM: Global Appearance + Global Shape	80.9	1.57
DNSM: Local Appearance + Global Shape	82.7	1.32
DNSM: Local Appearance + Local Shape	84.1	1.27
Level set: Appearance + Shape Priors [40]	77.8	-
Atlas: Rusu <i>et al.</i> in [46]	82.1	1.58
Interactive: RUNMC in [46]	80.8	1.83

TABLE V
PROSTATE SEGMENTATION QUANTITATIVE RESULTS

Method	Mean DICE
DNSM: No Shape Prior + Local Appearance	79.5
DNSM: Global Shape + Global Appearance	82.5
DNSM: Local Shape + Global Appearance	84.1
DNSM: Local Shape + Local Appearance	89.2
Level set: Shape Prior + Appearance Prior [40]	79.3
AAM: Vincent <i>et al.</i> in [47]	88.0
ConvNet: Yu <i>et al.</i> in [47]	89.4

E. Full Prostate Segmentation

We finally apply our algorithm to full prostate segmentation using the MICCAI PROMISE 2012 challenge dataset [47]. There are 50 training samples of T2-weighted MR images of the prostate with two different acquisition setups: with and without endorectal coil, each with 25 training samples. Since the two acquisition setups have different appearances, we decided to construct two training appearance models one

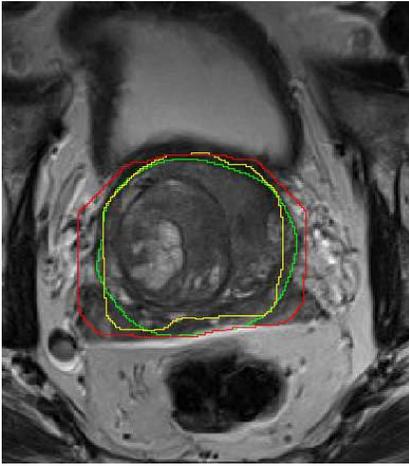


Fig. 13. Prostate segmentation results with local and global appearance priors in yellow and red respectively. Green is the ground truth.

for each setup. During segmentation (testing), first we detect the class of the image (whether it is with coil or without coil) automatically, and then use the training data of that particular class. The experiment is conducted using leave-one-out. We used $N = 8$ polytopes and $M = 32$ discriminants to represent a prostate shape with DNSM. Since the prostate has two distinct regions, the central gland and the peripheral region, a single global appearance prior has limited modeling capability. Learning local appearances in different parts of the prostate during training improves accuracy, as shown in Table V. Our approach, with local appearance and local shape priors, performs equivalent to the result from the top challenge participant. The method in Vincent *et al.* [47] is based on AAM, and Yu *et al.* [47] uses a volumetric convolutional neural network to segment the prostate. The table also compares our result with the level set based shape and appearance priors method [3], [40]. Figure 13 shows a sample segmentation result for one slice using local and global DNSM appearance models.

Note that, the computational time of our proposed algorithm is on the order of seconds for 3D dataset (e.g. prostate) and less than a second for 2D images (e.g. the spine segmentation). Our algorithm is implemented in C++ using the Insight Segmentation and Registration Toolkit (ITK) [52] on 2.5-GHz Intel Core i7, 8 GB RAM machine. The computational cost (C) of our algorithm is proportional to the number of DNSM parameters $M \times N \times D$; where N is number of polytopes, M is the number of discriminants per polytope, and D is the dimensionality of the image (2D or 3D). That is, DNSM that has N number of polytopes with M discriminants per polytope has $N \times M \times 3$ number of parameters to represent 2D shape, and $N \times M \times 4$ number of parameters to represent 3D shapes. Since $N = 8$ and $M = 16$ (or 32 for 3D) successfully represented all the shapes in our experiments, we believe that DNSM can represent complex shapes with relatively smaller number of parameters compared to active shape models (which use landmark points) or level set methods (which use signed distance function).

VI. CONCLUSION

In this paper, we presented priors-based image segmentation using the DNSM shape model in a Bayesian framework. The proposed segmentation framework solves some of the major drawbacks of the commonly used priors-based methods available in the literature for biomedical image segmentation; in particular, active shape and appearance models and level set based shape priors techniques. We have shown the potential of our segmentation framework by applying it to different 2D and 3D datasets. In the experiments, in addition to comparing our results with state-of-the-art methods, we have also shown the advantages of local shape and local appearance priors compared to the global priors. The experimental results further highlights the better suitability of our local shape prior method, particularly in situations when the training data is limited in size. Similarly, we have shown that learning appearance statistics at small local neighborhoods gives better results. Possible extensions of our work include coupled segmentation of multiple objects and object tracking.

APPENDIX

In order to learn the shape statistics of training samples, any pose (translation, scale, and rotation) variation has to be removed first. We needed to align the shapes to remove pose variations from the training samples before shape density estimation, and to register the candidate shape to the pose of the aligned training samples before modifying its shape using shape priors. The use of binary image registration techniques available in the literature is computationally expensive for shapes represented by the DNSM parameters, due to the required transformations between the image domain and the DNSM parameter domain.

In this Appendix, we show how the pose of a shape represented by the DNSM parameters can be changed without going to the image domain. For clarity of presentation, we focus on the translation and scaling for 2D shapes. Let a given shape with an initial (i) center-of-mass at (x_i, y_i) be represented by the DNSM parameters $W = \{w\}$. A discriminant line with a parameter $w = (w_x, w_y, w_b)$ has the x and y coordinate weights (w_x , and w_y) and the bias weight (w_b). Translation and scaling of this shape to a new location and size do not affect the orientations of the discriminants; hence only the bias weight w_b changes after the transformation. Let us say we wanted to change the pose of the current shape by translating it to a new center-of-mass location (x_f, y_f) and scale it by a factor C . Then, the final translated and scaled shape has the same w_x , and w_y parameters, but different (w_b). Therefore, during translation and scaling the only unknown parameter we need to find is the final bias weight (w_f).

Let d_i and d_f be the initial and final (after registering) distance of the center-of-mass point from the current discriminant line. If there is no scaling, then $d_i = d_f$. However, during scaling of the initial shape by C , then $d_f = \sqrt{C}d_i$. Using the point-to-line distance equation, the distance of the center-of-mass point (x_i, y_i) from the discriminant line $x.w_x + y.w_y + w_b$ is given as

$$d_i = \frac{|x_i.w_x + y_i.w_y + w_b|}{\sqrt{w_x^2 + w_y^2}} \quad (18)$$

After translating and scaling of the initial image, the final shape has a discriminant line given by $w_x + y \cdot w_y + w_f b$. $w_f b$ is the final bias weight, since the translated and scaled shape has the same w_x , and w_y parameters. Therefore, the distance of the new (final) center-of-mass point (x_f, y_f) from the discriminant line $x \cdot w_x + y \cdot w_y + w_f b$ is given as

$$d_f = \frac{|x_f \cdot w_x + y_f \cdot w_y + w_f b|}{\sqrt{w_x^2 + w_y^2}} \quad (19)$$

From equations (18) and (19), and the fact that $d_f = \sqrt{C} d_i$, it follows that the final bias weight, $w_f b$, is

$$w_f b = \sqrt{C} (|x_i \cdot w_x + y_i \cdot w_y + w_b|) - (|x_f \cdot w_x + y_f \cdot w_y|) \quad (20)$$

Using equation (20), therefore, the scaled and translated shape has discriminant weights $w_f = (w_x, w_y, w_f b)$. Pose variation due to rotation can similarly be derived using the observation that all the discriminant lines rotate at the same angle. For 3D shapes, the only differences are the use of the point-to-plane distance equation (in place of the point-to-line) and $d_f = \sqrt[3]{C} d_i$ (in place of $d_f = \sqrt{C} d_i$) in the above derivation. Because of the closed-form solution of (20), the back and forth registration of the candidate shape in our segmentation algorithm (for shape and appearance updates) is achieved at a very low computational cost.

ACKNOWLEDGMENT

This work is supported by NSF IIS-1149299(TT), NIH grant 1R01-GM098151-01, TUBITAK-113E603 and TUBITAK-2221. We also like to thank the Neuronal Structure and Function Lab., Champalimaud Center for providing the 2-photon microscopy data used in one of our experiments.

REFERENCES

- [1] A. Tsai, Jr. Yezzi, A., W. Wells, C. Tempny, D. Tucker, A. Fan, W.E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *Medical Imaging, IEEE Transactions on*, 22(2):137–154, Feb 2003.
- [2] K. Michael, W. Andrew, and T. Demetri. Snakes: Active contour models. *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 1(4):321–331, 1988.
- [3] J. Kim, M. Cetin, and A.S. Willsky. Nonparametric shape priors for active contour-based image segmentation. *Signal Processing*, 87(12):3021 – 3044, 2007.
- [4] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38 – 59, 1995.
- [5] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In *Computer Vision ECCV 2008*, volume 5305 of *Lecture Notes in Computer Science*, pages 504–513. Springer Berlin Heidelberg, 2008.
- [6] M. de Bruijne, B. van Ginneken, M. A. Viergever, and W. J. Niessen. Adapting active shape models for 3d segmentation of tubular structures in medical images. In *Information Processing in Medical Imaging*, volume 2732 of *Lecture Notes in Computer Science*, pages 136–147. Springer Berlin Heidelberg, 2003.
- [7] W. Wang, S. Shan, W. Gao, B. Cao, and B. Yin. An improved active shape model for face alignment. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, ICMI '02*, pages 523–, Washington, DC, USA, 2002. IEEE Computer Society.
- [8] T. Heimann and H. P. Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13(4):543 – 563, 2009.
- [9] T.F. Cootes and C.J. Taylor. A mixture model for representing shape variation. *Image and Vision Computing*, 17(8):567 – 573, 1999.
- [10] P. Etyngier, F. Segonne, and R. Keriven. Shape priors using manifold learning techniques. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, Oct 2007.
- [11] M. Kirschner, M. Becker, and S. Wesarg. 3d active shape model segmentation with nonlinear shape priors. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2011*, volume 6892 of *Lecture Notes in Computer Science*, pages 492–499. Springer Berlin Heidelberg, 2011.
- [12] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis*, 16(1):265 – 277, 2012.
- [13] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 484–498. Springer, 1998.
- [14] M.E. Leventon, W.E.L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 316–323 vol.1, 2000.
- [15] D. Cremers, S.J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International Journal of Computer Vision*, 69(3):335–351, 2006.
- [16] T. Chan and W. Zhu. Level set based shape prior segmentation. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 1164–1170 vol. 2, June 2005.
- [17] S. Chen and R.J. Radke. Level set segmentation with both shape and intensity priors. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 763–770, Sept 2009.
- [18] E. Horbert, K. Rematas, and B. Leibe. Level-set person segmentation and tracking with multi-region appearance models and top-down shape information. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1871–1878, Nov 2011.
- [19] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *International Journal of Computer Vision*, 72(2):195–215, 2007.
- [20] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis*, 16(1):265 – 277, 2012.
- [21] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D.N. Metaxas, and X. S. Zhou. Sparse shape composition: A new framework for shape prior modeling. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1025–1032, June 2011.
- [22] G. Hamarneh and X. Li. Watershed segmentation using prior shape and appearance knowledge. *Image and Vision Computing*, 27(12):59 – 68, 2009. Canadian Robotic Vision 2005 and 2006.
- [23] H. Wang, H. Zhang, and N. Ray. Adaptive shape prior in graph cut image segmentation. *Pattern Recognition*, 46(5):1409 – 1414, 2013.
- [24] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson. A bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*, 56(3):907 – 922, 2011.
- [25] H. V. Nguyen and F. Porikli. Support vector shape: A classifier-based shape representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(4):970–982, April 2013.
- [26] S. M. A. Eslami, N. Heess, C. K. I. Williams, and J. Winn. The shape boltzmann machine: A strong model of object shape. *International Journal of Computer Vision*, 107(2):155–176, 2014.
- [27] F. Chen, H. Yu, R. Hu, and X. Zeng. Deep learning shape priors for object segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1870–1877, June 2013.
- [28] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I. C. Chang. Deep learning of feature representation with multiple instance learning for medical image analysis. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 1626–1630, May 2014.
- [29] M. Ramesh, F. Mesadi, M. Cetin, and T. Tasdizen. Disjunctive normal shape model. In *ISBI, IEEE International Symposium on Biomedical Imaging*, 2015.
- [30] M. U. Ghani, F. Mesadi, S. Demir Kank, A. . Argunah, I. Israely, D. nay, T. Tadizen, and M. etin. Dendritic spine shape analysis using disjunctive normal shape models. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 347–350, April 2016.
- [31] M. U. Ghani, F. Mesadi, S. Demir Kank, A. . Argunah, A. Hobbiss, I. Israely, D. nay, T. Tadizen, and M. etin. Dendritic spine classification using shape and appearance features based on two-photon microscopy. *Journal of Neuroscience Methods*, 279:13 – 21, 2017.

- [32] F. Mesadi, M. Cetin, and T. Tasdizen. Disjunctive normal shape and appearance priors with applications to image segmentation. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pages 703–710. Springer International Publishing, 2015.
- [33] F. Mesadi, M. Cetin, and T. Tasdizen. Disjunctive normal level set: An efficient parametric implicit method. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4299–4303, Sept 2016.
- [34] F. Mesadi, M. Cetin, and T. Tasdizen. Disjunctive normal parametric level set with application to image segmentation. *IEEE Transactions on Image Processing*, 26(6):2618–2631, June 2017.
- [35] M. Hazewinkel. *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia*. Number v. 1 in *Encyclopaedia of Mathematics*. Springer, 1997.
- [36] J.C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.
- [37] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893 vol. 1, June 2005.
- [38] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Comput.*, 1(4):541–551, December 1989.
- [39] D. Cremers and M. Rousson. *Deformable Models: Theory and Biomedical Applications*, chapter Efficient Kernel Density Estimation Of Shape And Intensity Priors For Level Set Segmentation, pages 447–460. Springer New York, New York, NY, 2007.
- [40] A. Soglan, M. G. Uzunbas, and M. Cetin. Combining learning-based intensity distributions with nonparametric shape priors for image segmentation. *Signal, Image and Video Processing*, 8(4):789–798, 2014.
- [41] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076, 09 1962.
- [42] R.M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, May 1979.
- [43] T.F. Chan and L.A. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266–277, Feb 2001.
- [44] E. Erdil, A. O. Argunsah, T. Tasdizen, D. Unay, and M. Cetin. A joint classification and segmentation approach for dendritic spine segmentation in 2-photon microscopy images. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 797–800, April 2015.
- [45] B. Menze, M. Reyes, and K. Van Leemput. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, page 33, 2014.
- [46] NCI-ISBI Challenge. Automated segmentation of prostate structures, 2013.
- [47] MICCAI Grand Challenge. Prostate mr image segmentation, 2012.
- [48] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. Jodoin, and H. Larochelle. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35:18 – 31, 2017.
- [49] D. Zikic et al. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, volume 7512 of *Lecture Notes in Computer Science*, pages 369–376. Springer Berlin Heidelberg, 2012.
- [50] S. Bauer, T. Fejes, J. Slotboom, R. Wiest, L. P. Nolte, and M. Reyes. Segmentation of brain tumor images based on integrated hierarchical classification and regularization. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, volume 7512 of *Lecture Notes in Computer Science*, pages 10–13. Springer Berlin Heidelberg, 2012.
- [51] E. Geremia, B. H. Menze, and N. Ayache. Spatial decision forests for glioma segmentation in multi-channel mr images. In *MICCAI 2012*, volume 7512 of *Lecture Notes in Computer Science*, pages 14–18. Springer Berlin Heidelberg, 2012.
- [52] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide: The Insight Segmentation and Registration Toolkit (version 1.4)*. Kitware, 2003.