DTI Segmentation Using an Information Theoretic Tensor Dissimilarity Measure

Zhizhou Wang, *Member, IEEE,* and Baba C. Vemuri*, *Fellow, IEEE*

Abstract—In recent years, diffusion tensor imaging (DTI) has become a popular *in vivo* diagnostic imaging technique in Radiological sciences. In order for this imaging technique to be more effective, proper image analysis techniques suited for analyzing these high dimensional data need to be developed.

In this paper, we present a novel definition of tensor "distance" grounded in concepts from information theory and incorporate it in the segmentation of DTI. In a DTI, the symmetric positive definite (SPD) diffusion tensor at each voxel can be interpreted as the covariance matrix of a local Gaussian distribution. Thus, a natural measure of dissimilarity between SPD tensors would be the Kullback-Leibler (KL) divergence or its relative. We propose the square root of the J-divergence (symmetrized KL) between two Gaussian distributions corresponding to the diffusion tensors being compared and this leads to a novel closed form expression for the "distance" as well as the mean value of a DTI. Unlike the traditional Frobenius norm-based tensor distance, our "distance" is affine invariant, a desirable property in segmentation and many other applications. We then incorporate this new tensor "distance" in a region based active contour model for DTI segmentation. Synthetic and real data experiments are shown to depict the performance of the proposed model.

Index Terms—Diffusion tensor MRI, image segmentation, Kullback-Leibler divergence, J-divergence, Mumford-Shah functional, active contour.

I. INTRODUCTION

IN [8], Basser *et al.* presented their seminal work on diffusion tensor magnetic resonance imaging (DT-MRI). Since this new MRI modality can be used to quantify anisotropic properties of the imaged tissue by characterizing the water diffusion through the same, it has became a powerful technique to investigate the tissue microstructure *in vivo*, e.g., white matter connectivity in the brain or in the spinal cord. Investigating the tissue microstructure and changes therein (due to pathology) is achieved via DTI analysis. This in general includes a broad spectrum of interesting problems namely, optimal DT-MR imaging protocol design, DTI restoration, DTI segmentation, DTI registration, DTI visualization and fiber track mapping.

In DT-MRI, what is measured is the diffusion weighted echo intensity image (DWI) S_l. It can be related to the diffusion tensor $\mathbf{D}$ through the Stejskal-Tanner equation [8] given by

$$S_l = S_0 e^{-\mathbf{b}_l : \mathbf{D}} = S_0 e^{-\sum_{i=1}^{3} \sum_{j=1}^{3} b_{l.ij} D_{ij}} \tag{1}$$

where $\mathbf{b}_l$ is the diffusion weighting of the l-th diffusion encoding magnetic gradient, ":" denotes the generalized inner product for matrices. Given several noncollinear diffusion weighted intensity measurements, $\mathbf{D}$ can be estimated via multivariate regression techniques.

One key factor in DTI analysis is a proper choice of diffusion tensor distance that measures the similarity or dissimilarity between the tensors and is particularly important in the aforementioned tasks. In the following, we will present a brief overview of different tensor distance measures used in DTI analysis and various techniques currently in vogue in using tensor-based information for segmenting DTI.

In general, any kind of matrix norm can be used to measure the distance between two 2-tensors. One such example is the tensor Euclidean distance obtained by using the *Frobenius norm*. Due to its simplicity, tensor Euclidean distance has been used extensively in DTI restoration [11], [12], [32]. Alexander *et al.* [2] considered a number of tensor similarity measures for matching diffusion tensor images and concluded empirically that the Euclidean difference measure yields the best results. Though not many sophisticated tensor distances have been proposed in DTI analysis, there are quite a few in the context of machine learning. Miller *et al.* [22] proposed an interesting measure on transformation groups to design an invariant kernel for nonparametric density estimation. What is most closely related to our present work was proposed by Tsuda *et al.* [29]. They introduced information geometry in the space of positive definite matrices to derive a Kullback-Leibler divergence between them and then used it in an "em" (not the well-known expectation-maximization) algorithm to approximate an incomplete kernel.

In the context of DTI segmentation, in [33], Wiegell *et al.* applied the K-means clustering (involves tensor Frobenius norm) technique for segmentation of thalamic nuclei from DTI. Simultaneously, Zhukov *et al.* [35] proposed a level set segmentation method to segment scalar anisotropic images derived from the diffusion tensor. The fact that such scalar fields are used implies that they have ignored the direction information contained in the DTI. Thus, this method will fail if two homogeneous regions of DTI have the same anisotropy property but are oriented in a totally different direction.

Prior to our published work in [31], the only published work on the segmentation of matrix-valued images was reported in Feddern *et al.* [13]. In their work [13], Feddern *et al.* extended the geodesic active contours model for application to tensor fields, specifically to the structure tensor field computed from gray scale images. The stopping criterion in this case is chosen as a decreasing function of the trace of the sum of the structure tensors formed from individual elements of the given tensor in the tensor field. This amounts to taking the Frobenius norm of the tensors in the tensor field formed by the gradient magnitude of the individual channels/components of the given tensor. This scheme is a gradient based active contour(snake) whose performance is lacking in absence of significant gradient in the data. Moreover, the norm used here is not invariant to affine transformations of the input coordinates on which the original tensor field is defined. Affine invariance is a desirable property for the segmentation scheme when dealing with data sets obtained from different hardware.

In [31], we applied a *region-based active contour model* for tensor field segmentation by incorporating a tensor distance based on matrix Frobenius norm. Simultaneously, Rousson *et al.* [25] extended the classical surface evolution segmentation model by incorporating region statistics defined on the tensor fields for DTI segmentation. In both works [25], [31], different components in the diffusion tensor is treated in the same way. However, the diffusion tensor is actually the covariance matrix of a local diffusion process and different components have different importance/weight. The importance of this fact in DTI segmentation was first recognized by Wang and Vemuri in [30]. Specifically, they proposed an information theoretic "distance" for rank 2 tensors and incorporated it in a region-based active contour model to segment tensor fields. Soon after, Lenglet *et al.* [16], [17] extended the concepts presented in [30] for probability density field segmentation in a Bayesian framework with application to DTI segmentation.

In this paper, we tackle the DTI segmentation problem using a region-based active contour model by incorporating an information theoretic tensor dissimilarity measure. *This paper is a significant extension of our preliminary work reported in* [30]. Geometric active contour models have long been used in scalar and vector-valued image segmentation [9], [15], [18]–[20]. Our work is based on the active contour models derived from the Mumford-Shah functional [23], and can be viewed as a *significant extension* of the work on active contours without edges, by Chan and Vese [10] and the work on curve evolution implementation of the Mumford-Shah functional by Tsai *et al.* [28] to diffusion tensor images. Our key contributions are: 1) the definition of a new affine invariant discriminant of tensors based on information theory; 2) a theorem and its proof that allows for the computation of the mean value of an SPD tensor field—required in the piecewise constant version of the region-based active contour model—in closed form and facilitates the efficient segmentation of the DTI; 3) Derivation of an analytic expression for the derivatives of the energy function in the piecewise smooth region-based segmentation case; and 4) extension of the popular region-based active contour model in level-set form to handle matrix-valued images, e.g., DTI.

The remainder of this paper is organized as follows: in Section II, the new definition of tensor distance is introduced

and its properties are discussed in detail. Then, in Section III-A, the piecewise constant region-based active contour model for DTI segmentation is discussed. We present the associated Euler-Lagrange equation, the curve evolution equation, the level set formulation and the implementation details. In Section III-B, we present the piecewise smooth region-based active contour model for DTI segmentation. Section IV contains experiments on application of our model to synthetic tensor fields as well as real DTIs and Section V the conclusion.

II. A NEW TENSOR DISTANCE AND ITS PROPERTIES

We now present the definition and properties of our new "distance" measure between diffusion tensors.

A. Definitions

Given a diffusion tensor $\mathbf{T}$, the displacement $\mathbf{r}$ of water molecules starts from a given location at time t is a random variable with the following probability density:

$$p(\mathbf{r}\,|\,t, \mathbf{T}) = \frac{1}{\sqrt{(2\pi)^n |2t\mathbf{T}|}} e^{\frac{-\mathbf{r}^T \mathbf{T}^{-1} \mathbf{r}}{4t}} \qquad (2)$$

where n is the size of the square matrix $\mathbf{T}$. Note that the mean of $\mathbf{r}$, $E(\mathbf{r}) = \mathbf{0}$ and the covariance matrix of $\mathbf{r}$ is $\Lambda(\mathbf{r}) = 2t\mathbf{T}$ [26]. It follows that a natural distance between diffusion tensors can be derived from an information theoretical distance measure between Gaussian distributions. The first choice could be the Kullback-Leibler (KL) divergence given as

$$\mathrm{KL}(p\,\|\,q) = \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \qquad (3)$$

for a pair of probability density functions $p(\mathbf{x})$ and $q(\mathbf{x})$.

However, KL divergence is not symmetric and the most frequently used way to symmetrize it is the J-divergence given by

$$\mathrm{J}(p, q) = \frac{1}{2}(\mathrm{KL}(p\,\|\,q) + \mathrm{KL}(q\,\|\,p)) \qquad (4)$$

We propose a novel definition of diffusion tensor "distance" as the square root of the J-divergence of the corresponding Gaussian distributions, i.e.,

$$d(\mathbf{T}_1, \mathbf{T}_2) = \sqrt{\mathrm{J}(p(\mathbf{r}\,|\,t, \mathbf{T}_1), p(\mathbf{r}\,|\,t, \mathbf{T}_2))} \qquad (5)$$

For $p(\mathbf{r}\,|\,t, \mathbf{T}_1)$, and $p(\mathbf{r}\,|\,t, \mathbf{T}_2)$ given as in (2), it is known that [34]

$$\mathrm{J}(p(\mathbf{r}\,|\,t, \mathbf{T}_1), p(\mathbf{r}\,|\,t, \mathbf{T}_2)) = \frac{1}{4} \left[\mathrm{tr}\left(\mathbf{T}_1^{-1}\mathbf{T}_2 + \mathbf{T}_2^{-1}\mathbf{T}_1 \right) - 2n \right]$$

where $\mathrm{tr}(\,\cdot\,)$ is the matrix trace operator.

Thus, (5) has a very nice closed form given by

$$d(\mathbf{T}_1, \mathbf{T}_2) = \frac{1}{2}\sqrt{\mathrm{tr}\left(\mathbf{T}_1^{-1}\mathbf{T}_2 + \mathbf{T}_2^{-1}\mathbf{T}_1 \right) - 2n}. \qquad (6)$$

We use the quotation marks on the distance here because, what we have defined in (6) is not a true distance since it violates the triangle inequality. We could use the Rao's distance [6] between the Gaussian distributions $p(\mathbf{r}\,|\,t, \mathbf{T}_1)$ and $p(\mathbf{r}\,|\,t, \mathbf{T}_2)$ to define a true distance between the tensors $\mathbf{T}_1$ and $\mathbf{T}_2$. However,

Rao's distance between tensors will pose a computational difficulty for DTI segmentation in that it does not yield a closed form expression for the tensor field mean value and this will cause the computational cost to increase steeply. Moreover, there is a close relationship between our tensor "distance" and the aforementioned Rao's distance between infinitesimally close tensors. We will now proceed to derive this relationship. Following the notation given in Lenglet *et al.* [17], let $(\mathcal{M}, \mathcal{G})$ be a Riemannian Manifold, where $\mathcal{M}$ consists of a family of probability density functions parameterized by $\boldsymbol{\theta}$, $\mathcal{G} = (g_{ij})$ is the Fisher or the information metric [4]. Now, for two nearby distributions $p(\mathbf{r}, \boldsymbol{\theta})$, and $p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})$ on $\mathcal{M}$, the squared geodesic distance between them can be approximated by

$$ds^2(p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) = \sum_{i,j} g_{ij}(\boldsymbol{\theta}) d\theta_i d\theta_j. \quad (7)$$

similarly we have

$$\begin{aligned} ds^2(&p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta})) \\ &= \sum_{i,j} g_{ij}(\boldsymbol{\theta} + d\boldsymbol{\theta})(-d\theta_i)(-d\theta_j) \\ &= \sum_{i,j} (g_{ij}(\boldsymbol{\theta}) + o(\|d\boldsymbol{\theta}\|)) d\theta_i d\theta_j \\ &= \left(\sum_{i,j} g_{ij}(\boldsymbol{\theta}) d\theta_i d\theta_j \right) + o(\|d\boldsymbol{\theta}\|^3) \\ &= ds^2(p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) + o(\|d\boldsymbol{\theta}\|^3). \quad (8) \end{aligned}$$

Using the Taylor expansion of the KL-divergence between $p(\mathbf{r}, \boldsymbol{\theta})$ and $p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})$ as in [17], we have

$$\begin{aligned} \text{KL}(&p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) \\ &= \frac{1}{2} ds^2(p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) + o(\|d\boldsymbol{\theta}\|^2). \quad (9) \end{aligned}$$

Similarly

$$\begin{aligned} \text{KL}(&p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta})) \\ &= \frac{1}{2} ds^2(p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta})) + o(\|d\boldsymbol{\theta}\|^2) \\ &= \frac{1}{2}[ds^2(p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) + o(\|d\boldsymbol{\theta}\|^3)] \\ &\quad + o(\|d\boldsymbol{\theta}\|^2) \\ &= \frac{1}{2} ds^2(p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) + o(\|d\boldsymbol{\theta}\|^2). \quad (10) \end{aligned}$$

Thus we have

$$\begin{aligned} J(&p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) \\ &= \frac{1}{2}[\text{KL}(p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) \\ &\quad + \text{KL}(p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta}))] \\ &= \frac{1}{2} ds^2(p(\mathbf{r}, \boldsymbol{\theta}), p(\mathbf{r}, \boldsymbol{\theta} + d\boldsymbol{\theta})) + o(\|d\boldsymbol{\theta}\|^2). \quad (11) \end{aligned}$$

So, the J-divergence between two nearby distributions approximates half of the squared geodesic distance between them. This holds true for the special case of Gaussian distributions and thus our "distance" approximates the Rao's distance between two infinitesimally close tensors (up to a constant scaling factor)

and it is also computationally efficient for the purpose of DTI segmentation.

B. Affine Invariant Property

When an affine transformation is applied to the coordinate system on which the DTI is defined, the diffusion tensors will also be transformed. For a transformation given by $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$, the transformation of the displacement $\mathbf{r}$ of a water molecule is given by $\hat{\mathbf{r}} = \mathbf{Ar}$. Since the the distribution of $\mathbf{r}$ is a Gaussian with zero mean and covariance matrix $2t\mathbf{T}$, the transformed displacement $\hat{\mathbf{r}}$ will have a Gaussian distribution with zero mean and covariance matrix given by, $2t\mathbf{ATA}^T$. It follows that the tensor field is transformed as: $\hat{\mathbf{T}}(\mathbf{y}) = \mathbf{AT}(\mathbf{x})\mathbf{A}^T$.

The tensor "distance" we defined earlier is invariant to affine transformations, i.e., $d(\mathbf{T_1}, \mathbf{T_2}) = d(\mathbf{AT_1A}^T, \mathbf{AT_2A}^T)$. The diffusion tensors actually undergo a congruent transformation however, this is simply an outcome of the affine transformation of the coordinate system on which they are defined. Thus, the above invariance is still referred to as "affine" invariance and its importance in DTI segmentation will be illustrated in Section III-A. It is apparent that the tensor Euclidean distance based on matrix Frobenius norm commonly used in published work [2], [11], [12], [32] is not invariant under scaling as well as affine transformations.

C. Mean Value of an SPD Tensor Field

We now develop a theorem that allows us to compute the mean value of an symmetric positive definite (SPD) tensor field. This is essential for the piece-wise constant region-based active contour model used in the DTI segmentation algorithm, where the constant is the mean value taken over the region.

Theorem 1: The mean value of an SPD tensor field $\mathbf{T}$ can be defined as

$$\bar{\mathbf{M}}(\mathbf{T}, R) = \min_{\mathbf{M} \in \text{SPD}(n)} \int_R d^2[\mathbf{M}, \mathbf{T}(\mathbf{x})] d\mathbf{x} \quad (12)$$

where $\text{SPD}(n)$ is the set of all symmetric positive definite matrices of size n. This mean value can be computed analytically as

$$\bar{\mathbf{M}} = \sqrt{\mathbf{B}^{-1}} \sqrt{\sqrt{\mathbf{B}} \mathbf{A} \sqrt{\mathbf{B}}} \sqrt{\mathbf{B}^{-1}} \quad (13)$$

where $\mathbf{A} = \int_R \mathbf{T}(\mathbf{x}) d\mathbf{x}$ and $\mathbf{B} \int_R \mathbf{T}^{-1}(\mathbf{x}) d\mathbf{x}$.

Proof: Let $E(\mathbf{M}) = \int_R d^2[\mathbf{M}, \mathbf{T}(\mathbf{x})] d\mathbf{x}$, we have

$$\begin{aligned} E(\mathbf{M}) &= \int_R d^2[\mathbf{M}, \mathbf{T}(\mathbf{x})] d\mathbf{x} \\ &= \int_R \left[\frac{1}{4} \text{tr}(\mathbf{M}^{-1} \mathbf{T}(\mathbf{x}) + \mathbf{T}^{-1}(\mathbf{x})\mathbf{M}) - \frac{n}{2} \right] d\mathbf{x} \\ &= \frac{1}{4} \text{tr} \left[\mathbf{M}^{-1} \left(\int_R \mathbf{T}(\mathbf{x}) d\mathbf{x} \right) \right. \\ &\quad \left. + \left(\int_R \mathbf{T}^{-1}(\mathbf{x}) d\mathbf{x} \right) \mathbf{M} \right] - \frac{n}{2}|R| \\ &= \frac{1}{4} \text{tr}(\mathbf{M}^{-1} \mathbf{A} + \mathbf{BM}) - \frac{n}{2}|R|. \end{aligned}$$

For a small perturbation in the neighborhood $N(M) \subset \text{SPD}(n)$ represented by $\mathbf{M} + t\mathbf{V}$, where t is a small enough positive

number, $\mathbf{V}$ is symmetric matrix, we have $(\mathbf{M} + t\mathbf{V})^{-1} = \mathbf{M}^{-1}(\mathbf{I} - t\mathbf{V}\mathbf{M}^{-1} + o(t))$, thus

$$
\begin{aligned}
E(\mathbf{M} + t\mathbf{V}) \\
&= \frac{1}{4}\mathrm{tr}[(\mathbf{M} + t\mathbf{V})^{-1}\mathbf{A} + \mathbf{B}(\mathbf{M} + t\mathbf{V})] - \frac{n}{2}|R| \\
&= \frac{1}{4}\mathrm{tr}[\mathbf{M}^{-1}(\mathbf{I} - t\mathbf{V}\mathbf{M}^{-1} + o(t))\mathbf{A} \\
&\quad + \mathbf{B}(\mathbf{M} + t\mathbf{V})] - \frac{n}{2}|R| \\
&= \frac{1}{4}\mathrm{tr}(\mathbf{M}^{-1}\mathbf{A} + \mathbf{B}\mathbf{M}) - \frac{n}{2}|R| \\
&\quad + \frac{1}{4}\mathrm{tr}(\mathbf{B}\mathbf{V} - \mathbf{M}^{-1}\mathbf{V}\mathbf{M}^{-1}\mathbf{A})t + o(t) \\
&= E(\mathbf{M}) + \frac{1}{4}\mathrm{tr}(\mathbf{B}\mathbf{V} - \mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{V})t + o(t).
\end{aligned}
$$

Thus at the critical point, we need

$$
\mathrm{tr}(\mathbf{B}\mathbf{V} - \bar{\mathbf{M}}^{-1}\mathbf{A}\bar{\mathbf{M}}^{-1}\mathbf{V}) = 0, \quad \forall \text{ symmetric } \mathbf{V}. \quad (14)
$$

This is actually equivalent to $\bar{\mathbf{M}}\mathbf{B}\bar{\mathbf{M}} = \mathbf{A}$ and can be solved in closed form [5] yielding the desired result in (13).

It is not hard to show that: 1) the inverse and the square root of an SPD matrix are SPD matrices; 2) if $\mathbf{M}_1$ and $\mathbf{M}_2$ are two SPD matrices, then $\mathbf{M}_1\mathbf{M}_2\mathbf{M}_1$ is an SPD matrix. Since $\mathbf{A}$ and $\mathbf{B}$ are both SPD matrices, we can show that the following hold: a) $\mathbf{B}^{-1}$ and $\sqrt{\mathbf{B}}$ are SPD matrices; b) $\sqrt{\mathbf{B}^{-1}}$ and $\sqrt{\mathbf{B}}\mathbf{A}\sqrt{\mathbf{B}}$ are SPD matrices; c) $\sqrt{\sqrt{\mathbf{B}}\mathbf{A}\sqrt{\mathbf{B}}}$ is an SPD matrix; and d) $\bar{\mathbf{M}} = \sqrt{\mathbf{B}^{-1}}\sqrt{\sqrt{\mathbf{B}}\mathbf{A}\sqrt{\mathbf{B}}}\sqrt{\mathbf{B}^{-1}}$ is an SPD matrix. As $\bar{\mathbf{M}}$ is also an SPD matrix, it is indeed a solution to the minimization equation (12). ∎

In general, $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$ and (13) cannot be simplified further, therefore, one needs to resort to matrix diagonalization methods for computation.

III. THE REGION-BASED DTI SEGMENTATION MODEL

Given a noisy DTI $\mathbf{T}_0$ on a domain $\Omega \subset \Re^2$, our model for DTI segmentation is posed as a minimization of the following variational principle based on the Mumford-Shah functional [23]

$$
\begin{aligned}
E(\mathbf{T}, \mathbf{C}) = &\alpha \int_\Omega d^2(\mathbf{T}(\mathbf{x}), \mathbf{T}_0(\mathbf{x}))d\mathbf{x} \\
&+ \int_{\Omega/\mathbf{C}} p(\mathbf{T})(\mathbf{x})d\mathbf{x} + \beta|\mathbf{C}| \quad (15)
\end{aligned}
$$

where the curve $\mathbf{C}$ is the boundary between segmented regions and $\mathbf{T}$ is a piecewise smooth approximation of $\mathbf{T}_0$ with discontinuities only along $\mathbf{C}$. The first term measures the difference between the approximation $\mathbf{T}$ and the original DTI $\mathbf{T}_0$, the second term measures the lack of smoothness of the field $\mathbf{T}$ using the Dirichlet integral [14] and the third term $|\mathbf{C}|$ is the arclength of the curve $\mathbf{C}$. α and β are control parameters that can be used to balance the importance of various terms in the functional minimization to yield the desired result.

We now proceed to give more details on the second term in the above variational principle. Let $\mathcal{S}$ be the Riemannian manifold of SPD matrices of size m equipped with a metric g induced

by the Rao's distance for SPD matrices, the dimension of $\mathcal{S}$ is $m(m+1)/2$—the total number of independent components in $m \times m$ SPD matrices. An SPD tensor field, e.g., a slice of a DTI, is a map from $\Re^2$ to $\mathcal{S}$, thus $T(\mathbf{x})$ is a point on $\mathcal{S}$ and can have a small neighborhood of on $\mathcal{S}$ with local coordinate $\mathbf{u} = (u_1, \ldots, u_{m(m+1)/2})$. We can now define

$$
p(\mathbf{T})(\mathbf{x}) = \sum_{1 \leq k \leq n} \sum_{1 \leq i,j \leq m(m+1)/2} g_{ij}(u)\frac{\partial u^i}{\partial x_k}\frac{\partial u^j}{\partial x_k} \quad (16)
$$

where n is the dimension of the domain Ω. Equation (15) can be easily modified to accommodate segmentation in $\Re^3$ where the curve $\mathbf{C}$ is replaced with a surface $\mathbf{S}$ and the implementation in three dimensions is similar to that in two dimensions.

A. The Piecewise Constant DTI Segmentation Model

As in [23], (15) can be simplified to separate piecewise constant regions when α goes to zero. In particular, in the same spirit as the active contour without edges by Chan and Vese [10] for scalar value image segmentation, we consider the following binary segmentation model for DTI

$$
\begin{aligned}
E(\mathbf{C}, \mathbf{T}_1, \mathbf{T}_2) = &\int_R d^2(\mathbf{T}_0(\mathbf{x}), \mathbf{T}_1)d\mathbf{x} \\
&+ \int_{R^c} d^2(\mathbf{T}_0(\mathbf{x}), \mathbf{T}_2)d\mathbf{x} + \beta|\mathbf{C}| \quad (17)
\end{aligned}
$$

where $\mathbf{T}_1$ is the mean value of DTI in the region R enclosed by the curve $\mathbf{C}$ and $\mathbf{T}_2$ is the mean value of the DTI for the rest of the domain denoted by R^c.

We incorporate our new tensor "distance" in the above active contour model to achieve DTI segmentation. It can be proved that our model yields affine invariant segmentation when $\beta = 0$. When β is small enough, this model will still exhibit close to affine invariant segmentation. Using a tensor distance based on the Frobenius norm in (17), however, will not yield an affine invariant segmentation. A detailed proof is beyond the scope of this paper, we will instead present an experiment (subsequently) to support this statement. For a discussion on affine invariant segmentation of scalar-valued images, see [7].

1) Curve Evolution Equation: In the variational principle (17), $\mathbf{T}_1$ and $\mathbf{T}_2$ are just variables and they can be solved directly as

$$
\mathbf{T}_1 = \bar{\mathbf{M}}(\mathbf{T}_0, R), \quad \mathbf{T}_2 = \bar{\mathbf{M}}(\mathbf{T}_0, R^c).
$$

The Euler Lagrange equation for the variational principle (17) with fixed $\mathbf{T}_1$ and $\mathbf{T}_2$ is

$$
[d^2(\mathbf{T}_0, \mathbf{T}_1) - d^2(\mathbf{T}_0, \mathbf{T}_2) + \beta k]\mathbf{N} = 0
$$

where k and $\mathbf{N}$ is the curvature and the outward normal of the curve $\mathbf{C}$, respectively. The above equations can be solved iteratively. At each iteration, $\mathbf{T}_1$ and $\mathbf{T}_2$ can be computed analytically according to (13), and we have the following equation for updating the segmentation boundary $\mathbf{C}$:

$$
\begin{aligned}
\frac{\partial \mathbf{C}}{\partial t} = &-\left[d^2(\mathbf{T}_0, \mathbf{T}_1) - d^2(\mathbf{T}_0, \mathbf{T}_2) + \beta k\right]\mathbf{N} \\
&\text{where } \mathbf{T}_1 = \bar{\mathbf{M}}(\mathbf{T}_0, R), \quad \mathbf{T}_2 = \bar{\mathbf{M}}(\mathbf{T}_0, R^c). \quad (18)
\end{aligned}
$$

Algorithm 1 Two Stage Piecewise Constant Segmentation of DTIs

1: Set the initial curve $\mathbf{C}_0$ and compute its signed distance function ϕ_0.
2: Compute $\mathbf{T}_1$ and $\mathbf{T}_2$ according to equation (13).
3: Update signed distance function ϕ according to equation (19).
4: Reinitialize ϕ using the updated zero level set.
5: Stop if the solution is achieved, else go to step 2.

---

2) Level Set Formulation: Let ϕ be the signed distance function of $\mathbf{C}$ and choose it to be negative inside $\mathbf{C}$ and positive outside $\mathbf{C}$, then the curve evolution equation (18) can be reformulated within a levelset framework as

$$\frac{\partial \phi}{\partial t} = \left[d^2(\mathbf{T}_0, \mathbf{T}_1) - d^2(\mathbf{T}_0, \mathbf{T}_2) + \beta \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} \right] |\nabla \phi|. \quad (19)$$

3) Implementation and Numerical Methods: We developed a modified version of the Chan and Vese [10] implementation of the piecewise constant region-based scalar image segmentation model. As in [10], we solve the system iteratively where each iteration involves a two stage implementation. In the first stage, the embedding function ϕ is evolved according to (19) for a fixed $\mathbf{T}_1$ and $\mathbf{T}_2$. In the second stage, the mean values $\mathbf{T}_1$ and $\mathbf{T}_2$ are computed for a fixed ϕ. Summarizing, we have the procedure given in Algorithm 1.

The key to the computation of $\mathbf{T}_1$ and $\mathbf{T}_2$ is the computation of the square root of an SPD matrix. We use the matrix diagonalization to achieve this. A real symmetric matrix $\mathbf{A}$ can be diagonalized yielding $\mathbf{A} = \mathbf{O}\mathbf{T}\mathbf{O}^T$, where $\mathbf{O}$ is an orthogonal matrix and $\mathbf{T} = \text{diag}\{t_1, t_2, \ldots, t_n\}$ is a diagonal matrix. Then the polynomial form of $\mathbf{A}$ is $\mathbf{A}^\alpha = \mathbf{O}\mathbf{T}^\alpha\mathbf{O}^T$, where $\mathbf{T}^\alpha = \text{diag}\{t_1^\alpha, t_2^\alpha, \ldots, t_n^\alpha\}$. Note that in (13), $\sqrt{\sqrt{\mathbf{B}}\mathbf{A}\sqrt{\mathbf{B}}} \neq \mathbf{B}^{(1/4)}\mathbf{A}^{(1/2)}\mathbf{B}^{(1/4)}$, instead it has to be computed as in Algorith 2.

Equation (19) can be solved using a simple explicit Euler scheme. We can assume the spatial grid size to be 1, then using finite differences on the partial derivatives leads to

$$\Delta^i \phi_{i,j} = \frac{1}{2}(\phi_{i+1,j} - \phi_{i-1,j})$$

$$\Delta^j \phi_{i,j} = \frac{1}{2}(\phi_{i,j+1} - \phi_{i,j-1})$$

$$\Delta^{ii} \phi_{i,j} = \phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j}$$

$$\Delta^{jj} \phi_{i,j} = \phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1}$$

$$\Delta^{ij} \phi_{i,j} = \frac{1}{4}(\phi_{i+1,j+1} - \phi_{i+1,j-1} - \phi_{i-1,j+1} + \phi_{i-1,j-1}).$$

In this case, we have the following update equation at iteration $n+1$:

$$\frac{\phi_{i,j}^{n+1} - \phi_{i,j}^n}{\Delta t} = \left[d^2\left(\mathbf{T}_{0:i,j}, \mathbf{T}_1^n\right) - d^2\left(\mathbf{T}_{0:i,j}, \mathbf{T}_2^n\right) + \beta k_{i,j}^n \right]$$
$$* \sqrt{(\Delta^i \phi_{i,j})^2 + (\Delta^j \phi_{i,j})^2} \quad (20)$$

where the curvature $k_{i,j}^n$ of ϕ^n can be computed as

$$k_{i,j}^n =$$
$$\frac{\Delta^{jj}\phi_{i,j}^n\left(\Delta^i \phi_{i,j}^n\right)^2 - 2\Delta^i \phi_{i,j}^n \Delta^j \phi_{i,j}^n \Delta^{ij}\phi_{i,j}^n + \Delta^{ii}\phi_{i,j}^n\left(\Delta^j \phi_{i,j}^n\right)^2}{\left[\left(\Delta^i \phi_{i,j}^n\right)^2 + \left(\Delta^j \phi_{i,j}^n\right)^2\right]^{3/2}}.$$

Algorithm 2 Computing $\sqrt{\sqrt{\mathbf{B}}\mathbf{A}\sqrt{\mathbf{B}}}$

1: Diagonalize $\mathbf{B}$ as $\mathbf{O_B}\mathbf{T_B}\mathbf{O_B}^T$.
2: Compute $\mathbf{P} = \sqrt{\mathbf{B}}$ as $\mathbf{O_B}\sqrt{\mathbf{T_B}}\mathbf{O_B}^T$.
3: Compute $\mathbf{Q} = \mathbf{PAP}$.
4: Diagonalize $\mathbf{Q}$ using $\mathbf{O_Q}\mathbf{T_Q}\mathbf{O_Q}^T$
5: Compute $\sqrt{\mathbf{Q}} = \mathbf{O_Q}\sqrt{\mathbf{T_Q}}\mathbf{O_Q}^T$

The time complexity to update the signed distance function ϕ according to (20) on the whole domain Ω is $O(|\Omega|)$, which leads to poor performance when the the domain is large. However, the efficiency can be dramatically improved by using the narrow band method [1], [21] around the zero level set as we are only interested in the evolving zero level set. In addition, ϕ has to remain as a signed distance function and this can be done by reinitializing ϕ within a narrow band of the zero level set [27]. Finally, one can apply other numerical methods like the multi-grid scheme [28] to boost the speed. The solution is achieved when the change of the signed distance function in the narrow band is below certain threshold. Our current algorithm with explicit Euler scheme in conjunction with the narrow band method yields segmentation results within 3–5 s for the two-dimensional (2-D) synthetic data examples and within 2–10 min for the three-dimensional (3-D) real DTI examples on a 1-GHz Pentium-3 CPU, which is reasonably fast considering the huge amount of tensor field data processed.

B. The Piecewise Smooth DTI Segmentation Model

For certain regions in the DTI data sets, the piecewise constant assumption does not hold. In such cases, one has to employ the full power of the piecewise smooth model (15). In particular, we design a two-stage scheme following the curve evolution implementation of the Mumford-Shah functional proposed by Tsai *et al.* [28]. We start with a smoothing stage by fixing the curve and smooth within each disconnected regions while preserving the discontinuity across the curve that separate these regions. Then we freeze the piecewise smooth DTI computed previously and allow the curve to move in accordance with the governing evolution equation described in the following.

1) Piecewise Smooth DTI Approximation: When the curve $\mathbf{C}$ is not allowed to evolve, (15) is reduced to the following energy functional:

$$E_\mathbf{C}(\mathbf{T})$$
$$= \alpha \int_\Omega d^2(\mathbf{T}(\mathbf{x}), \mathbf{T}_0(\mathbf{x})) \, d\mathbf{x} + \int_{\Omega/\mathbf{C}} p(\mathbf{T})(\mathbf{x}) \, d\mathbf{x}. \quad (21)$$

As the Rao's distance $d_R(\cdot, \cdot)$ between tensors at two nearby locations is given by

$$d_R^2(T(\mathbf{x}+h\mathbf{dx}_k), T(\mathbf{x})) = \sum_{1 \le i,j \le m(m+1)/2} g_{ij}(u) \frac{\partial u^i}{\partial x_k} \frac{\partial u^j}{\partial x_k} h^2.$$

Equation (16) leads to

$$p(\mathbf{T})(\mathbf{x}) = \frac{1}{h^2} \sum_k d_R^2(T(\mathbf{x} + h\mathbf{dx}_k), T(\mathbf{x})). \quad (22)$$

The tensor "distance" we defined approximates Rao's distance (up to a constant factor) between infinitesimally close tensors and is computationally very attractive, hence, we use it in-

stead of the Rao's distance in the above equation. We can assume $h = 1$ and discretize equation (21) as

$$E_{\mathbf{C}}(\mathbf{T}) = \alpha \sum_{\mathbf{x}} d^2(\mathbf{T}(\mathbf{x}), \mathbf{T}_0(\mathbf{x}))$$

$$+ \sum_{(\mathbf{x},\mathbf{y}) \in N_C} d^2(\mathbf{T}(\mathbf{x}), \mathbf{T}(\mathbf{y})) \quad (23)$$

where N_C is a collection of neighboring pixels $(\mathbf{x}, \mathbf{y})$ that will not cut across the boundary.

We now have a function whose variables are all the independent components of the discretized DTI and we can therefore compute the gradient of this function with respect to the discrete DTI. Though this can be done directly using simple multivariate calculus by treating all the independent components of the DTI as components of a huge vector, the form of the gradient is tedious and is difficult to understand. In order to get a compact formulation, we resort to the derivative of a matrix function $f(\mathbf{A})$ defined as $\partial f(\mathbf{A})/\partial \mathbf{A} = [\partial f(\mathbf{A})/\partial \mathbf{A}(i,j)]$, where

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}(i,j)} = \lim_{dt \to 0} \frac{f(\mathbf{A} + dt\mathbf{E}_{ij}) - f(\mathbf{A})}{dt}. \quad (24)$$

In (24), matrix $\mathbf{E}_{ij}$ has value 1 at location (i, j) and 0 elsewhere.

Given a perturbation $\mathbf{V}$ on $\mathbf{T}(\mathbf{x})$, the variation of $E_{\mathbf{C}}(\mathbf{T})$ can be computed as follows:

$$E_{\mathbf{C}}(\mathbf{T}(\mathbf{x}) + \mathbf{V}) - E_{\mathbf{C}}(\mathbf{T}(\mathbf{x}))$$
$$= \alpha d^2(\mathbf{T}(\mathbf{x}) + \mathbf{V}, \mathbf{T}_0(\mathbf{x})) + \sum_{\mathbf{y} \in N_C(\mathbf{x})} d^2(\mathbf{T}(\mathbf{x}) + \mathbf{V}, \mathbf{T}(\mathbf{y}))$$
$$- \alpha d^2(\mathbf{T}(\mathbf{x}), \mathbf{T}_0(\mathbf{x})) - \sum_{\mathbf{y} \in N_C(\mathbf{x})} d^2(\mathbf{T}(\mathbf{x}), \mathbf{T}(\mathbf{y}))$$
$$= \frac{1}{4}\alpha \mathrm{tr}[(\mathbf{T}_0^{-1}(\mathbf{x}) - \mathbf{T}^{-1}(\mathbf{x})\mathbf{T}_0(\mathbf{x})\mathbf{T}^{-1}(\mathbf{x}))\mathbf{V}]$$
$$+ \frac{1}{4} \sum_{\mathbf{y} \in N_C(\mathbf{x})} \mathrm{tr}[(\mathbf{T}^{-1}(\mathbf{y}) - \mathbf{T}^{-1}(\mathbf{x})\mathbf{T}(\mathbf{y})\mathbf{T}^{-1}(\mathbf{x}))\mathbf{V}]$$
$$= \frac{1}{4}\mathrm{tr}[(\mathbf{B}_{\mathbf{T}}(\mathbf{x}) - \mathbf{T}^{-1}(\mathbf{x})\mathbf{A}_{\mathbf{T}}(\mathbf{x})\mathbf{T}^{-1}(\mathbf{x}))\mathbf{V}] \quad (25)$$

where $N_C(\mathbf{x})$ uses a 4-neighborhood system in 2-D and 8-neighborhood system in 3-D, and

$$\mathbf{A}_{\mathbf{T}}(\mathbf{x}) = \sum_{\mathbf{y} \in N_C(\mathbf{x})} \mathbf{T}(\mathbf{y}) + \alpha\mathbf{T}_0(\mathbf{x})$$
$$\mathbf{B}_{\mathbf{T}}(\mathbf{x}) = \sum_{\mathbf{y} \in N_C(\mathbf{x})} \mathbf{T}^{-1}(\mathbf{y}) + \alpha\mathbf{T}_0^{-1}(\mathbf{x}).$$

Let $\mathbf{V} = dt\mathbf{E}_{ij}$, we have from the above (25)

$$E_{\mathbf{C}}(\mathbf{T}(\mathbf{x}) + dt\mathbf{E}_{ij}) - E_{\mathbf{C}}(\mathbf{T}(\mathbf{x}))$$
$$= \frac{1}{4}\mathrm{tr}[dt(\mathbf{B}_{\mathbf{T}}(\mathbf{x}) - \mathbf{T}^{-1}(\mathbf{x})\mathbf{A}_{\mathbf{T}}(\mathbf{x})\mathbf{T}^{-1}(\mathbf{x}))\mathbf{E}_{ij}]$$
$$= \frac{1}{4}dt[\mathbf{B}_{\mathbf{T}}(\mathbf{x}) - \mathbf{T}^{-1}(\mathbf{x})\mathbf{A}_{\mathbf{T}}(\mathbf{x})\mathbf{T}^{-1}(\mathbf{x})](i,j)$$

then from (24), we have the gradient of $E_{\mathbf{C}}$ as

$$\frac{\partial E_C}{\partial \mathbf{T}(\mathbf{x})} = \frac{1}{4}[\mathbf{B}_{\mathbf{T}}(\mathbf{x}) - \mathbf{T}^{-1}(\mathbf{x})\mathbf{A}_{\mathbf{T}}(\mathbf{x})\mathbf{T}^{-1}(\mathbf{x})]. \quad (26)$$

So, a necessary condition for the minimizer of the discrete variational principle (23) is given by

$$\mathbf{B}_{\mathbf{T}}(\mathbf{x}) = \mathbf{T}^{-1}(\mathbf{x})\mathbf{A}_{\mathbf{T}}(\mathbf{x})\mathbf{T}^{-1}(\mathbf{x}). \quad (27)$$

Equation (27) is difficult to solve directly, since $\mathbf{A}_{\mathbf{T}}(\mathbf{x})$ and $\mathbf{B}_{\mathbf{T}}(\mathbf{x})$ depend on $\mathbf{T}$. Currently, we use the gradient descent method based on (26) to compute the piecewise smooth approximation $\mathbf{T}$ numerically.

2) Curve Evolution Equation and Level Set Formulation: Once a piecewise smoothed DTI is computed from previous step, we fix the the DTI and let the curve $\mathbf{C}$ evolve to minimize the following energy functional:

$$E_{\mathbf{T}}(\mathbf{C}) = \int_R [\alpha d^2(\mathbf{T}_R(\mathbf{x}), \mathbf{T}_0(\mathbf{x})) + p(\mathbf{T}_R)(\mathbf{x})]d\mathbf{x}$$
$$+ \int_{R^c} [\alpha d^2(\mathbf{T}_{R^c}(\mathbf{x}), \mathbf{T}_0(\mathbf{x})) + p(\mathbf{T}_{R^c})(\mathbf{x})]d\mathbf{x}$$
$$+ \beta|\mathbf{C}|. \quad (28)$$

The corresponding curve evolution equation for the above energy functional is given by

$$\frac{\partial \mathbf{C}}{\partial t} = \{[\alpha d^2(\mathbf{T}_{R^c}, \mathbf{T}_0) + p(\mathbf{T}_{R^c})]$$
$$- [\alpha d^2(\mathbf{T}_R, \mathbf{T}_0) + p(\mathbf{T}_R)] - \beta k\}\mathbf{N}. \quad (29)$$

Again to facilitate computation, we have

$$\frac{\partial \mathbf{C}}{\partial t} = \left[\alpha d^2(\mathbf{T}_{R^c}, \mathbf{T}_0) + \sum_{\mathbf{y} \in N_{R^c}(\mathbf{x})} d^2(\mathbf{T}_{R^c}, \mathbf{T}_{R^c}(\mathbf{y}))\right]\mathbf{N}$$
$$- \left[\alpha d^2(\mathbf{T}_R, \mathbf{T}_0) + \sum_{\mathbf{y} \in N_R(\mathbf{x})} d^2(\mathbf{T}_R, \mathbf{T}_R(\mathbf{y}))\right]\mathbf{N}$$
$$- \beta k\mathbf{N}. \quad (30)$$

The level set form of (30) is given by

$$\frac{\partial \phi}{\partial t} = \left[F + \beta\nabla \cdot \frac{\nabla\phi}{|\nabla\phi|}\right]|\nabla\phi| \quad (31)$$

where ϕ is the signed distance function of $\mathbf{C}$ and the data dependent speed F is given by

$$F = \left[\alpha d^2(\mathbf{T}_R, \mathbf{T}_0) + \sum_{\mathbf{y} \in N_R(\mathbf{x})} d^2(\mathbf{T}_R, \mathbf{T}_R(\mathbf{y}))\right]\mathbf{N}$$
$$- \left[\alpha d^2(\mathbf{T}_{R^c}, \mathbf{T}_0) + \sum_{\mathbf{y} \in N_{R^c}(\mathbf{x})} d^2(\mathbf{T}_{R^c}, \mathbf{T}_{R^c}(\mathbf{y}))\right]\mathbf{N}.$$

3) Implementation and Numerical Methods: Similar to the Algorithm 1 and also as in Tsai *et al.* [28], we have the two-stage implementation as in Algorithm 3. The major difference here lies in the discontinuity preserving smoothing. Currently, we use gradient descent (due to its simplicity) with adaptive step size directly based on (26) to solve $\mathbf{T}_R$ and $\mathbf{T}_{R^c}$. As we have a simple analytical form of the gradient, it is plausible to apply more sophisticated and efficient numerical techniques like the

Algorithm 3 Two Stage Piecewise Smooth Segmentation for DTI

1: Set initial curve $\mathbf{C}_0$ and compute its signed distance function as ϕ_0.
2: Compute $\mathbf{T}_R$ and $\mathbf{T}_{R^c}$ according to equation (26).
3: Update level set ϕ according to equation (31).
4: Reinitialize ϕ using the updated zero level set.
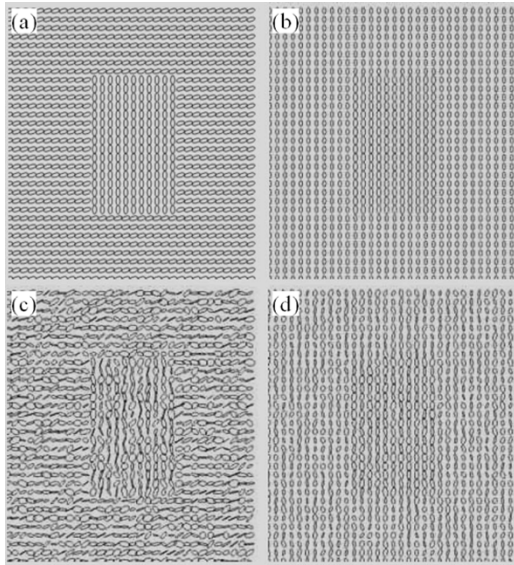5: Stop if the solution is achieved. Otherwise repeat from step 2.

Fig. 1. Synthetic tensor fields with two regions. (a) Two homogeneous regions differ only in the orientations, (b) two homogeneous regions differ only in scale, (c) is (a) with additive noise, and (d) is (b) with additive noise.
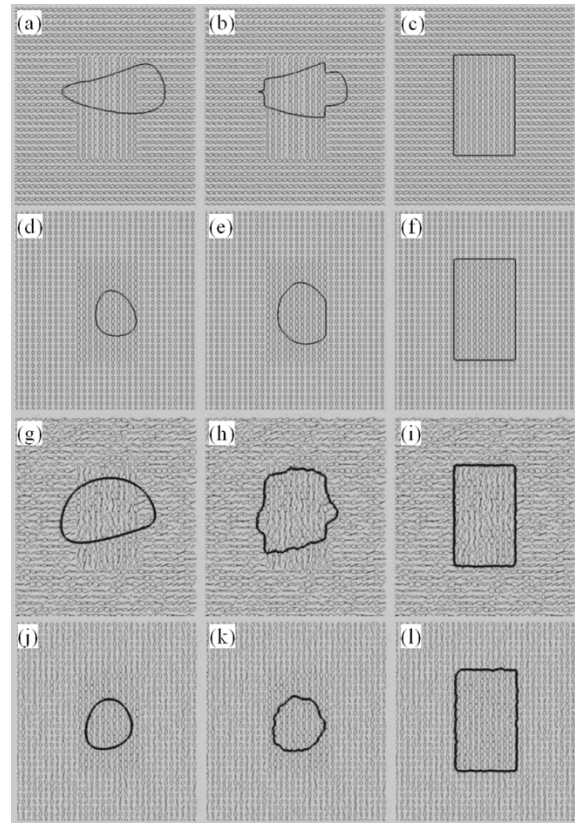

Fig. 2. Segmentation of synthetic tensor fields. Top to bottom: Segmentations of Fig. 1(a)–(d), respectively. Left to right are the initial, intermediate, and final steps of the segmentation procedure.

implicit Euler with preconditioned conjugate gradient and this will be the focus of our future research.

IV. EXPERIMENTAL RESULTS

In this section, we present application of our DTI segmentation algorithm to several datasets. The first one is the application to 2-D synthetic data sets, the second one depicts the affine invariant segmentation using our model, the third one is application to single slices of real DTIs estimated from DWIs and the last one is the application to 3-D real DTIs. In the experiments below, we use the piecewise constant model in (17) unless specified otherwise.

A. Synthetic Tensor Field Segmentation

The experiments with synthetic datasets are first used to demonstrate that full tensor information must be used in achieving quality segmentation for tensor fields. For this purpose, we create two synthetic tensor fields on a 128×128 lattice, each is a 2×2 SPD matrix valued image and has two homogeneous regions of different values. In the first tensor field, the two regions differ only in the orientations of the tensor field while the two regions differ only in scale in the second tensor field.

These two datasets are visualized as ellipses as shown in Fig. 1(a) and (b), respectively. Each ellipse is constructed by using eigenvectors of the corresponding tensor as axes, and then

scaled by the corresponding eigenvalues. As the two regions in the first tensor field have same scalar anisotropic properties of tensors, they cannot be separated using the method given in [35]. Similarly, the second tensor field cannot be segmented using only the dominant eigenvectors of the tensors. However, our proposed model can yield high quality segmentation results with an arbitrary initialization for both cases as shown in Fig. 2(a)–(f) where the evolving boundary is shown as black curves. It is evident that one has to use the full information contained in tensors for segmentation purpose as opposed to using either scalar maps or eigen vectors computed from the tensors. In addition, since our model is a region-based segmentation method, it is more resistant to noise than the gradient-based snakes. This follows from the well known fact that gradient-based snakes use local gradient information—which can be very noise sensitive—in segmenting the desired shapes. Fig. 2 third row and last row depict the segmentation process for synthetic noisy tensor fields shown in Fig. 1(c) and (d), and the results are as expected. In order to generate noisy tensor fields, we add Gaussian noise to the tensors at each location $\mathbf{x} = (x, y)$, i.e., the noisy tensor field is given by

$$\tilde{D}_{ij}(\mathbf{x}) = D_{ij}(\mathbf{x}) + n_{ij} \quad ij = 11, 12, 22,$$
$$\tilde{D}_{21}(\mathbf{x}) = \tilde{D}_{12}(\mathbf{x})$$

where n_{ij}s are independent Gaussian noises. As we need the noisy tensor field to be an SPD tensor field, we enforce this by setting $\tilde{D}_{21} = \tilde{D}_{12}$ and regenerating the tensor if it is not
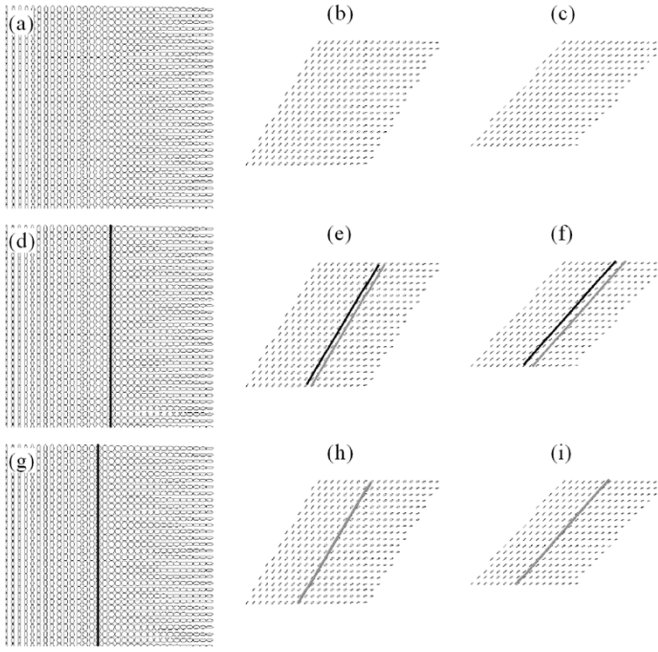
Fig. 3. Comparison of tensor field segmentation. Top to bottom: Original tensor field, result using tensor Euclidean distance and result using our new tensor "distance."
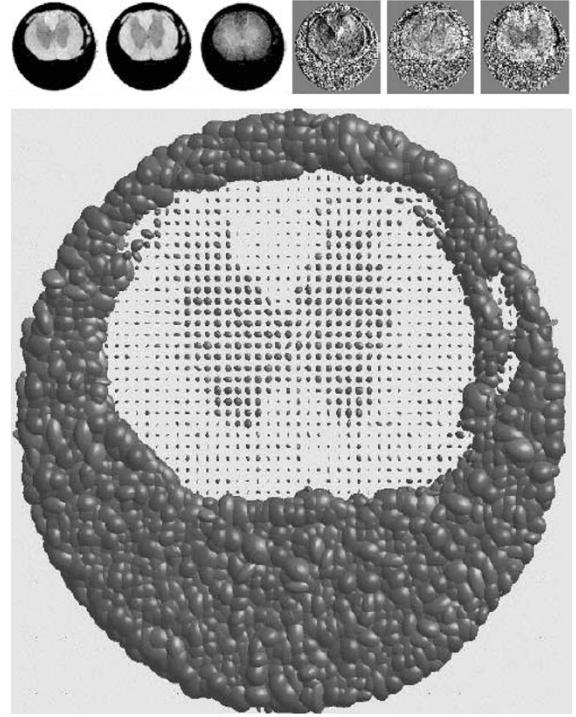

Fig. 4. A slice of the DTI of a normal rat spinal cord. Top row: viewed channel by channel as gray scale images. Bottom row: viewed using ellipsoids.

positive definite. This method of generation of the noisy tensor field is realistic as discussed in Pajevic *et al.* [24].

B. Affine Invariant Segmentation

When the curve smoothing term in the piecewise constant Mumford-Shah model using our tensor "distance" is zero, we have an affine invariant segmentation model. Using tensor distance based on Frobenius norm, however, does not have such a nice property. We now demonstrate an example in Fig. 3 to support this statement. Fig. 3(a) shows a tensor field of size 128×128, it contains two regions where the region on the left has vertical orientation while the one on the right has a horizontal orientation and there is a smooth transition between these two regions. This is an example where there is no "edge" between two regions and Mumford-Shah based segmentation models are particularly effective in such cases. Though there is no agreement on what the groundtruth is here, we could still assess the performance of segmentation models. This is done by checking whether a model can yield affine invariant segmentation. Let $\mathbf{C}$ be the segmentation of the original data, $\mathbf{A}(\mathbf{C})$ is the transformed segmentation and $\tilde{\mathbf{C}}$ be the segmentation of the transformed data. A model is affine invariant if $\mathbf{A}(\mathbf{C}) = \tilde{\mathbf{C}}$, when the original data undergoes an affine transformation.

Fig. 3(b) and (c) shows the transformed tensor fields obtained by applying two distinct affine transformations to the tensor field in Fig. 3(a). The segmentation results using a model based on tensor Euclidean (Frobenius norm based) distance in (17) are shown in Fig. 3(d)–(f). Fig. 3(d) shows the segmentation result of the original tensor field in (a) where the line in black separates the two regions. Fig. 3(e) shows the resulting boundary from segmenting the transformed tensor field in (b) in gray while the line in black is the transformed result from Fig. 3(d). Fig. 3(f) is

similar to Fig. 3(e) except that the transformation is a different one. We can see there is a big difference between the black line and the gray line in both Fig. 3(e) and Fig. 3(f) which shows that using tensor distance based on matrix Frobenius norm in piecewise constant Mumford-Shah is not affine invariant. On the contrary, our proposed method is affine invariant as shown in Fig. 3(g)–(i). The organization of Fig. 3(g)–(i) is the same as that of Fig. 3(d)–(f). We can see that the gray line coincides with the black line in (h) and (i) which means that the segmentation is affine invariant. We use $\beta = 0.0$ so that the curve smoothing term is negligible in all the above results.

C. Single Slice DTI Segmentation

Fig. 4 shows a slice of the DTI estimated from the DWIs of a normal rat spinal cord where the diffusion tensors in the white matter have similar orientations. Each of the six independent components of the individual symmetric positive definite diffusion tensors in the DTI is shown as a scalar image in the top row. The arrangements of the components from left to right are—$D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{yz}$, and D_{xz}. The off diagonal terms have been greatly enhanced by brightness and contrast factors for better visualization. An ellipsoid visualization of same slice as the top row is shown in the bottom row in Fig. 4. Each ellipsoid's axes correspond to the eigenvector directions of the 3×3 diffusion tensor and are scaled by the corresponding eigenvalues. For example, diffusion tensors in free water region are shown as large round ellipsoids (almost spherical). Fig. 5 depicts the segmentation process of the gray and white matter in a normal rat spinal cord with the evolving segmentation boundary curve in black superimposed on the ellipsoid visualization of the DTI.
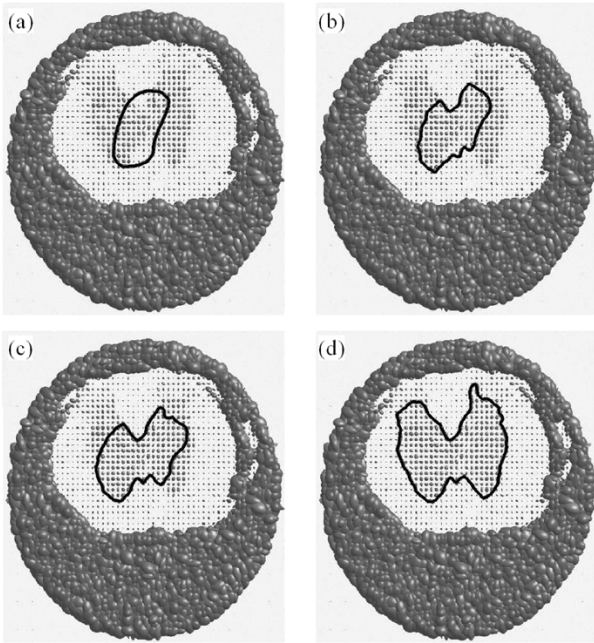
Fig. 5.   Segmentation of the slice of DTI shown in Fig. 4 (a)–(d) are initial, intermediate, and final steps in separating the gray and white matter inside the rat spinal cord.
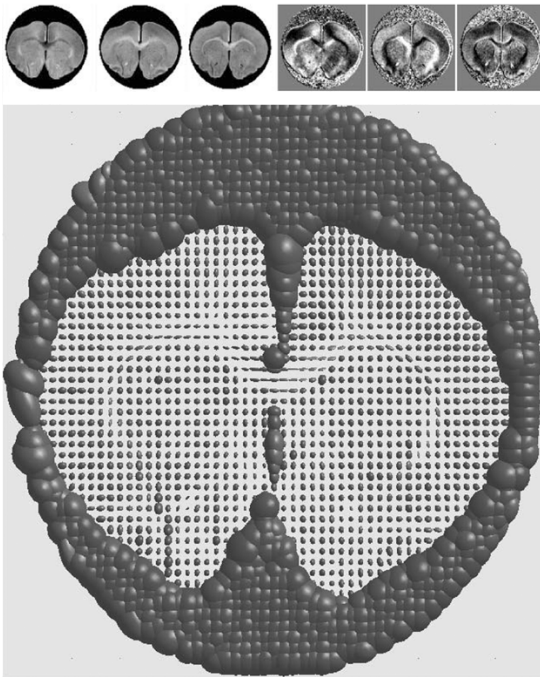

Fig. 6.   A slice of the DTI of a normal rat brain. Top row: viewed channel by channel as gray scale images. Bottom row: viewed using ellipsoids.

Similarly, Fig. 6 shows a slice of DTI of a normal rat brain and Fig. 7 demonstrates the segmentation of the corpus callosum using the piecewise constant DTI segmentation model. Though the majority of the corpus callosum is captured in the final step, the horns of the corpus callosum are not captured well. This problem is readily remedied by further applying the piecewise smooth DTI segmentation model [see (15)]. Starting from
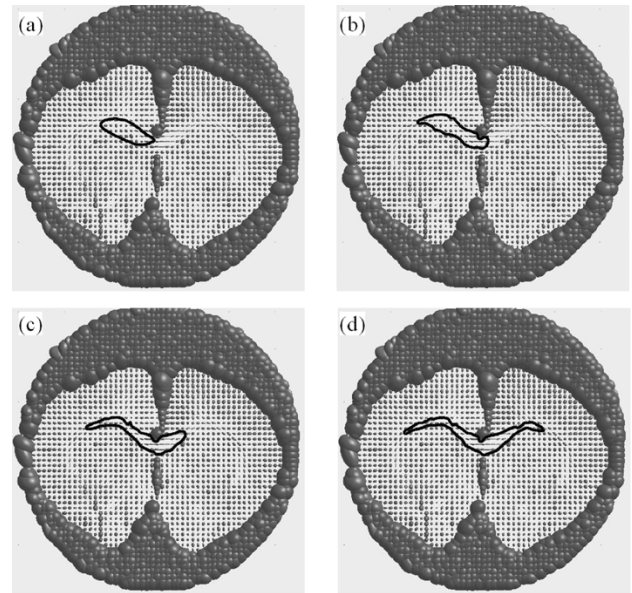

Fig. 7.   Segmentation of the corpus callosum from a slice of DTI shown in Fig. 6 (a)–(d) are initial, intermediate, and final steps of the segmentation procedure.
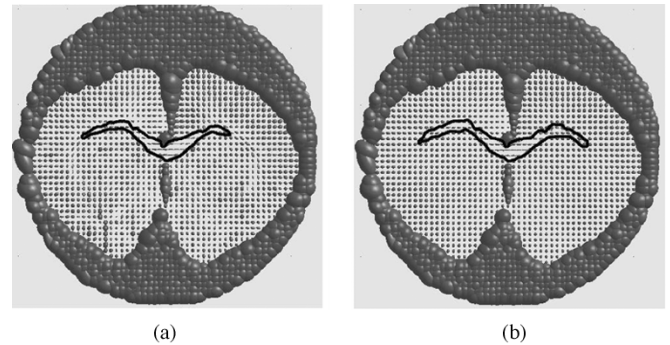

Fig. 8.   Segmentation of corpus callosum from a the slice of the DTI in Fig. 6 using piecewise smooth model. (a) Initial and (b) final steps of the segmentation procedure.

the segmentation result of the piecewise constant model, piecewise smooth model achieved a significant refinement as shown in Fig. 8.

D. DTI Segmentation in 3-D

Now we demonstrate segmentation results for two 3-D DTI data sets. Fig. 9 depicts the results of applying the segmentation algorithm to a normal rat spinal cord DTI of size $108 \times 108 \times 10$. Fig. 9(a)–(d) clearly depicts the surface evolution in 3-D and Fig. 9(e)–(h) depicts the intersection of the propagating surface in (a)–(d) with a slice of the D_{xx} component of the DTI. The separation of the gray and the white matter (where diffusion tensors have similar orientations) inside the normal rat spinal cord is achieved with ease.

Similarly, Fig. 10(a)–(h) depicts the segmentation procedure of the corpus callosum in a normal rat brain DTI of size $114 \times 108 \times 12$. The effectiveness of our algorithm is again evident, since the dominant part of the corpus callosum is well separated from the rest of the volume. Note that in all the real data experiments, we exclude the free water regions (using
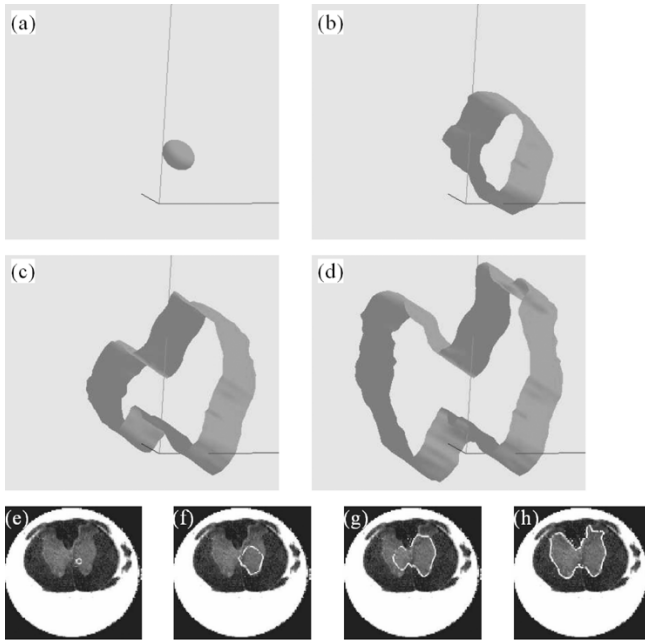
Fig. 9.   Segmentation of the 3-D DTI of a normal rat spinal cord. (a)–(d) Initial, intermediate and final steps in separating the gray and white matter inside the rat spinal cord. (e)–(h) A 2-D slice of the corresponding evolving surface in (a)–(d) superimposed on the D_{xx} component.
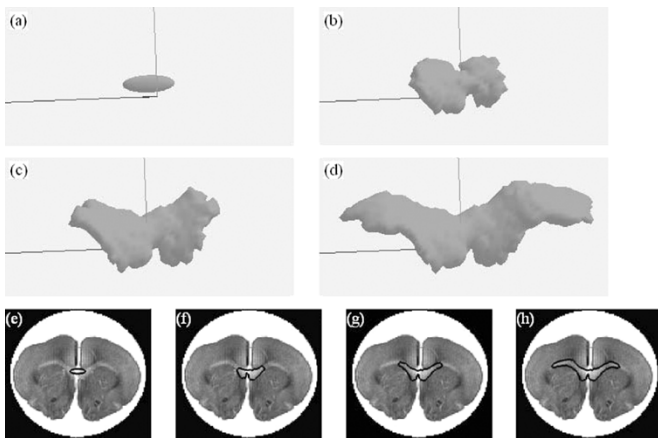

Fig. 10.   Segmentation of the corpus callosum from the 3-D DTI of a normal rat brain. (a)–(d): initial, intermediate and final steps of the segmentation procedure. (e)–(h): a 2-D slice of the corresponding evolving 3-D surface in (a)–(d) superimposed on the D_{xx} component.

simple thresholding on the apparent diffusion coefficient values and removing small disconnected regions) which are not of interest in the biological context.

V. CONCLUSION

In this paper, we presented a *novel DTI segmentation method* by incorporating an information theory-based tensor discriminant into the region based active contour models [10], [23], [28]. The particular tensor "distance" we employed naturally follows from the physical phenomena of diffusion interpreted using concepts of information theory. In addition, it has a simple form that makes our DTI segmentation model in both the piecewise constant and the piecewise smooth case computationally efficient. Specifically, *we provided a new theorem that allows for the com-*

putation of the mean of an SPD tensor field in closed form and also an analytical form of gradient for piecewise smooth tensor field approximation. Moreover, our tensor "distance" is affine invariant and can lead to affine invariant tensor field segmentation, a property that the tensor Euclidean distance based on matrix Frobenius norm—prevalent in current literature—does not possess.

We applied our segmentation model to real as well as synthetic data sets yielding promising results. There maybe situations where our segmentation model may be challenged and we envision the use of shape priors in such situations to assist our DTI segmentation model to achieve the desired results.

REFERENCES

[1] D. Adalsteinsson and J. A. Sethian, "A fast level set method for propagating interfaces," *J. Comput. Phys.*, vol. 118, no. 2, pp. 269–277, 1995.
[2] D. C. Alexander, J. C. Gee, and R. Bajcsy, "Similarity measures for matching diffusion tensor images," in *Proc. BMVC*, T. P. Pridmore and D. Elliman, Eds, 1999, pp. 93–102.
[3] S. Amari, "Information geometry on hierarchy of probability distributions," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 1701–1711, May 2001.
[4] S. Amari and H. Nagaoka, *Methods of Information Geometry*.   Oxford, U.K.: Oxford University Press, 2000.
[5] T. Ando, "Concavity of certain maps and positive definite matrices and applications to hadamard products," *Linear Alg. Applicat.*, no. 26, pp. 203–241, 1979.
[6] C. Atkinson and A. F. S. Mitchell, "Rao's distance measure," *Sankhya (The Indian J. Statist.)*, vol. 43, pp. 345–365, 1981.
[7] C. Ballester, V. Caselles, and M. Gonzalez, "Affine invariant segmentation by variational method," *SIAM J. Appl. Math.*, vol. 56, no. 1, pp. 294–325, 1996.
[8] P. J. Basser, J. Mattiello, and D. Lebihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *J. Magn. Reson.*, vol. 103, no. 3, pp. 247–254, 1994.
[9] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in *Proc. 5th Int. Conf. Comput. Vision*, 1995, pp. 694–699.
[10] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
[11] C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras, "Constrained flows of matrix-valued functions: Application to diffusion tensor regularization," in *Proc. 7th Eur. Conf. Comput. Vision*, vol. 2350, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., 2002, pp. 251–265.
[12] ——, "Regularizing flows for constrained matrix-valued images," *J. Math. Imag. Vis.*, vol. 20, no. 1–2, pp. 147–162, 2004.
[13] C. Feddern, J. Weickert, and B. Burgeth, "Level-set methods for tensor-valued images," in *Proc. 2nd IEEE Workshop Variational, Geometric and Level Set Methods in Computer Vision*, 2003, pp. 65–72.
[14] F. Hèlein, *Harmonic Maps, Conservation Laws and Moving Frames*, 2nd ed.   Cambridge, U.K.: Cambridge University Press, 2002.
[15] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Gradient flows and geometric active contour models," in *Proc. 5th Int. Conf. Comput. Vision*, 1995, pp. 810–815.
[16] C. Lenglet, M. Rousson, and R. Deriche, "Toward segmentation of 3d probability density fields by surface evolution: Application to diffusion MRI," in *Proc. 7th Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, 2004, Extended version as INRIA Reseach Report RR-5243 [Online]. Available: http://www-sop.inria.fr/rapports/sophia/RR-5243.html, pp. 18–25. .
[17] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. (2004) Statistics on Multivariate Normal Distributions: A Geometric Approach and Its Application to Diffusion Tensor MRI. INRIA, Paris, France. [Online]. Available: http://www-sop.inria.fr/rapports/sophia/RR-5242.html

[18] R. Malladi, J. A. Sethian, and B. C. Vemuri, "A topology independent shape modeling scheme," in *Proc. SPIE Conf. Geometric Methods in Computer Vision II*, vol. 2031, Jul. 1993, pp. 246–256.

[19] ——, "Evolutionary fronts for topology-independent shape modeling and recovery," in *Proc. 3rd Eur. Conf. Comput. Vision*, vol. 800, May 1994, pp. 3–13.

[20] ——, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 2, pp. 158–175, Feb. 1995.

[21] ——, "A fast level set based algorithm for topology-independent shape modeling," *J. Math. Imag. Vis.*, vol. 6, no. 2/3, pp. 269–289, 1996.

[22] E. G. Miller and C. Chefd'hotel, "Practical nonparametric density estimation on a transformation group for vision," in *IEEE Comp. Soc. Conf. Comput. Vision and Patt. Recog.*, vol. II, 2003, pp. 114–121.

[23] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational-problems," *Comm. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.

[24] S. Pajevic and P. J. Basser, "Parametric and nonparametric statistical analysis of DT-MRI data," *J. Magn. Reson.*, vol. 161, no. 1, pp. 1–14, 2003.

[25] M. Rousson, C. Lenglet, and R. Deriche, "Level set and region based surface propagation for diffusion tensor MRI segmentation," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer, 2004, vol. 3117, International Workshop Compute Vision Approaches to Medical Image Analysis (CVAMIA) and Mathematical Methods in Biomedical Image Analysis (MMBIA), pp. 123–134.

[26] G. A. F. Seber, *Linear Regression Analysis*. New York: Wiley, 1977.

[27] J. A. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge, U.K.: Cambridge University Press, 1996.

[28] A. Tsai, A. Yezzi, and A. S. Willsky, "Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification," *IEEE Trans. Image Processing*, vol. 10, no. 8, pp. 1169–1186, Aug. 2001.

[29] K. Tsuda, S. Akaho, and K. Asai, "The em algorithm for kernel matrix completion with auxiliary data," *J. Machine Learn. Res.*, vol. 4, pp. 67–81, 2003.

[30] Z. Wang and B. C. Vemuri, "An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation," in *IEEE Comp. Soc. Conf. Computation Vision and Pattern Recognition*, vol. I, 2004, pp. 228–233.

[31] ——, "Tensor field segmentation using region based active contour model," in *Lecture Notes in Computer Science*, T. Pajdla and J. Matas, Eds. Berlin, Germany, 2004, vol. 3024, Proc. 8th Eur. Conf. Comput. Vision (4), pp. 304–315.

[32] J. Weickert, "Diffusion and regularization methods for tensor-valued images," presented at the 1st SIAM-EMS Conf. Appl. Math. Our Changing World, Berlin, Germany, Sep. 2001.

[33] M. R. Wiegell, D. S. Tuch, H. W. B. Larson, and V. J. Wedeen, "Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging," *NeuroImage*, vol. 19, pp. 391–402, 2003.

[34] S. Yoshizawa and K. Tanabe, "Dual differential geometry associated with the Kullback-Leibler information on the Gaussian distributions and its 2-parameter deformations," *SUT J. Math.*, vol. 35, no. 1, pp. 113–137, 1999.

[35] L. Zhukov, K. Museth, D. Breen, R. Whitaker, and A. Barr, "Level set modeling and segmentation of DT-MRI brain data," *J. Electron. Imag.*, vol. 12, no. 1, pp. 125–133, 2003.