

Consistent Information Content Estimation for Diffusion Tensor MR Images

Brian G. Booth and Ghassan Hamarneh
Medical Image Analysis Lab
School of Computing Science, Simon Fraser University
Burnaby, BC, Canada
{bgb2, hamarneh}@sfu.ca

Abstract—We propose novel information content estimators for diffusion tensor images using binless approaches based on nearest-neighbour distances. Combining these estimators with existing tensor distance metrics allows us to generate entropy estimates that are consistent and accurate for diffusion tensor data. Further, we are able to obtain such estimators without having to reduce the dimensionality of the tensor data to the point where a binning estimator can be reliably used. We test our estimators in the context of noise estimation, image segmentation, and image registration. Results on 12 datasets from LBAM and 50 datasets from LONI show our estimators more accurately reflect the underlying DTI data and provide faster convergence rates for image segmentation and registration algorithms.

Keywords—diffusion tensor imaging; entropy; mutual information; noise estimation; image segmentation; image registration

I. INTRODUCTION

The advent of diffusion tensor imaging (DTI) has provided clinicians with the ability to assess the integrity of the brain’s neural pathways in a non-invasive manner. This unique ability has quickly allowed DTI to become an established imaging protocol generating images that, like other medical images, benefit from having computational methods for processing and analysis [1]. Yet unlike other medical image data, diffusion tensor images contain at each voxel a 3×3 symmetric positive definite matrix. As such, the computational analysis of these images is complicated both by the manifold-valued nature of the data as well as its high dimensionality. The result of these complications is that, frequently, analysis is performed only on a single dimension of the data (*e.g.*, fractional anisotropy). Working directly with the tensor data, on the other hand, typically requires adjusting existing medical image analysis algorithms.

Information theory in particular is a key component of computer vision and has seen uses in image compression, feature detection, segmentation, and registration (see [2] for a recent survey). Attempts have been made to apply information theory to diffusion tensor images in the contexts of image segmentation [3], [4], image registration [5], [6], [7], and feature detection [8], [9], [10]. The success of these applications, and other potentially similar work, relies on

being able to estimate entropy in a manner that is consistent¹ and accurate.

For scalar data, the most popular method for entropy estimation relies on histogram binning [11]. For this estimator, a normalized histogram is generated from a set of samples by discretizing the domain of the sample set. The resulting normalized histogram is then used as the probability distribution from which entropy can be estimated.

Binning estimators work well if the dimensionality of our samples is low (*e.g.*, $d = 1, 2$). However, for higher dimensional samples, the binning used for the estimator becomes an issue. In practice, we often have a fixed set of samples and as the size of the space increases, we require exponentially more bins to cover said space [11]. The combination of more bins and a fixed set of samples leads to biases and increasing mean-squared error as a result of undersampling [12]. Increased computational costs also result from higher bin counts. Bin sizes can be increased to overcome these limitations, but doing so introduces significant quantization effects that also increase the estimator’s mean-squared error [13].

The limitations of binning-based entropy estimators make them a poor choice for such tensor image data as the tensors make up a non-linear, convex half-cone in \mathbb{R}^6 [12]. As a result, the estimation of information content in diffusion tensor images has, to date, focused on reducing the dimensionality of the data to the point where a binning estimator can be used. In [8], [9], features of the tensors (*e.g.*, fractional anisotropy) are computed and used for entropy estimation. In [5], [6], [7], the information content of each tensor channel is computed separately, then averaged. In [6], all tensor channels are grouped together into a single one-dimensional histogram for entropy estimation. Reducing the dimensionality of the tensor data using such methods does not preserve uniqueness; multiple tensors can be mapped to the same point in the lower-dimensional space. As such, a set of non-homogeneous tensors could be reduced to same point in the lower-dimensional space, thereby generating an erroneous estimate of zero entropy. Information-theoretic measures have also been applied to individual tensors [3],

¹Consistent here refers to the formal mathematical notion that as the number of samples an estimator uses approaches infinity, the generated estimates converge to the exact value.

[4], [10], but these methods do not naturally extend to tensor sets.

We propose that the entropy of tensor-valued images should be computed using a different estimator that does not require reducing the data's dimensionality. In particular, we contribute an entropy estimator based on nearest-neighbor tensor distances. By using one of the established tensor distance metrics [3], [14], [15], [16], [17], we are able to obtain nearest neighbor distances that respect the underlying manifold of 3×3 symmetric positive definite matrices. Using these nearest neighbor distances for entropy estimation provides us with entropy estimates that are known to be consistent and accurate for high-dimensional data [12]. The contribution of this estimator also allows us to avoid the lossy dimensionality reductions used in existing binning-based entropy estimators for tensor-valued data. To our knowledge, this is the first work on computing information content of a set of unaltered tensor values.

In the following two sections, we present our nearest-neighbor entropy estimator for tensor-valued data and place it in the context of existing binning estimators. We further show how the same nearest-neighbor approach can be used to estimate mutual information, a similarity measure commonly used in image registration. We compare our estimator to existing binning estimators in Section IV. Specifically, we show how our proposed estimator behaves more consistently in the presence of noise and captures information in the tensor data that existing estimators miss. We further show that in the context of image segmentation, our entropy measure shows better agreement with the homogeneity properties of various anatomical regions. Finally, we show that for image registration, our estimator provides a more robust estimate of mutual information than existing binning estimators. We conclude with discussion in Section V.

II. BACKGROUND: BINNING ESTIMATORS

For the purposes of later comparisons, we first summarize the existing binning-based entropy estimators for tensor-valued data. Our nearest neighbor distance-based estimator will follow in Section III.

The Shannon entropy of a scalar-valued dataset is described as

$$H(\mathbf{X}) = - \int p(x) \log_2(p(x)) \quad (1)$$

where $p(x)$ is the probability of the sample x occurring within the dataset \mathbf{X} . For typical image processing applications, the distribution $p(x)$ does not have a parametric form, thereby forcing us to estimate both $p(x)$ and its entropy.

The most common approach to estimating entropy is to discretize the domain of $p(x)$ into a finite set of bins. The sample data is binned and the resulting normalized histogram is used as $p(x)$. The integration in (1) is then converted to a sum and the resulting estimate obtained as

$$H(\mathbf{X}) \approx - \sum_{x \in \mathbf{X}} p(x) \log_2(p(x)) \quad (2)$$

The same histogram binning concept also applies to joint entropy estimation between two sets \mathbf{X} and \mathbf{Y} .

$$H(\mathbf{X}, \mathbf{Y}) = - \int_x \int_y p(x, y) \log_2(p(x, y)) \quad (3)$$

$$\approx \sum_{x \in \mathbf{X}} \sum_{y \in \mathbf{Y}} p(x, y) \log_2(p(x, y)) \quad (4)$$

From the above definitions of entropy and joint entropy, we can estimate many information theoretic measures, including (but not limited to) conditional entropy, mutual information, and variation of information as follows

$$H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{X}, \mathbf{Y}) - H(\mathbf{X}) \quad (5)$$

$$MI(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{X}, \mathbf{Y}) \quad (6)$$

$$VI(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}) + H(\mathbf{Y}) - 2MI(\mathbf{X}, \mathbf{Y}) \quad (7)$$

The consistency of these estimators is inherently linked to that of their underlying entropy estimators and further highlights the necessity of having a consistent and accurate entropy estimator for tensor-valued data.

A. Extensions to Tensor-Valued Data

For tensor-valued data, the above estimators have been used following a scalarization of the data. Different scalarization schemes have led to three main entropy estimators:

- *FA Estimator*: The entropy estimation in (2) is performed on the fractional anisotropy (FA) of the tensors and not on the tensors themselves [8]. The FA can be computed from each tensor \mathbf{D} as

$$FA(\mathbf{D}) = \sqrt{\frac{3}{2}} \frac{|\mathbf{D} - \frac{1}{3}\text{trace}(\mathbf{D})\mathbf{I}|}{|\mathbf{D}|}. \quad (8)$$

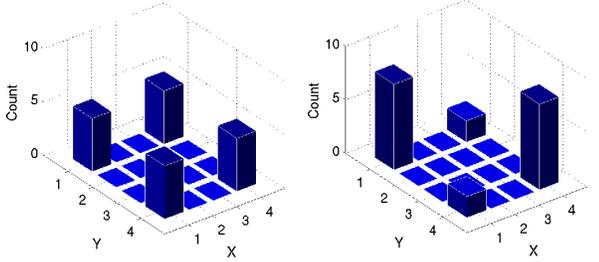
- *Averaging Estimator*: Given the six unique elements $\{d_1, \dots, d_6\}$ of each tensor \mathbf{D} , the entropy is estimated for each element independently using (2), then averaged [5], [6]. Specifically, given a set of tensors \mathbf{X}_D , the entropy is estimated as

$$H(\mathbf{X}_D) \approx \frac{1}{6} \sum_{i=1}^6 H(\mathbf{X}_{d_i}) \quad (9)$$

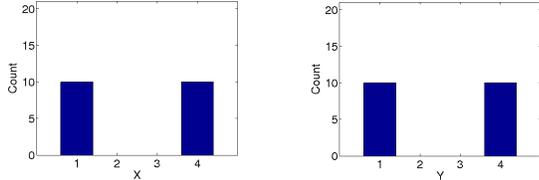
where \mathbf{X}_{d_i} is the set of scalars from channel i of the tensors in \mathbf{X}_D .

- *Grouping Estimator*: Samples in the six dimensions of \mathbf{D} are considered as six samples in one dimension and a single one-dimensional histogram is used to estimate entropy [6]. Given the scalar sets \mathbf{X}_{d_i} for each unique tensor channel i , this approach uses (2) to estimate entropy as

$$H(\mathbf{X}) \approx H(\mathbf{X}_{d_1} \cup \dots \cup \mathbf{X}_{d_6}). \quad (10)$$



(a) Example 1 ($H = 2$ bits) (b) Example 2 ($H = 1.7219$ bits)



(c) X Dimension (d) Y Dimension

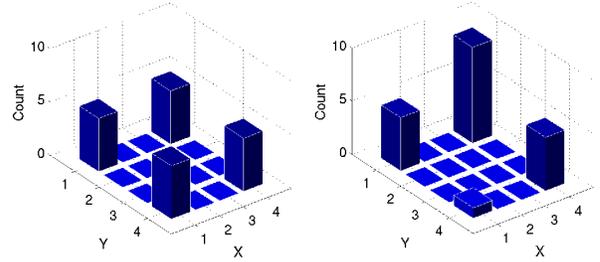
Figure 1. A simple 2D example showing the limitations of the averaging estimator. The two unique distributions in (a) and (b) have the same 1D histograms for each of their channels (shown in (c) and (d) respectively). An averaging estimator would obtain the same entropy estimate for both examples ($H = 1$ bit) despite the fact that (a) has higher entropy than (b).

B. Limitations of Binning Estimators

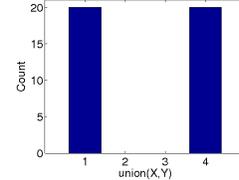
The grouping, averaging, and FA estimators have shown promise in various areas of application [5], [6], [8], yet they have their limitations. Clearly, the FA estimator does not take into account the orientation or size of the tensors, thereby omitting these aspects of the tensors from their analysis. The limitations of the grouping and averaging estimators are more subtle and, therefore, we examine them here.

Figure 1 shows a simple example where the averaging estimator in (9) would have difficulty. The two discrete 2D distributions in Figures 1(a) and 1(b) are clearly different with different entropy values. Yet, when looking at the X and Y channels independently in Figures 1(c) and 1(d), we see that data from both channels are equivalent for both 2D distributions. The averaging estimator would be unable to detect any difference between the information content of these two distributions.

The example in Figure 1 also provides us with some intuition as to where the averaging estimator might be insensitive to information content. The particular reason why the distribution in Figure 1(a) gives the same entropy estimate as that of the distribution in Figure 1(b) is that there is a correlation between the X and Y dimensions. In general, an N -dimensional distribution can cause difficulties for the averaging estimator if there is a negative correlation between different dimensions of the distribution. In such a situation, a distribution can be altered along one dimension (e.g., $X = 1$ in Figure 1) only to have that alteration be canceled out by a related alteration in a second dimension (e.g., $X = 4$ in Figure 1). Further, an increase in entropy



(a) Example 1 ($H = 2$ bits) (b) Example 2 ($H = 1.5754$ bits)



(c) X and Y Grouped

Figure 2. A simple 2D example showing the limitations of the grouping estimator. The two unique distributions in (a) and (b) have the same 1D histogram when grouping the samples from the different channels (shown in (c)). The grouping estimator obtains the same entropy estimate for both examples ($H = 1$ bit) despite the fact that (a) has higher entropy than (b).

along one dimension can be offset by decrease in entropy along a second dimension without affecting the average entropy of all channels.

Meanwhile, Figure 2 shows a simple example where the grouping estimator in (10) would have difficulty. The two discrete 2D distributions in Figures 2(a) and 2(b) are again unique with different entropy. Grouping the samples from the X and Y channels in each of the two distributions results in identical 1D histogram as shown in Figure 2(c). The grouping estimator would provide the same estimate for both 2D distributions despite their differences in true entropy.

Unlike the averaging estimator, the properties of a distribution that can expose the limitations of the grouping estimator cannot be concisely expressed. Yet, Figure 2 shows that such limiting situations exist and would affect further analysis based on these entropy estimates.

III. PROPOSED: NEAREST NEIGHBOR ESTIMATORS

We propose the use of the nearest neighbor entropy estimator to determine the information content of diffusion MRI volumes. The nearest neighbor distances, denoted by η , can be calculated using an appropriate diffusion MRI distance metrics.

The nearest neighbor entropy estimator was first presented by Kozachenko and Leonenko [18] and is given as

$$H_{nn}(\mathbf{X}) = \frac{d}{N} \sum_{j=1}^N \log_2(\|\eta_j\|) + \log_2 \left[\frac{\pi^{\frac{d}{2}}(N-1)}{\Gamma(\frac{d}{2}+1)} \right] + \frac{\gamma}{\ln(2)} \quad (11)$$

where N is the number of samples in set \mathbf{X} , η_j is the distance vector from sample x_j to its nearest neighbor in set \mathbf{X} , d is the number of distributions in the entropy estimation (*i.e.*, $d = \text{length}(\eta_j)$), Γ is the standard Gamma function, and γ is the Euler-Mascheroni constant ($\gamma = 0.5772$). The error of this estimator is known to be bounded as $|H_{nn}(\mathbf{X}) - H(\mathbf{X})| \sim O(N^{-1/2})$ [13]. Given that a typical diffusion tensor image of an adult brain has $N > 10^5$ tensors, this error, in practice, is rather small. Further, the estimator is known to be consistent, meaning that $\lim_{N \rightarrow \infty} H_{nn}(\mathbf{X}) = H(\mathbf{X})$ [18].

The nearest-neighbor estimator, as presented in (11) does have the undesirable property of generating an estimate of negative infinity when presented with non-unique data (*i.e.*, $\eta_j = 0$). This is not a significant concern as two methods have been proposed in the literature to address these infinite estimates. First, we can slightly perturb the data prior to computing the nearest-neighbor distances to ensure uniqueness of the samples in \mathbf{X} [19]. An alternative approach is to simply add a small constant ϵ to each nearest-neighbor distance η_j [13]. We implement that latter approach here and note that in practice, non-uniqueness occurs for a very small percentage of the data (*e.g.*, for our data of twelve adult brain DT images, this happened on average for 0.0041% of the tensors per image).

While the guaranteed consistency of the estimator is appealing, so too is the estimator's flexibility when it comes to measuring distance. We are free to calculate η_j using any appropriate distance metric. In our context of tensor-valued data, this allows us the ability to use any established tensor distance metric, including the Euclidean metric [15], the log-Euclidean metric [14], Riemannian metric [16], [17], and J-Divergence [3] (A comparison of these distance metrics can be found in [15] and is beyond the scope of this paper). These distance metrics are shown in Equations (12 - 15) respectively and their use in computing η_j in (11) give us four different entropy estimators for tensor-valued data.

$$d_E(\mathbf{D}_1, \mathbf{D}_2) = \|\mathbf{D}_1 - \mathbf{D}_2\|_2 \quad (12)$$

$$d_{LE}(\mathbf{D}_1, \mathbf{D}_2) = \|\log(\mathbf{D}_1) - \log(\mathbf{D}_2)\|_2 \quad (13)$$

$$d_R(\mathbf{D}_1, \mathbf{D}_2) = \|\log(\mathbf{D}_1^{-1/2} \mathbf{D}_2 \mathbf{D}_1^{-1/2})\|_2 \quad (14)$$

$$d_J(\mathbf{D}_1, \mathbf{D}_2) = \frac{1}{2} \sqrt{\text{tr}(\mathbf{D}_1^{-1} \mathbf{D}_2 + \mathbf{D}_2^{-1} \mathbf{D}_1) - 6} \quad (15)$$

A. Extension to Mutual Information

Note that (11) can also be used to estimate the joint entropy of two distributions from sample sets \mathbf{X} and \mathbf{Y} . If we consider the joint random variable $Z = (X, Y)$ and its sample set $\mathbf{Z} = \{(x_1, y_1), \dots, (x_N, y_N)\}$, it has been proposed in [12] that $H_{nn}(\mathbf{X}, \mathbf{Y}) = H_{nn}(\mathbf{Z})$. This joint entropy estimate could then be used to compute the additional information-theoretic measures in (5 - 7). It has been noted, however, that the differences in scale between the joint space (\mathbf{Z}) and marginal spaces (\mathbf{X} and \mathbf{Y}), combined

with competing bias correction terms (the last two terms in (11)) from the entropy estimates, can lead to biased estimates of the measures in (5 - 7) [20]. We propose instead the use of α -Mutual Information (α -MI) first presented in [21]. Like our proposed entropy estimator, the α -MI estimator relies on nearest-neighbor distances to obtain an estimate. Unlike our proposed entropy estimator, α -MI is based on Renyi entropy, a generalization of the Shannon entropy seen in (1). The nearest-neighbor estimator for α -MI is

$$MI(\mathbf{X}, \mathbf{Y}, \alpha) = \frac{1}{\alpha - 1} \log \frac{1}{N^\alpha} \sum_{i=1}^N \left(\frac{\eta_i(\mathbf{Z})}{\sqrt{\eta_i(\mathbf{X})\eta_i(\mathbf{Y})}} \right)^{2\beta} \quad (16)$$

where $\mathbf{Z} = \{(x_1, y_1), \dots, (x_N, y_N)\}$, $\eta_i(\mathbf{X})$ is, as in (11), the nearest-neighbor distance of $x_i \in \mathbf{X}$ using the tensor distance metrics in (12 - 15), and $\beta = d(1 - \alpha)$. In this paper, as in [21], we used $\alpha = 0.99$ to approximate the Shannon entropy case.

IV. RESULTS: COMPARISON OF ENTROPY ESTIMATORS

For the following experiments, we estimate entropy using seven different estimators from the aforementioned two estimator families:

- *Binning Estimators*: (a) The FA Estimator, (b) The Averaging Estimator, (c) The Grouping Estimator. Unless otherwise stated, bin widths are chosen using Scott's rule

$$h = \frac{3.5\sigma}{N^{1/3}} \quad (17)$$

where σ is the standard deviation of the scalar data while N is the number of scalar samples [11].

- *Proposed Nearest-Neighbor Estimators*: Nearest Neighbour distances computed using the (d) Euclidean distance metric, (e) log-Euclidean distance metric, (f) Riemannian distance metric, and (g) J-Divergence.

We aim to show how our proposed nearest-neighbor entropy formulation provides more robust and consistent behavior than existing binning estimators in the application areas of noise estimation, image segmentation, and image registration.

A. Noise Estimation

The objective of this experiment is to test whether the given entropy estimators behave consistently with respect to noise levels within a tensor image. We expect that, on average, as noise increases, entropy also increases.

For this experiment, two-dimensional tensor images of size 32×32 pixels were generated by sampling from a Wishart distribution. The Wishart distribution is defined over a non-negative-definite matrix-valued random variable with a given scale matrix \mathbf{V} and degree of freedom k [22]. The degree of freedom of the Wishart distribution follows the convention of statistical estimation; as the degree of freedom increases, the variability of the samples produced by the

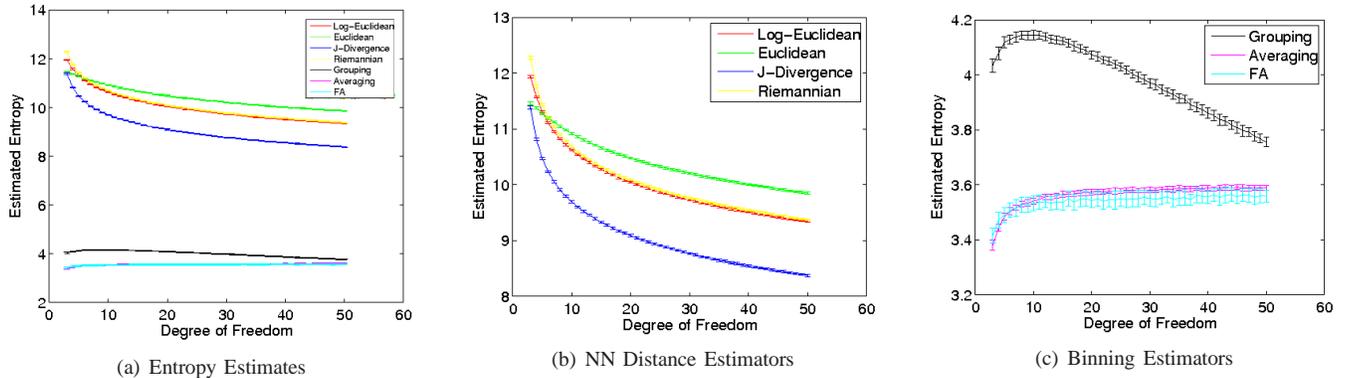


Figure 4. Entropy estimates for noisy synthetic data. Entropy estimates using various estimators shown in (a) and grouped into distance-based (b) and binning-based (c) methods for a close up view. Note that, as expected, the distance-based estimates decrease as noise decreases, while the binning-based estimates behave inconsistently with respect to the underlying uncertainty in the data.

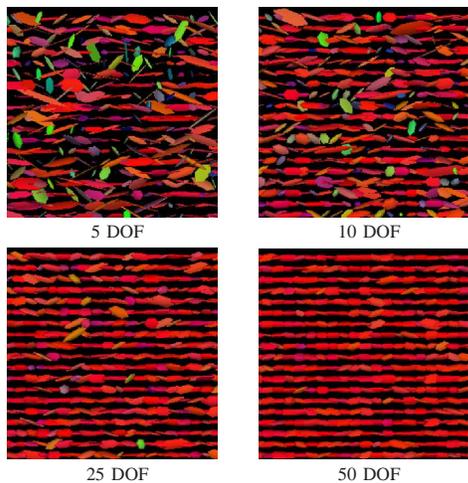


Figure 3. Synthetic tensor data sampled from Wishart distributions of different degrees of freedom and same mean tensor $\text{diag}(3, 1, 1)$. Note that as the degree of freedom increases, the variability of the sample data decreases.

distribution *decreases*. Sample data for various degrees of freedom are shown in Figure 3. We vary the degree of freedom parameter between 3 and 50 while keeping the mean tensor fixed at $\text{diag}(3, 1, 1)$ (other choices of mean tensor gave similar results). Fifty synthetic images are generated for each degree of freedom and entropy estimates are computed for each estimator-image pair.

The mean entropy estimates for each noise level are shown in Figure 4(a). Note how the nearest-neighbor estimators provide a much higher entropy estimate than the binning estimators. This result stems directly from our proposed estimator avoiding the dimensionality reduction used by the binning estimators. By reducing the size of the space that the sampled data can occupy, the binning estimators reduce the unpredictability of the sampled tensor space. It is precisely this unpredictability that entropy measures. By avoiding this

dimensionality reduction, our proposed estimators do not suffer from this entropy loss.

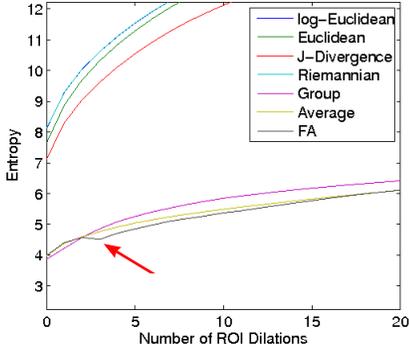
Figure 4 also shows the entropy estimates grouped by estimator families. For our proposed nearest-neighbor estimators in Figure 4(b), we see that, as expected, entropy monotonically decreases as noise in the images decreases. However, we notice erroneous behavior for the binning estimators in Figure 4(c). At high noise levels, the entropy estimates provided by the binning estimators increase as noise decreases. Further, the averaging and FA estimators effectively plateau around a degree of freedom of 20. These inconsistencies appear as a direct result of the dimensionality reductions performed prior to entropy estimation. As a result of the dimensionality reductions, the fidelity of the entropy estimates to the original data is lost, thereby producing inconsistent estimator behavior.

B. Image Segmentation

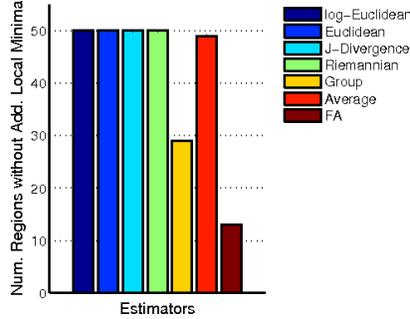
The problem of image segmentation can be described as the problem of splitting an image into two or more regions that are homogeneous in some fashion. As entropy is related to variability in a set of samples, segmentation methods have been proposed that define an optimal partition of an image as that which results in homogeneous regions with minimal entropy values in each region [23]. The estimators described herein can be used in such a context and we therefore examine the properties of our proposed estimators for image segmentation.²

To examine the potential of our proposed entropy estimators in this context, we estimate the entropy of the fifty labeled white matter fiber bundles of the ICBM DTI-81 Atlas [24]. Using these ground truth atlas segmentations, we

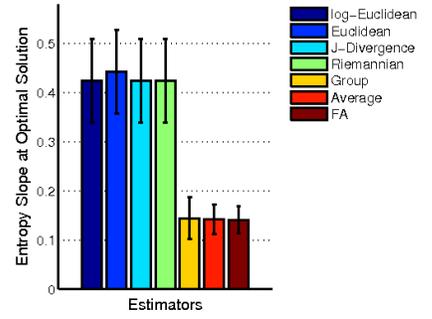
²We emphasize that our goal is not to perform and validate segmentation, but rather to study the effects of entropy estimation on a key aspect of segmentation algorithms (the objective function landscape). We do not carry out the optimization leading to segmentation in order to avoid polluting the comparisons with other optimization-related parameters (e.g. step size in gradient descent) that are irrelevant to the contribution of the paper.



(a) Entropy as a function of ROI Dilatation (\propto segmentation error) for the Left Cingulum



(b) Number of Regions without Superfluous Local Minima



(c) Slope of Entropy Function near Optimal Segmentation

Figure 5. Entropy estimates from various estimators for different quality segmentations of white matter fiber bundles in the ICBM DTI-81 Atlas. Figure (a) shows the entropy estimates as a function of ground truth ROI dilation for the left Cingulum. Note the slight dip in the entropy estimate for the FA estimator at around 3 ROI dilations (red arrow). To better quantify non-convexity, we study the existence of local minima. Figure (b) shows, for each estimator, the number of regions in this experiment where we avoid generating misleading local minima. Note that our proposed estimators are the only ones that, in all cases, did not generate additional local minima, thereby demonstrating their greater suitability in a gradient-based optimization of a DTI segmentation problem. Figure (c) shows the average slope of the entropy vs. ROI dilation curves near the optimal segmentations. The greater slope provided by our proposed estimators suggest faster convergence in the optimization step of an image segmentation problem.

proceed to gradually perturb and worsen the segmentation. We expect that for a measure of entropy to be effective for segmentation, the measure must increase as the image partitions include more and more voxels from other anatomical regions with different diffusion properties. To gradually worsen the segmentation one iteration at a time, we morphologically dilate the segmented regions with a spherical structuring element of radius $r = 1$ voxel. Entropy is then estimated using each estimator after each dilation step. A maximum of seventy-five region dilations were performed for each of the fifty segmentation fiber bundles.

Figure 5 displays how the different estimators behave with respect to different quality segmentations. A representative result is shown for the left Cingulum in Figure 5(a). As expected, each entropy estimator provides a minimum entropy value at the ground truth segmentation with higher entropy values for poorer quality segmentations. However, for the FA estimator, the increase in entropy was not monotonic as the segmentation quality got worse. Instead, we see a local minimum develop at around 3 ROI dilations. Such local minima are a common cause of difficulties for gradient-based optimization schemes commonly used to solve image segmentation problems. Figure 5(b) shows the number of anatomical regions (out of 50) for which no local minima was created as the segmentation was perturbed. The higher the number, the more suitable the measure is for segmentation. We see that, for all regions, our proposed entropy estimators do not generate additional local minima whereas the binning-based estimators generate non-convexity of the optimization landscape for one or more of the regions in this experiment.

Figure 5(c) displays the mean slope of the entropy versus ROI dilation curves for each estimator around the ground

truth segmentations. We see a significantly higher slope for our proposed estimators than for existing binning-based estimators. Such high slopes are beneficial for optimization methods used in image segmentation as they increase the rate of convergence of the optimizer, thereby generating a solution in fewer computational steps. These results provide evidence that our proposed estimators would be better suited for energy-minimizing segmentation by generating a more favorable objective function landscape.

C. Image Registration

Finally, we aim to show improvements in image registration using the mutual information estimates provided by our nearest-neighbor estimators³. We perform this task by computing the image registration energy landscape around the optimal alignment for given pairs of tensor-valued images. We expect that our proposed estimators would generate a more robust and distinguishable optimum value than the existing binning-based estimators. Twelve diffusion tensor MR images publicly available from the Laboratory of Brain Anatomical MRI (LBAM) at John Hopkins Medical Institute were used in this experiment [25]. For each of the twelve diffusion tensor MR images in the dataset, we generate candidate registered (*i.e.*, transformed) images by applying a deformation to the original image followed by the addition of additive noise. For our deformation step, we apply rigid rotations about the inferior-superior axis (the image Z-axis) followed by tensor reorientation. The image then is decomposed into its six diffusion weighted images using the methods in [1]. Rician noise of zero mean and a

³As in the image segmentation experiment, our goal is to study the effects of entropy estimation on the objective function landscape, not to perform the registration itself.

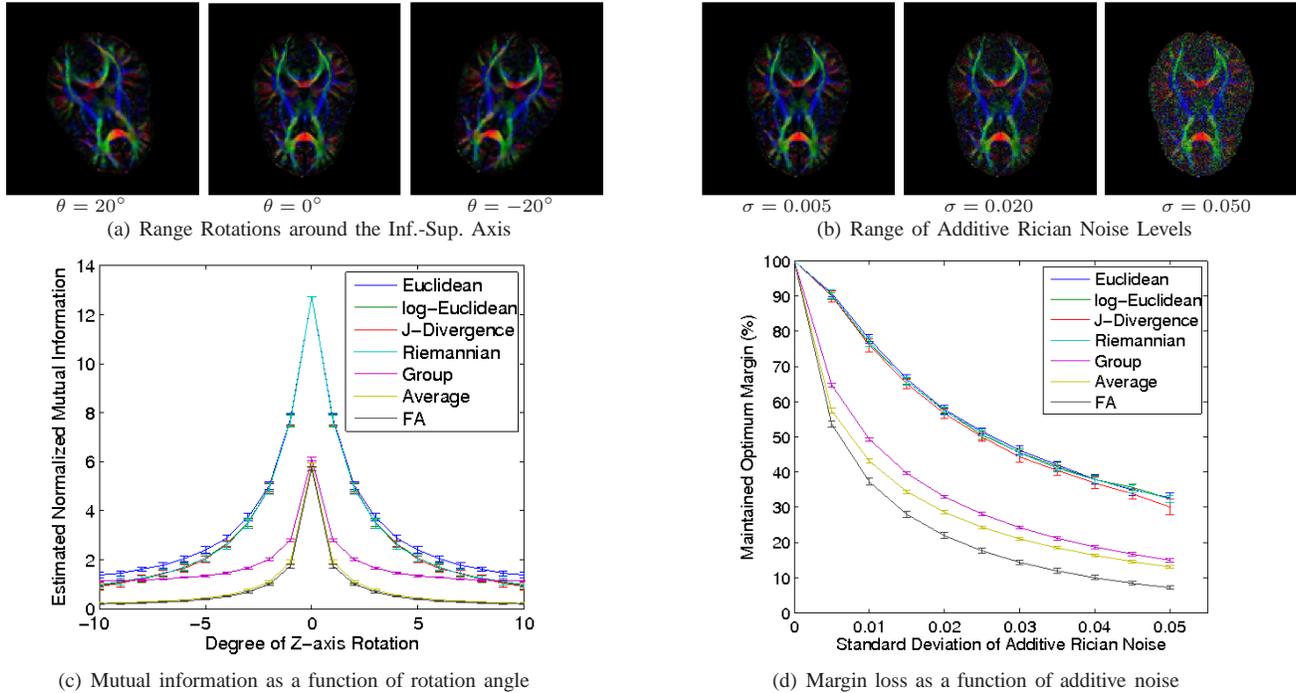


Figure 6. Mutual information estimates using various tensor-valued estimators for rotational misalignments with additive noise. Figure (a) shows the range of z-axis rotations tested while figure (b) shows the range of additive Rician noise values. Mutual information for different rotational alignments is shown in (c) for the case without additive Rician noise. Note the significantly higher slope around the optimum for our proposed estimators. Figure (d) shows how the margin between the optimal mutual information value and the baseline changes as noise increases. Note the added robustness of our nearest-neighbor estimators in the presence of additive noise.

given standard deviation is added to each diffusion weighted image. The tensors are then recomputed using the log-least-squares approach in [1] to generate our candidate image. We compute the mutual information using (6) and (16) between the original image and the candidate image for various degrees of rotation (from $\theta = -20^\circ$ to $\theta = 20^\circ$) and different levels of additive Rician noise (standard deviation from $\sigma = 0.005$ to $\sigma = 0.050$). The range of rotations and noise levels are displayed in Figures 6(a) and 6(b) respectively.

Figure 6(c) shows the resulting energy landscape for the case without additive noise. We note that all estimators obtain an optimum at the correct alignment, as expected, but that the slope around the optimal solution is clearly much higher for our nearest-neighbour estimators than for the binning estimators. Again, this greater slope improves the convergence rate of many optimizers used in image registration.

Figure 6(d) displays how each estimator behaves as we add noise to the candidate images. For each noise level, we compute the optimization *margin*: the mutual information at the optimum minus the asymptotic baseline mutual information value. We see that as we add noise, the margin for the binning estimators decrease rapidly while the margin from our proposed estimators decrease at a significantly slower rate. This sharper decrease in margin for the bin-

ning estimators demonstrate that they are more sensitive to slight differences in data than our proposed nearest-neighbor estimators.

V. CONCLUSIONS AND FUTURE WORK

We have presented for the first time estimators for entropy and mutual information of tensor-valued images. Using nearest-neighbor distances computed with metrics that respect the manifold of the tensor data, we generate entropy estimators that work directly in the tensor space and require no dimensionality reduction. Our proposed estimators are guaranteed to behave consistently with low error.

We have shown that our entropy estimators behaves as expected in the presence of varying levels of noise, a result not seen with state-of-the-art methods [5], [6], [8]. Further, we applied our estimators to DTI image segments that mismatch, to various degrees, their underlying anatomical regions. Results show that our estimators use in image segmentation can lead to objective function landscapes that are easier to optimize (with higher slopes and devoid of local minima). Finally, we have shown that using nearest-neighbor distances, we can obtain estimates of mutual information between two images that are more robust to noise, a desirable and key result for image registration algorithms.

While the results we have shown are not an exhaustive presentation of computer vision methods that rely on infor-

mation theory, we believe that our tensor-valued information content estimators can show further benefits in other areas of application (e.g., feature detectors, recognition) [2]. We also note that our approach extends naturally to higher order diffusion MRI data by using existing distance metrics [26], [27]. Future work will focus on incorporating our estimators into existing image segmentation and registration algorithms.

REFERENCES

- [1] C.-F. Westin, S. Maier, H. Mamata, A. Nabavi, F. Jolesz, and R. Kikinis, "Processing and visualization for diffusion tensor MRI," *Medical Image Analysis*, vol. 6, pp. 93–108, 2002.
- [2] F. Escolano, P. Suau, and B. Bonev, *Information Theory in Computer Vision and Pattern Recognition*. Springer, 2009.
- [3] Z. Wang and B. C. Vemuri, "An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation," in *Proceedings of Computer Vision and Pattern Recognition – CVPR 2004*, vol. 1, 2004, pp. 228–233.
- [4] —, "DTI segmentation using an information theoretic tensor dissimilarity measure," *IEEE Transactions on Medical Imaging*, vol. 24, no. 10, pp. 1267–1277, 2005.
- [5] C. Studholme, "Dense feature deformation morphometry: Incorporating DTI data into conventional MRI morphometry," *Medical Image Analysis*, vol. 12, pp. 742–751, 2008.
- [6] W. V. Hecke, A. Leemans, E. D'Agostino, S. D. Backer, E. Vandervliet, P. M. Parizel, and J. Sijbers, "Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information," *IEEE Transactions on Medical Imaging*, vol. 26, no. 11, pp. 1598–1612, 2007.
- [7] A. Leemans, J. Sijbers, S. D. Backer, E. Vandervliet, and P. M. Parizel, "Affine coregistration of diffusion tensor magnetic resonance images using mutual information," in *Advanced Concepts for Intelligent Vision Systems*, ser. LNCS, vol. 3708, 2005, pp. 523–530.
- [8] T. Ahearn, R. Staff, L. Whalley, and A. Murray, "Entropy of FA maps identifies healthy older individuals at risk of dementia," in *Organization for Human Brain Mapping 2009 Annual Meeting*, vol. 47, July 2009, p. S104.
- [9] T. Neuvonen and E. Salli, "Characterizing diffusion tensor imaging data with directional entropy," in *Proceedings of IEEE Engineering in Medicine and Biology – EMBS*, 2005, pp. 5798–5801.
- [10] E. Ozarslan, B. C. Vemuri, and T. H. Mareci, "Generalized scalar measures for diffusion MRI using trace, variance, and entropy," *Magnetic Resonance in Medicine*, vol. 53, pp. 866–876, 2005.
- [11] D. W. Scott, *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley, 1992.
- [12] J. D. Victor, "Binless strategies for estimation of information from neural data," *Physical Review E*, vol. 66, 2002.
- [13] J. Kybic, "High-dimensional mutual information estimation for image registration," in *Proceedings of IEEE International Conference on Image Processing – ICIP*, 2004.
- [14] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic Resonance in Medicine*, vol. 56, pp. 411–421, 2006.
- [15] O. Pasternak, N. Sochen, and P. J. Basser, "The effect of metric selection on the analysis of diffusion tensor MRI data," *NeuroImage*, vol. 49, pp. 2190–2204, 2010.
- [16] P. G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly, "A rigorous framework for diffusion tensor calculus," *Magnetic Resonance in Medicine*, vol. 53, pp. 221–225, 2005.
- [17] X. Pennec, P. Fillard, and N. Ayache, "A riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006.
- [18] L. Kozachenko and N. Leonenko, "On statistical estimation of entropy of a random vector," *Problems in Information Transmission*, vol. 23, no. 2, pp. 95–101, 1987.
- [19] J. Kybic, "Incremental updating of nearest neighbor-based high-dimensional entropy estimation," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing – ICASSP*, vol. III, 2006, pp. 804–807.
- [20] A. Kraskov, H. Stöbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, p. 16, 2004.
- [21] H. Neemuchwala and A. Hero, *Entropic Graphs for Registration*, ser. Signal Processing and Communications. CRC Press, 2005, ch. 6.
- [22] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20A, no. 1-2, pp. 32–52, 1928.
- [23] L. Hermes and J. Buhmann, "A minimum entropy approach to adaptive image polygonization," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1243–1258, 2003.
- [24] U. L. of Neural Imaging, "ICBM DTI-81 atlas," <http://www.loni.ucla.edu/Atlases>, accessed February 2009.
- [25] S. Mori, "John Hopkins Medical Institute: Laboratory of Brain Anatomical MRI, in vivo human database," <http://lbam.med.jhmi.edu/>, accessed February 2009.
- [26] A. Barmpoutis, M. S. Hwang, D. Howland, J. R. Forder, and B. C. Vemuri, "Regularized positive-definite fourth order tensor field estimation from DW-MRI," *NeuroImage*, vol. 45, pp. S153–S162, 2009.
- [27] M.-C. Chiang, A. D. Klunder, K. McMahon, G. I. de Zubicaray, M. J. Wright, A. W. Toga, and P. M. Thompson, "Information-theoretic analysis of brain white matter fiber orientation distribution functions," in *Proc. of Info. Proc. for Med. Imag.*, vol. 20, 2007, pp. 172–182.