A New Distribution Metric for Image Segmentation

Romeil Sandhu[a] Tryphon Georgiou[b] Allen Tannenbaum[a]

[a]School of Electrical & Computer Engineering, Georgia Institute of Technology,
Atlanta, GA 30332-0250
[b]Department of Electrical & Computer Engineering, University of Minnesota,
Minneapolis, MN 55455

ABSTRACT

In this paper, we present a new distribution metric for image segmentation that arises as a result in prediction theory. Forming a natural geodesic, our metric quantifies "distance" for two density functionals as the standard deviation of the difference between logarithms of those distributions. Using level set methods, we incorporate an energy model based on the metric into the Geometric Active Contour framework. Moreover, we briefly provide a theoretical comparison between the popular Fisher Information metric, from which the Bhattacharyya distance originates, with the newly proposed similarity metric. In doing so, we demonstrate that segmentation results are directly impacted by the type of metric used. Specifically, we qualitatively compare the Bhattacharyya distance and our algorithm on the Kaposi Sarcoma, a pathology that infects the skin. We also demonstrate the algorithm on several challenging medical images, which further ensure the viability of the metric in the context of image segmentation.

Keywords: Geometric Active Contours, Distributions, Segmentation, Metrics

1. INTRODUCTION

A well-studied problem in computer vision is the fundamental task of segmenting or partitioning an image into disjoint regions with applications ranging from medical image analysis, quality control, or military surveillance and tracking. Although the general segmentation problem involves separating N distinct partitions, a piecewise assumption of two sets is generally made. That is, the image is assumed to be comprised of two homogeneous regions, often referred to as "Object" and "Background". The goal of segmentation is to accurately capture these regions. Specifically, the use of active contours has been proven to be quite successful in accomplishing this task.[1–6]

Employing the methodology of geometric active contours (GAC), a curve is represented as the zero level-set of a higher dimensional surface[7,8] A common choice is the signed distance function. Although this implicit representation of a curve is computationally more expensive than parametric approaches, it allows for the contour to naturally undergo topological changes. In the GAC framework, a curve is evolved to minimize an image based energy functional, typically via gradient descent. Several energy functionals have been proposed in literature: Some models use local information and features such as edges,[3] while other methods use regional information by discriminating on a photometric variable of interest (e.g., gray-scale intensity, color, tensors).[5, 9–11] Recently, hybrid models, which incorporate regional statistics in a localized fashion, have been proposed.[12] It should also be noted that one can constrain the evolution of a curve by incorporating shape information, and we refer the interested reader for more details to the following references.[13–15]

Region based approaches have been popularized due to a higher level of robustness to noise and initialization when compared to models based on local information. In region-based segmentation, energy models employ the use of image statistics dependent on the segmenting curve using parametric[9, 16, 17] and non-parametric methods.[18] In this work, we will restrict most of our discussion to approaches that generalize the statistical inference beyond first and second moments to entire probability density functions (pdf). From this, segmentation can be reinterpreted as measuring the "distance" between two distributions via a similarity metric.

Further author information: (Send correspondence to R.S.)
R.S.: E-mail: rsandhu@gatech.edu, Telephone: 1 404 385 5062

By directly measuring the discrepancy of pixel intensities, Rousson and Deriche proposed to maximize the L_2-distance between the log-likelihood of two distributions defined by the interior and exterior regions of the segmenting curve.[16] Although a Gaussian assumption is made in their work, an extension to entire probability distributions is straightforward. Interestingly, the metric proposed in this paper, when mapped to a linear space, results in a similar energy.[19] Using an information theoretic approach, mutual information between region labels and image intensities has also been proposed. Recently, Freedman *et. al* introduce the Bhattacharyya distance and Kullback-Leibler divergence for segmentation by maximizing the similarity between the density of a region enclosed by a curve $\mathcal{C}$ and a known pdf that is learned *a-priori*.[20] By relaxing the assumption of *a-priori* knowledge, Rathi *et. al* derived a flow that optimally separates distributions via the Bhattacharyya measure.[21]

Although these measures are based on varying disciplines and are motivated by a specific problem, they can be extended to a general class of densities. Moreover, as stated above, they can be applied to visual information. However, we highlight and note that segmentation results are *directly impacted* by the type of similarity measure used. From this, we introduce a new distribution metric for image segmentation. Based on prediction theory, our intrinsic metric quantifies the "distance" between two density functions as the standard deviation of the difference between the logarithms of the two density functions.[19, 22] In previous work and similarly to that of Freedman *et al.*, we have proposed to match the distribution of the region interior to the segmenting curve with that of known density (learned *a-priori*).[23] However, the emphasis was for visual tracking. Specifically, we incorporated the metric in the deformation functional of a particle filtering scheme.[24] In this paper, we relax the assumption of *a-priori* knowledge, and focus on the accurate segmentation of biological structures or pathologies found in medical imagery. A qualitative comparison between the Bhattacharyya distance and our newly proposed metric is then done to highlight the notion that similarity measures can not be applied equally for image segmentation. In particular situations, one distribution metric may prevail while another measure may be highly dependent on the initialization or statistical information. To the best of our knowledge, a qualitative comparison of segmentation results has not been made.

This paper is organized as follows: In the next section, we provide a theoretical overview of our metric as well as review certain distribution measures popularized by the computer vision community. In Section 3, we incorporate the proposed metric within the GAC framework, and derive the corresponding curve evolution that maximizes the distance between the region interior and exterior to the segmenting curve. Experimental results are given in Section 4. We conclude with a summary in Section 5.

2. PRELIMINARIES

In this section, we provide a brief overview of the metric proposed for image segmentation as well as its comparisons to measures derived from the Fisher Information metric. In addition, we highlight interesting parallels between certain classical distances between distribution and our new metric.

2.1 Metric based on Prediction Theory

Typically, we seek to maximize the distance of two distributions defined by a certain photometric variable so as to identify the distribution that best characterizes an object or the foreground. To this end we motivate our metric that quantifies (dis)similarity between spectral distributions using basic concepts of prediction theory. So let $f_1(\theta)$ and $f_2(\theta)$, with $\theta \in [0, \pi]$, denote the power spectral densities of two discrete-time, zero-mean, random processes $u_k^{f_i}$ with $i \in \{1, 2\}$ and $k \in \mathbb{Z}$. The variance of the optimal, linear, one-step-ahead prediction error for the process corresponding the density $f_i(\theta)$ is

$$\mathcal{E}\{|u_0^{f_i} - \hat{u}_{0|past}^{f_i}|^2\} = \mathcal{E}\{|u_0^{f_i} - \sum \alpha_k^{f_i} u_{-k}^{f_i}|^2\} \tag{1}$$

with $k > 0$. Here, $\alpha_k^{f_i}$ denote the coefficients that minimize the linear prediction error variance for the specific $f_i(\theta)$. Since we want to quantify dissimilarity between $f_1(\theta)$ and $f_2(\theta)$, we use $f_2(\theta)$ to design a linear predictor (i.e., coefficients $\alpha_k^{f_2}$, $k \geq 0$), and then compare how well this predictor performs when used to predict a random process with spectral density $f_1(\theta)$ **against the optimal predictor** for the corresponding process, i.e., the optimal one that is based on $f_1(\theta)$ instead of $f_2(\theta)$. That is, we quantify the *degradation of predictive error variance* for the case where the predictor is optimal for one process and then applied to the other. The

degradation of predictive error variance turns out to be equal to the ratio between the arithmetic over the geometric means of the fraction f_1/f_2, namely

$$\rho(f_1, f_2) := \frac{\mathcal{E}\{|u_0^{f_1} - \sum_{k=1}^{k=\infty} \alpha_k^{f_2} u_{-k}^{f_1}|^2\}}{\mathcal{E}\{|u_0^{f_1} - \sum_{k=1}^{k=\infty} \alpha_k^{f_1} u_{-k}^{f_1}|^2\}} = \frac{\int \frac{f_1(\theta)}{f_2(\theta)} d\theta}{\exp \int \log \frac{f_1(\theta)}{f_2(\theta)} d\theta}. \tag{2}$$

The quantity $\rho(f_1, f_2)$ can be viewed as analogous to "divergences" such as the Kullback-Leibler relative entropy used in Information Theory. The differential form $\rho(f, f+\Delta)$, for Δ a "small" perturbation (i.e., neglecting third order terms and above), gives rise to the following Riemannian metric on the cone of spectral density functions

$$g_f(\Delta) := \int_\chi \left(\frac{\Delta(x)}{f(x)}\right)^2 dx - \left(\int_\chi \frac{\Delta(x)}{f(x)} dx\right)^2. \tag{3}$$

Geodesics turn out to be exponential families of distributions and geodesic distances can be computed in closed form.[19, 22] In fact, the geodesic distance provides the sought metric between distributions and this is given in Equation (9), tailored for the level set framework of the present paper. It is interesting to note that geodesic distances (as in Equation 9) are is insensitive to scaling and can be seen as giving rise to a "shape" comparator (and hence, actually, a pseudo-metric).

2.2 Fisher Metric, Hellinger Discrimination, Bhattacharyya Distance

We now discuss and compare the Riemannian structure of the well-known Fisher information metric with the structure of our metric that was proposed in the previous section. It is interesting to note that the Kullback-Leibler divergence and the Fisher metric can be motivated in a manner completely analogous to our metric, using a paradigm from source-coding instead of the degradation of predictive error variance that we saw earlier. Thus, in the present context two probability distributions p_1 and p_2 are compared based on "coding efficiency" of a design that is based on one of the two and then applied to a source generating symbols according to the other distribution. More specifically, let $\mathcal{P}$ denote the finite dimensional probability simplex

$$\mathcal{P} := \{p(k) \text{ where } k \in \{1, ..., n\}, p(k) \ge 0, \sum_x p(x) = 1\}.$$

Figure 1b displays the 2-dimensional simplex as a red triangular surface in $\mathbb{R}^3$. The optimal code-length for independent symbols generated according to $p_1 \in \mathcal{P}$ is precisely the entropy $-\sum_x p_1(x) \log(p_1(k))$ (e.g., see[25]). Now, if the code is designed **based** on p_2 while the symbols are generated according to p_1, the degradation of coding efficiency is given by the Kullback-Leibler (KL) divergence

$$K(p_1, p_2) := \sum_k p_1(k) \log(\frac{p_1}{p_2}). \tag{4}$$

The differential form of the Kullback-Leibler divergence is a Riemannian metric known as the Fisher information metric:

$$g_{fisher,p}(\Delta) = \sum_\chi \frac{\Delta(k)^2}{p(k)} \tag{5}$$

The mapping $p(k) \mapsto u(x) := \sqrt{p(x)}$ maps the probability simplex onto the (positive quadrant of the) sphere $\sum_\chi u(x)^2 = 1$. This is exemplified in Figure 1 where the simplex (red triangular surface) is mapped onto the blue spherical surface. Interestingly, under this map, the Fisher metric correspond to the Euclidean distance on the surface of the sphere.[26, 27] Thus, geodesics map onto great circles and the geodesic distance is precisely the arclength on the sphere. The well-known Hellinger discrimination[28]

$$H(p_1, p_2) = \sum_\chi \left(\sqrt{p_1(k)} - \sqrt{p_2(k)}\right)^2 \tag{6}$$

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \mapsto \begin{pmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \sqrt{p_3} \end{pmatrix}$$
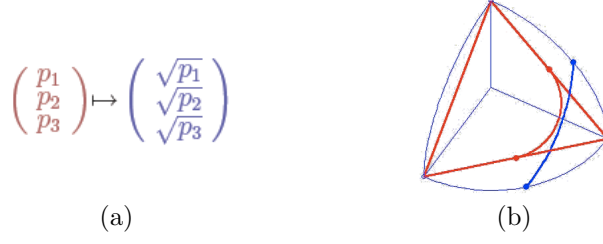
(a)   (b)

Figure 1. (a) The map $p \mapsto \sqrt{p}$ for each point on the simplex (b) The simplex as a triangular red surface taken onto the orthant of the blue spherical surface

as well as the famed Bhattacharyya distance[29]

$$B(p_1, p_2) = \sum_{\chi} \sqrt{p_1(k)p_2(k)} \tag{7}$$

can be directly related to the arclength and induce the same topology. Indeed, the Bhattacharyya distance is precisely the cosine of the geodesic arc between the two image points under the mapping $p \mapsto \sqrt{p}$. For continuous random variable and probability density functions (as opposed to probability vectors) summation is replaced by integration. In the next section we highlight the parallels between the two metrics.

2.3 Parallels and Comparison the Predictive metric and the Fisher metric

Given that we seek to measure the distance between two density functionals, it has been shown that the notion of degradation of performance is a powerful tool in forming a measure of (dis)similarity. That is, the performance that is measured when the wrong choice between the two alternatives is made. Moreover, both degradation measures induce a Riemannian metric in the space of the respective distributions. However, the manner in which the respective metrics penalize perturbations and distances are quite different as seen Table 1. Also, while the $p \mapsto \sqrt{p}$ maps the probability simplex onto the sphere, the mapping $f \mapsto \log f$ takes the cone of spectral densities into a linear space. In other words, we are able to map geodesics defined by our Riemannian metric into straight lines. In doing so, one should note the connection with the popular energy functional proposed by Rousson and Deriche. This can be now reformulated and is similar to that of the mapped version of our Riemannian metric, in the linear L_2 sense. However, our metric naturally incorporates a term that makes it scale-invariant and a "shape recognizer" as noted earlier. Table 1 below highlights differences between the Fisher metric and the metric proposed in this paper.

Type of Comparison	Information-Based Metric	Prediction-Based Metric
metric	$\int \frac{\Delta^2}{p}$	$\int (\frac{\Delta}{f})^2 - (\int \frac{\Delta}{f})^2$
mapping	$p \mapsto \sqrt{p}$	$f \mapsto \log f$
geodesic distance	great circles	logarithmic families

Table 1. Theoretical Comparison between the Fisher Information Metric and our proposed Metric

From the above table, it should be no surprise that segmentation results will indeed be different. In the next section, we cast the geodesic distance as an energy functional in the GAC framework for image segmentation.

3. LEVEL SET FORMULATION OF PROPOSED SIMILARITY METRIC

We consider the problem of segmenting an image I. That is, we first assume the image is composed of two homogeneous regions referred to as "Object" and "Background". From this, the goal of segmentation is to capture these two regions. To do so, we enclose a curve $\mathcal{C}$, represented as the zero-level set of a signed distance function $\phi : \Re^2 \to \Re$, such that $\phi < 0$ represents the inside of $\mathcal{C}$ and $\phi > 0$ represents the outside of $\mathcal{C}$. Our goal is to evolve the curve $\mathcal{C}$, or equivalently ϕ, so that the interior matches the Object and the exterior matches the

Background. As a result, the curve $\mathcal{C}$ would then match the boundary of the curve $\partial\Omega$ separating the Object and Background. The general minimization is performed by evolving C according to the flow:

$$\frac{\partial\phi}{\partial t} = \nabla_\phi E_{image} + \lambda \cdot \delta(\phi) \cdot \mathrm{div}\left(\frac{\nabla(\phi)}{|\nabla(\phi)|}\right) \tag{8}$$

where a regularization term is added to the image based energy. We now propose an energy functional based on our metric discussed previously, and derive the corresponding partial differential equation (PDE) that describes its curve evolution in the level set framework. Moreover, because our metric measures similarity or "distance" as the standard deviation between the log-likelihood of two distributions, p_{in} and p_{out}, we seek to maximize the following energy functional

$$E_{image}(z,\phi) = \sqrt{\mathcal{E}\left\{\left(\log\frac{p_{in}(z,\phi)}{p_{out}(z,\phi)}\right)^2\right\} - \mathcal{E}\left\{\left(\log\frac{p_{in}(z,\phi)}{p_{out}(z,\phi)}\right)\right\}^2} \tag{9}$$

where $\mathcal{E}\{f(z)\}$ is the expected value of the functional $f(z)$ with respect to the random photometric variable $z \in \mathcal{Z}$, and p_{in} and p_{out} are the pdf's defined on the random variable z. In the present work, we restrict the variable z to set of gray level values $\{1,2,...,256\}$. Moreover, let $I : \Re^2 \to \mathcal{Z}$ be a mapping of the image defined over the domain Ω, to the photometric variable z, and $x \in \Re^2$ be the image coordinates. Then the pdf inside the curve $\mathcal{C}$ can be formulated as such

$$p_{in}(z,\phi) = \int_\Omega \frac{K(z-I(x))H_\epsilon(-\phi)}{H_\epsilon(-\phi)}dx \qquad p_{out}(z,\phi) = \int_\Omega \frac{K(z-I(x))H_\epsilon(\phi)}{H_\epsilon(\phi)}dx \tag{10}$$

where $K(z-I(x))$ is a specified kernel. For numerical experiments, we have used $K(z-I(x)) = \delta_\epsilon(z-I(x))$. Also $H_\epsilon : \mathbf{R} \mapsto \{0,1\}$ denotes the smoothed Heaviside step function with the corresponding derivative δ_ϵ. These are both given as follows

$$H_\varepsilon(\phi) = \begin{cases} 1 & \phi > \epsilon \\ 0 & \phi < -\epsilon \\ \frac{1}{2}(1 + \frac{\phi}{\epsilon} + \frac{1}{\pi}\sin(\frac{\pi\phi}{\epsilon})) & \text{otherwise} \end{cases} \qquad \delta_\epsilon(\phi) = \begin{cases} 0 & \phi > \epsilon, \quad \phi < -\epsilon \\ \frac{1}{2\epsilon}(1 + \cos(\frac{\pi\phi}{\epsilon})) & \text{otherwise} \end{cases} \tag{11}$$

The gradient $\nabla_\phi T$ can be computed using the calculus of variations. Taking the first variation with respect to ϕ, we arrive at the following PDE

$$\nabla_\phi E_{image} = -\frac{\delta_\epsilon(\phi)}{E_{image}} \cdot [\mathcal{E}\{B \cdot G\} - \mathcal{E}\{B\} \cdot \mathcal{E}\{G\}] \tag{12}$$

with B and G given as

$$B = \log\frac{p_{in}(z,\phi)}{p_{out}(z,\phi)} \qquad G = \left[\left(\frac{1}{A_{in}} + \frac{1}{A_{out}}\right) - K(z-I(x))\left(\frac{1}{A_{in}p_{in}(z,\phi)} + \frac{1}{A_{out}p_{out}(z,\phi)}\right)\right] \tag{13}$$

Also noting that A_{in} is given by $\int_\Omega H_\epsilon(\phi)dx$, Equation (12) is a PDE that describes the evolution of the curve $\mathcal{C}$ that optimally maximizes the "distance" between the distributions which correspond to the exterior and interior regions of segmenting curve.

4. EXPERIMENTS

In this section, we present experimental results obtained from evolving a curve $\mathcal{C}$ according to Equation (8). Moreover, we provide a qualitative comparison between our metric and that of the result obtained with the Bhattacharyya distance. This is done for segmenting the Kaposi Sarcoma, a pathology that infects the skin. We also demonstrate our algorithm on segmenting medical structures and the classic image of a zebra.
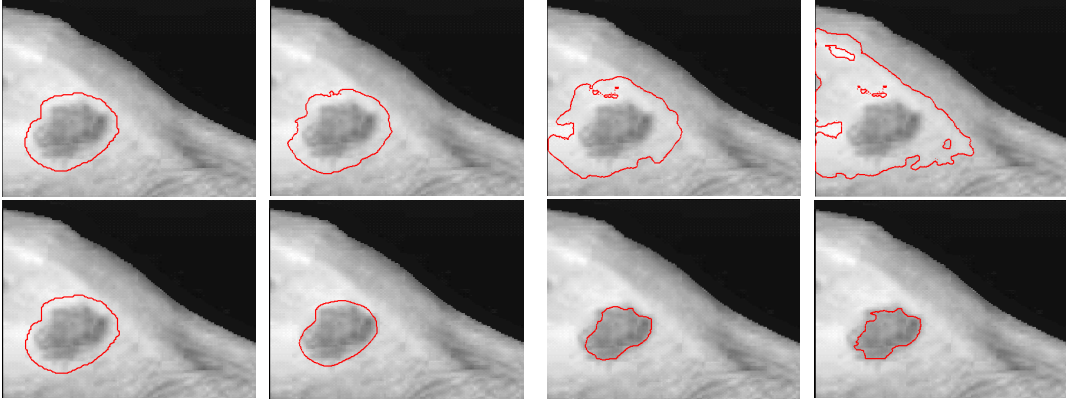
Figure 2. Case one of the Kaposi Sarcoma. Top Row: shows the evolution of a curve according to the Bhattacharyya distance resulting in an unsuccessful segmentation. Bottom Row: Successful segmentation when using the proposed similarity metric

4.1 Comparative Segmentation: Kaposi Sarcoma

Although various metrics and distributional functionals have been proposed for segmentation in the GAC framework, a qualitative comparison to demonstrate the varying behavior has not been done (to the best of our knowledge). The goal of this experiment is not to claim the ideal energy model for distinguishing between two distributions, but to add our metric with differing properties to a general class of models discriminating on probability densities. Hence, we would like show that with the same distribution, two different metrics can provide drastically different results. Note, no smoothing or regularizing term is used in these comparisons, and the results are obtained strictly on the energy describing the respective similarity measure.

In Figure 2, we demonstrate a segmentation comparison between the flow derived from the Bhattacharyya distance and the similarity metric proposed in this paper. Note, the same initialization is used, and hence, the same initial pdf's. Here, the segmentation result obtained by discriminating distributions with the Bhattacharyya distance fails to capture the infected portion of the skin. Moreover, it favors to segment an entirely different region. However, an acceptable segmentation result is obtained by evolving the curve according to Equation (12). Initial, intermediate, and final segmentation results are shown.

On a different case of the Kaposi Sarcoma, Figure 3 shows that the Bhattacharyya distance is again unable to capture the infected portion of the skin while the metric used in this paper results in a successful segmentation. From these experiments, we believe (without proof) that a major qualitative difference between the metric proposed and the Bhattacharyya distance, is the ability to capture objects under low contrast. As present in both of the Kaposi Sarcoma images, the infected portion's gray-scale intensity is not entirely different from the region surrounding it, when compared to region exterior of the initial curve. This will be a subject of future work. Several stages of the segmentation are given.

4.2 Segmentation Results: Medical Structures, Classic Zebra

In this section, we test our region-based segmentation model on several images, which further demonstrates the viability and possible advantage of our distribution metric for image segmentation. A common example that is often tested with energy models that discriminate on probability distributions is the zebra image. The goal here is to capture the entire zebra by separating the distributions so as to obtain a bimodal "Object" with a unimodal background. We note that several segmentation methods have been able to capture this image. However, for the sake of completeness, we show results in Figure 4. Stages of the segmentation are shown along with the corresponding plots of the probability distributions.

Segmenting biological structures from medical images is often a challenging task. This is due to the inherent inhomogeneous distribution of a photometric variable as well as the low contrast and noise (as seen in the Kaposi Sarcoma). In the remaining examples, we segment both the corpus callosum and an MRI image of a heart.
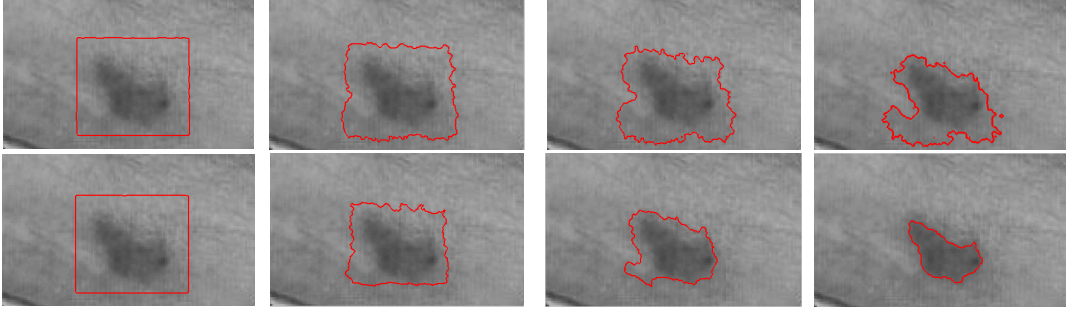
Figure 3. Case two of the Kaposi Sarcoma. Top Row: shows the evolution of a curve according to the Bhattacharyya distance resulting in an unsuccessful segmentation. Bottom Row: Successful segmentation when using the proposed similarity metric

With different types of initialization, including small seed points used for the corpus callosum, we are able to capture the finer details as shown in Figure 5. It should be noted that without discriminating on the entire probability distribution (e.g., making a Gaussian assumption), one would not be able to segment the corpus callosum. Finally, in Figure 6, we demonstrate our algorithm by segmenting an MRI image of a heart with a different type initialization compared to other initializations seen in this paper.

5. CONCLUSION

In this paper, we introduce a new metric for image segmentation that quantifies the "distance" between two distributions as the standard deviation of the difference between logarithms of those densities. While several metrics and measures have been proposed for image segmentation, results often vary drastically. Specifically, although separating distributions using the Bhattacharyya distance as a measure has resulted in the successful segmentation on some challenging imagery, in other cases such as the Kaposi Sarcoma, the Bhattacharyya - based algorithm fails to capture the infected portion of the skin while the energy model proposed in this paper results in a successful segmentation. The differing segmentation results can be traced to how the respective metrics penalize perturbations on a manifold of distributions. Thus, a subject of future work is to investigate and clarify how differences in metrics affect segmentation.

ACKNOWLEDGMENTS

REFERENCES

1. A. Kass and D. Terzopoulos, "Snakes: active contour models," *IJCV*, pp. 321–331, 1987.
2. S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Y. Jr., "Conformal curvature flows: From phase transitions to active vision," *Arch. Ration. Mech. Anal.* **134**, pp. 275–301, Sept. 1996.
3. V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.* **22**, pp. 61–79, Feb. 1997.
4. A. Blake and M. Isard, *Active Contours*, Springer, Cambridge, 1998.
5. S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, pp. 884–900, Sept. 1996.
6. D. Mumford, "A bayesian rationale for energy functionals," in *Geometry Driven Diffusion in Computer Vision*, B. Romeny, ed., *Kluwer Academic, Dordrecht*, pp. 141–153, 1994.
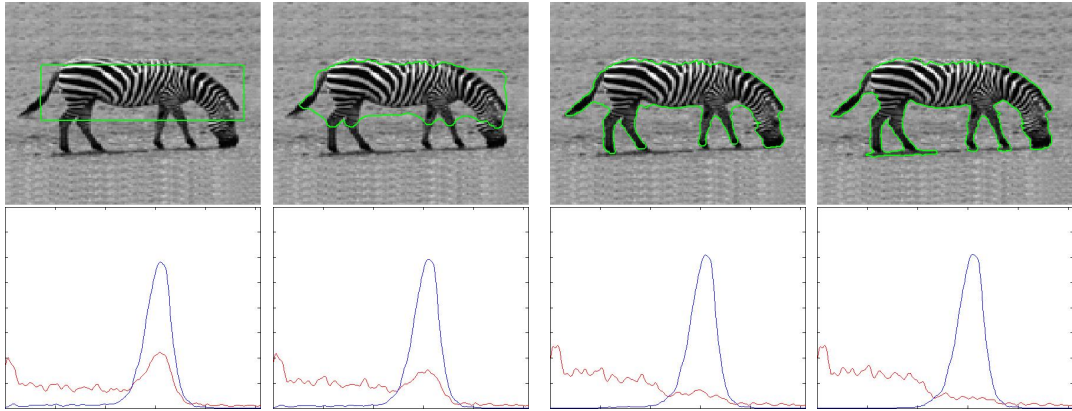
Figure 4. The Classic Zebra. Top Row: Several stages of the segmentation in which we capture the bimodal object Bottom Row: Corresponding distribution plots of both interior (red) and exterior (blue) regions if segmenting curve
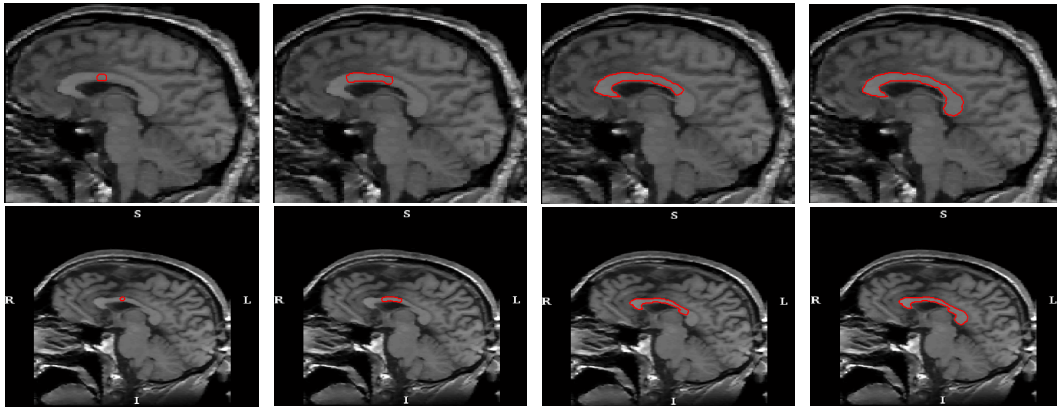

Figure 5. Corpus Callosum. Top Row: Successful Segmentation of a Corpus Callosum, a generally challenging task without discriminating on the entire pdf . Bottom Row: Successful segmentation of another case of the Corpus Callosum.
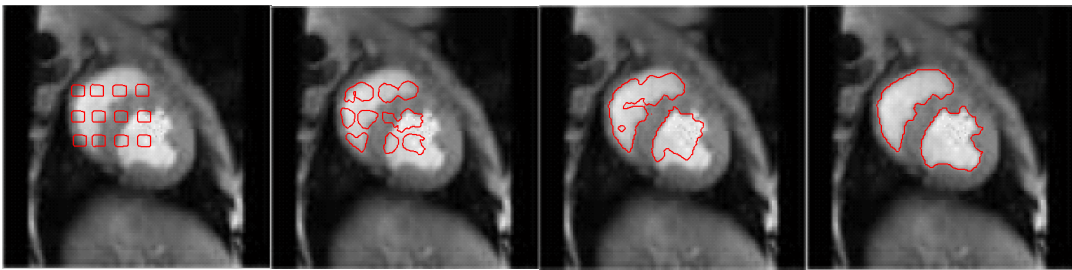

Figure 6. Successful segmentation of an MRI image of a heart using a different initialization.

7. J. Sethian, *Level Set Methods and Fast Marching Methods, Second Edition*, Springer, New York, NY, 1999.

8. S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Cambridge University Press, New York, NY, 2003.

9. T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Process.* **10**, pp. 266–277, Feb. 2001.

10. A. Yezzi, A. Tsai, and A. Willsky, "Medical image segmentation via coupled curve evolution equations with global constraints," in *Wkshp. Math. Meth. Biomed. Imag. Anal.*, pp. 12–19, 2000.

11. S. Dambreville, M. Niethammer, A. Yezzi, and A. Tannenbaum, "A variational framework combining level-sets and thresholding," in *Proceedings in British Machine Vision Conference.*, 2007.

12. S. Lankton, D. Nain, A. Yezzi, and A. Tannenbaum, "Hybrid geodesic region-based curve evolutions for image segmentation," in *Proc. SPIE: Med. Imag.*, **6510**, p. 65104U, Mar. 2007.

13. S. Dambreville, Y. Rathi, and A. Tannenbaum, "Shape-based approach to robust image segmentation using kernel pca," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 997–984, 2006.

14. D. Cremers, T. Kohlberger, and C. Schnoerr, "Shape statistics in kernel space for variational image segmentation," in *Pattern Recognition*, **36**, pp. 1292–1943, 2003.

15. M. Leventon, E. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1316–1324, 2000.

16. M. Rousson and R. Deriche, "A variational framework for active and adaptive segmentation of vector valued images," in *Wkshp. Motion Vid. Computi.*, p. 56, 2002.

17. A. Yezzi, A. Tsai, and A. Willsky, "A statistical approach to snakes for biomodal trimodal imagery," in *ICCV*, pp. 898–903, 1999.

18. J. Kim, J. Fisher, A. Yezzi, M. Cetin, and A. Willsky, "A nonparametric statistical method for image segmentation using information theory and curve evolution," *IEEE Trans. Image Process.* **14**, pp. 1486–1502, Oct. 2005.

19. T. Georgiou, "An intrinsic metric for power spectral density functions," *IEEE Trans. on Signal Processing Letters* (8), pp. 561–563, 2007.

20. D. Freedman and T. Zhang, "Active contours for tracking distributions," *IEEE Transactions on Image Processing* **13**(4), 2004.

21. Y. Rathi, O. Michailovich, J. Malcolm, and A. Tannenbaum, "Seeing the unseen: Segmenting with distributions," in *Proc. Int. Conf. Sig. Imag. Proc.*, 2006.

22. T. Georgiou, "Distances and riemannian metrics for spectral density functions," *IEEE Trans. on Signal Processing* (8), pp. 3395–4003, 2007.

23. R. Sandhu, T. Georgiou, and A. Tannenbaum, "Tracking with a new distribution metric in a particle filtering framework," in *IST and SPIE Symposium on Electronic Imaging*, 2008.

24. Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi, "Particle filtering for geometric active contours with application to tracking moving and deforming objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

25. T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

26. L. Campbell, "The relation between information theory and differential geometry approach to statistics," **35**, pp. 199–210, 1985.

27. S. Armari, "Differential-geometrical methods in statistics, lecture notes in statistics," 1985.

28. E. Hellinger, "Neaue begr undung der theorie der quadratischen formen von unendlichen vielen ver ander-lichen," **136**, pp. 210–271, 1909.

29. A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Calcutta Math. Soc.* **35**, pp. 99–110, 1943.