Available online at www.sciencedirect.com
ScienceDirect
Linear Algebra and its Applications 425 (2007) 663–672
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Distribution metrics and image segmentation ☆

Tryphon Georgiou [a,*], Oleg Michailovich [b,1], Yogesh Rathi [b], James Malcolm [b], Allen Tannenbaum [b,2]

[a] *Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, United States*
[b] *School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States*

Received 28 October 2006; accepted 4 March 2007
Available online 25 March 2007
Submitted by A.C. Antoulas

This paper is dedicated to our friend and mentor, Professor Paul Fuhrmann, on the occasion of his 70th birthday

Abstract

The purpose of this paper is to describe certain alternative metrics for quantifying distances between distributions, and to explain their use and relevance in visual tracking. Besides the theoretical interest, such metrics may be used to design filters for image segmentation, that is for solving the key visual task of separating an object from the background in an image. The segmenting curve is represented as the zero level set of a signed distance function. Most existing methods in the geometric active contour framework perform segmentation by maximizing the separation of intensity moments between the interior and the exterior of

☆ This work was supported in part by Grants from NSF, AFOSR, ARO, MURI, MRI-HEL as well as by a Grant from NIH (NAC P41 RR-13218) through Brigham and Women's Hospital. This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from `http://nihroadmap.nih.gov/bioinformatics`.

* Corresponding author.

E-mail addresses: georgiou@ece.umn.edu (T. Georgiou), olegm@ece.ualberta.ca (O. Michailovich), yogesh.rathi@gatech.edu (Y. Rathi), malcolm@ece.gatech.edu (J. Malcolm), tannenba@ece.gatech.edu (A. Tannenbaum).

[1] Present address: Department of Electrical and Computer Engineering, University of Alberta, Canada T6G 2E1.

[2] He is also with the Department of Electrical Engineering, Technion, Israel where he is supported by a Marie Curie Grant through the EU.

doi:10.1016/j.laa.2007.03.009

an evolving contour. Here one can use the given distributional metric to determine a flow which minimizes changes in the distribution inside and outside the curve.

Keywords: Distributions; Distance measures; Segmentation; Tracking

1. Introduction

We present certain natural distance measures between distributions and discuss their relevance in the context of visual tracking. These measures have their origin in different disciplines and were devised for different purposes. Yet, as it turns out, they are both quite effective in discerning differences between distributions and can be equally well applied for tracking using visual information.

Image segmentation has been a topic of extensive research in the computer vision community; see [1–6] and the references therein. In particular, geometric active contours (GAC) have been successfully used for this task. Most of these methods use photometric information such as intensity, color or texture to segment an object. In the GAC framework, an image based energy functional, typically a function of the image intensity moments, is minimized to separate an object from the background. More specifically, a closed curve C, represented implicitly as the zero level set of a signed distance function [7,8], is evolved so that it minimizes an image based energy functional. In this work, we propose to minimize a novel functional which represents the distance between the probability distribution function (pdf) inside and outside the curve C, i.e., the distance between the pdf of the object and the background. This will allow us to combine statistical and geometric information into one dynamic segmentation framework based on the theory of curve evolution.

The contents of this note is as follows. In Section 2, we review some of the key metrics for distributions, analyze their properties, and describe a possible classification. Next in Section 3, we describe some of the necessary ideas in curve evolution theory and level sets which will need for image segmentation. Accordingly, in Section 4 we describe our geometric statistical segmentation scheme, for which we perform some experiments on real data in Section 5. Finally, in Section 6 we draw some conclusions and directions for future research.

2. Metrics

We consider bounded distributions $f_i(x)$, $i = 1, 2$, with $x \in \mathscr{X} \subset \mathbb{R}^k$ and $k \in \{1, 2, 3\}$, or with $\mathscr{X}$ being a discrete set (finite). These may represent power spectral or probability densities of multi-dimensional random processes or variables. We assume that they are either integrable or summable, depending on the nature of their support set. In visual tracking these distributions typically characterize the spatial/texture profile of parts of an image. In order to identify and track objects based on visual information, one needs a natural distance measure that may capture relevant changes, at different parts of the viewing field.

Although a wide range of distance measures are available to metrize spaces of functions, when it comes to non-negative distributions, it is often of essence to acknowledge their structure as a non-negative cone. Two main alternative possibilities are appropriate and will be considered herein. They originate in statistics and prediction theory, respectively.

2.1. Fisher metric, Hellinger discrimination and Bhattacharyya distance

Let for now $\mathscr{X}$ be a discrete finite set and let

$$p_i(x) := \frac{f_i(x)}{\sum_{\mathscr{X}} f_i(x)}$$

be considered as probability densities on $\mathscr{X}$. The relative information between p_1 and p_2 ([9], see also [10])

$$S(p_1 \| p_2) := \sum_{\mathscr{X}} p_1 \log\left(\frac{p_1}{p_2}\right)$$

has been introduced as an appropriate measure for discriminating between the two. Alternatively, $S(p_1\|p_2)$ also represents the degradation of coding efficiency when selecting an optimal code based on a random source with probability distribution p_2 when the actual distribution is in fact p_1. It is well known that $X(p_1\|p_2) \geqslant 0$ and vanishes only when $p_1 = p_2$. However, it is not a metric.

The Kullback–Leibler divergence (as it is usually called in information theory) induces a Riemannian metric on the manifold of probability densities on $\mathscr{X}$. This is known as the Fisher information metric

$$S(p\|p + \Delta) \simeq g_{\mathrm{Fisher},p}(\Delta) = \sum_{\mathscr{X}} \frac{\Delta(x)^2}{p(x)} \tag{1}$$

(with $\int_{\mathscr{X}} f(x)\mathrm{d}x = 1$ and $\int_{\mathscr{X}} \Delta(x)\mathrm{d}x = 0$ since both f, $f + \Delta$ need to be probability densities). Consider the map

$$p(x) \mapsto u(x) := 2\sqrt{p(x)}. \tag{2}$$

Clearly, this maps the probability simplex (of non-negative functions on $\mathscr{X}$ which sum to 1) onto the first orthant of the sphere $\sum_{\mathscr{X}} u(x)^2 = 4$, and in fact turns out to correspond to distances defined on the probability simplex measured by the Fisher metric to Euclidean distances (e.g., see [11,12]).

Thus, it turns out that geodesics map to great circles, and distances in the Fisher metric on the probability simplex correspond to lengths of arcs on the sphere. The so-called Bhattacharyya distance [13]

$$B(p_1\|p_2) := \sum_{\mathscr{X}} \sqrt{p_1(x)p_2(x)}$$

is the cosine of the geodesic arc between the two image points under (2) of the two distributions. The length of the geodesic arc defines a metric on the probability simplex, and so does the angle

$$D(p_1, p_2) := \arccos(B(p_1\|p_2)).$$

These are equivalent. Yet another alternative equivalent metric is the Euclidean distance between the end points, which is what is known as the Hellinger discrimination [14]

$$H(p_1, p_2) := \sum_{\mathscr{X}} \left(\sqrt{p_1(x)} - \sqrt{p_2(x)}\right)^2.$$

The study of the differential structure induced by the Fisher information metric on probability simplices is the subject of Information Geometry [11]. For our purposes we explore these and their continuous counterparts

$$d(f_1, f_2) := \arccos \int_{\mathscr{X}} \sqrt{\frac{f_1}{\int f_2 \mathrm{d}x}} \sqrt{\frac{f_1}{\int f_2 \mathrm{d}x}} \mathrm{d}x \tag{3}$$

and

$$h(f_1, f_2) := \int_{\mathscr{X}} \left(\sqrt{\frac{f_1}{\int_{\mathscr{X}} f_1}} - \sqrt{\frac{f_2}{\int_{\mathscr{X}} f_2}} \right)^2. \tag{4}$$

as distances between distributions $f_i(x)$, $x \in \mathscr{X}$. In general, $d(f_1, f_2)$ defines a pseudo-metric since, due to normalization, it is insensitive to scaling.

2.2. Degradation of predictive variance and generalized means

In a manner completely analogous an alternative differential structure can be placed on distribution functions that has its origin in prediction theory. This we explain next.

Consider initially the case where $\mathscr{X} = [-\pi, \pi]$ and f_i, $i = 1, 2$, represent power spectral densities of discrete-time zero-mean stochastic processes $u_{f_i}(k)$ ($i \in \{1, 2\}$ and $k \in \mathbb{Z}$), and let $a_{f_i}(\ell)$ ($\ell \in \{1, 2, 3, \ldots\}$) represent values for the coefficients that minimize the linear prediction error variance

$$\mathscr{E} \left\{ \left| u_{f_i}(0) - \sum_{\ell=1}^{\infty} a(\ell) u_{f_i}(-\ell) \right|^2 \right\}.$$

Thus, the optimal set of coefficients depends on the power spectral density function of the process, a fact which is acknowledged by the subscript in the notation $a_{f_i}(\ell)$. It is reasonable to consider as a distance between f_1 and f_2 the degradation of predictive error variance when the coefficients $a(\ell)$ are selected assuming one of the two, and then used to predict a stochastic process corresponding to the other spectral density function.

The ratio of the “degraded” predictive error variance over the optimal error variance

$$\rho(f_1, f_2) := \frac{\mathscr{E}\{|u_{f_1}(0) - \sum_{\ell=1}^{\infty} a_{f_2}(\ell) u_{f_1}(-\ell)|^2\}}{\mathscr{E}\{|u_{f_1}(0) - \sum_{\ell=1}^{\infty} a_{f_1}(\ell) u_{f_1}(-\ell)|^2\}}$$

turns out to be equal to the ratio of the arithmetic over the geometric means of the fraction of the two spectral density functions, namely

$$\rho(f_1, f_2) = \frac{\left(\int_{\mathscr{X}} \frac{f_1(x)}{f_2(x)} \mathrm{d}x \right)}{\exp\left(\int_{\mathscr{X}} \log\left(\frac{f_1(x)}{f_2(x)} \right) \mathrm{d}x \right)},$$

see [15]. The derivation is based on classical least-variance prediction theory. Obviously, the above ratio is always greater than or equal to 1 (see also e.g., [16]). But then, the logarithm

$$\log \rho(f_1, f_2) =: \delta(f_1, f_2)$$

represents a measure of dissimilarity between the “shapes” of f_1 and f_2 and, can be viewed, as analogous to “divergences” of Information Theory (such as the Kullback–Leibler relative entropy discussed earlier). Indeed,

$$\delta(f_1, f_2) = \log \left(\int_{\mathscr{X}} \frac{f_1(x)}{f_2(x)} \mathrm{d}x \right) - \int_{\mathscr{X}} \log \left(\frac{f_1(x)}{f_2(x)} \right) \mathrm{d}x \tag{5}$$

vanishes only when f_1/f_2 is constant on $[-\pi, \pi]$ and is positive otherwise.

If we consider the distance $\delta(f, f+\Delta)$ between a nominal power spectral density f and a perturbations $f+\Delta$, and if we eliminate cubic terms and beyond, we are led (modulo a scaling factor of 2) to the Riemannian pseudo-metric

$$g_f(\Delta) := \int_{\mathscr{X}} \left(\frac{\Delta(x)}{f(x)}\right)^2 \mathrm{d}x - \left(\int_{\mathscr{X}} \frac{\Delta(x)}{f(x)} \mathrm{d}x\right)^2 \tag{6}$$

on density functions. It is a pleasant surprise to realize that, geodesic paths $f_\tau (\tau \in [0, 1])$ connecting spectral densities f_0, f_1 and having minimal length

$$\int_0^1 \sqrt{\delta(f_\tau, f_{\tau+\mathrm{d}\tau})} = \int_0^1 \sqrt{g_{f_\tau}\left(\frac{\partial f_\tau}{\partial \tau}\right)} \mathrm{d}\tau$$

can be explicitly computed [15]. These turn out to be logarithmic intervals (also referred to as exponential families),

$$f_\tau(x) = f_0^{1-\tau}(x) f_1^\tau(x) \quad \text{for } \tau \in [0, 1] \tag{7}$$

between the two extreme points. Furthermore, the length along such geodesics can be *explicitly* computed in terms of end points

$$d_{\text{geodesic}}(f_0, f_1) := \sqrt{\int_{\mathscr{X}} \left(\log \frac{f_1(x)}{f_0(x)}\right)^2 \mathrm{d}x - \left(\int_{\mathscr{X}} \log \frac{f_1(x)}{f_0(x)} \mathrm{d}x\right)^2}. \tag{8}$$

Clearly, the latter expression is the "standard-deviation" of the difference

$$\log(f_1) - \log(f_0).$$

This is a pseudo-metric, since this too does not account for constant multiplicative factors.

It is interesting to compare the differential structure on power spectral density functions with the corresponding differential structure of Information Geometry introduced earlier. Direct comparison reveals that the powers of $f(x)$ in (6) and in the Fisher information metric (1) are different. The latter, in case $\mathscr{X} \subset \mathbb{R}^k$, can be written in the form

$$g_{\text{Fisher}, f}(\Delta) = \int_{\mathscr{X}} \frac{\Delta(x)^2}{f(x)} \mathrm{d}x \tag{9}$$

(with $\int_{\mathscr{X}} f(x)\mathrm{d}x = 1$ and $\int \Delta(x)\mathrm{d}x = 0$ since both $f, f+\Delta$ need to be probability densities). Thus, it is curious and worth underscoring that in either differential structure, geodesics and geodesic lengths can be computed explicitly in analytic form. Such cases are rare indeed.

It is rather straightforward to extend the above concepts to instances where $\mathscr{X}$ is a subset of $\mathbb{R}^k$, with $k > 1$, or a discrete set. In particular, the analog of (8) is again a pseudo-metric.

3. Curve evolution and level sets

As we alluded to above, we will use the above metrics on distributions to propose novel algorithms for image segmentation and tracking. This will be based on proposing an energy functional formed from such metrics, and minimizing the function via gradient descent. Since the boundary of the object we wish to extract may be defined by a planar curve, this leads naturally to the notion of curve evolution. This will be briefly described in this section together with the corresponding level set implementations. Details about such methods may be found in [7,8].

The level set methods of Osher and Fedkiw [7] and Sethian [8] offer a natural and numerically robust implementation of such curve evolution equations. Level sets have the advantage of being parameter independent (i.e. they are implicit representations of the curve) and can handle topological changes naturally. The basic idea of the level set approach is to embed the contour C as the zero level set of a graph $\Phi : \mathbb{R}^{\nvDash} \to \mathbb{R}$, and then evolve the graph so that this level set moves according to the prescribed flow. In this manner, the level set may develop singularities and change topology while Φ itself remains smooth and maintains the form of a graph. Typically, Φ is chosen to be a signed distance function.

More precisely, let $C(s,t) : S^1 \times [0,\tau) \to \mathbb{R}^2$ be a family of curves satisfying the following evolution equation:

$$\frac{\partial C}{\partial t} = \beta \mathscr{N}, \tag{10}$$

where $\mathscr{N}$ denotes the unit inward normal. Then as we will show below, we can embed this family of curves in a two dimensional surface, and then solve the resulting equations of motion using a combination of straightforward discretization, and numerical techniques derived from hyperbolic conservation laws and Hamilton–Jacobi theory.

The embedding step is done in the following manner: The curve $C(s,t)$ is represented by the zero level set of a smooth and Lipschitz continuous function $\Phi : \mathbb{R}^2 \times [0,\tau) \to \mathbb{R}$. Assume that Φ is negative in the interior and positive in the exterior of the zero level set. (In what follows below $C(t)$ refers to the point-set defined by the parametrized curve $C(s,t)$.) We consider the zero level set, defined by

$$\{X(t) \in \mathbb{R}^2 : \Phi(X,t) = 0\}. \tag{11}$$

We have to find an evolution equation of Φ, such that the evolving curve $C(t)$ is given by the evolving zero level $X(t)$, i.e.,

$$C(t) \equiv X(t). \tag{12}$$

By differentiating $\Phi(X(t),t) = 0$ we obtain

$$\nabla\Phi(X,t) \cdot X_t + \Phi_t(X,t) = 0. \tag{13}$$

Note that for the zero level, the following relation holds:

$$\frac{\nabla\Phi}{\|\nabla\Phi\|} = \mathscr{N}. \tag{14}$$

In this equation, the left side uses terms of the surface Φ, while the right side is related to the curve C. Combining equations (10)–(14) gives

$$\Phi_t + \beta(\kappa)\|\nabla\Phi\| = 0 \tag{15}$$

and the curve C, evolving according to (10), is obtained by the zero level set of the function Φ, which evolves according to (15).

4. Image segmentation

Our main point is to explain how the earlier distances between distributions may be used in a segmentation scheme. As explained in the previous section, we consider objects as represented by closed curves enclosing their boundary. We use the level set methodology [7,8] which is based on implicit representations of contours. Thus, a curve C is represented as the zero-level set of a higher dimensional function, typically a signed distance function $\phi : \mathbb{R}^2 \to \mathbb{R}$. The advantage of this representation is that it allows for natural breaking and merging of curve topologies.

Differentiating between the interior and exterior of curves, which represent objects, can be done in a variety of ways. Typically, the intensity or color may be used to differentiate an object from its background. However, this is not always possible and one has to rely on more subtle cues. Indeed, it is often the case that a probability or spectral distribution of various photometric features that are important. Thus, the idea is to map the interior of a curve into a distribution function P_{in} and, similarly, map the exterior into a second distribution function P_{out}. These may be chosen to represent histograms of various photometric parameters as they occur in the respective domains (e.g., probability density functions p_i, $i = 1, 2$ as in Section 2.2), or may represent the spectral power of spacial variations of such photometric parameters in the two domains (e.g., spectral density functions f_i, $i = 1, 2$ as in Section 2.2). Then, the task of segmentation amounts to drawing a curve for which the distance between P_{in} and P_{out} is maximal. Similarly, the task of tracking an object amounts to re-drawing a curve in subsequent frames in such a way so that the distance between P_{in} and P_{out} remains maximal. Thus, the rôle of a distance metric between distributions is evident.

Below we explain how such distributions can be obtained in simple cases, and then how to determine the steepest-decent perturbation of a curve in the direction of maximizing the distance between P_{in} and P_{out}. This leads to a partial differential equation which performs tasks of segmentation and tracking.

Let us begin with the simplest case where $x \in \mathbb{R}^2$ specifies the coordinates in the image plane of a pixel. Let $I : \mathbb{R}^2 \to \mathscr{Z}$ be a mapping from the image plane to the space of a certain photometric variable such as intensity, a color vector, a texture vector, etc. Define P_{in} (and similarly P_{out}) via

$$P_{\text{in}}(z) = \frac{\int_\omega K(z - I(x))\mathrm{d}x}{\int_\omega \mathrm{d}x} = \frac{\int_\omega K(z - I(x))\mathrm{d}x}{A_{\text{in}}}, \tag{16}$$

which is the non-parametric kernel density estimate of the pdf of z for a given kernel K. Typical choices for K are the Dirac delta function $\delta(\cdot)$ and the Gaussian kernel given by $K(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-y^2/2\sigma^2)$. The rest of the derivation is independent of the choice of the kernel K. For the case of curve evolution, P_{in} is the density of the region inside the curve C. Thus, ω is the region enclosed by C, and A_{in} is the area of ω. Writing (16) in terms of the level set function ϕ, we get

$$P_{\text{in}}(z) = \frac{\int_\Omega K(z - I(x))H(-\phi(x))\mathrm{d}x}{\int_\Omega H(-\phi(x))\mathrm{d}x}, \tag{17}$$

where H is the Heaviside step function and Ω is the whole image domain. Similarly, $P_{\text{out}}(z)$ can be written as

$$P_{\text{out}}(z) = \frac{\int_\Omega K(z - I(x))H(\phi(x))\mathrm{d}x}{\int_\Omega H(\phi(x))\mathrm{d}x}. \tag{18}$$

A analogous formalism may be used to generate bounded distributions which represent spectral content of spacial variabilities in the two domains. Thus, in general, $z \in \mathscr{Z}$ is a vector and may include spacial frequencies as well. Then, as a distance functional

$$d(P_{\text{in}}, P_{\text{out}}), \tag{19}$$

we may select any of $B(P_{\text{int}}\|P_{\text{out}})$, $H(P_{\text{int}}\|P_{\text{out}})$, $\delta(P_{\text{int}}\|P_{\text{out}})$, or $d_{\text{geodesic}}(P_{\text{int}}\|P_{\text{out}})$ that we introduced earlier.

For concreteness, we outline the steps for the case of the Bhattacharyya distance $B(P_{\text{int}}\|P_{\text{out}})$ (following [17]). Computing the first variation of the distance functional we get

$$\nabla_\phi B = \frac{1}{2}\int_{\mathscr{Z}} (P_{\text{in}}(z)P_{\text{out}}(z))^{-1/2}\left(\frac{\partial P_{\text{in}}(z)}{\partial \phi}P_{\text{out}}(z) + P_{\text{in}}(z)\frac{\partial P_{\text{out}}(z)}{\partial \phi}\right)\mathrm{d}z,$$

where

$$\begin{aligned}\frac{\partial P_{\text{in}}(z)}{\partial \phi} &= \frac{\delta_\epsilon(\phi)}{A_{\text{in}}}(P_{\text{in}}(z) - K(z - I(x))),\\ \frac{\partial P_{\text{out}}(z)}{\partial \phi} &= \frac{\delta_\epsilon(\phi)}{A_{\text{out}}}(K(z - I(x)) - P_{\text{out}}(z))\end{aligned}$$

and $\delta_\epsilon(\phi)$ denotes a Dirac delta function at $\phi(x) = \epsilon$. Combining the above, we obtain the following partial differential equation:

$$\begin{aligned}\frac{\partial\phi(x,t)}{\partial t} = &-\frac{B\delta_\epsilon(\phi)}{2}\left(\frac{1}{A_{\text{in}}} - \frac{1}{A_{\text{out}}}\right)\\ &-\frac{\delta_\epsilon(\phi)}{2}\int_{\mathscr{Z}} K(z - I(x))\left(\frac{1}{A_{\text{out}}}\sqrt{\frac{P_{\text{in}}(z)}{P_{\text{out}}(z)}} - \frac{1}{A_{\text{in}}}\sqrt{\frac{P_{\text{out}}(z)}{P_{\text{in}}(z)}}\right)\mathrm{d}z.\end{aligned} \tag{20}$$

Throughout, A_{in} and A_{out} signify the areas inside and outside the curve, respectively. The first term in the above partial differential equation determines the "global" direction in which the entire curve moves, whereas the second term determines the "local" evolution direction.

5. Experiments

To test the feasibility of our distribution distance approach to image segmentation, we applied the flows to some medical imagery. Specifically, we employed the Bhattacharyya statistical flow given by Eq. (20) using the Dirac delta function kernel as explained above. In both cases presented in this section, the photometric variable employed was intensity. In principle, the flow may be applied to textures, color, and in fact to any type of vector-valued imagery.

The results of our experiments are shown in Figs. 1 and 2. In both sequences, we show an initial active contour on the left hand side and the progressive evolution from left to right. In Fig. 1, the active contour extracts the left ventricle from a magnetic resonance (MR) cardiac image. In Fig. 2, we show how the contour can be used to segment the (hip) bone from a computed

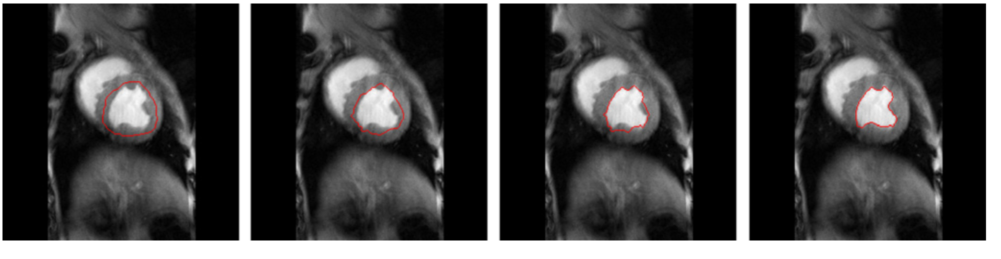

Fig. 1. Active contour capturing left ventricle in MRI image.

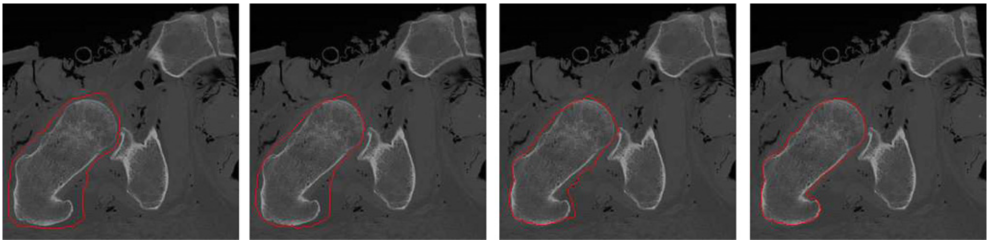

Fig. 2. Active contour capturing bone in CT image.

tomographic (CT) image. The convergence was very fast and clearly very accurate, taking less than a second on a simple Dell laptop.

6. Concluding remarks

In this work, we have described some common metrics on the space of distributions as well as the novel one from [15]. These metrics may be used to define statistically based energy functions on photometric image domains to be used to formulate an interesting class of segmentation and tracking algorithms. Indeed, one can segment an image in the geometric active contour framework by minimizing such an energy which is constructed from the distance measure between the pdf inside and outside the evolving contour C.

We demonstrated the utility of this methodology on some real images following [17]. The method is capable of separating regions even with the same mean and variance but differing only in the third and higher order moments. In future work, we will investigate and compare some of the other metrics for visual tracking as well as for segmentation in noisy and cluttered domains. We believe that by adding statistics to the geometric active contour method, one can utilize both geometric and statistical information in order to forge much more robust and versatile segmentation and tracking schemes.

References

[1] T. Zhang, D. Freedman, Tracking objects using density matching and shape priors, in: Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003, pp. 1950–1954.
[2] D. Freedman, T. Zhang, Active contours for tracking distributions, IEEE Trans. Image Process., 13 (4) (2004).
[3] S. Osher, N. Paragios, Geometric Level Set Methods in Imaging, Vision and Graphics, Springer-Verlag, 2003.
[4] N. Paragios, R. Deriche, Geodesic active contorus and level sets for the detection and tracking of moving objects, Trans. Pattern Anal. Machine Intell. 22 (3) (2000) 266–280.
[5] M. Rousson, D. Cremers, Efficient kernel density estimation of shape and intensity priors for level set segmentation, Med. Image Comput. Comput. Assist. Interven. 1 (2005) 757–764.
[6] G. Sapiro, Geometric Partial Differential Equations and Image Analysis, Cambridge University Press, 2000.
[7] S. Osher, R.P. Fedkiw, Level set methods: an overview and some recent results, J. Comput. Phys. 169 (2001) 463–502.
[8] J.A. Sethian, Level Set Methods and Fast Marching Methods, second ed., Cambridge University Press, 1999.
[9] S. Kullback, R. Leibler, On information and sufficiency, Ann. Math. Statist. 22 (1951) 79–86.
[10] T. Cover, J. Thomas, Elements of Information Theory, Wiley, 1991.
[11] S. Amari, H. Nagaoka, Translations of Mathematical Monographs, AMS, Oxford University Press, 2003.
[12] L. Campbell, The relation between information theory and the differential geometry approach to statistics, Inform. Sci. 35 (1985) 199–210.

[13] A. Bhattacharyya, Some analogues of the amount of information and their uses in statistical estimation, Sankhya 8 (1946) 1–14.
[14] E. Hellinger, Neue begr undung der theorie der quadratischen formen von unendlichen vielen ver anderlichen, J. Reine Aug. Math. 136 (1909) 210–271.
[15] T. Georgiou, Distances between power spectral densities, IEEE Trans. on Signal Process., in press.
[16] E. Beckenbach, R. Bellman, Inequalities, Springer-Verlag, 1965.
[17] A.T.O. Michailovich, Y. Rathi, Image segmentation using active contours driven by information-based criteria, IEEE Trans. Image Process., submitted for publication.