NON-PARAMETRIC INFORMATION GEOMETRY AND MULTI-SCALE MODELS OF TEXTURE

Washington Mio, Dennis Badlyans*, and Xiuwen Liu†*

* Department of Mathematics, Florida State University
† Department of Computer Science, Florida State University
Tallahassee, FL 32306, USA
e-mail: mio@math.fsu.edu, dbadlyan@math.fsu.edu, liux@cs.fsu.edu

ABSTRACT

We develop a novel algorithmic representation of textures using the statistics of multiple spectral components of images. Histograms of filter responses are viewed as elements of a non-parametric statistical manifold, and local texture patterns are compared using a geodesic metric derived from Riemannian information geometry. Several region-based image segmentation experiments are carried out to test the proposed representation and metric. This representation of textures is applied to the development of a *spectral cartoon model* of images.

1. INTRODUCTION

Since the introduction of the discrete *cartoon model* of images by Geman and Geman [6] and Blake-Zisserman [3], and its continuous analogue by Mumford and Shah [10], many variants followed and have been applied to a wide range of image processing tasks [4]. In these models, an image is typically viewed as composed of two basic elements: (i) a *cartoon* formed by regions bounded by sharp edges, within which the variation of pixel values is fairly smooth; (ii) a *texture pattern* within each region, which is frequently modeled as white noise. A drawback in such approaches is the texture model adopted; the view that texture is not noise, but some form of *"structured clutter"* is becoming prevalent. To address this problem, models such as the *spectrogram model* [9, 15] have been proposed (see also [7]), but they only capture texture properties partially. A common strategy in texture analysis has been to decompose images into their spectral components using bandpass filters and utilize histograms of filter responses to represent textures. Zhu et al. [14] have shown that marginal distributions of spectral components are sufficient to characterize homogeneous textures; other studies of the statistics of spectral components include [12, 5, 13]. Experiments reported in [8] offer empirical evidence that the same applies to non-homogeneous textures if boundary conditions are available; that is, enough pixel values near the boundary of the image domain are known.

In this paper, we model textures using histograms of spectral components viewed as elements of a non-parametric information manifold. Geodesic distances in this manifold are used to quantify texture variation and divergence. Several region-based image segmentation experiments are carried out to test the representation and metric. We also introduce a new multi-scale *spectral cartoon model* of images based on this information theoretical representation of textures. The paper is organized as follows. In Sec. 2, we give a brief description of the information manifold. Texture representation is discussed in Sec. 3, followed by experimental results in Sec. 4. The spectral cartoon model is introduced in the last section.

2. INFORMATION MANIFOLDS

We are interested in a non-parametric statistical manifold $\mathcal{P}$ formed by all positive probability density functions (PDFs) $p\colon I \to \mathbb{R}^+$ satisfying appropriate integrability conditions. Here, $I \subset \mathbb{R}$ is a fixed finite interval, which after normalization will be assumed to be $[0,1]$. The manifold $\mathcal{P}$ will be equipped with an information-theoretic geometric structure which, among other things, will allow us to quantify variations and dissimilarities of PDFs. Such infinite-dimensional statistical manifold has been constructed by Pistone and Sempi in [11].

Each tangent space $T_\varphi\mathcal{P}$ can be equipped with a natural inner product $\langle\, ,\rangle_\varphi$. Although a Hilbert-Riemannian structure might seem to be the natural geometric structure on $\mathcal{P}$ to expect, for technical reasons, one is led to a manifold locally modeled on Banach spaces. Since, in this paper, we are primarily interested in the computational aspects of information geometry, we construct finite-dimensional analogues of $\mathcal{P}$ by sampling probability density functions uniformly at a finite set of points under the assumption that they are continuous. Then, arguing heuristically, we derive an expression for the inner product on the tangent space $T_\varphi\mathcal{P}$, which induces a Riemannian structure on the finite-dimensional, non-parametric analogue of $\mathcal{P}$. Geodesic distances in these manifolds will be used to quantify divergence of PDFs. From the viewpoint of information theory, the geodesic distance can be interpreted as an intrinsic measurement of the uncertainty or unpredictability in a density function relative to another. We abuse notation and refer to both continuous and discrete models with the same symbols; however, the difference should be clear from the context.

Positive probability density functions will be represented via their log-likelihood $\varphi(x) = \log p(x)$. Thus, a function $\varphi\colon I \to \mathbb{R}$ represents an element of $\mathcal{P}$ if and only if it satisfies

$$\int_I e^{\varphi(x)}\, dx = 1\,. \tag{1}$$

Remark. In the discrete formulation, φ denotes the vector $(\varphi(x_1), \ldots, \varphi(x_n))$, where $0 = x_1 < x_2 < \ldots < x_n = 1$ are n uniformly spaced points on the interval I.

Tangent vectors to the manifold $\mathcal{P}$ at φ represent infinitesimal deformations of φ. Differentiating constraint (1) along a small path of PDFs through φ representing such deformation, it follows that a function $v\colon I \to \mathbb{R}$ represents a tangent vector at φ if and only if $\int_I v(x)e^{\varphi(x)}\, dx = 0$. This simply means that v has null expectation with respect to $e^{\varphi(x)}\,dx$. Thus, the tangent space $T_\varphi\mathcal{P}$ to the manifold $\mathcal{P}$ at φ can be described as

$$T_\varphi\mathcal{P} = \{v\colon I \to \mathbb{R}\colon \int_0^1 v(x)e^{\varphi(x)}\,dx = 0\}.$$

What is the natural inner product on $T_\varphi\mathcal{P}$ to consider? For finite-dimensional, parametric families of probability density functions, the Riemannian metric given by the positive definite quadratic form associated with the Fisher information matrix $g(\theta)$ has been studied extensively and is regarded as the most natural geometry to consider from the view point of information theory [1, 2]. Recall that if $\varphi(s;\theta)$ is the log-likelihood of a positive PDF parameterized by $\theta \in \mathbb{R}^k$, the (i,j) entry of $g(\theta)$, $1 \le i,j \le k$, is

$$g_{ij}(\theta) = \int_I \frac{\partial\varphi}{\partial\theta_i}(x;\theta)\,\frac{\partial\varphi}{\partial\theta_j}(x;\theta)\,e^{\varphi(x;\theta)}\,dx,$$

the covariance of $\partial\varphi/\partial\theta_i$ and $\partial\varphi/\partial\theta_i$ with respect to $e^{\varphi(x;\theta)}\,dx$. It is well-known that, infinitesimally, this Riemannian structure coincides with the double of the *Kullback-Leibler (KL) divergence*; that is, $KL(\theta+d\theta,\theta) = \frac{1}{2}\ {}_{i,j=1}^{k}\, g_{ij}(\theta)\,d\theta_i d\theta_j$. It can be shown that the natural extension to the non-parametric setting is the inner product on $T_\varphi\mathcal{P}$ given by

$$\langle v,w\rangle_\varphi = \int_I v(x)w(x)e^{\varphi(x)}\,dx,$$

which agrees with Fisher information on parametric submanifolds. A similar calculation with the *Jensen-Shannon (JS) entropy divergence*, which is a symmetrization of the *KL* divergence, leads to the same inner product, up to a multiplicative factor. This means that both *KL* and *JS*, which are extrinsic measures of divergence, essentially lead to the same inner product on $T_\varphi\mathcal{P}$. In the discrete formulation, we write $d(\varphi_1,\varphi_2)$ for the geodesic (or shortest path) distance between $\varphi_1,\varphi_2 \in \mathcal{P}$ with respect to this Riemannian structure. An algorithm to compute geodesics in $\mathcal{P}$ has been developed by the authors, but details will be presented elsewhere.

3. TEXTURE REPRESENTATION

Given a bank of filters $\mathcal{F} = \{F^j, 1 \le j \le K\}$ and an image I, let I^j be the associated spectral components obtained by applying filter F^j to the image. Assume that the histogram of the jth spectral component is modeled on a PDF with log-likelihood $\varphi_j \in \mathcal{P}$. The (texture of) image I will be represented by the K-tuple $\quad = (\varphi^1,\dots,\varphi^K) \in \mathcal{P}\times\ldots\times\mathcal{P} = \mathcal{P}^K$. We should point out that this is a global representation of the image I, but the same construction applied to local windows leads to *multi-scale representations* of texture patterns. If $\ _A,\ _B \in \mathcal{P}^K$ represent images I_A and I_B, respectively, let d_T be the root mean square geodesic distance

$$d_T(\ _A,\ _B) = \left(\frac{1}{K}\ {}_{j=1}^{K}\, d^2(\varphi_A^j,\varphi_B^j)\right)^{1/2}, \tag{2}$$

which defines a metric on the space $\mathcal{P}^K$ of texture representations.

Remark. In specific applications, one may wish to attribute weights to the various summands of $d_T(\ _A,\ _B)$ in order to emphasize particular filters.

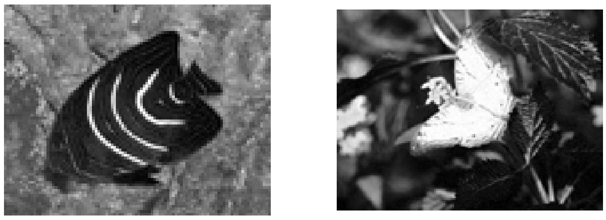

Figure 1: Original images.

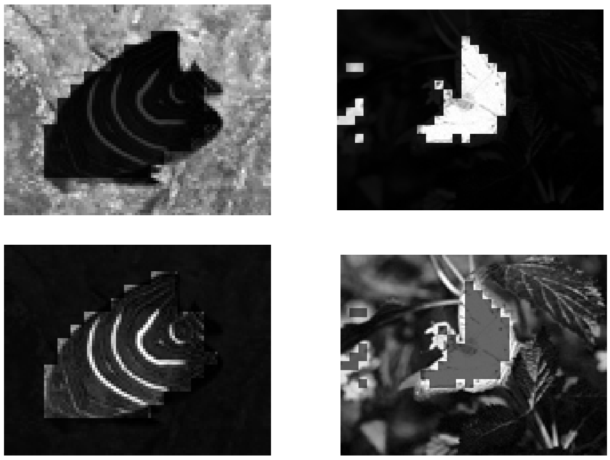

Figure 2: Regions obtained by clustering pixels into two clusters using histograms of local responses to 5 filters are highlighted in two different ways.

4. EXPERIMENTS

In this section, we discuss results obtained in image segmentation experiments with the ideas discussed above. To illustrate the ability of the metric d_T to discern and classify local texture patterns, we clustered the pixels of some images using local histograms associated with five spectral components and the metric d_T as measure of dissimilarity – a hierarchical "centroid" clustering was adopted.

In the first experiment, pixels of the images shown in Fig. 1 were grouped into two clusters. In Fig. 2, the clusters obtained are highlighted in two different ways; observe that clusters may be disconnected as in the image with a butterfly. Since clustering was performed at a low resolution, the boundaries of the regions are somewhat irregular. Note that because of the resolution and the local window size utilized in the spectral analysis, the relatively thin white stripes on the fish are clustered with the rest of the fish, not with the background.

The results of another set of experiments are displayed in Fig. 3. The original image is shown on the leftmost panel of the first row. The other images on the first row highlight the regions obtained by grouping the pixels into two clusters. The regions obtained in a similar experiment with three clusters are shown on the second row.

5. THE SPECTRAL CARTOON MODEL

To model texture patterns using *multi-resolution* spectral components of images, we localize the notion of appear-

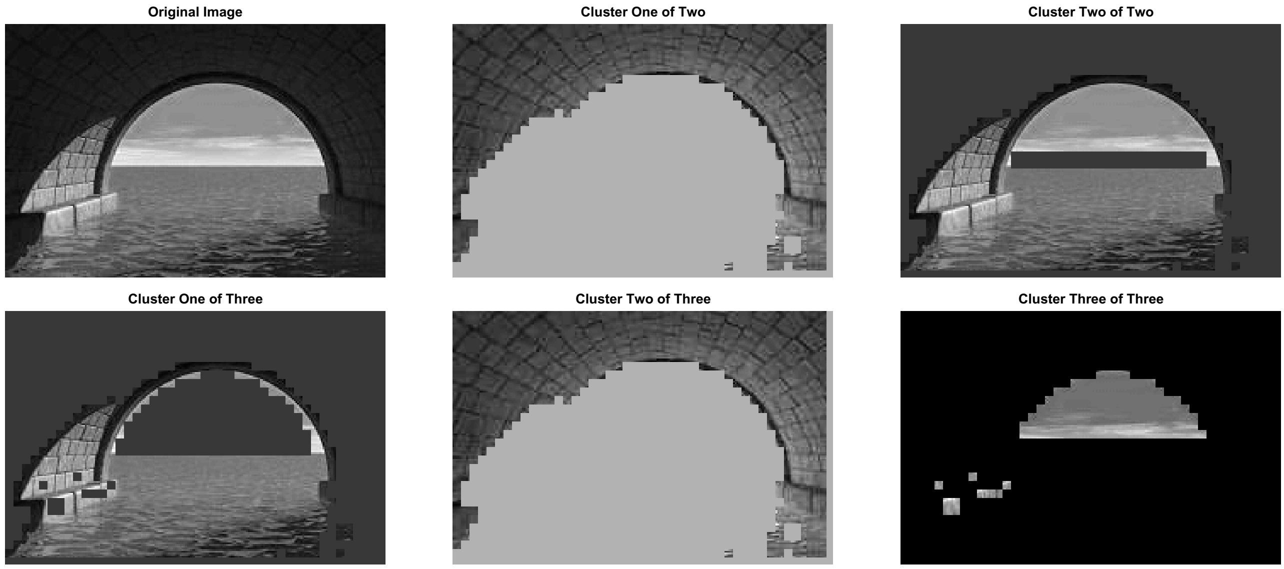

Figure 3: First row: the original image and the regions obtained by grouping the pixels into two clusters using histograms of local responses to 5 filters. Second row: results of a similar experiment with three clusters.

ance, as follows. Given a bank of filters $\mathcal{F} = \{F_1, \dots, F_K\}$ and an image I, let I^j, $1 \leq j \leq K$ be the associated spectral components. For a pixel p, consider a window of fixed size (this determines the scale) centered at p and let h_p^j be the histogram of I^j restricted to this window. After some processing (e.g., using kernel density estimators and appropriate normalizations), the histograms h_p^j yield an s-tuple $\ _p = \left(\varphi_p^1, \dots, \varphi_p^K\right) \in \mathcal{P}^K$, which encodes the local texture pattern near the pixel p. If p, q are two pixels, we use the distance $d_T(\ _p,\ _q)$ defined in (2) to quantify texture divergence.

To simplify the discussion, we consider a binary model and assume that the image consists of two main regions: a background and a single foreground element, which are separated by a closed contour C. The proposed model can be modified to allow more complex configurations as in [10]. A key difference to be noted is that unlike the classical Ising image model, where a binary cartoon is adopted (see e.g. [9]), we make a similar assumption at the level of spectral representations, so that even the cartoons can be non-trivially textured. Thus, variations of pixel values, usually treated as white noise, will be modeled on random fluctuations of more structured texture patterns.

Let $I\colon D \to \mathbb{R}$ be an image, where D is the image domain, typically a rectangle in $\mathbb{R}^2$. Consider triples $(\ _{in},\ _{out}, C)$, where C is a closed contour in D, and $\ _{in},\ _{out} \in \mathcal{P}^K$ represent models for the local texture patterns in the regions inside and outside C, respectively. We adopt a Bayesian model, with a Mumford-Shah type prior assumption that C is not "unnecessarily" long, so that the prior energy will be a multiple of the length $\ell(C)$. This can be easily modified to accommodate other commonly used terms such as the smoothness prior given by the elastic energy. The proposed data likelihood energy is of the form

$$
\begin{aligned}
E_d(I|\ _{in},\ _{out}, C) = \alpha \int_{D_{in}} d_T^2(\ _p,\ _{in})\, dp \\
+ \beta \int_{D_{out}} d_T^2(\ _p,\ _{out})\, dp\,,
\end{aligned}
\tag{3}
$$

where $\alpha, \beta > 0$, and D_{in}, D_{out} are the regions inside and outside C, respectively; the values of the parameters α, β are related to the variance of texture patterns in D_{in} and D_{out}. The idea is that E_d will measure the compatibility of local texture patterns in an image I with the texture of a proposed cartoon. The *spectral cartoon* of I is represented by the triple $(\ _{in},\ _{out}, C)$ that minimizes the posterior energy

$$
\begin{aligned}
E(\ _{in},\ _{out}, C|I) = \alpha \int_{D_{in}} d_T^2(\ _p,\ _{in})\, dp \\
+ \beta \int_{D_{out}} d_T^2(\ _p,\ _{out})\, dp + \gamma \ell(C)\,,
\end{aligned}
\tag{4}
$$

$\gamma > 0$.

Since the estimation of the triple $(\ _{in},\ _{out}, C)$ may be a costly task, we propose a simplification of the model. For a given curve C, the optimal $\ _{in}$ can be interpreted as the average value of $\ _p$ in the region D_{in}, and the integral $\int_{D_{in}} d_T^2(\ _p,\ _{in})\, dp$ as the total variance of $\ _p$ in the region. We propose to replace $d^2(\ _p,\ _{in})$, the distance square to the mean, with the average distance square from $\ _p$ to $\ _q$, for $q \in D_{in}$, $q \neq p$, which is given by

$$
\frac{1}{P_{in} - 1} \sum_{\substack{q \in D_{in} \\ q \neq p}} d^2(\ _p,\ _q)\,.
$$

Here, P_{in} is the number of pixels in D_{in}. Proceeding similarly for the region outside, the task is reduced to the simpler *maximum-a-posteriori* estimation of the curve C; that is, the curve that minimizes the energy functional

$$E(C|I) = \frac{\alpha}{P_{in}-1} \quad {}_{\substack{p,q\in D_{in}\\ q\neq p}} \; d^2(\;\; _p, \;\; _q)$$

$$+ \frac{\beta}{P_{out}-1} \quad {}_{\substack{p,q\in D_{out}\\ q\neq p}} \; d^2(\;\; _p, \;\; _q) + \gamma \ell(C)\,.$$

Acknowledgment

This work was supported in part by ARO grant W911NF-04-01-0268 and NSF grant IIS-0307998.

REFERENCES

[1] S. Amari. *Differential-Geometrical Methods of Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer, Berlin, 1985.

[2] S. Amari and H. Nagaoka. *Methods of Information Geometry*. AMS and Oxford University Press, New York, 2000.

[3] A. Blake and A. Zisserman. *Visual Reconstruction*. The MIT Press, 1987.

[4] T. Chan, J. Shen, and L. Vese. Variational PDE models in image processing. *Notices Amer. Math. Soc.*, 50:14–26, 2003.

[5] D. L. Donoho and A. G. Flesia. Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics? *Network: Computation in Neural Systems*, 12(3):371–393, 2001.

[6] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[7] T. Hoffmann and J. M. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19:1–14, 1997.

[8] X. Liu and L. Cheng. Independent spectral representations of images for recognition. *J. Optical Soc. of America*, 20(7), 2003.

[9] D. Mumford. The Bayesian rationale for energy functionals. In B. Romeny, editor, *Geometry-Driven Diffusion in Computer Vision*, pages 141–153. Kluwer Academic, 1994.

[10] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[11] B. Pistone and C. Sempi. An infinite-dimensional geometric structure on the space of all probability measures equivalent to a given one. *The Annals of Statistics*, 23(5):1543–1561, 1995.

[12] J. Portilla and E. P. Simoncelli. A parametric texture model based on join statistics of complex wavelet coeeficients. *International Journal of Computer Vision*, 40(1):49–70, 2000.

[13] A. Srivastava, X. Liu, and U. Grenander. Universal analytical forms for modeling image probability. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1200–1214, September 2002.

[14] Y. N. Wu, S. C. Zhu, and X. Liu. Equivalence of Julesz ensembles and FRAME models. *International Journal of Computer Vision*, 38(3):247–265, 2000.

[15] S. C. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy (FRAME). *International Journal of Computer Vision*, 27:1–20, 1998.