M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 185
Appeared: Soc. for Information Display 1992 Int. Symp. Digest, Vol XXIII, pp 685-688, Boston, MA, May 92.

35.4: Random Field Texture Coding

Rosalind W. Picard
M.I.T. Media Lab, E15-392; 20 Ames Street; Cambridge, MA 02139; picard@media.mit.edu

Abstract

Random field models are able to synthesize a large variety of complex patterns with a small number of parameters. This paper discusses the use of a Gibbs random field model as part of an image coding system. In particular, some semantic and perceptual attributes of this model are addressed.

1 Introduction

The Gibbs random field (GRF) model has been shown to produce a large variety of texture patterns [1]. Theoretically it is capable of synthesizing any pattern, including *both* random and regular textures. In practice, it models two-dimensional stochastic patterns very well with a small number of parameters.

In the next section some brief background is given on the GRF. Following that, I discuss its implementation as part of a texture coding system. In Sections 4 and 5 some perceptual and semantic issues concerning the model are addressed.

2 Background

A Gibbs random field is defined by

$$P(\mathbf{x}) = \frac{1}{Z}\exp(-\frac{1}{T}E(\mathbf{x})), \qquad (1)$$

where $\mathbf{x}$ is a vector representing the image, T is the "temperature" of the field, Z is a positive normalizing constant, and $E(\mathbf{x})$ is the Gibbs energy function. Textured image data is formed by synthesizing samples from this probability distribution.

There is some freedom in the choice of the Gibbs energy function, $E(\mathbf{x})$. Let each site s in an $M \times N$ image $\mathcal{S}$ contain a graylevel value $g \in \Lambda = \{0, 1, \ldots, n-1\}$. We say $x_s \in \Lambda$, $\forall s \in \mathcal{S}$. To each site in the image we also associate a symmetric neighborhood, $\mathcal{N}_s \subseteq \mathcal{S}$. The symmetry of the neighborhood implies that $\forall s, r \in \mathcal{S}$, $s \in \mathcal{N}_r$ if and only if $r \in \mathcal{N}_s$.

The Gibbs energy can now be defined by specifying interactions between sites in the image. In most of the image processing and computer vision literature, the Gibbs energy has been defined as the following sum:

$$E(\mathbf{x}) = \sum_{s\in\mathcal{S}} V_s(x_s) + \sum_{s\in\mathcal{S}}\sum_{r\in\mathcal{N}_s} V_{sr}(x_s, x_r), \qquad (2)$$

where the V_s's are the "single-site potentials" and the V_{sr}'s are the "two-site potentials". The single-site potentials are also called the "external field". Note that this energy only specifies interactions between at most two pixels in an image. The different models corresponding to this form of the energy are typically called "auto-"models, after Besag [2]. An example of a Gibbs model having an energy function of this form is the auto-binomial model used by Cross and Jain [1]. The homogeneous auto-binomial Gibbs energy is

$$-\sum_{s\in\mathcal{S}}\left(\alpha x_s + T\ln\binom{n-1}{x_s} + \sum_{r\in\mathcal{N}_s}\beta_r x_s x_r\right). \qquad (3)$$

This energy function has parameters α and β_r. The parameter α controls the influence of the external field, which allows one to impose structure on a pattern from an outside source (e.g. the force of gravity acting on each pixel.) The second parameter influences the "attraction" or "repulsion" between neighboring pairs of pixels in the image. This parameter is sometimes called a "bonding parameter."

3 Coding textures

Like many nonlinear systems the GRF produces complex patterns with a comparatively small number of parameters; consequently, it has potential for compressing images. Although theoretically able to take on any configuration, a given sample of the Gibbs distribution is most likely to have a stochastic appearance. Thus, it is currently easiest to apply the model to the stochastically textured parts of images. These textured regions are where the standard image coding tool, the Discrete Cosine Transform (DCT), tends to perform the worst; hence, this behavior of the GRF makes it a good candidate for pairing with a DCT.

The simplest place to begin studying the GRF for coding is in an assumed homogeneous textured region of an image. Note that the problem of image segmentation to obtain such a region is still an open one, and one which is also actively investigated using the GRF [3]. Also, truly homogeneous texture regions are not typical, complicating this assumption. The GRF can be made nonhomogeneous, but at the cost of additional parameters. (This is true for most models.)

One example of coding homogeneous texture in a real image is shown in Figure 1. In this figure two 64×64 patches of fur were removed from the original 512×512 image and replaced with samples of a second order anisotropic GRF. Only the top half of the original image is shown, to provide a closer look at the differences.

One patch was extracted from each side of the mandrill's face. For each patch, four bonding parameters were estimated at the shown image resolution. The estimation was performed using standard non-causal auto-regressive parameter estimation. The parameters estimated for the left patch were $\beta = 0.494$(E), -0.031(N), -0.003(NW), and 0.046(NE) (corresponding to eight compass directions since the neighborhood is symmetric). For the right patch they were slightly different, $\beta = 0.486$(E), -0.049(N), 0.072(NW), -0.004(NE).

These parameters were then inserted into the auto-binomial Gibbs energy. The temperature was set to a constant $T = .37$ and an initial random image configuration was chosen to have the same histogram as the extracted patch. Ten iterations of the Metropolis exchange algorithm on a periodic lattice were applied to the initial images, resulting in the synthetic fur patches[1]. The two patches were then inserted back into the image using the multiresolution spline of [4] to reduce perceptual boundary effects.

The estimation involved for the mandril fur was straightforward. However, in general this is not always the case; in particular, care must be exercised in choosing T [5]. Note that each patch was synthesized with five parameters (approximating 8192 pixels with ten parameters); on this pattern the GRF requires a couple of orders of magnitude fewer coefficients than a DCT. However, one cannot extrapolate this rate as the coding rate – one must additionally consider the overhead of modeling the image histogram, and of obtaining and representing the regions in the image corresponding to the texture. This overhead and the resulting bit rate performance of the GRF varies significantly with different images; there is not space here for a meaningful comparison. However, let us briefly consider two other issues which are as important as bit rate performance: perceptual and semantic performance.

4 Perceptual issues

It has been shown recently [6] that the second term in the Gibbs energy of (2) is equivalent to a linear combination of pairwise co-occurrence statistics (second order statistics of the graylevels). What makes this interesting from a perceptual standpoint is that co-occurrence statistics are a well-studied tool for texture analysis, and second-order statistics for perceptual analysis. Moreover, the Gibbs energy can be defined for any number of interacting pixels by including more terms in (2) corresponding to higher order statistics. Thus, the GRF is a flexible representation for studying perceptual issues within a coding model.

In Figure 2 the absolute difference is shown for the two images in Figure 1 (max value error unscaled = 204). Note that although the two images in Figure 1 are very hard to distinguish perceptually, their error in the replaced patches is significant. For modeling textures, one is generally not concerned with pixel-error; it is often desirable for two textures to be considered the same even if they differ in every position. Furthermore, the error will also tend to be stochastic, and it too may be modeled by a GRF. Thus the GRF can provide for controlled study of perceptual effects by varying its model parameters.

5 Semantic issues

As intelligent multimedia applications grow and memory becomes cheaper, it is reasonable to expect that the long-term emphasis of image coding may shift from pixel-level efficiency to "semantic" efficiency. For example, one may expect to see more applications that require the ability to search rapidly through a database of images for a "pattern that is similar to this one." The human is likely to describe the desired pattern semantically, as opposed to describing it by its DCT coefficients. Furthermore, two patterns that match semantically may not have the same DCT coefficients. In cases such as this, other features of the data are needed. In particular, it may help to represent pictures by different kinds of models – models whose parameters have a semantic interpretation for the given data.

Coming up with a universal semantic model is exceedingly difficult. What is more likely in the interim is the development of a menu of models, each with its own advantages. The GRF is one model (out of many) that could provide a more semantic description for some parts of an image, or for some classes of images. In [6] we showed that the auto-binomial GRF parameters correspond to rates of "mixing" and "separation" between different graylevels in an image. For images of fluids, or substances which behave similarly (e.g. desert sand) this model may be considered "semantic."

The rapidly growing interests in specialized databases may also be expected to turn into an important application area for semantic coding models. In such databases, there tends to be *a*

[1] This procedure is the same as in [1]; note that α can be ignored because of the choice of the synthesis method.

(a) Top half of Mandrill image, 256×512, $n = 16$ graylevels.

(b) Mandrill with two synthetic fur patches.

Figure 1: The 64×64 synthetic fur patches used in image (b) were made with a second order anisotropic GRF.

Figure 2: Absolute error between images on the previous page (black=0 error).

priori knowledge which can be exploited. For example, the GRF is known to be able to efficiently and accurately represent images of clouds [7]. If one had a database of weather cloud images, one might get excellent compression with the random field model. Even more importantly, one might be able to search for semantic features directly from the coded representations, without having to decompress all the images in the database each time there is a request to find something.

6 Summary

This paper has presented the basic Gibbs random field model for coding of stochastic textures in images. The model is shown to be very flexible with regard to perceptual tests that study statistical influences. Also, it is argued that semantic models will become increasingly important; some special cases where the GRF can be considered "semantic" are presented.

7 Acknowledgments

The author wishes to thank Alex Pentland for suggesting the Burt & Adelson multiresolution spline, and Fang Liu for her estimation code.

References

[1] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE T. Patt. Analy. and Mach. Intell.*, PAMI-5(1):25–39, 1983.

[2] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Stat. Soc., Ser. B*, 36:192–236, 1974.

[3] C. Bouman and B. Liu. Multiple resolution segmentation of textured images. *IEEE T. Patt. Analy. and Mach. Intell.*, PAMI-13(2):99–113, 1991.

[4] P. J. Burt and E. H. Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions On Graphics*, 2(4):217–236, 1983.

[5] R. W. Picard. Gibbs random fields: Temperature and parameter analysis. In *Proc. ICASSP*, pages III.45–III.48, San Francisco, CA, 1992.

[6] R. W. Picard, I. M. Elfadel, and A. P. Pentland. Markov/Gibbs texture modeling: Aura matrices and temperature effects. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 371–377, Maui, HI, June 1991.

[7] L. Garand and J. A. Weinman. A structural-stochastic model for the analysis and synthesis of cloud images. *J. of Climate and Appl. Meteorology*, 25:1052–1068, 1986.