Proceedings of the IEEE Workshop on Statistical and Computational Theories of Vision, Fort Collins, CO, June, 1999. (Published on Web).

Exploring Texture Ensembles by Efficient Markov Chain Monte Carlo

Song-Chun Zhu[1], Xiuwen Liu[1], Ying Nian Wu[2]

Abstract

This article presents a mathematical definition of texture pattern–the *Julesz ensemble*, which is the maximum set of images that share identical statistics ϕ defined on infinity lattice Z^2. Then texture modeling is posed as an inverse problem: given a set of images sampled from an unknown Julesz ensemble Ω_*, we search for the sufficient and necessary statistics ϕ_* which define Ω_*. Associated with a Julesz ensemble is a probability distribution $q(\mathbf{I};\phi)$ which is uniform over the images in the ensemble and has zero probability outside. It has been proven[35] that $q(\mathbf{I};\phi)$ is the *limit distribution* of the Gibbs distribution (FRAME model or Markov random field) derived by the minimax entropy principle[37]. Therefore, not only the engineering practice of synthesizing textures by matching statistics has been put on a firm mathematical foundation, but also we are released from the burden of learning the expensive FRAME model in feature pursuit, model selection and texture synthesis. The paper also proposed an algorithm for sampling Julesz ensembles by efficient Markov chain Monte Carlo, which generates texture images by moveing along filter coefficients and significantly extends the traditional single site Gibbs sampler. The paper shall also compare four popular statistical measures in the literature, namely, moments, rectified functions, marginal histograms and joint histograms of linear filter responses in terms of their descriptive abilities. We demonstrate that a small number of bins in the marginal histograms are sufficient in capturing a variety of texture patterns, so the full joint histograms appear to be an over-fit. We illustrate our theory and algorithm by successfully synthesizing a large number of natural textures.

1. Dept. of Computer and Info. Science, Ohio State Univ., Columbus, OH 43210.
2. Dept. of Statistics, Univ. of Michigan, Ann Arbor, MI 48109.

1 Introduction and motivations

In his seminar paper in 1962[22], Julesz initiated the research on texture by proposing the following fundamental question:

> *What features and statistics are characteristic of a texture pattern, so that texture pairs that share the same features and statistics cannot be told apart by pre-attentive human visual perception?*

Julesz's quest has been the scientific theme in texture modeling and perception in the last three decades, and his question raised two major challenges. The first is in psychology and neurobiology: what *features and statistics* are the *basic elements* in human texture perception? The second lies in mathematics and statistics: given a set of consistent statistics, how do we generate random texture images with identical statistics? In a mathematical language, we should be able to explore the *ensemble of texture images* that have exactly the same statistics. The first question cannot be answered without solving the second mathematical problem. As Julesz pointed out[24], the ideal theory of texture should be somewhat similar to the theory of "trichromacy", which states that any visible color is a linear combination of three basic colors: red, green, and blue. In the texture theory, this corresponds to the search for (i) the basic texture statistics and (ii) a method for mixing the exact amount of statistics with a given recipe.

The search for features and statistics went a long way beyond the Julesz 2-gon statistics conjecture. Examples include co-occurrence matrices[19], run-length statistics[34], sizes and orientations of various textons[33], cliques in Markov random fields[8], as well as dozens of other measures. All these features have rather limited expressive powers. A quantum jump occurred in the 1980s when the Gabor filters[11], filter pyramids, and wavelets transforms[10, 29] are introduced to image representation. Another advance occurred in the 1990s when simple statistics, e.g. first and second order moments, were replaced by histograms (either marginal or joint)[20, 36, 9], which contain all the higher order moments.

The research on mathematical methods for rendering texture pairs with identical statistics has also been extensive in the literature. The earliest work were task specific, for example, the "4-disk method" by Caelli et al. (1978) was designed for rendering random texture pairs sharing 2-gon statistics[5]. In computer vision, many systematic methods have been developed. For examples, 1). Gagalowizs and Ma for matching the intensity histogram and the two-pixel auto-correlation matrix using steepest descent minimization of the sum of squared errors[13], 2). Heeger and Bergen for matching marginal histogram by pyramid collapsing[20]. 3). Steepest descent minimization of match errors of rectified functions by Anderson and Langer[2]. 4). De Bonet and Viola for matching the full joint histogram using Markov tree[9]. 5). Simoncelli and Portilla for matching various correlations through repeated projections onto the statistics constraint surfaces[32].

Despite the recent successes of these methods in texture synthesis and their computational convenience, the following three fundamental questions remain unanswered.

1. What is the mathematical definition of texture adopted in the above methods? Under what conditions is it technically sound to synthesize textures by minimizing the statistical matching errors, without explicit statistical modeling?

2. How are these texture synthesis methods related to the rigorous theories on texture modeling, for example, Markov random field models (Cross and Jain) [8] and minimax entropy models (Zhu, Wu and Mumford)[37]?

3. In practice, these methods do not guarantee a close match of statistics, nor do they intend to **sample** the ensemble of images with identical statistics. Is there a general way of designing algorithms for efficient statistics matching?

In this paper, we provide answers to problem 1 and 3, while the detailed answer to problem 2 is referred to a companion paper[35].

Firstly, we study the mathematical issues in texture modeling and feature pursuit by defining the *Julesz ensemble*. Given a set of statistics and a set of observed images of a texture pattern, the Julesz ensemble is defined as the set of all images that share the same statistics as the observed. The Julesz ensemble leads to a mathematical definition of texture on the infinite lattice. In a companion paper[35], Wu and Zhu have proven that the Julesz ensemble is equivalent to the Gibbs ensemble which is the *limit* of the FRAME model as the image lattice goes to infinity. The ensemble equivalence reveals two significant facts in texture modeling.

1. *Feature pursuit, model selection, and texture synthesis can be done effectively by sampling from the Julesz ensemble, and thus we are released from the burden of learning the expensive FRAME models.*

2. *For a large (say* 256×256 *pixels) image sampled from the Julesz ensemble, any local image patch of the image given its environment follows the Gibbs distribution (or FRAME) derived by the minimax entropy principle.*

Secondly, we propose an effective algorithm for MCMC sampling. In the traditional single site Gibbs sampler[14, 37], the MCMC transition is designed in the following way. For each sweep, a pixel is chosen at random (or with a fixed scan order), and the intensity value of the pixel is updated according to its conditional distribution given neighboring pixels. The single site Gibbs sampler can be very inefficient in that the filters used are often quite large (e.g., 32×32 pixels), so flipping one pixel at a time has very little effect on the filter responses, therefore the formation and change of local features can take a long time.

We approach the efficiency problem by designing conditional moves of filter coefficients. For a linear filter, the window function [1] spans one dimension in the image space. Thus for a set of K filters in a pyramid, which do not have to be orthogonal to each other, we have K axes at each pixel. We propose moves along these dimensions separately or jointly, and thus the proposed moves update large patches of image so that local features can be formed and changed easily.

Third, we compare four popular statistical measures in the literatures: moments, rectified functions, marginal histograms and joint histograms of the Gabor filter responses. Our experiments show that moments and rectified functions are not sufficient for texture modeling. We also investigate what are the minimum set of statistics that can describe

[1] This is called impulse response in engineering, and it is called the receptive field in Neurosciences.

texture patterns, and demonstrate that a small number of bins in the marginal histograms are sufficient in capturing a wide variety texture patterns, so the full joint histogram may be just an over-fit.

We demonstrate our theory and algorithm by synthesizing many natural textures.

The paper is organized as follows. We start with a discussion of features and statistics in section (2) which is followed by a mathematical study of texture modeling in section (3). Section (4) shows a group of experiments on texture synthesis using the Gibbs sampler. Section (5) describe the feature pursuit experiments in selecting the histogram bins, and compares a variety of statistics. Section (6) presents a generalized Gibbs sampler for efficient MCMC sampling. Finally, we conclude the paper with a discussion in section (7).

2 Image features and statistics

In this section, we shall review some important image features and statistics that have been used in texture modeling, as a first step to pursue the "trichromacy" theory for texture.

Let $\mathbf{I}$ be an image defined on a finite lattice $\Lambda \subset \mathrm{Z}^2$, where for each pixel $v = (x, y) \in \Lambda$, the intensity value at v is denoted by $\mathbf{I}(v) \in S$, where S can be a finite interval on the real line, or a finite set of quantized grey levels. We define $\Omega_\Lambda = S^{|\Lambda|}$ to be the space of all images on Λ.

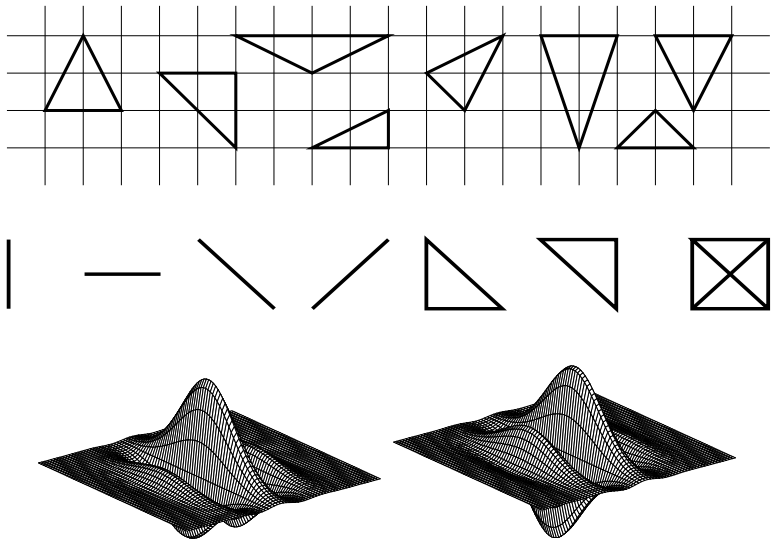


Figure 1: Choices of image features. Top row: various 3-polygons on the image lattice. Middle row: cliques in Markov random fields. Bottom row: Gabor cosine and Gabor sine filters.

In texture modeling and analysis, one starts with exploring a finite set of statistics, which are spatial averages of locally defined image features.

Figure 1 shows several important methods for extracting statistics from the image.

The first is the Julesz's k-gon statistics. A k-gon is a polygon of k pixels of a specific relative displacement, and can be indexed by $\alpha = (u_1, u_2, ..., u_k)$, which are the locations of the k pixels if we put the k-gon around the origin. Some constraints can be employed to make the index $(u_1, u_2, ..., u_k)$ unique for a k-gon. If we move this k-gon within the lattice, we will collect a set of k-tuples $\{(\mathbf{I}(v+u_1), \mathbf{I}(v+u_2), ..., \mathbf{I}(v+u_k)), v \in \Lambda\}$. The k-gon statistic is the k-dimensional histogram of these k-tuples, and it is also called the *co-occurence matrix*. To be more specific, suppose we quantize the intensity value into $1, 2, ..., g$, then the k-gon statistic collects $\{\phi^{(\alpha)}(b_1, b_2, ..., b_k; \mathbf{I}), \forall (b_1, b_2, ..., b_k) \in \{1, 2, ..., g\}^k\}$, with $\phi^{(\alpha)}(b_1, b_2, ..., b_k; \mathbf{I})$ being

$$\phi^{(\alpha)}(b_1, b_2, ..., b_k; \mathbf{I}) = \frac{1}{|\Lambda|} \sum_{v \in \Lambda} \delta(b_1 - \mathbf{I}(v + u_1)) \cdots \delta(b_k - \mathbf{I}(v + u_k)),$$

where $\delta()$ is the Dirac delta function with unit mass centered at zero. We assume that boundary conditions are properly handled (e.g., periodic boundary condition). Figure (1) displays a set of triangles in the top row. The k-gon statistics suffer from the curse of dimensionality for even a small k. For instance, for k as small as 4, and $g = 10$, the dimensionality of a 4-gon statistic is comparable to the size of the image. Summarizing the image into 4-gon statistics can hardly achieve any data reduction.

The second type of statistics is the clique statistics in Markov random fields (MRF). For a given neighborhood system of a MRF, a clique is a set of pixels that are neighbors of each other, so a clique is a special type of k-gon (see the middle row of figure 1). Let $\alpha = (u_1, u_2, ..., u_{k_\alpha})$ be the index for different types of cliques under a neighborhood system. According to the Hammersley-Clifford theorem[3], a Markov random field model has the Gibbs form

$$p(\mathbf{I}) = \frac{1}{Z} \exp\{-\sum_{\alpha} \sum_{v \in \Lambda} U_\alpha(\mathbf{I}(v + u_1), \mathbf{I}(v + u_2), ..., \mathbf{I}(v + u_{k_\alpha}))\}$$

where Z is the normalization constant or the partition function, and U_α are potential functions of k_α variables. The above Gibbs distribution can be derived from the maximum entropy principle under the constraints that $p(\mathbf{I})$ reproduces, on average, the co-occurence matrices $\phi^{(\alpha)}(b_1, ..., b_{k_\alpha}; \mathbf{I}), \forall \alpha$. *Therefore, the Gibbs model integrates all the co-occurence matrices for the cliques into a single probability distribution.* Unfortunately this general MRF model also suffers from the curse of dimensionality for even small cliques. The existing MRF texture models are much simplified so as to reduce the dimensionality of potential functions, such as the auto-binomial models[8], Gaussian MRF models[6] and ϕ-model[15].

The co-occurrence matrices (or joint intensity histograms) on the polygons and cliques are proven irrelevant to biologic vision systems. In the late 1980s, it was realized that the real world imagery is better represented on the spatial/frequency bases, such as Gabor filters[11], wavelet transforms[10], and filter pyramids. These filters are often called image *features*. Given a set of filters $\{F^{(\alpha)}, \alpha = 1, 2, ..., K\}$, a sub-band image $\mathbf{I}^{(\alpha)} = F^{(\alpha)} * \mathbf{I}(v)$ is computed for each filter.

Thus the third method in texture analysis extracts statistics on the sub-band images or pyramid instead of the original image. From a dimension reduction perspective, the

filters reduce a local image patch into a few filter responses, which can capture the spatial features of the local patch given that the filters are well tuned to the local spatial features. Because of this, very simple statistics of the sub-band images can capture information that otherwise would require k-gon or clique statistics of very high dimensions.

While the Gabor filters are well grounded on biologic vision[7], very little is known about how the visual cortices pool the statistics across the image. Figure 2 displays four popular choices of statistics in the literature.

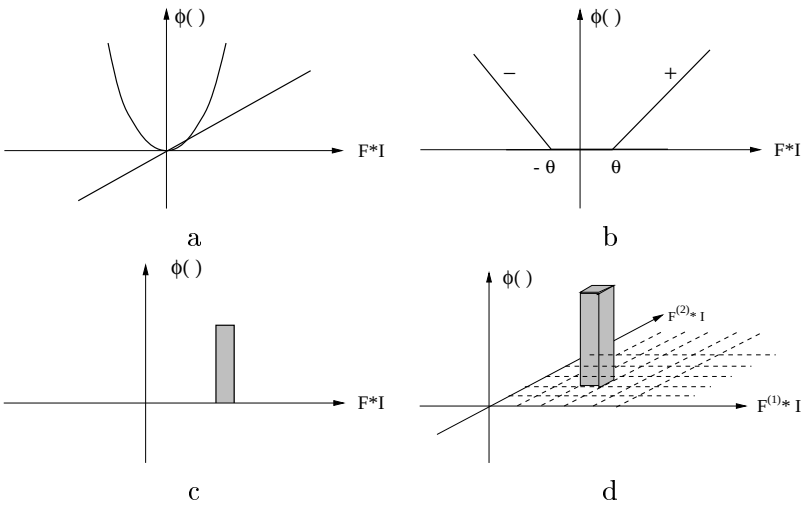


Figure 2: Four choices of statistics $\phi()$ in the feature spaces: a). Moments. b). Rectified functions for on/off cells. c). A bin of histogram (rectangle window or Gaussian window), d). a bin of the joint histogram. This figure is generalized from (Anderson etc 1996).

1. Moments of the single filter response, e.g. mean and variance of $\mathbf{I}^{(\alpha)}(v)$,

$$\begin{aligned} \phi^{(\alpha,1)}(\mathbf{I}) &= \frac{1}{|\Lambda|}\sum_{v\in\Lambda}\mathbf{I}^{(\alpha)}(v), \\ \phi^{(\alpha,2)}(\mathbf{I}) &= \frac{1}{|\Lambda|}\sum_{v\in\Lambda}(\mathbf{I}^{(\alpha)}(v)-\phi^{(\alpha,1)})^2. \end{aligned}$$

2. Rectified functions that model the responses of on/off cells[2]:

$$\begin{aligned} \phi^{(\alpha,+)}(\mathbf{I}) &= \frac{1}{|\Lambda|}\sum_{v\in\Lambda}R^{+}(\mathbf{I}^{(\alpha)}(v)), \\ \phi^{(\alpha,-)}(\mathbf{I}) &= \frac{1}{|\Lambda|}\sum_{v\in\Lambda}R^{-}(\mathbf{I}^{(\alpha)}(v)), \end{aligned}$$

where R^{+}, R^{-} are shown in figure 2b.

3. One bin in the empirical histogram of $\mathbf{I}^{(\alpha)}(v)$,

$$\phi^{(\alpha)}(b_i; \mathbf{I}) = \frac{1}{|\Lambda|} \sum_{v \in \Lambda} \delta(b_i - \mathbf{I}^{(\alpha)}), \quad \forall i, \ \forall \alpha.$$

4. One bin in the full joint histogram,

$$\phi(b_1, b_2, ..., b_k; \mathbf{I}) = \frac{1}{|\Lambda|} \sum_{v \in \Lambda} \delta(b_1 - \mathbf{I}^{(1)}(v)) \cdots \delta(b_k - \mathbf{I}^{(k)}(v)), \tag{1}$$

where $(b_1, b_2, ..., b_k)$ is the index for one bin in a k-dimensional histogram.

Perhaps the most general statistics are the co-occurence matrices (histograms) for polygons whose vertices are on the image pyramid across multiple layers, as displayed in Figure (3). For a k-gon in the image pyramid, a co-occurrence matrix (or joint histogram) can be computed,

$$\phi^{(\alpha)}(b_1, b_2, ..., b_k; \mathbf{I}) = \frac{1}{|\Lambda|} \sum_{v \in \Lambda} \delta(b_1 - \mathbf{I}^{(l_1)}(v + u_1)) \cdots \delta(b_k - \mathbf{I}^{(l_k)}(v + u_k)), \ \forall k, \forall \alpha, \tag{2}$$

where $\alpha = ((l_1, u_1), (l_2, u_2), ..., (l_k, u_k))$ is the index for the polygons in the pyramid, with l_i, u_i being respectively the index of sub-band and the displacement of the i-th vertex of the polygon. It is easy to see all the co-occurence matrices, histograms are specials case of the statistics in equation (2). For example, it reduces to the full joint histogram in equation (1) with the polygon being a straight line crossing the pyramid.

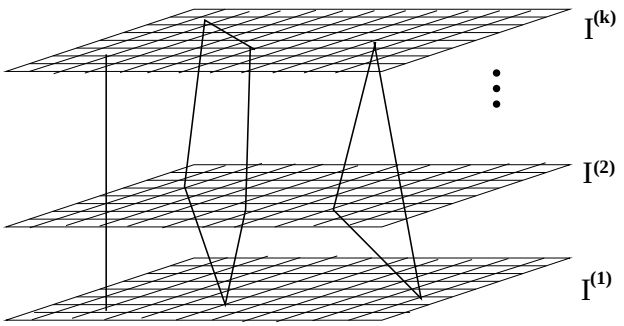


Figure 3: The more general statistics are co-occurence matrices for polygons in the image pyramid.

Obviously, the joint statistics are useful in characterizing features aligned across the spatial and frequency domain, such as edges and some elaborate details of texture elements. The question is how far one should pursue these high order statistics. The complexity of the statistics are limited by two aspects, one is the computational complexity, and the other is the statistical efficiency in estimating the model with finite data. After all, the texture model should be consistent with human texture perception.

In the literature, Heeger and Bergen made the first attempt to generate textures using marginal histograms[20], and Zhu, Wu, and Mumford studied a new class of Markov random field model that can reproduce the marginal statistics[37].[2] Recently Simoncelli, De Bonet and Viola have argued that joint statistics are needed for synthesizing some texture patterns[9, 4]. The argument is mainly based on the fact that matching the marginal statistics for the filters of steerable pyramid used by Heeger and Bergen[20] cannot reproduce some texture patterns. This argument is questionable because computationally, the Heeger and Bergen algorithm does not guarantee a close match of statistics, nor is it intended to *sample* from a texture ensemble. In theory, it is provable that marginal statistics are sufficient for reconstructing the full probability distribution of texture[37], therefore the question is about the efficiency of marginal statistics. We shall discuss the joint statistics in section (7).

3 The Julesz ensemble and MCMC sampling of texture

As a second step to answer the Julesz's quest, we should study efficient algorithms for matching statistics and for random sampling. We start with a mathematical definition of texture–the Julesz ensemble.

3.1 The Julesz ensemble - a mathematical definition of texture

Given a set of K statistics $\phi = \{\phi^{(\alpha)} : \alpha = 1, 2, ..., K\}$ which have been normalized with respect to the size of the lattice $|\Lambda|$, an image $\mathbf{I}$ is mapped into a point $\phi(\mathbf{I}) = (\phi^{(1)}(\mathbf{I}), ..., \phi^{(K)}(\mathbf{I}))$ in the *space of statistics* Ω_ϕ. Let

$$\Omega_\Lambda(\phi) = \{\mathbf{I} : \phi(\mathbf{I}) = \phi\}$$

be the set of images sharing the same statistics ϕ. Then the image space Ω_Λ is partitioned into equivalence classes with ϕ taking values in the statistics space,

$$\Omega_\Lambda = \cup_\phi \Omega_\Lambda(\phi).$$

Due to intensity quantization in finite lattice, in practice one may define the image set as

$$\Omega_\Lambda(\Phi) = \{\mathbf{I} : \phi(\mathbf{I}) \in \Phi\},$$

where $\Phi \subset \Omega_\phi$ is an open set around ϕ. Associated with each set $\Omega_\Lambda(\Phi)$ is a uniform distribution on the image space Ω_Λ.

$$q(\mathbf{I}; \Phi) = \begin{cases} \frac{1}{|\Omega_\Lambda(\Phi)|} & \text{for } \mathbf{I} \in \Omega_\Lambda(\Phi). \\ 0 & \text{otherwise.} \end{cases}$$

$|\Omega_\Lambda(\Phi)|$ is the volume of the set. It is easy to check that the entropy of this uniform model is $\log |\Omega_\Lambda(\Phi)|$.

[2] In the Zhu, Wu and Mumford framework, it is straight forward to derive Markov random field models that matches other statistics discussed in this section.

Definition *Given a set of normalized statistics* $\phi = \{\phi^{(\alpha)} : \alpha = 1, 2, ..., K\}$, *a Julesz ensemble with fixed* ϕ *is the limit of* $\Omega_\Lambda(\Phi)$ *as* $\Lambda \to \mathrm{Z}^2$ *and* $\Phi \to \phi$ *with some boundary condition.*

A Julesz ensemble is a mathematical idealization of $\Omega_\Lambda(\Phi)$ for a large Λ. We assume $\Lambda \to \mathrm{Z}^2$ in the sense of van Hove, i.e., the ratio between the boundary and the size of Λ goes to 0. As $\Lambda \to \mathrm{Z}^2$, it makes sense to let the normalized statistics $\Phi \to \phi$. Thus images in a Julesz ensemble share exactly the same normalized statistics. [3]

In the scientific meanings as expressed in the Julesz's quest, a Julesz ensemble defines a texture pattern on Z^2. Thus texture modeling is posed as an inverse problem. Suppose that we have observed a set of images $\Omega_{\text{obs}} = \{\mathbf{I}_1^{\text{obs}}, \mathbf{I}_2^{\text{obs}}, ..., \mathbf{I}_M^{\text{obs}}\}$ as samples from an unknown Julesz ensemble Ω_*, i.e. a texture pattern, and we assume that the image sizes are large enough. The goal of texture modeling is to search for the set of *sufficient* and *necessary* statistics ϕ so that $\Omega_\Lambda(\phi) = \Omega_*$.

Choosing a set of K statistics, we compute the normalized statistics over the observed images $\phi_{\text{obs}} = (\phi_{\text{obs}}^{(1)}, ..., \phi_{\text{obs}}^{(K)})$,

$$\phi_{\text{obs}}^{(\alpha)} = \frac{1}{M} \sum_{i=1}^{M} \phi^{(\alpha)}(\mathbf{I}_i^{\text{obs}}) \quad \alpha = 1, 2, ..., K. \tag{3}$$

Then we define the ensemble of texture images using the observed statistics

$$\Omega_{K,\epsilon} = \{\mathbf{I} : D(\phi^{(\alpha)}(\mathbf{I}), \phi_{\text{obs}}^{(\alpha)}) \leq \epsilon, \quad \forall \alpha\}, \tag{4}$$

where D is some distance, such as L_1 distance for histograms. Then $\Omega_{K,\epsilon} \to \Omega_K$ is a mathematical idealization as $\Lambda \to \mathrm{Z}^2, \epsilon \to 0$. Like $\Omega_\Lambda(\Phi)$, the ensemble Ω_K implies a uniform probability distribution $q(\mathbf{I}; \phi)$ over Ω_K, whose entropy is $\log |\Omega_K|$.

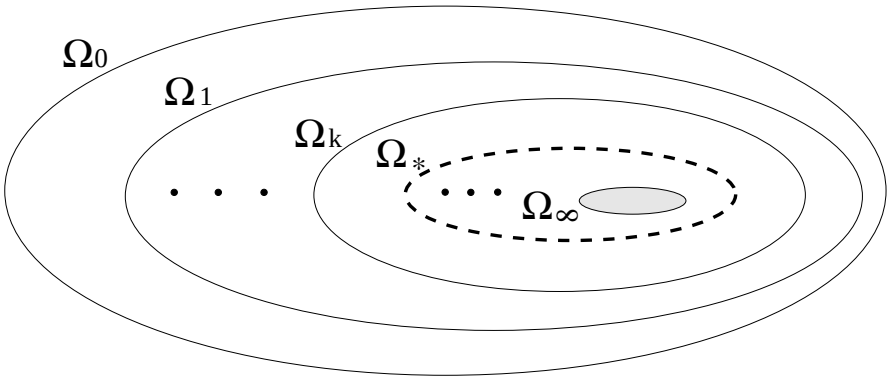


Figure 4: The volume of the Julesz ensemble decreases monotonically with more constraints added.

In search for the underlying Julesz ensemble Ω_*, one can adopt a pursuit strategy used by (Zhu, Wu, and Mumford[37]). When $K = 0$, we have $\Omega_0 = \Omega_\Lambda$. At the $k+1$st step, a

[3] We name this ensemble after Julesz to remember his pioneer work on texture. This does not necessarily mean that Julesz defined texture pattern in this mathematical formulation.

statistic $\phi^{(k+1)}$ is selected for the largest entropy decrease,

$$\phi^{(k+1)} = \arg\max_{\alpha} \Delta\text{entropy}(\phi^{(\alpha)}) = \log|\Omega_K| - \log|\Omega_{K+1}|. \tag{5}$$

We call $\Delta\text{entropy}(\phi^{(\alpha)})$ the *information gain* of $\phi^{(\alpha)}$.

As shown in figure 3.1, with more and more statistics added, the entropy or volume of the Julesz ensemble decreases monotonically

$$\Omega_\Lambda = \Omega_0 \subseteq \Omega_1 \subseteq \cdots \subseteq \Omega_K \subseteq \cdots$$

Obviously introducing too many statistics will lead to an "over-fit". In the limit, Ω_∞ is equal to Ω_{obs} subject to a translation invariant assumption.

Thus given observed finite images, the choice of a good model Ω_K is an issue of model complexity that has been extensively studied in the statistics literature. In the minimax entropy model[37, 38], an AIC criterion is adopted for model selection[1]. The intuitive idea is simple. With observed finite images, we should measure the fluctuation of a new statistics $\phi^{(\alpha)}$ over the sample. Thus when a statistic is added as constraint for the ensemble Ω_K, it brings information as well as estimation error. The feature pursuit process should stop when the estimation error brought by $\phi^{(\alpha)}$ is over its information gain.

3.2 The Gibbs ensemble and ensemble equivalence

Given a set of observed images Ω_{obs} and the statistics ϕ_{obs}, the other method for texture modeling is to pursue probability models, in particular the Gibbs distributions or MRF models.

One general class of MRF model is the minimax entropy model FRAME studied by Zhu, Wu, and Mumford[36, 37]. The FRAME model has the Gibbs form,

$$p(\mathbf{I};\boldsymbol{\beta}) = \frac{1}{Z(\boldsymbol{\beta})} \exp^{-\sum_{\alpha=1}^{K} <\beta^{(\alpha)}, \phi^{(\alpha)}(\mathbf{I})>} \tag{6}$$

where the parameters $\boldsymbol{\beta} = (\beta^{(1)}, \beta^{(2)}, ..., \beta^{(K)})$ are the Lagrange multipliers. The values of $\boldsymbol{\beta}$ are determined so that $p(\mathbf{I};\boldsymbol{\beta})$ reproduces the observed statistics,

$$E_p[\phi^{(\alpha)}(\mathbf{I})] = \phi^{(\alpha)}_{\text{obs}} \quad \alpha = 1, 2, ..., K.$$

As the image lattice becomes large enough, the large sampling effects take place, and the fluctuations of the normalized statistics diminishes at the speed of $\frac{1}{|\Lambda|}$. Thus as $\Lambda \to \mathrm{Z}^2$, the FRAME model converges to a *limiting random field* which is a uniform distribution over a *typical set* which we called the *Gibbs ensemble*, provided that there is no phase transition.

In a companion paper[35], we have proven that the Gibbs ensemble given by $p(\mathbf{I};\boldsymbol{\beta})$ is equivalent to the Julesz ensemble specified by $q(\mathbf{I};\phi)$, provided that the same set of observed statistics ϕ_{obs} is used in both models. For the equivalence between $q(\mathbf{I};\phi)$ and $p(\mathbf{I};\boldsymbol{\beta})$, the lattice has to be large enough so that large deviation takes effect. In practice, we fix $\Lambda = 256 \times 256$ pixels in our experiments, so that $\sqrt{|\Lambda|}$ is at least about 7 to 10 times larger than the biggest window size of the filters.

The ensemble equivalence reveals two significant facts in texture modeling.

1. *Feature pursuit, model selection, and texture synthesis can be done effectively by sampling from the Julesz ensemble, and thus we are released from the burden of learning the expensive FRAME models.*

2. *For a large (say* 256×256 *pixels) image sampled from the Julesz ensemble, any local image patch of the image given its environment follows the Gibbs distribution (or FRAME) derived by the minimax entropy principle.* That is, the Gibbs model $p(\mathbf{I}; \boldsymbol{\beta})$ provides a parametric form for the conditional distribution of $q(\mathbf{I}; \phi)$ for small image patches.

3.3 Sampling the Julesz ensemble

Sampling the Julesz ensemble is by no means a trivial task! As $|\Omega_k|/|\Omega_\Lambda|$ is exponentially small, the Julesz ensemble is almost invisible in the image space. The reject-sampling methods are inappropriate, thus we resort to Markov chain Monte Carlo methods.

First, we define an energy function

$$\mathcal{E}(\mathbf{I}) = \begin{cases} 0, & \text{if } D(\phi^{(\alpha)}(\mathbf{I}), \phi^{(\alpha)}_{\text{obs}}) \leq \epsilon, \quad \forall \alpha. \\ \sum_{\alpha=1}^{K} D(\phi^{(\alpha)}(\mathbf{I}), \phi^{(\alpha)}_{\text{obs}}), & \text{otherwise.} \end{cases}$$

Then the distribution

$$q(\mathbf{I}) = \frac{1}{Z_T} \exp(-\frac{\mathcal{E}(\mathbf{I})}{T}) \tag{7}$$

goes to a Julesz ensemble Ω_K, i.e. the uniform distribution over the minima of $\mathcal{E}(\mathbf{I})$, as the temperature T goes to 0. Different temperatures may be used for different α. The $q(\mathbf{I})$ can be sampled by the Gibbs sampler or other MCMC algorithms.

Algorithm I: Sampling the Julesz ensemble

Given texture images $\{\mathbf{I}^{\text{obs}}_i, \ i = 1, 2, ..., M\}$.
Given K statistics (filters) $\{F^{(1)}, F^{(2)}, ..., F^{(K)}\}$.
Compute $\{\phi^{(\alpha)}_{\text{obs}}, \quad \alpha = 1, ..., K\}$.
Initialize $\mathbf{I}^{\text{syn}}$ as any image (e.g. white noise).
$T \leftarrow T_0$.
Repeat
  Randomly pick a location $v \in \Lambda$,
  For $\mathbf{I}^{\text{syn}}(v) = 0, ..., G$ Do
    Calculate $q(\mathbf{I}^{\text{syn}}(v) \mid \mathbf{I}^{\text{syn}}(-v))$.
  Randomly flip one pixel $\mathbf{I}^{\text{syn}}(v)$ under $q(\mathbf{I}^{\text{syn}}(v) \mid \mathbf{I}^{\text{syn}}(-v))$.
  Reduce T gradually.
  Record samples when $D(\phi^{(\alpha)}(\mathbf{I}^{\text{syn}}), \phi^{(\alpha)}_{\text{obs}}) \leq \epsilon$ for $\alpha = 1, 2, ..., K$.
Until enough samples are collected.

In the above algorithm $q_T(\mathbf{I}^{\text{syn}}(v) \mid \mathbf{I}^{\text{syn}}(-\vec{v}))$ is the conditional probability of the pixel value $\mathbf{I}^{\text{syn}}(v)$ with intensities for the rest of the lattice fixed. We can pursuit the statistics in the above algorithm, as we did in [37].

3.4 About some theoretical issues

In this subsection, we briefly discuss some theoretical issues including boundary effects, equivalence of ensembles, and phase transition.

In the computation of a feature statistic $\phi^{(\alpha)}(\mathbf{I})$, we need to define boundary conditions so that the filter responses in Λ are well defined. It is very easy to show that the difference in $\phi^{(\alpha)}(\mathbf{I})$ with different boundary conditions diminishes at the speed of $\frac{|\partial\Lambda|}{|\Lambda|}$ in the Von Hove sense of converge [16]. Thus for large images, e.g. 256×256 pixels, the boundary effects for the value of $\phi^{(\alpha)}(\mathbf{I})$ are negligible.

Due to the equivalence between the Julesz ensemble and the Gibbs ensemble, the sampled images from $q(\mathbf{I}; \phi)$ and those from $p(\mathbf{I}; \boldsymbol{\beta})$ share the same statistics in that they produce not only the same statistics in ϕ, but also statistics extracted by any other filters, linear or nonlinear. Furthermore, pursuit of Julesz ensemble is also governed by the minimax entropy principle[37]. A in-depth discussion is referred to [35].

When there is phase transition, there can be more than one limiting random fields for the FRAME model, i.e., if we sample from the FRAME model for different large lattices with different boundary conditions, we will get images of different spatial patterns. If this is the case, the Julesz ensemble on Z^2 is not be unique either. We think that Gibbs ensembles or Julesz ensembles undergoing phase transition are not suitable for describing textures. Theoretical results suggest that phase transitions are exceptions rather than rules for Gibbs models with shift-invariant potential functions. See Georgii (1988) for details.

4 Experiment: sampling the Julesz ensemble

In our first set of experiments, we select all the 56 filters (Gabor filters at various scales and orientations and the small Laplacian of Gaussian filters) as discussed in [37]; the largest filter window size is 19×19 pixels. We choose ϕ to be the marginal histograms of these filters, and we sample the Julesz ensemble using algorithm I. For each texture pattern, it is often true that only a small subset of filters are necessary, but we choose to use a common filter set in this section, and we shall discuss the statistics pursuit issues in the next section. It is almost impractical to learn a FRAME model integrating all these 56 filters in our previous work[37], however the computation is much easier using the simple model $q(\mathbf{I})$.

We run the algorithm over a broad set of texture images collected from various sources. The results are displayed in Figures 5, 6, 7, 8, and 9. For these textures, the marginal statistics are matched closely (less than 1% error for each histogram) after about 50 to 100 sweeps.

These experiments show that the Gabor filters and marginal histograms are sufficient in capturing a wide variety of homogeneous texture patterns. For example, the cloth pattern in the middle row of figure 6 has very regular structures, which are reproduced fairly well in the synthesized texture image. This demonstrates that Gabor filters at various scales align up without using the joint histograms explicitly. The alignment or high order statistics are accounted for through the interaction of the filters in the model.

Figures 8 shows a periodic check board pattern and cheetah skin pattern. The synthesized check board is not strictly periodic, and this, we believe, is caused by the fast annealing process, i.e. the T in algorithm I decreases too fast, so that the long range effects did not propagate across the image before local patterns form. The synthesized cheetah skin pattern is homogeneous whereas the observed pattern is not.

Our experiments reveal two problems with algorithm I.

The first problem is demonstrated in the two failed examples in figure 9. The observed texture patterns have large structures whose periods are longer than the biggest Gabor filter windows in our filter set. As a result, these periodic patterns are scrambled in the two synthesized images.

The second problem is with the effectiveness of the Gibbs sampler. If we scale up the check board image, for example each square of the check board is of 15×15 pixel size, and we choose filters of large window sizes. Then it becomes infeasible to match the marginal statistics closely using the Gibbs sampler in algorithm I, for flipping one pixel at each time is inefficient for such large patterns. This suggests that we should search for more efficient sampling methods that can update large image patches. We believe this problem should occur for other statistics matching method, such as steepest descent[13, 2].

5 Searching for sufficient and efficient statistics

In this section, we compare various sets of statistics which we discussed in section (2).

Firstly, the full joint histogram defined in equation (1) appears to be an over-fit for most of the natural texture patterns, thus the MCMC algorithm matching the joint statistics generates texture images from a *subset* of the *true* Julesz ensemble. Our argument is based on two observations. I). Our experiments partially shown in section (4) demonstrate the sufficiency of marginal statistics. II). Suppose that one uses a modest number of filters, e.g. 10 filters, and suppose that each filter response is quantized into 10 bins, then the joint histogram has 10^{10} bins, whereas one often has only a 128×128 texture image as training data. There are far too few pixels to estimate the full joint statistics reliably.

Our conclusion about the sufficiency of marginal statistics should not be overstated. We believe that this conclusion is only valid for general texture appearance. For example, one may construct a texture pattern with hundreds of Asian faces as texture elements, thus the detailed alignment of eyes, mouths, and noses are semantically important. It seems unlikely that the marginal statistics of a small number of Gabor filters can recover these deliberate structures. Instead image features extracted by deformable face templates will become prominent. Indeed the statistics extracted by template can be considered as the general statistics extracted by some polygons across the image pyramid in equation (2). However, it seems inappropriate to call such Asian face images as texture anymore if one cares about the face expressions.

In general, we argue that it is still possible that joint statistics crossing a small number of (say $2 \sim 3$) filters at some bins (see figure 2.d), i.e., not the full joint histogram or full co-occurrence matrices, are important for some image features, especially for those texton patterns with deliberate details. Unfortunately, combinations of such statistics grow exponentially in the statistics pursuit procedure, we shall leave this topic for future research.

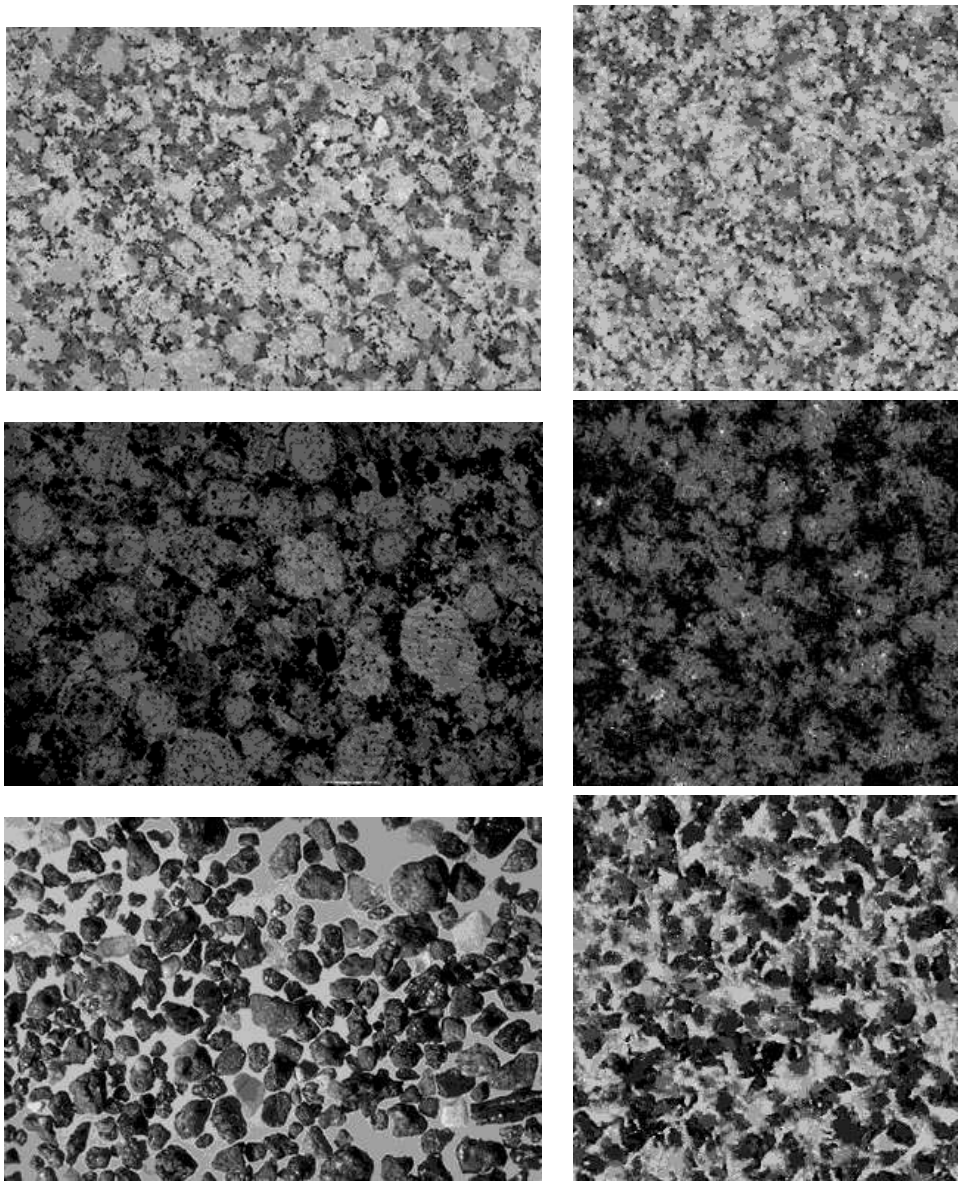


Figure 5: Left column: the observed texture images, right column: the synthesized texture images that share the exact histograms with the observed for 56 filters.

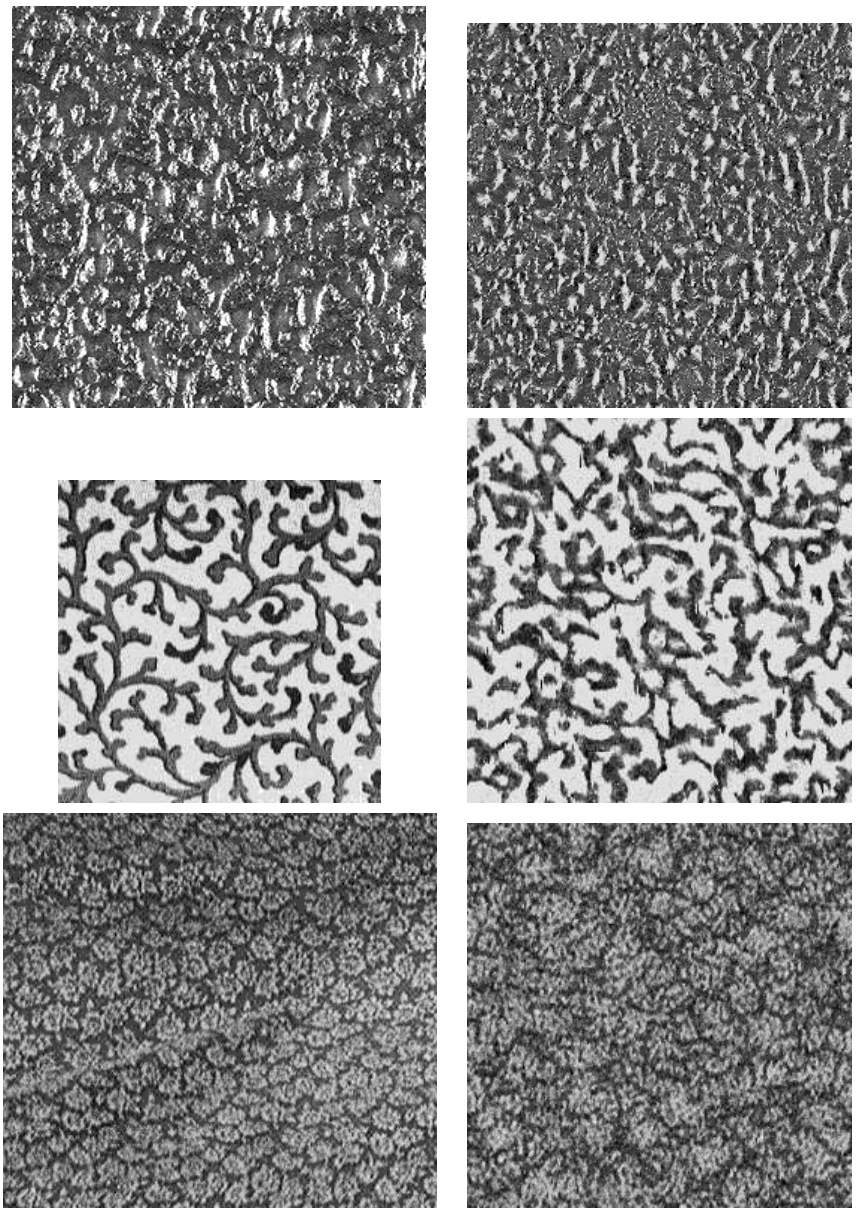


Figure 6: Left column: the observed texture images, right column: the synthesized texture images that share the exact histograms with the observed for 56 filters.

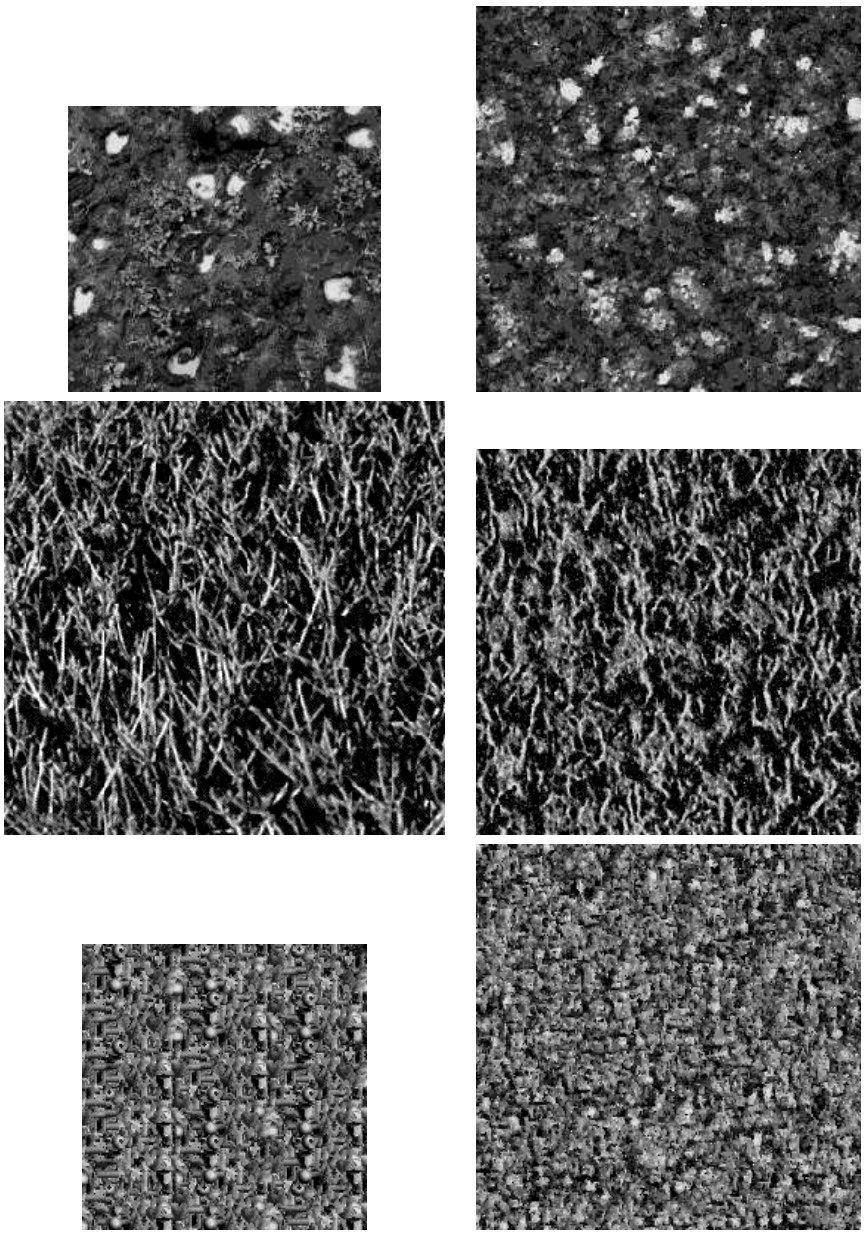


Figure 7: Left column: the observed texture images, right column: the synthesized texture images that share the exact histograms with the observed for 56 filters.

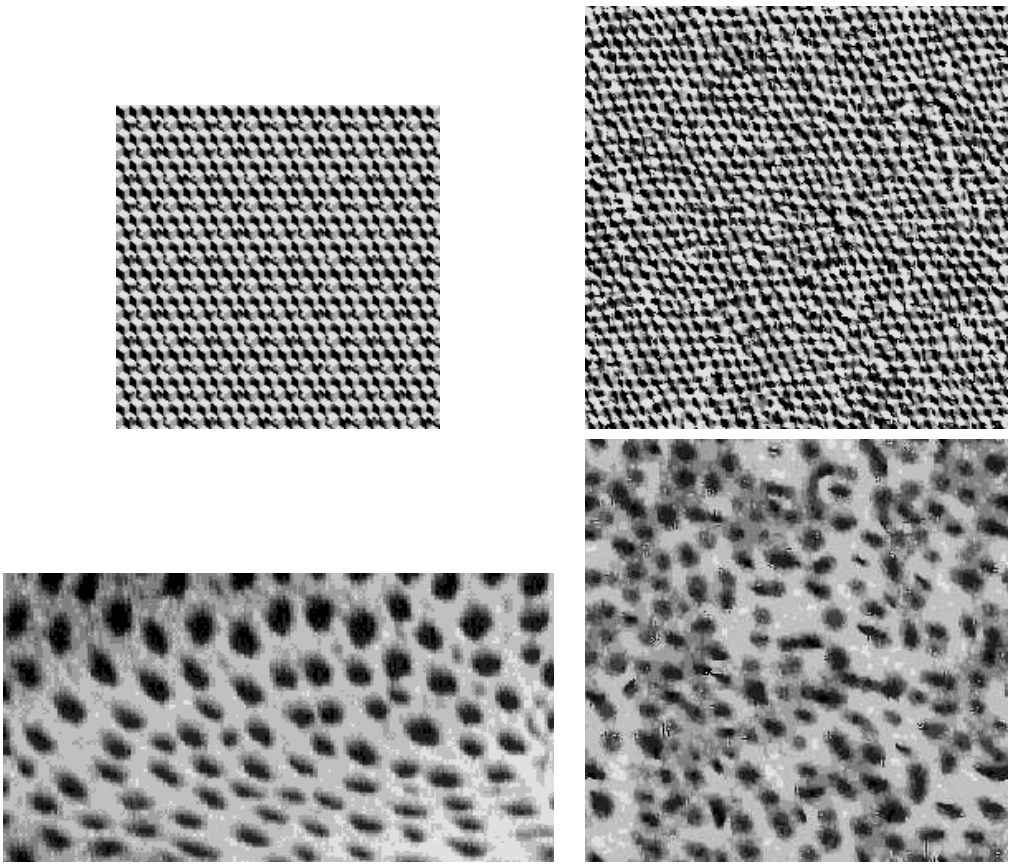


Figure 8: Left column: the observed texture images, right column: the synthesized texture images that share the exact histograms with the observed for 56 filters.

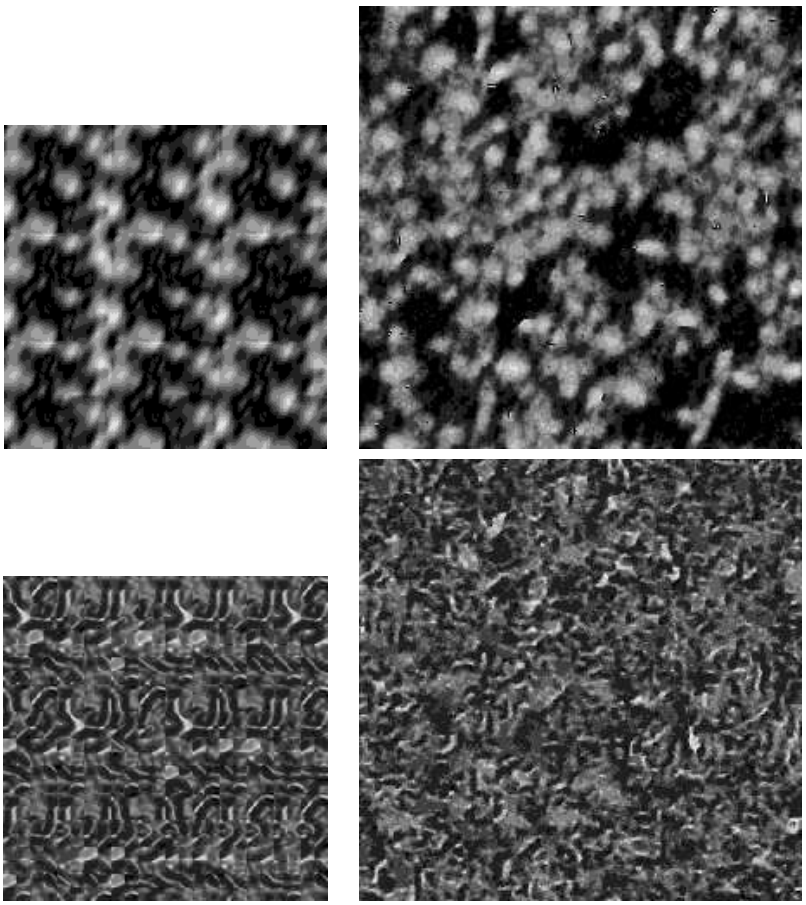
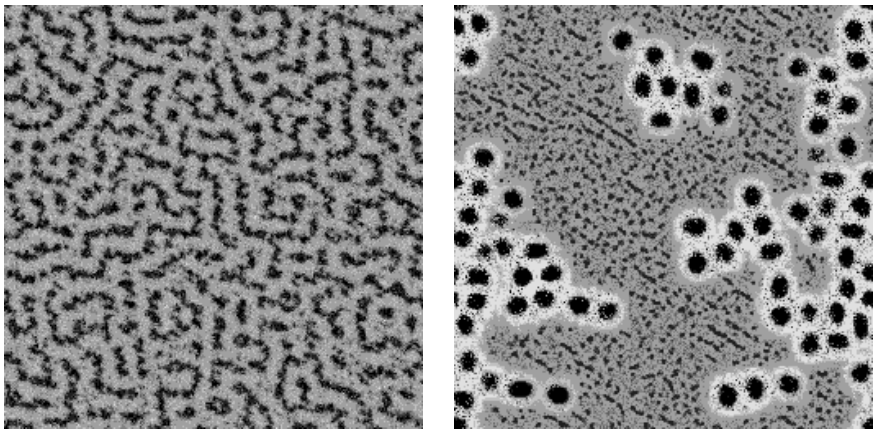


Figure 9: Left column: the observed texture images, right column: the synthesized texture images that share the exact histograms with the observed for 56 filters. The large periodic patterns are missing in the synthesized images due to the lack of large filters.

Secondly, we have computed the mean and variance for each sub-band images $\phi^{(\alpha,i)}$, $\alpha = 1, 2..., K, 1 = 1, 2$. As one may expect, the image matching the $2K$ statistics have noticeable differences from the observed ones.

Thirdly, we have also calculated the two rectified functions $\phi^{(\alpha,+)}, \phi^{(\alpha,-)}\alpha = 1, 2..., K$. Although the rectified functions emphasize the tails of the histograms, which often correspond to texture features with large filter responses, the random images from the Julesz ensemble have noticeable differences from the observed.



a b

Figure 10: Two sampled texture pattern from the Julesz ensemble matching the rectified functions of 56 filters. a). $\theta = 0.2$, b). $\theta = 0.8$.

Figure 10 displays an experiment of texture synthesis using the rectified functions. The observed texture is shown in Figure 11a, and we match $\phi^{(\alpha,+)}, \phi^{(\alpha,-)}$ for all 56 filters used before. We varied the parameter θ in R^-, R^+, so that $100 * \theta$ present of the pixels lies between $[-\theta, \theta]$. We display random examples from the Julesz ensembles with $\theta = 0.2, 0.8$ respectively.

We then proceeded to study statistics simpler than the marginal histograms. In particular, we are interested in knowing how many histogram bins are necessary for texture synthesis, and how the bins are distributed. We adopted the filter pursuit method developed by Zhu, Wu, and Mumford [37]. There are totally 56 filters, and each histogram has 11 bins except that the intensity filter has 8 bins. The algorithm selects one bin from the 605 bins each time according to an information criterion[4]. Suppose m bins have been selected, and $\mathbf{I}$ is a sample of the Julesz ensemble matching the m bins, then the information of the i-th bin of the histogram $\phi^{(\alpha)}(b_i; \mathbf{I})$ is measured by the matching error of this bin in a quadratic form,

$$d(\phi^{(\alpha)}(b_i)) = \frac{(\phi^{(\alpha)}(b_i) - \phi^{(\alpha)}_{\text{obs}}(b_i))^2}{2\sigma^2_\perp}. \tag{8}$$

[4]There is 1 redundant bin for each histogram.

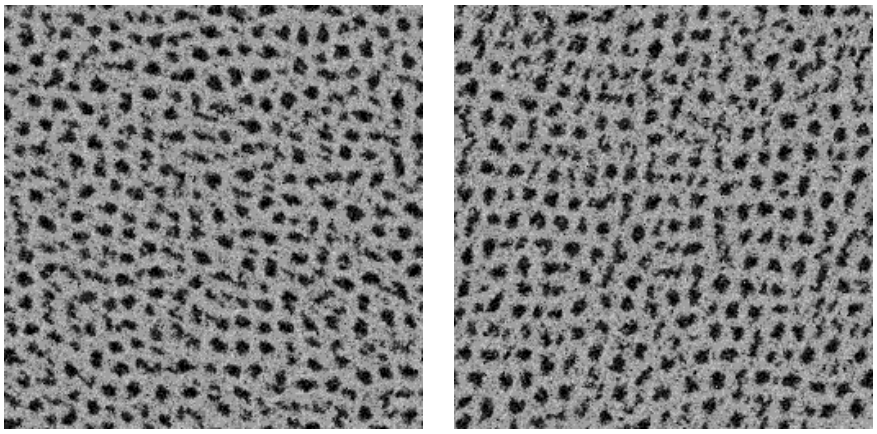


Figure 11: An example of texture synthesis by a number of bins. Left: the observed image. Right: the sampled image using 34 bins

In the above equation, $\sigma_\perp^2$ is the variance of $\phi^{(\alpha)}(b_i)$ de-correlated with the previously selected statistics. Intuitively the bigger the difference is, the more information the bin carries about the texture pattern. The $d(\phi^{(\alpha)}(b_i))$ is a second order approximation to the entropy decrease $\Delta\text{entropy}(\phi^{(\alpha)}(b_i))$ (see equation (5)) in the FRAME models in Taylor expansion. Because the entropy rate of the Julesz ensemble is the same as that of the FRAME model, $d(\phi^{(\alpha)}(b_i))$ also measures the entropy decrease in the Julesz ensemble. The details about the approximation and the computation of $\sigma_\perp^2$, is referred to [37]. Since $d(\phi^{(\alpha)}(b_i))$ is always positive, this means that the more statistics the more accurate the model is. This is not true when we only have finite observations, because we cannot estimate the observed histograms and variances exactly. Thus the information gain $d(\phi^{(\alpha)}(b_i))$ is balanced by a model complexity term that accounts for the statistical fluctuation, such as the AIC criterion discussed in [37].

Because the second order approximation is good only locally, in the first few pursuit steps, we used the L_1-norm distance $d(\phi^{(\alpha)}(b_i)) = |\phi^{(\alpha)}(b_i) - \phi^{(\alpha)}_{\text{obs}}(b_i)|$, and then use the quadratic distance after the sum of the matching errors for all 605 bins is below a certain threshold in terms of L_1-norm.

Figure 11 displays one example of bin selection. The left image is the observed image, and it was originally synthesized by matching the marginal histograms of 8 filters to a real texture pattern. We use this synthesized image as observed, for it has a ground truth in statistics selection. The right image is a sample from the Julesz ensemble using 34 bins. The selected bins for six filters are shown in figure 12. The height of each bin reflects the information gain, the higher bins are more important. The other two filters are the intensity filter (all 8 bins are chosen) and the ∇_x filter whose bin selection is very similar to ∇_y.

Our bin pursuit experiments reveal two interesting facts. First, only a small number of histogram bins are necessary for many textures. Second, the selected bins are roughly divided into two categories. One includes the bins near the center of histograms of some

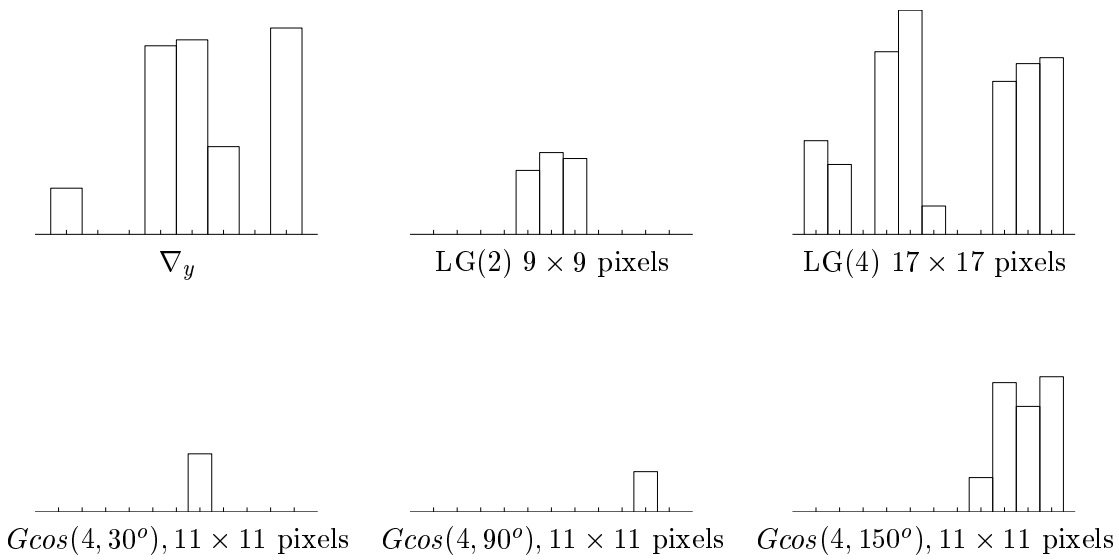


Figure 12: The selected bins for 6 filter histograms, and the height of each bin reflects the information gain.

small filters, such as the gradient filters and Laplacian of Gaussian filters. These central bins enforce "smoothness" appearances. The other bins are near the two tails of some large filters, which create image patterns. Such observations seems to confirm the design of Gibbs reaction-diffusion equations[38], where the central bins stand for diffusion effects and the tail bins are for reaction effects.

6 Designing efficient MCMC samplers.

In this section, we shall study methods for efficient Markov chain Monte Carlo sampling.

6.1 A generalized Gibbs sampler

Improving the efficiency of MCMC has been an important theme in statistics, and many strategies has been proposed in the literatures[17, 18]. However, there is no good strategies that are universally applicable, nor is there a satisfactory theory that can guide the search for better MCMC. In fact, the design of good MCMC moves often comes from heuristics of the problem domain.

In this paper, we start with an observation that the Gibbs sampler in algorithm I adopts the following proposed moves in the image space,

$$\mathbf{I} \longrightarrow \mathbf{I}' = \mathbf{I} + (\Delta - \mathbf{I}(v))\delta(v), \quad \text{for } \Delta \in \{1, 2, ..., g\}.$$

The location $v \in \Lambda$ is proposed (or visited) with a uniform distribution. $\delta(v)$ is an image matrix on Λ which has intensity one at v, and zero everywhere else. This algorithm is often called the single site Gibbs, as it moves along the dimensions of pixel intensities.

Since the Julesz ensemble is defined in the feature space through linear projections, it is natural to generalize the proposed moves to

$$\mathbf{I} \longrightarrow \mathbf{I}' = \mathbf{I} + \sum_{\alpha} (\Delta^{(\alpha)} - \mathbf{I}^{(\alpha)}(v)) W^{(\alpha)}(v)/c^{(\alpha)} \tag{9}$$

In equation (9), $W^{(\alpha)}(v)$ is a window function centered at $v \in \Lambda$ and $c^{(\alpha)}$ is a normalizing constant for the convolution of the window function with itself, i.e. $c^{(\alpha)} = W^{(\alpha)}(v) * W^{(\alpha)}(v)$.

In this paper, we choose $W^{(\alpha)}$ to be the window of filter $F^{(\alpha)}$ for convenience. In general it can be any filters, such as a Gaussian filter or a wavelet filter. In fact, if we use wavelet filters, we are actually doing a multi-resolution Gibbs sampler. In the new proposed moves, as long as the intensity filter $\delta(v)$ is chosen as one of the window function, the designed Markov chains are ergodic and aperiodic, therefore it has a unique invariant distribution on Ω_Λ. We call such generalized Gibbs sampler the *window Gibbs sampler*.

In the window Gibbs sampler, one sweep is to visit each pixel $v \in \Lambda$ once, and at each pixel, a window $W^{(\alpha)}$ is selected from a probability $q(\alpha)$. For example,

$$q(\alpha) = \frac{|\phi^{(\alpha)}(\mathbf{I}) - \phi^{(\alpha)}_{\text{obs}}|}{\sum_{\beta=1}^{k} |\phi^{(\beta)}(\mathbf{I}) - \phi^{(\beta)}_{\text{obs}}|}, \quad \alpha = 1, 2, ..., K. \tag{10}$$

The filters with large matching errors have more chances to update. Thus the designed Markov chain is inhomogeneous, as the Markov transition probability changes over time. Another choice for $q(\alpha)$ can be

$$q(\alpha) = \frac{|\log \phi^{(\alpha)}(\mathbf{I}) - \log \phi^{(\alpha)}_{\text{obs}}|}{\sum_{\beta=1}^{k} |\log \phi^{(\beta)}(\mathbf{I}) - \log \phi^{(\beta)}_{\text{obs}}|}, \tag{11}$$

which emphasizes matching the tails of the histogram $\phi^{(\alpha)}(\mathbf{I})$.

Once α is selected, the value of $\Delta^{(\alpha)}$ in equation (9) is sampled from the conditional distribution

$$\Delta^{(\alpha)} \sim p(\Delta^{(\alpha)}) = \frac{q(\mathbf{I} + (\Delta^{(\alpha)} - \mathbf{I}^{(\alpha)}(v) W^{(\alpha)}(v)/c)}{\sum_{\Delta} q(\mathbf{I} + (\Delta - \mathbf{I}^{(\alpha)}(v) W^{(\alpha)}(v)/c)}.$$

In the denominator the summation is over Δ being the center of each bin in the marginal histogram $\phi^{(\alpha)}_{\text{obs}}$.

It is interesting to see that the computational complexity for estimating $p(\Delta^{(\alpha)})$ is almost the same as computing $p(\mathbf{I}(v)|\mathbf{I}_{-v})$ in algorithm I. In the appendix, we briefly discuss the implementation details for computing the condition distributions $p(\Delta^{(\alpha)})$.

We briefly represent some results in the following section, before we discuss other ways of designing samplers.

6.2 Experiment on the window Gibbs sampler

In this section, we compare the performance of the window Gibbs sampler against the single site Gibbs sampler.

Figure 13a displays a cheetah skin texture. This image has been used in figure 8, but the image size in this experiment is as four times large as the one used before. The marginal histograms of eight filters have been chosen as the statistics in the Julesz ensemble.

Both the single site Gibbs and the window Gibbs are simulated with two initial conditions: $\mathbf{I}_u$ a uniform noise image and $\mathbf{I}_c$ a constant white image. We monitor the total matching error at each sweep of the Markov chain,

$$E = \sum_{\alpha=1}^{8} |\phi^{(\alpha)}(\mathbf{I}_{\text{syn}}) - \phi^{(\alpha)}_{\text{obs}}|_1.$$

Figure 13 displays the results for the first 100 sweeps. Figure 13b,d are the results of the single site Gibbs starting with $\mathbf{I}_u$ and $\mathbf{I}_c$ respectively. Figure 13c,e are the results of the window Gibbs starting with $\mathbf{I}_u$ and $\mathbf{I}_c$ respectively. The change of E is plotted in figure 14 against number of sweeps. The dash-dotted curve is for the single site Gibbs starting with $\mathbf{I}_u$, and the dotted curve is that for the single site Gibbs starting with $\mathbf{I}_c$. In both cases, the matching errors remain very high. The two dashed curve are for the window Gibbs starting with $\mathbf{I}_c$ and $\mathbf{I}_u$ respectively. For the latter two curves, the errors drop under 0.08. That means less than 1% error for each histogram on average. The one starting with $\mathbf{I}_u$ drops faster than the one from $\mathbf{I}_c$.

We now use another metric to demonstrate the effectiveness of the window Gibbs. Let $\mathbf{I}^n_{\text{syn}}$ be the image visited by a Markov chain (either single site Gibbs or generalized Gibbs) at sweep $n = 1, 2, ...$ We measure the Euclidean distance that the Markov chain travels during one sweep divided by the number of pixels,

$$D(n) = \sqrt{\frac{1}{|\Lambda|}\sum_{s\in\Lambda}(\mathbf{I}^{n+1}_{\text{syn}} - \mathbf{I}^n_{\text{syn}})^2}.$$

Figure 15 shows $D(n)$ for the single site Gibbs (dash-dotted and dotted curves) and the window Gibbs (solid and dashed curves) starting with $\mathbf{I}_u, \mathbf{I}_c$ respectively.

In summary, the window Gibbs algorithm outperforms the single site Gibbs in at least two aspects. 1). The window Gibbs can match statistics faster, particularly where statistics of long range features are involved. 2). The window Gibbs moves faster after statistics matching. Thus it can render texture images of difference appearance.

A rigorous theory for why the Window Gibbs is more effective has yet to be found. We only have some intuitive ideas. We believe that the window functions can be much better directions than the coordinate directions employed by the single site Gibbs for sampling $q(\mathbf{I})$ in at least two aspects. One is that it is easier to move towards a mode of $q(\mathbf{I})$ along these directions because they lead to substantial changes in local filter responses. So the local spatial features can be formed very quickly. The other aspect is that it is easier to escape from a local mode along these directions, which is important especially in the low temperature situation. Of course, it is likely that these two aspects may require the use of different window functions. Because $q(\mathbf{I})$ is too complicated for a theoretical analysis, intuitions and experiments are needed for finding good windows. It is our belief that the filters used for modeling should also be important for computation.

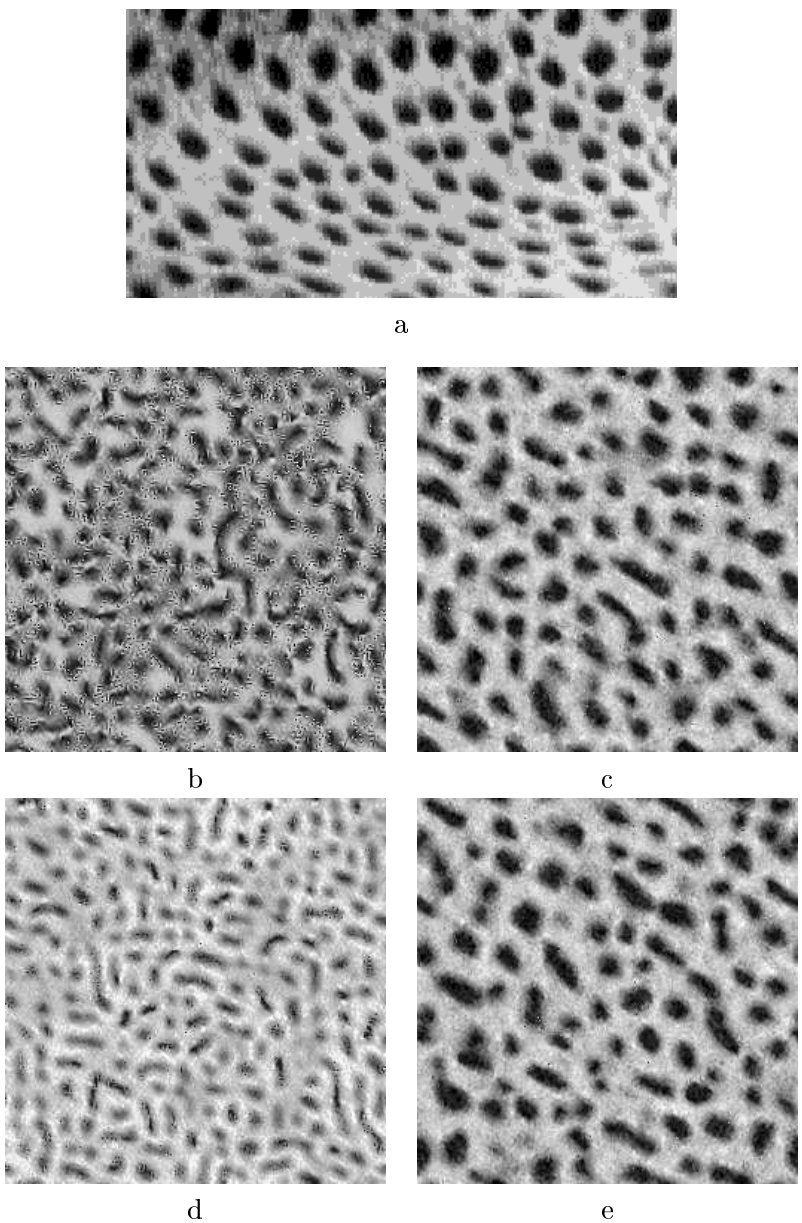


Figure 13: a) is the observed cheetah skin texture pattern, the image size of this image is as twice large as the one in figure 8. b) and d) are two sampled images using single site Gibbs sampler starting with a uniform noise image (b) and a constant image (d) respectively. c) and e) are the sampled images using the window Gibbs sampler starting with the same images as in b) and d) respectively..

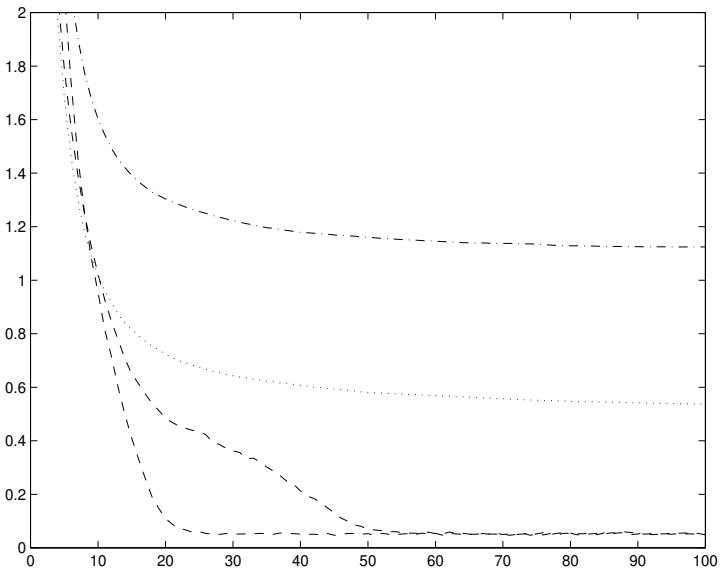


Figure 14: The statistics matching error in L_1 norm summed over 8 filters. The horizontal axis is the number of sweeps in MCMC.

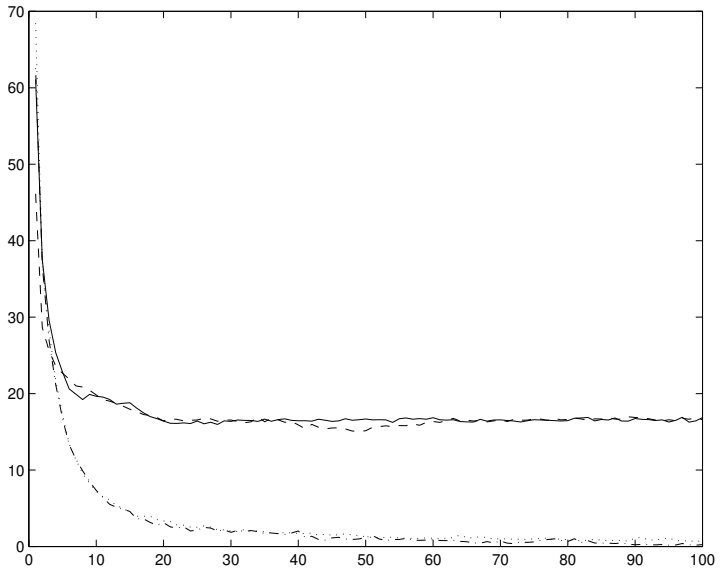


Figure 15: The averaged intensity differences by MCMC in one sweep.

6.3 Discussion on other MCMC methods

The equation (9) provides a way to design new families of Markov chains. We briefly discuss a few issues that are worth exploring in further research.

1. What are the optimal design for the directions of the Gibbs sampler? It was proven in [27] that these directions may not have to be linear, and that they can be any transformation groups.

2. In the window Gibbs, we adopted a simple heuristics, i.e. the probability $q(\alpha)$ to choose the directions of moves non-uniformly. In general, one can choose other non-uniform probabilities, and statistics other than those used in the Julesz ensemble to drive the Markov chain. For example, one may select the joint statistics (histograms) as the proposal probability $q(\Delta_1, \Delta_2, ..., \Delta_k)$ to move the coefficient jointly.

7 Conclusion

In this paper, each Julesz ensemble on Z^2 is said to be a texture pattern, and texture modeling is posed as an inverse problem. That is, given a set of images drawn from an unknown Julesz ensemble Ω_*, the goal of texture modeling is to search for the sufficient and necessary statistics that defined Ω_*. Also a rigorous mathematical foundation has been established for the engineering practice[13, 2, 32] of synthesizing textures from the Julesz ensembles, thanks for the equivalence between Julesz ensemble and Gibbs ensemble.

The remaining difficulty in texture modeling is to search for the efficient statistics that characterize the texture patterns.

Firstly, although section (2) provides a finite set of features and statistics, selecting the most efficient set of statistics within this dictionary is a tedious task. In particular, if we consider the general joint histogram bins in equation (9), we will be facing an exponential combinations. Furthermore as the filter set is often over complete, many different combinations of statistics can produce similar results.

Secondly, many texture patterns contain semantic structures presented in a hierarchic organization, as discussed in Marr's primal sketch[28]. For example, figure (16) shows three texture patterns: the hair of a woman, the twig of tree, and the zebra stripes. In these texture images, the basic texture elements and their spatial relationships are clearly perceivable. Indeed the perception of these elements plays important role for precise texture segmentation. We believe that modeling textures by filters and histograms is only a first order approximation, and further study has to be done so as to account for the hierarchic organization of the texture elements. This implies that we should look for meaning feature and statistics in a geometric hierarchy outside the current dictionary.

Acknowledgments

We would like to thank the support of an grant DAAH-04-95-1-0494 from the Army Research Office, a Microsoft research gift, and an OSU seed grant.

Figure 16: Texture patterns with flows.

Appendix: computing the conditional probability.

In this appendix, we specify the implementation details for estimating the conditional probability $p(\Delta^{(\alpha)})$ in the window Gibbs sampler.

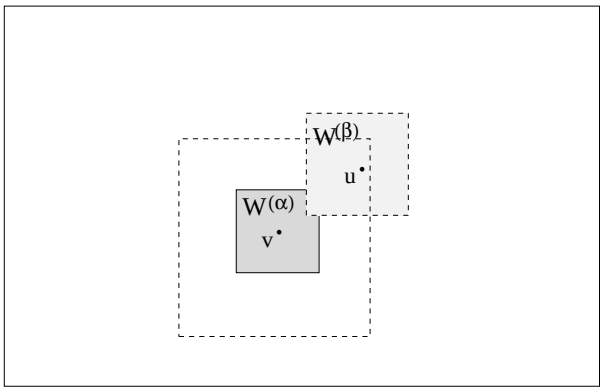


Figure 17: The computation in flipping a window function.

Suppose that at a certain step of the window Gibbs, a pixel $v \in \Lambda$ and a filter α are selected. The algorithm proposes the following moves that update an image patch covered by window $W^{(\alpha)}(v)$.

$$\mathbf{I}' = \mathbf{I} + (\Delta^{(\alpha)} - \mathbf{I}^{(\alpha)}(v))W^{(\alpha)}(v)/c^{(\alpha)} \tag{12}$$

Then to compute the probability $p(\Delta^{(\alpha)}$, we need to update the histograms $\phi^{(\beta)}$ for the sub-band image $\mathbf{I}^{\beta}, \beta = 1, 2, .., K$ with $\mathbf{I}$ being the current status of the Markov chain. As show in figure 17, the filter response $\mathbf{I}^{(\beta)}(u)$ at point $u \in \Lambda$ needs to be updated if the window $W^{(\beta)}(u)$ overlaps with $W^{(\alpha)}$.

The new filter response is

$$\mathbf{I}'^{(\beta)}(u) = \mathbf{I}^{(\beta)}(u) + ((\Delta^{(\alpha)} - \mathbf{I}^{(\alpha)}(v))W^{(\alpha)}(v) * W^{(\beta)}(u)/c^{(\alpha)}$$

Therefore, we need to compute the convolution $W^{(\alpha)}(v) * W^{(\beta)}(u)$ and save it in a table. Thus the updating of $\mathbf{I}^{(\beta)}(u)$ has $O(1)$ complexity. It is easy to check $\mathbf{I}'^{(\alpha)}(v) = \Delta^{(\alpha)}$.

References

[1] H. Akaike, "On entropy maximization principle", In *Applications of Statistics*, ed. Krishnaiah, P.R., p27-42, Amsterdam: North-Holland, 1977.

[2] C. H. Anderson and W. D. Langer, "Statistical models of image texture", *Unpublished preprint*, Washington University, St Louis, MO, 1996. (Send email to cha@shifter.wustl.edu to get a copy).

[3] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *Journal of the Royal Statistical Society, Series B,* vol 36, pp192-236, 1973.

[4] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain", TR 414, GRASP Lab. Upenn, 1997.

[5] T. Caelli, B. Julesz, and E. Gilbert, "On perceptual analyzers underlying visual texture discrimination: Part II," *Biological Cybernetics*, 29(4), p. 201-214 1978.

[6] R. Chellapa and A. Jain (eds) *Markov Random Fields*, 1993.

[7] C. Chubb and M. S. Landy, "Orthogonal distribution analysis: a new approach to the study of texture perception." *Comp. Models of Visual Proc.* Landy. etc (ed.) MIT press 1991.

[8] G. R. Cross and A. K. Jain, "Markov random field texture models." *IEEE, PAMI*, **5**, 25-39. 1983.

[9] J. S. De Bonet and P. Viola, "A non-parametric multi-scale statistical model for natural images", *Advances in Nueral Information Processing* 10, 1997.

[10] I. Daubechies, *Ten lectures on wavelets.* Society for Industry and Applied Math, Philadephia, PA 1992.

[11] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of Optical Soc. Amer.* Vol.2, No.7, 1985.

[12] P. Diaconis and D. Freedman, "On the statistics of vision: the Julesz conjecture", *Journal of Mathematical Psychology*, 24, 1981.

[13] A. Gagalowicz and S. D. Ma, "Model driven synthesis of natural textures for 3D scenes", *Computers and Graphics.* 10:161-170, 1986.

[14] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images". IEEE Trans. on PAMI 9(7), pp 721-741. 1984.

[15] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision." *Proc. Int. Congress of Math.*, Berkeley, CA, 1986.

[16] H. O. Georgii, *Gibbs Measures and Phase Transition.* de Gruyter, New York, 1988.

[17] W. R. Gilks and R. O. Roberts, "Strategies for improving MCMC", chapter 6 in W. R. Gilks et al (eds) *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1997.

[18] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1997.

[19] R. M. Haralick, "Statistical and structural approach to texture", *Proc. IEEE*, vol.67, May, 1979.

[20] D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis." *SIGGRAPHS*, 1995.

[21] E. T. Jaynes. "Information theory and statistical mechanics". *Physical Review.* **106**, pp620-630, 1957.

[22] B. Julesz, "Visual pattern discrimination", *IRE Transactions on Information Theory*, IT-8:84-92, 1962.

[23] B. Julesz, "Toward an axiomatic theory of preattentive vision", pp. 585-612 in *Dynamic aspects of neocortical function*, G. Edelman et al. (Eds.), Wiley, New York 1984.

[24] B. Julesz, *Dialogues on Perception*, 1995.

[25] S. Kullback, and R. A. Leibler, "On information and sufficiency", *Annual Math. Stat.* vol.22, pp79-86, 1951.

[26] J. T. Lewis, C. E., Pfister, and W. G. Sullivan, "Entropy, concentration of probability and conditional limit theorems". *Markov Processes Relat. Fields*, **1**, 319-396, 1995.

[27] J. S. Liu, C. Sabatti, and Y. N. Wu, "Haar measure and Gibbs sampler", preprint of the department of statistics, Univ. of Michigan, 1999.

[28] D. Marr, *Vision*, W.H. Freeman and Company, New York, 1982.

[29] S. Mallat. "Multi-resolution approximations and wavelet orthonormal bases of $L^2(R)$. Trans. Amer. Math. Soc. 315, pp 69-87. 1989.

[30] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087-92, 1953.

[31] K. Popat and R. W. Picard, "Cluster based probability model and its application to image and texture processing", *IEEE Trans. Information Processing*, 6(2), 1997.

[32] E. P. Simoncelli and J. Portilla, "Texture characterization via joint statistics of wavelet coefficient magnitudes", *Prof. 5th Int'l Conf. on Image Processing*, Chicago, IL, 1998.

[33] R. Vistnes, "Texture models and image measures for texture discrimination", *Int'l Journal of Computer Vision*, 3, 313-336, 1989.

[34] J. Weszka, C. R. Dyer, and A. Rosenfelf, "A comparative study of texture measures for terrain classification" *IEEE Trans. Sys. Man, Cybern.* Vol. SMC-6, April, 1976.

[35] Y. N. Wu and S. C. Zhu, "equivalence of image ensembles and fundamental bounds for texture discrimination", *IEEE Trans. on PAMI* (submitted) 1999.

[36] S. C. Zhu, Y. N. Wu, and D. Mumford. "FRAME: Filters, Random field And Maximum Entropy", *Proc. Int'l Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA 1996.

[37] S. C. Zhu, Y. N. Wu, and D. Mumford. "Minimax entropy principle and its application to texture modeling". Neural Computation, Vol. 9, no 8, Nov. 1997.

[38] S. C. Zhu and D. Mumford. "Prior learning and Gibbs reaction-diffusion." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.19, no.11, Nov. 1997.