Equivalence of Julesz and Gibbs Texture Ensembles

Ying Nian Wu[1], Song Chun Zhu[2], and Xiuwen Liu[2]
[1]Dept. of Statistics, Univ. of Michigan, Ann Arbor, MI 48109
[2]Dept. of Computer and Info. Science, Ohio State Univ., Columbus, OH 43210

Abstract

Research on texture has been pursued along two different lines. The first line of research, pioneered by Julesz (1962), seeks the essential ingredients in terms of *features and statistics* in human texture perception. This leads us to a mathematical definition of texture as a *Julesz ensemble.* A Julesz ensemble is the maximum set of images that share the same value of some basic feature statistics as the image lattice $\Lambda \to Z^2$, or equivalently it is a uniform distribution on this set. The second line of research studies *statistical models*, in particular, Markov random field (MRF) and FRAME models (Zhu, Wu, and Mumford 1997), to characterize texture patterns locally. In this article, we bridge the two lines by the fundamental principle of *equivalence of ensembles* in statistical mechanics (Gibbs, 1902). We prove that 1). The conditional probability of a arbitrary image patch given its environment, under the Julesz ensemble or the uniform model, is inevitably a FRAME (MRF) model, and 2). The limit of the FRAME (MRF) model, which we called the Gibbs ensemble, is equivalent to a Julesz ensemble as $\Lambda \to Z^2$. Thus the advantages of the two methodologies can be fully utilized.

1 Introduction

Texture modeling and synthesis has been intensively studied in computer vision and psychophysics in the past three decades. From a global view, the research has been pursued along two different lines.

Research along the first line, pioneered by Julesz (1962), studies the basic feature statistics that lead to human texture impression, so that images sharing the same values of feature statistics cannot be told apart in pre-attentive vision. Examples of feature statistics include co-occurrence matrices, clique statistics, and more recently, histograms of linear filter responses. For a set of feature statistics, as the image lattice $\Lambda \to Z^2$, we call the set of images sharing the same value of feature statistics, or more precisely, the uniform distribution over this set, the *Julesz ensembles* (Zhu, et al., 1999). Markov chain Monte Carlo (MCMC) can be used to synthesize texture images by sampling from the Julesz ensemble (Zhu, et al., 1999), and thus we can verify the sufficiency of the feature statistics. The Julesz ensemble is globally defined on Z^2, in the literature, it was unclear what local statistical properties the Julesz ensembles have when they are applied to tasks like texture segmentation and discrimination.

Research along the second line builds statistical models to characterize texture patterns. Among them, Markov random fields (MRF), or equivalently the Gibbs distributions are the most successful models (e.g., Besag, 1974; Cross and Jain, 1983; Geman and Geman, 1984). Recently, Zhu, Wu, and Mumford (1997) have shown that these models can be unified under a minimax entropy learning principle, and that MRF models incorporating statistics of filter responses (called FRAME) can model a wide variety of natural textures. We call the limit of the FRAME model as $\Lambda \to Z^2$ the *Gibbs ensemble.* The Markov property makes the Gibbs distributions suitable for image reconstruction and image segmentation, but it is necessary to know its global statistical property of the Gibbs ensembles for model verification and model selection purposes.

For a comparison between the Julesz ensemble and the Gibbs ensemble, the former is more fundamental scientifically and is defined by global hard constraints, whereas the latter is more elegant mathematically and is defined by local interactions or a "soft" constraint through maximum entropy (see Zhu, Wu and Mumford 1997). In this article, we unify the two research lines by showing the equivalence between the Julesz ensemble and the Gibbs ensemble, borrowing the fundamental principle of equivalence of ensembles in statistical mechanics. The equivalence of ensembles reveals two significant facts in texture modeling. 1). Locally, under the Julesz ensemble, the conditional distribution of an image patch of arbitrary shape given its environment is exactly the FRAME model. 2). Globally, on Z^2 (or large enough lattice) the Gibbs ensemble concentrates its probability mass uniformly

0-7695-0164-8/99 $10.00 © 1999 IEEE

over a set of images sharing the same value of feature statistics – the Julesz ensemble. Therefore, a Gibbs ensemble is also a Julesz ensemble.

The key to the equivalence of ensembles is the *probability rate function* in the large deviation theory (e.g., Lewis, Pfister, and Sullivan, 1995). The probability rate function describes the asymptotic behavior of the probabilities of different image sets, and sheds light on concepts like "typical" and "modeling". An important conclusion is that when we sample from the Julesz ensemble or the Gibbs ensemble, we will always get images with the same statistical property (and therefore, the same appearance).

2 Julesz ensemble and Gibbs ensemble

2.1 A simple example

In this subsection, we will use a simple example to demonstrate the important fact that *a statistical model defined on a large image lattice concentrates its probability mass uniformly on a set of images.* The key is the probability rate function in the large deviation theory, which is built on the simple fact that *the term with the largest exponential order dominates the sum, and the order of the sum is the largest order in the individual terms.* One can see this easily from the following simple example. Consider two terms, one is e^{5n}, and the other is e^{3n}. As $n \to \infty$, the sum $e^{5n}+e^{3n}$ is dominated by e^{5n}, and the order of this sum is still 5, i.e., $\log(e^{5n}+e^{3n})/n \to 5$.

Let $\mathbf{I}$ be an image defined on a finite lattice $\Lambda \subset Z^2$, and the intensity at pixel $v \in \Lambda$ is denoted by $\mathbf{I}(v) \in \mathcal{L} = \{1, 2, ..., L\}$. Thus $\Omega_\Lambda = \mathcal{L}^{|\Lambda|}$ is the space of images on Λ, with $|\Lambda|$ being the number of pixels in Λ.

let's consider a simple statistical model where the image intensities are independent and identically distributed (i.i.d.) with $P(\mathbf{I}) = \prod_{v\in\Lambda} P(\mathbf{I}(v))$, and $P(\mathbf{I}(v) = l) = p_l$ for $l = 1, ..., L$, and $\sum_l p_l = 1$. We write $p = (p_1, ..., p_L)$.

For each image $\mathbf{I} \in \Omega_\Lambda$, let the histogram of $\mathbf{I}$ be $h(\mathbf{I}) = (h_1(\mathbf{I}), ..., h_L(\mathbf{I}))$, where $h_l(\mathbf{I})$ is the proportion of pixels with level l in the image $\mathbf{I}$. Then $h(\mathbf{I})$ is the sufficient statistics for model P, i.e., if we denote by $\Omega_\Lambda(h)$ the set of images with $h(\mathbf{I}) = h$, then $P(\mathbf{I})$ assigns equal probabilities to images in $\Omega_\Lambda(h)$. Thus the image space is partitioned into equivalence classes

$$\Omega_\Lambda = \cup_h \Omega_\Lambda(h).$$

As shown in figure 1, each equivalence class $\Omega_\Lambda(h)$ is mapped into one point h on a *simplex* – a plane defined by $h_1 + \cdots + h_L = 1$ and $h_l \geq 0, \forall l$ in an L-dimensional space. We call images in $\Omega_\Lambda(h)$ as images of *type h*.

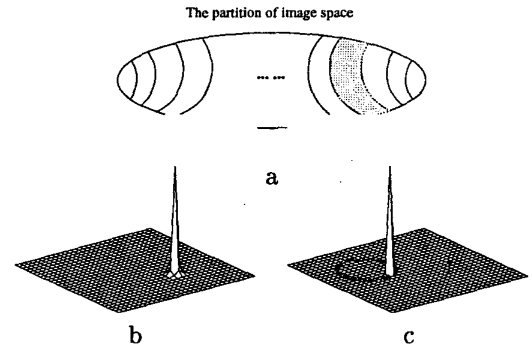

Figure 1: a). The partition of image space into equivalence classes, and each class corresponds to an h on the probability simplex in b). b). A function on the simplex with exponential fall-off, and it approaches a Dirac delta function as $\Lambda \to Z^2$. c). Zoom-in view at a disk area in b). The function at one point on the border of the circle dominates the entire integration.

For an image $\mathbf{I}$ of type h, the log-likelihood is

$$\log P(\mathbf{I}) = \log \prod_{l=1}^{L} p_l^{|\Lambda| h_l} = |\Lambda| \sum_l h_l \log p_l. \quad (1)$$

The number of images of type h, or the volume of the set $\Omega_\Lambda(h)$ is

$$|\Omega_\Lambda(h)| = \frac{|\Lambda|!}{(|\Lambda|h_1)! \cdots (|\Lambda|h_L)!}, \quad (2)$$

for which it is easy to prove that

$$\lim_{\Lambda \to Z^2} \frac{1}{|\Lambda|} \log |\Omega_\Lambda(h)| = -\sum_{l=1}^{L} h_l \log h_l = \text{entropy}(h). \quad (3)$$

Combining equations (3) and (1), the probability mass for the entire set $\Omega_\Lambda(h)$ of images of type h has an exponential rate,

$$\lim_{\Lambda \to Z^2} \frac{1}{|\Lambda|} \log p(\mathbf{I} \in \Omega_\Lambda(h)) = -D(h||p), \quad (4)$$

where

$$D(h||p) = \sum_{l=1}^{L} h_l \log \frac{h_l}{p_l} \geq 0,$$

is the Kullback-Leibler divergence from h to p.

Equation (4) tells us that the probability mass for each equivalence class $\Omega_\Lambda(h)$ is distributed in the order of $\exp\{-|\Lambda| D(h||p)\}$, i.e., the distribution of $h(\mathbf{I})$ under model P is in the order of $\exp\{-|\Lambda| D(h||p)\}$.

So we call $s_p(h) = -D(h||p)$ the probability rate function of $h(\mathbf{I})$ under model P. Clearly, $s_p(h)$ achieves its unique maximum 0 at $h = p$, and for $h \neq p$, the probability mass the model P assigns to $\Omega_\Lambda(h)$ becomes exponentially small as the image lattice Λ gets large. Therefore, when the image lattice is large, model P concentrates its probability mass on $\Omega_\Lambda(p)$. Because model P assigns equal probabilities to all images in $\Omega_\Lambda(p)$, model P on a large image lattice is essentially a uniform distribution over $\Omega_\Lambda(p)$. So $h = p$ is the typical value of histogram $h(\mathbf{I})$ under model P, i.e., if we randomly draw an image from model P defined on a large lattice, then essentially we will always get an image of type $h = p$. Thus, although locally the pixel intensities are randomly distributed, globally, we always observe the same value for the statistics $h(\mathbf{I})$. Figure 1.b) illustrates the intuitive interpretation of the exponential fall-off in the probability simplex. As $\Lambda \to Z^2$, the probability $P(\mathbf{I} \in \Omega_\Lambda(h))$ converges to a Dirac delta function centered at $h = p$.

Furthermore, as shown in Figure 1.c), for any set $\mathcal{H}$ on the simplex, let $\Omega_\Lambda(\mathcal{H}) = \{\mathbf{I} : h(\mathbf{I}) \in \mathcal{H}\}$. Then the probability

$$P(\mathbf{I} \in \Omega_\Lambda(\mathcal{H})) = \int_{h \in \mathcal{H}} P(\mathbf{I} \in \Omega_\Lambda(h)) dh.$$

In the above integral (which is a continuous version of sum), $P(\mathbf{I} \in \Omega_\Lambda(h))$ is of the order $\exp\{|\Lambda| s_p(h)\}$. Therefore, the integral is dominated by the $h_\star$ with the largest $s_p(h)$ in $\mathcal{H}$, and the order of the whole integral is still $s_p(h_\star)$. To be more specific, let

$$h_\star = \arg\max_{h \in \mathcal{H}} s_p(h) = \arg\min_{h \in \mathcal{H}} D(h||p),$$

then

$$\begin{aligned} & \lim_{\Lambda \to Z^2} \frac{1}{|\Lambda|} \log p(\mathbf{I} \in \Omega_\Lambda(\mathcal{H})) \\ = & \lim_{\Lambda \to Z^2} \frac{1}{|\Lambda|} \log p(\mathbf{I} \in \Omega_\Lambda(h_\star)) \\ = & \; s_p(h_\star) = -D(h_\star||p). \end{aligned}$$

2.2 Features and statistics

Recent approaches to texture modeling begin with introducing a set of features/filters $\{F^{(\alpha)}, \alpha = 1, 2, ..., A\}$, and computing the sub-band images $\mathbf{I}^{(\alpha)}$, with $\mathbf{I}^{(\alpha)}(v) = F^{(\alpha)} * \mathbf{I}(v)$ for linear filters. A general feature statistics can be computed as follows. First, choose a G-polygon whose G vertices lie on various sub-bands in the pyramid as displayed in Figure 2. So we can index the G-polygon by $\{(\alpha_g, u_g), g = 1, ..., G\}$, with α_g indexes the pyramid level, and u_g the displacement of the vertex. Because texture is a statistical property of local spatial structures, the u_g's should be close to each other. Then we can move this G-polygon over the image lattice, and collect a set of G-tuples of filter responses, $\{(\mathbf{I}^{(\alpha_1)}(v + u_1), ..., \mathbf{I}^{(\alpha_G)}(v + u_G)), v \in \Lambda\}$. Finally, the feature statistics for this polygon can be computed as the G-dimensional histogram of these G tuples,

$$H(\mathbf{I}) = \sum_{(h_1,...,h_G)} \prod_{g=1}^{G} 1_{\mathbf{I}^{(\alpha_g)}(v+u_g)=h_g},$$

where $(h_1, ..., h_G)$ runs through all possible values of the G-tuple, which are assumed to be suitably quantitized.

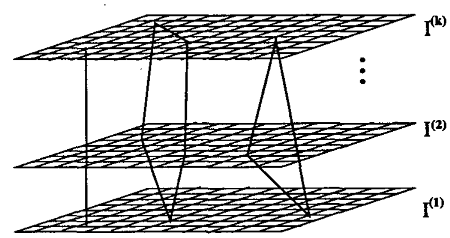

Figure 2: The general feature statistics are multidimensional histograms for polygons in the image pyramid.

If the pyramid has only one layer, i.e., the raw image $\mathbf{I}$, then the feature statistics reduce to co-occurrence matrices (Julesz, 1962; Gagalowicz and Ma, 1986). For a polygon with only one vertex, the feature statistics become the marginal histogram used in Heeger (1996) and Zhu et al. (1997). If the polygon is a straight line (see figure 2), the feature statistics become the joint histogram used by De Bonet and Viola (1999). The histograms can be further reduced to moments, rectified moments, or other more parsimonious statistics.

2.3 The Julesz ensemble

Our definition of the Julesz ensemble is motivated by Julesz's quest for a general "texton theory". In his seminal paper (Julesz, 1962), Julesz asked 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?

Suppose on the image pyramid, K polygons are used for texture modeling, which give K unnormalized histograms $H_1(\mathbf{I}), ..., H_K(\mathbf{I})$. We let $\mathbf{H}(\mathbf{I}) =$

$(H_1(\mathbf{I}), ..., H_K(\mathbf{I}))$, and $\mathbf{h}(\mathbf{I}) = \mathbf{H}(\mathbf{I})/|\Lambda|$ be the normalized histograms. Let

$$\Omega_\Lambda(\mathbf{h}) = \{\mathbf{I} : \mathbf{h}(\mathbf{I}) = \mathbf{h}\} \quad (5)$$

be the set of images sharing the same value $\mathbf{h}$ of feature statistics. The value $\mathbf{h}$ is often extracted from some observed images. For finite lattice Λ, the exact constraint $\mathbf{h}(\mathbf{I}) = \mathbf{h}$ may not be satisfied. So we relax this constraint a little bit, and replace $\Omega_\Lambda(\mathbf{h})$ by

$$\Omega_\Lambda(\mathcal{H}) = \{\mathbf{I} : \mathbf{h}(\mathbf{I}) \in \mathcal{H}\}$$

with $\mathcal{H}$ being an open neighborhood around $\mathbf{h}$. Then the associate uniform distribution is

$$q(\mathbf{I}; \mathcal{H}) = \begin{cases} 1/|\Omega_\Lambda(\mathcal{H})|, & \text{if } \mathbf{I} \in \Omega_\Lambda(\mathcal{H}), \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where $|\Omega_\Lambda(\mathcal{H})|$ is the volume of $\Omega_\Lambda(\mathcal{H})$.

Definition *Given a set of feature statistics* $\mathbf{h}(\mathbf{I}) = (h_1(\mathbf{I}), ..., h_K(\mathbf{I}))$, *a Julesz ensemble with parameter* $\mathbf{h}$ *is a limit of* $q(\mathbf{I}; \mathcal{H})$ *as* $\Lambda \to \mathrm{Z}^2$ *and* $\mathcal{H} \to \mathbf{h}$ *with some boundary condition.*

A Julesz ensemble is a mathematical idealization of $q(\mathbf{I}; \mathcal{H})$ for a large Λ with some boundary condition and with $\mathcal{H}$ close to $\mathbf{h}$. 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.

Then, we are ready to give a *mathematical definition* for texture.

Definition *A texture is the Julesz ensemble for the set of feature statistics* $\mathbf{h}(\mathbf{I})$ *employed by human vision in forming texture impression.*

Then, texture modeling can be posed as an inverse problem, i.e., given a set of observed images sampled by natural stochastic precesses (physical or chemical), find the largest Julesz ensemble, or more specifically, the minimal set of feature statistics, such that images sampled from this Julesz ensemble have the same texture appearance as the observed ones.

The verification of the sufficiency of feature statistics $\mathbf{h}(\mathbf{I})$ can be accomplished by first computing $\mathbf{h}_{\mathrm{obs}}$ from observed texture images, and then sampling from the Julesz ensemble with parameter $\mathbf{h}_{\mathrm{obs}}$, or more practically from $q(\mathbf{I}; \mathcal{H})$ on a large lattice with $\mathcal{H}$ close to $\mathbf{h}_{\mathrm{obs}}$, to see if the sampled images resemble the visual appearance of the observed ones. The necessity of feature statistics is a much more delicate issue as there are infinitely many ways to reduce $\mathbf{h}(\mathbf{I})$, and some reduction of $\mathbf{h}(\mathbf{I})$ may still be judged by human vision as being sufficient. For texture synthesis in computer graphics, the necessity is not an important issue. For texture modeling in computer vision, however, the necessity or the parsimony of feature statistics is very important.

The $q(\mathbf{I}; \mathcal{H})$ can be sampled by simulated annealing. We first define an energy function

$$\mathcal{E}(\mathbf{I}) = \begin{cases} 0, & \text{if } \mathbf{h}(\mathbf{I}) \in \mathcal{H}, \\ D(\mathbf{h}(\mathbf{I}), \mathbf{h}_{\mathrm{obs}}(\mathbf{I})), & \text{otherwise,} \end{cases}$$

where D is a suitably chosen distance (e.g., L_1 distance). Then the distribution

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

goes to $q(\mathbf{I}; \mathcal{H})$, i.e. the uniform distribution over the minima of $\mathcal{E}(\mathbf{I})$ as the temperature T goes to 0. We can sample $q(\mathbf{I})$ by the Gibbs sampler (Geman and Geman, 1984) or a generalized version of the Gibbs sampler (see Zhu, et al. 1999 and references therein).

In our experiments, we select all the 56 filters (Gabor filters at various scales and orientations and small Laplacian of Gaussian filters) used by Zhu, et al. (1997). We match the *marginal histograms* of the 56 filters all together. Some of the results are displayed in Figures 4 and 5. See Zhu, et al. (1999) for more details and discussions of the results. For simplicity, for the rest of the paper, we assume that the feature statistics are marginal statistics of filter responses, although our results apply to more general situations.

2.4 MRF models and the Gibbs ensemble

Statistical modeling of texture is motivated by vision problems such as texture clustering, discrimination and segmentation. Among all statistical models, the MRF models (e.g., Besag, 1974; Cross and Jain, 1983) are the most successful and the most elegant.

Recently, Zhu, et al. (1997) discovered the following general MRF model. Given statistics $\mathbf{H}(\mathbf{I}) = (H_1(\mathbf{I}), ..., H_K(\mathbf{I}))$, the MRF model for $\mathbf{I}$ is

$$\begin{aligned} p(\mathbf{I}; \beta) &= \frac{1}{Z_\Lambda(\beta)} \exp\{-\sum_{k=1}^{K} < \beta_k, H_k(\mathbf{I}) >\} \\ &= \frac{1}{Z_\Lambda(\beta)} \exp\{-|\Lambda| < \beta, \mathbf{h}(\mathbf{I}) >\}, \end{aligned}$$

where $Z_\Lambda(\beta)$ is the normalizing constant. This model is specified by the parameter $\beta = (\beta_1, ..., \beta_K)$, whose value is determined by the constraint

$$E_{p(\mathbf{I};\beta)}[\mathbf{h}(\mathbf{I})] = \mathbf{h}_{\mathrm{obs}},$$

where $\mathbf{h}_{\mathrm{obs}}$ is computed from observed images. We call the above constraint a soft constraint because it only requires that the statistics are matched on ensemble average. Among all the distributions $p(\mathbf{I})$ satisfying $E_p[\mathbf{h}(\mathbf{I})] = \mathbf{h}_{\text{obs}}$, $p(\mathbf{I};\beta)$ has the maximum entropy, so it integrates the observed statistics $\mathbf{h}_{\text{obs}}$ in the most unbiased way.

The $p(\mathbf{I};\beta)$ unifies all the MRF texture models, which are different only in their definitions of feature statistics $\mathbf{H}(\mathbf{I})$. Although the MRF models are less straightforward than the Julesz ensembles, they are much more analytically tractable due to the Markov property. More specifically, for any patch $\Lambda_0 \subset \Lambda$, the conditional distribution of $\mathbf{I}_{\Lambda_0}$ given the rest of the image $\mathbf{I}_{\Lambda/\Lambda_0}$ only depends on the pixels that can share the same filters with pixels in Λ_0. We call the set of such pixels the neighborhood of Λ_0, and denote it by $\partial\Lambda_0$. The condition distribution is

$$\begin{aligned} & p(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\Lambda/\Lambda_0};\beta) = p(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\partial\Lambda_0};\beta) \\ = & \frac{1}{Z_{\Lambda_0}(\beta)} \exp\{- < \beta, \mathbf{H}(\mathbf{I}_{\Lambda_0}|\mathbf{I}_{\partial\Lambda_0}) >\}, \end{aligned}$$

where $\mathbf{H}(\mathbf{I}_{\Lambda_0}|\mathbf{I}_{\partial\Lambda_0})$ is the statistics computed by filtering within $\Lambda_0 \cup \partial\Lambda_0$. Similar to the definition of the Julesz ensemble, we have

Definition *Given a set of feature statistics* $\mathbf{h}(\mathbf{I}) = (h_1(\mathbf{I}), ..., h_K(\mathbf{I}))$, *a Gibbs ensemble with parameter* β *is a limit of* $p(\mathbf{I};\beta)$ *as* $\Lambda \to Z^2$ *with some boundary condition.*

The Gibbs ensemble is a mathematical idealization of $p(\mathbf{I};\beta)$ on a large Λ with some boundary condition.

3 Equivalence of ensembles

3.1 Local Markov property of the Julesz ensemble

In this subsection, we derive the local Markov property of the Julesz ensemble, which is globally defined by $\mathbf{h}$. This derivation is adapted from the traditional argument in statistical mechanics (Gibbs, 1902), where the Julesz ensemble can be identified with the micro-canonical ensemble (an isolated system with fixed energy), and the Gibbs ensemble with the canonical ensemble (a system in equilibrium with a heat reservoir). To do this, we need to first derive the probability rate function of $\mathbf{h}(\mathbf{I})$ under the uniform distribution.

Let μ_Λ be the uniform distribution over the entire image space Ω_Λ, and let $\mu_\Lambda(\mathcal{H})$ be the probability that μ_Λ assigns to the image set $\Omega_\Lambda(\mathcal{H})$. Then the volume of $\Omega_\Lambda(\mathcal{H})$ is $L^{|\Lambda|}\mu_\Lambda(\mathcal{H})$. For μ_Λ, we have

Proposition 1 *The limit*

$$\lim_{\Lambda\to \mathbf{Z}^2} \frac{1}{|\Lambda|} \log \mu_\Lambda(\mathcal{H}) = s(\mathcal{H})$$

exists. Let $s(\mathbf{h}) = \lim_{\mathcal{H}\to\mathbf{h}} s(\mathcal{H})$, *then* $s(\mathbf{h})$ *is strictly concave, and* $s(\mathcal{H}) = \sup_{\mathbf{h}\in\mathcal{H}} s(\mathbf{h})$.

The probability rate function $s(\mathbf{h})$ tells us that the distribution $\mu_\Lambda(\mathbf{h})$ behaves like $\exp\{|\Lambda|s(\mathbf{h})\}$. The equation $s(\mathcal{H}) = \sup_{\mathbf{h}\in\mathcal{H}} s(\mathbf{h})$ can be understood in the same way as in the simple i.i.d. example we discussed, i.e., the term with the largest order dominates. See Lanford (1973) for a detailed analysis of the above result.

With $s(\mathbf{h})$, we are ready to derive the Markov property of the Julesz ensemble. Consider the model $q(\mathbf{I};\mathcal{H})$. For simplicity, we shall just take $\mathcal{H}$ to be $\mathbf{h}$, and assume that Λ is large, so $q(\mathbf{I};\mathbf{h})$ is uniform over $\Omega_\Lambda(\mathbf{h})$. First, we fix $\Lambda_1 \subset \Lambda$, and then fix $\Lambda_0 \subset \Lambda_1$. We are interested in the conditional distribution of the local patch $\mathbf{I}_{\Lambda_0}$ given its local environment $\mathbf{I}_{\Lambda_1/\Lambda_0}$ under the model $q(\mathbf{I};\mathbf{h})$ with a large Λ. We denote this distribution by $q(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\Lambda_1/\Lambda_0}, \mathbf{h})$. We assume that Λ_0 is sufficiently smaller than Λ_1 so that the neighborhood of Λ_0, $\partial\Lambda_0$, is contained in Λ_1.

Let $\mathbf{H}_0 = \mathbf{H}(\mathbf{I}_{\Lambda_0}|\mathbf{I}_{\partial\Lambda_0})$ be the statistics computed for $\mathbf{I}_{\Lambda_0}$ where filtering takes place within $\Lambda_0\cup\partial\Lambda_0$. Let $\mathbf{H}_{01}$ be the statistics computed by filtering inside the fixed environment Λ_1/Λ_0. Let $\Lambda_{-1} = \Lambda/\Lambda_1$ be the big patch outside of Λ_1. Then the statistics computed for Λ_{-1} is $\mathbf{h}|\Lambda| - \mathbf{H}_0 - \mathbf{H}_{01}$. Let $\mathbf{h}' = (\mathbf{h}|\Lambda| - \mathbf{H}_{01})/|\Lambda_{-1}|$, then the normalized statistics for Λ_{-1} is $\mathbf{h}' - \mathbf{H}_0/|\Lambda_{-1}|$.

For a certain image patch $\mathbf{I}_{\Lambda_0}$, the number of images in $\Omega_\Lambda(\mathbf{h})$ with such a patch $\mathbf{I}_{\Lambda_0}$ and its local environment $\mathbf{I}_{\Lambda_1/\Lambda_0}$ is $|\Omega_{\Lambda_{-1}}(\mathbf{h}' - \mathbf{H}_0/|\Lambda_{-1}|)|$. So if we sample an image from $\Omega_\Lambda(\mathbf{h})$ randomly, then the probability we observe $\mathbf{I}_{\Lambda_0}$ on Λ_0 with an environment $\mathbf{I}_{\Lambda_1/\Lambda_0}$ is

$$q(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\Lambda_1/\Lambda_0}, \mathbf{h}) \propto 1/|\Omega_{\Lambda_{-1}}(\mathbf{h}' - \frac{\mathbf{H}_0}{|\Lambda_{-1}|})|.$$

Note that as a distribution of $\mathbf{I}_{\Lambda_0}$, $q(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\Lambda_1/\Lambda_0}, \mathbf{h})$ is decided by $\mathbf{H}_0$, which is the sufficient statistics. Therefore, we only need to trace $\mathbf{H}_0$ while leaving other terms as constants. For large Λ,

$$\begin{aligned} & \log q(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\Lambda_1/\Lambda_0}, \mathbf{h}) = \text{const} + |\Lambda_{-1}|s(\mathbf{h}' - \frac{\mathbf{H}_0}{|\Lambda_{-1}|}) \\ = & \text{const} - < s'(\mathbf{h}'), \mathbf{H}_0 > + o(\frac{1}{|\Lambda|}), \end{aligned}$$

where the first equation follows from Proposition 1, and the second equation follows from a Taylor expansion at $\mathbf{h}'$. Letting $\beta = s'(\mathbf{h})$, then, as $\Lambda \to Z^2$, $\mathbf{h}' \to \mathbf{h}$, and

$$\begin{aligned} \log q(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\Lambda_1/\Lambda_0}, \mathbf{h}) & \to \text{const} - < s'(\mathbf{h}), \mathbf{H}_0 > \\ & = \text{const} - < \beta, \mathbf{H}_0 >, \end{aligned}$$

so

$$q(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\Lambda_1/\Lambda_0}, \mathbf{h}) \to \frac{1}{Z_{\Lambda_0}(\beta)} \exp\{- < \beta, \mathbf{H}(\mathbf{I}_{\Lambda_0} \mid \mathbf{I}_{\partial\Lambda_0}) >\},$$

which is exactly the Markov property that governs the Gibbs ensemble. This derivation shows that local computation using the MRF model is justified under the Julesz ensemble. It also reveals an important relationship, i.e., the parameter β can be identified as the derivative of the probability rate $s(\mathbf{h})$.

3.2 Global statistical property of the Gibbs ensemble

In this subsection, we shall start with the Gibbs ensemble defined by β, and show that it is essentially a Julesz ensemble.

Clearly, the MRF model $p(\mathbf{I};\beta)$ assigns equal probabilities to images in $\Omega_\Lambda(\mathbf{h})$ for any $\mathbf{h}$, because $\mathbf{h}(\mathbf{I})$ is the sufficient statistics. If we can show that $p(\mathbf{I};\beta)$ eventually focuses on a certain value of $\mathbf{h}(\mathbf{I})$, say, $\mathbf{h}_\star$, then for large lattice, $p(\mathbf{I};\beta)$ is essentially a uniform distribution over $\Omega_\Lambda(\mathbf{h}_\star)$, which leads to the equivalence of ensembles. For this purpose, we need to compute the probability rate function of $\mathbf{h}(\mathbf{I})$ under $p(\mathbf{I};\beta)$.

Because the number of images with $\mathbf{h}(\mathbf{I}) = \mathbf{h}$ is $|\Omega_\Lambda(\mathbf{h})| = L^{|\Lambda|}\mu_\Lambda(\mathbf{h})$, the probability distribution of $\mathbf{h}(\mathbf{I})$ under the MRF model $p(\mathbf{I};\beta)$ is

$$p(\mathbf{h};\beta) = \frac{1}{Z_\Lambda(\beta)} \exp\{-|\Lambda| < \beta, \mathbf{h} >\} L^{|\Lambda|}\mu_\Lambda(\mathbf{h}),$$

and the probability rate

$$\begin{aligned} s_\beta(\mathbf{h}) &= \lim_{\Lambda\to Z^2} \frac{1}{|\Lambda|} \log p(\mathbf{h};\beta) \\ &= - < \beta, \mathbf{h} > + \lim_{\Lambda\to Z^2} \frac{1}{|\Lambda|} \log \mu_\Lambda(\mathbf{h}) \\ &- (\lim_{\Lambda\to Z^2} \frac{1}{|\Lambda|} \log Z_\Lambda(\beta) - \log L). \end{aligned}$$

We already know that

$$\lim_{\Lambda\to Z^2} \frac{1}{|\Lambda|} \log \mu_\Lambda(\mathbf{h}) = s(\mathbf{h})$$

is the probability rate function of $\mathbf{h}(\mathbf{I})$ under the uniform model. For the last term in $s_\beta(\mathbf{h})$, we have

Proposition 2 *The limit*

$$\rho(\beta) = \lim_{\Lambda\to \mathbf{Z}^2} \frac{1}{|\Lambda|} \log Z_\Lambda(\beta) - \log L$$

exists and is independent of the boundary condition. ρ is strictly convex.

See Griffiths and Ruelle (1971) for a proof. Therefore, we have

Proposition 3 *The probability rate function $s_\beta(\mathbf{h})$ of the MRF model $p(\mathbf{I};\beta)$ is $s_\beta(\mathbf{h}) = s(\mathbf{h}) - < \beta, \mathbf{h} > -\rho(\beta)$.*

So we have the following theorem.

Theorem 1 *If there is a unique $\mathbf{h}_\star$ where $s_\beta(\mathbf{h})$ achieves its maximum 0, then $p(\mathbf{I};\beta)$ eventually concentrates on $\mathbf{h}_\star$, and therefore the Gibbs ensemble defined by β is equivalent to the Julesz ensemble defined by $h_\star$, and $s'(\mathbf{h}_\star) = \beta$.*

The uniqueness of $\mathbf{h}_\star$ holds under the condition that there is no phase transition at β. See the next subsection for a discussion.

When there is no phase transition, the Julesz ensemble or the corresponding Gibbs ensemble concentrates its probability mass on a set of typical images sharing the same statistical property. To see this fact, consider an arbitrary new statistics $h_0(\mathbf{I})$ not used for modeling. It can be shown that the Julesz (or Gibbs) ensemble concentrates on the unique $h_{0\star}$ that maximizes $s(\mathbf{h}_\star, h_0)$, where $s(\mathbf{h}, h_0)$ is the probability rate function for the enlarged statistics $(\mathbf{h}(\mathbf{I}), h_0(\mathbf{I}))$ under the uniform model. That means that almost all images in the Julesz (or Gibbs) ensemble produce $h_{0\star}$ for the statistics $h_0(\mathbf{I})$, i.e., if we sample from the Julesz (or Gibbs) ensemble, we will always observe $h_{0\star}$, which is the typical value of the statistics $h_0(\mathbf{I})$. Because $h_0(\mathbf{I})$ is arbitrary, if we sample from the Julesz (or Gibbs) ensemble, we will always get images with the same statistical property. Such images can be called typical images, which absorb all the probability mass of the Julesz (or Gibbs) ensemble.

The equivalence of ensembles also sheds new light on the minimax entropy principle Zhu, et al. (1997) introduced for texture modeling. The minimum entropy principle means that we should choose the feature statistics $\mathbf{h}_{\text{obs}}$ so that $s(\mathbf{h}_{\text{obs}})$ is the smallest, or the volume of $\Omega_\Lambda(\mathbf{h}_{\text{obs}})$ is the smallest, under constraint on model complexity. $s(\mathbf{h}_{\text{obs}})$ is a measure of entropy rate, or the randomness of the observed texture image. The maximum entropy principle means that we should put a uniform distribution over $\Omega_\Lambda(\mathbf{h}_{\text{obs}})$, so that sampling from this uniform distribution gives us typical images in $\Omega_\Lambda(\mathbf{h}_{\text{obs}})$, because almost all images in $\Omega_\Lambda(\mathbf{h}_{\text{obs}})$ are typical images.

3.3 Uniqueness of ensembles

Given a parameter β, $p(\mathbf{I};\beta)$ may go to different limits as $\Lambda \to Z^2$ under different boundary conditions.

Such a phenomenon is called phase transition in statistical physics, and it can manifest itself if we sample from $p(\mathbf{I}; \beta)$: we may get images of different statistical properties if we use different large Λ and different boundary conditions. It is also possible that a sampled image consists of large image patches of different statistical properties. Mathematically, phase transition reflects the fact that there is a cusp, and thus not differentiable in the function $\rho(\beta)$, or a flat top in the function $s_{\beta}(\mathbf{h})$. When there is no phase transition, there is only one Gibbs ensemble which is an ergodic random field.

For a given $\mathbf{h}$, it is also possible that the uniform distribution $q(\mathbf{I}; \mathcal{H})$ goes to different limits as $\Lambda \to Z^2$ and $\mathcal{H} \to \mathbf{h}$. This can manifest itself in a similar way as described above. See Martin-Lof (1979) for more details. Again, we consider such a Julesz ensemble unsuitable for texture modeling.

4 Discussion

There are two important goals in texture modeling. 1). Search for the sufficient and necessary statistics that define the underlying texture pattern. 2). Search for conditional probability of an arbitrary image patch given its environment.

The first goal leads us to the Julesz ensemble, and the second goal leads us to the Gibbs ensemble. In this paper, we establish the equivalence between the two ensembles, therefore justify the FRAME model of Zhu, et al. (1997) as an inevitable description of texture. Figure 3 summarizes the global picture for texture modeling. The dashed line (path 2) represents the research line which pursues the Julesz ensembles. The solid line represents the research which build minimax entropy models. The two lines are connected by the equivalence of ensembles. The advantages of both lines can now be better utilized, with the Julesz ensemble is much more efficient for texture synthesis, model verification, and statistics pursuit, and the Gibbs models provide precise local probability measures for image segmentation and classification.

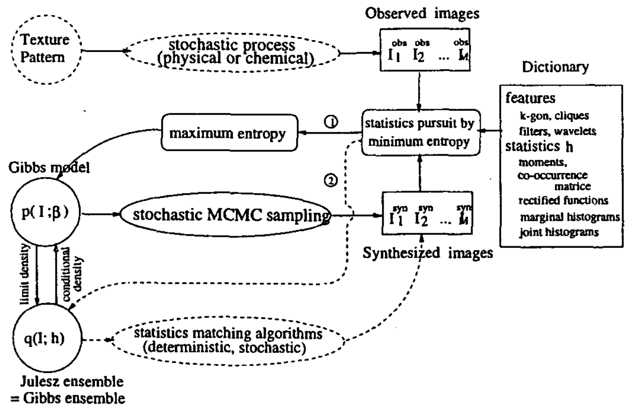

Figure 3: A global picture for texture modeling theories.

References

[1] J. Besag (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Royal Stat. Soc., B*, **36**, 192-236.

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

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

[4] S. Geman and D. Geman (1984) Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. PAMI*, **6**, 721-741.

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

[6] J. W. Gibbs (1902) *Elementary Principles of Statistical Mechanics*. Yale University Press.

[7] R. Griffiths and D. Ruelle (1971) Strict convexity ("continuity") of the pressure in lattice system. *Comm. Math. Phys.*,, **23**, 169.

[8] D. J. Heeger and J. R. Bergen (1996) Pyramid-based texture analysis/synthesis. *SIGGRAPHS*.

[9] O. E. Lanford (1973) Entropy and equilibrium states in classical mechanics. In *Statistical Mechanics and Mathematical Problems*, A. Lenard ed. Springer.

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

[11] A. Martin-Lof (1979) The equivalence of ensembles and Gibbs' phase rule for classical lattice systems. *J. Stat. Phys.*, **20**, 557-569.

[12] Y. Wu and S. C. Zhu (1999) Equivalence of Image Ensembles and Fundamental Bounds — A unified theory of texture modeling and analysis, *Preprint* (available via www.cis.ohio-state.edu/~szhu/publication.html).

[13] S. C. Zhu, X. Liu, and Y. Wu (1999) Exploring Texture Enesmbles by Efiicient Markov Chain Monte Carlo, *Proc. of Workshop on Statistical and Computational Theories of Vision*, June 1999.

[14] S. C. Zhu, Y. Wu, and D. B. Mumford (1997). Minimax entropy principle and its application to texture modeling, *Neural Computation*, **9**, 1627-1660.

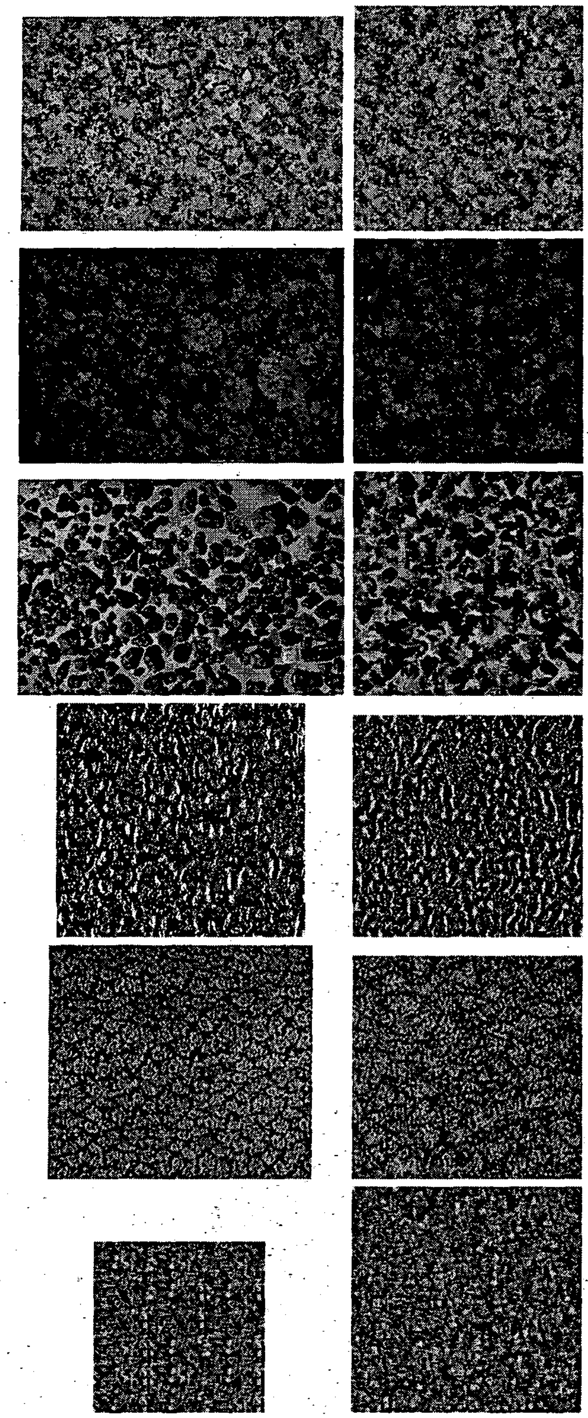

Figure 4: Left column: the observed texture images, right column: the synthesized texture images that share the exact histograms with the observed for 56 filters.

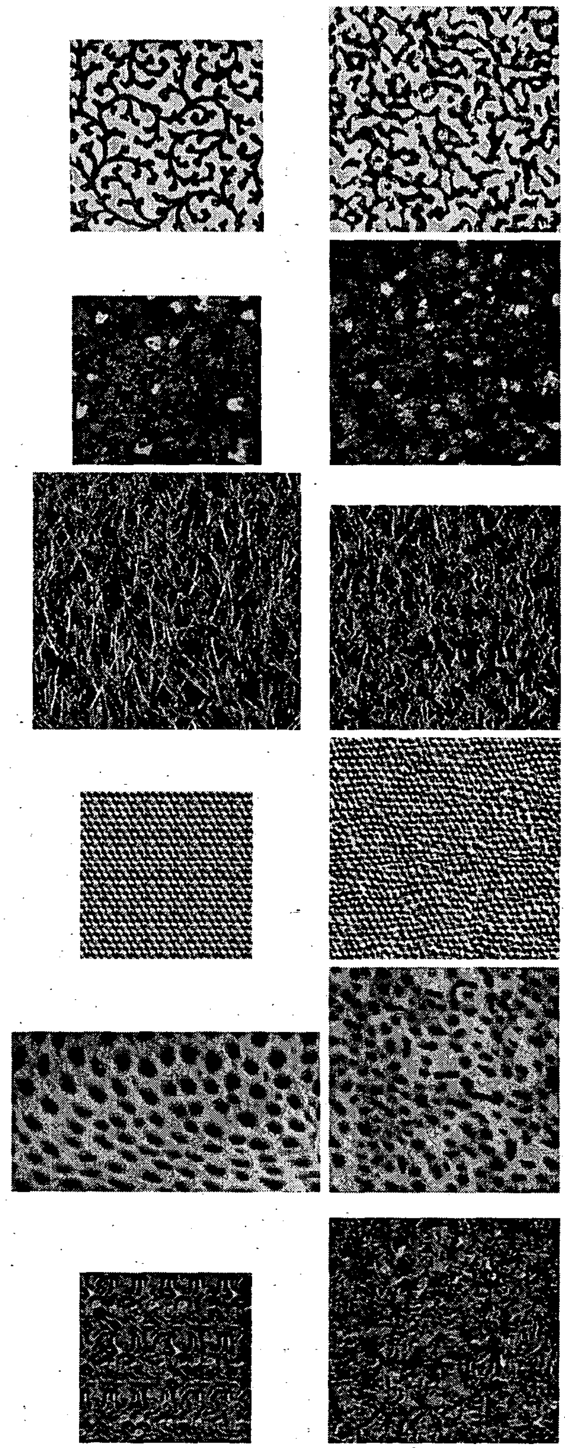

Figure 5: See caption of Figure 4.