Statistical Modeling of Visual Patterns: Part I

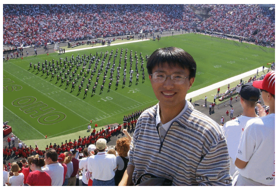
--- From Statistical Physics to Image Models

Song-Chun Zhu

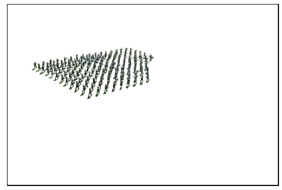
Departments of Statistics and Computer Science
University of California, Los Angeles

Natural images contain a wide variety of visual patterns

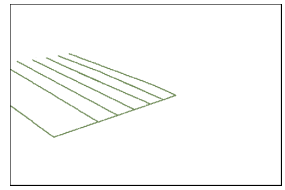
Decomposing Images Into Patterns



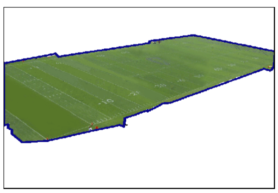
Input image



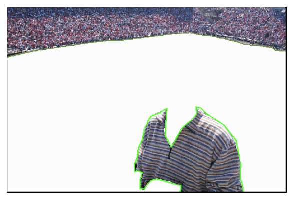
point process



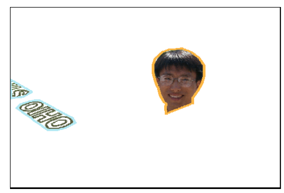
line process



region process



texture/curve process



face and words

Visual Modeling: Knowledge Representation

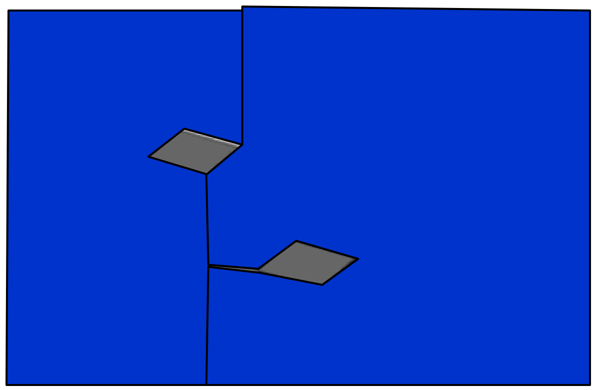
1. What is a mathematical *definition* and *model* of a visual pattern ?

2. What are the *vocabulary* for these visual patterns ?
   By analogy to language, what are the phonemes, words, phrases, sentences, …

3. Can these models and vocabulary be *learned* from
   natural images and video sequences?

How Human Vision Perceives a Visual Pattern

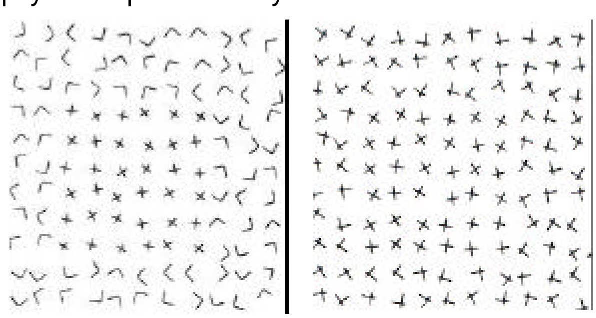
A demo by Adelson (MIT psychology)

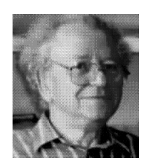


summary + details

How Human Vision Perceives a Visual Pattern

Psychophysics Experiments by Julesz 1960s at the Bell Labs



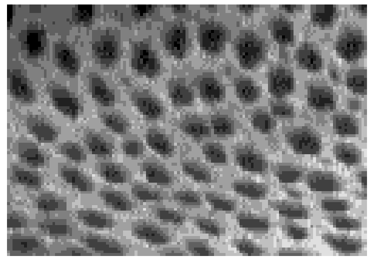


Clearly, human vision extract / summarize some statistics and ignore the other statistics properties associated with the instances.

At least this is true in the early vision stage (0.1-0.4sec).

How Human Vision Perceives a Visual Pattern

For example:



Summary : Some general “impression/statistics/properties” summarize from pixel intensities
--- yields the texture concept of the cheetah skin pattern .

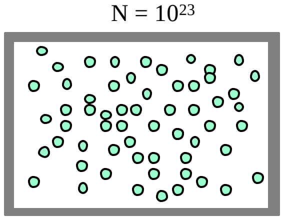
Instance: The specific arrangements of the blobs etc.
---associated with this special example.

Corresponding to Statistical Physics

Statistical physics studies macroscopic properties of systems that consist of massive elements with microscopic interactions.

e.g.: a tank of insulated gas or ferro-magnetic material

$N = 10^{23}$



Micro-canonical Ensemble

A state of the system is specified by the position of the N elements X^N and their momenta p^N

$$S = (x^N, p^N)$$

But we only care about some global properties
Energy E, Volume V, Pressure, ...

Micro-canonical Ensemble = $\Omega(N, E, V) = \{ s : h(S) = (N, E, V) \}$

Definition of a Visual Pattern

Our concept of a pattern is an *abstraction* for an ensemble of *instances* which satisfy some statistical description:

For a homogeneous signal s on 2D lattice Λ, e.g. s = I,

$$\text{a pattern} = \Omega(h_c) = \{ s:\ h(s) = h_c;\ \ \Lambda \to Z^2 \},\quad f(s) = 1/|\Omega(h_c)|.$$

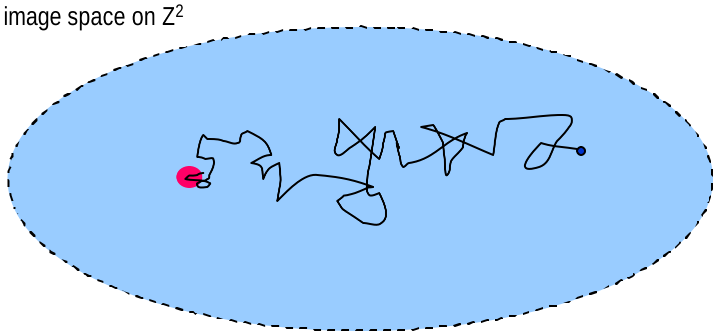
h_c is a summary and s is an instance with details.

This equivalence class is called a Julesz ensemble (Zhu et al 1999)

Simulation for Julesz Ensemble

Draw random samples from the ensemble by Markov chain Monte Carlo methods.

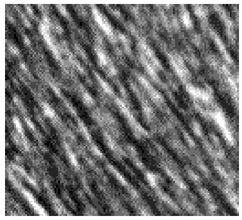
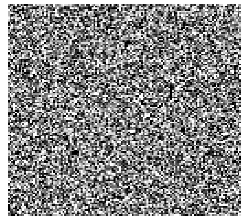
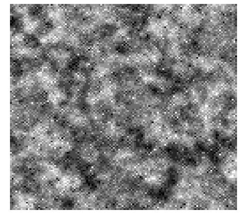
image space on Z^2



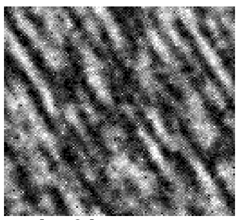
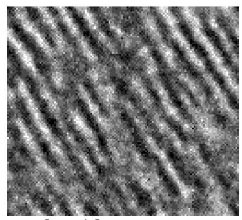
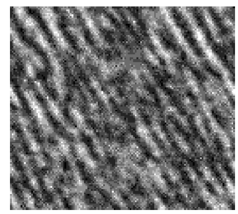
Each point in the space is a large image.

Los Alamos National Lab, Dec. 6, 2002. Song-Chun Zhu

An Example of Texture Pattern (Zhu,Wu,Mumford, 1996-98)

I^{obs} from a unknown $\Omega(h_c)$ $I^{syn} \sim \Omega(h)$ with h= ϕ I^{syn} with h= 1 histogram

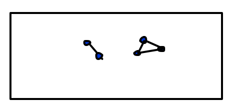




h= 2 histograms h= 4 histograms h= 7 histograms

What are the Essential Statistics: A Brief History

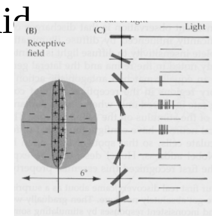
1. Multi-pixel co-occurrence, cliques

(k-gon statistics, Julesz et al. 1960s, 70s)



2. Linear filtering, Gabor, image pyramid

Huber and Weissel 1960s,
Bergen and Adelson 1986,
Turner 1986,
Malik and Perona 1990.
Simoncelli et al. 1992.



3. Histograms of Gabor filtering/wavelets

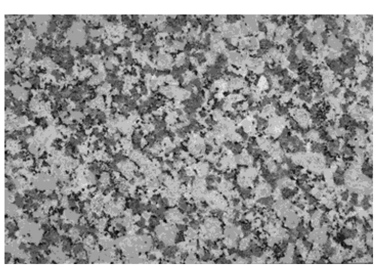
Heeger and Bergen, Siggraph 1995,
Zhu, Wu, and Mumford 1996, … …

Two Obstacles in Answering the Julesz Quest

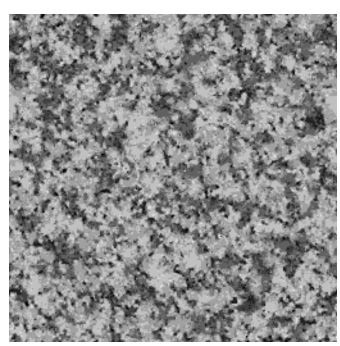
1. Given an arbitrary statistics h_c hypothetically, how can we generate texture pairs that share identical statistics --- no more and no less.

2. Texture is a spatial pattern, unlike color, it cannot be defined on a single pixel.
   --- if it cannot be defined on m x n pixels,
   then it cannot be defined on (m+1) x (n+1) pixels either.

More Examples of Texture



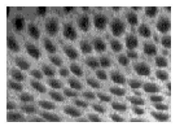
Observed

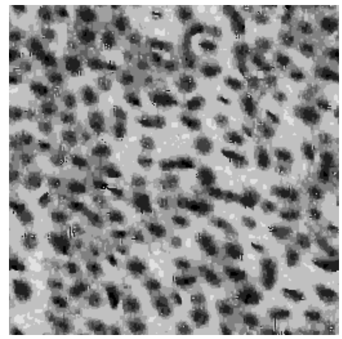


MCMC sample

(Zhu, Wu, Mumford 96-98)

More Examples of Texture

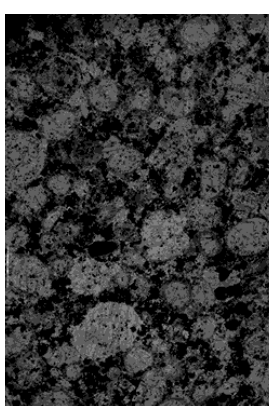


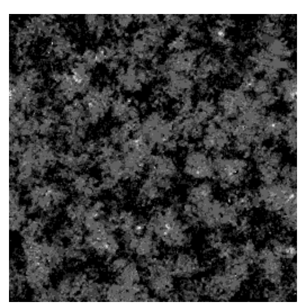


Observed MCMC sample

(Zhu, Wu, Mumford 96-98)

More Examples of Texture

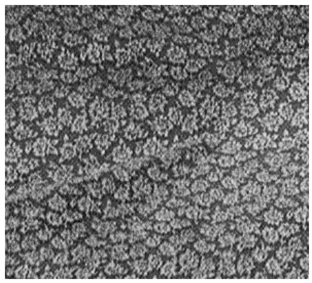


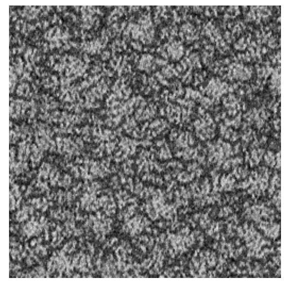


Observed MCMC sample

(Zhu, Wu, Mumford 96-98)

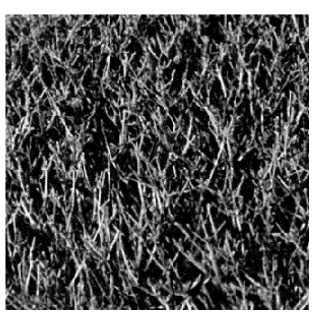
More Examples

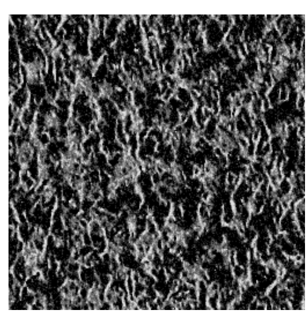




Observed $I^{obs} \sim \Omega(h_c)$ $I^{syn} \sim \Omega(h)$ by MCMC sampling

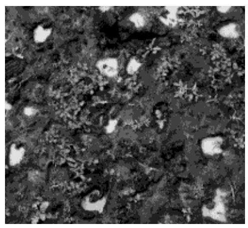
More Examples

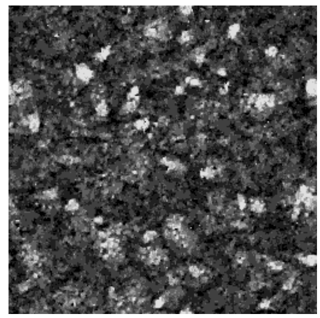




Observed $I^{obs} \sim \Omega(h_c)$ $I^{syn} \sim \Omega(h)$ by MCMC sampling

More Examples





Observed $I^{obs} \sim \Omega(h_c)$ $I^{syn} \sim \Omega(h)$ by MCMC sampling

How Can You Do It?

A Sketch of the work:

Firstly, I will talk about maximum entropy modeling
-- posed as a statistical learning problem.

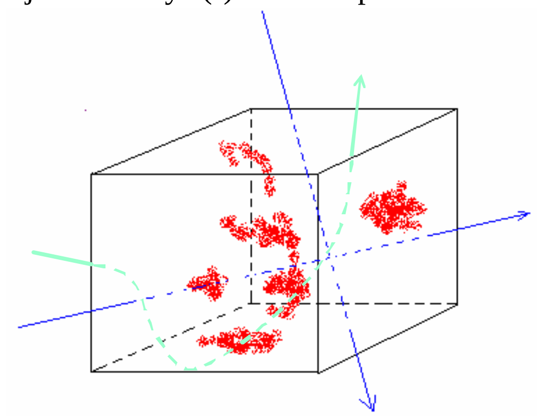
Suppose the ensemble of a visual pattern is governed by a unknown probability / frequency f(s), and our goal is to estimate f(s) by a model p(s).

Secondly, I will introduce the connection from the maximum entropy model to the Julesz ensembles from stat. physics.

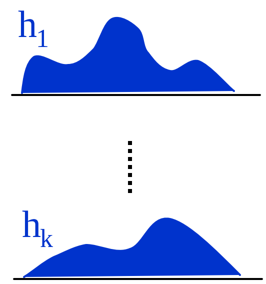
It shows the definition of texture is the limit distribution of model p(s) as the lattice goes to infinity, in the absence of phase transition.

Learning A High Dimensional Density

joint density f(I) in 256^2-space

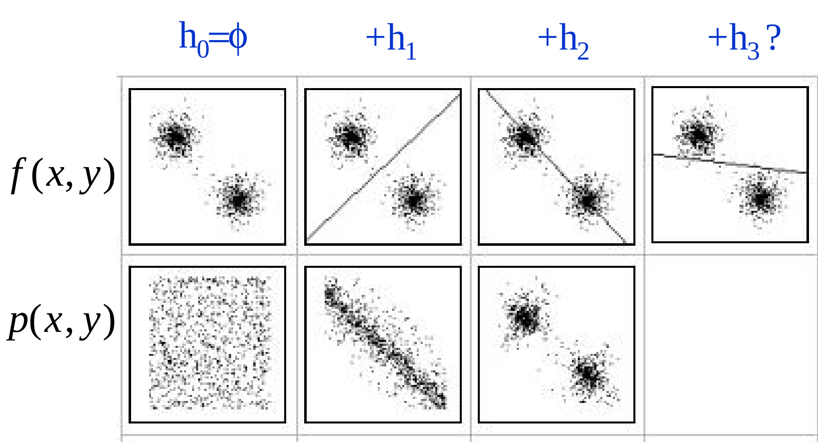


projected (marginal) densities 1D histograms



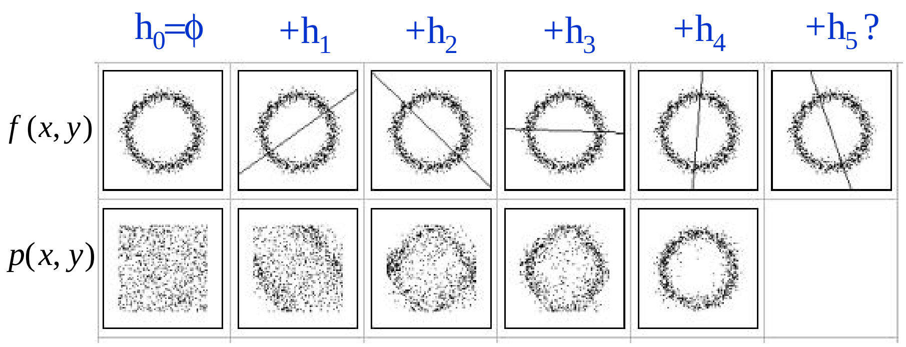
$h_c = (h_1, \ldots, h_k)$.

Toy Example I: Estimating 2D Distribution $f(x, y)$



Please note, it is not match pursuit !

Toy Example II: Estimating 2D Distribution $f(x,y)$



Cramer and Wold theorem: Any continuous density f(x) is a linear combination of its marginal distributions on the linear filter responses.

Probabilistic Modeling by Maximum Entropy
(Jaynes, 1957)

Among all model p that satisfy the constraints, we choose one that has maximum entropy (Jaynes 1957).

$$p = \arg\max_{p \in \Omega_p} - \int p(\mathrm{I}_\Lambda) \log p(\mathrm{I}_\Lambda)\, d\,\mathrm{I}_\Lambda$$

Find model p() Subject to constraints:

p and *f* *must have the same projected statistics*

$$E_{p(\mathrm{I})}[\mathrm{h}_i(\mathrm{I})] = E_{f(\mathrm{I})}[\mathrm{h}_i(\mathrm{I})] \approx \mathrm{h}_i^{obs}, \quad \forall i$$

$$\int p(\mathrm{I})\, d\,\mathrm{I} = 1$$

Maximum Entropy Model of Texture

Solving this constrained optimization problem yields:

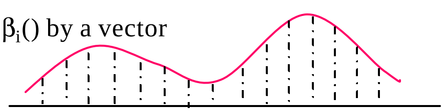
The FRAME model (Filters, Random fields And Maximum Entropy) (Zhu, Wu, Mumford, 1997)

$$p(\mathrm{I};\, \beta, \mathrm{F}) = \frac{1}{Z(\beta, \mathrm{F})} \exp\{-\sum_{j=1}^{K} \sum_{(x,y)} \beta_j(\mathrm{F}_j * \mathrm{I}(x, y))\}$$

$\mathrm{F} = \{ \mathrm{F}_1, \mathrm{F}_2, \ldots, \mathrm{F}_k \}$ are selected filters (wavelets)

$\beta = (\beta_1(), \beta_2(), \ldots, \beta_k())$ are 1D potential functions --- Lagrange multipliers

Approximating $\beta_i()$ by a vector



Los Alamos National Lab, Dec. 6, 2002. Song-Chun Zhu

Selecting Features F and Statistics h_c

Informative F and h_c are selected from a "dictionary" Ω_F to minimize a Kullback-Leibler divergence from *p to f*. The design of Ω_F is human *art*.

$$\mathrm{F}^* = \arg\min_{\mathrm{F}\in\Omega_F} D(f \,\|\, p)$$

$$= \arg\min_{\mathrm{F}\in\Omega_F} \int f(\mathrm{I}) \log \frac{f(\mathrm{I})}{p(\mathrm{I};\beta,\mathrm{F})} d\,\mathrm{I}$$

$$= \arg\min E_f[\log p] - E_f[\log p]$$

$$= \arg\min\, entropy(p) - enrtopy(f)$$

Los Alamos National Lab, Dec. 6, 2002. Song-Chun Zhu

Minimax Entropy Learning

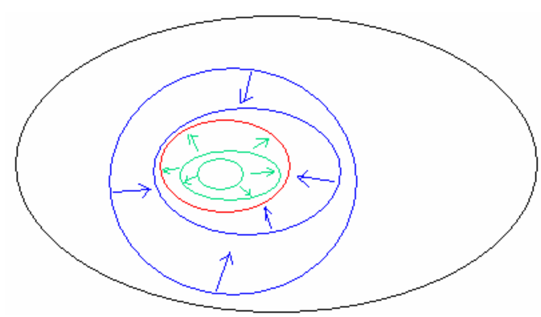
For a Gibbs (max. entropy) model *p*, this leads to the minimax entropy learning principle (Zhu,Wu, Mumford 96,97)

$$p^* = \arg\min_{F} \{ \max_{ß} \text{entropy}(p(\mathrm{I}\, ; ß, \mathrm{F})) \}$$

Actually, it is straightforward to show that the minimax entropy learning steps are related to the maximum likelihood estimation (MLE). But the minimax entropy brings some new perspectives and insights to the problem.

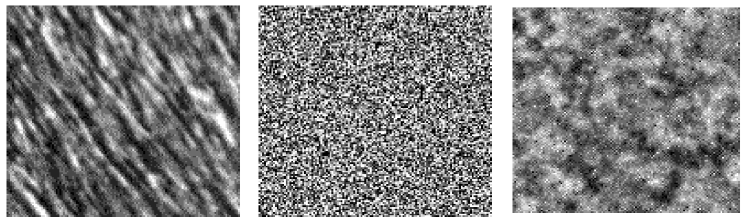
Minimax Entropy Learning

Intuitive interpretation of minimax entropy.

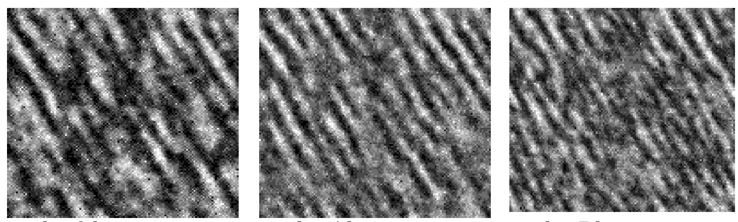


1. Choose *informative* features/statistics to minimize entropy (i.e. log volume or uncertainty).
2. Under the constraints, choose a distribution that has maximum entropy (i.e. *least bias*).

Revisit the example of Texture Modeling

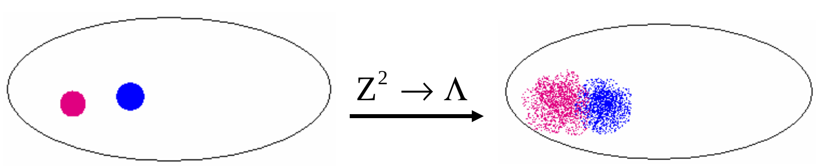


I^{obs} from a unknown $\Omega(h_c)$ | $I^{syn} \sim \Omega(h)$ with $h = \phi$ | I^{syn} with h= 1 histogram



h= 2 histograms | h= 4 histograms | h= 7 histograms

Relationship between Conceptualization and Modeling



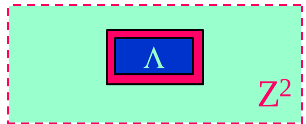
texture ensembles : $f(\mathrm{I};\ h_c)$

texture models : $p(\mathrm{I}_\Lambda \mid \mathrm{I}_{\partial\Lambda}; \beta)$

Markov random fields and FRAME models on finite lattice (Zhu, Wu, Mumford, 1997):

$$p(\mathrm{I}_\Lambda \mid \mathrm{I}_{\partial\Lambda}; \beta) = \frac{1}{Z(\beta)} \exp\{-\sum_{j=1}^{k} \beta_j h_j(\mathrm{I}_\Lambda \mid \mathrm{I}_{\partial\Lambda})\}$$

Equivalence of Julesz ensemble and FRAME models



Theorem 3

For a very large image from the Julesz ensemble $\mathbf{I} \sim f(\mathbf{I}; h_c)$ any local patch of the image $\mathbf{I}_\Lambda$ given its neighborhood follows a conditional distribution specified by a FRAME model $p(\mathbf{I}_\Lambda \mid \mathbf{I}_{\partial\Lambda} : \beta)$

Theorem 4

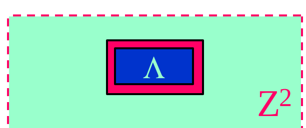
As the image lattice goes to infinity, $f(\mathbf{I}; h_c)$ is the limit of the FRAME model $p(\mathbf{I}_\Lambda \mid \mathbf{I}_{\partial\Lambda} : \beta)$, in the absence of phase transition.

A texture $\longleftrightarrow h_c \longleftrightarrow \beta$

Observation in Statistical Physics

The above theorems reflect an 100-year old observation by Gibbs in stat.physics

"*If a system of a great number of degrees of freedom is micro-canonically distributed in phase, any very small part of it may be regarded as canonically distributed*" --- Gibbs, 1902.



This shows us a truly origin of probability.

--- The reason why we need to play with probabilities in vision is not just because of image noise. With modern digital cameras, there are rarely any noises in images ! It is because of the relationship above !!!

Definition of Finite Patterns

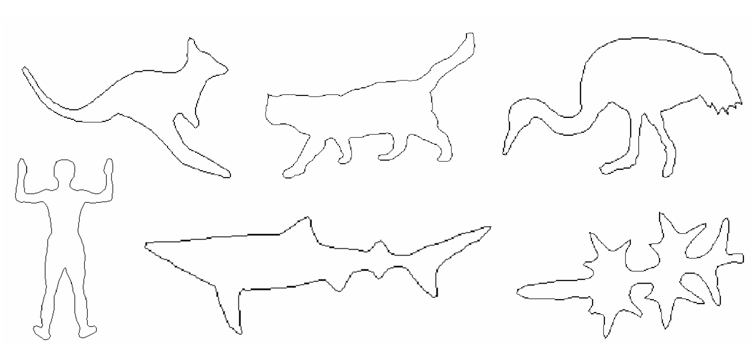
For patterns defined with finite structures: shapes, and faces, they are defined as a set associated with a probability --- “ensemble”.

$$\text{a pattern} = \Omega(h_c) = \{ (s,\ f(s)):\ \ E_f\,[h(s)] = h_c \}$$

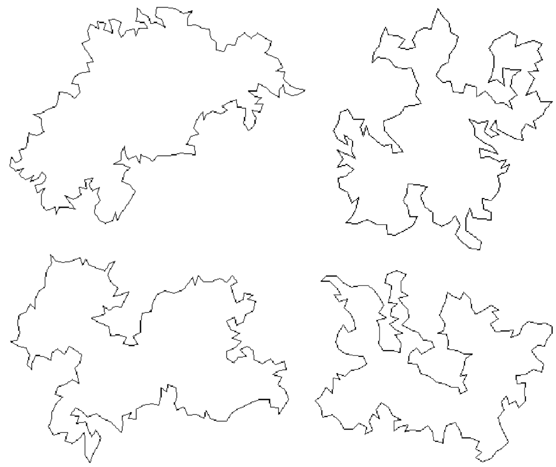
The probability f(s) is constrained by a number of observations, and is constructed by a maximum entropy principle.

Other Example: Closed Simple Curves

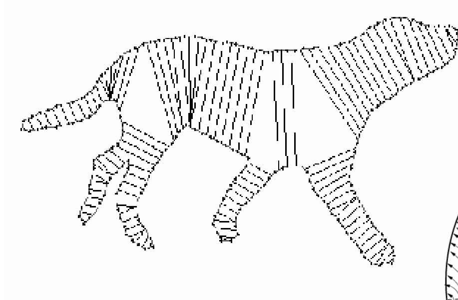
Example: 2D Flexible Shapes



Random Shapes by Markov Chain Walks

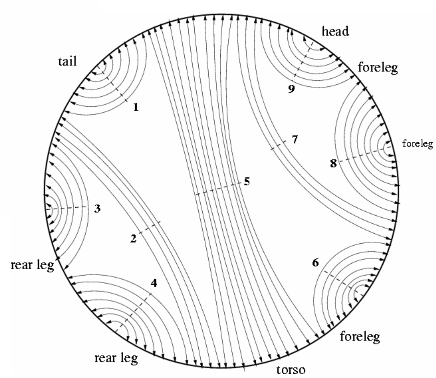


A Random Field for 2D Shape

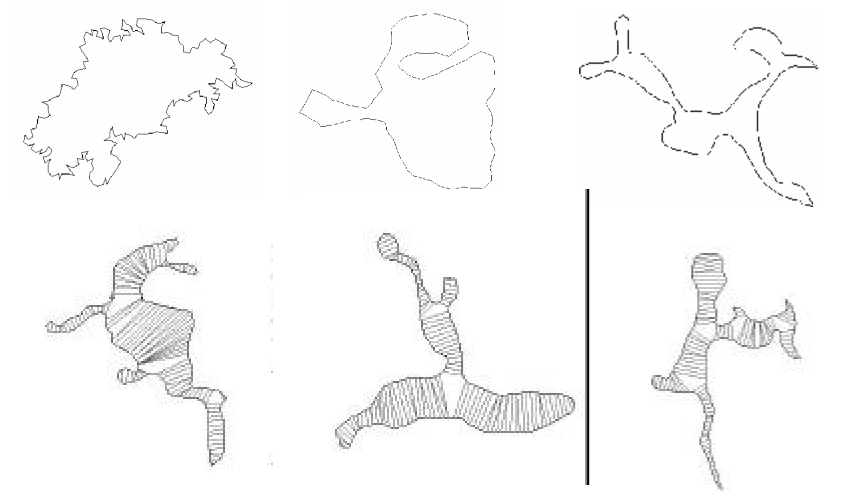


Co-linearity, co-circularity, proximity, parallelism, symmetry, …

The neighborhood

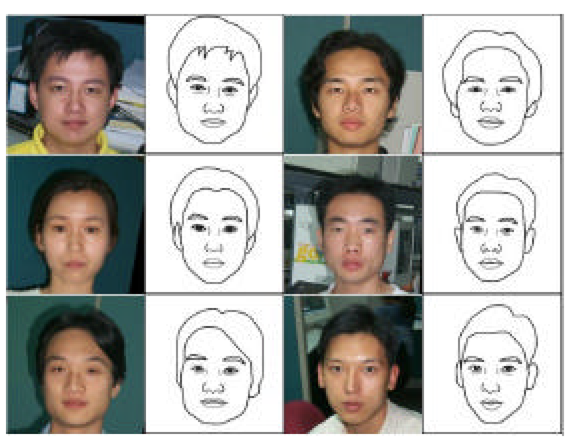


Example: Random Samples of 2D Shapes



(Zhu, 1999)

Face Sketch by Computer



(with H. Chen et al. at MSR, 2000-01)

Another example: Face Modeling

(Liu, Zhu, and Shum, ICCV, 2001)

Main References for This Lecture

Results presented in this lecture can be seen from the following papers.

1. S. C. Zhu, Y.N. Wu and D.B. Mumford, "Minimax Entropy Principle and Its Applications to Texture Modeling", *Neural Computation* Vol. 9, no 8, pp 1627-1660, Nov. 1997.

2. S.C. Zhu and D.B. Mumford, "Prior Learning and Gibbs Reaction-Diffusion", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol.19, no.11, pp1236-1250, Nov. 1997.

3. Y. N. Wu and S. C. Zhu, "Equivalence of Julesz Ensemble and FRAME models", *International Journal of Computer Vision*, 38(3), 247-265, July, 2000.

4. S. C. Zhu, "Embedding Gestalt Laws in Markov Random Fields", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, No.11, pp1170-1187, Nov, 1999.

A tutorial web page with many ref to other groups and texture results:
civs.stat.ucla.edu/Texture/General/Texture_general.htm