
A Mathematical Theory for Texture, Texton, Primal Sketch and Gestalt Fields

Song-Chun Zhu

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

Joint work with Y.Wu, C.Guo and D. Mumford, 1995-2003

UCLA Psychology, 2003.

Song-Chun Zhu

Julesz's Quest for texture perception



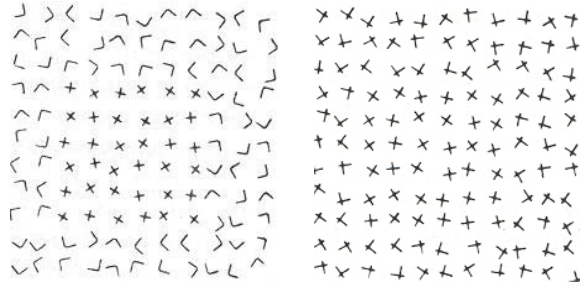
“What **features** and **statistics** are characteristics of a texture pattern, so that texture pairs that share the same features and statistics cannot be told apart by pre-attentive human visual perception?”

---1960s

UCLA Psychology, 2003.

Song-Chun Zhu

Psychophysics Experiments

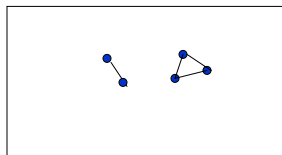


Texture discrimination in early vision (0.1-0.4sec).

Major Technical Difficulties

1. What kind of features and statistics should we search for?
2. A mathematical tool for mixing the exact amount of statistics for a given recipe.

Julesz and his school focused on k-gon statistics ($k=2,3$), which limited the scope of search. Unfortunately, our visual cortices do not compute k-gon statistics!



Neurophysiology Experiment



Huber and Weissel 1960s

Single neuron recording in the V1 area in cats and monkeys.

Return to Psychology

1. Bergen and Adelson 1991.
2. Chubb and Landy 1991.
3. Karni and Sagi 1991.

A conjecture:

“A texture pair cannot be told apart if they share the same histograms for a bank of Gabor response”

A reckless experiment by Heeger and Bergen 1995.

Two Technical 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 $m \times n$ pixels, then it cannot be defined on $(m+1) \times (n+1)$ pixels either.

Julesz ensembles

Given a set of statistics per pixel, $h = (h^{(i)} : i = 1, 2, \dots, k)$ for images on a finite lattice Λ , we define an equivalence class

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

H is an open set centered at h .

A Julesz ensemble $\Omega(h)$ is the limit of $\Omega_{\Lambda}(H)$ as $\Lambda \rightarrow Z^2$ under some boundary conditions.

Julesz Ensemble

As image lattice goes to infinity in the 2D plane,
statistical fluctuations diminishes and thus we obtain a
deterministic set.

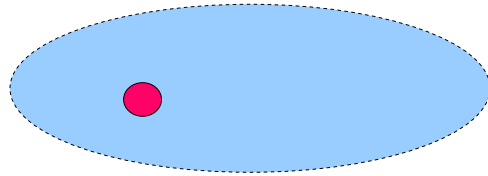
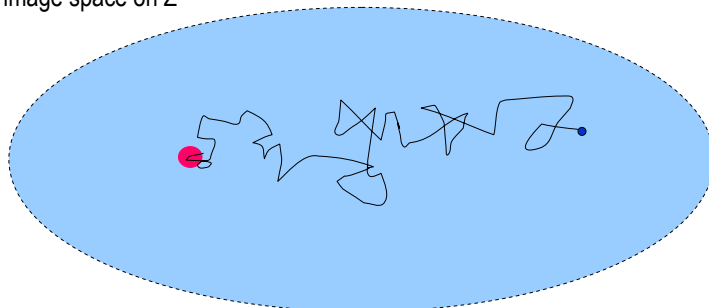


image space on Z^2

Simulation for Julesz Ensemble

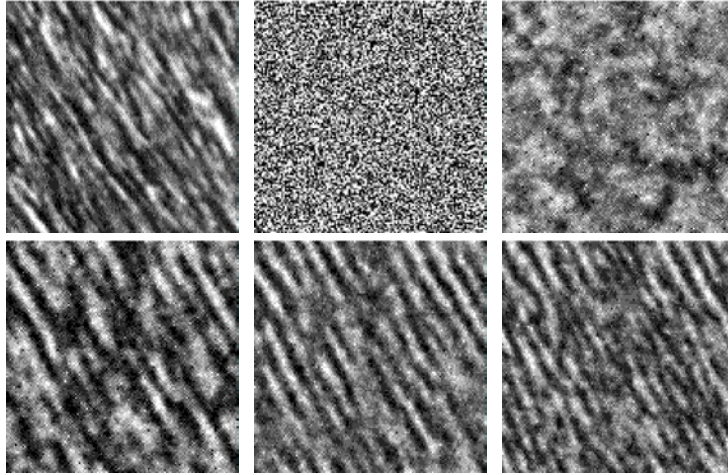
Draw random samples from the ensemble by
Markov chain Monte Carlo methods.

image space on Z^2



Each point in the space is a large image.

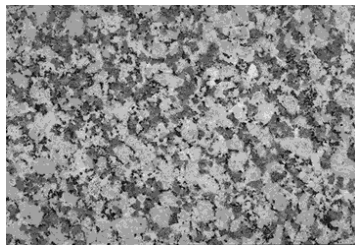
Example



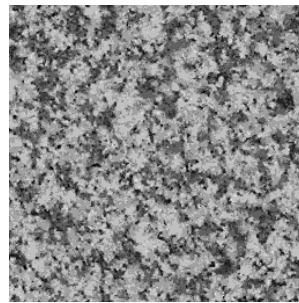
UCLA Psychology, 2003.

Song-Chun Zhu

Example: texture modeling



Observed

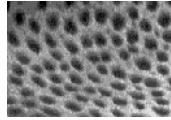


MCMC sample

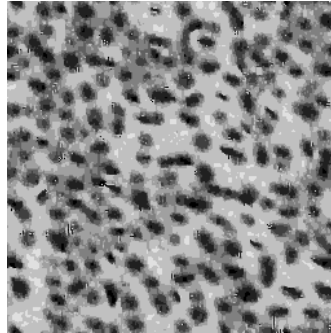
UCLA Psychology, 2003.

Song-Chun Zhu

Example: texture modeling

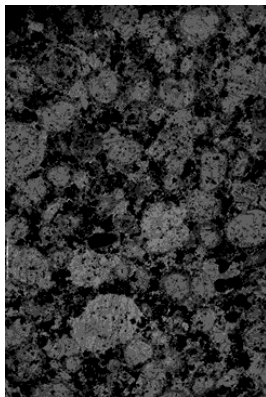


Observed

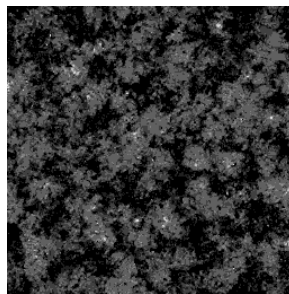


MCMC sample

Example: texture modeling

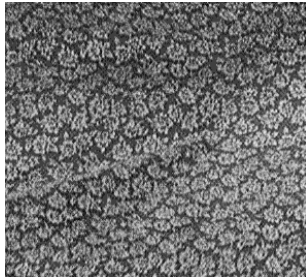


Observed

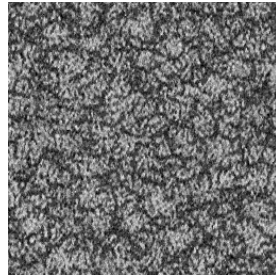


MCMC sample

Example: texture modeling

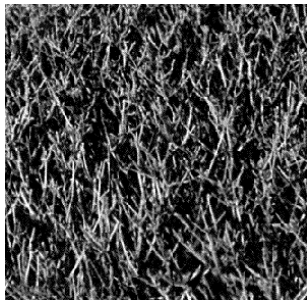


Observed

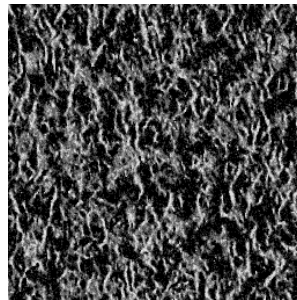


MCMC sample

Example: texture modeling

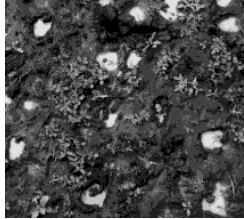


Observed

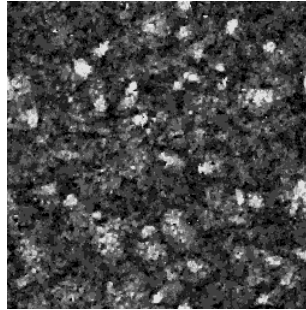


MCMC sample

Example: texture modeling



Observed



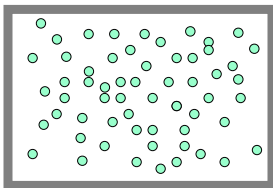
MCMC sample

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,

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

Definition of a Visual Pattern

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

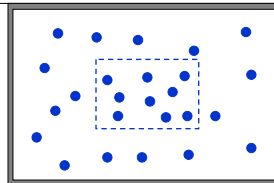
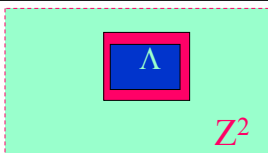
For a homogeneous signal I on 2D lattice Λ ,

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

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

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

Equivalence of Julesz ensemble and FRAME models



Theorem

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

Theorem

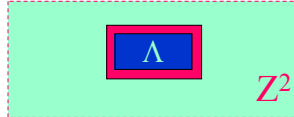
As the image lattice goes to infinity, $f(I; h_c)$ is the limit of the FRAME model $p(I_\Lambda | I_{\partial\Lambda} : \beta)$, in the absence of phase transition.

$$A \text{ texture} \longleftrightarrow h_c \longleftrightarrow \beta$$

Observation in Statistical Physics

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

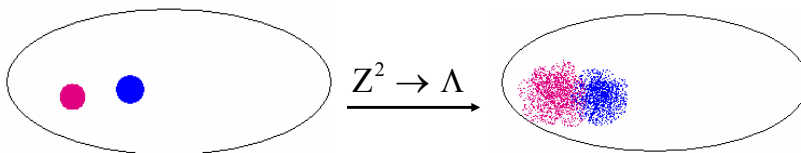
"If a system of a great number of degrees of freedom is micro-canonically distributed in phase, any very small part of it may be regarded as canonically distributed" --- Gibbs, 1902.



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 !!!

Relationship between Conceptualization and Modeling



texture ensembles :

$$f(I; h_c)$$

texture models :

$$p(I_\Lambda | I_{\partial\Lambda}; \beta)$$

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

$$p(I_\Lambda | I_{\partial\Lambda}; \beta) = \frac{1}{Z(\beta)} \exp\left\{-\sum_{j=1}^k \beta_j h_j(I_\Lambda | I_{\partial\Lambda})\right\}$$

Maximum Entropy Model of Texture

Solving this constrained optimization problem yields:

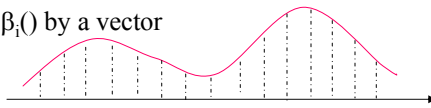
The FRAME model (Filters, Random fields And Maximum Entropy) (Zhu, Wu, Mumford, 1997)

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

$F = \{ F_1, F_2, \dots, F_k \}$ are selected filters (wavelets)

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

Approximating $\beta_i()$ by a vector



Minimax Entropy Learning

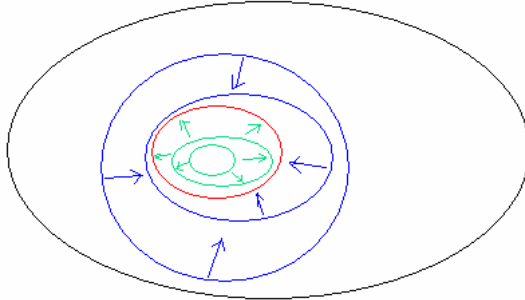
For a Gibbs (max. entropy) model p , this leads to the minimax entropy learning principle (Zhu, Wu, Mumford 96,97)

$$p^* = \arg \min_F \{ \max_{\beta} \text{entropy}(p(I; \beta, F)) \}$$

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*).

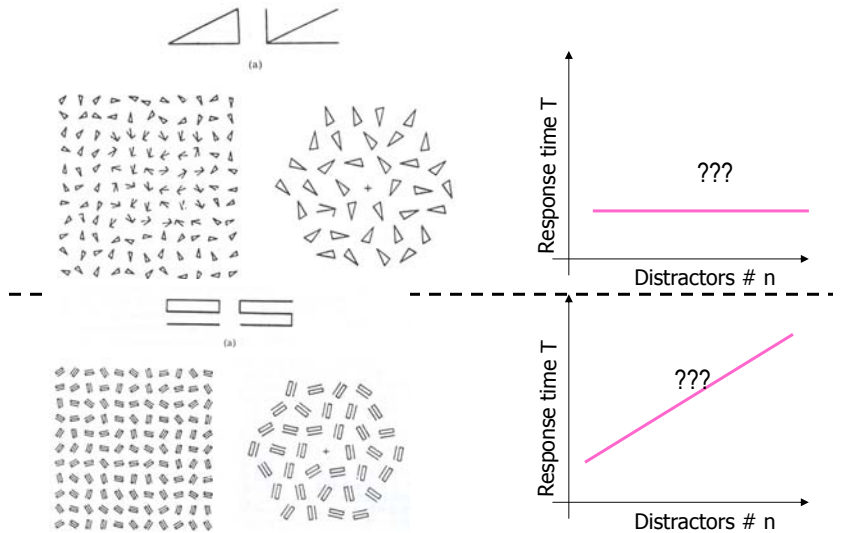
Textures vs Geometry:



Textures and Geometry are intimately blended in natural images and our perception Must be switching between them over scales.

Textons: Fundamental elements in Visual perception

(Julesz, 1981 Textons: the fundamental elements of visual perception, Nature)



UCLA Psychology, 2003.

Song-Chun Zhu

“Early Vision”

Julesz's heuristic (or axiom):

Textons are the fundamental elements in preattentive vision,

1. Elongated blobs
2. Terminators
3. Crossings

Texton plasticity:

Textons in human perception may change by training !

(Karni and Sagi, 1991)

What are the textons for the ensemble of natural images?

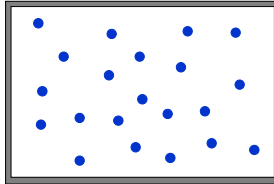
UCLA Psychology, 2003.

Song-Chun Zhu

Review: Ensembles in Statistical Mechanics

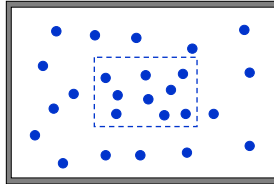
(See a stat. physics book by Chandler 1987)

$$N = 10^{23}$$



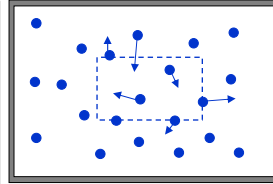
Micro-canonical Ensemble

$$N_1 = 10^{23}, N_2 = 10^{18}$$



Canonical Ensemble

$$N_1 + N_2 = 10^{23}$$



Grand-Canonical Ensemble

What are the basic elements in the ensemble of visual patterns?

We assumed pixels, points in typical Gibbs models. This should be generalized.

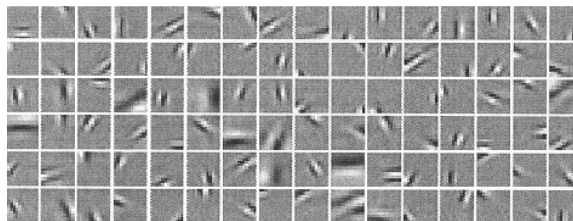
Very likely, the basic elements vary from different ensembles, and thus they need to be learned automatically from natural images --- for generic vision models.

Sparse coding (Olshausen and Fields, 95 nature).

Learning an over-complete image basis from natural images

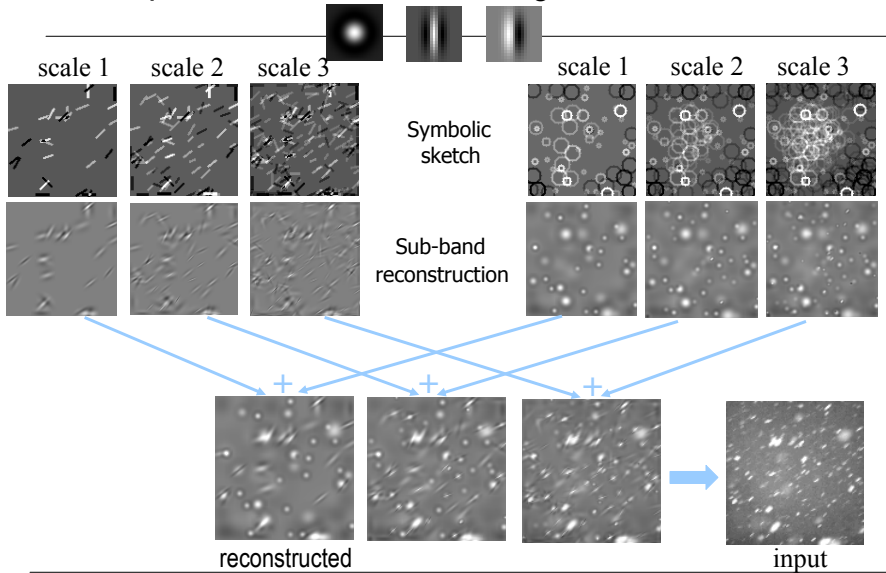
An image I is assumed to be a linear addition of some image bases ψ_i , $i=1,2, \dots, n$ which are selected from an over-complete basis (dictionary).

$$I = \sum_i \alpha_i \psi_i + n, \quad \alpha_i \sim p(\alpha) \text{ iid}$$



It was said that these learned bases resemble cells in primate V1.

Example of Coarse-to-fine image reconstruction



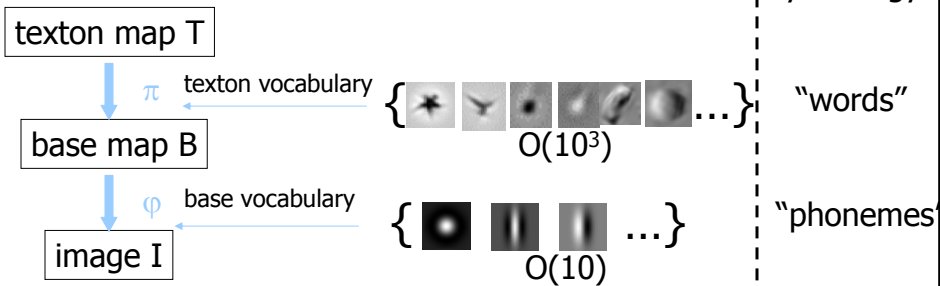
UCLA Psychology, 2003.

Song-Chun Zhu

A Three Level Generative image model

Textons are defined as a vocabulary associated with a generative image model.

A two level image model:



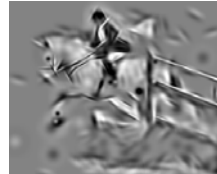
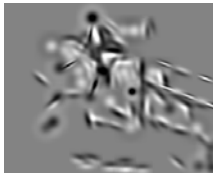
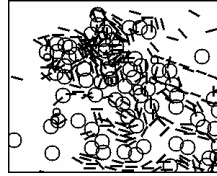
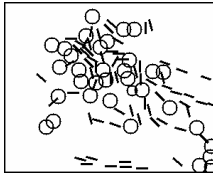
UCLA Psychology, 2003.

Song-Chun Zhu

Problems with linear additive models and wavelet dictionaries



Reconstruct image by
Gabor/LoG

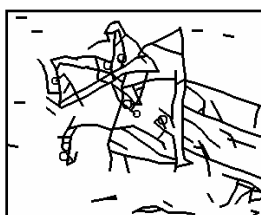


1. Missing the semantics structures
2. Not sparse enough!

Primal sketch of generic images [Guo, Zhu and Wu, 2003]



input image



Primal sketch



sketching pursuit

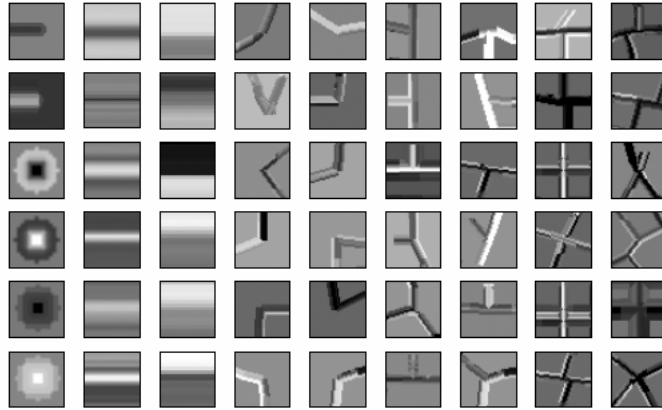


synthesized image



sketch image

Image primitives learned from natural images



UCLA Psychology, 2003.

Song-Chun Zhu

Parameters Used

Image Size	300*200
Sketchable Pixels	18,185 ~ 25%
Primitive Number	230
Primitive Width	7
Primitive Parameters	2,350 ~ 3.5%
MRF Parameters	$5*7*13=455$
Total Parameters	2,805 ~ 4.7%

UCLA Psychology, 2003.

Song-Chun Zhu

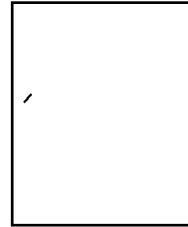
Primal sketch of generic images [Guo, Zhu and Wu, 2003]



input image



primal sketch



sketching pursuit



synthesized image



sketch image

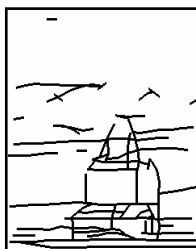
UCLA Psychology, 2003.

Song-Chun Zhu

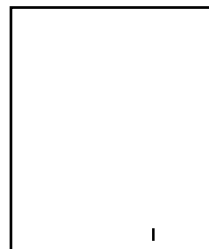
Primal sketch of generic images [Guo, Zhu and Wu, 2003]



input image



primal sketch



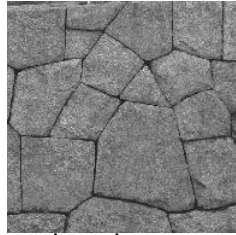
sketching pursuit



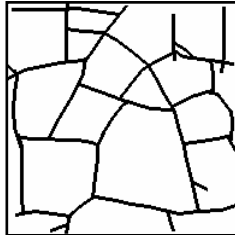
UCLA Psychology, 2003.

Song-Chun Zhu

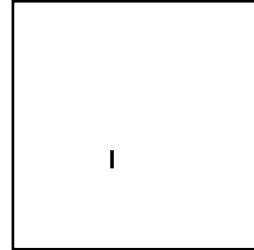
Primal sketch of generic images [Guo, Zhu and Wu, 2003]



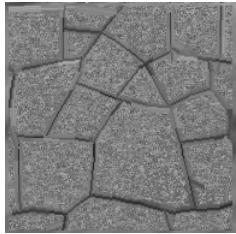
input image



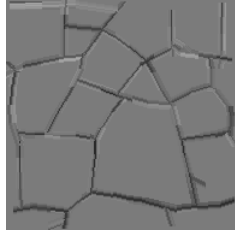
Primal sketch



sketching pursuit



synthesized image



sketch image

UCLA Psychology, 2003.

Song-Chun Zhu

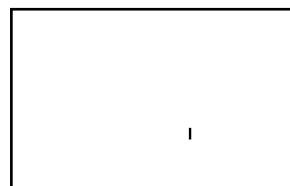
Primal sketch of generic images [Guo, Zhu and Wu, 2003]



input image



primal sketch



sketching pursuit



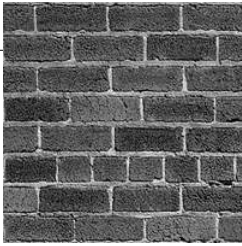
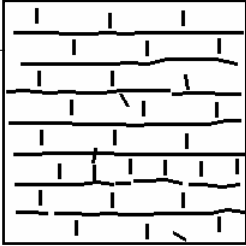
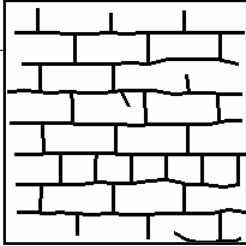
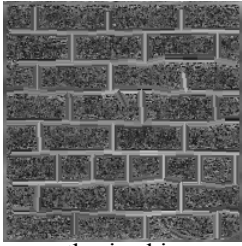
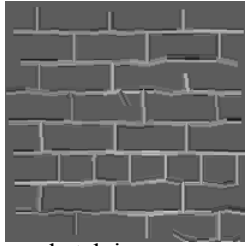
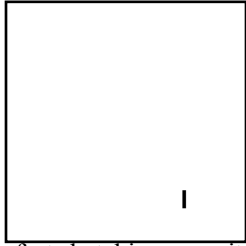
synthesized image



sketch image


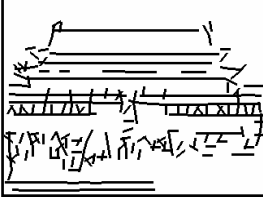


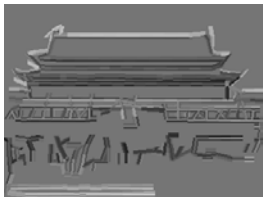

UCLA Psychology, 2003.

Song-Chun Zhu

		
input image	fast sketching pursuit result	fast sketching refinement result
		
synthesized image	sketch image	fast sketching pursuit and refinement procedure

UCLA Psychology, 2003. Song-Chun Zhu

Primal sketch of generic images [Guo, Zhu and Wu, 2003]

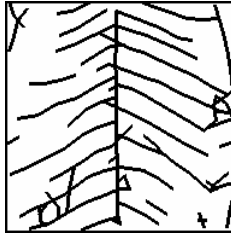
		
input image	fast sketching pursuit result	fast sketching refinement result
		
synthesized image	sketch image	fast sketching pursuit and refinement procedure

UCLA Psychology, 2003. Song-Chun Zhu

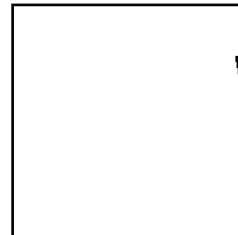
Primal sketch of generic images [Guo, Zhu and Wu, 2003]



input image



primal sketch



sketching pursuit



synthesized image

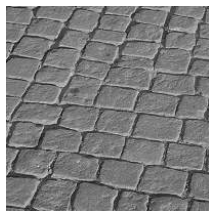


sketch image

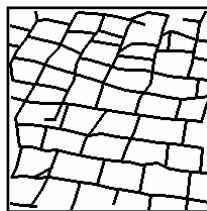
UCLA Psychology, 2003.

Song-Chun Zhu

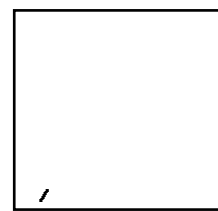
Primal sketch of generic images [Guo, Zhu and Wu, 2003]



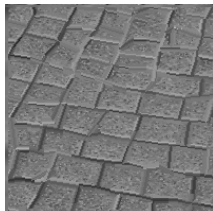
input image



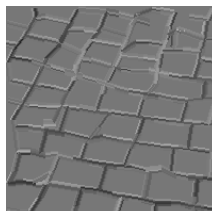
primal sketch



sketching pursuit



synthesized image



sketch image

UCLA Psychology, 2003.

Song-Chun Zhu

Primal sketch of generic images [Guo, Zhu and Wu, 2003]



input image



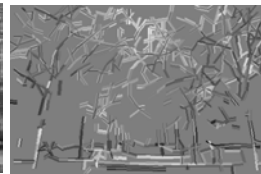
primal sketch



sketching pursuit



synthesized image



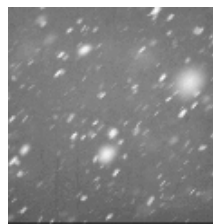
sketch image

UCLA Psychology, 2003.

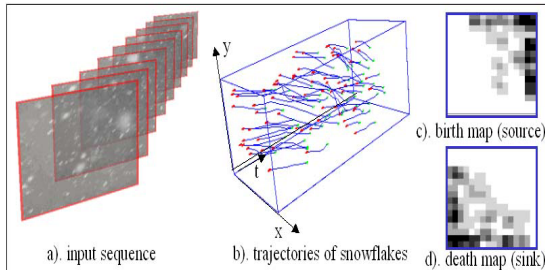
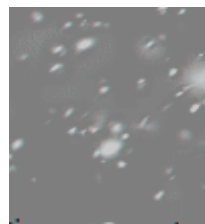
Song-Chun Zhu

“Motons” Moving textons---the snow flakes

Observed
Sequence



Synthesized
Sequence

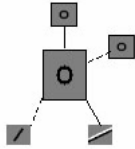


UCLA Psychology, 2003.

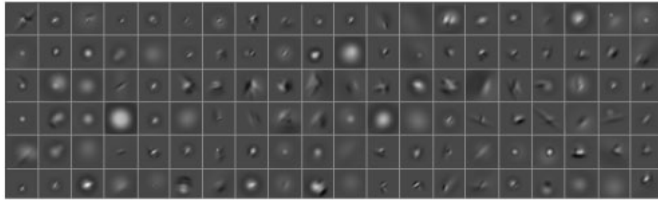
Song-Chun Zhu

Motons: snow flakes

For instance, a texton template π for the snowflake is shown below. Then 120 random snowflake instances are sampled randomly from π for a proof of variety.



a texton template π

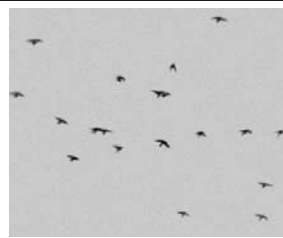


many texton instances randomly sampled from π

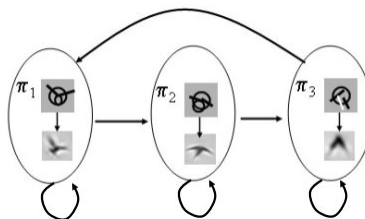
Motons: flying birds



Observed Sequence

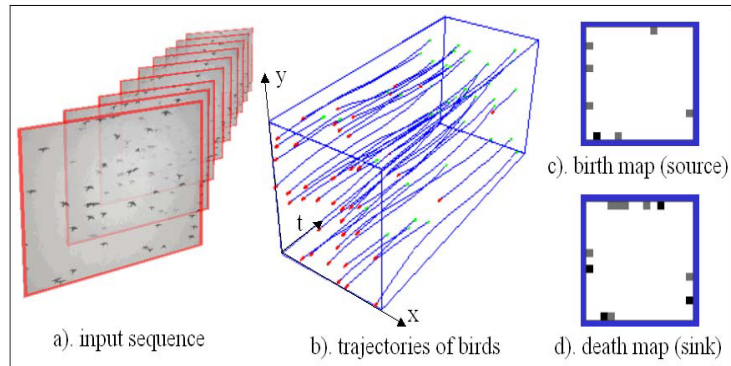


Synthesized Sequence



Experimental Results

- Flying Birds

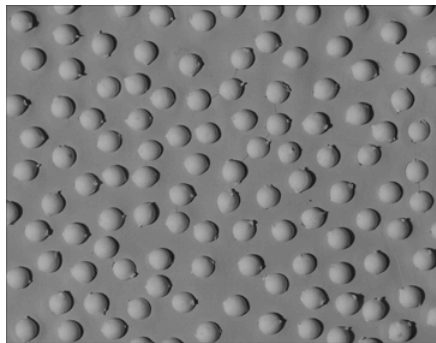


UCLA Psychology, 2003.

Song-Chun Zhu

“Lightons” ---Texon with lighting variations

Extend the generative model to 3D and lighting:
--- often accurate 3D depth is not recoverable
or unnecessary



Some observed images at varying lighting conditions

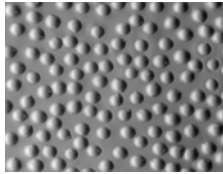
UCLA Psychology, 2003.

Song-Chun Zhu

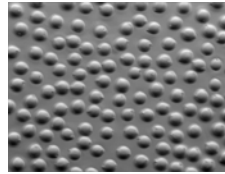
“Lightons”---Texton with lighting variations

By SVD, one can recover the three image bases b_1, b_2, b_3

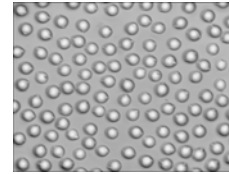
$$I = s_1 b_1 + s_2 b_2 + s_3 b_3 \quad (s_1, s_2, s_3) \text{ is the light.}$$



b_1



b_2



b_3

Each image base b is further decomposed into a linear sum of textons. Comparing with $I = \sum_i \alpha_i \psi_i$

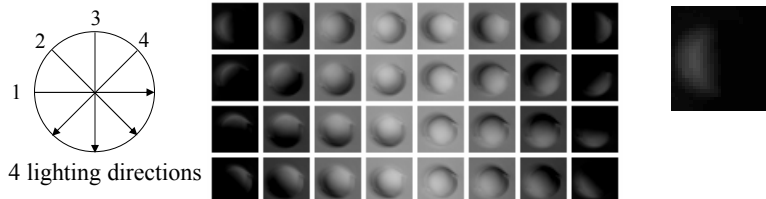
Texton with lighting variation

Each element is represented by a triplet of textons

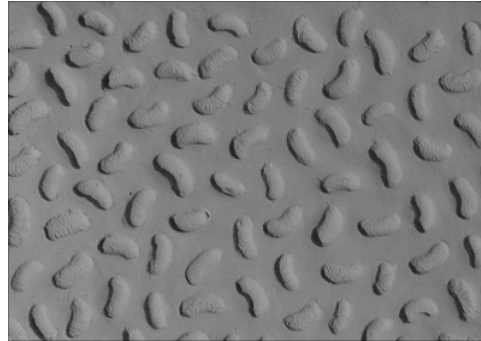
“lighton”? 

sketch 

Sampling the 3D elements under varying lighting directions



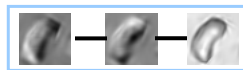
Example 2



input images

Example 2

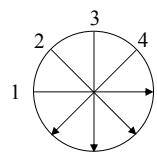
“Lighton”



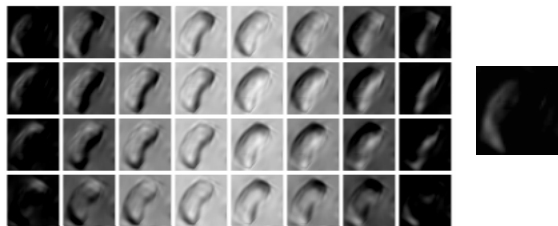
sketch



Sampling the 3D elements under varying lighting directions



4 lighting directions



Summary: From Bases to Textons

Textons are atomic structures in natural images
Mathematically, textons form a vocabulary associated with a generative image model $p(I; \Theta)$, each texton is a triplet specified by 4 groups of parameters:

$$\text{a texton} = \left\{ I = (s_1, s_2, s_3)' (b_1(\phi), b_2(\phi), b_3(\phi)) : \right. \\ \left. \begin{array}{l} (x, y, \sigma, \theta): \text{similarity transform,} \\ \pi: \text{geometric deformations,} \\ (A, B, C, D): \text{dynamics,} \\ \text{lighting variations } (s_1, s_2, s_3), \end{array} \right\}$$

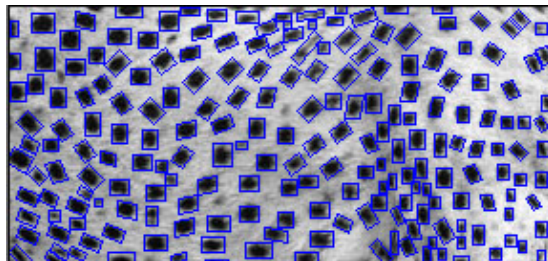
Treat a Texton Map as an Attributed Point Process

One layer of hidden variables: the texton map

Ψ --- a "texton" (a min-template, a mother wavelet)

$$T = \{n, (x_i, y_i, \theta_i, s_i, \alpha_i), i = 1, 2, \dots, n\}$$

For texton #, translation, rotation, scale, contrast, ...



Markov Random Field Model for Gestalt Fields

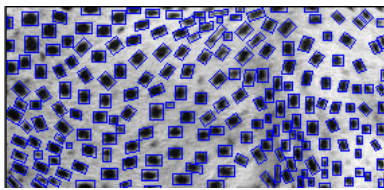
$$T = \{n, (x_i, y_i, \theta_i, s_i, \alpha_i), i = 1, 2, \dots, n\}$$

The texton map is governed by a Gibbs distribution

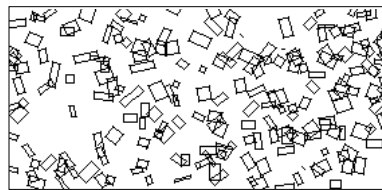
$$p(T; \beta) = \frac{1}{Z} \exp \left\{ -\beta_0 n - \sum_i \langle \beta_i, h_i(T) \rangle \right\}$$

Where β_0 controls the density of the textons in unit area and $h(T)$ captures some spatial statistics related to Gestalt psychology. The model can also be derived from maximum entropy principle. The parameters β are learned from MLE.

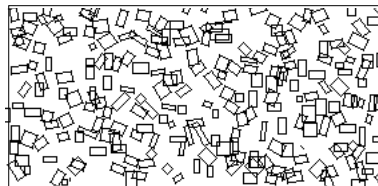
Experiment II: Modeling Texton Maps



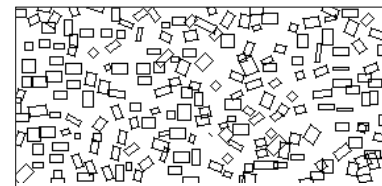
input



T=1



T=30

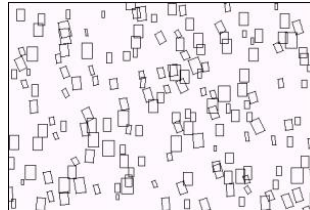


T=234

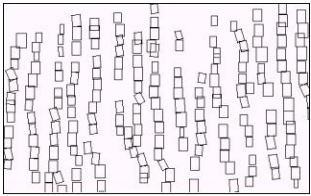
Experiment II: Modeling Texton Maps



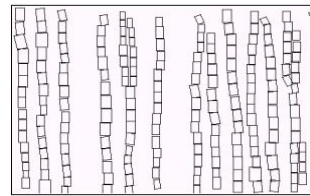
input



T=1



T=30

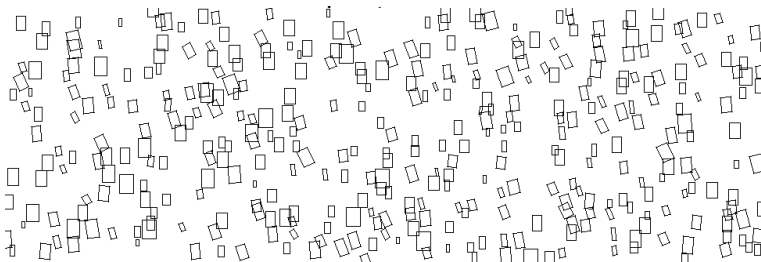


T=332

UCLA Psychology, 2003.

Song-Chun Zhu

Modeling Texton Maps



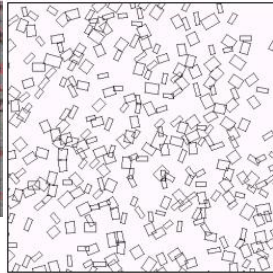
UCLA Psychology, 2003.

Song-Chun Zhu

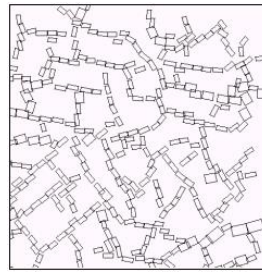
Experiment II: Modeling Texton Maps



input



T=1

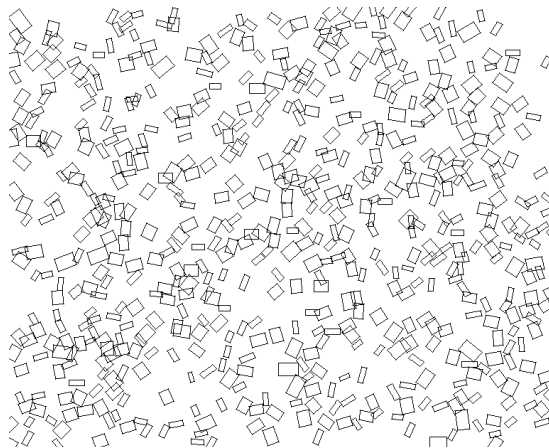


T=202

Modeling Texton Maps



input



A Unified Theory

Input: an ensemble of images

$$S = \{ I_1, I_2, \dots, I_N \} \sim f(I)$$

Output: a probabilistic model

$$p(I) \rightarrow f(I)$$

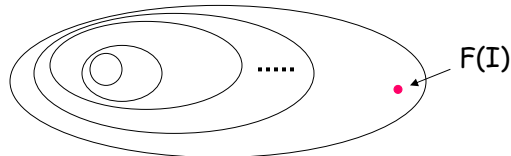
The model approaches the true density by minimizing the Kullback-Leibler Divergence:

$$p^* = \underset{p \in \Omega}{\operatorname{argmin}} D(f || p) = \underset{p \in \Omega}{\operatorname{argmax}} \sum_{j=1}^N \log(p(I_j)); I_j \sim f(I)$$

Pursuit of the probabilistic families

We pursue models over a sequence of nested probability families

$$\Omega_1 \subset \Omega_2 \subset \Omega_3 \cdots \subset \Omega_K$$



Ways to augment the probability families:

1. Adding more parameters in the Gibbs models
2. Adding more features
3. Introducing hidden variables (hierarchical graph) for larger structures ---textons
4. Introducing dynamic address variables for graph structures---Gestalt fields

All can be done in a unified and principled way !!!