Visual Learning with Explicit and Implicit Manifolds

--- A Mathematical Model of Texture, Texton, and Primal Sketch

Song-Chun Zhu

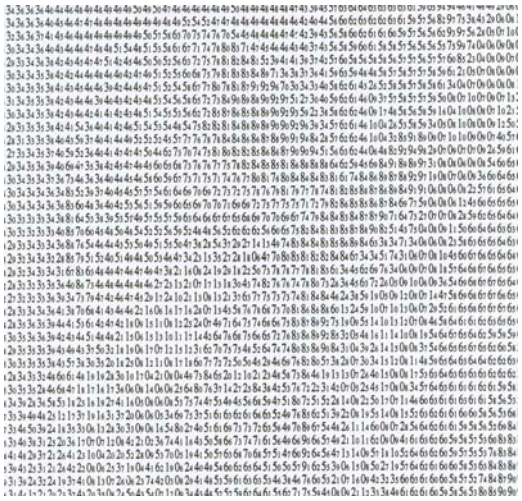
Departments of Statistics and Computer Science
University of California, Los Angeles

www.stat.ucla.edu/~sczhu

Vision is to understand the special language in images

A math/stat view
what are the **structures** and **models** in the image space?

Vision is becoming a **compiler theory**:
what are the **visual words** and **image grammar** ?



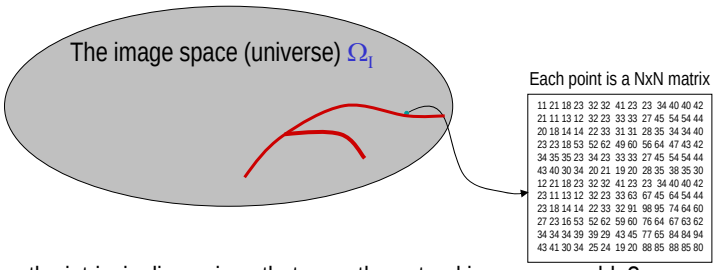
The image space

Consider an image I with 256 grey levels on a lattice Λ of NxN pixels, N=256.

$$\text{The volume of image space } |\Omega_I| = 2^{8\times256\times256} = 10^{157,830}$$

$$\text{The volume of natural image ensemble } |\Omega_f| \cong 2^{0.3\times256\times256} \cong 10^{5,718}$$

$$\text{The volume of images seen by humans } \le 10^{10}\times10^{10} = 10^{20}$$



1. What are the intrinsic dimensions that span the natural image ensemble?
2. What are the structures of the natural image ensemble?
   (related to neuron functions and perception)

Consider the image space of small image patches (i.e. 7x7 pixels)

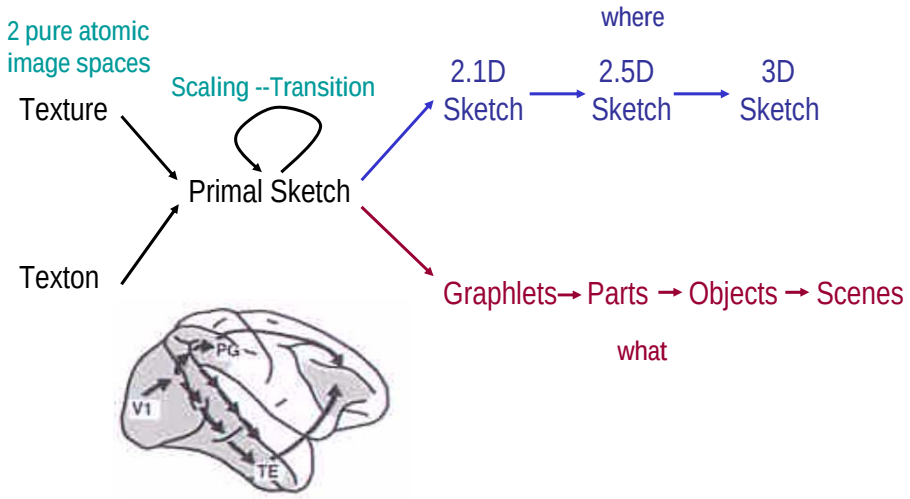
By analogy: A picture of the universe

At different temperatures we observe different entropy patterns.

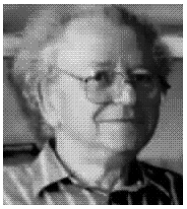
At different scales, different forces rule the systems.

A photo from Cosmology. Our image space of small patches is very much like this, it contains patterns of wide range of entropy regimes.

A road map for visual representation



What is a texture?



The Julesz Quest

"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?"

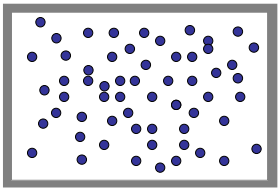
---1960--80s

This is similar 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

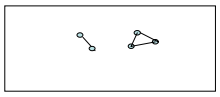
$$S = (x^N, p^N)$$

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

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

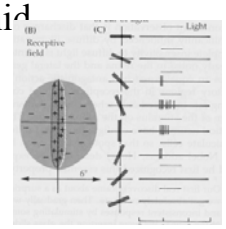
What are the Essential Statistics to vision?

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

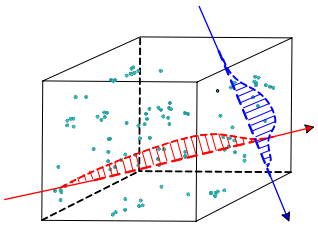
Heeger and Bergen, Siggraph 1995,
Zhu, Wu, and Mumford 1996, … …

Formulated in Statistical learning

Conceptualizing a visual concept (texture here) to a set of images that satisfy certain statistical constraints --- an implicit manifold.

$$\text{A texture class} = \Omega(h) = \{\mathbf{I} : H_i(\mathbf{I}) = h_i, i = 1, 2, ..., K\}.$$

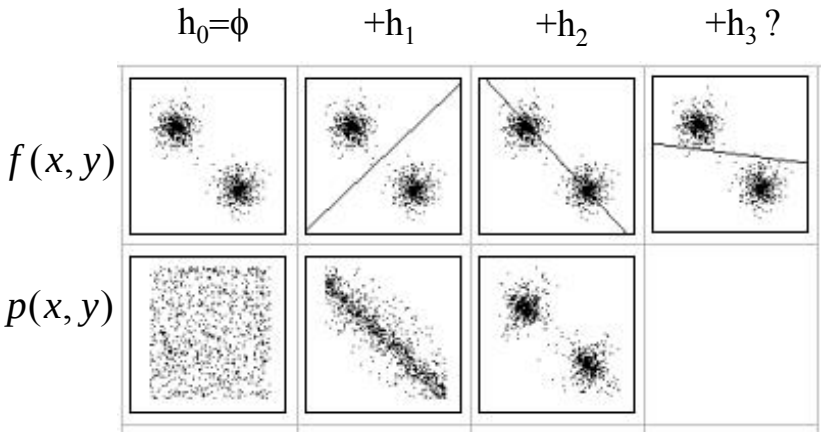
$$h = (h_1, h_2, ..., h_K) \text{ for large enough images}$$



Examples:
- Gibbs,
- MRF,
- FRAME,
- Mixed Markov model

Model pursuit: choose informative features and statistics to minimize the log-volume of the set or Shannon entropy of the model --- minimax entropy principle.

Toy Example: Estimating 2D Distribution

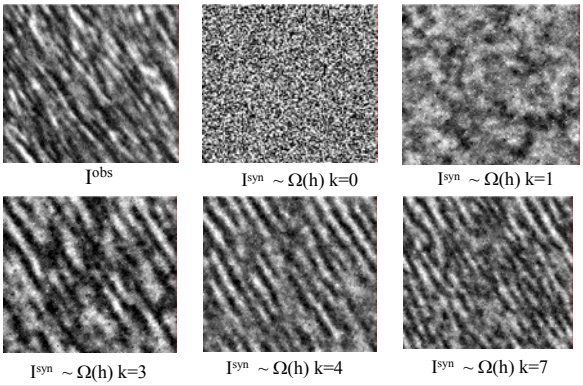


Please note, it is not match pursuit !

Example: a texture pattern is an implicit manifold

$$\text{a texture} = \Omega(h_c) = \{ I : \lim_{\Lambda \to Z^2} \frac{1}{|\Lambda|} \sum_{(i,j) \in \Lambda} h(I_{(i,j)}) = h_c , \ \ | h_c | = k \}$$

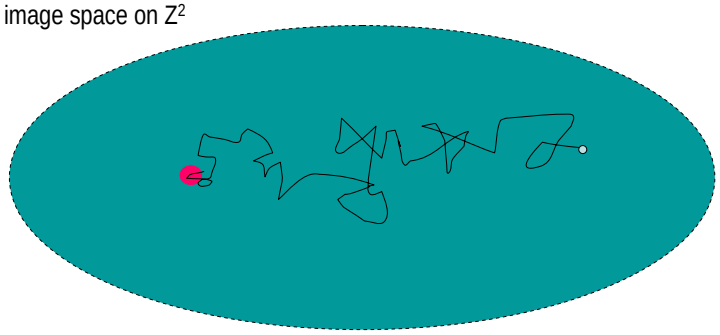
H_c are histograms of Gabor filters, i.e. marginal distributions of $f(I)$



(Zhu et al, 1996-01)

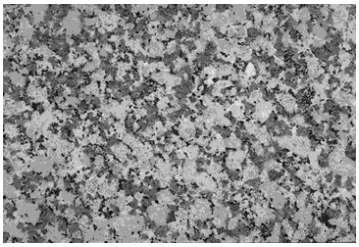
Simulation for Julesz Ensemble

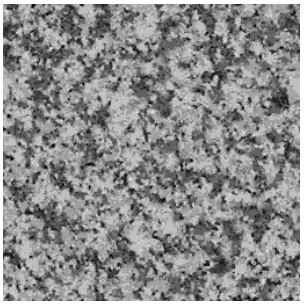
Draw random samples from the ensemble by Markov chain Monte Carlo methods.



Each point in the space is a large image.

Example 2: texture modeling

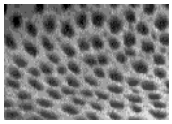


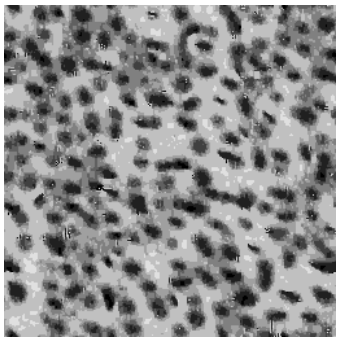


Observed MCMC sample

(Zhu et al, 1996-01)

Example 3: texture modeling

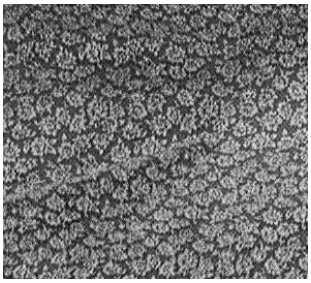




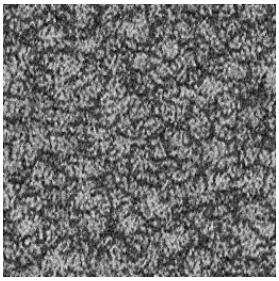
Observed MCMC sample

(Zhu et al, 1996-01)

Example 4: texture modeling



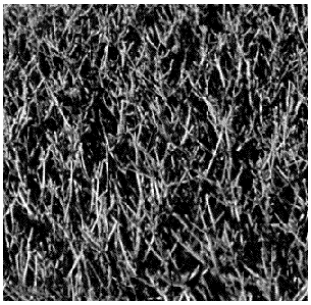
Observed



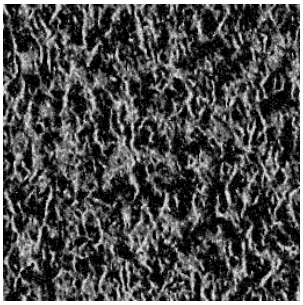
MCMC sample

(Zhu et al, 1996-01)

Example 5: texture modeling



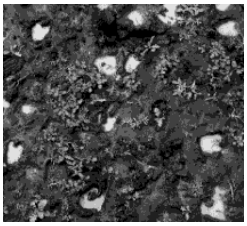
Observed

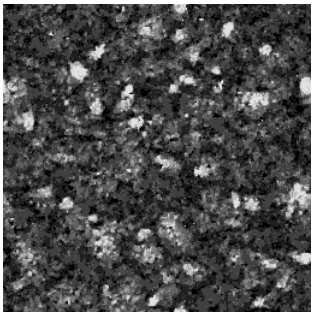


MCMC sample

(Zhu et al, 1996-01)

Example 5: texture modeling

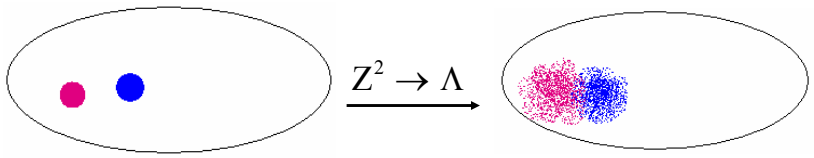




Observed MCMC sample

(Zhu et al, 1996-01)

Relationship between Conceptualization and Modeling



texture ensembles :

$f(I;\ h_c)$

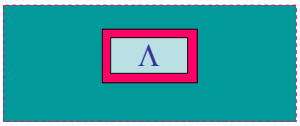
texture models :

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

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

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

Equivalence of Julesz ensemble and FRAME models



Theorem 3

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 $p(I_\Lambda \mid I_{\partial\Lambda} : \beta)$

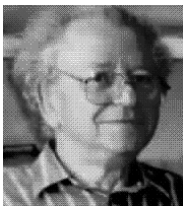
Theorem 4

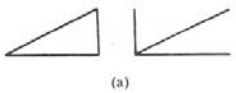
As the image lattice goes to infinity, $f(I; h_c)$ is the limit of the FRAME model $p(I_\Lambda \mid I_{\partial\Lambda} : \beta)$, in the absence of phase transition.

A texture $\Longleftrightarrow$ h_c $\Longleftrightarrow$ β

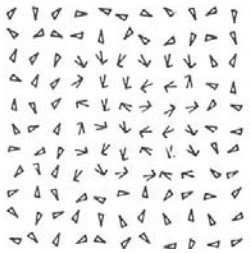
Texture vs Texton

While texture is a macroscopic (collective) concept, Texton refers to the individual atomic elements that we are sensitive to.

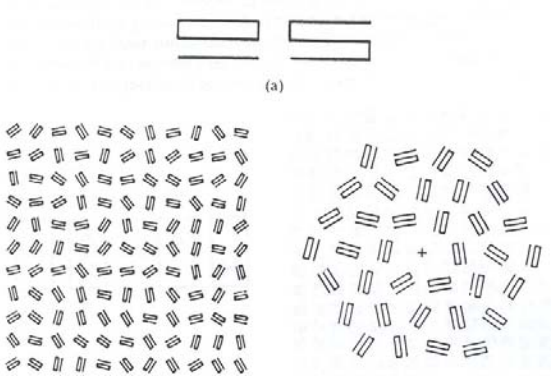




(a)



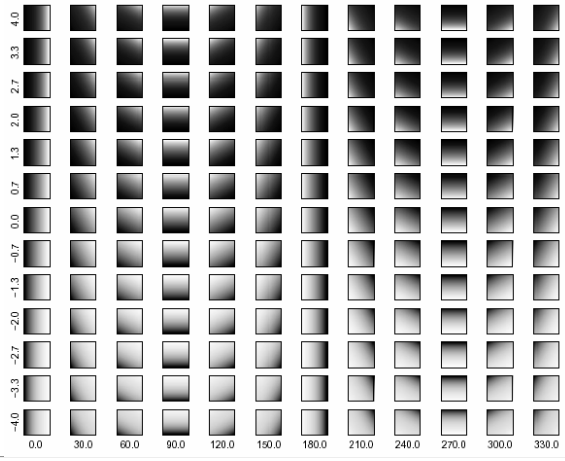
Texture vs Texton



In this example, the detection time is proportional to the number of background elements, And thus suggests that the subject is doing element-by-element scrutiny.

Example: A 3-dimen manifold of step edge

Each texton element spans a low dimensional manifold in the image space, whose density goes to infinity as we approach the manifold perpendicularly.



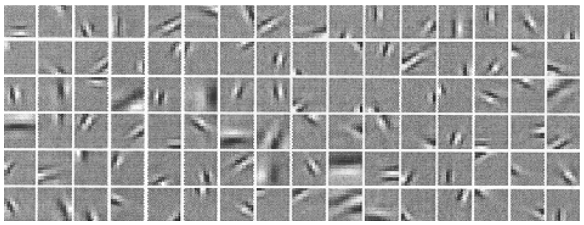
Lee, Peterson, Mumford 2000

Searching for image elements 1: sparse coding

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, ……,n which are selected from an over-complete basis (dictionary).

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



(Olshausen and Fields, 1995).

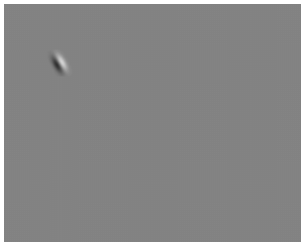
Sparse coding: generative model of image

Generative models: decompose the original signal into a set of components.

$$I = \bigoplus_{i=1}^{N} c_i \psi_i + \varepsilon, \psi_i \in \Delta$$

$\bigoplus$ Composition (linear/non-linear)

input image

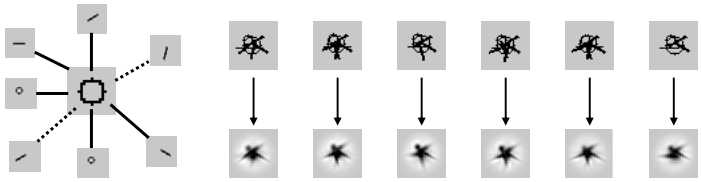


matching pursuit

Δ: a dictionary of over-complete Gabor, LoG bases.

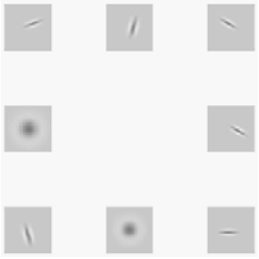
The sparse coding can be seen as a simple stochastic grammar.

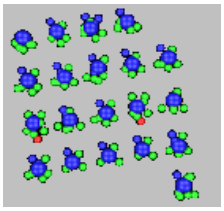
Searching for textons 2: aligning image bases



texton template

a sample of texton instances



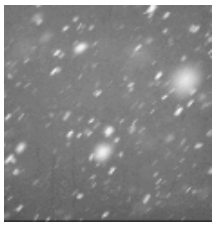


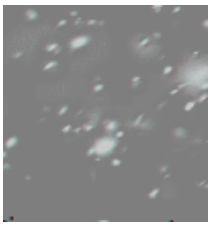
Guo and Zhu 2002

"atomic" model

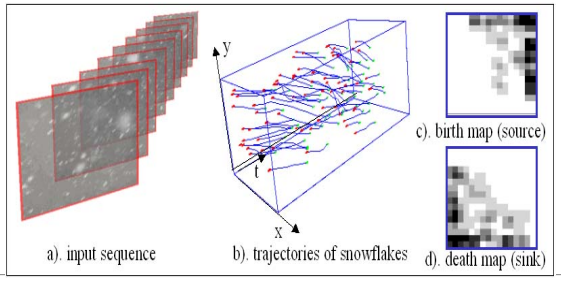
Identifying textons in motion

Observed Sequence





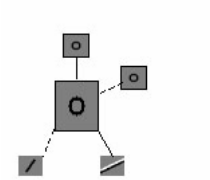
Synthesized Sequence



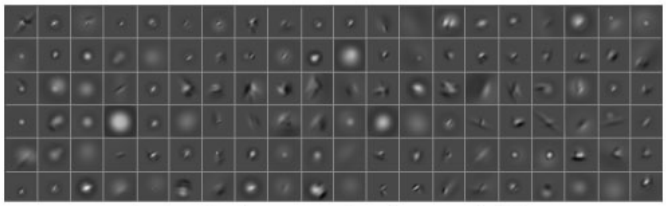
Wang,Zhu 2003

A texton class: snowflake

For instance, a texton class 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



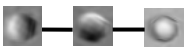
many texton instances randomly sampled from π

Wang/Zhu 2003

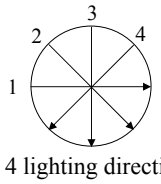
Texton class: lighting variations

Each element is represented by a triplet of textons

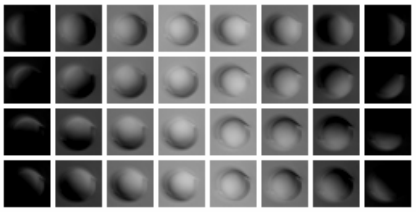
"lighton"?



Sampling the 3D elements under varying lighting directions



4 lighting directions



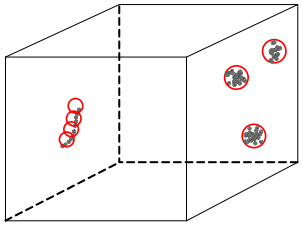
Wang,Zhu 2003

Summary on textons

A texton is an equivalence class of images, and it is defined explicitly

$$\text{A texton class} = \{I : I = f(w; \Delta),\ \ w \sim p(w)\}$$

$$\text{A texture class} = \{I : H(I) = h,\ \Lambda \to \infty\}.$$



W are the hidden variables spanning the dimensions of the manifold, and Δ is the vocabulary of the generative function.

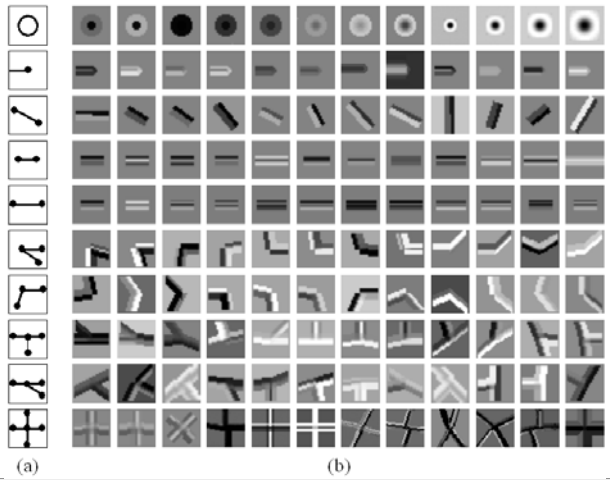
Model pursuit: choose optimal dictionary (*epsilon-balls*) to cover the maximum probability mass or minimize the Kolmogorov entropy

--- again, minimax entropy

It is closely related to binding with maximum mutual information.

Other manifolds of "hyper-sparse" image primitives

Learned texton/primitive dictionary with some landmarks that transform and warp the patches



Consider the image space of small image patches (i.e. 5x5 pixels)

By analogy: A picture of the universe

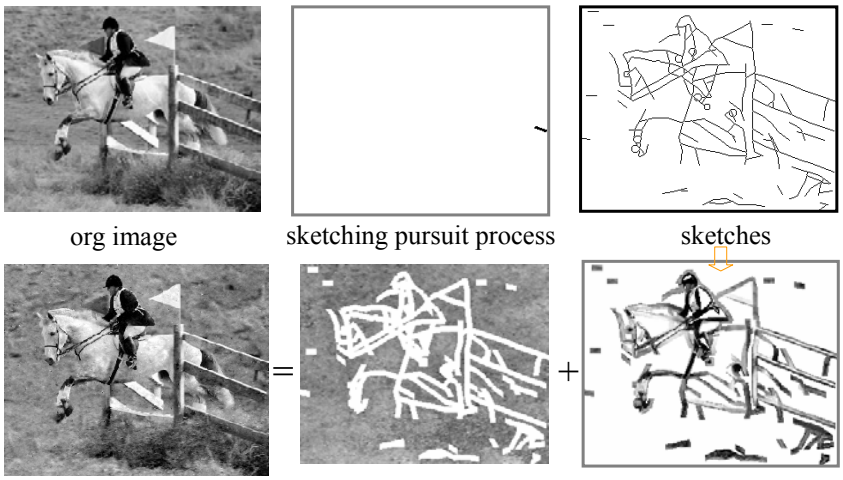
At different temperatures we observe different entropy patterns.

At different scales, different forces rule the systems.

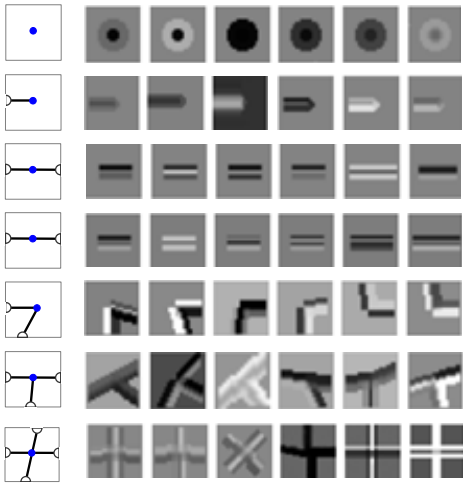
A photo from Cosmology. Our image space of small patches is very much like this, it contains patterns of wide range of entropy regimes.

Primal sketch: integrating the two regimes

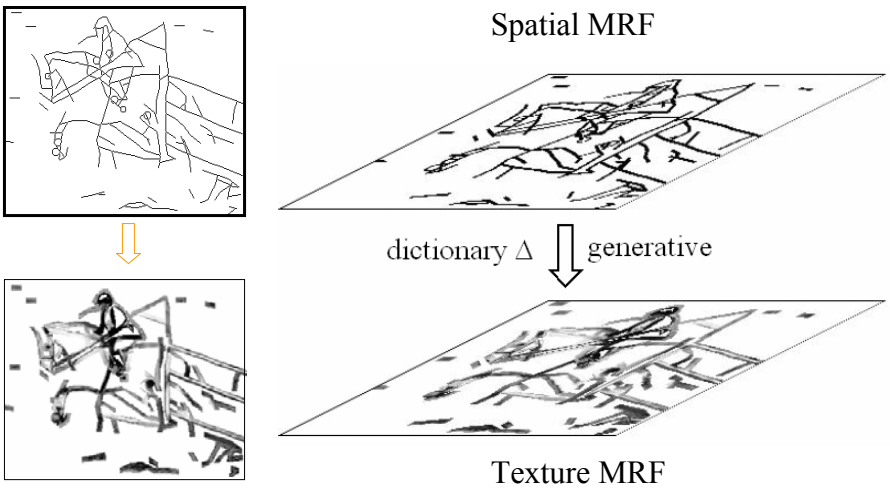
(Guo,Zhu,Wu, 2003-05)



Examples of the dictionary of image primitives

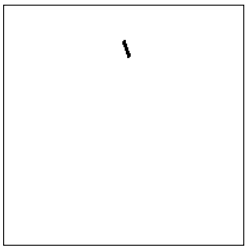


Primal Sketch is a two-level Markov random field model



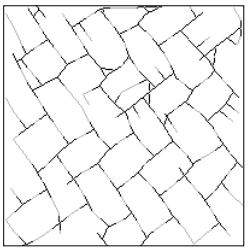
Primal sketch example

input image



sketching pursuit process

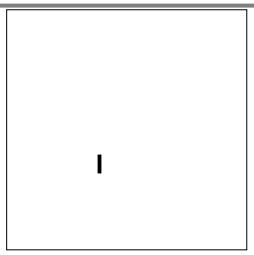
synthesized image



sketches

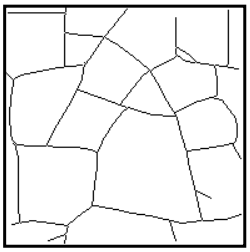
Primal sketch example

original image



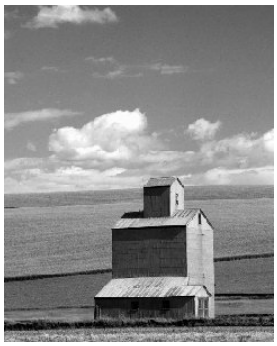
sketching pursuit process

synthesized image

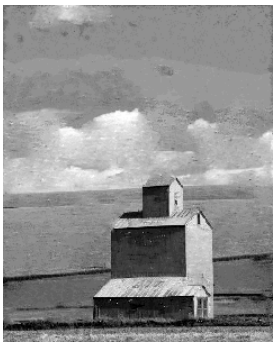


sketches

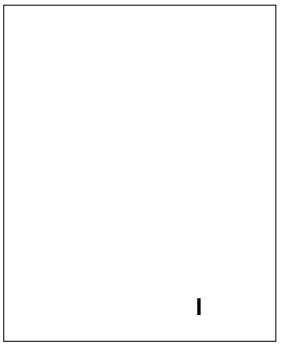
Primal sketch example



original image



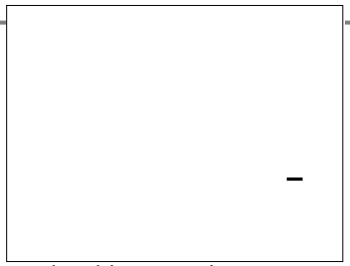
synthesized image



sketching pursuit process

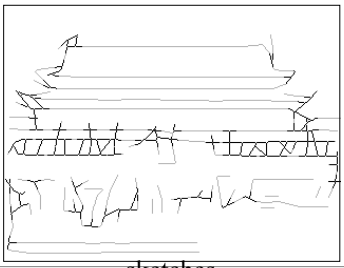
Primal sketch example

original image



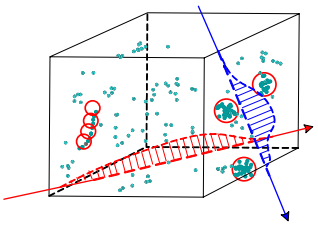
sketching pursuit process

synthesized image



sketches

Learning the primal sketch from natural images

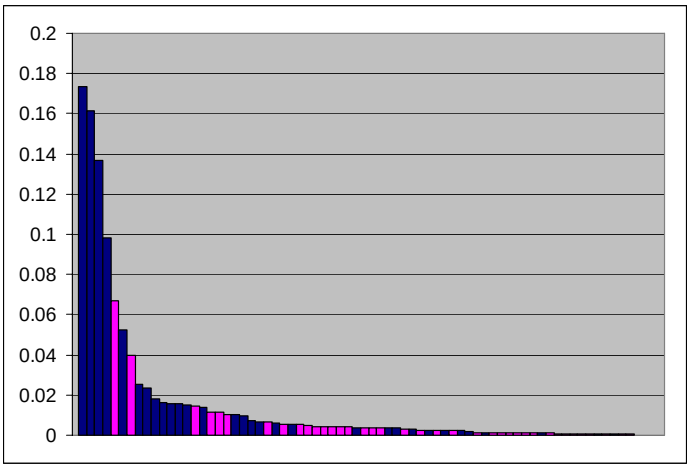


Frequency Table

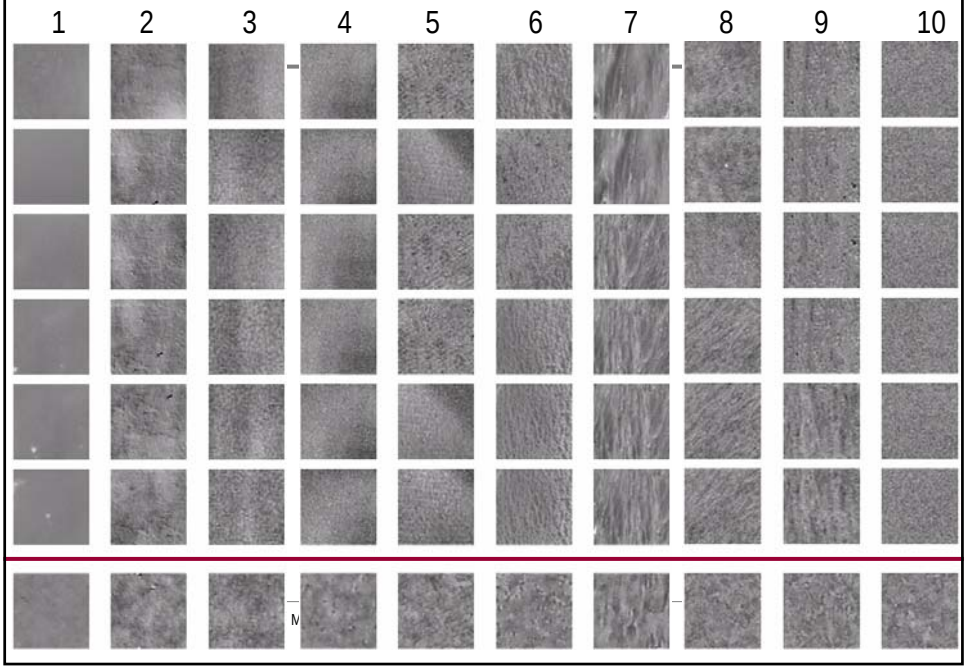
Texture	0.158944	Texture	0.020594	Y junction	0.005433
Texture	0.081222	Texture	0.019605	L junction	0.005254
Texture	0.065283	Texture	0.019399	L junction	0.00493
Texture	0.051938	Ridge	0.017631	L junction	0.004742
Edge	0.049656	Texture	0.017175	Y junction	0.004081
Texture	0.045718	Texture	0.017052	Ridge	0.003176
Texture	0.037522	Texture	0.016805	Y junction	0.002867
Edge	0.034838	Texture	0.016805	Ridge	0.001536
L junction	0.030281	Texture	0.016352	Terminator	0.001212
Texture	0.030026	Ridge	0.015033	Terminator	0.001212
L junction	0.029865	Texture	0.014786	Cross	0.000732
Texture	0.028914	Texture	0.011656	Terminator	0.000567
Texture	0.026319	Texture	0.011121	Terminator	0.000446
Texture	0.025948	L junction	0.010339	Cross	0.000439
Texture	0.024548	Texture	0.010215	Terminator	0.000352
Ridge	0.022903	Texture	0.009473	Cross	0.000329
Texture	0.02183	L junction	0.006512	Ridge	0.000254

Frequency plot of the ex/implicit manifolds in natural images

Blue is implicit texture clusters, pink is explicit primitives (textons)

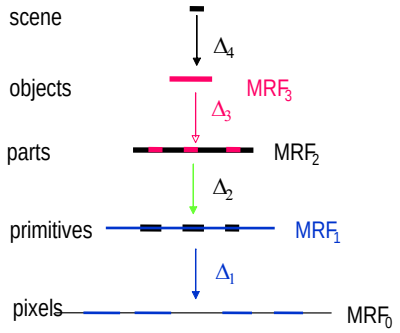


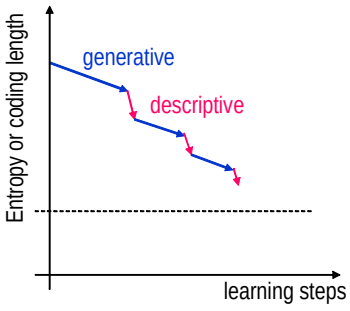
The top 10 texture clusters



Augmentation of the Integrated generative model

The pursuit of an integrated model is to minimize the Shannon and Kolmogorov entropy in turns,





Leaves at a range of scales

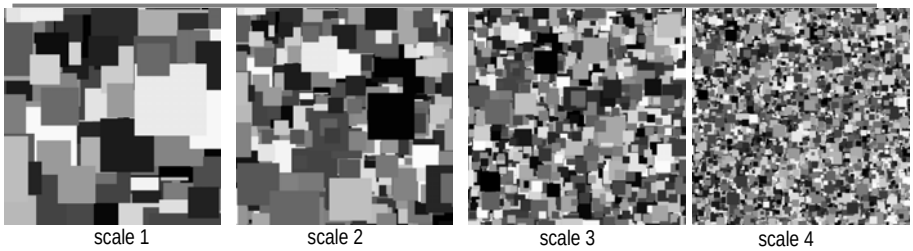
This picture contains trees/leaves at four ranges of distance, over which our perception changes.

A: see individual leaves with sharp edge/boundary (occlusion model)
B: see leaves but blurry edge (additive model)
C: see a texture impression (MRF)
D: see constant area (iid Gaussian)

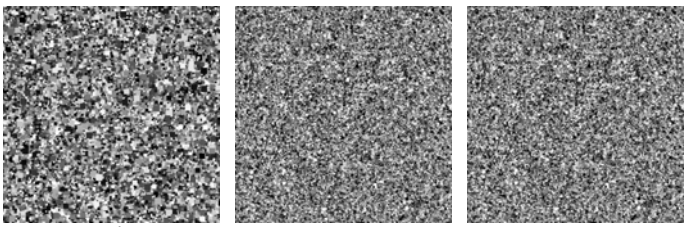
Intimate mixing of texture and textons, and perceptual transitions between them

We proposed a perceptual scale space theory (Wang and Zhu 2005)

Model regime transitions in scale space



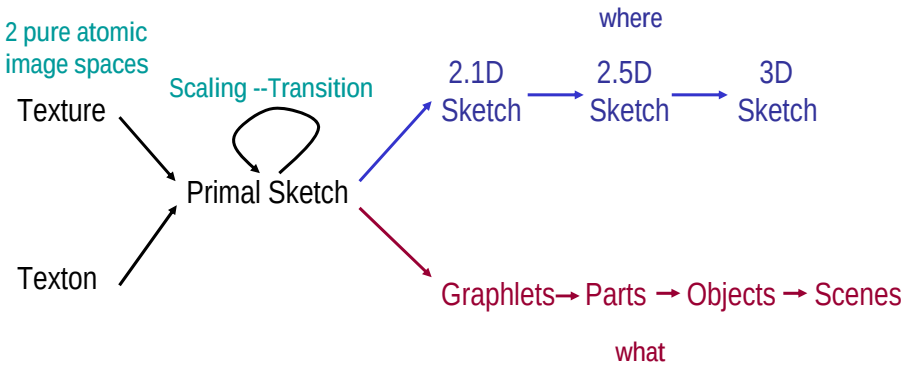
scale 1 scale 2 scale 3 scale 4



scale 5 scale 6 scale 7

We need a seamless transition between different regimes of models

Summary



What do these mean to biologic vision?