From Scaling Laws of Natural Images To Regimes of Statistical Models

Song-Chun Zhu

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

Joint work with

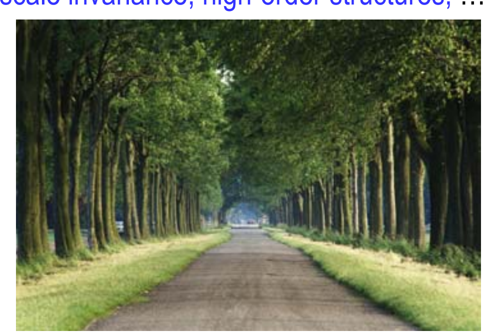
C.E. Guo

Y.N. Wu

Statistics of "Natural Images"

In recent years, there has been a growing interest in studying the statistics of natural images:
1/f-power law, high kurtosis, scale invariance, high-order structures, …

….
Ruderman and Bialek 87, 94
Fields 87, 94
Zhu and Mumford 95-96
Chi and Geman 97-98
Lee, Mumford and Gidas 00-02
Simoncelli etc 98-03
…..



Natural Images often refer to the scenes that contain objects in continuous scales.

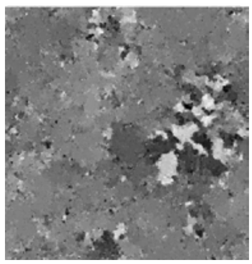
Research Stream 1. Seeking Scale Invariant Image Models

Rational:

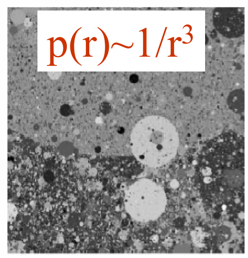
Let p be a probability measure on generic images I(x,y), then p should be scale-invariant, because images are observed at arbitrary scales, i.e.

$$p(I(x,y)) = p(I(\sigma x, \sigma y)) \qquad \sigma > 0$$

Such probability does not exist, instead people look for models that are approximately scale invariant, including Markov random fields and generative models.



(Zhu and Mumford 1996)



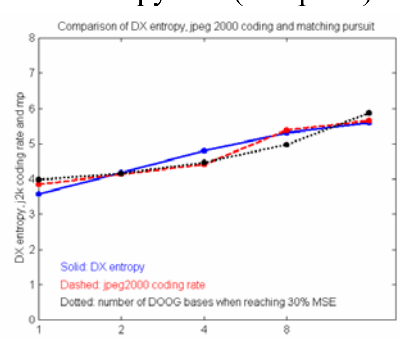
(Lee and Mumford 1997)

(Chi and Geman 1998)

But Statistics Change over Scales

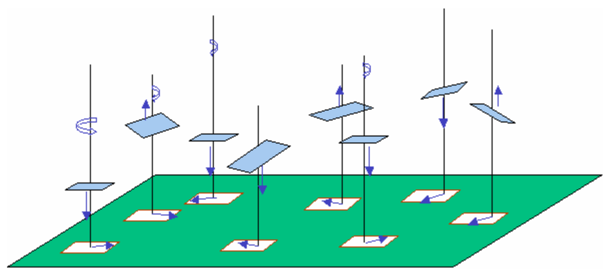
When we zoom out from a natural scene, the image entropy rate increases (assuming the images are renormalized).

Entropy rate (bits/pixel)

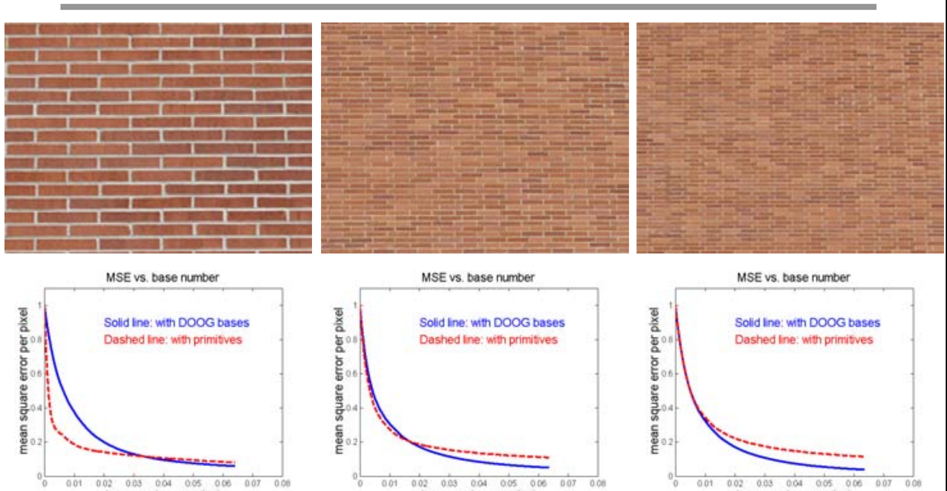


Research Stream 2. Seeking Fundamental Image Elements

Sparse coding, Olshausen and Felds, 95
X-lets, Donoho School 98-04
Transformed Component Analysis, Frey and Jojic 00
Textons, Leung and Malik 99, Guo, Zhu and Wu, 01,02, (Dated back to Julesz)
Image primitives, Guo, Zhu and Wu, ICCV, 03 (Dated back to Marr)
… …



But Image Elements (Dictionary) Change Over Scales



Two Dictionaries Used

DOOG bases

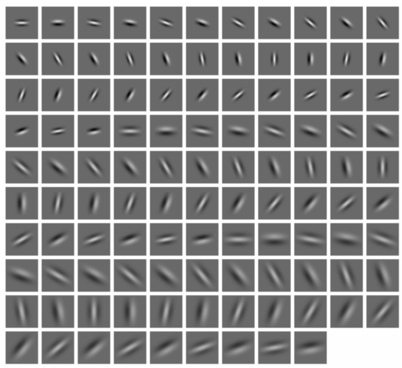
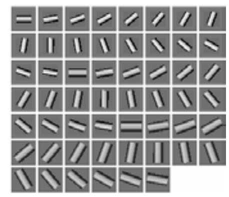


Image Primitives



Our Perception Changes Over Scales!

Our perception of visual patterns jumps (both mild and catastrophic) over scales, so should the statistical models!

Such transition is not accounted for in the scale-space theory or pyramids.

This picture contains 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)

A Theoretical Problem:
Gap Between Two Theoretical Foundations

How do we represent/mix two different patterns consistently?
and what trigger the perceptual transition (jump)?

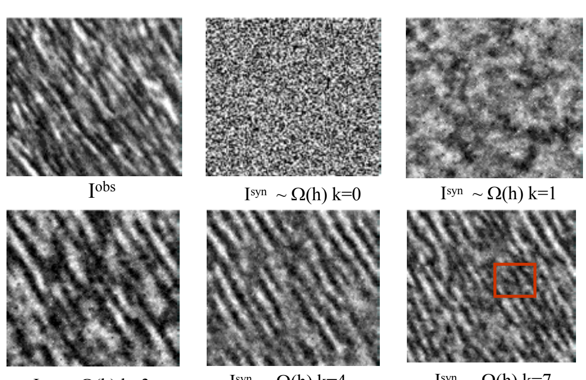
a). Markov random fields,
from statistical physics.

b). Image coding, wavelet et al,
from harmonic analysis.

Theory 1: MRF for Modeling Texture Patterns

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

H_c are histograms of Gabor filters, i.e. marginal distributions (Zhu, Wu, Mumford, 1996-01)

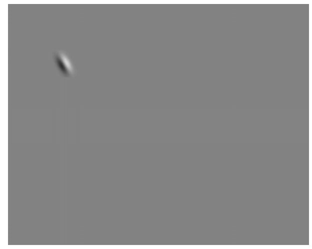


For images I drawn in the ensemble, any local patch follows a Markov random field model (FRAME),

Theory 2: Image Coding

$$\mathrm{I} = \sum_{j \in D} \alpha_j b_j + \mathrm{n}, \quad \text{D is a dictionary}$$

input image



matching pursuit (Mallat-Zhang 93)

500 bases

800 bases

Comparison of the Two Theories

1. Markov Random Fields --- Descriptive model $h(I)=h_o$
   - Statistical Physics
   - Effective on texture but not structures
   - Implies population coding (pooling)

2. Image Coding --- Generative model $I=g(W; D)$
   - Harmonic Analysis
   - Effective on structures but not texture
   - Implies winner-take-all (lateral inhibition)

They work on different entropy regimes!

Definition: Image Complexity

Let $I \sim p(I)$ defined on lattice Λ

the *image complexity* – $H(I)$:

is defined as the entropy of $p(I)$

$$H(I) = -\sum_{I} p(I)\log p(I)$$

Image Complexity Rate (per pixel) is:

$$\overline{H}(I) = \frac{H(I)}{|\Lambda|}$$

Down-scaling = local smoothing + down-sampling

Local Smoothing Theorem

Suppose we smooth image I to J by kernel K. $I \xrightarrow{K} J$

Theorem 1: The smoothing operator decreases the entropy rate

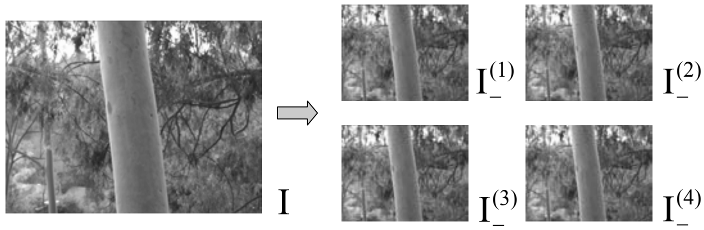
$$\overline{H}(J) - \overline{H}(I) \xrightarrow{\Lambda \to Z^2} \int \log |\hat{k}(w)| \, dw < 0$$

$\hat{k}(w)$ Fourier transform of kernel k

$$\int \log |\hat{k}(w)| \, dw \le 0$$

Image complexity rate is decreasing with local smoothing (by a constant related to the kernel).

Image Down-Sampling



$$I = (I_{-}^{(1)}, I_{-}^{(2)}, I_{-}^{(3)}, I_{-}^{(4)})$$

$$\frac{1}{4}\sum_{k=1}^{4} \overline{H}(I_{-}^{(k)}) - \overline{H}(I) = M(I_{-}^{(1)}, I_{-}^{(2)}, I_{-}^{(3)}, I_{-}^{(4)}) \geq 0$$

M(…): mutual information

Down-Sampling Theorem

Theorem 2:

Image complexity decreases with down-sampling.

$$H(I_{-}^{(k)}) \leq H(I), \quad k = 1,2,3,4$$

(a down-sampled image has less information than the original image)

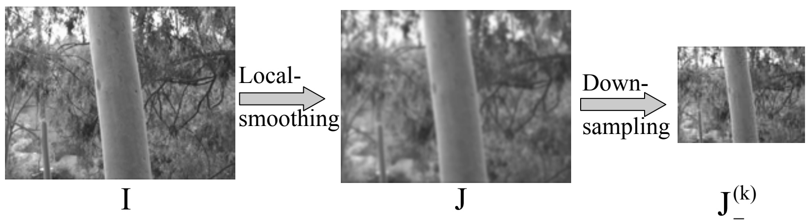
Image complexity rate increases with down-sampling.

$$\frac{1}{4}\sum_{k=1}^{4} \overline{H}(I_{-}^{(k)}) \geq \overline{H}(I)$$

(there is less mutual information between pixels in a down-sampled image, and thus it looks more random.)

Complexity Scaling Law

Image complexity rate changes by M-K with down-scaling.



$J_{-}^{(k)}$

$$\frac{1}{4}\sum_{k=1}^{4}\overline{H}(J_{-}^{(k)}) - \overline{H}(I) \xrightarrow{\Lambda \to Z^2} M - \hat{K}$$

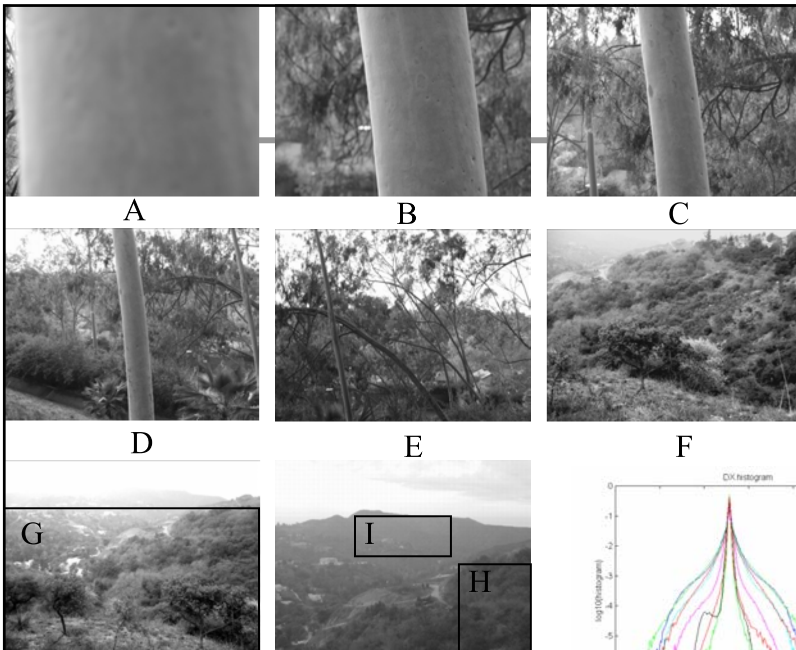
Scale Invariant if

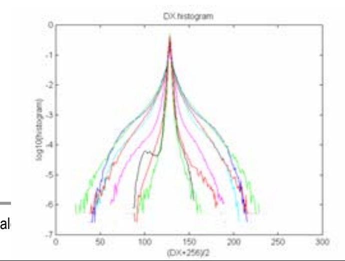
$$M \approx \hat{K}$$

$$M \equiv M(I_{-}^{(1)}, I_{-}^{(2)}, I_{-}^{(3)}, I_{-}^{(4)})$$

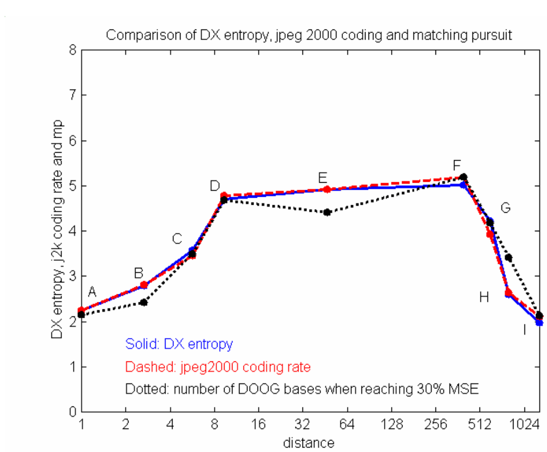
$$K \equiv -\int \log |\hat{k}(w)| \, dw$$

Example: zooming out with natural scene



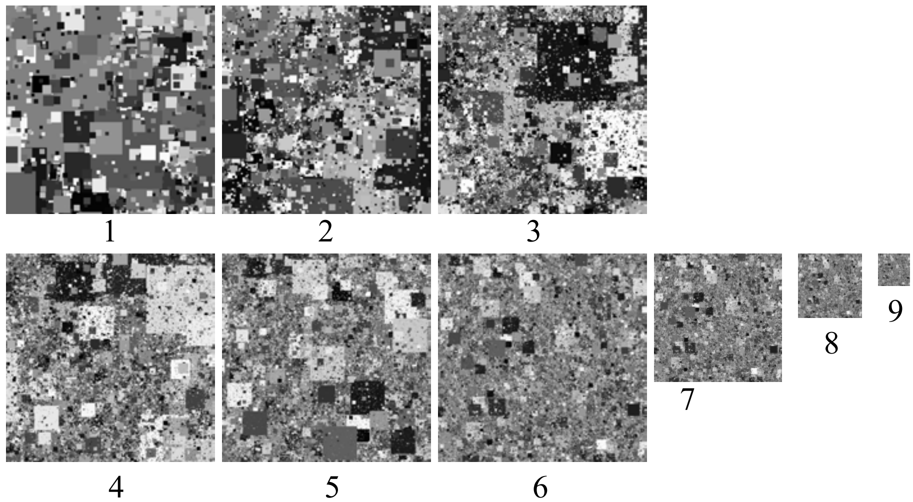


Entropy rate (bits/pixel) over distance

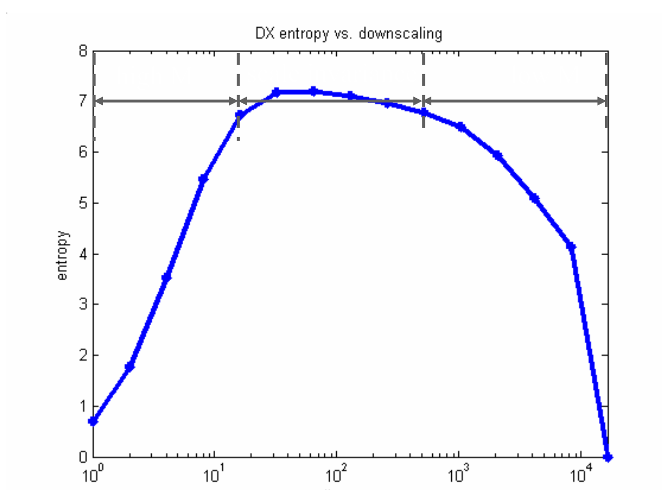


1. entropy of I_x
2. JPEG2000
3. #of DooG bases for reaching 30% MSE

Synthetic Example: Down Scaling



Synthetic Complexity Rate



Definition: Perceptibility Scaling

Let W be the description of the scene (world), $W \sim p(W)$

Assume: generative model $I = g(W)$

1. *Scene Complexity* is defined as the entropy of p(W)

$$H(W) = -\sum_{W} p(W) \log p(W)$$

2. *Imperceptibility* is defined as the entropy of posterior p(W|I)

$$H(W \mid I) = -\sum_{W} p(W) \log p(W \mid I) = H(W) - H(I)$$

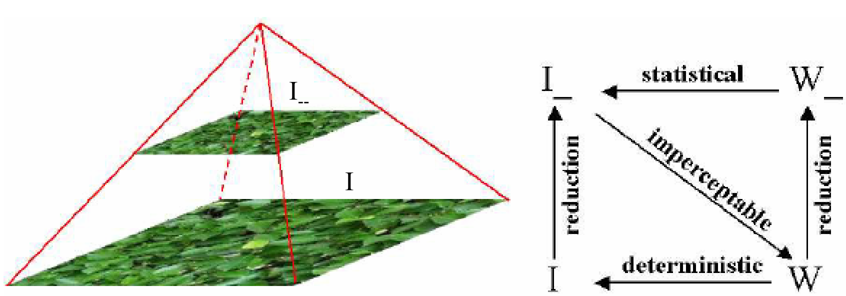
Imperceptibility = Scene Complexity – Image complexity

Perceptibility Scaling Law

Theorem 3: Imperceptibility increases with down-scaling.

If $W \sim p(W)$, $I = g(W)$, $I\_ = R(I)$ by down-scaling

Then $H(W \mid I\_) \geq H(W \mid I)$



A Simplified Physical Model

Let a scene W consists of N iid distributed planar objects, N=10^{23}.

This is called a *micro-canonical ensemble* in physics and a *marked point process* in statistics

$$\Pi = \{(x_i, y_i, r_i, I_i), \quad i = 1,2,\ldots,N. \},$$
$$(x, y, I) \sim \text{unif}, \; A_i = |r_i| \sim f(A), \quad A \in [A_{\min}, A_{\max}]$$

Suppose we view the scene at scale σ,

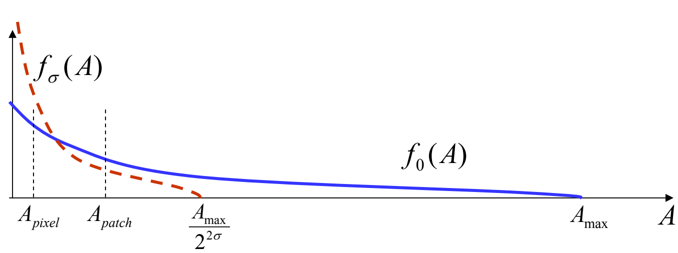
a pixel r covers a domain r(x,y) with a resolution σ^2.

a patch R covers a domain with 5x8 pixels,

The effects of scaling

After scaling and down-sampling, we obtain a new process.

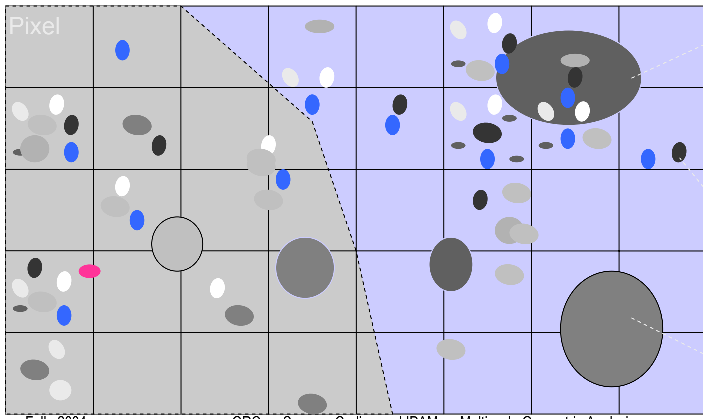
$$\Pi_\sigma = \{(x_i, y_i, r_i, I_i), \quad i = 1,2,\ldots,N\}, \qquad (x, y, I) \sim \text{unif},$$
$$A_i = |r_i| \sim f_\sigma(A) = 2^{2\sigma} f_0(2^{2\sigma} A)$$



A 5x8 Patch Model

$$I = \sum_{j \in D} \alpha_j b_j, \quad D = \{k: \ R \cap r_k \neq \phi\},$$

$$\alpha_j = I_j, \ b_k(x,y) = |r(x,y) \cap r_k| \ is \ a \, base$$



b_k
$$\begin{pmatrix} 0.00, & 0.00, & 0.00 \\ 0.19, & 0.65, & 0.08 \\ 0.10, & 0.27, & 0.04 \\ 0.00, & 0.00, & 0.00 \end{pmatrix}$$

b_k
$$\begin{pmatrix} 0.0, & 0.0, & 0.0 \\ 0.0, & 0.05, & 0.0 \\ 0.0, & 0.0, & 0.0 \end{pmatrix}$$

b_k
$$\begin{pmatrix} 0.0, & 0.0, & 0.0, & 0.0 \\ 0.0, & 0.41, & 0.28, & 0.0 \\ 0.0, & 0.20, & 0.17, & 0.0 \\ 0.0, & 0.0, & 0.0, & 0.0 \end{pmatrix}$$

A Patch Model

Now we can divide the set D (objects overlapping patch R) into three subsets

$$I = \sum_{i \in D_1} \alpha_i b_i + \sum_{j \in D_2} \alpha_j b_j + \sum_{k \in D_3} \alpha_k b_k \qquad D = D_1 \cup D_2 \cup D_3$$

1. D_1 includes objects whose size (at least in 1 dimension) is larger than the patch.

$$D_1 = \{i: \ \dim(r_i) > \dim(R), \ r_i \in \Pi\}, \quad n_1 = |D_1|$$

If $n_1<2$, then it is non-sketchable. If $n_1>=2$, then sketchable n_1 is the degree of the primitive

2. D_2 includes objects whose size is smaller than the patch but bigger than pixel

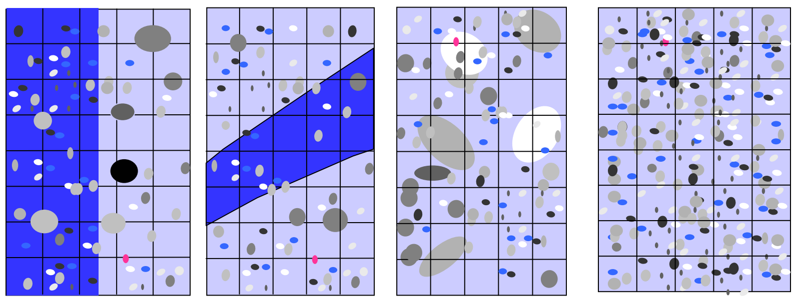
$$D_2 = \{j: \ \dim(r) < \dim(r_j) < \dim(R), \ r_i \in \Pi\}, \quad n_2 = |D_2|$$

Such objects may not cause noticeable structures, but generate pixel correlations and textures.

3. D_3 includes objects whose size is smaller pixel, these objects produce the iid noises.

$$D_3 = \{k: \ \dim(r_k) < \dim(r), \ r_i \in \Pi\}, \quad n_3 = |D_3|$$

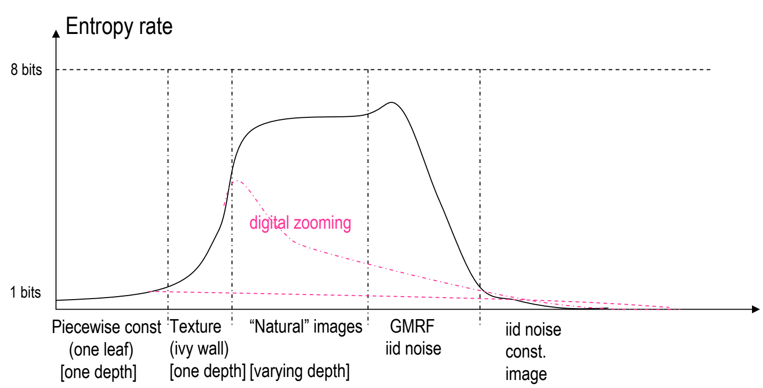
Image patches at various entropy regimes



(a) a step edge (b) a bar (c) texture (d) noise

Summary: Regimes of Image Models

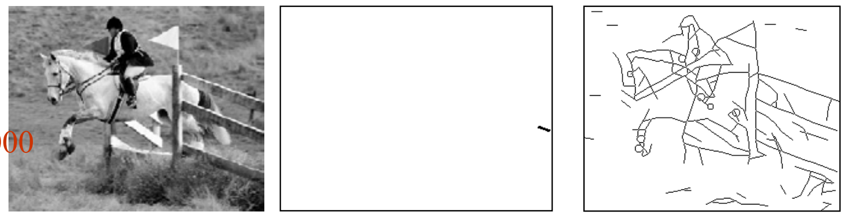
Zooming out, we see more objects in a window until the biggest object size R_{max} is smaller than the window size.



Primal Sketch : mixing three regimes of models

(Guo, Zhu and Wu iccv03)

50KB
JPG2000



org image | sketching pursuit process | sketches

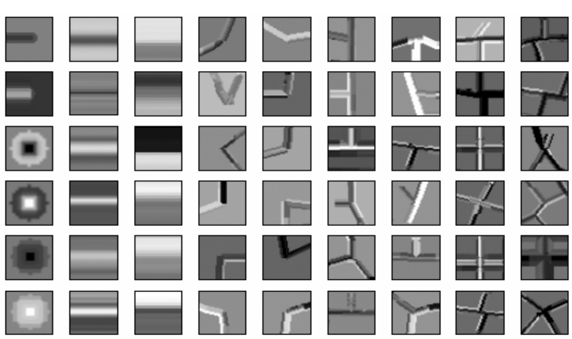
2KB

syn image = synthesized textures + sketch image

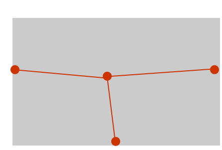
Learned Image Primitive (non-linear)

Part of the image primitive dictionary

(Guo, Zhu and Wu, ICCV 03)

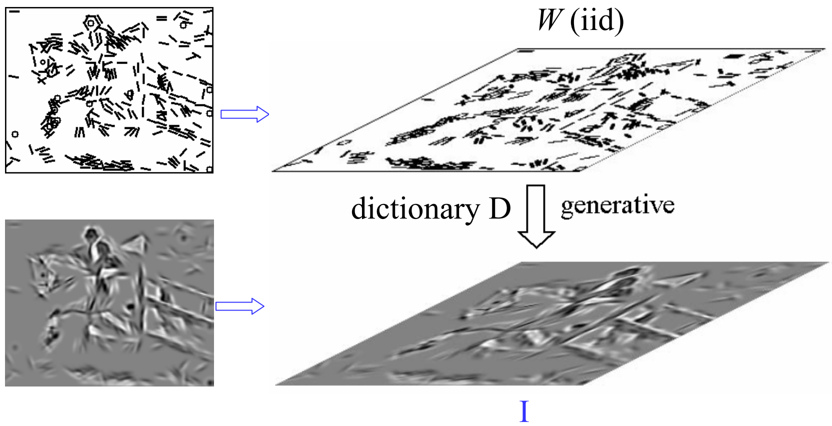


Each primitive has degree=d control points warping the patch.

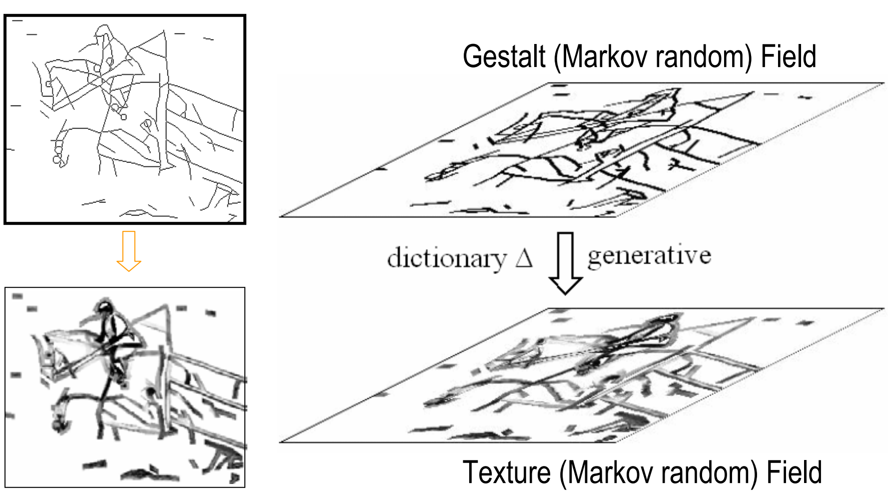


Comparison with Image Coding

Symbolic sketch of Gabor/LoG bases



Primal Sketch: two-level MRF



The Primal Sketch Model

$$\Lambda = \Lambda_{nsk} \cup \Lambda_{sk}, \quad \Lambda_{sk} = \bigcup_{i=1}^{K} \Lambda_{sk}^{i}(\theta_{geo})$$

$$I_{\Lambda_{sk}}(x, y) = B_{\ell}(x - u, y - v; \ \ \theta_{geo}, \theta_{pho}), \quad (x, y) \in \Lambda_{sk}^{i}(\theta_{geo}).$$

ℓ indexes the type.

$$I_{\Lambda_{nsk}} \sim p(I_{\Lambda_{nsk}} \mid I_{\Lambda_{sk}}; \beta)$$

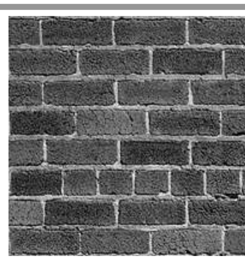
is a FRAME model (MRF)

The sketch is a mixed Markov field with dynamic neighborhood

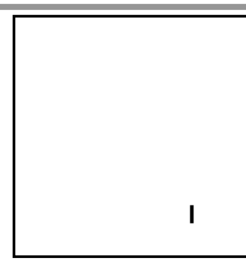
$S = (V, E) \ \sim p(V, E)$

Primal Sketch Experiments

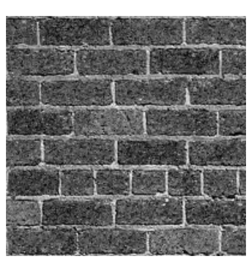
input



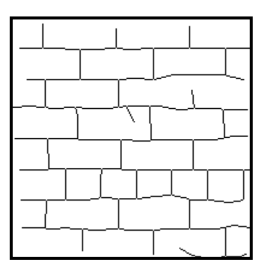
sketching pursuit



synthesized



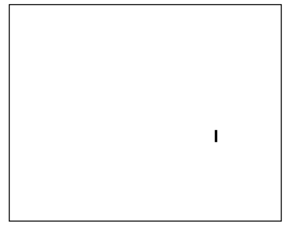
primal sketches



Primal Sketch Experiments

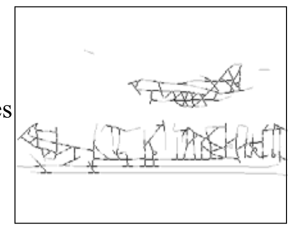
input

sketching pursuit

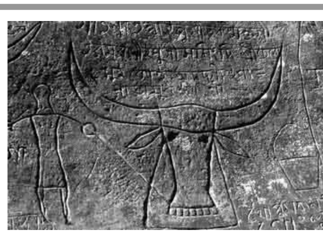


synthesized

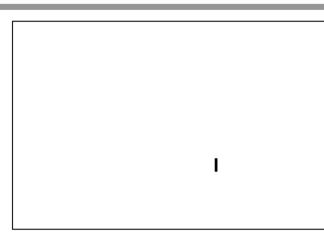
primal sketches



Primal Sketch Experiments

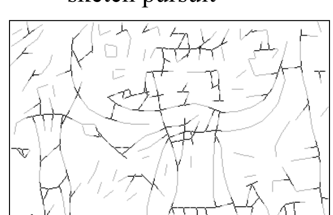


input image



sketch pursuit

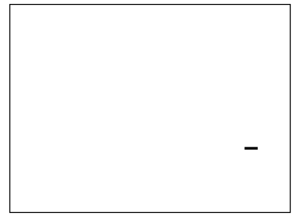
synthesized



primal sketches

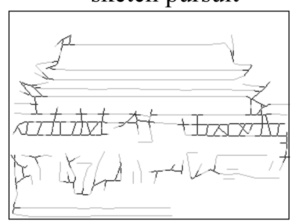
Primal Sketch Experiments

original image



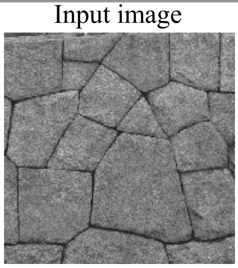
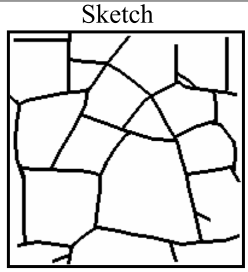
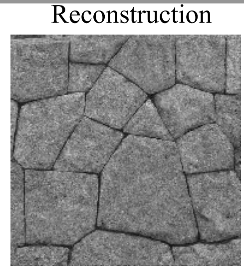
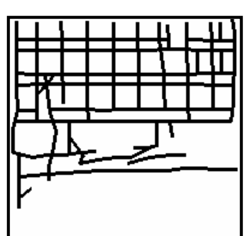
sketch pursuit

synthesized image



sketches

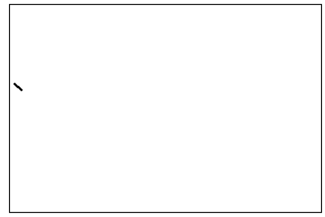
Primal Sketch Experiments

Input image	Sketch	Reconstruction
		
		

sketch Old syn

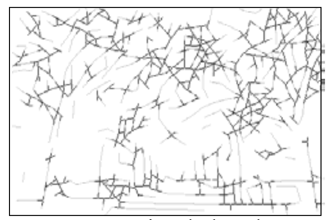
Primal Sketch Experiments

original image



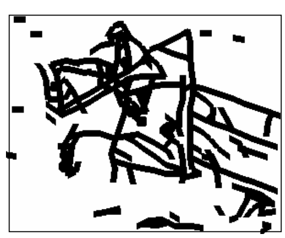
sketch pursuit

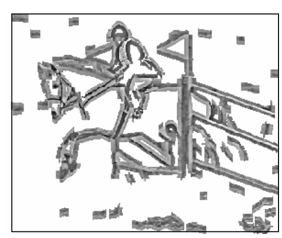
synthesized image

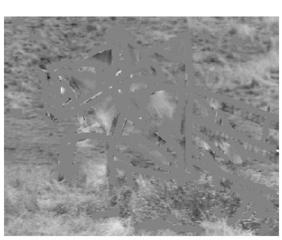


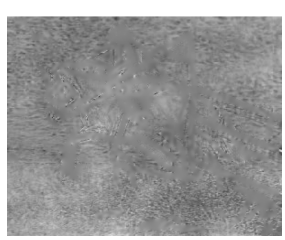
Primal sketch

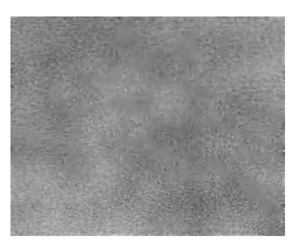
Removing the Components in the horse riding image





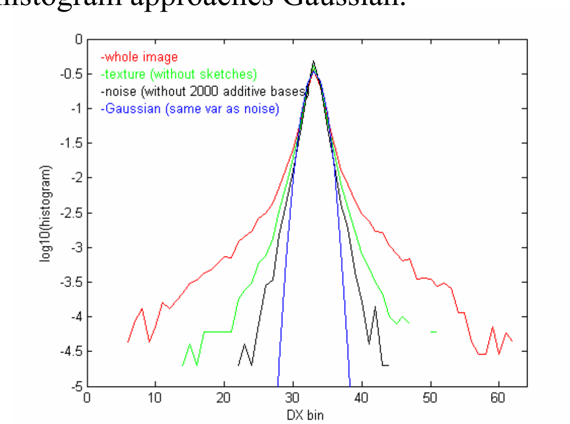




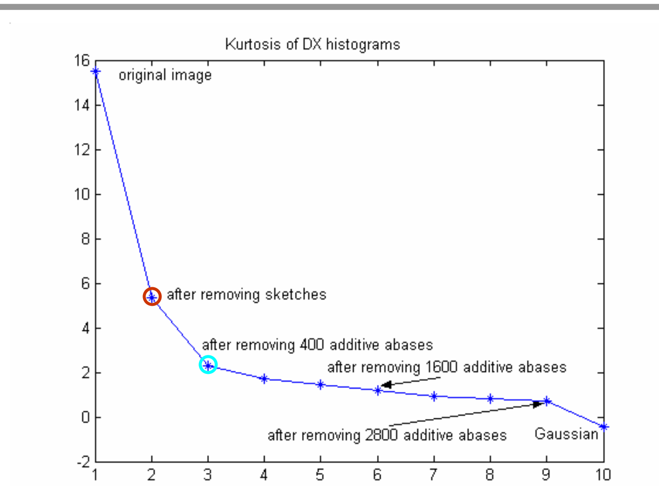


The Histogram Changes in Horse riding

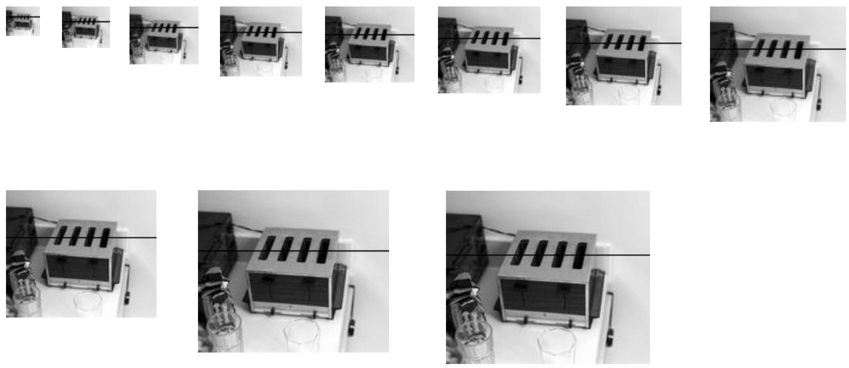
After removing the structured (low entropy rate) patches, the I_x histogram approaches Gaussian.



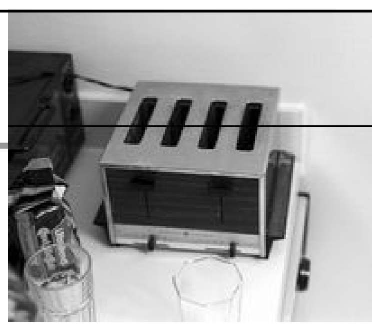
Kurtosis Changes in Horse riding Image

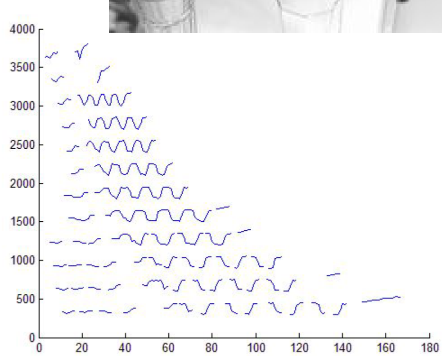


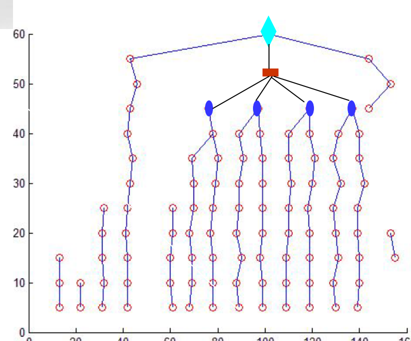
What Occurs in Image Scaling?



Experiment by Y.Z. Wang

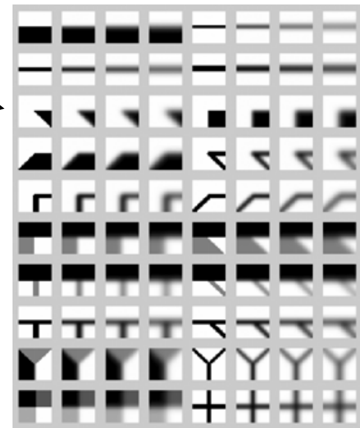






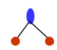
What occurs in perception when up-scaling?

1. Image sharpening on boundaries



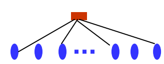
2. Mild jumps

   e.g. birth of a sketch, or split a bar to 2 edges
   ---- handled by graph grammar.



3. Catastrophic transition

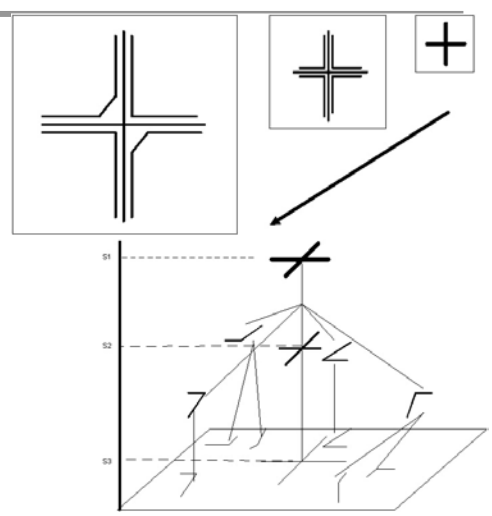
   e.g. from texture to 100s primitives



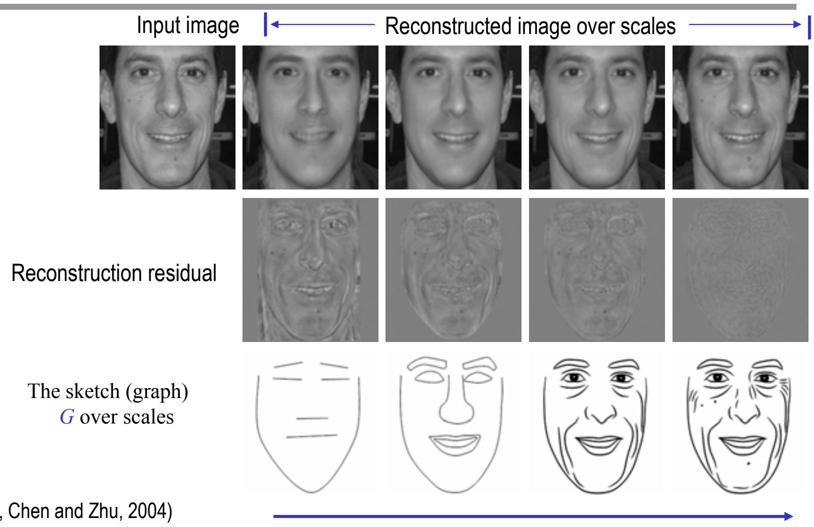
Topologic changes over multi-scales

The current scale-space theory is based on continuous Gaussian --Laplacian pyramids. While it is suitable for the retina and LGN, it is wrong for V1.

We need a new scale-space theory which is multi-layer of primal sketches



Example of hierarchic graph of face



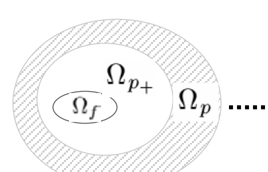
(Xu, Chen and Zhu, 2004)

Theory 1, Minimizing Shannon Entropy

$$H_p = \int p(I;\beta)\log\frac{1}{p(I;\beta)}dI = \log|\Omega_p|$$

The models are augmented by pursuing *best features h_+*
so as to minimize the entropy or volume,

$$h_+ = \arg\max H_p - H_{p_+} = \log\frac{|\Omega_p|}{|\Omega_{p_+}|}$$



Until the information gain of the best feature is statistically insignificant.