Pursuing Explicit and Implicit Manifolds by Information Projection

Song-Chun Zhu

University of California, Los Angeles, USA

Lotus Hill Research Institute, China（中国莲花山研究院）

Joint work with Yingnian Wu, Kent Shi, ….

1, Background on visual (appearance) manifolds

Image patches from a single object category are often found to form low dimensional manifolds.

e.g. ISOMAP, LLE:
Saul and Roweis, 2000.

But, people found that image patches of generic natural images do not follow this observation.

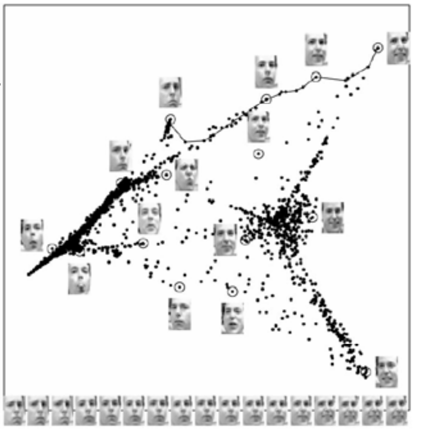
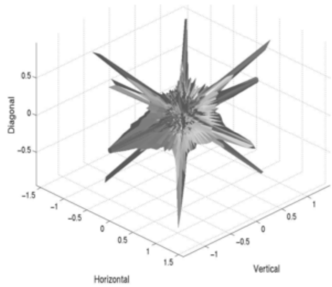


Fig. 3. Images of faces (11) mapped into the embedding space described by the first two coordinates of LLE. Representative faces are shown next to circled points in different parts of the space. The bottom images correspond to points along the top-right path (linked by solid line), illustrating one particular mode of variability in pose and expression.

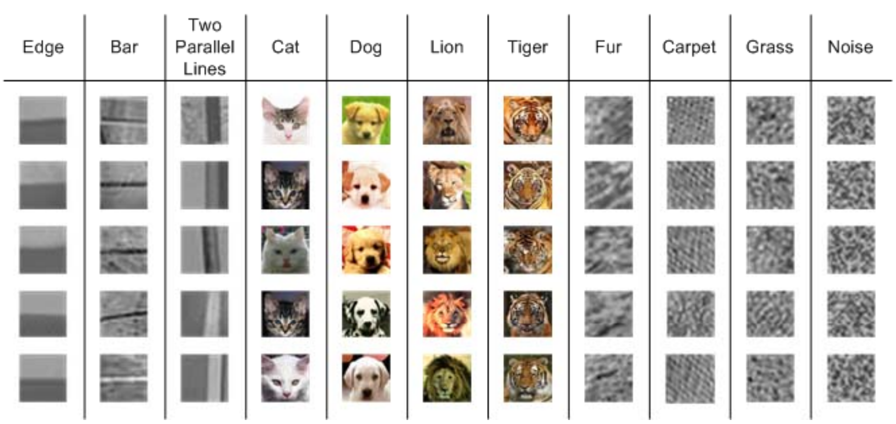
Looking at local, generic natural image statistics

….
Ruderman and Bialek 87, 94
Fields 87, 94
Zhu and Mumford 95-96
Chi and Geman 97-98
Huang, 2000
Simoncelli etc 98-03
…..



Here is an example of how real world data can be truly complex – non-Gaussian and highly kurtotic. This is an iso-density contour for a 3D histogram of log(range) images (2x2 patches minus their means) (Brown range image database, thesis of James Huang)

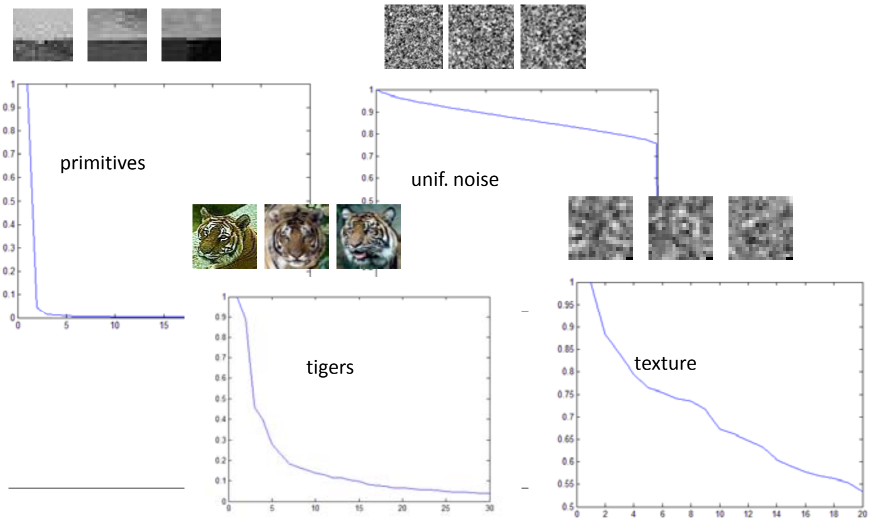
A wide spectrum of categories from low to high entropy



Entropy ~ Dimension ~ Log volume(manifold)

Visual manifolds have varying dimensions

Take 16x16 image patches (256-space), run PCA for each category, and plot the eigen-values in decreasing order.



By analogy: pictures of our universe

entropy (temperature) regimes.

compositional structures.

How do we learn these manifolds?
Can we do it by K-mean clustering?
3 modeling theories in vision:
(1) Markov random fields, (2) Sparse coding, (3) Grammar and Composition

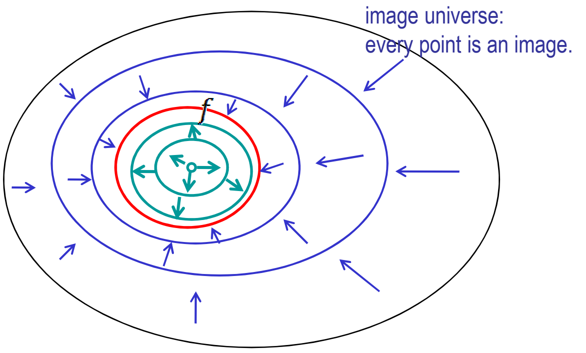
2, Manifold pursuit in the universe of image patches

f : target distribution; *p*: our model; *q*: initial model

$$q = p_0 \to p_1 \to \cdots \to p_k \quad to \quad f$$

1, *q = unif()*

2, *q =* $\delta()$



model ~ image set ~ manifold ~ cluster

Intuitive idea: a professor grading an exam

The full score (like dimension in our case) is 100. You have two ways:

For top students (high dimensional manifolds), you start from 100 and deduct points :

100 – 2 – 0 – 0 – 3 – 0– 2 – 0 – 0 – 0 – 0 – 0 – 1 = 92

For bottom students (low dimensional manifolds), you start from 0 and add points

0 + 8 + 0 + 0 + 3 + 0 + 2 + 0 + 0 + 5 + 0 + 0 + 1 = 19

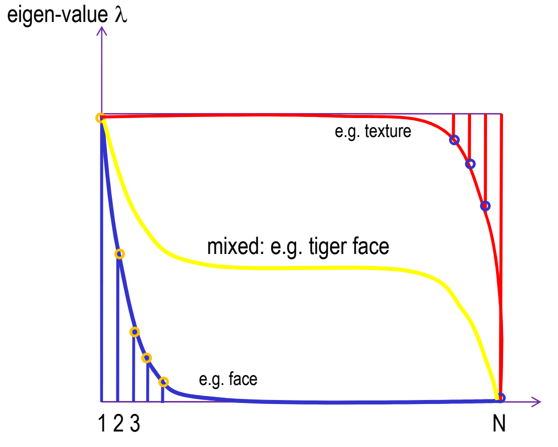
In reality, suppose the exam is very long (just like the large image has >1M pixels), a student may have mixed performance, e.g. doing excellent in the 1st half and doing poorly in the 2nd half. Thus a most effective way is to use the two methods for different sections of the exam.

(50 – 2 – 0 – 0 – 3 – 0) + (0 + 5 + 3 + 0 + 0 + 2) = 45 + 10 = 55

In fact, most of the object categories are middle entropy manifolds and have mixed structures.

Manifold pursuit in the image universe

In a simple case: *f* is a Gaussian distribution



Manifold pursuit by information projection

Given only positive examples from a class c

$$\Omega_c^+ = \{\mathrm{I}_i^{obs}; \quad i = 1,2,\ldots, M^+\} \sim f(\mathrm{I})$$

We pursue a series of models *p* to approach a underlying "true" probability *f*

$$q = p_0 \to p_1 \to \cdots \to p_k \quad to \quad f$$

At each step, we augment the current model *p* to a new model p_+

$$h_+^* = \arg\max \ KL(f \mid p) - KL(f \mid p_+)$$
$$= \arg\max KL \ (p_+ \mid p)$$

Subject to a projection constraint:

$$E_{p_+}[\mathrm{h}_+(\mathrm{I})] = E_f[\mathrm{h}_+(\mathrm{I})] \cong \bar{h}_+$$

$\mathrm{h}_+(\mathrm{I})$ is a feature statistics of image **I**

Manifold pursuit by information projection

Solving the constrained optimization problem leads to the Euler-Lagrange equation

$$p_+^* = \arg\min \int p_+(I) \log \frac{p_+(I)}{p(I)} \, dI + \lambda_+ \left[\int p_+(I) h_+(I) dI - \bar{h}_+\right] + \lambda_o \int p_+(I) dI - 1]$$

$$p_k(I; \Theta_k) = \frac{1}{Z_{+,k}} p_{k-1}(I; \Theta_{k-1}) e^{-\lambda_k h_k(I)}$$

$$= \frac{1}{Z_k} q(I) \exp\{-\sum_{i=1}^{k} \lambda_i h_i(I)\}$$

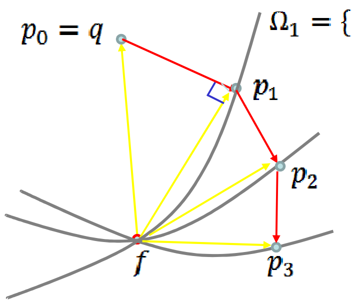
where

$$Z_k = Z_{+,1} \cdot Z_{+,2} \cdots Z_{+,k} \qquad \Theta_k = (\lambda_1, \lambda_2, \cdots \lambda_k)$$

For q being a uniform distribution, we have $q(I) = \frac{1}{Z_o}$

Information projection

DellaPietra, DellaPietra, Lafferty, 97
Zhu, Wu, Mumford, 97



$$\Omega_1 = \{ p: \; E_p[h_1(I)] = E_f[h_1(I)] \}$$

$$\Omega_2 = \{ p: \; E_p[h_2(I)] = E_f[h_2(I)] \}$$

$$KL(f \mid p) = KL(f \mid p_+) + KL(p_+ \mid p)$$

So the KL-divergence decreases monotonically.

A Maximin Learning Principle

max-step: choosing a distinct feature and statistics

$$h_+^* = \arg\max KL\ (p_+ \mid p)$$

min-step: given the selected feature constraint, computing the parameter

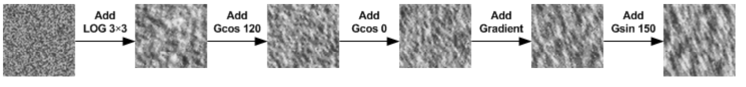
$$\lambda_+^* = \arg\min KL(p_+ \mid \ p)$$

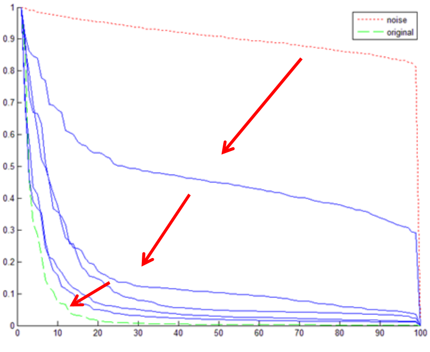
Claim: this learning procedure unifies almost all we know in visual modeling

- PCA, sparse coding,
- MRF, Gibbs, FRAME,
- Adaboost (when h() is binary),
- Stochastic grammar

3, Case studies:

Case 1: Pursuing texture models by compression from white noise

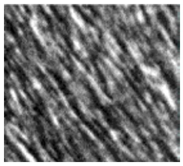




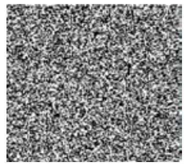
A texture pattern is an "implicit manifold"

$$\text{a texture} = \Omega(h_c) = \{\, I : \; h_i(I) = h_{c,i} \;\; , i = 1,2,\ldots, K \,\}$$

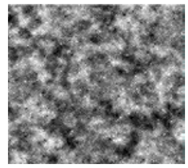
H_c are histograms of Gabor filters, i.e. marginal distributions of $f(I)$



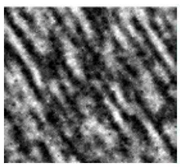
I^{obs}



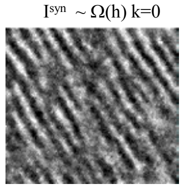
$I^{syn} \sim \Omega(h)$ k=0



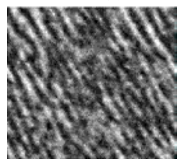
$I^{syn} \sim \Omega(h)$ k=1



$I^{syn} \sim \Omega(h)$ k=3



$I^{syn} \sim \Omega(h)$ k=4

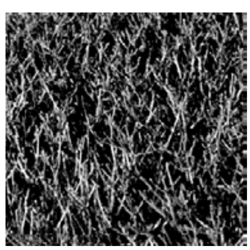


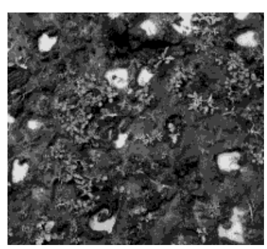
$I^{syn} \sim \Omega(h)$ k=7

(Zhu,Wu, Mumford 97,99,00)

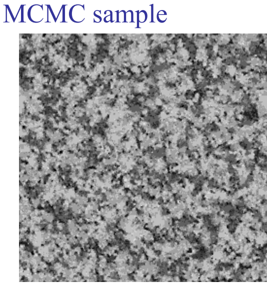
More examples of the texture manifold (implicit)

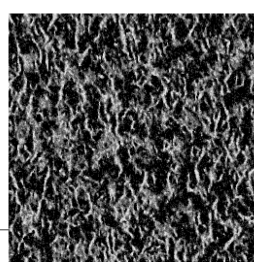
Observed

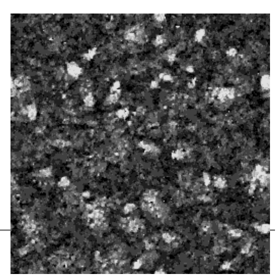




MCMC sample





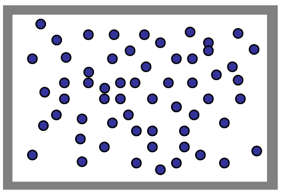


This is originally from statistical physics !

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

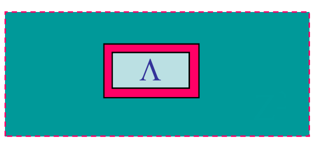
A state of the system is specified by the position of the N elements X^N and their momenta p^N

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

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

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

Equivalence of Julesz ensemble and FRAME/MRF models



Zhu, Wu, Mumford, 1997
Wu and Zhu, 1999

Theorem 1

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

Theorem 2

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

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

Case 2: A car pattern is an "explicit manifold"

Learning active basis as deformable template

A basis is an image space spanned by a number of vectors (e.g. Gabor/primitives)

$$B = (B_1, B_2, \ldots, B_k)$$

$$\text{A car} = \Omega = \{I: \; I = \sum_i \gamma_i B_{i,\delta} \}$$

A car template

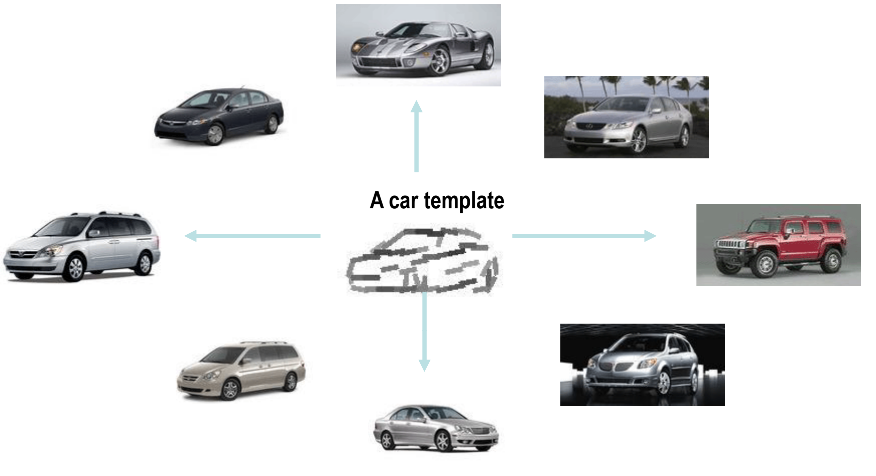
(Gabor elements represented by bar)

An incoming car image:

With slight modification, this model can handle multi-views

Wu, Si. Gong, Zhu, 2008

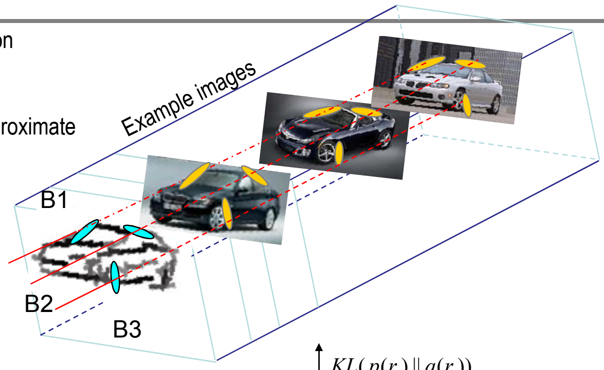
Deformed to fit many car instances

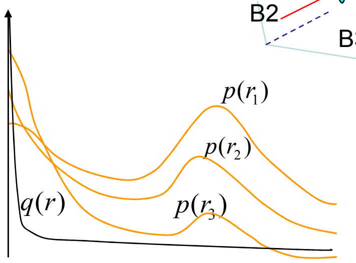


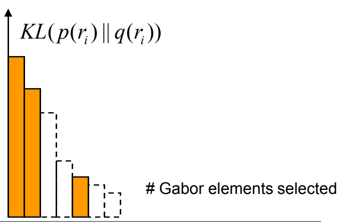
Pursuing the active basis model (explicit manifold)

q(I): background distribution (all natural images)

p(I): pursued model to approximate the true distribution.





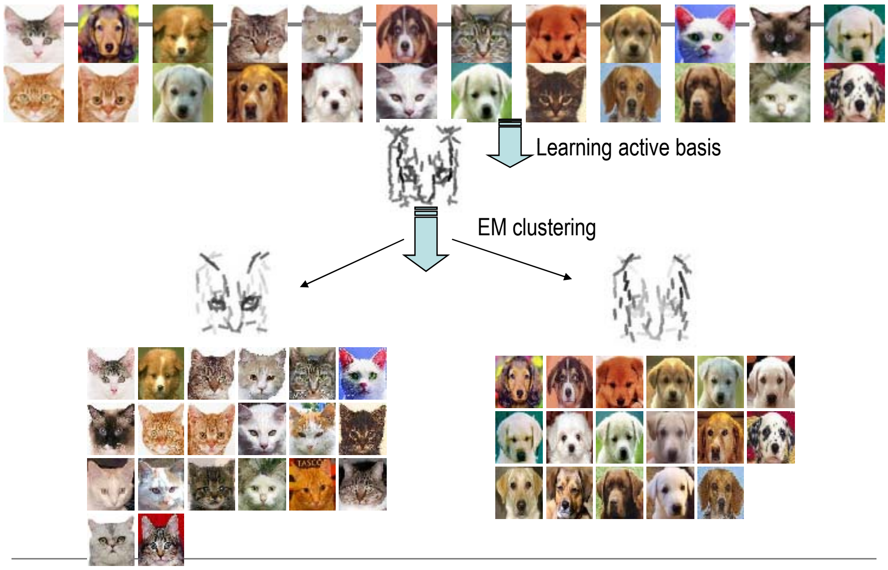


A running example

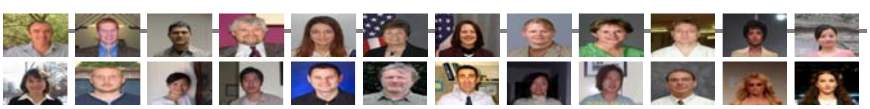
A car template consisting of 48 Gabor elements

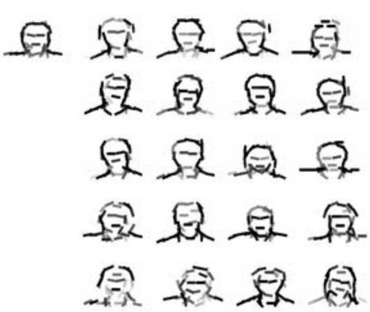
Car instances

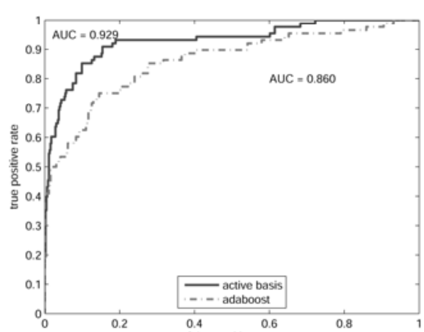
Experiment : learning and clustering



Experiment : learning and detection







Wu,Si,Fleming,Zhu,07

vs: Viola, Jones, 04

Template detection experiment

Wu,Si,Fleming,Zhu,07

Summary: two pure manifolds

implicit vs. *explicit*

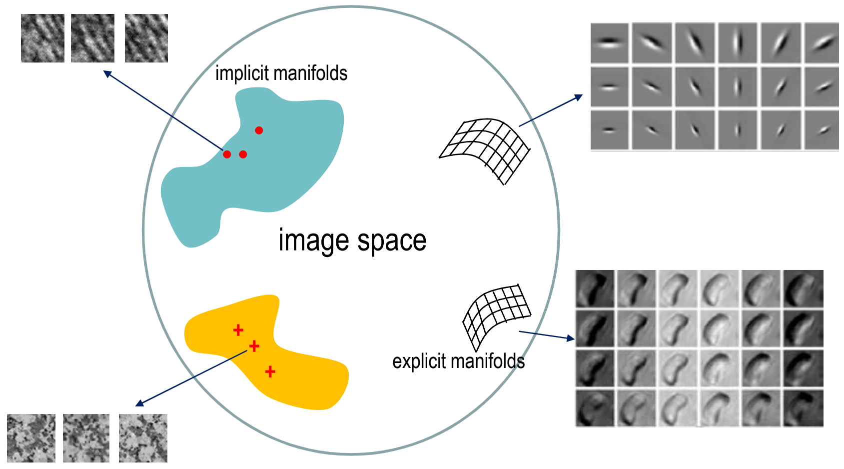
$$\Omega = \{I: \; h(I) = h_o \}$$

h(I) is some image feature/statistics

$$\Omega = \{ I: \; I = g(w; \Delta) \}$$

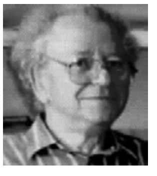
g is a generation function,
w is intrinsic dimension
Δ is a dictionary

Summary: a second look at the space of image patches

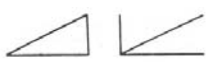


4, Relations to the literature: psychophysics

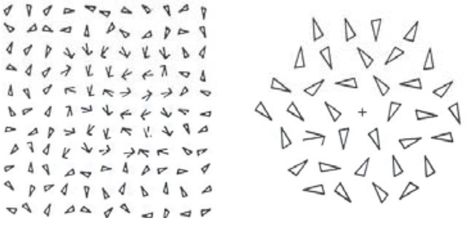
(1) textures vs textons (Julesz, 60-70s)

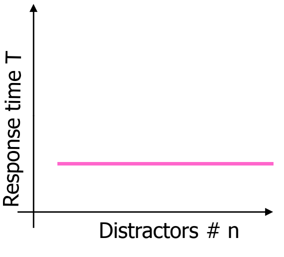


textons



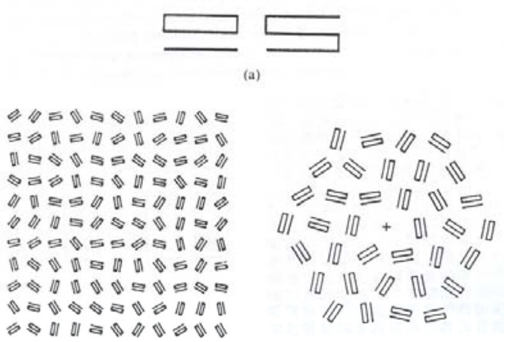
(a)

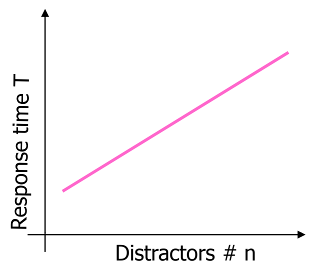




Textons vs. Textures

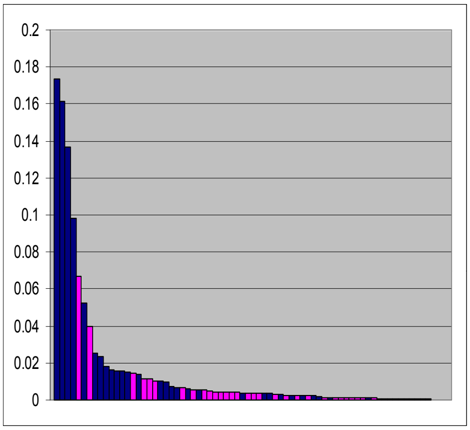
textures

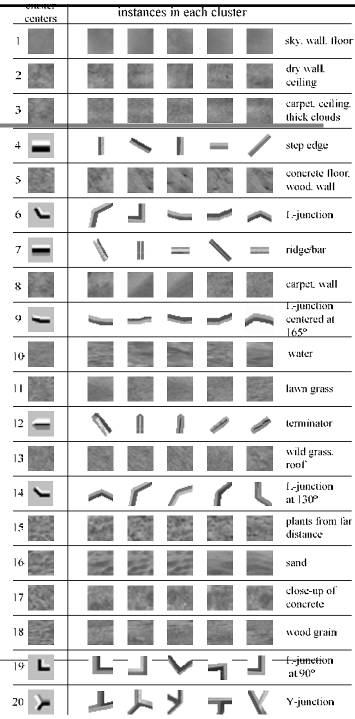




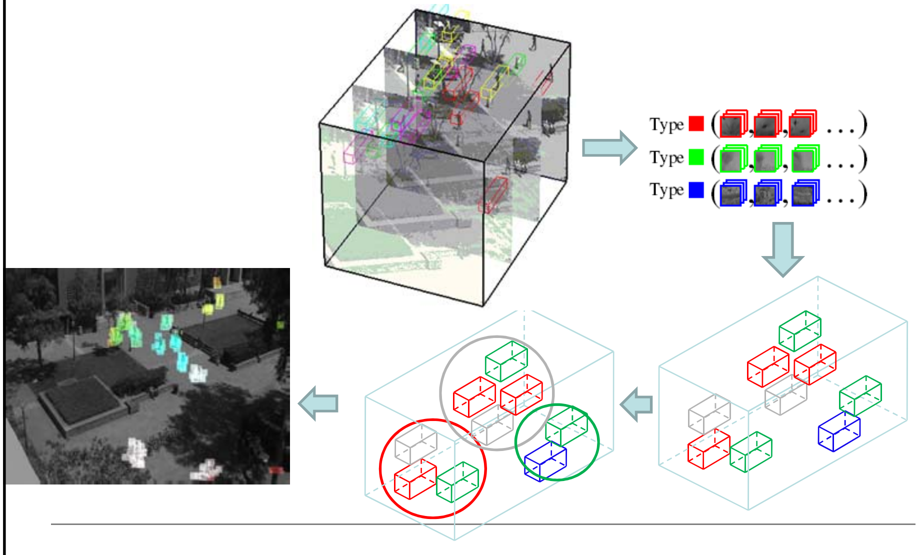
Frequency plot of the ex/implicit manifolds in natural images

implicit texture clusters (blue),
explicit primitive clusters (pink).

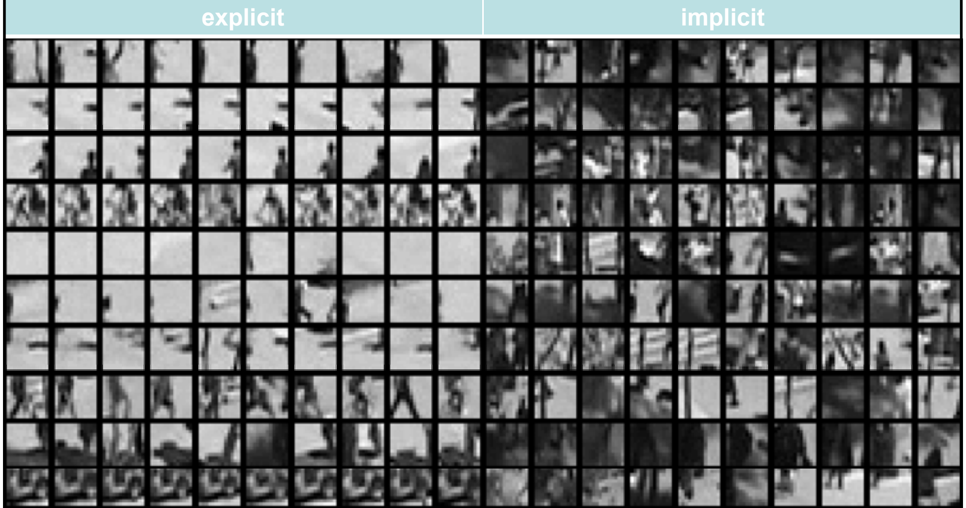




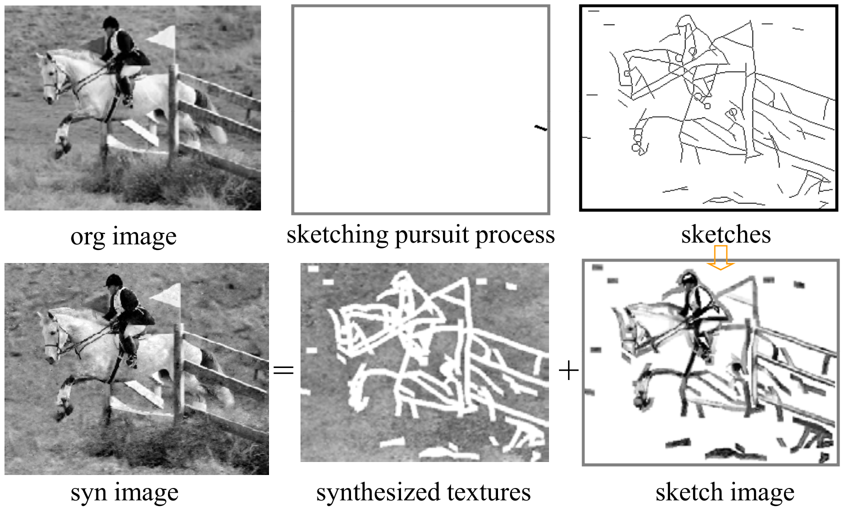
Clustering in video



Examples in video



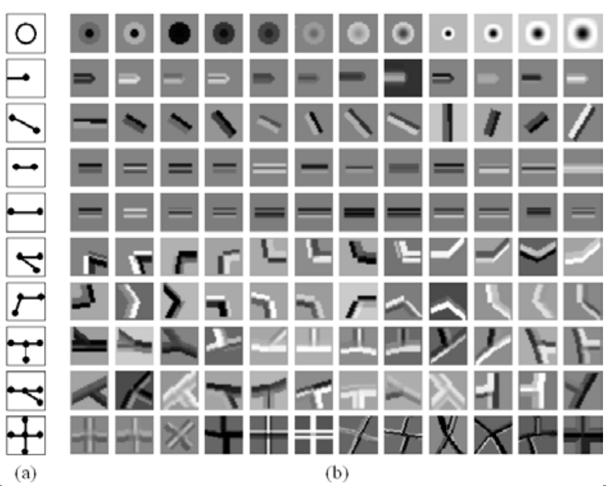
6, Primal sketch: integrating the two regimes



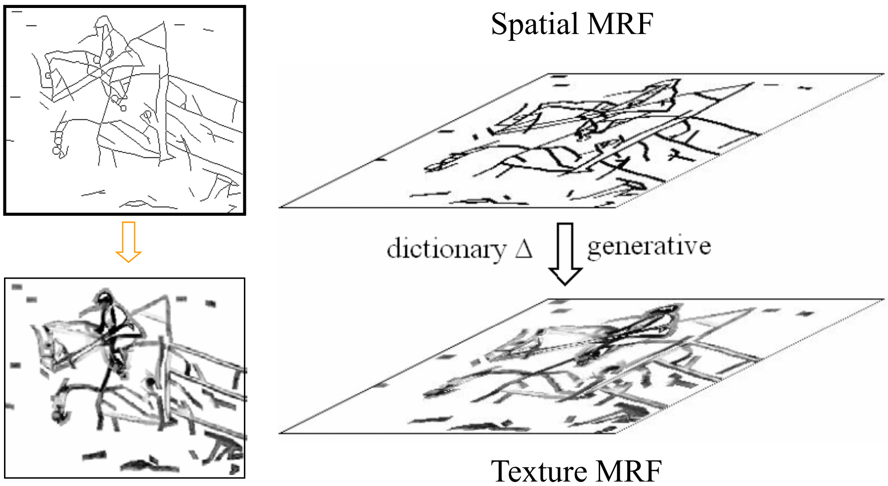
(Guo,Zhu,Wu, 2003-05)

manifolds of image primitives

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



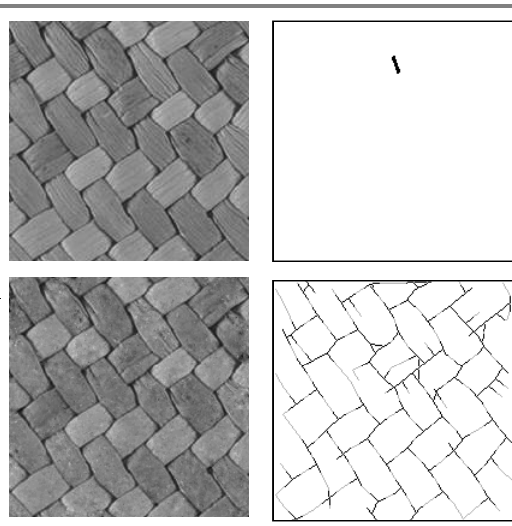
Primal Sketch is a two-level MRF model



Primal sketch example

input image

sketching pursuit process

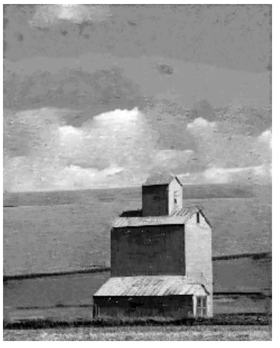


synthesized image

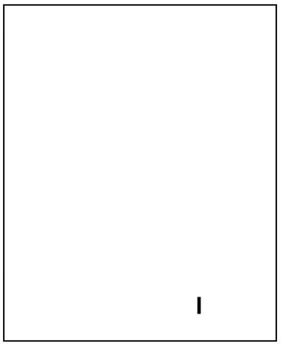
sketches

Primal sketch example

original image

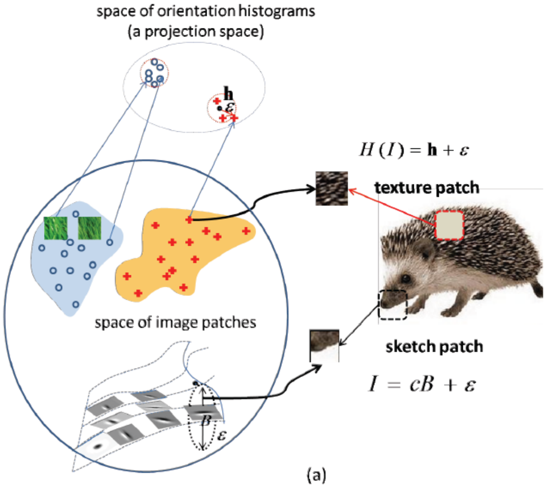


synthesized image



sketching pursuit process

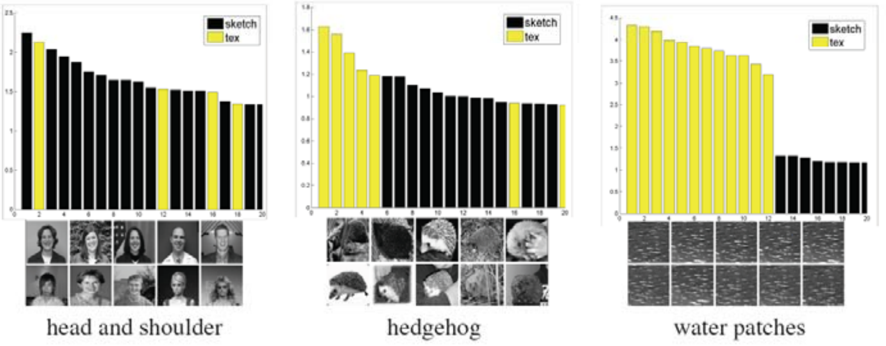
7, deformable template: mixing the im/explicit manifolds



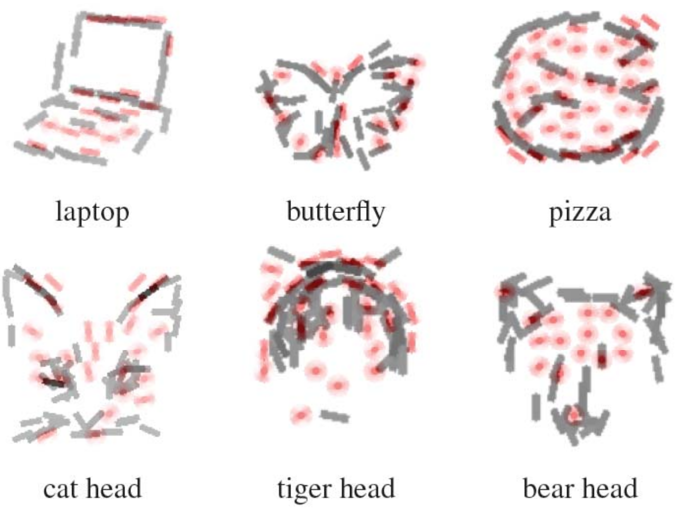
(a)

Si et al 2008

The two types of models compete in learning the templates



Some examples of learn object categories



8, Information scaling leads to manifold transitions !

Scaling (zoom-out) increases the image entropy (dimensions)

Wu, Zhu, Guo, 04,07

Transition of the manifolds through info. scaling

How are these manifolds related to each other ?

perceptual scale space theory (Wang and Zhu 2005)

Summary: understanding the “ingredients of our herbs” !

2 type manifolds, pursuit, integration, mixing, and transition