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:

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

<table>
<thead>
<tr>
<th>Edge</th>
<th>Bar</th>
<th>Two Parallel Lines</th>
<th>Cat</th>
<th>Dog</th>
<th>Lion</th>
<th>Tiger</th>
<th>Fur</th>
<th>Carpet</th>
<th>Grass</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
</tr>
</tbody>
</table>

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

\[ q = \mu_0 \rightarrow \mu_1 \rightarrow \cdots \rightarrow \mu_k \rightarrow f \]

1. \( q = \text{unif()} \)
2. \( q = \delta() \)

image universe: every point is an image.

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:

\[ 0 + 8 + 0 + 3 + 0 + 2 + 0 + 0 + 5 + 0 + 0 + 1 = 19 \]

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.

\[ (50 - 2 - 0 - 3 - 0) + (0 + 5 + 3 + 0 + 0 + 2) = 45 + 10 = 55 \]

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

\[
\Omega^c = \{ f^{(1)}, \ldots, f^{(M)} \} \sim f(I)
\]

eigen-value \( \lambda \)

e.g. texture

mixed: e.g. tiger face

e.g. face

Manifold pursuit by information projection

Given only positive examples from a class \( c \)

\[
q = p_0 \rightarrow p_1 \rightarrow \cdots \rightarrow p_\infty \rightarrow f
\]

At each step, we augment the current model \( p \) to a new model \( p_\infty \)

\[
h_\infty = \arg \max \quad KL( f \mid p) - KL( f \mid p_\infty)
= \arg \max \quad KL( p_\infty \mid p)
\]

Subject to a projection constraint:

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

\( h_+(I) \) is a feature statistics of image I
Manifold pursuit by information projection

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

\[ p^*_n = \arg \min_p \int \nu_n(x) \log \frac{p(x)}{\nu_n(x)} \, dx + \lambda \int \nu_n(x) h_{\nu_n}(x) \, dx - \sum \nu_n(x) + \lambda \int \nu_n(x) h_{\nu_n}(x) \, dx - I \]

where

\[ p_n^*(x) = \frac{1}{Z_n} p_{n-1}(x; \theta_{n-1}) e^{-\lambda \log \nu_n(x)} \]

\[ = \frac{1}{Z_n} q(x) \exp \left\{ - \sum_{i=1}^{k} \lambda_i h_i(x) \right\} \]

For \( q \) being a uniform distribution, we have

\[ q(x) = \frac{1}{Z_n} \]

---

Information projection

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

\[ p_0 = q \]

\[ \Omega_1 = \{ \mu : E_p[h_1(x)] = E_f[h_2(x)] \} \]

\[ \Omega_2 = \{ \mu : E_p[h_2(x)] = E_f[h_2(x)] \} \]

\[ KL(f \mid \mu) = KL(f \mid p_{\mu}) + KL(p_{\mu} \mid \mu) \]

So the KL-divergence decreases monotonically.
A Maximin Learning Principle

max-step: choosing a distinct feature and statistics

\[ h^*_x = \arg \max_{\mathcal{H}_x} KL(p_x | p) \]

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

\[ \lambda^*_x = \arg \min_{\mathcal{H}_x} KL(p_x | p) \]

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”

A texture = \( \Omega(h) = \{ 1: h_i(I) = h_{c,i}, i = 1,2,\ldots,K \} \)

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

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

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

Theorem 1

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

Theorem 2

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

\[ B = (B_1, B_2, \ldots, B_k) \]

\[ A \text{ car} = \Omega = \{ t : t = \sum_{i} y_i B_{i3} \} \]

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

Summary: two pure manifolds

\[ \Omega = \{ I : h(I) = h_0 \} \]

- \( h(I) \) is some image feature/statistics

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

- \( g \) is a generation function,
- \( w \) is intrinsic dimension
- \( \Delta \) 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 vs. 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

<table>
<thead>
<tr>
<th>explicit</th>
<th>implicit</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Explicit Examples" /></td>
<td><img src="image2.png" alt="Implicit Examples" /></td>
</tr>
</tbody>
</table>
6. Primal sketch: integrating the two regimes

![org image]  ![sketching pursuit process]  ![sketches]

![syn image]  ![synthesized textures]  ![sketch image]

(Guo, Zhu, Wu, 2003-05)

**manifolds of image primitives**

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

(a)  (b)
Primal Sketch is a two-level MRF model

Spatial MRF

Texture MRF

Primal sketch example

input image

synthesized image

sketching pursuit process

sketches
Primal sketch example

original image  synthesized image  sketching pursuit process

7. deformable template: mixing the im/explicit manifolds

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)

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