Ch2: Statistics of “Natural Images”

Topics in this chapter:

• 1/f-power law, high kurtosis, scale invariance,
• high-order structures,
• manifolds in the image universe,
• transitions in a spectrum of image spaces

Reasons for studying the natural image statistics

**Perspective 1:** Redundancy reduction and better image/video coding
e.g. best codebook and strategies in image/video compression.

**Perspective 2:** Ecologic influences on neural receptive fields
e.g. how are neurons in the visual cortical areas adapted to the visual environments.

**Perspective 3:** Regularities and models for solving ill-posed vision problems
e.g. image restoration (denoising, inpainting etc)
e.g. estimating surface (depth map) from stereo, motion, texture, shading etc.
e.g. developing concepts and generative models (as prior knowledge)
for scenes, objects, actions, events, and causality from images and video.
The space of Images:

--- “Natural image” are special and have regularities

Consider an image \( I \) with 256 grey levels on a lattice \( \Lambda \) of \( N \times N \) pixels, \( N = 256 \).

The volume of image space \( | \Omega_I | = 2^{256 \times 256} = 10^{157,830} \)

The volume of natural image ensemble \( | \Omega_{I_n} | \approx 2^{0.3 \times 256 \times 256} \approx 10^{5.718} \)

The volume of images seen by all humans \( \leq 10^{10} \times 10^{10} = 10^{20} \)

[10 billion people \( \times 100 \text{ yr} \times 365 \times 24 \times 3600 \times 20] \)

Scientific quest: what are the structures in this space?

A “Telescope” to Look Into The Image Universe

What are the structures in the space of images?
Ecologic influences on neural receptive fields

What the cortical cells (in V1) see?

Huber and Weissel 1960s on cat experiments Tai Sing Lee’s Monkey

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

Neurons in V1 seem to compute some basic image elements

Why and how is this response function developed?

It is not obvious what the local features are. People are still mining diligently …

This motivated many explanations and models: edge detection, texton, and Gabor filters in the 1980s. It also inspired wavelets in thinking.

This one-layer model has been drastically extended by multi-layered DNN.

Question:

Do the V1 neurons in monkey and rabbit differ in their response functions? since they live in different environments, i.e. data and face different tasks.

Note that, when we are talking about a model and representation, it is not merely decided by the data, but also by the tasks. This is different from physics.
Statistics of “Natural Images” from a math/stat perspective

Natural Images often refer to the scenes that contain natural objects with rich details and a continuum of scales and the optical axis of the camera is horizontal.

They are very different from artificial or noise images.

Ruderman and Bialek 87, 94
Fields 87, 94
Zhu and Mumford 95-96
Chi and Geman 97-98
Mumford and Gidas 00
Simoncelli etc 98-03
Wu et al 06-09
Hierarchical models

The ultimate goal is to learn a probability model $p(I)$ for images $I$.
The natural image statistics are marginal distributions of $p(I)$.

Examples of Natural Images

Natural images: intuitively they are images taken in real environments with handhold forward-looking cameras. They contains objects of varying sizes or similar sizes at varying distances.
Empirical observation 1: Redundancy in real world images

From D. Kersten 1987

Empirical observation 2: 1/f Power law

1/f power plot of six images (Fields 1994 paper), a randomized image with 1/f power (from Mumford’s paper)
Empirical observation 3: high kurtosis $\rightarrow$ sparsity

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)

What is your explanation?

Some notations

Some key notations discussed on the blackboard:

1. Fourier transform and power over a frequency band,
2. entropy and entropy rate,
3. redundancy,
4. high Kurtosis,
5. Gaussian, Super/sub-Gaussian, Laplacian, Cauchy distributions,
6. histograms and marginal distributions,
7. sparse vs compact coding.
Empirical observation 4: Generalized Laplacian (log-plot)

![Generalized Laplacian](image)

*Figure 3.* Estimated Bessel $K$ and generalized Laplacian densities compared to the observed histograms for the images in top. Plain lines are histograms, lines with large beads are Bessel $K$, and lines with small beads are generalized Laplacian. These densities are plotted on a log scale.


Empirical observation 5: scale invariant gradients histogram

![Scale Invariant Gradients](image)

Gradient histogram of Natural Images, log-plot

white noise images don't observe

The different curves are for 3 scales. (picture from Zhu and Mumford, 1997)
Example image which has scale invariant gradient histograms

(Zhu and Mumford, 1996)

Chi’s 3D Explanation on Scale Invariance

Simulated image using two primitives: cylinders and spheres with Lambertian shading and texture surface.
Chi's 3D Explanation on Scale Invariance

The simulated images exhibit scale invariance.

Scale invariance is not true for DC components

Intensity histograms of natural images

- one individual image
- averaged over 44 images
Natural images:
objects of varying sizes sitting on ground in depth
Observation 6: Edge manifold

The probability density increases exponentially as we move towards the manifold.
Example of a manifold for illumination variations

Each element is represented by a triplet of textons

“lighton”? ![Image]

Sampling the 3D elements under varying lighting directions

An example of application: night time surveillance

In surveillance video, often the camera view is fixed. In order to detect the foreground moving objects: cars, pedestrians, bikes, one needs a background model.

Look at the patch of 15x 15 pixels in the red window.

![Image]
Looking at a single patch over time

Projecting a 15x15 x 7 spatio-temporal patch ("brick") in first 2-PCA space

The red dots are the brick at silent night time, the green curve is the patch when a car drove through the scene.

Using the model for background modeling in surveillance video

Empirical observation 7: entropy rate over scales

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

Statistics Change over Scales

Entropy rate (bits/pixel)
Entropy rate (bits/pixel) over distance

1. entropy of $I_x$
2. JPEG2000
3. # of DooG bases for reaching 30% MSE
Scaling and model transition

Watch the texture-texton transitions over scale / distance

A photo at Bei Air

Sketchable vs. Non-sketchable

An example: the leaves

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

Image representations: sketchable vs. non-sketchable
Information scaling and regimes of models!

Issues: not only the perception changes, our models and coding mechanism changes as well.

Wu, Zhu, Guo, 04.07

Image patches at different resolutions

Wu, Zhu, Guo, 04.07
Coding efficiency and number of clusters over scales

**Definition: Image Complexity**

Let $I \sim p(I)$ defined on lattice $\Lambda$, the *image complexity* $H(I)$ is defined as the entropy of $p(I)$

$$H(I) = -\sum p(I) \log p(I)$$

*Image Complexity Rate* (per pixel) is:

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

Down-scaling = local smoothing + down-sampling

*Note that:* complexity $\neq$ information
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

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

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

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

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

---

Image Down-Sampling

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

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

$M(\ldots)$: 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.
\[ \frac{1}{4} \sum_{k=1}^{4} H(I^{(k)}) \geq H(I) \]
(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.

Scale Invariant if
\[ M \approx \hat{K} \]
\[ M \equiv M(I^{(1)}, I^{(2)}, I^{(3)}, I^{(4)}) \]
\[ K \equiv -\int \log |\hat{k}(w)| dw \]
Imperceptibility--- explaining perceptual transitions

Generative Model \( W \sim p(W), I = g(W) \)

Entropy measures the complexity and uncertainty
\[
H(p(I)) = E[-\log p(I)] = -\int p(I) \log p(I) dI
\]

\[
H(p(W \mid I)) = H(p(W)) - H(p(I))
\]

imperceptibility = Scene complexity – Image complexity

When the posterior probability \( p(W \mid I) \) has high entropy imperceptibility, it means that certain variables in \( W \) cannot be inferred (as uncertainty is too high), and thus are imperceptible. We need to reduce our representation of the world.

Perceptibility Scaling Law --- explaining perceptual transitions

**Theorem 3.** Imperceptibility increases with down-scaling.
If \( W \sim p(W), I = g(W), I_\downarrow = R(I) \) by down-scaling
Then \( H(W \mid I_\downarrow) \geq H(W \mid I) \)

One underlying assumption is that our perception changes tasks, and infer those representations that can be computed reliably.

Each visual concept has a finite “lifespan” in scales

Many vision people have noticed this phenomenon.

Different motion representations in computer vision

Video representation: trackable vs. non-trackable motion
Information scaling and perceptual transition in video

Density (occlusion) and motion dynamics also contribute to perceptual transitions.

Video representations: trackable vs. non-trackable

An example: the birds

Info. scaling can be triggered by scale (sizes), density (occlusion) and dynamics (speed).

Mapping the Image Universe at different entropy rate

Iso-contours of entropy rate

Landscape mapping by effective MCMC

Simulating MCMC to explore the space and identify the local basins and ridges.
Texture vs texton regimes in the Landscape

Discussion: Imperceptibility and Abstract Arts

Computer generated
Imperceptibility and perceptual paths

Controlling the Imperceptibility and perceptual path

- Assuming ambiguity is only in Object Recognition
  \[ \mathcal{L} = (\ell_1, \ell_2, \cdots, \ell_K) \]
- Uncertainty in \( p(\mathcal{L}|\mathbf{I}) \): Shannon Entropy
  \[ \mathcal{H}(\mathcal{L})|\mathbf{I} = \sum_{\mathcal{L}} -p(\mathcal{L}|\mathbf{I}) \log p(\mathcal{L}|\mathbf{I}) \]
Controlling the Imperceptibility and perceptual path

(a) $\tilde{H} \approx 0.25$  (b) $\tilde{H} \approx 0.5$  (c) $\tilde{H} \approx 0.75$