Statistics of Natural Images

Course Material for IT 530
(Image Representation and Analysis) at
DA-IICT.

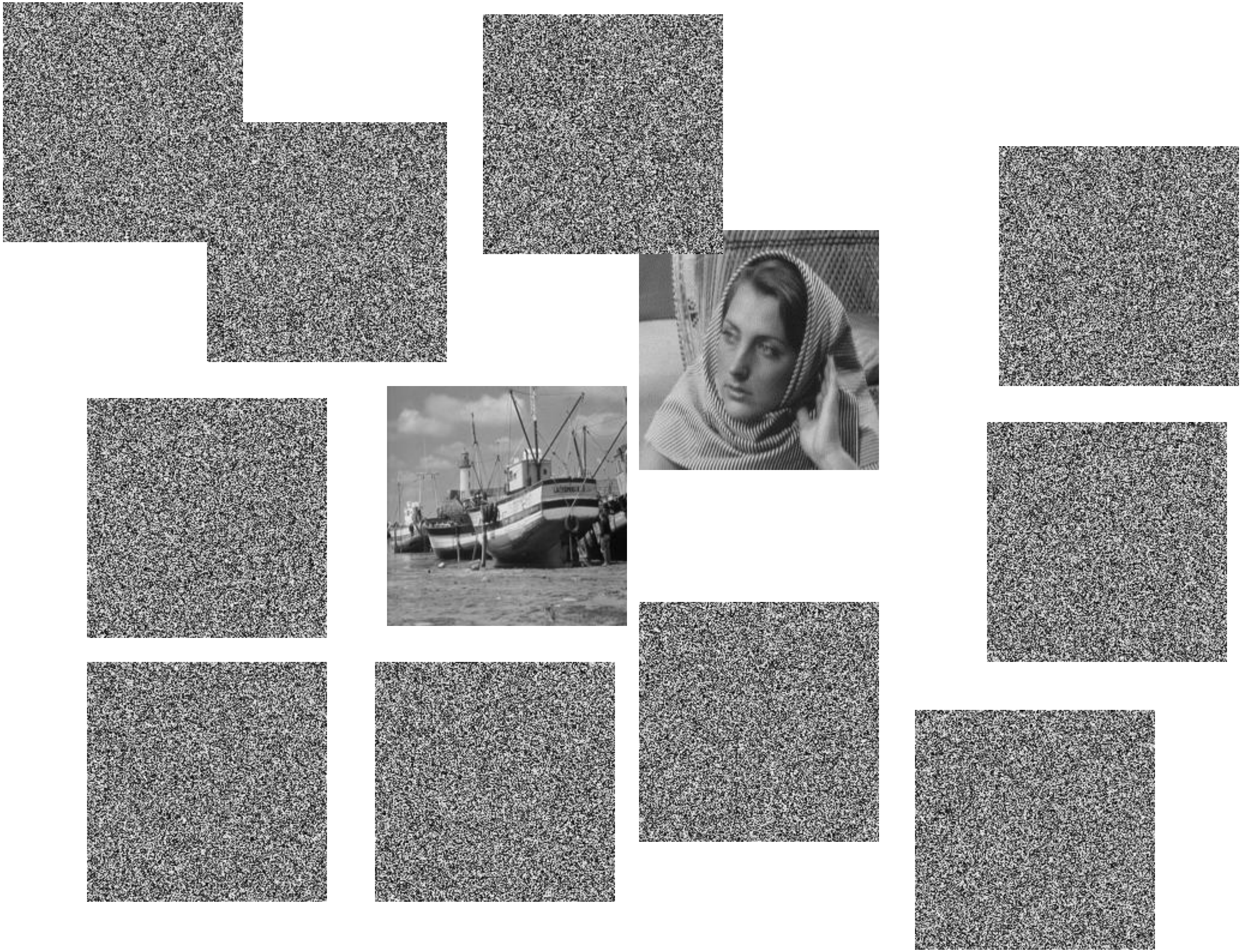
Course Instructor: Ajit Rajwade

Motivation

- Number of possible 200 x 200 images (of 256, i.e. 8 bit intensity levels) = 256^40000 = 2^320000 = 10^110000.

- This is several trillion times the number of atoms in the universe (10^90).

- Only a tiny subset of these are plausible as natural images.

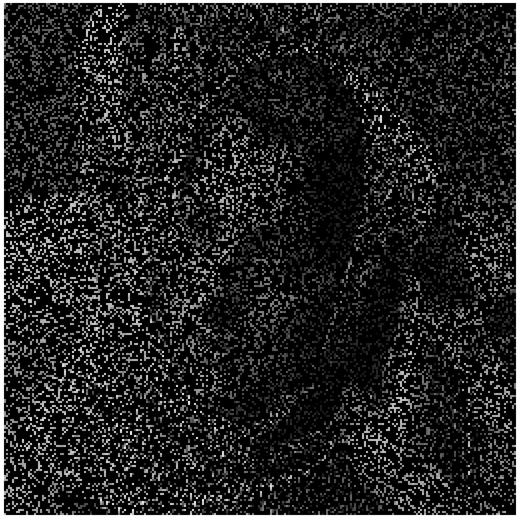


Why study statistics of natural images?

- Useful for many computer vision applications

1. Image denoising, deblurring, filling of missing pixels in images (inpainting)
2. Image compression
3. Segmentation of images
4. Synthesizing textures
5. Classification into image categories


Since 1699, when French explorers
landed at the great bend of the
Mississippi River and celebrated
the first Mardi Gras in North America,
New Orleans has brewed a fascinating
melange of cultures. It was French,
then Spanish, then French again, then
sold to the United States. Through all
these years, and even into the 1900s,
others arrived from everywhere:
Acadians (Cajuns), Africans, indige-




Why study statistics of natural images?

- Natural image statistics also help us understand the human visual system better.

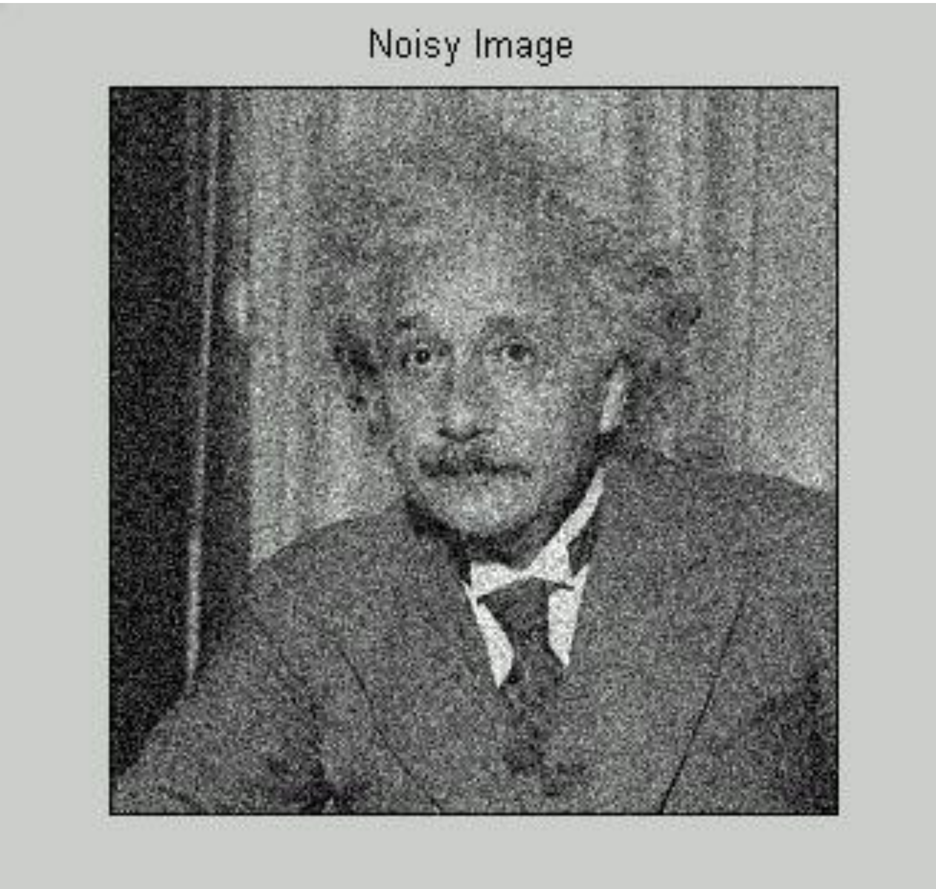
Real-world signal (scene – static or dynamic)

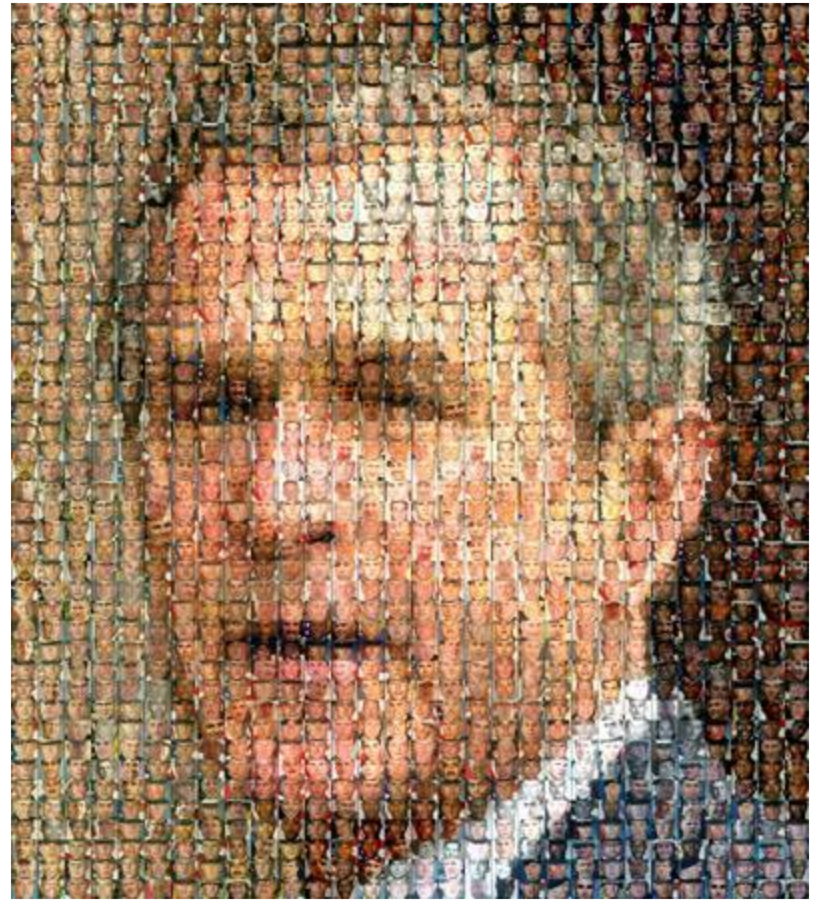
Eye (Retina)

Visual Cortex

Convolution with Blur Kernel + Noise

Brain is solving an inverse problem!


Noisy Image



Why study Natural Image Statistics: Bayesian Framework

$$y = g * x + \varepsilon$$

Observation

Known Blur kernel

Unknown (to be determined)

noise

$p(y \mid x)$ → Likelihood

$p(x)$ → **Prior Model**

Posterior probability

Bayes rule

$$p(x \mid y) = \frac{p(y \mid x)\, p(x)}{p(y)} \propto p(y \mid x)\, p(x)$$

Why study Natural Image Statistics: Bayesian Framework

$$y = g * x + \varepsilon$$

Assuming zero-mean i.i.d. (additive) Gaussian noise model, the likelihood is as follows:

NOTE: Likelihood derived from the assumed noise model

$$p(y \mid x) \propto \exp\left(-\frac{\left\| y - g * x \right\|^2}{2\sigma^2} \right)$$

Bayesian Framework: To Estimate x

- Maximum a posteriori (MAP) estimate:

$$\widehat{x} = \arg\max_x \; p(x \mid y) = \arg\max_x \; p(y \mid x)\, p(x)$$

- Minimum mean square error (MMSE) estimate:

Prior is important!

$$\widehat{x} = \arg\min_z \int \|x - z\|^2 p(x \mid y)\, dx$$

$$\therefore \; \widehat{x} = \frac{\int x p(y \mid x)\, p(x)\, dx}{\int p(y \mid x)\, p(x)\, dx} = E(x \mid y)$$

Simple Example: 1

Prior

$$p(x=10)=0.7$$

$$p(x=15)=0.3$$

Likelihood

$$y=x+\varepsilon$$

$$\varepsilon \sim N(0,\sigma),\sigma=2$$

Observed Value of y = 14. Determine x given y and the knowledge of the noise model (likelihood) and prior on x.

Simple Example :1

$$\widehat{x}_{MAP} = \arg\max_x \ p(x \mid y) = \arg\max_x \ p(y \mid x)\, p(x)$$

$$= \arg\max[\,0.7 \times e^{-(14-10)(14-10)/(2\times 4)}, 0.3 \times e^{-(14-15)(14-15)/(2\times 4)}\,]$$

$$= \arg\max[\,0.0947, 0.2647]$$

$$= 15$$

$$\widehat{x}_{MMSE} = \frac{\int x p(y \mid x)\, p(x) dx}{\int p(y \mid x)\, p(x) dx}$$

$$= \frac{10(e^{-2})(0.7) + 15(e^{-1/8})(0.3)}{(e^{-2})(0.7) + (e^{-1/8})(0.3)} = 13.6825$$

Simple Example: 2

$$x \sim N(1,2), \mu_x = 1, \sigma_x = 2$$

$$y = x + \varepsilon$$

$$\varepsilon \sim N(0,3), \mu_y = 0, \sigma_y = 3$$

Observed Value of y = 2. Determine x given y and the knowledge of the noise model (likelihood) and prior on x.

Simple Example: 2

$$\widehat{x}_{MAP} = \arg\max_x p(x \mid y) = \arg\max_x p(y \mid x) p(x)$$

$$= \arg\max_x e^{-(y-x)^2/(2*3*3)} e^{-x^2/(2*2*2)}$$

$$= \arg\max_x e^{-(x-\hat{\mu})/(2\hat{\sigma}^2)}$$

$$= \hat{\mu}(why?)$$

$$\hat{\mu} = \frac{\mu_x {\sigma_y}^2 + \mu_y {\sigma_x}^2}{{\sigma_x}^2 + {\sigma_y}^2} = \frac{1*4+0}{4+9} = \frac{4}{13}$$

Simple Example: 2

$$\widehat{x}_{MMSE} = \frac{\int xp(y \mid x)p(x)dx}{\int p(y \mid x)p(x)dx}$$

$$= \frac{\int xe^{-(x-\hat{\mu})/(2*\hat{\sigma}^2)}dx}{\int e^{-(x-\hat{\mu})/(2*\hat{\sigma}^2)}dx} = \frac{\sigma\sqrt{2\pi}mean(x)}{\sigma\sqrt{2\pi}} = \hat{\mu} = \frac{4}{13}$$

When the likelihood and prior are both Gaussian, the MAP and MMSE estimates are equal.

Maximum Likelihood Estimation

$$\widehat{x}_{ML} = \arg\max_x \; p(y \mid x)$$

If there is only one observation sample available (assume Gaussian noise), what is the maximum likelihood estimate of *x*?

If there are some *N* observation samples available (under Gaussian noise), what is the maximum likelihood estimate of *x*?

These two examples were very simple and involved scalar quantities. In future lectures, we will use more complex examples, where the unknown quantity x will be multivariate – in fact, it will be an image.

Power law for natural images

- The squared-amplitudes of the frequency components of a typical natural image undergo rapid decay w.r.t. frequency f

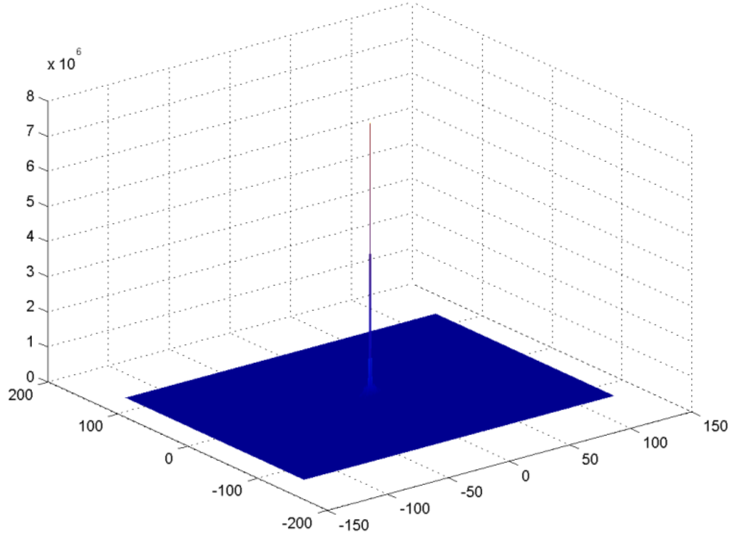
$$S^2(f) = Af^{\alpha-2}$$

where α is a small number between 0 and 1 (usually around 0.2), *A* is a constant.

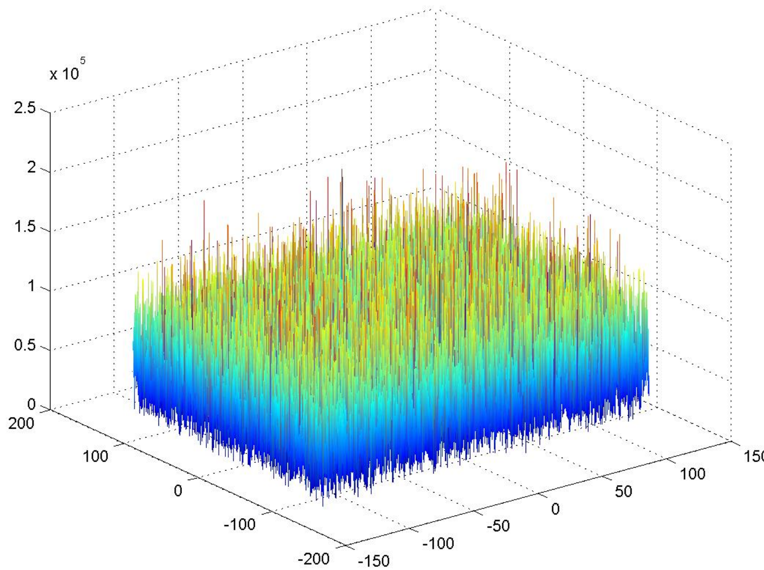
- This property holds across all different image scales (scale-invariance). **[Ruderman and Bialek, "Statistics of Natural Images: Scaling in the Woods"]**

- Why is the power law useful?

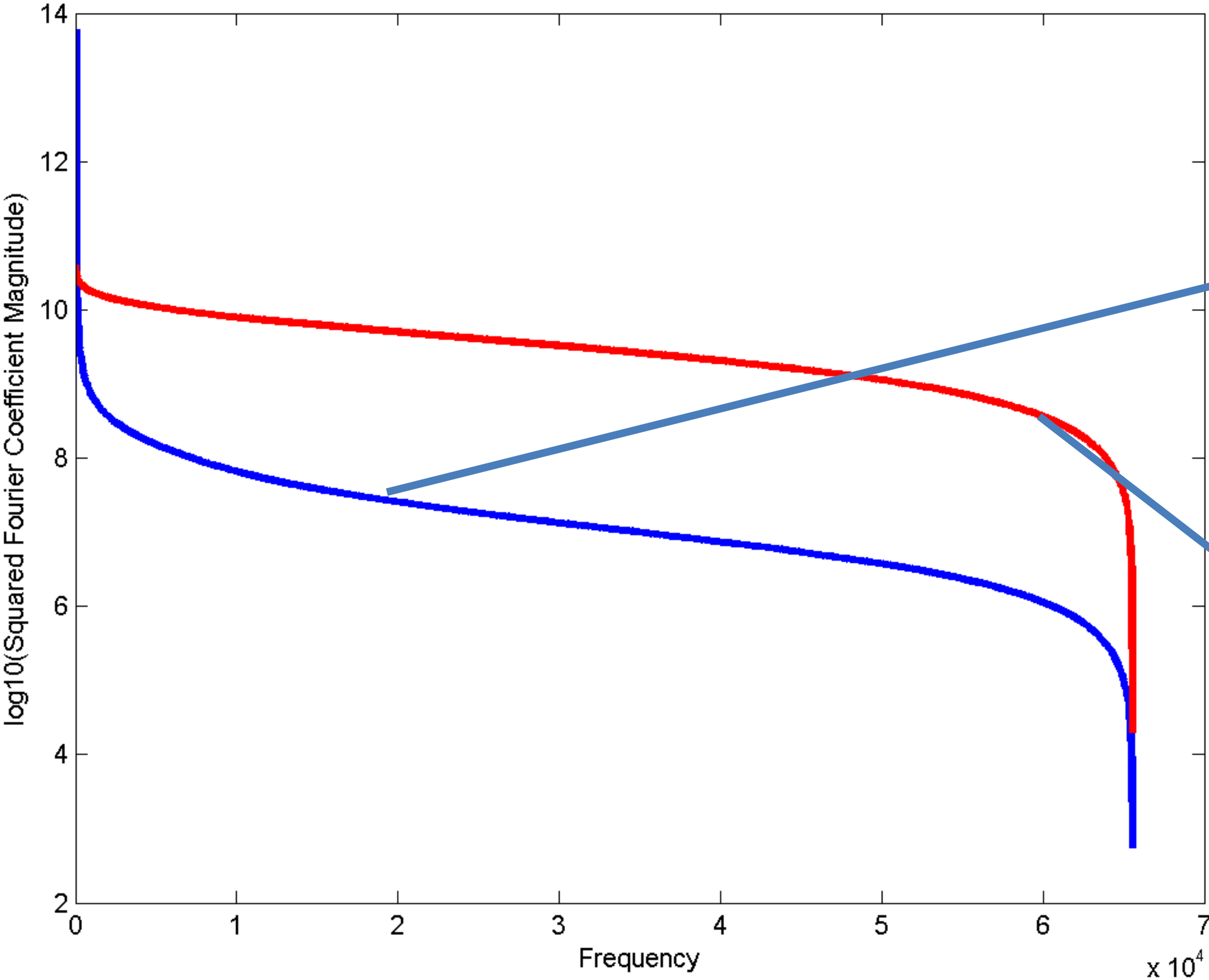
Spectrum of Barbara

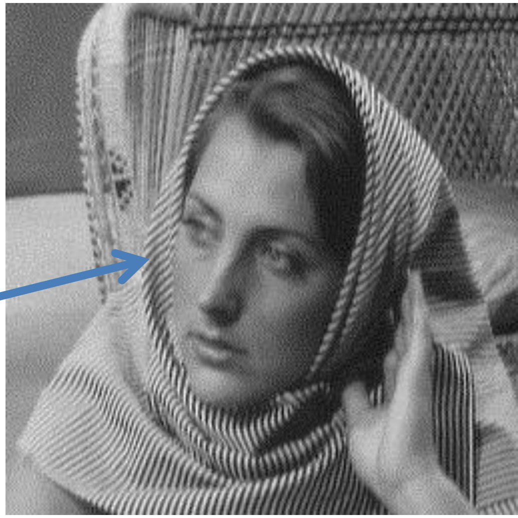


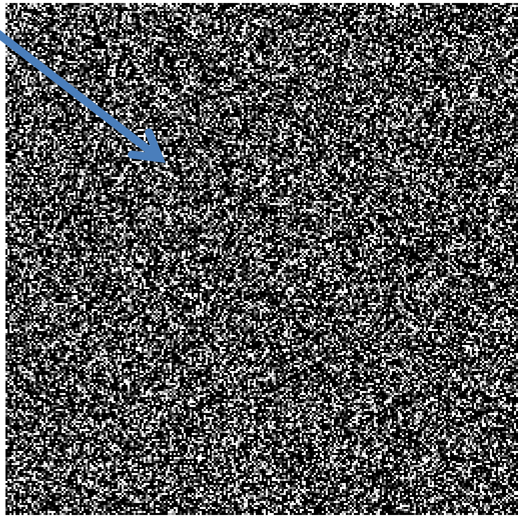
Spectrum of Noise



Power Law


log10(Squared Fourier Coefficient Magnitude)
14
12
10
8
6
4
2
0
1
2
3
4
5
6
7
Frequency
x 10^4





Power Law

- Power law is tremendously useful for (lossy) image compression.
- Image energy concentrated in just first few Fourier coefficients. Remaining coefficients are small can be ignored (i.e. considered to be 0).
- This principle is used by the JPEG algorithm (DCT coefficients), usually at a patch (block) level (patch-size of 8 x 8, usually).

Scale invariance: pixel contrasts

- Pixel contrasts are defined as follows:

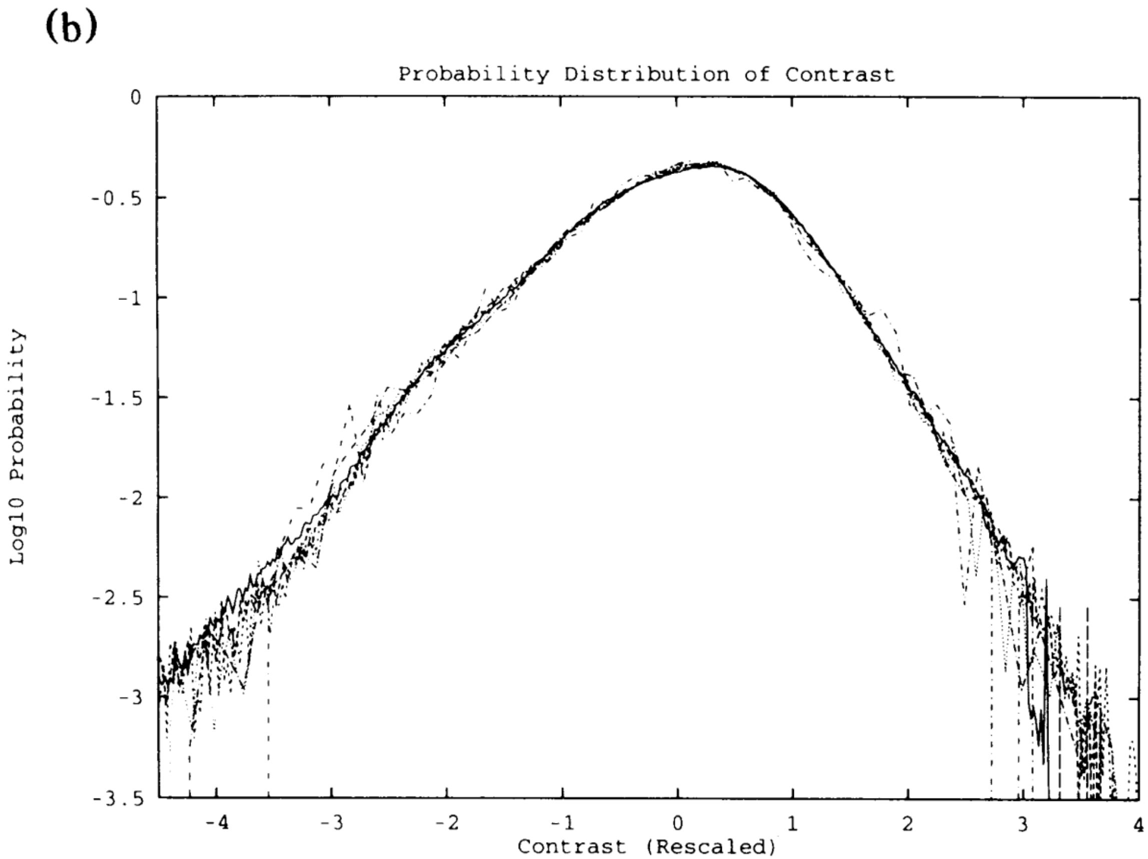
$$\phi(x) = \ln\left(\frac{I(x)}{I_0}\right)$$

A factor adjusted such that the average image contrast is 0.

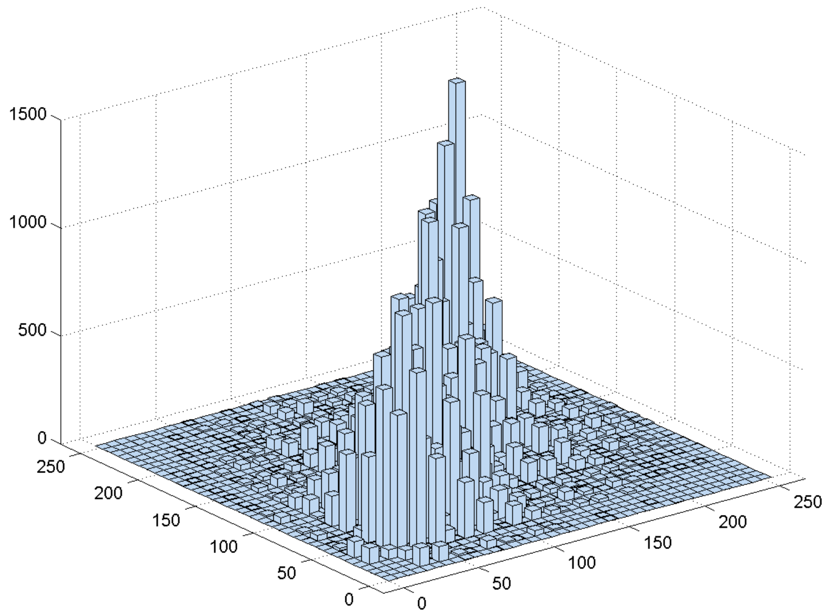
- The distribution of pixel contrasts is (almost) the same for pixels of various sizes from 1 x 1 to 32 x 32.
- Distribution is non-Gaussian (heavy tails).

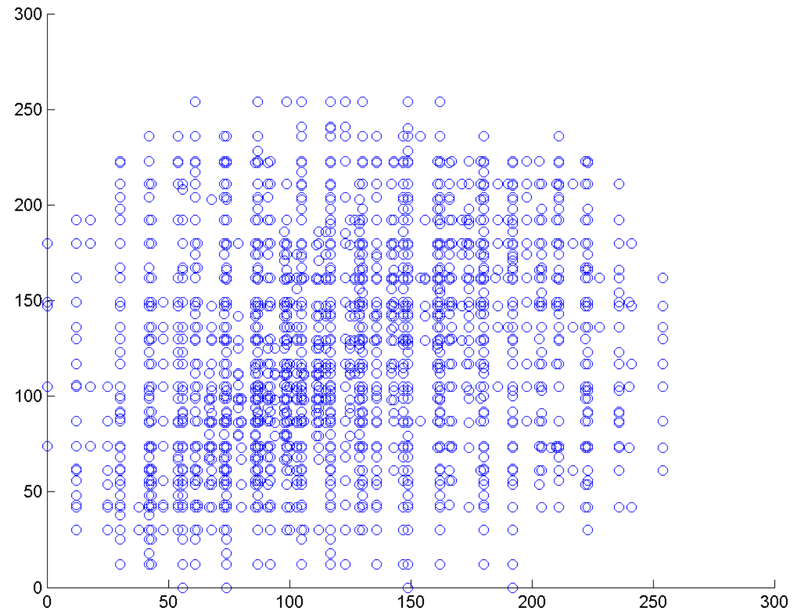
Scale invariance: pixel contrasts

Ref: Ruderman and Bialek, 1994.



Joint statistics: pixel and right neighbor

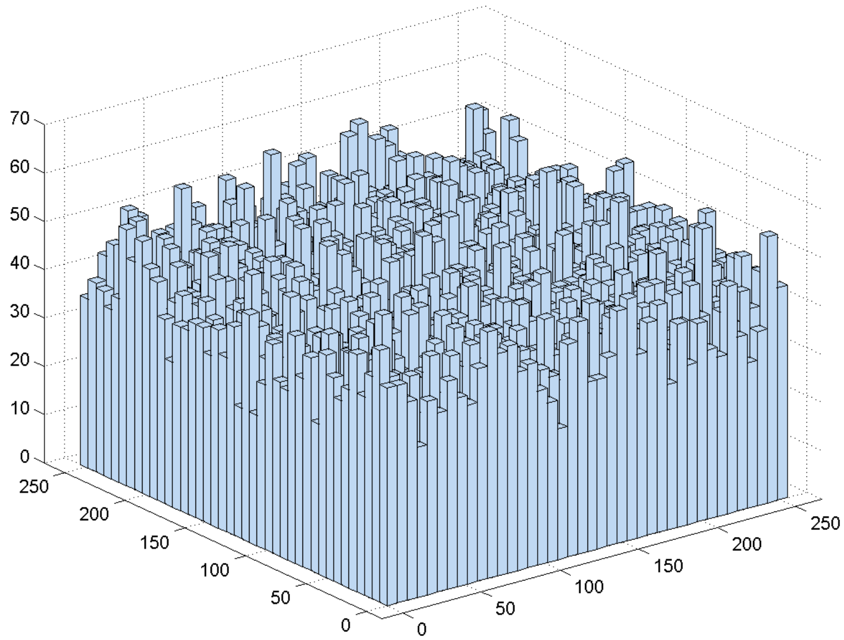


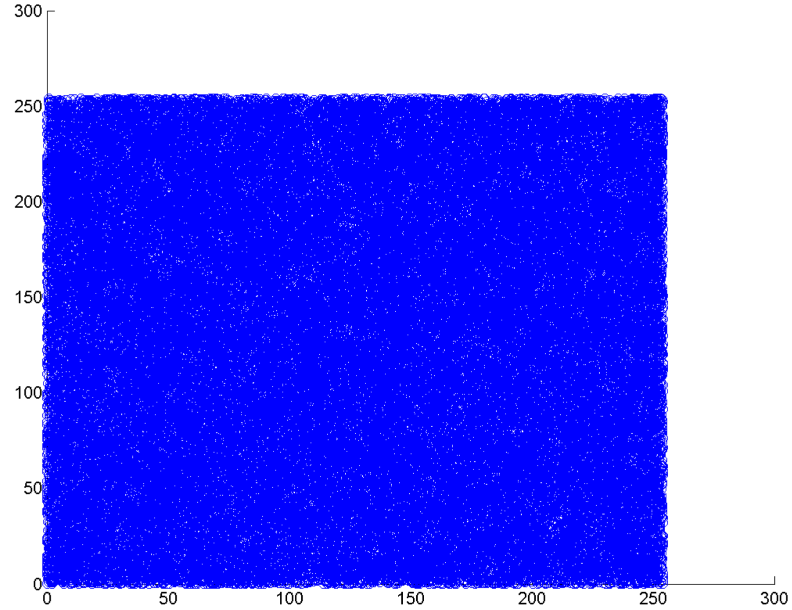


What do you think the corresponding plot for a pure noise image would look like?

Ref: Huang and Mumford, CVPR 99.

Joint statistics: pixel and right neighbor





Small segway: Generalized Gaussian Distribution

$$p(x;\mu,\sigma^2)=\frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

$$p(x;\mu,\sigma^2,\beta)=\frac{\beta e^{-(|x-\mu|/\sigma)^\beta}}{\sigma\Gamma(1/\beta)}$$

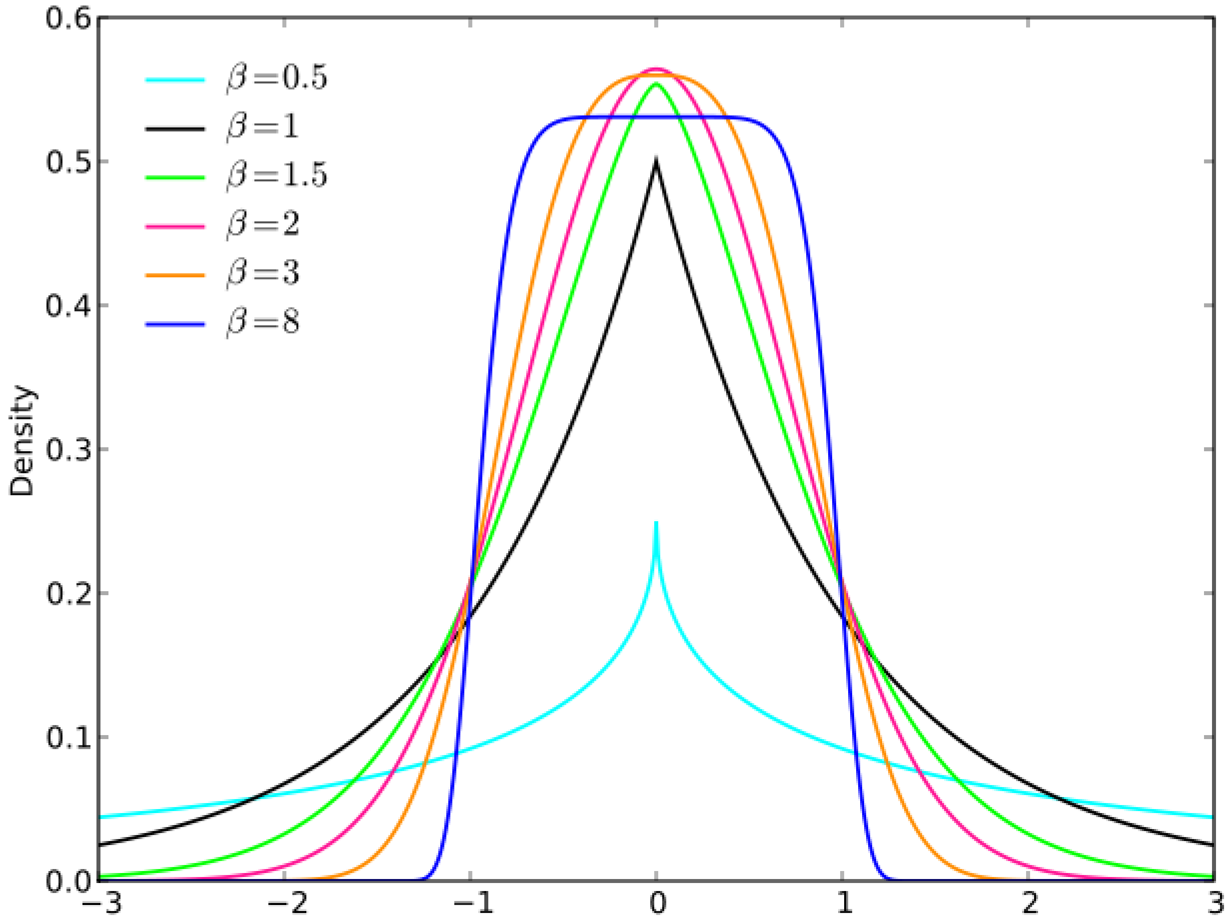
Shape parameter

Scale parameter

$\beta=2$ → Gaussian

$\beta=1$ → Laplacian

$\beta\to\infty$ → Uniform



0.6
0.5
0.4
0.3
0.2
0.1
0.0
Density
−3
−2
−1
0
1
2
3
$\beta = 0.5$
$\beta = 1$
$\beta = 1.5$
$\beta = 2$
$\beta = 3$
$\beta = 8$

Marginal statistics of wavelet coefficients

- Very peaky, heavy-tailed
- Approximated well by Generalized Gaussian Distributions.

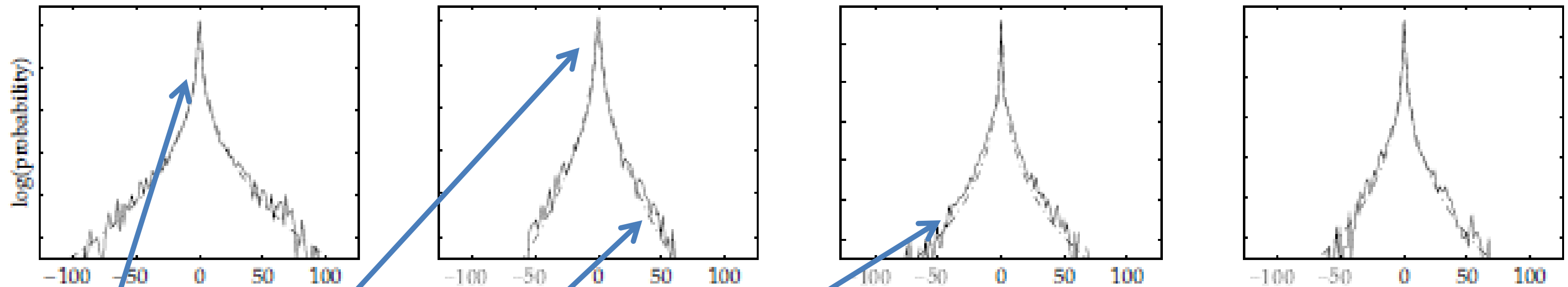
$$p(x;\mu,\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}}$$

Ref: Simoncelli, “Statistical Models for Compression, Restoration and Synthesis”

Scale parameter

Shape parameter (between 0.5 and 1 for density of wavelet coefficients of natural images)

Least squares fitting of Generalized Gaussian Distributions to wavelet coefficient histograms



$\kappa = 14.5, \ \ \Delta H = 0.079$ $\kappa = 18.9, \ \ \Delta H = 0.067$ $\kappa = 22.9, \ \ \Delta H = 0.25$ $\kappa = 23.0, \ \ \Delta H = 0.076$

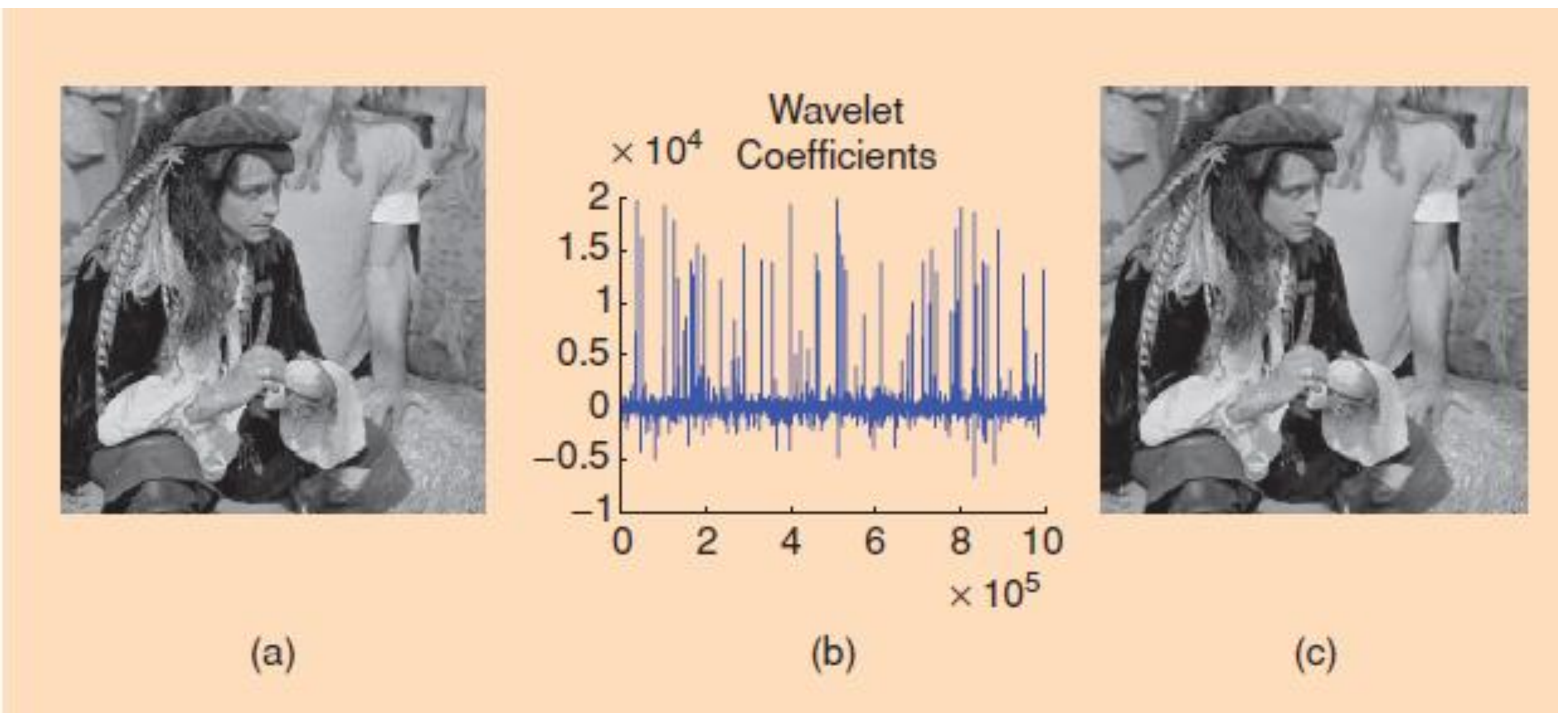
Figure 1. Examples of 256-bin coefficient histograms for vertical bands of four images ("Boats", "Lena", "CTscan", and "Toys"), plotted in the log domain. Also shown (dashed lines) are fitted model densities corresponding to equation (1). Below each histogram is the sample kurtosis (fourth moment divided by squared variance), and the relative entropy of the model.

Smoother regions: smaller coefficient values

Textured regions/edges: larger values

Use of marginal statistics of wavelet coefficients

- Follow exponential decay rule (like Fourier coefficients) – small coefficients can be ignored, the remaining can be coded.



(a) (b) (c)

Use of marginal statistics of wavelet coefficients

- The significant wavelet coefficients can be (say) Huffman encoded using their histogram.

- This can be used in image compression algorithms.

- BUT YOU CAN DO EVEN BETTER!!

Joint Statistics of Haar wavelet coefficients

Ref: Huang and Mumford, CVPR 99.

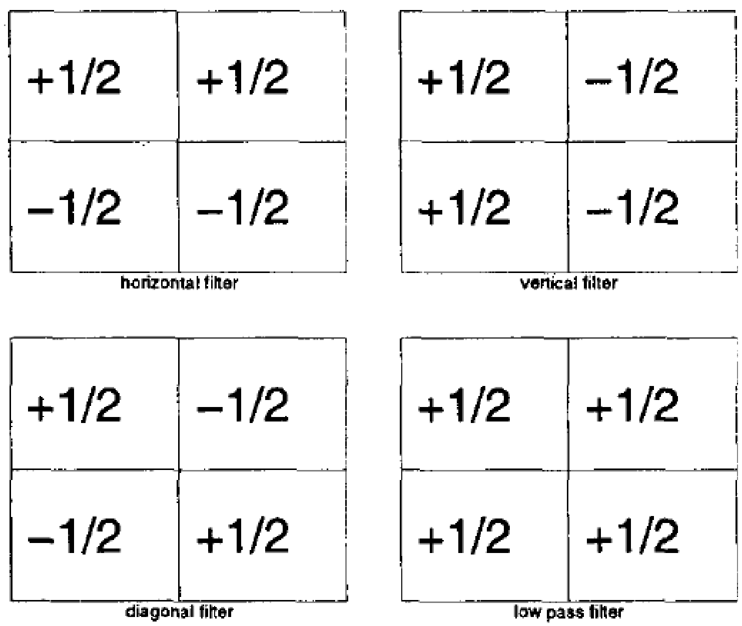
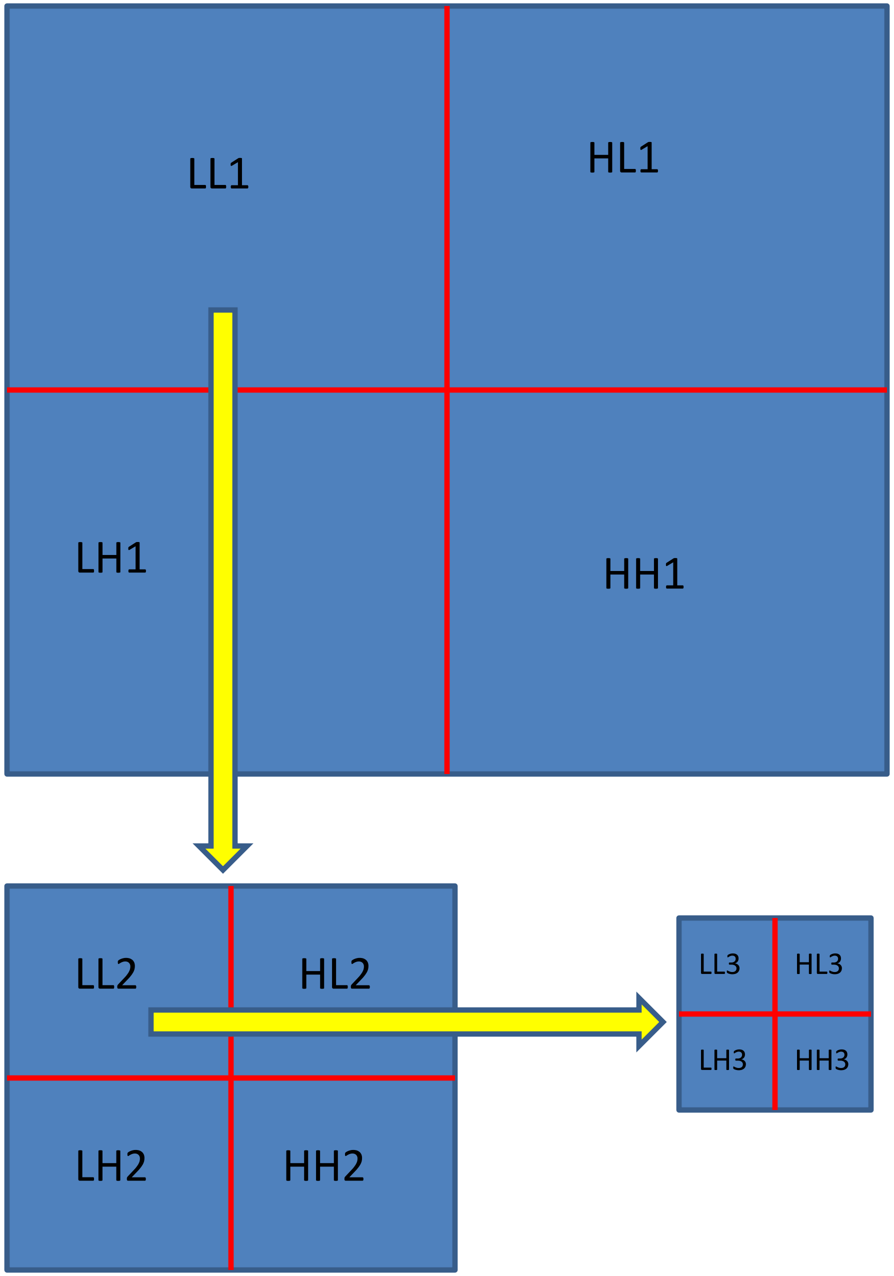


Figure 5: Haar Filters

Coefficients computed at multiples scales of the Haar wavelet pyramid

Concept of parent, child, sibling and cousin coefficients (all are called wavelet sub-bands)



Three-level wavelet decomposition of an image

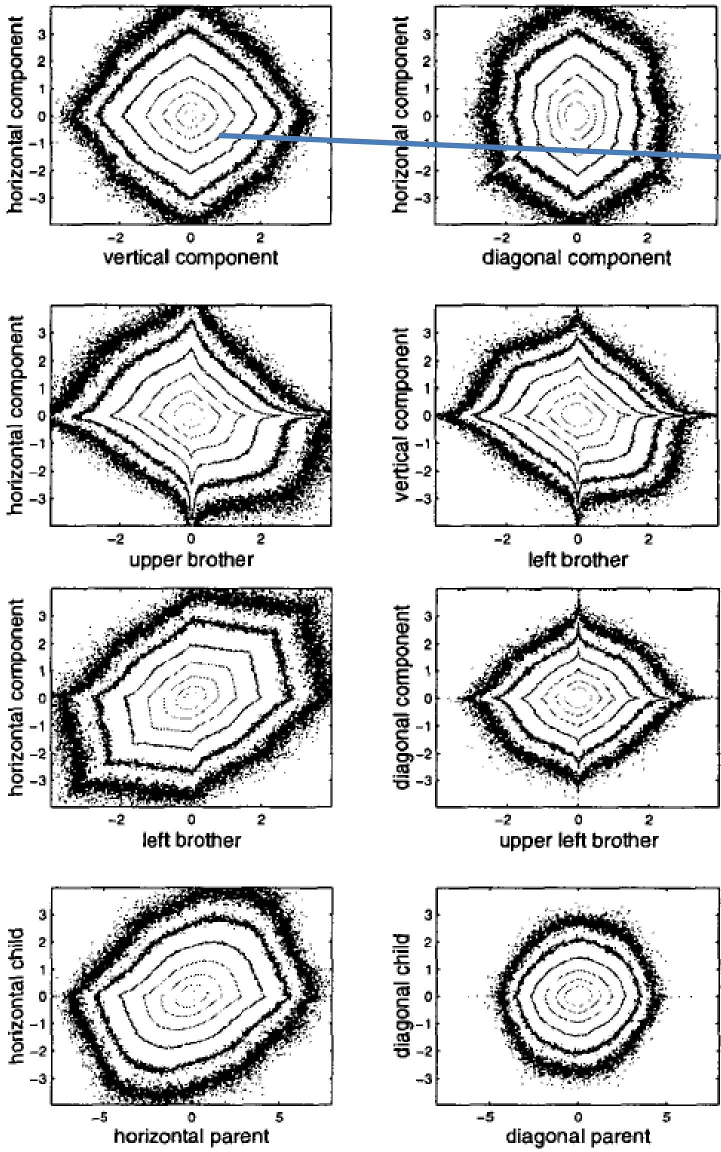


Figure 8: Contour Plot of the log(histogram) of several wavelet coefficient pairs

$|x|+|y|=1$

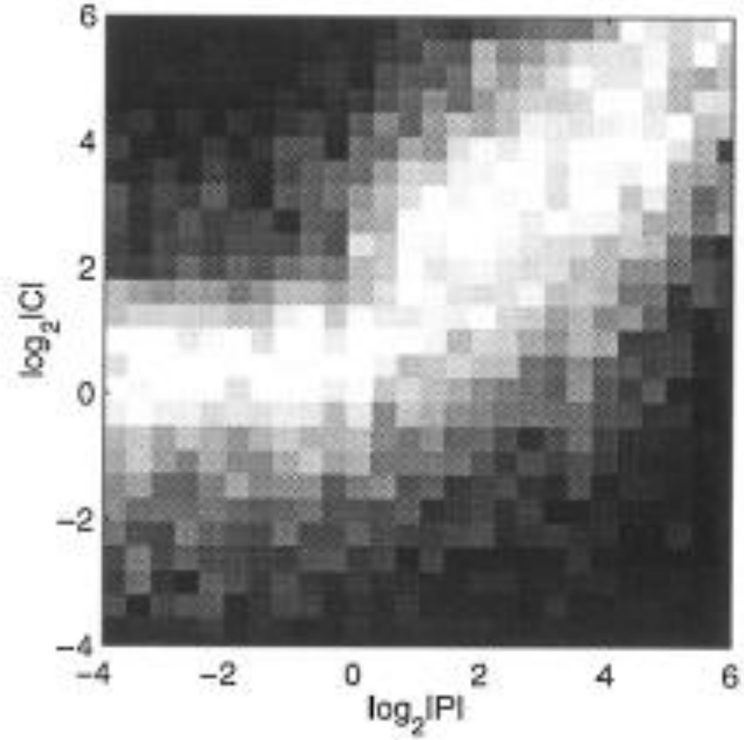
$$f(ch, cv) = e^{C_1 + C_2(|ch|+|cv|)^{\alpha}}$$

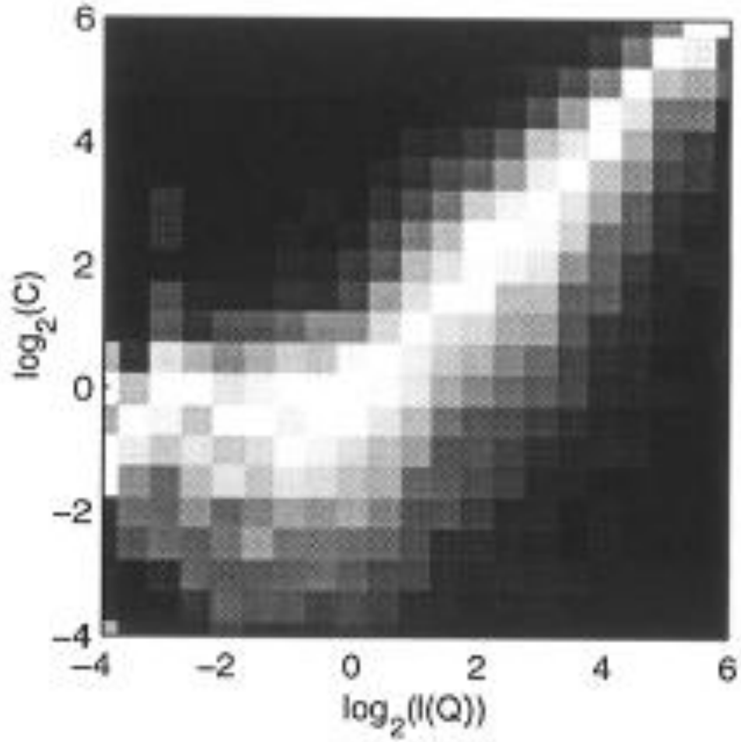
Joint statistics reveal that wavelet coefficients are NOT independent

Ref: Buccigrossi et al, 1997

Fig. 3. Coefficient magnitudes of a wavelet decomposition. Shown are absolute values of subband coefficients at three scales, and three orientations of a separable wavelet decomposition of the Einstein image. Also shown is the lowpass residual subband (upper left). Note that high-magnitude coefficients of the subbands tend to be located in the same (relative) spatial positions.

Large magnitude coefficients tend to occur at neighboring spatial locations within a sub-band, or at the same locations in sub-bands of adjacent scale/orientation



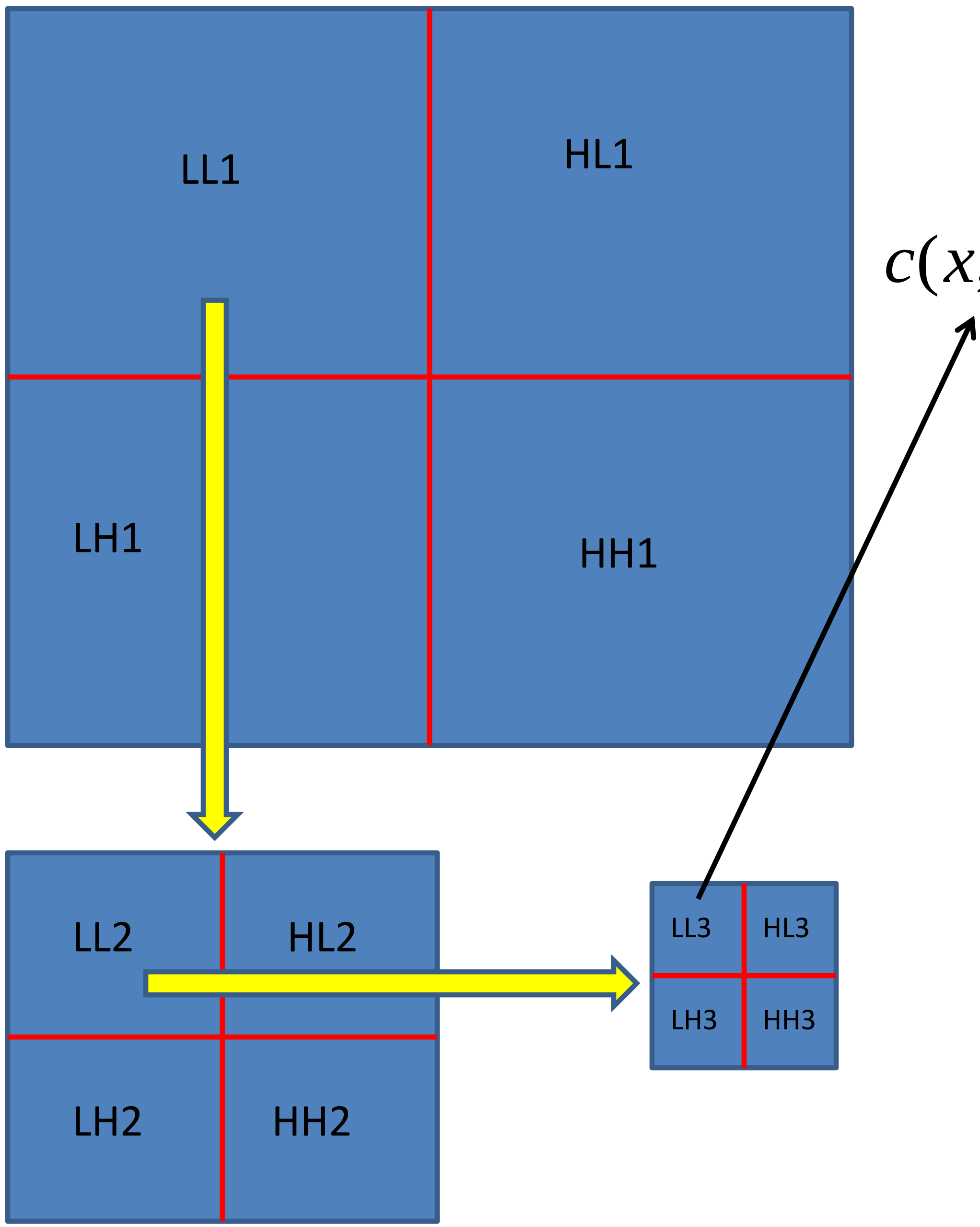


Joint histogram of logarithms of absolute value of child and parent coefficients (also true for coefficients at adjacent orientations) – features like prominent edges or textures have large magnitude coefficients at multiple scales

Joint histogram of logarithm of absolute value of child coefficient, and linear combination of logarithms of absolute values of coefficients from neighboring sub-bands

$$l(\vec{Q}) \equiv \vec{w} \cdot \vec{Q} = \sum_k w_k Q_k$$

$$\vec{w} = \mathcal{E}(\vec{Q}\vec{Q}^T)^{-1} \cdot \mathcal{E}(C' \cdot \vec{Q})$$



$$c(x, y) = \sum_{k=1}^{\#neighbors} w_k Q_k(x_k, y_k)$$

Wavelet coefficient from neighboring sub-band (at corresponding location)

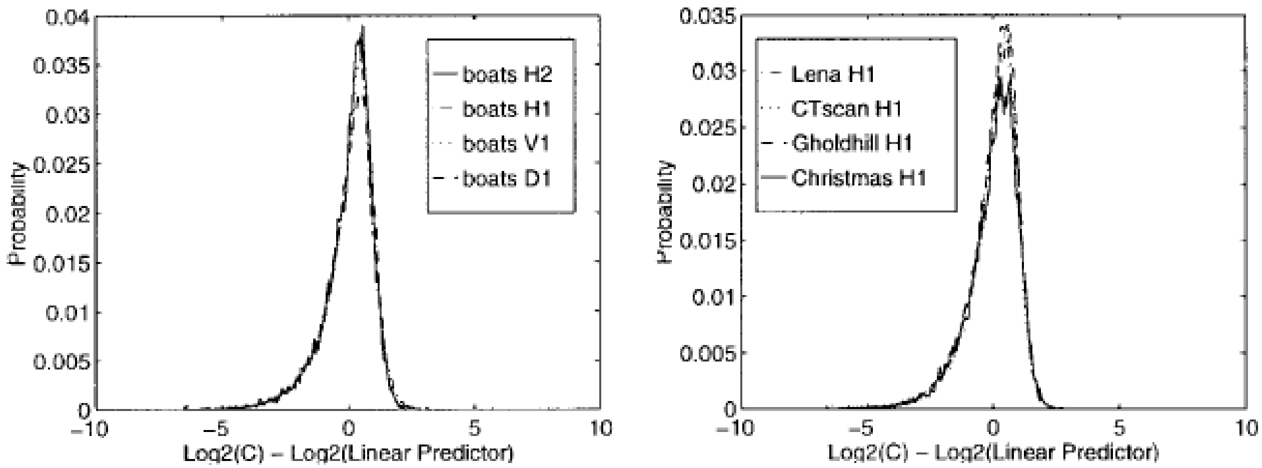


Fig. 6. Comparison of conditional distributions in the log domain of different subbands and images. Distributions were normalized (in the log domain) to have mean zero and variance one. Left: comparison of distributions for different subbands of the boats image. Right: comparison of distributions for different images (Lena, Goldhill, CT scan, Christmas).

Conditional distributions of a wavelet coefficient (log absolute) given linear combinations of its neighbors (log absolute) – the shapes are robust across images! The plots are mean and variance normalized.

This redundancy means that we need not store all the wavelet coefficients – we can just predict some coefficients directly given their neighbors using the linear model that was fit. **Useful for lossy image compression!**

The predictive model for a wavelet coefficient – given its neighbors – may be error-prone. But using this information does improve the compression quality IN PRACTICE!

$$H(p) = -\sum_{k=1}^{M} p_k \log(p_k)$$

$$= -E(\log(p))$$

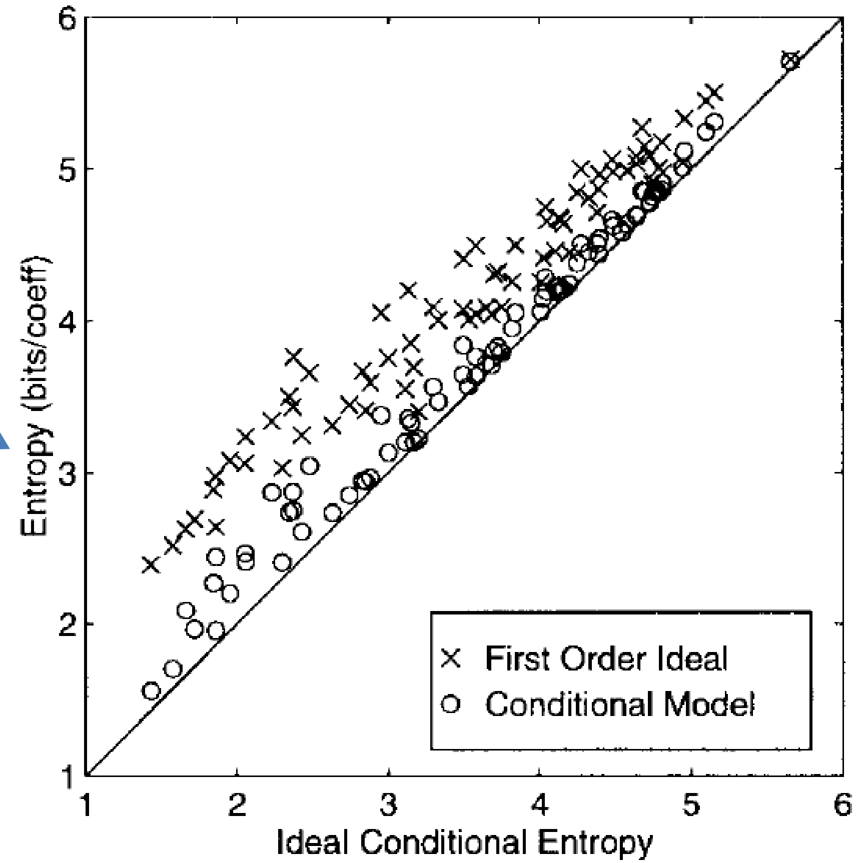


Fig. 8. Comparison of encoding cost using the conditional probability model of (5), and the encoding cost using the first-order histogram as a function of the encoding cost using a 256 × 256-bin joint histogram. Points are plotted for six bands (two scales, three orientations) of the 13 images in our sample set. The average relative entropy (Kullback–Leibler divergence) of the empirical marginal histogram is 0.548 b/coefficient, while the average relative entropy of the conditional model is 0.129 b/coefficient.

Summary

- Motivation for studying statistics of natural images – applications and Bayesian framework
- Statistics: power law, marginal and joint distributions of wavelet coefficients
- Implications for image compression