

Statistics of Natural Images and Models

by

Jinggang Huang

Sc.M., Brown University, 1997

Sc.M., University of Iowa, 1995

Sc.M., Nankai University, 1992

Sc.B., Nankai University, 1989

Thesis

Submitted in partial fulfillment of the requirements for
the Degree of Doctor of Philosophy
in the Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

May 2000

Abstract of “Statistics of Natural Images and Models,” by Jिंगgang Huang, Ph.D., Brown University, May 2000

We calculate several important statistics from two intensity image databases and a range image database. Mathematical models are fitted to these statistics, and some interesting features observed are explained. We examine the scale invariant property of images carefully, and find different scaling behaviors in different types of images. We also explain the well known ‘generalized Laplace’ distribution of mean-0 filter reactions under several simple assumptions, and propose an Ising-like model which duplicates some local statistics we observed in images. Finally, we show some applications using the statistics we calculate.

© Copyright 2000 Jinggang Huang

This dissertation by Jinggang Huang is accepted in its present form by the
Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of
Doctor of Philosophy

Date.....
David Mumford

Recommended to the Graduate Council

Date.....
Stuart Geman

Date.....
Donald M^cClure

Approved by the Graduate Council

Date.....

The Vita of Jinggang Huang

Jinggang Huang was born on April 28, 1967 in Tianjin, China. He entered Nankai University in 1985 where he received a B.S. in Mathematics in 1989 and a M.S. in Mathematics in 1992. In August of 1992 he came to the united States of America to continue his education at University of Iowa, Iowa. He graduated from Iowa University in May of 1995 with a M.S. in Mathematics. The same year, he was admitted to the Ph.D. program in Applied Mathematics at Brown University. He got a M.S. in Computer Science in 1997, and began his doctoral research under the supervision of Professor David Mumford, studying the statistics of natural images. He was elected a member of the Sigma Xi scientific society in 2000. Recent papers:

1. *Statistics of Natural Images and Models*, with D.Mumford in: Proc. IEEE Conference Computer Vision and Pattern Recognition, Fort Collins, CO,1999
2. *Statistics of Range Images*, with A.B.Lee and D.Mumford. Accepted by IEEE Conference Computer Vision and Pattern Recognition 2000.
3. *Random-Collage Model for Natural Images*, with A.B. Lee and D.Mumford. Submitted to International Journal of Computer Vision.

Acknowledgments

First and foremost, I would like to thank my advisor Professor David Mumford for his guidance. His insight and enthusiasm for the research were a continual inspiration to me. His patience and willingness to help went as far as spending many hours with me in programming and debugging. For that alone, I am deeply indebted.

Also, I wish to express my gratitude to Professor Ulf Grenander, Professor Stuart Geman and Professor Robert Pelcovits for stimulating discussions and valuable suggestions. I am grateful to Ann Lee for the collaboration in research. Part of the studies on range images are joint works with her. I would like to thank Daniel Potter who made the range image collecting project possible in the first place.

Thanks also go to Professor Stuart Geman and Professor Donald M^cClure who read this thesis and to Artur Fridman, Daniel Potter, Jin Shi and Tao Pang who helped in the development of the manuscript.

I would like to thank Hans van Hateren and Andy Wright for providing the intensity image databases and answering questions we had about them.

Last, but not least, I would like to thank my parents for their support and my wife Minhong Ji for her years of love, encouragement and sacrifice.

Contents

| | |
|--|------------|
| Acknowledgments | iii |
| 1 Introduction | 1 |
| 1.1 Overview | 2 |
| 1.2 Databases | 3 |
| 1.3 Symbols and some remarks | 4 |
| 1.4 Scale Invariance Property of Images | 7 |
| 2 Empirical Study | 10 |
| 2.1 Single Pixel Statistics | 11 |
| 2.1.1 Single Pixel Statistics of intensity image | 11 |
| 2.1.2 Single Pixel Statistics of different categories | 12 |
| 2.1.3 Single Pixel Statistics of Range Images | 14 |
| 2.2 Derivative Statistics | 18 |
| 2.2.1 Derivative Statistics For Intensity Images | 18 |
| 2.2.2 Derivative Statistics For Different Categories of Intensity Images | 21 |
| 2.2.3 Derivative Statistics for Range Images | 24 |
| 2.3 Two-point Statistics | 29 |
| 2.3.1 Intensity Images | 29 |
| 2.3.2 Range Images | 31 |
| 2.3.3 Covariance of nature images | 36 |
| 2.4 Joint Statistics of Haar Filters Reactions | 40 |
| 2.4.1 Haar Filters | 40 |

| | | |
|----------|---|------------|
| 2.4.2 | 2D Joint Statistics of Intensity Images | 41 |
| 2.4.3 | 2D Joint Statistics of Sowerby Images | 49 |
| 2.4.4 | 3D Joint Statistics for Haar Wavelet Coefficients | 56 |
| 2.4.5 | Statistics for Haar Wavelet from range images | 58 |
| 2.5 | Joint Statistics for Other Wavelet Coefficients | 64 |
| 2.5.1 | Joint Statistics for QMF | 64 |
| 2.5.2 | Joint Statistics for Steerable Pyramids | 64 |
| 2.5.3 | An interesting phenomenon | 66 |
| 2.6 | DCT Coefficients | 70 |
| 2.7 | Random Filters | 74 |
| 2.8 | 2X2 blocks | 77 |
| 2.9 | 8X8 patches | 80 |
| 3 | Simulating local statistics with a generalized Ising model | 84 |
| 3.1 | Basic Setup | 85 |
| 3.2 | Object generating process | 92 |
| 3.3 | Simulation of the Derivative Statistics | 97 |
| 3.4 | Simulation of the Haar Wavelet Coefficients | 98 |
| 4 | Applications | 102 |
| 4.1 | Image Compression | 103 |
| 4.2 | Noise removal | 105 |
| 4.3 | Texture Classification | 107 |
| 5 | Conclusion | 110 |

List of Tables

| | | |
|-----|---|----|
| 1.1 | Categories and Their Weights | 5 |
| 2.1 | Some constants associated to single pixel statistics from van Hateran database. The last column is the KL distance between the single pixel statistics and the normal distribution with the same variance. | 11 |
| 2.2 | Some constants associated to single pixel statistics of different categories. The last column is the KL distance between the single pixel statistics and the normal distribution with the same variance. | 13 |
| 2.3 | constants associated to the derivatives statistics at different scales , van Hateran Database. α_1, s_1 are the constants in the model 2.1 , calculated directly by minimizing the KL distance between empirical distribution and the model, α_2, s_2 the are the constants calculated from 2.2. p_0 is the empirical density function; p_1 and p_2 are the model with constants (α_1, s_1) and (α_2, s_2) respectively | 19 |
| 2.4 | constants associated to the derivatives statistics at different scales , Sowerby Database. α_1, s_1 are the constants in the model 2.1 , calculated by the optimal method. α_2, s_2 the are the constants calculated by direct parameter estimation. p_0 is the empirical density function; p_1 and p_2 are the models with constants (α_1, s_1) and (α_2, s_2) respectively | 19 |
| 2.5 | Some constants associated to derivative statistics of different categories. . . | 24 |
| 2.6 | constants associated to the derivative statistics of range images at different scales(Scaling down by block minimum). | 24 |
| 2.7 | Some interesting features in figure 2.25 and their explanation | 43 |

| | | |
|------|---|----|
| 2.8 | Some constants for different pairs. a_{11}, a_{12}, a_{22} are the entries of the covariance matrix. The last column is the KL distance between model 2.4 and the data. | 44 |
| 2.9 | Constants for the fits of the maximum entropy model to empirical distribution after 12 iterations. The second shows the KL distance between the 2D empirical distribution and the model, the last 4 columns shows the KL distance between the marginal of the empirical distribution and that of the model. | 48 |
| 2.10 | Some constants for different Haar Wavelet coefficient pairs calculated from Sowerby database. a_{11}, a_{12}, a_{22} are the entries of the covariance matrix. | 50 |
| 2.11 | constants of DCT coefficients(part 1) | 71 |
| 2.12 | constants of DCT coefficients(part 2) | 72 |
| 2.13 | constants of random mean-0 filter coefficients | 75 |
| 2.14 | constants of binary random mean-0 filter reactions | 76 |
| 3.1 | Definition of the potential function. a, b, c, d are entries of a 2×2 block from left to right and up to down. | 95 |

List of Figures

| | | |
|-----|---|----|
| 1.1 | Four images from the van Hateren data base | 4 |
| 1.2 | Two images from Sowerby Image Data Base and their segmentations. | 8 |
| 1.3 | A sample image from our range image data base | 8 |
| 1.4 | A laser range finder centered at O and two homothetic triangles, $\triangle ABC$ and $\triangle A'B'C'$ | 9 |
| 1.5 | Two ways to generate images of half dimensions. Left: By taking the center part of the original. Right: By taking disjoint $2X2$ block averaging | 9 |
| 2.1 | $\log(pdf)$ of single pixel intensity. Left: van Hateren Database; Right: Sowerby Database. Red, Green, Blue, Black and Yellow for scale 1,2,4,8 and 16 respectively | 12 |
| 2.2 | Comparison of single pixel statistics calculated from two databases. Red: van Hateren data base. Black:Sowerby data base | 12 |
| 2.3 | $\log(pdf)$ of single pixels statistics at different scales and categories: red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8 | 14 |
| 2.4 | Semi-log (left) and log-log (right) plots of the single pixel statistics (i.e. range statistics). <i>Solid</i> : distribution for the whole image. <i>Dotted</i> : distribution for the bottom half of a range image. <i>Dashed</i> : distribution for the top half of a range image. | 15 |
| 2.5 | Top view of a randomly generated forest scene. See text. | 17 |
| 2.6 | Ground model. See text. | 17 |
| 2.7 | Compare the derivative statistics from two databases. Red: van Hateren, Black: Sowerby | 18 |

| | | |
|------|--|----|
| 2.8 | Derivative statistics at different scales. Upper: van Hateren database, Lower: Sowerby database. Red, Green, Blue, Black and Yellow for scales 1,2,4, 8,16 respectively | 20 |
| 2.9 | The fit generalize Laplace model to derivative statistics at scale 2, van Hateren Database, Blue: Data, Red: Model | 21 |
| 2.10 | $\log(pdf)$ of D at different scales and categories: red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8 | 22 |
| 2.11 | Plot of \log_2 (standard deviation) vs. scale level and its linear fit in five cases: + = sky, * = vegetation categories, o = road, x = man-made categories and Δ = all pixels. The slopes are 0.15 for man-made, 0.02 for all, -0.11 for vegetation, -0.30 for road and -0.50 for sky. | 25 |
| 2.12 | $\log(pdf)$ of D at different scales and categories. Renormalized according to the standard deviations. red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8 | 26 |
| 2.13 | Derivative statistics for range images at different scales(Scaling down by block minimum). red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8 . | 27 |
| 2.14 | Left: $\log(pdf)$ of f_2 . Right: the fit of equation 2.3, red = $\log(f_0)$, blue = $\log((1 - \lambda) \cdot f_2)$ and green = $\log(\lambda \cdot f_1)$, and $\lambda = 0.988$ | 28 |
| 2.15 | Left figure: Joint Histogram of two adjacent pixels p_1 and p_2 , calculated from van Hateren database. Right figure: The product density function of $p_1 - p_2$ and $p_1 + p_2$ | 29 |
| 2.16 | Contour plots of the log histograms of pixel pairs for intensity images (left column) and the best bivariate fit to the random collage model (right column). x : distance between the two pixels, u : sum of the pixel values, v : difference of the two pixel values. | 32 |
| 2.17 | The values for λ and the 1D functions g_x, h_x, q_x in the best bivariate fit to the bivariate statistics of intensity images at distances $x = 1$ (yellow), 2 (magenta), 4 (cyan),8 (red), 16(green) 32(blue) and 128(dotted red). | 33 |

| | | |
|------|--|----|
| 2.18 | Contour plots of the log histograms of pixel pairs for range images (left column) and the best bivariate fit to the random collage model (right column). x : distance between the two pixels, u : sum of the pixel values, v : difference of the two pixel values. | 34 |
| 2.19 | The values for λ and the 1D functions g_x, h_x, q_x in the best bivariate fit to the bivariate statistics of range images at distances $x = 1$ (yellow), 2 (magenta), 4 (cyan), 8 (red), 16 (green) 32 (blue) and 128 (dotted red). | 35 |
| 2.20 | Difference Function | 37 |
| 2.21 | log-log plot of the horizontal cross section and some fitting constants, see text | 39 |
| 2.22 | Haar Filters | 40 |
| 2.23 | Relations Between Wavelet Coefficients | 41 |
| 2.24 | Mesh Plot of the $\ln(\text{Joint Histogram})$ of Horizontal Component and its Vertical Cousin | 42 |
| 2.25 | Joint Histogram of Haar Wavelet Coefficients calculated from van Hateran Database. Red: the level curves of the empirical distribution. Blue: the model 2.4. Ch, Cv, Cd stand for horizontal, vertical and diagonal components respectively; L stands for left brother; Ph, Pd for horizontal parent and diagonal parent; Dl for upper-left brother. | 51 |
| 2.26 | Maximum entropy model with the same marginals along axes and diagonal lines | 52 |
| 2.27 | Function f_i 's | 52 |
| 2.28 | Joint Histogram of Haar Wavelet Coefficients calculated from Sowerby Images | 53 |
| 2.29 | Joint Histogram of Haar Wavelet Coefficients calculated from the 'sky' category of Sowerby Images | 53 |
| 2.30 | Joint Histogram of Haar Wavelet Coefficients calculated from the 'vegetation' category of Sowerby Images | 54 |
| 2.31 | Joint Histogram of Haar Wavelet Coefficients calculated from the 'road' category of Sowerby Images | 54 |
| 2.32 | Joint Histogram of Haar Wavelet Coefficients calculated from the 'man made' category of Sowerby Images | 55 |

| | | |
|------|--|----|
| 2.33 | An equi-surface of 3d joint histogram of horizontal component, its vertical cousin and diagonal cousin, viewed from three different angles | 57 |
| 2.34 | Contour plots of the log(pdf) of wavelet coefficient pairs for range images. . | 58 |
| 2.35 | Contour plots of the log histogram of wavelet coefficient pairs, calculated from range images scaled down by taking the minimum of 2×2 blocks. . . . | 59 |
| 2.36 | Contour plots of the log histogram of wavelet coefficient pairs, calculated from range images scaled down by taking the average of 2×2 blocks. . . . | 60 |
| 2.37 | Contour plots of the log histogram of wavelet coefficient pairs, calculated from intensity images scaled down by taking the max or min, whichever closer to the mean | 61 |
| 2.38 | An equi-surface of a 3D joint histogram of horizontal, vertical and diagonal wavelet coefficients in range images, viewed from three different angles. . . . | 63 |
| 2.39 | QMF Filters(9-tap) | 65 |
| 2.40 | Joint Histogram of QMF Wavelet Coefficients calculated from natural images | 66 |
| 2.41 | 2-orientation steerable pyramid filters | 67 |
| 2.42 | Joint Histogram of Steerable Pyramid Coefficients calculated from natural images | 67 |
| 2.43 | The conditional histogram $H(\log hc \mid \log hp)$ calculated from our model. Bright parts mean high values | 69 |
| 2.44 | Joint Histogram of QMF Wavelet Coefficients calculated from natural images | 69 |
| 2.45 | log(pdf) of DCT coefficients | 73 |
| 2.46 | contour plot of the standard derivations of DCT coefficients | 74 |
| 2.47 | The empirical pdf of some random mean-0 filter reactions | 75 |
| 2.48 | The empirical pdf of binary random mean-0 filter reactions | 76 |
| 2.49 | k-means centers of 8×8 images patches. The number is the frequency that images patches are closest to the center. | 81 |
| 2.50 | Principal components around centers. The upper-left subplot is the center, and the others are the eigenvectors of the covariance matrix around the center, See text | 83 |

| | | |
|------|---|-----|
| 2.51 | Percentage of samples fall in percentage of the total volume. Bottom curve: assuming disk like regions, top curve: assuming elliptic regions. See Text . . . | 83 |
| 3.1 | The pdf of ‘normalized intensities’(see text). Left: for 2×2 blocks. Right: for 4×4 blocks | 87 |
| 3.2 | The distribution of the color factor η . Left: van Hateren. Right: Sowerby . . . | 91 |
| 3.3 | Simulation of the derivative statistics by assuming Gaussian object factor, Red: statistics from natural images. Blue: statistics from simulation. Left: simulation of the van Hateren database. Right: simulation of the Sowerby database. | 92 |
| 3.4 | Simulation of the derivative statistics by assuming Gaussian object factor. Red: statistics from natural images. Blue: statistics from simulation. From left to right and up to down: sky, vegetation, road, man-made | 93 |
| 3.5 | An vertical edge and an diagonal | 94 |
| 3.6 | A sample from our MRF at $T = 1.88$ | 96 |
| 3.7 | Left figure: Expectation of the energy of samples v.s. temperature. Right figure: Derivative of the expectation of the energy. The vertical dotted line is $T = 1.88$, the critical temperature. | 96 |
| 3.8 | The derivative statistics. Blue: from natural images. Red: simulation at scale 1($pixel = \mu(8 \times 8 subpixels)$). Green:simulation at scale 2($pixel = \mu(16 \times 16 subpixels)$). | 99 |
| 3.9 | Left: Variance of the derivative statistics. Right: KL distance between the derivative statistics calculated from natural images and that from model. Blue: from natural images. Red: simulation at scale 1($pixel = \mu(8 \times 8 subpixels)$). Green:simulation at scale 2($pixel = \mu(16 \times 16 subpixels)$). | 100 |
| 3.10 | The distribution of the object factor(for derivative filter) at the critical temperature and two scales. Red: scale 1. Blue: scale 2 | 100 |
| 3.11 | Haar Wavelet Coefficients of the Simulations at Temperature $T=1.88$ | 101 |
| 4.1 | Noise removal experiments. From up to down and left to right are the original image, the noisy image and the cleaned image | 106 |

| | | |
|-----|---|-----|
| 4.2 | The 70 textures we used in our experiment | 107 |
| 4.3 | three examples of texture classification results. The patch at the upper left corner is the picked patch, the others are the patches which are most ‘similar’ to the picked patch according to our algorithm. | 109 |

Chapter 1

Introduction

1.1 Overview

Knowledge about statistics of natural images is crucial in many applications such as image compression, image denoising and target recognition. They are also important in understanding the biological vision system.

Due to the complex and specific structures in images, simple parametric statistics models such as Gaussian Models, fail to describe natural images accurately. As was remarked by Rosenfeld in the 80's, the "noise" in images is more often "clutter", and the statistics of clutter are not Gaussian at all.

There has been much attention recently to the statistics of natural images. For example, Field [7] discussed the scale invariance property of natural images and linked the design of the biological vision system to the statistics of natural images, Ruderman [17] proposed a model to explain why images are scale invariant. Zhu *et al.* [23] set up a general frame work for natural image modeling via exponential models. Buccigrossi *et al.* [3] and Wainwright *et al.* [22] uncovered significant dependencies of wavelet coefficients in natural image statistics.

Many of these papers base their calculation on a small set of images, casting doubt on how robust their results are. Also, because of the small sample sets, rare events (e.g. strong contrast edges) which are important visually may not show up frequently enough to stabilize the corresponding statistics. We tried to overcome these problems by using large natural image databases and systematically investigating the statistics underlying *all* images.

We also studied the statistics of range images, which by themselves are important in applications, for example they give priors for stereo algorithms ([1],[2]). We are interested in them because they lead to a direct understanding of the stochastic nature of the geometry of the world and make it possible to construct more realistic models of intensity images. For example, authors in [17], [4] and [14] have modeled intensity images as a perspective view of the 3D world, with objects of random geometry (size, shape, position) and intensity. The object geometries in these models are usually based on assumptions which have not been directly verified in real scenes. There is no doubt that with a fairly large data base of range images, we will better understand the scene geometry of the 3D world, and thus be able to develop more realistic models for intensity images.

In the remaining part of this chapter, we will introduce the databases from which we

calculated all the statistics and some notations we will use later. Chapter 2 gives detailed results of the statistics we calculated and mathematical models for some of them. In chapter 3, we discuss an ‘Ising-like’ Markov Random field model which simulates some of the statistical properties. In Chapter 4, we show some simple applications of the statistics we observed.

1.2 Databases

We calculated all our statistics from three image databases.

1. **Van Hateren** image database. This image database contains 4000 calibrated images, each image has 1024×1536 pixels, and each pixel is 12-bits deep. Figure 1.1 shows some sample images from this data base. These images are calibrated, being proportional to light energy received from the world up to an unknown multiplicative constant. More details about this database can be found in [10].
2. **Sowerby** image database. This database was provided by British Aerospace. There are 214 RGB images in this database, with 512×768 pixels each. Each pixel is 8 bits deep. Since we only work with grey levels, we took the weighted average of the three components, with widely used weights: 0.299, 0.587, 0.114. The images have been laboriously segmented into pixels representing 11 different categories of the scene. Table 1.1 shows the categories and the frequencies(in percentage) of each of them. Figure 1.2 shows two images and their segmentations.

We found that the ‘manmade’ categories 6,8,9 and 11 have similar histograms and scaling behavior, so we put such categories together. Likewise, the ‘vegetation’ categories 2 and 7 behave similarly (note that category 7 contains vegetation like hedges, as well as fences, etc.). Together with the ‘sky’(category 1) and ‘road’ (category 4) we have 4 larger categories from which we will calculate different statistics.

3. **Range** image database. We have collected 205 panoramic range images from varying environments (both outdoor and indoor scenes) — 54 of these images were taken in different forests in Rhode Island and Massachusetts during August-September. In



Figure 1.1: Four images from the van Hateren data base

this paper, we will focus on the forest scenes, because the statistics of these images appear more stable than those of other categories (for example, residential and interior scenes). Figure 1.2 shows a sample. We used a laser range-finder with a rotating mirror¹ to collect the range images. Each image contains 444×1440 measurements with an angular separation of 0.18 deg. The field of view is thus 80^0 vertically and 259^0 horizontally. Each measurement is calculated from the time of flight of the laser beam. The operational range of the sensor is typically 2-200m. The laser wavelength of the range-finder is $0.9\mu m$, which is in the near infra-red region.

1.3 Symbols and some remarks

In this section, we explain some symbols we will frequently use in the paper:

¹3D imaging sensor LMS-Z210 by Riegl

| Category | Frequency | Description |
|----------|-----------|--------------------------------|
| 1 | 10.87 | sky, cloud, mist |
| 2 | 37.62 | trees, grass, bush, soil, etc. |
| 3 | 0.20 | road surface marking |
| 4 | 36.98 | road surface |
| 5 | 6.59 | road border |
| 6 | 3.91 | building |
| 7 | 2.27 | bounding object |
| 8 | 0.11 | road sign |
| 9 | 0.28 | telegraph pole |
| 10 | 0.53 | shadow |
| 11 | 0.64 | car |

Table 1.1: Categories and Their Weights

| | |
|---|--|
| μ | mean |
| σ | standard deviation |
| \mathcal{S} | skewness |
| κ | kurtosis |
| $\mathcal{H}(\mathcal{X})$ | differential entropy(in bits) of random variable X |
| $\mathcal{I}(\mathcal{X}, \mathcal{Y})$ | mutual information between X and Y |
| $D(p \parallel q)$ | Kullback Leibler distance between two probability mass functions p and q . |

The skewness and kurtosis of a random variable X are defined below.

$$\kappa = \frac{E(X - \mu)^4}{\sigma^4} \quad \mathcal{S} = \frac{E(X - \mu)^3}{\sigma^3}$$

Since the kurtosis is very sensitive to outliers, direct calculation of this statistic would result in unstable results. To solve this problem, we will only calculate the ‘trimmed’ kurtosis, i.e. we will discard the probability density outside $[-14\sigma, 14\sigma]$ (where σ is the standard deviation of the variable) before we calculate statistics. For the definition of other symbols, see [5].

Most our figures of probability distributions (or of normalized histograms) will be shown with the vertical scale not probability but *log of probability*: this is very important as it

shows the non-Gaussian nature of these probability distributions more clearly and shows especially the nature of the tails.

We will regard each image I in the data set as a $M \times N$ matrix, and $I(i, j)$ represents the intensity or range at pixel (i, j) . As mentioned in the previous section, all intensity images measure the light in the world up to an unknown multiplicative constant, i.e. if we assume the physical intensity of a scene is I_0 , then the image of the scene is $I = C_I I_0$, where C_I is some unknown constant, different from image to image. In order to avoid such constants from entering our statistics, we work on the *log contrast* of the images, which is $\log(I(i, j)) - \text{mean}(\log(I))$. Notice that the multiplicative constant has been cancelled out in the log contrast. In the future, without specification, a image I always means the **log-contrast** of the original image in our databases.

For range images, we also work with $\log(\text{range})$ instead of range directly, because the former statistic is closer to being shape invariant. Figure 2 shows a top view of a laser range-finder (see circle) centered at O , and two homothetic triangles $\triangle ABC$ and $\triangle A'B'C'$ (P_A, P_B and P_C correspond to three pixels in the range image). Assume that the distances between O and the vertices of $\triangle ABC$ are r_A, r_B and r_C , respectively, and the distances between O and the vertices of $\triangle A'B'C'$ are R_A, R_B and R_C , respectively. Let

$$D = \text{range}(P_A) - \text{range}(P_B)$$

be the difference in range for pixels P_A and P_B , and let

$$\begin{aligned} \hat{D} &= \log(\text{range}(P_A)) - \log(\text{range}(P_B)) \\ &= \log\left(\frac{\text{range}(P_A)}{\text{range}(P_B)}\right) \end{aligned}$$

be the difference in *log* range for the same two pixels. Then, a scene with $\triangle ABC$ and a scene with triangle $\triangle A'B'C'$ will lead to different values for D ($r_A - r_B$ vs. $R_A - R_B$) but the *same* value for \hat{D} ($\log(\frac{r_A}{r_B}) = \log(\frac{R_A}{R_B})$). Hence, $\log(\text{range})$ is appropriate if we want the difference statistics (or any mean-0-filter reaction) to be “shape invariant”.

1.4 Scale Invariance Property of Images

Scale invariance is a striking property of natural images. We give a simple explanation of the meaning of this property for digital images.

Suppose there is a random variable Ψ_0 generating $N \times N$ *natural images* (like those in our databases) for us. We can assume that Ψ_0 has a density function $f_{\Psi_0}(\mathbf{x})$ on $R^{N \times N}$. For each $N \times N$ image, there are two natural but quite different ways to get $\frac{N}{2} \times \frac{N}{2}$ images as shown in Figure 1.5, one is to take the center $\frac{N}{2} \times \frac{N}{2}$ patch of the $N \times N$ image, the other is to scale the original image down to an image of size $\frac{N}{2} \times \frac{N}{2}$ by taking disjoint 2×2 block average. The two methods correspond to two projections

$$\Pi_1 : R^{N \times N} \rightarrow R^{\frac{N}{2} \times \frac{N}{2}}$$

$$\Pi_2 : R^{N \times N} \rightarrow R^{\frac{N}{2} \times \frac{N}{2}}$$

Then we have two random variables defined on $R^{\frac{N}{2} \times \frac{N}{2}}$,

$$\Psi_1 = \Pi_1 \cdot \Psi_0$$

$$\Psi_2 = \Pi_2 \cdot \Psi_0$$

and each has an induced density function on $R^{\frac{N}{2} \times \frac{N}{2}}$, denoted by f_{Ψ_1} and f_{Ψ_2} respectively. We call Ψ_0 is scale invariant if $f_{\Psi_1} = f_{\Psi_2}$.

In practice, it is impossible to calculate the high dimensional distributions f_{Ψ_1} and f_{Ψ_2} directly. To verify the scale invariant property, we usually check to see whether some marginal distributions of f_{Ψ_1} and f_{Ψ_2} (e.g. distribution of linear filter reactions) match. Also if we assume that the image generating process is stationary, then marginal distributions calculated from f_{Ψ_1} should be the same as those calculated from f_{Ψ_0} . So in the future we verify the scale invariant property by checking marginal distributions of f_{Ψ_0} and f_{Ψ_2} . We will use $I^{(k)}$ to denote the $k \times k$ disjoint block average of I .

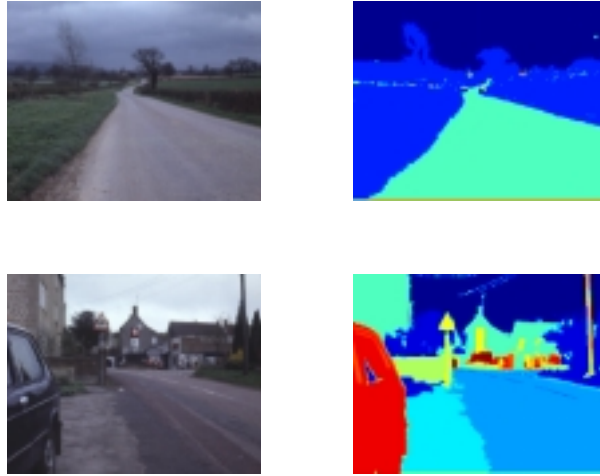


Figure 1.2: Two images from Sowerby Image Data Base and their segmentations.

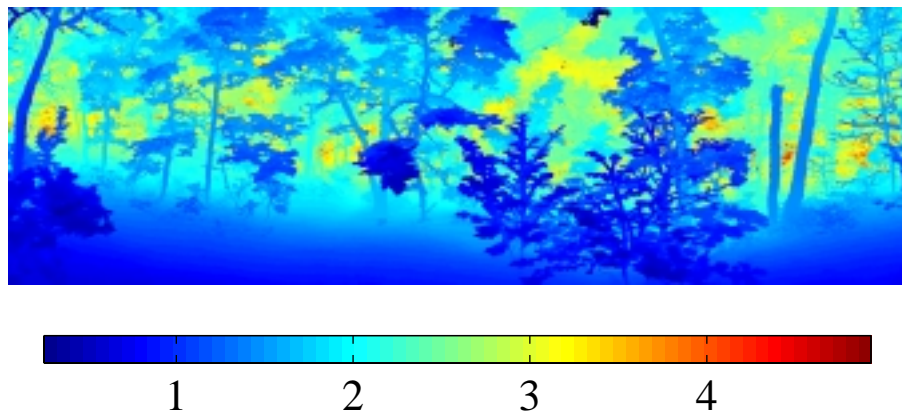


Figure 1.3: A sample image from our range image data base

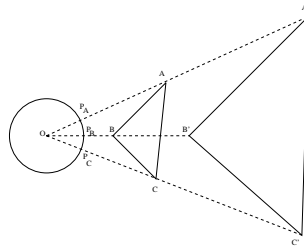


Figure 1.4: A laser range finder centered at O and two homothetic triangles, $\triangle ABC$ and $\triangle A'B'C'$.

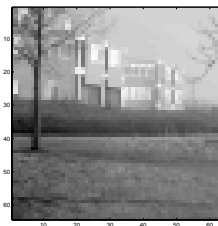
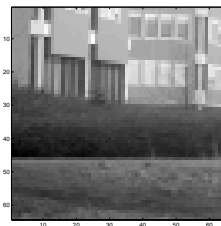
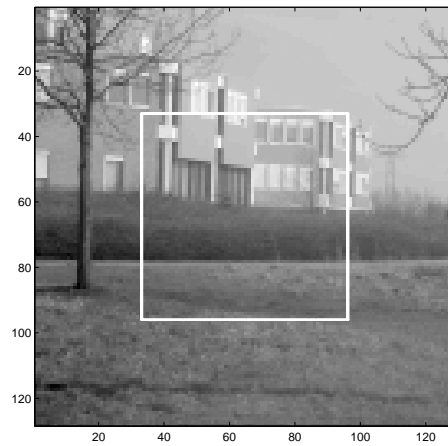


Figure 1.5: Two ways to generate images of half dimensions. Left: By taking the center part of the original. Right: By taking disjoint 2×2 block averaging

Chapter 2

Empirical Study

2.1 Single Pixel Statistics

2.1.1 Single Pixel Statistics of intensity image

The red curve in figure 2.1 shows the $\log(\text{empirical pdf})$ of the single pixel statistics (log contrast) calculated from images of the two intensity image database. We can see that this statistic is slight asymmetric. One important reason is the presence of a portion of sky in many images, which is quite different from other parts of images, always with a high intensity value. Another interesting feature is the linear portion in both of the $\log(\text{pdf})$. Obviously, this statistic is not Gaussian.

Figure 2.1 also shows this statistic at different scales and table 2.1 gives some constants associated to them. We can see single pixel intensity is roughly scale invariant. The last field in the table is the KL distance between the empirical distribution and the Gaussian distribution with the same standard deviation. From point view of information theory, the single pixel intensity is roughly Gaussian. This fact can also be seen from the small kurtosis 4.6-4.9(3 for Gaussian distribution) and the parabola shape around 0 in Figure 2.1. We should point out that this statistic is not very stable and may be quite different in different databases. In figure 2.2 we super imposed the $\log(\text{pdf})$ of single pixel statistics calculated from the two intensity images to show the difference.

| $Scale$ | σ | \mathcal{S} | κ | \mathcal{H} | $D(p \parallel n_p)$ |
|---------|----------|---------------|----------|---------------|----------------------|
| 1 | 0.788 | 0.218 | 4.56 | 1.66 | 0.0469 |
| 2 | 0.767 | 0.253 | 4.49 | 1.62 | 0.0489 |
| 4 | 0.744 | 0.253 | 4.55 | 1.57 | 0.0541 |
| 8 | 0.719 | 0.252 | 4.67 | 1.51 | 0.0628 |
| 16 | 0.693 | 0.256 | 4.83 | 1.44 | 0.0746 |

Table 2.1: Some constants associated to single pixel statistics from van Hateran database. The last column is the KL distance between the single pixel statistics and the normal distribution with the same variance.

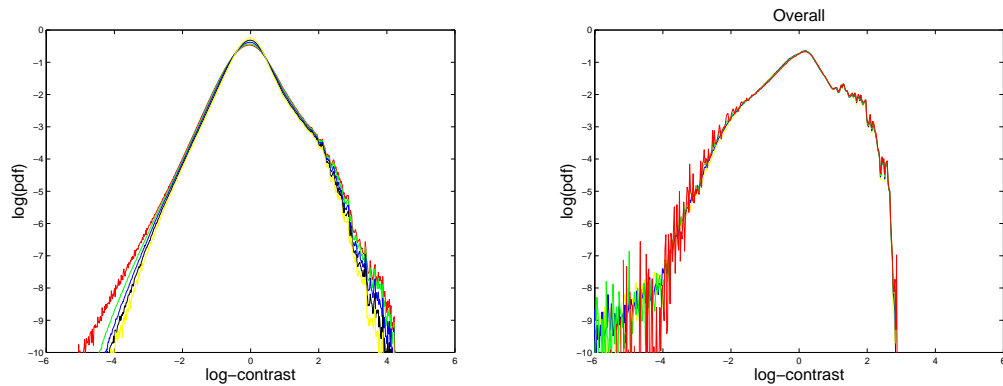


Figure 2.1: $\log(pdf)$ of single pixel intensity. Left: van Hateren Database; Right: Sowerby Database. Red, Green, Blue, Black and Yellow for scale 1,2,4,8 and 16 respectively

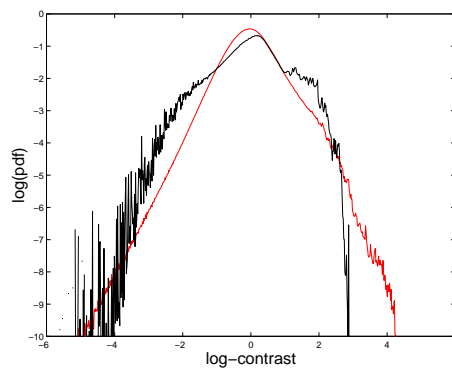


Figure 2.2: Comparison of single pixel statistics calculated from two databases. Red: van Hateren data base. Black: Sowerby data base

2.1.2 Single Pixel Statistics of different categories

As mentioned in the introduction, images in the Sowerby database are segmented, this allows us to make a study of statistics of different categories. Figure 2.3 shows the distributions of the single pixel statistics of different categories at different scales and table 2.2 shows some constants associated to the distributions. For each category, the single pixel statistics is roughly scale invariant.

| <i>Category</i> | <i>Scale</i> | σ | \mathcal{S} | κ | \mathcal{H} | $D(p \parallel n_p)$ |
|-----------------|--------------|----------|---------------|----------|---------------|----------------------|
| Sky | 1 | 0.887 | -0.686 | 5.33 | 1.78 | 0.0979 |
| | 2 | 0.867 | -0.586 | 4.93 | 1.75 | 0.0953 |
| | 4 | 0.855 | -0.528 | 4.77 | 1.72 | 0.0964 |
| | 8 | 0.841 | -0.458 | 4.58 | 1.7 | 0.0983 |
| Vegetation | 1 | 1.89 | -0.0459 | 3.28 | 2.93 | 0.0265 |
| | 2 | 1.79 | -0.132 | 3.35 | 2.87 | 0.0195 |
| | 4 | 1.69 | -0.222 | 3.45 | 2.79 | 0.022 |
| | 8 | 1.61 | -0.307 | 3.55 | 2.7 | 0.0279 |
| Road | 1 | 1.03 | 0.24 | 3.47 | 2.07 | 0.0213 |
| | 2 | 1.01 | 0.315 | 3.39 | 2.03 | 0.0244 |
| | 4 | 0.989 | 0.37 | 3.32 | 2 | 0.0279 |
| | 8 | 0.976 | 0.41 | 3.29 | 1.98 | 0.0321 |
| Man-made | 1 | 1.9 | -0.346 | 2.97 | 2.94 | 0.0373 |
| | 2 | 1.84 | -0.373 | 3.08 | 2.9 | 0.0307 |
| | 4 | 1.77 | -0.391 | 3.23 | 2.84 | 0.0294 |
| | 8 | 1.67 | -0.39 | 3.38 | 2.75 | 0.0356 |
| Overall | 1 | 2.07 | -0.244 | 3.21 | 3.05 | 0.0485 |
| | 2 | 2.04 | -0.225 | 3.24 | 3.03 | 0.0447 |
| | 4 | 2.01 | -0.2 | 3.26 | 3.01 | 0.0421 |
| | 8 | 1.98 | -0.173 | 3.28 | 2.99 | 0.0397 |

Table 2.2: Some constants associated to single pixel statistics of different categories. The last column is the KL distance between the single pixel statistics and the normal distribution with the same variance.

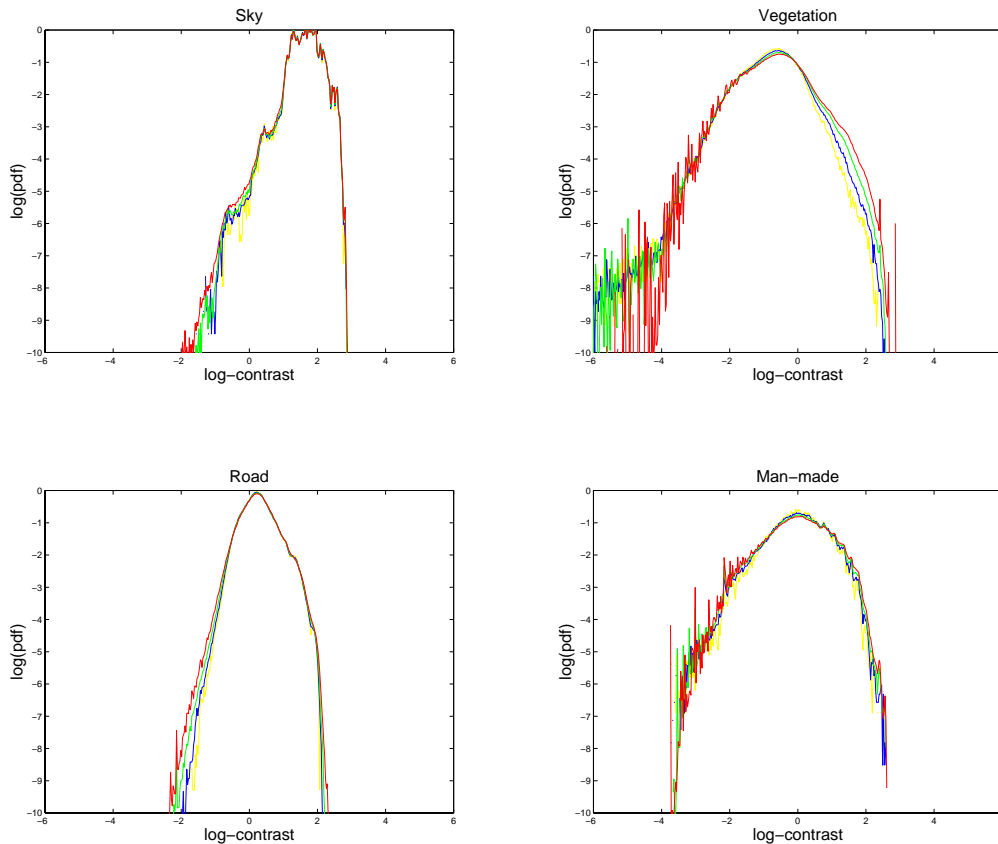


Figure 2.3: $\log(pdf)$ of single pixels statistics at different scales and categories: red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8

2.1.3 Single Pixel Statistics of Range Images

The solid line in Figure 2.4 shows the single-pixel statistics of $\log(range)$ images. We observe a sudden change in slope in the log-log plot at a range of about 20 meters (or $\log(range) \approx 3$; see vertical line in figure) — this may be related to the accumulation of occlusion effects. In Figure 2.4, we have also plotted the $\log(range)$ histograms for the top half (dashed line) and bottom half (dotted line) of a range image separately. The two halves correspond to different distributions of objects — mainly ground for the bottom half and mainly trees for the top half — and display quite different statistics. The distribution from the top half has an approximately linear tail in a semi-log plot (indicating an exponential decay $\sim e^{-0.12r}$),

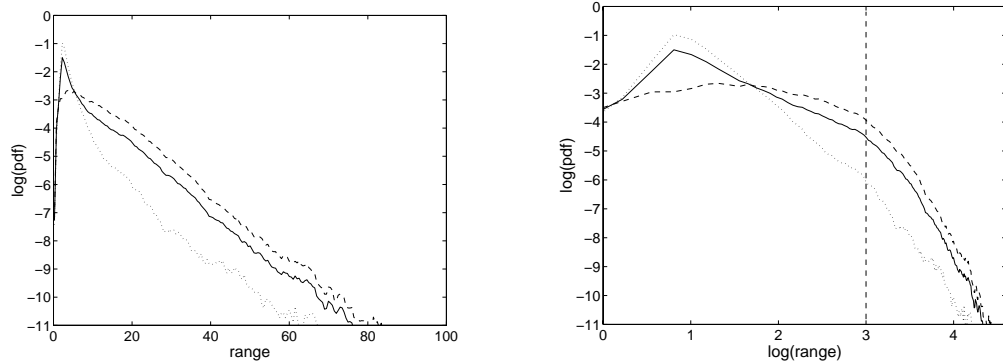


Figure 2.4: Semi-log (left) and log-log (right) plots of the single pixel statistics (i.e. range statistics). *Solid*: distribution for the whole image. *Dotted*: distribution for the bottom half of a range image. *Dashed*: distribution for the top half of a range image.

while the bottom half shows an approximately linear tail in a log-log plot (indicating a power law $\sim r^{-2.6}$). We can qualitatively explain the two different behaviors with the following simplified models:

For the top half, we assume tree trunks (cylinders) uniformly distributed on a plane, according to a Poisson process with density λ . Figure 2.5 shows a top view of a randomly generated “forest” scene. Each disk represents a cross section of the trunk of a tree.

If we assume all disks are of diameter D , and assume that the probability that a horizontal beam from the laser range-finder first hits a tree at distance r is given by an distribution $f(r)$. Let $g(r_0)$ be the probability that $r > r_0$, then

$$g(r_0) = 1 - \int_0^{r_0} f(r) dr$$

hence,

$$f(r) = -\frac{dg(r)}{dr}$$

From r_0 to $r_0 + dr$, trees in the annulus $r_0 < r < r_0 + dr$ will block η percent of the beams which reach the distance r_0 , since the number of trees in the annulus is

$$\lambda \cdot 2\pi r_0 dr$$

and we have

$$\eta = \frac{\lambda \cdot 2\pi r_0 dr D}{2\pi r_0} = \lambda D dr$$

By definition of $g(r)$, $g(r + dr) = g(r) \cdot (1 - \eta)$. Hence

$$\log(g(r + dr)) - \log(g(r)) = \log(1 - \eta) = \log(1 - \lambda D dr) = -\lambda D dr$$

which indicates that

$$\log(g(r)) = C - \lambda D r$$

and $g(r) = e^{C - \lambda D r}$. So

$$f(r) = -\frac{dg(r)}{dr} = \lambda D e^{C - \lambda D r}$$

which means that $f(r)$ decays exponentially.

For the bottom half, we assume flat ground only. Let the height of the sensor from the ground be H , as shown in Figure 2.6. Then at angle θ , the distance between the sensor and the ground is $r = \frac{H}{\sin\theta}$. The laser range-finder samples evenly with respect to the polar coordinate(s) θ (and ϕ), i.e. the density function w.r.t. θ is some constant,

$$h(\theta) = C$$

for $0 < \theta < \pi/2$ Then the density function with respect to r is,

$$g(r) = \frac{d\theta}{dr} h(\theta) \sim \frac{H}{r^2 \sqrt{(1 - (H/r)^2)}} \sim \frac{1}{r^2}$$

for large r .

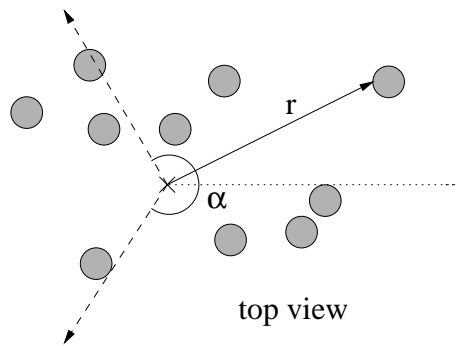


Figure 2.5: Top view of a randomly generated forest scene. See text.

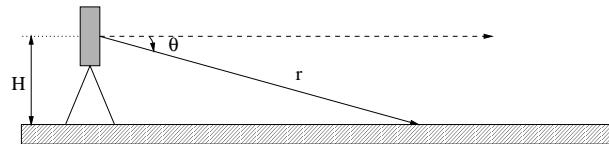


Figure 2.6: Ground model. See text.

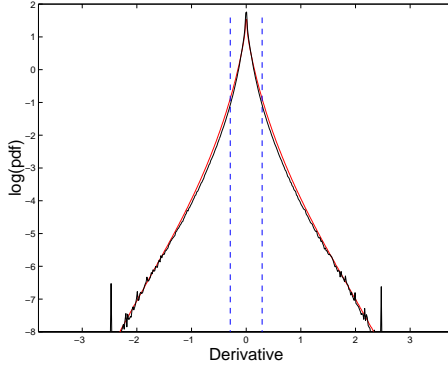


Figure 2.7: Compare the derivative statistics from two databases. Red: van Hateren, Black: Sowerby

2.2 Derivative Statistics

2.2.1 Derivative Statistics For Intensity Images

We now look at the marginal distribution of horizontal derivatives, which in discrete case, is simply the difference between two adjacent pixels in a row, i.e.

$$D = I(i, j) - I(i, j + 1)$$

Figure 2.7 shows the $\log(pdf)$ of D for two different databases.

Compared to the statistics of single pixel intensity (figure 2.2 on page 12), the derivative statistic is much more consistent through different databases. Let $D^{(k)}$ be the derivative at scale k , i.e.

$$D^{(k)} = I^{(k)}(i, j) - I^{(k)}(i, j + 1)$$

Figure 2.8 shows the empirical distributions of $D^{(k)}$ calculated from the two intensity image databases, for $k = 1, 2, 4, 8, 16$. We can see that, except for far in the tails, the distributions of $D^{(k)}$ match reasonably well over different scales. The scaling behavior is a little different between the two image databases, the tail of $D^{(k)}$ from Sowerby database goes up (becomes heavier), and that from van Hateren database goes down. Also, for both image databases,

the difference between scale 1 and scale 2 is relatively larger than that between scale 2 and other scales.

| scale | σ | \mathcal{S} | κ | \mathcal{H} | α_1 | s_1 | $D(p_0 \parallel p_1)$ | α_2 | s_2 | $D(p_0 \parallel p_2)$ |
|-------|----------|---------------|----------|---------------|------------|-------|------------------------|------------|-------|------------------------|
| 1 | 0.259 | 0.077 | 15.4 | -0.24 | 0.69 | 0.077 | 0.00211 | 0.604 | 0.051 | 0.00722 |
| 2 | 0.277 | 0.030 | 12.8 | -0.18 | 0.616 | 0.060 | 0.00119 | 0.655 | 0.072 | 0.00229 |
| 4 | 0.293 | 0.014 | 11.1 | -0.0563 | 0.64 | 0.072 | 0.00335 | 0.7 | 0.093 | 0.00542 |
| 8 | 0.292 | 0.011 | 10.4 | -0.0216 | 0.68 | 0.080 | 0.00233 | 0.722 | 0.10 | 0.00323 |
| 16 | 0.282 | 0.014 | 10.3 | -0.0487 | 0.72 | 0.098 | 0.000929 | 0.725 | 0.10 | 0.000943 |

Table 2.3: constants associated to the derivatives statistics at different scales , van Hateren Database. α_1, s_1 are the constants in the model 2.1 , calculated directly by minimizing the KL distance between empirical distribution and the model, α_2, s_2 the are the constants calculated from 2.2. p_0 is the empirical density function; p_1 and p_2 are the model with constants (α_1, s_1) and (α_2, s_2) respectively

| scale | σ | \mathcal{S} | κ | \mathcal{H} | α_1 | s_1 | $D(p_0 \parallel p_1)$ | α_2 | s_2 | $D(p_0 \parallel p_2)$ |
|-------|----------|---------------|----------|---------------|------------|-------|------------------------|------------|-------|------------------------|
| 1 | 0.248 | -0.018 | 15.1 | -0.338 | 0.624 | 0.056 | 0.00396 | 0.609 | 0.051 | 0.00422 |
| 2 | 0.259 | 0.002 | 15.8 | -0.403 | 0.533 | 0.032 | 0.0013 | 0.597 | 0.049 | 0.00562 |
| 4 | 0.263 | 0.007 | 16.7 | -0.501 | 0.465 | 0.017 | 0.00124 | 0.583 | 0.046 | 0.0187 |
| 8 | 0.274 | 0.022 | 16.3 | -0.502 | 0.428 | 0.012 | 0.00262 | 0.589 | 0.050 | 0.0362 |
| 16 | 0.305 | 0.050 | 14 | -0.319 | 0.432 | 0.014 | 0.00622 | 0.627 | 0.069 | 0.0508 |

Table 2.4: constants associated to the derivatives statistics at different scales , Sowerby Database. α_1, s_1 are the constants in the model 2.1 , calculated by the optimal method. α_2, s_2 the are the constants calculated by direct parameter estimation. p_0 is the empirical density function; p_1 and p_2 are the models with constants (α_1, s_1) and (α_2, s_2) respectively

Table 2.3 and 2.4 shows some constants related to each statistics. (The last 6 columns will be explained later). Notice that the shape of the histogram has a distinct peak at 0, and a concave tail. Writing the density function for D as $f_0(x)$, we consider the popular ‘generalized Laplace’ model f_{model} for f_0

$$f_{model}(x) = \frac{1}{Z} \cdot e^{-|x/s|^\alpha} \quad (2.1)$$

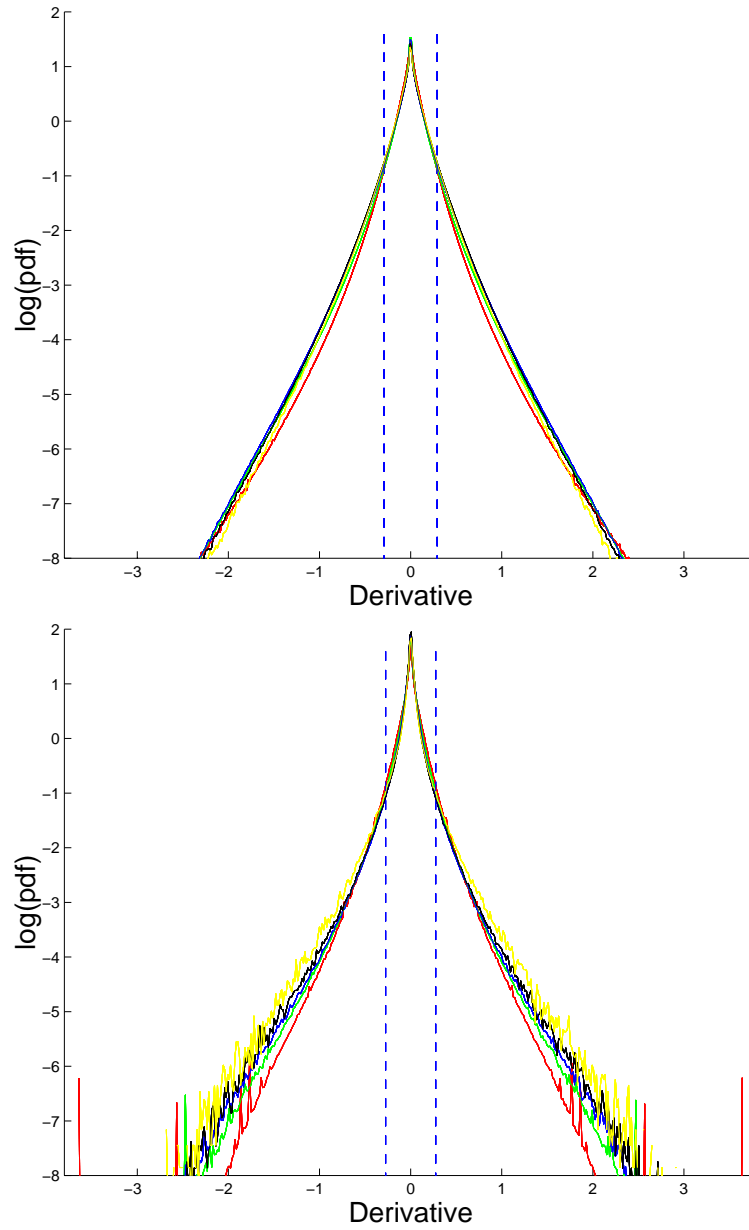


Figure 2.8: Derivative statistics at different scales. Upper: van Hateren database, Lower: Sowerby database. Red, Green, Blue, Black and Yellow for scales 1,2,4, 8,16 respectively

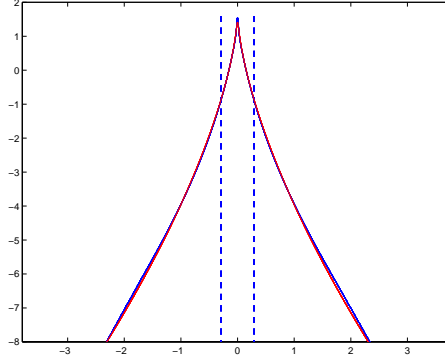


Figure 2.9: The fit generalize Laplace model to derivative statistics at scale 2, van Hateren Database, Blue: Data, Red: Model

where s, α are constants, they can be directly related to the variance and kurtosis by:

$$\sigma^2 = \frac{s^2 \Gamma(\frac{3}{\alpha})}{\Gamma(\frac{1}{\alpha})} \quad \text{and} \quad \kappa = \frac{\Gamma(\frac{1}{\alpha}) \Gamma(\frac{5}{\alpha})}{\Gamma^2(\frac{3}{\alpha})} \quad (2.2)$$

There are two ways to fit the model.

1. *The optimal method* is to find constants α and s which will minimize $D(f_0 || f_{model})$.
2. *Direct parameter estimation* is to solve the equation 2.2 for α and s directly.

The second method is much easier to calculate, and has been used in [3] successfully for parameter estimations.

We calculated the estimation using both methods. The 8th and 11th column of table 2.3 and table 2.4 shows the KL distance between empirical distribution and the models calculated by the optimal and the direct parameter estimation methods respectively. We can see that the direct method is almost as good as the optimal method in most of the cases. Figure 2.9 shows an example of the fit of the model to empirical distribution by using the optimal method.

2.2.2 Derivative Statistics For Different Categories of Intensity Images

Figure 2.10 shows the density functions for the derivative statistics of different categories and at different scales and table 2.5 shows constants associated to them. The man-made

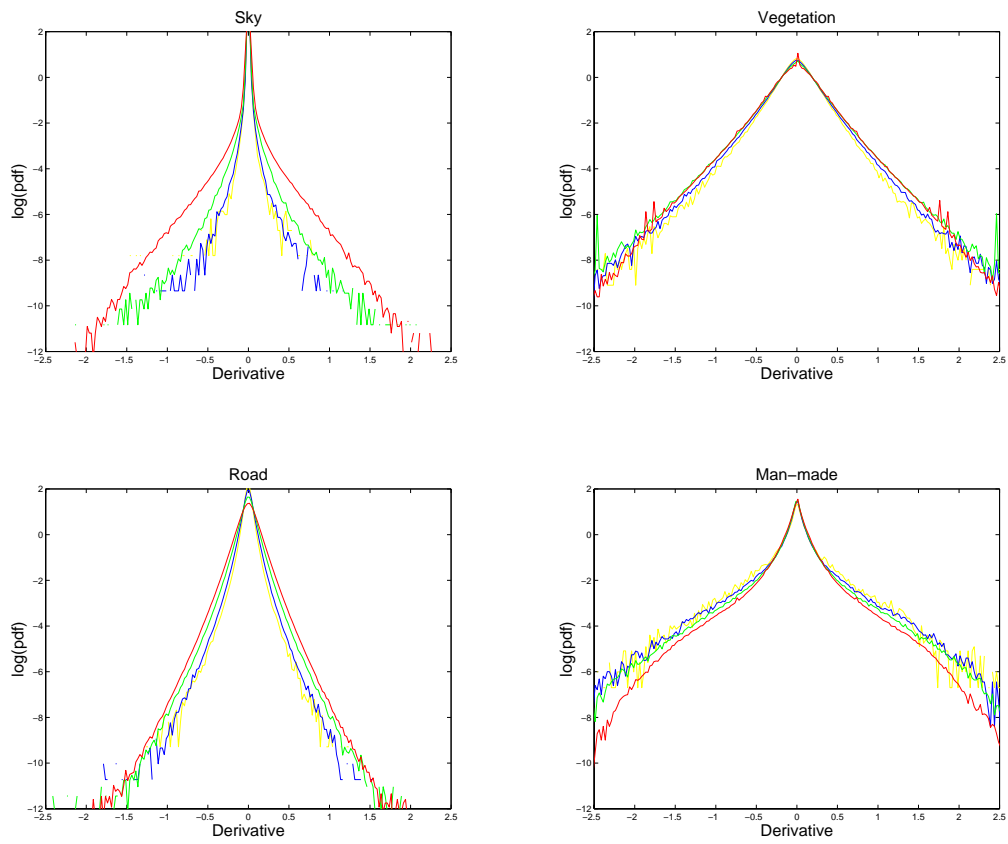


Figure 2.10: $\log(\text{pdf})$ of D at different scales and categories: red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8

and vegetation categories scale fairly well while the sky and road categories do not scale at all well. Next, we compare the scale behavior of different categories. For each category, let the standard deviation sd_ℓ of D at scale level $\ell = \log_2(k)$. In figure 2.11 where \log_2 of the standard deviation is plotted against ℓ , the negative of the slope of the fitting line gives us an approximation to half of what the physicists call the ‘anomalous dimension’ η . In other words we are fitting $sd_\ell \approx 2^{-\frac{\eta}{2}\ell} sd_0$. Equivalently, this is fitting a model for the second order statistics in which the power spectrum scales as $1/\text{freq}^{(2-\eta)}$.

Here are some observations we made for each category:

1. The vegetation category looks linear in the log plot of the histogram. It scales well although the power spectrum is modeled by $1/\text{freq}^{1.8}$ which is very close to what Ruderman and Bialek found [18]. The $\log(\text{histogram})$ can be modeled by $C_1 - C_2|x|$, the ‘double exponential’ distribution.
2. The man-made category has a histogram with big ‘shoulders’ in the log plot. The center parts of the histograms match well for different scales, but the tails go up with increasing scale. We believe this phenomena is caused by large objects and their edges. Along an edge the total number of pixel pairs goes down by the factor of 2 when scaling, while the overall number of pairs goes down by the factor of 4. As a result, the frequency of edge pixels increases.
3. For the sky category, the density of the distribution mainly concentrated around 0, and shifts further to the center with increasing scale. The scaling fit gives an power spectrum like $1/\text{freq}^{1.0}$.
4. In the road category, the log histogram is slightly concave, and scales badly. However, if we correct for the changing variance, using the assumption that its power spectrum is $1/\text{freq}^{(1.4)}$, we get a much improved fit as shown in figure 2.12.

To summarize, it appears that the multi-scale families of histograms for each category can be modeled with three parameters: a) their standard deviation, b) the anomalous dimension η and c) a ‘shape’ parameter for the histogram which has been identified in [16] as the parameter in an infinitely divisible family of probability distributions.

| <i>Category</i> | <i>Scale</i> | σ | κ | \mathcal{H} | α_1 | $D(p_0 \parallel p_1)$ | α_2 | $D(p_0 \parallel p_2)$ |
|-----------------|--------------|----------|----------|---------------|------------|------------------------|------------|------------------------|
| Sky | 1 | 0.069 | 59 | -3 | 0.165 | 0.0158 | 0.386 | 0.12 |
| | 2 | 0.037 | 57.3 | -3.57 | 0.147 | 0.009 | 0.389 | 0.128 |
| | 4 | 0.024 | 46.8 | -3.83 | 0.159 | 0.006 | 0.412 | 0.104 |
| | 8 | 0.022 | 46.3 | -3.83 | 0.17 | 0.007 | 0.413 | 0.0837 |
| Vegetation | 1 | 0.34 | 15.1 | 0.328 | 0.937 | 0.0151 | 0.607 | 0.0698 |
| | 2 | 0.34 | 10.3 | 0.323 | 0.95 | 0.00377 | 0.726 | 0.0238 |
| | 4 | 0.31 | 9.37 | 0.197 | 0.956 | 0.0033 | 0.762 | 0.0172 |
| | 8 | 0.28 | 9.01 | 0.0697 | 0.966 | 0.0044 | 0.778 | 0.0168 |
| Road | 1 | 0.14 | 7.12 | -0.87 | 1.08 | 0.00389 | 0.892 | 0.0121 |
| | 2 | 0.12 | 9.09 | -1.19 | 0.942 | 0.00423 | 0.774 | 0.0129 |
| | 4 | 0.10 | 11.3 | -1.61 | 0.856 | 0.00505 | 0.693 | 0.0137 |
| | 8 | 0.08 | 15.4 | -1.93 | 0.767 | 0.00823 | 0.602 | 0.0173 |
| Man-made | 1 | 0.30 | 12.6 | -0.221 | 0.522 | 0.00832 | 0.659 | 0.0294 |
| | 2 | 0.35 | 12.3 | -0.105 | 0.461 | 0.00855 | 0.666 | 0.0561 |
| | 4 | 0.39 | 10.8 | 0.0595 | 0.443 | 0.011 | 0.709 | 0.0814 |
| | 8 | 0.41 | 9.3 | 0.259 | 0.491 | 0.0155 | 0.765 | 0.0717 |

Table 2.5: Some constants associated to derivative statistics of different categories.

2.2.3 Derivative Statistics for Range Images

| <i>scale</i> | σ | \mathcal{S} | κ | \mathcal{H} |
|--------------|----------|---------------|----------|---------------|
| 1 | 0.203 | -0.0453 | 48.9 | -2.89 |
| 2 | 0.225 | -0.113 | 39.8 | -2.6 |
| 4 | 0.246 | -0.217 | 32.3 | -2.19 |
| 8 | 0.266 | -0.322 | 25.9 | -1.69 |

Table 2.6: constants associated to the derivative statistics of range images at different scales(Scaling down by block minimum).

We calculated the histogram of the Derivative statistics from range images. The red line in figure 2.13 shows the log probability density function of D . As in the studies of intensity images, this distribution has a high kurtosis with large tails, and a peak at 0. It is closest to the statistics for intensity images of man-made environments(see figure 2.10). but has an even higher peak at 0. This strongly suggests a mixture of two different type of distributions, one is highly concentrated around 0, and the other has heavy tails. We believe this corresponds to the cases when the two adjacent pixels are on the same object or different objects respectively. Let x_1, x_2 be two adjacent pixels, and $a = I(x_1), b = I(x_2)$

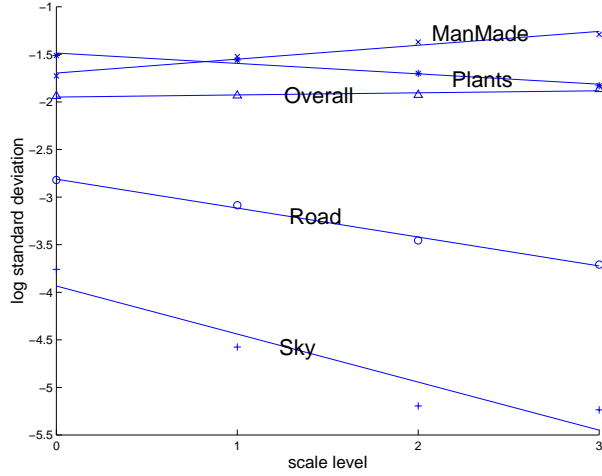


Figure 2.11: Plot of $\log_2(\text{standard deviation})$ vs. scale level and its linear fit in five cases: + = sky, * = vegetation categories, o = road, x = man-made categories and Δ = all pixels. The slopes are 0.15 for man-made, 0.02 for all, -0.11 for vegetation, -0.30 for road and -0.50 for sky.

be the log-range at the two pixels. Let f_0 be the probability density of D , f_1 be that of D given that x_1, x_2 are on the same object, and f_2 be that of D given that x_1, x_2 are on different objects. Let $0 < \lambda < 1$ be the probability that x_1, x_2 are on the same object, then

$$f_0 = \lambda \cdot f_1 + (1 - \lambda) \cdot f_2 \quad (2.3)$$

Since pixels from the same object are likely have similar values ($a \approx b$), it's reasonable to assume that f_1 concentrates around 0. On the other hand, when two pixels are from different objects, a and b should be independent. Let g_0 be the marginal distribution of a and b , which has been discussed in section 2.1.3, we may calculate f_2 by

$$f_2 = \text{convolve}(g_0(x), g_0(-x)) \quad (2.4)$$

f_2 is shown in figure 2.14(left). Compare the curve in figure 2.14 and the red curve in figure 2.13, we see that they have similar tails, indicating that the outliers of f_0 come from those of f_2 . Figure 2.14, right, shows the fit of the model given in equation 2.3. We will explore similar mixture models in section 2.3 when we discuss the joint distribution of pixel

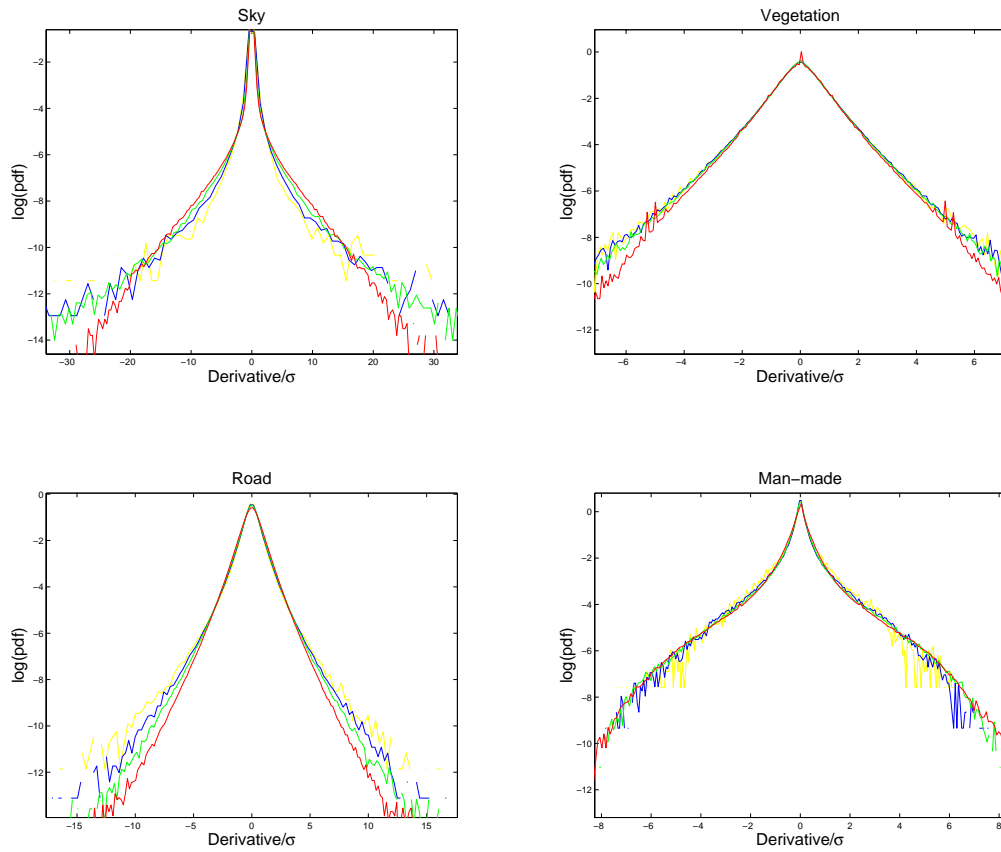


Figure 2.12: $\log(pdf)$ of D at different scales and categories. Renormalized according to the standard deviations. red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8

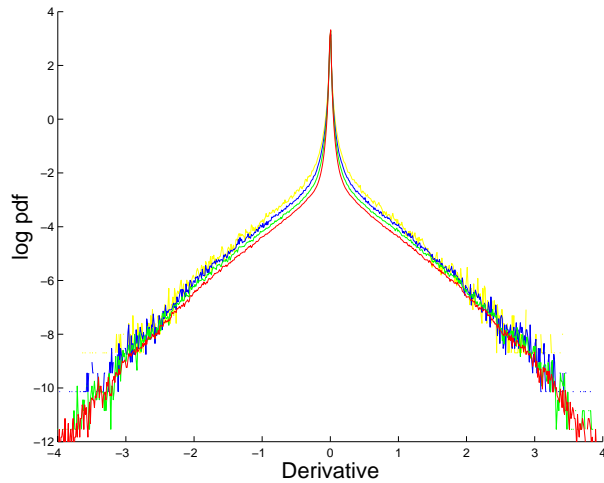


Figure 2.13: Derivative statistics for range images at different scales(Scaling down by block minimum). red = scale 1, green = scale 2, blue = scale 4, yellow = scale 8

pairs.

We also calculated the distribution of D at different scales. Here we scale down the range images by taking the minimum, instead of the average of $N \times N$ blocks. This is the appropriate renormalization for range images, because laser range finders measure the distance to the nearest object in each solid angle. Figure 2.13 shows the results, which indicate that range images scale fairly well. Table 2.6 shows constants associated to this statistic.

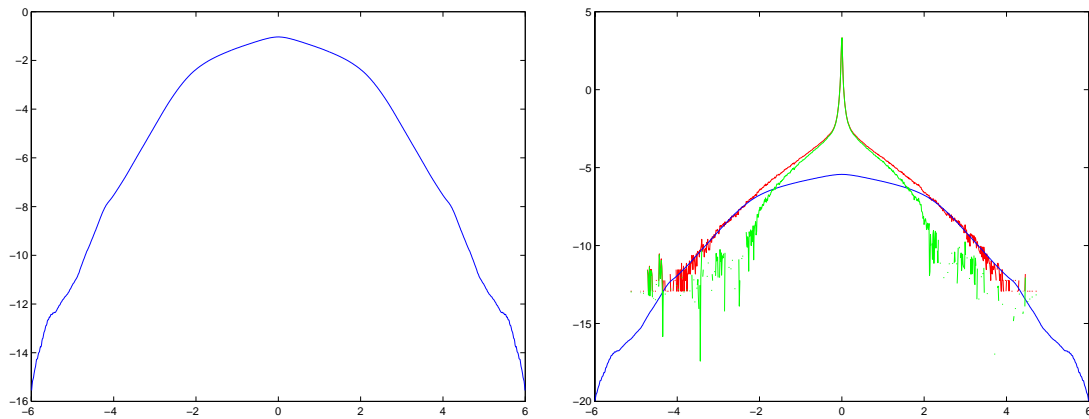


Figure 2.14: Left: $\log(\text{pdf})$ of f_2 . Right: the fit of equation 2.3, red = $\log(f_0)$, blue = $\log((1 - \lambda) \cdot f_2)$ and green = $\log(\lambda \cdot f_1)$, and $\lambda = 0.988$

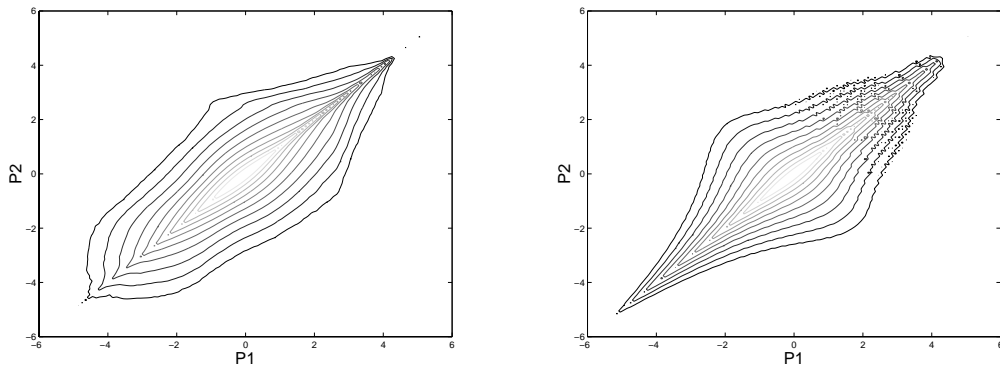


Figure 2.15: Left figure: Joint Histogram of two adjacent pixels p_1 and p_2 , calculated from van Hateren database. Right figure: The product density function of $p_1 - p_2$ and $p_1 + p_2$.

2.3 Two-point Statistics

2.3.1 Intensity Images

Figure 2.15(left) shows the joint distribution of the intensities a and b at two horizontally adjacent pixels x_1 and x_2 , calculated from van Hateren database, where $a = I(x_1)$, $b = I(x_2)$. The differential entropy of this joint distribution is: $\mathcal{H}(a, b) = 1.51$ and the mutual information between a and b is: $\mathcal{I}(a, b) = \mathcal{H}(a) + \mathcal{H}(b) - \mathcal{H}(a, b) = 1.80$. Notice that the mutual information \mathcal{I} between adjacent pixels is a large number, indicating that adjacent pixels are highly correlated. On the other hand, we can see from the contour plot, that there is some symmetry along $a = b$, and a rough symmetry along $a = -b$, we may guess that the sum and the difference of two adjacent pixels are more likely independent. Figure 2.15(right) shows the direct product of the marginals of $a + b$ and $a - b$ (still plotted according to a and b). Comparing the two contour plots, we can see that at the center part (where the density is much higher than other places) the product distribution and the original distribution are very similar, but the shape of the level curves away from $(0, 0)$ becomes quite different. The mutual information between $a + b$ and $a - b$ is $\mathcal{I}(a + b, a - b) = 0.0255$. Compared to that of a and b , it's very small, indicating a rough independence between $a + b$ and $a - b$ from information theory point of view. On the other hand, if the two pixels x_1 and x_2 are far away, they are then likely on different objects. As in section 2.2.3, it will be reasonable

to assume that the intensities a, b are independent. It is interesting to see how the joint distribution of a and b shifts from one model to another as the distance between the two pixels becomes larger. Follow the same idea presented in section 2.2.3, we propose the following model, Let

$$K(a, b|x) = Pr\{I(\mathbf{x}_1) = a, I(\mathbf{x}_2) = b \mid \|\mathbf{x}_1 - \mathbf{x}_2\| = x\} \quad (2.5)$$

where $I(\mathbf{x}_1)$ and $I(\mathbf{x}_2)$ represent the log contrast at pixels \mathbf{x}_1 and \mathbf{x}_2 . The odd columns of Figure 2.16 show the contour plots of $K(a, b|x)$ for separation distances $x = 1, 2, 4, \dots, 128$. As we saw in Figure 2.15, the joint distribution aligned along $u = a + b$ and $v = a - b$, so we plot the distributions under this coordinate. As in [14], we use the following model to fit the bivariate statistics of images:

$$\tilde{K}(a, b|x) = [1 - \lambda(x)]q_x(a)q_x(b) + 2\lambda(x)h_x(a + b)g_x(b - a) \quad (2.6)$$

where q_x is the marginal distribution for a single pixel, h_x is some distribution similar to q_x , and g_x is a distribution highly concentrated at 0. The first term models the case where the two pixels belong to different objects (we assume that different object are statistically independent), the second term represents the case where they are on the same object (assume that the difference of the pixels is some noise, which does not depend on the average of the two pixels), and $\lambda(x)$ is the probability of their being on the same object. Under the coordinate (u, v) , the model becomes:

$$Q(u, v|x) = \frac{1}{2}[1 - \lambda(x)]q_x\left(\frac{u+v}{2}\right)q_x\left(\frac{u-v}{2}\right) + \lambda(x)h_x(u)g_x(v) \quad (2.7)$$

The even columns of figure 2.16 shows the best fit of the model 2.6 to the empirical statistics (shown to the left). The fits are good around the center, but do not match the empirical distribution well in the outliers. Figure 2.17 shows the λ values and functions q_x, h_x and g_x in the best fittings for different x . We can see that q_x and h_x are almost the same for all x .

2.3.2 Range Images

For two adjacent pixels in range images, the mutual information $\mathcal{I}(a + b, a - b) = 0.0631$ is also a small number, indicating a similar mixture model for range images. Figure 2.18 shows the joint distributions of pixel pairs of range images and the best fits to them. Figure 2.19 shows the λ values and functions f_x, g_x and h_x of the models, as in Figure 2.17.

We find that the model fits *better* to the bivariate statistics of range images than to that for intensity images (compare to Figure 2.16). This indicates that range images present a simpler, cleaner problem than intensity images. For example, the concept of objects is better defined for range images where we do not need to take lighting, color variations, texture etc. into account.

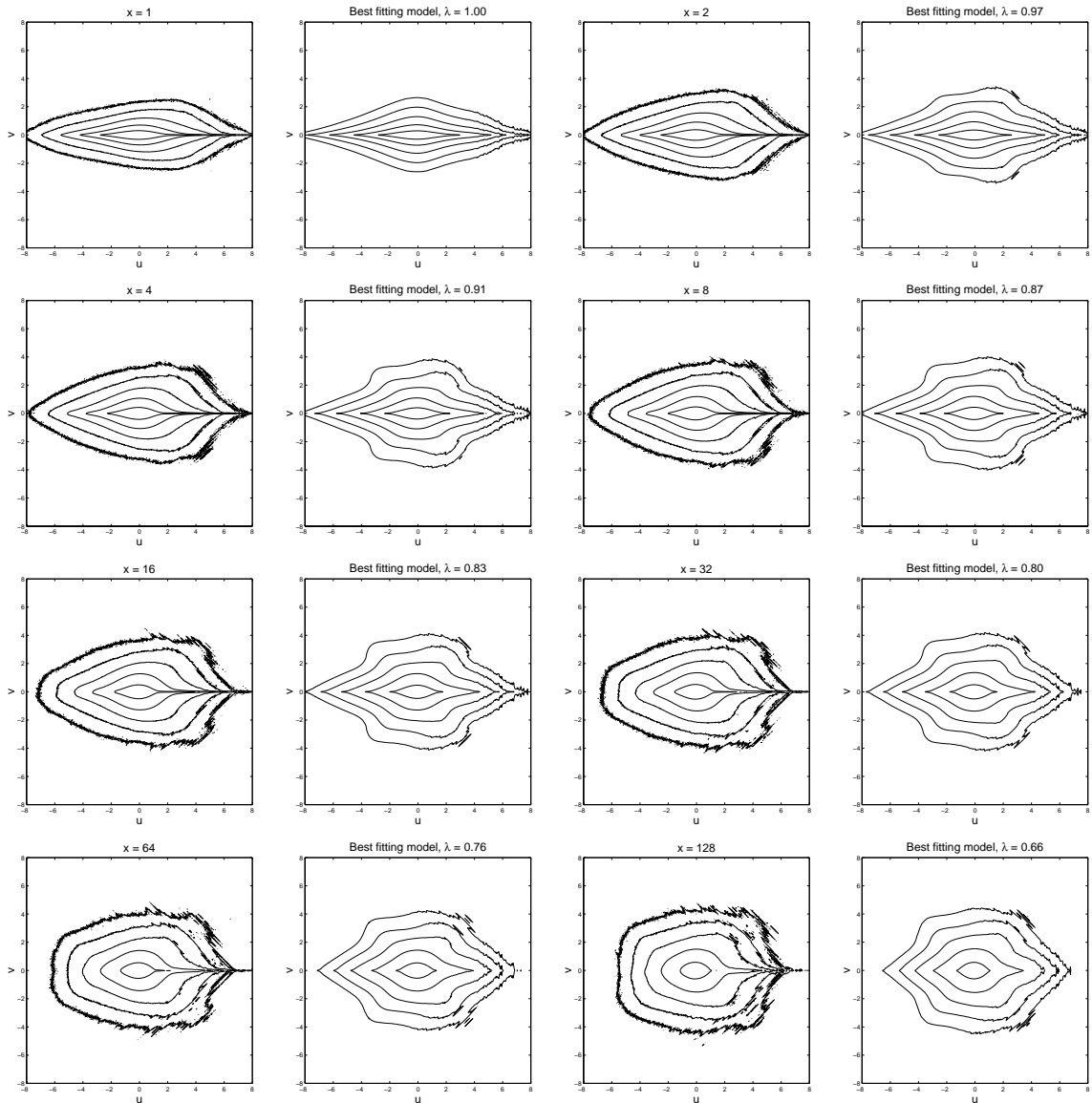


Figure 2.16: Contour plots of the log histograms of pixel pairs for intensity images (left column) and the best bivariate fit to the random collage model (right column). x : distance between the two pixels, u : sum of the pixel values, v : difference of the two pixel values.

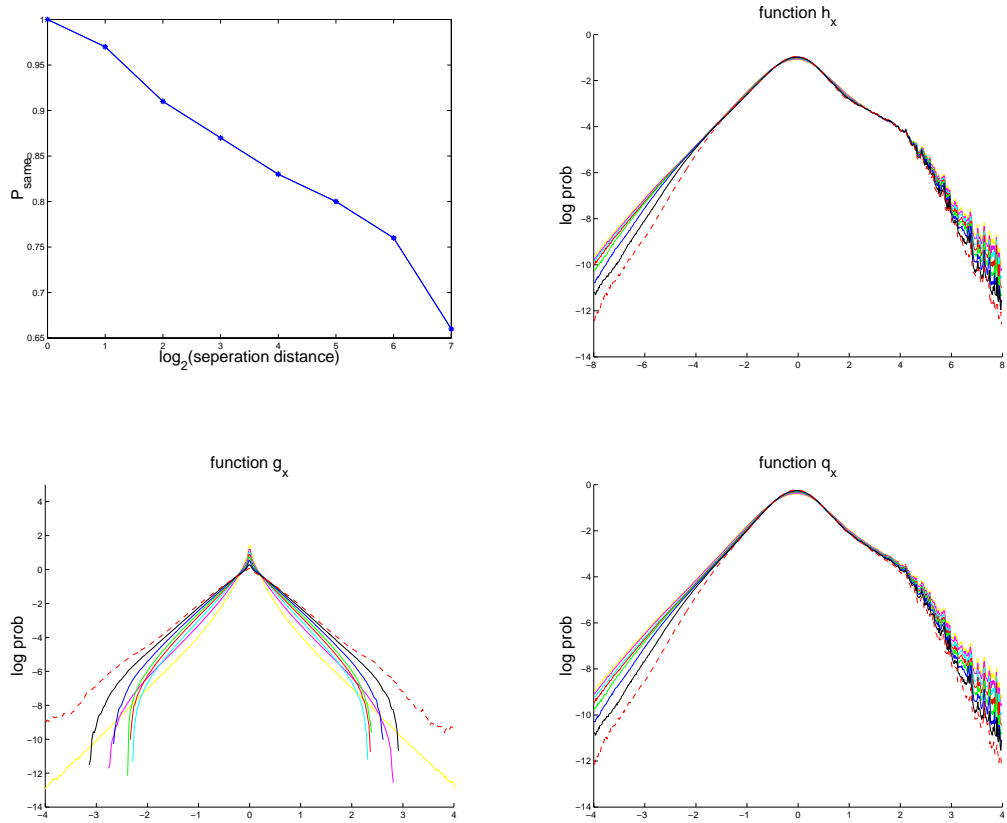


Figure 2.17: The values for λ and the 1D functions g_x , h_x , q_x in the best bivariate fit to the bivariate statistics of intensity images at distances $x = 1$ (yellow), 2 (magenta), 4 (cyan), 8 (red), 16 (green), 32 (blue) and 128 (dotted red).

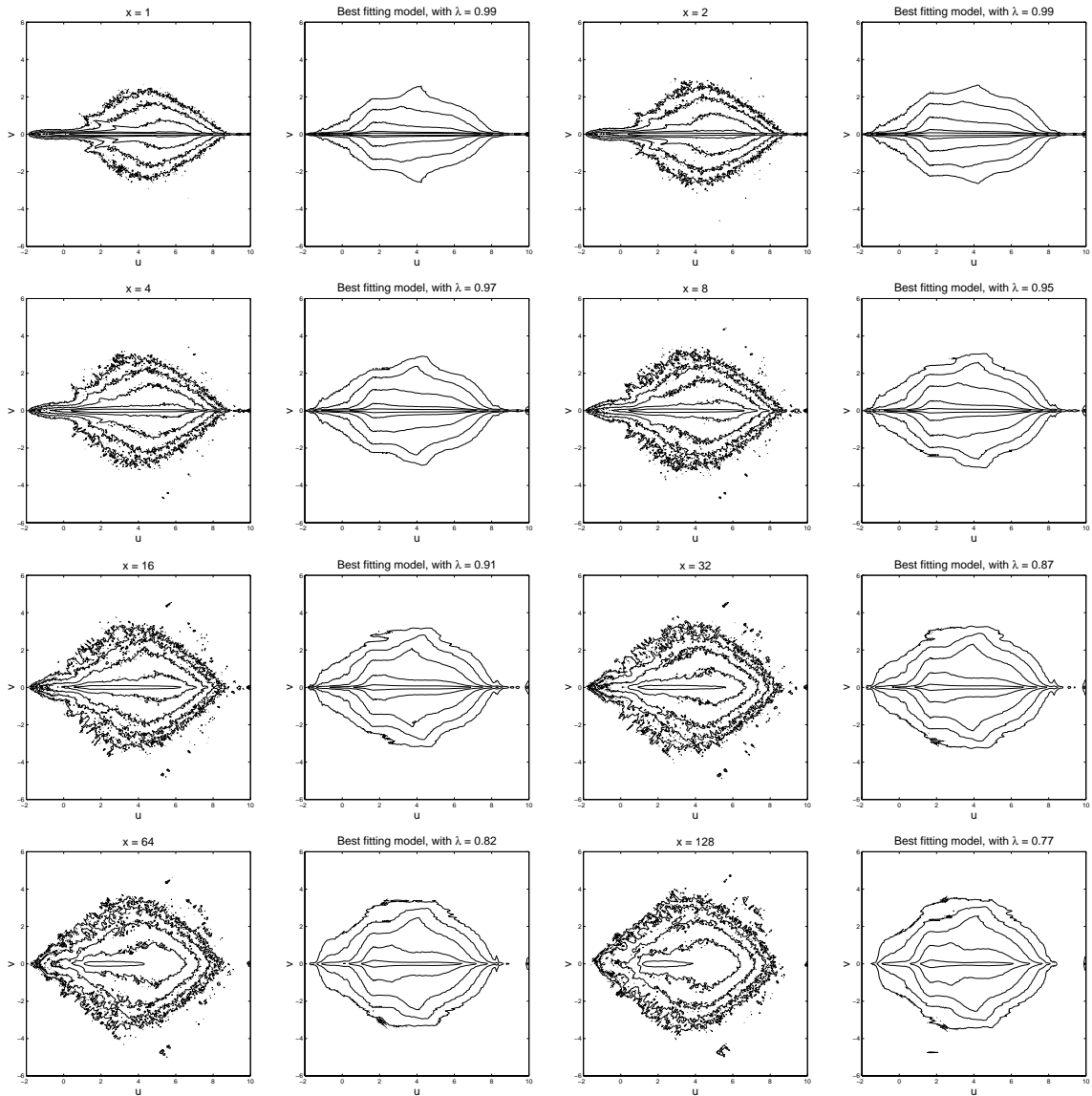


Figure 2.18: Contour plots of the log histograms of pixel pairs for range images (left column) and the best bivariate fit to the random collage model (right column). x : distance between the two pixels, u : sum of the pixel values, v : difference of the two pixel values.

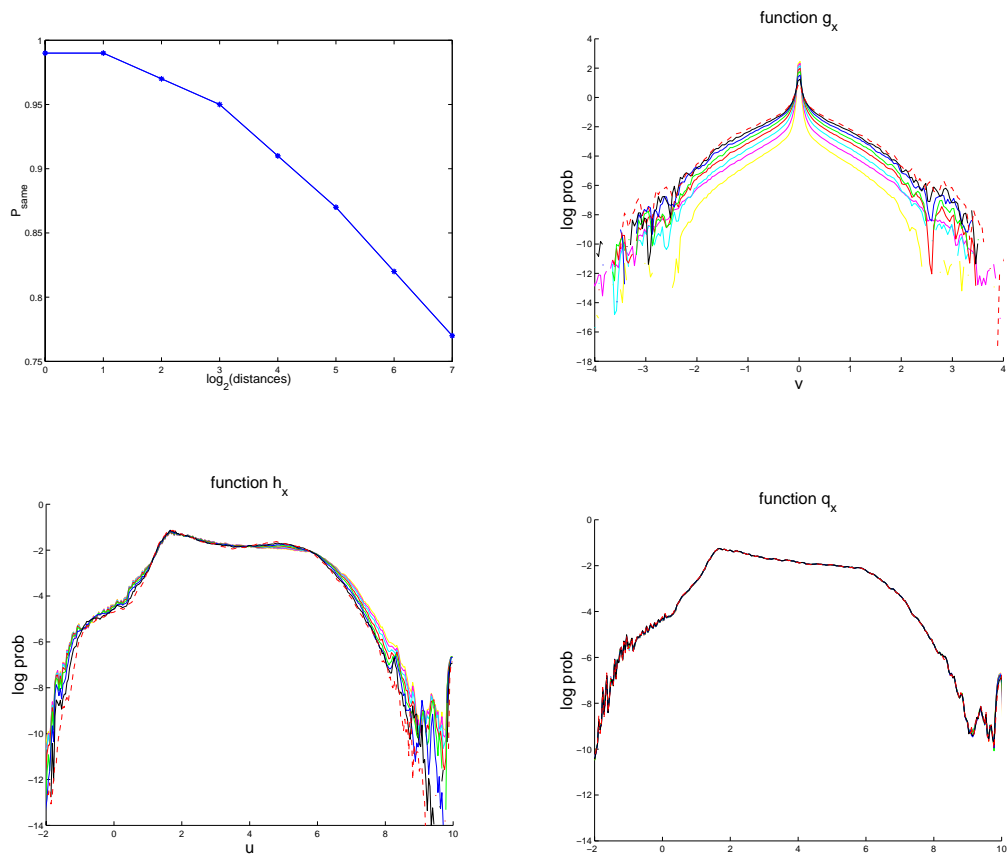


Figure 2.19: The values for λ and the 1D functions g_x , h_x , q_x in the best bivariate fit to the bivariate statistics of range images at distances $x = 1$ (yellow), 2 (magenta), 4 (cyan), 8 (red), 16 (green), 32 (blue) and 128 (dotted red).

2.3.3 Covariance of nature images

The covariance of images are defined as,

$$C(x, y) = \langle I(x, y)I(0, 0) \rangle$$

Here $\langle \rangle$ is the expectation, taking over all the images. However, our images are samples of a distribution which is only well-defined up to an additive constant, so we replace this statistic by the ‘difference function’:

$$D(x, y) = \langle |I(x, y) - I(0, 0)|^2 \rangle$$

which is related to the covariance by

$$D(x, y) + 2C(x, y) = \text{constant}$$

when both are well defined.

In section 2.3.1, we studied the bivariate statistics, which contain all the information about the difference function. The goal of this section is to check the correlations along different directions and the scale invariance property.

In [17], Ruderman calculated the ‘one-dimensional’ difference function, i.e., he took the average of $D(x, y)$ over all directions, and got a one dimensional function $D_1(x)$ to which he fit a scaling model:

$$D_1(x) = C_1 + C_2|x|^{-\eta} \tag{2.1}$$

These covariance models correspond in frequency domain to the power spectrum $\sim \frac{1}{f^{2-\eta}}$. If η goes to 0, note that $1 - r^{-\eta} = 1 - e^{-\eta \log r} \approx \eta \log r$ giving us the model

$$D_1(x) = C_1 + C_2 \log(|x|) \tag{2.2}$$

which is the model implied by the assumption that 2×2 block averages of the image I have the same distribution as I [21]. The best fitting constants Ruderman found from his image dataset are: $C_1 = 0.79$, $C_2 = -0.64$ and $\eta = 0.19$.

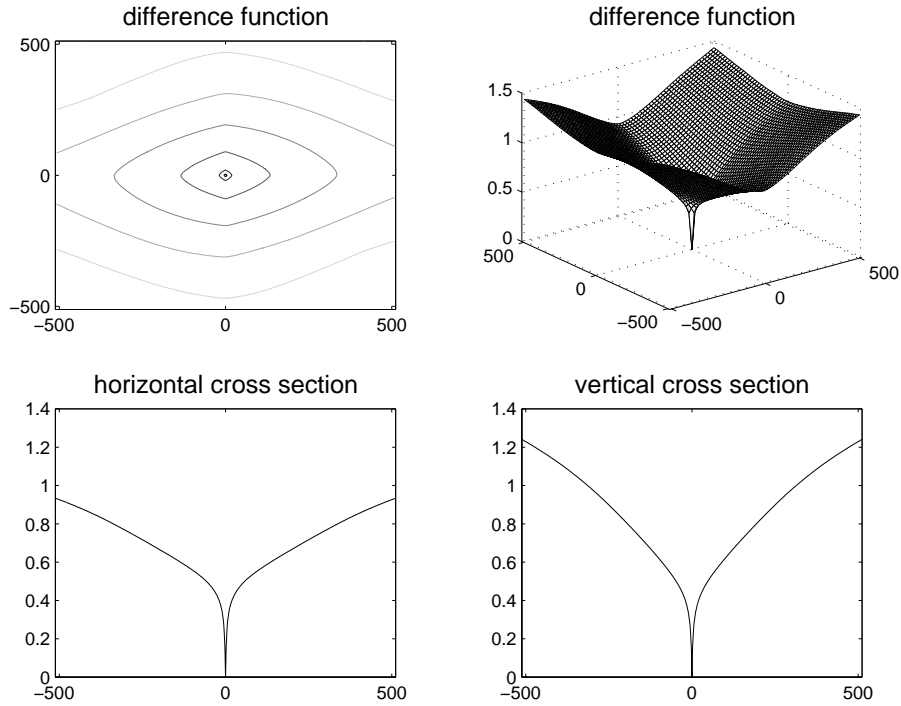


Figure 2.20: Difference Function

We calculated the two dimensional $D(x, y)$ from our data set. Using a Fourier transformation technique, we actually took into our calculation *all* the possible pixel pairs within distance of 500 pixels. The statistics we got are very stable, and we can look more closely at the tail of the statistics, and even take delicate operations like derivatives on them. The upper two images in figure 2.20 show the contour and mesh plot of $D(x, y)$ we got. The lower two show the two cross sections along horizontal and vertical direction. We can see, the cross section along vertical direction grows faster than that along the horizontal direction. We believe the main reason is that, in many images, there is a portion of sky at top, and ground at bottom and the large difference between them will contribute a lot to the difference function along the vertical direction.

The upper left image in figure 2.21 shows the log-log plot of the derivative of the positive part of horizontal cross section. The base we used when we took the log operation is 2. We see that, between 2 and 5 (corresponding to distances of 4 and 32 pixels), the derivative is

close to a straight line, with a slope -1.19. If we use model (6.1), then $\eta = -(-1.19+1) = 0.19$, which is exactly what Ruderman got. But notice how the log-log plot begins to turn and becomes almost a horizontal line around 8. This clearly indicates that there exists a linear term, i.e. we can model it as:

$$D_1(x) = C_1 + C_2|x|^{-\eta} + C_3|x|$$

Generalizing it to $D(x, y)$, we seek a model:

$$D(x, y) = C_1(\theta) + C_2(\theta)r^{-\eta} + C_3(\theta)r$$

where, $r = \sqrt{x^2 + y^2}$, and $\theta = \tan^{-1}(\frac{y}{x})$. The best fitting η we got is 0.32, and the best fitting $C_1(\theta)$, $C_2(\theta)$ and $C_3(\theta)$ are shown in figure 2.21. The maximum fitting error we got is 0.0035, which is very small, considering the range of $D(x, y)$ is between 0 and 0.8, and the large area on which we fit the model (an annulus with $4 < r < 200$).

One interesting observation we make is that $C_1(\theta) + C_2(\theta)$ is almost zero, hence we may fit our model with one less parameter:

$$D(x, y) = C_2(\theta)(1 - r^{-\eta}) + C_3(\theta)r$$

Since $C_3(\theta)$ is very small, the linear term can be omitted when r is small, we get (for r small):

$$D(x, y) \approx C(\theta)\ln(r)$$

This shows that while random images seem very close to samples from a scale-invariant process, there are also systematic deviations from scale invariance on a large scale.

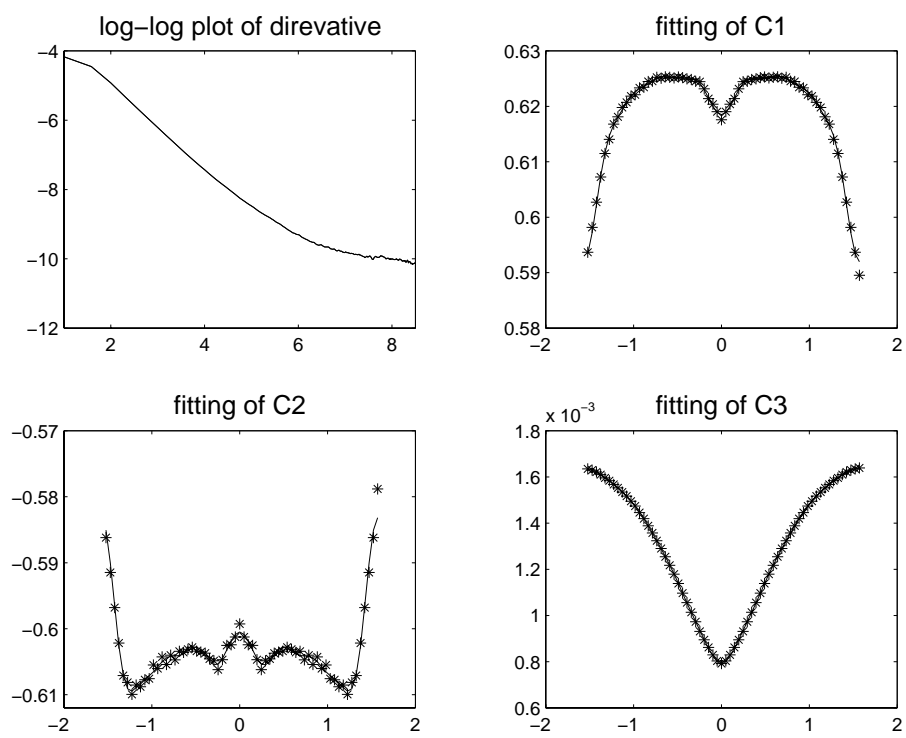


Figure 2.21: log-log plot of the horizontal cross section and some fitting constants,see text

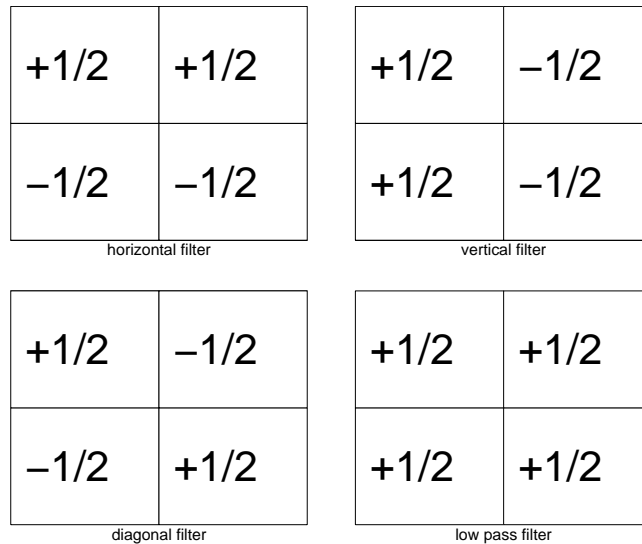


Figure 2.22: Haar Filters

2.4 Joint Statistics of Haar Filters Reactions

2.4.1 Haar Filters

The idea of applying wavelet pyramids to image processing has proven to be very successful. We choose the Haar wavelet for its simplicity: any structure in the statistics can be directly related to pixel values. Figure 2.22 shows the four basic Haar filters in 2D. Assuming we have a 8×8 image, we can apply the horizontal filter on non overlapping 2×2 blocks of the image to get a 4×4 matrix of responses, which called the first level horizontal sub band. Similarly, we have the first vertical and diagonal sub bands. Next, apply the low pass filter on the image, getting a 4×4 image, and repeat the above procedure on this smaller image to get the second level horizontal, vertical and diagonal sub bands, all of them with dimension 2×2 . This procedure can go on till we get a 1×1 image. This way we get the wavelet pyramid of sub bands whose statistics we wish to study. Figure 2.23 shows the sub bands of the first two levels. In order to describe the relative positions of wavelet coefficients in sub bands, we borrow some of the definitions given in [3]: we call the coefficients at adjacent spatial locations in the same sub band *brothers*, (left, right, upper, lower brothers according to the relative positions), call the coefficients in same level, same position, but different

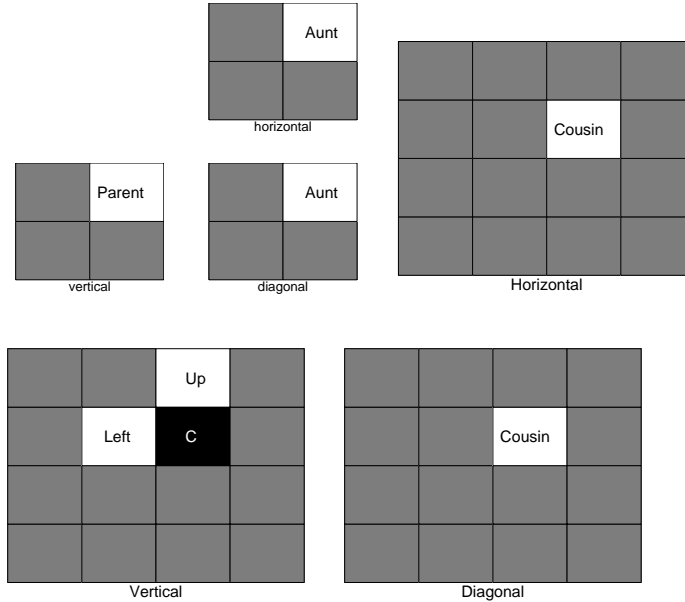


Figure 2.23: Relations Between Wavelet Coefficients

orientations *cousins*. And call the relationship of coarse-scale coefficients and finer-scale coefficients *parent* and *child*. Figure 2.23 shows a coefficient **C** in the vertical sub band, and its relatives.

2.4.2 2D Joint Statistics of Intensity Images

Figure 2.24 shows the mesh plot of the joint statistics of the pair horizontal component(*hc*), and its vertical cousin(*vc*). Figure 2.25 shows the contour plot(red curves) of the joint density functions of different pairs. Table 2.8 shows some constants about each pair. It's clear that the pdf of each pair is roughly symmetrical with respect to the origin. All contour plots show some polyhedra-like structures, each vertex on the level curve represents a direction along which the probability density is higher than neighboring directions. It is interesting to see what kind of image patches these vertices correspond to. For example, look at the vertex marked by ▼ in the upper left subplot of figure 2.25. Assume it corresponds

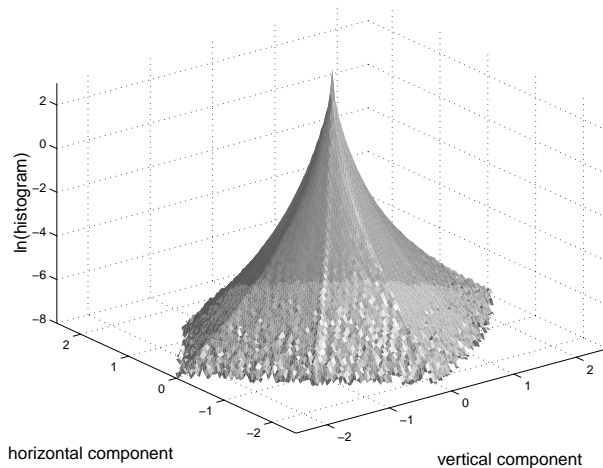


Figure 2.24: Mesh Plot of the $\ln(\text{Joint Histogram})$ of Horizontal Component and its Vertical Cousin

to a 2×2 image patch

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then $a + b - c - d = a - b - c + d$, hence $b = d$. So this vertex corresponds to T-junctions. Likewise, we can explain other vertices(see table 2.7). In section 2.4.5 we will discuss more about the vertices and their meaning in detail about range images.

We tried two methods to model these statistics. First we made the observation from figure 2.24 that, for any angle $0 \leq \theta \leq \pi$, the cross section along $hc = \tan(\theta)vc$ is similar to the derivative statistics we got in section 2.2.1, and the ‘generalized Laplacian’ model can be easily used to fit the cross sections here. This suggests, as a first step that we fit the following model to the joint distribution of the pair (hc, hv) .

$$f(x, y) = e^{C_1(\theta) + \left(\frac{r}{s(\theta)}\right)^{\alpha(\theta)}} \quad (2.3)$$

where $r = \sqrt{hc^2 + hv^2}$ and θ is the angle of (hc, hv) . From figure 2.24, $\log(f(x, y))$ is continuous at 0, hence $C_1(\theta)$ should be a constant. We will also assume α to be constant

| Symbol | Typical image patch | Description |
|----------------------|--|----------------------|
| Δ | $\begin{bmatrix} a & a \\ b & b \end{bmatrix}$ | horizontal edge |
| \blacktriangle | $\begin{bmatrix} a & b \\ a & b \end{bmatrix}$ | vertical edge |
| \blacktriangledown | $\begin{bmatrix} a & b \\ a & d \end{bmatrix}$ | T-junction |
| \square | $\begin{bmatrix} a & a & a & a \\ b & b & b & b \end{bmatrix}$ | continuation in edge |
| \blacksquare | $\begin{bmatrix} a & a & c & c \\ b & b & c & c \end{bmatrix}$ | T-junction |
| \diamond | $\begin{bmatrix} a & b & b & b \\ a & b & b & b \end{bmatrix}$ | vertical edge |
| \blacklozenge | $\begin{bmatrix} a & b & b & a \\ a & b & b & a \end{bmatrix}$ | vertical bar |

Table 2.7: Some interesting features in figure 2.25 and their explanation

to simplify our model, so we fit the following model to data.

$$f(x, y) = e^{C_1 + (\frac{r}{s(\theta)})^\alpha} \quad (2.4)$$

In our calculation, we approximate $s(\theta)$ by using a finite set of terms (first 10) of their Fourier expansions, which reduces the number of parameter to finitely many,

$$s(\theta) = a_0 + a_1 \cos(\theta) + b_1 \sin(\theta) + a_2 \cos(2\theta) + b_2 \sin(2\theta) + \dots + a_{10} \cos(10\theta) + b_{10} \sin(10\theta) \quad (2.5)$$

Here is the procedure we followed to fit the model to data,

1. Let α runs from 0.01 to 1 with step 0.01
2. For each α , we find the best C , a 's and b 's by regression in an appropriate region, calculate the mean-square error and make a record.

We then pick the α which minimize the mean-square error. In figure 2.25, the blue curves are for the model, which fit closely to the corresponding data curves. The KL distance between the model can be found in table 2.8 which are all small. This indicates that model 2.4 is an appropriate one.

| <i>Pairs</i> | \mathcal{H} | \mathcal{I} | a_{11} | a_{12} | a_{22} | α | $D(f \parallel f_{model1})$ |
|--------------|---------------|---------------|----------|-----------|----------|----------|-----------------------------|
| ChCv | -1.03 | 0.118 | 0.0582 | -0.000116 | 0.0469 | 0.47 | 0.0348 |
| ChCd | -1.51 | 0.119 | 0.0582 | -0.000179 | 0.0227 | 0.46 | 0.0579 |
| ChL | -0.981 | 0.206 | 0.0582 | 0.0198 | 0.0582 | 0.39 | 0.0615 |
| CvL | -1.18 | 0.123 | 0.0469 | -0.0076 | 0.0469 | 0.39 | 0.0855 |
| ChU | -0.896 | 0.121 | 0.0582 | -0.011 | 0.0582 | 0.43 | 0.0567 |
| ChPh | -0.931 | 0.229 | 0.0582 | 0.0203 | 0.0638 | 0.42 | 0.0259 |
| CdPd | -2.19 | 0.101 | 0.0228 | 0.00165 | 0.0212 | 0.46 | 0.0337 |
| CdDl | -2.09 | 0.0843 | 0.0227 | 0.00155 | 0.0227 | 0.43 | 0.0876 |

Table 2.8: Some constants for different pairs. a_{11}, a_{12}, a_{22} are the entries of the covariance matrix. The last column is the KL distance between model 2.4 and the data.

For the other approach, we apply the maximum entropy technique described in [23]. Let $\tilde{p}(x, y)$ be the empirical pdf of a pair of wavelet coefficients (X, Y) . Let \tilde{p}_{ξ_i} be the marginal distributions of ξ_i where $i = 1, 2, \dots, K$ and ξ can be any statistics of (X, Y) . We are seeking the *simplest* pdf $p(x, y)$ which has the same marginal distributions. As described by Zhu *et al.* [23], we have the following constrained optimization problem,

$$\begin{aligned} \hat{p}(x, y) &= \operatorname{argmax} H(p(x, y)) \\ \text{subject to } \hat{p}_{\xi_i}(x) &= \tilde{p}_{\xi_i}(x), \text{ for } i = 1, 2, \dots, K \end{aligned}$$

As it has been proved in [23], there exists a unique solution $p(x, y)$ to the above optimization problem, and it can be written in the following form:

$$\hat{p}(x, y) = e^{C + \sum_{i=1}^K f_i(\xi_i)} \quad (2.6)$$

To choose the statistics ξ_i , we observed that the level curves have polyhedra-like structures and all vertices lie on one of the two axes or one of the two diagonal lines. This suggests that we may be able to reconstruct the two dimensional distributions from the distributions along these special directions fairly well. Hence we will consider the following four statistics: $\xi_1 = X$, $\xi_2 = X - Y$, $\xi_3 = Y$ and $\xi_4 = X + Y$. We denote the marginal distributions of these statistics by p_{ξ_i} . With a little abuse of symbols, we also use ξ_i to represent the corresponding function of ξ_i w.r.t (x, y) , e.g. $\xi_4(x, y) = x + y$.

The following algorithm is used in our calculation of the model.

1. let p^{old} be the uniform distribution.
2. let

$$p^{new}(x, y) = C \prod_{i=1}^4 \left(\frac{\tilde{p}_{\xi_i}(\xi_i(x, y))}{p_{\xi_i}^{old}(\xi_i(x, y))} \right)^{w_i} p^{old}(x, y)$$

where w_i are some positive constants, and $\sum w_i = 1$. C is the constant that makes the right hand a probability distribution.

3. $p^{old} = p^{new}$
4. go to step 2.

We now show why the above procedure works.

Lemma 2.1 *Let p, p_1, p_2, \dots, p_n be any probability density functions, let $w_i > 0$, $i = 1 \dots n$ and $\sum_{i=1}^n w_i = 1$ Then*

$$D(p \| c \prod_{i=1}^n p_i^{w_i}) \leq \sum_{i=1}^n w_i D(p \| p_i)$$

where c makes the product a probability density function.

Proof.

$$D(p \| c \prod_{i=1}^n p_i^{w_i}) = \int p \log \frac{p}{c \prod_{i=1}^n p_i^{w_i}} = \sum_{i=1}^n w_i D(p \| p_i) - \log(c)$$

So we only have to prove that $c \geq 1$. Actually,

$$1 = c \int \prod_{i=1}^n p_i^{w_i}$$

Let $u_i = p_i^{w_i}$, and by the Hölder's inequality we have

$$\begin{aligned} \text{above} &= c \int \prod_{i=1}^n u_i \leq c \prod_{i=1}^n \left(\int u_i^{\frac{1}{w_i}} \right)^{w_i} \\ &= c \prod_{i=1}^n \left(\int p_i \right)^{w_i} = c \end{aligned}$$

Hence $c \geq 1$ ■

Lemma 2.2 $D(\tilde{p} \| p^{new}) \leq D(\tilde{p} \| p^{old}) - \sum_{i=1}^K w_i D(\tilde{p}_{\xi_i} \| p_{\xi_i}^{old})$
and equality holds iff $p^{new} = \hat{p}$

Proof. From the above lemma, we have

$$\begin{aligned} D(\tilde{p} \| p^{new}) &= D(\tilde{p} \| c \prod_{i=1}^n \left(\frac{\tilde{p}_{\xi_i}}{p_{\xi_i}^{old}} \right)^{w_i} p^{old}) \leq \sum_{i=1}^n D(\tilde{p} \| \frac{p_{\xi_i}}{p_{\xi_i}^{old}}) \\ &= \sum_{i=1}^n w_i \left(\int \tilde{p} \left(\log \frac{\tilde{p}}{p^{old}} - \log \frac{\tilde{p}_{\xi_i}}{p_{\xi_i}^{old}} \right) \right) = \sum_{i=1}^n w_i (D(\tilde{p} \| p^{old})) - \sum_{i=1}^n w_i (D(\tilde{p}_{\xi_i} \| p_{\xi_i}^{old})) \end{aligned}$$

where we have use the fact that $\frac{\tilde{p}_{\xi_i}}{p_{\xi_i}^{old}} p^{old}$ integrate to 1. ■

Let $p^{(n)}$ be the result at n th iteration.

Lemma 2.3 $D(\hat{p}_{\xi_i} \| p_{\xi_i}^{(n)}) \rightarrow 0$, as $n \rightarrow \infty$

Proof.

In the above lemma, the same conclusion holds if we substitute \tilde{p} by \hat{p} . Hence $D(\hat{p} \| p^{(n)})$ decreases to a number $c \geq 0$. It follows that $D(\hat{p}_{\xi_i} \| p_{\xi_i}^{(n)})$ tends to 0 as $n \rightarrow \infty$. ■

Proposition 2.1 $D(\hat{p} \| p^{(n)}) \rightarrow 0$, as $n \rightarrow \infty$

Proof. In our numerical calculations, we quantized each variable ξ_i into M bins $B_{i1}, B_{i2}, \dots, B_{iM}$. Let ϕ_{im} be the characteristic function of B_{im} ($\phi_{im}|_{B_{im}} = 1$, $\phi_{im}|_{B_{im}^c} = 0$). Then the marginal distribution of ξ_i is piecewise constant,

$$\tilde{P}_{\xi_i}(x) = \sum_{m=1}^M \tilde{h}_{im} \phi_{im}(x)$$

where $\tilde{h}_{im} = \frac{E_{\hat{p}}(\phi_{im}(\xi_i))}{\mu(B_{im})}$. Let $\varphi_{im} = \phi_{im}(\xi_i)$. We can now reformulate the optimization problem as

$$\begin{aligned} \hat{p}(x, y) &= \operatorname{argmax} H(p(x, y)) \\ \text{subject to } E_{\hat{p}}(\varphi_{im}) &= E_{\tilde{p}}(\varphi_{im}) \end{aligned}$$

By Lagrange multipliers,

$$\hat{p}(x, y) = e^{c + \sum_{k=1}^K \sum_{m=1}^M \hat{\alpha}_{km} \varphi_{km}(x, y)}$$

and $\hat{\alpha}$'s are chosen s.t.

$$E_{\hat{p}}(\varphi_{km}) = E_{\tilde{p}}(\varphi_{km})$$

In order to simplify our notation, we merge the (k, m) index pair to a single index $i = (k - 1)M + m$, and let $N = K \cdot M$. In general, let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ be a group of constants, and define

$$P_\alpha(x, y) = e^{c + \sum_{i=1}^N \alpha_i \varphi_i(x, y)}$$

where c is a constant that makes P_α a probability. Then

$$\begin{aligned} \frac{\partial}{\partial \alpha_i} (E_{P_\alpha}(\varphi_i)) &= \frac{\partial}{\partial \alpha_i} \int \int \varphi_i e^{c + \sum_{i=1}^N \alpha_i \varphi_i} dx dy \\ &= \int \int \varphi_i \varphi_j e^{c + \sum_{k=1}^N \alpha_k \varphi_k(x, y)} + \varphi_i e^{c + \sum_{k=1}^N \alpha_k \varphi_k(x, y)} \frac{\partial c}{\partial \alpha_i} dx dy \end{aligned}$$

Notice that

$$c = -\log\left(\int \int e^{\sum_{k=1}^N \alpha_k \varphi_k} dx dy\right)$$

and

$$\frac{\partial c}{\partial \alpha_i} = -\frac{1}{\int \int e^{\sum_{k=1}^N \alpha_k \varphi_k} dx dy} \int \int e^{\sum_{k=1}^N \alpha_k \varphi_k} \varphi_i dx dy = E_{P_\alpha}(\varphi_i)$$

which follows that,

$$\frac{\partial}{\partial \alpha_i} (E_{p_\alpha}(\varphi_i)) = E_{p_\alpha}(\varphi_i \varphi_j) - E_{p_\alpha}(\varphi_i) E_{p_\alpha}(\varphi_j) = \text{cov}_{p_\alpha}(\varphi_i, \varphi_j) \quad (2.7)$$

Now, let's consider the map $T : R^N \rightarrow R^N$

$$T(\alpha_1, \alpha_2, \dots, \alpha_N) = (E_{p_\alpha}(\varphi_1), E_{p_\alpha}(\varphi_2), \dots, E_{p_\alpha}(\varphi_N))$$

by equation 2.7, the Jacobian $D(T) = (\text{cov}_{p_\alpha}(\varphi_i, \varphi_j))$, where $i, j = 1, 2, \dots, N$. By definition of φ_i 's, they are linearly independent, hence the covariance matrix $(\text{cov}_{p_\alpha}(\varphi_i, \varphi_j))$ is positive definite. Hence $\det(D(T)) > 0$, which implies that $T : R^N \rightarrow T(R^N)$ is a homeomorphism.

By lemma 2.2,

$$\begin{aligned} T(\alpha_1^{(n)}, \alpha_2^{(n)}, \dots, \alpha_N^{(n)}) &= (E_{p^{(n)}}(\varphi_1), E_{p^{(n)}}(\varphi_1), \dots, E_{p^{(n)}}(\varphi_N)) \rightarrow \\ &= (E_{\hat{p}}(\varphi_1), E_{\hat{p}}(\varphi_1), \dots, E_{\hat{p}}(\varphi_N)) = T(\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_N) \end{aligned}$$

as $n \rightarrow \infty$, so

$$(\alpha_1^{(n)}, \alpha_2^{(n)}, \dots, \alpha_N^{(n)}) \rightarrow (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_N)$$

and $D(\hat{p}||p^{(n)}) \rightarrow 0$, as $n \rightarrow \infty$ ■

| <i>Pairs</i> | $D(\tilde{p} p)$ | $D(\tilde{p}_1 p_1)$ | $D(\tilde{p}_2 p_2)$ | $D(\tilde{p}_3 p_3)$ | $D(\tilde{p}_4 p_4)$ |
|--------------|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| ChCv | 0.0144 | 2.02e-08 | 1.05e-08 | 5.66e-09 | 1.71e-08 |
| ChCd | 0.0114 | 8.11e-07 | 2.59e-07 | 1.49e-07 | 3.12e-07 |
| ChL | 0.0138 | 4.02e-09 | 3.62e-09 | 3.35e-10 | 2.04e-09 |
| CvL | 0.0139 | 7.97e-08 | 3.28e-08 | 3.85e-08 | 5.34e-08 |

Table 2.9: Constants for the fits of the maximum entropy model to empirical distribution after 12 iterations. The second shows the KL distance between the 2D empirical distribution and the model, the last 4 columns shows the KL distance between the marginal of the empirical distribution and that of the model.

Figure 2.26 and figure 2.27 shows the level curves and the functions f_i 's of model 2.6.

Table 2.9 shows some constants associated to the fit, we can see from the second column that for all pairs, the KL distance between the data and the model is small. But there are significant difference between the level curves shown in figure 2.26 and those in the corresponding plots in figure 2.25. For example, in the ChCv pair, the model has strong vertices on the diagonal lines. This is not the case in the data. So we may conclude that the maximum entropy method based on marginal distributions do not predict the high dimensional distribution precisely.

2.4.3 2D Joint Statistics of Sowerby Images

We did the same thing on different categories of Sowerby Images. Figure 2.28–2.32 shows the results, constants associated to them are shown in table 2.10. We use the same contour levels in all these figures. There are some similarity between level curves of the same coefficient pairs calculated from different categories. The Man-made category show strong vertices in the level curves of the distribution of different coefficient pairs, which we believe correspond to some large scale structures in images. For the other categories (more noise like), there are some similarity between the shape of the level curves of the same coefficient pairs. It seems that any pair of coefficients (ξ_a, ξ_b) calculated from these noise like categories can be written as

$$(\xi_a, \xi_b) = \eta \cdot (\xi_{a_0}, \xi_{b_0}) \tag{2.8}$$

where η is some scalar random variable, which is decided by the category but independent of the choice of (ξ_a, ξ_b) , whereas (ξ_{a_0}, ξ_{b_0}) is independent of the category. Further discussions will show that (ξ_{a_0}, ξ_{b_0}) is related to the geometry aspect of the world, and η to the color or intensity of the world. (see section 3.1 for more details).

| <i>Category</i> | <i>Pairs</i> | \mathcal{H} | \mathcal{I} | a_{11} | a_{12} | a_{22} |
|-----------------|--------------|---------------|---------------|----------|-----------|----------|
| Sky | ChCv | -7.29 | 0.0828 | 0.00121 | -2.7e-06 | 0.00104 |
| | ChCd | -7.59 | 0.0804 | 0.00121 | 1.6e-05 | 0.000433 |
| | ChL | -7.32 | 0.169 | 0.00119 | 0.000519 | 0.00121 |
| | CvL | -7.32 | 0.0589 | 0.00103 | -0.000109 | 0.00104 |
| Vegetation | ChCv | 0.145 | 0.0782 | 0.0866 | 1.69e-05 | 0.0806 |
| | ChCd | -0.533 | 0.092 | 0.0866 | -9.49e-05 | 0.0331 |
| | ChL | 0.147 | 0.132 | 0.0867 | 0.0272 | 0.0866 |
| | CvL | 0.0989 | 0.0716 | 0.0806 | -0.015 | 0.0806 |
| Road | ChCv | -2.11 | 0.0468 | 0.0225 | 1.55e-05 | 0.0121 |
| | ChCd | -2.41 | 0.0402 | 0.0225 | 2.24e-06 | 0.00764 |
| | ChL | -1.8 | 0.138 | 0.0225 | 0.00788 | 0.0225 |
| | CvL | -2.54 | 0.0626 | 0.0121 | -0.00236 | 0.0121 |
| Man-made | ChCv | -0.762 | 0.0811 | 0.0947 | 0.000332 | 0.0742 |
| | ChCd | -1.85 | 0.14 | 0.0947 | 0.000179 | 0.0108 |
| | ChL | -1.01 | 0.656 | 0.0941 | 0.069 | 0.0947 |
| | CvL | -1.16 | 0.15 | 0.0746 | -0.0122 | 0.0742 |
| Overall | ChCv | -1.12 | 0.151 | 0.0563 | -4.81e-06 | 0.0459 |
| | ChCd | -1.7 | 0.153 | 0.0563 | -3.41e-05 | 0.0174 |
| | ChL | -1.03 | 0.277 | 0.0563 | 0.0234 | 0.0563 |
| | CvL | -1.35 | 0.163 | 0.046 | -0.008 | 0.0459 |

Table 2.10: Some constants for different Haar Wavelet coefficient pairs calculated from Sowerby database. a_{11}, a_{12}, a_{22} are the entries of the covariance matrix.

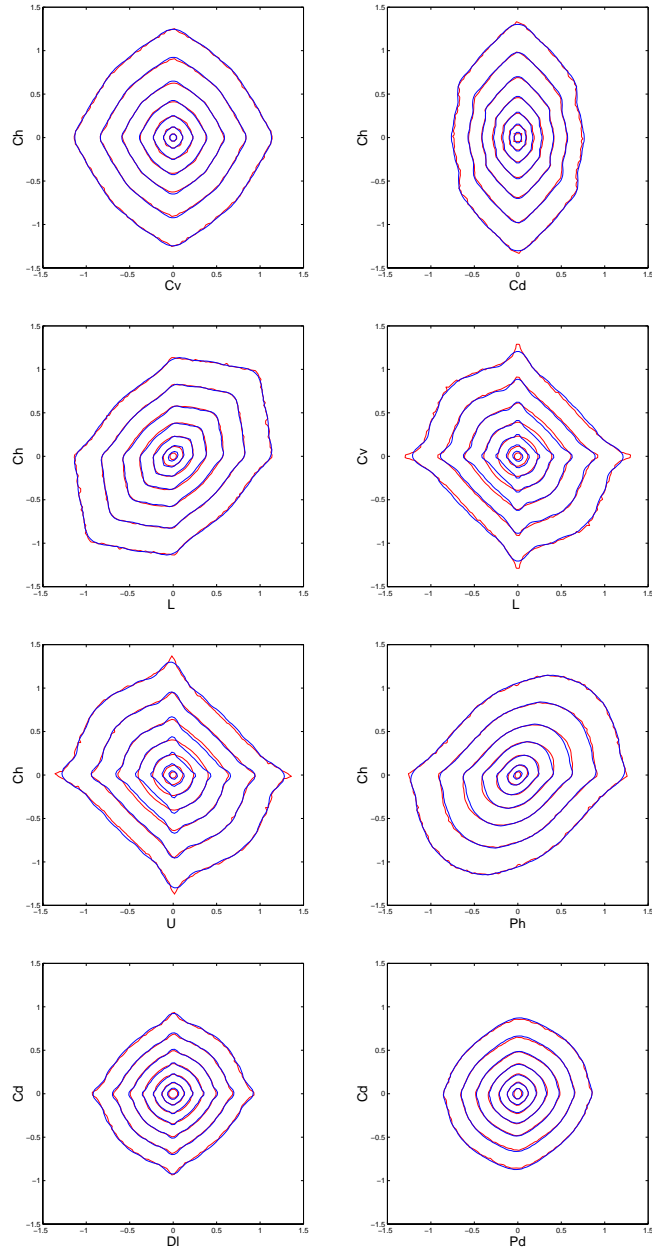


Figure 2.25: Joint Histogram of Haar Wavelet Coefficients calculated from van Hateran Database. Red: the level curves of the empirical distribution. Blue: the model 2.4. Ch,Cv,Cd stand for horizontal, vertical and diagonal components respectively; L stands for left brother; Ph, Pd for horizontal parent and diagonal parent; Dl for upper-left brother.

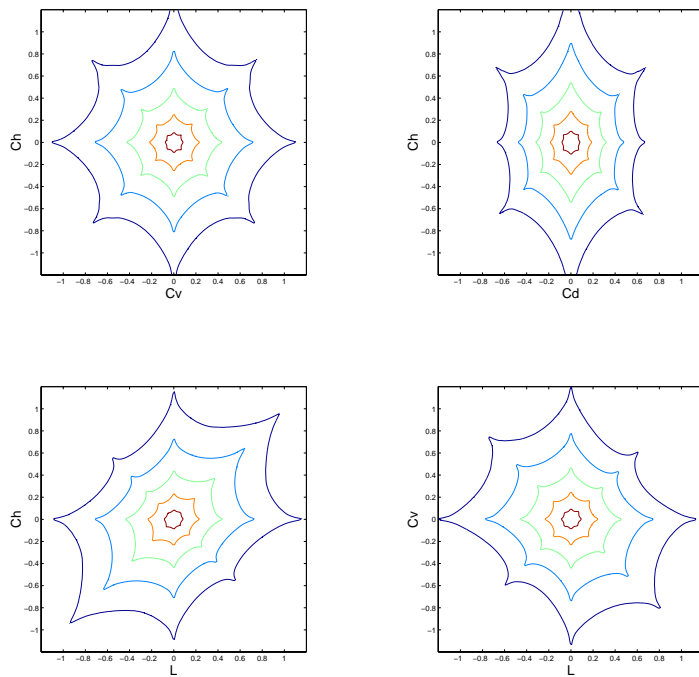


Figure 2.26: Maximum entropy model with the same marginals along axes and diagonal lines

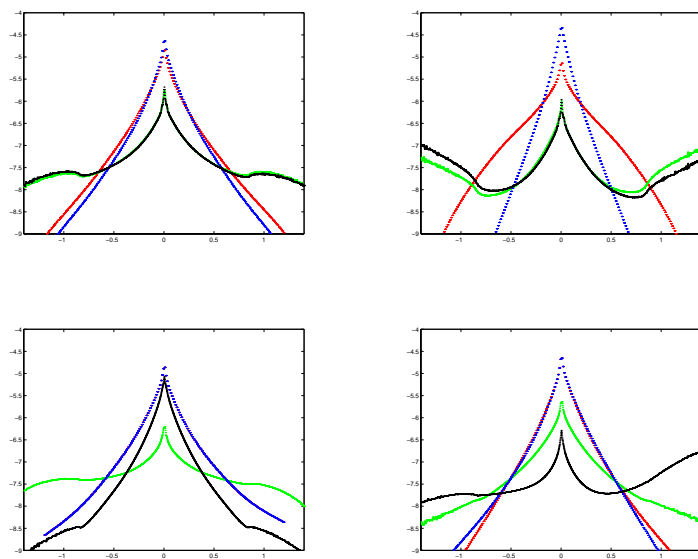


Figure 2.27: Function f_i 's

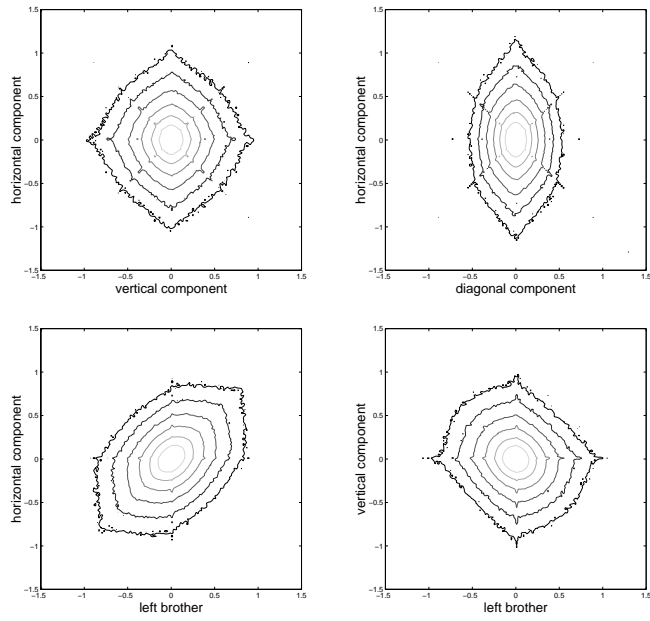


Figure 2.28: Joint Histogram of Haar Wavelet Coefficients calculated from Sowerby Images

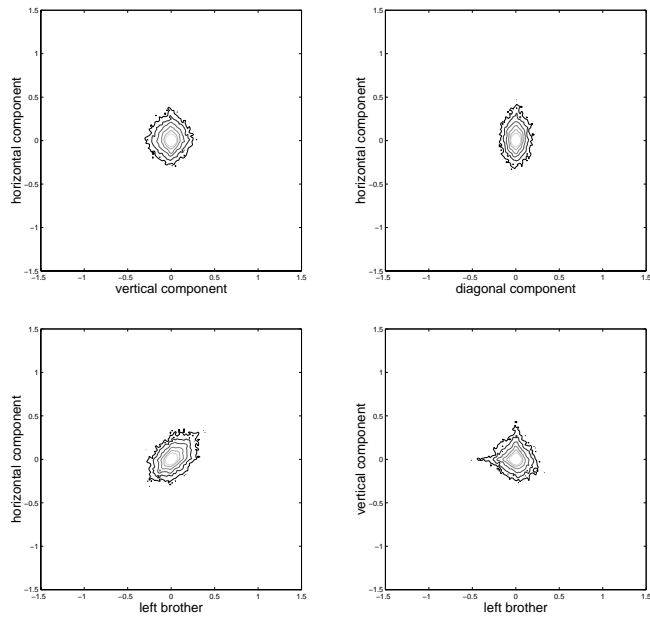


Figure 2.29: Joint Histogram of Haar Wavelet Coefficients calculated from the 'sky' category of Sowerby Images

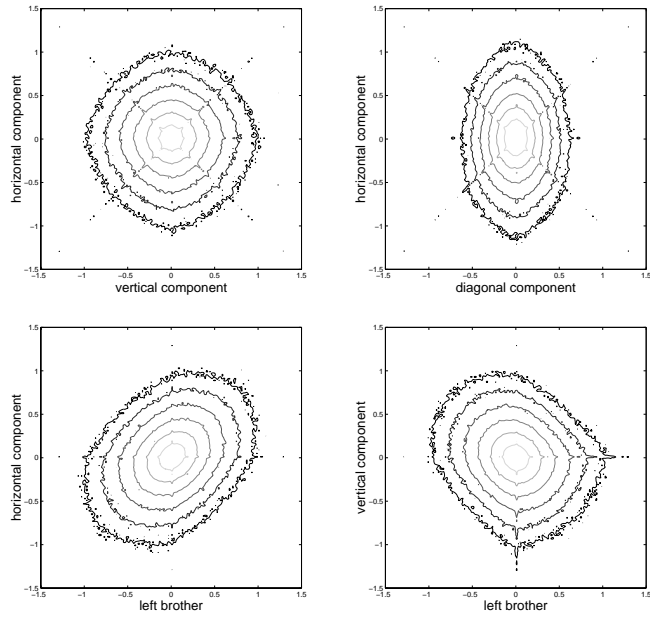


Figure 2.30: Joint Histogram of Haar Wavelet Coefficients calculated from the ‘vegetation’ category of Sowerby Images

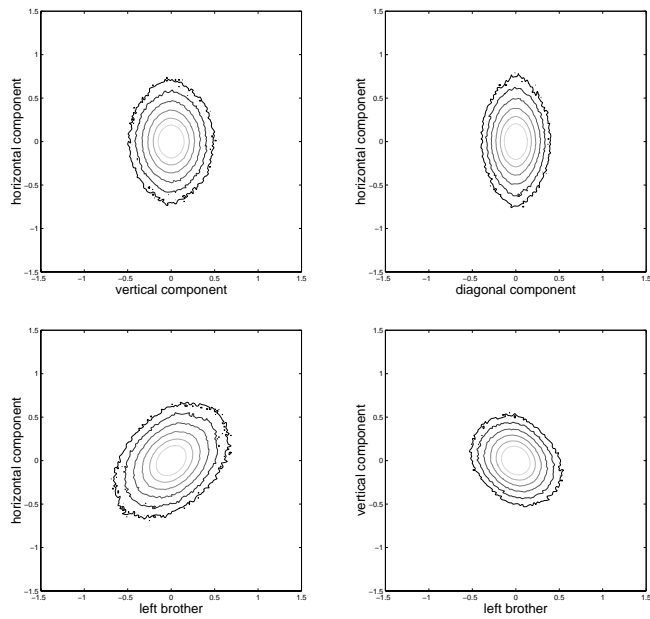


Figure 2.31: Joint Histogram of Haar Wavelet Coefficients calculated from the ‘road’ category of Sowerby Images

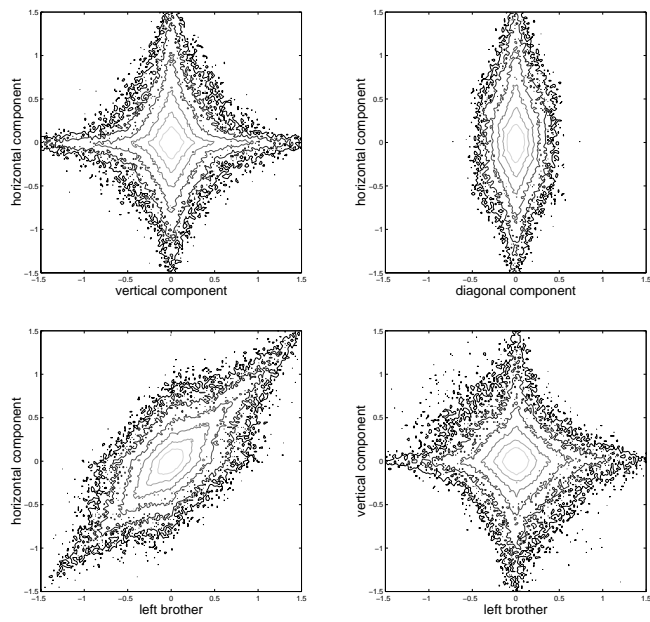


Figure 2.32: Joint Histogram of Haar Wavelet Coefficients calculated from the ‘man made’ category of Sowerby Images

2.4.4 3D Joint Statistics for Haar Wavelet Coefficients

Here we just present the joint statistics for the triple: horizontal component, its vertical cousin and diagonal cousin. Figure 2.33 shows a level surface(counterpart of a single level curve in Figure 2.25) viewed from different angles. We can see more structures here. There are four corners on the ‘horizontal’ and ‘vertical’ axes, which imply that the probability density heavily concentrates on these two axis. We will explain what kind of image patches correspond to these corners in section 2.4.5.

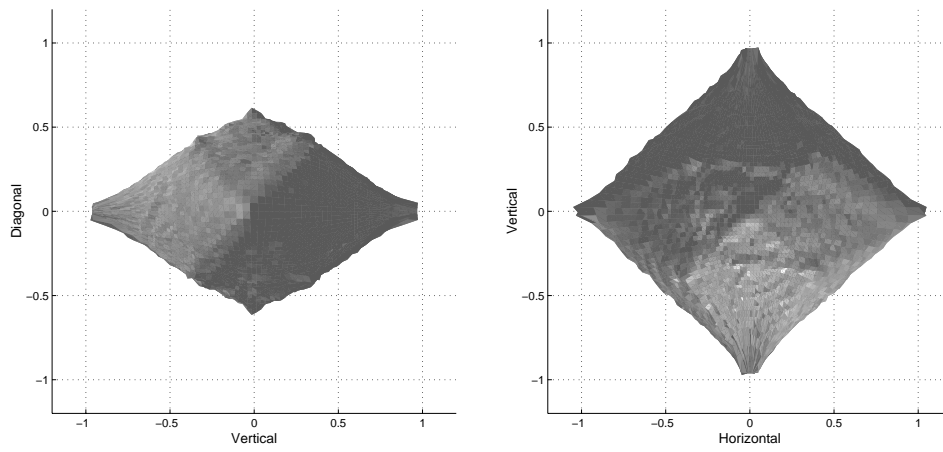
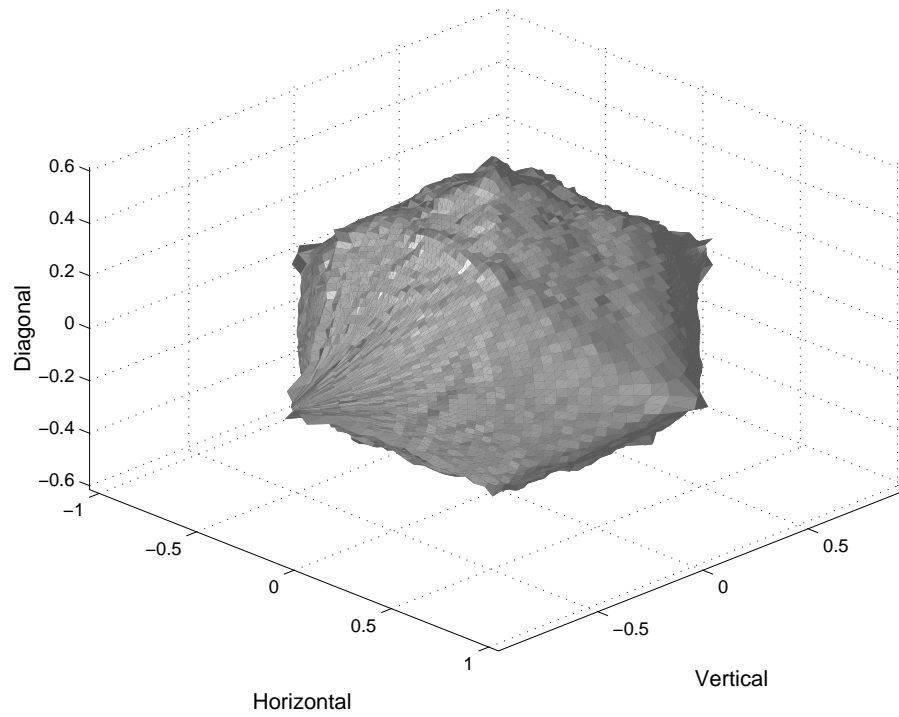


Figure 2.33: An equi-surface of 3d joint histogram of horizontal component, its vertical cousin and diagonal cousin, viewed from three different angles

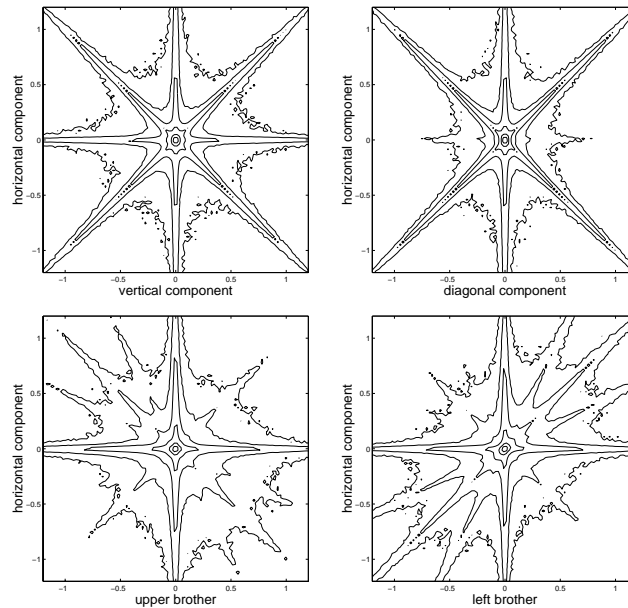


Figure 2.34: Contour plots of the $\log(\text{pdf})$ of wavelet coefficient pairs for range images.

2.4.5 Statistics for Haar Wavelet from range images

We calculated similar joint distributions from range images. Figure 2.34 shows contour plots of the joint density functions for different wavelet coefficient pairs. We observe strong cusps in all contour level curves; most of which lie on the lines $x = 0$, $y = 0$, $x = \pm y$ and $x = \pm 2y$. As a comparison, look at Figure 2.25 where the corresponding contour plots are calculated for optical images in van Hateren image database. We see that many cusps occur along the same lines, but are not as peaked.

Furthermore, we did the following experiments on the range images:

1. We scaled the images down by taking the block minimum and calculated the joint histograms. Figure 12 shows the result.
2. We scaled the images down by taking the block average and calculated the joint histograms, Figure 13 shows the result.

When we scale a range image down by taking block minimum, the resulting image is

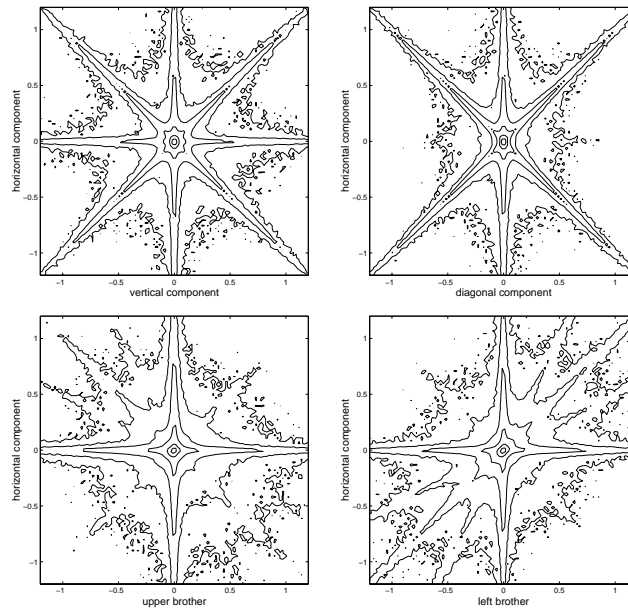


Figure 2.35: Contour plots of the log histogram of wavelet coefficient pairs, calculated from range images scaled down by taking the minimum of 2×2 blocks.

approximately the same as the image taken in the same environment at half resolution (as mentioned in section 2.2.3, range images are roughly scale invariant under block minimum). The second method (block averaging) is, however, closer to how a digital camera for optical images works: The intensity at each pixel is the average (or some weighted mean) of the intensity at points covered by that pixel. This explains why the statistics shown in Figure 2.35 is similar to that in Figure 2.34, and the statistics in Figure 2.36 is similar to that in Figure 2.25. We can do the experiment the other way around. For each intensity image, we scaled it down by taking the max or min of each block, whichever closer to the mean. This way, every pixel in the resulting image is closer to the point intensity instead of average of blocks. Figure 2.4.5 shows the joint distributions of wavelet coefficients, which look like those calculated from range images.

Note that the observed cusps in Figure 2.34 are not caused by noise, but correspond to some special structures in the images. It is important to see what these structures are:

For the cousin pairs (horizontal and vertical; horizontal and diagonal) this is relatively straight-forward — because the joint distribution of horizontal, vertical and diagonal wavelet

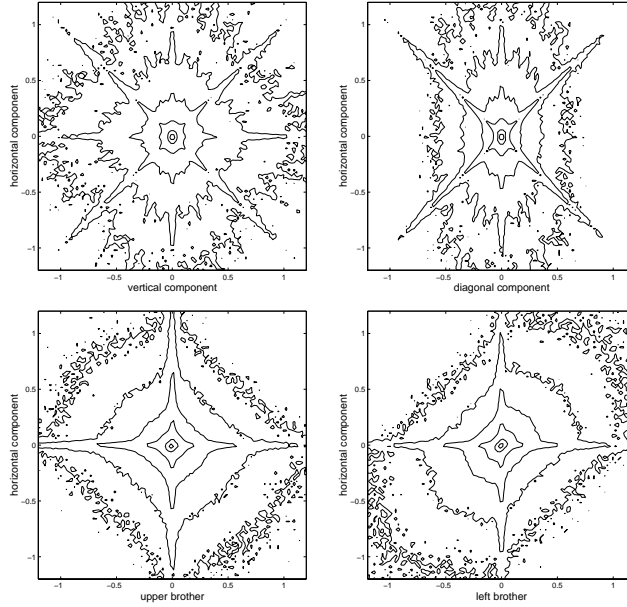


Figure 2.36: Contour plots of the log histogram of wavelet coefficient pairs, calculated from range images scaled down by taking the average of 2×2 blocks.

coefficients is a sufficient statistics for 2×2 blocks modulo mean (3 variables). Figure 2.38 shows an equiprobable surface of the 3D joint distribution (horizontal, vertical, diagonal). We see vertices along the lines $y = z = 0$, $x = z = 0$, $x = y = 0$ and $x = \pm y = \pm z$. Simple calculations show that these vertices correspond to the following 2×2 blocks and their rotations:

$$\begin{pmatrix} a & a \\ b & b \end{pmatrix}, \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \begin{pmatrix} a & b \\ b & b \end{pmatrix}$$

Notice that the bar-like patch (on the line $x = y = 0$) happens less frequently than the other types of patches. This can be seen from the lower-left plot of figure 2.38

For the *horizontal filter and left brother* pair (see lower right subplot of figure 2.34), we see cusps along $y = 0$, $x = 0$, $y = x$, $y = 2x$ and $y = \frac{1}{2}x$. To find the 2×4 patches that correspond to these vertices, we sample randomly from our database for patches with strong filter reactions along these directions. The 8 range values in the typical patches fall clearly

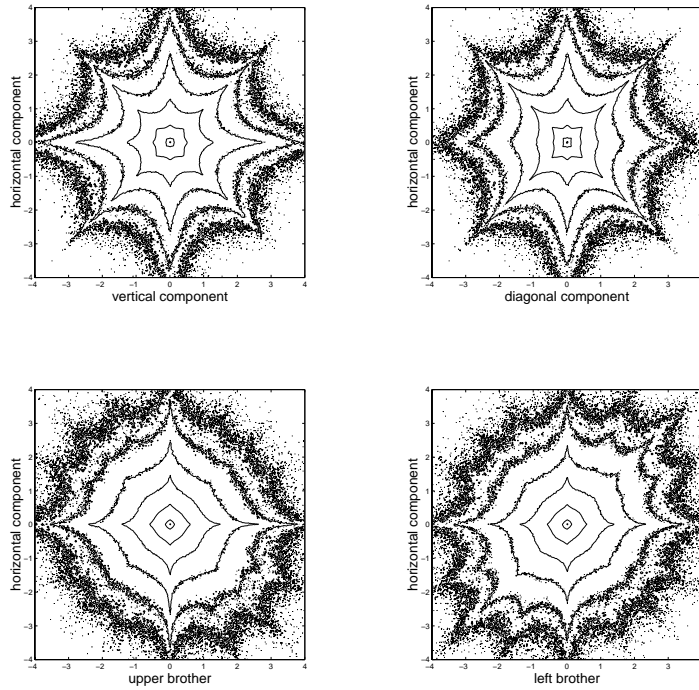


Figure 2.37: Contour plots of the log histogram of wavelet coefficient pairs, calculated from intensity images scaled down by taking the max or min, whichever closer to the mean

into 2 and occasionally 3 tight clusters. Calling the clusters a,b, and c, we get for example:

direction

typical patches

$$y = 0 \quad \begin{pmatrix} a & b & b & c \\ b & b & b & c \end{pmatrix}, \begin{pmatrix} a & a & b & c \\ b & b & b & c \end{pmatrix}, \begin{pmatrix} a & b & a & a \\ a & a & a & a \end{pmatrix}$$

$$y = x \quad \begin{pmatrix} a & b & b & b \\ a & a & a & b \end{pmatrix}, \begin{pmatrix} a & b & b & a \\ a & a & a & a \end{pmatrix}, \begin{pmatrix} a & a & a & a \\ b & b & b & b \end{pmatrix}, \begin{pmatrix} a & a & a & a \\ b & b & b & b \end{pmatrix},$$

$$y = \frac{1}{2}x \quad \begin{pmatrix} a & a & a & b \\ b & b & b & b \end{pmatrix}$$

For the *horizontal filter and upper brother* pair(see the lower left subplot of figure 2.34),

we see cusps along $y = 0$, $x = 0$, $y = \pm x$, $y = -2x$ and $y = -\frac{1}{2}x$. The typical patches in the database, that lead to high filter reactions along these lines are:

direction typical patches

$$y = 0 \quad \begin{pmatrix} a & a \\ b & b \\ b & b \\ b & b \end{pmatrix}$$

$$y = -x \quad \begin{pmatrix} a & a \\ b & a \\ b & a \\ a & a \end{pmatrix}, \quad \begin{pmatrix} a & a \\ b & a \\ b & a \\ a & a \end{pmatrix}, \quad \begin{pmatrix} a & a \\ b & b \\ b & b \\ a & a \end{pmatrix},$$

$$\begin{pmatrix} a & b \\ b & b \\ b & b \\ b & a \end{pmatrix}$$

$$y = -2x \quad \begin{pmatrix} a & b \\ b & b \\ b & b \\ a & a \end{pmatrix}, \quad \begin{pmatrix} a & a \\ b & a \\ b & b \\ a & a \end{pmatrix}$$

$$y = x \quad \begin{pmatrix} a & a \\ b & b \\ a & a \\ b & b \end{pmatrix}, \quad \begin{pmatrix} a & a \\ a & b \\ a & b \\ b & b \end{pmatrix}, \quad \begin{pmatrix} a & b \\ b & b \\ a & a \\ a & b \end{pmatrix}$$

We see that all the striking cusps in the contour plots in the Haar wavelet domain relate to the piecewise constant geometric structure in range images.

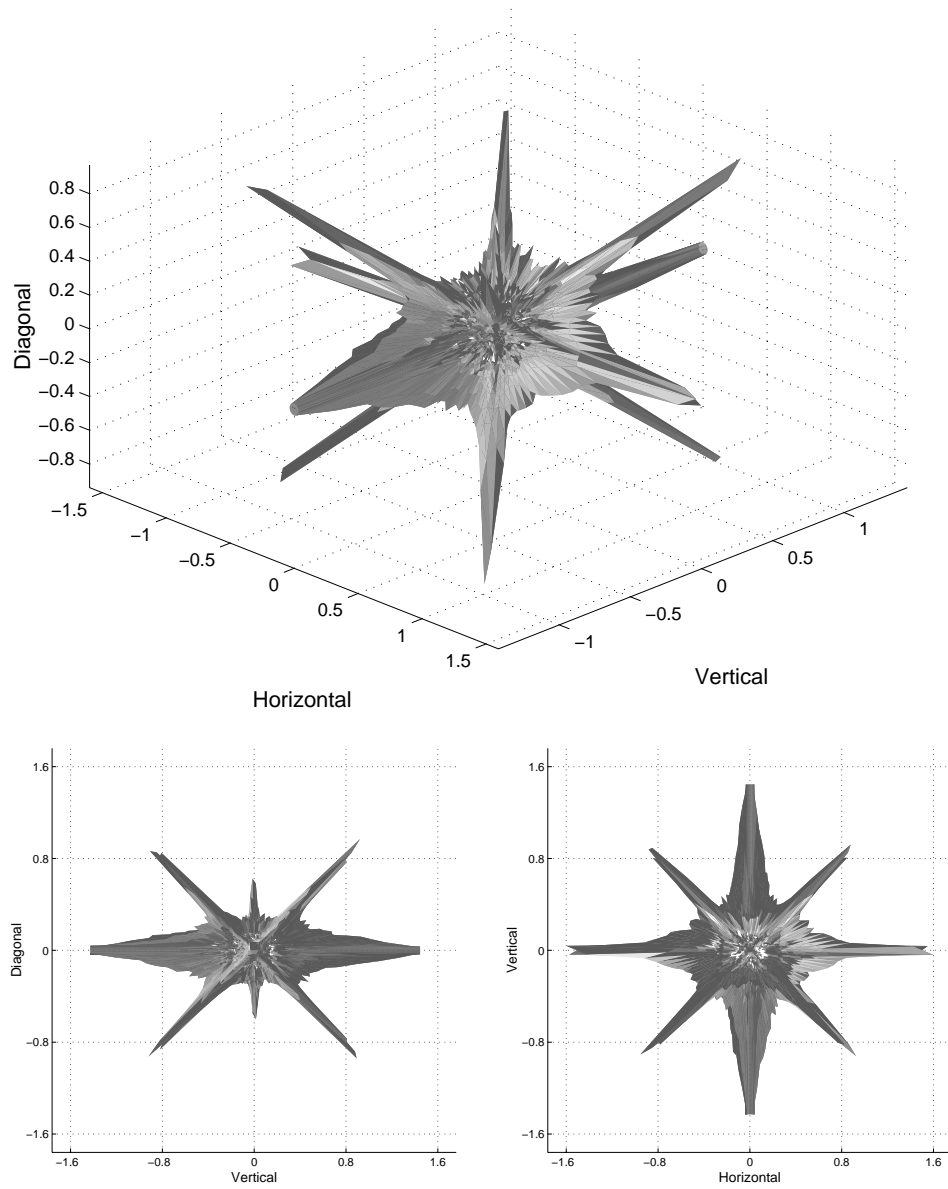


Figure 2.38: An equi-surface of a 3D joint histogram of horizontal, vertical and diagonal wavelet coefficients in range images, viewed from three different angles.

2.5 Joint Statistics for Other Wavelet Coefficients

Since Haar filters are not well localized in the frequency domain, they may get block effects which is perceptually undesirable. In applications such as noisy removal and image restoration, more sophisticated filters are often used instead. Here we studied the joint statistics for the coefficients of two types of popular wavelets.

2.5.1 Joint Statistics for QMF

Buccigrossi *et. al* [3] utilized a recursive pyramid decomposition based on separable Quadrature Mirror Filters(QMFs) in image compression. Figure 2.39 shows the horizontal, vertical, diagonal and low pass filters. Details can be found in [19]. Figure 2.40 shows empirical joint distribution of different QMF filter pairs. The horizontal-vertical coefficient pair of QMF and that of Haar filters(see 2.25) have very similar distributions. But the contour plots of the other three pairs show different level curves, especially the horizontal-diagonal pair.

2.5.2 Joint Statistics for Steerable Pyramids

Steerable pyramid decomposition was developed by Freeman and Adelson [8], which is quite different from the Haar Wavelet and QMF Wavelet. The following paragraph is a simple introduction(Copied from the manual of EPWIC package, which is available at <http://www.cns.nyu.edu/~eero/>). For more details, see [20].

The steerable pyramid is a multi-scale representation that is translation-invariant, but that also includes representation of orientation. Furthermore, the representation of orientation is designed to be rotation-invariant. The basis/projection functions are oriented (steerable) filters, localized in space and frequency. It is over-complete to avoid aliasing. And it is "self-inverting" (like the QMF/Wavelet transform).

The system diagram for the steerable pyramid is as follows:

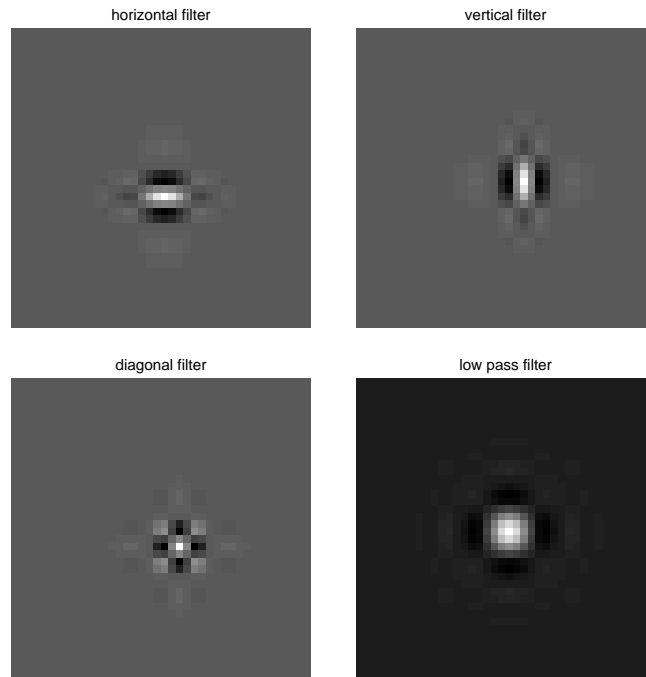
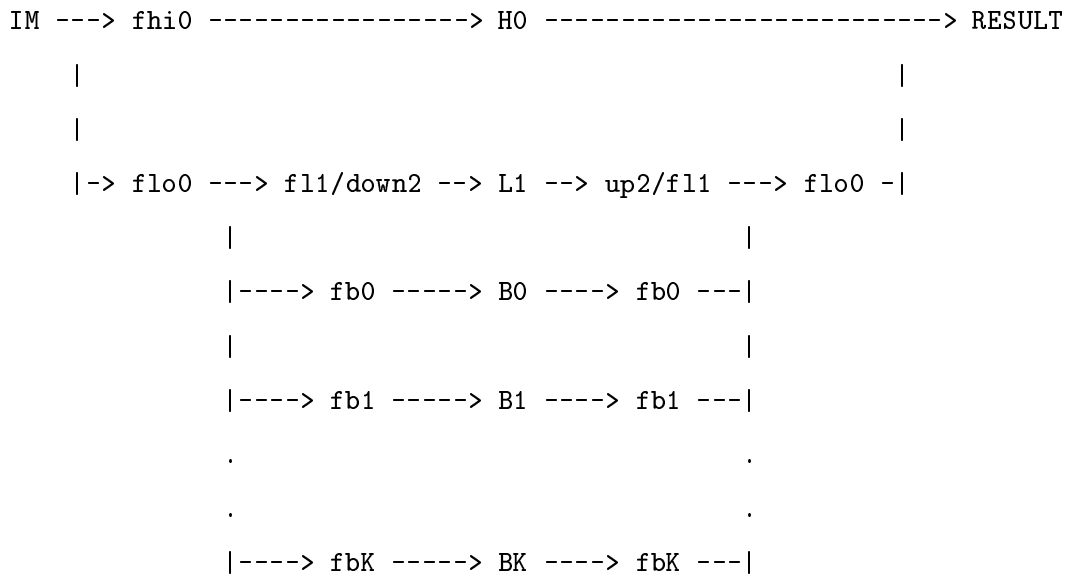


Figure 2.39: QMF Filters(9-tap)



The filters {fhi0,flo0} are used to initially split the image into

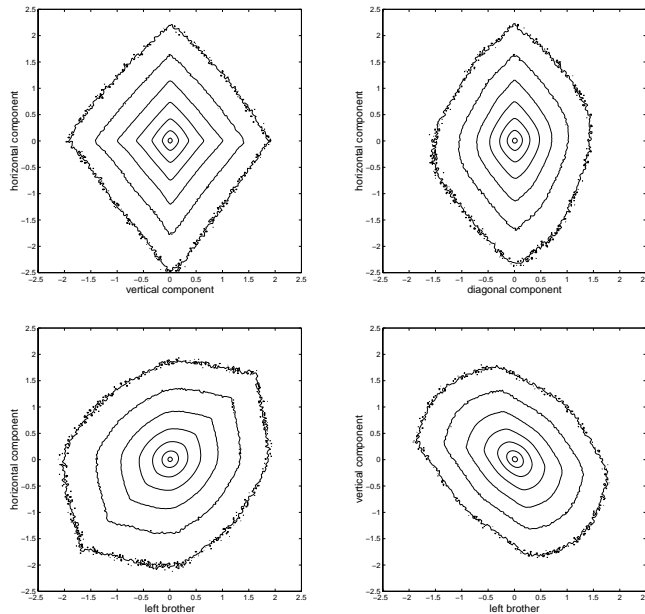


Figure 2.40: Joint Histogram of QMF Wavelet Coefficients calculated from natural images

a highpass residual band H_0 and a lowpass subband. This lowpass band is then split into a low(er)pass band L_1 and $K+1$ oriented subbands $\{B_0, B_1, \dots, B_K\}$. The representation is substantially overcomplete. The pyramid is built by recursively splitting the lowpass band (L_1) using the inner portion of the diagram (i.e., using the filters $\{f_{l1}, f_{b0}, f_{b1}, \dots, f_{bK}\}$).

We use the 2-orientation band filters ($K = 1$ in the above diagram) The filters f_{b0} and f_{b1} are shown in figure 2.41. Figure 2.42 shows the empirical distributions of different coefficient pairs. We see that for all the pairs, the level curves around the origin are nearly elliptic, especially those for the horizontal and vertical pair where they are almost circular.

2.5.3 An interesting phenomenon

In [3], the authors observed an interesting phenomena in the joint histograms of the QMF wavelet coefficients of images. Let h_p represents horizontal coefficient at parent level (we normalized both h_c , h_p by their L_1 norm). Hence the variance of the coefficients should

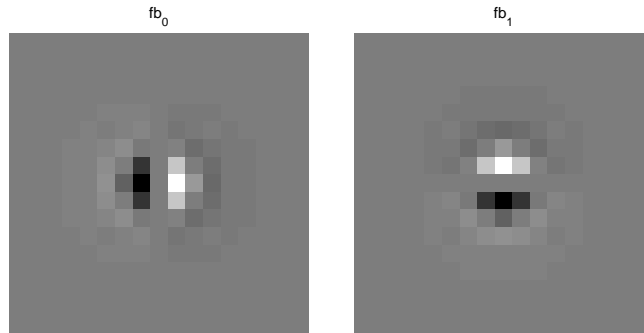


Figure 2.41: 2-orientation steerable pyramid filters

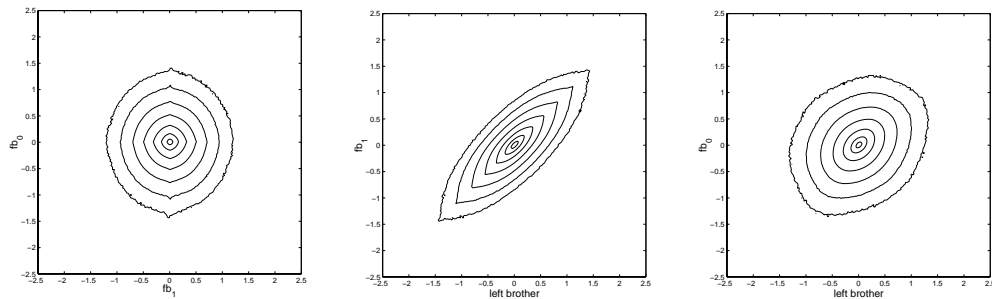


Figure 2.42: Joint Histogram of Steerable Pyramid Coefficients calculated from natural images

be the same, if the images are scale invariant), and hc represents horizontal coefficient at child level, they observed that the conditional histogram $H(\log|hc| \mid \log|hp|)$ has a shape shown in Figure 2.43 (we will explain how we get this figure next). We tried to explain this phenomena. First, figure 2.44 shows the joint distribution of hc and hp . We found that the shape of level curves of joint histogram is close to a circle. So our general model 2.4 may reduce to:

$$f(hc, hp) = C_1 e^{-C_2 r^\alpha} \quad (2.9)$$

where $r = \sqrt{hc^2 + hp^2}$, and C_1, C_2, α are just constants, independent of θ . We found that this model fits fairly well to the data for $C_2 = 1$ and $\alpha = 0.5$. From the phenomena they observed, the authors concluded in [3] that, the conditional expectation, $E(|hc| \mid |hp|)$ is approximately proportional to P , and derive a very simple linear predictor from it. For

the left part, they suspected these low-amplitude coefficients are dominated by quantization and other noise sources.

This phenomena can be actually explained mathematically, if the joint distribution of the two variables are in the form of (2.9). Since $f(hc, hp)$ is symmetric, the density function for $(|hc|, |hp|)$ should have the same expression, only that C_1 is different. So we just assume $hc > 0, hp > 0$. Let $x = \log(hc)$ and $y = \log(hp)$, then the joint density function for x, y is:

$$g(x, y) = C_1 e^x e^y e^{-C_2 r^\alpha}$$

where $r = \sqrt{e^{2x} + e^{2y}}$. The conditional density function will be:

$$g(x|y) = C(y) e^x e^{-C_2 r^\alpha}$$

Figure 2.43 shows the numerically calculated $g(x | y)$. Next, we explain analytically this phenomenon. For fixed y , let's find x which maximize $g(x | y)$. Set the derivatives of $g(x | y)$ to zero, and substitute C_2 by 1, we get the equation:

$$e^{2y} = e^{2x} (\alpha^{\frac{1}{1-\alpha/2}} e^{2x \frac{\alpha}{2-\alpha}} - 1) \quad (2.10)$$

From this equation, it is easy to see that:

$$\begin{aligned} x &\approx (1 - \frac{\alpha}{2})y - \frac{1}{2} \log \alpha, & \text{if } y \rightarrow \infty \\ x &\rightarrow -\frac{1}{\alpha} \log \alpha, & \text{if } y \rightarrow -\infty. \end{aligned}$$

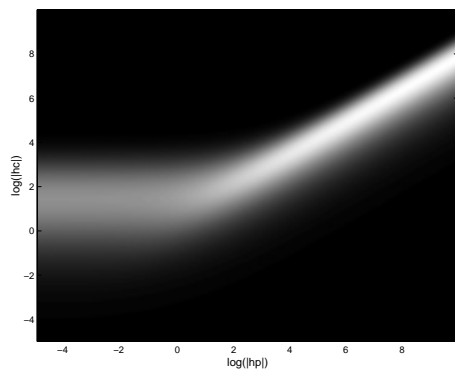


Figure 2.43: The conditional histogram $H(\log|hc| \mid \log|hp|)$ calculated from our model. Bright parts mean high values

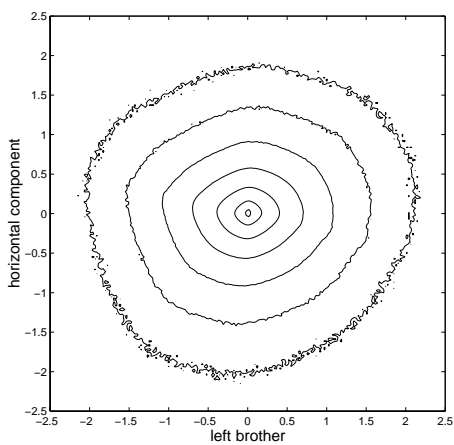


Figure 2.44: Joint Histogram of QMF Wavelet Coefficients calculated from natural images

2.6 DCT Coefficients

DCT-based method has been used in JPEG image compression standard. In this section, we study the statistics of DCT coefficients. First, we review the definition of (2 dimensional) DCT. Assume A is a $M \times N$ Matrix, with entries A_{mn} for $0 \leq m \leq M-1$ and $0 \leq n \leq N-1$. Then the discrete cosine transformation of A is,

$$B_{uv} = \alpha_u \alpha_v \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)u}{2M} \cos \frac{\pi(2n+1)v}{2N}$$

for $0 \leq u \leq M-1, 0 \leq v \leq N-1$ where

$$\alpha_u = \begin{cases} 1/\sqrt{M}, & u = 0 \\ \sqrt{2/M}, & 1 \leq u \leq M-1 \end{cases} \quad \alpha_v = \begin{cases} 1/\sqrt{N}, & v = 0 \\ \sqrt{2/N}, & 1 \leq v \leq M-1 \end{cases}$$

As in standard JPEG, we work on 8×8 blocks, i.e. $M = N = 8$. Figure 2.45 shows the empirical distribution of all the DCT coefficients. B_{00} is the DC coefficient, the others are AC coefficients. Since DC coefficient is a measure of the average value of the 8×8 block, its distribution (shown in the first subplot of figure 2.45) is similar to that of single pixel statistics (with different standard deviation). All the AC coefficients are the filter reactions of mean-0 filters, and their distributions can also be described by generalized Laplace model. An interesting fact here is that, although the coefficients have different variance, they all have quite similar shapes. Table 2.11 shows constants associated to these coefficients, we can see that the kurtoses of all the DCT coefficients are close to each other. As formula 2.2 suggests, the kurtosis decides the constant α in the generalized Laplace distribution, which in turn decides the shape of the $\log(\text{pdf})$. Another interesting phenomenon we observed here is that the DC terms with the same $u+v$ value have similar standard deviations. Figure 2.46 shows the contour plot of the standard deviations of DCT coefficients, with respect to u and v . We can see that the contour lines are roughly parallel to the $u+v = \text{constant}$ lines.

| u | v | σ | \mathcal{S} | κ | \mathcal{H} | α_1 | $D(p_0 \parallel p_1)$ | α_2 | $D(p_0 \parallel p_2)$ |
|-----|-----|----------|---------------|----------|---------------|------------|------------------------|------------|------------------------|
| 0 | 0 | 5.54 | 0.256 | 4.83 | 4.44 | 1.14 | 0.00905 | 1.18 | 0.00934 |
| 0 | 1 | 1.06 | -0.014 | 9.66 | 1.87 | 0.729 | 0.00254 | 0.75 | 0.00273 |
| 0 | 2 | 0.671 | -0.133 | 10.3 | 1.18 | 0.679 | 0.00582 | 0.725 | 0.00677 |
| 0 | 3 | 0.465 | -0.0171 | 11 | 0.645 | 0.675 | 0.00454 | 0.703 | 0.0049 |
| 0 | 4 | 0.348 | -0.0481 | 11.3 | 0.217 | 0.669 | 0.00441 | 0.693 | 0.00467 |
| 0 | 5 | 0.27 | -0.0342 | 12.1 | -0.162 | 0.66 | 0.00315 | 0.67 | 0.00321 |
| 0 | 6 | 0.218 | -0.00888 | 13.3 | -0.492 | 0.65 | 0.00245 | 0.643 | 0.00255 |
| 0 | 7 | 0.185 | -0.0192 | 14.5 | -0.763 | 0.626 | 0.00189 | 0.619 | 0.00198 |
| 1 | 0 | 1.07 | 0.0108 | 10.7 | 1.83 | 0.673 | 0.00366 | 0.711 | 0.00443 |
| 1 | 1 | 0.673 | 0.00351 | 9.29 | 1.21 | 0.689 | 0.00978 | 0.765 | 0.0123 |
| 1 | 2 | 0.505 | 9.73e-05 | 9.51 | 0.783 | 0.681 | 0.00958 | 0.756 | 0.012 |
| 1 | 3 | 0.38 | 0.000789 | 9.77 | 0.372 | 0.68 | 0.00803 | 0.745 | 0.00992 |
| 1 | 4 | 0.296 | -0.00122 | 10.1 | 0.0086 | 0.682 | 0.00619 | 0.733 | 0.00735 |
| 1 | 5 | 0.236 | -0.00475 | 10.5 | -0.325 | 0.678 | 0.00425 | 0.719 | 0.00503 |
| 1 | 6 | 0.195 | -0.00177 | 11.3 | -0.621 | 0.669 | 0.00259 | 0.693 | 0.00289 |
| 1 | 7 | 0.169 | -0.0016 | 12.5 | -0.855 | 0.649 | 0.00163 | 0.662 | 0.00171 |
| 2 | 0 | 0.672 | -0.207 | 11.4 | 1.14 | 0.639 | 0.00953 | 0.69 | 0.0109 |
| 2 | 1 | 0.497 | 0.00392 | 9.87 | 0.753 | 0.674 | 0.0107 | 0.741 | 0.0127 |
| 2 | 2 | 0.403 | 0.0436 | 9.97 | 0.45 | 0.673 | 0.0096 | 0.737 | 0.0114 |
| 2 | 3 | 0.323 | 0.00598 | 10.1 | 0.126 | 0.674 | 0.00768 | 0.733 | 0.00925 |
| 2 | 4 | 0.26 | 0.0141 | 10.3 | -0.183 | 0.679 | 0.00568 | 0.726 | 0.0067 |
| 2 | 5 | 0.213 | 0.00668 | 10.6 | -0.476 | 0.678 | 0.00378 | 0.716 | 0.00446 |
| 2 | 6 | 0.179 | 0.0142 | 11.2 | -0.739 | 0.674 | 0.00224 | 0.697 | 0.0025 |
| 2 | 7 | 0.158 | 0.0077 | 12.3 | -0.947 | 0.658 | 0.00138 | 0.666 | 0.00141 |
| 3 | 0 | 0.456 | 0.00171 | 12.2 | 0.567 | 0.635 | 0.00654 | 0.669 | 0.0072 |
| 3 | 1 | 0.365 | 0.00243 | 10.3 | 0.297 | 0.666 | 0.00865 | 0.724 | 0.0103 |
| 3 | 2 | 0.313 | -0.00501 | 10.4 | 0.0763 | 0.667 | 0.00783 | 0.722 | 0.00923 |
| 3 | 3 | 0.263 | 0.00264 | 10.4 | -0.172 | 0.672 | 0.0061 | 0.72 | 0.00718 |
| 3 | 4 | 0.221 | -0.00739 | 10.5 | -0.423 | 0.679 | 0.0044 | 0.717 | 0.00507 |
| 3 | 5 | 0.186 | 0.00244 | 10.7 | -0.67 | 0.683 | 0.00288 | 0.712 | 0.00327 |
| 3 | 6 | 0.16 | 0.00268 | 11.2 | -0.894 | 0.681 | 0.00171 | 0.696 | 0.00182 |
| 3 | 7 | 0.144 | -0.00142 | 12.2 | -1.07 | 0.668 | 0.00113 | 0.668 | 0.00114 |

Table 2.11: constants of DCT coefficients(part 1)

| u | v | σ | \mathcal{S} | κ | \mathcal{H} | α_1 | $D(p_0 \parallel p_1)$ | α_2 | $D(p_0 \parallel p_2)$ |
|-----|-----|----------|---------------|----------|---------------|------------|------------------------|------------|------------------------|
| 4 | 0 | 0.331 | -0.0861 | 12.7 | 0.0956 | 0.628 | 0.00555 | 0.657 | 0.00603 |
| 4 | 1 | 0.277 | -0.000888 | 10.8 | -0.113 | 0.66 | 0.00606 | 0.71 | 0.00729 |
| 4 | 2 | 0.245 | 0.0238 | 10.7 | -0.286 | 0.665 | 0.00546 | 0.711 | 0.00649 |
| 4 | 3 | 0.213 | 0.00406 | 10.7 | -0.482 | 0.673 | 0.00417 | 0.711 | 0.00487 |
| 4 | 4 | 0.184 | 0.0122 | 10.8 | -0.685 | 0.682 | 0.00301 | 0.708 | 0.00333 |
| 4 | 5 | 0.16 | 0.00156 | 11 | -0.887 | 0.688 | 0.00195 | 0.703 | 0.00205 |
| 4 | 6 | 0.142 | 0.00468 | 11.4 | -1.07 | 0.689 | 0.00122 | 0.69 | 0.00123 |
| 4 | 7 | 0.13 | 0.00206 | 12.3 | -1.22 | 0.678 | 0.000953 | 0.667 | 0.00106 |
| 5 | 0 | 0.248 | 0.0197 | 13.3 | -0.34 | 0.619 | 0.00315 | 0.643 | 0.00351 |
| 5 | 1 | 0.214 | -0.0037 | 11.3 | -0.499 | 0.651 | 0.00353 | 0.694 | 0.00453 |
| 5 | 2 | 0.194 | -0.00129 | 11.1 | -0.634 | 0.66 | 0.00318 | 0.7 | 0.00401 |
| 5 | 3 | 0.173 | -0.000753 | 11 | -0.789 | 0.672 | 0.00244 | 0.702 | 0.00289 |
| 5 | 4 | 0.154 | 0.00321 | 11.1 | -0.948 | 0.685 | 0.00172 | 0.7 | 0.00184 |
| 5 | 5 | 0.138 | 0.00225 | 11.2 | -1.11 | 0.694 | 0.00119 | 0.695 | 0.0012 |
| 5 | 6 | 0.125 | 0.00382 | 11.6 | -1.25 | 0.697 | 0.000848 | 0.686 | 0.000945 |
| 5 | 7 | 0.116 | -0.00292 | 12.4 | -1.37 | 0.689 | 0.000949 | 0.663 | 0.0014 |
| 6 | 0 | 0.192 | -0.0414 | 14.3 | -0.726 | 0.612 | 0.00161 | 0.622 | 0.00167 |
| 6 | 1 | 0.17 | 0.00287 | 12 | -0.849 | 0.642 | 0.00159 | 0.674 | 0.00219 |
| 6 | 2 | 0.157 | 0.00348 | 11.7 | -0.955 | 0.654 | 0.00147 | 0.682 | 0.00192 |
| 6 | 3 | 0.143 | 0.00431 | 11.5 | -1.07 | 0.669 | 0.00115 | 0.686 | 0.0013 |
| 6 | 4 | 0.131 | 0.00407 | 11.5 | -1.19 | 0.685 | 0.00089 | 0.687 | 0.000893 |
| 6 | 5 | 0.12 | -0.000198 | 11.6 | -1.32 | 0.696 | 0.000768 | 0.683 | 0.000872 |
| 6 | 6 | 0.111 | 0.00379 | 12 | -1.43 | 0.7 | 0.000834 | 0.674 | 0.00125 |
| 6 | 7 | 0.105 | -0.00181 | 12.8 | -1.52 | 0.693 | 0.00114 | 0.654 | 0.00212 |
| 7 | 0 | 0.158 | -0.000334 | 15.8 | -1.05 | 0.593 | 0.00175 | 0.597 | 0.00178 |
| 7 | 1 | 0.143 | 0.00587 | 13.2 | -1.14 | 0.623 | 0.000714 | 0.644 | 0.00105 |
| 7 | 2 | 0.133 | 0.00169 | 12.9 | -1.22 | 0.637 | 0.000661 | 0.652 | 0.00081 |
| 7 | 3 | 0.124 | -0.00168 | 12.6 | -1.31 | 0.656 | 0.000676 | 0.657 | 0.000678 |
| 7 | 4 | 0.115 | 0.00114 | 12.5 | -1.4 | 0.672 | 0.000714 | 0.66 | 0.000806 |
| 7 | 5 | 0.108 | -0.00298 | 12.6 | -1.49 | 0.686 | 0.00092 | 0.659 | 0.00139 |
| 7 | 6 | 0.102 | -0.00801 | 12.9 | -1.57 | 0.691 | 0.00124 | 0.652 | 0.00221 |
| 7 | 7 | 0.0982 | -0.0116 | 13.6 | -1.63 | 0.692 | 0.00217 | 0.636 | 0.00427 |

Table 2.12: constants of DCT coefficients(part 2)

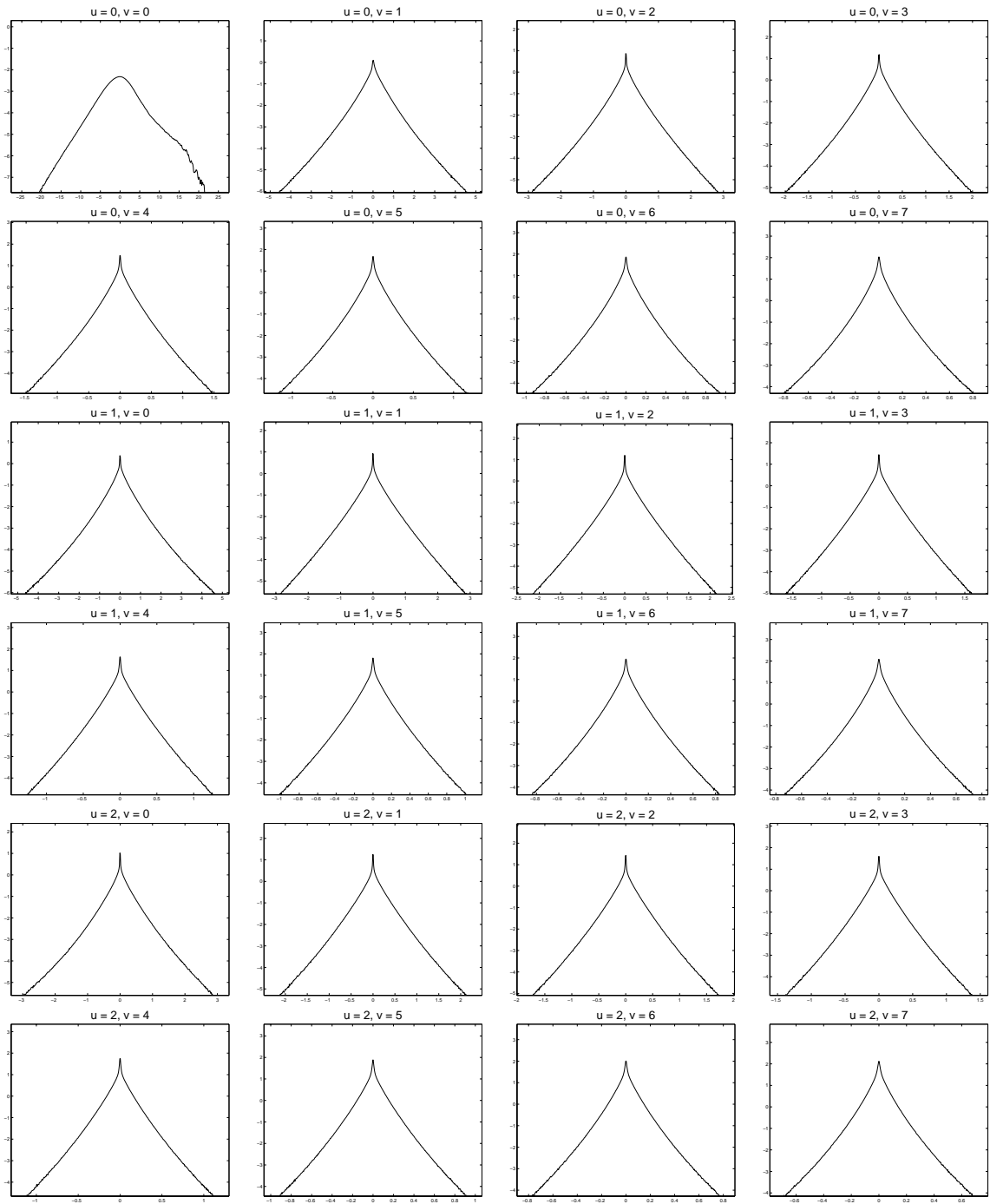


Figure 2.45: log(pdf) of DCT coefficients

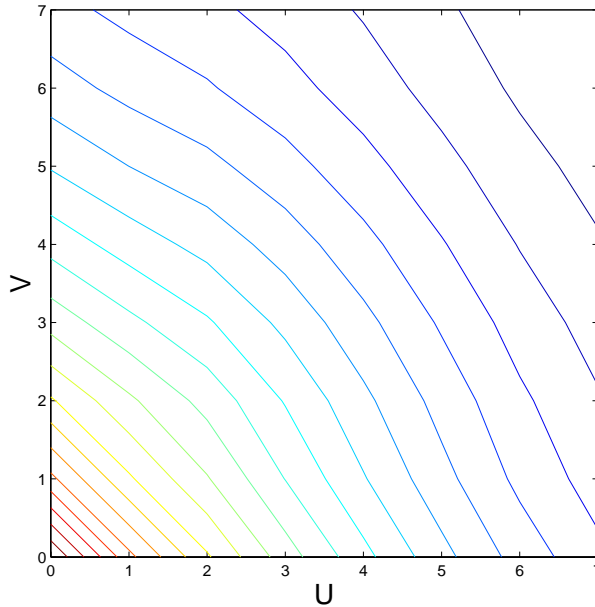


Figure 2.46: contour plot of the standard deviations of DCT coefficients

2.7 Random Filters

In the previous section, we observed an interesting fact that the $\log(pdf)$ of all the DCT coefficients have similar shapes. We want to make sure that this fact is caused by some property of natural images not the special design of the DCT filters.

For the first experiment, we generate random mean-0 filters in the following way: Get a random 8×8 matrix F_0 whose 64 entries are independent and are sampled uniformly from [01], and let

$$F = \frac{F_0 - \mu(F_0)}{\|F_0 - \mu(F_0)\|_2}$$

We can regard F as a mean-0 filter and apply it to images. Figure 2.47 shows the distribution of these filters' reactions for several random mean-0 filters and table 2.13 shows some constant associated to them.

In the second experiment, we generate random filters in a different way: we pick randomly 32 of the 64 entries of a 8×8 matrix, assign 1 to these picked entries and -1 to the

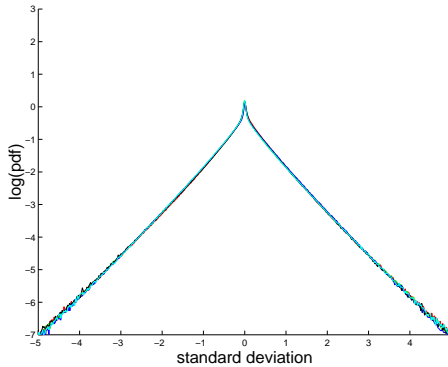


Figure 2.47: The empirical pdf of some random mean-0 filter reactions

| <i>filterid</i> | σ | \mathcal{S} | κ | \mathcal{H} | α_1 | $D(p_0 \parallel p_1)$ | α_2 | $D(p_0 \parallel p_2)$ |
|-----------------|----------|---------------|----------|---------------|------------|------------------------|------------|------------------------|
| 1 | 0.316 | -0.0145 | 7.95 | 0.194 | 0.806 | 0.00278 | 0.835 | 0.00304 |
| 2 | 0.302 | -0.00828 | 7.39 | 0.132 | 0.798 | 0.00469 | 0.872 | 0.00644 |
| 3 | 0.352 | -0.0112 | 6.99 | 0.376 | 0.848 | 0.00303 | 0.903 | 0.00391 |
| 4 | 0.244 | -0.0151 | 7.72 | -0.179 | 0.793 | 0.00396 | 0.849 | 0.00499 |
| 5 | 0.297 | 0.021 | 7.42 | 0.111 | 0.808 | 0.00479 | 0.87 | 0.00599 |
| 6 | 0.292 | -0.0112 | 7.58 | 0.0814 | 0.797 | 0.00503 | 0.859 | 0.00624 |
| 7 | 0.445 | 0.0288 | 8.47 | 0.651 | 0.753 | 0.00268 | 0.805 | 0.00375 |
| 8 | 0.33 | 0.0554 | 7.54 | 0.263 | 0.808 | 0.00415 | 0.862 | 0.00508 |
| 9 | 0.343 | 0.00999 | 7.38 | 0.319 | 0.81 | 0.00436 | 0.872 | 0.00558 |
| 10 | 0.317 | -0.0218 | 7.43 | 0.205 | 0.812 | 0.00327 | 0.869 | 0.00435 |

Table 2.13: constants of random mean-0 filter coefficients

others. We called such random filters the binary random mean-0 filters. Figure 2.48 and Table 2.14 show the results for some binary random mean-0 filters.

We can see that in both experiments, the distributions of the filter reactions are very similar to each other. Also, the $\log(\text{pdf})$ of all the distributions show linear tail, as we observed in the derivative statistics calculated from the vegetation category (see section 2.2.2), this fact can also be confirmed by the α values in the two tables where they are all close to 1.

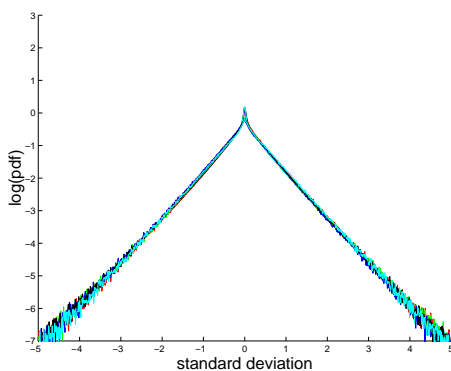


Figure 2.48: The empirical pdf of binary random mean-0 filter reactions

| <i>filterid</i> | σ | \mathcal{S} | κ | \mathcal{H} | α_1 | $D(p_0 \parallel p_1)$ | α_2 | $D(p_0 \parallel p_2)$ |
|-----------------|----------|---------------|----------|---------------|------------|------------------------|------------|------------------------|
| 1 | 4.69 | 0.018 | 6.51 | 4.15 | 0.949 | 0.00162 | 0.946 | 0.00163 |
| 2 | 4.75 | 0.00752 | 6.51 | 4.17 | 0.944 | 0.00146 | 0.945 | 0.00146 |
| 3 | 4.83 | -0.00886 | 6.55 | 4.2 | 0.959 | 0.00187 | 0.942 | 0.00196 |
| 4 | 4.79 | -0.0335 | 6.42 | 4.18 | 0.946 | 0.00107 | 0.954 | 0.00108 |
| 5 | 4.1 | 0.00689 | 6.14 | 3.97 | 0.963 | 0.00253 | 0.983 | 0.00261 |
| 6 | 4.04 | -0.0107 | 6 | 3.95 | 0.97 | 0.00325 | 0.999 | 0.00342 |
| 7 | 3.87 | 0.0058 | 5.9 | 3.88 | 0.943 | 0.00514 | 1.01 | 0.00605 |
| 8 | 4.07 | 0.00537 | 6.06 | 3.95 | 0.944 | 0.00358 | 0.993 | 0.00407 |
| 9 | 4.23 | -0.00779 | 6.05 | 4.02 | 0.972 | 0.00225 | 0.994 | 0.00234 |
| 10 | 4.39 | 0.000961 | 6.28 | 4.06 | 0.944 | 0.00231 | 0.968 | 0.00244 |

Table 2.14: constants of binary random mean-0 filter reactions

2.8 2X2 blocks

The joint statistics of horizontal, vertical and diagonal Haar wavelet coefficients at the same level and same location almost describe the probability distribution of 2X2 blocks, except for the mean value or the DC coefficient. In this section we extend the discussion in the previous section, and check some simple properties of 2×2 blocks by testing some simple hypotheses.

First, we calculated the joint empirical distribution of all 2X2 blocks. Let $[X_1 X_2 X_3 X_4]$ be the pixels values, and a linear transformation:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

i.e. Y_1, Y_2, Y_3, Y_4 are the filter reactions of the low-pass, horizontal, vertical and diagonal Haar filters. Let the empirical distribution of the random variables under the x_i coordinate be $f_0(x_1, x_2, x_3, x_4)$, and the same thing under the y_i coordinate be $\tilde{f}_0(y_1, y_2, y_3, y_4)$. We now check according to the information theory, how good are the following hypotheses:

Hypothesis 1 X_1, X_2, X_3, X_4 are independent.

Let $f_1(x_1, x_2, x_3, x_4) = g_1(x_1) * g_1(x_2) * g_1(x_3) * g_1(x_4)$, where g_1 is the marginal distribution of single pixels. we have:

$$D(f_0 || f_1) = 3.9$$

Obviously, this is a bad hypothesis, adjacent pixels are closely related to each other. But in applications, independent assumptions are usually made to simplify algorithms. The test on the following two hypotheses show that under the y_i 's are more independent to each other.

Hypothesis 2 Y_1 is independent from Y_2, Y_3, Y_4 .

Let

$$\tilde{f}_1(y_1, y_2, y_3, y_4) = \tilde{g}(y_1) \times \tilde{h}(y_2, y_3, y_4)$$

where \tilde{g}, \tilde{h} are the marginal distributions of Y_1 and $[Y_2, Y_3, Y_4]$ respectively. We have:

$$D(\tilde{f}_0 || \tilde{f}_1) = 0.078$$

Hypothesis 3 Y_1, Y_2, Y_3, Y_4 are independent.

Let

$$\tilde{f}_2(y_1, y_2, y_3, y_4) = \tilde{g}_1(y_1)\tilde{g}_2(y_2)\tilde{g}_3(y_3)\tilde{g}_4(y_4)$$

where \tilde{g}_i is the marginal distribution of Y_i . We have

$$D(\tilde{f}_0 || \tilde{f}_2) = 0.316$$

In [13], Koloydenko has assumed that the statistics of 2×2 blocks are invariant under rotations, flips and inversions. Next, we check from our data how good these assumptions are. Let $\tilde{f}_{transpose}, \tilde{f}_{rflip}, \tilde{f}_{udflip}, \tilde{f}_{inv}$ be the pdf of the transposition, left-right flipping, up-down flipping and inversion of the original 2×2 blocks. Since there are a lot of empty bins in our histogram, the *KL* distance between these and the original distribution maybe infinity. A more reasonable way to is to check the KL distance between the original and the *average* of the original and transformed pdf.

Hypothesis 4 The joint distribution of X_1, X_2, X_3, X_4 is invariant under rotations, flips

and inversion.

$$\begin{aligned}
D(f_0 || \frac{f_0 + f_{transpose}}{2}) &= 0.0036 \\
D(f_0 || \frac{f_0 + f_{trflip}}{2}) &= 0.0020 \\
D(f_0 || \frac{f_0 + f_{udflip}}{2}) &= 0.0024 \\
D(f_0 || \frac{f_0 + f_{inv}}{2}) &= 0.0241
\end{aligned}$$

It is not a surprise that $D(f_0 || \frac{f_0 + f_{trflip}}{2})$ is the smallest among these KL distances, because it should be true that an left-right flipped natural image is another perfect natural image. The fact that $D(f_0 || \frac{f_0 + f_{transpose}}{2})$ and $D(f_0 || \frac{f_0 + f_{udflip}}{2})$ are also very small indicates that some large structures(e.g. the horizon) is not clearly reflected by this local statistics. On the other hand, some large structures(e.g. portion of sky) do make $D(\tilde{f}_0 || \frac{\tilde{f}_0 + \tilde{f}_{inv}}{2})$ significantly larger than the other three values.

2.9 8X8 patches

In this section, we try to describe some important statistics structures in the 8×8 image patches. Since now we are dealing with such a high dimensional space, the direct histogram methods used in previous sections do not work here, instead we apply some non-parametric statistics technique, e.g. k-means centers method. Similar methods was used by Malik *et al.* in [15] in analyzing ‘textons’.

But first, we follow the ideas in [15] and preprocess our data. For each 8×8 patch, we calculate the 3-level Haar Wavelet transformation to get 63 AC coefficients, and we renormalized the coefficients at level 2 and 3 by dividing 2 and 4 respectively. This way, we can assume all filters are of the same L1 norm(for scale invariance). So we have a linear map from the space of 8×8 patches to R^{63} ,

$$T_0(P) \in R^{63}$$

where P is an 8×8 image patch. The authors in [15] further renormalized the above map by log-function to create local maxima in the distribution, away from the origin. We here use a more direct approach, we simply project all the non-zero vectors to the unit sphere

$$T_1(P) = \frac{T_0(P)}{\|T_0(P)\|}$$

for $\|\mathbf{T}_0(\mathbf{P})\| > 0$. So given a non-constant P , we get a sample point $T_1(P)$ on the unit sphere S^{62} . For each image I , let P runs through all the 8×8 patches, and we get a set of samples on S^{62} . We then run the k-means algorithm on these samples to get 128 k-means centers. We put together all these centers calculated from all 4000 images in van Hateren database, and do k-means again to get the final 128 centers, which are shown in figure 2.49. We can see most of the centers show edge or bar structures.

We want to see how well the centers represent the sample points on S^{62} . Let’s define:

$$S_r = \{v \in S^{62} | \text{angle}(v, C) < r\}$$

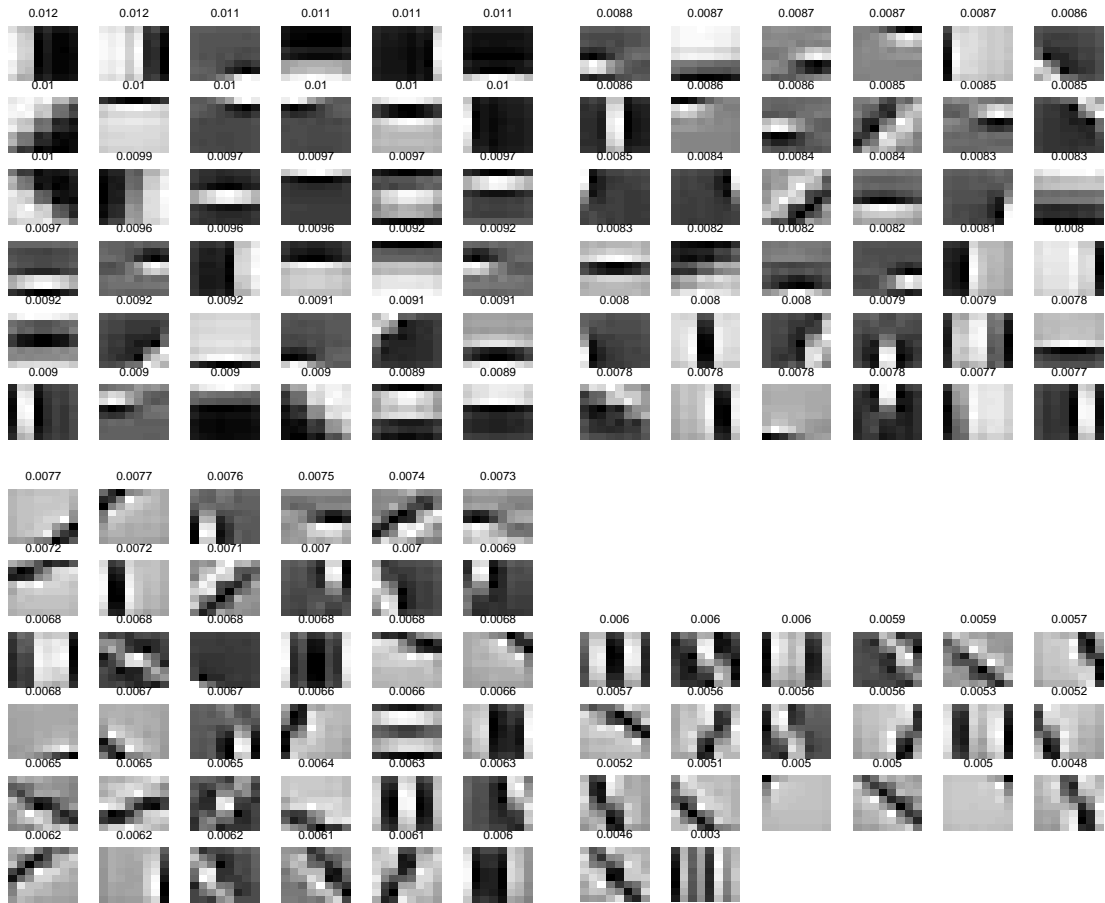


Figure 2.49: k-means centers of 8X8 images patches. The number is the frequency that images patches are closest to the center.

where C is the set of all the 128 centers, and

$$angle(v, C) = \min_{c \in C} (angle(v, c))$$

We can calculate the percentage of samples fall in S_r for different r 's. Also, we can calculate the volume of S_r numerically. The lower curve of figure 2.51 is the plot of the percentage of samples fall in S_r against $\frac{vol(S_r)}{vol(S^{62})}$. We can see from the curve that about 90% of sample points fall in a region of 10% of the total volume.

From our experience with low dimensional natural scene statistics (e.g. 3D joint statistics of Haar coefficients), we know that the density function may have different rates of change

along different directions. To get a rough description of this, we define for each $c \in C$ a map $P_c : S^{62} \rightarrow R^{62}$ by:

$$P_c(v) = \frac{v - \langle v, c \rangle c}{\|v - \langle v, c \rangle c\|} \text{angle}(v, c) \quad (2.11)$$

where v are samples that are closer to c than any other centers in C . This way, we map all samples around c to vectors in R^{62} , and we can calculate the covariance matrix Q_c of all the samples $P_c(v)$, the eigenvectors and eigenvalues of which roughly tell us along which direction the value of the density function on S^{62} concentrates. Figure 2.50 shows two examples of the principle components around two centers. For each $c \in C$, let Q_c be the covariance matrix of the samples $P_c(v)$, we define the *pca* normalized angle as

$$\widetilde{\text{angle}}(v, c) = \sqrt{P_c(v) Q_c^{-1} P_c(v)^T}$$

and similarly define

$$\tilde{S}_r = \{v \in S^{62} | \widetilde{\text{angle}}(v, C) < r\}$$

Again, we plot the percentage of samples fall in \tilde{S}_r against $\frac{\text{vol}(\tilde{S}_r)}{\text{vol}(S^{62})}$. The top curve of figure 2.51 shows the plot. Now we can see that 90% of the samples fall in only a region of 2% of the total volume, and we can claim from this that the k-means centers shown in figure 2.49 do represent some common structures in natural images.

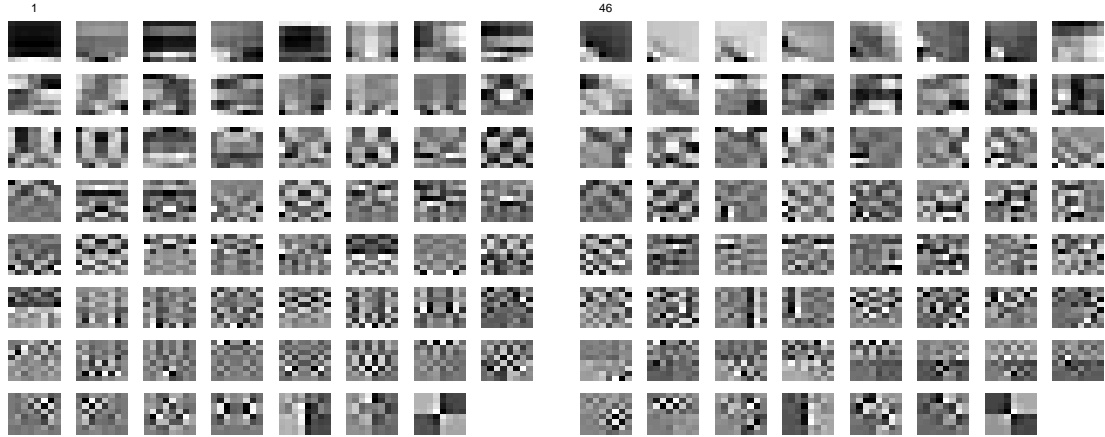


Figure 2.50: Principal components around centers. The upper-left subplot is the center, and the others are the eigenvectors of the covariance matrix around the center, See text

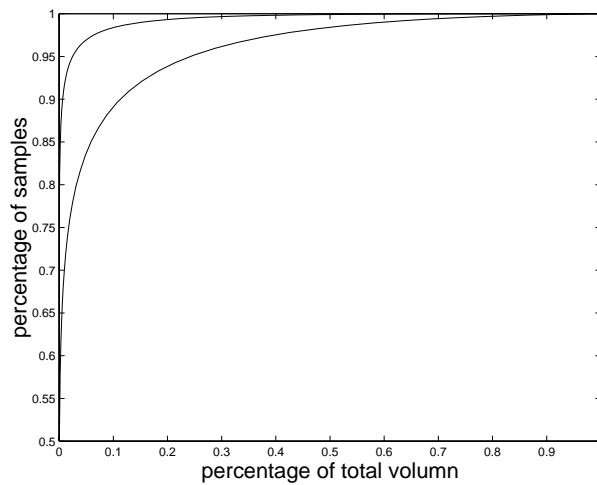


Figure 2.51: Percentage of samples fall in percentage of the total volume. Bottom curve: assuming disk like regions, top curve: assuming elliptic regions. See Text

Chapter 3

Simulating local statistics with a generalized Ising model

3.1 Basic Setup

In his paper [17], Ruderman tried to explain the scale invariance property by assuming that the world can be broken down into discrete objects of different scales. More recently, Lee *et al* [14] have developed the ‘Random Collage Model’, which simulates some statistics of natural images as well as the scale invariance property. A problem of such approaches is that images synthesized by this model is difficult to analyze. Still, it’s surprising that complicated structures like those observed in figure 2.25 can be simulated under some quite seemingly irrelevant assumptions. This lead us to believe that the statistical features we observed in natural images are caused by some simple stochastic process. Grenander [9] suggested a model using the product of Gaussian random variables, and from this model he derived a family of distributions for derivative statistics. Here we introduce a different approach.

First, as in [4], we assume a *real scene* $\phi(x, y)$ is a function (or more properly a generalized function, see [16]) defined on R^2 . A (digitized) image I of ϕ is a function defined on integer grid Z^2 , such that the value of $I(i, j)$ is the average of $\phi(x, y)$ over the square (or pixel) $S_{ij} = [i, (i + 1)] \times [j, j + 1]$, i.e.

$$I(i, j) = \int_{S_{ij}} \phi(x, y) dx dy$$

According to Ruderman’s idea, an real scene consists of objects, the color (or intensity) of objects are independent of each other, i.e.

$$\phi(x, y) = \sum_{i=1}^N \alpha_i Ind_{A_i}$$

where A_i ’s (objects) are some non-overlapping sets, and α_i is the intensity of the corresponding object. Here we make a further assumptions about the local structures.

Assumption 1 On a small patch of real scene Ω (say a region covered by 2×2 pixels), there are at most two ‘objects’ A and B . And the intensity α on A and β on B are constants,

i.e.

$$\phi|_{\Omega} = \alpha Ind_A + \beta Ind_B$$

where $A \cup B = \Omega$, and Ind_A and Ind_B are the identification functions.

We can argue that on a small non-homogeneous image patch, we are likely to observe an edge or a bar, where the above assumption is reasonable. T-junctions are definitely counter examples, but they are rare events compared to edges and bars. To test this assumption, We did an experiment following the procedure given below.

1. For each image I , we scale it down by taking the maximum or minimum of each 4×4 blocks, whichever closer to the mean value. Let's denote the resulting image by I' .
2. For each $M \times M$ block $B_{M \times M}$ of I' , normalize all pixels in that block by:

$$f(x) = \frac{x - \min(B_{M \times M})}{\max(B_{M \times M}) - \min(B_{M \times M})}$$

3. Collect all the normalized pixel values $f(x)$ except the 1's(maximum) and 0's(minimum) and calculate the histogram of $f(x)$.

Step 1 in the above procedure is to make I' more like a 'real scene' by reducing the averaging effects in the pixels of digital images. Figure 3.1 shows the result of the experiment for $M = 2$ and $M = 4$. We can see that in both cases, the distributions have two maximum values at the two ends, which indicates that the above assumption is reasonable.

In assumption 1, when Ω is fixed, $\phi|_{\Omega}$ (we will just write ϕ in the future) is generated by two processes: an object generating process ζ_0 which generates A (and hence B), and a 'coloring process', $\eta_0 = (\alpha, \beta)$ which 'paints' A and B with 'colors' α and β respectively. The following two assumptions are adopted by Ruderman [17] and Lee *et al.* [14].

Assumption 2 The object generating process ζ_0 and coloring process η_0 are independent.

Assumption 3 α_0 and β_0 are independent.

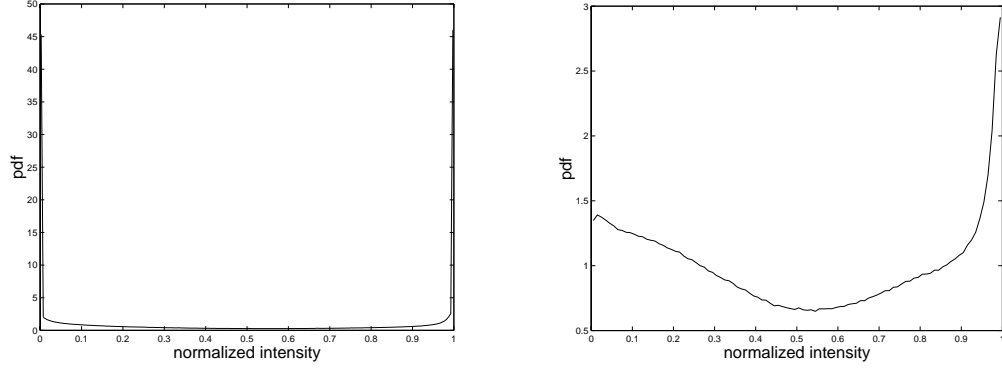


Figure 3.1: The pdf of ‘normalized intensities’(see text). Left: for 2×2 blocks. Right: for 4×4 blocks

Let’s see what we can conclude from these assumptions. Let ϕ_0 be the ‘characteristic function’ of the ‘objects’.

$$\phi_0|_A = 1, \quad \phi_0|_B = -1$$

Then

$$\phi = \frac{\phi_0 + 1}{2}\alpha + \frac{1 - \phi_0}{2}\beta$$

Assume we have a group of filters supported by Ω : $f_1, f_2 \dots f_n$. Each filter reaction of ϕ will be:

$$\begin{aligned} \langle f_i, \phi \rangle &= \int f_i(x, y)\phi(x, y)dxdy = \int f_i(x, y)\left(\frac{\phi_0 + 1}{2}\alpha + \frac{1 - \phi_0}{2}\beta\right)dxdy \\ &= \frac{\alpha - \beta}{2} \int f_i(x, y)\phi_0(x, y)dxdy + \frac{\alpha + \beta}{2} \int f_i(x, y)dxdy \end{aligned}$$

If f_i is a mean-0 filter(i.e $\int f_i dxdy = 0$), we have

$$\langle f_i, \phi \rangle = \frac{\alpha - \beta}{2} \int f_i(x, y)\phi_0(x, y)dxdy$$

With a linear transformation, we can assume any group of n linear filters consists of n

zero-mean filters or n-1 zero-mean filters and one non zero-mean filter. For the former case, the joint reactions of all the filters can be written as:

$$\langle f_1, \phi \rangle, \dots, \langle f_n, \phi \rangle = \frac{\alpha - \beta}{2} \langle f_1, \phi_0 \rangle, \dots, \langle f_n, \phi_0 \rangle \quad (3.1)$$

So we have factored the joint filter reactions into two simpler *independent* processes. We call $\eta = \frac{\alpha - \beta}{2}$ the *color factor*, call $\zeta = \langle f_1, \phi_0 \rangle, \dots, \langle f_n, \phi_0 \rangle$ the *object factor*. Hence the filter reactions $\gamma = \langle f_1, \phi \rangle, \dots, \langle f_n, \phi \rangle = \eta \cdot \zeta$ is the independent product of the color factor and the object factor. Assume the pdf of the two factors are $h_\eta(y)$ and $h_\zeta(u_1, u_2, \dots, u_n)$ respectively, so the joint density function of the two factors (η, ζ) (with the infinitesimal elements) is

$$h_\eta(y) \cdot h_\zeta(u_1, u_2, \dots, u_n) dy du_1 du_2 \dots du_n$$

let $(x_1, x_2, \dots, x_n) = y \cdot (u_1, u_2, \dots, u_n)$, so $(u_1, u_2, \dots, u_n) = (x_1, x_2, \dots, x_n)/y$ and the probability density function of (η, γ) will be:

$$h_\eta(y) \cdot h_\zeta\left(\frac{x_1}{y}, \frac{x_2}{y}, \dots, \frac{x_n}{y}\right) dy \frac{dx_1 dx_2 \dots dx_n}{|y|^n}$$

By integrating w.r.t y , we get the density function of h_γ :

$$h_\gamma(x_1, x_2, \dots, x_n) = \int h_\eta(y) \frac{1}{|y|^n} h_\zeta\left(\frac{x_1}{y}, \frac{x_2}{y}, \dots, \frac{x_n}{y}\right) dy \quad (3.2)$$

Let's call the above operation the independent mixing of h_η and h_ζ , and denote it by:

$$h_\gamma = h_\eta \otimes h_\zeta$$

Notice that independent mixing to independent product of two random variables is what convolution to independent summation of two random variables. There are some interesting properties of independent product of two random variables.

Proposition 3.1 *If X and Y are independent, then $\mu(X \cdot Y) = \mu(X) \cdot \mu(Y)$*

Proposition 3.2 *If X and Y are independent, and $\mu(X) = \mu(Y) = 0$, then:*

$$\begin{aligned} \text{var}(X \cdot Y) &= \text{var}(X) \cdot \text{var}(Y) \\ \text{kurtosis}(X \cdot Y) &= \text{kurtosis}(X) \cdot \text{kurtosis}(Y) \\ \text{skewness}(X \cdot Y) &= \text{skewness}(X) \cdot \text{skewness}(Y) \end{aligned} \tag{3.3}$$

proof

We only show the proof for kurtosis, others are similar. If $\mu(X) = \mu(Y) = 0$, we have:

$$\begin{aligned} \text{kurtosis}(X \cdot Y) &= \frac{\mu((X \cdot Y)^4)}{\mu((X \cdot Y)^2)^2} \\ &= \frac{\mu(X^4)\mu(Y^4)}{\mu(X^2)^2\mu(Y^2)^2} \\ &= \text{kurtosis}(X) \cdot \text{kurtosis}(Y) \end{aligned}$$

■

Proposition 3.3

$$\text{kurtosis}(X \cdot Y) \geq \max(\text{kurtosis}(X), \text{kurtosis}(Y))$$

This is a direct conclusion given the following lemma.

Lemma 3.1 *For any random variable X , $\text{kurtosis}(X) \geq 1$, and equality holds if and only if the density function $p(x)$ of X can be written (symbolically) as:*

$$p(x) = \frac{1}{2}\delta(x - \mu - a) + \frac{1}{2}\delta(x - \mu + a)$$

where μ is the mean of X and $a \neq 0$ is some constant.

proof

Without loss of generality, we can assume that $\mu = 0$. Notice that,

$$\left(\int x^2 p(x) dx\right)^2 = \left(\int x^2 p(x)^{\frac{1}{2}} p(x)^{\frac{1}{2}} dx\right)^2 \leq \int x^4 p(x) dx \cdot \int p(x) dx = \int x^4 p(x) dx$$

The Cauchy-Schwartz inequality is used above, and the equality holds if and only if

$$x^2 p(x)^{\frac{1}{2}} = C \cdot p(x)^{\frac{1}{2}}$$

all most everywhere for some constant C . This can only be true when

$$p(x) = \frac{1}{2}\delta(x - a) + \frac{1}{2}\delta(x + a)$$

where $a = \sqrt{C}$. ■

These conclusions explain why we observed high kurtosis in the distributions of filter reactions of images. More exactly, we believe the independent mixing of two density functions is responsible(at least partially) for the ‘generalized Laplace’ distributions of local filter reactions of images.

Since we assume α and β are sampled independently from the single pixel empirical distribution(with pdf $f_0(x)$, see section 2.1.1 on page 11), the density function $h_\eta(x)$ of the color factor η can be calculated by,

$$h_\eta(x) = 2 \cdot \text{convolve}(f_0(y), f_0(-y))(2x) \tag{3.4}$$

which is shown in figure 3.2. The object factor ζ is more difficult to model. Noticed that the color factor does not depend on the choice of filters f_1, f_2, \dots, f_n , but the object factor does. Now let’s look at a example of independent mixing.

Consider the simplest mean-0 filter reaction: the derivative D . In section 2.2.1, we have calculated the empirical distribution of derivative statistics from van Hateren database. Let’s use the statistics at scale 2(which is more stable than that of scale 1) as an example. We have:

$$D = \eta \cdot \zeta$$

where ζ is the object factor and η is the color factor. By proposition 3.2,

$$\sigma(\zeta) = \sigma(D)/\sigma(\eta) = 0.25$$

$$\mathcal{K}(\zeta) = \mathcal{K}(\mathcal{D})/\mathcal{K}(\eta) = 3.43$$

Since $\mathcal{K}(\zeta)$ is close to 3, we may guess that ζ is close to Gaussian, so we simply assume $\zeta \sim N(0, 0.25^2)$. Figure 3.3(left) shows the $\log(\text{pdf})$ of D and that of $\eta \cdot \zeta$. Look how close they are to each other. Results of the same experiment on the statistics calculated from different categories of Sowerby database are shown in figure 3.3(right) and Figure 3.4. The experiment on the sky category is obviously a failure. This may be related to the fact that there are large portions of homogeneous regions in this category, as a result, the object factor is far from Gaussian (look at the big kurtosis values for the sky category in table 2.5).

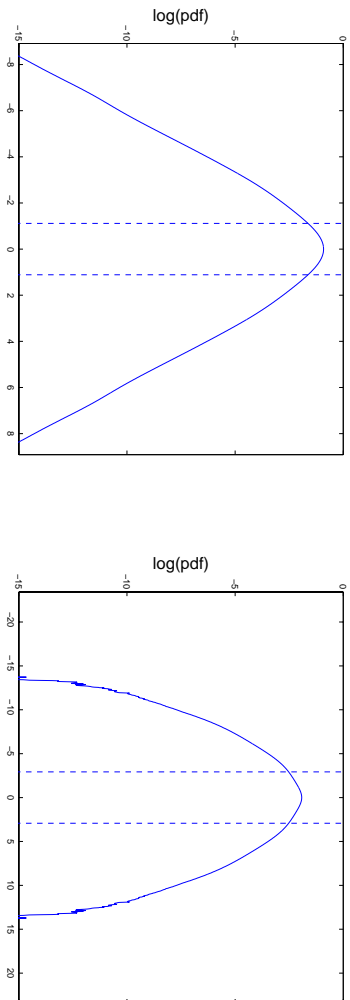


Figure 3.2: The distribution of the color factor η . Left: van Hateren. Right: Sowerby

Remark. Wainwright *et al.* [22] have proposed the Gaussian scale mixture(GSM) model. Let $\gamma = (\langle f_1, \phi \rangle, \dots, \langle f_n, \phi \rangle)$, they assumed that

$$\gamma = {}^d zU$$

where $z \geq 0$ is a scalar random variable; $U \sim N(0, Q)$ is a Gaussian random vector, z and U are independent. The above example shows that this assumption is a good one for $n = 1$, and z is nothing else than $C|\eta|$, the absolute value of the color factor multiplied by some constant.

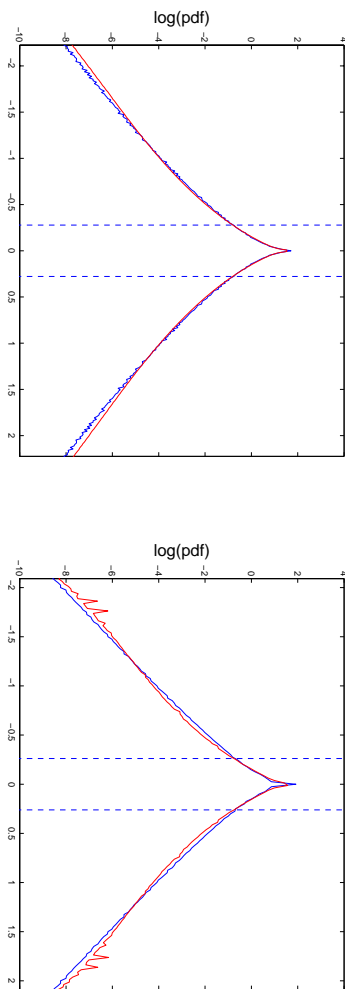


Figure 3.3: Simulation of the derivative statistics by assuming Gaussian object factor, Red: statistics from natural images. Blue: statistics from simulation. Left: simulation of the van Hateren database. Right: simulation of the Sowerby database.

Finally, in equation 3.1 we have assumed that all the filters are mean-0. If this is not the case, by a linear transformation (and by adding an extra filter if necessary), we can assume f_1, \dots, f_{n-1} are mean-0, and $f_n \equiv 1$. Then

$$\begin{aligned} \langle < f_1, I \rangle, \dots, \langle f_n, I \rangle \rangle &= \frac{\alpha - \beta}{2} \langle < f_1, I_0 \rangle, \dots, \langle f_{(n-1)}, I_0 \rangle, 0 \rangle + \\ & (0, 0, \dots, 0, \frac{\alpha - \beta}{2} (\mu(A) - \mu(B))) + \frac{\alpha + \beta}{2} (\mu(A) + \mu(B)) \end{aligned}$$

3.2 Object generating process

The coloring process η is straight forward, whose density function has been related to the distribution of single pixels in equation 3.4.

The object generating process ζ generates subsets of R^2 . To simulate this process numerically, we assume that each pixel has $K \times K$ subpixels, and a ‘real scene’ ϕ is a function defined on these subpixels (instead of continuous space). Each subpixel belongs to either A or B , and the pixel intensity p_i is the average of the values of ϕ at all the K^2 subpixels. Assume the small ‘real scene’ patch Ω cover $M \cdot M$ pixels (and hence covers $(M \cdot K) \times (M \cdot K)$ subpixels). Now the object generation process should sample the binary valued function ϕ_0 on the $(M \cdot K) \cdot (M \cdot K)$ grid. As in natural images, we want a ‘object A ’ so

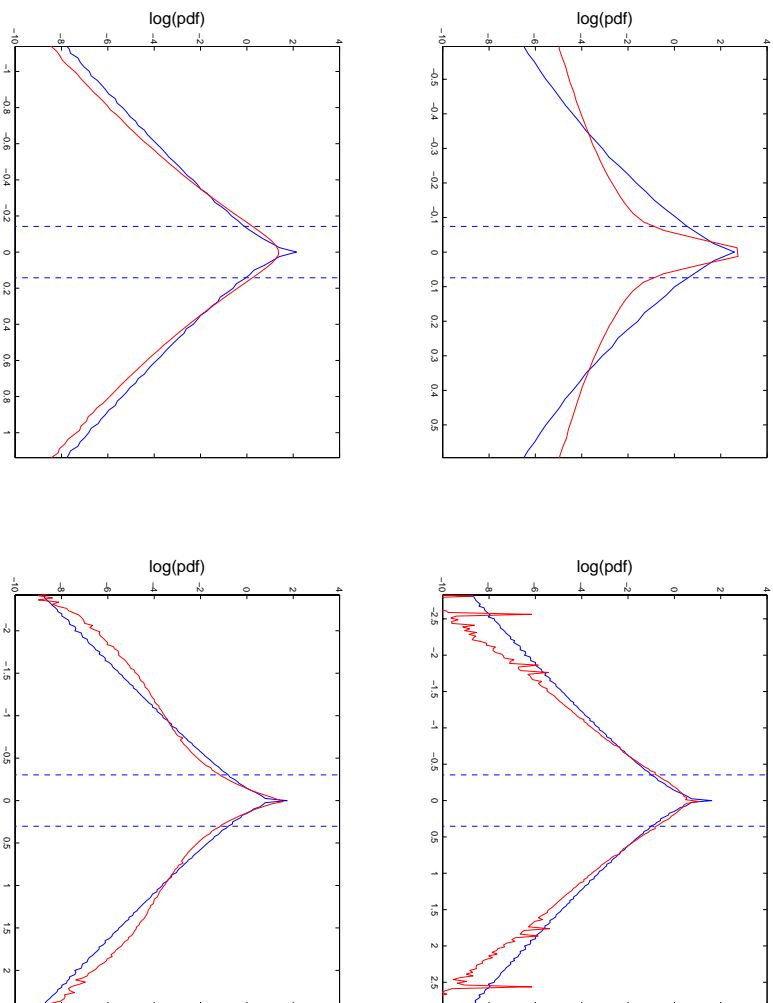


Figure 3.4: Simulation of the derivative statistics by assuming Gaussian object factor. Red: statistics from natural images. Blue: statistics from simulation. From left to right and up to down: sky, vegetation, road, man-made

generated to have some connectivity and some smoothness in the boundary. The simplest process we can think of will be the Ising model without external magnetic field. But the Ising model will favor horizontal and vertical edges against diagonal ones, and instead we used a Markov Random Field which is slightly more complicated than the Ising Model. We assume the neighborhood of a site(subpixel) are all the 8 neighbor sites, including diagonal neighbors. So the maximum cliques will be all the 2×2 blocks, and each 2×2 block has 16 possible configurations. To define the Markov Random Field, we only have to define a potential function E of all these 16 configurations. To simplify our notation, let (a, b, c, d) represents the patch

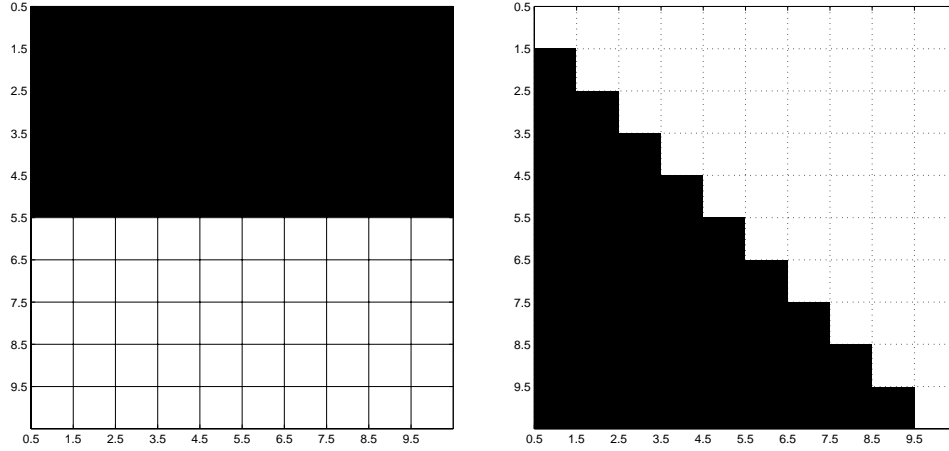


Figure 3.5: An vertical edge and an diagonal

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

It is reasonable to assume that E should be invariant under flipping, rotating and inverting(1 to -1 and -1 to 1), so we only have to define E on the following 4 cases:

$$(1, 1, 1, 1), (1, 1, 1, -1), (1, 1, -1, -1), (1, -1, -1, 1)$$

$(1, 1, 1, 1)$ is a constant patch, hence $E(1, 1, 1, 1) = 0$. Without loss of generality, we can define $E(1, 1, -1, -1) = 1$. To find an appropriate $x = E(1, 1, 1, -1)$, consider the two images shown in figure 3.5. Both of them are of the same size(say $n \times n$). The left figure has a horizontal edge. There are $n - 1$ 2×2 blocks of type $(1, 1, -1, -1)$ and all others 2×2 blocks are constant patches. So the total energy for this image is $n - 1$. The right figure has a diagonal edge, and there are $2n - 3$ 2×2 blocks of type $(1, 1, 1, -1)$, so the total energy will be $(2n - 3) \cdot x$. Since we want to treat horizontal and diagonal edges equally, and notice that the length of the diagonal edge is $\sqrt{2}$ that of the horizontal edge, we have

$$(2n - 3) \cdot x = \sqrt{2}(n - 1)$$

when n large, $x \approx \frac{\sqrt{2}}{2}$. Hence we define $E(1, 1, 1, -1) = \frac{\sqrt{2}}{2}$. We can think $(1, -1, -1, 1)$ as a patch with two edges, and define $E(1, -1, -1, 1) = 2 \cdot E(1, 1, 1, -1) = \sqrt{2}$. Table 3.1 gives the exact definition of E .

| a | b | c | d | E |
|-----|-----|-----|-----|----------------------|
| -1 | -1 | -1 | -1 | 0 |
| -1 | -1 | -1 | 1 | $\frac{\sqrt{2}}{2}$ |
| -1 | -1 | 1 | -1 | $\frac{\sqrt{2}}{2}$ |
| -1 | -1 | 1 | 1 | 1 |
| -1 | 1 | -1 | -1 | $\frac{\sqrt{2}}{2}$ |
| -1 | 1 | -1 | 1 | 1 |
| -1 | 1 | 1 | -1 | $\sqrt{2}$ |
| -1 | 1 | 1 | 1 | $\frac{\sqrt{2}}{2}$ |
| 1 | -1 | -1 | -1 | $\frac{\sqrt{2}}{2}$ |
| 1 | -1 | -1 | 1 | $\sqrt{2}$ |
| 1 | -1 | 1 | -1 | 1 |
| 1 | -1 | 1 | 1 | $\frac{\sqrt{2}}{2}$ |
| 1 | 1 | -1 | -1 | 1 |
| 1 | 1 | -1 | 1 | $\frac{\sqrt{2}}{2}$ |
| 1 | 1 | 1 | -1 | $\frac{\sqrt{2}}{2}$ |
| 1 | 1 | 1 | 1 | 0 |

Table 3.1: Definition of the potential function. a, b, c, d are entries of a 2×2 block from left to right and up to down.

Once the potential function is given, the Markov Random Field is defined as:

$$P(\mathbf{x}) = \frac{1}{Z} e^{-\sum_{\{x_1, x_2, x_3, x_4\}} E(x_1, x_2, x_3, x_4)/T}$$

where $\{x_1, x_2, x_3, x_4\}$ runs through all 2×2 blocks and T is the temperature.

In our simulation, we pick T evenly between $[0.1 \ 5]$ and generalized 512×512 binary images for different temperature T by Gibbs Sampler. Figure 3.6 shows a sample of the MRF. Let $U = \sum_{\{x_1, x_2, x_3, x_4\}} E(x_1, x_2, x_3, x_4)$ be the energy, Figure 3.7 shows $\mu(U)$ as a function of temperature T as well as the derivative $\frac{d\mu(U)}{dT}$ against T , and we can see that the critical temperature T_0 is around 1.88.

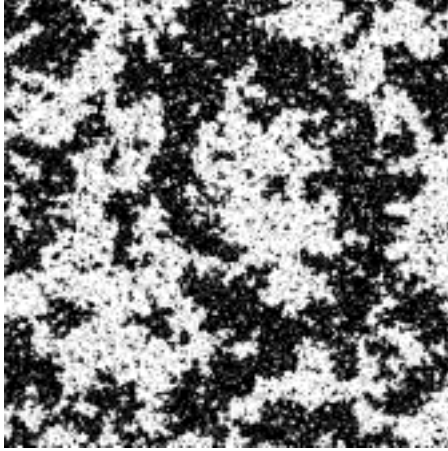


Figure 3.6: A sample from our MRF at $T = 1.88$.

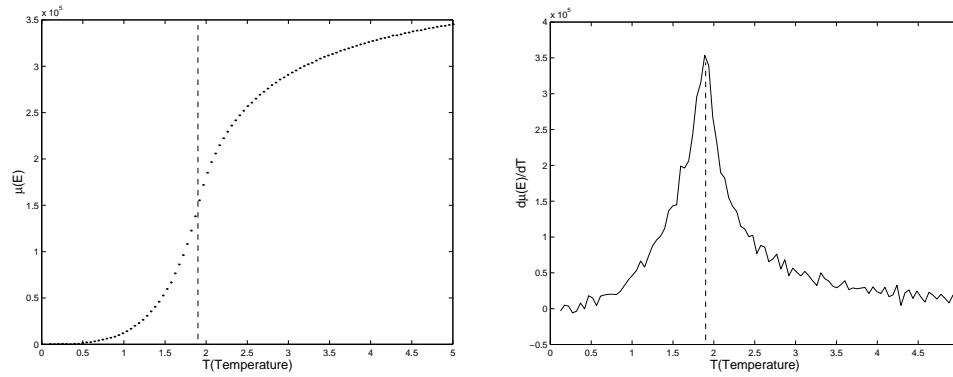


Figure 3.7: Left figure: Expectation of the energy of samples v.s. temperature. Right figure: Derivative of the expectation of the energy. The vertical dotted line is $T = 1.88$, the critical temperature.

3.3 Simulation of the Derivative Statistics

Here is the procedure to get one derivative sample from ϕ_0 (a 512×512 sample from the MRF at fixed temperature T). Again we assume each pixel contains $K \times K$ subpixels,

(1) Get a $K \times 2K$ patch Ω from ϕ_0 . Let A, B be the objects in Ω

$$A = \{x \in \Omega | \phi_0(x) = 1\}$$

$$B = \{x \in \Omega | \phi_0(x) = -1\}$$

(2) Sample randomly two values α and β from the empirical distribution of single pixel statistics. As we calculated in section 2.1.1.

(3) Let $\phi_1 = \alpha Ind_A + \beta Ind_B$, which is a $K \times 2K$ patch at subpixel resolution.

(4) Take $K \times K$ block average of ϕ_1 to get two pixel intensities, and the difference of them gives a sample of the derivative statistics.

Repeat the above procedure for different patches Ω and different ϕ_0 (samples for a fixed temperature T), and we can collect a histogram of the derivative statistics at temperature T .

A more efficient way was used in our calculation. In equation 3.1, let $n = 1$ and f_1 be defined on a $K \times 2K$ grid, with $f_1|_{left\ K \times K\ block} = \frac{1}{K^2}$ and $f_1|_{right\ K \times K\ block} = -\frac{1}{K^2}$. We only have to calculate the distribution of the object factor $\langle f_1, \phi_0 \rangle$, because the distribution of the derivative can then be calculated from equation 3.2. Here is the procedure to get a sample of the object factor:

1. Get a $K \times 2K$ patch Ω from ϕ_0 .
2. Take $K \times K$ block average of the patch to get two pixels ‘intensities’, whose difference gives a sample of the object factor.

We run this algorithm for samples at different temperature T separately, and also we use 2 different K values to simulate 2 scales: $K = 8$ and $K = 16$. Figure 3.8 shows the result of the simulation of the derivative statistics at different temperatures and scales. Figure 3.9 shows the KL distance between the empirical derivative statistics calculated from natural image database and that from our model, also the comparison of the variances. It is interesting to observe that around the critical temperature, everything comes together: the derivative statistics calculated from our model are roughly scale invariant and they are close to the empirical distribution calculated from natural images. It's well known in physics that the Ising model is scale invariant at the critical temperature. Notice that, in most of the literatures in physics, the system is scaled down by taking block majority instead of block average. Nevertheless, it may be not a surprise that we observed a rough scale invariance at the critical temperature. On the other hand, the fact that *single pixel statistics + the Ising-like model + scale invariance* predicates distribution the the derivative statistics is interesting, it indicates that our model does capture some stochastic structures in natural image.

Figure 3.10 shows the distribution of the object factor at the critical temperature. Notice that this variable is close to but not exactly scale invariant. Since the object factor can have 129 possible values for $K = 8$ and 513 possible values for $K = 16$, we can not define scale invariance exactly for finite K in the first place. We may overcome the problem by studying the limit behavior when $K \rightarrow \infty$, but this is not our concern here.

3.4 Simulation of the Haar Wavelet Coefficients

Similar to the derivative statistics, we can calculate Haar Wavelet Coefficients of the simulation. Since the simulation works best at the critical temperature for derivative statistics, we will only simulate at the critical temperature here.

Figure 3.11 shows the joint statistics of some 2d Haar wavelet coefficients calculated from our simulation. Compare that to figure 2.25, we can see our simulation captures the polyhedra structures in the level curves.

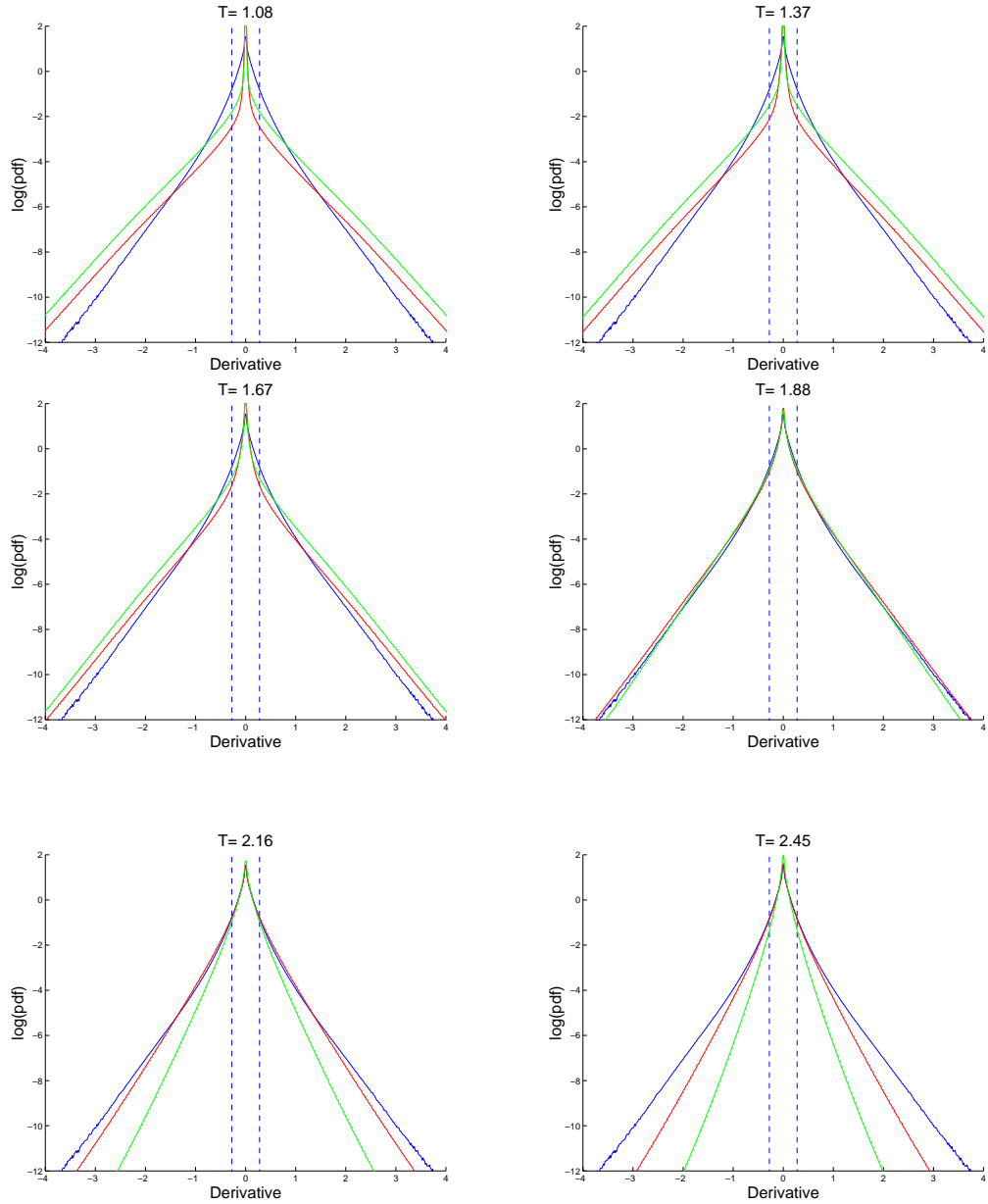


Figure 3.8: The derivative statistics. Blue: from natural images. Red: simulation at scale 1 ($\text{pixel} = \mu(8 \times 8 \text{ subpixels})$). Green: simulation at scale 2 ($\text{pixel} = \mu(16 \times 16 \text{ subpixels})$).

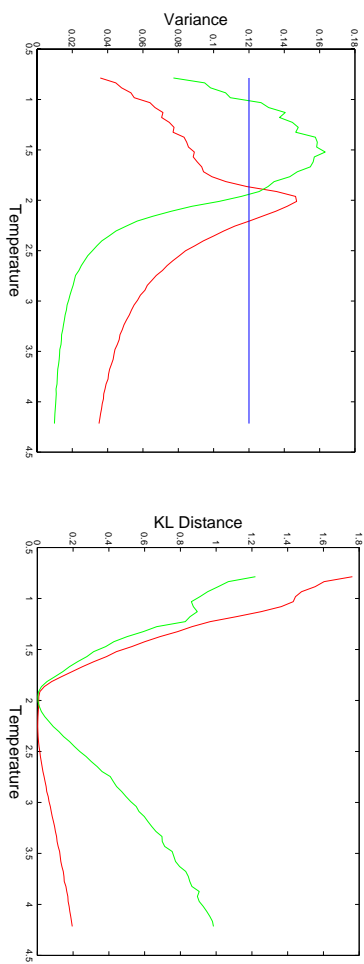


Figure 3.9: Left: Variance of the derivative statistics. Right: KL distance between the derivative statistics calculated from natural images and that from model. Blue: from natural images. Red: simulation at scale 1 ($pixel = \mu(8 \times 8 \text{ subpixels})$). Green: simulation at scale 2 ($pixel = \mu(16 \times 16 \text{ subpixels})$).

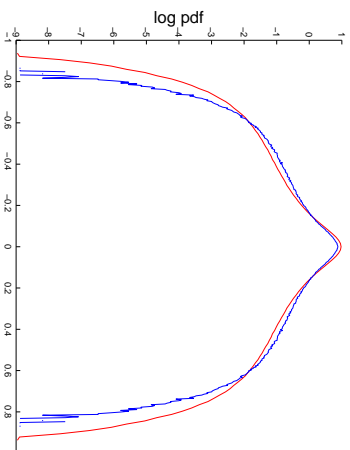


Figure 3.10: The distribution of the object factor (for derivative filter) at the critical temperature and two scales. Red: scale 1. Blue: scale 2

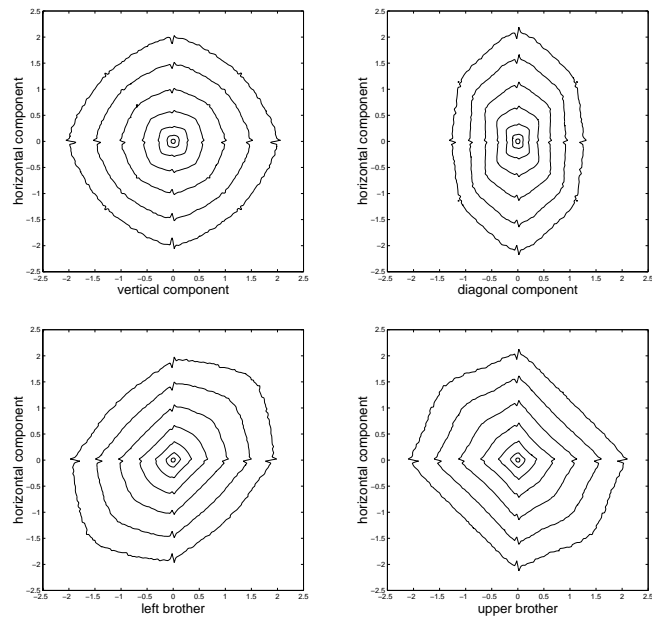


Figure 3.11: Haar Wavelet Coefficients of the Simulations at Temperature $T=1.88$

Chapter 4

Applications

4.1 Image Compression

We estimate the limit of (lossless) image compression from the statistics we calculated.

(1)DCT

The standard JPEG image compression algorithm utilizes the marginal distribution of single AC coefficients. For the DC coefficient, the distribution of the difference of adjacent DC coefficients is used. The total bits needed to code a 8×8 patch of an image according to this compression method is:

$$\sum_{i=1}^{63} H(p_i) + H(q_0)$$

where $H(p_i)$'s are the differential entropy of AC coefficients(see the last column of table 2.11 at page 71). $H(q_0)$ is the entropy of the difference of two adjacent DC coefficients. Notice that the DC coefficient of DCT is simply 8 times the average of the 8×8 block. The entropy of the difference between two adjacent 8×8 blocks is 0.02 (see section 2.3 on page 19). Hence $H(q_0) = 0.02 + \log_2(8) = 3.02$. Adding these numbers up, we get that the total number of (differential)bits needed for coding a 8×8 block is: -24.34 bits, or -0.38 bit/pixel. Notice a single pixel statistics has the entropy 1.66(see table 2.1), so the number of true bits saved is: $1.66 - (-0.38) = 2.04$ bit/pixel

(2)Haar 3D

The differential entropy of the horizontal, vertical and diagonal filter reactions is $E_0 = -2.48$, calculated from the 3D joint distribution of the Haar filter reactions(see section 2.4.4). Let's assume that the wavelet coefficients at different scales or different locations are independent. For a $N \times N$ image, the total entropy for the wavelet coefficient at first level will be:

$$\frac{N \times N}{4} E_0$$

For the second scale, if we assume scale invariance property of images, and assume the Haar filters are normalized so that the L_1 norms are 1, then we should have the same differential entropy E_0 . Since Haar filters are actually normalized according to L_2 norm, the differential

entropy is $E_0 + \log_2(8)$, so the total entropy for the wavelet coefficients at second level will be:

$$\frac{N \times N}{4^2}(E_0 + \log_2(8))$$

Repeating this procedure, we get the total entropy at level k is:

$$\frac{N \times N}{4^k}E_0 + \log_2(8)(k - 1)$$

Hence the average entropy of each pixel will be

$$\frac{1}{4}E_0 + \frac{1}{4^2}(E_0 + 3) + \frac{1}{4^3}(E_0 + 3 \cdot 2) + \dots = \frac{1}{3}E_0 + \frac{1}{3}$$

which is about -0.498 bit/pixel, and the number of bits saved: $1.66 - (-0.498) = 2.16$ bit/pixel, which is slightly larger than that from DCT coefficients.

4.2 Noise removal

In this section, we apply the statistics of wavelet coefficients to the application of noise removal of images. Next figure shows an image of boats (we have taken the logarithm of the original intensity). We add an iid Gaussian noise to each pixel to get a ‘dirty’ picture. Our goal is to reduce the noise in the picture. We used the ordinary MAP method, explained next.

First, let’s assume X is a single dimensional random variable, with density $f(x) \sim \exp(h(x))$, and $Y = X + N$, where N is a Gaussian noise $\sim N(0, \sigma)$. Suppose we observed Y , and want to make a ‘best’ guess of X . One way is to calculate the $\operatorname{argmax}_x f(x|y)$, where $f(x|y)$ is the posterior density of X given Y . Since

$$f(x|y) \sim f(x)f(y|x) = f(x)g(y-x)$$

we have the MAP estimation of x :

$$\tilde{x} = \operatorname{argmax}_x (f(x|y)) = \operatorname{argmax}_x (h(x) + \frac{(y-x)^2}{2\sigma^2}) \quad (4.1)$$

Similar method applies when X is a random vector instead of a random variable,

$$\tilde{\mathbf{x}} = \operatorname{argmax}_{\mathbf{x}} (f(\mathbf{x}|\mathbf{y})) = \operatorname{argmax}_{\mathbf{x}} (h(\mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{x})^T) \quad (4.2)$$

where Q is the covariance matrix of the noise N .

Now, given an noise image, we calculate the steerable pyramid decomposition of it with two oriented subbands B_0, B_1 at each scale (as in section 2.5.2). At each scale, let $Y_{i,j} = ((B_0(i,j), B_1(i,j)))$. We assume these coefficient pairs are independent of each other, and apply the MAP estimation given in 4.2 separately on each pair. Here we use the empirical distribution calculated in section 2.5.2 as the prior h . The covariance matrix Q for the Gaussian noise can be calculated from the definition of the filters. Once we get the MAP estimation \tilde{X} for each coefficient pairs, we can then reconstruct the image from them to get the ‘cleaned image’. The result of our experiment is shown in the lower left subplot of figure 4.1.

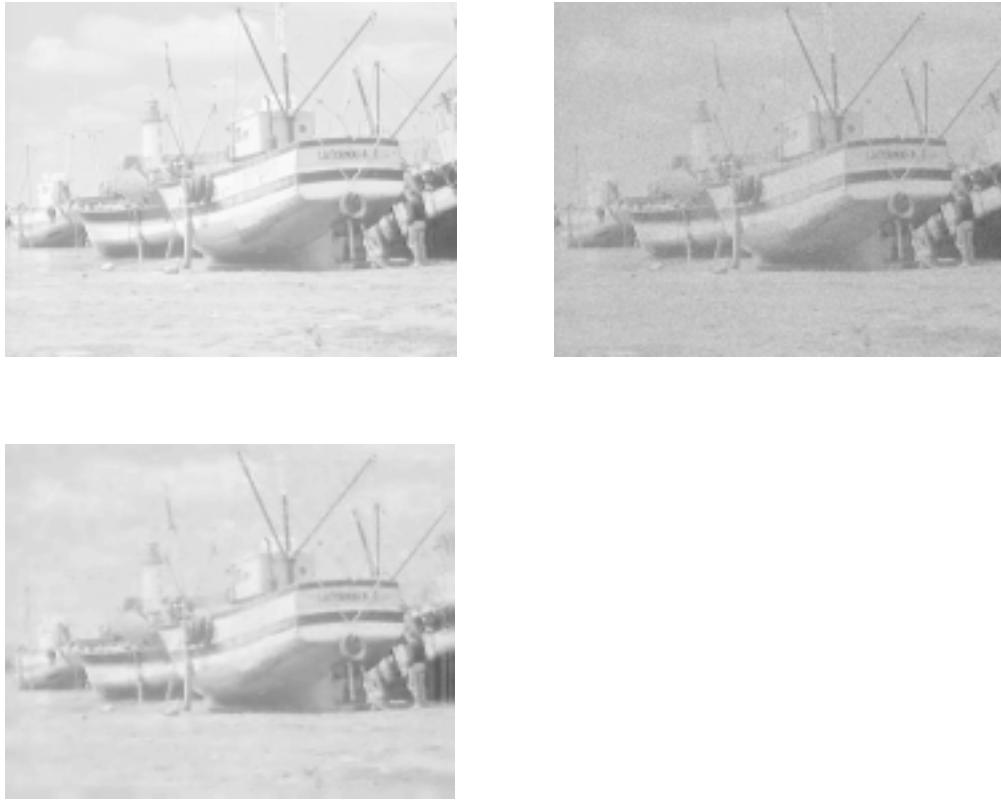


Figure 4.1: Noise removal experiments. From up to down and left to right are the original image, the noisy image and the cleaned image

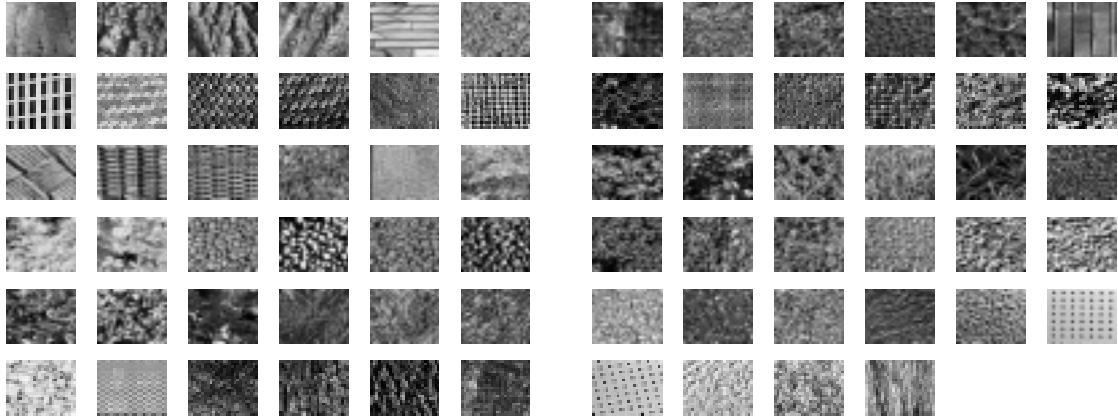


Figure 4.2: The 70 textures we used in our experiment

4.3 Texture Classification

To see how well the statistics we got in section 2.9 capture the important features of vision, we utilize these statistics to do some experiments on texture classification. Figure 4.2 shows the 70 textures used in this experiment.

To measure the similarity of textures, we adopt the idea ‘textons’ in [15]. In our case, we fixed the textons to be the 128 k-means centers we calculated in section 2.9, and denote them by C_1, C_2, \dots, C_{128} . Each one of the texton can also be regarded as a 63-dimensional vector. For a $N \times N$ texture T , we calculated a texton histogram F_T as the follows,

1. Let F_T be a 128 dimensional vector, which is initialized as 0.
2. For each 8×8 patch P of T , there are 63 AC wavelet coefficients in the Haar Wavelet decomposition of P . We normalize these coefficients in the way described in section 2.9. Thus P corresponds to a vector \tilde{P} in R^{63} .
3. Find the texton C_K which is closest to \tilde{P} , and increase the K th bin of F_T by 1.
4. Repeat step 2 and 3 for all 8×8 patches P of T , and we get the histogram F_T .

We measure the similarity of two textures T_1 and T_2 by the ‘distance’ between the two

texton histograms F_{T_1} and F_{T_2} . As in [15], we choose the chi-square test as the ‘distance’:

$$\chi^2(F_{T_1}, F_{T_2}) = \frac{1}{2} \sum_{k=1}^{128} \frac{(F_{T_1}(k) - F_{T_2}(k))^2}{F_{T_1}(k) + F_{T_2}(k)} \quad (4.3)$$

For the first experiment, we subdivide each texture into 16 patches to get in total $16 \times 70 = 1120$ patches. Then pick a single patch from them, we order the other patches according to the similarity(defined above) to the picked patch. If the similarity we defined works perfectly, then the picked patch and the first 15 ‘most similar’ patches should all come from the same texture. Figure 4.3 shows three examples of the result of our experiments. Our algorithm works perfectly in one example, but only works partially well in another , and fails totally in the third example. We should point out that situations as in the third example are very rare in our experiments.

In another experiment, we pick randomly a patch from a texture, and using the same method to see whether we can find which texture it belongs to. The rate that this method succeeds in finding the correct texture is 98.3%.

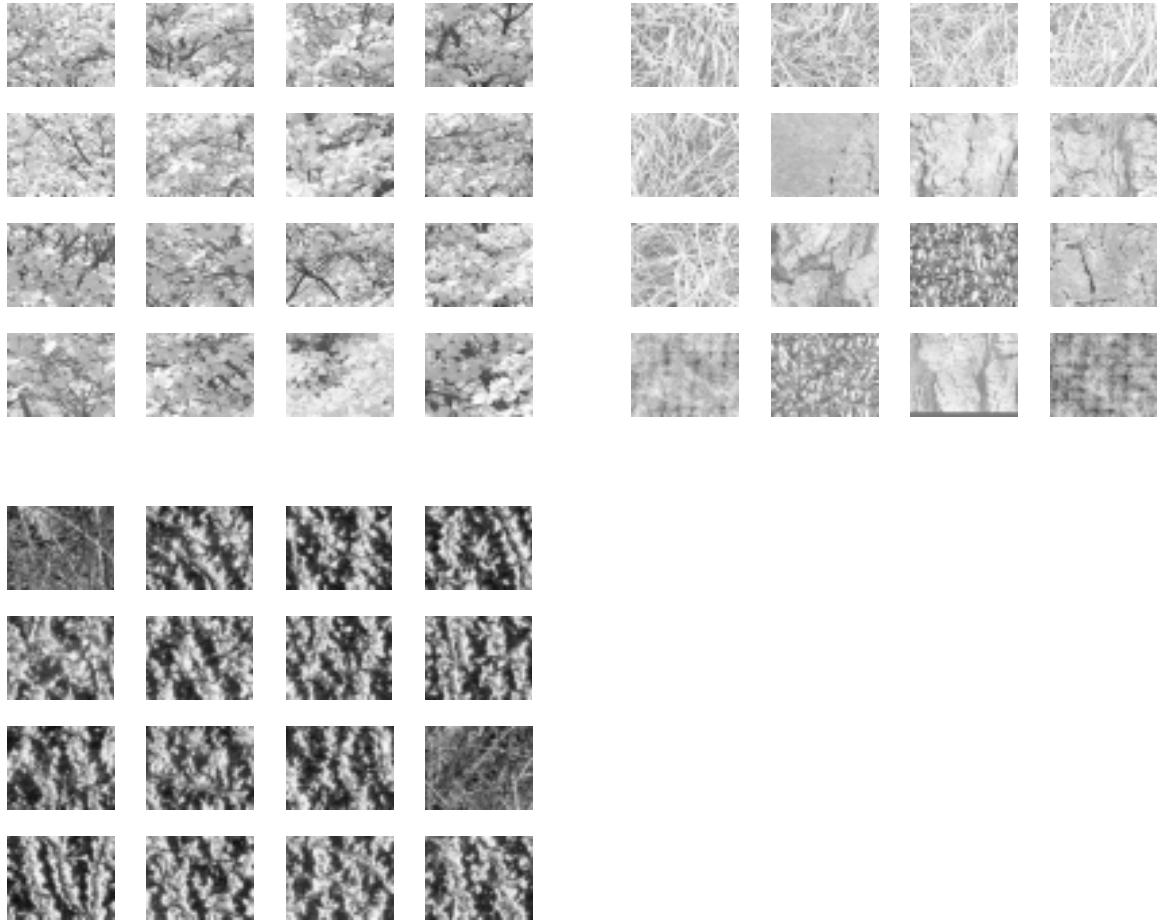


Figure 4.3: three examples of texture classification results. The patch at the upper left corner is the picked patch, the others are the patches which are most 'similar' to the picked patch according to our algorithm.

Chapter 5

Conclusion

In this thesis, we have presented several important statistics from intensity images and range images. We fitted mathematical models to these statistics and explained some phenomena we observed. To see how useful are the statistics we calculated, we applied them to some simple applications.

We found that statistics involving local filters are approximately scale invariant, but there are systematic deviations from scale invariance on a large scale. The scaling behavior of different types of images is quite different, and can be characterized by their ‘anomalous dimension’.

Statistics of any mean-0 filters can be well modeled by ‘generalized Laplace’ distributions and the common statistical features (highly peaked center and heavy tails) are related to the fact that these statistics can be decomposed to two independent factors, one corresponds to the distributions of objects of the world, one corresponds to the intensity or color of objects.

There are significant dependencies between filter reactions at adjacent locations, scales or orientations. The joint distributions of such filter reactions can be quite complicated, yet can still be explained by assuming the independent product of a object factor and a color factor.

The statistics calculated from range images are closely related to those calculated from intensity images. Some of the striking features observed in the Haar filter reactions indicate that range images are much simpler to analyze than intensity images. We believe that segmenting range images from natural scenes and a thorough analysis of the results will lead to a better understanding of the scene geometry of the 3D world, as well as more realistic statistical models for intensity images.

Our simulations of natural images by using Ising-like MRF are successful in duplicating some of the local filter reactions. To make the model work well for more complicated filters, we must reconsider the assumptions we made in section 3.1. For example, assumption 1 is a very strong statement, and it holds well only for small (e.g. 2×2 or 4×4) image patches. One possible way to release the restrictions is to assume 3 or more objects in a image patch, and use more general Potts-like MRF to generate objects.

Bibliography

- [1] P. N. Belhumeur and D. Mumford. "A Bayesian treatment of the stereo correspondence problem using half-occluded regions". *Proc. IEEE Conference Computer Vision and Pattern Recognition, 1992* 506-512
- [2] P. N. Belhumeur. "A Binocular Stereo Algorithm for Reconstructing Sloping, Creased, and Broken Surfaces in the Presence of Half-Occlusion". *Proc. Fourth International Conference on Computer Vision* 431-438
- [3] R. W. Buccigrossi and E. P. Simoncelli, "Image Compression via Joint Statistical Characterization in the Wavelet Domain". *GRASP Laboratory Technical Report #414, Univ. of Pennsylvania*, available at:
<http://www.cns.nyu.edu/~eero/publications.html>
- [4] Z. Chi. "Probability Models for Complex Systems". Ph.D. Thesis, Div. of Applied Math., Brown Univ. 1998
- [5] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1985
- [6] D.J. Field "Relations between the statistics of natural images and the response properties of cortical cells." *Journal of the Optical Society of America A*, 4.2379-2394, Dec 1987
- [7] B.A. Olshausen and D.J. Field "Natural image statistics and efficient coding". *Network: Computation in Neural Systems* 7:333-339,1996
- [8] W.T.Freeman and E.H.Adelson. "The design and use of steerable filters".*IEEE pat. Anal. Mach. Intell.* 13(9):891-906,1991.
- [9] U.Grenander. "Clutter Research 5". Internal Brown Memorandum.
- [10] J.H. van Hateren and A. van der Schaaf. "Independent Component Filters of Natural Images Compared with Simple Cells in Primary Visual Cortex" *Proc.R.Soc.Lond. B* 265:359-366,1998
- [11] J. Huang and D. Mumford. "Statistics of Natural Images and Models". *Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1999* 541-547
- [12] J.Huang, Ann.B.Lee and D. Mumford. "Statistics of range images". Accepted by IEEE Conference on Computer Vision and Pattern Recognition, 2000.

- [13] A.Koloydenko “Modeling Natural Microimage Statistics”. Ph.D. Thesis, Dept. of Mathematics and Statistics, Univ. of Massachusetts at Amherst. 2000
- [14] A. B. Lee, J. Huang and D. Mumford. “Random-Collage Model for Natural Images”. Submitted to *International Journal of Computer Vision*
- [15] J.Malik, S.Belongie, J.Shi and T.Leung. “Textons, Contours and Regions: Cue Integration in Image Segmentation”. Int. Conf. Computer Vision, Sept 1999, Corfu, Greece
- [16] D. Mumford and B. Gidas. “Stochastic Models for Generic Images”. Submitted to *Quarterly Journal of Mathematics*.
- [17] D.L. Ruderman “Origins of Scaling in Natural Images”. *Vision Research* vol 37, No23, 3385-3395, 1997
- [18] D.L. Ruderman and W. Bialek. “Statistics of natural images: scaling in the wood s”. *Phys. Rev. Letter.*, 73:814-817,1994
- [19] E.P.Simoncelli and E.H.Adelson. “Subband Transforms”. *Subband Coding*,P379-383, edited by John Woods, Lluwer Academic Press, 1990
- [20] E.P.Simoncelli and W.T.Freeman. “The Steerable Pyramid: A Flexible Architecture For Multi-Scale Derivative Computation”. 2nd Annual IEEE International Conference on Image Processing. Washington, DC. Oct, 1995
- [21] Ya. G. Sinai. “Self-Similar Probability Distributions”. *Theory of Probability and Its Applications* Volume XXI, 64-80, 1976
- [22] M.J.Wainwright, E.P. Simoncelli. “Scale Mixtures of Gaussians and the Statistics of Natural Images” *Advances in Neural Information Processing Systems 12* S.A. Solla, T.K. Leen an K.-R. Muller(eds.), pages 855–861, MIT Press(2000)
- [23] S Zhu, Y Wu and D Mumford. “Filters, random fields and maximum entropy(FRAME) -towards the unified theory for texture modeling”. *IEEE Conf. Computer Vision and Patt Rec* June 1996