Proceedings of the IEEE Workshop on Statistical and Computational Theories of Vision, Fort Collins, CO, June, 1999. (Published on Web).

Invariant Statistics and Coding of Natural Microimages

Donald Geman[1] and Alexey Koloydenko[2]

Abstract

We search for universal characteristics of the microstructure of natural images. Our data consist of a very large set of 3×3 patches randomly extracted from indoor and outdoor grey level scenes. The patches are grouped into natural equivalence classes ("patterns") based on photometry, "complexity" and geometry. We analyze the stability of the pattern statistics over image sets, resolutions and grey scale distortions. Important aspects of the probability distribution of the patterns, e.g., the dominant masses, are stable in our experiments. We also compare the statistics of the natural patch world with those of artificially generated images; the results are consistent with recently proposed "scaling laws" for the sizes of objects in natural images. These results suggest that well-chosen patch labels might serve as elementary features in pattern recognition and other imaging problems in which the fine structure of the grey level configurations is not essential, and we sketch a computationally efficient way to carry this out using tree-structured vector quantization.

1,2. Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003; Supported in part by the ONR under contract N00014-97-1-0249 and the ARO under MURI grant DAAH04-96-1-0445.

1 Introduction

We investigate the microstructure of natural grey scale images, focusing on 3×3 subimages ("patches") of ordinary photographs of indoor and outdoor 3D scenes - landscapes, urban sights, portraits, etc. Our experiments are based on a huge collection of patches randomly extracted from two databases. Our objective is to determine if there are statistical characteristics of this microworld which are "universal" in the sense of being stable across scenes types, resolutions, and other global image attributes.

One motivation for this work is finding stable, elementary features for image analysis. Decisions about how the patches are grouped - what type of information is preserved - are driven by applications to object recognition and other high-level vision problems. Different objectives might lead to somewhat different coding schemes, say for fractal image compression [13] in the spirit of the "archetype classification" in [4]. In [1] we coded 5×5 patches of *binary* images of handwritten digits. The basic idea was to vector quantize a large random sample by examining just a few of the 25 pixels. The order in which the pixels were observed was determined by maximizing the gain in information of each query. The internal nodes of the resulting binary tree served as local features for recognition. We began this study by wondering how this idea might be extended to greyscale images and one such program is outlined based on tree-structured vector quantization.

We are also motivated by related work on the "statistics of natural images" [7],[14] and on scale invariance in natural images [6],[8],[12], wherein evidence and explanations are provided for universal scaling laws, for instance for the probability distribution of object sizes. In fact, we compared microimage statistics for natural images with those generated artificially from Poisson disk models, sampling the disk radii according to several different densities. After adjusting for the amount of "background," the statistics match surprisingly well in the case of inverse cube laws, as discussed in [6]. There is also work similar in spirit to ours about how edges and wavelets might "emerge" from series expansions for small subimages and comparisons are drawn with biological operations in the primary visual cortex ([2], [9]); in particular, work on Independent Component Analysis (ICA) (see, e.g., [3]) characterizes edges as "the most statistically independent" image features. A rather different characterization emerges from our study: Roughly speaking, edges are "the most probable non-background" microimage configurations.

Suppose all grey level images are quantized to L values. Then there are L^9 different patch types (grey level configurations). But these are useless as image features because their frequencies are extremely low and unstable. Hence some form of grouping is unavoidable. Now from a purely information-theoretic point of view, code lengths should be proportional to negative log frequencies, and the Huffman code achieves this by successively aggregating masses, starting with the rarest ones. However, grouping patches in this way would be quite unnatural, at least in a perceptual sense, and certainly inefficient for vision tasks. There is no reason the resulting classes would capture important invariances, nor would the code length necessarily reflect actual computation.

Instead we group the patches into equivalence classes ("patterns"), first based on photometry, then a measure of complexity, and finally with respect to geometrical symmetries. For example, at a coarse level it is natural to combine patches with similar grey levels at each pixel, and to combine these into larger clusters under some form of rotational symmetry. Mathematically, this corresponds to an action of a symmetry group, whose *orbits* are more statistically stable than individual elements. This results in a coding hierarchy determined by invariance in addition to patch frequencies.

Experiments were performed in order to determine how the frequencies - and other statistics - of the resulting patterns change from one image set or resolution to another. Our main conclusion is that these "statistics" are very stable. For example, regarding the binary Huffman code for the probability distribution of the dominant classes as a summary statistic, we find that this tree does not change if all the images are downsampled, and remains nearly the same from one data set to another.

2 Organization of the Paper

The first step is to factor out redundancy in photometry. Since our primary goal is image interpretation (rather than restoration, enhancement, compression, etc.), we argue that the essential information is preserved under rather coarse quantization. The details are in Section 3. This results in 3×3 matrices assuming only a small number of integer values; we refer to these as "patterns."

Next, in Section 4, we identify equivalence classes of patterns of approximately equal "complexity," measured by the number of distinct values. Ideally, this mapping preserves textures, flat regions, smooth boundaries, etc. Our experiments show that, in nearly all natural scenes, the lowest complexity class ("background") is by far the most probable. Moreover, conditioned on non-background, one class of "moderate complexity" consistently appears to absorb most of the remaining mass. In Section 5, we aggregate patterns of this class based on pseudo-circular rotation and intensity inversion.

We then measure the stability of the statistics of the classes mentioned above. Data on the frequencies of occurrence of the dominant classes obtained from two large databases is presented in Section 6 and in Section 7 we compute a Huffman code for an approximation to the distribution of the random variables naturally associated with the equivalence classes. We compare statistics from one database to another, and with respect to downsampling and photometric transformations, in Section 8. We also compare natural statistics with those corresponding to artificial images in Section 9. Finally, computational issues are discussed in Section 10 and some concluding remarks are made in Section 11.

3 Quantization

Let $G = \{0, .., L-1\}$ and $Q = \{0, .., L_0-1\}$, $L_0 < L$, be the original and quantized intensity scales, respectively. The set of 3×3 matrices with entries in G (resp. Q) is denoted $M(G)$ (resp. $M(Q)$). At minimum, quantization should preserve *relative brightness* and any such map $F : M(G) \to M(Q)$ will be called a *quantization*. Thus, if $\omega \in M(G)$, then

$$\omega_s \leq \omega_t \Rightarrow (F\omega)_s \leq (F\omega)_t,$$

where s and t denote pixels. Obviously any quantization F defines an equivalence relation on $M(G)$; the equivalence classes are $F^{-1}(\eta), \eta \in M(Q)$. We will refer to the elements of

$M(Q)$ as "patterns."

Example 1: Uniform Quantization. Some of our experiments are based on the standard quantization F_u: G is partitioned into L_0 intervals of length L/L_0 and $(F_u\omega)_s = l$ $(l \in Q)$ if ω_s falls in interval $l+1$. Generally, our original images have $L = 256$ grey levels and we take $L_0 = 16$. An example is provided in Figure 1; the left matrix is $\omega \in M(G)$ and the middle matrix is $F_u\omega \in M(Q)$.

$$\begin{array}{|c|c|c|}\hline 37 & 40 & 41 \\ \hline 38 & 80 & 62 \\ \hline 81 & 60 & 63 \\ \hline \end{array} \xrightarrow{F_u} \begin{array}{|c|c|c|}\hline 2 & 2 & 2 \\ \hline 2 & 5 & 3 \\ \hline 5 & 3 & 3 \\ \hline \end{array} \xrightarrow{\pi} \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 0 & 2 & 1 \\ \hline 2 & 1 & 1 \\ \hline \end{array}$$

Figure 1: Uniform quantization followed by factorization

Example 2: Alternate Quantization. We obtained a better coding starting with a different scheme in which the new value given to pixel s depends on all the (nine) grey levels in ω rather than only ω_s. Assign the darkest pixel(s) the value 0, then assign the next brightest pixel(s) the label 0 if the difference is less than L/L_0 and the label 1 otherwise, and so forth. More precisely, suppose again $L = 256$ and $L_0 = 16$, and label the pixels $s_1, s_2, ..., s_9$ to satisfy $\omega_{s_1} \leq \omega_{s_2} \leq \cdots \leq \omega_{s_9}$. Put $(F_a\omega)_{s_1} = 0$. For $i = 2, ..., 9$, set $(F_a\omega)_{s_i} = (F_a\omega)_{s_{i-1}}$ if $\omega_{s_i} - \omega_{s_{i-1}} < 16$ and set $(F_a\omega)_{s_i} = (F_a\omega)_{s_{i-1}} + 1$ otherwise. Unlike uniform quantization, if $(F_a\omega)_s \neq (F_a\omega)_t$ then $|\omega_s - \omega_t| \geq L/L_0$. Figure 2 contrasts F_u with F_a; left to right the three matrices are ω, $F_a\omega$ and $F_u\omega$. More details comparing the two schemes can be found in [10].

32	8	17
11	15	3
29	65	67

0	0	0
0	0	0
0	1	1

2	0	1
0	0	0
1	4	4

Figure 2: Comparison of the two quantization schemes. Left: Original grey levels; Middle: Alternate; Right: Uniform.

4 Complexity

The number of distinct intensities in $F(\omega) \in M(Q)$ is a measure of the complexity of ω. Let $z : M(Q) \to Z \overset{\text{def}}{=} \{0, 1, ..., 8\}$ denote this number minus one. We now consider only "dense" quantization schemes for which z is in fact the maximal value of each quantized subimage, i.e., there are no "gaps" in the quantized intensity scale of any individual subimage. The scheme F_a has this property whereas F_u does not in general. Evidently, any F can be made to have this property simply by replacing the quantized values of each subimage by their natural enumerations $0, .., z$. We refer to this post-quantization procedure as "factorization" and denote it by $\pi : M(Q) \to M(Q)$. For example, the center matrix in Figure 1 is mapped to the righthand one. Obviously $\pi \circ F$ is another quantization. In fact, another characterization of the equivalence relation "$\sim$" on $M(Q)$ determined by π is as follows. For $\eta, \zeta \in M(Q)$:

$$\eta \sim \zeta \Longleftrightarrow \eta_s < \eta_t \Leftrightarrow \zeta_s < \zeta_t \;\; \forall s, t.$$

Finally, notice that z remains well-defined on the quotient space $\overline{M}(Q) \stackrel{\text{def}}{=} M(Q)/\sim$. Henceforth we will assume π is incorporated into F.

Clearly information is lost during quantization; given F, the original representation of a monochrome image by a gray map is now irreversibly replaced by the values of F computed for all (or for a large sample of) pixels. However, we will argue that, for reasonable choices of F and L_0, such as $F = F_a$ and $L_0 = 16$, the gain in generality and simplicity leads to stable microimage coding. Moreover, the amount of remaining information is still extremely rich, hopefully sufficient for recognition and similar tasks. Of course discerning between essential and redundant information for recognition is related to invariance; for instance, certain transformations of the gray scale should not affect the scene interpretation. Later on we will be more precise about measuring photometric invariance and in next section we consider another type of invariance.

We wish to develop a tree-structured representation for $M(G)$ based on the equivalence classes in $\overline{M}(Q)$. In order to be computationally efficient, this entails measuring the frequencies of the patterns and assigning short code words to the most frequent values of F. We will refer to $\{z = 0\} = \{\omega \in M(G) : z(F\omega) = 0\}$ as the "background class"; equivalently, this is the set of patches mapped by F to the null pattern. The patterns in $\{z = 1\}$ are binary matrices; there are $2^9 - 2 = 510$ of them since $\eta_s \equiv 1$ is identified with $\eta_s \equiv 0$ under π. In terms of frequencies, the classes $\{z = 0\}$ and $\{z = 1\}$ dominate the natural microworld (see Section 6).

5 Geometric Invariance

Most of the non-null patterns in $\overline{M}(Q)$ (even those with $z = 1$) are individually very rare in most images. Since there are a great many of these, a direct, bottom-up Huffman coding of the distribution of patterns is impractical. Also, this distribution is likely to vary drastically from image to image. Hence it does not seem feasible to code patches based solely on photometry and we will further aggregate patterns (and hence patches) based on geometry in the hope of obtaining a more stable and invariant representation.

We identify two patterns in $\overline{M}(Q)$ that can be obtained from each other by one of the eight "pseudo-circular" rotations. Motivated by the conservation of image information under intensity inversion, we also identify two patterns which match under the map $\{\eta_s\} \longmapsto \{z(\eta) - \eta_s\}$. In Figure 3 we display three patterns in $\{z = 1\}$ which are identified by rotation and inversion. The quotient set resulting from rotation and inversion is denoted by Y and the map from $\overline{M}(Q)$ to Y is denoted by y.

Some of the equivalence classes corresponding to binary patterns in $\overline{M}(Q)$ correspond roughly to "edges" (see Figure 5). Obviously, if such patterns are coded independently of "orientation", i.e., if only the elements of Y are labeled, there will be a severe loss of information for most applications. Consequently, a finer coding will eventually be indispensable, even if the statistics of underlying classes are less robust than those of the compound classes represented by Y.

0	0	1
1	0	1
1	0	1

1	0	0
1	0	1
0	1	1

1	1	0
0	1	0
0	1	0

Figure 3: Illustration of the rotation (lower left) and inversion (lower right) symmetries

6 Frequency Data

We are going to estimate the probability laws of various random variables. What are the underlying probability spaces? At the image level, it is $(\mathcal{I}, P_{im})$, where $\mathcal{I}$ is the sample space of all digital grey scale images of natural scenes (say currently stored on the world wide web) and P_{im} is the empirical probability measure. In effect, P_{im} is uniform on $\mathcal{I}$. At the patch level, the space is $(M(G), P)$, where P denotes the probability distribution on $M(G)$ induced by P_{im} by taking all patches from all images in $\mathcal{I}$. Given a quantization F, let $\mathcal{Z}(\omega) = z(F\omega)$ and $\mathcal{Y}(\omega) = y(F\omega)$. (Recall that we assume factorization π is incorporated into F.) The random variable $\mathcal{Z}$ assumes values in $Z = \{0, ..., 8\}$ and the random variable $\mathcal{Y}$ assumes values in Y, the set of equivalence classes in $\overline{M}(Q)$ after factoring out rotation and inversion.

One could "standardize" $(\mathcal{I}, P_{im})$, and hence $(M(G), P)$, in various ways in the hope of increasing stability. Specifically, any preprocessing (e.g., histogram equalization) of the grey scale images will in general alter P, and this might stabilize the distribution on patterns, and in particular the laws of $\mathcal{Z}$ and $\mathcal{Y}$. For example, we rescaled the intensity values to the full dynamic range after cutting off the tails. Although we observed a better match in comparing certain point estimates between data sets and after downscaling, a corresponding increase in the estimation errors rendered the overall results inconclusive. Thus, all the experiments are based on the raw image data - there is no preprocessing.

Our experiments involve two collections $\mathcal{D}_1$ and $\mathcal{D}_2$ of photographic images, each containing forty pictures originally quantized to $L = 256$ grey levels. The photographs are diverse in origin, scale, and quality; there are landscapes, urban sights, people, animals and even a galaxy. Collection $\mathcal{D}_2$ consists of 196×128 thumbnail pictures and is part of the database used in [9]. It should be noted that $\mathcal{D}_2$ is somewhat special (e.g., all the images have very low contrast) and perhaps not very representative of natural images. Nonetheless, we assume the 80 images represent a random sample from $\mathcal{I}$.

We first gathered a large sample of patches from each collection and computed the relative frequencies of all the basic photometric equivalence classes, i.e., of all the patterns in $\overline{M}(Q)$. We did this for both uniform quantization, F_u, and the alternate scheme, F_a. Qualitatively, the results are similar, but there are some differences: F_a gives somewhat more stable statistics and the classes generated are more in keeping with our intuitive notions of "background" and "edge." For example, a patch cut from a slowly varying planar image surface may not fall into the background class $\{z = 0\}$ under F_u. But it would under F_a provided every pixel was within L_0 grey levels of at least one other pixel. In contrast, two pixels differing by only one grey level might be discriminated under F_u. Consequently, many more patches are regarded as background under F_a; specifically, the null pattern has (estimated) mass .22 (the second highest value after $P(\mathcal{Z} = 1)$) under F_u and .65 (the highest value) under F_a. Figure 4 displays 36 randomly chosen "background" patches under F_a and F_u. There is clearly a broader spectrum of "structureless" subimages in the F_a collection.

Also, the set of patches corresponding to "edge" patterns (e.g., $\eta_s = 1$ for pixels in the top row and $\eta_s = 0$ elsewhere) seem to be more coherent under F_a. Figure 5 helps to visualize this difference by displaying randomly selected patches corresponding to the "edge" pattern described above. From now on, we will focus on F_a only.

The estimated distribution of $\mathcal{Z}$ based on patches from $\mathcal{D}_1$ is presented in Table 1. The frequencies for each $z = 0, ..., 8$ are averages μ over all forty images, but the individual forty averages allow us to compute a standard error σ. The *relative* errors (σ/μ) range from 0.28 to 0.86 for all but the rarest values. Conditioning on $\{\mathcal{Z} > 0\}$ resulted in a significant error decrease.

z	0	1	2	3	4	5	6	7	8
$\hat{P}(\mathcal{Z} = z)$	0.6461	0.1733	0.0936	0.0515	0.0241	0.0088	0.0022	0.0004	0
$\hat{P}(\mathcal{Z} = z \mid \mathcal{Z} > 0)$	0	0.5239	0.2599	0.1327	0.0588	0.0194	0.0047	0.0006	0

Table 1: Estimated distribution of $\mathcal{Z}$

Consider now the 510 patterns in $\overline{M}(Q)$ with $\mathcal{Z} = 1$. Of these, 114 have contiguous sets of 1's (and 0's) around the center pixel; for example, there are 16 patterns with exactly one 1 or exactly one 0 on the perimeter, another 16 with exactly two 1's or two 0's, and so forth. Apart from the background, these patterns (individually as well as cumulatively) totally dominate the largest frequencies.

For example, in one sample, eleven of the sixteen *contiguous* patterns with two 1's or 0's on the perimeter appear in the 40 patterns with the highest frequencies, whereas only one "disconnected" pattern (all 1's except two 0's in diagonal corners) is among them.

Let $\mathcal{Y} = 1, ..., 8$ denote the eight equivalence classes in Y displayed in Figure 6. In this way, the class index corresponds to the number of pixels on the perimeter with values different from the center.

The estimated probabilities of $P(\mathcal{Y} = k)$ for $k = 1, ..., 8$ are given in Table 2. The

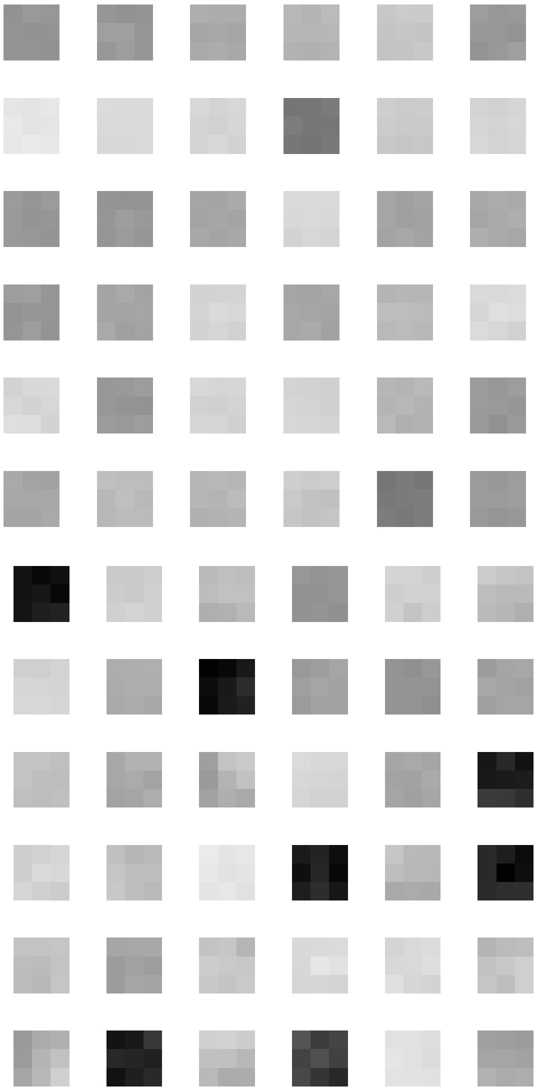

Figure 4: “Background” patches. Top: Uniform quantization. Bottom: Alternate quantization.

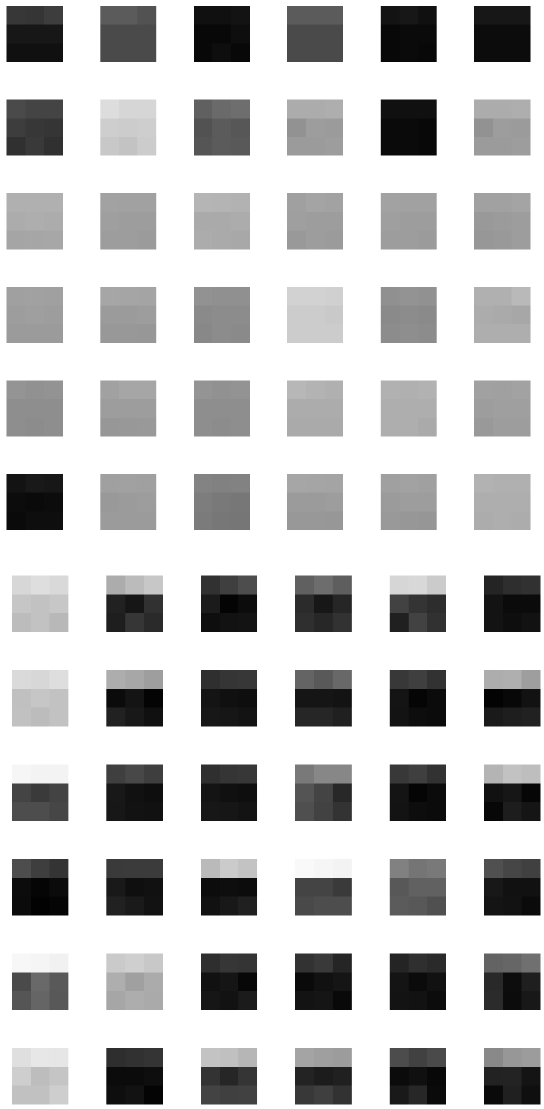

Figure 5: “Edge” patches. Top: Uniform quantization. Bottom: Alternate quantization.

$$\begin{array}{|c|c|c|}\hline 0&0&0\\\hline 0&0&1\\\hline 0&0&0\\\hline\end{array}\quad\begin{array}{|c|c|c|}\hline 0&0&1\\\hline 0&0&1\\\hline 0&0&0\\\hline\end{array}\quad\begin{array}{|c|c|c|}\hline 0&1&1\\\hline 0&0&1\\\hline 0&0&0\\\hline\end{array}\quad\begin{array}{|c|c|c|}\hline 1&1&1\\\hline 0&0&1\\\hline 0&0&0\\\hline\end{array}$$

$$\begin{array}{|c|c|c|}\hline 1&1&1\\\hline 1&0&1\\\hline 0&0&0\\\hline\end{array}\quad\begin{array}{|c|c|c|}\hline 1&1&1\\\hline 1&0&1\\\hline 1&0&0\\\hline\end{array}\quad\begin{array}{|c|c|c|}\hline 1&1&1\\\hline 1&0&1\\\hline 1&1&0\\\hline\end{array}\quad\begin{array}{|c|c|c|}\hline 1&1&1\\\hline 1&0&1\\\hline 1&1&1\\\hline\end{array}$$

Figure 6: Representatives of the eight classes $\mathcal{Y} = 1, .., 8$

relative errors (apart from the rarest classes) are 0.48-0.76; again, they are significantly reduced by conditioning. Conditioned on non-background, the eight classes consume about 40% of the total mass of the distribution of $\mathcal{Y}$.

k	1	2	3	4	5	6	7	8
$\hat{P}(\mathcal{Y} = k)$	0.0706	0.0183	0.0333	0.0049	0.0025	0.0016	0.0044	0.0033
$\hat{P}(\mathcal{Y} = k \mid \mathcal{Z} > 0)$	0.2225	0.0596	0.1152	0.0159	0.0079	0.0042	0.0111	0.0087

Table 2: Estimated distribution of $\mathcal{Y}$ for the first eight classes

7 A Coarse Representation

In this section we present a coarse Huffman coding of the joint distribution of $(\mathcal{Z}, \mathcal{Y})$. We regard the Huffman tree as a summary statistic for measuring the stability of the resulting patch labeling across images and transformations; for example, one can compare the trees obtained before and after downsampling the images. Short code words in the Huffman tree do not correspond to a small amount of computation in determining the label of one patch or all the patches in an image. Rendering the labeling computationally efficient is quite another matter; see Section 10.

We assign distinct labels to the elements of a partition of the range of $Z \times Y$ into ten subsets. The first five are determined by $\mathcal{Z}$ alone: $\mathcal{Z} = 0, 2, 3, 4$ and $\mathcal{Z} \geq 5$. The second five partition the event $\mathcal{Z} = 1$ into five groups: $(\mathcal{Z}, \mathcal{Y}) = (1,1), (1,2), (1,3)$, $(\mathcal{Z}, \mathcal{Y}) \in \{1\} \times \{4,5,6,7,8\}$ and $(\mathcal{Z}, \mathcal{Y}) \in \{1\} \times Y \setminus \{1,2,3,4,5,6,7,8\}$. Thus the tree in Figure 7 has ten terminal nodes. For example, the code word 0 is assigned to $\{\mathcal{Z} = 0\}$, the event with the largest mass, and $(1,0,0)$ is assigned to $\{\mathcal{Z} = 1, \mathcal{Y} = 1\}$.

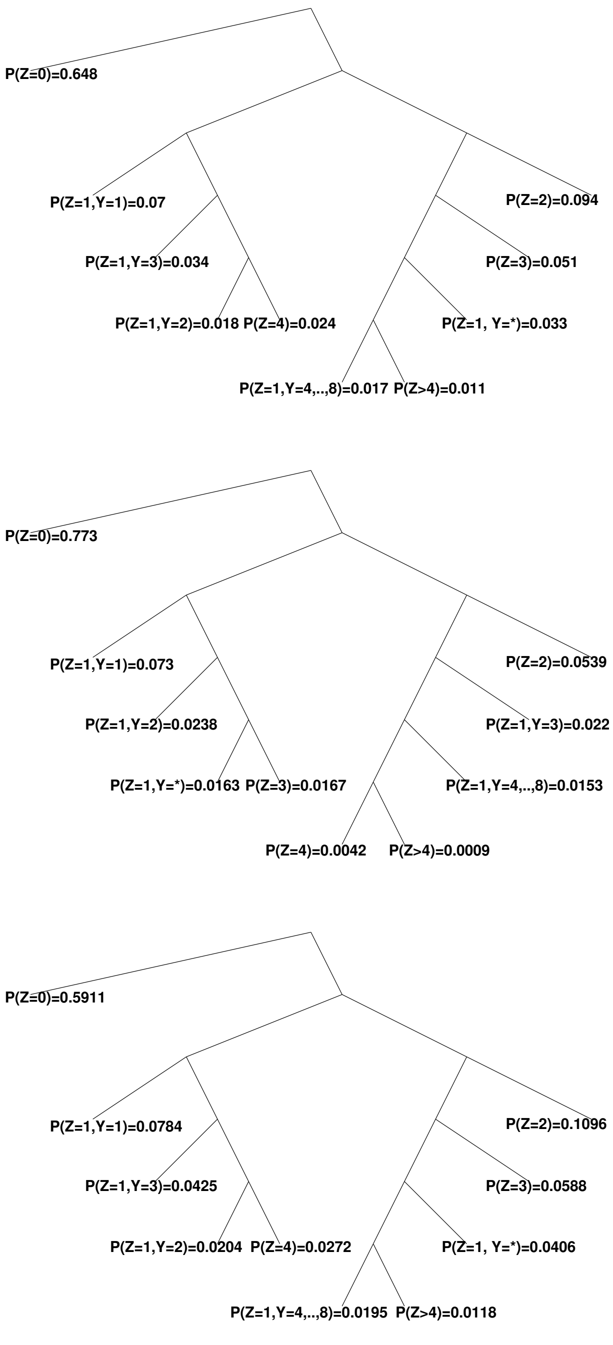

Figure 7: A Huffman coding of the $(\mathcal{Z}, \mathcal{Y})$ distribution. Top: The first data set. Middle: The second data set. Bottom: The first data set after downscaling.

8 Experiments on Invariant Statistics

8.1 Stability Across Images

We repeated all the computations with the second image set, $\mathcal{D}_2$. Typical results are presented in Figure 8; further ones can be found in [10]. The graph of the conditional distribution of $\mathcal{Z}$ given $\mathcal{Z} > 0$ is visually similar to that of $\mathcal{Z}$, and similarly the marginal distribution of $\mathcal{Y}$ is similar to its conditional distribution given $\mathcal{Z} > 0$. As one can see, stability extends beyond the order statistics to the individual numerical values. We found this to be rather surprising in view of the relatively small sample of images.

We also computed the Huffman code based on the coarse partition of the range of values of the pair $(\mathcal{Z}, \mathcal{Y})$ described in the previous section. The result is given in the middle of Figure 7 and should be compared with the tree on top. The tree is *not* the same, but closer inspection reveals many similarities.

8.2 Downscaling

Natural images are commonly believed to possess statistical scale invariance as, for example, discussed in [6]. To assess this claim, and the sensitivity of our coding to image resolution, we repeated the computations with the first image set after downscaling the forty images to half their sizes in both dimensions (by uniform averaging over disjoint 2×2 blocks and subsampling). In Figure 9 we illustrate the effect on the distributions of $\mathcal{Z}, \mathcal{Z}|\mathcal{Z} > 0$ and $\mathcal{Y}$. There is clear invariance to scaling. (The considerable increase in the standard error of $P(\mathcal{Y} = 3)$ is an inevitable consequence of the downscaling procedure.) In fact, the Huffman code is identical at the two scales; see Figure 7.

8.3 Photometric Invariance

Given a greyscale transformation, it is useful to distinguish between two types of photometric invariance: *strong sense*, meaning that individual patches retain their code values and *weak sense*, meaning that only the *probability distribution* of the microimage coding remains the same. Strong invariance is perhaps an unrealistic goal. We made the (global) linear greyscale transformation

$$I_s \longrightarrow \alpha I_s + \beta$$

(truncating at 0 and 255) for all the images I in the database for selected values of α. The quantization F_a is (nearly) independent of β. In Figure 10 we display the fraction of patch codes which change as well as the estimated probabilities of the events $\{\mathcal{Z} = 0\}$ and $\{\mathcal{Z} = 1\}$. As expected, many individual codes change as we shrink or expand the greyscale, but the probabilities of these events (i.e., the total number of each type) remains fairly stable.

9 Disk-Based Image Models

We made several sets of 40 artificial 256×256 images according to a stochastic image formation model similar to the ones proposed in [11] and [12] in order to help explain the

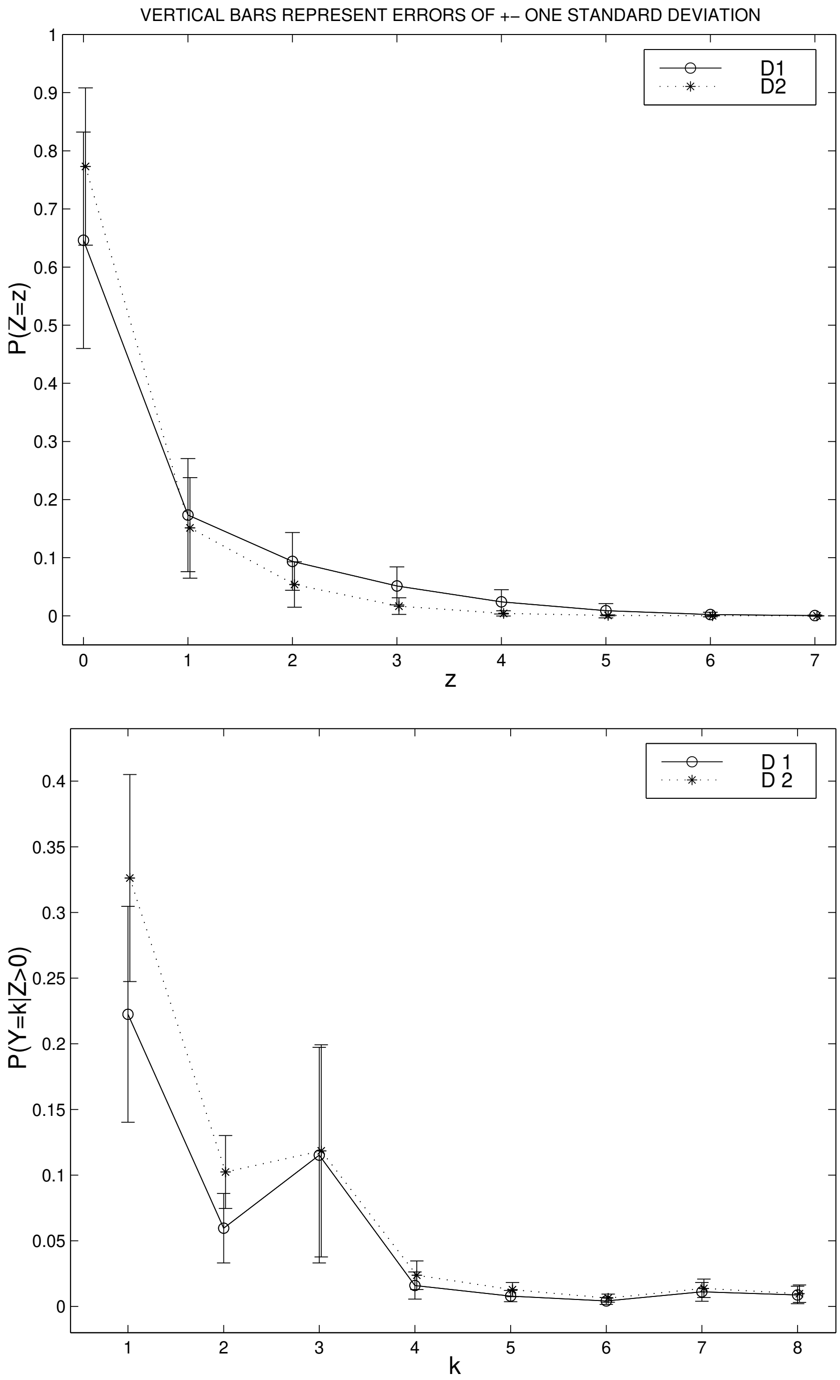

Figure 8: Comparison of statistics for databases $\mathcal{D}_1$ and $\mathcal{D}_2$. Top: Distribution of $\mathcal{Z}$. Bottom: Conditional distribution of $\mathcal{Y}$ given non-background.

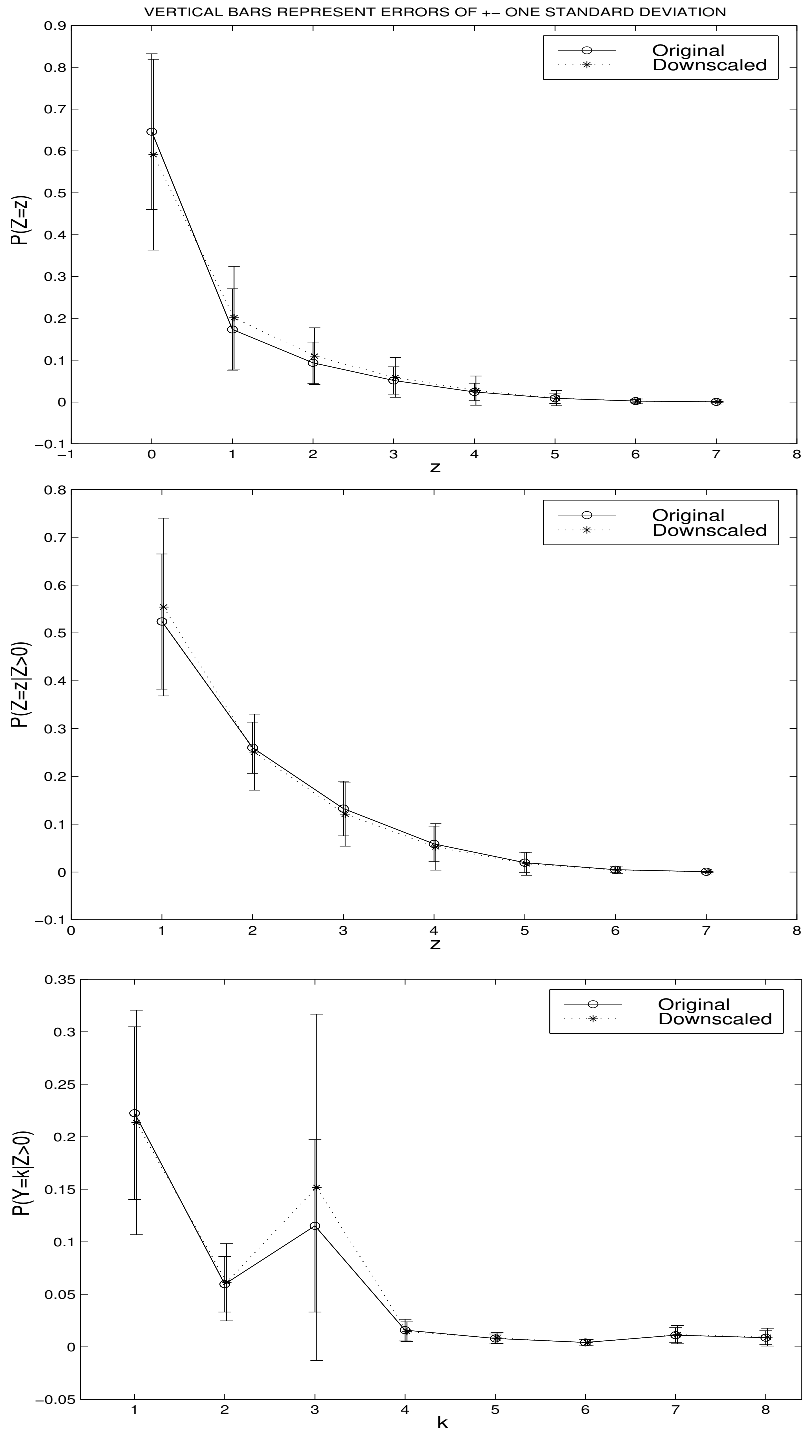

Figure 9: Comparison of original and downscaled statistics. Top: Distribution of $\mathcal{Z}$. Middle: Conditional distribution of $\mathcal{Z}$ given non-background. Bottom: Conditional distribution of $\mathcal{Y}$ given non-background.

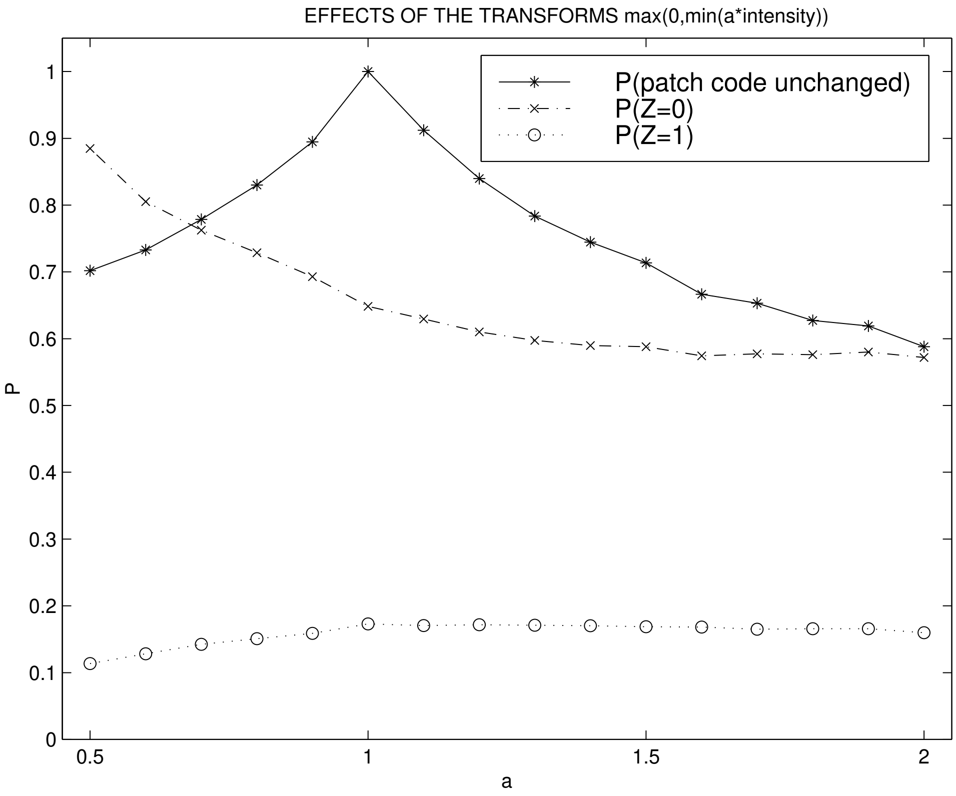

Figure 10: Sensitivity to re-scaling the grey levels

origins of scaling in natural images. First, points are uniformly scattered over the lattice with density equal to one point per square pixel. Each point serves as the center of a disk whose radius R is randomly selected according to the density

$$f(r;\theta) = \begin{cases} (\theta - 1)r^{-\theta} & r \geq 1, \\ 0 & r < 1, \end{cases}$$

where $\theta > 1$ is the parameter of interest. Finally, disks are "colored" with an intensity value chosen uniformly from $\{0, ..., 255\}$. An example is given in Figure 12 for $\theta = 3$.

There are several convincing arguments that the sizes of objects in natural images scale like r^{-3}; see the analyses in [6] and [8]. We computed the same statistics for $\mathcal{Z}$ and $\mathcal{Y}$ for the image model with $\theta \in \{2, 2.5, 3, 3.5\}$ that we computed earlier for real images. The best match occurred with $\theta = 3$, and some comparisons of the $\theta = 3$ data with natural data are displayed in Figure 11. Given the simplicity of the model, and *after conditioning on non-background,* the resemblance is rather striking.

10 Coding: Vector Quantization of Patches

One motivation for this work was to find informative primitive features for applications (e.g., shape classification) in which the fine greyscale structure needn't be preserved. Suppose every patch ω has been assigned a unique label $L(\omega)$ and that $L(\omega)$ depends only on the photometric equivalence class of the patch ω, i.e., only on the pattern $\eta = F_a\omega$ in $\overline{M}(Q)$. For example, the labels might correspond to the partition of size ten of the $(\mathcal{Z}, \mathcal{Y})$ values mentioned in Section 7, or to some other decomposition of $\overline{M}(Q)$ based on geometry. Similar ideas were very effective in classifying binary images of handwritten digits [1] and preliminary experiments on classifying grey level images of leaves are promising [10].

However, assigning a label to every 3×3 patch in a large image might be computationally difficult. The Huffman code tree was merely a summary statistic in order to measure the stability of the patch labels; the expected depth in that tree is not a measure of efficiency; indeed, the amount of computation required to perform each split is not even incorporated. In practice, we want to determine $L(\omega)$ by asking a small number of very simple questions about the grey level configuration ω.

Consider queries of the form $X_{st}(\omega) = 1$ if $\omega_s - \omega_t \geq L/L_0$ and $X_{st}(\omega) = 0$ otherwise, where s and t are two distinct pixels in the 3×3 neighborhood which defines the patch. Let $\mathcal{X}$ denote this set of $9 \times 8 = 72$ such binary functions. It is intuitively clear, and easy to show, that $F_a\omega$ is determined by $\mathcal{X}(\omega)$ (but not vice-versa). We could then quantize the patch world by constructing a binary code tree based on $\mathcal{X}$. Still, the amount of computation necessary to determine a patch label might be rather large, although much smaller than *separately* checking for each label type directly from the definitions.

Instead, we consider a lossy scheme and regard patch labeling as a classification problem itself in which the "true" classes are the labels and some errors are tolerated. For example, the exact labels might be replaced by the terminal nodes of a binary decision tree T using the queries in $\mathcal{X}$ as splitting rules. Ideally, the mean path length of T (which is now truly proportional to computation) would be small and the terminal nodes of T would

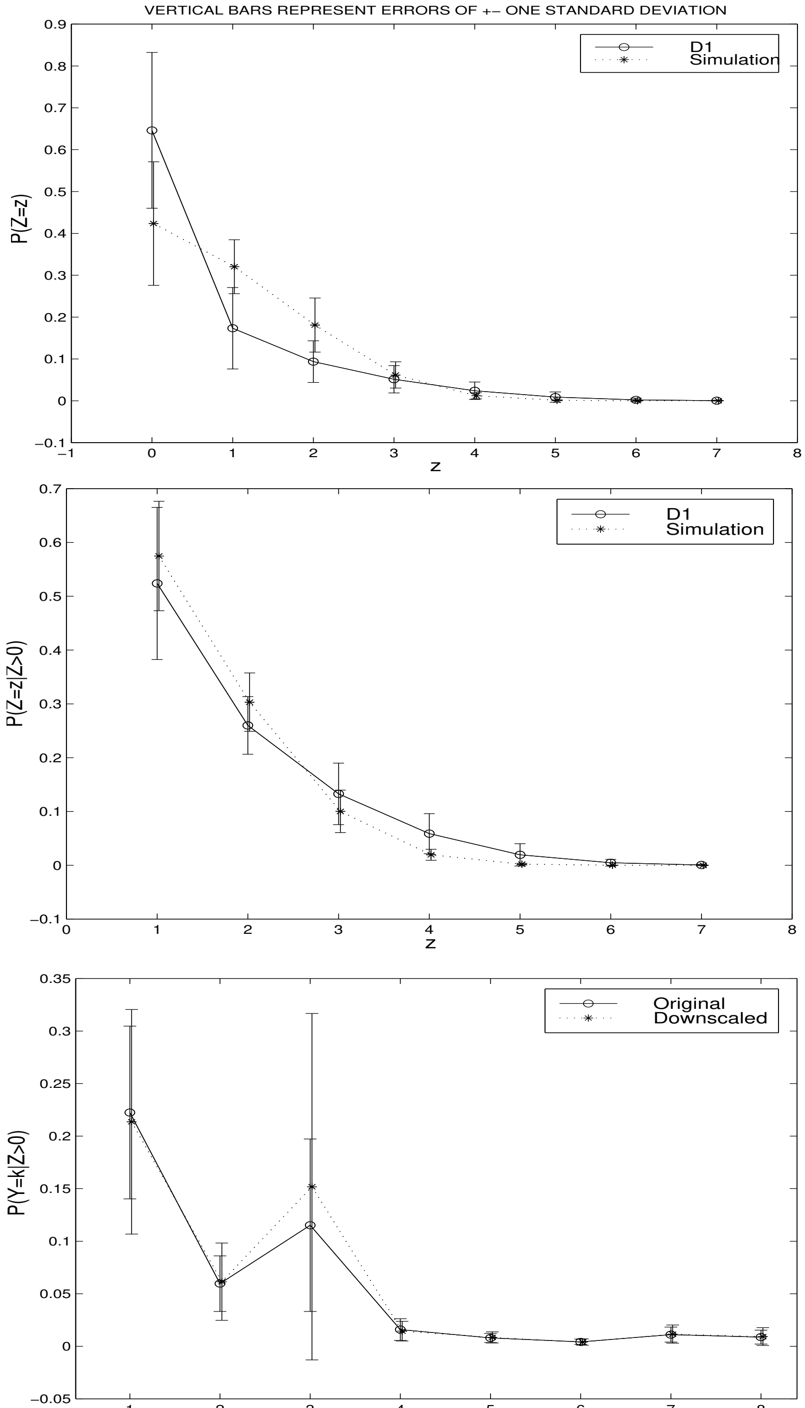

Figure 11: Comparison of statistics for real and simulated images. Top: Distribution of $\mathcal{Z}$. Middle: Conditional distribution of $\mathcal{Z}$ given non-background. Bottom: Conditional distribution of $\mathcal{Y}$ given non-background.

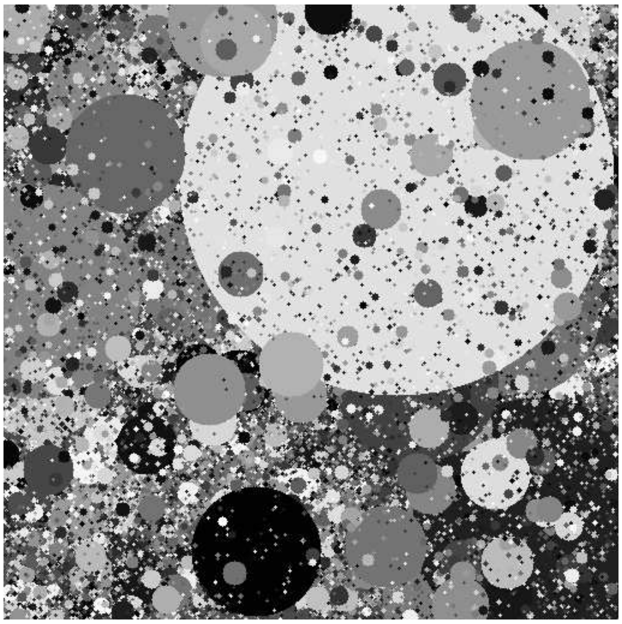

Figure 12: A 512×512 simulated "disk" image

be nearly homogeneous with respect to L. One could construct T with standard greedy algorithms based on stepwise entropy reduction (e.g., CART [5]) or, more ambitiously, by accounting for both mean accuracy *and* mean path length using global optimization. In either case, each ω is finally assigned a variable-length binary "code" $T(\omega)$ determined by the terminal node in which it lands. Retaining the internal nodes which are visited leads to a hierarchical scheme. Experiments are underway to assess this program.

11 Conclusion

We have presented a scenario for coding natural microimages based on mapping 3×3 greyscale patches to equivalence classes determined by photometric and geometric criteria. In order to assess the universality of the coding, we extracted a large set of 3×3 patches from two natural image databases. We regarded the set of microimages as a random sample with respect to the empirical probability measure on the space of all patches from "all" natural images. The mappings into equivalence classes are then random variables, whose dominant masses were found to be reasonably insensitive to the particular images from which the patches are drawn and to vagaries in the original greyscale representation. The distribution is also nearly invariant to resolution, allowing one to use the same coding scheme at multiple scales.

These properties suggest that some type of efficient, formal coding of small patches using standard information-theoretic principles might yield valuable elementary features for image analysis. We presented one example of a partition of the patch world and one example of how in practice these labels might be efficiently approximated with tree-structured vector quantization based on simple queries about intensity differences. This is only a first step. More finely drawn and informative groupings are likely to be needed for any particular application; in particular, the orientation of "edge-like" patterns must eventually play an important role. Perhaps a practical coding involves two phases, one coarse-grained but "universal," and the other dedicated to the particular imagery or application. Also, the tradeoff between accuracy and efficiency must be explored in the context of lossy approximations to the true codes. Some of these issues are currently being investigated in the context of classifying natural shapes [10].

References

[1] Y. Amit, D. Geman, and K. Wilder. Joint induction of shape features and tree classifiers. *IEEE Trans. PAMI*, 19:1300–1306, 1997.

[2] H. Barlow. Unsupervised learning. *Neural Computation*, 1:295–311, 1989.

[3] A. J. Bell and T. J. Sejnowski. Edges are the "independent components" of natural scenes. Technical report, Computational Neurobiology Laboratory, The Salk Institute, La Jolla, CA 92037, 1996.

[4] R. D. Boss and E. W. Jacobs. Archetype classification in an iterated transformation image compression algorithm. In Y. Fisher, editor, *Fractal Image Compression - theory and Application*, pages 79–90. Springer-Verlag, New York, 1994.

[5] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA., 1984.

[6] Z. Chi. *Probability Models for Compex Systems*. PhD thesis, Brown University, 1998.

[7] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Optical Soc. America*, 4, 1987.

[8] S. Geman. Invariant binding in composition systems. Technical report, Division of Applied Mathematics, Brown University, 1998.

[9] 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.

[10] A. Koloydenko. *Statistical Learning and Classification of Natural Shapes*. PhD thesis, University of Massachusetts, 1999.

[11] D. Mumford. Stochastic models for generic images. Technical report, Division of Applied Mathematics, Brown University, 1998.

[12] D. L. Ruderman. Origins of scaling in natural images. *Vision Research*, pages 814–817, 1996.

[13] D. Saupe, R. Hamzaoui, and H. Hartenstein. Fractal image compression - an introductory overview. In J. Hart, editor, *Fractal Models for Image Synthesis, Compression, and Analysis*, ACM SIGGRAPH'96 Course Notes. 1996.

[14] S. C. Zhu and D. Mumford. Prior learning and gibbs reaction-diffusion. *IEEE Trans. PAMI*, 19:1236–1250, 1997.