The Nonlinear Statistics of High-Contrast Patches in Natural Images

Ann B. Lee
Department of Physics, Box 1843, Brown University, Providence, RI 02912, USA

Kim S. Pedersen
Department of Computer Science, University of Copenhagen, Universitetsparken 1, DK-2100 Copenhagen Ø, Denmark

David Mumford
Division of Applied Mathematics, Box F, Brown University, Providence, RI 02912, USA

June 2001 (Revised December 2001)

Abstract. Recently, there has been a great deal of interest in modeling the non-Gaussian structures of natural images. However, despite the many advances in the direction of sparse coding and multi-resolution analysis, a description of the full probability distribution of pixels in a neighborhood is still missing. In this study, we explore the space of data points representing the values of 3×3 high-contrast patches from optical and 3D range images. We find that the distribution of data is extremely "sparse" with the majority of the data points concentrated in clusters and non-linear low-dimensional manifolds. Furthermore, a detailed study of probability densities allows us to systematically distinguish between images of different modalities (optical versus range), which otherwise display similar marginal distributions. Our work indicates the importance of studying the full probability distribution of natural images, not just marginals, and the need to understand the intrinsic dimensionality and nature of the data. We believe that object-like structures in the world and the sensor properties of the probing device generate observations that are concentrated along predictable shapes in state space. Our study of natural image statistics accounts for local geometries (such as edges) in natural scenes, but does not impose such strong assumptions on the data as independent components or sparse coding by linear change of bases.

Keywords: Natural image statistics, non-linear sparse coding, pixel-based image models, microimages, clutter, higher-order statistics, geometrically based statistics, high-dimensional probability density estimation

Submitted to the International Journal of Computer Vision, April 28, 2001. Revised, December 12 2001.

1. Introduction

A number of recent attempts have been made to describe the non-Gaussian statistics of natural images (Field, 1987; Ruderman and Bialek,

Division of Applied Mathematics Brown University Technical Reports on Applied Probability, Pattern Theory, and Statistics. APPTS Report #01-3. Available online at http://www.dam.brown.edu/ptg/publications.html

1994; Olshausen and Field, 1996; Huang and Mumford, 1999; Simoncelli, 1999b; Grenander and Srivastava, 2001). The interest for these studies in the computer vision community has been motivated by the search for more realistic priors for applications as diverse as object localization (Sullivan et al., 1999), segmentation (Malik et al., 2001; Tu et al., 2001), image reconstruction (Nielsen and Lillholm, 2001), denoising (Zhu and Mumford, 1998; Simoncelli, 1999a) and compression (Buccigrossi and Simoncelli, 1999).

The research in natural image statistics can roughly be divided into two related directions: Some studies involve analyzing 1D or 2D marginal statistics with respect to some fixed linear basis. Grenander and Srivastava (2001), for example, have shown that one can use a family of Bessel functions to model the 1D marginals of band-pass filtered data from a variety of different types of images. In Wegmann and Zetzsche (1990) and Simoncelli (1999b), the authors use a wavelet basis and uncover complex dependencies between pairs of wavelet coefficients at nearby spatial positions, orientations, and scales. On the other side, there are studies of image statistics which try to find an "optimal" set of linear projections or basis functions in the state space defined by the image data (8×8 patches, for example, define a distribution in $\mathbb{R}^{64}$). The directions in state space are usually chosen according to some higher-order statistical measure reflecting the non-Gaussianity or multi-modality of the projected data density; see e.g. projection pursuit (Huber, 1985; Friedman, 1987), sparse coding (Olshausen and Field, 1996) and ICA (see Hyvärinen (1999) for a survey of ICA and related methods).

Despite the many advances in sparse coding and multi-resolution analysis, we are still short of a description of the *full* probability distribution (as opposed to marginal distributions) of pixels in a neighborhood. Furthermore, so far, there have been few attempts to make the connection between the object structure in the world, and the probability distribution of natural images, more precise.

In this paper, we analyze the state space of local patterns of pixel values. More precisely, we look at the probability distribution of 3×3 patches of optical and range images. These two types of images reflect different aspects of generators (or objects) in the world, as well as differences in image sensor properties. After preprocessing (consisting of subtracting the mean of each patch and then whitening the data), the extracted 3×3 image patches define a distribution in $\mathbb{R}^7$. We address the following questions: "How is the data distributed in the 7-dimensional state space of contrast-normalized patches?" and "Are there any clear qualitative differences between the distributions of data from images of different modalities, e.g. optical versus range images?".

To develop statistically efficient image representations, one needs to better understand how natural data is distributed in higher-dimensional state spaces. It seems that without this knowledge of natural images we are not able to fully take advantage of the sparseness of the state space of the data. From a sparse coding point of view, high-density clusters and low-dimensional manifolds are especially interesting. These type of structures greatly reduce the dimensionality of the problem.

In ICA and related methods, one assumes that there exists a *linear* change of basis (into independent components) that sparsifies the image data. We believe that an analysis of the probability distribution of natural images should be free from such strong assumptions as independent components, or linear decompositions of an image into a few dominant basis images. In reality, the most common rule for image formation is occlusion — which is non-linear — and the state space of image patches is rather complex with many more high-density directions than the dimension of the state space. The complexity of the data can partially be seen in the Haar wavelet statistics of natural images (Huang and Mumford, 1999; Huang et al., 2000). Take, for example, the 3D joint distribution of horizontal, vertical, and diagonal wavelet coefficients of natural images [1]. Figure 1 shows that the equi-probable surface of this distribution has several "hot spots" with 6 vertices along the axes $y = z = 0$, $x = z = 0$, $x = y = 0$ and $x = \pm y = \pm z$; and 8 local maxima around the shoulders $x = \pm y = \pm z$. These cusps are even more striking for range images (Figure 2).

The observed cusps in Figure 1 and Figure 2 show very clearly frequently occurring *local geometric patterns* in pixelated natural images. A simple calculation gives the following 2×2 blocks and their rotations:

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

A similar analysis can be done empirically for Haar wavelet coefficients at adjacent spatial locations in the same subband. Such a study of the 2D joint histogram of these so called wavelet "brothers" will reveal frequent occurrences in 2×4 patches of more complex local geometries best described as blobs, T-junctions, edges and bars (Huang et al., 2000).

We believe that object-like structures in the world and the sensor properties of the probing device generate observations that are concentrated along predictable shapes (manifolds or clusters) in state space. We want to better understand how edges and other image "primitives" (see David Marr's primal sketch (Marr, 1982)) are represented geometrically in the state space of image data. Furthermore, we want to study how empirical data from natural images is distributed statistically with

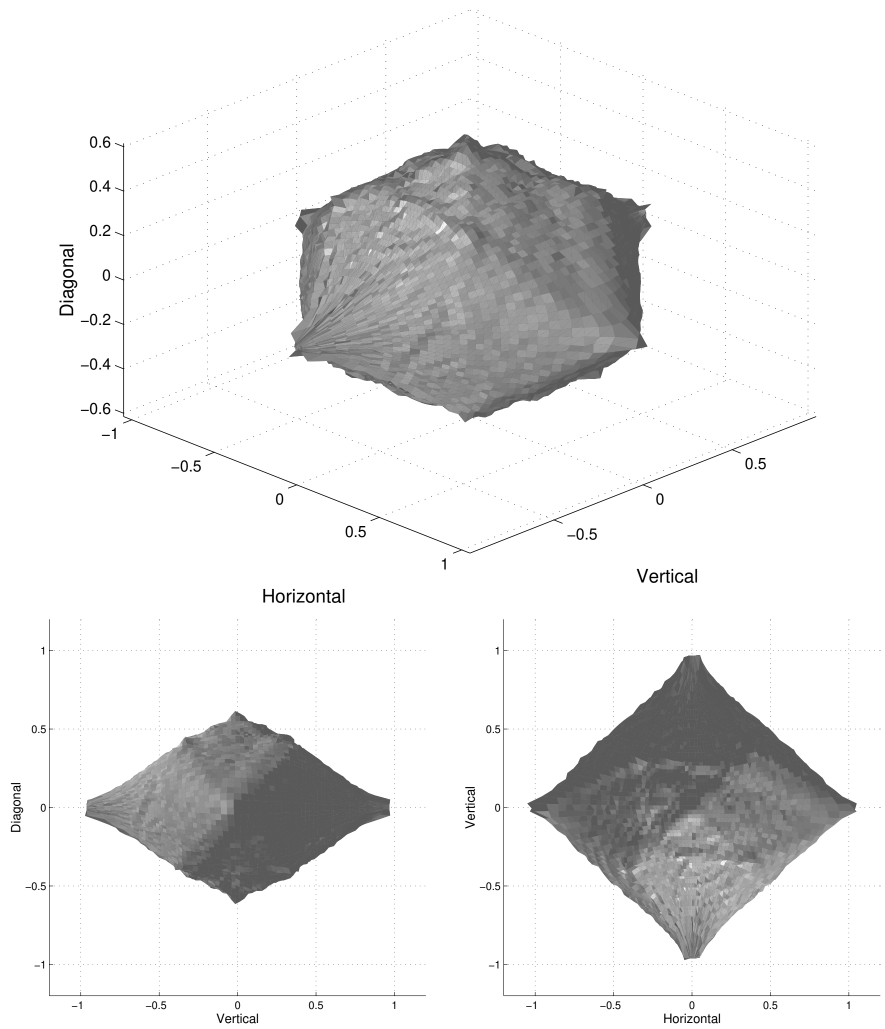

Figure 1. An equi-probable surface of the joint distribution of horizontal, vertical, and diagonal wavelet coefficients in optical images, viewed from three different angles.

respect to the predicted clusters and manifolds. We are, in other words, searching both for a *geometric* and *probabilistic* model in state space of the *basic* primitives of generic images.

In this study, we focus on high-contrast [2] data. It is commonly believed that image regions with high contrast carry the most important content of a scene. Reinagel and Zador (1999) have shown, for natural images with a variety of cognitive content, that humans tend to focus

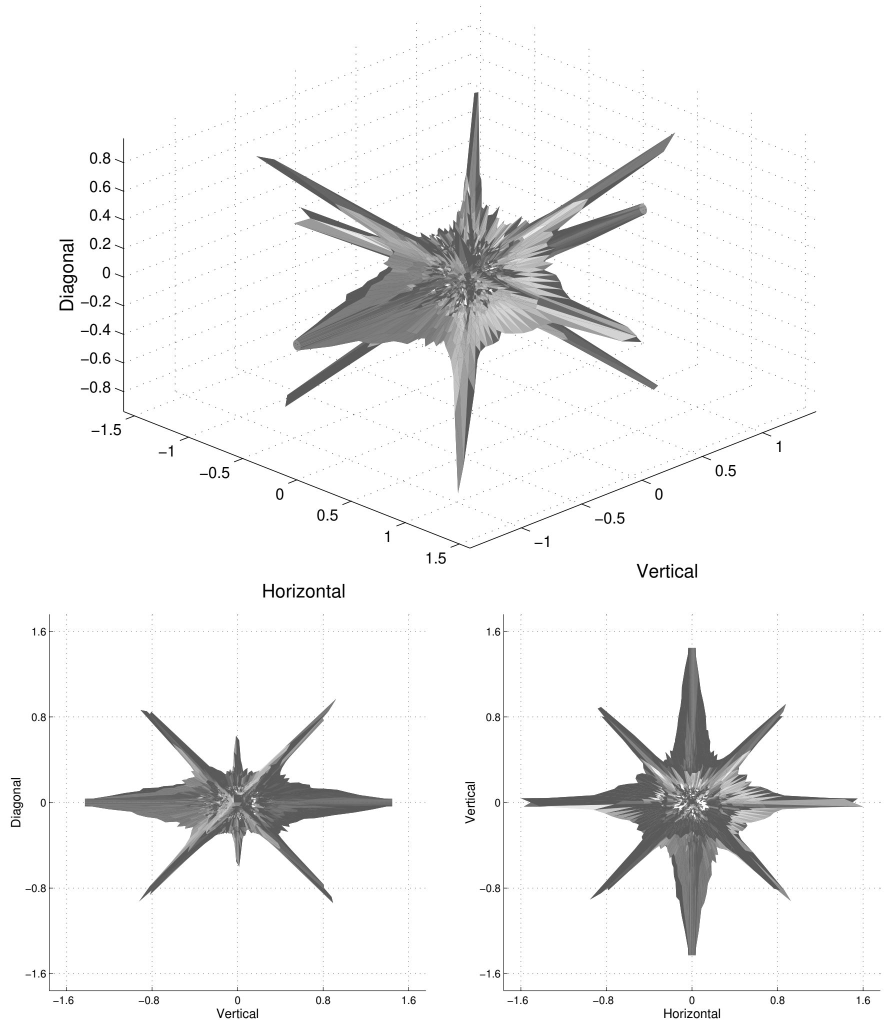

Figure 2. An equi-probable surface of the joint distribution of horizontal, vertical, and diagonal wavelet coefficients in range images, viewed from three different angles.

their eye movements around high-contrast regions — thus significantly biasing the effective input that reaches the early stages of the visual system towards these types of image regions. Furthermore, we tend to believe that high-contrast and low-contrast regions follow qualitatively different distributions, and should be modeled separately. The equi-probable contours mentioned above (see Figure 1 and Figure 2) are highly irregular and star-shaped in the regions far from the origin of the plot. This clearly indicates the non-Gaussian statistics of high-contrast data. The contours near the center part of the plot look different. These contours are more ellipsoidal, which could mean that fluctuations around low-contrast image regions are Gaussian in nature.

This study deals with local patterns of pixel values, although it should be noted that the more general results (regarding the intrinsic dimension and shape of structures in state spaces) generalize to larger image patches and collections of filter responses (see the discussion in Section 7). Deriving *pixel-level* models — for example, through an iterative coarse-to-fine scheme using 3×3 or 5×5 patch structures — is also interesting by itself: What makes denoising and many computer vision application very hard, is that a natural image often contains many irrelevant, often partially resolved, objects. This type of "noise" is highly non-Gaussian and sometimes referred to as "clutter". To develop better image enhancement algorithms that can deal with structured noise, we need explicit models for the many regularities and geometries seen in local pixel patterns.

This work has some similarities to work by Geman and Koloydenko (1999). The latter study also concerns geometrical patterns of 3×3 patches but in image space as opposed to state space. The authors quantize 3×3 blocks according to a modified order statistic, and define "equivalence classes" based on photometry, complexity and geometry in image space. One of their goals is data classification for object recognition applications.

The organization of the paper is as follows: In Section 2, we describe the two data sets extracted from an optical image database by van Hateren and van der Schaaf (1998) and a range image database by Huang et al. (2000). In Section 3, we describe the preprocessing of the data sets. Our analysis is divided into three parts: In Section 4 we study the distribution of our data with respect to a Voronoi tessellation of the space of data points. This first part, is a *model-free* first exploration of the state space of contrast-normalized patches. We proceed in Section 5 with a study of the probability density of high-contrast optical image patches around a 2D manifold. The manifold represents the loci in state space of blurred step edges of different orientations and positions. Finally, in Section 6 we analyze the probability density of range data around clusters corresponding to binary patches. The results are discussed in Section 7.

Figure 3. Samples from the van Hateren optical image database. The gray values code for log-intensity values.

2. Optical and Range Data Sets

We extract two data sets of high-contrast 3×3 patches from optical and range images, respectively. These patches are then preprocessed according to Section 3.

- The *optical* data set contains about $4.2 \cdot 10^6$ high-contrast log-intensity patches. These are extracted from van Hateren's still image collection (van Hateren and van der Schaaf, 1998) of 4167 linear [3] calibrated 1024×1536 images; see Figure 3 for samples. From each image, we randomly select 5000 3×3 patches, and keep the top 20 percent, i.e. 1000 patches, with the highest contrast in log-intensities (see Section 3 for definition of contrast or "D-norm").

- The *range* data set contains about $7.9 \cdot 10^5$ high-contrast log-range patches. These are extracted from the Brown database by Huang and Lee (Huang et al., 2000) of around 200 444×1440 range images with mixed outdoor and indoor scenes; see Figure 4 for samples. We divide each image into disjoint 3×3 patches, and discard all patches with out-of-range data. From the remaining

Figure 4. Samples from the Brown range image database by Huang and Lee. The gray values code for log-range values.

patches in each image, we randomly select 20000 patches and keep the top 20 percent with highest contrast in log-range values.

The optical and range images have quite different subresolution properties. In optical images, the subpixel details are *averaged* by the point-spread function of the camera. The pixel values in a range image, on the other hand, usually correspond to the *minimum* of the subresolution details. The field of view of the scanner is 80° vertically and 259° horizontally. The beam divergence of the laser range finder is approximately 3 mrad. If the laser footprint hits two targets with a range difference larger than 3 meters, the returned value is the range to the nearest target. If the range difference is less than 3 meters, the returned value is roughly a weighted average of both ranges where the weight depends on the reflectivity of the two targets.

3. Preprocessing

We want to compare and group image patches based on their *geometrical structure*. In a natural scene, the reflectance and the shape of a surface are usually fixed quantities, while the absolute distance to, and the illumination of, a point in a scene can vary widely. We thus work with the *logarithm*, rather than the absolute values, of intensity [4] or range. Furthermore, before analyzing the data, we subtract the mean and contrast-normalize each image patch.

Let $\mathbf{x} = [x_1, \ldots, x_9]^T = [I_{11}, I_{21}, I_{31}, I_{12}, \ldots, I_{33}]^T \in I\!R^9$ be a non-constant vector with the log-values of the original patch. Subtracting the mean and contrast normalizing lead to a new vector

$$\mathbf{y} = \frac{\mathbf{x} - \frac{1}{9}\sum_{i=1}^{9} x_i}{||\mathbf{x} - \frac{1}{9}\sum_{i=1}^{9} x_i||_D} \ . \tag{1}$$

The contrast $||\mathbf{x}||_D$, or "D-norm", is here calculated by summing the differences between 4-connected neighbors ($i \sim j$) in a 3×3 patch and then taking the square root, i.e.

$$||\mathbf{x}||_D = \sqrt{\sum_{i \sim j} (x_i - x_j)^2}. \tag{2}$$

In matrix form, we have

$$||\mathbf{x}||_D = \sqrt{\mathbf{x}^T D \mathbf{x}}, \tag{3}$$

where

$$D = \begin{bmatrix} 2 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 3 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 3 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 2 \end{bmatrix} . \tag{4}$$

The preprocessed data points lie on a 7-dimensional ellipsoid $\tilde{\mathrm{S}}^7 \subset I\!R^9$, where

$$\tilde{\mathrm{S}}^7 = \{\mathbf{y} \in I\!R^9 \ : \ \sum_{i=1}^{9} y_i = 0 \ , \ \mathbf{y}^T D \mathbf{y} = 1\} \ . \tag{5}$$

For convenience, we make a change of basis to a coordinate system where the data points lie on a Euclidean sphere. In the case of scale-invariant images, this is exactly equivalent to whitening the data [5]. The 2-dimensional Discrete Cosine Transform (DCT) basis of a 3×3 image patch diagonalizes the matrix D. In vector form, we write the 8 non-constant DCT basis vectors as

$$\begin{array}{lllllllllll}
\mathbf{e}_1 & = \frac{1}{\sqrt{6}}[& 1, & 0, & -1, & 1, & 0, & -1, & 1, & 0, & -1]^T \\
\mathbf{e}_2 & = \frac{1}{\sqrt{6}}[& 1, & 1, & 1, & 0, & 0, & 0, & -1, & -1, & -1]^T \\
\mathbf{e}_3 & = \frac{1}{\sqrt{54}}[& 1, & -2, & 1, & 1, & -2, & 1, & 1, & -2, & 1]^T \\
\mathbf{e}_4 & = \frac{1}{\sqrt{54}}[& 1, & 1, & 1, & -2, & -2, & -2, & 1, & 1, & 1]^T \\
\mathbf{e}_5 & = \frac{1}{\sqrt{8}}[& 1, & 0, & -1, & 0, & 0, & 0, & -1, & 0, & 1]^T \\
\mathbf{e}_6 & = \frac{1}{\sqrt{48}}[& 1, & 0, & -1, & -2, & 0, & 2, & 1, & 0, & -1]^T \\
\mathbf{e}_7 & = \frac{1}{\sqrt{48}}[& 1, & -2, & 1, & 0, & 0, & 0, & -1, & 2, & -1]^T \\
\mathbf{e}_8 & = \frac{1}{\sqrt{216}}[& 1, & -2, & 1, & -2, & 4, & -2, & 1, & -2, & 1]^T
\end{array} \tag{6}$$

where the normalization is chosen such that $||\mathbf{e}_1||_D = \ldots ||\mathbf{e}_8||_D = 1$.

Let the DCT basis vectors above be the columns of a 9×8-matrix $A = [\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_8]$, and introduce a diagonal 8×8-matrix Λ with the diagonal elements equal to $1/||\mathbf{e}_1||^2$, $1/||\mathbf{e}_2||^2$, …, $1/||\mathbf{e}_8||^2$. A change of basis taking $\mathbf{e}_i$ to be the unit vectors according to

$$\mathbf{v} = \Lambda A^T \mathbf{y} \ , \tag{7}$$

or equivalently, $\mathbf{y} = A\mathbf{v}$, will then transform the ellipsoid $\tilde{\mathrm{S}}^7$ in Equation 5 to a 7-dimensional Euclidean sphere

$$\mathrm{S}^7 = \{\mathbf{v} \in I\!\!R^8 \ : \ \sum_{i=1}^{8} v_i = 0 \ , ||\mathbf{v}|| = 1\} \ . \tag{8}$$

The 7-sphere S^7 is the *state space* of our preprocessed data points.

We would like to be able to measure the distance between two 3×3 image patches P_1 and P_2. Since the contrast-normalized data is on a sphere, we simply calculate the *angular distance* between the corresponding two points $\mathbf{v}_1, \mathbf{v}_2 \in I\!\!R^8$ on the sphere. In other words, our distance measure is given by

$$\begin{aligned}
\mathrm{dist}(P_1, P_2) &= \arccos(\mathbf{v}_1 \cdot \mathbf{v}_2) \\
&= \arccos\left(\frac{B\mathbf{y}_1 \cdot B\mathbf{y}_2}{||B\mathbf{y}_1|| \ ||B\mathbf{y}_2||}\right) \ ,
\end{aligned} \tag{9}$$

where the matrix $B = \Lambda A^T$, and the vectors $\mathbf{y}_1, \mathbf{y}_2 \in I\!\!R^9$ represent the centered 3×3 image patches P_1 and P_2.

4. A First Exploration of the 7-Sphere

As a first exploration of our two data sets, we divide the 7-sphere $S^7 \subset \mathbb{R}^8$ into Voronoi cells, and analyze how the data points are distributed with respect to the tessellation. This is a model-free exploration of the state space, where we derive non-parametric probability distributions.

Assume a discrete collection of sampling points $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N\}$ in S^7. A Voronoi cell Ω_i around a sampling point $\mathbf{P}_i$ is defined as the set of all points $\mathbf{x} \in S^7$ that are at least as close to $\mathbf{P}_i$ as to any other sampling point $\mathbf{P}_j$; that is,

$$\Omega_i = \{\mathbf{x} \in S^7 \mid \mathrm{dist}(\mathbf{x}, \mathbf{P}_i) \leq \mathrm{dist}(\mathbf{x}, \mathbf{P}_j) \text{ for any } \mathbf{P}_j \in \mathcal{P}\}, \tag{10}$$

where $\mathrm{dist}(\cdot, \cdot)$ is the angular distance, as defined in Equation 9, between two points on the sphere.

The problem of choosing a dense set of sampling points on a sphere $S^7 \subset \mathbb{R}^8$ is analogous to the problem of packing spheres in $\mathbb{R}^8$ itself. For a fixed number of sampling points, we seek a set of points such that equal non-overlapping spheres centered at the points cover the sphere "efficiently", in the sense that the space not covered by these spheres is minimal. This is a non-trivial problem in the general n-dimensional case (see (Conway and Sloane, 1988) for an in depth exposé of sphere packing in higher dimensions). In the case of 8-dimensional lattices, however, there exists an optimal solution given by the so called E_8 lattice. There are several possible coordinate systems for E_8. Using the "even" coordinate system, we obtain

$$E_8 = \{(x_1, \ldots, x_8) \mid \text{all } x_i \in \mathbb{Z} \text{ or all } x_i \in \mathbb{Z}+1/2, \textstyle\sum x_i \equiv 0 \pmod 2\}. \tag{11}$$

Suppose that there are N points in the E_8 lattice at a distance u from the origin. Then these points, when rescaled by dividing them by u, form a dense set of sampling points $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N\}$ of S^7. The first spherical shell with $u = \sqrt{2}$ and $N = 240$ is the unique solution to the "kissing number" problem in $\mathbb{R}^8$, where one wants to arrange the maximum number of non-overlapping spheres of radius 1 so that they all touch the unit sphere. For our Voronoi sampling, we have chosen to take the 4:th spherical shell of E_8 with $u = \sqrt{8}$. After normalization, this gives us a total of 17520 Voronoi cells on the 7-sphere with roughly the same size. The sampling points $\mathbf{P}_i$ are given by the permutations and sign changes of the following five 8-vectors:

1. The 112 permutations and sign changes of $[2, 2, 0, 0, 0, 0, 0, 0]^T/\sqrt{8}$.
2. The 8960 permutations and sign changes of $[2, 1, 1, 1, 1, 0, 0, 0]^T/\sqrt{8}$.

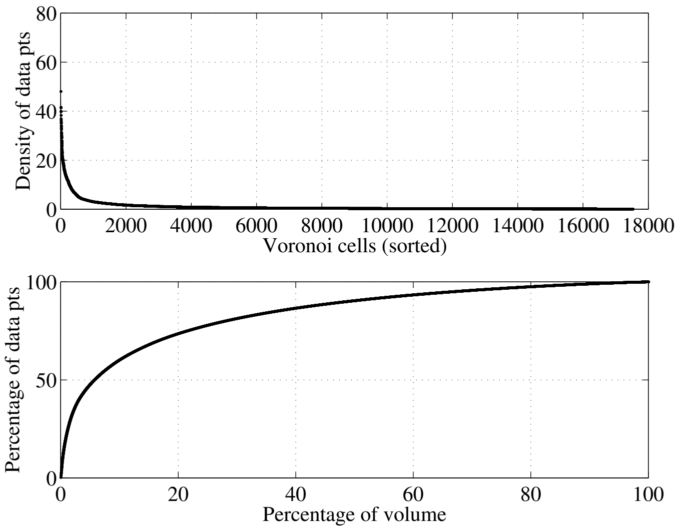

Figure 5. (Top) Density $\rho(\Omega_i)$ of high-contrast optical patches in Voronoi cells that are sorted according to decreasing density. (Bottom) Cumulative percentage of optical patches in the Voronoi cells above versus the cumulative percentage of volume in S^7 that are occupied by these cells.

3. The 256 permutations and sign changes of $[1,1,1,1,1,1,1,1]^T/\sqrt{8}$.

4. The 7168 permutations and sign changes with the constraint that the number of minus signs is an odd number; $\left[\frac{3}{2},\frac{3}{2},\frac{3}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}\right]^T/\sqrt{8}$.

5. The 1024 permutations and sign changes with the constraint that the number of minus signs is an even number; $\left[\frac{5}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}\right]^T/\sqrt{8}$.

From a Monte Carlo simulation (with 5 million random points on the 7-dimensional unit sphere), we get that the volumes of the 5 types of Voronoi cells above are approximately $6.3\cdot10^{-3}$, $1.8\cdot10^{-3}$, $4.1\cdot10^{-3}$, $1.8\cdot10^{-3}$, and $1.8\cdot10^{-3}$, respectively.

We bin our high-contrast optical and range patches into the 17520 Voronoi cells using the definition given in Equation 10. We define the density of data points in the Voronoi cell around sample point $\mathbf{P}_i$ as

$$\rho(\Omega_i) = \frac{N(\Omega_i)/\sum_i N(\Omega_i)}{\mathrm{vol}(\Omega_i)/\mathrm{vol}(\mathrm{S}^7)}, \tag{12}$$

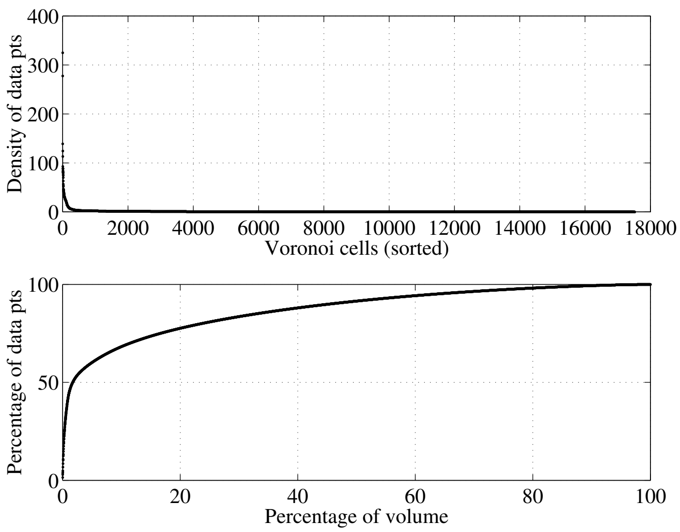

Figure 6. (Top) Density $\rho(\Omega_i)$ of high-contrast range patches in Voronoi cells that are sorted according to decreasing density. (Bottom) Cumulative percentage of range patches in the Voronoi cells above versus the cumulative percentage of volume in S^7 that are occupied by these cells.

where $N(\Omega_i)$ is the number of patches in the Voronoi cell Ω_i associated with the sampling point $\mathbf{P}_i$, $\mathrm{vol}(\Omega_i)$ is the volume of that cell, and $\mathrm{vol}(\mathrm{S}^7) = \sum_i \mathrm{vol}(\Omega_i) = \pi^4/3$ is the total volume of the 7-sphere.

In Figure 5 and 6 we show the density ρ of the Voronoi cells for high-contrast optical and range patches, together with the percentage of volume occupied by the percentage of patches. We find that the distribution of data on S^7 is extremely "sparse", with the majority of data points concentrated in a few high-density regions on the sphere. For both the optical and range data, half of the patches can be divided into an optimal set of Voronoi cells that occupies less than 6% of the total volume of the 7-sphere.

We can use the Kullback-Leibler distance or relative entropy to get an information-theoretic measure of the deviation of the probability distributions of our data from a uniform distribution. Note that a Gaussian assumption on natural images corresponds to a uniform distribution in state space after whitening.

We estimate the probability density functions $p_{\mathrm{o}}(\Omega_i)$ and $p_{\mathrm{r}}(\Omega_i)$ for optical and range data, respectively, by calculating

$$p(\Omega_i) = \frac{N(\Omega_i)}{\sum_i N(\Omega_i)} \tag{13}$$

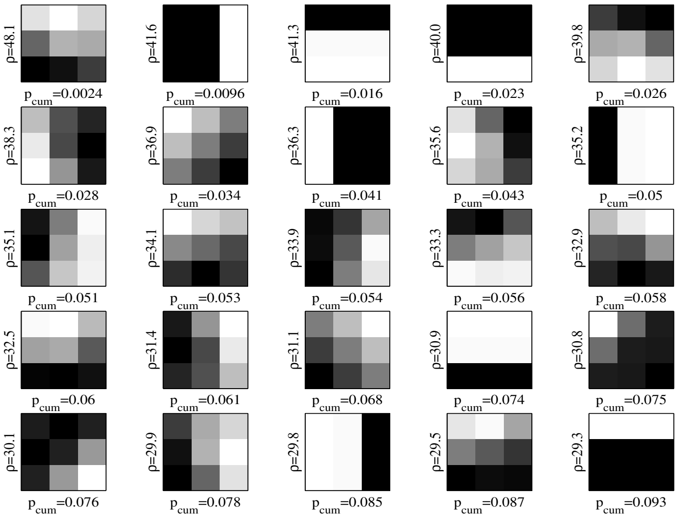

Figure 7. The first 25 Voronoi patches ordered after their densities $\rho(\Omega_i)$ in Equation 12 for optical images. For each Voronoi cell, we display the 3×3 patch corresponding to the sample point $\mathbf{P}_i$. The cumulative sum of $p(\Omega_i)$ (Equation 13) over the ordered patches is shown as p_{cum}.

for $i = 1, \ldots, 17520$. As before $N(\Omega_i)$ is the number of optical or range data points in Voronoi cell Ω_i. The corresponding probability density function for a uniform distribution is defined as

$$q_{\text{u}} = \frac{\text{vol}(\Omega_i)}{\text{vol}(\text{S}^7)} , \tag{14}$$

where $\text{vol}(\Omega_i)$ is the volume of Voronoi cell Ω_i, and $\text{vol}(\text{S}^7) = \pi^4/3$ is the total volume of the 7-sphere. The KL-distance between the empirical probability density of the data and a uniform density on the 7-sphere is given by

$$D(p\|q_{\text{u}}) = \sum_i p(\Omega_i) \, \log_2 \left(\frac{p(\Omega_i)}{q_u(\Omega_i)} \right) . \tag{15}$$

This gives us a measure of the number of excess bits we incur by coding the data points distributed by $p_{\text{o}}(\Omega_i)$ or $p_{\text{r}}(\Omega_i)$ with a code book based on the uniform distribution $q_{\text{u}}(\Omega_i)$ (Cover and Thomas, 1991). For our two datasets, we have that $D(p_{\text{o}}\|q_{\text{u}}) = 1.62$ bits and $D(p_{\text{r}}\|q_{\text{u}}) = 2.59$ bits. These numbers indicate that the distribution of optical or range data points in the state space of the contrast-normalized data is *highly* non-uniform.

In Figure 7 and Figure 8, we display the first 25 sampling points $\mathbf{P}_i$ of the Voronoi cells ordered after their densities $\rho(\Omega_i)$ (defined according

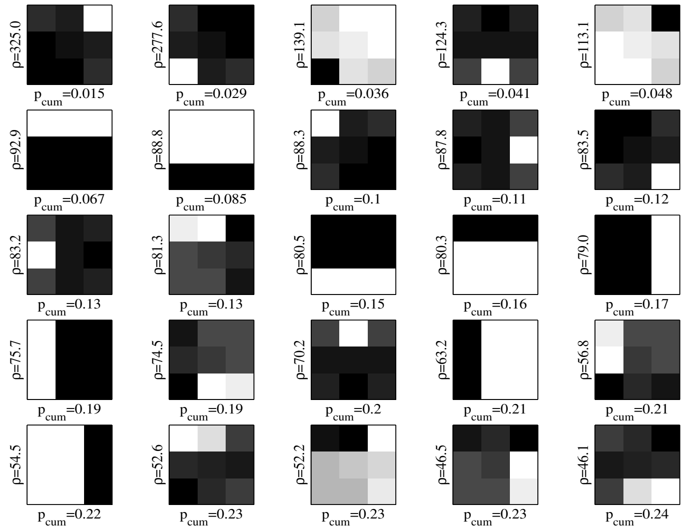

Figure 8. The first 25 Voronoi patches ordered after their densities $\rho(\Omega_i)$ in Equation 12 for range images. For each Voronoi cell, we display the 3×3 patch corresponding to the sample point $\mathbf{P}_i$. The cumulative sum of $p(\Omega_i)$ (Equation 13) over the ordered patches is shown as p_{cum}.

to Equation 12) for optical and range data, respectively. (The pixel patterns in these two figures depend of course on the exact choice of sampling scheme, and can look very different if one were to choose basis functions from a different lattice.) In Figure 7 for optical patches, the centers of the Voronoi cells with highest densities are close to blurred step edges (see Section 5.1). For high-contrast range patches, the cells with highest densities resemble binary patches (compare Figure 8 with Figure 18). Note in particular that some of the first 25 Voronoi cells here are similar to binary symmetry classes 1', 2', and 5'.

5. Optical Data: Probability Density Around a 2D Manifold of Step Edges

The analysis of Voronoi cells on the 7-sphere indicates that blurred step edges are common high-contrast patterns in optical patches. In Section 5.1, we present an ideal model for edges in optical images that takes into account the *averaging* effects of the optics of the camera. The model predicts a 2-dimensional continuous manifold in state space, parametrized by the orientation α and position l of an edge. In Sections 5.2 and 5.3, we test the model with optical data from natural images. Finally, in Section 5.4 we apply this model to range data and discuss the differences in the results.

5.1. The Ideal Manifold of Edges

We represent the 3×3 image patch by a square $SQ = \{(x, y) : -3/2 \leq x, y \leq 3/2\}$. The pixels in the patch are given by

$$S_{ij} = \{(x, y) : \ (j - 3/2) \leq x \leq (j - 1/2), \ (1/2 - i) \leq y \leq (3/2 - i)\} \ , \tag{16}$$

where $i, j = 0, 1, 2$.

The pixel value $I(i, j)$ in an optical image is approximately an average of the underlying scene $\phi(x, y)$ recorded at each pixel S_{ij}, i.e. the pixel intensity

$$I(i, j) = \int_{S_{ij}} \phi(x, y) dx dy \tag{17}$$

where $\phi(x, y)$ is the scene luminance. For an ideal step edge,

$$\phi_{\alpha,l}(x, y) = \begin{cases} a \ \text{ if } \ -x \sin\alpha + y \cos\alpha > l \\ b \ \text{ if } \ -x \sin\alpha + y \cos\alpha < l \end{cases} \tag{18}$$

where $a > b$. The parameter $\alpha \in [0, 2\pi)$ is the angle that the direction perpendicular to the edge makes with the y-axis, and the parameter $l \in (-3/2, 3/2)$ is the displacement of the edge from the origin (Figure 9). Thus, pixels in $I_{\alpha,l}(i, j)$ strictly above the edge have intensities a, pixels strictly below the edge have intensities b, and pixels along the edge is a weighted average of a and b. After subtracting the mean and contrast normalizing each edge patch (see Section 3), we arrive at a set of points $\mathbf{v}_{\alpha,l} \in \mathrm{S}^7 \subset \mathbb{R}^8$. It can be shown that the loci of these points, with $\alpha \in [0, 2\pi)$ and $l \in (-3/2, 3/2)$, define a C^1 2-dimensional manifold, M^2, embedded in the 7-dimensional sphere.

Because of the centering and the contrast normalization, some (α, l)-values are *degenerate*, i.e. they lead to the same patch or point $\mathbf{v}_{\alpha,l} \in \mathrm{M}^2$ after preprocessing. This situation occurs for patches with "glancing edges". Assume for example that the surface parameter $l > 0$. Then consider all edges that connect the points $(-1.5, y)$ and $(x, 1.5)$ on the border of the patch, where $x \in (-1.5, \infty)$ is *fixed* and y can take *any* value in the interval $[0.5, 1.5)$. Simple trigonometry shows that for a fixed value of x, this defines a set of (α, l)-values that correspond to the *same* contrast-normalized $\mathbf{v}_{\alpha,l}$-patch. Rotations by $\pi/2$, reflections and contrast sign inversions (see the 16 symmetries in Equation 28) give the full family of equivalent edge patches. Each line in Figure 10 represents one set of equivalent (α, l)-values in this family. There are two special cases of the example above with edges through the points $(-1.5, y)$

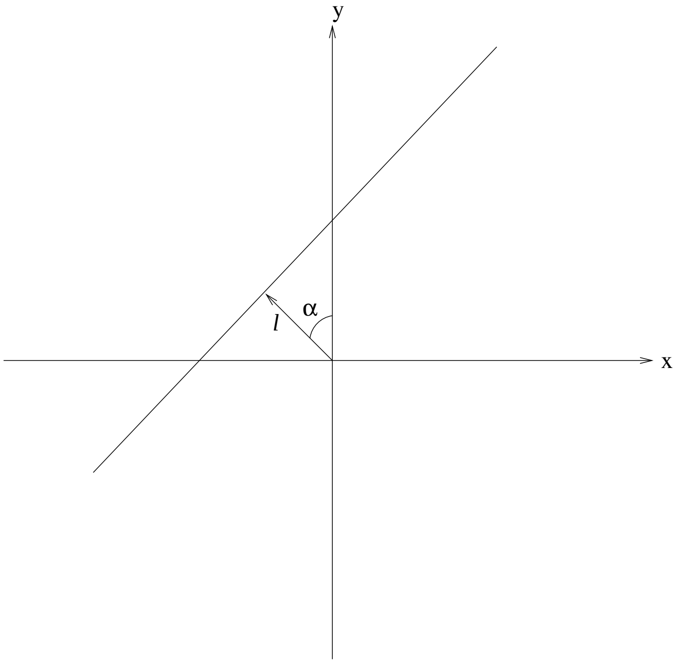

Figure 9. The parameter $\alpha \in [0, 2\pi)$ is the angle that the direction perpendicular to the edge makes with the y-axis, and the parameter $l \in (-3/2, 3/2)$ is the displacement of the edge from the origin.

and $(x, 1.5)$. One special case is when $x \in (-1.5, -0.5]$: *All* edges with $-1.5 < x \leq -0.5$ and $0.5 \leq y < 1.5$ lead to the same contrast-normalized patch. The light shaded regions in Figure 10 represent these "corner patches" and their symmetries. Another special case is when $x \to \infty$: This limit case and its symmetries "converge" to the set of equivalent (α, l)-values where $0.5 \leq l < 1.5$ or $-1.5 < l \leq -0.5$ and $\alpha = 0, \pi/2, \pi$ or $3\pi/2$ (horizontal and vertical edges). *Non-degenerate* edge patches are given by $0 \leq \alpha \leq \pi/4$ and $0 \leq l < (1.5 \sin\alpha + 0.5 \cos\alpha)$, and its 16 symmetries. In the figure, these (α, l)-values are represented by the dark shaded region.

Figure 11 shows a few examples of ideal edge patches that correspond to different (α, l)-values between $0 \leq \alpha \leq 180$ degrees and $1.5 < l < -1.5$. The step edges are here chosen on a triangular grid with the spacing $\Delta\alpha = 15$ degrees and $\Delta l = 1/4$ pixel units. Numbers between patches represent the angular distances in degrees between nearest neighbors. Although we use an even sampling of grid points in the (α, l)-coordinate system, the distances between the nearest neighbors vary widely.

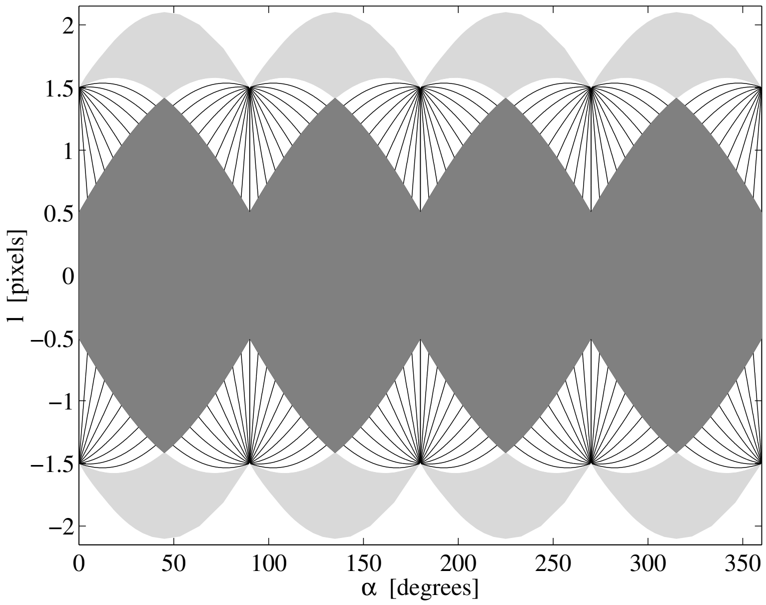

Figure 10. Because of the contrast normalization, some (α, l)-values are degenerate, i.e. they correspond to the same point $\mathbf{v}_{\alpha,l}$ on the 7-dimensional unit sphere S^7. Each line in the figure is an example of a set of (α, l)-values that lead to the same contrast-normalized 3×3 patch. The light shaded region corresponds to degenerate values for corner edges. The dark shaded region in the interior of the graph shows all (α, l)-values that are well-defined, i.e. non-degenerate.

Figure 12 shows the geometry of the surface M^2 of step edges of different orientations and positions more clearly. Here we estimate the surface metric $f(\alpha, l) = \frac{dA}{d\alpha\, dl}$ numerically with a triangulated mesh that is much finer spaced than in the example above: We first divide the surface into rectangular bins with widths $\Delta\alpha = \pi/48$ (3.75 degrees) and $\Delta l = 1/16$ pixels, hereafter we discard all bins that are completely outside the interior region of the (α, l)-plot (see Figure 10), that is, bins with degenerate points. Each of the remaining bins is then split into 4 spherical triangles, where the vertices represent blurred step edges in M^2. The mesh is finally successively refined, where needed, until the distance between any two vertices in a triangle is less than 8 degrees. The final mesh contains 14376 spherical triangles. The area of a rectangular bin in Figure 12 corresponds to the sum of the areas of spherical triangles [6] inside the bin.

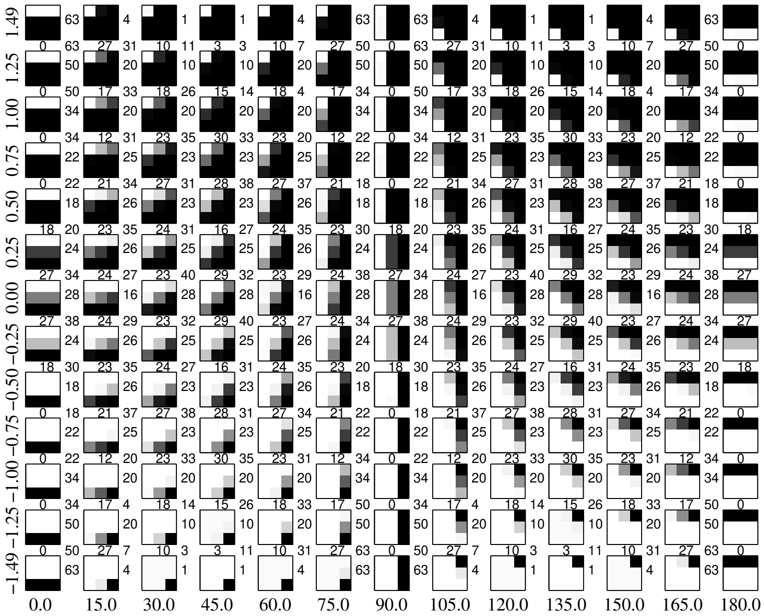

Figure 11. Examples of ideal step edges between $0 \leq \alpha \leq 180$ degrees (horizontal axis) and $1.5 < l < -1.5$ (vertical axis). The step edges are here chosen on a triangular grid with sides $\Delta\alpha = 15$ degrees and $\Delta l = 1/4$ pixel units. Numbers between patches represent the angular distances in degrees between neighbors in the horizontal, vertical and (lower left - upper right) diagonal directions in the (α, l)-grid.

5.2. Density of Optical Data as a Function of the Distance to the Ideal Edge Manifold

We next try to get a numerical estimate of the probability density of high-contrast optical data around the surface of step edges. In the following experiment, we use the "optical dataset" with $N_{\text{tot}} = 4.2 \cdot 10^6$ high-contrast patches described in Section 2. As above, we model the surface with a mesh with about 14000 spherical triangles, where the vertices of the triangles are blurred step edges in M^2.

For each optical data point $\mathbf{x}_n$ $(n = 1, 2, \ldots, N_{\text{tot}})$, we calculate the distances to the centers of the spherical triangles in the mesh. We assume that the distance to the ideal edge manifold M^2 is approximately the same as the distance to the closest triangle center $\mathbf{v}_{\alpha_n, l_n}$, i.e. we assume that

$$\theta = \mathrm{dist}(\mathbf{x}_\mathrm{n}, \mathrm{M}^2) \approx \min_{\mathbf{v}_{\alpha_\mathrm{n}, \mathrm{l}_\mathrm{n}}} \mathrm{dist}(\mathbf{x}_\mathrm{n}, \mathbf{v}_{\alpha_\mathrm{n}, \mathrm{l}_\mathrm{n}}). \tag{19}$$

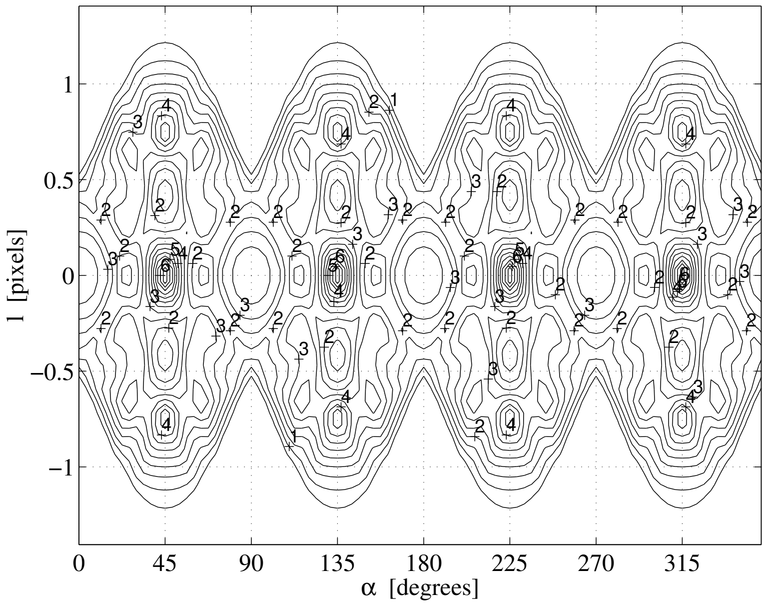

Figure 12. Contour plot of the metric $f(\alpha, l) = \frac{dA}{d\alpha\, dl}$ for a surface of ideal step edges. The rectangular bins in the figure have widths $\Delta\alpha = \pi/48$ (3.75 degrees) and $\Delta l = 1/16$ pixels. For the calculation of the area A of a bin, we add up the areas of 4 or more spherical triangles inside the bin (see text).

The error is largest for very small θ, where we sometimes get an over-estimate of θ due to the finite grid spacing.

Figure 13(a), top, shows a normalized histogram of the number of data points as a function of the estimated distance θ. Let

$$N(\theta) = \#\left\{n : \theta - \frac{\Delta\theta}{2} \leq \operatorname{dist}(\mathbf{x}_{\mathrm{n}}, \mathrm{M}^2) < \theta + \frac{\Delta\theta}{2}\right\} , \tag{20}$$

where $\Delta\theta$ is the bin width of the histogram. Linear regression (Figure 13(b), top) gives that $N(\theta) \propto \theta^{1.4}$ for small θ.

For the density estimate, we also need to calculate the volume of the set

$$B_\theta = \left\{\mathbf{x} \in \mathrm{S}^7 : \theta - \frac{\Delta\theta}{2} \leq \operatorname{dist}(\mathbf{x}, \mathrm{M}^2) < \theta + \frac{\Delta\theta}{2}\right\} . \tag{21}$$

Figure 13(a), bottom, and Figure 13(b), bottom, show the results from a Monte Carlo simulation with $V_{\mathrm{tot}} = 10^7$ sample points, $\mathbf{s}_n$ ($n = 1, 2, \ldots, V_{\mathrm{tot}}$), that are uniformly randomly distributed on a 7-dimensional unit sphere. The number of sample points in B_θ, i.e.

$$V(\theta) = \#\left\{n : \theta - \frac{\Delta\theta}{2} \leq \operatorname{dist}(\mathbf{s}_n, \mathrm{M}^2) < \theta + \frac{\Delta\theta}{2}\right\} , \tag{22}$$

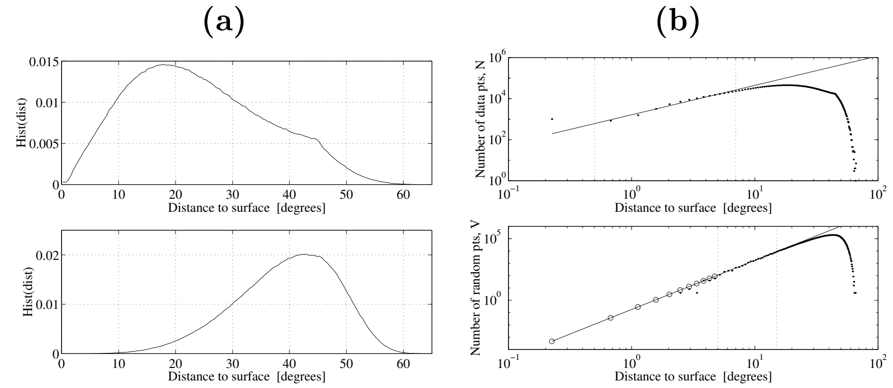

Figure 13. (a) *Top*: Normalized histogram of the number of optical high-contrast patches, N, versus the distance, θ, to the surface. *Bottom*: Normalized histogram of the number of random Monte Carlo samples, V, versus θ. (b) *Top*: Log-log plot of N versus θ. Linear regression in the interval between 0.5 and 7 degrees gives $N \sim \theta^{1.4}$ (solid line). *Bottom*: Log-log plot of V versus θ. Linear regression in the interval between 5 and 15 degrees gives $V \sim \theta^{4.0}$ (solid line). The circles represent extrapolated values of V for $\theta \leq 5$ degrees. The total number of optical data points is $N_{\text{tot}} \approx 4 \cdot 10^6$. The total number of random samples is $V_{\text{tot}} = 10^7$. The random samples are uniformly distributed on a 7-dimensional unit sphere, and give a Monte Carlo estimate of the volume of the space occupied by the histogram bins.

is directly proportional to the *volume* of B_θ. As expected [7], the number of random points V increases approximately as $V(\theta) \propto \theta^4$ for small θ. The curve for $V(\theta)$ has a maximum around $\theta = 43$ degrees, after which it drops. The drop may indicate folds in the surface where part of the 7-sphere are at equal distance to different points of the surface. Furthermore, the plot shows that all points in S^7 are within approximately 60 degrees of the edge manifold M^2. If the surface were flat, we would expect the corresponding distance to be 90 degrees (for the two antipodal points on the sphere).

Figure 14 shows the density function

$$p(\theta) = \frac{N(\theta)/N_{\text{tot}}}{V(\theta)/V_{\text{tot}}} \, . \tag{23}$$

For $\theta \lesssim 10$ degrees,

$$p(\theta) \sim \theta^{-2.5} \, . \tag{24}$$

This result strongly supports the idea that there exists a 2-dimensional manifold in the 7-sphere where the data points are concentrated. In fact, we here see evidence of an *infinite* probability density at the ideal edge manifold (where $\theta = 0$).

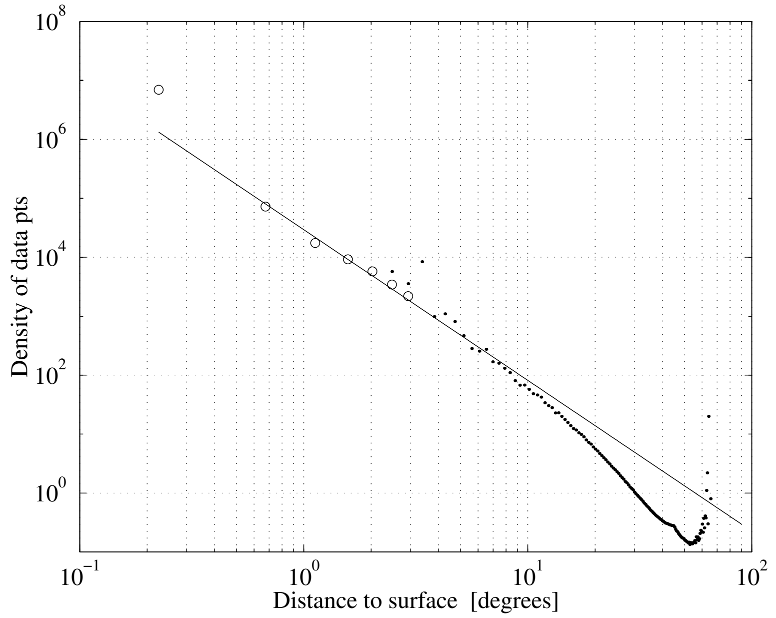

Figure 14. Log-log plot of the average density function $p(\theta)$ for optical data. For $\theta \lesssim 10$ degrees, $p(\theta) \sim \theta^{-2.5}$ (solid line). The circles represent the values of the ratio $\frac{N(\theta)/N_{\text{tot}}}{V(\theta)/V_{\text{tot}}}$ for extrapolated values of $V(\theta)$ when θ is small (Figure 13b, bottom).

In Figure 15, top, we have plotted the percentage of data points that are within a certain distance θ of the surface. The curve shows that about 50% of the data points are within a tubular neighborhood

$$K_\theta = \{\mathbf{x} \in \mathrm{S}^7 : \mathrm{dist}(\mathbf{x}, \mathrm{M}^2) \leq \theta\} \tag{25}$$

of the surface with width $\theta = 26$ degrees. This neighborhood corresponds to only 9% of the total volume of S^7 (Figure 15, bottom).

5.3. Density of Optical Data as a Function of the Surface Parameters

We next study how the high-contrast optical data are spread out *along* the surface of step edges. For the position calculation, we only include data points that are very close to the surface. For each data point $\mathbf{x}_n$ in our data set, we find the closest center point $\mathbf{v}_{\alpha_n, l_n}$ in the triangulated mesh. We compute the 2D histogram

$$\begin{aligned} N(\alpha, l) = \; & \#\{n : \mathrm{dist}(\mathbf{x}_n, \mathbf{v}_{\alpha_n, l_n}) \leq \theta_{\max}, \\ & \alpha_n \in \left[\alpha - \tfrac{\Delta\alpha}{2}, \alpha + \tfrac{\Delta\alpha}{2}\right), \; l_n \in \left[l - \tfrac{\Delta l}{2}, l + \tfrac{\Delta l}{2}\right)\right\} \;, \end{aligned} \tag{26}$$

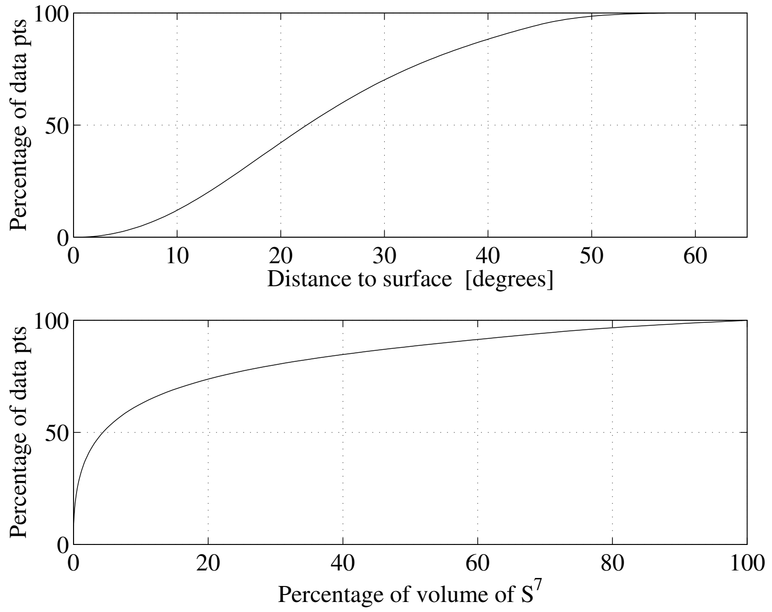

Figure 15. (a) Percentage of optical data points that are within a tubular neighborhood K_θ of the surface with width θ ("distance to surface"). (b) Same curve plotted versus the ratio $\frac{100\cdot\mathrm{vol}(K_\theta)}{\mathrm{vol}(S^7)}$ ("percentage of volume of S^7").

where $\theta_{\max} = 20$ degrees, and the bin widths $\Delta\alpha = \pi/48$ radians and $\Delta l = 1/16$ pixels.

We define the density $p(\alpha, l)$ of data points along the surface as

$$p(\alpha, l) = \frac{N(\alpha, l)/\sum_{\alpha,l} N(\alpha, l)}{f(\alpha, l)/\sum_{\alpha,l} f(\alpha, l)}\ , \tag{27}$$

where the sum $\sum_{\alpha,l} N(\alpha, l) \approx 1.3\cdot 10^6$, and $f(\alpha, l)$ is given in Section 5.1.

Figure 16 shows the results for regions where $f(\alpha, l) > 0.5$ (see Figure 12 for the surface metric). Although the data points are spread out along the whole surface, there is a clear concentration of data points around $\alpha = 0$, 90, 180, and 270 degrees (vertical and horizontal edges) and the (α, l)-values near the border of degenerate edges.

5.4. Range Data Comparison: Probability Density as a Function of the Distance to the Edge Manifold

A probability density estimate of high-contrast *range* data leads to very different results around the surface of blurred step edges.

In Figure 17 we show a normalized histogram $N(\theta)$ of range patches as a function of the distance θ to the surface; compare with Fig-

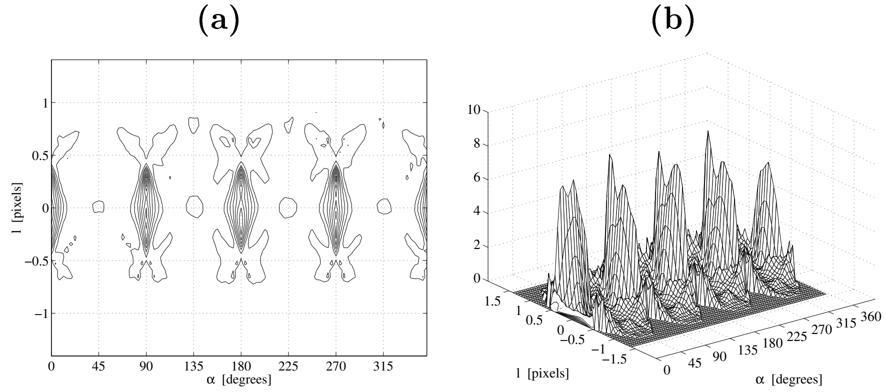

Figure 16. (a) Two-dimensional contour plot and (b) three-dimensional mesh of the density $p(\alpha, l)$ of high-contrast optical data along the surface of step edges. In the density calculation, we only include data points that are, at a position (α, l) where $f(\alpha, l) > 0.5$, and in a tubular neighborhood of the surface with width $\theta_{\max} = 20$ degrees. The bin widths $\Delta\alpha = \pi/48$ (3.75 degrees) and $\Delta l = 1/16$ pixel units.

ure 13(a), top, for optical patches. The histogram for range data has sharp peaks at $\theta \approx 0, 11, 22, 24, \ldots$ degrees. These peaks indicate the presence of *high-density clusters* of data points in S^7. A more detailed analysis shows that the positions of the local maxima correspond closely to the distances between binary patches and the edge manifold.

For 3×3 patches, there are 510 binary patches. These can be divided into 50 equivalence classes with respect to the 16 elements in the product group [8]

$$G = \{-1\} \times C_{4v} \;, \tag{28}$$

where $\{-1\}$ represents sign inversion, and the point group C_{4v} (Schönflies notation (Elliott and Dawber, 1979)) is generated by rotations through $\pi/2$ around the center pixel, and reflections across a plane containing the rotation axis.

We denote the set of 510 binary patches in $S^7 \subset I\!R^8$ by

$$\mathrm{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{510}\} \tag{29}$$

and the set of 50 distinct equivalence classes of the binary patches with respect to the symmetry group G by

$$\mathcal{E} = \{[\mathbf{b}_{1'}]_G, [\mathbf{b}_{2'}]_G \ldots, [\mathbf{b}_{50'}]_G\} \;. \tag{30}$$

We use primed indices to denote binary patches that are grouped into equivalence classes, and unprimed indices to denote the 510 original binary patches.

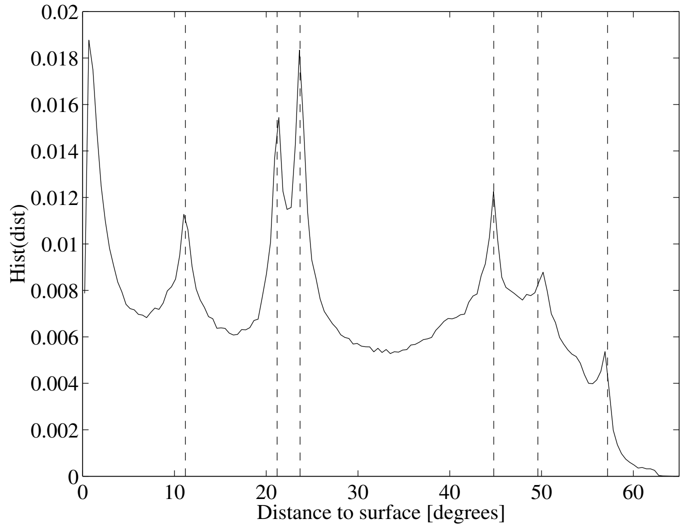

Figure 17. Normalized histogram of $N(\theta)$ for range patches versus the distance θ to the surface of ideal step edges. Dashed lines corresponds to binary symmetry classes $3'$, $4'$, $5'$, $14'/18'$, $25'$, and $42'$ (see Figure 18).

In Figure 18, we have sorted the 50 symmetry equivalence classes according to their distances to the surface of step edges. Note that the binary patches in $[\mathbf{b}_{1'}]$ and $[\mathbf{b}_{2'}]$ are exactly on the surface. Patches in $[\mathbf{b}_{3'}]$ are at the same distance from the surface as the second local maximum at $\theta \approx 11$ (Figure 17), patches in $[\mathbf{b}_{4'}]$ and $[\mathbf{b}_{5'}]$ correspond to the third and fourth local maxima at $\theta \approx 22$ and $\theta \approx 24$, respectively. The peak at $\theta \approx 45$ may be due to, for example, the symmetry classes $[\mathbf{b}_{14'}]$ (blobs) and $[\mathbf{b}_{18'}]$ (horizontal and vertical bars). The peaks at $\theta \approx 50$ and $\theta \approx 57$ could be signs of patterns similar to $[\mathbf{b}_{25'}]$ (diagonal bars) and $[\mathbf{b}_{42'}]$ (single dots), respectively. The displacement of the first peak from 0 in Figure 17 (compare $[\mathbf{b}_{1'}]$ and $[\mathbf{b}_{2'}]$) may be due to the overestimate of the distance θ for points that are very close to the surface.

6. Range Data: Probability Density Around Binary Patches

In the previous section we saw that high-contrast range image patches are concentrated in high-density clusters both on and around the surface of step edges. Furthermore, these clusters appear to be centered around binary patches. This motivates us to investigate the density of high-contrast range patches around the 510 possible binary patches.

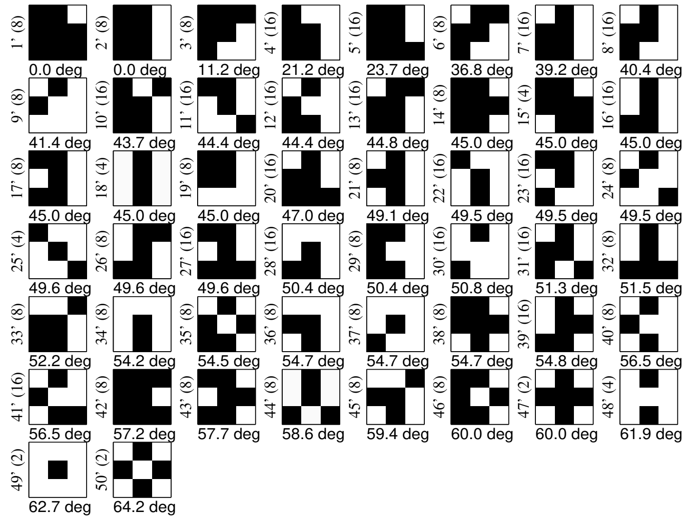

Figure 18. The 50 symmetry classes of binary patches arranged according to increasing θ-values, where θ is the angular distance to the closest point on the surface of ideal step edges. The number below each displayed patch represents this distance. The number in parenthesis left of each patch represents the number of binary patches in each equivalence class.

6.1. Density as a Function of Distance to Nearest Binary Patch

We start our analysis by calculating the density of the $N_{\text{tot}} = 7.9 \cdot 10^5$ high-contrast range patches as a function of the angular distance to the nearest binary patch:

For each high-contrast range patch we compute the angular distances to each of the 510 binary patches in the set $\mathrm{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{510}\}$, and find the *nearest* binary patch $\mathbf{b}_k \in B$. Let

$$N_k(\theta) = \# \left\{ n : \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\mathbf{x}_{\mathrm{n}}, \mathbf{b}_{\mathrm{k}}) < \theta + \frac{\Delta\theta}{2} \right\} , \qquad (31)$$

where $k = 1, 2, \ldots, 510$, be the histogram of range patches $\mathbf{x}_n$ ($n = 1, 2, \ldots, N_{\text{tot}}$) that are closest to, and at a distance $\theta - \Delta\theta/2 < \phi \leq \theta + \Delta\theta/2$ from, the binary patch $\mathbf{b}_k$. The volume

$$V(\theta) = \mathrm{vol}\{\mathbf{x} \in \mathrm{S}^7 : \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\mathbf{x}, \mathbf{b}_{\mathrm{k}}) < \theta + \frac{\Delta\theta}{2}\} \qquad (32)$$

is given by

$$V(\theta) = \frac{16\pi^3}{15} \int_{\theta - \Delta\theta/2}^{\theta + \Delta\theta/2} \sin^6(\phi)\, d\phi. \qquad (33)$$

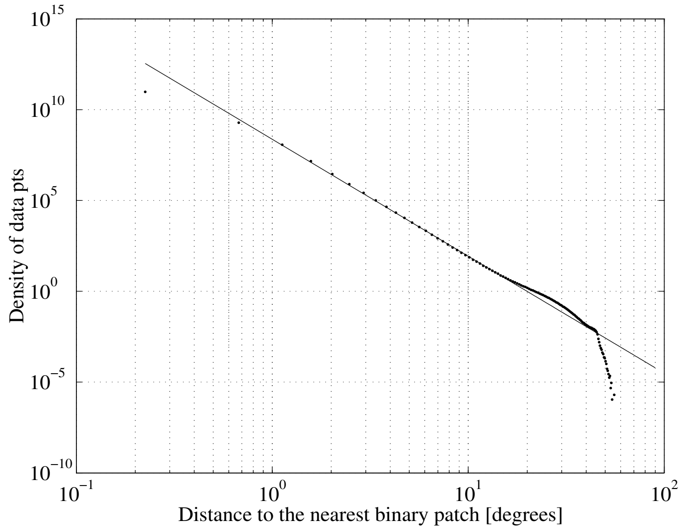

Figure 19. Log-log plot of the density $p(\theta)$ (Equation 34) of high-contrast range patches with respect to θ, the distance to the nearest binary patch. Linear regression in the interval 0.6 and 10 degrees gives $p(\theta) \sim \theta^{-6.4}$ (see solid line).

We define the average density of high-contrast range patches as a function of the angular distance θ to the nearest binary patch as

$$p(\theta) = \frac{\sum_{k=1}^{510} N_k(\theta)/N_{\text{tot}}}{510 \cdot V(\theta)/\text{vol}(\mathrm{S}^7)}, \tag{34}$$

where $\text{vol}(\mathrm{S}^7) = \pi^4/3$. In Figure 19 we show a log-log plot of $p(\theta)$. The graph is almost straight for more than a decade of distances, from 0.6 to 10 degrees. Linear regression gives that

$$p(\theta) \sim \theta^{-6.4}. \tag{35}$$

In Figure 20 we show the cumulative percentage

$$P_{\text{cum}}(\theta) = \sum_{k,\, \phi \leq \theta} \frac{N_k(\phi)}{N_{\text{tot}}} \tag{36}$$

of the number of patches with respect to the distance θ to the nearest binary patch, as well as the cumulative volume

$$V_{\text{cum}}(\theta) = \sum_{k,\, \phi \leq \theta} \frac{V(\phi)}{\text{vol}(\mathrm{S}^7)} \tag{37}$$

versus $P_{\text{cum}}(\theta)$. We see that 80% of the high-contrast range patches are within a spherical neighborhood of 30 degrees from one of the 510

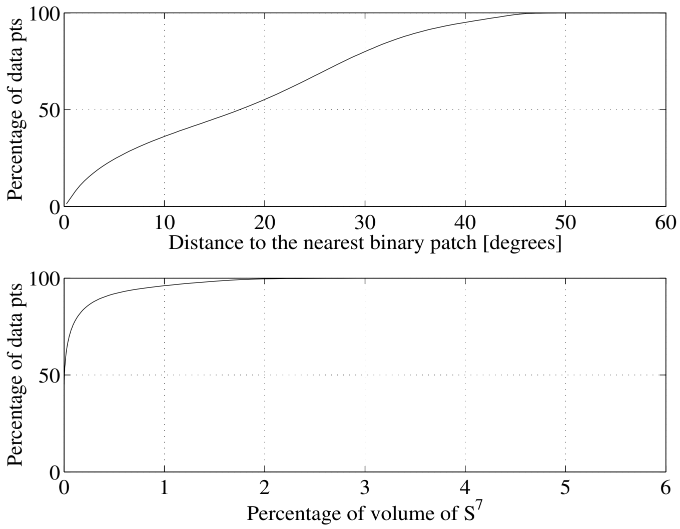

Figure 20. (Top) The cumulative percentage $P_{\mathrm{cum}}(\theta)$ of the number of patches $N(\theta)$ with respect to the distance θ to the nearest binary patch. (Bottom) The cumulative volume $V_{\mathrm{cum}}(\theta)$ versus $P_{\mathrm{cum}}(\theta)$.

binary patches. The neighborhoods of these 80% patches occupy only 0.14% of the total volume of S^7.

These results show that 3×3 high-contrast range patches are densely clustered around the 510 binary patches, and that the probability density is infinite at the positions of these binary patches.

6.2. Distribution of Range Patches Across the 50 Binary Symmetry Classes

We next study how the range patches are distributed among the 50 symmetry classes for binary patches (see Section 5.4). This will give us an idea of the typical geometrical patterns for high-contrast range patches in image space.

As before, we find the binary patches that are nearest to the data points $\mathbf{x}_1 \ldots \mathbf{x}_{N_{\mathrm{tot}}}$. We then group the data points together depending on which equivalence classes (with respect to the symmetry group G) the closest binary patches belong to. We define the number of range patches in symmetry class $[\mathbf{b}_{k'}]_G$ (as a function of the distance θ to the nearest binary patch) by

$$N_{k'}(\theta) = \sum_{j \text{ s.t. } \mathbf{b}_j \in [\mathbf{b}_{k'}]_G} N_j(\theta) \ , \tag{38}$$

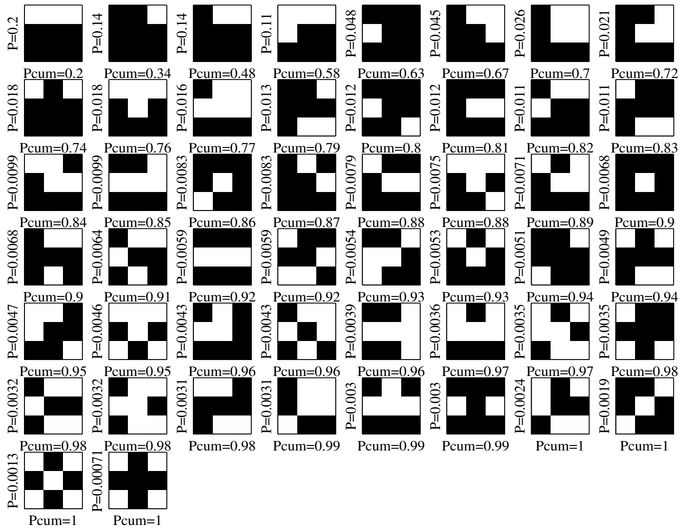

Figure 21. The 50 symmetry classes of binary patches ordered after the percentage $P_{k'}$ of high-contrast range patches closest to one of the binary patches in the equivalence class $[\mathbf{b}_{k'}]_G$. The cumulative percentage $P_{k',\mathrm{cum}}$ is also indicated. For each class, we display the binary patch which is most common among the high-contrast range patches.

where N_j is given in Equation 31. Furthermore, the density of range patches in the symmetry class $[\mathbf{b}_{k'}]_G$ is defined by

$$p_{k'}(\theta) = \frac{N_{k'}(\theta)/N_{\mathrm{tot}}}{\mathrm{size}([\mathbf{b}_{k'}]_G) \cdot V(\theta)/\mathrm{vol}(\mathrm{S}^7)} , \tag{39}$$

where $V(\theta)$ is given by Equation 33, and $\mathrm{size}([\mathbf{b}_{k'}]_G)$ is the number of equivalent patches in the class $[\mathbf{b}_{k'}]_G$.

Figure 21 displays the 50 symmetry classes of binary patches ordered after the percentage of range patches $P_{k'} = \sum_\theta N_{k'}(\theta)/N_{\mathrm{tot}}$ in each class. For each symmetry class $[\mathbf{b}_{k'}]_G$, we show the binary patch $\mathbf{b}_j \in [\mathbf{b}_{k'}]_G$ that is most common among the range patches. The figure also shows the cumulative percentage $P_{k',\mathrm{cum}} = \sum_{j' \leq k'} P_{k'}$. From these numbers we conclude that most high-contrast range patches cluster around binary patches that belong to only a few of the 50 symmetry classes; in fact, 70% of the patches are closest to the first 7 symmetry classes. The most common structures among high-contrast patches are horizontal and vertical edges followed by slanted edges, corner- and T-junction–like structures. The least probable structures are checkerboard and cross patches. Figure 21 agrees with our Voronoi results for range patches (Figure 8), as the patterns of the 25 most frequent Voronoi

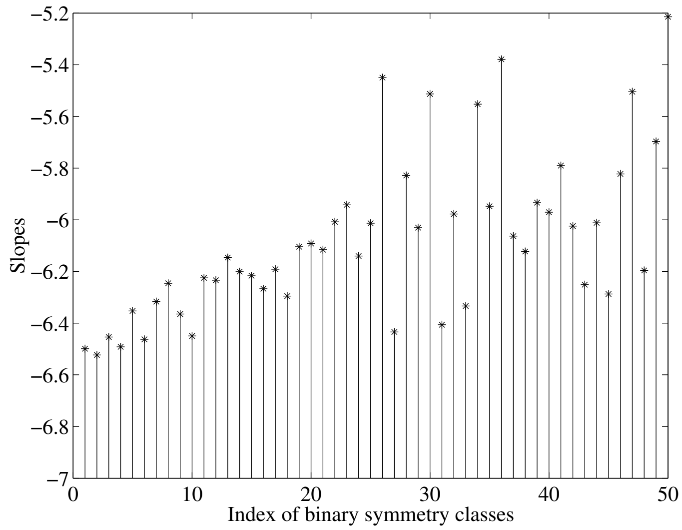

Figure 22. Slopes in a log-log plot of the densities $p_{k'}(\theta)$ for each of the 50 binary symmetry classes $[\mathbf{b}_{k'}]_G$ ordered after decreasing cumulative percentage $P_{k',\mathrm{cum}}$.

cells resemble the patterns of the patches in the 5 most frequent binary symmetry classes.

The graphs of the densities $p_{k'}(\theta)$ (Equation 39) for the 50 symmetry classes are similar in appearance to Figure 19. In Figure 22, we show the slopes obtained by linear regression in a log-log plot of $p_{k'}(\theta)$ ordered after decreasing cumulative percentage $P_{k',\mathrm{cum}}$. The most frequent symmetry groups have very steep density curves, and there is a gradual decrease in the slopes of the curves for the less frequent symmetry classes. This is consistent with the result that most patches are close to binary patches which belong to the 7 most frequent classes.

7. Summary and Conclusions

In this work, we have taken a somewhat different approach to natural image statistics: Most of the work in image statistics focus either on modeling 1D or 2D histograms of linear filter reactions or on finding a linear change of basis that sparsifies the data. Few attempts have been made to understand the *full* probability distribution of natural images and the *intrinsic dimension* of the state space of generic image data. It seems that we cannot take full advantage of the sparseness of the state space of the data without this knowledge of natural images.

In this study, we have analyzed the local geometric patterns seen in *generic* images. We believe that simple geometric structures in the world and the sensor properties of the probing device generate observations that are concentrated along predictable shapes (manifolds or clusters) in state space. The basic vocabulary of images (with edges, bars, blobs, terminations etc.) seems to be the same — whether one studies all types of natural images as here, or specific classes of images; e.g. medical images or images of just trees, indoor scenes etc.

Optical and range images measure different aspects of generators (objects) in the world; scene luminance versus distances to the nearest objects in a natural scene, respectively. Thus, there are bound to be differences in the statistics of these two types of images. However, we tend to believe that the main qualitative differences between optical and range images are due to differences in subresolution properties. It seems that basic primitives (such as edges) in the world and *morphological* or ordering filters (such as median and mean filters) lead to *compact clusters* in state space. On the other hand, the same primitives and *averaging* filters lead to *continuous submanifolds* in state space.

In this paper, we have analyzed the probability distribution of 3×3 high-contrast patches from natural images of different modalities (optical versus 3D range images). In the preprocessing stage, we subtracted the mean and contrast-normalized the log-values of each image patch. The state space of the preprocessed image data (from optical or range images) is a 7-dimensional unit sphere in $\mathbb{R}^8$.

As a first exploration, we examined how the data distribute with respect to an approximately uniform Voronoi tessellation of the 7-sphere. The analysis showed that both optical and range patches occupy a very small amount of the total surface area (volume) of the state space: In both cases, half of the data can be divided into a set of Voronoi cells with a total volume of less than 6% of the volume of the 7-sphere. For optical patches, the centers of the most densely populated Voronoi cells resemble blurred step edges. For range patches, they resemble binary patterns.

A more detailed analysis showed clear differences in the probability distributions of optical and range patches:

The majority of high-contrast optical patches are concentrated around a 2-dimensional C^1 *submanifold* embedded in the 7-dimensional sphere. This surface is highly *non-linear* and corresponds to ideal step edges blurred by the optics of the camera. A density calculation showed that the probability density of optical patches is infinite on this ideal surface. About 50% of the optical data points are within a tubular neighborhood of the surface with a width that corresponds to only 9% of the total volume of the state space.

The majority of range patches, on the other hand, are concentrated in *compact clusters*, rather than on a smooth manifold. The centers of these clusters seem to correspond to binary patches, i.e. patches with only two range values. A density calculation around the 510 possible binary patches indicated an infinite probability density at these "hot spots" of the image space. About 80% of the high-contrast patches are in a neighborhood of these spots that correspond to only 0.14% of the total volume of the 7-dimensional unit sphere. The most frequent binary patches are horizontal edges followed by slanted edges and corner- and T-junction–like structures.

Although the analysis in the current paper only deals with 3×3-patches, we believe that the more general results apply to *larger* patches and even *general* filter responses. The picture that seems to emerge is that basic image primitives — such as edges, blobs, and bars — generate low-dimensional and (in general) non-linear structures in the state space of image data. The dimension of the state space is determined by the number of filters or pixels in the analysis; this number is usually very large. The thing to realize is that the intrinsic dimension of the manifolds generated by different primitives is *fixed* and determined by the complexity of the primitives only. The edge manifold we found for optical data is continuous and 2-dimensional. This is because an ideal edge can be characterized by two parameters: the orientation α and the position l of the edge. For optical data and bar structures which can be parameterized by 3 parameters (the orientation, position, and width of a bar), we would expect a 3-dimensional submanifold in state space.

More generally, we believe that one can define a dictionary of probability models on representations of *general* primitives parameterized by $\Phi = \{\phi_1, \phi_2, \ldots\}$ for *any* set of filters $f_1, \ldots, f_N$. In the N-dimensional state space of the filter-based image representations (and averaging linear filters), the image primitives will define manifolds of the general form

$$M(\Phi) = [f_1(\cdot) * I(\cdot; \Phi), \ldots, f_N(\cdot) * I(\cdot; \Phi)]^T ,$$

where $*$ denotes a convolution.

Our empirical results for the edge manifold and binary patches (of optical and range data) show that, when studying natural image statistics, it is best not to put such strong assumptions on the data analysis as independent components or sparse coding by linear change of basis. In this work, we have furthermore seen clear differences between different types of images (optical versus range) that otherwise have seemingly similar statistics; e.g. 1D marginal distributions of filter reactions, second-order statistics, scaling properties etc. When one looks at the *full* probability distribution of small patches, the differences are very clear. Analyzing 3×3 patches, for example recursively, could thus offer a systematic and precise way of distinguishing and comparing image data. A complete description of the probability distribution of natural image patches, however, requires also modeling low-contrast patches, and high-density regions that lie outside the ideal manifold of step edges (for optical patches) and binary clusters (for range patches). The Voronoi tessellation in Section 4 offers an automatic way of characterizing the full state space of 3×3 pixel data, while the "geometry-based" methods in Section 5 and Section 6 may lead to a better understanding and parametric probability models of natural image data.

Notes

[1] This joint distribution is a sufficient statistic for 2×2 blocks modulo mean (3 variables)

[2] We here choose the top 20 percent highest-contrast image patches. The qualitative results of our study, however, seem very robust to the exact choice of the cut-off.

[3] The pixel values in these images are approximately linearly proportional to the scene luminance.

[4] The logarithmic compression of intensities is consistent with the inverse relationship between ambient illumination and human sensitivity. According to Weber's law, the ratio $\Delta L/L$ of the just noticeable difference ΔL and the ambient luminance L, is constant for a wide range of luminances. It is argued that the human visual system uses this adaptation scheme to preserve the relative reflectances (brightness ordering) of an array of surfaces in a scene.

[5] The "D-norm" $||I||_D^2 = \sum_{i \sim j}(I_i - I_j)^2$ is a finite version of the unique scale-invariant norm $\int\int ||\nabla I||^2 dxdy \propto \int\int (\xi^2 + \eta^2)||\hat{I}||^2 d\xi d\eta$ on images $I(x, y)$ (with Fourier transforms $\hat{I}(\xi, \eta)$).

[6] The area of a spherical triangle with radius R is equal to $A = R^2\epsilon$, where $\epsilon = \alpha + \beta + \gamma - \pi$ is the so called spherical excess. The three angles α, β and γ can be directly related to the lengths of the sides (great circular arcs) of the triangle (Bronshtein and Semendyayev, 1998).

[7] For a single point $\mathbf{x}_0 \in \mathrm{S}^7$, the volume

$$\mathrm{vol}\{\mathbf{x} \in \mathrm{S}^7 : \theta \leq \mathrm{dist}(\mathbf{x}, \mathbf{x}_0) \leq \theta + d\theta\} = \mathrm{C}_6\,(\sin\theta)^6 d\theta\;,$$

where C_6 is the volume (surface area) of a 6-dimensional unit sphere. A 2-dimensional surface in S^7 (such as the manifold $\mathrm{M}^2 \subset \mathrm{S}^7$ or the envelop of overlapping spherical caps around the centers of the triangles in the mesh) has 5 normal directions. Hence, for smaller θ, where the sub-manifold or surface is locally linear, $\mathrm{vol}(X_\theta) \propto (\sin\theta)^4\, d\theta$.

[8] The statistics of optical patches from natural images are, to a first approximation, invariant under the operations in G (Geman and Koloydenko, 1999).

References

Bronshtein, I. N. and K. A. Semendyayev: 1998, *Handbook of Mathematics*. Springer-Verlag, 3rd edition.

Buccigrossi, R. W. and E. P. Simoncelli: 1999, 'Image Compression via Joint Statistical Characterization in the Wavelet Domain'. *IEEE Trans Image Processing* **8**(12), 1688–1701.

Conway, J. H. and N. J. A. Sloane: 1988, *Sphere Packings, Lattices and Groups*, No. 290 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag.

Cover, T. M. and J. A. Thomas: 1991, *Elements of Information Theory*. New York: John Wiley & Sons.

Elliott, J. P. and P. G. Dawber: 1979, *Symmetry in Physics*, Vol. 1. New York: Oxford University Press.

Field, D. J.: 1987, 'Relations between the statistics of natural images and the response proporties of cortical cells'. *Journal of Optical Society of America* **4**(12), 2379–2394.

Friedman, J. H.: 1987, 'Exploratory Projection Pursuit'. *Journal of the American Statistical Association* **82**(397), 249–266.

Geman, D. and A. Koloydenko: 1999, 'Invariant Statistics and Coding of Natural Microimages'. In: *Proc. of the IEEE Workshop on Statistical and Computational Theories of Vision*. Published on the Web.

Grenander, U. and A. Srivastava: 2001, 'Probability Models for Clutter in Natural Images'. *IEEE Trans. PAMI* **23**(4), 424–429.

Huang, J., A. B. Lee, and D. Mumford: 2000, 'Statistics of Range Images'. In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1. Hilton Head Island, SC, pp. 324–331.

Huang, J. and D. Mumford: 1999, 'Statistics of Natural Images and Models'. In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*.

Huber, P. J.: 1985, 'Projection Pursuit'. *The Annals of Statistics* **13**(2), 435–475.

Hyvärinen, A.: 1999, 'Survey on Independent Component Analysis'. *Neural Computing Surveys* **2**, 94–128.

Malik, J., S. Belongie, T. Leung, and J. Shi: 2001, 'Contour and Texture Analysis for Image Segmentation'. *International Journal of Computer Vision* **43**(1), 7–27.

Marr, D.: 1982, *Vision*. New York: W. H. Freeman.

Nielsen, M. and M. Lillholm: 2001, 'What Do Features Tell about Images?'. In: M. Kerckhove (ed.): *Scale-Space and Morphology in Computer Vision*. pp. 39–50.

Olshausen, B. A. and D. J. Field: 1996, 'Natural image statistics and efficient coding'. *Network: Computation in Neural Systems* **7**(2), 333–339.

Reinagel, P. and A. M. Zador: 1999, 'Natural Scene Statistics at the Centre of Gaze'. *Network: Computation in Neural Systems* **10**(4), 341–350.

Ruderman, D. L. and W. Bialek: 1994, 'Statistics of Natural Images: Scaling in the Woods'. *Physical Review Letters* **73**(6), 814–817.

Simoncelli, E. P.: 1999a, 'Bayesian Denoising of Visual Images in the Wavelet Domain'. In: P. Müller and B. Vidakovic (eds.): *Bayesian Inference in Wavelet Based Models*. New York: Springer-Verlag, pp. 291–308.

Simoncelli, E. P.: 1999b, 'Modeling the Joint Statistics of Images in the Wavelet Domain'. In: *Proc. SPIE, 44th Annual Meeting*, Vol. 3813. pp. 188–195. Denver, CO, July 99.

Sullivan, J., A. Blake, M. Isard, and J. MacCormick: 1999, 'Object Localization by Bayesian Correlation'. In: *Proc. Int. Conf. Computer Vision*. Corfu, Greece, pp. 1068–1075.

Tu, Z. W., S. C. Zhu, and H. Y. Shum: 2001, 'Image Segmentation by Data Driven Markov Chain Monte Carlo'. In: *Proc. of International Conference on Computer Vision*. Vancouver, Canada.

van Hateren, J. H. and A. van der Schaaf: 1998, 'Independent component filters of natural images compared with simple cells in primary visual cortex'. In: *Proc.R.Soc.Lond.*, Vol. B 265. pp. 359–366.

Wegmann, B. and C. Zetzsche: 1990, 'Statistical Dependence Between Orientation Filter Outputs Used in an Human Vision Based Image Code'. In: *Proc. SPIE Visual Comm. and Image Processing*, Vol. 1360. Lausanne, Switzerland, pp. 909–922.

Zhu, S. C. and D. Mumford: 1998, 'GRADE: Gibbs Reaction And Diffusion Equations. — a framework for pattern synthesis, image denoising, and removing clutter'. In: *Proc. of International Conference on Computer Vision*.