Chapter 7

Scale Invariance of Natural Images

7.1 Introduction

Scale invariance refers to the phenomenon that the marginal distributions of many statistics of natural images are unchanged after the images get scaled. From the information point of view, scale invariance implies that even though individual natural images do change after being scaled, the information from the population of all the scaled natural images is no different than from the population of the original ones. As we shall see, scale invariance is a very robust property of natural images, and despite its simple form, we will argue that it is a non-trivial characteristic of natural images, and therefore an interesting natural phenomenon in its own right.

Scale invariance of natural images is of great interest in vision. It is widely believed that the statistical properties of natural images determine the basic aspects of the visual system. Scale invariance is among the most prominent statistical characteristics of natural images people have ever found. In Knill *et al.* [1], it was demonstrated that human visual system, when discriminating fractal images, is most sensitive to those which are approximately scale invariant. Because of the fractal nature of many texture images, this result suggests the role of scale invariance in texture discrimination by the visual system. Scale invariance is also an important ingredient in various theories on sensory coding. In Field [3], it was proposed that the visual system adopts a sparse coding scheme. One of the reasons that sparse representation is effective, the author argued, is that natural scenes are scale invariant.

Scale invariance has a lot of applications in computer science, especially in computer vision and image compression, and the reasons for this are very much the same as in vision.

This article is concerned with scale invariance of natural images itself rather than its importance to other areas of science. We will consider the following fundamental problem about scale invariance: Why are natural images scale invariant? To pursue an answer to this problem not only helps better understanding scale invariance, but also gives insight into statistical characterization of natural images. In the next sections, efforts are devoted to establishing a model on the origin of scale invariance of natural images.

A remark follows. Natural images are always presumed to be translation invariant, or

stationary. Stationarity means that over the ensemble of natural images, the statistics at one location are the same as any other location. This is a reasonable assumption because, intuitively speaking, we can not observe "special" locations in images where the statistics tend to be peculiar. It implies that over the ensemble of natural images, all features have the same probability of occuring in one location versus another. From now on, when we say scale invariance of natural images, we always mean scale and translation invariance of these images.

There have been only a few models on the origin of scale invariance of natural images. One recent example is given in Rudderman [3]. According to this model, images are generated randomly by superimposing "objects" at random locations on a plane. Objects are planar patches with independent random shapes and sizes. Each object is also independently painted by a single random color. It was argued that if sizes of objects are distributed by a power-law, then images of the plane, when the plane is fully covered by the objects, have some scale invariant statistics.

Another model is presented in Mumford [4]. As the model in [3], images are made up of independent objects. Unlike that model, however, objects are patches of patterns, shadows, textons, etc., which means that within each object, the color is not a constant, but a function of location inside the object. The formation of images is by superimposing independent randomly scaled objects on a plane at random locations. The biggest difference between these two models lies in their explanations of the cause of scale invariance. In [3], it is the occlusion that is the main reason for scale invariance. On the other hand, in [4], only when occlusion is ignored, can images obtained in the above way be considered as scale invariant. The first model can get scale invariance only for some statistics, while the second one, when ignoring occlusion, guarantees scale invariance of all statistics.

Different as they are, both models consider objects as patches distributed on a plane. Such objects can only be considered as intermediate because they do not have clear physical meaning. After all, natural images are perspective projections of the real world, which is three dimensional, onto a planar surface. With high order approximation, it can be assumed that the projection is through an ideal camera in which the effects of diffraction, aberrations, and discrete sampling are absent. Because the world can be broken up into physical objects, it is therefore reasonable to presume that images consist of perspective projections, or 2D views, of the objects. A Poisson law is proposed as the law of distribution of objects in the three dimensional world. It is argued that the Poisson law of distribution of objects and the perspective projection of objects onto the camera image plane lead to approximate scale invariance of natural scenes. As in [4], only when the effects of occlusion are neglectable, can this argument be correct.

A by-product of our model is the representation of natural scenes as sums of wavelets. This representation was also proposed in [4]. As said earlier, the model proposed in this article gives wavelets a natural explanation. Another representation of natural scenes by sums of wavelets was given in [3]. However, it lacks the randomness which characterizes the wavelet representation derived from the Poisson model.

The article proceeds as follows. Section 7.2 discusses some evidence of scale invariance of natural images. We will argue that scale invariance is a very special property which separates natural images from other visual signals. Then we will formulate scale (and translation)

invariance mathematically. In order to establish a model on the origin of scale invariance, we will first study some simple properties of scale invariant images. Section 7.3 gives one of such properties, which is the law of size of object in scale and translation invariant images. We will motivate the law by two arguments. Section 7.4 gives details of our model and section 7.5 concludes by showing the numerical results.

7.2 Evidence of Scale Invariance of Natural Images

This section presents evidence of scale invariance of natural images. But we will start by making clear what scaling for images is.

7.2.1 Scaling of Images

A (digitized) image I on an $M \times N$ lattice is simply a matrix with M rows and N columns. We adopt the convention of C language, so that the elements of I are represented by $I(i, j)$, where i is the row number running from 0 to $M - 1$, and j is the column number running from 0 to $N - 1$.

Scaling is achieved in the following way. To scale down an $M \times N$ image I by factor k, we take the disjoint $k \times k$ blocks $B_{ij} = [ik, (i+1)k-1] \times [jk, (j+1)k-1]$ in I and compute the average intensity value of each block. The average intensity of the block B_{ij} is taken as the intensity value at (i, j) in the down-scaled image. In mathematical terms, if $I^{(k)}$ denotes the down-scaled image, then it is a $\lfloor M/k \rfloor \times \lfloor N/k \rfloor$ matrix such that for $0 \le i \le \lfloor M/k \rfloor - 1$ and $0 \le j \le \lfloor N/k \rfloor - 1$,

$$I^{(k)}(i, j) = \frac{1}{k^2} \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} I(ik + n, jk + m). \tag{7.1}$$

Why is scaling defined in this way? Naturally, we can imagine that every finite image is part of an infinite image, still denoted I, which is defined on the whole integer grid. Assume for each infinite image I, there is an underlying function $\phi(x, y)$ defined on $\mathbf{R}^2$, such that the value of $I(i, j)$ is the average of $\phi(x, y)$ over the square $S_{ij} = [id, (i+1)d] \times [jd, (j+1)d]$, where $d > 0$ is a constant, i.e.

$$I(i, j) = \frac{1}{d^2} \int_{S_{ij}} \phi(x, y) dx dy. \tag{7.2}$$

In other words, I is a digitized version of ϕ at "sampling rate" $1/d$. In order that the average of ϕ over S_{ij} makes sense, we assume ϕ is "regular", e.g., measurable. This is an ideal model, because in real images, pixel intensity values are responses of complicated filters to the visual signal. The filters may not be distributed on a square lattice, and their supports can overlap with each other.

Under the ideal model given by (7.2), for each $k \ge 1$, define an infinite image $I^{(k)}$ by (7.1), with i, j running through all integers. Then

$$I^{(k)}(i, j) \quad = \quad \frac{1}{k^2} \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} I(ik + n, jk + m)$$

$$
\begin{aligned}
&= \frac{1}{k^2d^2}\sum_{n=0}^{k-1}\sum_{m=0}^{k-1}\int_{S_{ik+n,jk+m}}\phi(x,y)dxdy \\
&= \frac{1}{k^2d^2}\int_{[ikd,(i+1)kd]\times[jkd,(j+1)kd]}\phi(x,y)dxdy \\
&= \frac{1}{d^2}\int_{S_{ij}}\phi^{(k)}(x,y)dxdy,
\end{aligned}
$$

where

$$
\phi^{(k)}(x,y) = \phi(kx,ky),
$$

which demonstrates that $I^{(k)}(i,j)$ is the average of $\phi^{(k)}$ on S_{ij}. By the definition of scaling for functions defined on continuum, $\phi^{(k)}$ is the down-scaled by factor k version of f. It is therefore natural to define $I^{(k)}$ as the down-scaled by factor k version of the image I.

Scaling simulates two situations. Firstly, suppose a natural scene produces a (continuous) image $\phi(x,y)$ on a camera's image plane. Usually the distance between a natural scene and the camera is much larger than the focal distance of the camera, and therefore the camera image plane is almost located right at the focus. As the focal distance of the camera changes while the camera itself stands still, in order to get focused images of the same scene, the camera image plane needs to move closer or farther away from the camera lens, depending on whether the focal distance decreases or increases. In this case, in first order approximation, images produced on the image plane are scaled versions of each other. If the focal distance is k times smaller or k times larger, then the images are down-scaled or up-scaled by factor k, respectively. The error of the approximation lies in the fact that a natural scene is composed of objects with different distances from the camera. Only objects at a specific distance can produce truly focused images on the camera image plane. All other objects only produce blurred images. However, since both the diameter of the camera lens and the focal distance are much smaller than the distances of the objects from the camera, the blurring can be ignored. The second situation is called aperture imaging and is less familiar. The apparatus for aperture imaging is almost identical to a camera except that there is a tiny hole instead of a convex lens in the front of the apparatus to let light in. As the apparatus stands still while its image plane moves forward and backward, the images generated on the image plane, instead of being approximately scaled, as in the first situation, are truly scaled versions of each other.

One may think that if a natural scene is viewed from different distances, the images that it produces on the observer's retina or the camera's image plane will be scaled versions of each other. This is however incorrect. Because of perspective effects, as the observer or the camera gets closer to the scene, the nearer objects get larger faster than the farther objects. On the other hand, when the observer or the camera moves away from the scene, the nearer objects get smaller faster than the farther objects. Mathematically, if an object is originally at distance d, then as the observer or the camera moves farther away by distance x, the image of the object is down-scaled by factor $d/(d+x)$ which is a variable in d instead of a constant. This implies that images of objects are not scaled by a common factor, and hence the whole images are not scaled versions of each other.

7.2.2 Experiments

The images that we use are collected from the Internet. All the images are 256×256 matrices with integer intensity values between 1 and 256. Figure 6 shows six pictures in the collection.

It was reported in Zhu *et al.* [5] that the marginal distributions of x and y-derivatives of natural images are scale invariant. We conduct an experiment on our images and confirm the result. For digitized images, derivatives at a pixel are approximated by differences between the intensity values of the pixel and its neighboring pixels. For instance, at a pixel with location (i, j) in an image I, ∇_x and ∇_y are computed by

$$\nabla_x I(i,j) = I(i,j+1) - I(i,j)$$
$$\nabla_y I(i,j) = I(i+1,j) - I(i,j),$$

Notice that i corresponds to the y coordinate while j corresponds to the x coordinate.

In order to get the empirical marginal distribution of derivatives, we first compute the histogram of derivatives for each image. Each histogram has 101 bins evenly dividing the interval $[-255, 255]$ and is normalized so that the sum of the histogram is 1. The average normalized histogram over all the images is then the empirical marginal distribution.

The results are presented in Figure 7.2. To demonstrate that the marginal distributions are really close to each other after images are scaled, we plot the logarithms of the marginal distributions. Figures 7.2a and b plot those of $\nabla_x I^{(k)}$, for $k = 2$ to 5, against $\nabla_x I$. Figures 7.2c and d plot those of $\nabla_y I^{(k)}$, for $k = 2$ to 5, against $\nabla_y I$.

From Figure 7.2, we can clearly see that the marginal distributions are almost unchanged to scaling. Notice the symmetry of the marginal distribution of $\nabla_x I$. The symmetry can be explained as the nature lacks obvious preference of the left over the right or vice versa. There is, however, no such apparent reason for the symmetry of the marginal distribution of $\nabla_y I$.

It is also noticeable that even many individual images have scale invariant marginal distribution. Figure 7.3 shows logarithms of normalized histograms of $\nabla_x I$ for the images in Figure 7.1. As can be seen, some of the histograms have strong scale invariance.

To see if the scale invariance we have observed is approximately independent of calibration, we generate, for each image I, a new image J by the following formula

$$J(i,j) = \log I(i,j).$$

Then we compute the marginal distributions of $\nabla_x J$. The results are given in Figure 7.4. Still, we observe strong scale invariance.

Not only differentiations, but also many other linear filterings produce responses that have scale invariant marginal distributions. We have tested two other kinds of linear filters. The first one is the isotropic center-surround filters, i.e., the Laplacian of Gaussian filters,

$$LG(x,y,s) = C \cdot (x^2 + y^2 - s^2) \exp\left(-\frac{x^2+y^2}{s^2}\right),$$

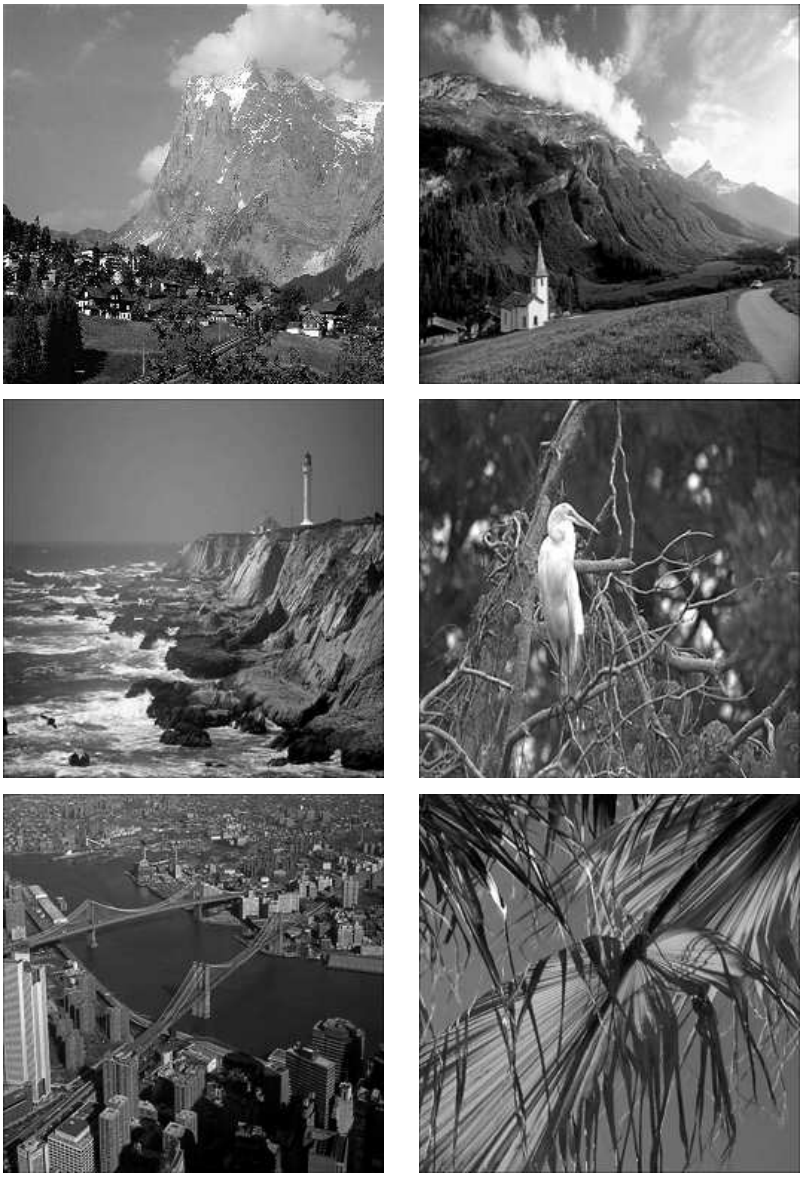

Figure 7.1: 6 out of the 30 collected images

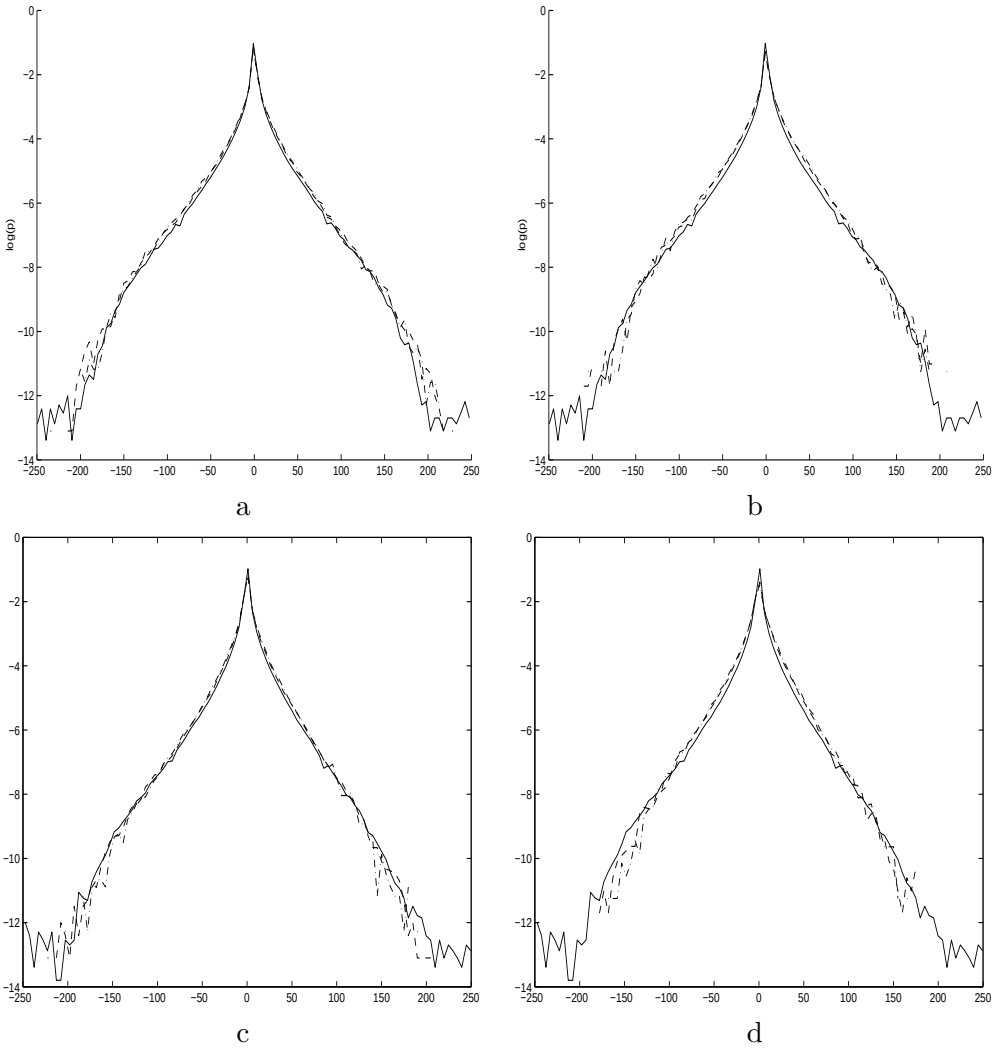

Figure 7.2: Logarithms of marginal distributions of derivatives, solid curves are for unscaled images a. $\nabla_x I^{(k)}$, $k = 2$ (dashed), $k = 3$ (dash-dotted); b. $\nabla_x I^{(k)}$, $k = 4$ (dashed), $k = 5$ (dash-dotted); c. $\nabla_y I^{(k)}$, $k = 2$ (dashed), $k = 3$ (dash-dotted); d. $\nabla_y I^{(k)}$, $k = 4$ (dashed), $k = 5$ (dash-dotted).

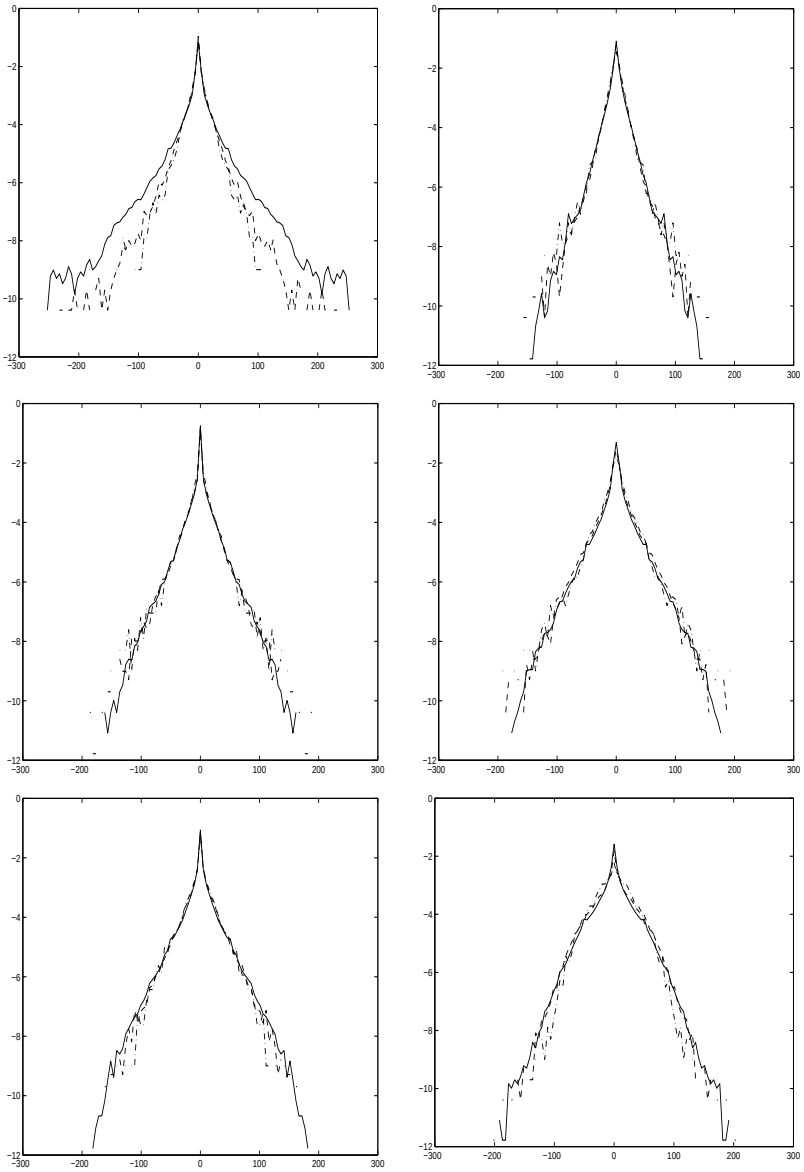

Figure 7.3: Logarithms of normalized histograms of $\nabla_x I^{(k)}$ for images in Figure 7.1, $k = 1$ (solid), $k = 2$ (dashed), and $k = 4$ (dash-dotted)

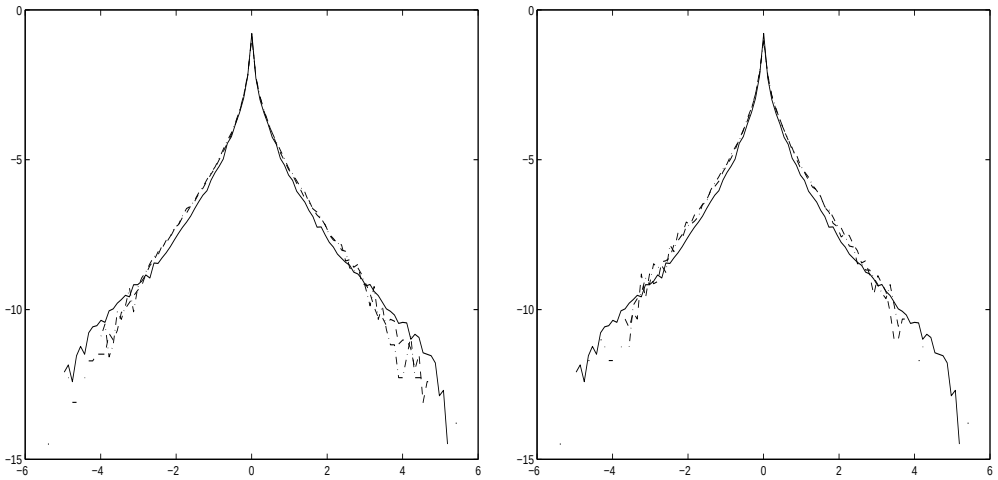

Figure 7.4: Logarithms of marginal distributions of $\nabla_x J^{(k)}$, $J = \log I$. a. $\nabla_x I^{(k)}$, $k = 2$ (dashed), $k = 3$ (dash-dotted); b. $\nabla_x I^{(k)}$, $k = 4$ (dashed), $k = 5$ (dash-dotted);

where C is a constant and s stands for the scale of the filter. We denote these filters by $LG(s)$. The second one is Gabor filters, which is defined as

$$G(x, y, s, \theta) = C \cdot \exp\left(-\frac{4z^2 + w^2}{2s^2}\right) \exp\left(-i\frac{2\pi z}{s}\right),$$

where

$$\begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The real and image parts of the filters are denoted by $Gcos(s, \theta)$ and $Gsin(s, \theta)$, respectively.

In Figure 7.5, we plot logarithms of marginal distributions of responses to these two kinds of filters with different parameters, and again we observe scale invariance of the marginal distributions.

7.2.3 Discussion

That natural images have rich structures and scale invariant distributions makes them distinguished from noise signals. Firstly, Cauchy noise images do scale. Indeed, for a Cauchy noise image I, $I(i, j)$ are i.i.d. random variables with density function,

$$f(x) = \frac{1}{\pi}\frac{1}{1 + x^2}, \quad -\infty < x < \infty,$$

and characteristic function $\psi(u) = e^{-|u|}$. For each $k \geq 1$, $I^{(k)}(i, j)$ is the average of k^2 independent random variables from f. It is then seen that the characteristic function of $I^{(k)}(i, j)$ is still $e^{-|u|}$, implying $I^{(k)}$ and I have the same distribution, and therefore Cauchy

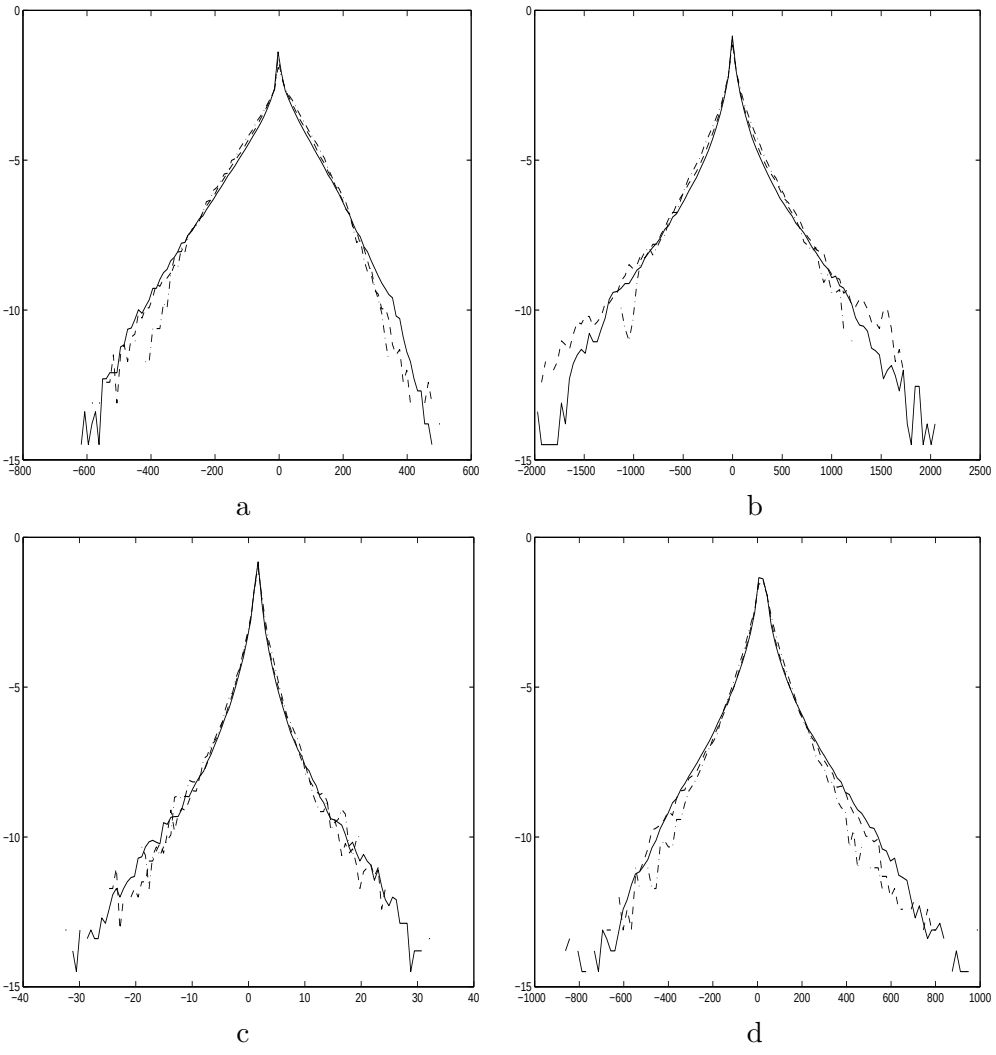

Figure 7.5: Logarithms of marginal distributions of $F * I^{(k)}$, $k = 1$ (solid), $k = 2$ (dashed), $k = 4$ (dash-dotted). a. $F = LG(2.5)$, b. $F = Gsin(4, 0)$, c. $F = Gcos(4, \pi/2)$, d. $F = Gcos(3, \pi/4)$

noise images are scale invariant. However, it is very easy to distinguish natural images from Cauchy noise images because the former ones always contain much richer structures.

Secondly, the marginal distribution of derivatives in white noise images is not scale invariant. Indeed, if $I(i,j)$ are i.i.d. $\sim \mathcal{N}(0,1)$, then for each $k \geq 1$, the marginal density function of $\nabla_x I^{(k)}$ is $\mathcal{N}(0,\sigma_k^2)$ with $\sigma_k = \sqrt{2}/k$. Having decreasing variance, the normalized histogram of $I^{(k)}$ becomes "narrower" as k increases. To see this, first note that the values of $\nabla_x I$ are not independent, because for each (i,j), $\nabla_x I(i,j) = I(i,j+1) - I(i,j)$ and $\nabla_x I(i,j+1) = I(i,j+2) - I(i,j+1)$ have dependency. However, $\nabla_x I(i,1)$, $\nabla_x I(i,3), \ldots \nabla_x I(i,1+2s), \ldots$ are independent to each other. By the law of large numbers, the normalized histogram of $\{\nabla_x I(i,1+2s)\}$ converges to the marginal distribution of $\nabla_x I(i,j)$ as the size of I goes to infinity. Similarly, the normalized histogram of $\{\nabla_x I(i,2s)\}$ converges to the marginal distribution of $\nabla_x I(i,j)$ as the size of I goes to infinity. The normalized histogram of $\nabla_x I$ is the average of the two histograms and therefore tends to the marginal distribution of $\nabla_x(i,j)$. Since the marginal distribution is $\mathcal{N}(0,\sigma_k^2)$ with $\sigma_k = \sqrt{2}/k$, then the normalized histogram is increasingly concentrated around 0 as k increases.

One may point out that the numerical results we have shown do not directly involve intensity values of images and there might be some distribution μ, such that if $I(i,j)$ are i.i.d. $\sim \mu$, then I's are perceptually similar to natural images, and, even though I's themselves are not scale invariant, we still can get the same numerical results. However, we argue that this is unlikely to be true. We observed that the normalized histogram of $\nabla_x I$ is scale invariant, which, by an argument similar to last paragraph, implies the marginal distribution of $\nabla_x I(i,j)$ is scale invariant. Since

$$\nabla_x I(i,j) \stackrel{\mathcal{D}}{=} \nabla_x I^{(k)}(i,j),$$

rewriting both sides in terms of differences between pixel values, there is

$$I(i,j+1) - I(i,j) \stackrel{\mathcal{D}}{=} \frac{1}{k^2} \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} [I(ik+n,(j+1)k+m) - I(ik+n,jk+m)].$$

Because $I(ik+n,(j+1)k+m) - I(ik+n,jk+m)$, $0 \leq n,m \leq k-1$ are independent to each other and have the same distribution as $\nabla_x I(i,j) = I(i,j+1) - I(i,j)$, the distribution of $\nabla_x I(i,j)$ is not only infinitely divisible but also a Cauchy distribution. We then get that the sub-image $\{\nabla_x I(i,1+2j)\}$ is a Cauchy noise image. However, this is not the case for natural images. Because even in the subsample $\{\nabla_x I(i,1+2j)\}$ of a natural image, we can observe a lot of structures.

All the experiments we have conducted are on images defined on finite lattice. However, we have seen it is convenient and natural to consider images as defined on $\mathbf{R}^2$. From now on, we use $\phi(x,y)$ to represent an image defined on $\mathbf{R}^2$ and $I(i,j)$ a digitized image defined on a finite or infinite integer lattice.

For digitized images $I(i,j)$, only down-scaling by an integer factor is appropriate. Up-scaling and scaling by a non-integer factor are not well defined. However, because across the ensemble of digitized natural images, we observe scale invariance of many filter responses, no matter how high the image sampling rate is, it makes sense to think that the underlying continuous images have marginal distributions invariant to scaling.

We formulate scale invariance of natural images as follows. Recall we always implicitly require stationarity of images.

Definition. Let $\mathcal{E}$ be a space of functions defined on $\mathbf{R}^2$ (think of $\mathcal{E}$ as the ensemble of natural images), such that for any $\phi \in \mathcal{E}$, any $\lambda > 0$ and any $(a, b) \in \mathbf{R}^2$, $\phi(\lambda x+a, \lambda y+b) \in \mathcal{E}$. A probability distribution on $\mathcal{E}$ is scale and translation invariant if for any $\lambda > 0$ and $(a, b) \in \mathbf{R}^2$,

$$\phi(\lambda x + a, \lambda y + b) \stackrel{\mathcal{D}}{=} \phi(x, y).$$

We need to fill a gap between the observation and our claim. We have observed that many filterings produce responses that have scale invariant marginal distributions. It is natural then to speculate that *all* filterings produce responses which have scale invariant marginal distributions [1]. But why does this imply that the distribution of natural images itself is scale invariant? Indeed, if ϕ is an image and F is a linear filter, than the filter response of ϕ to F is the convolution $F * \phi$ on $\mathbf{R}^2$,

$$F * \phi(x, y) = \int \phi(u, v)F(x - u, y - v)du\ dv.$$

Assuming ergodicity of the distribution of natural images, with probability one, the histogram of $F * \phi$ is the distribution of $F * \phi(0, 0)$, in the sense that, for any $a < b$, as $M \to \infty$,

$$\frac{1}{4M^2} m(\{(x, y) \in [-M, M]^2 : F * \phi(x, y) \in [a, b)\}) \to \text{Prob}(\langle \phi, \bar{F} \rangle \in [a, b)),$$

where $m(\cdot)$ is the Lebesgue measure. But $F * \phi(0, 0) = \langle \phi, \bar{F} \rangle$, where $\bar{F}(x, y) = F(-x, -y)$. Since the histogram of $F * \phi$ is scale invariant, the distribution of $\langle \phi, \bar{F} \rangle$ is scale invariant. Together with the always implicitly assumed stationarity, this leads to

$$E(e^{i\langle \phi(x,y), \bar{F}(x,y) \rangle}) = E(e^{i\langle \phi(\lambda x+a, \lambda y+b), \bar{F}(x,y) \rangle}).$$

If this is true for all filters, then the *characteristic functional* of the probability distribution on images is scale and translation invariant. Since a probability distribution on images is uniquely determined by its characteristic functional, the distribution is scale and translation invariant.

In section 7.2.2 we mentioned that scale invariance of marginal distribution of derivatives be approximately independent of calibration. Indeed, if the sampling rate of a digitized image I is high, for each (i, j), the underlying continuous image ϕ is about constant over the square $S_{ij} = [id, (i + 1)d] \times [jd, (j + 1)d]$, where d is the inverse of the sampling rate. Thus, for any smooth calibration κ,

$$\kappa(I(i, j)) = \kappa \left(\frac{1}{d^2} \int_{S_{ij}} \phi(x, y)dx\ dy \right) \approx \frac{1}{d^2} \int_{S_{ij}} \kappa \circ \phi(x, y)dx\ dy.$$

[1] Strictly speaking, the "raw intensity" of images is not expected to be scale invariant.

If $\phi(x,y) \stackrel{\mathcal{D}}{=} \phi(\lambda x, \lambda y)$, then $\kappa \circ \phi(x,y) \stackrel{\mathcal{D}}{=} \kappa \circ \phi(\lambda x, \lambda y)$. Letting $J = \kappa(I)$, from the above approximation, we get

$$\left\{ J^{(k)}(i,j) \right\} \approx \left\{ \frac{1}{d^2} \int_{S_{ij}} \kappa \circ \phi(kx, ky) dx \; dy \right\} \stackrel{\mathcal{D}}{=} \left\{ \frac{1}{d^2} \int_{S_{ij}} \kappa \circ \phi(x,y) dx \; dy \right\} \approx \{ J(i,j) \} .$$

Therefore, the distribution of J is approximately scale invariant, verifying our suggestion that the scale invariance of marginal distributions is approximately independent of calibration.

Finally, we establish a connection between the version of scale invariance given in the above definition and a result on scale invariance in the literature. It is well known that natural images have power spectrum of the form [3]

$$S(k) = \frac{A}{k^{2-\eta}},$$

where k is the spatial frequency, A is a constant, and η is close to 0. $S(k)$ is defined as

$$S(k) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_{\mathbf{R}^2} \langle \phi(\mathbf{x}) \phi(\mathbf{x}+\mathbf{y}) \rangle e^{-ik\mathbf{v}(\theta)\cdot\mathbf{y}} d^2 y,$$

where for fixed $\mathbf{y}$, $\langle \phi(\mathbf{x}) \phi(\mathbf{x}+\mathbf{y}) \rangle$ is the average of $\phi(\mathbf{x})\phi(\mathbf{x}+\mathbf{y})$ over all $\mathbf{x}$ and all ϕ, and $\mathbf{v}(\theta) = (\cos\theta, \sin\theta)$. The ideal case is that $\eta = 0$. To see the reason for this, note that under the assumption of ergodicity of the distribution of images,

$$\langle \phi(\mathbf{x}) \phi(\mathbf{x}+\mathbf{y}) \rangle = E(\phi(0)\phi(\mathbf{y}))$$

where E is over all ϕ. Therefore, by scale invariance,

$$\begin{aligned}
S(k) &= \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_{\mathbf{R}^2} E(\phi(0)\phi(\mathbf{y})) e^{-ik\mathbf{v}(\theta)\cdot\mathbf{y}} d^2 y \quad (\mathbf{z} = k\mathbf{y}) \\
&= \frac{1}{k^2} \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_{\mathbf{R}^2} E(\phi(0)\phi(k^{-1}\mathbf{z})) e^{-i\mathbf{v}(\theta)\cdot\mathbf{z}} d^2 z \quad (\phi(k^{-1}\mathbf{z}) \sim \phi(\mathbf{z})) \\
&= \frac{1}{k^2} \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_{\mathbf{R}^2} E(\phi(0)\phi(\mathbf{z})) e^{-i\mathbf{v}(\theta)\cdot\mathbf{z}} d^2 z \\
&= \frac{S(1)}{k^2}.
\end{aligned}$$

7.3 The $\frac{1}{r^3}$ Law of Size of Object

Our goal is to build a model on the origin of scale (and translation) invariance of natural images. As a first step to this goal, we consider the distribution of sizes of objects in images. Let r be the one dimension size of object, such as diameter and periphery. As a first order approximation, the density function of r is Cr^{-3}, where C is a constant. There are several arguments to get this the result. We will demonstrate two of them. A third argument, which is based on compositional rules, can be found in Geman [6]. We start from Mumford's line segment argument.

7.3.1 Poisson Line Segment Argument

We consider images composed only of straight line segments with finite lengths. A random image is generated in the following way. First, we produce a sample $\{(x_i, y_i)\}$ from a homogeneous Poisson point process on $\mathbf{R}^2$. With probability 1, the sample is countable. For each (x_i, y_i), we independently sample an r from a distribution with density function $f(r)$ and an angle θ uniformly from $[0, \pi]$. We then put a line segment with length r and orientation $(\cos\theta, \sin\theta)$ at (x_i, y_i), with (x_i, y_i) being its middle point. All the random line segments then compose an image I.

Now suppose I produced by the above random procedure are distributed by a scale and translation invariant law. We want to know the form of $f(r)$, which is the law of size of object in this case.

In order to get $f(r)$, define a function $N(a, b; R)$ such that for any $0 < a < b$, $N(a, b; R)$ is the expected number of line segments of I with midpoints falling into the square $S_R = [0, R] \times [0, R]$ and lengths between a and b. Define $N_2(a, b; R)$ similarly for $I^{(2)}$. Since $I \sim I^{(2)}$, we get

$$N(a, b; R) = N_2(a, b; R).$$

Because $I^{(2)}$ is a down-scaled by factor 2 version of I, any line segment of $I^{(2)}$ contained in S_R is a down-scaled by factor 2 version of a line segment of I in the square S_{2R}, and the latter line segment has length twice larger than the first one. We get

$$N_2(a, b; R) = N(2a, 2b; 2R).$$

The square S_{2R} consists of four disjoint squares, each being identical to S_R. Because the random processes involved to generate the images are homogeneous, we get

$$N(2a, 2b; 2R) = 4N(2a, 2b; R) \Rightarrow N(a, b; R) = 4N(2a, 2b; R).$$

On the other hand, we have

$$N(a, b; R) \propto \int_a^b f(r)dr, \quad a < b.$$

Therefore,

$$\int_a^b f(r)dr = 4\int_{2a}^{2b} f(r)dr.$$

More generally, we can replace 2 by any positive number s to get

$$\int_a^b f(r)dr = s^2\int_{sa}^{sb} f(r)dr.$$

Differentiating with respect to b, we finally get

$$f(r) = s^3 f(sr) \Rightarrow f(r) = \frac{C}{r^3}.$$

It is obvious that $1/r^3$ can not be a density function because it is singular at 0. The problem comes from the assumption that the images can be up-scaled by any factor. This assumption implies that the density of the Poisson point process must be infinity which is impossible. On way to fix the problem is to require that the line segments have lengths larger than a threshold, say, ϵ and only down-scaling of images be allowed. When an image is down-scaled, all line segments with lengths less than ϵ are thrown away. Then with the same argument, we still can get the $1/r^3$ law, except that r should be larger then ϵ.

7.3.2 Coding Theory Argument

Still, we consider images composed of line segments and assume that the images are scale and translation invariant. Our first step is to discretize the images so that end points of digitized line segments are on the lattice $\{(nd_N, md_N)\}_{n,m\in\mathbf{Z}}$, where $d_N = 1/N$ is the resolution.

We want to compare the probability of a digitized line segment l and the probability of its down-scaled by factor k version $l^{(k)}$. To this end we code digitized line segments by a k-ary code. A k-ary code is of the form $a_{n-1}\dots a_1a_0$, where $a_i \in \{0,\dots,k-1\}$ and $a_{n-1} > 0$. For each line segment l, let $c(l;k)$ be its k-ary code. If the coding is optimal, then by Shannon's theorem,

$$\text{Prob}(l) = \frac{1}{k^{|c(l;k)|}},$$

where $|c(l;k)|$ is the length of the code $c(l;k)$.

Suppose the end points of $l^{(k)}$ are (m_1d_N, n_1d_N) and (m_2d_N, n_2d_N). Then each of the k^4 line segments with end points $((km_1+i_1)d_N, (kn_1+j_1)d_N)$ and $((km_2+i_2)d_N, (kn_2+j_2)d_N)$, $i_1, i_2, j_1, j_2 = 0,\dots,k-1$, has $l^{(k)}$ as its down-scaled by factor k version. As N is large enough, $d_N = 1/N$ is small and all the k^4 line segments are spatially close to each other. By continuity, the information about these k^4 line segments is evenly distributed. Therefore, given the k-ary code of $l^{(k)}$, in order to get the whole information about l, we need $\log_k k^4 = 4$ extra bits. This implies

$$|c(l;k)| \approx 4 + |c(l^{(k)};k)| \Rightarrow \text{Prob}(l) \approx \frac{1}{k^4}\text{Prop}(l^{(k)}).$$

For any line segment ℓ on $\mathbf{R}^2$, let l_N be its digitized version at resolution d_N. Then $\ell^{(k)}$ is digitized as $l_N^{(k)}$ at resolution $1/d_N$. We then have

$$\lim_{N\to\infty} \frac{\text{Prob}(l_N)}{\text{Prob}(l_N^{(k)})} = \frac{p(\ell)}{p(\ell^{(k)})} \Rightarrow \frac{p(\ell)}{p(\ell^{(k)})} = \frac{1}{k^4},$$

where $p(\ell)$ is the density of ℓ. The above relation holds for any line segment and any positive integer scaling factor. By continuity, it holds for an arbitrary positive scaling factor. Because orientations of line segments are uniformly distributed and the distribution of images is translation invariant, for any line segments l_r and l_{sr} with lengths r and sr, respectively,

$$p(l_r) = s^4 p(l_{sr}) \Rightarrow p(l_r) = \frac{C}{r^4}.$$

In order to get the marginal distribution of r, we integrate the density function over all line segments with length r and with one end point at the origin. The other end points of these line segments are on a circle with radius r. Therefore the marginal distribution of r is

$$f(r) = 2\pi r \cdot \frac{C}{r^4} = \frac{C'}{r^3}.$$

7.4 The Poisson Model

In this section we build a Poisson model on the origin of scale and translation invariance of natural images. The model has two components (1) distribution of objects in the 3D world, and (2) surface processes that describe intensity distributions inside 2D views of objects. We start by modeling objects of 3D world and after establishing the Poisson model, we will discuss implications and limitations of the model.

7.4.1 Modeling Objects

As discussed in section 7.1, natural images are perspective projections of the 3 dimensional world on an image plane. World breaks up into physical objects with different shapes, surface colors and sizes and so natural images also break up into the viewed surfaces of objects. Thus the first problem that comes up is how to model physical objects.

As a coarse approximation, objects in our model are independent rigid planar templates parallel to the image plane. Each template has a reference point. The position of an object is the spatial location of its reference point.

Several correlated important aspects of real objects are ignored in our model.

(1) Occlusion and orientation of object. The surface of a 3D object always has several different aspects. Because of occlusion by the other aspects of the same surface, an aspect which is visible when the object is at one place can become invisible when the object moves to another place. Even when objects are modeled as planar templates, if an object is not parallel to the camera image plane, then because of perspective effect, when the object moves on a plane parallel to the camera image plane, the farther away it moves from the camera laterally, the larger its 2D view becomes. This effect is not accounted for by our argument. In real situation, since the angle of view of a camera is usually small, the effects incurred by occlusion and orientation are small. In our model, however, we allow an arbitrary large, but fixed angle of view. In order to avoid complications, we require templates be parallel to the image plane.

(2) Dependence between objects. It often happens that in a large region of the world, objects have long range dependence. For example, each window on a building can be considered as an object. The position of a window is obviously not independent to the positions of the other windows on the same building. A solution to this is to define two objects as two parts of a larger object whenever they have dependence. But this can also cause problem. For instance, houses built by a street stand approximately along a straight line. If all the houses are put together as a single object, then the perspective effect on the 2D view of the

object, as mentioned in (1), can not be ignored. Another example is a long river flowing on a plain. It is even hard to model it as a template parallel to the image plane.

(3) Volume of object. Because a real object has a certain volume, when it occurs at a location in the space, other objects can not occupy space arbitrarily close to it. On the other hand, since a template is a planar shape with zero volume, other templates can get arbitrarily close to it.

Let us describe the 3D world by a Euclidean coordinate system. Every point in the world is represented by (x, y, z). Assume that the camera lens is at the origin of the coordinate system. The direction of view of the camera is along the positive x-axis. Suppose the distance between the camera image plane and the lens is 1. Then the image plane is the plane $\{(-1, y, z) : \ y, z \in \mathbf{R}\}$. We also need a coordinate system for the image plane. Let every point on the image plane be represented by (u, v), and let the origin of the image plane coordinate system be the intersection point of the x-axis and the image plane, which is the space point $(-1, 0, 0)$. Because the perspective projection of an object is upside down and left and right reversed, we define the direction of the u-axis as the opposite direction of the y-axis, and the direction of the v-axis as the opposite direction of the z-axis. Then the projection of a spatial point (x, y, z), $x > 0$ is $(y/x, z/x)$ on the image plane.

7.4.2 Distribution of Objects

Under the set-ups of section 7.4.1, we make the following assumption on the distribution of objects in the world.

Assumption 1. Objects are distributed by a homogeneous Poisson law.

We must decide the support of the Poisson law, i.e., the region in which an object can be any where with positive probability. Let us show that the support can not be the whole 3D space. For simplicity of discussion, for now we assume all the objects are identical, i.e., they share the same template. Refer to Figure 7.6. By our set-ups, D in the figure is 1. If the distance between an object with size R and the camera lens is d, then the size of the projection of the object is $RD/d \propto 1/d$. Letting r be the size of the projection, we have

$$r \propto \frac{1}{d}.$$

Now we derive the probability density function $f(r)$. As in section 7.3.1, fix a finite square on the image plane. Images in the square are projections of an infinite cone in the space, illustrated as the shaded area in Figure 7.6. All objects in the cone with distance d from the camera are on a planar region with area proportional to d^2. Assume the distribution of objects is homogeneous 3D Poisson, then the density $g(d)$ of objects in the cone with distance d is also proportional to d^2. From $r \propto 1/d$ and $g(d) \propto d^2$, we get the law of size of object $f(r) \propto 1/r^4$. This is inconsistent with the result in section 7.3, where $f(r) \propto r^{-3}$. The heuristic argument suggests that the support of the Poisson law be assumed other than the whole 3D space. In other words, objects should be modeled as being distributed in a sub-region in the 3D space by a homogeneous Poisson law.

We have to look at the nature more closely. Natural images are taken on the earth. The surface of the earth, within our visible distance, is flat. Although objects can be any where

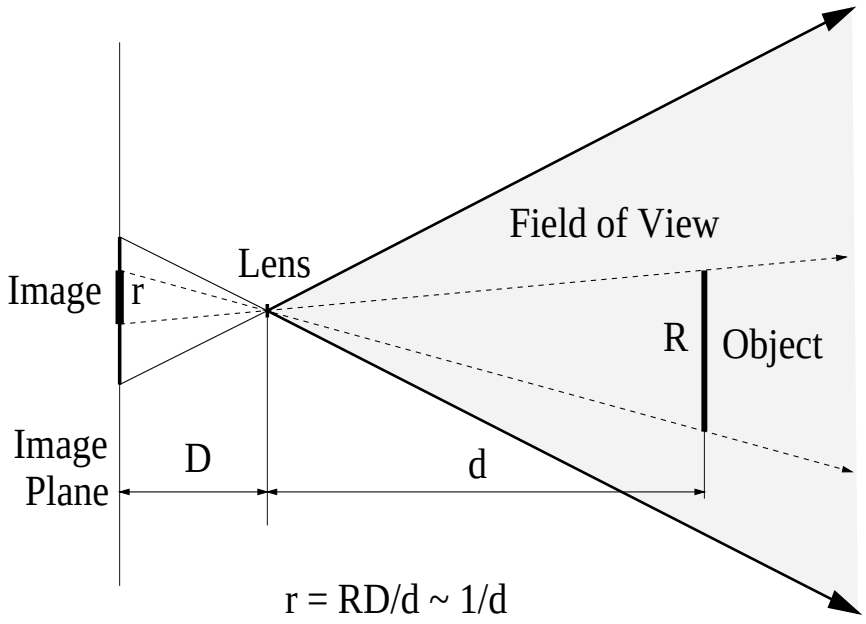

Figure 7.6: Perspective view of an object

above the ground, they are overwhelmingly distributed below a certain altitude. Therefore, we modify our previous assumption to the following.

Assumption 2. There is a constant $H > 0$, such that objects are distributed by a homogeneous Poisson law in the region between the earth and the height H.

Let the plane $z = 0$ represent the earth. Let $\{(x_i, y_i, z_i)\}$ be the positions of objects. Then the assumption means that $\{(x_i, y_i, z_i)\}$ is a sample from a homogeneous Poisson process in the region $\mathbf{R}^2 \times [0, H] = \{(x, y, z) : \ x, y \in \mathbf{R}, 0 \le z \le H\}$.

To see this assumption is consistent with the r^{-3} law of size of object, refer to Figure 7.6 again. When d is large enough, any object in the shaded area with distance d is on a rectangular planar region with width proportional to d and with fixed altitude H. Thus $g(d) \propto d$ and this together with $r \propto 1/d$ leads to $f(r) \propto 1/r^3$.

7.4.3 Surface Processes of Objects

At different distances, the 2D view of an object not only has different sizes, but also shows different surface colors, textures, etc. The distribution of intensity values inside the 2D view of an object, which we want to call "surface process", can be very complicated. It can be not only a smooth function, but also an "irregular" function, e.g., a sample from a random process like white noise. Such irregular functions are called generalized functions in mathematical terms. In any case, the surface process inside the projection of an object is a function of $\mathbf{u} = (u, v)$ on the image plane, with support inside the projection. We use $\psi(\mathbf{u}; \mathbf{x}, T)$ to represent the surface process inside the projection of a template T which is located at $\mathbf{x}$.

Given a template T, suppose P is a point on T with location relative to the reference point of T being $\mathbf{v} = (a, b)$. If T is located at $\mathbf{x} = (x, y, z)$, then the spatial location of P is $\mathbf{y} = (x, y + a, z + b)$ and its projection on the image plane is

$$\mathbf{u} = \left(\frac{y+a}{x}, \frac{z+b}{x} \right).$$

Under ideal conditions, where the effects of decay, scattering, interference, diffraction, etc., as light travels in the space, are absent, and where the camera is ideal, if there are no other objects between P and $\mathbf{u}$, the intensity at $\mathbf{u}$ equals the intensity of the light setting off from P to $\mathbf{u}$. The intensity of the light depends not only on P's own physical condition, which we assume to be unchanged no matter where T is, but also on the lighting condition around the spatial location of P as well as the direction the light goes. We assume that the lighting condition is uniform all over the space. We also assume that the intensity of light from P is constant on all directions. Then the light that goes from P to $\mathbf{u}$ has intensity depending only on P but not on its location in the space. This implies that the light intensity is a function only on the relative location of P on T, and therefore the intensity at $\mathbf{u}$ is determined by $\mathbf{v}$. By our notations, this can be written as

$$\psi(\mathbf{u}; \mathbf{x}, T) = I(\mathbf{v}; T).$$

Write $\mathbf{p} = (y, z)$. Then $\mathbf{x} = (x, \mathbf{p})$ and $\mathbf{u} = x^{-1}(\mathbf{p} + \mathbf{v})$. Therefore,

$$\psi(\mathbf{u}; (x, \mathbf{p}), T) = I(x\mathbf{u} - \mathbf{p}; T).$$

From the equation it is seen that the surface process of a template object T located at $(x, \mathbf{p})$ in space is a scaled and translated version of I. We call this change of surface process by location "color rendering".

We now consider the whole picture. Because the distribution of objects is homogeneous Poisson in $\mathbf{R}^2 \times [0, H]$, with probability one, there are countably many objects in the region $\mathbf{R}^+ \times \mathbf{R} \times [0, H]$. Let the positions of these objects be $\{(x_i, \mathbf{p}_i)\}$, $\mathbf{p}_i = (y_i, z_i)$. At each location $(x_i, \mathbf{p}_i)$, a template T_i is independently selected from a certain distribution. Writing $I_i(\mathbf{u}) = I(\mathbf{u}; T_i)$, then I_i are i.i.d. *If we ignore occlusion, then the whole image I is the arithmetic sum of the projections of all the objects and can be written as*

$$I(\mathbf{u}) = \sum_i I_i(x_i\mathbf{u} - \mathbf{p}_i), \quad I_i \text{ i.i.d.} \tag{7.3}$$

7.4.4 Discussion

The first consequence of the above results is as follows. Fix a large enough rectangular image I. For $0 < a < b$, define

$$A_a^b = \text{Expected total area of regions in } I \text{ which are covered by projections of objects with distance from the camera between } a \text{ and } b.$$

If we ignore occlusion, then for any $\lambda > 0$, $A_{\lambda a}^{\lambda b} = A_a^b$.

Call an object a T-object if the object is a template T. If a T-object is x away from the lens, then its projection has area proportional to x^{-2}. By the Poisson distribution, the density

of T-objects showing up in I with distance from the camera being x is proportional to x. Without occlusion, the total area of the projections of T-objects with distance between a and b is proportional to

$$\int_a^b \frac{1}{x^2} x dx = \log \frac{b}{a}.$$

Integrating over all possible templates, we get, without occlusion,

$$A_a^b \propto \log \frac{b}{a} \Rightarrow \text{for all } \lambda > 0, A_{\lambda a}^{\lambda b} = A_a^b.$$

The same argument also applies to explain the scale invariance of marginal distribution of derivatives. Without loss of generality, consider ∇_u. Given a template T, let

$$\begin{aligned} D_a^b(x;T) &= \text{Total area of regions in the projection of } T \text{ where values of } \nabla_u \text{ are} \\ &\quad \text{between } a \text{ and } b \text{ when the distance between } T \text{ and the camera is } x. \end{aligned}$$

Since the surface process inside the projection of a T-object with location $\mathbf{x} = (x, \mathbf{p})$ is $\psi(\mathbf{u}; \mathbf{x}, T) = I(x\mathbf{u} - \mathbf{p}; T)$, the derivative at $\mathbf{u}$ is $x\nabla_u I(x\mathbf{u} - \mathbf{p}; T)$, therefore

$$\nabla_u \psi(\mathbf{u}; (x, \mathbf{p}), T) \in [a, b] \Leftrightarrow \nabla_u I(x\mathbf{u} - \mathbf{p}; T) \in [ax^{-1}, bx^{-1}].$$

By the fact that the area of the projection of the T-object is proportional to $1/x^2$,

$$D_a^b(x;T) = \frac{1}{x^2} D_{a/x}^{b/x}(1;T).$$

Let D_a^b be the expected total area of regions in I where $\nabla_u I$ is between a and b. Neglecting occlusion and integrating the above equation over all $x \in (0, \infty)$ and all T, D_a^b is proportional to

$$\begin{aligned} \int d\mu(T) \int_0^\infty D_a^b(x;T) g_T(x) dx &= \int d\mu(T) \int_0^\infty \frac{1}{x^2} D_{a/x}^{b/x}(1;T) c_T x dx \\ &= \int_0^\infty \frac{1}{x} K\left(\frac{a}{x}, \frac{b}{x}\right) dx, \end{aligned}$$

where $d\mu$ is the distribution of T and $g_T(x)$ is the density of T-objects with distance x. By section 7.4.2, $g_T(x) = c_T x$, where c_T is a constant depending only on T. For any $\lambda > 0$,

$$\int_0^\infty \frac{1}{x} K\left(\frac{a}{\lambda x}, \frac{b}{\lambda x}\right) dx = \int_0^\infty \frac{1}{x} K\left(\frac{a}{x}, \frac{b}{x}\right) dx \Rightarrow D_{a/\lambda}^{b/\lambda} = D_a^b.$$

If I is scaled by factor λ, then in the scaled image $I^{(\lambda)}$, the derivative at $\mathbf{u}$ equals λ times the derivative at $\lambda^{-1}\mathbf{u}$ in I. Thus the area of regions in $I^{(\lambda)}$ where derivatives are between a and b, denoted $\bar{D}_a^b$, is proportional to $D_{a/\lambda}^{b/\lambda} = D_a^b$. Therefore, for any $a < b$,

$$\frac{\bar{D}_b^a}{Area(I^{(\lambda)})} = \frac{D_b^a}{Area(I)}.$$

Under the assumption of ergodicity of images,

$$\frac{D_b^a}{Area(I)} = \text{marginal propability that } \nabla_u I \in [a, b],$$

and

$$\frac{\bar{D}_b^a}{Area(I^{(\lambda)})} = \text{marginal propability that } \nabla_u I^{(\lambda)} \in [a, b].$$

Then it is seen the marginal distribution of derivatives of $I^{(\lambda)}$ is the same as I.

The expression (7.3) strongly suggests using randomly scaled and translated "template functions" to represent images. These template functions, as called in [4], are random wavelets. There random wavelets are explained as random patches superimposed on a planar region. Here we find a natural explanation for random wavelets: they are the projections of objects randomly distributed in the 3D world. It is interesting that by modeling images in different ways, the same form of random wavelet representation is obtained. Further study of using random wavelet expansion to construct scale and translation invariant distributions on images is given in next chapter.

7.5 Numerical Experiment

For real images, occlusion can not be ignored. Unfortunately, there are few methods to analyze the effects of occlusion on our model. We resort to numerical experiments to check how well the Poisson model approximates scale invariance.

We simulate putting objects in the spatial region $\mathbf{R} \times \mathbf{R} \times [0, H]$ and projecting them on a finite rectangle camera film. To prevent images from being covered by the projections of only a few objects which are very close to the camera, the simulation only allows objects with distance from the camera larger than a lower bound. An upper bound is also selected for the distance, so that if an object has distance larger than the maximum value, its 2D view is smaller than a pixel. Only objects with distance between the lower and upper bounds are generated.

Figure 7.7 illustrates the side view of the camera as well as the "world" in the simulation. The lens is located on the earth, i.e., with z-coordinate equal to 0. Note that to project objects, which are distributed above the earch, onto the film, the film has to be put "under" the earth, as show in the picture.

The actual implementation does not involve sampling objects in space. When plotting 2D views of objects, we need first know their positions on the image as well as their scaling factors. The positions and scaling factors can be sampled based on the following observation. Let the film be the rectangle $[-1, 1] \times [0, 1]$. Given the distance x of an object, the scaling factor of its 2D view is x and the positions $\{(u_i, v_i)\}$ of all the 2D views with scaling factor x that occur on the image film compose a sample from a Poisson point process with density λx on the region $[-1, 1] \times [0, H/x]$.

A pseudo-code for sampling positions and scaling factors of 2D views is as follows. Note that the scaling factors are discretized.

POSITION-SCALING

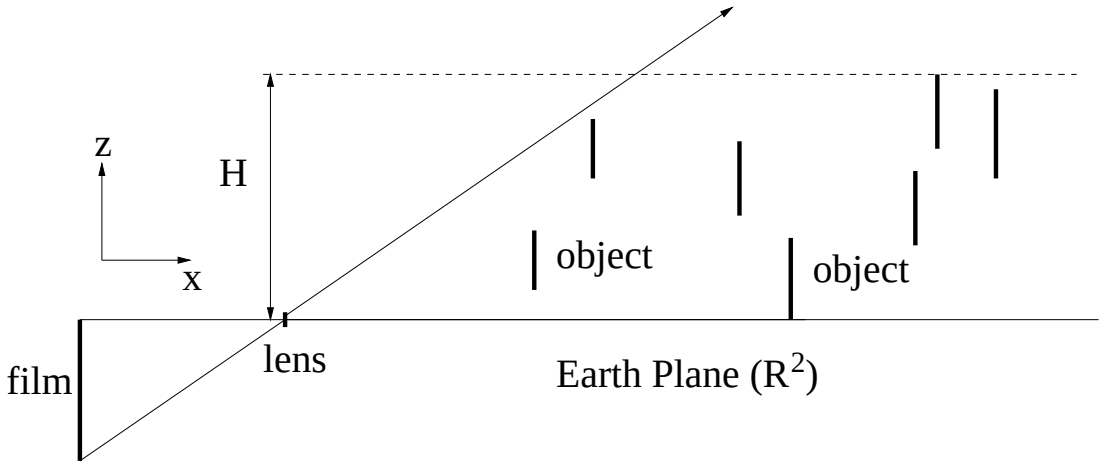

Figure 7.7: Side View of the "World" and the Camera

fix the film as the plane region $[-1,1] \times [0,1]$
fix $D_{\min}$, $D_{\max}$ and H
fix density λ and step size ϵ
$P \leftarrow \emptyset$
$D \leftarrow D_{\min}$
while $D \leq D_{\max}$ **do**
  sample N from the Poisson distribution $\text{Prob}(N = n) = \dfrac{e^{-D\lambda}}{n!}(D\lambda)^n$
  sample N (u, v)'s independently from $[-1,1] \times [0, H/D]$
  $P \leftarrow P \cup \{(D, u_i, v_i)\}_{i=1}^{N}$
  $D \leftarrow d + \epsilon$
return P

Given their positions and scaling factors, the second step is to plot the 2D views of the objects. To simulate occlusion, we start from those with the largest scaling factor, which corresponds to the largest distance from the camera. When two 2D views overlap, the one with larger scaling factor is overwritten by the other one. The input of the following subroutine is $P = \{(D_i, y_i, z_i)\}_{i=1}^{n}$ with $D_1 \geq D_2 \geq \ldots D_n > 0$.

DRAW(P)
  for $i \leftarrow 1$ to n **do**
    SUPPER-IMPOSE(D_i, y_i, z_i)

To display the image, we digitize the film $[-1,1] \times [0,1]$ by dividing it into $2N \times N$ squares indexed by (i, j), $i = -N, -N+1, \ldots, N-1$, $j = 0, \ldots, N-1$. The value at pixel (i, j) is the average intensity value of the image over the square $S_{ij} = [id, (i+1)d] \times [jd, (j+1)d]$, where $d = 1/N$.

The templates we use are rectangles and circles with random sizes. The surface processes are smooth functions plus white noise. Suppose we want to plot the 2D view of a T-object located in the space at $(x, \mathbf{p})$ with surface process $I(x\mathbf{u} - \mathbf{p}; T)$. Then $I(\mathbf{u}; T) = I_s(\mathbf{u}; T) + W(\mathbf{u}; T)$, where I_s is a smooth function and W is a white noise with variance σ^2.

Therefore

$$\frac{1}{d^2}\int_{[0,d]\times[0,d]} W(\mathbf{u})d^2u$$

is Gaussian random variable with distribution $N(0,\sigma^2/d^2)$. Then for a pixel (i,j) with the square S_{ij} being inside the 2D view of the T-object, its intensity value is

$$\begin{aligned} & \frac{1}{d^2}\int_{S_{ij}} I_s(x\mathbf{u}-\mathbf{p};T)d^2u + \frac{1}{d^2}\int_{S_{ij}} W(x\mathbf{u}-\mathbf{p};T)d^2u \\ = \quad & \frac{1}{d^2}\int_{S_{ij}} I_s(x\mathbf{u}-\mathbf{p};T)d^2u + \frac{1}{x}\xi(i,j), \end{aligned}$$

where $\xi(i,j)$ are i.i.d. $\sim N(0,\sigma^2/d^2)$.

A pseudo-code for the above procedure is as follows. For simplicity, we only show how to plot a scaled disc with a random surface process. In addition, the constant d is assumed to be 1 in the code and therefore $\sigma^2/d^2 = \sigma^2$.

SUPPER-IMPOSE(x,y,z)
  pick s randomly from {"disc", "rectangle", ...}
  if $s =$"disc" **then**
    sample a random radius r, a smooth function I_s and a variance σ from certain distributions
    for $i \leftarrow -N$ to $N-1$
      for $j \leftarrow 0$ to $N-1$
        if $|(id,jd)-(y,z)| \le \frac{r}{x}$ **then**
          sample a ξ from $N(0,\sigma^2)$
          $I(i,j) \leftarrow \frac{1}{d^2}\int_{S_{ij}} I_s(x\mathbf{u}-\mathbf{p})d^2u + \frac{1}{x}\xi$
  else if $s =$"rectangle" **then**
    ...

Figure 7.8 plots logarithms of marginal distributions of $\nabla_x I$ for images generated by the simulation. It shows good scale invariance. In Figure 7.9, we present a sampled scene. From the picture we see that the "color rendering" makes 2D views of closer objects look like having more details and 2D views of farther objects look smoother.

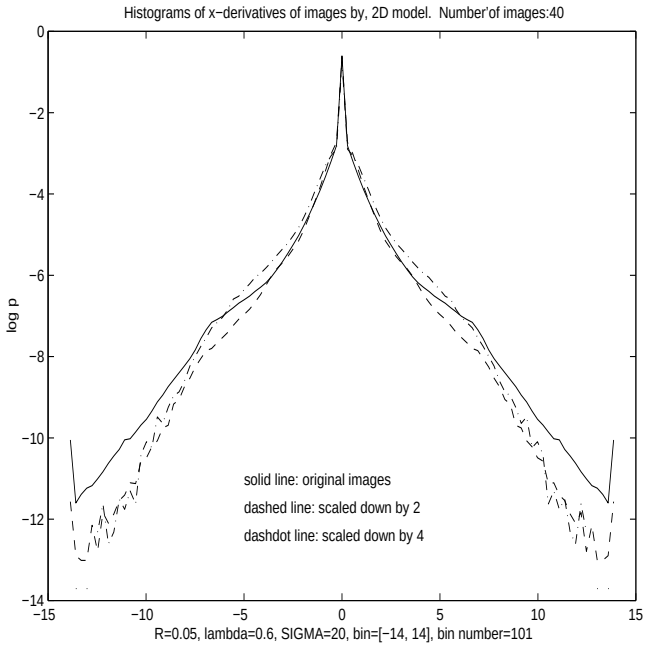

Figure 7.8: Logarithms of the marginal distributions of $\nabla_x I$, $i = 1$ (solid), $i = 2$ (dashed), and $i = 3$ (dash-dotted)

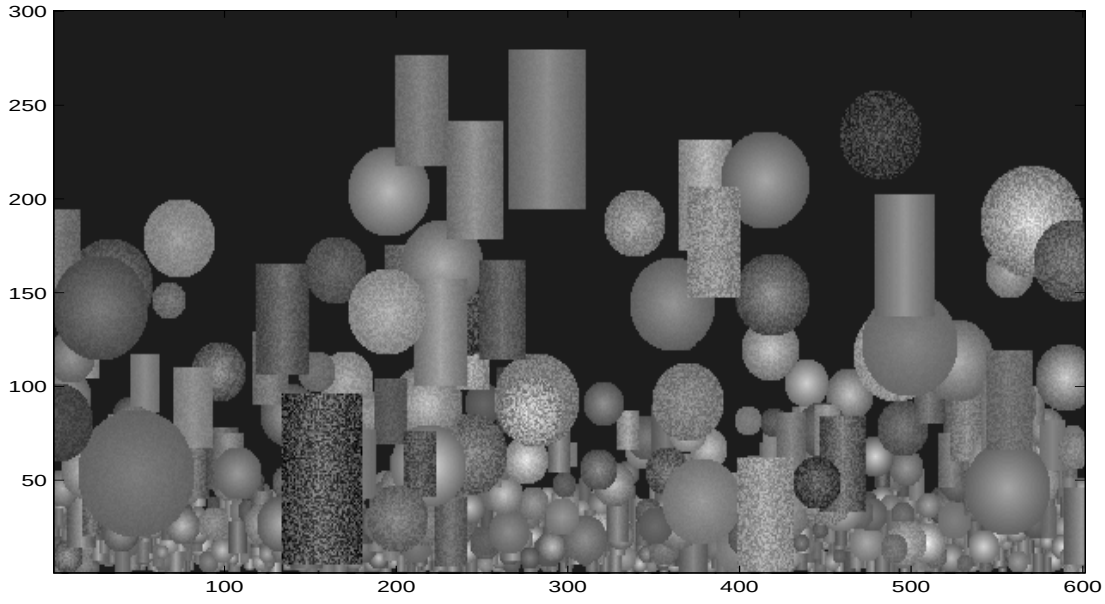

Figure 7.9: A sample scene.

Bibliography

[1] D. C. Knill, D. J. Field, and D. Kersten. Human Discrimination of Fractal Images. *Journal of the Optical Society of America A, Optics and Image Science*, Vol. 7, No. 6. June, 1990.

[2] D. J. Field. What Is the Goal of Sensory Coding? *Neural Computation* **6**, 559-601. 1994.

[3] D. L. Rudderman. Origins of Scaling in Natural Images. *Vision Research*. December, 1996.

[4] D. B. Mumford. Stochastic Models for Generic Images. Division of Applied Mathematics, Brown University. 1998. In preparation.

[5] S. C. Zhu and D. B. Mumford. Prior Learning and Gibbs Reaction-Diffusion. *IEEE Transactions on PAMI*. 1997.

[6] S. Geman. Invariant Binding in Composition Systems. Technical report, Division of Applied Mathematics, Brown University. 1998. In preparation.