

## 自然图像统计：树林中的缩放

Daniel L. Ruderman\*

William Bialek

翻译：林绪虹†

### 摘要

为了更好地理解视觉系统，我们应该尝试描述它处理的自然图像。我们从树林中收集图像，发现这些场景具有整体尺度不变性。此外，它们是高度非高斯的，这种非高斯特性无法通过局部线性滤波去除。我们发现，在滤波过程中加入简单的“增益控制”非线性使得滤波器输出非常高斯，这意味着信息在固定通道方差下最大化。最后，我们使用测量的功率谱来设置受体阵列传达的有关自然场景的信息的上限。

*Keywords:* 自然图像统计.

## 1. 介绍

自然刺激在我们理解感官处理方面发挥着越来越重要的作用。这是因为感官系统执行任务的能力是一个统计量，取决于信号和噪声特性。最近，有几种方法探索了与自然图像相关的视觉处理（Atick & Redlich '90、Bialek *et al* '91、van Hateren '92、Laughlin '81、Srinivasan *et al* '82）。然而，对自然场景的良好描述却非常缺乏。在本文中，我们分析了来自树林的图像，以努力弥补这一差距。我们进一步尝试了解生物视觉系统应如何最好地编码这些图像。

## 2. 图像

我们的图像由  $256 \times 256$  像素  $I(x)$  组成，这些像素已根据亮度进行校准（见附录）。我们将图像对比度对数定义为

$$\phi(\mathbf{x}) = \ln(I(\mathbf{x})/I_0),$$

其中  $I_0$  是针对每幅图像定义的参考强度。我们选择此常数，使得  $\sum_{\mathbf{x}} \phi(\mathbf{x}) = 0$ ；也就是说，每幅图像的平均对比度为零。我们的分析针对的是对比度数据  $\phi(\mathbf{x})$ 。

### 3. 缩放

最近的测量（Field '87、Burton & Moorhead '87）表明，自然场景的集合具有尺度不变性。这意味着，任何在给定尺度上定义的量都具有不受该尺度变化影响的统计数据。考虑到图像由所有距离的物体组成，因此没有特定的角度尺度应该突出，这似乎是合理的。（请注意，这并不意味着任何特定图像都是分形的！相反，场景的集合具有不受尺度影响的统计数据。）

#### 3.1. 对比度分布

我们可以通过观察各种数量的统计数据如何随尺度变化来直接测试这一缩放假设。我们将大小为  $N \times N$ （像素）的框的平均对比度定义为

$$\phi_N = \frac{1}{N^2} \sum_{i,j=1}^N \phi(i,j)$$

我们现在问：“概率  $P(\phi_N)$  随  $N$  如何变化？”

在图 1 的左图中，我们绘制了  $\log(P(\phi_N/\phi_N^{RMS}))$  for  $N = 1, 2, 4, 8, 16, 32$  以及与方差相同的高斯分布相对应的抛物线。通过除以 RMS 值，我们只需将所有图形绘制在同一对比度尺度上即可。所有图形都位于彼此之上，这意味着对比度尺度 - 分布的形状不受角度尺度变化的影响。请注意，概率远非高斯分布，因为图形具有线性而不是抛物线尾部。即使在平均近 1000 个像素（在  $32 \times 32$  的情况下）之后，它仍然是非高斯分布。中心极限定理的这种分解意味着像素在很长的距离上是相关的。这类似于临界点的热力学系统的物理学。

#### 3.2. 梯度分布

作为缩放的另一个例子，我们考虑图像梯度的概率分布。我们通过离散近似来定义梯度的幅度，使得

$$G(\mathbf{x}) = |\mathbf{G}(\mathbf{x})| \approx |\nabla\phi(\mathbf{x})|.$$

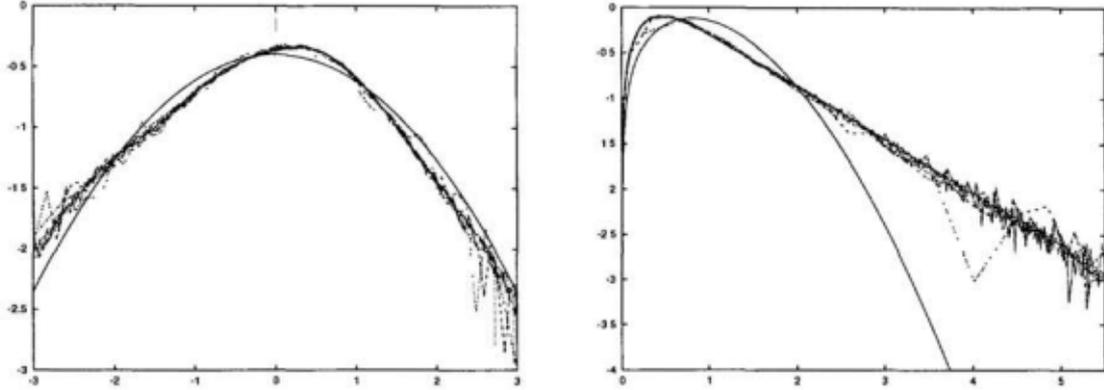


图 1: 左图:  $P(\phi_N/\phi_N^{RMS})$  for  $N = 1, 2, 4, 8, 16, 32$  的半对数图, 与方差相同的高斯分布进行比较 (实线)。右图:  $P(G_N/\bar{G}_N)$  的半对数图, 与同一组  $N$  的瑞利分布进行比较 (实线)。

我们通过首先按上述方式重新缩放图像, 然后在新比例下评估梯度, 以在不同比例下检查此数量。我们在图 1 的右图中绘制了  $\log(P(\phi_N/\phi_N^{RMS}))$  for  $N = 1, 2, 4, 8, 16, 32$ , 以及瑞利分布  $P \approx G \exp(-\alpha G^2)$ 。如果图像具有高斯统计量, 则局部梯度将呈瑞利分布。再次注意分布的缩放。

### 3.3. 功率谱

缩放也可以在功率谱层面上得到证明。如果集合是尺度不变的, 那么频谱应该是这样的形式

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

其中  $k$  以周期/度为单位,  $S$  是按方向平均的功率谱。

图 2 中的光谱以对数对数轴显示。它显示了两个焦距的重叠数据, 并显示光谱在空间频率上扩展了约 2.5 个十年。我们将参数确定为  $A = (6.47 \pm 0.13) \times 10^{-3} \text{deg.}^{(0.19)}$  和  $\eta = 0.19 \pm 0.01$ 。高达 60 个周期/度 (人类分辨率极限) 的积分功率谱给出了约 30% 的 RMS 对比度。

## 4. 本地过滤

视觉的早期阶段由对局部图像块作出反应的神经元组成。这些局部处理单元的统计数据是什么样的？我们使用图 3 左侧所示的滤波器对图像进行卷积，并在图右侧以半对数刻度绘制其输出的直方图。

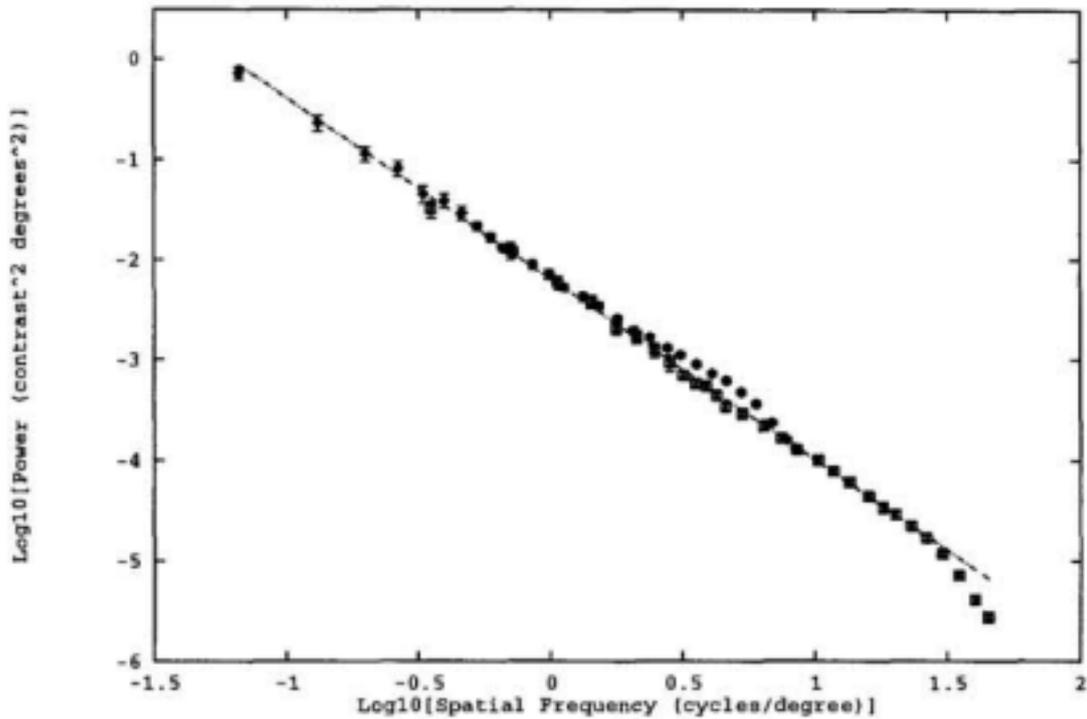


图 2: 自然场景对比度的功率谱（对数对数图）。

在近 40 年的概率中，该分布呈指数级增长。事实上，几乎任何不通过 DC 的局部线性滤波器都具有此属性，包括中心环绕接收场。信息论告诉我们，最好将具有高斯统计的信号发送到具有功率限制的信道上。因此，找到某种类型的滤波方法将我们发现的指数分布转换为高斯量是很有意义的。

事实证明，音乐具有一些类似的特性。图 4 左侧显示了 5 分钟的“蓝色多瑙河”的振幅直方图。它在 40 年的概率上几乎精确呈指数增长。我们可以猜测是什么导致了峰值和尾部超过高斯分布；这是动态。当播放安静的段落时，振幅仅接近零，并在峰值中产生过量。当音乐响亮时，波动很大，从而产生尾部。最重要的是，这些安静和响亮的段落的时间上连贯地延伸；因此，要消除峰值和尾部，我们可以简单地慢慢调整“音量旋钮”来使波动正常化。图像是由在空间上具有连贯结构的物体组成的，并且会发生类似的局部动

态。要消除它，我们需要某种增益控制。

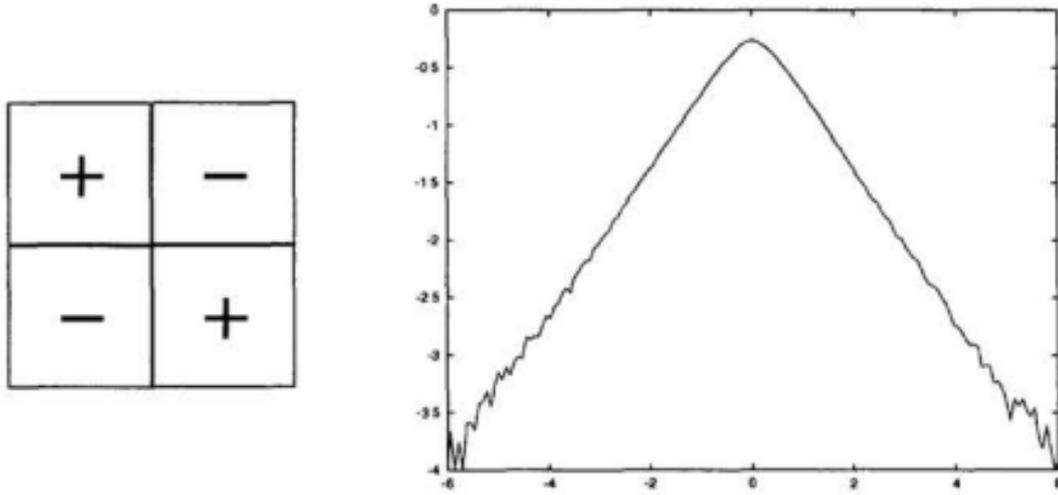


图 3: 左:  $2 \times 2$  局部滤波器。右: 过滤自然场景时其输出直方图的半对数图。

为此，我们将图像传递到局部过滤器，然后根据图像的局部标准差进行标准化（类似于声音段落的音量）：

$$\psi(\mathbf{x}) = \frac{\phi(\mathbf{x}) - \bar{\phi}(\mathbf{x})}{\sigma(\mathbf{x})}.$$

这里  $\bar{\phi}(\mathbf{x})$  是  $x$  周围  $N \times N$  个区域中图像对比度的平均值，而  $\sigma(\mathbf{x})$  是同一区域内的标准差（见图 4 右侧）。

我们发现，当值  $N = 5$ （负周围与正中心的比率）时， $\psi$  的直方图最接近高斯（参见图 5 的左侧）。

此外， $\psi$  的梯度直方图非常接近瑞利分布（参见图 5 的右侧）。这些都是高斯分布的特征。从功能上讲，这种“方差归一化”过程类似于视网膜和 LGN 中的对比度增益控制 (Benardete *et al.*, '92)。它的作用可能是“高斯化”图像统计数据吗？

## 5. 视网膜中的信息

根据测量的统计数据，我们可以确定一组光受体所传递的有关自然图像的信息量的上限。我们做出以下假设：

- 图像具有测量的功率谱，符合高斯分布。这为自然场景的熵设置了上限，从而为所

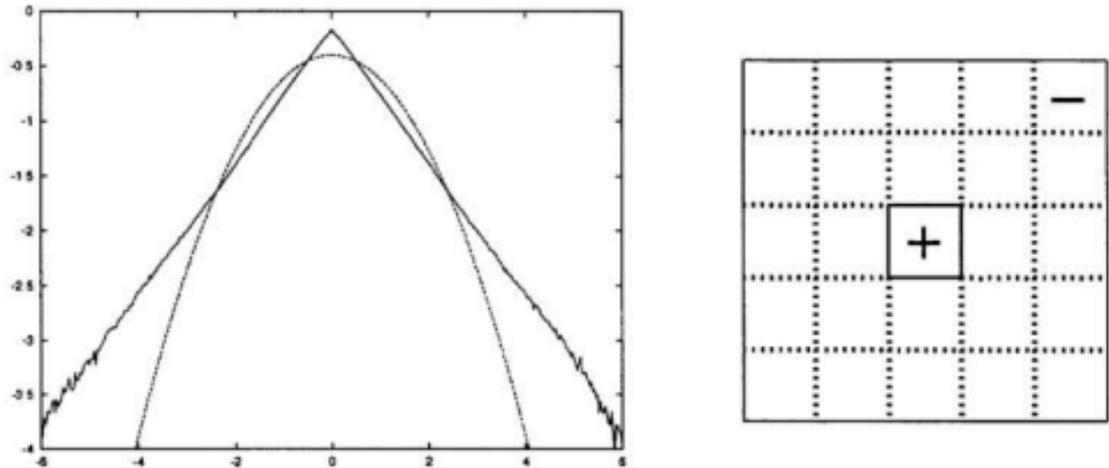


图 4: 左:《蓝色多瑙河》的半对数直方图, 与高斯分布进行比较(虚线)。右:  $5 \times 5$  中心环绕滤波区域。

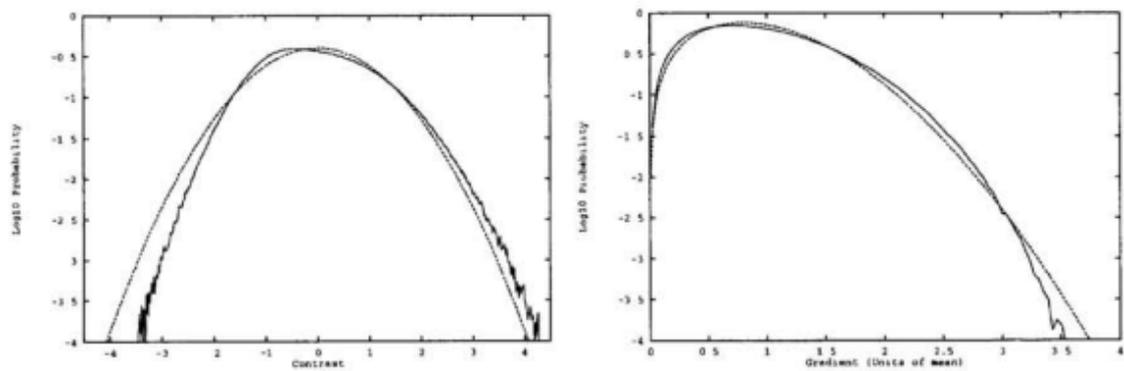


图 5: 左:  $\psi$  直方图的半对数图, 与高斯分布进行比较(虚线)。右:  $\psi$  梯度直方图的半对数图, 与瑞利分布进行比较(虚线)。

表示的信息设置了上限。

- 接收器采用衍射极限光学元件以六边形阵列采集图像。没有混叠。
- 噪声是加性的、高斯的、白色的、与图像无关的。

因此，第  $n^{\text{th}}$  个受体的输出由下式给出：

$$y_n = \int d^2x \phi(\mathbf{x}) M(\mathbf{x} - \mathbf{x}_n) + \eta_n,$$

其中  $X_n$  是接收器的位置， $M(x)$  是光学器件的点扩展函数， $\eta_n$  是噪声。对于衍射受限光学器件，

$$M(\mathbf{k}) \approx 1 - |\mathbf{k}|/k_c,$$

其中  $k_c$  是截止频率为 60 个周期/度。

在无限晶格的极限中，傅里叶分量是独立的，总信息是每个分量中信息的总和：

$$\mathcal{J} = \frac{A_c}{4\pi} \int_0^{k_c} dk k \log \left[ 1 + \frac{1}{A_c \sigma^2} |M(k)|^2 S(k) \right].$$

这里  $\mathcal{J}$  是每个受体的信息， $A_c$  是晶格中单元格的面积， $\sigma^2$  是噪声的方差。

我们取  $S(k) = A/k^{2-\eta}$ ，其中  $A$  和  $\eta$  取其测量值，并以受体中的信噪比表示噪声水平。在图 6 中，我们绘制了每个受体的信息与  $SNR$  的关系，以及该  $SNR$  下光感受器晶格的信息容量（每个受体），即

$$\mathcal{C} = \frac{1}{2} \log [1 + SNR].$$

即使在  $SNR = 1000$  的情况下，每个图像每个接收器所传递的信息也少于 2 位。从曲线之间的间隙可以看出，这种表示的冗余度相当高；至少浪费的信息容量与使用的信息容量一样多。

## 6. 结论

我们已经证明，森林中的图像具有尺度不变、高度非高斯的统计数据。非高斯直方图的缩放和功率谱的幂律形式证明了这一点。局部线性滤波产生的值具有相当指数的概率分布。为了“高斯化”，我们必须使用充当增益控制的非线性滤波器。这类似于哺乳动物

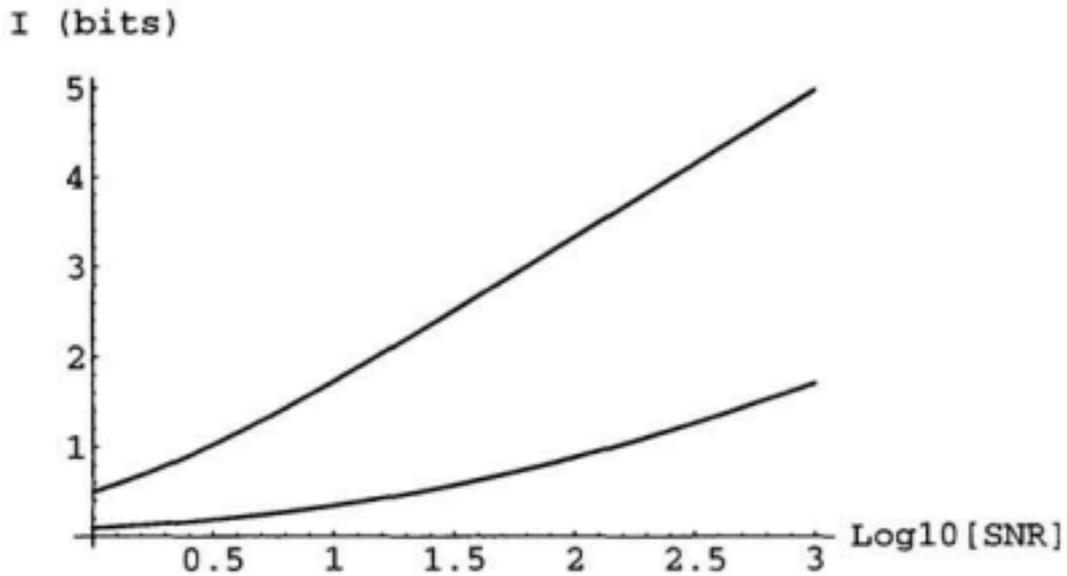


图 6: 每个受体每幅图像的信息量（以位为单位）与  $\log(\text{SNR})$  的关系（下线）。每个受体的信息容量（上线）。

视网膜中的对比度增益控制。最后，编码这些自然图像的受体阵列最多只能为每个受体传递几位信息，即使在高  $\text{SNR}$  下也是如此。在每秒 50 张的图像速率下，这对中央凹神经节细胞的信息要求不到每秒约 100 位。

## 附录

使用配备 9.5-123.5mm 变焦镜头的 Sony Mavica MVC-5500 静态摄像机收集快照。根据标准 CIE 公式  $Y = 0.59G + 0.30\bar{R} + 0.11B$  组合红、绿、蓝信号，以在每个像素处产生灰度值。量  $Y$  根据入射亮度进行校准，以产生图像强度  $I(x)$ 。图像被裁剪到中央  $256 \times 256$  区域。

数据集包括 45 张焦距为 15 毫米（视角为  $15^\circ$ ）的图像和 25 张焦距为 80 毫米（视角为  $3^\circ$ ）的图像。所有图像均为远距离物体，以避免对焦问题。选择图像的方法是将相机放置在路径上的随机点，然后旋转视野，直到画面中没有出现附近的物体。相机上下倾斜不超过  $100$  度，以避开天空和地面。森林环境（新泽西州春季的树林）主要由树木、岩石、山坡和溪流组成。

## 致谢

我们感谢 H. B. Barlow、B. Gianulis、A. J. Libchaber、M. Potters、R. R. de Ruyter van Stevenink 和 A. Schweitzer。本研究部分由 Fannie and John Hertz 基金会 (D.L.R.) 提供奖学金资助。

## 参考文献

J .J. Atick and N. Redlich. Towards a theory of early visual processing *Neural Computation*, 2:308, 1990.

E. A. Benardete, E. Kaplan, and B. W. Knight. Contrast gain control in the primate retina: P cells are not X-like, some M-cells are. *Vis. Neurosci.*, 8:483-486, 1992.

W. Bialek, D. L. Ruderman, and A. Zee. The optimal sampling of natural images: a design principle for the visual system?, in *Advances in Neural Information Processing systems*, 3, R. P. Lippman, J. E. Moody and D. S. Touretzky, eds., 1991.

G. J. Burton and I. R. Moorhead. Color and spatial structure in natural scenes. *Applied Optics*, 26:157-170, 1987.

D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *I. Opt. Soc. Am. A*, 4:2379, 1987.

J. H. van Hateren. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *I. Compo Physiol. A*, 171:157-170, 1992.

S. B. Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.*, 36c:910-912, 1981.

M. V. Srinivasan, S. B. Laughlin, and A. Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B*, 216:427-459, 1982.

**Affiliation:**Daniel L. Ruderman<sup>1</sup>

William Bialek

翻译：林绪虹<sup>2</sup>E-mail: [linxuhong@yahoo.com](mailto:linxuhong@yahoo.com)

---

Silkman Statistical Journal

published by the Funny Project of Silkman Press

May 2024, Volume 99, Issue 2

[doi:10.18637/ssj.2024.test](https://doi.org/10.18637/ssj.2024.test)

---

<http://cookwhy.com/>

<http://cookwhy.com>

*Submitted:* 2024-04-12

*Accepted:* 2024-05-12

---

<sup>1</sup>NEC Research Institute 4 Independence Way Princeton, N.J. 08540

<sup>2</sup>软件工程师, 数学史爱好者, 本文基于原作者 1993 年发表于 *Advances in neural information processing systems* 的原文翻译, 翻译时间 2024 年 5 月)