

自然图像的统计数据告诉我们有关视觉编码的什么信息

David J. Field*

翻译：林绪虹†

摘要

为什么哺乳动物的视觉系统会这样表示信息？如果我们假设视觉系统已经进化到可以应对自然环境，那么我们可能会认为视觉系统的编码特性与我们环境的统计结构有关。事实上，自然环境的图像没有随机统计数据。

之前已经讨论过自然图像的一阶统计量（例如像素值的分布）和二阶统计量（例如功率谱），它们与视觉编码有着重要的关系。高于二阶的统计量很难测量，但提供了有关图像的重要信息。例如，可以证明自然图像中的线和边是这些高阶统计量的函数。在本文中，这些高阶统计量将与哺乳动物视觉系统的编码特性一起讨论。有人提出，皮质“过滤器”的空间参数（例如简单和复杂细胞的带宽）与这些高阶统计量密切相关。特别是，将证明皮质复合细胞表现出的间隙非线性为早期视觉系统提供了了解这些统计数据所需的信息。

Keywords: 自然图像统计.

1. 介绍

自从 Hubel 和 Wiesel 于 1962 年首次绘制出哺乳动物视觉皮层细胞的图谱以来，它们的确切功能一直困扰着研究人员。关于这些细胞为何如此表现的理论多种多样，从边缘和条带检测到傅里叶分析。最近，有人提出，通过更好地了解视觉环境的统计数据，可以深入了解哺乳动物视觉通路细胞的行为 [2,3]。本文将继续采用这种方法。我们将考虑自然图像的一个特殊属性，以及该属性与皮层细胞行为的关系。特别是，我们希望强调一

种被描述为“复杂”的细胞类型 [1,4]。尽管有人认为这种细胞占初级视觉皮层的一半 [5]，但目前还没有广泛接受的关于其功能的理论。复杂细胞表现出一种特殊的空间非线性，这使它们有别于更线性和研究更广泛的“简单”细胞。

本文主要有两点。首先，自然图像表现出一种特殊形式的冗余，这种冗余在二阶统计的测量中不会出现。这种冗余导致局部特征，如边缘和线条。在频域中，边缘或条可以描述为相位的局部相关性：在边缘的位置，傅立叶分量处于同一相位。然而，当图像在不同位置包含大量边缘和线条时，全局傅立叶系数之间的相关性就会丢失。可能仍然有冗余，但现在是高阶统计的形式。它可能被粗略地描述为图像不同尺度上的冗余；不同频带中存在的非正式现象不是独立的。

在本文的第二部分，我们将描述一个视觉模型，该模型用“过滤器”来表示图像，这些过滤器具有与皮质感受野相同的许多特性。将要研究的最重要的特性是皮质复杂细胞所表现出的空间非线性。通过操纵该模型中的带宽，我们将尝试更好地了解皮质细胞为何具有其特定参数以及这些参数与上述统计数据的关系。

1.1. 自然图像的统计

自然界中随机选择的图像没有随机统计数据。此类图像不符合随机生成的像素的描述。自然图像的许多规律最好用“n 阶”统计数据来描述。例如，一阶统计是指每个像素值的概率分布（即不同点的亮度）。如果我们假设图像中的所有像素都代表来自相同分布的样本，那么就可以汇集单个图像中的像素以获得一阶统计量的估计值。

二阶统计量是指像素对之间的联系关系。例如，自相关函数是像素之间相关性的度量，它是像素之间距离的函数。功率谱也是二阶统计量的度量（它是自相关函数的傅里叶变换），因此提供了另一种度量。二阶和二阶统计量提供了有关图像的重要信息，并提供了有关视觉编码的重要线索。自然场景的二阶统计量表现出相当一致的行为 [3]（即，功率谱作为频率（f）的函数下降大约 $1/f^2$ ）。这些统计数据对于理解视觉编码至关重要。

然而，自然图像除了一阶和二阶统计量所提供的特性之外，还具有其他一致的特性。图 1 a 显示了自然场景（F.W. Campbell 教授）。图 1 b 显示了具有相同功率谱的图像。该图是通过随机化奇数的傅立叶系数的相位创建的。因此，这两幅图像具有相同的相关结构（二阶统计量）。但是，这两幅图像明显不同。

其中一个重要区别是，当相位随机化时，图像中没有任何清晰的边缘和线条。这些边缘和线条的存在无法用二阶统计来衡量。要理解这两幅图像之间的差异，我们必须考虑高阶统计。

1.2. 站点统计和本地能源

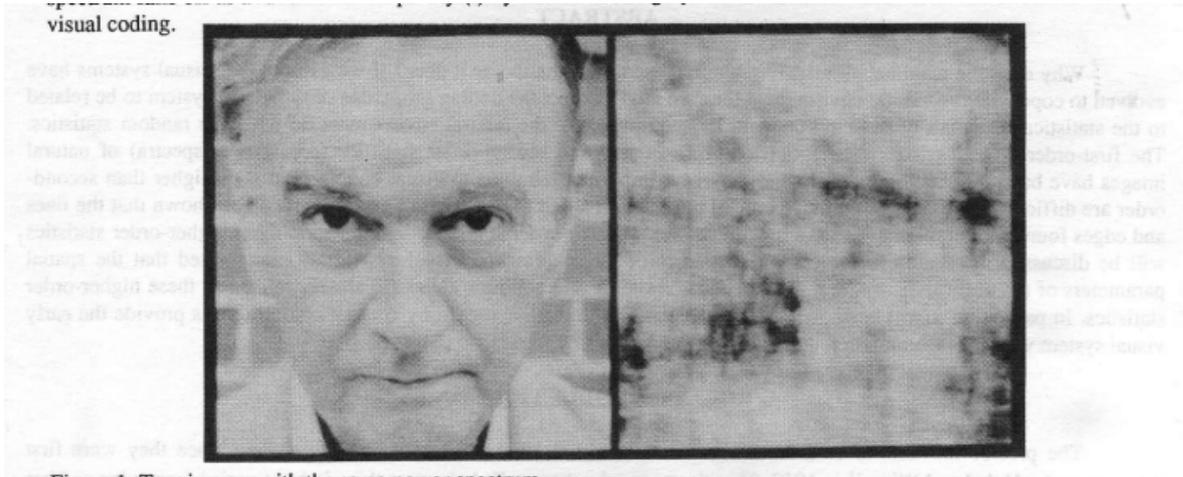


图 1: 两张具有相同功率规格的图片

本节将做出的一般假设是自然场景的统计数据是“平稳的”。这意味着，如果我们考虑一个非常大的图像集合，一个位置的统计数据与其他位置的统计数据没有区别。例如，如果垂直边缘倾向于只出现在所有图像的上半部分，那么统计数据将是平稳的。如果自然场景的统计数据是平稳的，那么可以证明所有实部和虚部系数（正弦和余弦）的振幅将不相关。如果图像的所有位置都具有相同的相位，那么任何特定频率的所有相位都是同样可能的。因此，可以通过对每个位置的所有振幅的乘积进行积分来确定任何两个频率（频率 n 和 m ）的相关。相关性与乘积之和成比例，因此：

$$\begin{aligned}
 r &= k \int_0^{2\pi} e^{inx} e^{imx} dx \\
 &= k \int_0^{2\pi} e^{i(n+m)x} dx \\
 &= k [e^{i(n+m)2\pi} \cdot 1] / i(n+m) \quad n \text{ 对于离散傅里叶变换，并且 } m \text{ 必须是整数，因此} \\
 e^{i(n+m)2\pi} &= 1 \quad \text{并且 因此:}
 \end{aligned} \tag{1}$$

$$r = 0$$

因此，如果所有相位都同样可能，那么所有频率的正弦和余弦将不相关。但是，两个频率不相关并不意味着系数是独立的。问题可以归结为两个向量。考虑 @fig-02 中所示的两个向量，它们表示两个频率的复振幅。让我们假设这两个向量的长度完全相关，而它们的方向 (0) 是随机的。

在这种情况下，以下预测将彼此不相关：

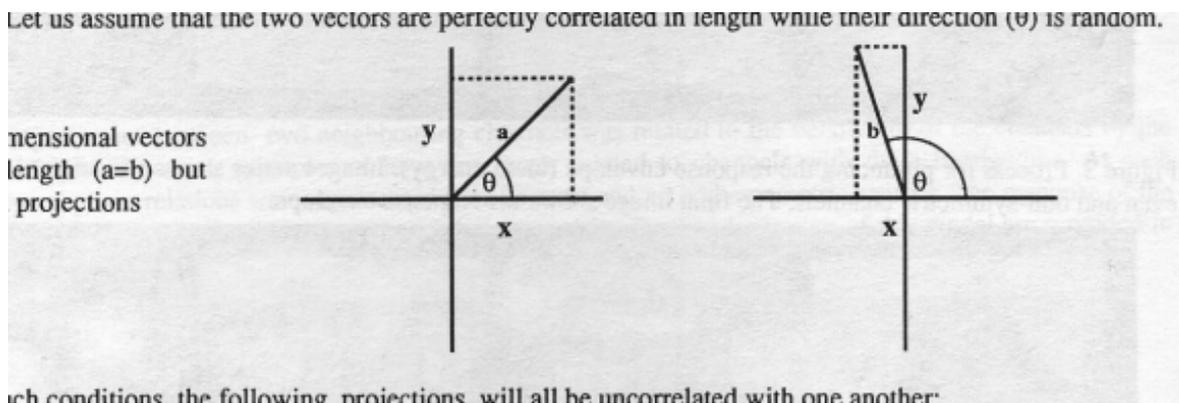


图 2: 长度相同 ($a=b$) 但投影不相关的二维向量

$$x_1 = a \cdot \cos(\theta_1) \quad y_1 = a \cdot \sin(\theta_1) \quad x_2 = b \cdot \cos(\theta_2) \quad y_2 = b \cdot \sin(\theta_2) \quad (2)$$

换句话说, 二阶统计量不提供任何关于二维向量冗余的信息。有人可能会说, 该信息是由四阶统计量提供的, 因为第四维可以通过了解其他三个维度来预测。但是, 可以使用非线性变换将该信息转换为二阶统计量。转换为极坐标系 $z = (x^2 + y^2)^{1/2}$ 是显而易见的选择。在这样的变换下, 人们会发现 z_1 和 z_2 是相关的。

类似的过程可应用于二维空间滤波器。[3,6,7] 考虑通过线性滤波器过滤的图像, 该滤波器对有限的频带和方向进行选择。图 3 中显示了一个示例。图 3 a 显示了原始图像。图 3 b 显示了将图像与线性滤波器进行卷积的结果, 该滤波器不会在频谱中产生相移 (偶对称滤波器)。图 3 c 是将图像与类似滤波器进行卷积的结果, 该滤波器将每个系数的相位偏移 90 度。这两个滤波器是正交的, 被描述为“正交”。与上面的矢量一样, 空间中每个点的矢量幅度可以计算为这两个正交矢量的矢量和。结果显示在图 3 b 中。图 4 显示了当滤波器涉及不同频带时相同的过程。要点是, 像图 3b、3c、4b 和 4c 这样的线性滤波器的输出将与原始图像的统计数据无关。然而, 图 3 d 和图 4 d (响应包络) 中显示的矢量幅度之间的相关性将提供有关原始图像冗余的重要信息。

1.3. 细胞

这种反应包络与“复杂”细胞行为之间的相似性此前已被注意到 [3,7]。与上述功能一样, 此类细胞对空间局部区域中有限的方向和频率带敏感。在此区域内, 此类细胞表现出空间非线性, 这似乎与上述功能非常相似。此外, Pollen 和 Ronner[8] 的研究表明, 皮层中相邻的简单细胞可能相差 90° , 这是正交对所必需的。因此, 可能能够产生具有上述许多特性的细胞, 其形式为相邻简单细胞的非线性和。但这并不能告诉我们为什么视觉系

统需要这样的细胞。由于此类细胞占初级视觉皮层的一半，因此它们似乎在视觉编码中发挥着重要作用。在以下章节中，我们将提出一种可能的解释。

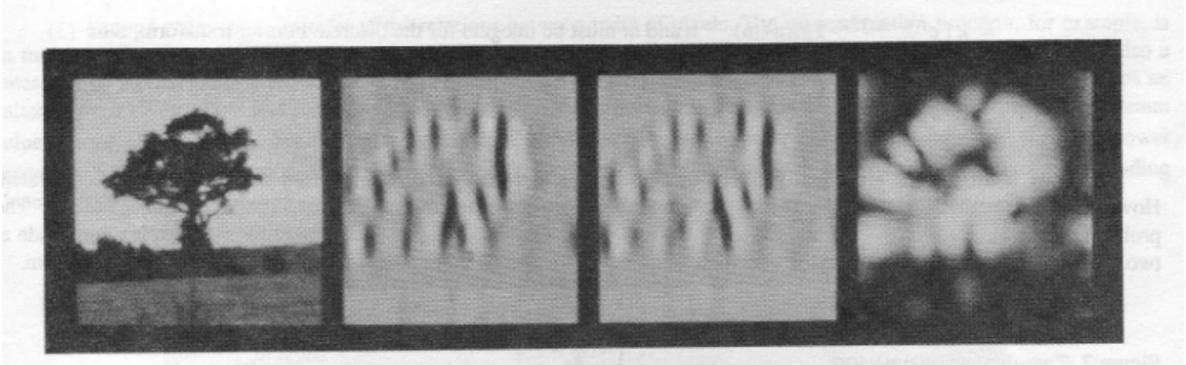


图 3: 生成响应包络（局部能量）的过程。图像系列显示原始图像，后面是偶数和奇数对称通道。最终图像显示响应包络。

even and odd-symmetric channels. The final image shows the response envelope.

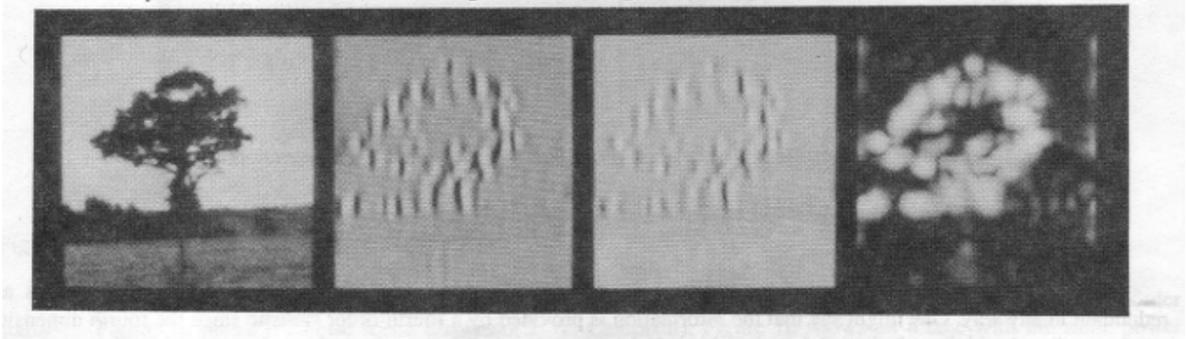


图 4: 与图 3 相同，但滤波器对较高频段有选择性。

2. 视觉模型

在下一节中，我们将比较模型神经元之间的相关性。为此，我们使用一个代码来尝试模拟视觉皮层的许多特性。然而，这绝不是视觉皮层的完整模型。相反，它是一种在图像中表示信息的方法，可以让我们探索视觉系统的各种参数，并间接测试图像的各种属性。该模型的一般形式已由许多先前的研究人员开发 [9,10]，并在作者的先前论文 [3] 中进行了更详细的讨论。主要特征概述如下。必须定义两个术语。传感器是指位于空间中的单个过滤器，代表单个假设细胞。术语通道是指调整到相同频率和方向的传感器的空间阵列。

- 1) 在频域中，相邻通道之间的距离由空间频率带宽决定，这也决定了单个传感器的宽度，从而决定了空间中相邻传感器之间的距离。

- 2) 在频域中，相邻方向通道之间的距离由齿状带宽决定，这也决定了各个传感器的长度，从而决定了传感器长度上的间距。
- 3) 对于给定代码的所有传感器，空间频率带宽以倍频程为单位是恒定的，方向带宽以度为单位也是恒定的。但是，可以自由更改每个代码的这些带宽。
- 4) 每个位置都有两个相互正交且相位关系为正交的传感器，局部能量（即假设的复杂单元）表示为这两个传感器的矢量和。
- 5) 传感器的总数与空间频率或方向带宽的特定选择无关（即由 65,536 个像素组成的图像由 65,536 个传感器表示）。

将使用的特定函数与 Gabor 表示法类似，但不同之处在于频率响应为对数正态。这种函数的优点在前面已经讨论过。对于图 7 所示的垂直方向传感器，其响应可以用函数来描述：

$$G(f, O) = \exp(-(\ln(f/f_0))^2/2\sigma^2) * \exp(-(O - O_0)^2/2\sigma^2)$$

不同传感器之间的采样距离与不同通道之间的采样距离相同。对于本文所需的所有通道，采样距离对应的距离为：

$$D = \sigma^*(2\pi)^{1/2}$$

两个相邻通道之间的采样距离与通道带宽的关系由相同的因子决定。图 5 提供了三个示例，说明该因子与三个带宽的通道之间的关系。在每个位置，都表示两个正交相位关系（即，一个偶数对称传感器和一个奇数对称传感器）。复杂单元（局部能量）的响应只是这两个单元的矢量和。早期的论文 [3] 对该模型进行了进一步讨论。

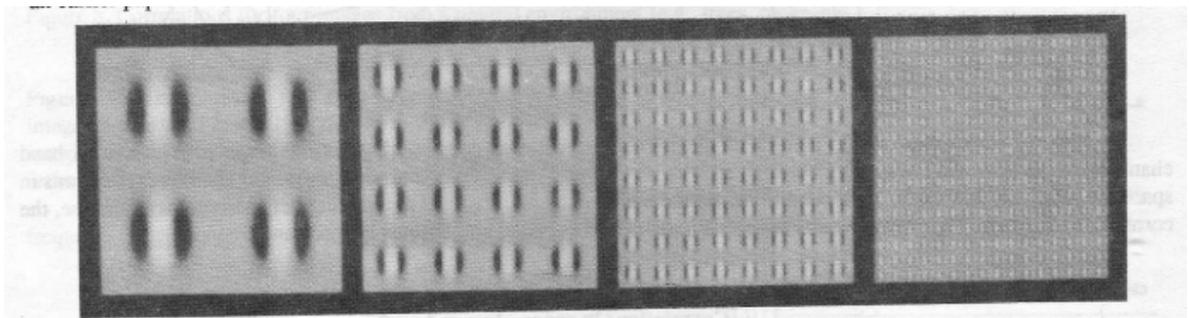


图 5: 频率带宽为 1 个八度的代码中垂直方向传感器（均匀对称）的示例。为了清晰起见，传感器之间的采样距离增加了大约 2 倍。

3. 估计相关性的方法

这种类型的模型使我们能够确定空间中相邻传感器之间的相关性以及不同通道（即不同频带）之间的相关性。对于每个带宽，频率相关性的估计涉及四个步骤：

- (1) 图像通过两个通道（公式 1）进行过滤，两个通道之间的距离如公式 2 中所述，并在图 3 和图 4 中进行了演示。
- (2) 复振幅按上述方法确定。
- (3) 确定整个中心区域的两个复振幅之间的相关性（图像的边缘将产生虚假相关性）。
- (4) 对所有相邻通道重复步骤 (1) - (3)。

空间相关性的估计采用类似的程序：

- (1) 对图像进行单通道滤波（公式 1）。
- (2) 复振幅按上述方法确定。
- (3) 确定了复杂通道与同一通道的位移版本之间的相关性。位移的程度是该特定通道中传感器大小的函数（即，距离对应于该通道中下一个传感器的距离 - 如图 5 所示）。
- (4) 对所有相邻通道重复步骤 (1) - (3)。

4. 结果

图 6 显示了图 1 中图像的相邻通道之间的相关性结果。当通道为窄带且相邻通道之间的间隔相对较小时，通道之间的相关性较大。随着通道带宽和通道之间的间隔增加，相关性会下降。这意味着对于图 1 中的图像，相关性跨越不同的频率。但是，相关性并不完整，频率越远，相关性越低。

图 7 显示了这些结果以及空间中相邻传感器之间的相关性：对于窄带信道，传感器和传感器之间的间距相对较大。在这种情况下，空间相关性相对较低似乎并不奇怪。随着带宽增加和传感器间距减小，相关性增加。因此，这些结果表明空间相关性和频率相关性之间存在权衡。

图 8 显示了另外四幅图像的结果。权衡似乎是一个相当一致的特征。然而，最有趣的是，权衡的区域大致在皮质细胞带宽范围内，覆盖 1-2 个八度音阶的范围。结果的一致性可能会让人质疑权衡是否仅仅是代码而不是图像的函数。空间和频率的相关性必须在一定程度上权衡，这当然是真的，但它们的交叉点确实是图像相关结构的函数。为了证明这一点，分析了图 1 b 中的图像。在该图像中，傅立叶系数的相位是故意随机化的。在该图像中，人们不会期望通道之间有任何高相关性。结果显示在图 9 中。在这些图中，可以看到空间相关性保持不变（因为功率谱不变），但频率相关性被消除（残余相关是由于两个函数在频率上在一定程度上重叠）。

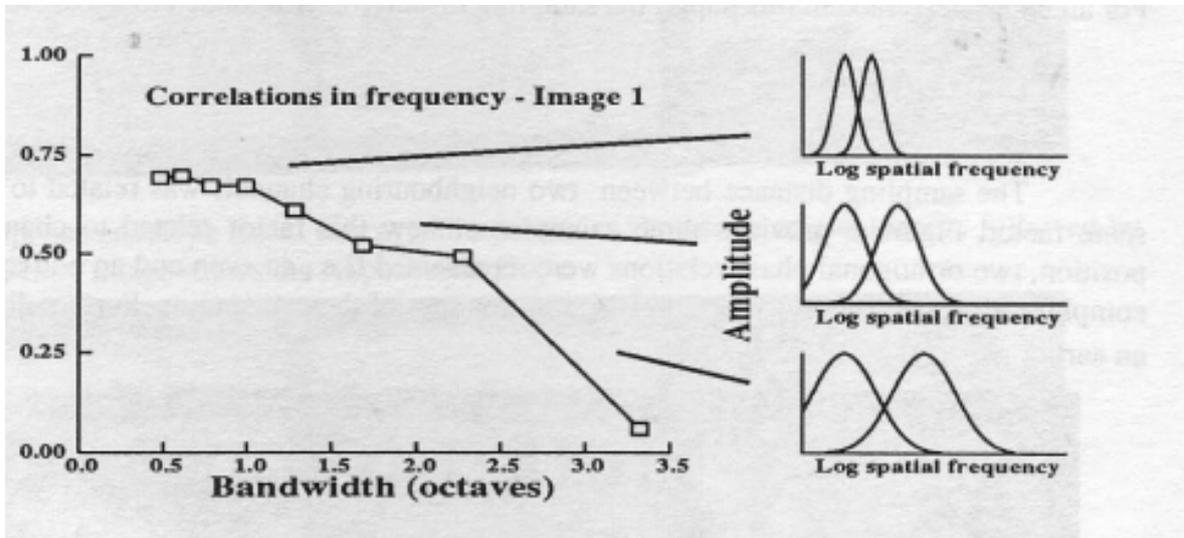


图 6: 相邻通道之间的相关性。

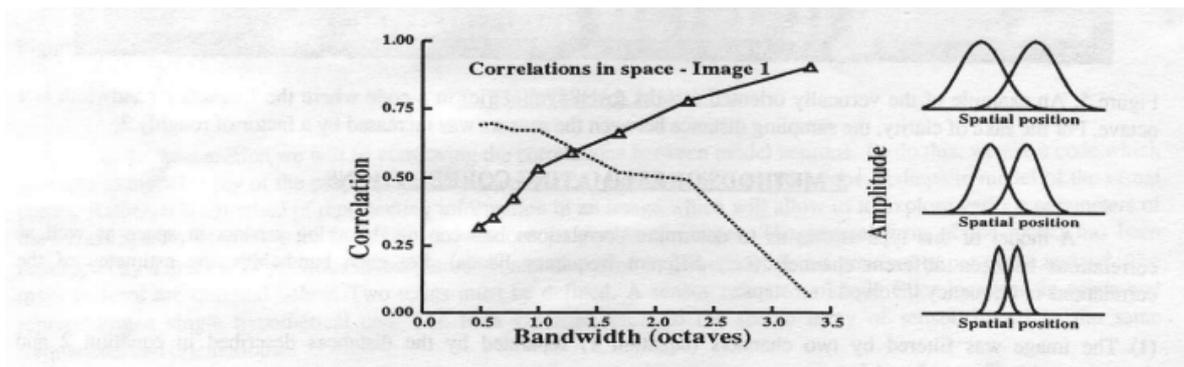


图 7: 此图显示了代码中相邻传感器之间的关系。在所有情况下, 采样距离与传感器尺寸成正比, 与通道尺寸成反比。

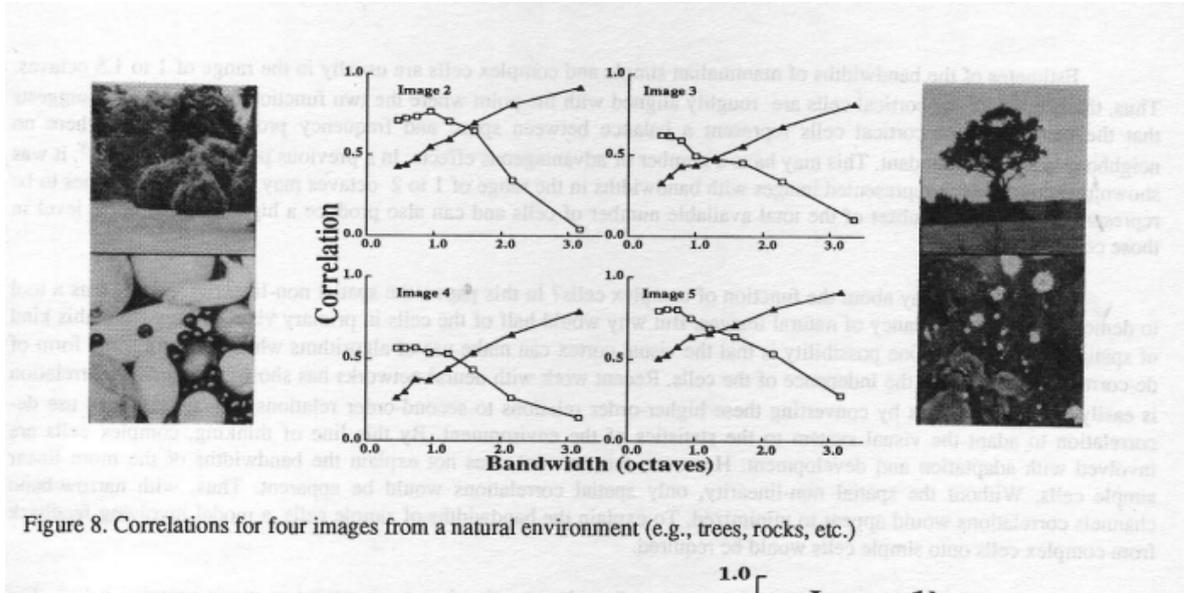


Figure 8. Correlations for four images from a natural environment (e.g., trees, rocks, etc.)

图 8: 来自自然环境的四幅图像的相关性 (例如费用、岩石等)

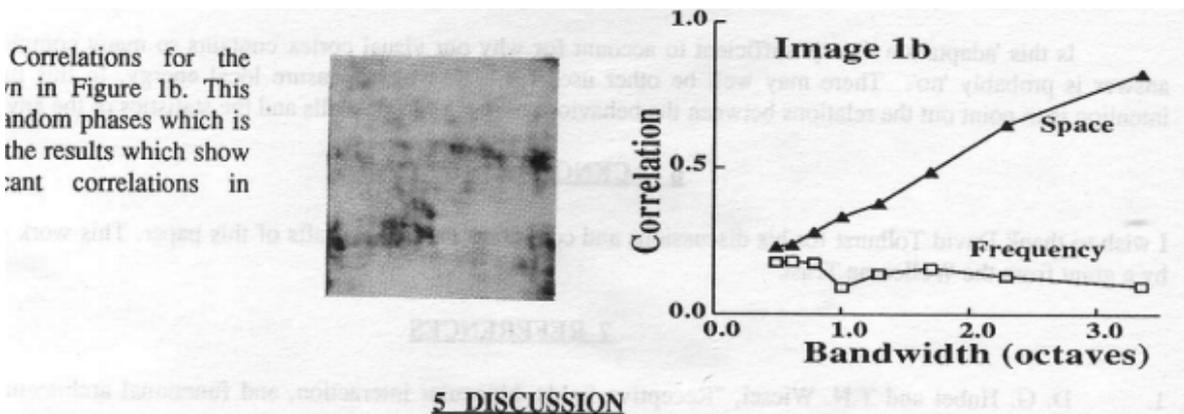


图 9: 图 1b 中所示图像的相关性。该图像具有随机相位, 这反映在结果中, 表明频率没有明显的相关性。

5. 讨论

如前所述，本文分为两部分。第一部分讨论了两种类型的冗余。第一种冗余由空间相关性表示，并以二阶统计量（如反相关函数和功率谱）来衡量。第二种冗余在自然场景的讨论中很少考虑。这种冗余存在于图像的不同尺度（即不同的频带）之间。这种冗余不是二阶的，因此不会出现在相关中。然而，结果表明，一种特定的非线性变换能够将这些高阶统计量转换为传感器之间的二阶相关，然后可以用相关来衡量。描述了一个模型，该模型允许用线性阵列的理论单元或“传感器”来表示图像，这些单元或“传感器”对不同频率和方向的局部波段具有选择性。从这些线性“传感器”中，我们展示了如何获得局部能量的估计值（该测量方法与圆锥形“复杂”细胞的行为有有趣的相似之处）。

在第二部分中，我们研究了视觉编码模型与这两种冗余形式之间的关系。结果表明，这两种冗余形式都存在于自然场景中。不同的代码似乎会促进不同数量的这两种类型的冗余。如果用窄带宽函数表示自然场景，则相邻通道之间的冗余度会很高。如果用宽带宽函数表示自然场景，则空间中的相关性会很高。任何试图最大化代码单元（即视觉系统中的传感器或细胞）独立性的代码都必须考虑这两种类型的冗余。

哺乳动物简单和复杂细胞的带宽估计通常在 1 到 1.5 个倍频程范围内。因此，皮层细胞的带宽大致与两种功能的交叉点一致 [5,11]。这表明皮质细胞的带宽代表了空间和频率之间的平衡，从而产生了一种没有邻居高冗余的状态。这可能带来许多有利的影响。在作者之前的一篇论文 [3] 中，作者表明，表示带宽在 1 到 2 个倍频程范围内的图像的代码可能允许仅使用总可用细胞数的子集来重新表示自然场景，并且还可以在這些细胞中产生高信噪比水平。

这说明了复杂细胞的功能吗？在本文中，空间非线性被用作展示自然图像冗余的工具。但是，为什么初级视觉皮层中有一半的细胞会表现出这种空间非线性呢？一种可能性是，视觉系统可以利用执行某种形式的去相关的算法来增加细胞的独立性。最近对神经网络的研究表明，这种去相关是很容易实现的 [12]。因此，通过这些高阶关系转换为二阶关系，视觉细胞可以使用去相关来使视觉系统适应环境的统计数据。按照这种思路，复杂细胞参与了适应和发展。然而，这种解释并不能解释更线性的简单细胞的带宽。如果没有空间非线性，则只会出现空间相关性。因此，在窄带通道中，相关性似乎被最小化了。为了解释简单细胞的带宽，需要一个涉及从复杂细胞到简单细胞的反馈的模型。

这种“适应理论”是否足以解释为什么我们的视觉皮层包含如此多的复杂细胞？答案可能是肯定的。测量局部能量的细胞可能还有其他用途。本文的唯一目的是指出这些皮层细胞的行为与环境统计数据之间的关系。

6. 致谢

我要感谢 David Tolhurst 对本文早期草稿的讨论和评论。这项工作得到了 Wellcome Trust 的资助。

7. 参考文献

[1] D. G. Hubel and T.N. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," *J. Physiol.* 160, 106-154 (1962).

[2] S.B. Laughlin, "Matching coding scenes to enhance efficiency" in *Physical and Biological processing of Images*, O.J. Braddick and A.C. Sleight, eds. (Springer, Berlin, 1983), pp.42-72.

[3] D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4 2379 -2394 (1987).

[4] A.F. Dean and D.J. Tolhurst. On the distinctness of simple and complex cells in the visual cortex of the cat. *J. Physiol., London* 344,305-325 (1983).

[5] D.J. Tolhurst, and I.D. Thompson. On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proc. Roy. Soc. London B* 213, 183-189 (1982);

[6] E.H. Adelson, and J.B. Bergen, Spatio-temporal energy models for the perception of motion. *J. Opt. Soc Am. A* 2 2U-299. (1985).

[7] M.C. Morrone and D.C. Burr, "Feature Detection in human vision: A phase-dependent energy model". *Proc. Roy. Soc London Ser B* 235,221-245 (1988), [8] D.A. Pollen and S.F. Ronner, "Phase relationships between adjacent simple cells in the cat," *Science* 212, 1409-1411 (1981).

[9] J.G. Dairgman. "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters" *J. Opt Soc. Am. A* 2, 1160-1169 (1985).

[10] A.B. Watson. Efficiency of a model human image code *J. Opt. Soc. Am. A* 4 -2'378 (1987).

[11] R.L. DeValois, D.G. Albrecht, and L.G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Research*, 22 ,545-559 (1982).

[12] Barlow, H.B. and Foldiak, P.2., *Adaptation and de-correlation in the conex,*"

To appear in the computing Neuron (Miall, Durbin and Mitchison, eds.) Addison Wesley.

Affiliation:

David J. Field¹

翻译: 林绪虹²

E-mail: linxuhong@yahoo.com

Silkman Statistical Journal
published by the Funny Project of Silkman Press

<http://cookwhy.com/>
<http://cookwhy.com>

May 2024, Volume 99, Issue 2
[doi:10.18637/ssj.2024.test](https://doi.org/10.18637/ssj.2024.test)

Submitted: 2024-04-12
Accepted: 2024-05-12

¹Physiological Laboratory, University of Cambridge Cambridge England. CB2 3EG

²软件工程师, 数学史爱好者, 本文基于原作者 1989 年发表于 SPIE Vol. Io77 Human Vision, Visual Processing, and Digital Display (1989) 的原文翻译, 翻译时间 2024 年 5 月)