

# 视皮层中的视差计算

孔庆群<sup>1</sup> 明雁声<sup>1</sup> 胡占义<sup>1</sup>

**摘要** 对应问题是从图像恢复物体三维几何形状的核心问题,也是计算机视觉领域长久以来没有很好解决的问题. 尽管文献中有大量对应点匹配算法报道,但这些主体基于数学和工程的方法,在鲁棒性方面离人类视觉还有很大的差距. 所以,建立基于生理机理的立体感知模型,可望有效提高匹配的鲁棒性. 然而,要建立这种基于生理机制的计算模型,需要对生物立体感知的机理有所了解. 基于此,本文对近年来关于生物立体感知的进展和主要研究成果进行了总结,包括初级区域 V1, V2, V3 区,以及背部通道的 MT, MST, IPS 和腹部通道的 V4, IT 区. 除了对这些区域关于神经生理的成果进行介绍外,我们还简单介绍了著名的视差能量模型及其推广模型. 本文对研究计算机视觉的人员有一定的参考价值.

**关键词** 立体感知, 初级区域, 背部通道, 腹部通道, 视差能量模型

**DOI** 10.3724/SP.J.1004.2011.00645

## Disparity Computation in the Visual Cortices

KONG Qing-Qun<sup>1</sup> MING Yan-Sheng<sup>1</sup> HU Zhan-Yi<sup>1</sup>

**Abstract** The correspondence problem is essential for recovering the 3D shape of object from its images, however, this problem has not been solved satisfactorily in the computer vision field. Although numerous mathematical and engineering methods were reported in the literature, they were unable to perform as the human stereo vision does in terms of robustness. Therefore, a physiology based computational model is solicited, and this in turn requires some knowledge about biological stereo vision. To this end, we reviewed the relevant progresses in physiology about stereo depth perception, including the low level areas V1, V2, V3, dorsal areas MT, MST, IPS, and ventral areas V4, IT. In addition to these physiological studies, we also introduced the well known disparity energy model and its generalized model. This article could be of reference to computer vision researchers.

**Key words** Stereo depth perception, low level visual areas, dorsal areas, ventral areas, disparity energy model

立体视觉 (Stereopsis) 是人类视觉系统的一个重要组成部分,其功能是通过视差感知三维信息,包括物体的空间位置和三维形状. 最近 20 年来,神经生理学家们在哺乳动物视皮层做了大量实验,力图揭示立体视觉的神经机理. 很多重要的成果,比如视差能量模型的提出,很大程度上丰富了人们对于视皮层立体视觉机制的认识. 虽然目前人们还未彻底掌握立体视觉的奥秘,但现有的发现已足以勾画出立体视觉这样一个复杂神经网络的大致轮廓,其视差计算的多样性、鲁棒性和精密性让人叹为观止. 本节将介绍一些背景知识,随后的章节将介绍视皮层相关区域的主要研究成果. 最后,本文将列出几个在这个方面做出系统性工作的实验室网址,以方便读者进一步了解本领域的进展.

首先,什么是视差? 三维空间中的一个点,在左右眼的视网膜上各成一个像,这两个像在视网膜上

的位置差异就是视差. 注视点在左右眼的图像均位于视网膜的中央凹处,视差是零. 如图 1 所示,  $F$  是注视点,对应的视差为零,其他点的视差可以通过注视点计算出来,比如  $P$  点的视差是  $\omega - \theta$ . 上面说的视差又称为绝对视差. 从图 1 可以看出,绝对视差和注视点的深度有关. 如果发生眼动,绝对视差就会变化. 而相对视差(也就是两个绝对视差的差)只反映空间中相邻物体的位置关系,和注视点的位置无关. 心理学实验证明,相对视差和人的深度感知有更加密切的联系.

除了视差之外,生物立体视觉系统还可以利用其他一些信息来感知深度,比如遮挡、阴影、纹理梯度等. Julesz 设计了随机点立体图 (Random dot stereograms, RDS),如图 2 所示,左右两图方框内是相同的,其他部分均是随机点. 随机点立体图除了视差之外不含任何与深度有关的信息. Julesz 的实验发现人可以轻易地从 RDS 中感知三维形状,于是证明了视差信号自身就足以引起三维感知. Julesz 随后的实验说明了视差是最重要的立体视觉信息来源. 生物立体视觉系统依赖于视差的可能原因是:视差信息是普遍存在的,而其他单眼信息只是存在于特定的场景中. 食肉动物,比如老鹰,就常常依赖视差信息捕猎. 它们的猎物,比如兔子,进化出有保

收稿日期 2010-09-03 录用日期 2011-02-25  
Manuscript received September 3, 2010; accepted February 25, 2011

国家自然科学基金 (60820012, 60973005) 资助  
Supported by National Natural Science Foundation of China (60820012, 60973005)

1. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190  
1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

护作用的毛皮纹理和颜色, 可以和背景混在一起, 就如图 2 中的方块一样. 这时候食肉动物只能通过计算视差把猎物从背景中分割出来. 为了发挥立体视觉的功能, 食肉动物的两只眼睛通常长在脸的前方, 有着很大的双目可见区域.

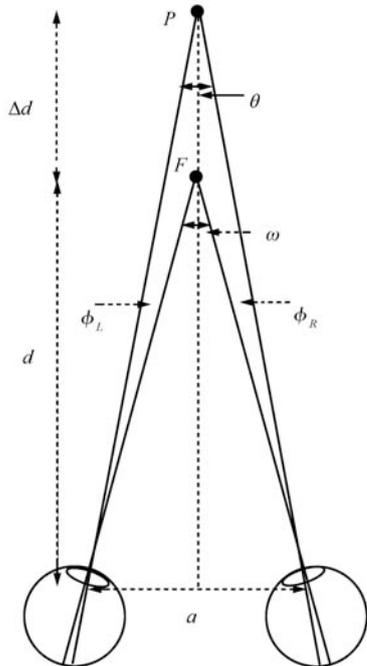


图 1 视差的定义<sup>[1]</sup> ( $F$  为注视点, 视差为零.  $P$  点的绝对视差为  $\omega - \theta$ )

Fig. 1 The definition of disparity ( $F$  is the fixation point with zero disparity. The absolute disparity of  $P$  is  $\omega - \theta$ .)

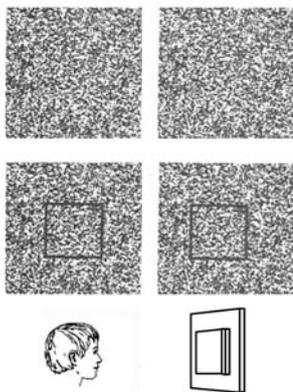


图 2 随机点立体图

Fig. 2 The random dot stereogram

其次, 立体视觉是由怎样的神经机制实现的? 现在还不能确切回答这个问题. 但可以肯定的是, 立体视觉的各项功能是靠视皮层多个区域大量神经元的合作来完成的. 生理学实验发现对视差敏感的神经元遍布于视皮层的各个区域, 并非限制在某个

特定的区域之中. 图 3 是视皮层中视觉通道的示意图, 腹部通道从 V1 经 V3, V4 等区域到达 IT 区 (Inferior temporal cortex), 背部通道从 V1 经 V2, MTA (Middle temporal area), MST (Middle superior temporal area) 等区域到 IPS (Intraparietal sulcus) 区. 视差处理的广泛性也间接反映了视差计算的艰巨性. 当然还有一种可能: 视皮层的一两个核心区域完成了立体视觉的主要工作, 其他区域仅仅是负责为核心区域传输来自视网膜的信息, 或者使用核心区域的输出信号. 然而, 一些生理学发现表明这种方式的可能性不大. 一个原因是视差处理的递进性. 左右眼的信息在 V1 区就得到了融合, 其中的复杂细胞被认为是最早编码视差的单元, 但是立体视觉计算直到视觉通道末端区域才最终完成. 并且, 中间区域神经元的视差选择性显示出从低级到高级、从简单到复杂的递进性特点. 如果立体视觉真的存在所谓“核心区域”的话, 那么神经元的视差选择特性在视觉通道中的某个阶段会发生突然变化, 这与实验发现的递进式发展规律相左. 视皮层视差处理还具有多样性. 不同区域的视差选择特性与该区域复杂的视觉任务有关. 三维信息可以指导人完成不同的任务, 包括眼动、识别、用手抓取物体等. 一般来说腹部通道被认为是负责形状识别, 背部通道负责指导运动. 实验发现这两个通道的高级区域的视差选择特性确实有助于完成各自的任务, 可以推测, 两个通道内的视差处理策略也是不同的. 考虑到视差处理的广泛性、递进性和多样性, 我们认为生物立体视觉是由一个贯穿背部和腹部通道的神经网络实现的.

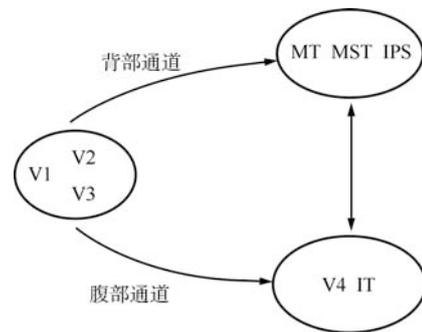


图 3 视觉通道示意图

Fig. 3 The visual pathway

## 1 低级区域

### 1.1 V1 区

V1 区 (又称初级视皮层或条纹皮层) 是目前为止研究最为透彻的区域, 既有生理学研究成果, 又有相应的数学解释模型.

### 1.1.1 生理学发现

V1 区是双目信号最初融合的地方, 相对高级区域, 功能较为简单, 有最早的视差编码单元, 但还不能很好地解释人的深度感知。

自 Barlow 等在猫初级视皮层中发现视差敏感的神经元<sup>[2]</sup> 后, 神经生理学家在一些高等动物的初级视皮层中同样发现了对视差敏感的神经元. 根据感受野的特征, Hubel 等将 V1 区的神经元划分为简单细胞和复杂细胞两类<sup>[3-4]</sup>, 简单细胞的单目感受野轮廓呈现对光的兴奋区域和抑制区域交替排列的特性. 数学上, Gabor 函数被认为是描述简单细胞单目感受野响应轮廓的最优模型<sup>[5]</sup>, 如图 4 所示, 白色代表兴奋区域, 黑色代表抑制区域. 而复杂细胞的感受野, 对光的兴奋区域和抑制区域重叠在一起, 并且对刺激物的位置具有较宽的选择性, 即对刺激物的对比度和位置都不敏感, 因而复杂细胞更适合于编码视差, 这点在后面的视差能量模型中也得到了验证.

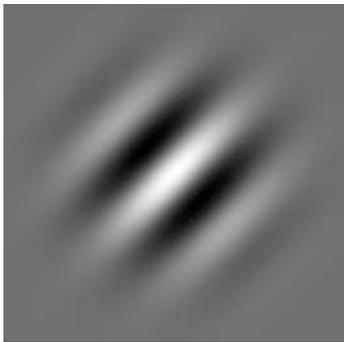


图 4 Gabor 函数近似简单细胞感受野

Fig. 4 The Gabor function, approximating simple cell RF

需要指出的是视差编码并不等同于深度感知. 例如视差信号可以仅仅用来指导眼动, 或者用于场景分割等. 大量的心理物理学观察表明相对视差在深度感知中起重要作用, 然而 V1 区编码的是绝对视差<sup>[6]</sup>. 对于感受野内刺激物相同而感知深度不同的窗口信号 (图 5), V1 区神经元的发放是相同的<sup>[7]</sup>; 反相关随机点图 (RDS 中一幅图像的对比度取反, 后面简称 aRDS) 不能引起深度感知, 但 V1 区神经元对 aRDS 的视差仍然有选择性<sup>[8]</sup>, 这些结果说明 V1 区神经元的性质类似于局部滤波器, 不同于深度感知单元.

V1 区对视差的编码还具有低通性, Nienborg 等<sup>[9-10]</sup> 利用视差呈现波浪形起伏变化的皱型随机点图测试了 V1 区神经元的发放率变化, 发现随着视差变化的空间频率或时间频率的增加, 神经元的相对调谐率 (幅值/均值) 都会下降, 也就是说神经元会逐渐感受不到高于一定频率的视差变化. V1 区神

经元的这个特点与人类的心理学感知是相符的, 因为当视差变化的空间频率或时间频率很大时, 人类是无法感知到视差变化的.

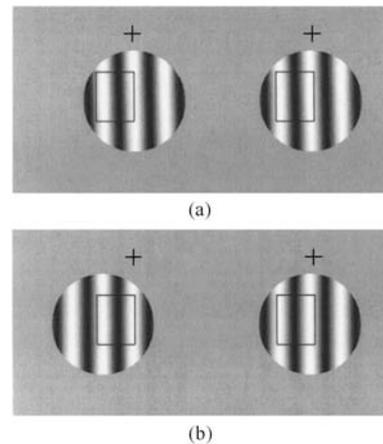


图 5 窗口信号<sup>[7]</sup> (方框代表感受野. (a) 中刺激物的对应部分刚好落在感受野内; (b) 中感受野内的刺激物与 (a) 中相同, 但不是对应部分)

Fig. 5 Windowed grating stimuli (The rectangle represents the location of the receptive field. In (a), the corresponding parts of the stimuli overlies the receptive field; but in (b), the noncorresponding parts within the receptive field are identical to (a).)

V1 区不仅是编码绝对视差 (0 阶视差) 的最早单元, 而且可能是高级区域编码三维方向的信息来源. Sanada 和 Ohzawa<sup>[11]</sup> 在猫视皮层 17 和 18 区发现 30% 的双目神经元的倾向平面相对于额平行平面 (Frontoparallel plane) 有倾斜, 其倾斜程度足以表示典型视觉场景中平面的倾斜角, 所以他们认为这部分神经元可能用于编码斜面的方向. 同时, 他们发现这部分神经元的左右眼感受野存在频率差异, 这也为视差梯度计算提供了一种可能的机制.

V1 区是双目信号最早融合的地方, 作为立体视觉的开端区域, V1 区对视差的处理比较简单, 但是它为后面的区域提供输入, 所以 V1 区仍然是立体视觉通路中重要的组成部分.

### 1.1.2 视差能量模型及其推广模型

尽管 1968 年就在猫初级视皮层区发现了对视差敏感的神经元<sup>[2]</sup>, 但直到 1990 年, Ohzawa 等才提出能够定量描述复杂细胞视差检测机制的视差能量模型<sup>[12]</sup>, 结构如图 6 所示. 其中,  $S$  代表简单细胞, 而  $C_X$  代表复杂细胞. 图 6 清晰地表示出, 视差能量模型中有一个复杂细胞, 它接受 4 个简单细胞的输入. 其中, 4 个简单细胞分为两对, 每对内两个细胞感受野相位相差  $180^\circ$ , 两对中的细胞感受野相位则相差  $90^\circ$ . 视差能量模型能够很好地反映视差检测

机制, 当两眼感受野的差异与刺激物的视差相同时, 响应达到最大, 如图 7 所示. 然而视差能量模型并不能完全解释 V1 区的生理数据. 比如视差能量模型不能很好地解释某些复杂细胞为什么对 aRDS 视差选择性比 RDS 弱的原因, 另外一点就是该模型不能反映神经元对环境的适应性. 比如, 尽管左右两眼存在相位差的刺激物在自然界中是不会存在的, 然而视差能量模型却仍然可以编码其视差. 针对这些问题, Haefner 等在视差能量模型的基础上做了一些改进, 提出了广义的视差能量模型<sup>[13]</sup>.

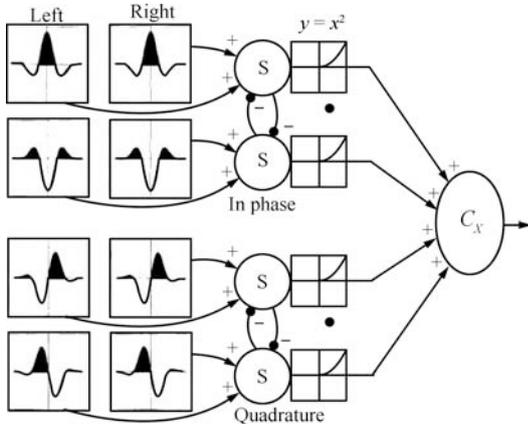


图 6 视差能量模型的神经网络结构图<sup>[5]</sup>  
Fig.6 Schematic representation of the disparity energy model

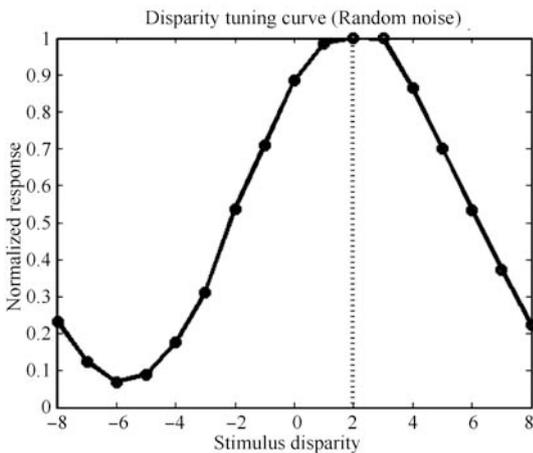


图 7 视差调谐曲线<sup>[1]</sup>  
Fig.7 The disparity tuning curve

在阐述此模型之前, 先简单介绍一下位置差和相位差这两个概念. 双目神经元对信号视差的检测是通过左右眼感受野之间的差异实现的, 根据两眼感受野的轮廓是否相同, 两者之间的差异分为位置差和相位差, 数学上表示为

$$f_L(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \cos(2\pi fx + \phi)$$

$$f_R(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_R^2}{2\sigma^2}\right) \cos(2\pi fx_R + \phi + \Delta\phi)$$

其中

$$x_R = x - \Delta x$$

$f_L(x)$  和  $f_R(x)$  分别是简单细胞的左右眼感受野.  $\Delta x$  是位置差,  $\Delta\phi$  是相位差. 有相位差的神经元对刺激物的相位差敏感. 但是我们左右两眼看到的图像不可能存在相位差, 只能是位置差. 关于两眼感受野为什么会存在相位差一种可能的解释是为了排除假目标, 从而解决匹配问题<sup>[14]</sup>.

下面介绍广义的视差能量模型, 结构框图如图 8 所示, 图中 C1, C2 代表两个复杂细胞. 此模型提出的背景是:

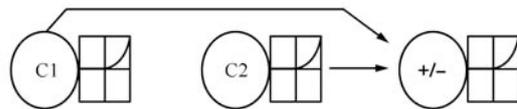


图 8 广义视差能量模型结构框图<sup>[13]</sup>  
Fig.8 The generalized energy model

视差能量模型的任务就是编码视差, 不管左右眼的图像对是否是自然界真实存在的. 举个例子加以说明, 对于两个频率的复合正弦光栅信号 ( $f_2/f_1 = 2$ ), 给定位置差  $\Delta x$ , 左右两眼相位差与频率成正比, 即  $\Delta\phi_i = \Delta x \cdot f_i$ <sup>[15]</sup>, 所以  $\Delta\phi_2/\Delta\phi_1 = 2$ , 如图 9 中的黑实线 (被称为真实视差线). 视差能量模型对于复合光栅信号的响应是单频信号响应的和, 所以引起神经元最大发放率的刺激物和引起最小发放率的刺激物在各个频率上的相位差均为  $180^\circ$ , 于是, 最大值点与最小值点在图 9 中连线 (如灰色斜线所示) 的斜率为 1. 然而 V1 区的实验数据表明, 只有 42% 的神经元响应可以用如上的视差能量模型近似. 还有将近一半神经元的响应的最大最小值都集中在真实视差线附近, 如图 10 所示. 这说明神经元集中编码真实刺激物的视差. Haefner 等针对这种神经元响应, 提出了一种广义的视差能量模型. 该模型中, 两个传统的视差能量模型经过非线性化后相减, 再经过非线性化最终输出. 神经元响应最大值的位置由 C1 决定, 而最小值的位置则由 C2 决定, 这样保证编码的是真实刺激物的视差. 该模型不仅能够解释神经元对环境的适应性, 同时还能解释神经元对 aRDS 视差较小的选择性, 如图 10 所示, 虚线代表反相关刺激物视差线, 与真实视差线的相位相差  $180^\circ$ . 因为神经元响应最大最小值都集中在真实视差线附近, 所以对 aRDS 的响应弱. 另外, 该模型还可用来解释纹状体外皮层 (Extrastriate cortex (V1 区后面的区域)) 中视差调谐曲线奇对称占优势的现象<sup>[16]</sup>.

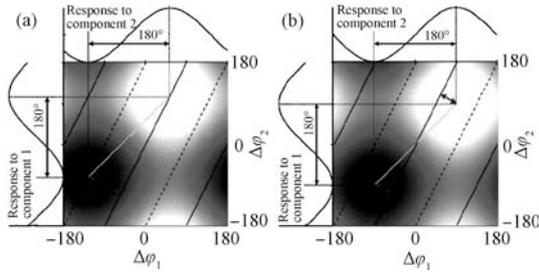


图 9 纯位置差 (a) 或纯相位差 (b) 的能量模型响应图<sup>[13]</sup>  
 Fig. 9 The response of the disparity energy model with a pure position shift (a) or phase shift (b)

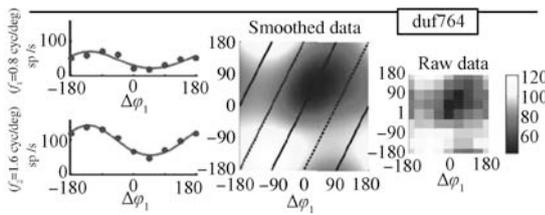


图 10 V1 某个神经元发放示意图 (其响应都在真实视差线附近<sup>[13]</sup>)  
 Fig. 10 A sample neuron's response in V1 (The neuron concentrates its response near the physical disparity line)

尽管广义的视差能量模型比传统的视差能量模型能解释更多的神经生理实验数据, 然而视差能量模型是构成广义视差能量模型的基础, 在视差能量模型提出的 20 多年里, 其正确性受到不同实验室生理学实验数据的支持, 到目前为止仍是解释 V1 区复杂细胞视差选择特性的主要模型. 目前 V1 区之外的有视差选择性的神经元还没有形成领域内比较认可的数学模型.

### 1.2 V2 区

V2 区和 V3 区 (见第 1.3 节) 同样是视觉通道的低级区域, 它们接收 V1 区的输入, 并且有大量向背部和腹部通道的输出. 虽然紧邻 V1 区, 但是 V2 区的视差选择性和 V1 区有以下几个方面的重要区别.

首先, V2 区中的一部分神经元开始对相对视差有响应<sup>[17]</sup>. 一个理想的对相对视差有选择性的神经元对感受野中心和边缘处视差的响应应该如图 11 所示. 横轴是感受野中心处的视差, 纵轴是感受野边缘参与的视差. 当感受野边缘处的视差改变时, 神经元的视差调谐曲线应该发生平移, 并且平移的大小应该和边缘视差变化的大小相等. 于是, Thomas 等用中心区域视差调谐曲线平移量和边缘区域视差变化量之比 (Shift-ratio) 来度量神经元对相对视差选择性的强弱<sup>[17]</sup>, 理想的对相对视差有选择性的神经元的 Shift-ratio 是 1, 完全没有相对视差选择性

的值是 0. 图 12 是他们实验中得到的 Shift-ratio 分布. 可以看出 V2 区中的一部分神经元的 Shift-ratio 值在 0 和 1 之间, 显示出不完全的相对视差选择性. 该图同时显示了 V1 区和 V2 区的区别, V1 区中的神经元几乎没有相对视差选择性.

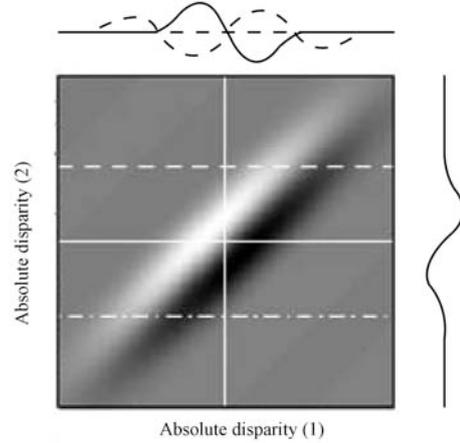


图 11 一个对相对视差有选择性的神经元的响应随刺激物的中心视差和边缘视差的变化规律<sup>[17]</sup>

Fig. 11 The response of a neuron selective for relative disparity to stimuli with varying central and surround disparities

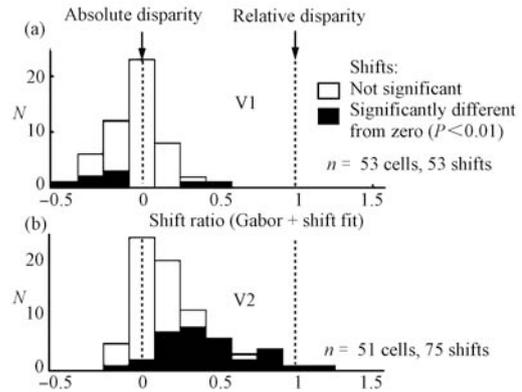


图 12 V1, V2 区神经元相对视差选择性比较<sup>[17]</sup>

Fig. 12 The shift ratios were compared for neurons in V1 (a) and V2 (b)

另外, V2 区中的神经元开始对由深度变化产生的边缘 (Cyclopean edge) 敏感<sup>[18]</sup>. Qiu 等的实验显示了 V2 区的神经元能够整合视差信息和格式塔信息 (Gestalt factors), 编码图像边缘的归属 (Border-ownership coding)<sup>[19]</sup>. 值得注意的是, 他们的实验中用了含有格式塔信息的刺激物. Bredfeldt 等随后使用只含有视差边缘信息、没有格式塔信息的 RDS 立体图同样进行了实验<sup>[20]</sup>, 发现 V2 区神经元对这类简单的刺激物的响应完全可以用一个简单的前馈模型解释. 这说明虽然 V2 区神经元可能参与前景

背景分割 (Figure-ground segmentation), 但是还不能理想地表示边界。

和 V1 区的另一个区别在于, V2 区神经元的发放和人的深度感知存在相关性<sup>[21]</sup>。这种相关性是通过计算神经元的选择概率 (Choice probability, CP) 来刻画的<sup>[22]</sup>。这类实验的大致思路是让猴子观看 RDS 随机点图, 并训练猴子回答 RDS 中心区的视差和注视点的远近关系。当同一个 RDS 反复出现时, 猴子的深度感知并不总是正确的, 同时, V2 区神经元的发放率也是变化的。Nienborg 等发现猴子对同一刺激物的感知波动和 V2 区的神经元的发放率有相关性, 而和 V1 区的神经元没有这种相关性。随后的实验证明了这种相关性受感知策略 (Perceptual strategy) 的影响<sup>[23]</sup>。应该怎样看待 V1 区和 V2 区的这个区别呢? 必须先分析造成这种相关性的原因。一种可能性是 V2 区的神经元发放直接或间接地引起了深度感知。当然, V2 区神经元的噪声也会引起感知的波动。另一种解释是 V2 区的神经元接收高层的反馈, 即 V2 区神经元的发放会随着感知的变化而变化。Nienborg 等最近的实验表明, 这种相关性更可能来自高层的反馈, 至少不是来自单纯的因果关系<sup>[24]</sup>。而 V1 区缺乏这种相关性的原因可能是缺乏高层负责深度感知的神经元的反馈。高层反馈到达 V2 区而没有到达 V1 区, 说明 V2 区在视差处理方面是比 V1 区更高级的区域。

最后, 和 V1 区相比, V2 区中有视差选择性的神经元形成了更规则的组织结构。光学成像 (Optical imaging) 实验显示 V2 区的粗条纹区 (Thick stripes) 有视差选择相似性的神经元以柱状结构组织在一起<sup>[25]</sup>, 而在 V1 区没有发现相似的结构。该实验表明 V2 区可能进一步处理了 V1 区的视差信号, 虽然现在对这种柱状结构的具体功能还没有定论。随后, 生理学家在 MT (Middle temporal area) 区 (见第 2.1 节) 等区域也发现了这样的柱状结构, V2 区的这种结构很有可能是更高级区域视差细胞聚集结构的起源。

以上的实验从多方面证明, V2 区进一步处理了 V1 区输入的视差信号, 从而得到更有价值的三维信息, 尽管 V2 区还不足以独立完成更高级的任务。

### 1.3 V3 区

实验表明 V3 及其临近区域在立体视觉中起着重要作用, Adams 等发现, 实验中 61% 的 V3 神经元都对视差有选择性, 并且 V3 中视差选择相似性的神经元以柱状结构集合在一起<sup>[26]</sup>。同时, V3 又是对三维形状有反应的最低级区域<sup>[27]</sup>。然而 V4, MT 这些接收 V3 输出的区却对三维形状的没有反应。目前 V3 及其临近区域在立体视觉通道中所处的位置

及其作用还没有更系统的报道。

## 2 背部通道高级区域

### 2.1 MT 区

MT 区接受多个低级区域, 如 V1, V2, V3 等的输出, 但主要接受 V1 区的输出<sup>[28]</sup>。与 V1 区不同, MT 区有规则的视差组织结构, 同时出现了一些更加复杂的特性, 如编码斜面的三维方向。此外, 一些实验发现 MT 区神经元发放与深度感知有紧密的关系。

MT 区大多数神经元具有视差选择性<sup>[29]</sup>, 编码的视差信号噪声比 V1 区要小, 并且编码范围更广。同时 MT 区视差选择特性相似的神经元是以柱状结构组织在一起的<sup>[30]</sup>, 利用群响应发现具有视差选择性的区域内的神经元的自身倾向视差是逐渐变化的。虽然运动方向选择相似的神经元也是以柱状结构组织在一起的<sup>[31]</sup>, 然而这种结构与视差选择的组织结构没有必然联系。MT 区在运动感知和立体视觉的作用这两者之间有没有联系, 目前还没有定论。

其次, MT 区能够对三维表面倾斜方向进行编码。Nguyenkim 等利用 RDS 斜面发现了大多数神经元 (72/97) 能够编码倾斜轴, 并且不随斜面的平均视差和倾斜程度的变化而变化<sup>[32]</sup>。他们认为神经元对倾斜轴的编码可能是由感受野内视差调谐的异质性决定的。而有这种特性的 MT 区神经元或许对相对视差具有一定的选择性, 但使用中心-外围随机点刺激物 (Center-surround RDS) 发现 MT 区不能编码相对视差<sup>[33-34]</sup>。

最后, MT 区神经元的发放与深度感知存在紧密联系。Uka 等训练猴子判断随机点刺激物与注视点的远近关系<sup>[35]</sup>, 通过记录神经元的发放, 发现 MT 区神经元的发放与猴子的判断有一定的相关性, 他们认为 MT 区对粗深度判断有贡献。然而, 这个实验并不能排除 MT 区神经元的发放是来自高层反馈信息的可能性。利用微电刺激 (Microstimulation) 实验<sup>[36-37]</sup>, 给某些 MT 区神经元一个弱电流刺激, 观察猴子深度感知的变化。此实验中猴子要完成两个任务: 1) 判断随机点刺激物与注视点的远近关系 (粗深度感知); 2) 判断刺激物中心和外围的远近关系 (细深度感知)。实验发现用微电极增加某些神经元的发放, 会影响猴子的粗深度感知, 而对细深度感知基本没有影响, 这与前面 MT 对绝对视差而非相对视差有选择性的结论是一致的。该实验通过测量 Shift-ratio (定义见第 1.2 节), 也进一步验证了 MT 区的神经元对相对视差不具有选择性。然而反相关模式 (Anti-correlated pattern) 视差变化时, 利用 Dodd 等的实验<sup>[38]</sup> 得到的 CP 很大的神经元的响应

也会随之变化, 即对不能引起深度感知的反相关模式的视差有选择性<sup>[39]</sup>, 这说明 MT 区的神经元发放跟深度感知还是有差别的. 因此, MT 区还是处在立体视觉模块比较低级的位置, 与它在视觉通道所处的位置是相一致的.

## 2.2 MST 区

MST 区 (Middle superior temporal area) 直接接受 MT 区的输出信息, 出现了比 MT 区更高级的功能.

MST 区大多数神经元是具有 (绝对) 视差选择性的远神经元 (Far neurons) 或近神经元 (Near neurons)<sup>[40]</sup>, 这一性质说明了 MST 区更适合于粗立体视觉 (Coarser stereopsis). MST 区神经元的发放还与眼动响应存在相关性. Takemura 等发现汇聚眼动 (Vergence response) 随视差变化的响应以及时间域的响应, 均可以用 MST 区神经元的群响应之和来近似<sup>[41]</sup>. 他们通过模糊分类方法, 根据 MST 区神经元的视差调谐曲线形状将神经元分为 4 类, 如图 13 所示, 在零视差附近, 神经元可以近似为一个线性伺服系统, 最下行是眼动响应随视差的变化曲线. 进一步的分析发现, 每一类神经元在近似眼动响应中都起着重要的作用. 每个神经元只能表示眼动响应的某个方面, 而群响应能很好地近似眼动响应, 这可能是由不同神经元的异质性决定的. 虽然 aRDS 不能引起深度感知<sup>[42]</sup>, 但与 RDS 一样能引起眼动响应, 只是方向相反, 这时眼动响应随视差的变化曲线仍可以用细胞群的视差调谐曲线的和来近似, 这些发现说明眼动响应可能只需要局部信息, 也说明了在 MST 区对应点匹配问题仍没有解决. Ilg 等通过微电刺激实验<sup>[43]</sup>, 进一步证实了 MST 区神经元在眼动响应中的重要作用, 给神经元增加微电流, 发现猴子的后扫视 (Post-saccade) 运动速度会增加, 并且眼动的方向倾向于神经元所倾向的运动方向. 该实验说明了 MST 区神经元的发放与眼动存在着因果关系. 因此, MST 区在控制眼动的通道中起着重要的作用.

其次, MST 区神经元的发放与手动存在着紧密联系<sup>[43]</sup>. 实验过程大致为: 猴子食指按住按钮, 双眼保持注视点不变, 运动目标出现后, 猴子松开按钮, 目标消失, 猴子食指触摸屏幕. 通过微电刺激 MST 区的神经元发放, 发现不管目标的运动方向如何, 猴子食指的位置集中于受刺激的神经元所倾向的运动方向附近, 这说明 MST 区神经元的发放影响到了猴子手指的位置, 其原因可能是 MST 区的信号被 IPS (Intraparietal sulcus) 区 (见第 2.3 节) 控制手动的神经元利用.

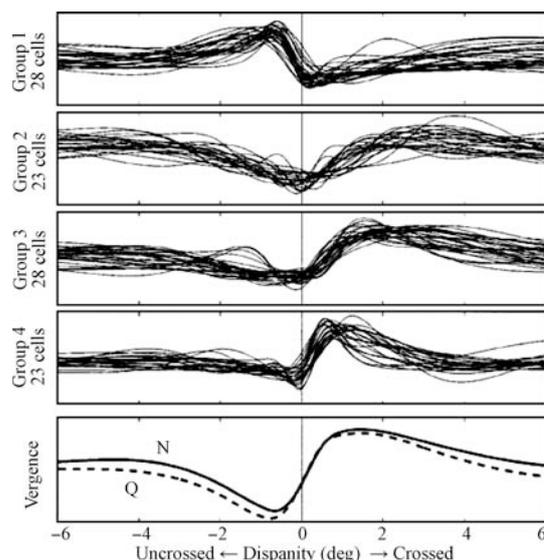


图 13 MST 不同神经元的视差调谐曲线之和可以近似眼动响应<sup>[41]</sup>

Fig. 13 The vergence eye movement could be approximated by the summed responses of different groups of MST neurons

以上的实验说明, MST 区相对 MT 区对视差信号做了进一步的处理, 神经元的发放与眼动响应有紧密联系, 但对应问题在 MST 区仍没有解决.

## 2.3 IPS 区

IPS 区是背部通道的高级区域, 能为某些行为提供视觉信息的指导, 如伸手-抓取运动、扫视眼动 (Saccade) 等. IPS 区包含几个小区, 每个小区都有不同的功能, 如 AIP (Anterior intraparietal area), LIP (Lateral intraparietal area) 和 MIP (Medial intraparietal area) 区分别控制抓取运动、眼动和伸手运动. Durand 等<sup>[44]</sup> 利用核磁共振成像技术 (Functional magnetic resonance imaging, fMRI), 发现 IPS 区对物体的三维形状, 空间位置以及 2D 形状具有选择性, 如表 1 所示. 实验结果表明各个子区域处理视差信号的方式支持了其负责的功能, 比如控制抓取运动的 AIP 区需要物体的三维形状信息, 而 MIP 区控制伸手运动, 控制手的运动轨迹, 需要物体的空间位置. 值得注意的是, Durand 等又在人脑的 IPS 区做了类似的实验<sup>[45]</sup>, 发现了与猴子存在功能相似的子区域, 例如 DIPSA (The dorsal IPS anterior area) 和 DIPSM (The dorsal IPS medial area) 区分别对应着 AIP 区的后部, LIP 的前部, 对应区的功能也存在着微小的差别, 对于 VIPS/V7\* (Ventral IPS area) 可能与 CIP (Caudal intraparietal area) 有同族关系, 但这种对应关系目前还不清楚.

表 1 IPS 区各个区对物体三维形状、空间位置及 2D 形状的选择性<sup>[44]</sup> (“√”- 选择性, “×”- 没有选择性)

Table 1 Possible selectivities of areas in IPS for 3D shape, position, and 2D shape (“√”-selectivity, “×”-nonselectivity)

	深度 位置	三维 形状	三维 曲面	三维 方向	大的 2D 形状
AIP 及其临近的区域	×	√	√	√	√
LIP 的后部	√	×	×	×	√
CIP	√	√	×	×	√
PIP, MIP	√	√	×	×	×

从上面的结果可以看出, AIP 区在获取物体的三维形状中起重要作用, Srivastava 等对 AIP 区神经元做了进一步的研究<sup>[46]</sup>, 发现 AIP 区的神经元对三维形状的选择性不随位置、大小的变化而变化. IPS 区对三维形状的选择具有快速性 (Fast)、粗略性 (Coarse) 和度量性 (Metric). 在刺激物出现 60~80 ms 后, AIP 区就能区分倾向 (Preferred) 与非倾向 (Non-preferred) 的三维形状. 并且当三维曲面用斜面或平面来近似时, AIP 区神经元的发放基本保持不变, 体现了 AIP 区视差处理的快速性和粗略性. 图 14 所示是神经元响应随曲面凹凸程度的变化曲线, 横坐标表示三维形状的凹凸程度, “+”与“-”分别代表该神经元倾向与非倾向的三维形状. 可以看出神经元响应是单调变化的, 并且在“0”值处连续, 体现了 AIP 区视差处理的度量性, 这也说明了 AIP 区对凹凸方向变化临界处不敏感, 同时 AIP 神经元不能很好地区分三维曲面与其近似, 即 AIP 区对三维形状粗略的选择性. 抓取运动需要快速得到物体粗略的三维形状信息, 与 AIP 区对三维形状的选择特性是相符的. 腹部通道的 TEs 也对三维形状具有选择性 (见第 3.2 节), 然而腹部通道与物体识别相关联, 任务的不同使得 TEs 的三维形状选择特性与 AIP 区有比较大的差别<sup>[47-48]</sup>.

Taira 等利用不含透视信息的 SFS (Solid figure stereogram) 和 RDS 立体图对进行了实验<sup>[49]</sup>, SFS 只是在表面轮廓含有视差梯度信息, 而 RDS 则是在表面含有视差梯度信息. 结果发现 CIP 区几乎所有的具有方向选择性的神经元对这两种三维斜面均具有选择性. 实验结果似乎与 Durand 等<sup>[44]</sup>的结果相矛盾, 这是因为两个实验的刺激物大小以及轮廓均是不一样的. Taira 实验的刺激物大小为  $6.3^\circ \times 6.3^\circ$ , 倾斜角为  $45^\circ$  的方形刺激物, 注视点的位置为 44 cm, 比 Durand 等<sup>[47]</sup>的实验刺激物要大, 而 CIP 一般对大刺激物而非小的刺激物有响应. 这

两个实验的对比说明 CIP 在处理三维形状的过程中其作用是有限的.

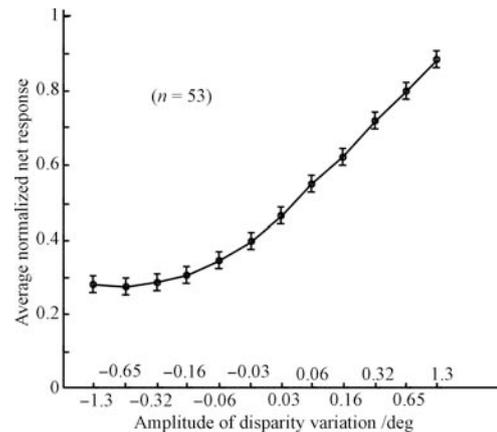


图 14 AIP 区对三维方向选择的度量性<sup>[46]</sup>

Fig. 14 Metric selectivity of 3D shape in AIP

上面的实验表明, IPS 区在伸手-抓取物体的运动中起重要作用, AIP 区对三维形状的选择性以及 TEs 的关系有待进一步的研究.

### 3 腹部通道高级区域

#### 3.1 V4 区

V4 区是腹部通道的中间区域, 它与 V2, V3, MT, IT 等区域均有连接. 一系列实验说明, V4 区神经元视差选择性介于低级区域 V1, V2 和高级区域 IT 之间, 符合该区在腹部通道中的解剖学位置.

首先, Hinkle 等研究了 V4 区神经元视差调谐曲线的分布, 他们发现这个分布是连续的<sup>[50]</sup>. 大多数神经元的调谐曲线是单调曲线和单峰曲线的结合, 但一部分曲线仅仅是单调曲线. 本区神经元的视差调谐曲线相对于 V1 区神经元更宽, 常常是单峰而非多峰, 而 IT 区大多数是单调曲线. 所以, V4 区的上述性质介于 V1 区和 IT 区之间. 与 V1 区相比, 神经元发放携带的视差信息的歧义性减少了. 与 V1, V2 区的神经元不同的是, V4 区开始出现对高阶视差敏感神经元. 对于光条刺激物, 大量的神经元对其在视网膜上像的二维方向有选择性. 在这些 V4 区神经元中, 1/3 的神经元对光条的三维方向产生了选择性<sup>[51]</sup>. 图 15 是 Hinkle 等的实验示意图. 第 1 行是光条俯视图, 第 2 行是光条正视图, 第 3 行是某个神经元随光条的朝向变化产生的调谐曲线. 可以看出这个神经元倾向左端在前、右端在后的倾斜光条. 值得注意的是, 在不同倾向角度下光条的二维形状是不变的, 因此, 神经元的发放变化不是由二维图像变化引起的. 进一步的实验表明这种三维选择性不随光条深度和视网膜上位置的变化而变化, 神

神经元主要是从光条的边缘处的视差变化来感知三维朝向的, 而光条内部的纹理并不是必需的. 那么 V4 区神经元对纯粹由视差定义的三维方向是否有相似的选择性呢? Hedge 等<sup>[52]</sup> 的实验用由 RDS 定义的平面、斜面和凹凸曲面, 对 V4 区的神经元是否能编码由视差定义的斜面或者曲面进行了测试. 他们发现虽然大量的神经元对斜面的倾斜方向或者曲面的凹凸方向敏感, 但是没有细胞可以称得上是真正编码了三维形状. 这两个实验说明 V4 区神经元只对特定刺激物的三维方向有选择性.

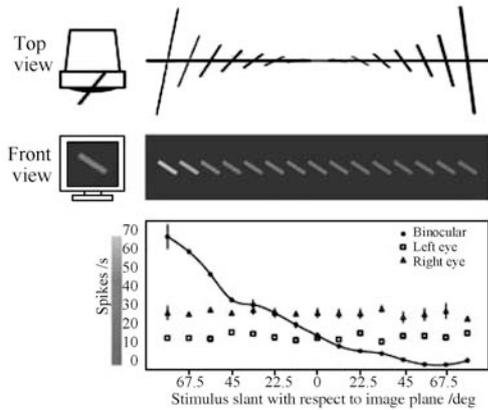


图 15 Hinkle 实验示意图<sup>[51]</sup>

Fig. 15 Sketch map of Hinkle's experiment

其次, 和 V2 区相似, V4 区的神经元也对相对视差敏感. Umeda 等发现与 V2 区相比, V4 区中有更高比例的神经元对相对视差敏感, 并且敏感性更强<sup>[53]</sup>. 同时他们也使用了 Shift-ratio 这个参数来度量相对视差的敏感性, 并与其他区域的神经元进行了比较. 结果如图 16 所示, 圆圈代表 Shift-ratio 的均值, 水平条的长度代表方差. 这个实验支持了 Thomas 等发现的关于 V2 区和 V1 区对相对视差选择性的区别<sup>[17]</sup>. V4 区的相对视差敏感性高于 V2 等几个区域, 但是其 Shift-ratio 指数只是理想值的一半, 所以 V4 区仍然没有达到对相对视差的完美表示. 除了上述定量的区别外, V4 区中相对视差敏感性的产生机制似乎也与 V2 区不同. V2 区神经元对相对视差的敏感性可以用前馈模型来解释, 但 V4 区神经元的特性与前馈模型并不完全相符<sup>[53]</sup>. 总之, 虽然 V4 区还不能完全地表示相对视差, 但 V4 区无疑对视差信号进行了进一步处理, 使之更接近于人的深度感知.

最后, 大多数对 RDS 视差有选择性的 V4 区神经元减弱或者完全消除了对 aRDS 视差的选择性<sup>[54]</sup>. 因为对 aRDS 视差没有选择性的神经元才有可能解决对应问题, 这个实验也反映出 V4 是比 V1, V2 更高级的视差处理区域. 除了 V4 区之外, 只有

IT 区和 IPS 区的某些神经元完全不对 aRDS 的视差有选择性. 那么 V4 区的这个特性是否来自于更高级区域, 比如 IT 区的反馈呢? Tanabe 等分析了 V4 区神经元发放随时间变化的规律. 他们发现 V4 区神经元从发放的开始就能区分 RDS 和 aRDS. 因为接收 IT 区域反馈是需要一定时间的, 所以他们认为这一特性不太可能来自高层反馈, 而是可能起源于 V4 和 V3 间的前馈连接. Kumano 等<sup>[55]</sup> 发现有较宽频率带宽的神经元对 aRDS 视差的选择性较弱, 因此, 他们推测这些神经元可能通过整合多个频段的信息来减弱对假目标的反应. 值得注意的是, 直接汇聚多频率通道信息并不能减弱对 aRDS 的选择性. Haefner 等的 2SU 模型<sup>[13]</sup> 能够解释 V1 区神经元对多频率反相关随机点立体图视差选择性比对单频率反相关光栅立体图更弱的现象<sup>[16]</sup>. V4 区神经元是否用到了相似的机制减弱了对 aRDS 的响应还有待进一步研究.

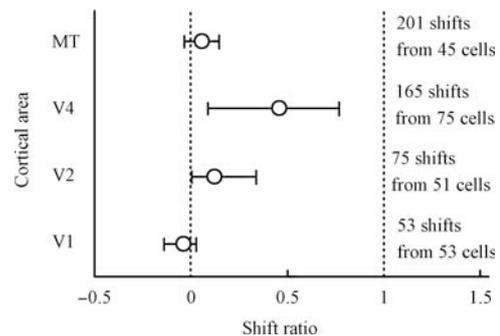


图 16 V4 区神经元相对视差选择性与其他区域的比较<sup>[53]</sup>

Fig. 16 The comparison of selectivities for relative disparity

### 3.2 TE 区

TE 是 IT 区中的一个区, 从解剖学的角度可以把它分成几个小区, 但最初人们并不清楚这些子区域功能上的区别. Jassen 等发现 TE 中对三维形状有选择性的神经元都集中在 TEs (Superior temporal sulcus), 而非 TE 的侧部 (Lateral TE), 结合 TEs 独有的与 IPS 区的连接, 他们认为 TE 至少分为两个部分, 即 TEs 区和侧部区域<sup>[47]</sup>. TEs 区和立体视觉密切相关, 因为它是腹部通道中唯一能够编码三维形状的区域, 其编码方式精细且鲁棒, 并且和深度感知有密切联系. 具体体现在:

Jassen 等发现 TEs 中除了对绝对视差有选择性的神经元外, 还有大量的神经元对一阶视差或者二阶视差有选择性<sup>[56]</sup>. 这种选择性是不随刺激物相对于感受野的位置、大小变化的, 且很多神经元对高阶视差的选择性十分精细. 图 17 所示的神经元对光滑曲面的凹凸方向有选择性, 该图的第 1、2 行显示

的是不同种类的 RDS 立体图和其对应的空间结构, 第 3 行显示的是神经元对凹曲面和其各种近似曲面的响应, 第 4 行显示的是对凸曲面和其各种近似曲面的响应. 我们可以发现这个神经元倾向于凹曲面, 并且这个神经元能够区分凹曲面和它的各种近似曲面. 进一步的实验发现 TEs 神经元对曲面凹凸的方向十分敏感, 超过其对凹凸程度的敏感性. 这一点和一些关于形状识别的心理学实验相吻合.

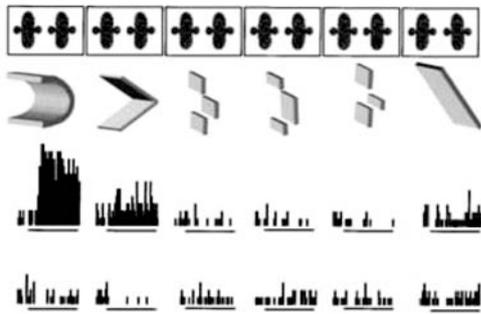


图 17 一个只对曲面响应的 TEs 神经元<sup>[57]</sup>

Fig. 17 This neuron is only responsive to curved surface

以上的实验中用到的刺激物既含有三维表面, 又含有三维轮廓的信息. TEs 神经元是靠哪种信息感知三维形状的呢? Jassen 等的实验使用了只含有其中一种信息的刺激物<sup>[57]</sup>. 其中, 无纹理曲面填充了不相关随机点, 不含有三维表面信息, 而用 RDS 定义的半球则不含有三维的轮廓信息. 在这个实验中, 很多 TEs 神经元对只含有轮廓信息的刺激物的三维形状有选择性. 相反地, 也有很多神经元只需要表面纹理信息就能产生三维形状的选择性. 因此, Jassen 等认为三维表面和轮廓的信息都在 TEs 中得到了表示. 具有能够从不同类型的信息推断三维形状的能力说明 TEs 神经元有较高的鲁棒性. 这一点也能够用 aRDS 的实验佐证. TEs 中无论是对高阶还是零阶视差有选择性的神经元, 都已经不对 aRDS 的视差有选择性. 从神经元发放率随时间变化的趋势分析, 这种特性不太可能是来自更高级区域的反馈. 所以 Jassen 等认为对应问题在 IT 区或者之前就解决了<sup>[58]</sup>.

上面的实验虽然表明 TEs 神经元能对三维形状进行编码, 但是并不能确定一个神经元最倾向的曲面形状. 困难在于曲面有太多可能的形状, 不可能记录神经元对每个曲面的发放率. Yamane 等用了一种进化搜索的策略快速地找到了神经元所倾向的曲面<sup>[59]</sup>. 图 18 显示了实验中搜索出的使某个 TEs 神经元有强烈发放的 50 个三维形状. 直观上可以发现这些形状都有共同的几何特征. Yamane 等对这些 TEs 神经元的选择性进行了数学建模. 他们发现一

个 TEs 神经元的发放率并不取决于整个三维形状, 而是取决于多个局部的曲面片. 对每一个曲面片, 神经元的发放依赖于该面片与中心的相对位置、曲率、表面朝向等 7 个参数. 当一个神经元所倾向的几个曲面片同时在一个三维形状中出现时, 神经元的发放就会很大. 与人们的直觉相反, 一个神经元所编码的不是整个三维形状, 也不是某个孤立的局部形状, 而是局部曲面片的组合. 这可以看成是一种形状编码的折中方案, 其好处在于既使得单个神经元的信息量比较大, 又能保证所需的神经元数量不会太多.

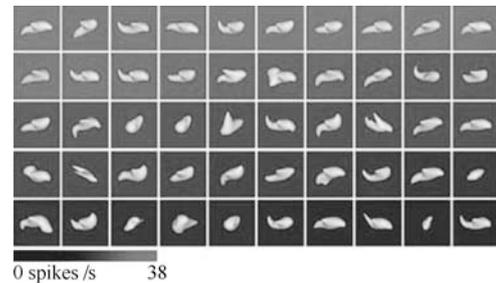


图 18 引起 TEs 某个神经元最大发放率的 50 个三维形状<sup>[59]</sup>

Fig. 18 Fifty 3D shapes causing an TEs neuron's highest response

总之, TEs 神经元所表现出来的编码三维形状的鲁棒性和精密性能满足人类判断三维形状、识别物体的要求. 这些性质是腹部通道的其他区域所不具备的. TEs 中大量的神经元对二维形状也有选择性, 二维信息和三维信息的结合使得这个区成为实现物体识别功能的理想场所. 这些研究表明三维形状在物体识别中也起到一定的作用. 需要说明的是, TEs 的这些特性并不一定完全来自它本身, 它和 IPS 区域还存着一些连接关系. 生理学实验确实也发现了这两个区域中三维形状选择性的相似之处.

## 4 总结

本文总结了近 20 年神经生理学在立体视觉方面的主要发现, 这些发现揭示了神经系统如何准确快速地从二维的视网膜图像中感知物体的空间位置和三维形状, 让我们不得不感叹大脑视皮层功能的强大. 但要彻底弄清视皮层的机制, 还需要长期艰苦的努力. 可以预见随着实验手段的进步, 特别是各种成像手段的进步, 不久的将来还会有更多有价值的成果问世. 这些神经科学的成果不但能增加人们对生物视觉系统的了解, 还能为计算机立体视觉的研究提供新的视角. 所以本文希望通过梳理这些成果, 为对相关领域感兴趣的读者提供一些帮助. 在结束本文之前, 还有下列几点需要指出:

1) 目前的神经生理结果, 使用的刺激基本上是正弦光栅和随机点立体图对. 利用这些简单刺激得到的结果, 也许与复杂的自然场景下的结果有本质的区别. 再加上目前基于电极的测量神经元发放的手段, 也仅仅能同时测量几十个神经元. 对如此复杂庞大的神经网络, 目前的结果会不会都是“盲人摸象”呢? 非创伤性观测手段的发展, 为同时观测大量神经元提供了可能, 可望一些全新的成果, 甚至对目前理论带有颠覆性的成果即将问世.

2) 关于立体感知模型, 本文仅仅介绍了解释初级视皮层生理数据的能量视差模型及其扩展模型. 这主要是因为对于 V1 区后续的区域, 人们对视差计算的生理机制还不是很清楚, 文献中还没有大家比较认可的模型. 值得庆幸的是, Jassen 等的研究表明, 对应问题至少在 IT 区已经解决<sup>[58]</sup>, 考虑到神经反应时间, 可能对应问题并不需要高层信息的反馈, 这一结论无疑为建立鲁棒的计算模型提供了生理基础. 但是, 没有高层信息反馈, 绝不意味着层与层之间没有反馈. 事实上, 可能正是由于层与层之间的反馈, 才使得人类视觉在对应问题上表现出无比的鲁棒性. Grossberg 研究组的 Laminar cortical model (LCM)<sup>[60]</sup>, 包含层与层之间的前向连接和反向反馈. 从生理学的观点, 目前人们对于高层区域对低层区域的反馈机制还知之甚少, 如反馈什么, 如何反馈. 所以, 我们没有对 LCM 进行介绍, 感兴趣的读者可以参阅相关文献.

3) 另外一个问题是, 人类在解决对应问题时, 是否真的需要高层信息的反馈, 特别是具有语义的高层信息. Jassen 等的实验<sup>[58]</sup>用的刺激物是随机点图, 对于自然场景, 对应问题是否至少在 IT 区已解决了呢? 纯粹的视差信息是否能够获取自然场景准确的三维信息呢? 视觉系统和认知的关系又是怎样的呢? 现实生活中, 视觉信号和其他信息来源 (如听觉信号) 结合, 共同为生物的感知世界服务, 所以在大脑皮层的某个阶段, 多种信号发生了融合. 另一方面, 视觉系统本身也会受到高层知识 (如先验知识) 的影响. 我们期待在将来能够有新的发现来进一步揭示两者的关系.

4) 目前对视觉的研究, 特别是定量研究, 在感知层面的比较多. 尽管文献中有 Visual cognition 这样的研究分支, 但从认知角度看, 纯粹视觉认知, 特别是一些比较复杂的认知现象, 仅仅是视觉刺激的结果吗? 我们认为认知一定是大脑不同模态共同作用的结果, 不能单单强调视觉的作用. 鉴于此, 我们认为, 在近期内如何建立视觉 V2, V3 区的计算模型, 应该是研究重点. V3 区以后, 如何建立背向或腹部通道不同区域的计算模型, 还有很长的路要走, 在短期内不太可能有大的进展.

最后我们列出在本领域有系统性研究的实验室网址, 感兴趣的读者可以从这些网站进一步了解更详细的信息.

Bruce G. Cumming 研究组:

<http://www.nei.nih.gov/intramural/lsr/cumming/publications.asp>

Izumi Ohzawa 研究组:

<http://www7.bpe.es.osaka-u.ac.jp/index-e.html>

Guy A. Orban 研究组:

<http://www.labome.org/expert/belgium/katholieke/orban/guy-a-orban-953443.html>

Ralph D. Freeman 研究组:

<http://vision.berkeley.edu/VSP/content/faculty/facprofiles/freeman.html>

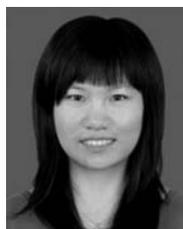
## References

- 1 Yu Lian-Qing. Stereo Matching with Cortical Disparity Detection Mechanisms [Ph. D. dissertation], Institute of Automation, Chinese Academy of Sciences, China, 2008 (于连庆. 基于视皮层视差检测机制的立体视觉技术研究 [博士学位论文], 中国科学院自动化研究所, 中国, 2008)
- 2 Barlow H B, Blakemore C, Pettigrew J D. The neural mechanism of binocular depth discrimination. *The Journal of Physiology*, 1967, **193**(2): 327–342
- 3 Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 1962, **160**(1): 106–154
- 4 Hubel D H, Wiesel T N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 1968, **195**(1): 215–243
- 5 Jones J P, Palmer L A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 1987, **58**(6): 1233–1258
- 6 Cumming B G, Parker A J. Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *The Journal of Neuroscience*, 1999, **19**(13): 5602–5618
- 7 Cumming B G, Parker A J. Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *The Journal of Neuroscience*, 2000, **20**(12): 4758–4767
- 8 Cumming B G, Parker A J. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, 1997, **389**(6648): 280–283
- 9 Nienborg H, Bridge H, Parker A J, Cumming B G. Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *The Journal of Neuroscience*, 2004, **24**(9): 2065–2076
- 10 Nienborg H, Bridge H, Parker A J, Cumming B G. Neuronal computation of disparity in V1 limits temporal resolution for detecting disparity modulation. *The Journal of Neuroscience*, 2004, **25**(44): 10207–10219

- 11 Sanada T M, Ohzawa I. Encoding of three-dimensional surface slant in cat visual areas 17 and 18. *Journal of Neurophysiology*, 2006, **95**(5): 2768–2786
- 12 Ohzawa I, DeAngelis G C, Freeman R D. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 1990, **249**(4972): 1037–1041
- 13 Haefner R M, Cumming B G. Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model. *Neuron*, 2008, **57**(1): 147–158
- 14 Read J C, Cumming B G. Sensors for impossible stimuli may solve the stereo correspondence problem. *Nature Neuroscience*, 2007, **10**(10): 1322–1328
- 15 Fleet D J, Wagner H, Heeger D J. Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Research*, 1996, **36**(12): 1839–1857
- 16 Tanabe S, Cumming B G. Mechanisms underlying the transformation of disparity signals from V1 to V2 in the macaque. *The Journal of Neuroscience*, 2008, **28**(44): 11304–11314
- 17 Thomas O M, Cumming B G, Parker A J. A specialization for relative disparity in V2. *Nature Neuroscience*, 2002, **5**(5): 472–478
- 18 Heydt V D R, Zhou H, Friedman H S. Representation of stereoscopic edges in monkey visual cortex. *Vision Research*, 2000, **40**(15): 1955–1967
- 19 Qiu F T, Heydt V D R. Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. *Neuron*, 2005, **47**(1): 155–166
- 20 Bredfeldt C B, Read J C A, Cumming B G. A quantitative explanation of responses to disparity-defined edges in macaque V2. *Journal of Neurophysiology*, 2009, **101**(2): 701–713
- 21 Nienborg H, Cumming B G. Macaque V2 neurons, but not V1 neurons, show choice-related activity. *The Journal of Neuroscience*, 2006, **26**(37): 9567–9578
- 22 Britten K H, Newsome W T, Shadlen M N, Celebrini S, Movshon J A. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, 1996, **13**(1): 87–100
- 23 Nienborg H, Cumming B G. Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nature Neuroscience*, 2007, **10**(12): 1608–1614.
- 24 Nienborg H, Cumming B G. Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, 2009, **459**(7243): 89–92
- 25 Chen G, Lu H D, Roe A W. A map for horizontal disparity in monkey V2. *Neuron*, 2008, **58**(3): 442–450
- 26 Adams D L, Zeki S. Functional organization of macaque V3 for stereoscopic depth. *Journal of Neurophysiology*, 2001, **86**(5): 2195–2203
- 27 Georgieva S, Peeters R, Kolster H, Todd J T, Orban G A. The processing of three-dimensional shape from disparity in the human brain. *The Journal of Neuroscience*, 2009, **29**(3): 727–742
- 28 Born R T, Bradley D C. Structure and function of visual area MT. *Annual Review of Neuroscience*, 2005, **28**(1): 157–189
- 29 DeAngelis G C, Uka T. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology*, 2003, **89**(2): 1094–1111
- 30 DeAngelis G C, Newsome W T. Organization of disparity-selective neurons in macaque area MT. *The Journal of Neuroscience*, 1999, **19**(4): 1398–1415
- 31 Albright T D, Desimone R, Gross C G. Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology*, 1984, **51**(1): 16–31
- 32 Nguyenkim J D, DeAngelis G C. Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *The Journal of Neuroscience*, 2003, **23**(18): 7117–7128
- 33 Uka T, DeAngelis G C. MT neurons do not signal relative disparity. *Journal of Vision*, 2002, **2**(7): 37
- 34 Neri P, Bridge H, Heeger D J. Stereoscopic processing of absolute and relative disparity in human visual cortex. *Journal of Neurophysiology*, 2004, **92**(3): 1880–1891
- 35 Uka T, DeAngelis G C. Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *The Journal of Neuroscience*, 2003, **23**(8): 3515–3530
- 36 Uka T, DeAngelis G C. Linking neural representation to function in stereoscopic depth perception: roles of the middle temporal area in coarse versus fine disparity discrimination. *The Journal of Neuroscience*, 2006, **26**(25): 6791–6802
- 37 DeAngelis G C, Cumming B G, Newsome W T. Cortical area MT and the perception of stereoscopic depth. *Nature*, 1998, **394**(6694): 677–679
- 38 Dodd J V, Krug K, Cumming B G, Parker A J. Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *The Journal of Neuroscience*, 2001, **21**(13): 4809–4821
- 39 Krug K, Cumming B C, Parker A J. Comparing perceptual signals of single V5/MT neurons in two binocular depth tasks. *Journal of Neurophysiology*, 2004, **92**(3): 1586–1596
- 40 Roy J P, Komatsu H, Wurtz R H. Disparity sensitivity of neurons in monkey extrastriate area MST. *The Journal of Neuroscience*, 1992, **12**(7): 2478–2492
- 41 Takemura A, Inoue Y, Kawano K, Quaia C, Miles F A. Single-unit activity in cortical area MST associated with disparity-vergence eye movements: evidence for population coding. *Journal of Neurophysiology*, 2001, **85**(5): 2245–2266
- 42 Masson G S, Busetini C, Miles F A. Vergence eye movements in response to binocular disparity without depth perception. *Nature*, 1997, **389**(6648): 283–286
- 43 Ilg U J, Schumann S. Primate area MST-l is involved in the generation of goal-directed eye and hand movements. *Journal of Neurophysiology*, 2007, **97**(1): 761–771

- 44 Durand J B, Nelissen K, Joly O, Wardak C, Todd J T, Norman J F, Janssen P, Vanduffel W, Orban G A. Anterior regions of monkey parietal cortex process visual 3D shape. *Neuron*, 2007, **55**(3): 493–505
- 45 Durand J B, Peeters R, Norman J F, Todd J T, Orban G A. Parietal regions processing visual 3D shape extracted from disparity. *Neuroimage*, 2009, **46**(4): 1114–1126
- 46 Srivastava S, Orban G A, De Maziere P A, Janssen P. A distinct representation of three-dimensional shape in macaque anterior intraparietal area: fast, metric, and coarse. *The Journal of Neuroscience*, 2009, **29**(34): 10613–10626
- 47 Janssen P, Vogels R, Orban G A. Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science*, 2000, **288**(5473): 2054–2056
- 48 Janssen P, Vogels R, Orban G A. Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**(14): 8217–8222
- 49 Taira M, Tsutsui K I, Jiang M, Yara K, Sakata H. Parietal neurons represent surface orientation from the gradient of binocular disparity. *Journal of Neurophysiology*, 2000, **83**(5): 3140–3146
- 50 Hinkle D A, Connor C E. Quantitative characterization of disparity tuning in ventral pathway area V4. *Journal of Neurophysiology*, 2005, **94**(4): 2726–2737
- 51 Hinkle D A, Connor C E. Three-dimensional orientation tuning in macaque area V4. *Nature Neuroscience*, 2002, **5**(7): 665–670
- 52 Hegde J, Van Essen D C. Role of primate visual area V4 in the processing of 3-D shape characteristics defined by disparity. *Journal of Neurophysiology*, 2005, **94**(4): 2856–2866
- 53 Umeda K, Tanabe S, Fujita I. Representation of stereoscopic depth based on relative disparity in macaque area V4. *Journal of Neurophysiology*, 2007, **98**(1): 241–252
- 54 Tanabe S, Umeda K, Fujita I. Rejection of false matches for binocular correspondence in macaque visual cortical area V4. *The Journal of Neuroscience*, 2004, **24**(37): 8170–8180
- 55 Kumano H, Tanabe S, Fujita I. Spatial frequency integration for binocular correspondence in macaque area V4. *Journal of Neurophysiology*, 2008, **99**(1): 402–408
- 56 Janssen P, Vogels R, Orban G A. Three-dimensional shape coding in inferior temporal cortex. *Neuron*, 2000, **27**(2): 385–397
- 57 Janssen P, Vogels R, Liu Y, Orban G A. Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. *The Journal of Neuroscience*, 2001, **21**(23): 9419–9429
- 58 Janssen P, Vogels R, Liu Y, Orban G A. At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron*, 2003, **37**(4): 693–701

- 59 Yamane Y, Carlson E T, Bowman K C, Wang Z H, Connor C E. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*, 2008, **11**(11): 1352–1360
- 60 Grossberg S, Howe P D. A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research*, 2003, **43**(7): 801–829



**孔庆群** 中国科学院自动化研究所博士研究生. 2008 年获得中国海洋大学工程学院自动化系学士学位. 主要研究方向为基于人类视觉机制的计算理论和模型. 本文通信作者.

E-mail: qqkong@nlpr.ia.ac.cn

(**KONG Qing-Qun** Ph.D. candidate at the Institute of Automation,

Chinese Academy of Sciences. She received her bachelor degree from Ocean University of China in 2008. Her research interest covers computational theories and models based on human vision mechanism. Corresponding author of this paper.)



**明雁声** 澳大利亚国立大学博士研究生. 2007 年获得浙江大学控制科学与工程系学士学位. 2010 年获得中国科学院自动化研究所硕士学位. 主要研究方向为基于人类视觉机制的计算理论和模型.

E-mail: mingyansheng@hotmail.com

(**MING Yan-Sheng** Ph.D. candidate at the College of Engineering and

Computer Science, Australian National University. He received his bachelor degree in the Department of Control Science and Engineering, Zhejiang University in 2007. He received his master degree in pattern recognition and machine learning from the Institute of Automation, Chinese Academy of Sciences in 2010. His research interest covers computational theories and models based on human vision mechanism.)



**胡占义** 中国科学院自动化研究所研究员. 主要研究方向为摄像机标定、三维重建和视觉机器人导航.

E-mail: huzy@nlpr.ia.ac.cn

(**HU Zhan-Yi** Professor at the Institute of Automation, Chinese Academy of Science. His research interest covers camera calibration, 3D reconstruction,

and vision guided robot navigation.)