# 北京理工大学本科生毕业设计（论文）外文翻译

**外文原文题目：** Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?

**中文翻译题目：** 灵长类视觉皮层中的深层次结构：我们可以为计算机视觉学习什么？

**毕业设计（论文）题目：** 基于自然场景统计的立体图像质量评价

**姓 名：** 张久发

**学 院：** 计算机学院

**班 级：** 07111305

**指导教师：** 刘利雄

**灵长类视觉皮层中的深层次结构：我们可以为计算机视觉学习什么？**

From：Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?，IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 35, NO. 8, AUGUST 2013.

摘要：灵长类视觉系统的计算模型的产生与计算机视觉正在面临的一些挑战具有潜在的相关性，比如对象的识别与分类、运动检测和动作识别、或者基于视觉的导航和操纵。本文回顾了通常被认为是灵长类视觉皮层的一些功能原理和结构，并且尝试得到可以进一步推进计算机视觉研究的生物学原理。针对计算机视觉的相关读者，我们考虑到最近在神经生理学上的发现，提出了在灵长类视觉系统存在的处理层次的功能原理。与现在使用的主流的计算机视觉的平面视觉结构相对比，灵长类视觉系统的层次处理的特征为构成了深层次的不同级别的处理序列（大约10个数量级）。我们希望在灵长类视觉系统中实现的深层次的功能描述能够在为计算机视觉算法的设计提供有价值的参考，促进生物学和计算机视觉研究之间更加有效地互动。

目录

[本科生毕业设计（论文）外文翻译 1](#_Toc482894232)

[第1章 引言 2](#_Toc482894233)

[第2章 视觉皮层结构的相关概念 6](#_Toc482894234)

[2.1 层次结构 6](#_Toc482894235)

[2.2 不同视觉区域的基本情况 8](#_Toc482894236)

[第3章 皮质下视觉系统 9](#_Toc482894237)

[3.1 基本层次：视网膜光感受器 10](#_Toc482894238)

[3.2 神经节细胞和LGN 11](#_Toc482894239)

[3.2.1 中心-环绕接收场 12](#_Toc482894240)

[3.2.2 单向对称细胞 12](#_Toc482894241)

[第4章 枕叶皮质中的一般场景表示 13](#_Toc482894242)

[4.1 V1区域 13](#_Toc482894243)

[4.1.1 边缘，条和光栅 14](#_Toc482894244)

[4.1.2 点特征 15](#_Toc482894245)

[4.1.3 绝对差异 15](#_Toc482894246)

[4.1.4 局部运动 16](#_Toc482894247)

[4.1.5 双向对称细胞 17](#_Toc482894248)

[4.2 V2区域 18](#_Toc482894249)

[4.2.1 纹理定义和虚像（主观）轮廓 19](#_Toc482894250)

[4.2.2 边界所有权 19](#_Toc482894251)

[4.2.3 相对差异 20](#_Toc482894252)

[4.3 V4区域 20](#_Toc482894253)

[4.3.1 曲率选择性 21](#_Toc482894254)

[4.3.2 色相和亮度不变性 21](#_Toc482894255)

[4.4 MT区域 22](#_Toc482894256)

[4.4.1 2D运动 22](#_Toc482894257)

[4.4.2 运动渐变和运动定义的形状 23](#_Toc482894258)

[第5章 目标识别与分类：腹测流 23](#_Toc482894259)

[5.1 TEO区域 24](#_Toc482894260)

[5.2 TE区域 24](#_Toc482894261)

[第6章 动作视觉：背侧流 26](#_Toc482894262)

[6.1 MST 26](#_Toc482894263)

[6.2 尾部壁内区（Caudal Intraparietal Area，CIP） 27](#_Toc482894264)

[6.3 侧顶叶（Lateral Intraparietal Area，LIP） 28](#_Toc482894265)

[6.4 腹侧壁内区（Ventral Intraparietal area，VIP） 29](#_Toc482894266)

[6.5 内侧壁内区（Medial Intraparietal Area，MIP） 30](#_Toc482894267)

[6.6 前壁内区（Anterior Intraparietal Area，AIP） 30](#_Toc482894268)

[第7章 垂直视图：处理不同的视觉模式 31](#_Toc482894269)

[7.1 颜色 32](#_Toc482894270)

[7.2 二维形状 33](#_Toc482894271)

[7.3 三维形状 34](#_Toc482894272)

[7.4 运动 35](#_Toc482894273)

[7.5 对象识别 37](#_Toc482894274)

[7.6 动作表示支持（Action Affordances） 38](#_Toc482894275)

[第8章 我们可以在视觉系统中为计算机视觉学习什么？ 39](#_Toc482894276)

[8.1 分层处理 39](#_Toc482894277)

[8.2 信息通道的分离 40](#_Toc482894278)

[8.3 反馈 41](#_Toc482894279)

[8.4 视觉处理层次结构的发展与学习 42](#_Toc482894280)

[第9章 总结 42](#_Toc482894281)

[致谢 43](#_Toc482894282)

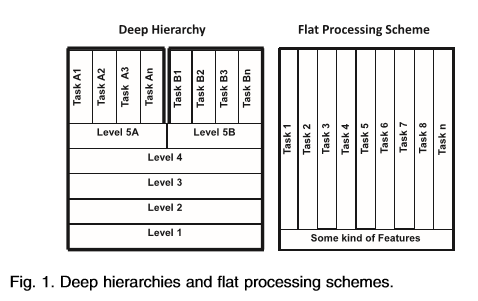
[参考文献 43](#_Toc482894283)

# 第1章 引言

计算机视觉的研究历史已经有半个多世纪了。然而，对于一些重要的问题比如大规模物体、场景和动作的识别与分类，以及基于视觉的操纵的一般的、稳定的、完整的、令人满意的解决方案仍然是现在的机器视觉系统所不能做到的。生物视觉系统，尤其是灵长类的视觉系统，看起来可以轻松地完成这些人物，也因此经常被用来作为计算机视觉研究人员的灵感来源。

“生物视觉”和“计算机视觉”的交互强度在计算机视觉的发展历史中不尽相同，并且在一定程度上反映了机器视觉研究者变化的研究重点[32]。毫无疑问，Hubel和Wiesel[72]通过Marr[109]建立类似于灵长类视觉系统的视觉层次的工作而进行的突破性工作，给计算机视觉研究群体带来了巨大的推动力。然而，当时缺少可用的计算资源，并且对灵长类视觉系统的处理阶段缺少更加详细的了解，这给在这一方向上未来的研究过程带来了2个难以对付的障碍。

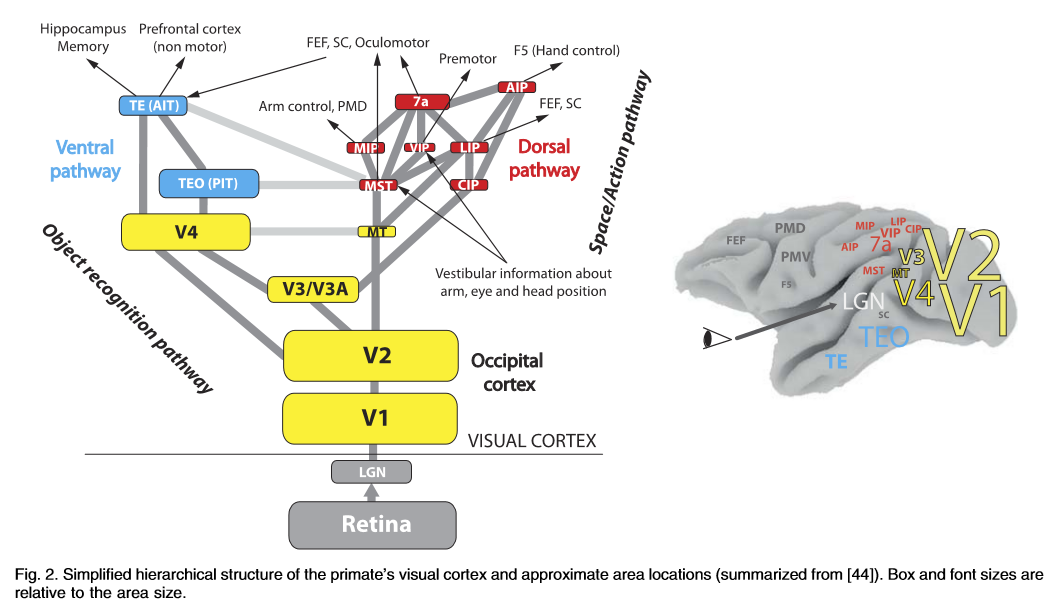
接下来是主流计算机视觉的重定位，从尝试解决一般性问题到更多地关注与具体任务相关的具体方法。这在平面处理方案中最常实现（参见Fig.1，右），把简单的基于特征的描述作为输入，然后由与任务相关的学习算法对它们进行处理。这样与生物视觉的联系不那么紧密了，如果这里还有与生物相关的机制的参考，通常都被局限在个别功能模块或者特征选择中，比如Gabor小波。



虽然在一些特定机器视觉问题和问题领域取得了巨大进展（在一些任务中，这些系统可以轻易地超越人类的能力），人工系统依然缺少灵长类视觉系统具有的一般性和稳定性。随着我们越来越深刻地理解视觉皮层的功能机制（主要由于在神经科学中使用的成像技术），现在我们可以重新尝试寻找可以使人工视觉在一般性和稳定性的能力更接近生物系统的能力的机制。从计算的角度看，这也可能是一个可行的事业，尤其是考虑到计算机体系结构的新的发展，比如GPU和多核系统。

在本文中，我们将首先关注灵长类的层次表示和功能机制。我们将看看不同的处理层次和不同的信息信道（例如，形状、颜色、动作），并且讨论出现在层次结构中的信息抽象。

我们都知道有大约55%的灵长类大脑皮层与视觉有关[44]，有一个跨越8到10个级别的处理管道的层次组织（参见Fig.2）。有清楚的证据表明，早期视觉区域中的神经元在视觉空间的小的局部区域提取简单的图像特征（例如，方向、动作、视差等），这些信息之后被传输到响应覆盖更大的视觉区域的更复杂的特征的更高视觉区域的神经元中。与所谓的平面处理模式（参见Fig.1，右）相比，这样的分层结构，我们称为深层次结构（参见Fig.1，左）表现出了许多的计算优势。



计算效率和泛化是两个重要的方面：随着层次级别的在彼此间的建立，他们利用共享机制来有效地到达更复杂的信息单元。这样的设计原理有助于通用计算（在学习和推理期间），将导致在更低存储需求上的更高效的处理。此外，反复利用在不同视觉实体间存在的共性，并且是用于实现不同任务的重要的感知构建块，导致泛化能力和知识的传递。例如，有很强的神经生理学证据表明，根据各种视觉特性的一般描述是在区域V1-V4和MT中计算的，覆盖了大约60%的灵长类皮层中的视觉处理量（参见[44]和Fig.2，其中视觉区域与它们的实际尺寸成比例绘制）。这些区域携带可完成不同任务的大量重要信息，比如对象识别和分类、掌握、操纵、路径规划等。

此外，在灵长类视觉系统中存在处理不同类型的视觉信息（颜色、形状、运动、纹理、3D信息）的分离的（尽管高度互联）的通道，这有助于表达的效率（避免集成表示的组合爆炸）和稳定性（相对于可用信息）。这些优势涵盖多个方面，将在第八节中进行详细讨论。

然而，尽管所有的神经生理学证据表明在灵长类视觉系统中实现了相当多的层次级别，大多数现有的计算机视觉系统却是“平面的”因而不能充分利用与深层次相关的优势。尤其是泛化和可扩展能力对于任何认知智能都是至关重要的。事实上，有压倒性的神经生理学证据表明认知和深层次的概念是相关联的。因此，我们将建立深层次结构作为我们建立人工认知系统的一个主要挑战。

Bengio[9]讨论了深层次的潜力和与深层次学习的相关问题。特别的，他强调了由于大量的层次级别而需要被探究的巨大的参数空间的问题。学习问题可以通过以下方式缓解：1）把处理中间表示作为独立的学习问题，2）根据在深层次机构的独立单元的层次级别或局部连接所表示的基本连接结构引入偏差。我们相信本文可以通过在灵长类视觉系统中给出恰当的中间表示的指示来帮助指导深层次视觉系统的学习过程。此外，我们相信可以在生物模型中得到合适的连接模式的指导，通过神经元合适的感知域大小，生物模型正在处理的层次的数量，以及被灵长类视觉皮层面积大小所指示的在一个特定层次级别的单位数量。

尽管存在与深层次学习相关的挑战，在计算机视觉中有一系列的工作在理解和建立层次模型时有重要贡献。由于缺少空间，更加彻底的回顾不在本文范围内，下面的列表也远不完整。从计算复杂性的角度来看，Tsotsos[188]已经证明了无限视觉搜索是NP完全的，层次结构可能是解决问题的最有前途的方案。一些工作表明，有效匹配只能在几个层次阶段进行，包括Ettinger[42]，Geman等[59]，[60]，Mel和Fiser[114]，Amit[1]，[2]，Hawkins[69]，Fidler等[45]，Scalzo和Piater[152]，Ullman和Epshtein[191]，DiCarlo和Cox[31]，Ommer和Buhmann[126]，Serre和Poggio[156]，Pugeault等[207]和Rodrı ´guez-Sa ´nchez和Tsotsos[143]。更多已知的层次模型有Neocognitron[54]，HMAX[140]，[157]，LHOP[46]，2DSIL[144]，和ConvolutionalNets[101]。最近，Bengio[9]发表了一片关于人工智能的深层结构的详细文章。

总之，在本文中我们想要表明深层次结构是实现一个一般的、稳定的、通用的计算机视觉系统的合适概念。更重要的是，我们希望能以一种可理解的方式为计算机视觉科学家提出灵长类视觉系统层次组织的相关见解。我们意识到我们的抽象概念从神经生理学观点来看是相当粗糙的，我们也已经忽略了在在不同层次进行处理时的重要细节，但我们希望这种抽象和本文提出的整体图片能有助于促进两个领域间的有效交流。

本文组织如下：在第2节中，我们将涉及与理解和对处理层次建模相关的灵长类视觉系统的相关方面。然后从2个方面概述灵长类视觉系统的层次结构。在横向视角（第3、4、5、6节）中，我们给出了在Fig.2中所示的不同区域的处理的描述。在第7节中，我们给出了在不同区域处理不同视觉模式的垂直视角。在第8节中，我们得到有关在深层次结构下的人工视觉系统建模与学习的结论。

# 第2章 视觉皮层结构的相关概念

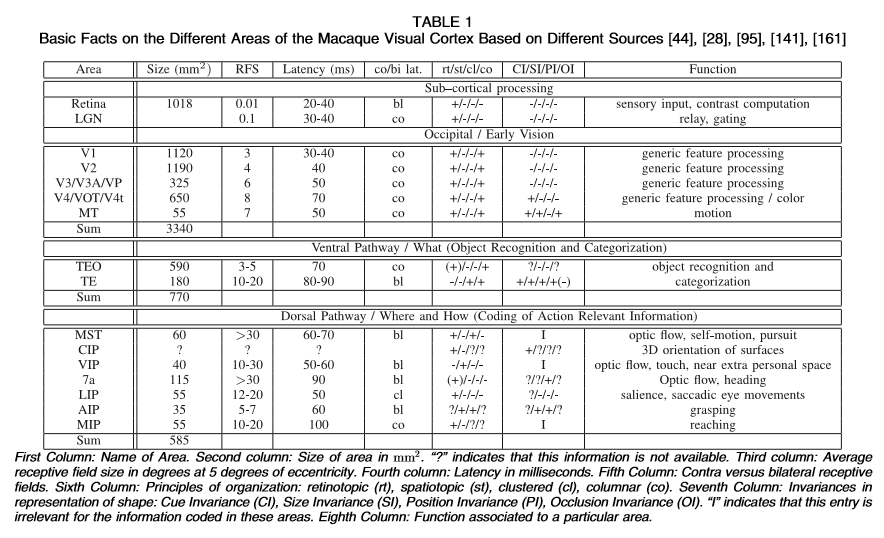
在2.1节中，我们提出关于灵长类视觉系统深层次结构的概述。在2.2节中，我们也将给出在下面小节中使用的基本术语（主要是生物学的）的直观知识。我们下面提供的大多数数据是从猕猴得到的，因为大多数神经生理学知识也来源于对它们的调查。

虽然灵长类大脑由大约100个皮层区域组成，但人类大脑可能含有多达150个区域。有普遍共识的是，猴子中的主要感觉和运动区域与人类大脑中的相应区域同源。此外，猴子中的几个其他皮层区域在人类中具有已鉴定的同源物（例如，MT / MST，前足内侧区域（AIP））。这些区域被视为可用于将人类中的其他皮层区域与猴子中的已知区域相关联的界标。

应该提到的是，视觉皮质区域由六个层组成，其不对应于人工深度模型中的层。通常，第4层是输入层，其中来自较早级的输入到达这一层。第4层之上的层（第2、3层）通常向下游视觉区域（例如，从V1到V2）发送前馈连接，而层5和6向上游区域或结构（例如从V1到LGN和the Superior Colliculus-参见第3.2节）。在视觉层次中的较高阶段，连接几乎总是双向的。目前，关于皮质微电路在这些不同层中的确切作用的详细知识是缺乏的。

## 2.1 层次结构

在这里，我们给出了在灵长类动物视觉系统中实现的处理层次的粗略和直观的总结。 更详细的描述可以在第3,4,5和6节中找到。Table 1中提供了关于不同区域的大小，接受字段大小，等待时间，组织等的基本数据。



视觉信息的神经元处理过程在左眼和右眼的视网膜中开始。几乎所有的连接在到达视觉皮层之前会被投射到称为LGN的视觉区域。我们将这些阶段称为皮质前处理，对这些区域的处理将在第3节中描述。​​视觉皮层通常分为三部分（Fig.2和Table 1中）：枕部给背侧和腹侧流提供输入。枕部包括V1-V4区域和MT。这些区域被视网膜性质地组织，即，视觉皮层中的附近神经元具有附近的接受场（参见Table 1，第六列），并且接受场的大小从V1增加到V4（参见Table 1，第三列）。有强烈的迹象表明，这些领域在处理视觉信息的不同方面时要计算通用场景表示[84]。然而，在不同级别编码的特征的复杂性随着层级的级别而增加，如将在第4节中详细描述的。还有值得注意的是，枕部的大小超过其他两个部分，占据超过62％的视觉皮层，相比于22％的腹侧和11％的背侧通路占比[44]（见Table 1，第二列）。在下文中，我们称为在枕部部分早期视觉中建立的功能过程，指示在复杂特征结构中执行通用场景分析。

腹侧路径覆盖了涉及对象识别和分类的区域TEO和TE，其接受领域的大小也通常明显大于枕部。在区域TEO中存在在区域TE中没有观察到的弱的视网膜组织（rt）。神经元的接受场通常包括中央凹（具有最高空间分辨率的视网膜的中心部分）。在腹侧路径中，特征的复杂性增加到特定对象类（例如面）的对象级别[127]；然而，大多数神经元响应在这些对象级别以下的特征，指示使用这些编码符中的多个来编码对象和场景的编码方案[172]。

背侧路径由运动区域MST和后顶叶皮层中的视觉区域组成。背侧流从事空间分析和行动规划。类似于腹测流，接收场的大小沿着背侧路径增加，并且刺激特征的复杂性逐渐增加（例如，从MT中的简单运动到MST和腹部内部区域（VIP）中的更复杂的运动模式）。此外，接受视野与视网膜位置的关系变弱；相反的，较高的区域编码空间或头部固定坐标中刺激的位置。

除了分为两个路径（腹侧和背侧），值得注意的是还有两个流要区分：巨细胞（M-）和细胞（P-）流[73]。这种区别已经存在于神经节细胞水平，即存在于视网膜的输出水平。P神经节细胞是颜色敏感的，具有小的接受场并且负责中心视野中的高视力。M神经节细胞具有比P神经节细胞更低的空间但更高的时间分辨率。P和M细胞之间的区别通过LGN到整个视皮质。对于第一个近似，P路径被认为是负责形状和对象感知，而M路径可以解释运动和突变的感知[84]。此外，从中央凹到视觉外围的强烈的空间变化分辨率携带通过视觉皮层的大多数区域。

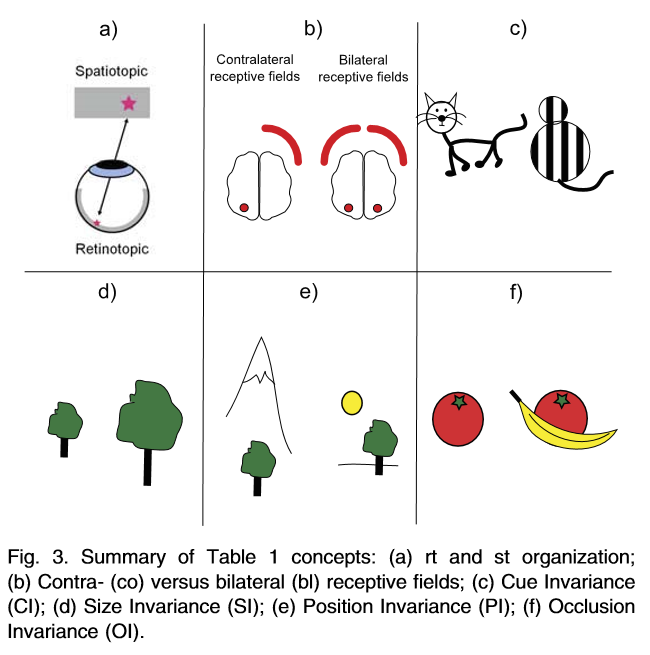
值得注意的是，在视觉层级中的每个阶段，神经元还表现出存在于层级的较早阶段的选择性（例如，可以观察到知道TEO层次的方向选择性）。

一般认为，外在信息对大脑中视觉表示的影响随着其在等级中的水平而增加。例如，没有关于视网膜中的任何学习或适应过程的报告，以及相当一些关于对V1中的定向图的遗传预构造的高影响的证据（参见，例如[63]）。另一方面还已经表明，学习可以改变神经元的视觉特征选择性。然而，通过学习诱导的单细胞水平的可测量变化在视觉分级结构中较早的水平（例如V1 [154]）与稍后的阶段（例如V4 [139]或IT [104]）相比要小得多。

## 2.2 不同视觉区域的基本情况

Table 1给出了视觉系统不同区域的基本数据。第一列指示区域的名称，第二列指示以mm2为单位的大小（也参见Fig.2，其中区域与它们的区域大小成比例地绘制）。第三列指示在5度偏心处的平均接受场大小。第四列指示对视网膜上的刺激的第一反应的潜伏期。

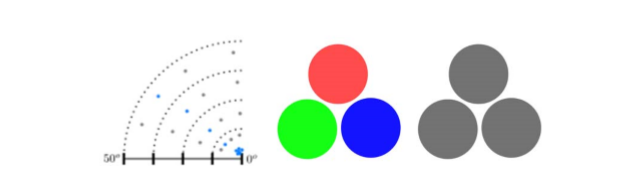
Fig.3提供了在第5列到第7列中的大多数术语的概述。第5列区分了对侧和双侧受体场。对侧（Table 1中的co）仅接受来自一个半球的信息，而双侧（Table 1中的b1）接受场覆盖两个半球（Fig.3b）。第6列表示不同的组织模式：rt组织表示来自视网膜的输入的空间排列被保持，每当我们移动我们的眼睛时改变；st指示在真实世界坐标中的世界的表示（参见Fig.3a ）；聚类组织（cl）表示存在具有相似函数的更大的子区域；列组织（co）表示根据一些组织方案（主要与视觉特征或视网膜相关）存在列中的系统组织。第7列表示不同种类的不变性（参见Fig.3c，3d，3e和3f）：提示不变性（CI）指从不同提示获得相同类型的信息的能力; 对独立于其大小的对象进行响应的单元称为大小不变（SI），类似于PI。最后，与对象相似地响应的单元，不管它是否完全或部分存在，对于遮挡（OI）是不变的。



# 第3章 皮质下视觉系统

在这一节，我们描述灵长类皮质下视觉系统。我们从视网膜光感受器作为视觉处理的第一阶段（第3.1节）开始，并通过外侧膝状核（LGN）遵循来自眼睛的视觉信号（第3.2节）。 对于所有领域，我们首先提供神经生理学，然后是有关其功能上的观点。

## 3.1 基本层次：视网膜光感受器

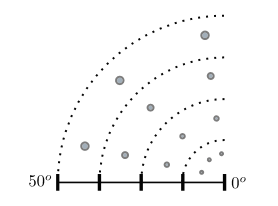


视网膜位于眼睛的内表面上，并且包含仅对电磁谱的特定间隔敏感的光感受器，以及将视觉信息转换为神经信号的细胞。上图左边的象形图表示出了如下所述的杆（灰色）和锥体（蓝色）的空间变化的视网膜密度，以及均匀小的接收场尺寸（大约0.01度的视角）。 这将与我们在以下部分中始终给出的相应象形图进行比较。

神经生理视图。有两种类型的感光器：杆和锥（这里说的应该是眼睛中感光的杆型和锥型细胞）。杆具有对低亮度水平的高敏感度（参见左图）。锥则对高亮度水平较为敏感。我们可以将锥依照它们的波长吸收率的函数进行分类，如S（短波长=蓝色），M（中间波长=绿色）和L（长波长=红色）。这三种锥可以感知颜色[13]。分辨率（即每mm2的受体数目）随着与中央凹的距离而急剧减小。这适用于杆和锥，除了在中央凹中没有杆。大多数锥集中在中心凹中和凹陷周围，而杆在高偏心率下构成感光体的主体。

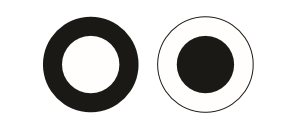
功能视图。因为只有小部分视网膜具有高空间分辨率（中央凹），所以需要凝视控制来引导眼睛，使得感兴趣的场景特征投射到中央凹。因此，灵长类动物具有主动控制眼睛运动的广泛系统（涉及额叶中的FEF，顶叶中的侧脑内区域（LIP）和中脑中的上丘脑）。它受到反射，信号驱动和有意识、认知驱动的注意机制的影响，并且涉及整个视觉等级。注意力模型计算在哪里固定[135]，[142]，一些工作甚至涉及学习控制凝视，例如，以最小化跟踪不确定性[6]。然而，在计算机视觉中，认知驱动的注意机制仍然在很大程度上还未进行开发。

## 3.2 神经节细胞和LGN



从视网膜的光感受器上接收到的信息通过神经节细胞和LGN到达初级视觉皮层。左LGN接收来自双眼的右视觉半视野的输入，右LGN接收来自双眼的左视觉半视野的输入。然而，来自两只眼睛的信息在LGN的六个不同神经元层（四个P加两个M层，三个层接收来自左眼的输入，另三个层接收来自右眼的输入）仍然保持完全分离; 在这个级别上没有进行双眼集成。关于空间分析，在视网膜神经节细胞和它们的LGN对应之间没有显著差异（甚至在视网膜神经节和LGN细胞之间几乎一一对应）[95]。在运动分析中，LGN神经节细胞具有较低的最佳时间频率，在视网膜神经节细胞中，4-10Hz对应20-40Hz，这表明在视网膜神经节细胞中存在低通滤波[95]。在这个层面上出现的两个突出的新特征是中心-环绕接收场和颜色对立。 视觉皮层也被划分成不同的层，其中大多数前馈连接（即，到层级中的较高级的连接）源自表面层，并且大多数反馈连接源自较深层。然而，关于这些不同的皮质层在刺激处理中的作用的研究还不深入。

### 3.2.1 中心-环绕接收场



神经生理视图。具有中心-环绕接收场的亮度敏感单元有两种类型：中心/非围绕单元对黑暗背景上的亮点敏感；偏心/环绕单元对逆模式敏感。两者对均匀亮度不敏感。这些细胞是巨细胞（M）神经元，并参与时间分析。

功能视图。中心-环绕接收场可以通过高斯差分建模，并类似于用于边缘检测的拉普拉斯滤波器[68]。因此它们强调亮度的空间变化。这些细胞也对时间变化敏感，并且形成运动处理的基础。值得注意的是，强调空间和时间变化的表示的变换在非常早的阶段执行，紧接在受体水平之后，在任何其它视觉处理之前。

大多数当前的计算机视觉技术还涉及最早阶段的类梯度计算，这是检测器/描述符的重要部分，例如SIFT，HOG / HOF等。

### 3.2.2 单向对称细胞



神经生理视图。单向对称细胞是颜色敏感的并且计算色差，名为L-M（L对于长波长，M对于中间波长，符号“ - ”表示对立）和S-（L + M）（S表示短波长） 并建立红-绿和蓝-黄色轴。它们具有用于亮度（灰度值）刺激的带通滤波特性，但是对于单色（纯色）刺激具有低通特性。这些细胞是细胞（P）神经元，并且具有比巨细胞神经元更小、更慢的接受场，即更高的空间分辨率。它们对中心视野中的高视力视觉特别重要。

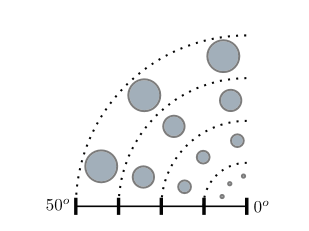
功能视图。可以通过对一个颜色通道中的高斯滤波（例如L）和相对颜色通道中相反符号的另一个高斯滤波（即-M）来对单向细胞进行建模。这会导致每个颜色通道中的低通滤波。颜色对立为亮度变化提供一定程度的不变性，是朝向颜色恒常性的一步。

# 第4章 枕叶皮质中的一般场景表示

枕叶皮质中的所有区域（除了MT）以视网膜方式组织构成，并且以朝向柱作为基本单位（参见Table 1，第6列）。MT区域也是以视网膜方式组织构成的，但具有深度和运动列。需要注意的是，视觉系统不是以严格顺序的层次结构组织的，而在层次结构的层次之间存在“捷径”。存在一个流V1→V2（→V35）→V4到腹侧途径和另一个流V1→V2→MT到背侧途径（Fig.2）。 然而，V4和MT之间也存在交叉连接。

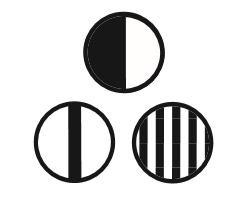
视觉信号的等待时间随着每个层次增加大约10ms，并且接受字段大小逐渐增加（参见Table 1，第3和第4列）。一般来说，巨细胞通路提供背侧视觉流的大部分输入，而细胞旁通路提供腹侧视觉流的大部分输入，但这不是绝对的。

## 4.1 V1区域



V1是处理视觉信息的第一个皮层区域。因此，其特征的敏感程度比在LGN中更复杂，但是保持相对简单：边缘，光栅，线结束，运动，颜色和视差。

### 4.1.1 边缘，条和光栅



神经生理视图。V1包含优先响应边缘，条纹和光栅的细胞单元，即线性定向的图案。它们对于图案的取向以及在光栅的情况下的空间频率敏感（综述参见[127]）。一些细胞对边缘或单条更敏感，而其他细胞更喜欢光栅。有两种类型的这种细胞，简单细胞和复杂细胞。 前者对光栅的相位（或条的精确位置）敏感，后者不是并且具有更大的接收场。

功能视图。Hubel和Wiesel提出的实现复杂细胞的相位不变取向调谐特性的最初提议是简单地增加简单细胞单元沿垂直于其取向的轴的响应，参见[166]计算模型。后来的作者将复杂细胞的行为归因于类似MAX的操作[48]（产生类似于与个体刺激相关的较大响应的振幅的响应，参见例如[140]）或者一个无定向LGN细胞池的非线性积分[115]。在计算模型中，简单细胞单元可以通过优化来自自然图像的稀疏的线性变换来自发组织工作，只有少数单元会在任意给定时间[125]，或统计独立性[8]时有很强的响应。然而，线性模型可能不足以对简单细胞建模[148]。复杂细胞单元可以通过优化慢速的二次变换从图像序列中学习，即单元的输出应该尽可能慢地随时间变化[38]、[10]。一个更加技术性的实验表明，Gabor小波是简单细胞的合理近似，而Gabor正交对的幅度类似于复杂细胞的响应[82]。Gabor小波在图像压缩[29]，图像检索[108]和面部识别[201]等应用中也非常成功。实际上，使用Gabor小波（和V1中的简单细胞单元）的图像统计显示可以用来构建图像的有效编码 [163]。

### 4.1.2 点特征



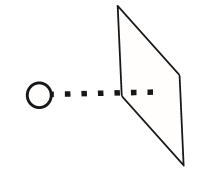
神经生理视图。V1还包含对条或边缘的末端或光栅边界敏感的细胞单元。这样的细胞被称为终止或超复合物[127]。

功能视图。在V1中，终端细胞可以帮助解决系统在运动以及差异处理（参见第4.1.3节）中面临的孔径问题，因为它们也可以沿边缘的方向检测位移[127]。像复杂细胞一样，可以通过优化慢度从图像序列学习超复合细胞[10]。

在计算机视觉中，已经使用了各种类型的感兴趣点检测器（其不受孔径问题的影响，由于它们分析具有不同取向的局部区域的事实）[107]，[116]，这对于匹配任务和许多计算机视觉问题（姿态估计，对象识别，立体声，运动结构等）中的任务是重要的。在这方面有趣的是，V1由用于线性特征（边缘，条纹，光栅）的检测器（简单和复杂细胞单元）支配。一个可能的原因可能是自然场景中的最有意义的特征实际上是边缘，并且也允许输入信号的完全重建（参见例如[39]）。

在这种视觉处理的低水平阶段对点特征敏感的神经元的非常罕见的暗示：对于自下而上的视觉处理来说，初级视觉系统不依赖点特征。基于边缘和线特征的立体和运动处理进一步表明孔径问题不是由V1解决，而是涉及用于空间积分的后续皮质层。

### 4.1.3 绝对差异

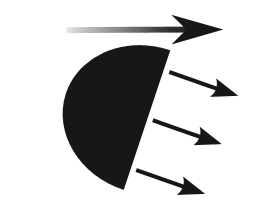


神经生理视图。V1是包含从两只眼睛接收输入的神经元的第一区域[84]（LGN中的神经元仍然是单眼的），并且能够计算视差。在V1中，这仍然是绝对误差（即，参考中央凹的在左视网膜和右视网膜上的点的投影之间的角度差）。计算差异和深度可以在V1中没有单眼轮廓在图像中完成，这从我们解释随机点立体图的容易性中可以发现 [83]。在V1中也存在对反相关立体图中的视差敏感的神经元[26]，其中一只眼睛中的点的对比度极性与另一只眼睛相反。然而，这些神经元不具备深度感知但是可以具有其他功能。

功能视图。V1中视差估计的突出模型是能量模型，它基于具有轻微相位或位置偏移的Gabor小波[50]。误差在早期（在处理和开发方面，见[87]）和在近距离操作上面当然只是深度感知的一个提示，在更高的层次和更远的距离上使用的特征，如遮挡、运动视差等[84]，在灵长类脑的背侧和腹侧视觉流的更高级别的区域中处理（见第4.4节）。此外从发展的角度看，这里的误差与大约6个月后才出现的图像深度线索存在显著差异[87]。这与观察自然场景的统计与知觉组织的规律有关，这是一个由Brunswick和Kamiya [17]首先提出的观点，后来被计算确认（[199]评论）。这种想法表面了一个观点，用来制定在计算机视觉系统中作为统计学习问题导出图像深度线索的问题。视差不仅对于深度感知很重要，而且对于注视控制[84]，对象抓取和对象识别是重要的。通过使相邻单元之间的相互信息最大化，可以从立体图像学习视差调整单元，因为深度是跨空间相当稳定的特征[7]。

在计算机视觉中，立体图像是具有基于点特征的许多方法的研究领域，它们的匹配固定所有自由度（例如，参见[16]）。然而，在计算机视觉中也有使用Gabor小波的相位差的方法[49]。

### 4.1.4 局部运动

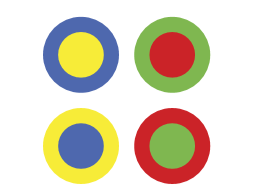


神经生理视图。区域V1和V2中的神经元不仅涉及静态场景分析，而且涉及运动分析。V1中的一小部分简单复杂细胞是方向选择性的，这意味着它们只有在刺激模式（光栅）在一个方向上移动而不在另一个方向移动时才响应[127]。然而，只有复杂的细胞具有时空频率调谐。方向选择细胞属于M-途径并且主要投射到区域MT [118]。在该处理阶段孔径问题没有得到解决。

功能视图。估计运动或光流实际上与估计误差相关，因为后者可以被视为前者的特殊情况：只有两个帧在空间中而不是在时间上移位。计算机视觉中的算法以及V1的模型一般对应地类似为估计视差所讨论的那些（参见第4.1.3节）。对于V1（主要是简单细胞单元），运动处理通常被概念化并由时空接受域建模[194]，[178]。通过优化慢度学习的复杂细胞单位是运动方向选择性的，很像生理神经元[10]。

有趣的是，时空特征例如运动已经被证明是人类发展存在于识别物体（甚至早于颜色和取向）的第一个特征[203]。

### 4.1.5 双向对称细胞

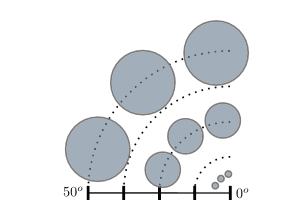


神经生理视图。约5%-10%的V1区域细胞是专用的颜色编码细胞（综述参见[23]、[160]）。除了对应于局部颜色（在蓝色-黄色或红-绿色轴上）的与LGN类似的单向对称细胞之外，V1还具有双向对称细胞。除了不同颜色通道之间的对立之外，这些细胞，其存在曾被辩论并且现在被越来越多的证据支持（例如，[22]）在每个颜色通道内具有空间对立结构。这样的细胞对其对应颜色的背景上的一种颜色的斑点具有特别好的反应，并且被认为在感知颜色恒常性中起关键作用。因此，在V1中已经观察到的颜色对比效应，即刺激的感知颜色偏离背景颜色的变化是不足为奇的[127]。这些细胞的接受场很少是圆形对称的，因此也显示一些定向调谐，但它们的空间分辨率低。一些双对称细胞单元也是取向选择性的。另一方面，简单和复杂的细胞单元虽然不被认为对颜色编码，但是通常对均衡刺激的方向敏感，即仅由颜色对比度而不是亮度对比度定义的边缘或光栅。这表明它们对颜色敏感，但它们不编码的颜色极性只有方向。因此，我们看到的颜色和形式处理在V1中大部分（但不是完全）分离。

功能视图。双向对称细胞单元是形成颜色对比和颜色恒常性的基础，因为它们允许系统在确定所感知的颜色时考虑颜色上下文[23]。有趣的是，通过优化统计独立性可以从自然彩色图像中学习双重感受性领域[20]，这表明它们是由信息优化过程组织的，因此是功能驱动的。

与计算机视觉中的低级颜色归一化相反，计算机视觉主要基于对每个像素以相同方式应用操作（参见例如[47]），从人类颜色感知明显看出，颜色恒常性的实现涉及局部以及跨越层级的所有级别的全局过程，如Helmholtz已经指出的（参见[198]和第7.1节）。

## 4.2 V2区域



V2是视网膜形式组织的区域，大多从V1区域接受输入。在V2中，M和P途径之间的分离尽管不完全，但是很大程度上保留[84]。像V1一样，V2包含调整方向、颜色和视差的细胞单元。然而，V2细胞单元的一部分对相对差异敏感（与在V1中出现的绝对差异相反），这意味着它们表示相对于另一平面的深度而不是绝对深度。V2的主要新功能是更复杂的轮廓表示，包括纹理定义轮廓，虚像（主观）轮廓和边界所有权的轮廓。

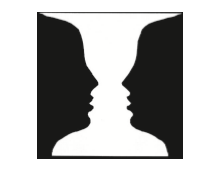
### 4.2.1 纹理定义和虚像（主观）轮廓



神经生理视图。一些V2细胞单元对纹理定义的轮廓敏感，定向调谐类似于亮度定义的轮廓的定向调谐[127]。V2细胞对虚象轮廓也很敏感[84]。 这些可以出现在各种上下文中，包括纹理或视差不连续性，或者关于图像地面效应，例如Kanizsa三角形（参见左边的图标）。

功能视图。这是朝向形状感知的更大不变性的步骤，因为轮廓可以由更多种类的线索定义。

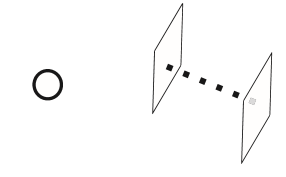
### 4.2.2 边界所有权



神经生理视图。边界（即轮廓）主要由两个或更多个表面的投影形成，所述表面在3D中相交或在它们之间具有间隙。在大多数情况下，这种边界仅属于在边界处相遇的表面中的一个，并且边界所有权属于边界所属的表面（或区域）的分配。边界所有权已被确定为重要的视觉信息[90]，虽然具有不同的术语：belongingness。边界所有权，在视觉计算方法中很大程度上被忽略，但是它对于扩散和填充机制十分重要，可以在很大程度上减少和纠正丢失和模糊的视觉信息。最近发现了对边界所有权敏感的细胞。在2000年，Zhou等人[205]发现，V1中18％的细胞和V2、V4中超过50％的细胞（沿着腹侧途径）根据边界所有者的方向进行响应或编码。然而，我们还不了解这种神经元确定所有权的机制。

功能视图。事实上，边界所有权敏感的神经元在响应开始后10-25毫秒将所有者的方向区分开，并且边界所有权敏感性在V1就有出现（尽管在较小程度上），表明边界所有权可以使用局部线索，可以通过沿着边界的横向长程相互作用而被积分。然而，如Fang[43]等人描述的，该过程也可能受到更高水平的皮层区域的注意调制或影响。

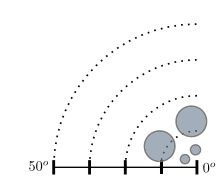
### 4.2.3 相对差异



神经生理视图。V2还包括视差敏感单元。然而，与V1中的视差敏感单元相反，V2中的视差敏感单元对相对视差敏感，相对视差是空间中两个点的绝对差异之间的差。相对视差是，例如，固视平面上的点（零视差）和更接近观察者的点（近视差）之间的视差的差异。立体视觉主要依赖于相对视差的处理[130]。

功能视图。通过对V2中相对视差的敏感性，可以比较对象的深度和3D空间关系形成的原因。

## 4.3 V4区域



与似乎由M途径输入主导的MT（参见第4.4节）相反，V4似乎将来自M和P途径的输入组合，因为阻断M或P途径都会降低V4中的大多数细胞的活性[84]。

V4神经元有选择地响应方向，颜色，差异和简单的形状。它们继续将较低级别的响应集成到较高级别和增加不变性的过程。例如，V4单元响应由速度和/或运动方向的差异定义的轮廓，其方向选择性匹配对亮度限定的轮廓的选择性[127]（在V1和V2中也发现几个这样的单元，但是具有更长的延迟，这再次表明它们由来自V4的反馈驱动）。V4中的突出新特征是曲率选择性和色调的亮度不变编码。

### 4.3.1 曲率选择性



神经生理视图。一些V4单元被调谐到具有特定曲率（具有朝向凸轮廓的偏置[131]）或具有特定角度的顶点[127]的轮廓。该选择性甚至特定于轮廓段相对于所考虑的形状的中心的位置，从而产生以物体为中心的形状表示。V2还具有响应曲线（不是直线的轮廓）的细胞，但是它们的响应可以通过它们单独调整到边缘来解释，这不是V4神经元的情况。

功能视图。在V4被消融的猴子中的实验显示V4对于形式和模式/形状辨别的感知是重要的。V4神经元对可通过曲率角位置函数拟合的群体代码的简单形状做出反应[131]。在该表示中，物体的曲率附接到相对于物体质心的某个角位置。大多数V4神经元代表单独的部分或轮廓片段。

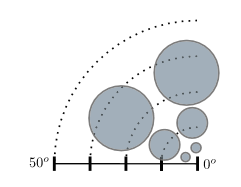
### 4.3.2 色相和亮度不变性



神经生理视图。V4中的颜色编码细胞单元与V2中的编码单元不同之处在于它们编码色调不是沿着两个主色轴的颜色对立，并且对色调的调整不依赖于亮度[24]。即使专门用于颜色，这些细胞中的许多也显示出突出的定向调谐。

功能视图。亮度不变调节到色调已经是颜色恒常性的形式，并且颜色编码单元的取向调谐指示颜色和形式感知之间的一定程度的积分，尽管V4神经元被清楚地分离为两个群体，一个用于颜色，一个用于形式处理 [176]。

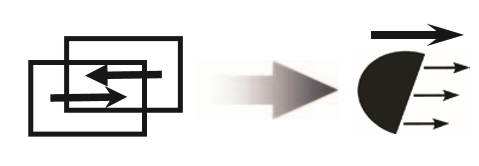
## 4.4 MT区域



中间区域（MT）专用于视觉运动和双目深度处理。区域MT中的绝大多数神经元对移动刺激敏感。神经元被调整到对运动的方向和速度（敏感）[112]。MT神经元接收场比V1大大约10倍，这使得MT神经元可以在更大面积上集成来自V1的运动信号。接收场在接收场的不同部分显示不同运动灵敏度的特征子结构[106]。许多MT神经元也对双眼视差敏感[30]。 MT中的活动直接涉及感知运动[151]和深度[14]判断。MT区域被视网膜形式地组织，具有类似于V1中的方向和眼优势列的运动和深度列。

MT不仅对于知觉而且对于运动控制也很重要，特别是对于平滑追赶眼睛的运动。MT与MST一起在反馈控制回路中提供主要的速度信号[37]，[94]通过输出连接到脑干中的运动结构。

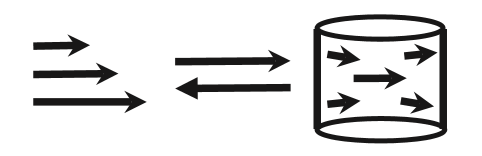
### 4.4.1 2D运动



神经生理视图。MT神经元通过组合响应局部运动的V1神经元的输入来计算运动的中间水平表示[164]，[118]。一些MT细胞可以解决孔径问题，并编码独立于移动刺激方向的运动方向[117]。MT细胞编码速度，而不是V1细胞编码的时空频率[133]。在计算运动信号时，MT神经元遵循粗略到精细的策略，其中对运动刺激的响应是快速的，但是不精确的，并且随时间变得更精细[129]。

功能视图。在局部时空能量的初始测量（在V1中）之后，需要运动测量的组合来求解孔径问题，导出2D运动的方向并估计速度。这导致视场中的运动的中间级表示，它比早先的视觉区域（例如V1和V2）更忠实于真实运动并且对噪声的响应更稳定。在大的接收场上的运动的组合中固有的空间平滑被运动信号的组合中的视差信息部分地减少[99]。

### 4.4.2 运动渐变和运动定义的形状



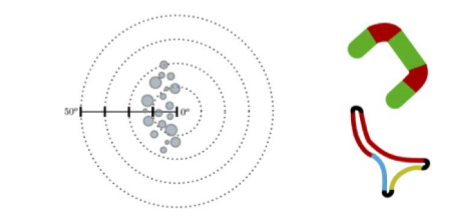
神经生理视图。一些MT细胞对运动的高阶特征（例如运动梯度，运动限定的边缘，局部相反的运动和运动定义的形状）是选择性的[127]。这些选择性由视差灵敏度辅助。视差有助于从不同距离处的对象中分离运动信号，保留运动视差，并计算透明运动和三维运动表面。

功能视图。MT构造运动定义的表面和表面上的运动表示。

# 第5章 目标识别与分类：腹测流

损伤研究已经证明，腹侧通路对于对象辨别是关键的[192]，而后顶叶皮层对于空间视觉是重要的。基于在TEO中但不在TE中的粗视网膜位点的存在，下颞皮质（IT）的最广泛使用的分割在更后部分TEO和更前部分TE区域之间（参见Table 1），在TE区域的神经元大于TEO的接受场大小。在IT中已经鉴定出两种类型的神经元[173]：初级细胞响应特征的简单组合，并且是TEO中的多数；精细细胞响应面部，手和复杂的特征配置，并且在区域TE中具有高的存在。

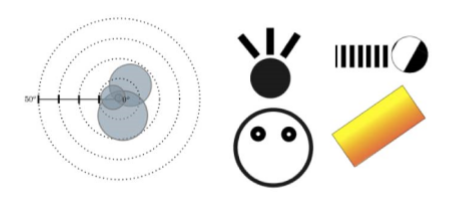
## 5.1 TEO区域



神经生理视图。TEO（也称为PIT for Posterior IT）神经元是定向和形状选择性的。 已经表明的是，TEO神经元主要响应非常简单的形状元件。TEO和TE之间的主要区别是TEO中的粗视网膜组织，其在TE中不存在。TEO神经元的接受场仍然相对较小（3-5度）并且位于中央凹周围或对侧半球体中。

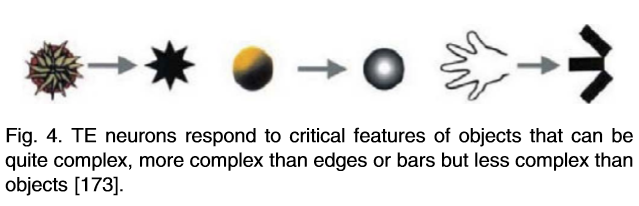
功能视图。TEO负责中等复杂性的特征，它集成了关于多个轮廓元素的形状和相对位置的信息。TEO集成的轮廓元素与V4相比具有更高的复杂性。这种积分是非线性的，它包括抑制性输入（除了兴奋性输入）。形状调整是位置和尺寸不变，并且支持基于部分的形状理论[127]。

## 5.2 TE区域



神经生理视图。TE区域（也称为Anterior IT的AIT）的特征在于相对于腹侧途径中的先前区域，驱动神经元的视觉特征的复杂性的显着增加（第4节）。形状选择性TE神经元整合来自先前区域的输出。TE中视觉神经元的接受场范围为10至20度的视角，平均响应延迟为约70-80ms。

尽管2D形状是TE神经元响应的主要刺激，但是其他对象属性也在TE中进行编码：颜色[182]，视差[182]，纹理[182]和3D形状[81]。至少对于颜色和3D形状，已经证明这些对象属性的处理很大程度上局限于TE中的特定子区域[80]，[183]。



Tanaka[173]通过开发刺激-减少方法做出了关键的贡献（见Fig.4）。在测量TE神经元对真实世界对象的响应之后，他系统地减少最有效对象的图像，以努力识别TE神经元响应的关键特征。对于许多TE神经元，关键特征是中等复杂的，即比整个图像更简单，但比简单的条或点更复杂（Fig.4）。在一些情况下，由关键特征驱动的神经元聚集在可能被认为是皮质柱的位置[182]。这些发现导致了假设：TE神经元不显式地编码整个对象，只有部分对象。因此，TE的读出需要组合来自许多TE神经元的信息以建立显式对象表示。

功能视图。TE神经元的许多性质（例如不变性，参见Table 1，第7列）与视觉对象识别的性质可以良好地对应起来。一些研究表明，TE神经元发烧率的试验变化与猕猴在各种任务中的感知报告相关，包括物体识别[119]，颜色辨别[111]和3D形状辨别[195]。

能够进行对象识别的神经系统必须满足两个看似冲突的要求，即选择性和不变性。一方面，神经元必须区分不同的对象，以通过对区分对象的视网膜图像中的特征的敏感性向系统的其余部分提供关于对象身份（以及在分类的情况下的对象类别）的信息。另一方面，该系统还必须处理相同对象的高度不相似的视网膜图像，因此必须对视网膜图像中在自然视觉中发生的变换敏感（例如位置，亮度，视网膜尺寸的变化等等）。这可以通过导出对于某些变化高度鲁棒的不变特征通过丢弃视觉数据的某些方面（例如SIFT描述符[107]）来实现。然而，从系统观点来看，不丢弃信息，而是表示信息，使得不变量的方面与变化部分分离，可以使两种信息可以被有效地使用（例如参见[31]和第8.2节）。

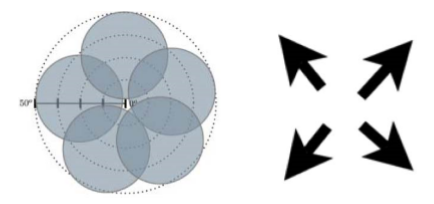
TE神经元通常偏好对大范围的刺激转化的不变性显示形状（尽管通常不是在绝对响应水平）。最广泛研究的TE神经元的不变性包括位置（PI，参见Table 1，第7列）和大小（SI）的不变性，但是其他刺激转化也可以引起不变形状偏好：定义形状的视觉线索 （CI，[182]），部分遮挡（遮挡不变性OI，[182]），深度位置[79]，照明方向[92]和杂乱（重叠形状[182]）。深度旋转引起对象的视网膜图像中最剧烈的变化以及TE中最弱的不变性，因为即使在广泛训练之后，大多数TE神经元也显示强烈的视图相关响应。唯一的例外可能是面孔，其中视图依赖和视图不变响应都已被记录[182]。

TE神经元通常响应同一类别的几个但不是所有的样本，并且许多TE神经元也响应不同类别的样本[197]。因此，在TE中没有明确表示对象类别。然而，最近的读出实验已经证明，可以训练统计分类器（例如支持向量机）以基于少量TE神经元的响应来分类对象[182]，[88]。因此，TE神经元群体可以通过它们的组合活动可靠地对信号对象进行分类。令人惊讶的是，相对较少的视觉训练对单个细胞水平以及fMRI的视觉感知具有显着的生理效应[93]。例如，将对象变形到彼此增加它们感知的相似性，这被认为是学习不变性的有用机制[51]。

# 第6章 动作视觉：背侧流

背侧视觉流（见Fig.2）包含多个从诸如MT和V3A的区域接收的视觉信息的区域，并且主要投射到额叶中的预运动区域，桥接在视觉和运动系统之间。位于背侧流中的区域在功能上与不同的效应器相关：LIP涉及眼睛运动，手臂运动中的内侧腹膜内区域（MIP），手部运动中的AIP（抓握）以及身体运动中的MST和VIP（自运动）。

## 6.1 MST



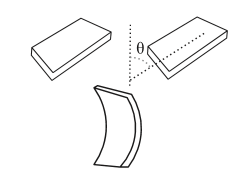
神经生理视图。MST区域从MT接收其主要输入（见Fig.2）。像MT一样，MST有许多响应视觉运动的神经元。MST中的接受场比MT的接受场大得多，通常覆盖视野的大部分而没有清晰的视网膜位置排列。许多MST神经元选择性地响应全局运动模式，如大场扩张或旋转[175]。因此，MST神经元在视野内在不同方向上整合运动。然而，接受场的结构是非常复杂的，并且通常不直观地与图案选择性相关[34]。MST神经元被调谐到光流场中的自运动或前进方向上[132]，[100]。MST神经元携带视差信号[147]和接收前庭输入[15]，[67]，这与他们参与自我运动估计一致。

MST区域也参与平滑追踪眼睛运动[37]，采用非视觉（extraretinal）输入[122]。使用这种视野外信息，一些MST神经元可以消除眼睛运动的视网膜效应，并响应客观世界中的运动，而不是视网膜上的运动[41]。这也可以在区域V3A中看到[57]。

功能视图。MST区域涉及自运动，用于头部（或身体）在空间中的运动和眼睛在头部中的运动。MST神经元对光流模式的选择性产生了基于群体的MST方向图[100]。不是在视域的视网膜图中表示特定特征的分布，而是像V1，V2，V4或MT这样的较低区域，MST创建表示空间中不同方向上的自我运动的新参考系。MST不是按照视网膜形式组织的，但其主题可以表示在视网膜坐标上，即相对于凝视方向的左或右。访问外部眼球运动信息使得MST能够在身体运动和眼睛运动的组合中估计相关运动信息。

实现从光流的自运动的估计是机器人技术中的常见要求。这个问题的解决方案依赖于来自视场的不同部分以及来自与视线方向估计相关的非视觉区域的许多运动信号的组合。

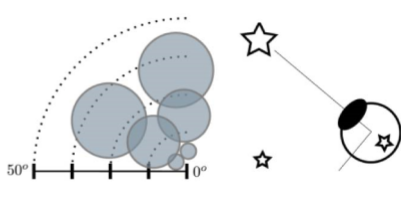
## 6.2 尾部壁内区（Caudal Intraparietal Area，CIP）



神经生理视图。CIP从V3A区域接收预测（投影），并投射到LIP和AIP[121]。在[162]，[169]的报道中，CIP神经元选择性响应由双眼视差（一阶差异）定义的倾斜平面表面。一些CIP神经元对于细长刺激的3D定向也是具有选择性的[150]。CIP神经元可以为平面表面的倾斜显示CI，这意味着当不同的深度线索表示倾斜（视差，纹理和透视[189]）时，CIP保持对特定倾斜的偏好。结果[187]显示出了对CIP神经元的零级差异（深度位置）的选择性。最近，[86]还报道了对一个猴子的曲面选择性（二阶差异）的实验。CIP神经元在眼跳运动期间不反应。实验没有关于CIP接受字段的大小和形状以及CIP神经元的响应等待时间的数据。

功能视图。区分不同层次的深度信息和视差是很方便的[71]。零阶差异是指平面表面的深度位置（或绝对差异，沿着表面没有视差变化，参见第7.3节），一阶差异是指倾斜表面（倾斜和倾斜，沿着表面的视差的线性变化），二阶视差是指曲面（凹面或凸面，表面上的视差变化的变化）。CIP包含编码零阶，第一和可能的二阶视差的神经元，这表明它是可以向诸如LIP和AIP的视觉运动领域提供输入的重要的视觉中间区域。我们对CIP的内部组织知之甚少。

## 6.3 侧顶叶（Lateral Intraparietal Area，LIP）



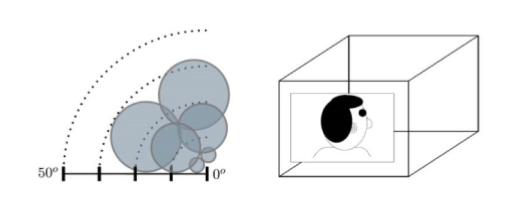
神经生理视图。LIP位于视觉区域和运动系统之间，接收来自背侧和腹侧流的信息并投射到额叶（FEF）和上丘中的其他眼动电位控制中心[103]。LIP神经元在眼跳运动进入接受场之前发生反应，LIP的电微小刺激可以引起眼球运动[180]。

LIP中的视觉反应与刺激的显著性相关[65]，这导致LIP包含视野的显著图，引导注意并决定对相关刺激的扫视[11]。 此外，LIP已经牵涉到其他几种认知过程：决策形成[159]，奖励处理[136]，时间[76]和分类[52]。最近的一系列研究也表明，LIP神经元可以在被动固定中选择性地对简单的二维形状做出反应[155]，这种特性主要分配给腹侧视觉流。

功能视图。LIP中的空间表示揭示了背侧流中的空间处理的几个关键性质。LIP神经元具有代表视网膜上的位置的视觉接受场，即它们表示以视网膜为中心的坐标系中的刺激。然而，在眼跳运动前的几毫秒，一些LIP神经元变得对在其将在其接受场扫视后的位置的刺激敏感[36]。在当前和未来接受场之间的活动的这种重新映射看起来像在扫视之前接受场的瞬态转变。此外，虽然LIP接受场基本上在视网膜中心坐标中，但是细胞的活性由眼睛位置调节，即当动物向右看时，当LIP看起来，一些细胞对它们的接受场中的刺激反应比向左边看时更强烈。反之亦然[4]。以视网膜为中心的接收场和眼睛位置调制的组合提供了LIP中的人口码，其可以表示以头部为中心的坐标中的刺激的位置，即可以执行坐标变换[206]，[137]。这种变换允许视觉与听觉空间输入的组合以用于景物和声音的定位[3]。

LIP是背侧流中研究最多的区域之一。尽管已经有超过二十年的单细胞研究，但是对LIP在高级认知控制过程，如运动规划，注意，决策形成等的作用仍然存在相当大的争议。然而，LIP被认为是行为相关刺激的空间表示的核心领域。视觉（和听觉）输入被变换成空间表示，其中每个神经元使用眼睛为中心的坐标，但是整个群体形成以头部为中心的表示来编码刺激位置，即使当眼睛位置改变时。在单个神经元水平，不管眼睛运动，对跨越等活动的重新映射确保视觉表示的连续性。

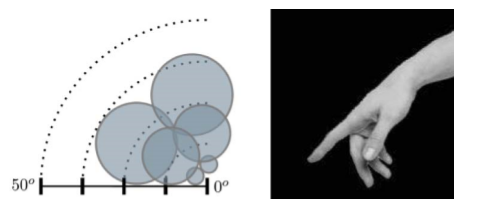
## 6.4 腹侧壁内区（Ventral Intraparietal area，VIP）



神经生理视图。VIP区域与广泛的视觉，体感和运动（嘴部表示）区域连接。VIP神经元是多模态的，它们可以通过视觉，触觉，前庭和听觉刺激以及平滑追踪眼睛运动来激活[21]。触觉接受场通常位于头部和脸部的皮肤上，并且视觉和触觉接受场在尺寸和位置上经常匹配：响应于围绕嘴部的区域的触觉刺激的神经元也将响应于接近嘴部区域的刺激。研究已经提出VIP编码近外部空间[21]。VIP神经元的接受场从纯视网膜中心变化到纯中心[35]，也包括在视网膜中心和头中心之间的接受场。此外，一些VIP神经元响应复杂的运动刺激，例如光流显示器中的前进方向。

功能视图。VIP区域可能涉及自我运动，头部运动的控制，以及链接触觉和视觉领域的近外部（头部中心）空间的编码。

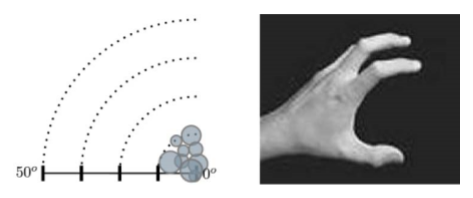
## 6.5 内侧壁内区（Medial Intraparietal Area，MIP）



神经生理视图。MIP主要投射到背部运动前皮层（PMd）。该区域中的神经元通常在延迟到达任务期间选择性地响应，指示猴子在一定时间延迟之后到达触摸屏上的目标以接收奖励。MIP神经元将响应特定的到达方向，这种神经选择性主要是眼睛为中心。当猴子自由选择目标时，MIP和PMd显示增加的尖峰场相干性，表明这些脑区之间的直接通信[134]。

功能视图。MIP神经元的活动主要反映朝向目标的运动计划，而不仅仅是目标的位置或目标外观引起的视觉注意[55]。当动物选择到达时，与动物选择朝向目标的扫视相比，MIP神经元也反应更多，表明MIP编码自主选择的运动计划[25]。

## 6.6 前壁内区（Anterior Intraparietal Area，AIP）

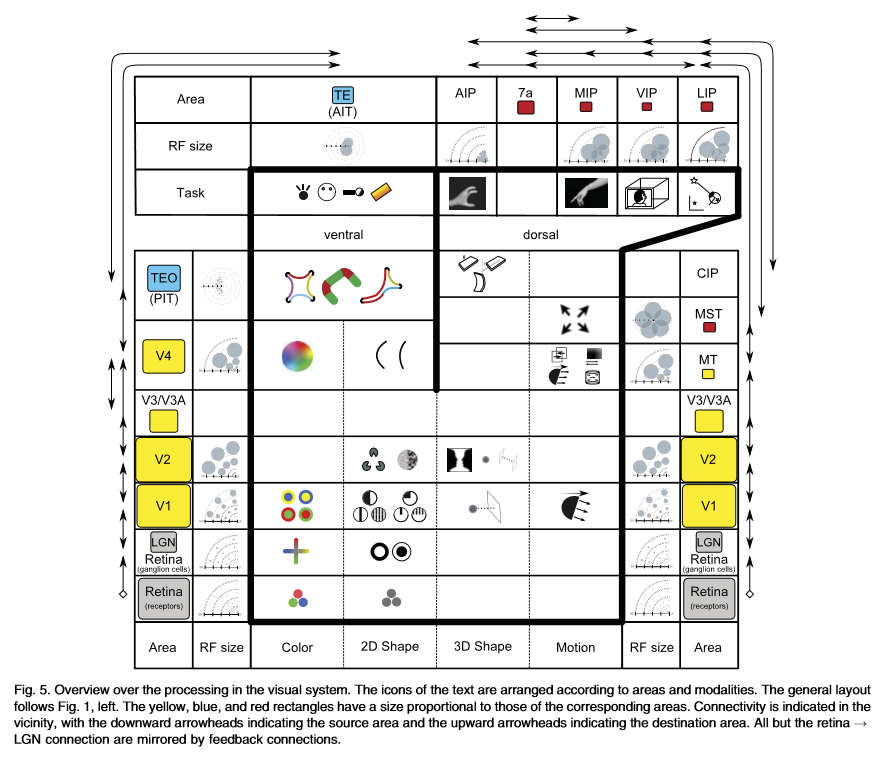


神经生理视图。AIP的主要输入发生在LIP，CIP和腹侧途径[12]，而AIP的输出指向腹侧前运动区F5，这也涉及手部运动。AIP的可逆失活导致对侧手深的抓握缺陷[56]。Sakata[150]等人的研究表明，AIP神经元在对象抓取期间经常放电，并且对一些对象优先。一些AIP神经元在对象固定和抓握期间响应，但不是在黑暗中抓取（视觉优势神经元）；其他AIP神经元在对象固定期间不响应，但只有当对象被抓住时，甚至在黑暗中（优势神经元）才会响应；而第三类AIP神经元在对象固定和抓握期间以及在黑暗中抓握期间响应（视觉运动神经元，[120]）。 AIP对曲面视差定义的3D结构进行编码[167]。然而，猴子的实验表明AIP中3D形状的神经编码与3D形状的感知分类无关[195]。相比之下，大多数3D形状选择性AIP神经元也在对象抓取期间响应[179]，这表明AIP表示3D对象属性是用于抓取（即掌握可见性）的目的。

功能视图。AIP中的神经元对于与抓握相关的手的物体和形状（在明亮或黑暗环境中）的2D和3D特征敏感。换句话说，AIP可能涉及将对象的抓取能力与其2D和3D特征相关联的能力。从视觉信息中提取掌握能力也是目前机器人技术领域的一个高度研究领域，因为拾取未知物体是自主和服务机器人技术中的一个常见任务。

# 第7章 垂直视图：处理不同的视觉模式

基于第3,4,5和6节中我们在涉及视觉信息处理的大脑区域获得的知识，现在我们可以总结对不同视觉模式的处理，如颜色（7.1节），2D和3D形状（7.2节和7.3节），运动（7.4节），以及对象识别处理（7.5节）和行为（7.6节），这一部分强调对视觉信息处理的层次方面。Fig.5给出了该垂直（每种模式）以及水平（每个区域）视图的概述。



## 7.1 颜色

颜色是一个具有很多信息的提示，并一直被用于心理和物理视觉搜索实验的基本特征之一。一旦它们在颜色空间中被广泛分离，就可以用异质颜色（多达九种干扰）执行有效搜索[202]。

在神经生理学上，颜色处理的特征在于朝向颜色恒定性的稳定进展（参见Fig.5，第3列）。三种锥型细胞（L，M，S）具有宽泛且大部分重叠的调谐波长，并且它们的发射率很大程度上受到亮度的影响。LGN中的单向对称细胞单元建立红-绿和蓝-黄的两个颜色轴，从而锐化波长调谐并实现一些不变的亮度。双向对称细胞单元提供了考虑颜色对比度的附近颜色的手段。在V4中，色调是编码的，并延展至整个颜色空间。最后一步是IT，其中存在颜色和形式的关联[204]。在TEO（更接近V4）中，大多数神经元通过特征的简单组合被最大化激活，例如尺寸，方向和颜色不同的条形或盘形组合[174]。精细细胞（TE中的大部分）对不同特征（形状和纹理，形状和颜色，纹理和颜色，纹理、颜色和形状）的组合做出响应[174]。

有一些相关的见解可以从前面部分中提到的神经生理学证据中得出。颜色处理发生在一个很大程度上分离的路径中，该路径仅合并在TE水平上的一般形状表示。颜色是计算机视觉可以利用的易获得但也脆弱的特征。它对于对象识别的有效使用取决于实现颜色恒定性，这仍然可以被视为计算机视觉应用中的挑战。在灵长类动物视觉系统中，这仅在相当晚的阶段（V4及以后）才实现，因此它涉及到视觉层次的大部分。这与计算机视觉中占优势的像素级的颜色归一化方案非常不同。分层表示可能能够提供用于提供中级和高级提示以实现颜色恒定性的手段。

## 7.2 二维形状

通过增加接受场尺寸，增加相关特征的复杂性和增加不变性程度，在整个视觉系统中表征2D形状的处理（参见Fig.5，第4列）。

接受场尺寸在视网膜中很小，但是在IT中可以大到视野的一半（见Table 1，第2列）。但它不仅是尺寸增加; 接受的领域也变得更加复杂和动态。在早期，感受性领域倾向于显示线性响应。从V1开始，细胞具有非经典的接受场，即这些细胞的响应由周围环境调节，这意味着上下文效应。在V4中，强烈的注意效应已经显示，导致如果任务和注意状态改变，细胞对相同刺激会有不同反应[84]。在IT中，接受场非常大，但是场景中的杂波和其他物体会影响选择性[146]。对于空白背景上的孤立对象，接受字段非常大；对于杂乱的背景上或许多干扰物中的对象，接收场相对较小，这暗示了检测和分割之间的紧密连接。

驱动视觉系统中的细胞的特征也逐渐增加复杂性。最开始，它们只是视网膜和LGN中的斑点，是V1中的条或边缘，V4中的特定曲线，然后是TEO和TE中更复杂的图案和对象部分。一般的概念是，更复杂的特征可以从更简单的特征构建，例如，对条纹敏感的简单细胞可以从对斑点敏感的细胞和偏心细胞组合而得到。

早期的视觉系统的确试图对频繁发生的做出不变的反应但不相关的变化[200]。这已经开始在视网膜中，其中几种机制实现高度的亮度不变性，使得我们可以在夜晚的黑暗和晴朗的日子的亮度中看到。首先通过复杂细胞在V1中实现一些位置不变性（PI），其对某些取向的条和边缘敏感，如简单细胞，但是对刺激的位置不太敏感。PI在整个视觉系统中增加，并且在IT中，对象可以移动10度或甚至更多，而不降低一些IT细胞的选择性[75]，[184]。还存在尺寸不变性（SI）的增加。除了照明的不变性和几何变换之外，还针对用于定义对象的线索实现不变性（见Table 1，第7列）。边缘似乎是用来表示对象的主要特征。在V1中，它们被定义为暗和亮之间或不同色调之间的边界；在V2中，轮廓也可以由纹理边界限定，并且这些单元响应虚幻轮廓；在V4中，轮廓甚至可以由运动的差异限定。

表示和识别2D形状不仅仅需要边缘的集合。边缘必须以某种方式集成到一个相干感知中。这在神经科学中称为约束问题[185]，[84]。这种观点认为必须有一种机制，将基本特征绑定在一个对象上，否则将混合一个对象的特征与另一个对象的特征，并感知不存在的东西（这实际上发生在人类遇到快速显现时间的情况下[186]）。绑定问题的可能解决方案是将细胞调谐到特征的连接，空间注意和时间同步。后者的想法假设视觉系统管理同步的那些神经元的代表相同的对象，并使它们与其他神经元同步[165]，这也可以解释为何我们可以同时处理的对象数量相当有限。绑定问题与分割有关，同样表示边界所有权的响应，如在V2中，以及特定于边缘相对于对象中心的相对位置的响应（如V4中）可能与这两个进程相关。

## 7.3 三维形状

大脑从大量的深度线索计算第三维（深度）。双目视差是最强大的深度线索之一。重要的是，只有二阶差异（见6.2节）与眼睛位置（聚散角）和距离无关[71]，从而构成一个非常鲁棒的参数来估计环境的3D布局。

3D形状的神经表示在视觉系统中逐渐出现（参见Fig.5，“3D形状”列）。可以确定几个一般原则。首先，在视觉系统逐渐更高的阶段，从V1的绝对差异开始，神经元变得调谐到更复杂的深度特征[27]。沿着腹侧流，对V2的相对差异[181]，V4[70]的一阶差异以及IT中的二阶差异[70]，[80]出现了新的选择性。沿着背侧流，V3和V3A主要编码绝对差异[5]，MT编码绝对、相对和一阶差异[97],[190],[123],CIP主要编码一阶差异[189]，AIP主要编码二阶差异[167]。与每个其他视觉特征表示一样，神经元的接受场变得更大，并且延迟变得更长。第二，在层次中的每个级别上，重复先前级别的神经选择性，使得在各个级别（例如，IT皮层）中的最高级别，对于零阶，一阶和二阶视差的选择性是可以测量的[81]。

第三，在视觉层次中，似乎存在相当多的3D形状信息的并行处理。因此，腹侧和背侧视觉流（AIP）的末期区域各自包含3D形状的单独表示[79]，[167]。这些表示是不同的，因为3D形状选择性神经元的属性在IT和AIP之间显著不同：AIP中的3D形状的编码更快（更短的延迟），更粗糙（对表面中的不连续性的敏感性较低），更不分类，更加基于边界（表面信息的影响较小）[179]，[77]。最后，两个神经表示变得更加适合两个处理流支持的行为目标：在IT中，3D形状表示支持3D形状的分类[196]，但在AIP中，大多数3D形状选择性神经元也在抓取期间做出反应[179]。相反，在V1[26]，MT[96]和MST[170]中存在抗相关差异的选择性（其中一只眼睛中的每个黑点对应于另一只眼睛中的白点，并且没有深度信息感知），在V4[171]中较弱，在IT[78]或AIP[179]中不存在，可能是因为后者的区域不参与眼球运动，而这种运动受到反相关差异的强烈调制[110]。

## 7.4 运动

当运动通过环境时在视网膜上诱导的运动模式提供关于运动和关于环境的结构的信息[61]。运动路径从光流提取该信息。

V1中的运动分析的第一步涉及从视网膜图像的动力学计算局部时空运动能量[194]，[178]。MT中的中间级表示基于V1输入计算基本运动特征，例如2D方向和速度[164]。这个计算需要解决几个困难的问题。首先，V1中的时空接受场的局部运动能量计算仅测量垂直于移动条或光栅的取向的方向（孔径问题）。第二，时空接受场不能计算速度，而只能计算时空频率。速度调整对应于时空频率中的取向。大多数V1神经元响应空间和时间频率的特定组合，而真正的速度调节神经元响应于在空间和时间频率s.t：v = df/dt的范围上的优选速度v。这两个问题可以在MT中通过整合来自许多不同的V1细胞的信号解决[164]，[133]。然而，V1中的一些复杂细胞也已经被发现可以解决这些问题[128]。

如Fig.5（第6列）所示，执行该整合所需的空间需要MT中较大的接受场，因此具有空间平滑运动模式的效果。然而，这种平滑很好地适应于光流的结构，并保留自运动信息[18]。此外，MT接受场内的输入的不同加权允许计算新的运动特征，例如差分运动（视场中相邻位置处的运动方向之间的差异），运动边缘和梯度[127]。这些较高阶运动信号与场景中的表面的性质直接相关。携带该信息的重要信号是运动视差，即，在距移动观察者不同距离处的两个物体的速度差。MT神经元对运动边缘和局部相反运动的灵敏度可用于从光流提取运动视差。运动处理与MT中的视差分析相结合，从不同深度分离运动信号[99]。

关于自运动的信息的提取是MST区域的功能。MST神经元具有非常大的双边接受场并且响应运动模式。这些模式包括膨胀，收缩，旋转，更一般来说的话，螺旋[175]，[34]。因此，一种观察MST的方法是模式分析。然而，MST在自运动分析方面更好理解[100]。自运动描述了观察者的眼睛在空间中的平移和旋转，即任何刚体运动6个自由度。单个MST神经元被调谐到特定的自运动，即，到特定的平移方向（例如，向前或向右）和旋转，以及旋转和平移的组合[100]，[67]。 MST因此包含自运动的表示。

运动处理链接到平滑追求的眼睛运动。当用眼睛跟踪移动目标时，目标在视网膜上稳定，而背景正在扫过视网膜。然而，目标被感知为移动，并且背景被感知为稳定。MST中的一些细胞通过组合视觉信息与关于正在进行的眼睛运动的远视信息来响应世界中的运动而不是视网膜上的运动[41]。这种视觉和外部信号的组合对于自我运动分析也是有用的，特别是当人们不注视运动的方向，但是固定在视野中并且跟踪视野中的对象时[99]。关于自运动状态的前庭输入也与MST中的视觉效应相结合[67]。

总之，灵长类视觉系统中的运动分析以从V1（局部时空滤波）到MT（2D运动）到MST（自运动，世界坐标中的运动）的层次结构进行。沿着这个层次，几个计算问题被解决，特征变得更复杂，接受场变得更大，并且运动信号的空间整合程度增加。该表示从视域（V1，MT）中的运动之一移动到世界中的运动和世界中的运动（MST）之一。同样沿着该层次，视觉运动处理与视差（MT，MST），眼睛运动信息（MST）和前庭信号（MST）组合。因此，该表示变得较少地束缚图像并且更多地束缚身体的动作。

## 7.5 对象识别

对象识别在几个方面超越了简单的二维形状感知：不同的线索和模态的整合，不变的深度旋转和关节运动，使用上下文。区分类间辨别（对象分类）和对象的类内辨别也很重要。

关于2D形状的感知已经完成了一些不同线索的集成。例如，可以通过V1中的亮度，V2中的纹理以及V4中的运动差异来定义边缘。然而，颜色和形状似乎被相当独立地处理，直到在高级的层次结构中。运动在早期被处理，但是它以与形状感知不同的方式用于对象识别。例如，人们可以通过他们的特征步态识别远距离的熟悉的人。其他模式，如声音和气味，显然也有助于对象识别。

IT中的单元将中间复杂性的各种特征从腹侧流中的较低级中整合到一起以构建对象部分的模型。这些部分的精确粒度目前还没有建立，虽然有迹象表明它们跨越不同大小的接收场，并且可能调谐到不同水平的特征不变性（抽象）[173]。可以主要被描述为组成层次结构（分类表示的分层组织）的计算模型定义/学习与这些发现不矛盾的单元。例如，已经表明，在大致对应于IT级别的已经学习的（以无监督的方式）的特征包含用于对象类别的可靠分类的足够信息（这可以与读出实验相关[74]）。一些相关的计算模型也可以帮助做出关于大量反馈（从IT到LGN/V1）的需要的预测，并缓解刺激减少方法的问题，因为这些刺激可以通过学习过程产生[46]。

深度的旋转通常会相当剧烈地改变对象的形状。然而，IT神经元的一小部分可以表现出一些旋转不变性，并且熟悉对象的识别速度不依赖于旋转角度[79]。特定情况是面部敏感的神经元，其可以对深度旋转显示相当大的不变性。在不同角度下的相同对象的表示，可能被组合成类似简单细胞反应的旋转不变表示可以被组合成复合细胞反应。从不同角度比较不熟悉的对象似乎需要心理旋转，需要额外的时间与旋转角成比例[153]。

上下文在对象识别中起着主要作用[124]，并且可以具有不同的性质——语义，空间配置或姿态，并且至少部分地由IT之外的更高区域提供。一个简单的例子是单词“THE”和“CAT”，它们可以在中心用相同的字符写成在“H”和“A”之间的形状。我们识别这些同一的形状所采用的适当的方式取决于周围两个字母的上下文。有趣的是，对象也有助于识别上下文，上下文可以在粗略的统计级别上定义[124]。

一些人具有完全良好的对象识别能力，但不能识别面部，称为prosopagnosia，虽然他们可以通过他们的衣服或声音识别人。 FFA（梭形面部区域）似乎是具有面部识别功能的脑结构[85]。有证据表明，prosopagnosia不仅影响面部处理，而且从同一类别分离实例的一个缺陷。例如，一个具有prosopagnosia的鸟类观察者者不能分辨鸟，汽车专家不能做出精细的汽车分辨[58]。

有趣的是，在人类受试者中已经描述了可以支持对象识别的高选择性神经元。例如，癫痫患者在内侧颞叶的记录显示单个神经元可靠地响应特定的对象，如无论什么图像中的比萨斜塔[138]。

## 7.6 动作表示支持（Action Affordances）

为了向动作的规划和控制提供可视信息，视觉系统沿着枕部和背侧路径提取分级处理中的特定动作相关特征。该处理的特征在于逐渐增加复杂性，多感觉集成，以及从一般视觉表示向特定效果器和动作特定的表示的转变。此外，这种处理在某种程度上独立于有意识感知，使得病变患者可能能够与他们不能识别的对象正确地交互，反之亦然。

背侧流层级（V1，V2，MT）的早期阶段涉及视觉特征提取（位置，取向，运动）以及来自不同线索的动作相关对象特征（例如表面取向）的估计（运动：MT，立体声：CIP）。这些特征被编码在视网膜形式组织的参考系中。层次上更高的区域在空间或头中心参考帧中编码信息，有时在单细胞水平（如在区域VIP[35]中），通常在群体细胞水平上（区域MST，LIP，7A，MIP）编码[137]。因此，背侧流的主要功能在于坐标变换。

这些变换是必要的，因为使用不同效应器的动作规划需要考虑不同参考帧中的目标。眼睛运动最好以视网膜中心表示来编码，但到达运动需要转换到手臂坐标，因此需要转换到空间中的目标表示。我们并不总是清楚什么是特定动作的最佳编码，但是顶叶皮层中的区域为不同的任务提供了许多并行编码。

这些变换的另一个问题在于视觉与其他感觉或运动信号的组合。在背侧流中的处理中，视觉信息是前庭（在MST，VIP中），听觉（在LIP中），体感（在VIP中）和本体感觉或运动反馈信号（用于平滑眼睛运动的MST和VIP，扫视，MST / VIP / 7A / MIP眼睛位置）等多种信号的组合。由于这些信号具有不同的感觉表示，所以与视觉的组合需要大量的空间变换。

最终，背侧流中的较高区域构造专门用于提供用于特定动作的信息的空间表示：LIP表示视觉场景中的突性，作为用于眼睛运动的目标信号，MIP和AIP提供用于到达（目标信号）和抓取（形状信号）的信号。LIP和VIP提供用于控制自运动的信息。因此，背侧流中与动作相关的视觉信息的处理的特征在于功能的分离，这不同于在腹侧流中的处理，后者集中于对象的感知。

# 第8章 我们可以在视觉系统中为计算机视觉学习什么？

我们可以从灵长类动物的视觉系统中学习计算机视觉系统以及深层次的学习吗？我们认为前者的至少四个设计原则对于后者是有利的：分层处理，信息通道的分离，反馈，以及在先前编码结构和学习之间的适当平衡。

## 8.1 分层处理

灵长类动物视觉系统的一个突出特征是其分层结构，其由许多区域组成，可以在序列中大致排序，首先是共同的早期加工，然后是分裂成两个相互作用的途径；见Fig.2和Fig.5。每个路径逐渐计算更复杂和不变的表示。这种架构的可能优点是什么？

计算效率。大脑是具有大量相对简单和缓慢的处理单元，神经元的机器。因此，如果计算系统有效地分布，就能实现显著的性能。视觉分层网络在空间上和顺序上都是这样。空间分割导致局部化的接受场，并且顺序分割导致逐渐计算越来越复杂的特征的不同区域。因此，计算被大量并行化和流水线化。在PC上，这不是一个问题，因为它只有一个或几个非常快的处理单元。然而，这可能随着GPU或其他计算机体系结构的出现在将来改变，然后层级网络的高度并行化可能变成一个真正的优势。

灵长类动物视觉系统中的计算效率也源自许多被重复用于若干不同目的的处理这一事实。枕部构成视觉皮层的大部分，提供了用于对象识别，导航，抓取等的通用表示。这节省了大量的计算空间。

学习效率。同样重要的是，在推理过程中的计算效率即是学习效率。分层处理有助于提供已经被证明在某些任务中有用和鲁棒的几个不同级别的特征。学习新任务可以基于这些并且可以快速实现，因为在相对高级别的适当特征已经可用。例如，不变性属性可以简单地从特征继承，并且不必再次学习。

分层处理，特别是与感知场尺寸的流行（参见Table 1，第3列）相结合，提供了可以缓解过度拟合问题的机制。即，较低层级中的小尺寸接收场限制了接收场内的特征的潜在变化性，并且因此限制单元来学习低维特征，其可以用相对少的训练样本来采样[46]。该过程递归地应用于整个层次结构，导致较高层上的单元的总体复杂度的受控进展。这对应于隐式正则化。

重要的是，生物视觉系统在充分考虑复杂度和复杂的发展过程（通过增长神经基质）和学习（调整神经单元）等一系列阶段中发展成熟。从计算的角度来看，这有重要的影响，值得在未来更多的关注。

世界是分层的。在大脑中，视觉系统是极端的，因为它有这样深的层次结构。这可能与视觉问题的复杂性或重要性视觉有关。但它也可能是一个事实的结果，即我们周围的（视觉）世界是具有空间布局和层次结构的。对象可以自然地分成部分和子部分，复杂的特征和简单的特征，这使得分层处理有用。视野中的附近点比远点更相关，这使得在有限的接受场中的局部处理在较低层次水平有效。

## 8.2 信息通道的分离

视觉系统的另一个显著特征是信息通道的分离。颜色，运动，形状等在它们集成在较高区域之前，甚至在单独的解剖结构中被分开处理相当一段时间。这些特征中的一些甚至在背侧和腹侧途径中重复，但具有不同的特征并用于不同的目的。我们认为这至少有两个原因：信息的可用性和表示的效率。

信息的可用性。根据实际情况，某些信道可能无法随时使用。如果我们看照片，深度和运动信息不可用。如果它是黑暗的，颜色不可用。如果它是模糊的，高分辨率形状信息不可用，并且运动和颜色可以是更可靠的线索。如果一个线索丢失，那么同时集成所有线索的表示将被严重损害。分离信息信道提供了关于不同信息提示的可用性的鲁棒性。

表示的效率。分离信息通道自然导致了阶乘码；集成表示将产生已知遭受组合爆炸的组合代码，并且也不能很好地推广到新对象。如果我们分别表示四种颜色和四种形状，我们可以更有效地表示16个不同的对象，即与如果将每个对象表示为唯一的颜色/形状组合相比，将具有更少的单元。而且，如果我们只看到16种可能的组合中的几种，我们仍然可以容易地学习和表示未看见的组合。

已经提出的是，不同的神经元处理相同对象的不同视觉特征（例如，颜色和形状）而出现的绑定问题可以通过在时域中的神经元同步来解决[40]，[145]。在这种“同步绑定”假设中，贯穿皮层的相同对象的编码特征的神经元将显示同步活动，其将充当指示不同特征属于同一对象的“标签”。然而，同步假设的实验支持已经被混合到[33]，[98]，[158]，没有实验明确证明同步是必要的绑定。

## 8.3 反馈

虽然我们在本文中概述了视觉处理的层次前馈视图，但重要的是，在视觉皮层内，通常有比前向连接更多的反馈连接。侧向连接也起着重要的作用。

这暗示了诸如注意，期望，自顶向下推理，想象和填充的过程的重要性。许多计算机视觉系统试图以纯粹前馈的方式工作。然而，视觉本质上是模糊的，并且可以从任何现有的知识中获益。这甚至可以暗示比萨斜塔的外观如何影响V1的水平上的边缘的感知。这也意味着系统应该能够产生同时考虑并且可能未解决的几个假设[102]。

## 8.4 视觉处理层次结构的发展与学习

在本文中，我们专注于讨论从最终产品，成人灵长类动物的功能视觉系统得到的描述和经验。我们在这里没有空间来讨论关于生物视觉处理层次结构的发展[193]和学习的知识（例如，[105]，[113]）。然而，有一些相当明显的结论与计算机视觉有关。

首先，与最深层次的计算机视觉系统相反，灵长类视觉处理分层结构不包括从下到上或以统一方式训练的类似层的同类堆栈。相反，它由异构和专门的（水平）层和（垂直）流组成，它们的功能有很大不同。因此，概念上简单的通用视觉系统可能是不可实现的。可能生物学已经选择优化专门的功能和把它们融入感知的整体。然而，仍然需要看到的是，皮层区域的专业化于否是由于根本上不同的机制，或由于输入和特定组合非常小的学习原则的差异造成的（例如[31]，[91]）。

计算机视觉应当感兴趣的这些异质层和流的一个方面是这些不同的功能单元和中间表示为分层学习系统的设计提供结构指导。如Bengio[9]所讨论的，这构成了将巨大的端对端学习问题分解成一系列更简单的问题的方式（见引言）。

其次，生物视觉系统由于遗传编码的结构偏差和暴露于视觉信号之间的相互作用而产生。人们可能会认为，这正是今天的计算机视觉系统的构想：计算程序是手动设计的，其参数使用训练数据调整。然而，非均匀处理层次需要在各个阶段的专用学习方法。成人皮质可塑性的成熟证据表明学习对皮质处理的影响比调整固定神经结构内的突触强度更深远[62]，[66]。

# 第9章 总结

我们主要在与视觉处理相关的功能水平上回顾了关于灵长类视觉系统的基本事实。我们认为，作为原则的证明和建立人工视觉系统的灵感来源，视觉系统仍然是非常有价值的。我们特别主张在较低层次上分离信息通道的分层处理。此外，可以从生物模型中激发对于深层次结构（例如中间表示的结构，层的数量和层之间的基本连接结构）至关重要或潜在地促进深层次结构学习的具体设计选择。然而，主流计算机视觉似乎遵循与我们从灵长类中知道的设计原理完全不同的设计原理。我们希望本文提供的审查和想法有助于重新考虑这一总体趋势，并鼓励开发灵活的多用途视觉模块，从而为人工视觉系统的分层架构做出贡献。

# 致谢

这项工作由欧盟项目Xperience（FP7-ICT-270273）资助。作者还要感谢与Michael D'Zmura进行了富有成果的讨论。

# 参考文献

[1] Y. Amit, 2D Object Detection and Recognition: Models, Algorithms and Networks. MIT Press, 2002.

[2] Y. Amit and D. Geman, “A Computational Model for Visual Selection,” Neural Computation, vol. 11, no. 7, pp. 1691-1715, 1999.

[3] R. Andersen, A. Batista, L. Snyder, C. Buneo, and Y. Cohen, “Programming to Look and Reach in the Posterior Parietal Cortex,” The New Cognitive Neurosciences, M. Gazzaniga, ed., chapter 36, second ed., pp. 515-524, MIT Press, 2000.

[4] R. Andersen and V.B. Mountcastle, “The Influence of the Angle of Gaze upon the Excitability of the Light-Sensitive Neurons of the Posterior Parietal Cortex,” J. Neuroscience, vol. 3, no. 3, pp. 532-548, 1983.

[5] A. Anzai, S. Chowdhury, and G. DeAngelis, “Coding of Stereoscopic Depth Information in Visual Areas V3 and V3a,” J. Neuroscience, vol. 31, no. 28, pp. 10270-10282, 2011.

[6] L. Bazzani, N. Freitas, H. Larochelle, V. Murino, and J.-A. Ting, “Learning Attentional Policies for Tracking and Recognition in Video with Deep Networks,” Proc. 28th Int’l Conf. Machine Learning, pp. 937-944, L. Getoor and T. Scheffer, eds., June 2011.

[7] S. Becker and G.E. Hinton, “A Self-Organizing Neural Network that Discovers Surfaces in Random-Dot Stereograms,” Nature, vol. 355, no. 6356, pp. 161-163, 1992.

[8] A.J. Bell and T.J. Sejnowski, “The ‘Independent Components’ of Natural Scenes Are Edge Filters,” Vision Research, vol. 37, no. 23, pp. 3327-3338, 1997.

[9] Y. Bengio, “Learning Deep Architectures for AI,” Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1-127, 2009.

[10] P. Berkes and L. Wiskott, “Slow Feature Analysis Yields a Rich Repertoire of Complex Cell Properties,” J. Vision, vol. 5, no. 6, pp. 579-602, http://journalofvision.org/5/6/9/, 2005, doi:10.1167/5.6.9.

[11] J.W. Bisley and M.E. Goldberg, “Attention, Intention, and Priority in the Parietal Lobe,” Ann. Rev. Neuroscience, vol. 33, pp. 1-21, 2010.

[12] E. Borra, A. Belmalih, R. Calzavara, M. Gerbella, A. Murata, S. Rozzi, and G. Luppino, “Cortical Connections of the Macaque Anterior Intraparietal (AIP) Area,” Cerebral Cortex, vol. 18, no. 5 pp. 1094-1111, 2008.

[13] J. Bowmaker and H. Dartnall, “Visual Pigments of Rods and Cones in a Human Retina,” J. Physiology, vol. 298, pp. 501-511, 1980.

[14] D.C. Bradley, G.C. Chang, and R.A. Andersen, “Encoding of Three-Dimensional Structure-from-Motion by Primate Area MT Neurons,” Nature, vol. 392, pp. 714-717, 1998.

[15] F. Bremmer, M. Kubischik, M. Pekel, M. Lappe, and K.-P. Hoffmann, “Linear Vestibular Self-Motion Signals in Monkey Medial Superior Temporal Area,” Ann. New York Academy of Sciences, vol. 871, pp. 272-281, 1999.

[16] M. Brown, D. Burschka, and G. Hager, “Advances in Computational Stereo,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pp. 993-1008, Aug. 2003.

[17] E. Brunswik and J. Kamiya, “Ecological Cue-Validity of ‘Proximity’ and of Other Gestalt Factors,” Am. J. Psychologie, vol. 66, pp. 20-32, 1953.

[18] D. Calow, N. Kru ¨ger, F. Wo ¨rgo ¨tter, and M. Lappe, “Biologically Motivated Space-Variant Filtering for Robust Optic Flow Processing,” Network, vol. 16, no. 4, pp. 323-340, 2005.

[19] M. Carandini and D.J. Heeger, “Normalization as a Canonical Neural Computation,” Nature Neuroscience, vol. 13, pp. 51-62, 2012.

[20] M.S. Caywood, B. Willmore, and D.J. Tolhurst, “Independent Components of Color Natural Scenes Resemble V1 Neurons in Their Spatial and Color Tuning,” J. Neurophysiology, vol. 91, no. 6, pp. 2859-2873, June 2004.

[21] C. Colby and M. Goldberg, “Space and Attention in Parietal Cortex,” Ann. Rev. Neuroscience, vol. 22, no. 1, pp. 319-349, 1999.

[22] B. Conway, “Spatial Structure of Cone Inputs to Color Cells in Alert Macaque Primary Visual cortex (V-1),” The J. Neuroscience, vol. 21, no. 8, pp. 2768-2783, 2001.

[23] B.R. Conway, “Color Vision, Cones, and Color-Coding in the Cortex,” Neuroscientist, vol. 15, pp. 274-290, June 2009.

[24] B.R. Conway, S. Moeller, and D.Y. Tsao, “Specialized Color Modules in Macaque Extrastriate Cortex,” Neuron, vol. 56, pp. 560573, Nov. 2007.

[25] H. Cui and R. Andersen, “Posterior Parietal Cortex Encodes Autonomously Selected Motor Plans,” Neuron, vol. 56, no. 3, pp. 552-559, 2007.

[26] B. Cumming and A. Parker, “Responses of Primary Visual Cortical Neurons to Binocular Disparity without Depth Perception,” Nature, vol. 389, no. 6648, pp. 280-283, 1997.

[27] B. Cumming and A. Parker, “Binocular Neurons in V1 of Awake Monkeys Are Selective for Absolute, Not Relative, Disparity,” J. Neuroscience, vol. 19, no. 13, pp. 5602-5618, 1999.

[28] C. Curcio and K. Allen, “Topography of Ganglion Cells in Human Retina,” J. Comparative Neurology, vol. 300, no. 1 pp. 5-25, 1990.

[29] J. Daugman, “Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression,” IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36, no. 7 pp. 1169-1179, July 1988.

[30] G.C. De Angelis, B.G. Cumming, and W.T. Newsome, “Cortical Area MT and the Perception of Stereoscopic Depth,” Nature, vol. 394, pp. 677-680, 1998.

[31] J.J. DiCarlo and D.D. Cox, “Untangling Invariant Object Recognition,” Trends in Cognitive Sciences, vol. 11, no. 8, pp. 333-341, 2007.

[32] S. Dickinson, “The Evolution of Object Categorization and the Challenge of Image Abstraction,” Object Categorization: Computer and Human Vision Perspectives, S. Dickinson, A. Leonardis, B. Schiele, and M.J. Tarr, eds., pp. 1-37, Cambridge Univ. Press, 2009.

[33] Y. Dong, S. Mihalas, F. Qiu, R. von der Heydt, and E. Niebur, “Synchrony and the Binding Problem in Macaque Visual Cortex,” J. Vision, vol. 8, no. 7, article 30, 2008.

[34] C.J. Duffy and R.H. Wurtz, “Sensitivity of MST Neurons to Optic Flow Stimuli. II. Mechanisms of Response Selectivity Revealed by Small-Field Stimuli,” J. Neurophysiology, vol. 65, pp. 1346-1359, 1991.

[35] J. Duhamel, F. Bremmer, S. BenHamed, and W. Graf, “Spatial Invariance of Visual Receptive Fields in Parietal Cortex Neurons,” Nature, vol. 389, no. 6653, pp. 845-848, 1997.

[36] J. Duhamel, C. Colby, and M. Goldberg, “The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye Movements,” Science, vol. 255, no. 5040 pp. 90-92, 1992.

[37] M.R. Du ¨rsteler and R.H. Wurtz, “Pursuit and Optokinetic Deficits Following Chemical Lesions of Cortical Areas MT and MST,” J. Neurophysiology, vol. 60, pp. 940-965, 1988.

[38] W. Einha ¨user, C. Kayser, P. Ko ¨nig, and K.P. Ko ¨rding, “Learning the Invariance Properties of Complex Cells from Their Responses to Natural Stimuli,” Euro J. Neuroscience, pp. 475-486, Feb. 2002.

[39] J.H. Elder, “Are Edges Incomplete?” Int’l J. Computer Vision, vol. 34, pp. 97-122, 1999.

[40] A. Engel, P. Roelfsema, P. Fries, M. Brecht, and W. Singer, “Role of the Temporal Domain for Response Selection and Perceptual Binding,” Cerebral Cortex, vol. 7, no. 6, pp. 571-582, 1997.

[41] R.G. Erickson and P. Thier, “A Neuronal Correlate of Spatial Stability during Periods of Self-Induced Visual Motion,” Experimental Brain Research, vol. 86, pp. 608-616, 1991.

[42] G.J. Ettinger, “Hierarchical Object Recognition Using Libraries of Parameterized Model Sub-Parts,” technical report, MIT, 1987.

[43] F. Fang, H. Boyaci, and D. Kersten, “Border Ownership Selectivity in Human Early Visual Cortex and Its Modulation by Attention,” J. Neuroscience, vol. 29, no. 2, pp. 460-465, 2009.

[44] D. Felleman and D.V. Essen, “Distributed Hierarchical Processing in Primate Cerebral Cortex,” Cerebral Cortex, vol. 1, pp. 1-47, 1991.

[45] S. Fidler, M. Boben, and A. Leonardis, “Similarity-Based CrossLayered Hierarchical Representation for Object Categorization,” Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, 2008.

[46] S. Fidler, M. Boben, and A. Leonardis, “Learning Hierarchical Compositional Representations of Object Structure,” Object Categorization: Computer and Human Vision Perspectives, S. Dickinson, A. Leonardis, B. Schiele, and M.J. Tarr, eds., pp. 196-215, Cambridge Univ. Press, 2009.

[47] G. Finlayson and S. Hordley, “Color Constancy at a Pixel,” J. Optical Soc. Am. A, vol. 18, pp. 253-264, 2001.

[48] I.M. Finn and D. Ferster, “Computational Diversity in Complex Cells of Cat Primary Visual Cortex,” J. Neuroscience, vol. 27, no. 36, pp. 9638-9648, 2007.

[49] D.J. Fleet, A.D. Jepson, and M.R. Jenkin, “Phase-Based Disparity Measurement,” CVGIP: Image Understanding, vol. 53, no. 2, pp. 10, 1991.

[50] D.J. Fleet, H. Wagner, and D.J. Heeger, “Neural Encoding of Binocular Disparity: Energy Models, Position Shifts and Phase Shifts,” Vision Research, vol. 36, no. 12, pp. 1839-1857, June 1996.

[51] M. Franzius, N. Wilbert, and L. Wiskott, “Invariant Object Recognition and Pose Estimation with Slow Feature Analysis,” Neural Computation, vol. 23, no. 9, pp. 2289-2323, 2011.

[52] D. Freedman and J. Assad, “Experience-Dependent Representation of Visual Categories in Parietal Cortex,” Nature, vol. 443, no. 5502, pp. 85-88, 2006.

[53] D. Freedman, M. Riesenhuber, T. Poggio, and E. Miller, “Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex,” Science, vol. 291, no. 5502, pp. 312-316, 2001.

[54] K. Fukushima, S. Miyake, and T. Ito, “Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition,” IEEE Systems, Man, and Cybernetics, vol. 13, no. 3, pp. 826-834, Sept./Oct. 1983.

[55] A. Gail and R. Andersen, “Neural Dynamics in Monkey Parietal Reach Region Reflect Context-Specific Sensorimotor Transformations,” J. Neuroscience, vol. 26, no. 37, pp. 9376-9384, 2006.

[56] V. Gallese, A. Murata, M. Kaseda, N. Niki, and H. Sakata, “Deficit of Hand Preshaping after Muscimol Injection in Monkey Parietal Cortex,” Neuroreport, vol. 5, no. 12, pp. 1525-1529, 1994.

[57] C. Galletti, P.P. Battaglini, and P. Fattori, “‘Real-Motion’ Cells in Area V3a of Macaque Visual Cortex,” Experimental Brain Research, vol. 82, pp. 67-76, 1990.

[58] I. Gauthier, P. Skudlarski, J.C. Gore, and A.W. Anderson, “Expertise for Cars and Birds Recruits Brain Areas Involved in Face Recognition,” Nature Neuroscience, vol. 3, no. 2, pp. 191-197, 2000.

[59] S. Geman, “Hierarchy in Machine and Natural Vision,” Proc. 11th Scandinavian Conf. Image Analysis, 1999.

[60] S. Geman, D. Potter, and Z. Chi, “Composition Systems,” Quarterly of Applied Math., vol. 60, no. 4, pp. 707-736, 2002.

[61] J. Gibson, “The Perception of Visual Surfaces,” Am. J. Psychology, vol. 63, pp. 367-384, 1950.

[62] C. Gilbert and W. Li, “Adult Visual Cortical Plasticity,” Neuron, vol. 75, no. 2, pp. 250-264, 2012.

[63] I. Go ¨decke and T. Bonhoeffer, “Development of Identical Orientation Maps for Two Eyes without Common Visual Experience,” Nature, vol. 379, pp. 251-255, 1996.

[64] M.A. Goodale and A.D. Milner, “Separate Visual Pathways for Perception and Action,” Trends Neuroscience, vol. 15, no. 1, pp. 2025, 1992.

[65] J.P. Gottlieb, M. Kusunoki, and M.E. Goldberg, “The Representation of Visual Salience in Monkey Parietal Cortex,” Nature, vol. 391, pp. 481-484, 1998.

[66] E. Gould, A. Reeves, M. Graziano, and C. Gross, “Neurogenesis in the Neocortex of Adult Primates,” Science, vol. 286, no. 5439, pp. 548-552, 1999.

[67] Y. Gu, P. Watkins, D. Angelaki, and G. DeAngelis, “Visual and Nonvisual Contributions to Three-Dimensional Heading Selectivity in the Medial Superior Temporal Area,” J. Neuroscience, vol. 26, no. 1, pp. 73-85, 2006.

[68] M. Hawken and A. Parker, “Spatial Properties of Neurons in the Monkey Striate Cortex,” Proc. Royal Soc. London, Series B, Biological Sciences, vol. 231, pp. 251-288, 1987.

[69] J. Hawkins and S. Blakeslee, On Intelligence. Times Books, 2004.

[70] D. Hinkle et al., “Three-Dimensional Orientation Tuning in Macaque Area v4,” Nature Neuroscience, vol. 5, no. 7, pp. 665670, 2002.

[71] I. Howard and B.J. Rogers, Seeing in Depth, Vol. 1: Basic Mechanisms. Univ. of Toronto Press, 2002.

[72] D. Hubel and T. Wiesel, “Receptive Fields, Binocular Interaction and Functional Architecture in the Cat’s Visual Cortex,” J. Phyiology, vol. 160, pp. 106-154, 1962.

[73] D. Hubel and T. Wiesel, “Receptive Fields and Functional Architecture of Monkey Striate Cortex,” J. Physiology, vol. 195, no. 1, pp. 215-243, 1968.

[74] C. Hung, G. Kreiman, T. Poggio, and J. DiCarlo, “Fast Readout of Object Identity from Macaque Inferior Temporal Cortex,” Science, vol. 310, no. 5749, pp. 863-866, 2005.

[75] M. Ito, H. Tamura, I. Fujita, and K. Tanaka, “Size and Position Invariance of Neuronal Responses in Monkey Inferotemporal Cortex,” J. Neurophysiology, vol. 73, no. 1, pp. 218-226, Jan. 1995.

[76] P. Janssen and M. Shadlen, “A Representation of the Hazard Rate of Elapsed Time in Macaque Area Lip,” Nature Neuroscience, vol. 8, no. 2, pp. 234-241, 2005.

[77] P. Janssen, R. Vogels, Y. Liu, and G. Orban, “Macaque Inferior Temporal Neurons Are Selective for Three-Dimensional Boundaries and Surfaces,” J. Neuroscience, vol. 21, no. 23, pp. 9419-9429, 2001.

[78] P. Janssen, R. Vogels, Y. Liu, and G. Orban, “At Least at the Level of Inferior Temporal Cortex, the Stereo Correspondence Problem Is Solved,” Neuron, vol. 37, no. 4, pp. 693-701, 2003.

[79] P. Janssen, R. Vogels, and G. Orban, “Macaque Inferior Temporal Neurons Are Selective for Disparity-Defined Three-Dimensional Shapes,” Proc. Nat’l Academy of Sciences, vol. 96, no. 14 p. 8217, 1999.

[80] P. Janssen, R. Vogels, and G. Orban, “Selectivity for 3D Shape that Reveals Distinct Areas within Macaque Inferior Temporal Cortex,” Science, vol. 288, no. 5473, pp. 2054-2056, 2000.

[81] P. Janssen, R. Vogels, and G. Orban, “Three-Dimensional Shape Coding in Inferior Temporal Cortex,” Neuron, vol. 27, no. 2, pp. 385-397, 2000.

[82] J. Jones and L. Palmer, “An Evaluation of the Two Dimensional Gabor Filter Model of Simple Receptive Fields in Striate Cortex,” J. Neurophysiology, vol. 58, no. 6, pp. 1223-1258, 1987.

[83] B. Julesz, Foundations of Cyclopean Perception. Univ. Chicago Press, 1971.

[84] Principles of Neural Science, E.R. Kandel, J.H. Schwartz, and T.M. Jessel eds., fourth ed. McGraw-Hill, 2000.

[85] N. Kanwisher, J. McDermott, and M.M. Chun, “The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception,” J. Neuroscience, vol. 17, no. 11, pp. 4302-4311, 1997.

[86] N. Katsuyama, A. Yamashita, K. Sawada, T. Naganuma, H. Sakata, and M. Taira, “Functional and Histological Properties of Caudal Intraparietal Area of Macaque Monkey,” Neuroscience, vol. 167, no. 1, pp. 1-10, 2010.

[87] P. Kellman and M. Arterberry, The Cradle of Knowledge. MIT Press, 1998.

[88] R. Kiani, H. Esteky, K. Mirpour, and K. Tanaka, “Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex,” J. Neurophysiology, vol. 97, no. 6, pp. 4296-4309, 2007.

[89] C. Koch, Biophysics of Computation: Information Processing in Single Neurons. Oxford Univ. Press, 1999.

[90] K. Koffka, Principles of Gestalt Psychology. Routledge, 1955.

[91] P. Ko ¨nig and N. Kru ¨ger, “Perspectives: Symbols as Self-Emergent Entities in an Optimization Process of Feature Extraction and Predictions,” Biological Cybernetics, vol. 94, no. 4, pp. 325-334, 2006.

[92] K. Ko ¨teles, P. De Maziere, M. Van Hulle, G. Orban, and R. Vogels, “Coding of Images of Materials by Macaque Inferior Temporal Cortical Neurons,” European J. Neuroscience, vol. 27, no. 2, pp. 466482, 2008.

[93] Z. Kourtzi and J.J. DiCarlo, “Learning and Neural Plasticity in Object Recognition,” Current Opinion Neurobiology, vol. 16, pp. 152158, Apr. 2006.

[94] R.J. Krauzlis and S.G. Lisberger, “A Model of Visually-Guided Smooth Pursuit Eye Movements Based on Behavioral Observations,” J. Computational Neuroscience, vol. 1, no. 4, pp. 265-283, 1994.

[95] J. Kremers, The Primate Visual System: A Comparative Approach. Wiley, 2005.

[96] K. Krug, B. Cumming, and A. Parker, “Comparing Perceptual Signals of Single V5/MT Neurons in Two Binocular Depth Tasks,” J. Neurophysiology, vol. 92, no. 3, pp. 1586-1596, 2004.

[97] K. Krug and A. Parker, “Neurons in Dorsal Visual Area V5/MT Signal Relative Disparity,” J. Neuroscience, vol. 31, no. 49, pp. 17892-17904, 2011.

[98] V. Lamme et al., “Neuronal Synchrony Does Not Represent Texture Segregation,” Nature, vol. 396, no. 6709, pp. 362-366, 1998.

[99] M. Lappe, “Functional Consequences of an Integration of Motion and Stereopsis in Area MT of Monkey Extrastriate Visual Cortex,” Neural Computation, vol. 8, no. 7, pp. 1449-1461, 1996.

[100] M. Lappe, F. Bremmer, M. Pekel, A. Thiele, and K.P. Hoffmann, “Optic Flow Processing in Monkey STS: A Theoretical and Experimental Approach,” J. Neuroscience, vol. 16, no. 19, pp. 6265-6285, 1996.

[101] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-Based Learning Applied to Document Recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[102] T. Lee and D. Mumford, “Hierarchical Bayesian Inference in the Visual Cortex,” J. Optical Soc. Am., vol. 20, no. 7, pp. 1434-1448, 2003.

[103] J. Lewis and D. Van Essen, “Corticocortical Connections of Visual, Sensorimotor, and Multimodal Processing Areas in the Parietal Lobe of the Macaque Monkey,” J. Comparative Neurology, vol. 428, no. 1, pp. 112-137, 2000.

[104] N. Li and J. DiCarlo, “Unsupervised Natural Visual Experience Rapidly Reshapes Size-Invariant Object Representation in Inferior Temporal Cortex,” Neuron, vol. 67, no. 6, pp. 10621075, 2010.

[105] S. Li, S.D. Mayhew, and Z. Kourtzi, “Learning Shapes Spatiotemporal Brain Patterns for Flexible Categorical Decisions,” Cerebral Cortex, to be published, 2011.

[106] M.S. Livingstone, C.C. Pack, and R.T. Born, “Two-Dimensional Substructure of MT Receptive Fields,” Neuron, vol. 30, no. 3, pp. 781-793, 2001.

[107] D.G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” Int’l J. Computer Vision, vol. 2, no. 60, pp. 91-110, 2004.

[108] B. Manjunath and W. Ma, “Texture Features for Browsing and Retrieval of Image Data,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 8, pp. 837-842, Aug. 1996.

[109] D. Marr, Vision: A Computational Investigation into the Human Representation and Processing of Visual information. Freeman, 1977.

[110] G. Masson, C. Busettini, and F. Miles, “Vergence Eye Movements in Response to Binocular Disparity without Depth Perception,” Nature, vol. 389, no. 6648, pp. 283-286, 1997.

[111] T. Matsumora, K. Koida, and H. Komatsu, “Relationship between Color Discrimination and Neural Responses in the Inferior Temporal Cortex of the Monkey,” J. Neurophysiology, vol. 100, no. 6, pp. 3361-3374, 2008.

[112] J.H.R. Maunsell and D.C. van Essen, “Functional Properties of Neurons in Middle Temporal Visual Area of the Macaque Monkey. I. Selectivity for Stimulus Direction, Speed, and Orientation,” J. Neurophysiology, vol. 49, no. 5, pp. 1127-1147, 1983.

[113] S.D. Mayhew, S. Li, and Z. Kourtzi, “Learning Acts on Distinct Processes for Visual Form Perception in the Human Brain,” J. Neuroscience, vol. 32, no. 3, pp. 775-786, 2012.

[114] B.W. Mel and J. Fiser, “Minimizing Binding Errors Using Learned Conjunctive Features,” Neural Computation, vol. 12, no. 4, pp. 731762, 2000.

[115] B.W. Mel, D.L. Ruderman, and K.A. Archie, “TranslationInvariant Orientation Tuning in Visual ‘Complex’ Cells Could Derive from Intradendritic Computations,” J. Neuroscience, vol. 18, no. 11, pp. 4325-4334, 1998.

[116] K. Mikolajczyk and C. Schmid, “A Performance Evaluation of Local Descriptors,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630, Oct. 2005.

[117] J.A. Movshon, E.H. Adelson, M.S. Gizzi, and W.T. Newsome, “The Analysis of Moving Visual Patterns,” Pattern Recognition Mechanisms, C. Chagas, R. Gattass, and C. Gross, eds., pp. 117-151, Springer, 1985.

[118] J.A. Movshon and W.T. Newsome, “Visual Response Properties of Striate Cortical Neurons Projecting to Area MT in Macaque Monkeys,” J. Neuroscience, vol. 16, no. 23, pp. 7733-7741, 1996.

[119] R. Mruczek and D. Sheinberg, “Activity of Inferior Temporal Cortical Neurons Predicts Recognition Choice Behavior and Recognition Time during Visual Search,” J. Neuroscience, vol. 27, no. 11, pp. 2825-2836, 2007.

[120] A. Murata, V. Gallese, G. Luppino, M. Kaseda, and H. Sakata, “Selectivity for the Shape, Size, and Orientation of Objects for Grasping in Neurons of Monkey Parietal Area AIP,” J. Neurophysiology, vol. 83, no. 5, pp. 2580-2601, 2000.

[121] H. Nakamura, T. Kuroda, M. Wakita, M. Kusunoki, A. Kato, A. Mikami, H. Sakata, and K. Itoh, “From Three-Dimensional Space Vision to Prehensile Hand Movements: The Lateral Intraparietal Area Links the Area V3a and the Anterior Intraparietal Area in Macaques,” J. Neuroscience, vol. 21, no. 20, pp. 8174-8187, 2001.

[122] W.T. Newsome, R.H. Wurtz, and H. Komatsu, “Relation of Cortical Areas MT and MST to Pursuit Eye Movements. II. Differentiation of Retinal from Extraretinal Inputs,” J. Neurophysiology, vol. 60, no. 2, pp. 604-620, 1988.

[123] J. Nguyenkim and G. DeAngelis, “Disparity-Based Coding of Three-Dimensional Surface Orientation by Macaque Middle Temporal Neurons,” J. Neuroscience, vol. 23, no. 18, pp. 71177128, 2003.

[124] A. Oliva and A. Torralba, “The Role of Context in Object Recognition,” Trends Cognitive Sciences, vol. 11, pp. 520-527, 2007.

[125] B.A. Olshausen and D.J. Field, “Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images,” Nature, vol. 381, no. 6583, pp. 607-609, June 1996.

[126] B. Ommer and J.M. Buhmann, “Learning the Compositional Nature of Visual Objects,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2007.

[127] G.A. Orban, “Higher Order Visual Processing in Macaque Extrastriate Cortex,” Physiological Rev., vol. 88, pp. 59-89, Jan. 2008.

[128] C. Pack, M.S. Livingstone, K. Duffy, and R. Born, “End-Stopping and the Aperture Problem: Two-Dimensional Motion Signals in Macaque V1,” Neuron, vol. 39, pp. 671-680, 2003.

[129] C.C. Pack and R.T. Born, “Temporal Dynamics of a Neural Solution to the Aperture Problem in Visual Area MT of Macaque Brain,” Nature, vol. 409, no. 6823, pp. 1040-1042, 2001.

[130] A. Parker, “Binocular Depth Perception and the Cerebral Cortex,” Nature Rev. Neuroscience, vol. 8, no. 5, pp. 379-391, 2007.

[131] A. Pasupathy and C. Connor, “Responses to Contour Features in Macaque Area V4,” J. Neurophysiology, vol. 82, no. 5 pp. 2490-2502, 1999.

[132] M. Pekel, M. Lappe, F. Bremmer, A. Thiele, and K.-P. Hoffmann, “Neuronal Responses in the Motion Pathway of the Macaque Monkey to Natural Optic Flow Stimuli,” NeuroReport, vol. 7, no. 4, pp. 884-888, 1996.

[133] J.A. Perrone and A. Thiele, “Speed Skills: Measuring the Visual Speed Analyzing Properties of Primate MT Neurons,” Nature Neuroscience, vol. 4, no. 5, pp. 526-532, 2001.

[134] B. Pesaran, M. Nelson, and R. Andersen, “Free Choice Activates a Decision Circuit between Frontal and Parietal Cortex,” Nature, vol. 453, no. 7193, pp. 406-409, 2008.

[135] R. Peters, A. Iyer, L. Itti, and C. Koch, “Components of Bottom-Up Gaze Allocation in Natural Images,” Int’l J. Neural Systems, vol. 45, no. 18, pp. 2397-2416, 2005.

[136] M. Platt and P. Glimcher, “Neural Correlates of Decision Variables in Parietal Cortex,” Nature, vol. 400, no. 6741, pp. 233-238, 1999.

[137] A. Pouget and T.J. Sejnowski, “Spatial Transformations in the Parietal Cortex Using Basis Functions,” J. Coginitive Neuroscience, vol. 9, no. 2, pp. 222-237, 1997.

[138] R.Q. Quiroga, L. Reddy, C. Koch, and I. Fried, “Decoding Visual Inputs from Multiple Neurons in the Human Temporal Lobe,” J Neurophysiology, vol. 98, no. 4, pp. 1997-2007, 2007.

[139] S. Raiguel, R. Vogels, S. Mysore, and G. Orban, “Learning to See the Difference Specifically Alters the Most Informative V4 Neurons,” J. Neuroscience, vol. 26, no. 24, pp. 6589-6602, 2006.

[140] M. Riesenhuber and T. Poggio, “Hierarchical Models of Object Recognition in Cortex,” Nature Neuroscience, vol. 11, no. 2, pp. 1019-1025, 1999.

[141] K. Rockland, J. Kaas, and A. Peters, Cerebral Cortex: Extrastriate Cortex in Primates, vol. 12. Springer, 1997.

[142] A. Rodrı ´guez-Sa ´nchez, E. Simine, and J. Tsotsos, “Attention and Visual Search,” Int’l J. Neural Systems, vol. 17, no. 4, pp. 275-288, 2007.

[143] A.J. Rodriguez-Sanchez and J.K. Tsotsos, “The Importance of Intermediate Representations for the Modeling of 2D Shape Detection: Endstopping and Curvature Tuned Computations,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 43214326, 2011.

[144] A.J. Rodriguez-Sanchez and J.K. Tsotsos, “The Roles of Endstopped and Curvature Tuned Computations in a Hierarchical Representation of 2D Shape,” PLoS ONE, vol. 7, no. 8, pp. 1-13, 2012.

[145] P. Roelfsema, “Solutions for the Binding Problem,” Zeitschrift fu ¨r Naturforschung. C, J. Biosciences, vol. 53, no. 7/8, pp. 691-715, 1998.

[146] E. Rolls, B. Webb, and M.C.A. Booth, “Responses of Inferior Temporal Cortex Neurons to Objects in Natural Scenes,” Soc. Neuroscience Abstracts, vol. 26, p. 1331, 2000.

[147] J.-P. Roy and R.H. Wurtz, “Disparity Sensitivity of Neurons in Monkey Extrastriate Area MST,” J. Neuroscience, vol. 12, no. 7, pp. 2478-2492, 1992.

[148] N. Rust, O. Schwartz, J. Movshon, and E. Simoncelli, “SpikeTriggered Characterization of Excitatory and Suppressive Stimulus Dimensions in Monkey V1,” Neurocomputing, vol. 58, pp. 793799, 2004.

[149] U. Rutishauser, R.J. Douglas, and J.J. Slotine, “Collective Stability of Networks of Winner-Take-All Circuits,” Neural Computation, vol. 23, no. 3, pp. 735-773, 2011.

[150] H. Sakata, M. Taira, M. Kusunoki, A. Murata, and Y. Tanaka, “The Parietal Association Cortex in Depth Perception and Visual Control of Hand Action,” Trends in Neurosciences, vol. 20, no. 8, pp. 350-357, 1997.

[151] C.D. Salzman, K.H. Britten, and W.T. Newsome, “Cortical Microstimulation Influences Perceptual Judgements of Motion Direction,” Nature, vol. 346, no. 12, pp. 174-177, 1990.

[152] F. Scalzo and J.H. Piater, “Statistical Learning of Visual Feature Hierarchies,” Proc. IEEE Workshop Learning Computer Vision and Pattern Recognition, 2005.

[153] C.E. Stern and H.E. Schendan, “Mental Rotation and Object Categorization Share a Common Network of Prefrontal and Dorsal and Ventral Regions of Posterior Cortex,” Neuroimage, vol. 35, pp. 1264-1277, 2007.

[154] A. Schoups, R. Vogels, N. Qian, and G. Orban, “Practising Orientation Identification Improves Orientation Coding in V1 Neurons,” Nature, vol. 412, no. 6846, pp. 549-53, 2001.

[155] A. Sereno and J. Maunsell, “Shape Selectivity in Primate Lateral Intraparietal Cortex,” Nature, vol. 395, no. 6701, pp. 500-503, 1998.

[156] T. Serre and A.T. Poggio, “Neuromorphic Approach to Computer Vision,” Comm. ACM, vol. 53, pp. 54-61, 2010.

[157] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, “Object Recognition with Cortex-Like Mechanisms,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 3, pp. 411-426, Mar. 2007.

[158] M. Shadlen and J. Movshon, “Synchrony Unbound: Review a Critical Evaluation of the Temporal Binding Hypothesis,” Neuron, vol. 24, pp. 67-77, 1999.

[159] M. Shadlen and W. Newsome, “Motion Perception: Seeing and Deciding,” Proc. Nat’l Academy of Sciences, vol. 93, no. 2 pp. 628633, 1996.

[160] R. Shapley and M.J. Hawken, “Color in the Cortex: Single- and Double-Opponent Cells,” Vision Research, vol. 51, no. 7, pp. 701717, Apr. 2011.

[161] X. Shi, N. Bruce, and J.K. Tsotsos, “Fast, Recurrent, Attentional Modulation Improves Saliency Representation and Scene Recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Biologically-Consistent Vision, pp. 1-8, 2011.

[162] E. Shikata et al., “Selectivity of the Parietal Visual Neurones in 3D Orientation of Surface of Stereoscopic Stimuli,” Neuroreport, vol. 7, no. 14, pp. 2389-2394, 1996.

[163] E. Simoncelli, “Vision and the Statistics of the Visual Environment,” Current Opinion in Neurobiology, vol. 13, no. 2, pp. 144-149, 2003.

[164] E.P. Simoncelli and D.J. Heeger, “A Model of Neuronal Responses in Visual Area MT,” Vision Research, vol. 38, no. 5, pp. 743-761, 1998.

[165] W. Singer, “Consciousness and the Binding Problem,” Ann. New York Academy of Sciences, vol. 929, no. 1, pp. 123-146, 2001.

[166] H. Spitzer and S. Hochstein, “A Complex-Cell Receptive-Field Model,” J. Neurophysiology, vol. 53, no. 5, pp. 1266-1286, 1985.

[167] S. Srivastava, G. Orban, P. De Mazie `re, and P. Janssen, “A Distinct Representation of Three-Dimensional Shape in Macaque Anterior Intraparietal Area: Fast, Metric, and Coarse,” J. Neuroscience, vol. 29, no. 34, pp. 10613-10626, 2009.

[168] S. Swaminathan and D. Freedman, “Preferential Encoding of Visual Categories in Parietal Cortex Compared with Prefrontal Cortex,” Nature Neuroscience, vol. 15, pp. 315-320, 2012.

[169] M. Taira, K. Tsutsui, M. Jiang, K. Yara, and H. Sakata, “Parietal Neurons Represent Surface Orientation from the Gradient of Binocular Disparity,” J. Neurophysiology, vol. 83, no. 5 pp. 31403146, 2000.

[170] A. Takemura, Y. Inoue, K. Kawano, C. Quaia, and F. Miles, “Single-Unit Activity in Cortical Area MST Associated with Disparity-Vergence Eye Movements: Evidence for Population Coding,” J. Neurophysiology, vol. 85, no. 5, pp. 2245-2266, 2001.

[171] S. Tanabe, K. Umeda, and I. Fujita, “Rejection of False Matches for Binocular Correspondence in Macaque Visual Cortical Area V4,” J. Neuroscience, vol. 24, no. 37, pp. 8170-8180, 2004.

[172] K. Tanaka, “Neuronal Mechanisms of Object Recognition,” Science, vol. 262, pp. 685-688, 1993.

[173] K. Tanaka, “Inferotemporal Cortex and Object Vision,” Ann. Rev. of Neuroscience, vol. 19, no. 1, pp. 109-139, 1996.

[174] K. Tanaka, H. Saito, Y. Fukada, and M. Moriya, “Coding Visual Images of Objects in the Inferotemporal Cortex of the Macaque Monkey,” J. Neurophysiology, vol. 66, no. 1, pp. 170189, 1991.

[175] K. Tanaka and H.-A. Saito, “Analysis of Motion of the Visual Field by Direction, Expansion/Contraction, and Rotation Cells Clustered in the Dorsal Part of the Medial Superior Temporal Area of the Macaque Monkey,” J. Neurophysiology, vol. 62, no. 3, pp. 626641, 1989.

[176] H. Tanigawa, H.D. Lu, and A.W. Roe, “Functional Organization for Color and Orientation in Macaque V4,” Nature Neuroscience, vol. 13, pp. 1542-1548, Dec. 2010.

[177] J.B. Tenenbaum, C. Kemp, T.L. Griffiths, and N.D. Goodman, “How to Grow a Mind: Statistics, Structure, and Abstraction,” Science, vol. 331, pp. 1279-1285, 2011.

[178] F.E. Theunissen, S.V. David, N.C. Singh, A. Hsu, W.E. Vinje, and J.L. Gallant, “Estimating Spatio-Temporal Receptive Fields of Auditory and Visual Neurons from Their Responses to Natural Stimuli,” Network: Computation in Neural Systems, vol. 12, no. 3, pp. 289-316, 2001.

[179] T. Theys, S. Srivastava, J. van Loon, J. Goffin, and P. Janssen, “Selectivity for Three-Dimensional Contours and Surfaces in the Anterior Intraparietal Area,” J. Neurophysiology, vol. 107, no. 3, pp. 995-1008, 2012.

[180] P. Thier and R. Andersen, “Electrical Microstimulation Suggest Two Different Kinds of Representation of Head-Centered Space in the Intraparietal Sulcus of Rhesus Monkeys,” Proc. Nat’l Academy of Sciences USA, vol. 93, pp. 4962-4967, 1996.

[181] O. Thomas, B. Cumming, and A. Parker, “A Specialization for Relative Disparity in V2,” Nature Neuroscience, vol. 5, no. 5, pp. 472-478, 2002.

[182] T. Tompa and G. Sa ´ry, “A Review on the Inferior Temporal Cortex of the Macaque,” Brain Research Rev., vol. 62, no. 2, pp. 165-182, 2010.

[183] R. Tootell, K. Nelissen, W. Vanduffel, and G. Orban, “Search for Color Center(s) in Macaque Visual Cortex,” Cerebral Cortex, vol. 14, no. 4, pp. 353-363, 2004.

[184] M.J. Tove ´e, E.T. Rolls, and P. Azzopardi, “Translation Invariance in the Responses to Faces of Single Neurons in the Temporal Visual Cortical Areas of the Alert Macaque,” J. Neurophysiology, vol. 72, no. 3, pp. 1049-1060, Sept. 1994.

[185] A. Treisman, “The Binding Problem,” Current Opinion in Neurobiology, vol. 6, no. 2, pp. 171-178, 1996.

[186] A. Treisman and H. Schmidt, “Illusory Conjunctions in the Perception of Objects,” Cognitive Psychology, vol. 14, no. 1, pp. 107-141, 1982.

[187] D. Tsao et al., “Stereopsis Activates V3a and Caudal Intraparietal Areas in Macaques and Humans,” Neuron, vol. 39, no. 3, pp. 555568, 2003.

[188] J.K. Tsotsos, “Analyzing Vision at the Complexity Level,” Behavioral and Brain Sciences, vol. 13, no. 3, pp. 423-469, 1990.

[189] K. Tsutsui, H. Sakata, T. Naganuma, and M. Taira, “Neural Correlates for Perception of 3D Surface Orientation from Texture Gradient,” Science, vol. 298, no. 5592, pp. 409-412, 2002.

[190] T. Uka and G. DeAngelis, “Linking Neural Representation to Function in Stereoscopic Depth Perception: Roles of the Middle Temporal Area in Coarse versus Fine Disparity Discrimination,” J. Neuroscience, vol. 26, no. 25, pp. 67916802, 2006.

[191] S. Ullman and B. Epshtein, “Visual Classification by a Hierarchy of Extended Features,” Towards Category-Level Object Recognition. Springer-Verlag, 2006.

[192] L.G. Ungerleider and M. Mishkin, “Two Cortical Visual Systems,” Analysis of Visual Behavior, D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield, eds., pp. 549-586, MIT Press, 1982.

[193] C. van den Boomen, M.J. van der Smagt, and C. Kemner, “Keep Your Eyes on Development: The Behavioral and Neurophysiological Development of Visual Mechanisms Underlying Form Processing,” Front Psychiatry, vol. 16, no. 3, 2012.

[194] J.H. van Hateren and D.L. Ruderman, “Independent Component Analysis of Natural Image Sequences Yields Spatio-Temporal Filters Similar to Simple Cells in Primary Visual Cortex,” Proc. Biological Sciences, vol. 265, no. 1412, pp. 2315-2320, 1998.

[195] B. Verhoef, R. Vogels, and P. Janssen, “Contribution of Inferior Temporal and Posterior Parietal Activity to Three-Dimensional Shape Perception,” Current Biology, vol. 20, no. 10, pp. 909-913, 2010.

[196] B. Verhoef, R. Vogels, and P. Janssen, “Inferotemporal Cortex Subserves Three-Dimensional Structure Categorization,” Neuron, vol. 73, pp. 171-182, 2012.

[197] R. Vogels, “Categorization of Complex Visual Images by Rhesus Monkeys. Part 2: Single-Cell Study,” European J. Neuroscience, vol. 11, no. 4, pp. 1239-1255, 1999.

[198] Handbuch der Physiologischen Optic, H. von Helmholtz, ed. Hamburg and Leipzig, 1866.

[199] J. Wagemans, J. Elder, M. Kubovy, S. Palmer, M. Peterson, M. Singh, and R. von der Heydt, “A Century of Gestalt Psychology in Visual Perception: I. Perceptual Grouping and Figure-Ground Organization,” Psychological Bull., to be Published.

[200] L. Wiskott, “How Does Our Visual System Achieve Shift and Size Invariance?” 23 Problems in Systems Neuroscience, chapter 16, J.L. van Hemmen and T.J. Sejnowski, ed., pp. 322-340, Oxford Univ. Press, 2006.

[201] L. Wiskott, J.-M. Fellous, N. Kru ¨ger, and C. von der Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775-779, July 1997.

[202] J.M. Wolfe, “Visual Search,” Attention, H. Pashler, ed., Univ. College London Press, 1998.

[203] F. Xu and S. Carey, “Infants’ Metaphysics: The Case of Numerical Identity,” Cognitive Psychology, vol. 30, no. 2, pp. 111-153, Apr. 1996.

[204] S. Zeki, S. Aglioti, D. McKeefry, and G. Berlucchi, “The Neurobiological Basis of Conscious Color Perception in a Blind Patient,” Proc. Nat’l Academy of Sciences, vol. 96, pp. 14124-14129, 1999.

[205] H. Zhou, H. Friedman, and R. Von Der Heydt, “Coding of Border Ownership in Monkey Visual Cortex,” J. Neuroscience, vol. 20, no. 17, pp. 6594-6611, 2000.

[206] D. Zipser and R.A. Andersen, “A Back-Propagation Programmed Network That Simulates Response Properties of a Subset of Posterior Parietal Neurons,” Nature, vol. 331, no. 6158, pp. 679-684, 1988.

[207] N. Pugeault, F. Wo ¨rgo ¨tte, and N. Kru ¨ger, “Visual Primitives: Local, Condensed, and Semantically Rich Visual Descriptors and Their Applications in Robotics,” Int’l J. Humanoid Robotics, special issue on cognitive humanoid vision, vol. 7, no. 3, pp. 379-405, 2010.