

# 视觉运动信息加工模型

丁锦红

王东晖

中国科学院心理研究所(北京 100012)

[摘要]视觉运动信息加工是一个复杂的过程。研究者们用不同的方法,分别从各自的角度来研究这一过程,同时也建立了一系列模型。其中早期模型通过研究昆虫和兔子等动物的视觉特征,从神经生理学的角度来描述视觉运动信息加工过程(Reichardt 和 Barlow 等的模型);在早期模型基础上,近期模型则采用较为复杂的数学计算,把运动信息量化,以便能够在计算机上实现。这些模型包括时间协变模型、空间-时间梯度模型、线性运动感受器模型和空间-时间能量模型等。

关键词 模型,视觉运动信息,感受器,时间-空间整合

看电影时,实际上我们所看到的是一系列位置上的物体的图象。尽管每一幅图象都是静止的,且呈现时间都固定不变,但我们仍然觉得物体在不断地运动。其原因在于视觉系统把相邻图象内对应点相匹配,然后根据图象在一定时间间隔内所移动的距离来推断运动速度,这样就可以把一系列静止的图象知觉为连续运动的物体,这就是视觉运动知觉过程。视觉运动信息加工是一个复杂的过程,许多研究者用不同的方法,从各自的角度进行了研究,同时也建立了一系列模型。本文旨在粗略介绍几个有代表性的模型。

## 1 早期模型

### 1.1 Reichardt 模型

1961 年 Reichardt 首先提出了关于运动觉察的模型<sup>[1]</sup>。这一模型主要是用来说明昆虫的视觉运动反应特性的。运动觉察器包含两个互为镜像的子单元,输入到这些子单元的信息是对某一位置( $L(x, t)$ )的左、右( $X_{左}$ 和 $X_{右}$ )同时取样所获得的信号( $L(X_{左}, t)$ ,  $L(X_{右}, t)$ )。即输入通道拥有被两个子单元共享的点感受野。这些子单元分别接收反向运动信息。例如,右边单元对来自左边的运动敏感,通过线性时间过滤后,其输出就可以反映出 $L(X_{左}, t)$ 和 $L(X_{右}, t)$ 的匹配情况了。左边单元的活动与之类似。运动觉察最基本的单元是一对感受器,其中一个感受器的延时输出与另一个感受器的输出非线性相乘。子单元输出之间的差异决定了觉察器的最后输出结果,输出结果的符号就是运动的方向。从理论上讲,这种运动信号所具有的许多特征与昆虫研究所得到的结果相一致:首先,亮度和运动信号之间是平方关系。

其次,如果一个光栅的半周期小于昆虫小眼之间的距离;或者单步运动的每一步都伴随着相反的对比时,这两种条件下都可以看到方向相反的运动。最后,它可以预测出使得运动机制产生适当方向性反应的空间与时间频率范围。

总的来说,对于包含不同方向与空间频率的复杂视觉刺激的加工经过两个阶段:一是测量一定方向上的空间频率成份的速度;二是把前一阶段中的结果加以综合,产生一个在一定范围内与其成份的运动速度和方向一致的运动知觉<sup>[2]</sup>。

Reichardt 的运动觉察模型是局部模型。模型中涉及到的无限时间平均过程(infinite time averaging)在实际系统中是不可能的,因为它要求时间无穷延伸。而且,人和昆虫的视觉系统有着本质差别,这一模型并不一定适合人的视觉运动信息加工的特点。尽管如此,它对后来的运动加工模型仍具有十分重要的影响。

## 1.2 神经生理学模型

Barlow 和 Levick 于 1965 年在对兔子的视觉特征研究的基础上提出了神经生理学模型<sup>[1]</sup>。他们认为,在兔子的视网膜上可能有两种具有方向敏感性的机制。一是具有方向偏好,只对单一方向的侧向产生兴奋反应;另一个是无方向偏好,对单一方向的侧向产生抑制反应。在第一种情况下,一个运动点将要落向的部位的反应已被以前的刺激所在的相邻区域易化。在第二种条件下,与第一种情况相反,这些部位的反应已被以前的刺激所在的相邻区域抑制。模型认为,视觉系统把位置 A 的明度和时间  $T_1$  分别与 B 位置的明度和时间  $T_2$  相比较,从而产生运动知觉。这一过程与短程运动分析(只对明亮的点反应)相似,而与长程运动分析(只对某种边界等反应)不同。模型的上述两种特性或单个或同时产生作用,使系统的方向敏感性提高。例如,把两个闪光点在呈现时间和位置作为变量,并且分别对有方向偏好和无方向偏好的细胞呈现。结果表明,不论是在快速运动条件下,还是在慢速运动条件下,无方向偏好机制的反应都是抑制性的。皮层细胞的感受野与视网膜上的感受野区域对单个闪光刺激的反应相同,都接近线性乘积。这种反应只发生在该区域的偏好方向上。而在非偏好方向上的反应则是非线性的。神经生理模型也是建立在对动物视觉系统研究的基础上,很难推广到人类的运动视觉加工过程。当然,该模型的优点在于它很容易在计算机上模拟实现。

上述两种模型虽然有些差别,但也有相似之处。(1)它们都涉及到乘法过程,至少在部分程度上是这样。(2)每个模型都只说明了如何确定运动方向,但并未说明速度大小是如何确定的。

除上述两种模型以外,在早期模型中还有两个运动加工模型也具有一定代表性。一个是建立在工程应用基础上的工程应用模型。这一模型是用来测量运动物体速度的大小而不是方向。另一个是空间-时间梯度模型,但这一模型不能很好地说明哺乳动物视觉系统的运动敏感性。由于这两个模型有很大的局限性,所以,后来并没有引起研究者更大兴趣。

## 2 近期模型

### 2.1 时间协变模型(Temporal Covariance Model)

这是 van Santen 和 Sperling<sup>[3]</sup>对 Reichardt 模型的修正。他们通过阈限实验来说明模型是如何进行工作的。在阈限实验中,让被试看相邻的垂直线段。这些线段的明度都不同,且随时间变化而变化。被试的任务是报告这样的刺激模式是向左运动还是相右运动。该模型

认为,基本的觉察单元包括两个对相反运动敏感的子单元。每个子单元对输入的信息进行空间和时间的线性过滤,其输出结果相乘积,被乘积后的结果又被整合,整合时间要比明度变化所需时间长一些。然后再把所有有关边界类型的知识与图象明度的局部变化时间相结合,就可以推断运动方向。

该模型认为,视觉系统能够进行一些局部的空间-时间富立叶分析。所有的富立叶成份既包含着不同的时间频率,也包含着相同的时间频率。通过空间-时间的富立叶分析,把不同时间频率分离开来,然后进行线性相加。因此,(1)通过空间-时间的富立叶分析,可以对复杂模式识别的作业成绩进行预测。(2)在特定情况下,作业成绩取决于相邻线段的明度变化大小,这一点支持了乘法原则。(3)加入静止图象模式不影响作业成绩。(4)即使加入同质闪烁刺激,也不产生影响;但在特定条件下,将使得知觉到的方向变反。由此可见,运动知觉不包括逐幅图形的比较,因为加入静止模式不影响作业成绩。运动机制不是比较两个同时出现的空间亮度模式,而是比较两个不同位置上不同时刻的亮度模式。

模型的行为不依赖于某些细节(例如,比较的具体类型- - 乘积的或是延迟抑制的、选择规则以及感受野和时间过滤器的特性等)而是由其深层结构所引起(包括对相邻位置上的同一时刻明度变化的非线性比较)。这一模型的主要优点在于它消除了原 Reichardt 模型中的空间不同频率的波在重叠点的混淆现象(aliasing),使得该模型的特点与人的运动知觉特性更加接近。

## 2.2 空间-时间梯度模型(Spatio-temporal Gradient Model)

Marr-Ullman 在 1981 提出这一模型<sup>[1]</sup>。他们用空间频率过滤过程来说明视觉运动加工,使得空间-时间梯度模型更加合理化。其中包括两个成份:一是空间成份- - 运动目标的边界;另一个是时间成份- - 觉察边界的净明暗程度。当物体运动时,其明暗程度将发生变化。他们认为,图象是被感受野中的感受器所卷积的( $\text{del}^2 G$ , 高斯-拉普拉斯变换),卷积

产生一个新的图象强度分布  $L$ , 则速度用式  $V_x = \frac{\partial \nabla^2 G \times \frac{1}{\alpha}}{\partial \nabla^2 G \times \frac{1}{\alpha}}$  计算。

为了减少计算,每一次的测量都被限定在  $L$  的零交叉处(zero crossing),因为零交叉处对应于图象明度发生明显变化的区域。上述公式使得算法更加简单,它在计方向时很经济。公式中只考虑分子和分母的符号,对两个有符号的信号进行初步地比较就可以计算出运动方向。这就大大地减轻计算负担。但从计算中却得不到精确的速度大小信息。Marr 等还认为, Y 细胞可以进行必要的时间微分。然而,实际上,由于这些细胞在对子单元整流时已达到饱和,所以不能传递  $\partial \nabla^2 G \times \frac{1}{\alpha}$  的符号。尽管这种说法与 Marr 对 Y 细胞作用的假设冲突,但 Marr 等的模型在计算理论方面做出新的贡献。Marr 等的模型的优点在于:它只需局部测量,没有时间延时参与计算,计算结果非常可靠,能够对很大范围内的运动速度反应。

## 2.3 线性运动感受器模型(linear Model of human visual-motion sensing)

上述几种模型都从运动的非线性特征出发。但, Watson 和 Ahumada<sup>[4]</sup>认为,一个运动的刺激也可用适当的线性成份来描述。在一定的空间-时间频率范围内,一个运动的光栅可以用与原先相对称的两个谱点来描述。

该模型有两个基本来源,一是利用对人的心理物理特性方面的限制的认识,尤其是运动

觉察元素只对两维的空间频率敏感的特性;二是运动图象的频谱,即在一个二维平面内,可以用空间频率和图象速度来描述时间频率。此模型认为,在整个空间和时间频率范围内都存在着可以进行空间调节的感受野,这两个平行通道的输出结果被简单地相加,而不是相乘。在模型的第一阶段中,那些具有方向敏感性的线性感受器对特定空间频率反应。同时,每一个感受器还能够对特定的时间频率反应,即对感受器方向上的图像速度成份进行编码。第二阶段,所有成份被非线性地组合在一起,以测量图象在每一个空间位置和空间频率上的运动速度。模型可以用来解释许多运动现象,包括似动、光栅运动和自然图象序列等。模型性质与人类视觉系统的知觉特性比较接近。

#### 2.4 空间-时间能量模型(Spatiotemporal energy model)

这是由Adelson和Bergen<sup>[5]</sup>提出的。模型是建立在理想的机制上的。它包含两个重要的概念,即单元(unit)和通道(channel)。单元是指一个或一小组能够提取视野中某一位置简单特性的细胞。通道则由分布在特定的视野中一组单元构成。单元把其感受野内的光学感受器所获得的信息收集起来,再把不同时刻所获得的信息综合起来,从而产生适当的时间脉冲反应。具有方向选择性的过滤器将有选择地对某一方向的运动反应。这种过滤器由视觉系统中可分离机制构成。

运动刺激可以被看成是一个三维空间,其中的X、Y是两个空间维度,T是时间维度。一个静止线段可以激活若干单元,其区域形状就象在X、Y坐标中的一个线段一样。一个运动点所激活的皮层单元区域形状也和静止线段所激活的皮层单元区域形状相似。觉察运动过程就是对两个空间维度在不同的时刻进行取样。当时间取样很慢时(如早期的电影),则觉察到的运动是跳跃的,但仍然能够看到。因为连续的和按一定时间间隔取样的刺激具有基本相同的空间-时间能量。所以,要获得运动印象,所取样刺激并非都必须相同。如果对运动的取样频率很高,则所觉察到的运动的连续性也高。在具有适当的空间-时间方向能量的条件下,也可以从连续的或离散的呈现方式中知觉到运动。为了提取空间-时间能量,这些过滤器被组合成若干个正交对(quadrature pairs),对其所包含的过滤器的输出结果进行平方和。再对两个线性过滤器(其中每一个过滤器都对同一方向的运动敏感,但敏感的相位相差90度)的输出进行组合,就可以得到一个相位独立的运动反应能量。然后进入非线性压缩阶段。这一阶段中,左、右能量觉察器相结合,构成了反向能量觉察器。该觉察器对向右的运动作正反应,而对向左的运动作负反应。对一个稳定的运动所产生的反应也是稳定的。符号取决于运动方向,而不取决于刺激的极性(polarity of stimulus)。

这一模型与Reichardt的修正模型(由van Santen和Sperling提出)关系密切,甚至相同。它们是用不同的方法来计算同一东西。它可以用来解释许多基本运动现象。例如,连续运动知觉、似动以及许多运动错觉现象(流动方波、倒转的 $\phi$ 现象等)。该模型中的运动觉察器不需要对每一时刻的位置变化进行计算。在运动能量提取过程中,不需要识别边界,不需要确定峰值的位置,也不需要划分界限等。空间-时间的方向是空间-时间刺激的局部特征,其提取机制与提取空间方向的机制相同。

空间-时间能量模型也有不少局限性。(1)它不适合用来解释B radick以及其他研究者所提出的长程运动过程,而用传统的匹配概念更容易解释。(2)它也不能解释无能量的运动(两个频率和方向不相同的光栅叠加时产生的运动)。但它的确可以使人们对一些基本运

动现象的低水平运动知觉有所了解。它也为分析运动知觉中的许多问题提供了概念性的工具。

### 3 小结

上述模型主要集中在对运动方向觉察的问题上,但不能精确地测量速度的大小。线性感受器只能简单地反映出特定通道整合而得出的谱能量的数值。这些通道的输出结果可以和许多其它线性通道的输出结果相比较,进而提供运动的测量结果。通道间的比较对于从多个通道获得速度大小信息也非常重要。另外,高阶单元(high-order units)通过对具有共同空间-时间特征的过滤器的输出结果进行求和,也可以获得速度信息。

这几种模型大都是对一阶运动的描述,基本上都把运动觉察器看成是一对空间-时间觉察器。两个相邻位置点的刺激含有许多不同模式,其中有适合视觉系统的最优刺激,也有非最优刺激。运动觉察器有能力从两个以上的空间-时间位置上进行空间与时间取样。

上述模型在一定程度上可以解释一些运动现象,许多模型主要是针对计算机的实现而提出的,它具有一定的量化特征,为计算机模拟提供了理论依据。模型的不断出现标志着对运动知觉研究从抽象的、模糊的阶段正走向成熟。由于人的视觉系统非常复杂,要想真正地模拟人的视觉功能,还必须对人的视觉系统的信息输入、加工等特点有更加充分地认识,进一步完善模型的理论基础。也许将来的运动模型需要一组输入觉察器,而不只是一对觉察器。

### 参考文献

- [1] Nakayama K. Biological Image Motion Processing: A Review. *Vision Research*, 1985, 25(5): 625- 660
- [2] Derrington A M, Suero M. Motion of complex patterns is computed from the perceived motions of their components. *Vision Research*, 1991, 31(1): 139- 149
- [3] van Santen J P H, Sperling G. Temporal covariance model of human motion perception. *Journal of the Optical Society of America* 1984, A, 1(5): 451- 473
- [4] Watson A B, Ahumada A J Jr. Model of human visual- motion sensing. *Journal of the Optical Society of America* 1985, A, 2(2): 322- 342
- [5] Adelson E H, Bergen J R. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America* 1985, A, 2(2): 284- 299

(上接第 15 页)

- [20] Fawzy FI, Fawzy NW, Hyun Cs, et al. Malignant melanoma: Effect of an early structured psychiatric intervention, coping, and affective state on recurrence and survival 6 year later. *Arch Gen Psychiatry*, 1993, 50(9): 681- 689
- [21] 吴传恩. 癌症病人的疼痛问题. *肿瘤*, 1991, 11(7): 64- 66