

视觉选择性注意的模型化计算及其应用前景

刘 琼, 秦世引, 李志成

北京航空航天大学自动化科学与电气工程学院, 北京 100191

摘要 视觉选择性注意 (Visual Selective Attention, VSA) 的模型化计算是当今认知信息处理领域的热点前沿研究方向, 在计算机智能视觉感知领域有着广阔的应用前景。对近年来视觉选择性注意研究领域的成果进行了回顾, 着重对中央区自下而上兴趣圈的产生、视觉任务和物体知识的表达、自上而下与自下而上的双向信息流的汇合以及显性注意分支中凝视点的转移控制等问题中具有较高影响力的模型化计算方法的效能和特点进行了分析与探讨, 对相关争论进行了归结和评述。进而通过 VSA 模型化计算在工程应用领域特别是视频压缩、目标检测、广告设计等方面的成功应用范例展示了其工程应用价值, 最后对其应用前景进行了展望。

关键词 认知信息处理; 主动视觉; 视觉选择性注意; 模型化计算

中图分类号 TP183

文献标识码 A

文章编号 1000-7857(2010)01-0107-09

The Modeling of Visual Selective Attention and Its Applications

LIU Qiong, QIN Shiyin, LI Zhicheng

School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

Abstract Visual Selective Attention (VSA) modeling is an advanced topic in cognitive information perception. It bridges neurophysiology, psychology and computer vision, and it aimed at developing a human-like visual system for computer. It shows a wide and promising application in intelligent computer vision from the late of last century. Firstly, recent VSA research topics in VSA modeling area are introduced in this paper, while the important and key problems are analyzed and discussed in details. In these topics, origin of bottom-up interesting map, representation of visual task and object, combination of bottom-up and top-down information, and fixation shifting in overt attention are taken as the highly influential aspects and thus described elaborately. The controversial issues in these topics are discussed as well. Secondly, in the view of application, the significance of VSA modeling are demonstrated by introducing and analyzing its successful application instances in video compression, target detection, and advertisement design. The inevitable problems and challenges of VSA modeling are summarized and its possible future applications from VSA modeling are prospect in the end.

Keywords cognitive information processing; active vision; visual selective attention; modeling

0 引言

人脑是一个内部结构极为复杂的系统, 它由 1000 亿个不同功能的神经元构成, 其中大量神经元通过多重前馈与反馈联系形成复杂的信息流网络, 协同完成记忆、注意、意识三大认知功能^[1]。“注意”在记忆与意识的引导下决定人类视觉感知的主动特性, 其中最为核心的视觉选择性注意 (Visual Selective Attention, VSA) 机制, 使人具备从复杂环境中搜索感兴趣目标的能力。同时, 这种主动行为特性也是当前机器视

觉区别于人类视觉的沟壑所在^[2-3]。人眼具有 VSA 特性主要源于大脑中可用资源的限制, 尤其是视觉系统接受的信息量与大脑中神经细胞的数目相差无几。其次, 外界环境的所有信息对于观察者来说并不是同等重要, 因此大脑只需要对部分重要信息做出响应。

融合更多的智能化因素是信息科学未来发展的必然趋势。所谓智能化信息处理即是指利用某种关于信息的经验和知识并将上层知识与下层处理相结合的信息处理方式^[4]。人

收稿日期: 2009-10-27

作者简介: 刘琼, 博士研究生, 研究方向为视觉选择性注意机理的模型化计算、图像处理与模式识别, 电子信箱: liuqiong1q@126.com; 秦世引 (通信作者), 教授, 研究方向为图像处理与模式识别、大规模复杂系统、多机器人混杂系统, 电子信箱: qsy@buaa.edu.cn

类视觉系统能瞬息感知外部世界,其主要原因是人脑中关于物体知识的记忆以及意识的引导与环境刺激驱动相结合所引发的视觉选择性注意起着重要的指向与汇聚作用,极大地提高了视觉感知的有效性^[5-6]。人类完美的视觉系统是智能化视觉信息处理系统的典范。近年来,随着脑科学、信息科学、计算神经科学和神经生理学等学科的发展,理解人类的视觉选择性注意机制,并应用信息科学领域的相关工具将其转化成可计算模型移植于计算机、赋予计算机智能化的视觉感知能力已成为可能。因此,对 VSA 进行模型化计算将为智能化视觉感知的研究奠定重要的理论与应用基础。

1 从 Marr 视觉计算理论到视觉选择性注意的模型化计算

1.1 Marr 视觉计算理论的历史性贡献

20 世纪 80 年代初,美国麻省理工学院的 Marr^[7]综合信息处理、心理学以及神经生理学的研究成果,从信息科学的角度提出了第一个较为完整的视觉计算理论,极大地推动了计算机视觉的发展与应用。这一理论的提出突破了以往仅在心理与生理学领域对人类视觉系统的定性描述,首次将视觉系统的计算提高到数学理论的严密水平。其次,理论的视觉处理三层次架构首次阐述了视觉系统机制,使得计算机视觉的研究有了—个比较明确的体系。因此,从提出到发展至今的 20 余年里,Marr 的计算理论一直是该领域研究者所尊崇的权威与主流。

Marr 视觉理论的核心问题是设法从图像结构推导出外部世界的三维结构,视觉从图像开始,经过—系列的处理和转换,最后达到对外部现实世界的认识。因此,图像的符号表征、图像处理算法以及硬件实现构成了计算理论中的三个主要部分。然而,图像是真实世界中物体的二维投影,物体在图像中的表征虽然仅是灰度的分布,却由几何特征、光照、物体材料表面性质、物体的颜色、摄像机参数等多种因素所决定。因此,由物体的二维图像投影反推其三维结构,若缺乏足够的约束条件,三维重建是一个病态问题,这也是计算理论中最富争议的论点^[8]。此外,视觉计算框架中的信息流通路是自下而上单一方向的,对高层知识的引导和反馈缺乏足够的重视,与人类视觉系统有目的的、主动的认知过程相距甚远。随着研究的深入,研究者逐步发现 Marr 视觉理论框架中存在的这些不足之处,尤其是其被动式的视觉感知^[9]。因此,知识引导下的主动视觉感知成为 Marr 理论较为完美的补充。同时,主动选择行为的引入,使得传统的 Marr 视觉框架中信息表示的计算约束变得易于解决^[11,10],并逐渐引起了众多研究者的注意。

1.2 VSA 模型化计算是认知信息处理的关键性命题

心理学领域有关“注意”的研究从 19 世纪 50 年代到 20 世纪 20 年代的 James 时代便已开始,期后提出的六大经典理论与模型仍然为“注意”的心理学研究者所推崇^[11-18]。除此之外,心理学研究者陆续对图形信息编码与表征^[19],记忆系统中图形不同特征的提取^[20]以及图画知觉过程的眼动研究^[21]等都

为 VSA 模型化计算提供了有利条件。信息科学领域若寻求构建与人类视觉系统相比拟的通用计算机视觉系统,需要综合利用心理学与神经生理学领域对 VSA 的研究成果并将其转化成可计算模型。

视觉是人类获取外界信息的主要渠道。人类的主动视觉 (Active Vision, AV) 行为是在大脑意识驱动下、记忆与注意共同引导的视觉感知方式。其中 VSA 作为 AV 的核心,最重要的功能在于协调外界信息量与人脑资源的不平衡,实现大脑资源的合理分配。目前,对 VSA 的研究已成为认知科学领域—个重要命题。其主要原因可以归结于以下几点:① VSA 强调有意识有目的的行为,是人类视觉系统的一种重要机制,同时也是揭开人类大脑奥秘的一扇窗户;② 视觉作为人类最重要的感知手段,与听觉、嗅觉等感知方式相比,可以提供更丰富的信息;③ 计算机视觉的应用领域日益增多,迫切需要一种可以和人类相比拟的通用视觉系统,因此需要对人类视觉系统有更深刻的认知;④ 除了视觉具有选择注意行为,在听觉中也同样存在该机制,对 VSA 的模型化计算可以为听觉选择注意行为的建模提供必要经验。由此可见,VSA 的模型化计算毫无疑问将是认知计算领域的关键性命题,对于该领域的长远发展具有不可估量的研究意义与价值。

1.3 从生物实验到模型化计算

纵观人类发展的历史长河,可以发现许多给生活带来极大便利的工具大部分都是人类受大自然万千生物的某种生理机制、结构或功能启发的结果。在现代信息科学领域,任何一种仿生功能的实现,从机制的发现到最终为人类效力,都离不开由生物实验、临床试验到模型化计算的发展历程,人类视觉系统机制的模型化计算也不例外。

选择性注意机制的大脑成像研究先驱是心理学家 Michael Posner 和神经学者 Marcus Raichle^[22]。他们与生理学、临床神经病学相交融的研究很快激起了整个神经科学界的兴趣,并且随着新技术测量大脑局部活动能力的发展,神经科学家对这类结合了认知心理学和大脑成像的新技术的研究也日趋增多。20 世纪 90 年代,神经科学家开始使用正电子发射计算机断层扫描 (PET) 以及功能性核磁共振成像 (fMRI) 技术,得到各种刺激条件下大脑的医学影像及相应的行为反应,推想大脑完成“注意”任务的工作机制。目前生理学普遍认为视觉信息处理通路分为三个阶段:最早的处理阶段包括视网膜、侧膝体 (Lateral Geniculate Nucleus, LGN)、主视皮层区 (Primary Visual Cortex),在灵长目动物中也称为 V1 区,都是通过提取简单的局部特征如中心-外周感受野以及具有朝向的线与边缘进行编码来表达图像。这种编码来源于去相关与冗余的计算准则或者采用稀疏编码对输入图像可靠的重建。早期处理过后,有效且适度的复杂特征在 V4 以及与其相邻的 TEO 区表达,最后,部分或完整的感兴趣物体视图在 IT 皮层的前区表达^[23]。VSA 在该处理通路中发挥了重要的作用,心理学在这一方面的研究主要采用眼跟踪设备与各种刺激呈现相结合的方法实时跟踪被试者眼球运动轨迹及其注视

区域,并通过对不同条件下得到的数据进行分析,统计决策人眼的选择注意行为与不同刺激的关系^[24-26]。

神经生理学及心理学领域的实验结果给 VSA 模型化计算的研究提供了可借鉴的理论依据,在某种程度上可以说,生理实验研究工具、方法以及实验手段等各方面的进步是 VSA 模型化计算具有可验证性的重要保障,为模型化计算提供可靠的生理理论依据,并推动其大踏步的发展。从生物实验到模型化计算,将生物生理机制在心理生理科学的定性描述提升到信息科学数理水平的定量描述,是技术变革的重要源泉,同时还还原论方法的观点也对计算模型的验证提供了支持。在从生物实验到模型化计算这一转变过程中涌现出了大量的研究成果,其中较为有名的当属 1965 年 Rodieck^[27]的视觉感受野模型。他提出利用两个高斯型函数的差即 DoG (Difference of Gaussian) 函数描述视网膜和侧膝体上同心圆型感受野,如下式所示:

$$DoG(r)=A\exp(-r^2/\sigma_A^2)-B\exp(-r^2/\sigma_B^2)$$

式中, r 为感受野中任一点到中心点的距离,并由 A 、 B 及 σ_A 、 σ_B 的关系模拟 ON 中心型感受野或 OFF 中心型感受野,该模型被后续诸多研究者采纳^[28]。其他的如对朝向选择性细胞的模拟^[29]和对纹理选择性细胞的模拟^[30]等等。诸如此例,都为 VSA 的模型化计算提供了可借鉴的经验。

2 VSA 模型化计算的发展历史和研究现状

VSA 的模型化计算从 1985 年 Koch 和 Ullman^[31]在 *Human Neurobiology* 上的一篇探讨视觉选择注意与转移的文章为开端至今,陆续引起众多学者的关注。虽然后续提出的各个模型所偏倚的重点有所区别,但核心还是围绕如何构建由刺激引发的自底向上的数据驱动与任务引导下的自顶向下的目标驱动的视觉注意的数学模型所展开,以及由此而引发的信息流通路的分叉与汇合问题的研究。

2.1 Koch-Ullman 模型的突破性贡献及其局限性

Koch 和 Ullman 于 1985 年提出的视觉选择注意模型开创了 VSA 模型化计算的新纪元。显著图(Saliency Map)的概念也是由他们在那时提出。Koch-Ullman 模型第一次提出了先计算后选择的观点,并提出利用“赢者全取”(Winner Take All, WTA)与抑制返回(Inhibit of Return, IOR)机制,实现凝视觉点的选取与转移,且转移必须遵循两条规则,一是距离优先,二是特征相似优先。该模型奠定了后续模型的理论根基,并蕴含了许多对 VSA 模型化计算具有重要影响的关键思想。

Koch-Ullman 模型是对 Marr 计算理论的有效补充,特别是在单帧图像的感兴趣信息获取机制上,更接近于人类的主动选择行为特性,并通过显著图理论在计算层次上对其补充。但由于受当时实验条件所限,主要还是注重于理论上对 VSA 模型化计算的研究,而且相关文献中也只展示了对简单的人工图像进行的仿真实验。

2.2 20 世纪 80 年代中后期以来的竞争性研究历史回顾

随着 Koch-Ullman 模型的提出以及研究者对 Marr 计算

理论中的“Shape from X”的质疑呼声高涨,主动视觉的研究走进了计算机视觉领域,并日渐引起大家的关注,兴起了一股研究的热潮。

20 世纪 90 年代初期 Tsotsos^[32]主要从复杂性分析与最小代价耗费的角度对视觉感知的有效性进行衡量,视觉感知的可计算性分析是作者在 VSA 模型化计算领域做出的最重要的工作。该模型中体现的主要思想是选择调制以及自顶向下、由粗到精、分等级的赢者全取机制增强注意区域的选择与定位,对没有被选择的连接采取抑制的方式。Olshausen、Anderson 与 van Essen^[33]在 1993 年提出的模型,其主要思想体现在注意机制使得物体在视野中具有方位与尺度不变特性,从而实现识别。该模型依赖于一系列的控制神经元动态的修改皮层间的连接权值,使得 V1 区的有限信息有选择地流至高级皮层区。神经元的动态性能由简单的微分方程模拟神经生物学的类似循环机制实现。该模型第一次将视觉选择注意与识别联系,并认为在识别阶段,控制神经元主要由记忆与视网膜信息的交互而驱动。其后, Wolfe^[34]在 GS 2.0 模型中,突出前期刺激在广泛调制的类别通道中被过滤,并通过比较自下而上和自上而下两者的局部差异输出激活特征图,根据对激活特征图的加权和形成激活图,按照活跃度的强弱分配后续有限的资源。模型中自上而下信息流对自下而上任务的权值分配按照特征的唯一性与可辨别性两种能力分配。这一模型最早引进了自上而下对特征的调制,并给出不可能同时拥有两个甚至多个注意凝视觉点的关键思想。

国内郑南宁^[1]在 20 世纪 90 年代中期也在 VSA 的模型化计算领域做出了显著贡献。他提出的 VSA 计算模型主要强调自下而上与自上而下注意的融合,通过对视觉信息传播通路的分析,采用非线性动力系统方法对双向信息流的闭环控制机制建模,其建模的动力学思想为视觉信息处理的非线性过程的进一步研究打下了基础。

2.3 Itti-Koch VSA 计算模型的里程碑作用

回顾 20 世纪 80 年代中后期以来 VSA 模型化计算竞争性研究的历史局面,绝大部分模型只是提出建模与计算的思想并对其进行定性的分析,因此将其提高到数理水平的定量计算主导了其 VSA 模型化计算的研究方向。Itti 和 Koch^[28]于 1998 年在 *IEEE Transactions on PAMI* 上发表了一篇关于视觉场景快速分析的文章,其中最主要的贡献就是将 VSA 机制从本质上推向了模型化定量计算的新阶段,使 VSA 在从神经生理学的机制理论研究到信息科学领域的可计算性分析,进而走向定量计算的工程应用的发展过程中发挥了重要的里程碑作用。该模型模拟了人类的视觉选择注意机制,其算法结构如图 1 所示。根据人脑 V1 区视觉神经对颜色、亮度和方向等特征较为敏感的特点,模型依次通过提取上述 3 种特征通道并对 3 通道采用多尺度分解运算与多尺度差异运算仿生人眼的多分辨率分析特性与视觉神经的“中央-外周”(Center-Surround, CS)机制,最后将多尺度下的多组特征图进行竞争合并得到显著图。在显著图中,依赖 WTA 竞争机制和

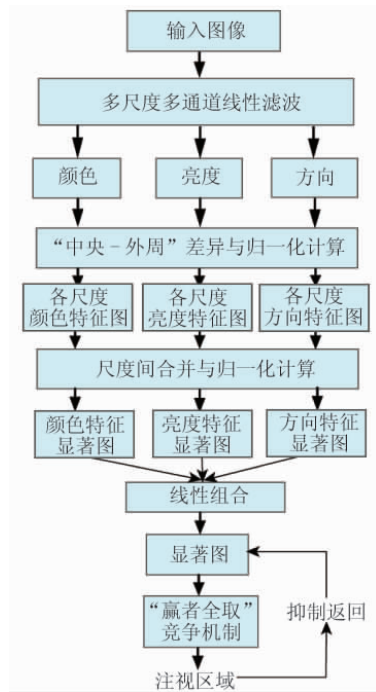


图 1 Itti-Koch 模型结构

Fig. 1 General architecture of the Itti-Koch model

抑制返回方式确保模型能在不同显著程度的区域间实现注视转移。

该模型继承了特征整合理论的思想并结合感受野的 CS 机制、WTA、IOR，运用合理的数学工具对上述机制进行定量计算。相对已有的模型客观分析，Itti-Koch 模型在生理机制上并没有加入新的成分，且采用的数学工具与计算方法也较为经典，但最为巧妙之处是模型首次运用数学工具对已有生理机制的仿生建模，沟通了神经生理学、心理学与计算机科学、数理科学的桥梁，突破了信息科学领域长期以来只是在理论与仿真层面对 VSA 研究的局面。将该模型应用于真实场景实验并结合眼球跟踪实验，结果表明该模型在预测视觉注意方面具有较高的准确性和鲁棒性，这有力地促进了 VSA 模型化计算在信息科学领域的工程化应用。受到当初生理学研究进展的影响，模型没有考虑自上而下的信息流及其所涉及与影响的各个方面，这也是文献作者目前在积极探讨与寻求解决的问题。

2.4 进一步的争论和发展

通过 VSA 的模型化计算实现人类具有选择性注意机制的主动视觉行为，如同 Marr 视觉计算理论的提出与发展，是一次艰难的摸索与实践。由于神经生理学以及心理学领域对于人类视觉系统的研究并没有唯一性，缺乏评判的标准，而且人类的行为特性也因人而异，因此上升到 VSA 的模型化计算会存在类似的难题，但这是一个不断反馈修正的研究过程，随着各种先进工具与知识的出现，通过 VSA 的模型化计算以高可靠性仿生人类的视觉选择性注意机制应该不会只是痴人说梦。

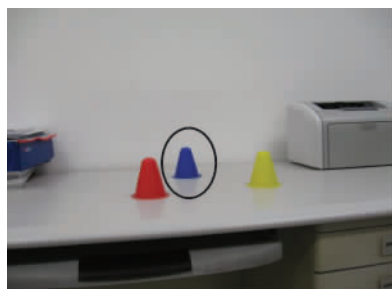
通过对 VSA 的模型化计算进行分析与细化，得出关键环节主要突出在以下 4 个方面：如何产生中央区自下而上兴趣图；视觉任务和物体知识的表达；自上而下与自下而上双向信息流的汇合以及显性注意中凝视点的转移控制。这 4 个方面也是当前 VSA 计算领域发展中的主要研究内容。

中央区自下而上兴趣图的产生在 VSA 的研究中，其核心问题在于如何选择一种合适的显著性度量方法，使得选择的区域能够在其他区域中突显 (pop-out)^[31]，即找到输入图像中的感兴趣区域。目前针对自下而上兴趣图的产生主要以目标先验信息是否存在为算法研究的依据。当存在先验目标信息时，由于人脑执行基于内容语义的高级视觉注意，与视觉任务、物体和环境的知识有关，并且与模式识别和匹配密不可分，因此较多的文献中采用的方法主要通过提取目标图像与当前图像的兴趣点^[39]（如 SIFT 特征点、Harris 角点等）或其他局部特征^[35-38]，再通过图像匹配以及学习的方法获取 ROI。当目标先验信息不存在的时候，就像把人置于完全陌生的环境中，脑认知功能通过对于“pop-out”研究的实验结果认为此时刺激特征主要表现在当刺激物之间由低对比度转向高对比度时能引起感受野细胞的空间重组，从而引起观察者的注意^[40-41]。因此，采用对比度衡量显著性将是自底向上模型合理仿生的途径。目前已有 Wolfe^[41]提出的 GS 2.0 模型，Itti^[28]提出的 Saliency 模型等都体现了对比即突显的思想。此外也有模型采用香农信息熵度量显著性^[42]以及信息熵与尺度相结合的方法^[43-46]等。这些方法大部分都通过信息熵来衡量图像区域中信息量的大小，从而衡量该区域的复杂性程度，目前这种方法已经在识别中也得到了越来越广泛的应用^[47-48]。

视觉任务和物体知识的表达虽然在当前信息科学领域中已有许多主流的方法，如结构描述法、几何约束法、多维特征空间表达法以及近似特征空间表达法等等，但是与人类大脑记忆中任务与物体知识存储相切合的表达模式，是尚待解决的问题。采用语义网络模型与多维特征空间表达相结合的方法，能够在仿生学与生理学的交叉领域找到一个较好的折中点。语义网络模型的引入可以用来描述物体之间的关系，而多维特征空间能表达物体的特征，因此两者的结合能够多方位地体现大脑中存储的知识。语义内容已经被证实是最佳搜索路径的预测线索^[49]，因此，建立大规模的真实世界语义模型对选择性注意机制建模具有重要的意义。目前，大部分语义模型的构建都是基于科林斯和奎连的方法和理论模型^[50]。他们以层次网络作为语义记忆的核心，主张认知经济原则为主要的组织原则。如奎连层次模型采用层次结构的形式，名词概念在语义层次中一层一层排列，属性贮存在具有这种属性的最高或最一般的节点上。搜索过程所需时间与层次间差异成正比。目前，运用语义网络模型表达的调节控制注意选择区域在文献[51]中已有描述。但是该文中的语义模型缺乏对抽象物体的连接，因而模拟实现人类的工作记忆尚须完善。

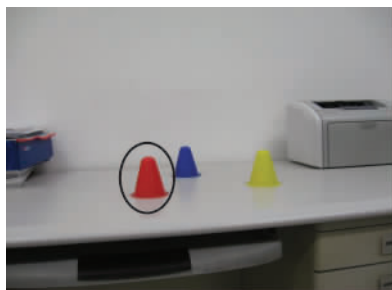
VSA 是在自上而下与自下而上双向信息流融合的基础

上产生的结论已经得到了充分的证明,并早在 James 时代就已有描述^[52]。目前在 VSA 模型化计算的研究领域,对于双向信息流的汇合,常见的方法主要是通过得到目标先验信息,再加以对底层特征的调制,即信息流在显著图汇合(激活)^[28-31]。至于两者之间,究竟谁执导谁的问题阐述,较为有名的是 Wolfe 在特征整合理论的基础上提出的导向搜索理论(Guided Search)^[34]。该理论假设存在两个加工阶段,首先是特征加工阶段,类似于特征整合理论的第一阶段;在第二阶段上,Wolfe 认为在特征识别之后,来自自上而下和自下而上的信息可以联合起来对注意进行引导,在这个过程中那些和被试期望相匹配的奇异刺激最有可能获得注意,两种信息的结合可以使对复合刺激的搜索快速完成。这就澄清了特征整合理论没能解释的对复合特征的搜索也可以平行进行,该思想在文献^[53]中也有体现。文献^[54]则通过试验更进一步证实自上而下的线索对底层视觉特征的调制是精细粒度,即目标与分心物在单一的特征维内所在的区间也能影响搜索的速度^[55]。如图 2 是在白天拍摄的一幅简单的室内图像,图 2(a)是在没有目标导引的情况下对灰度、颜色和方向特征分配同等权值时得到的显著目标,图 2(b)是考虑在白天光照条件下,根据色彩理论对三者权值的分配遵循高斯分布所得到的显著目标,其中颜色信息得到峰值权值。选取 10 个事先没有任何先验信息且视觉正常的被试者参与实验,根据被试的试验结果,80% 选择红色路标模型为显著物体。由此可见,对底层特征各特征维部署不同的特征权重得到的显著区域也会大相径庭。同时也有文献中的方法通过最优化策略得到不同特征维的自



(a) 无权重分配时的计算结果

(a) No priority assigned result



(b) 颜色特征占较大权重时的计算结果

(b) Color priority oriented result

图 2 特征分配不同权值时对同一副图像得到的不同结果

Fig. 2 Experiment results under different feature priority

下而上显著图的相关权值,并通过学习得到的统计信息优化自底向上信息流中各底层特征信息的权重,提升目标显著性,抑制背景显著度^[56]。

从信息科学领域的角度审视两者结合的问题,针对目前生理学的研究成果,权值调制是一种可行的途径^[57]。但从权值调制中引发的权值设置也是信息科学领域中机器学习研究的热点问题,目前已有基于 Bayes 网络的统计学方法^[58],基于场景内容的方法^[59]等用来解决 VSA 模型化计算中的该类问题。

注意可区分为外显性注意与内隐性注意。外显性注意是指直接把感觉器官转向外界刺激来源的动作,而内隐性注意是对几个可能的感觉刺激中的一个产生知觉集中的行为,发生在固定的视场范围内^[59]。目前在 VSA 模型化计算研究领域更多的研究都关注于内隐性注意,即缺乏眼动的注意,而实际上人眼正常的生理现象既包含显性注意也包含隐性注意。对于内隐性注意的转移控制较为经典的实现来源于最初 Koch 在 1985 年发表的文献,其中采用返回抑制的竞争网络根据显著程度依次扫视相关区域^[31]。由于 IOR 机制得到了神经生理学的实验证明^[26,60-62],此后众多的关于隐性注意转移控制的模型化计算工作也都采用与 IOR 竞争网络相似的方法实现^[52,63]。对于外显性注意转移控制的计算建模其最大的挑战在于当眼动发生时如何实现显著图发生转移的补偿机制,以及由此滋生出来的扫视抑制、显著图动态构建、返回抑制、发生显性扫视前的隐性注意目标选择并由此上升的复杂视觉场景的在线理解等^[64-65]。

3 VSA 模型化计算的工程范例和应用前景

随着信息科学领域对计算机智能化水平的发展要求以及随之而来的相关应用对算法智能化性能要求的不断提高,赋予计算机主动视觉行为是 VSA 模型化计算研究的最终目标。工程化背景下如何在不同的应用场合中,满足特定的可靠性与时效性要求是 VSA 模型化计算取得成功的基础与重要前提。目前,已经在目标检测^[57,66-67]、目标识别^[68-72]、图像压缩^[57,73]、视频压缩^[74]、机器人自主定位^[75-76]等领域都获得了较好的应用。

3.1 在目标检测中的应用

VSA 通过引导高效可靠的视觉感知成为众多灵长目类动物一项重要的心理调节机制,以过滤筛选的方式提取视觉感知中的有效信息即目标相关信息,从而合理分配大脑视皮层中的信息加工资源^[77]。因此,其应用的直观体现即是目标检测。如图 3 是文献^[57]中的目标检测实例结果。复杂场景中具有丰富的纹理信息,在局部范围内亮度信息、阴影与反射都有较大差异,大量物体都发生了部分遮挡。被试(人)找到车辆平均花费时间为 2.6s。图 3(b)为运用模型找到的车辆位置,模型的第一个显著目标为车辆,花费时间为 2.2s。由此可见,人眼的定性分析与模型的定量计算相比较,计算机在运算能力上比人类更具优势。

3.2 在图像压缩中的应用

VSA 吸引研究者目光的优点主要表现在能根据事先设



(a) 原始图像
(a) Original image



(b) 目标检测结果
(b) Object detection result

图3 VSA 模型化计算在目标检测中的应用

Fig. 3 VSA computational model applied in object detection

定的显著性准则突显场景中的有效信息,因此将其应用于图像压缩领域具有重要的社会价值与研究意义。图像压缩的本质就是利用图像数据的相关性,充分去除图像数据的各种冗余,如信息熵冗余、视觉冗余等,用尽可能少的位数表示和重建原始图像^[78]。目前,图像压缩主要采用的是基于变换域编码的 JPEG 算法,以消除图像数据的空间相关性。根据人们日常观察图像的经验得知,有时候人们关注的仅仅是图像中子区

域的信息,因此采用 VSA 模型化计算对图像中的次要信息采用较高的压缩比能在传输信道的节省上做出重要贡献。图 4 是文献[57]中 VSA 计算模型在图像压缩中的应用,采用的是非标准的 JPEG 压缩算法。很明显,标准的 JPEG 压缩效果使得图像有较强的块状效应,且图像结构已经变得混乱。而文献中模型对图像显著区域保留了更多的信息,因此在视觉效果上,保真度更高。



(a) 原始图像
(a) Original



(b) JPEG 压缩效果
(b) JPEG



(c) 基于 VSA 显著性压缩
(c) Saliency-based

图4 VSA 模型化计算在图像压缩中的应用

Fig. 4 VSA computational model applied in image/video compression

3.3 在广告设计中的应用

VSA 模型化计算应用于广告设计主要是通过突出模型计算出的感兴趣区域吸引大众的注意。同时,还能将模型用于检验广告的设计效果。图 5 是将文献[57]中的 Saliency 模型用于杂志的封面设计。图中黄色圆圈标注的是封面中的感兴趣区域。区域的标注基于纯粹的自下而上的线索,没有图像语义内容知识的引导。

VSA 模型化计算用于艺术与广告设计领域,其最大的优点是能评价广告产品的艺术效果。但是同样不能忽略的是,没有自上而下信息的引导,不能排除观察者对于广告的首次扫视将落在模型预测的结果之外。因此,不论是评价还是设计,应该同平均的人眼注视区域相比较而不是同带有典型目

的的人眼观测结果比较。

VSA 模型化计算除了在目标检测、图像/视频压缩以及广告艺术设计领域取得了重要的应用,同时目标识别、机器人自主定位等应用研究领域也广泛吸收 VSA 的选择注意特性实现显著区域的检测,快速识别目标及定位机器人,在实时性能和鲁棒性能上得到提高。此外,如智能人机交互等其他方面的应用不再赘述。

3.4 来自应用需求的发展动力

视觉是生物感知外界信息的主要渠道,让机器人也拥有视觉是人类长久以来的梦想。由于主动视觉具有选择性注意的特性,因此相对被动视觉行为来说,能够更加自主且有效地获取感兴趣信息。在人无法靠近的危险环境中,机器人可



图 5 VSA 模型化计算在广告设计与评价中的应用

Fig. 5 VSA computational model applied in advertisement and art design and evaluation

以通过主动视觉方式并根据先验的目标信息甚至在缺乏先验信息的情况下自主找到目标或异常物体；在军事战争中，无人驾驶装备可依赖主动视觉系统准确并快速发现敌方目标；在日常生活中，无人车可依赖主动视觉系统识别交通标志及各种异常情况，达到安全行驶。总之，凡是一切需要利用视觉信息的场合都可以采纳 VSA 计算模型快速有效地获取感兴趣信息，尤其是在信息量日益膨胀的今天，VSA 机制用于信息检索领域会大量减轻网络压力，提高检索效率。

硬件设计与软件研发是当前技术变革与知识创新的重要手段。主动视觉认为在视觉信息获取过程中，应更主动地调整摄像机的参数，如方向、焦距、孔径，等并能使摄像机迅速对准感兴趣的物体。更一般地说，它强调注意机制 (Attention)，强调对分布于不同空间范围和时间段上的信号采用不同的分辨率有选择性地感知^[79]。因此，VSA 模型化计算是实现计算机视觉系统整体主动选择行为的关键部分，具有举足轻重的地位。

3.5 应用前景

由于 VSA 模型化计算综合利用神经生理学、心理学及信息科学领域的知识，是将生物生理特性应用于实际工程的重要探索，符合人工智能发展的长期目标^[80]，基于人类感知与认知机制的智能化信息处理将对整个自然科学和技术科学产生深远和重大的影响，其将可能领导整个信息科学领域的重大变革^[4]。

展望未来，VSA 模型化计算面临巨大的应用需求。视觉感知是任何视觉系统设备不可缺少的信息获取途径，特别是随着视觉传感器设备的发展与成本降低，视觉感知成为了主流的信息获取渠道。现在，在国计民生的众多行业和领域都有机器视觉技术的应用，如智能交通安防、医疗制药行业、纺织行业以及电子行业等。此外，在军事领域也不乏机器视觉技术的用武之地，如军事目标检测、高精度目标跟踪、成像制导、救援机器人视觉导航、基于视觉导航的无人机自主着陆、

天文导航等等。因此，在已有应用领域提高视觉感知的主动性是未来 VSA 模型化计算应用的主要方面；其次，将 VSA 模型化计算理论拓展到听觉、嗅觉选择性注意，综合利用视、听、嗅觉信息完成环境感知是 VSA 模型化计算应用的重要方面；此外，在信息技术发展日益强大的当今社会，运用 VSA 模型化计算及其延伸的视听觉选择性注意研究成果，可以有效地防控海量信息拥堵导致的系统崩溃，是关系到国计民生的重要举措，具有重要的社会价值。

4 结论

回顾 Koch 与 Ullman 在 1985 年开创 VSA 模型化计算到至今发展的 20 余年间，视觉选择性注意机制突破了传统的只重视视觉表象开展视觉信息研究的框架，充分模拟灵长目类动物主动的、有目的的获取任务目标的选择性机制实现智能化的信息处理，在信息科学领域的研究取得了突飞猛进的增长，并日益得到专家与学者的重视。

从目前对选择性注意成功的应用例证中可以看出，视觉选择性注意能够有效地捕捉重要信息，且实时性和鲁棒性较传统的算法都有较大的改进。但是它所面临的挑战也体现在以下 3 点：① 如何合理地融入记忆与意识对自底向上视觉搜索的影响；② 如何找到衡量模型性能的标准；③ 如何将显性的眼动注意予以考虑，并将其与隐性注意相融合，找到它们之间的切入点。同时，在生理学的论证上也急需得到一些问题的答案，比如大脑中信息的传播是否只是简单的循环，到底是分层的等级竞争还是只在形成显著图时存在竞争等。

本文从信息科学领域的角度出发针对模型化计算中存在的若干疑难问题并结合国内外对 VSA 相关研究的已有研究成果及大量现有的模型进行分析，针对经典模型中所涉及的关键技术进行了深入探讨，同时也给出了现有的视觉选择注意的模型化计算中所面临的问题与挑战。最后在成功应用范例的基础上对 VSA 模型化计算的应用前景进行了长远展望。

参考文献 (References)

- [1] 郑南宁. 认知过程的信息处理和新型人工智能系统[J]. 中国基础科学. 2000(8): 9-18.
Zheng Nanning. *China Basic Science*, 2000(8): 9-18.
- [2] Bajcsy R. Active perception [J]. *Proceedings of the IEEE*, 1988, 76(8): 966-1005.
- [3] 张广军. 机器视觉[M]. 北京: 科学出版社, 2005.
Zhang Guangjun. *Machine vision*[M]. Beijing: Science Press, 2005.
- [4] 史忠植. 智能科学[M]. 北京: 清华大学出版社, 2006.
Shi Zhongzhi. *Intelligence science* [M]. Beijing: Tsinghua University Press, 2006.
- [5] 罗四维, 等. 视觉感知系统信息处理理论 [M]. 北京: 电子工业出版社, 2006.
Luo Siwei, *et al.* *Information processing theory of visual perception*[M]. Beijing: Publishing House of Electronics Industry, 2006.
- [6] 危辉, 潘云鹤. 形象思维的视表象基础 [J]. 浙江大学学报: 工学版, 2001, 35(2): 152-156.
Wei Hui, Pan Yunhe. *Journal of Zhejiang University: Engineering Science Edition*, 2001, 35(2): 152-156.
- [7] Marr D. 视觉计算理论[M]. 姚国正, 刘磊, 汪云九, 译. 北京: 科学出版社, 1988.
Marr D. *Computational theory of vision* [M]. Yao Guozheng, Liu Lei, Wang Yunjiu, trans. Beijing: Science Press, 1988.
- [8] Tarr M J, Black M J. Dialogue: A computational and evolutionary perspective on the role of representation in vision [J]. *CVGIP: Image Understanding*, 1994, 60: 66-75.
- [9] Aloimonos Y, Rosenfeld A. A response to "ignorance, myopia, and navieté in computer vision systems" by R. C. Jain and T. O. Binford[J]. *CVGIP: Image Understanding*, 1991, 53(1): 120-124.
- [10] Wolfe J M. Opinion: What attributes guide the deployment of visual attention and how do they do it[J]. *Nature Neuroscience*, 2004, 5: 1-7.
- [11] Broadbent D E. Perception and communication [M]. New York: Pergamon Press, 1958: 19-58.
- [12] Oswald I, Taylor A, Treisman M. Discriminative responses to stimulation during human sleep[J]. *Brain*, 1960, 83: 440-453.
- [13] Duetsch J A, Duetsch D. Attention: Some theoretical considerations[J]. *Psychological Review*, 1963, 70: 80-90.
- [14] Kahneman D. Attention and effort [M]. New Jersey: Prentice Hall, 1973: 240.
- [15] Treisman A M, Gelade G. A feature-integration theory of attention[J]. *Cogn Psychol*, 1980, 12(1): 97-136.
- [16] Treisman A M. Feature binding, attention and object Perception [J]. *Phil Trans R Soc Lond*, 1998, 353: 1295-1306.
- [17] Desimone R, Duncan J. Neural mechanism of selective visual attention [J]. *Annu Rev Neurosci*, 1995, 18: 193-222.
- [18] Kastner S, Ungerleider L G. The neural basis of biased competition in human visual cortex[J]. *Neuropsychologia*, 2001, 39(12): 1263-1276.
- [19] 丁锦红, 林仲贤. 图形的信息编码与表征[J]. 心理学动态, 1999, 7(1): 1-5.
Ding Jinhong, Lin Zhongxian. *Journal of Developments in Psychology*, 1999, 7(1): 1-5.
- [20] 丁锦红, 林仲贤. 记忆系统中图形不同特征的提取 [J]. 心理科学. 2001, 24(3): 273-275.
Ding Jinhong, Lin Zhongxian. *Psychological Science*, 2001, 24(3): 273-275.
- [21] 陶云. 图画知觉过程的眼动研究进展[J]. 心理科学. 2001, 24(2): 194-196.
Tao Yun. *Developments in the study of eye movements in scene perception*[J]. *Psychological Science*, 2001, 24(2): 194-196.
- [22] 维基百科-"注意"[EB/OL]. <http://en.wikipedia.org/wiki/Attention>, 2009.
Wikipedia - "attention" [EB/OL]. <http://en.wikipedia.org/wiki/Attention>, 2009.
- [23] 郝晶. Functional MR imaging in human visual system [J]. 中国医学影像技术, 2004, 20(4): 634-636.
Hao Jing. *Chinese Journal of Medical Imaging Technology*, 2004, 20(4): 634-636.
- [24] Solomon J A, Morgan M J. Facilitation from collinear flanks is cancelled by non-collinear flanks[J]. *Vision Research*, 2000, 40(3): 279-286.
- [25] Sceniak M P, Hawken M J, Shapley R. Contrast-dependent changes in spatial frequency tuning of macaque V1 neurons: Effects of a changing receptive field size[J]. *Journal of Neurophysiology*, 2002, 88(3): 1363-1373.
- [26] Dukewich K R, Boehnke S E. Cue repetition increases inhibition of return[J]. *Neuroscience Letters*, 2008, 448(3): 231-235.
- [27] Rodieck R W, Stone J J. Analysis of receptive fields of cat retina ganglion cells[J]. *Journal of Neurophysiology*, 1965, 28: 833-849.
- [28] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(11): 1254-1259.
- [29] Hubel D H, Wiesel T N. Receptive fields, binocular interaction, and functional architecture in cat's visual cortex [J]. *Journal of Physiology*, 1962, 160: 106-154.
- [30] Kruizinga P, Petkov N. Nonlinear operator for oriented texture [J]. *IEEE Trans on Image Processing*, 1999, 8(10): 1395-1407.
- [31] Koch C, Ullman S. Shifts in selective visual attention: Towards the underlying neural circuitry[J]. *Hum Neurobiol*, 1985, 4(4): 219-227.
- [32] Tsotsos J K. Analyzing vision at the complexity level [J]. *Behavioral and Brain Sciences*, 1990, 13(3): 423-445.
- [33] Olshausen B E. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information[J]. *J Neuroscience*, 1993, 13(1): 4700-4719.
- [34] Wolfe J M. Guided search 2.0: A revised model of visual search[J]. *Psych Bull Rev*, 1994, 1: 202-228.
- [35] Oliva A, Torralba A, Castelano M S, *et al.* Top-down control of visual attention in object detection[C]//*Proceedings of International Conference on Image Processing*. 2003: 253-256.
- [36] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope [J]. *International Journal of Computer Vision*, 2001, 42(3): 145-175.
- [37] Yaoru S. Hierarchical object-based visual attention for machine vision [D]. Oxford: University of Edinburg, 2003.
- [38] Torralba A. Contextual priming for object detection [J]. *International Journal of Computer Vision*, 2003, 53(2): 169-191.
- [39] Wang F, Wei J, Zheng N, *et al.* Automatic ROI selection for JPEG 2000 compression of remote sensing images [C]//*Proceedings of International Conference on Semantic Computing*. 2007: 615-621.
- [40] Solomon J A, Morgan M J. Facilitation from collinear flanks is cancelled by non-collinear flanks[J]. *Vision Research*, 2000, 40(3): 279-286.
- [41] Mareschal I, Henrie J A, Shapley R M. A psychophysical correlate of contrast dependent changes in receptive field properties [J]. *Vision Research*, 2002, 42(15): 1879-1887.
- [42] Gilles S. Robust description and matching of images [D]. Oxford: University of Oxford, 1998.
- [43] Jagersand M. Saliency maps and attention selection in scale and spatial coordinates: an information theoretic approach [C]//*Proceedings of*

- International Conference on Computer Vision. 1995.
- [44] Ferraro M, Boccignone G, Caelli T. On the representation of image structures via scale space entropy conditions [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(11): 1199–1203.
 - [45] Kadir T. Scale, saliency and scene description[D]. Oxford: University of Oxford, 2002.
 - [46] 刘昌. 人类工作记忆的某些神经影像研究 [J]. *心理学报*, 2002, 34(6): 634–642.
Liu Chang. *Acta Psychologica Sinica*, 2002, 34(6): 634–642.
 - [47] Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification [J]. *Nature Neuroscience*, 2002, 5(7): 682–687.
 - [48] Epshtein B, Ullman S. Feature hierarchies for object classification[C]// *Proceedings of International Conference on Computer Vision*. 2005.
 - [49] Cerf M, Frady E P, Koch C. Using semantic content as cues for better scanpath prediction [C]// *Proceedings of the eye tracking research and applications symposium*. 2008: 143–146.
 - [50] 王益文, 林崇德. 信息保持、短时存贮与执行控制的脑模型[J]. *心理科学进展*, 2004, 12(5): 661–671.
Wang Yiwen, Lin Chongde. *Advances in Psychological Science*, 2004, 12(5): 661–671.
 - [51] Navalpakkam V, Itti L. A goal oriented attention guidance model[C]// *Biologically Motivated Computer Vision, Proceedings*. 2002, 2525: 453–461.
 - [52] James W. The principles of psychology [M]. New York: Henry Holt, 1890.
 - [53] Wolfe J M, Butcher S J, Lee C, et al. Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons [J]. *Journal of Experimental Psychology-Human Perception And Performance*. 2003, 29(2): 483–502.
 - [54] Navalpakkam V, Itti L. Top-down attention selection is fine grained[J]. *Journal of Vision*, 2006, 6(11): 1180–1193.
 - [55] Navalpakkam V, Itti L. Search goal tunes visual features optimally[J]. *Neuron*, 2007, 53(4): 605–617.
 - [56] Navalpakkam V, Itti L. An integrated model of top-down and bottom-up attention for optimal object detection [C]// *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2006.
 - [57] Itti L. Models of bottom-up and top-down visual attention [D]. Pasadena, California: California Institute of Technology, 2000.
 - [58] Mozer M C, Shettel M, Vecera S. Top-down control of visual attention: A rational account[C]// *Proceedings of the Neural Information Processing Systems*. 2005: 923–930.
 - [59] Wright R D, Ward L M. Orienting of attention [M]. Oxford: Oxford University Press, 2008.
 - [60] Klein R M. Inhibition of return [J]. *Trends in Cognitive Sciences*, 2000, 4(4): 138–147.
 - [61] Chica A B, Lupianez J. Effects of endogenous and exogenous attention on visual processing: an Inhibition of Return study[J]. *Brain Res*, 2009, 1278: 75–85.
 - [62] Ludwig C J H, Farrell S, Ellis L A, et al. The mechanism underlying inhibition of saccadic return [J]. *Cognitive Psychology*, 2009, 59 (2): 180–202.
 - [63] Yaoru S, Fisher R. Object-based visual attention for computer vision[J]. *Artificial Intelligence*, 2003, 146(1): 77–123.
 - [64] Rao R, Zelinsky G J, Hayhoe M M, et al. Eye movements in iconic visual search[J]. *Vision Research*, 2002, 42(11): 1447–1463.
 - [65] 龙甫荟, 郑南宁. 一种引入注意机制的视觉计算模型[J]. *中国图象图形学报*, 1998, 03(7): 592–595.
Long Fuhui, Zheng Nanning. *Journal of Image and Graphics*, 1998, 03 (7): 592–595.
 - [66] 汪哲慎. 复杂场景下交通标志检测技术研究[D]. 厦门: 厦门大学信息科学与技术学院, 2006.
Wang Zhesen. Technologies of traffic sign detection in complex scenes [D]. Xiamen: School of Information Science and Technology, Xiamen University, 2006.
 - [67] 桑农, 李正龙, 张天序. 人类视觉注意机制在目标检测中的应用[J]. *红外与激光工程*. 2004, 33(01): 38–42.
Sang Nong, Li Zhenglong, Zhang Tianxu. *Infrared and Laser Engineering*, 2004, 33(1): 38–42.
 - [68] Walther D. Interactions of visual attention and object recognition: Computational modeling, algorithms and psychophysics [D]. Pasadena, California: California Institute of Technology, 2006.
 - [69] Salah A A, Alpaydin E, Akarun L. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(3): 420–425.
 - [70] 吴田富. 视觉选择性注意机制计算模型及其在物体识别中的应用[D]. 合肥: 合肥工业大学计算机与信息学院, 2005.
Wu Tianfu. A computational model of visual selective attention mechanism and its application in object recognition[D]. Hefei: School of Computer and Information, Hefei University of Technology, 2005.
 - [71] Walther D, Itti L, Riesenhuber M, et al. Attentional selection for object recognition—A gentle way [J]. *Biologically Motivated Computer Vision, Proceedings*. 2002, 2525: 472–479.
 - [72] Schiele B, Crowley J L. Transinformation for active object recognition[C]// *Proceedings of International Conference on Computer Vision*. 1998: 249–254.
 - [73] 陈安宏. 基于生理视觉特性的压缩编码方法在航天遥测图像传输中的应用[J]. *计算机测量与控制*. 2008, 16(10): 1475–1477.
Chen Anhong. *Computer Measurement and Control*, 2008, 16 (10): 1475–1477.
 - [74] Itti L. Automatic foveation for video compression using a neurobiological model of visual attention [J]. *IEEE Transactions on Image Processing*, 2004, 13(10): 1304–1318.
 - [75] Se S, Lowe D G, Little J. Vision-based global localization and mapping for mobile robots[J]. *IEEE Transactions on Robotics*, 2005, 21(3): 364–375.
 - [76] Ulrich I, Nourbakhsh I. Appearance-based place recognition for topological localization [C]// *IEEE International Conference on Robotics and Automation*. 2000: 1023–1029.
 - [77] Nothdurft H C. Attention shifts to salient targets [J]. *Vision Research*, 2002, 42(10): 1287–1306.
 - [78] 孙即祥, 等. 图像压缩与投影重建[M]. 北京: 科学出版社, 2005.
Sun Jixiang, et al. Image compression and projective reconstruction[M]. Beijing: Science Press, 2005.
 - [79] 郑南宁. 计算机视觉与模式识别[M]. 北京: 国防工业出版社, 1998.
Zheng Nanning. Computer vision and pattern recognition [M]. Beijing: National Defense Industry Press, 1998.
 - [80] Nilsson N J. 人工智能: 计算机科学 [M]. 郑扣根, 庄越挺, 潘云鹤 译. 北京: 机械工业出版社, 2000.
Nilsson N J. Artificial intelligence: A new synthesis[M]. Zheng Kougen, Zhuang Yueting, Pan Yunhe, trans. Beijing: China Machine Press, 2000.

(责任编辑 杨书卷)