

一种引入注意机制的视觉计算模型

龙甫荃 郑南宁

(西安交通大学人工智能与机器人研究所, 西安 710049)

摘要 提出了一种基于注意机制的视觉模型,其特点是:将注意过程分为3个层次,分别模拟生物的瞳孔聚焦、眼动和头动;提出一种新的可变结构的非均匀采样映射来模拟生物视网膜特性;强调了数据驱动的 bottom-up 过程与知识驱动的 top-down 过程的融合;提出新颖的类似树形的知识表示方法和基于深度优先搜索树的注意点转移控制机制。

关键词 注意, 对数极坐标变换, 非均匀采样, 中央区, 周边区, 知识粒度

1 引言

经典的通用视觉模型下的一些快速算法不能从根本上解决视觉计算的有效性问题。生理学研究表明,生物视觉系统基于复杂的生理结构,采用各种策略解决计算复杂性问题,其中很重要的一点就是选择注意和主动感知能力。生物视觉并非被动地接受外界光刺激,而是能够根据视觉空间的各种突变信息如运动、闪烁、锐变边缘以及有关外部世界和视觉任务的知识主动地观察感兴趣物体,而对无关或不重要的细节进行大量的信息压缩,并可以通过眼动、头动寻找新的感兴趣区域。基于这一点,人们开始将注意机制引入视觉计算模型^[1~3]。但以往的大多数模型都是在均匀采样的信息表示结构下建立的,而且它们或者只基于数据驱动的 bottom-up 过程,或者只基于知识驱动的 top-down 过程。本文则基于以往的工作提出一种相对完整的基于注意机制的计算模型。

2 模型

本模型中注意机制体现在3个层次上:(1)通过摄像机的转动在景物空间中搜索感兴趣的视野范围,以模拟生物身体及头部的移动;(2)在确定的视

野范围中,通过非均匀采样模拟生物视网膜的非均匀信息表示和对数压缩特性,将视野范围划分成中央区 and 周边区,通过非均匀采样中心坐标的改变模拟眼动;(3)在中央区进一步对感兴趣的区域做精细的模式匹配与识别,模拟瞳孔的聚焦作用。图1(见下页)给出了模型框架。

摄像机摄取外界景物形成一定的视野范围,视野范围经可变结构的非均匀采样映射被划分成高分辨率的均匀采样中央区和低分辨率的非均匀采样周边区。这种非均匀采样映射基于传统的对数极坐标变换^[4],但同时引入了变焦特性。图象中的任一象素 $z = x + yi = |z|e^{i\theta}$, 定义如下映射:

$$w = \begin{cases} x + yi & |z| < b \\ \log[k(|z| - b) + a]e^{i\theta} & |z| \geq b \end{cases} \quad (1)$$

其中, k 控制径向采样率的变化, b 是大于等于0的实数, a 取大于0的实数(一般取1)以有效地去除奇异点。对于 $|z| \geq b$ 的区域,式(1)将矩形栅格图象 $f(x, y)$ 映射成极坐标对数图象 $L(r, \theta)$, 其中, $x \in \{0, \dots, M-1\}$, $y \in \{0, \dots, N-1\}$, $r \in \{0, \dots, R-1\}$, $\theta \in \{0, \dots, S-1\}$, R 是最大同心圆数, S 是各同心圆上的最大划分数, 设原点为 (x_0, y_0) , 当 $|z| = b$ 时, $\log[k(|z| - b) + a]$ 取得最小值 $r_{min} = \log a$, 当 $|z| = r_0$ 时(见公式(3)), $\log[k(|z| - b) + a]$ 取得最大值 $r_{max} = \log[k(r_0 - b) + a]$, 故:

$$r = L_r(x, y) = \frac{\log[k(\sqrt{(x - x_0)^2 + (y - y_0)^2} - b) + a] - r_{\min}}{r_{\max} - r_{\min}} \cdot R \quad (2)$$

$$r_a = \max(\sqrt{x_0^2 + y_0^2}, \sqrt{(x_0 - (M - 1))^2 + y_0^2}, \sqrt{x_0^2 + (y_0 - (N - 1))^2}, \sqrt{(x_0 - (M - 1))^2 + (y_0 - (N - 1))^2}) \quad (3)$$

$$\theta = L_\theta(x, y) = \frac{\arctan(x - x_0, y - y_0)}{2\pi} \cdot S \quad (4)$$

设 $a(r, \theta)$ 是极坐标对数图上某一象素在矩形栅格图象上的区域大小, 则:

$$a(r, \theta) = \sum_{x, y} 1 |L_r(x, y) = r, L_\theta(x, y) = \theta \quad (5)$$

对数极坐标正变换图象和反变换图象可以分别用公式(6), (7)定义:

$$L(r, \theta) = \frac{1}{a(r, \theta)} \sum_{x, y} f(x, y) |L_r(x, y) = r, L_\theta(x, y) = \theta \quad (6)$$

$$L^{-1}(x, y) = L(L_r(x, y), L_\theta(x, y)) \quad (7)$$

当 $b=r_a$ 时, 式(1)变为矩形栅格的均匀采样图象表示, 当 $b=0$ 时, 式(1)变成类似 Schwarz^[5]的对数极坐标表示, 一般情况下 $0 < b < r_a$, 式(1)将视野范围划分成均匀采样的中央区和对数压缩的周边区, 故 b 的引入提供了一种灵活的变结构的图象表示方法, 可根据不同的视觉任务选择不同的 b 。

视野范围通过非均匀采样映射被划分成中央区和周边区, 对中央区进行高维特征提取, 包括边缘、轮廓、方向、曲率、颜色、运动、对称性等; 而视野周边

区因极坐标对数压缩而发生形变, 使得基于形状描述的一些精细特征失去意义, 故对周边区进行低维特征包括边缘、广义对称性和运动等信息的提取。上述特征对视觉注意能够提供有利线索。记所得的 k 维特征图为 $F_{x, y}^k$ (中央区的 k 值大于周边区的 k 值), 类似 Milanese 的方法我们用一组方向高斯差分滤波器模拟中央开周边闭(center-on, surround-off)感受野, 在多尺度上对特征图的各位置实现合作与竞争, 计算出感性趣测度 $I_{x, y}^k$, 再将 k 维测度值 $I_{x, y}^k$ 进行线性加权融合, 生成单一的 bottom-up 注意图 $B_{x, y}$:

$$B_{x, y} = \sum_{i=1}^k w_i I_{x, y}^i \quad (8)$$

其中, w_i 是测度权重, $0 \leq w_i \leq 1$, 且 $\sum w_i = 1$ 。在不知哪种特征更重要时, 可利用空间统计分布假设, 对存在于较大面积的特征赋予较小的权重, 也可假设权重 $w_1 = w_2 = \dots = w_k = 1/k$, 此外还可利用视觉任务的知识分配权重, 这体现了 bottom-up 与 top-down 结合的必要性。

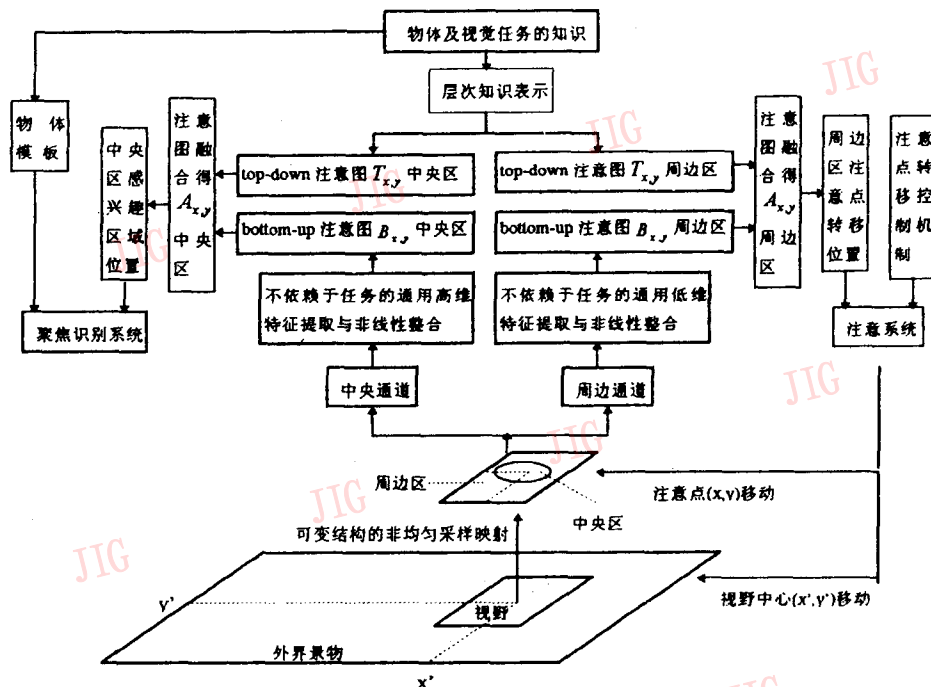


图 1 基于注意机制的视觉计算框架

top-down 过程是由知识驱动的。在本系统中, 物体的知识以两种形式表示: 基于物体属性的层次知识表示及具有不变性的物体模板表示, 两者通过类似树形的结构统一在一起。层次知识表示将物体的属性按照其对该物体识别所起作用的大小分为多个层次, 越靠近根节点的属性其类间分类能力越强, 我们称之为粗粒度知识, 越远离根节点的属性类内分类能力越强, 我们称之为细粒度知识, 叶节点则是具有不变性的物体模板表示, 它对物体在全分辨率下进行模板匹配。由于周边区分辨率较低, 对于周边区感兴趣物体的查找仅利用高层属性(如第一、二层属性), 而在中央区定位感兴趣的物体需用到低层属性, 一旦感兴趣物体被定位, 则用叶节点即物体模板对之进行匹配识别, 不变性模板匹配作为识别系统的一部分, 不参与 top-down 注意图的生成。设描述物体的属性被分为 n 层, 而前 m ($m < n$) 层对周边区有意义, 当前 i 层属性上均有至少一个属性得以匹配, 而第 $i+1$ 层上无一属性能匹配时, 则对中央区而言, 该区域质心的可信度值为 i/n ; 而对周边区而言, 可信度值为 i/m 。检查所有可能的区域, 可得到视野范围内各点处的物体可信度值, 由这些可信度值构成 top-down 注意图 $T_{x,y}$ 。

在视觉注意中, 单靠基于各种通用的统计或结构特征的 bottom-up 过程会增大搜索空间并为注意过程带来噪声; 而单靠有目的的 top-down 过程, 视觉系统将无法看到一些意想不到的(unexpected)、不受欢迎的(unwelcome)或不大可能的(improbable)东西。为此, 我们将 bottom-up 注意图 $B_{x,y}$ 与 top-down 注意图 $T_{x,y}$ 融合, 产生最终的单一注意图 $A_{x,y}$, 它为中央区提供了感兴趣物体的位置, 系统将对这些位置通过聚焦过程逐个进行模板匹配, 同时 $A_{x,y}$ 为周边区提供了注意点移动的位置信息。我们采用线性加权求和的方法:

$$A_{x,y} = \alpha_1 B_{x,y} + \alpha_2 T_{x,y}, \quad \alpha_1 + \alpha_2 = 1 \quad (9)$$

α_1, α_2 是权系数, 当 $\alpha_1 > 1/2$ 时, bottom-up 起主要作用, $\alpha_1 = 1$ 就是纯粹的 bottom-up 系统; 当 $\alpha_2 > 1/2$ 时, top-down 起主要作用, $\alpha_2 = 1$ 就是知识系统。

聚焦识别系统在中央区根据 $A_{x,y}$ 对感兴趣的位置做进一步的精细观察, 而对各感兴趣区域, 我们采用类似^[5]的不变性模板匹配的方法进行物体识别。首先对图象进行对数极坐标变换, 使图象具有比例和旋转不变性, 然后用一组 LOG 滤波器对变换后的图象滤波, 使图象对外界亮度的变化具有不变性, 再施加傅立叶变换, 而傅氏变换的幅值具有平移不

变性, 匹配针对傅氏变换的幅值进行, 由于对数极坐标变换对中心点的选取极为敏感, 一般将中心点取为物体或区域的质心。

本模型采用深度优先搜索树控制注意点的转移, 假设当前注意点在 z_0 ($z_0 = (x_0, y_0)$), 可变结构的非均匀采样映射将以 z_0 为中心 b_0 为半径的圆形区域映射成中央区, 该圆形区以外的部分为周边区, 在周边区内取 $A_{x,y}$ 值最大的 n 个点作为待注意点 z_1, \dots, z_n , 即:

$$\begin{aligned} [z_1, \dots, z_n] &= \max_n AT_{x,y}, \\ \sqrt{(x - x_0)^2 + (y - y_0)^2} &> b_0 \end{aligned} \quad (10)$$

其中 \max_n 表示前 n 个最大值。将 z_0 作为父节点, z_1, \dots, z_n 作为子节点, 并按从左到右的顺序排列, 如果 $z_i = z_j$, 则将与 z_0 距离小的点排在左边。对中央区的识别与匹配完成后, 选择 z_0 最左边的子节点 z_1 作为下一个注意点, 并根据以 z_1 为中心的区域的大小确定 b_1 , 进行新的非均匀采样变换并根据采样结构计算新的 $A_{x,y}$, 重复以上过程便可形成一深度优先搜索的注意树。一种特殊情况是每一父节点均只取一个子节点, 将形成一个线形的注意链, 使问题大大简化, 但注意机制的抗噪能力也随之下降。为了防止对注意过的区域重复注意引起注意点转移过程的振荡现象, 我们对注意过的区域进行标记并赋予时变衰减因子。

3 对比与结论

本文提出了一个基于注意机制的视觉计算模型, 与其它模型相比, 本文提出的模型具有以下几个特点。它将注意过程分为 3 个层次, 分别模拟了生物的瞳孔聚焦、眼动和头动; 充分强调了 bottom-up 与 top-down 的互动; 提出了一种可变结构的非均匀采样映射, 它将视野分成均匀采样的中央区和对数压缩的周边区, 中央均匀采样区域的半径灵活可调; 结合注意过程的 3 个层次, 提出了一种新颖的知识表示方法, 它将层次属性与物体模板相结合, 构成一种类似树形的结构, 周边区注意点的查找、中央区感兴趣物体的定位及中央区的物体识别与匹配分别采用不同粒度的知识; 采用一种基于深度优先搜索树的注意点移动控制机制。本文对深化视觉计算模型的研究, 尤其是通用视觉与目的视觉的研究具有重要的理论意义。

参考文献

- 1 Koch C, Ullman S. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. In L. M. Vaina (edt), Matters of Intelligence, Reidel Publishing, 1987;115~141.
- 2 Mozer M C. The Perception of Multiple Objects: A Connectivist Approach, MIT Press, Cambridge, MA,1991.
- 3 Burt P J. Attention Mechanisms for Vision in a Dynamic World, Proc. 9th Int. Conf. on Pattern Recognition, 1988;977~987.
- 4 Schwartz E L. Computational Anatomy and Functional Architecture of Striate Cortex: Spatial Mapping Approach to Perceptual Coding, Vision Reaserch. 1980,20: 645~669.
- 5 Poltzleitner W, Wechsler H. Selective and Focused Invariant Recognition Using Distributed Associative Memories. IEEE Trans. on PAMI, 1990,12(8);809~814.



龙甫荟 1970年生,1995年获西北工业大学计算机应用专业硕士学位,现于西安交通大学攻读博士学位,主要从事计算机视觉、图象处理和模式识别方面的研究工作,曾发表多篇论文。



郑南宁 教授,博士生导师。1952年生,1985年获日本庆应大学工学博士学位,1996年被评为中国青年科学家,曾发表多篇文章。现主要从事计算机视觉与模式识别、神经网络、智能系统并行处理的研究。

A Visual Computing Model Based on Attention Mechanism

Long Fuhui ,Zheng Nanning

(Inst. of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xian 710049)

Abstract A visual computing model based on attention mechanism with the following characters is proposed in this paper. First, it imitates the characteristic of zoom-lens, eye movement and head movement of human vision system. Second, the retina character is modeled by a new nonuniform sampling method with changeable structure. Third, the bottom-up process driven by data and top-down process driven by knowledge are combined. A novel knowledge representation method and a fixation movement controlling mechanism based on depth first search tree are proposed as well.

Keywords Attention, Log-polar transform, Nonuniform sampling, Fovea, Periphery, Knowledge grain.

IBM 推广 Java 乐此不疲

IBM 公司日前又推出一系列的 Java 新品,其中包括 ServletExpress 软件插入程序、建立 Java 企业应用软件“旧金山”框架所需要的订购和仓库应用软件工具、Txseries 分布式交易处理环境、Component Broker 分布式对象服务器以及 Visual Age Java 开发工具新版本等。另外还开通了一个具有搜索引擎的 Jcentral 新网点,Java 开发人员可以直接从网页上“拉”所需的 Java 信息,为 Java 软件开发人员更快、更有效地创建实用、强大的商业应用软件提供得心应手的帮助。