

计算机视觉

计算机科学

人工智能

计算机科学1999Vol. 26No. 3

## 计算机视觉的理论框架研究

Study on the Theoretical Framework of Computer Vision

刘会杰 王学孝

(哈尔滨工业大学 哈尔滨 150001)

**Abstract** The theoretical framework of computer vision, D. Marr's vision computing theory, is concluded in this paper. With regard to the presence of computer vision, flaws of D. Marr's vision theory are also specified, additionally, this paper emphasizes several directions of computer vision. In the final section, partial relevant practical works supported by the author are given.

**Keywords** Computer vision, Image understanding, Machine vision, Vision computing theory

## 1. 引言

计算机视觉是计算机科学和人工智能的一个重要分支,它是工业生产自动化、机器人智能化、自主车导航、目标跟踪以及工业检测、医疗和军事应用的核心,也是实现机器智能及第五代计算机的关键因素之一<sup>[1]</sup>。

计算机视觉的研究目的和内容有两个方面,一是用计算机部分实现人类视觉的功能;二是理解人类视觉机理<sup>[2,3]</sup>。此二方面使计算机视觉的研究既带有基础性,又有很强的应用特征和工程性质。就人类视觉而言,其功能集中于识别与理解周围环境内对象物坐标、物体间相对位置及颜色等。仿人类视觉机制,利用计算机实现从二维或系列图象中构造、理解三维世界模型的功能就称之为计算机视觉。由此亦知,计算机视觉是一综合性学科,它涉及图象处理、模式识别、图象理解、计算机科学等诸多领域。

计算机视觉的研究内容十分广泛,根据侧重点的不同,可作如下划分<sup>[2,4,5]</sup>:

1) 计算视觉,即计算机视觉算法,它重点研究生物视觉信息处理的计算模型,定位于计算机科学同生理解剖学的交叉和边缘点。

2) 机器视觉,强调的是实际应用,视觉过程的实质就是从图象中提取信息并作出决策。在特定问题中,根据结构和限制条件不同,可由部分可用信息作出依赖于问题结构的结论。因为其鲜明的应用特征与新一代智能机器人紧密结合,故又称机器人视觉。

3) 图象理解,侧重于实现方法,即对景物的几何结构进行恢复以及对物体进行识别的方法。

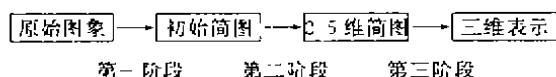
值得指出的是,作上述划分是为了更好地了解、认识计算机视觉这一庞杂的研究领域。事实上,以上三个侧面相互补充、相互渗透,共同构成一个完善的计算机视觉世界。

## 2. 视觉计算理论框架—D. Marr 理论

本世纪七十年代,美国 MIT 的 D. Marr 教授提出了人类的视觉计算理论,极大地推动了计算机视觉研究工作的发展,同时在概念上统一了模式识别、图象处理、图象分析、图象理解等术语,并最终行成这一领域中的主导思想。D. Marr 视觉理论是该领域中迄今为止最为完善的理论,它将视觉过程看作一信息处理流程,并在三个不同的层次上进行研究,即计算理论层次、表示与算法层次以及硬件实现层次。在计算理论层次上,视觉过程规定为从二维图象中定量恢复图象所反映景物中三维物体的形状和位置,也就是要确定:为什么要进行计算?为什么采用某种计算?怎样进行计算?更进一步,需确定输入、输出以及由输入和输出之间的内在约束从输入确定输出。表示和算法层次上的问题具体化为:怎样实现计算理论?怎样表示输入和输出?采用何种算法使输入转换为输出?硬件实现层次则需解决物理上如何实现特定的表示和算法<sup>[2~4]</sup>。该种不同层次的划分,一方面佐证了视觉计算理论的合理性,另一方面也

有助于深入理解人类视觉同计算机是视觉的关系。

人类生活在三维世界中,而由视觉感知的世界则体现为二维图象,基于此,视觉系统的功能可归结为从所感知的二维图象中提取出有关二维世界的信息(指图象中所反映场景的二维结构和空间位姿)。为完成上述生物学意义上的视觉功能,D Marr 的视觉计算模型将视觉信息处理过程分为三个阶段,即:



并以零交叉(zero-crossing)为图象特征,在计算机上实现了这一理论。

与三个阶段相对应,在这一框架之下,又可划分为低层(早期)视觉、中层(中期)视觉以及高层(后期)视觉<sup>[2,4]</sup>。低层视觉的任务在于图象传感和预处理,即清楚地表示原始二维图象中的重要信息,如零交叉、角点、边界、线段等,它们对应图象中灰度急剧变化的位置及其几何分布和组织结构。中层视觉则由输入图象和低层视觉输出的初始简图得到2.5维简图,如在以观察者为中心的坐标系中可见表面法向、大致深度及不连续轮廓等。按 Marr 理论,中层视觉处理又可归为 shape-from-X,此处 X 指代 motion, stereo, shading, contour 以及 texture 等。高层视觉涉及由输入图象、初始简图以及2.5维简图获得物体的三维表示,即在以物体为中心的坐标系中获得物体的几何结构和空间位姿关系。

由前述叙述知道,Marr 理论将视觉计算过程看成几何成像过程的逆问题,为求解视觉过程,首先分析几何成像过程,其中的问题主要表现为<sup>[6,4,7]</sup>:

1) 三维场景投影为二维图象,丢失了第三维信息,从而使同一物体在不同视角下所成图象有很大不同,并且还存在因遮挡关系而造成的信息损失问题。

2) 多元迭加效应,使得场景中照明、物体结构以及物理性质、相机特性、及光源、相机和物体的空间关系等诸多因素影响综合为单一的图象中象元灰度值。

3) 成像过程中畸变和噪声的影响。

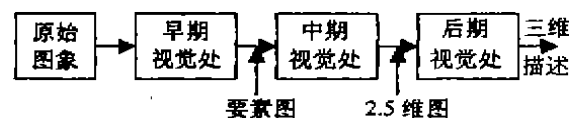
因此,作为成像过程逆过程的视觉处理,其目的就是带畸变和噪声的、二维的、单一的象元灰度值中尽可能提取出不带畸变和噪声的、三维场景中的

物体形状及空间位置信息,特别地,当要寻求通用的、定量的精确结果时,这一不适定(ill-posed)问题的求解将更为困难。

### 3. 计算机视觉的研究现状及 D. Marr 理论的不足之处

D. Marr 的视觉计算理论自七十年代末提出以来逐步为广大计算机视觉研究者所接受并主导了这一领域的发展趋势。然而,从目前状况而言,计算机视觉还存在应用研究欠缺及实验标准不完善等问题<sup>[3,4]</sup>,具体来说有:1)对应用研究的重视,就目前技术水平而言,短期内不可能实现全自动的通用二维视觉系统,应结合实际的应用条件,开发各种专用视觉系统,且允许以人工交互方式改善系统性能,其中典型的有各种文字识别和印章识别等应用系统,2)对实验的重视,因视觉过程求解的复杂性,很难进行深入的理论分析,因而用实验手段来检验和评价各种理论和算法就极为可取。为此,系统的实验方法的研究、实验系统的设计和实验以及各种测试标准的建立也成为近期计算机视觉的研究课题。

随着各相关学科理论和实验研究的深入开展,D. Marr 视觉理论框架也表现出了局限性。针对这一情况,许多研究人员提出了不同的观点,为便于比较和说明,对 D. Marr 理论框架重述如下:



在此框架中存在的不足之处为:

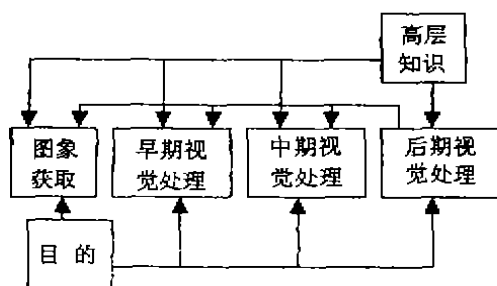
1) 输入为被动的,提供何种图象就处理何种图象,无主动性和目的性可言,这根本不同于生物视觉。

2) 整个处理过程基本上自底向上,没有反馈,这种纯粹数据驱动的视觉机理往往会导致相应系统功能的失败。

3) 处理目的局限于场景中物体的三维形状及空间位姿关系。

为解决上述不足,众多计算机视觉研究人员所提出的不同观点可综合为如下框架<sup>[2]</sup>。

这一框架充分体现了生物视觉的特点:1)从输入来说,它嵌入了主动性、选择性机制(主动性意指



视觉系统可根据已有分析结果和视觉的当前要求,确定相机的运动以便从合适的视角获得相应输入图象;而选择性意指对注意点或兴趣点的控制而非相机的角变化)。2)从目的来说,本框架在原有基础上集成了“目的性视觉”和“定性视觉”,即视觉系统的功能取决于应用目的,这一方法论指导无论对视觉系统的开发还是应用探索都极为重要。3)高层知识的利用,目前计算机视觉同生物视觉的一个根本不同点在于生物在生存活动中所积累的知识,生物的视觉过程常常隐含了大量相关知识,因而可从视网膜图象中得出可靠结果。类似地,对计算机视觉不适定视觉过程(表现为病态数学问题)的求解,也需要用各种形式的知识加以限定,以期得到合理的结果。4)控制策略,在D. Marr 视觉理论中,处理流程总是自下而上,各阶段之间不存在反馈,因而基本不存在控制问题;而在本框架中,仿照生物机制在前后处理之间引入了交互作用。

此外,在具体的处理阶段,也存在一些不同的观点,这种状况表明,计算机视觉尽管已有三十多年的历史,然而同生物视觉相比仍处于相当初级的阶段。D. Marr 的理论对计算机视觉的发展产生了深远的影响,它为计算机视觉规定了明确的研究内容,并使之形成一个较完整的体系,而且它仍是迄今为止这一领域中最为完善的理论,但从计算机视觉的研究目的而言,D. Marr 理论框架和上述不同观点的共存又促使计算机视觉不断向前发展并趋于成熟。

#### 4. 计算机视觉的发展趋势

如前一部分所述,因D. Marr 视觉理论同生物视觉机制之间的差异,因而引发了各种不同的观点,这些观点的综合以及众多研究人员在D. Marr 视觉理论框架下卓有成效的工作构成了计算机视觉领域未来的发展趋势<sup>[8~16]</sup>。

·图象分割,在计算机视觉流程中,图象分割的

重要性不言而喻。然而,已有的关于计算机视觉的研究工作大都基于良好分割的前提之下,而事实上,这一假设本身就需要相当复杂和细致的工作,这源于图象分割问题所固有的困难。传统的分割方法只简单地利用图象中象素特性的相似或不相似信息,因而易受噪声影响而发生错误。另一方面,利用图象中目标的先验信息可有助于图象分割,但可资利用的先验信息又受到一定的限制,因此,结合待解决问题,在特定的视觉流程中引入先验信息进行图象分割将成为计算机视觉的一个研究方向。

·从距离图象中提取三维信息,距离传感器技术的发展使场景距离信息的直接获取成为可能,从而避开了立体视觉中对应性问题的困难。这里,问题转化为如何从距离图象中提取深度信息,也就是对距离图象的处理和理解。关于这方面的工作,一方面可以利用目前已有的图象处理和理解技术,另一方面,较为流行的趋势乃是采用微分几何方法进行相应分割和理解。与此同时,计算机视觉中几何模型以及利用几何模型进行三维推理也得到了很多研究人员的重视。

·多传感(算子)信息的集成和融合,这一思想的发展一度较为缓慢。目前,信息融合理论以及特定问题中多源信息提取等已都成为计算机视觉领域的研究课题,并且这一思想同小波分析方法的结合已收到了一定的效果。

·动态视觉,在过去十多年里,动态视觉的研究目的一直是利用最小数量的图象帧确定光流由运动恢复结构,其中包含对图象灰度值进行三次求导的问题,当应用于真实图象时,该种计算受到很大限制,因而,目前的动态视觉分析主要集中于长序列图象分析,通过冗余信息的利用降低计算复杂性。

此外,定性视觉、目的性视觉以及知识的利用和推理等也是计算机视觉中很有潜力的发展方向,这些在前文已有论述。

#### 5. 部分工作

在计算机视觉研究中,由单幅灰度图象确定目标的三维位姿得到了广泛关注。关于这一问题,已有工作<sup>[17~19]</sup>均基于图象中目标的先验模型(知识),通过模型与图象间特征对应关系的建立进行求解。文献[17,18]中通过线特征对应,并经适当坐标转换,将问题归结为高( $\geq 4$ )次方程求解,其中亦涉及利用

假设-验证策略来排除伪解。本文工作利用点特征对应求解上述问题,其中深度信息嵌入比例因子中。

设  $I_m(u, v_i)$  为图象特征点,  $M_m(x_i, y_i, z_i)$  为模型特征点,  $i=1, 2, \dots, n$ , 且  $I_m$  与  $M_m$  相对应, 问题具体化为: 确定绕坐标轴  $x, y, z$  的旋转角  $\alpha, \beta, \gamma$ , 沿坐标轴  $x, y$  的平移  $d_x, d_y$  以及比例因子  $s$ , 由图象特征点与模型特征点对应关系, 可建立如下方程:

$$T_x \cdot T_y \cdot T_z \cdot R_x \cdot R_y \cdot R_z \cdot \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} x'_i \\ y'_i \\ z'_i \end{bmatrix}$$

其中:  $R_x, R_y, R_z$  分别为绕坐标轴  $x, y, z$  的旋转矩阵,  $T_x, T_y$  分别为沿坐标轴  $x, y$  的平移矩阵,  $T_z$  为比例矩阵。上式可进一步写为:

$$T \cdot \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} x'_i \\ y'_i \\ z'_i \end{bmatrix}$$

该式中,  $T = T_x \cdot T_y \cdot T_z \cdot R_x \cdot R_y \cdot R_z$ , 参数  $\alpha, \beta, \gamma, d_x, d_y, s$  这六个未知因子包含于矩阵  $T$  中。

本文工作中, 考虑到透视投影在目标距相机较远(或目标尺寸与该距离相比可以忽略)时可用平行投影近似, 故此处取平行投影, 即

$$P \cdot \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} -u_i \\ -v_i \end{bmatrix}$$

其中,  $P$  为投影变换矩阵, 在平行投影下,  $u_i = x'_i, v_i = y'_i$ 。

至此, 问题转化为: 确定矩阵  $T$  中的六个参数。因每一个特征点给出两个方程, 故六个未知数共需三个对应的特征点, 但这样建立的关于六个未知数的方程为非线性方程, 考虑到线性代数方程求解的有效性, 此处引入一额外对应特征点, 通过四个特征点对应, 建立八个方程, 从而完成求解, 为了减小误差影响, 又引入了附加的对应特征点, 利用最小二乘法对求解结果进行优化。实验结果表明, 该方法具有很好的三维位姿测定能力。(参考文献共19篇略)

(上接第81页)

### 参考文献

- King S, et al. CASE 2000: the future of CASE technology. *Software Engineering Journal*, 1994, 9(4): 138~140
- 扬美清. 青岛工程现状与发展. 96全国软件工程会议论文集. 1~8
- 冯玉琳, 等. 面向对象的组合软件工程研究. *计算机学报*, 1996, 19(3): 237~241
- Mih M, et al. Reusing software: issues and research directions. *IEEE Transactions on Software Engineering*, 1995, 21(6): 528~562
- Gacek C. Exploiting domain architectures in software reuse. *ACM Software Engineering*, 1995, 20(8): 229~232
- Garlan D. First international workshop on architectures for software systems workshop summary. *ACM Software Engineering*, 1995, 20(3): 84~89
- Koster M. Robots in the web: threat or treat? *ConneXions*, 1995, 9(4): 38~45
- Adler R M. Emerging standards for component software. *IEEE Computer*, 1995, 18(3): 68~77
- 蔡智明, 等. CASEE 的集成与开放. '96全国软件工程会议论文集. 79~84
- 蔡智明, 等. CASE 构件的表示与分层. *合肥工业大学学报*, 1996, 19(4): 60~63
- 蔡智明, 等. 多媒体虚拟现实技巧. 北京: 机械工业出版社, 1996
- 刘宗田. 通用历史数据库管理系统 GHDBC 的设计与实现. *计算机研究与发展*, 1997, 34(4): 281~286
- 鲍敬峰, 等. 软件复用库的模糊表示和查询方法. *小型微型计算机系统*, 1997, 18(4): 27~31
- 鲍敬峰, 等. 软件规格说明综述. *软件*, 1996, (6): 14~19
- 袁兆山, 等. 软件工程环境 EPOS 应用基础. 中国科技大学出版社, 1994