

三维物体识别研究进展

李庆¹⁾ 周曼丽²⁾ 柳健¹⁾

¹⁾(华中理工大学图象所, 图象信息处理与智能控制国家教委开放研究实验室, 武汉 430074)

²⁾(华中理工大学电子与信息工程系, 武汉 430074)

摘要 出于工业和医疗等领域大量现实应用的需要, 如今三维物体识别已成为一个很活跃的研究领域. 一般来说, 三维物体识别系统可以通过两个阶段的处理来完成三维物体的识别和定位, 首先用传感器获取的场景输入数据来得到场景的表达; 然后将它与数据库中存储的物体表达相匹配. 为了推动该领域研究进一步发展, 因而对近10a年中该识别过程中必须解决的传感器类型、三维物体表达方法和匹配策略等3个方面问题的研究成果进行了综述, 对主要方法进行了分类和总结; 并提出了一些三维视觉系统中还需要深入研究的问题, 包括对所研究物体形状的限制、复杂背景的影响和表达以及识别中的“整体和局部”的矛盾等.

关键词 计算机视觉 三维物体识别 深度图象 三维物体表达方法 匹配策略

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2000)12-0985-09

A Review on 3D Objects Recognition

LI Qing¹⁾, ZHOU Man-li²⁾, LIU Jian¹⁾

¹⁾(Institute for Pattern Recognition and Artificial Intelligence, State Education Commission Laboratory for Image Processing and Intelligence Control, Huazhong Univ. of Sci. & Tech., Wuhan 430074)

²⁾(Department of Electronics and Information, Huazhong Univ. of Sci. & Tech. Wuhan 430074)

Abstract For the requirement of many practical applications in the area of industry and medicine, the research in three-dimensional objects recognition is very active. In general, the dominant paradigm in 3D objects recognition system proposes to achieve recognition and localization of 3D objects from images by a two-stage process: first derive an internal representation of a scene from the sensed input data and then match it against stored representations of objects in the database. This paper presents a comprehensive survey of the achievement of 3D objects recognition system in the recent decade. The three problems are discussed which are the type of sensors used, 3D objects representation and match strategy. Furthermore, the paper classifies and summarizes the primary schemes. At last, we describe some problems requiring a thorough research on three-dimensional vision system, such as the limitation by the class of shapes that can be described in most of the representation schemes, the clustered background influence on the recognition, and the contradiction between global and local in the 3D objects representation and recognition.

Keywords Computer vision, 3D object recognition, Range image, 3D object representation, Match strategy

0 引言

计算机视觉系统的目标是解释已有的“视觉”数据, 并使用这些解释去完成任务. 出于机器人工件抓

取任务、自动导航、自动检测、装配任务与医学图象的分析等等大量现实应用的需要, 如今三维物体识别已成为一个很活跃的研究领域.

计算机视觉系统的一个典型范例就是要通过两个阶段的处理, 以完成图象中3D物体的识别和定

位^[1],即首先通过传感器获取的场景输入数据来得到场景的表达;然后将它与数据库中存储的物体表达相匹配.具体来说,三维物体识别系统必须解决如下几个问题^[2]:(1)获取原始数据的传感器类型;(2)三维物体模型构造方法;(3)描述原始数据和物体模型的工具;(4)从输入数据获得的物体描述和物体模型描述之间的匹配方法.其中,传感器决定了输入数据的分辨率、准确性和类型(2D或3D信息);而模型则提供了视觉系统的先验知识;表达则用于描述采集的数据和物体模型,是计算机视觉中的一个关键论题;另外,知识表达还决定了匹配策略、匹配策略的鲁棒性和系统效率,并且在匹配过程中描述还用来计算物体的不同特征.由于匹配策略是在识别系统运行时执行的,因此它必须解决数据和模型描述之间的二义性.一旦正确的匹配已确定,则根据物体模型,即将已定位物体的方向和变换计算出来,最终完成物体识别的任务.一个三维物体识别系统的组成如图1所示^[3].

的比较示意图.

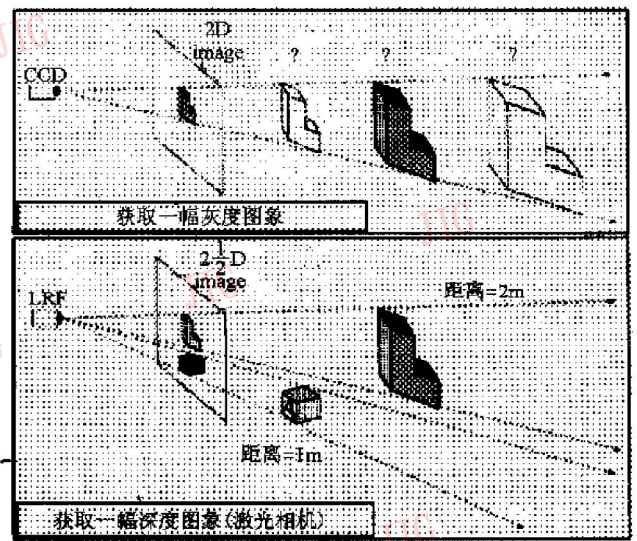


图2 深度图象与灰度图象获取方法的比较示意图

目前,主要有如下两种商业用距离相机:一种是激光雷达相机(Laser radar camera,或称Laser range finder);另一种是结构光扫描仪,而更多的关于深度图象和激光雷达相机的信息可以在文献[4]中找到.

2 三维物体表达方法

一般来说,识别系统的性能可以由物体的其它信息,例如颜色和纹理等来改善,但在本文中仅讨论物体的几何表达方法,而且物体表达方法必须具备下述条件^[2]:(1)明确性(没有两个物体有相同的表达);(2)唯一性(使用同一表达方法的每个物体必须只有一种描述);(3)不敏感性(物体在部分被遮挡的情况下会缺少部分数据);(4)方便匹配和存储.

2.1 三维物体表达方法分类

根据描述物体坐标系的不同,可以将三维物体表达方法分成以物体为中心的(Object centered)和以观察者为中心的(Viewer centered)^[5]两个主要类别.其中,以物体为中心的表达方法侧重在物体本身坐标系中描述物体,并使用与视点无关的本质特征(如拐角、孔洞、边等等)来描述物体;以观察者为中心的表达方法则依赖于从某一个或多个视角来确定物体的外观,并使用与视点相关的特征(例如,遮挡的边缘、轮廓和形状的T型连接等等).

根据描述3D物体所使用的几何特征粒度大小的不同,可以按照图3将表达方法从低到高分为如

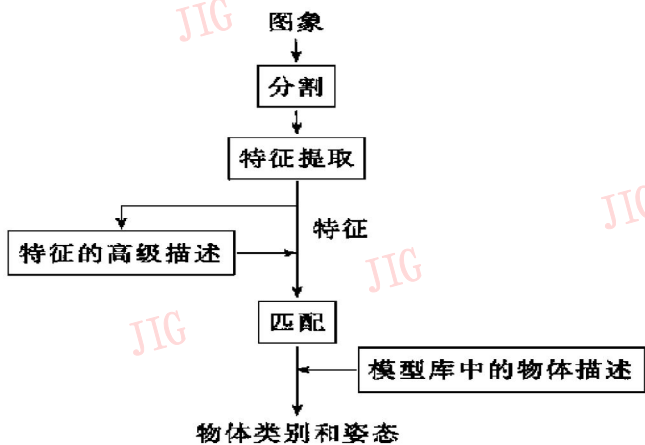


图1 三维物体识别系统的组成框图

1 传感器类型

三维物体识别中有如下两类最经常使用的传感器:一是灰度传感器(Intensity sensor),其获取图象的每个象素点对应于一个亮度测量;另一类是深度传感器(Range sensor),其象素点对应于从传感器到可视的物体表面的距离.由于灰度图象象素值的大小受环境光照强度、物体表面反射特性和纹理特征等多种因素的影响,而深度图象是直接反映物体的深度信息,则不受这些因素影响,因此在根据物体几何信息进行的三维物体识别中,深度图象得到了广泛的应用.图2为深度图象与灰度图象获取方法

下 6 类: ①基于三维点的表达方法, 主要使用物体的深度数据、表面点的法向方向和曲率大小等信息; ②基于凸点的表达方法, 主要利用物体的顶点、拐角点和曲率极大、极小值点等信息; ③基于轮廓的表达方法, 主要使用物体的边缘信息; ④基于表面的表达方法, 主要使用物体的表面信息, 例如平面、球面、二次曲面以及面与面之间的连接关系等; ⑤基于体的表达方法, 主要使用体素、椭球体、超级椭球体等体描述符; ⑥基于零件的表达方法, 主要分析物体的基本组成部分, 例如体元(geon)方法。

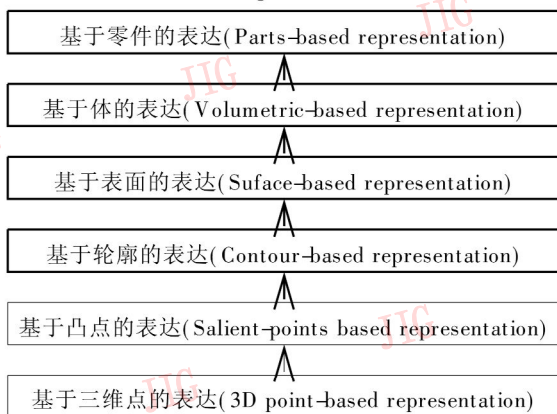


图 3 根据描述 3D 物体特征的粒度大小不同得到的 3D 物体表达方法分类

2.2 三维物体表达方法

下面按照选取特征的不同, 来介绍几类有代表性的三维物体表达方法, 并具体论述各方法的主要特点。

(1) 基于基本表面特征的方法

早期研究工作使用了物体表面的一些基本特征, 即将物体表达为面、边和顶点的队列, 且它们是在图象投影中与视角无关的特征, 如 Grimson 和 Lozano-peres 曾使用边、直线段和法向矢量来表达多面体^[6], 由于使用多面体来表达逼近曲面物体的表面, 需要大量空间, 所以有人提出使用平面和二次曲面方程来描述曲面表面^[7]; 还有一些研究是根据表面点的高斯曲率和平均曲率^[8], 将表面片分成峰(peak)、凹(pit)、鞍(saddle)等等一些主要的形状, 关于这方面的内容在文献[9]中有详细的论述; 同时 Stein 和 Medioni 提出了一种结构表达方法^[10], 这种方法根据表面法矢量的分布来确定边和局部表面片, 因此它能够处理任意形状的物体表面; 另外, Chitra Dorai 提出了 3D 自由形态物体的 COSMOS 表达方法^[11], 这种方法是先定义表面小区域上的归一化形状索引特征, 然后在全部表面上进行统计, 以

得到形状谱, 再根据形状索引特征得到最大一致形状表面片, 两者联合起来则构成对任意形状物体的表达。一般来说, 基于局部边界和表面等基本表面特征的方法, 对获取数据中的噪声敏感, 并且准确性要依赖于是否能够从输入图象来可靠地提取描述物体特征。

(2) 基于表面不连续性的表达(Discontinuity-based representation)法

还有一些方法是利用物体表面的不连续性来表达物体的, 如 Godin and Levine 使用物体的屋脊和阶越边缘来构造边缘链接图(edge-junction graph), 其中, 节点表示连接关系; 边表示连接到一起的边缘^[12]; 而 Chen 和 Stockman 则综合使用了表面和断点的信息, 来标记边缘^[13], 并定义了飞边(wing)的概念, 即每条飞边是一个面、边缘、面的三元组; 另外, Chen 和 Stockman 还使用二维轮廓特征来描述 3D 任意形态物体^[14]。由于基于表面不连续性的表达方法是利用物体的边缘信息, 因此降低了存储需求, 提高了高层处理的效率, 但在大多数情况下, 这种表达是不完整的, 因为它丢失了表面信息(例如凹凸特性)。

(3) 拟合表面参数表达法

新近的一种全局物体表达方法是将物体表面用参数方程拟合, 如 Umasuthan 分析了高阶代数方程某些参数具备的不变性^[15], 并用它们来表达 3D 物体的表面片; 李松涛提出了一种双二次变量正交多项式的曲面拟合方法^[16], 该方法可根据拟合得到的代数方程参数来计算物体表面特征; 另外使用 B 样条表面也可拟合 3D 表面^[17], 但表面拟合与识别的范围约束和收敛性都是需要彻底研究的领域。

另外, 超二次曲面特征最早由 Barr 引入计算机视觉领域^[18], 而从深度图象获取超二次曲面表达也是通过对输入数据进行隐含方程(implicit equation)拟合得到的^[19]。周林还提出了一种扩展超二次曲面的表达方法^[20], 它比超二次曲面具有更强的描述能力。

在这类方法中, 物体遮挡再次成为难以解决的问题, 因为由物体部分外观生成的代数多项式与由物体全部外观生成的多项式可能不同, 而且它首先需要进行区域分割, 然后通过表面拟合才能得到代数方程的参数, 这样得到的参数准确性将依赖于分割的结果。

(4) 基于定位的表达(Orientation-based

representation) 法

对物体表面上的任一点,都可以在高斯球(单位球)上找到与之表面方向相同的一点,这样就可以得到物体表面上的点到高斯球的映射.如果该物体表面是一凸性物体的表面,则上述映射是一一映射的,但高斯球表示的缺点是存在二义性,它不能保存物体的变换和尺寸信息.一些改进方法是将物体的表面法矢量根据支撑函数映射到单位球上,例如广义高斯图象(Generalized Gaussian image, GGI)^[21]由于存储了单位球上直接相邻点之间的连接信息,因此确保了对所有物体表达的唯一性;又如 Ghos 和 Kumar 对支撑函数进行了讨论^[21],并证明了一些看似差异很大的表达方法,其本质上都是相同类型的支撑函数表达,同时也讨论了表达非凸物体的支撑函数.

基于定位的表达方法主要描述物体的表面信息,虽然它是一种全局的表达方法,但从识别的角度看,它们太烦琐,且不能处理物体有被部分遮挡的情况,而且,从由多个物体的场景图象生成的高斯图中分割出单独属于不同物体的区域,也是非常困难的问题.

(5) 基于格网的表达法

格网表达法就是图形学中使用的三维任意形态物体描述方法,它使用多边形来表示物体的形状,常用的是四边形和三角形格网,由于其数据量大,因而存储、传输和运算都比较困难,很难直接用在三维物体识别中,因此经常要转换为其它表达.在这方面 Johnson 作了较有成效的研究^[23,24],他将格网中某个基准顶点与其它顶点的几何位置关系转换成一个二维图象,称为 spin-image,它可用来描述物体的三维特征,进而实现识别和定位等工作.虽然引入了简化算法,但使用 spin-image 的操作计算还是太复杂.

(6) 体素(voxel)法

体素表达^[25]主要用于描述物体的体特征,它类似于像素在平面中代表一小块面积,而体素在空间中则代表一小块体积,它将物体描述成一个互不重叠的立方体的集合,这些体素(立方体)分布在三维方形网格中紧密排列,并填充物体所占的空间,然而体素表达并不适用于 3D 物体识别,它更多的是用在物体的表面重建和建模上.

(7) 八叉树(octree)法

八叉树(octrees)法以分层的方式来描述物体,它是一种每个节点有 8 个分支(octant)的树型结

构,且它的根节点是一个能够完全包住物体的立方体,分层描述时它将物体所占的空间逐层迭代地划分为 8 个分量,迭代的停止条件是每个最终的子立方体在某种特征上是同质的,如 Chien 等提出了使用 3 对相互垂直的深度图象来生成物体 octree 模型的方案^[26];而 Li 则提出了一种利用任意视角深度图象来构造 octree 的方法^[27],它能够描述物体的凹面.虽然 octree 表达能描述物体全局性的体特征,但它不是一种精确的表示法,其描述物体的近似程度取决于分割的精度.

(8) 构造实体几何法(Constructive solid geometry, CSG)法

构造实体几何法是用系统定义的简单几何形体,经过并、交、差等布尔组合,来构造出所需要的复杂实体^[28],而每个物体表示成一棵树,树的叶节点是块、楔、柱等简单物体或它们的变形,其它节点是构造物体的布尔操作.一般构造实体几何法经常用在构造物体的模型上,特别是在 CAD/CAM 系统中,而用在物体识别的时候则较少.这种表达方法的缺点是不能精确地表达物体,即表达不唯一,且建造时所用的简单刚体知识对识别系统来说,效果不好,因为识别系统中只有物体表面是可视的,而物体的组成部件可能被遮挡.

(9) 体基元(geon)法

体基元表达可以作为物体的高层次表达方法. Biederman 提出人脑一定程度上可使用物体的组成部分来识别物体^[29],而且可以用二维线条来描述物体组成部分的形状,因此他根据 4 种二维属性,定义了 24 类体基元,以用来描述物体的组成部分. Nguyen 和 Levine 应用了 Biederman 提出的体基元种类^[30],并以物体的边缘连接图作为系统输入,然后将边缘连接图分解为子图,每个子图代表物体的一个组成部分,并对应于一个体基元,而体基元之间的连接关系由物体的凹边缘和 T 型连接确定.这种基本体基元的表达还可以使用超二次曲面和广义柱的方法.体基元表达方法的缺点是基元特性集合不一致,缺乏对这些集合存在的必要性、充分性和正确性的论证.另外,从物体的单幅外观表现来计算所有这些基元特性也非常困难.

(10) 扫描(sweeping)法

扫描法表达是将形状描述成一个 2D 函数在 3D 空间中的扫描,其扫描的方式有平移和旋转^[28]两种.这类方法中用得最广的是广义柱(Generalized

cylinder), 而最普通的广义柱是一个圆沿着与圆平面垂直的轴平移所得到的柱体, 而且如果圆的半径是轴位置的线性函数, 则平移的结果是广义锥. 为描述更多类型的形状, 人们定义了多种广义柱^[31], 例如, 直线均匀广义柱 (Straight homogeneous generalized cylinders) 和 曲面刚体旋转 (Curved solid revolution). 广义柱的缺点是表达可能不唯一, 因为为了表达数据, 其起始的横截面可能以不同方式定义, 因而致使相同的数据点有不同的表达方法.

(11) 以观察者为中心的表达法

以物体为中心的表达方法, 其目的都是在以一种本质的方式来表达物体的全部形状; 而以观察者为中心的表达方法则是将物体表达成一个 2D 外观的集合, 而不是单一的模型. 其中外观图方法 (The aspect graph)^[32] 就是将三维物体的所有二维外观,

分别组成有一定意义的外观聚类集合. 它是将物体的视点空间, 分割成一些具有相同外观表现 (appearance) 的区域, 称为 aspect, 并由发生在两类相邻观察 (view) 之间的视觉事件 (visual event) 分开, 当物体的轮廓拓扑发生变化时, 就产生一个视觉事件. 基于外观的表达是采取了一个连接图的方式, 其图中每个节点对应一个外观, 每条边即对应视觉事件.

大多数以观察者为中心的表达, 其主要缺点是描述物体不简洁, 因而如果物体由几个参数方程描述, 则以观察者为中心的表达就明显不合适, 但是, 在描述那些形状不能由简单分析方式或一套方程来获得的复杂物体时, 这种方法就扮演了一个重要的角色.

表 1 是对以上这些方法的总结.

表 1 三维物体表达方法小结

表达方法	形状描述符的类型和粒度	使用物体范围(试验中)	传感器类型
点、平面和边 ^[7]	局部; 轮廓和表面	多面体	深度图象
圆弧、直线、柱面和平面 ^[14]	局部; 轮廓和表面	平面和柱面物体	深度图象
溅射和超级多边形分割 ^[10]	局部; 表面	任意曲面	深度图象
COSMOS ^[11]	局部+ 整体; 表面	任意形状物体	深度图象
飞边(wing)表达 ^[13]	整体; 轮廓和表面	多面体	灰度图象
2D 外表的轮廓 ^[14]	整体; 轮廓	曲面物体	灰度图象
代数多项式拟合 ^[15]	整体; 表面	曲面物体	深度图象
超二次曲面 ^[19]	整体; 表面	曲面物体	深度图象
广义高斯图象 ^[21]	整体; 表面	凸面物体	灰度图象
SpinImage ^[24]	局部或整体; 三维点	任意形状物体	深度图象
体素表达 ^[25]	整体; 体特征	无孔洞物体	深度图象
八叉树 ^[27]	整体; 体特征	曲面物体	深度图象
构造实体几何法 ^[28]	局部; 体特征	曲面物体	深度图象
体基元(geon) ^[30]	基于部件; 零件特征	曲面和互相连接的物体	深度图象
广义柱(GC) ^[31]	整体; 体特征	广义柱面	灰度图象
外观图 ^[33]	整体	凸多面体和一类曲面	灰度图象

3 匹配策略

根据 3D 物体识别过程, 从输入图象中得到物体表达后, 就需要对场景中存在的物体进行识别和定位. 这种识别是通过将场景中得到的物体特征和物体模型中的特征进行匹配来完成的, 而各种识别方法的不同之处不仅在于使用不同的物体表达方法, 而且也在于如何完成场景特征和模型特征之间的匹配、如何从场景模型特征匹配中得到一致解释和如何从一致解释中得到姿态估计.

如今识别和定位三维物体的主要方法有: ①假

设验证法; ②解释树; ③基于图的识别方法; ④霍夫姿态聚类; ⑤几何哈西表; ⑥迭代模型拟合技术; ⑦基于神经网络的匹配方法. 通常一个识别系统可能会联合使用多种匹配方法.

(1) 假设验证法

假设验证法是基于模型知识, 并采用了自底向上(bottom-up)与自上而下(top-down)相结合的策略, 这样可以大幅度地减少图象低层次处理与识别的计算时间^[33].

该方法是先将假设的物体模型特征和部分场景特征联合起来以得到一超限定(over-constrained)线性方程或非线性方程, 然后通过解此方程组, 以得

到标志匹配质量平均误差最小的解,并将其作为从物体参考坐标系到场景参考坐标系的变换.验证的过程是将假设模型的其它特征变换到场景中,再利用原场景图象来检查这些特征的位置是否正确,从而确定假设的正确性.通过配准的识别方法(Recognition-by-Alignment)就非常贴切地使用了这一思路^[34,35].

假设验证法面临的问题有两个:其一是在场景复杂的情况下,由于生成的假设很多,因而验证计算的花费会很大^[36];另一个是数据中存在的噪声误差会传播给姿态估计,以致将降低验证阶段确定相关的质量^[37].现今研究人员针对这些困难已作了大量工作,如 Gandhi and Camps 先由一小部分特征得到姿态估计,之后选择待匹配的特征,其目的是为了找到能够使姿态验证中数据不确定性最小的模型点的子集^[37].Jurie 也提出了一种快速、鲁棒的假设验证方法^[36],即先根据高斯误差模型在姿态空间内设置一定范围,并计算变换后的姿态在此范围内特征的可能性估计,然后选取具有最大概率的物体作为匹配结果的方法,这种配准方法是相当费时的.Kenji 和 Grimson 指出在最差情况下,配准方法计算复杂度是多项式级的^[35],因而为提高配准速度,他们提出了以相关特征组为中心的配准方法.现在这种假设验证法常与其它匹配方法相结合来完成识别工作,例如,王新华曾使用假设验证法与特征关系图相结合来完成识别^[38];Yi 和 Chelberg 则将其与几何哈希表相结合来完成识别^[39].

(2) 解释树(Interpretation tree, IT)法

构成解释树的节点常用于表示场景特征和模型特征间可能的匹配,且树上从根到叶子的每条路径都表示相关问题的一个可能解.这种方法就是为了找到一条路径以便得到传感器数据和模型特征之间的一致相关.但实际上并不需要搜索整个树来得到一致的解,可根据局部几何约束,例如,特征之间的角度和距离,即可以丢弃和修剪一些模型特征和场景特征之间不一致的匹配.当找到一个完整的路径后,就可以计算出物体整体的变换,以确定和验证物体的姿态位置.该算法的整体控制结构是采用顺序假设-测试加回溯的方式.Grimson 很好地研究并公式化了解释树的方法^[40].Ikeuchi 使用解释树搜索的方法将场景中的物体分类成为物体模型存储的一个外观^[41],然后在这些外观组中估计场景物体的姿态;Umasuthan 和 Wallace 又用代数多项式参数

的不变量作为 3D 物体表面特征,且在模型库中以 k-d 树的结构来存储这些代数不变量^[15];Grimson 在假设了一些条件后,得到的解释树搜索复杂度为 $O(n^2)$ ^[42],其中, n 是模型和单物体场景中的场景特征数,但在较差情况下(例如多物体场景中),复杂度会呈指数上升.

(3) 基于图的识别方法(Graph-based recognition)

特征关系图(Attributed-Relational Graphs, ARG)中的节点对应于场景和模型中的主要特征,而节点之间的弧则表示两个特征之间的关系.这里,场景和模型物体都用 ARG 图来描述,且它们之间的匹配均使用图论的匹配技术(例如图论中的子图同构(isomorphism)、单一同态(monomorphisms)和同胚(homomorphism)等方法)来实现.

但使用相关图的识别非常困难,因为图匹配算法是 NP 完备的,而对特征关系图的扩展则是特征超图表达(Attributed hypergraph representation, AHR)^[43],它包含的超边和超节点本身就是 ARG.由于应用 AHR 方法可以使用分层匹配策略,因此复杂度大大降低,如 Kao 等就使用超级关系图(Super relational graph, SRG)来识别边界表达(B-rep)的三维物体特征^[44],且 SRG 是一种无方向图,其图的节点代表一个面,其节点之间有两种连接,分别对应它们定义的两种面与面之间的关系.王新华则定义了 3 种面之间的关系(邻接关系、临接交面关系和平行关系),并构造了 3 种关系下的关连图,且分别进行匹配识别^[38].而 Dorai 又将她定义的最大一致形状片(Constant-shape maximum patch, CSMP)组织成一个无向图,并使用分层的图同构方法来识别由深度数据表达的三维物体^[45].这种基于图的识别方法的主要困难在于难以处理物体遮挡问题和图匹配的大运算量问题.

(4) 姿态聚类法(Pose cluster)

该法也称作广义霍夫变换.它是根据局部几何约束(例如角度和距离度量)来减少不正确的匹配,即先从每一个成功的匹配计算出一个几何变换,然后将其存储成霍夫空间(变换参数空间)的一个点,并进行累加^[46].在霍夫空间中最大的聚类点就确定了场景中物体姿态的一致假设,如 Grimson 和 Huttenlocher 曾使用“统计占有模型”(Statistical occupancy model)分析了霍夫聚类的敏感性^[47],他们得出的结论是:在霍夫累加器中最大误差与场景

中物体的的杂乱程度无关,且参数空间中误差极大值出现的概率依赖于传感器噪声、遮挡和量化效果。

(5) 几何哈希表法(Geometric hashing)

也称索引表(indexing)法,这种方法是将搜索空间设定为由待匹配特征作为索引值组成的查找表,该表中还包含对应物体模型的索引值,这样确定特征相关和数据库搜索过程则转变成表查找的过程。不同几何哈希表法的区别在于表中索引值类型和表的排列方法的不同。Stein 和 Medioni 就使用一种自己定义的溅射(splash)特征和 3D 曲线来分别描述物体的表面和边缘,并对这两种特征进行编码,以作为索引值^[10];Flynn 和 Jain 则根据同时可见的表面片三元组来生成表面-表面的约束关系,并将其作为索引键值^[48]。为提高识别效率,人们还引入了统计方法^[14, 39, 49],如 Wheeler 和 Ikeuchi 从每个物体的多幅深度图象中离线地编译图象和物体特征,然后使用马尔可夫随机场来表示假设之间的相似性,并选择可能性大的假设^[49]。Yi 和 Chelberg 则使用局部表面组(LSG, Local surface group)作为索引值,并将 LSG 特征对各个模型的后验概率也存储在索引表中,来综合考虑特征的检测准确性概率和后验概率,然后选择可能性大的物体优先验证^[39]。

(6) 迭代模型拟合法

这类方法的一个明显特征是在计算上的迭代过程。当三维物体用参数进行表达描述,并在该参数既能确定形状,也能确定物体姿态的情况下,可以使用迭代模型拟合的方法。此时,物体识别和姿态估计可简化成估算图象数据中的模型(姿态)参数,且可与存储的模型参数表达相匹配,即通过迭代的方法解一个过约束线性方程或者非线性方程系统,并选择与场景有最小平方和误差的解,来作为估计参数^[19]。如 Besl 提出了最邻近点匹配算法^[50],其目的是为了找到两幅深度图之间的变换关系,其具体实现过程是通过迭代的方法找到平均平方误差最小的变换。该法也称作 ICP 算法(Iterative closest point),后来该法广泛地应用到 3D 物体识别中^[51, 52]。

(7) 基于神经网络的方法

由于神经网络有很好的分类能力和并行处理能力,因此有些研究人员将其运用到 3D 物体识别中。如 Chen 和 Lin 曾使用一个单层 Hopfield 网络来识别 CSG 法表达的 3D 物体^[53],但该系统只能识别表面为平面的物体和单物体;又如 Zha 等同样也使用

单层 Hopfield 网络完成识别工作^[54],他们的方法是将物体模型表示成基于表面片的描述(Patch-based description),并定义了能量函数,其最小化时的网络节点状态即为识别结果,但与文献[53]不同的是,该方法可以识别场景中的多个物体。Ham 和 Park 也曾使用表面类型、矩特征、表面面积和相邻表面关系来描述 3D 物体^[55],并使用隐马尔可夫模型(Hidden markov model, HMM)来匹配场景和模型物体表面,然后将得到匹配表面的概率值输入到三层 BP 神经网络,以最终识别出物体。由于 HMM 是单独训练每一个物体模型,所以这个系统比较适用于模型库经常更新的情况。

与传统的方法相比,基于神经网络的识别方法具有能并行运算、处理速度快和鲁棒性好等优点。

4 结 论

本文围绕着三维物体识别系统必须解决的传感器类型、三维物体表达方法和匹配策略等 3 个方面问题,对近 10a 来的研究成果进行了综述,并对主要方法进行了分类和总结。通过以上分析和综述可见,在过去的 10a 里,人们已经达到了对不同主题的深刻理解,并完成了对过去系统的巨大改进,其中,首先,深度传感器的精度和数据采集速度得到了大幅提高,然而价格还是居高不下;其次,已研究出大量新的知识表达方法和匹配策略,且每种都适合于不同的表面形状和传感器假设,但没有哪一种方法能适应所有的应用,这也印证了视觉系统知识表达的选择不能独立于现实应用的这个事实。

因而今后三维视觉系统中还将对以下问题作深入研究:首先是任意形状(free-form)物体的表示和识别方法问题。尽管有前面叙述的众多 3D 物体表达方案,但大部分方法对物体的形状均作了限制,因此对任意形状物体的表示和识别仍将是研究热点;其次是实际应用中的复杂背景问题,因为有的表达方法和匹配策略只作了单物体的假设,而且实际获取图象中不可避免的噪声会对识别过程有一定影响,因此,该问题也需进一步探讨;第三,三维物体的表达方法一直存在着“整体和局部”的矛盾,它一方面表现在,针对实际应用中,物体可能只有部分可见的实际情况,而只能使用物体的局部特征,虽然局部特征容易从场景数据中提取,但是它又不能描述出物体的全部形状,因此其不足以分辨物体;另一方面

表现在,虽全局表达能更好地描述物体,且有更好的区分能力,但是它不能用在场景中物体被遮挡的情况,而且计算复杂度高;第四,识别速度问题,这是关系到是否能够满足实际应用的一个关键问题,由于人们对识别速度和识别准确程度都有着很高的要求,但两者存在一定矛盾,因而实际的识别系统是这两者的折衷方案.其可能的解决办法有两个,或者在匹配策略和模型数据库中综合使用不同层次的表达方法,以实现分层次的识别;或者将并行处理机制有效地引入到识别过程中.

致谢: 华中理工大学图象所田金文副教授对本文的修改提出了宝贵意见,特此表示感谢.

参 考 文 献

- Jain A K, Flynn P J (Eds). 3D object recognition systems. Amsterdam: Elsevier Science Publishers B.V., 1993.
- Arman F, Aggarwal J K. Model-based object recognition in dense range images A review. ACM Computing Surveys, 1993, 25(1): 5~ 43.
- Suk M, Bhandarkar S M. Three-dimensional object recognition from range images. Tokyo: Springer-Verlag, 1992.
- Hoover A. The space envelope representation for 3D scenes. PhD. Dissertation. Univ. of South Florida, 1996.
- Arie J B, Nandy D. A volumetric/iconic frequency domain representation for objects with application for pose invariant face recognition. ITPAMI, 1998, 20(5): 449~ 457.
- Grimson W E L, Lozano-Peres T. Localizing overlapping parts by searching the interpretation tree. ITPAMI, 1991, 13(9): 920~ 934.
- Fan T J, Medioni G, Nevatia R. Recognizing 3D objects using surface descriptions. ITPAMI, 1989, 11(11): 1140~ 1157.
- Besl P J. Surfaces in range image understanding. New York: Springer-Verlag, 1988.
- 孙龙祥等. 深度图象分析. 北京: 电子工业出版社, 1996.
- Stein F, Medioni G. Structural indexing: Efficient 3-D object recognition. ITPAMI, 1992, 14(2): 125~ 145.
- Dorai C, Jain A K. COSMOS—A representation scheme for 3D free-form objects. ITPAMI, 1997, 19(10): 1115~ 1130.
- Godin G D, Levine M D. Structured edge map of curved objects in a range image. In: IEEE Conference on Computer Vision and Pattern Recognition, San Diego California, 1989: 276~ 281.
- Chen S, Stockman G. Object wings— $2\frac{1}{2}$ D primitives for 3-D recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, San Diego California, 1989: 535~ 540.
- Chen J L, Stockman G C. 3D free-form object recognition using indexing by contour features. CVIU, 1998, 71(3): 334~ 355.
- Umasuthan M, Wallace A M. Model indexing and object recognition using 3D viewpoint invariance. PR, 1997, 30(9): 1415~ 1434.
- 李松涛. 深度图象处理方法的研究[博士学位], 北京: 清华大学, 1999.
- Liao Ch W, Medioni G. Representation of range data with B-spline surface patches. In: Proc, International conference on Pattern Recognition, Atlantic Washington, New Jersey, 1992: 745~ 748.
- Barr A H. Superquadrics and anglepreserving transformations. CG&A, 1981, 1: 11~ 23.
- Solina F, Bajcsy R. Recovery of parametric models from range images: The case for superquadrics with global deformations. ITPAMI, 1990, 12(2): 131~ 147.
- 周林, 袁保宗. 扩展超二次曲面: 一种新的光滑变形曲面模型. 电子学报, 1998, 26(8): 46~ 50.
- Liang P, Taubes C H. Orientation-based differential geometric representations for computer vision applications. ITPAMI, 1994, 16(3): 249~ 258.
- Ghosh P K, Kumar K V. Support function representation of convex bodies. Its Application in Geometric Computing, and Some Related Representations. CVIU, 1998, 72(3): 379~ 403.
- Johnson A E, Hebert M. Surface matching for object recognition in complex three-dimensional scenes. IVC, 1998, 16: 635~ 651.
- Johnson A E *et al.* Toward a general 3-D matching engine: Multiple models, complex scenes, and efficient data filtering. In: Image Understanding Workshop, Monterey, California, 1998: 1097~ 1107.
- Best L, Magee M. Autonomous construction of three-dimensional models from range data. PR, 1998, 31(2): 121~ 136.
- Chien C H, Sim Y B, Aggarwal J K. Generation of volume/surface octree from range data. In: IEEE Conference on Computer Vision and Pattern Recognition, Ann Arbor Michigan, 1988: 254~ 260.
- Li A, Crebbin G. Octree encoding of objects from range images. PR, 1994, 27(5): 727~ 739.
- 马利庄, 王荣良. 计算机辅助几何造型技术及其应用. 北京: 科学出版社, 1997.
- Biederman I. Higher level vision, in An invitation to cognitive science. MA: MIT Press, 1989.
- Nguyen Q L, Levine M D. Representing 3-D objects in range images using geons. CVIU, 1996, 63: 158~ 168.
- Ponce J *et al.* Invariant properties of straight homogeneous generalized cylinders and their contours. ITPAMI, 1989, 11(9): 951~ 966.
- Plantinga W H, Dyer C R. Visibility, occlusion, and the aspect graph. IJCV, 1990, 5: 137~ 160.
- 马颂德, 张正友. 计算机视觉—计算理论与算法基础. 北京: 科学出版社, 1998.
- Huttenlocher D P, Ullman Sh. Recognizing solid objects by alignment with an image. IJCV, 1990, 5(2): 195~ 212.
- Nagao K, Grimson W E L. Object recognition by alignment

- using invariant projections of planar surfaces. MIT: AI Lab Tech. Report, 1994.
- 36 Jurie F. Robust hypothesis verification: application to model-based object recognition. PR, 1999, 32: 1069~ 1081.
- 37 Gandhi T, Camps O. Robust feature selection for object recognition using uncertain 2d image data. In: IEEE Conference on Computer Vision and Pattern Recognition, Seattle Washington, 1994: 281~ 287.
- 38 王新华. 三维工业物体识别与测量方法的研究[博士学位论文], 武汉: 武汉测绘科技大学, 1997.
- 39 Yi J H, Chelberg D M. Model-based 3D object recognition using bayesian undexing. CVIU, 1998, 69(1): 87~ 105.
- 40 Grimson W E L, Peres T L. Localizing overlapping parts by searching the interpretation tree. ITPAMI, 1991, 13(9): 920~ 934.
- 41 Ikeuchi K, Hong K S. Determining linear shape change: Toward automatic generation of object recognition programs. CVGIP, 1991, 53(2): 154~ 170.
- 42 Grimson W E L, Huttenlocher D P. On the verification of hypothesized matches in model-based recognition. ITPAMI, 1991, 13(12): 1201~ 1213.
- 43 Ki A *et al.* Recognition and shape synthesis of 3D objects based on attributed hypergraphs. ITPAMI, 1989, 11(3): 279~ 290.
- 44 Kao C Y, Kumara S R T, Kasturi R. Extraction of 3D object features from SAD boundary representation using the super relation graph method. ITPAMI, 1995, 17(12): 1228~ 1233.
- 45 Dorai Ch. COSMOS: A framework for representation and recognition of 3D free-from objects. PhD. Dissertation, Michigan State Univ., 1997.
- 46 Stockman G C. Object recognition and localization via pose clustering. CVGIP, 1987, 40: 361~ 387.
- 47 Grimson W E L, Huttenlocher D P. On the sensitivity of the hough transform for object recognition. ITPAMI, 1990, 12(3): 255~ 274.
- 48 Flynn P J, Jain A K. 3D object recognition using invariant feature indexing of interpretation tables. CVGIP: IU, 1992, 55(2): 119~ 129.
- 49 Wheeler M, Ikeuchi K. Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition. ITPAMI, 1995, 17(3): 125~ 145.
- 50 Besl P, McKay N. A method for registration of 3-D shapes. ITPAMI, 1992, 14(2): 239~ 256.
- 51 Zhang Z. Iterative point matching for registration of free-form curves and surfaces. IJCV, 1994, 13(2): 119~ 152.
- 52 方昭江. 基于模型的三维物体匹配方法研究[硕士学位论文], 武汉: 华中理工大学图象所, 1999.
- 53 Chen T W, Lin W Ch. A neural network approach to CSG-based 3D object recognition. ITPAMI, 1994, 16(7): 719~ 726.
- 54 Zha H B, Nanamegi H, Nagata T. Recognition 3-D objects by using a hopfield-style optimization algorithm for matching patch-based descriptions. PR, 1998, 31(6): 727~ 741.
- 55 Ham Y K, Park R H. 3D object recognition in range images using hidden markov models and neural networks. PR, 1999, 32: 729~ 742.



李 庆 1973 年生, 1996 年获华中理工大学通信与电子系统专业硕士学位, 现为华中理工大学图象所模式识别与智能控制专业博士研究生. 主要研究领域为三维物体识别与建模、深度图象处理、人工智能等.



周曼丽 1939 年生, 教授, 博士生导师, 1962 年毕业于华中工学院无线电技术专业. 主要从事模式识别与人工智能、多媒体通信和信息系统方面的研究. 在国内外有关刊物、会议发表论文 50 余篇.



柳 健 1939 年生, 教授, 博士生导师, 1961 年毕业于华中工学院无线电技术专业. 主要研究方向为数字图象处理与识别、计算机视觉和遥感图象分析等. 近几年来发表学术论文 80 余篇.