

基于多分辨率格网的三维物体识别方法

李 庆¹, 周曼丽², 柳 健¹

(1. 华中理工大学图像所, 图像信息处理与智能控制教育部重点实验室, 湖北武汉 430074; 2. 华中理工大学电信系, 湖北武汉 430074)

摘 要: 本文首先提出了一种改进的三维物体表达方法, 它将一个三维物体表面网格与其它表面网格的几何关系表示为一个二维矩阵, 称为距离角度图. 这种表达能够描述任意形态物体, 抑制杂乱背景和遮挡, 几何意义直观, 且适应不同分辨率、非规则的三角格网. 然后, 以这种表达方法为基础, 本文阐述了一种基于多分辨率格网的, 由粗到精的三维物体识别方法. 它先在场景和模型的低分辨率格网上进行粗匹配以得到模型候选集合, 之后在已匹配网格的高分辨率格网邻域上筛选模型候选集合, 最后综合考虑多个网格对应的模型候选以得到最终模型候选的确认和验证. 这种识别方法具有运算量小、准确可靠等优点, 实验证明该方法正确有效.

关键词: 计算机视觉; 三维物体识别; 多分辨率格网; 深度图像

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2001) 07-0891-04

Multi-Resolution Mesh Based 3D Object Recognition

LI Qing¹, ZHOU Man-li², LIU Jian¹

(1. State Key Laboratory for Image Processing and Intelligent Control, Inst. for Pattern Recognition and Artificial Intelligence, Wuhan, Hubei 430074, China; 2. Dept. of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China)

Abstract: In this paper, we proposed an improved 3D object representation first, which computes the local signatures of a given basic polygon on the surface mesh, and converts the signatures to a 2D array called the distance-angle (DA) images. This representation is adaptive to free-form objects, resistant to occlusion and clutter. Furthermore, it has distinct meaning, and adaptation to different resolution and irregular triangle mesh. Secondly, based on the representation, a novel 3D object recognition algorithm is presented, which is a multi-resolution mesh based, coarse-to-fine recognition. By matching the DA image of a polygon in the scene surface mesh with the DA images of models in low resolution, a model candidate set is obtained. The set is filtered in the neighborhood of the matched polygons in high-resolution mesh and verified by the model candidate sets of other polygons. Experiments show that this algorithm needs less computation, and is very accurate and robust.

Key words: computer vision; 3D object recognition; multi-resolution mesh; range image

1 引言

出于大量现实应用的需要, 例如机器人工件抓取任务 (bin-picking)、自动导航、自动检测、装配任务和显微图像与医学 X 光图像的分析等等, 三维物体识别成为一个很活跃的研究领域.

根据描述三维物体出发点的不同, 三维物体识别方法可以分为以物体为中心的 (object centered) 和以观察者为中心的 (viewer centered)^[1]. 根据所使用几何特征的不同, 可以分成基于表面的、基于不连续性的和基于体的等三种^[2]. 这些众多的三维物体识别方法都要面临以下难题: 首先, 如何描述任意形状 (free-form) 的物体; 其次, 如何处理物体被遮挡和自遮挡的情况; 最后, 算法的计算复杂性.

为较好地解决上述问题, Johnson 提出了一种称为 spirim-

ages 的三维物体表达和相应识别办法^[3]. 该方法可以描述任意形状的物体; 通过对支持区域大小的控制, 可以有效地抑制杂乱背景和遮挡在物体表达上的影响. 其不足是: 由于物体表达的计算由模型点集得到, 因此其对数据点的分辨率敏感, 格网分辨率 (Johnson 定义格网分辨率为格网中边的平均长度, 对于同一物体分辨率越高, 顶点和面的数目越多) 越低, 参加计算的数据点越少, 实际应用中要求模型和场景物体具备相同的格网分辨率^[3]. 另, 计算量太大, Johnson 提出了简化方法, 然而简化过程中要计算一个巨大矩阵的特征向量, 难以实现.

为解决 Johnson 算法中的困难, 本文提出了一种改进的三维物体表达方法, 能够部分克服不同格网分辨率对表面匹配带来的影响, 并以此表达提出了利用多分辨率格网进行三维物体识别的方法, 在本文的最后介绍了识别实验结果和结论.

2 一种改进的三维物体表达方法

首先申明下文中使用格网(mesh)对应物体的全部多边形表面,而网格(face)对应格网中的单个面。

2.1 距离角度映射的定义

已知一三维物体的格网 M 。以格网中的某网格作为基准,考虑此基准网格与其它网格的位置关系。这样在识别定位阶段,若能够在输入图像提取的格网中找到某个网格与其它网格的位置关系与前面的结果相同,即找到两个网格之间的对应关系,也即完成了识别定位任务。

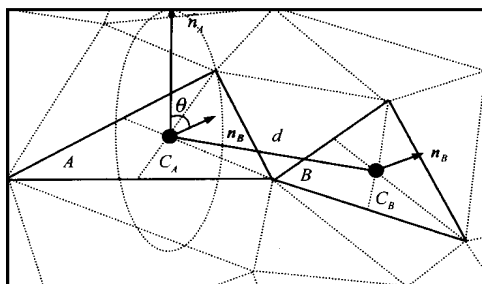


图1 距离角度映射示意图

在格网中任取一网格 A , 作为基准网格, 任取一网格 B , 如图1所示。定义距离角度映射 f_A 为:

$$f_A: R^3 \rightarrow R^2, f_A(B) (d, \theta) = (|C_A - C_B|, \arccos(n_A \cdot n_B)) \quad (1)$$

式中 C_A, C_B 分别是网格 A 与 B 的质心, n_A, n_B 分别是网格 A 与 B 的单位法矢量, 即 d 是网格 A 与 B 质心间的距离, θ 为网格 A 与 B 法矢量之间的夹角。则根据映射 f_A , 能将模型上所有网格映射到一个二维点的集合, 将此集合写作 $\{F_A(M)\}$, M 中的每个网格都对应一个集合, 为方便计算和存储, 需将此二维点集转换为矩阵形式, 称作距离角度图。

2.2 距离角度图的生成

首先需确定距离、角度坐标上每个矩阵元素对应的单元大小 bin 与 bin_d , 如式(2)。考虑到要使不同分辨率格网表达有相同的大小尺寸, 实验中将 J_{max} 与 I_{max} 固定为 18×32 , 这样的距离角度图既能充分反映数值差异, 数据量也不大。点集中任一个点 (d, θ) 对应到矩阵中位置 (i, j) 的计算如式(3):

$$bin = \frac{J_{max}}{\max(d)}, \quad bin_d = \frac{I_{max}}{\max(\theta)} \quad (2)$$

$$j = \text{floor}\left(\frac{d}{bin}\right) + 1, \quad i = \text{floor}\left(\frac{\theta}{bin_d}\right) + 1 \quad (3)$$

式中 $\text{floor}()$ 为向零方向的取整函数。本文采用加权双线性内插的方法计算单个点对它周围的行列值的贡献大小, 权值大小为网格的面积。距离角度图在生成方法见图2, 数据点上的线段高度对应权值的大小。

图2(b)为加权双线性内插的示意图。假设点 p 的坐标为 (d, θ) , 对应矩阵中的位置为 (i, j) 。 p 对应的网格为 S_p , 面积为 S_p , 距离角度图矩阵记作 O , 则点 p 对它周围行列值的贡献可以由下式描述:

$$O(i+1, j+1) = \frac{S_p}{S_M} (1 - \frac{d - d_{i,j}}{bin_d}) * (1 - \frac{\theta - \theta_{i,j}}{bin})$$

$$(-1) * d) * S_p, = P(0,1) \quad (4)$$

整理前面的叙述, 格网 M 中以网格 A 为基准的距离角度图生成算法总结如下:

- (1) 计算 A 网格的质心位置, M 中所有网格的质心位置集合 $\{C_M\}$ 与面积集合 $\{S_M\}$;
- (2) 按照式(1)计算 M 中所有网格的距离角度坐标集合 $\{DA_M\}$; 按照式(2)计算 bin 与 bin_d ;
- (3) 在 M 中取得任一个未计算距离角度图投票的网格 B ; 按照式(3)计算 B 对应的距离角度图矩阵的位置 (i, j) ; 按照式(4)计算 B 对距离角度图矩阵的双线性内插;
- (4) 若 M 中还有未经过计算的网格 B , 则返回式(3), 否则结束, 得到距离角度图。

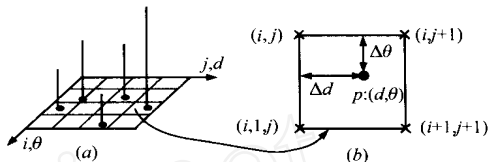


图2 距离角度图的生成。(a) Distance-angle images; (b) 加权双线性内插

图3是使用上述算法得到的距离角度图实例。三维物体上所有网格的距离角度图集合构成物体模型。距离角度图的相似性度量采用文献[3]中介绍的方法。

2.3 距离角度图的分析

由于从物体格网表示得到, 因此距离角度图可以描述任意形状的物体。若设定距离门限 D 和角度门限 θ , 规定只有小于门限的网格才能参与构造计算, 则就可以限制距离角度图只反映局部特征, 进而抑制杂乱背景和遮挡的影响。这两个门限称作距离角度图的生成参数。本文方法中考虑了网格的面积信息, 保证不同分辨率下的距离角度图累加和大致相同, 能够直接使用文献[3]中的简化匹配方法。另外两个生成参数几何意义直观, 例如若设定角度门限为 90° , 则意味着当视角与基准网格平行的时候, 参加计算仅仅是能够观察到的网格。

总的来说, 距离角度图有能够描述任意形态物体, 抑制杂乱背景和遮挡, 几何意义直观, 且适应不同分辨率、非规则的三角格网等优点。

3 利用多分辨率格网进行三维物体识别

在距离角度图的基础上, 本文提出了一个利用多分辨率格网进行三维物体识别的方法。

3.1 识别算法简介

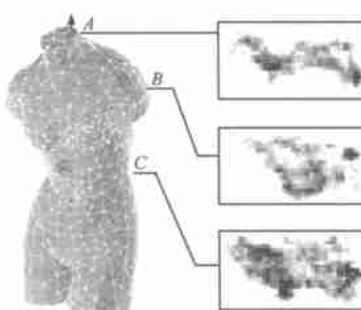


图3 一个距离角度图的例子, 左边为维纳斯模型的三角形格网, 右边由上到下分别是以黑边三角形 A, B, C 为基准生成的距离角度图

首先,先浅显地探讨该方法存在的视觉心理基础.这里只考虑对几何信息的感受.人观察一个物体并不是一次获得全部信息后才作出判断和处理,而是一个由粗到精的过程.先获得粗略的形状信息,再根据具体需要增加细节以改善对物体的判断.例如,观察图 4(a)的兔子,只根据外形轮廓便可分辨出来,而眼睛等细节特征则无太大关系.另外,三维物体能够以近似的形状表达出来. Hoover^[4]指出可以在尺度空间里由粗到精地表达物体表面.人们对格网简化算法也作了很多研究^[6],图 4(b)是兔子的简化格网表示,省略了很多表面细节.

简单地讲,本文的识别方法是先在低分辨率下进行场景中物体和模型的粗匹配,以此为基础进行细分率上的匹配,作为对粗匹配结果的完善和确认.

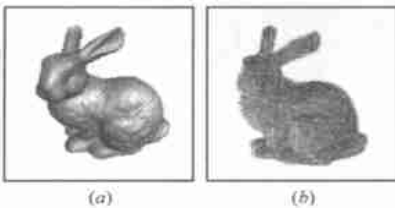


图 4 (a) 兔子模型的原始灰度图像;
(b) 简化后的格网表示

3.2 识别过程详解

3.2.1 模型库的建立 在我们的系统中,模型库中保存各个模型在多个分辨率下的格网表达,记做 M_i^j , i 对应不同的模型, j 对应不同的分辨率.低分辨率格网是由格网简化算法^[6]得到的.

3.2.2 建立场景中单个网格的模型候选集合 在匹配之前先进行预处理工作:计算所有模型在最低分辨率下所有网格的距离角度图,存入距离角度图堆栈.模型识别的具体过程是:在场景图像中取得最初的格网表达 s^0 ,此格网的分辨率是最高的.按照一定算法进行格网简化,得到 n 个不同分辨率等级的格网,分别记做 $s^{-1} \dots s^{-n}$.在 s^{-n} 中任取一个网格 p^{-n} ,计算 p^{-n} 的距离角度图.与模型库中所有模型最低分辨率格网距离角度图进行相似性比较,生成相似性直方图,选择相似性度量在一定门限以上网格加入对应网格候选集 $\{R_c\}$,涉及到的模型加入到模型候选集 $\{M_c\}$.

3.2.3 筛选单个网格的模型候选集 低分辨率下得到的距离角度图对三维物体形状描述的近似误差较大,因此,要在高分辨率上对模型候选集进行筛选和确认.在介绍此过程之前,先介绍两个定义.

定义 1 给定格网 s^{-n} 上的网格 p^{-n} ,则在高分辨率格网 s^{k-n} 中与 p^{-n} 的质心欧氏距离最近的网格 q^{k-n} 称为 p^{-n} 在 s^{k-n} 上的最邻近网格.

定义 2 假设网格 $p^j \in M^j$, $n > 0$,则在更高分辨率格网 M^{j+n} 中所有满足下式条件的网格,构成网格 p^j 在 M^{j+n} 中的邻域.式中 R 为邻域大小阈值,即以网格 p^j 的质心为球心,半径为 R 的球所包含的 M^{j+n} 中的网格构成网格 p^j 在 M^{j+n} 中的邻域.

$$\{q^{j+n} | d(C_q, C_p) \leq R, q^{j+n} \in M^{j+n}\} \quad (8)$$

首先,在更高分辨率的场景格网中,例如,在中 s^{-n+1} 中,找到 p^{-n} 的最邻近网格 p^{-n+1} ,计算 p^{-n+1} 的距离角度图;然后,选择对应网格候选集中的一个候选网格 r_i^1 ,在此网格所属的模型的更高分辨率网格中,例如在 M_i^2 中找到的 r_i^1 邻域,计

算邻域内网格的距离角度图,与 p^{-n+1} 的距离角度图进行相似性比较,得到相似性度量.如此处理对应网格候选集中的所有网格,生成相似性直方图.选取一定门限以上网格构成新的对应网格候选集,相关的模型构成新的模型候选集.

由于是在高分辨率上计算的距离角度图,对物体模型形状的近似误差减小,因此,去除初始对应网格候选集中不正确匹配的网格,达到筛选模型候选集的目的.另外,因为找到的是更小网格的对应关系,所以模型与场景物体的位置变换关系也更精确.经过这个步骤,我们得到了由场景中一个网格的对应网格和模型候选,还需要计算场景中其他网格,对模型候选做进一步的验证和确认.

3.2.4 验证模型候选集 经过前面的步骤我们得到了场景中单个网格的模型候选集合,仅通过单个网格反映的物体局部特征来确定模型是不完全的,且由于场景中存在的噪声影响可能会得到错误结果.因此,需要计算多个网格的模型候选集合进行验证和确认工作.

令 N_i 为所有网格候选集中模型 M_i 出现的次数, N_A 为全部经过识别计算的网格个数,有 $N_i \leq N_A$.我们采用了如下算法验证模型候选集:

- i 从单个网格的模型候选集合生成初始的 $\{N_i\}$ 和 N_A ;
 - ii 若 $\max(\{N_i\})/N_A > 0.8$, 且 $\max(\{N_i\}) > 3$, 则识别成功,场景中物体为 M_i , 否则执行步骤(i);
 - iii 在场景中再任选一个网格,计算它的模型候选集合,重新计算 $\{N_i\}$ 和 N_A . 若 $N_A > 20$, 则识别失败,否则执行步骤(i).
- 比起简单地取各个网格模型候选集的交集作为最后的识别结果,这样处理不会因为由于单个网格的错误结果不能得到识别结果,而且多个网格都支持同一模型,保证了识别结果的可靠性.

4 实验结果

在实验中使用的模型库是包含三个模型的小模型库.三个模型如图 5.

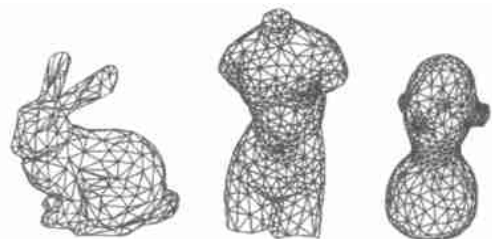


图 5 模型库中的模型

由于简化方法的计算量较大,且这里关心的只是识别而非定位,因此实验中使用单一分辨率的场景格网,高低两个分辨率的模型格网完成识别操作.初始场景格网 s^0 由文献[5]中介绍的算法获

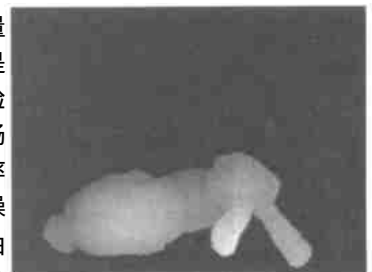


图 6 实验图像

取,直接使用格网 S^0 进行计算是不合适的,因为格网 S^0 中格网的数目很多,运算量太大,本文使用 Garland 的算法^[6]将其网格数目简化到 10^3 数量级。

我们暂时只对场景中有一单一物体的情况作了实验。下面给出一个识别的例子,使用的实验图像为 Stanford 大学计算机系图形研究组使用 Cyberware 3030MS 激光相机获取的深度图像,大小为 400×512 ,见图 6。

图 7 为识别过程中得到的部分距离角度图相似性直方图。(a, b, c) 为一个场景网格在模型低分辨率格网上粗匹配的相似性直方图。实验数据和识别结果见表 1。

由识别结果看出,从测试图像可以正确识别场景中的物体。我们另外还使用了兔子和木偶的多幅深度图像进行了测试,均能得到正确的识别结果,证实了该方法的有效性。

表 1 实验数据和识别结果

步数	筛选前的候选网格集合元素个数	筛选前的候选模型集合	筛选后的候选模型集合
1	bunny:8 venus:5 doll:7	{bunny, doll, venus}	{bunny}
2	bunny:5 venus:10	{bunny, venus}	{bunny, venus}
3	bunny:4	{bunny}	{bunny}
4	bunny:8	{bunny}	{bunny}
5	bunny:3	{bunny}	{bunny}
识别结果:	Bunny		

5 结束语

只根据物体的近似几何形状就可以粗略地识别物体,在需要的时候补充必要的细节几何特征。本文的实验结果证实了这个想法是可行的。与固定分辨率下的识别方法^[3]相比,本文提出的方法具有速度更快,准确性更高的优点。

致谢 本文中使用的兔子模型和深度图像来自 Stanford 大学的 Computer Graphic 研究组,木偶模型和深度图像来自 Washington 大学的 Intelligent System 实验室,在此向他们表示感谢。

参考文献:

- [1] J Berr-Arie, D Nandy. A Volumetric/ Iconic Frequency Domain Representaion for Objects With Application for Pose Invariant Face Recognition [J]. IEEE Trans. 1998, PAMF-20 (5): 449 - 457.
- [2] F Arman, J K Aggarwal. Model-Based Object Recognition in Dense-Range images, A Review [J]. ACM Computing Surveys, 1993, 25 (1): 5 - 43.
- [3] A E Johnson, M Hebert. Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes [J]. IEEE Trans. ,1999, PAMF-21 (5): 433 - 449.
- [4] A Hoover, et al. Dynamic-Scale Model Construction for Range Imagery [J]. IEEE Trans. 1998, PAMF-20 (12): 1352 - 1357.
- [5] G Turk, M Levoy. Zippered Polygon Meshes from Range Images [A]. SIGGRAPH94 [C].
- [6] M Garland, P S Heckbert. Surface Simplification Using Quadric Error Metrics [A], SIGGRAPH97 [C].

作者简介:



李 庆 男, 1973 年生于天津, 1996 年于华中理工大学获得通信与电子系统专业硕士学位, 现为华中理工大学图像所模式识别与智能控制专业博士研究生, 主要研究领域: 三维物体识别与建模、深度图像处理、人工智能等。



周曼丽 女, 1939 年生于江西上饶, 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 男, 1939 年出生于长沙, 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像处理与识别、计算机视觉和遥感图像分析等。近几年来发表学术论文 80 余篇。