

·信号与信息处理·

基于多视图 SIFT 特征的三维模型检索

江秋鑫¹, 华顺刚¹, 时树胜²

(1. 大连理工大学机械工程学院, 辽宁 大连 116024; 2. 大连理工大学教务处, 辽宁 大连 116024)

摘要: 研究了一种基于多视图 SIFT 特征的三维模型检索算法。首先对三维模型进行多视图投影, 得到其全方位的二维投影深度图, 并在各投影深度图上提取 SIFT 特征。分别利用按模型特征数比例分次建立码本及建立模型库整体码本这两种方式建立了模型库的码本, 继而将模型的 SIFT 特征聚类量化, 并用一个多维向量将其表示出来。通过计算三维模型特征向量的 KL 散度实现三维模型的相似性匹配。实验结果表明, 基于多视图 SIFT 特征的三维模型检索算法对于有关节模型和刚性模型均能获得较好的检索性能。

关键词: 三维模型检索; SIFT 特征; k-means; 多视图投影

中图分类号: TP391.41

文献标识码: A

文章编号: 1673-1255(2011)04-0056-05

Multi-View 3D Retrieval Using SIFT Feature

JIANG Qiu-xin¹, HUA Shun-gang¹, SHI Shu-sheng²

(1. School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China;

2. Teaching Affairs Division, Dalian University of Technology, Dalian 116024, China)

Abstract: A 3D model retrieval algorithm based on multi-view scale invariant feature transform (SIFT) feature is researched. By carrying on the multi-view projection of a 3D model, the omnidirectional 2D projection depth maps are obtained, from which SIFT features are extracted. Individual-k-means and Whole-k-means are used to establish the codebook for shape benchmark respectively. Using the feature quantization codebook, all the features are clustered and generated the histogram as the 3D model feature vector. The similarity matching of 3D model is realized by calculating Kullback-Leibler (KL) divergence according to the feature vector. Experiments show that the algorithm can be adequate to both articulated shapes and rigid shapes, and gain a better retrieval effect.

Key words: 3D retrieval; scale invariant feature transform(SIFT) feature; k-means; multi-view projection

三维模型由于能够提供比二维图像更丰富、直观的视觉效果, 而被广泛应用于人们生活、工作的各个领域, 由此产生了大量的三维模型。人们常常希望在已经存在的三维模型中寻找想要的模型实现重用。因此随着三维模型库的不断增大及数量的不断增多, 使得帮助用户快速准确地从海量三维模型库中获取所需三维模型, 实现资源重用的三维模型检索技术成为了计算机视觉、计算机图形学界的一个研究热点。

现有的三维模型检索技术可以分为基于统计特征、基于拓扑结构、基于几何结构分析和基于视觉投影四大类。基于统计特征的方法通过对三维模型的某一特征信息进行统计, 获得统计直方图后将其量化而得到三维模型特征向量。基于拓扑结构的方法通过提取三维模型的拓扑结构信息来得到三维模型的特征描述符。基于几何结构分析的方法对三维模型在几何空间中的分布特征进行了描述。基于视觉投影的三维模型检索算法先对三

收稿日期: 2011-07-12

基金项目: 国家部委预研项目(40401010105)

作者简介: 江秋鑫(1986-), 女, 山西太原人, 硕士研究生, 主要研究方向为计算机图形图像处理。

维模型进行二维投影,再对二维投影进行特征的提取用以描述三维模型。这种方法将复杂的三维问题转化为较为成熟的二维图像处理问题,对于三维模型的噪声、简化及细分均具有较好的鲁棒性。

目前,对于基于视觉投影类方法,已有多种算法被相继提出。普建涛^[1]等提出使用一个平面沿着切割方向逐步与模型进行相交求解,得到一系列交点组织成二维多边形集,采用形分布(shape distributions)思想对二维多边形集进行均匀采样,归一化后使用广义的欧氏距离作为相似性度量。Mohamed Chaouch^[2]等将三维模型在正二十面体的20个顶点处进行投影,从所得深度图上提取一系列深度线(depth line),将其序列化后得到一个深度序列(depth sequence)作为三维模型的特征描述符,并用动态规划距离(dynamic programming distance)进行模型的相似性计算。Thibault Napoleon^[3]等对经过CPCA(continuous principal component analysis)预处理后的三维模型进行二维投影,并对每一幅二维投影图进行外形描述,从而形成一种多尺度外形描述符对三维模型不同尺度的轮廓的凹凸性进行描述。

文中研究了一种基于视觉投影的三维模型检索算法。它从三维模型的一系列二维投影图像上提取出SIFT(scale invariant feature transform)^[4]特征,聚类后作为三维模型的特征向量,并利用KL散度(Kullback-Leibler divergence)进行模型相似性匹配,实现了对三维模型的检索。

1 算法描述

基于多视图SIFT特征的三维模型检索算法通过一系列局部多尺度视觉特征来对三维模型进行描述。首先获得三维模型多个视图方向的一系列投影深度图,然后从所有的投影深度图上提取出SIFT特征。SIFT特征是由David G Lowe提出的一种图像局部特征,其实质是在不同的尺度空间上查找出图像的关键点,并用一组向量将这些含有位置、尺度和方向信息的关键点描述出来。该特征对图像尺度、旋转具有不变性,对亮度和视角改变也能保持一定稳定性。因此,该特征被广泛应用于计算机视觉领域的图像匹配、模式识别等^[5,6]方面。

从一个三维模型的多幅投影深度图上提取出

的SIFT特征可高达几千个,若直接对2个三维模型的SIFT特征进行计算将十分耗时,所以,考虑将三维模型的所有SIFT特征量化,累积成一个直方图作为该模型的特征向量来减少计算量。其中,量化所需的码本(code book)是通过k-means聚类得到的。图1所示为文中算法中三维模型特征向量生成过程图。此三维模型检索算法的具体步骤如下:

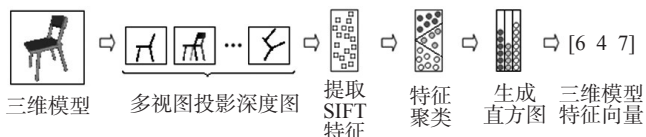


图1 SIFT特征向量生成过程

(1) 模型预处理

对于任意三维模型,其初始空间位置、尺寸及方向具有很大的随机性,为了对三维模型在相同尺度、相同规则下进行比较检索,首先对其进行位置和尺寸标准化预处理。

位置标准化是在求出三维模型的重心后将其作为坐标系原点更新三维模型上所有点的坐标。使用的三维模型重心公式为

$$x_c = \frac{1}{N} \sum_{i=1}^N x_i, y_c = \frac{1}{N} \sum_{i=1}^N y_i, z_c = \frac{1}{N} \sum_{i=1}^N z_i \quad (1)$$

其中, N 为三维模型上点的个数; (x_i, y_i, z_i) 为三维模型上各点的原始坐标; (x_c, y_c, z_c) 为模型的重心坐标。

尺寸标准化是在找到一个以三维模型重心为球心的能够包含该三维模型的最小球面后,利用其半径对三维模型进行归一化处理

$$\begin{aligned} x_i' &= \tilde{x}_i / (2 \times r_{\max}), y_i' = \tilde{y}_i / (2 \times r_{\max}), \\ z_i' &= \tilde{z}_i / (2 \times r_{\max}) \end{aligned} \quad (2)$$

其中, $(\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)$ 表示位置标准化后的点的坐标; (x_i', y_i', z_i') 为尺寸标准化后的三维模型的点的坐标; $r_{\max} = \max_{1 \leq i \leq N} \sqrt{\tilde{x}_i^2 + \tilde{y}_i^2 + \tilde{z}_i^2}$ 为最小球面半径。

(2) 三维模型多视图投影

通常来说,如果2个三维模型从所有的视角看上去都相似,则这两个模型相似。所以,可以用三维模型的多个视角投影深度图的相似性来判断三维模型的相似性。综合考虑描述完整性和计算复杂度后,采用80面体的42个顶点作为投影视点对三维模型进行投影,得到其42个不同视角的投影深度图(图2)。文中各投影深度图的尺寸设定为 $256 \times$

256。

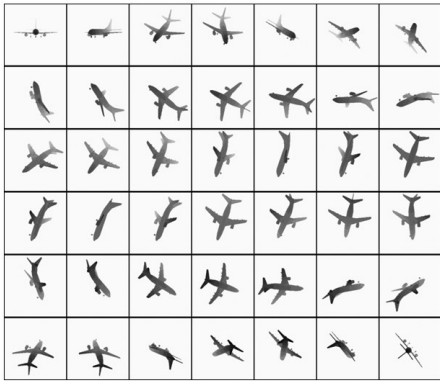


图2 飞机模型的42幅投影深度图

(3) 三维模型 SIFT 特征提取

得到三维模型多个视图的投影深度图后,就可以从各投影深度图上提取 SIFT 特征。分别对 McGill (McGill university benchmark database)、PSB (princeton shape benchmark) 和 ESB (purdue engineering shape benchmark) 3 个模型库进行了 SIFT 特征提取实验。McGill 内每个模型可提取大约 1 500 个 SIFT 特征, PSB 内的每个三维模型可提取大约 1 000 个 SIFT 特征, ESB 内的每个三维模型可提取大约 550 个 SIFT 特征。其中每个 SIFT 特征均为 1 个 128 维的向量。

(4) 三维模型特征描述符

首先利用 k-means 算法对 3 个模型库分别聚类生成各自的码本。在执行 k-means 程序时,需要对聚类个数进行设定,所以考虑了 2 种方式对其进行设定并对 2 种方式的聚类效果进行了比较讨论。

①按模型特征数比例分次建立码本:将模型库内所有模型的 SIFT 特征根据模型类别分为 n 类,并统计出各个类的 SIFT 特征个数: $N_i (i=1, 2, \dots, n)$ 。则第 i 次执行 k-means 程序的聚类中心数为 $k_i =$

$$\left(N_i / \sum_{i=1}^n N_i \right) \times k。其中, k 为模型库码本的维数。$$

②建立模型库整体码本:将模型库码本的维数作为聚类中心个数 k ,则执行一次 k-means 程序就可得到模型库的码本。

在建立了模型库的码本后,将三维模型的每一个 SIFT 特征与码本进行比对,用码本中与该特征最近的码字对其进行表示,这样,三维模型的每一个 SIFT 特征均被编码成码本中的一个码字,即:三维

模型的 SIFT 特征被聚成了 k 个类。然后,统计出每个类中 SIFT 特征的个数,生成一个含有 k 个类的直方图,并将该直方图表示成一个 k 维的向量作为三维模型的特征向量。于是,一个三维模型便由一个 k 维的向量表示出来。

(5) 三维模型相似度计算

采用 KL 散度进行模型的相似度计算。KL 散度是描述 2 个概率分布差异关系的一种测度,当两个概率分布相同时,其 KL 散度为 0;当 2 个概率分布的差别增加时,其 KL 散度也增加。文中所使用的是 KL 散度用于 2 个向量的相似性度量的一种变形公式^[7]

$$D(U, V) = \sum_{i=1}^n (v_i - u_i) \ln \frac{v_i}{u_i} \quad (3)$$

其中, $U = (u_i)$, $V = (v_i)$ 是 2 个三维模型的特征向量; n 是向量的维度。

2 三维模型检索实验

2.1 实验环境

硬件中的具体配置为 CPU: Intel(R) Core(TM) 2 Duo E7300 2.66 GHz、2 G 内存。开发平台为 Microsoft Windows XP Professional。所用开发工具是 Visual C++ 6.0。模型库为 McGill 中抽取 427 个模型,分为诸如“蚂蚁”、“人”、“钳子”等 17 个类,该模型库中的模型均为有关节模型;PSB 内所有 1 814 个模型,包括一个训练集和一个测试集,训练集中含有 90 个类共 907 个模型,测试集中含有 92 个类共 907 个模型,该模型库中的模型均为刚性模型;ESB 中共有 866 个模型,被分为 45 个类。

2.2 码本建立

码本维数 k 的取值在 1 000~1 500 之间时,可获得较好的模型检索效果^[7],所以,文中设定码本维数 $k = 1 000$ 。分别利用 2 种方式来对码本的建立进行实验。其中,按模型特征数比例分次建立码本对于每个模型类都要执行一次 k-means 程序,然后将得到的各个聚类中心整合到一起成为模型库的码本。表 1 所示为 McGill 模型库建立码本时执行 k-means 程序的聚类中心数 k_i 。

对于建立模型库整体码本,聚类中心数即为码

表1 McGill模型库SIFT特征数及按模型特征数比例分次建立码本的 k_c 值

模型类别	SIFT特征数(N_i)	聚类中心数(k_c)
ant	42 879	102
crab	44 447	106
hand	14 781	35
human	28 987	69
cuttle	37 785	90
pliers	12 250	29
snake	9 888	23
glasses	17 918	43
bug	51 812	123
bear	19 184	46
plane	24 225	58
bird	17 232	41
chair	29 494	70
cup	9 939	24
dinosaur	13 789	33
fish	28 439	68
table	17 749	42

本维数,执行一次k-means程序就可得到模型库的码本。

利用这2种方式分别为McGill库、PSB训练集和ESB库建立码本,所耗费的时间如表2所示。从中可以看出,2种方式建立码本的时间消耗相差悬殊:对于McGill,建立模型库整体码本所耗费的时间是按模型特征数比例分次建立码本所耗费的时间的近15倍;对于PSB,二者比值约为61;对于ESB,二者比值约为46。

表2 2种方式建立码本所耗费的时间

模型库	按模型特征数比例分次建立码本 /h	建立模型库整体码本 /h
McGill	0.72	10.75
PSB	0.50	30.66
ESB	0.34	15.63

为了分析2种码本建立方式对三维模型检索效果的影响,分别对相应模型库进行模型检索测试,所得查准率-查全率曲线如图3所示。从该曲线可以看出,2种建立码本方式的检索性能相近。由此可以得出结论:按模型特征数比例分次建立码本与建立模型库整体码本相比,检索效果相近,时间成本大幅降低,即:按模型特征数比例分次建立码本更加高效。

2.3 三维模型检索算法的性能比较实验

实验选用D2(D2 shape distribution)算法^[8]、MODFD(multiple orientations depth fourier descrip-

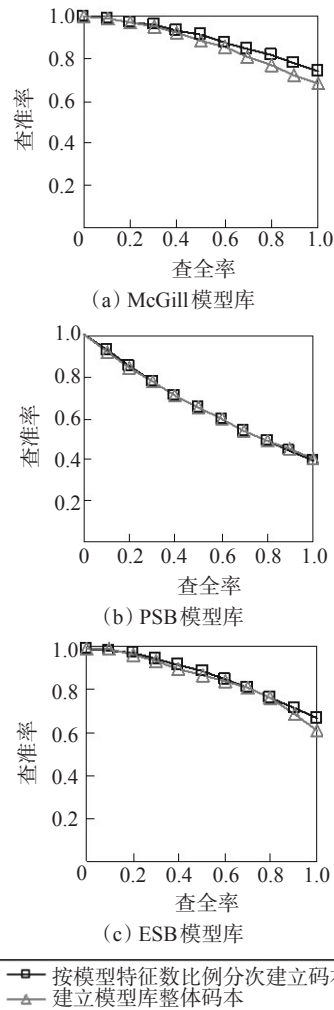


图3 2种方法建立码本进行模型检索实验所得查准率-查全率曲线

tor)算法^[9]和AAD(absolute angle distance histogram)算法^[10]与文中的多视图SIFT特征算法进行检索性能比较。D2算法是普林斯顿大学Robert Osada提出的一种统计三维模型表面随机点几何关系的形分布方法的形状函数,统计的是2个随机点的欧式距离。MODFD算法将三维模型在42个视点方向上做正交投影,计算每个正交投影深度图的通用傅里叶描述子并构建三维模型的特征描述符。AAD算法是一种基于外形的三维模型特征提取算法,它首先将含有表面信息的三维模型转换为一个有向点集模型,计算所有的点对之间的距离并用这些距离计算出二维直方图。分别利用这3种算法对McGill、PSB和ESB进行模型检索实验,得到其查准率-查全率曲线。利用按模型特征数比例分次建立码本方式为3个模型库建立码本并用文中算法对各模型库进行模型检索实验,将得到的查准

率-查全率曲线

与3种算法的查准率-查全率曲线进行比较,结果如图4所示。

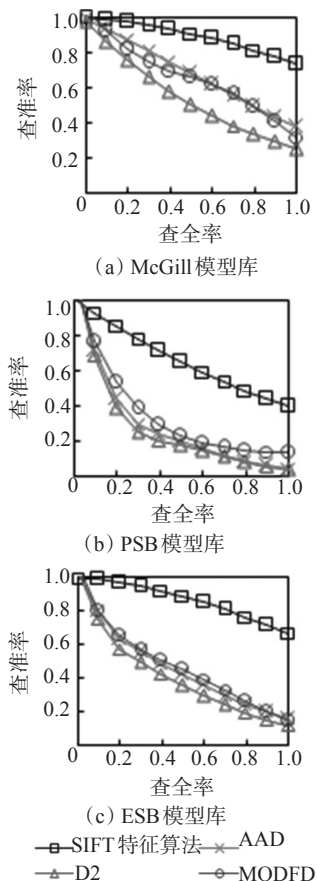


图4 模型检索测试所得查准率-查全率曲线

从上述结果可以看出,对于McGill中的有关节模型,文中算法的查准率-查全率为89%,D2、MODFD、AAD分别为55%、67%、68%;对于PSB中的刚性模型,该值分别为:67%、30%、37%、32%;对于ESB来说,该值分别为:86%、42%、49%和47%。这说明,文中算法对于有关节模型、刚性模型和机械模型均适用。

根据文中的算法开发了一个三维模型检索系统,该系统首先提取出用户输入的检索模型的特征向量,然后将其与模型库中各模型的特征向量进行相似度计算,并返回最相似的20个模型。图5所示为使用该系统进行三维模型检索的一个实例,左上角为用户输入的检索模型“钳子”,返回的相似模型按照KL散度值从小到大依次排列在“NO.1”“NO.2”……“NO.20”处。



图5 基于多视图的SIFT特征检索算法检索三维模型示例

3 总结与展望

研究了一种基于多视图SIFT特征的三维模型检索算法。该算法将复杂的三维模型特征提取问题通过多视图转换为较简单的二维图像处理问题,从而可以利用二维图像领域的SIFT算法对图像进行局部特征的提取,将这些局部特征量化聚类生成直方图并表示成一个特征向量来对三维模型进行描述。这样不仅保证了描述的完整性而且实现了计算的低复杂度。文中对码本的建立进行了实验和对比研究,结果表明,按模型特征数比例分次建立码本与建立模型库整体码本相比,在最终检索性能相近的基础上,码本的建立速度快了近几十倍。最后,分别利用文中算法及D2算法、MODFD算法和AAD算法对McGill、PSB和ESB进行了模型检索实验,结果表明,文中算法对于有关节模型、刚性模型和机械模型均适用。

今后的研究可以考虑将基于内容的三维模型检索算法与基于文本的三维模型检索算法结合起来,并且可以引进仿生学等其他学科方法改进三维模型的特征提取算法,进一步提高三维模型检索的成功率等。

参考文献

- [1] 普建涛,刘一,辛谷雨,等.一种基于二维多边形集相似性的三维模型检索方法[J].中国图像图形学报,2004,9(12):1437-1442.
- [2] Mohamed Chaouch, Anne Verroust-Blondet. A new descriptor for 2D depth image indexing and 3D model re-

(下转第80页)

法。但以上方法只是从颜色分布的相似程度来分析伪装效果,伪装效果的客观评价需要综合多个图像特征,才能更加准确地给出符合人眼视觉特性的评价结果。

参考文献

- [1] 汪东,吕绪良,许卫东,等.基于灰度直方图分析技术的伪装应用模型[J].解放据理工大学学报(自然科学版),2004,5(3):74-77.
- [2] Smith A R. Color gamut transformation pairs[J]. Comput-

(上接第60页)

- rieval[C]// Proc. ICIP(6),2007:373-376.
- [3] Thibault Napoleon, Tomasz Adamek, Francis Schmitt, et al. SHREC'08 entry: Multi-view 3D retrieval using multi-scale contour representation[C]//2008 IEEE International Conference on Shape Modeling and Applications,2008.
- [4] David G Lowe. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [5] 王瑞瑞,马建文,陈雪.多传感器影像配准中基于虚拟匹配窗口的SIFT算法[J].武汉大学学报·信息科学版,2011,36(2):163-166.
- [6] 钟金琴,檀结庆,李莹莹等.基于二阶矩的SIFT特征匹配算法[J].计算机应用,2011,31(1):29-32.
- [7] Ryutarou Ohbuchi, Kunio Osada, Takahiko Furuya, et

(上接第63页)

法的优点,在于它结合了接收方对压缩的主观要求,实现了感兴趣区动态编码。此外文中算法与标准JPEG2000相比,需要分别对ROI和背景确定截止门限,由于确定截止门限采用二分法搜索方法,计算量不会成倍增加,只是略有增加,满足了系统实时性的需求。

参考文献

- [1] Zhou Wang, Bovik Alan C. Bitplane-by-Bitplane Shift (BbBShift)-A Suggestion for JPEG 2000 Region of Interest Image Coding [J]. IEEE Signal Processing Lett (S1070-9908),2002,9(5):160-162.
- [2] Zhou Wang, Serene Banerjee, Brian L, et al. Generalized Bitplane-by-Bitplane Shift Method for JPEG2000 ROI Coding[C]//Proceedings of 2002 International Conference on Image Processing. Rochester, NY, USA, III81-III84,2002:22-25.
- [3] Taubman D, Marcellin M W. JPEG2000: Image Com-

er Graphics,1978,12(3):12-19.

- [3] B S Manjunath, J R Ohm, V V Vasudenva, et al. Color and Texture Descriptor s[J]. IEEE Trans. on CSVT, 2001,11(6):703-715.
- [4] 王涛,胡事民,孙家广.基于颜色-空间特征的图像检索[J].软件学报,2002,13(10):2031-2036.
- [5] 沈云涛.基于视觉感知特性的图像检索研究[D].西安:西北工业大学,2005:52-63.
- [6] 徐英.基于背景代表色提取的迷彩伪装颜色选取算法[J].光电工程,2007,34(1):100-104.
- [7] 许卫东,吕绪良,陈兵.一种基于纹理分析的伪装器材效果评价模型[J].兵工学报,2002,23(3):329-331.

al. Salient Local Visual Features for Shape-Based 3D Model Retrieval[C]// Proc. IEEE Shape Modeling International(SMI). NY, USA, Stony:BrookJune,2008:4-6.

- [8] Robert Osada, Thomas Funkhouser, Bernard Chazelle, et al. Shape Distributions[J]. ACM Transactions on Graphics,2002, 21(4):807-832.
- [9] Ryutarou Ohbuchi, Masatoshi Nakazawa, Tsuyoshi Takei. Retrieving 3D Shapes Based on Their Appearance [C]//5th ACM SIGMM international workshop on multimedia information retrieval. Berkeley, California, USA, 2003:39-45.
- [10] Ryutarou Ohbuchi, Takahiro Minamitani, Tsuyoshi Takei. Shape-Similarity Search of 3D Models by using Enhanced Shape Functions[J]. International Journal of Computer Applications in Technology (IJCAT), 2003, 23(3/4/5):70-85.

pression Fundamentals, Standards, and Practice[M]. Boston:Kluwer Academic Publishers,2002.

- [4] Nguyen A, Chandran V, Sridharan S. Important Prioritization Coding in JPEG2000 for Interpretability with Applications to Surveillance Image[C]//Proc. of SPIE Visual Communication and Image Processing, 2003, 5150: 806-817.
- [5] Sanchez V, Basu A, Mandal M K. Prioritized Region of Interest Coding in JPEG2000[J]. IEEE Trans. on Circuits and Systems for Video Tech., 2004, 14(9):1149-1155.
- [6] Rosenbaum R, Schumann H. Flexible, Dynamic and Compliant Region of Interest Coding in JPEG2000[C]// Proc. of IEEE Int. Conf. on Image Processing. New York:Rochester,2002:101-104.
- [7] Tammam Tillo, Gabriella Olmo. A Novel Multiple description coding scheme compatiblewith the JPEG2000 decoder[J]. IEEE signal processing lett, 2004, 11(11) 908-911.
- [8] 侯俊,方向忠,李继良.多码率JPEG2000的感兴趣区域编码[J].计算机工程,2007,33(5):200-201.