

球面深度全景图表示下的三维形状识别

冯元力¹⁾, 夏梦¹⁾, 季鹏磊¹⁾, 周潇¹⁾, 曾鸣²⁾, 刘新国^{1)*}

¹⁾ (浙江大学 CAD&CG 国家重点实验室 杭州 310058)

²⁾ (厦门大学软件学院 厦门 361005)

(xgliu@cad.zju.edu.cn)

摘要: 三维形状识别是近年来较为热门的研究方向, 针对其中的三维模型形状的表达方法和识别问题, 提出一种多分支卷积神经网络下的三维模型识别方法. 该方法通过对三维模型进行球面深度投影得到球面全景图; 为了提高识别精度, 将每个模型的球面全景图从多个角度展开, 创建多幅平面图像作为识别系统的输入; 识别系统使用多分支的卷积神经网络, 并将多幅全景图进行整合分析, 最终得到一个三维模型的识别结果. 对三维模型进行分类和检索的实验结果表明, 文中方法的识别效果优于近年来的前沿方法, 对三维模型进行检索的准确度甚至超过了多视图识别方法.

关键词: 卷积神经网络; 形状识别; 球面投影; 全景图

中图法分类号: TP391.41

Deep Spherical Panoramic Representation for 3D Shape Recognition

Feng Yuanli¹⁾, Xia Meng¹⁾, Ji Penglei¹⁾, Zhou Xiao¹⁾, Zeng Ming²⁾, and Liu Xinguo^{1)*}

¹⁾ (State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058)

²⁾ (Software School of Xiamen University, Xiamen 361005)

Abstract: 3D shape recognition is a hot topic in recent years. This paper proposed a 3D model recognition method with multi-branch convolutional neural network (CNN) to address the problems of 3D shape representation and recognition. The inputs of the proposed method are spherical panoramas by deep spherical projection of 3D models; to improve recognition accuracy, the spherical panorama of the shape first unfolded on various orientations to produce multiple rectified images as input of recognition frame; the recognition system consists of a multi-branch CNN, which analyzes the panoramas as a whole to produce the final recognition result. The experiment results of retrieval and classification on various of 3D dataset showed that the performance of our method is better than the state-of-the-art methods, and the retrieval accuracy outperforms that of multi-view method.

Key words: convolutional neural network; shape recognition; spherical projection; panorama

三维模型识别是一个重要的研究课题, 在医学影像学、CAD 及虚拟现实等领域都有着广泛的应用. 三维模型识别方法可分为 2 类: 一类是基于

三维数据的形状识别, 另一类是基于二维数据的视图识别. 基于三维数据的方法包括 3DShapeNet 方法^[1]及球面调和函数^[2]等; 基于二维数据的方法

收稿日期: 2016-07-10; 修回日期: 2017-07-07. 基金项目: 国家自然科学基金(61379068). 冯元力(1991—), 男, 博士研究生, 主要研究方向为三维模型处理与识别; 夏梦(1993—), 女, 硕士研究生, 主要研究方向为计算机图形学、模式识别; 季鹏磊(1994—), 男, 博士研究生, 主要研究方向为三维模型识别和实时重建; 周潇(1992—), 男, 硕士研究生, 主要研究方向为三维重建; 曾鸣(1985—), 男, 博士, 助理教授, CCF 会员, 主要研究方向为三维建模、几何分析; 刘新国(1972—), 男, 博士, 教授, 博士生导师, 论文通讯作者, 主要研究方向为计算机图形学、计算机动画、几何造型、三维重建、虚拟现实.

包括基于草图的深度检索方法^[3]、基于柱面全景图的 DeepPano 方法^[4]及基于多视图的深度学习方方法^[5]等。

传统的基于视图的形状识别方法是一种较为有效的方法,其根据人眼识别物体的原理,将模型的投影视图传入到识别系统中进行特征计算。但这类方法具有局限性:1) 视图本身是对模型外表面的描述,无法表示模型内部的结构,因而会造成数据丢失;2) 基于视图的形状识别过程是根据视图中的局部特征及拓扑关系进行的,事实上,单凭模型的一个线描图即可得到足够的特征数据,相比之下视图图像中的大部分像素对识别过程几乎没有贡献,这些冗余数据大大降低了识别系统的运算效率。

本文提出一种基于球面全景图的深度识别方法,其是一种基于二维视图描述符的方法。本文将一个三维模型投影到球面上,再将球面展开到平面上得到模型的全景图。为了提高识别精度,本文从多个不同的角度来展开球面全景图,从而对每个模型生成多幅全景图,然后使用全景图样本对基于多视图的卷积神经网络^[5]进行训练,最终得到一个高精度的多视图描述符。由于全景图中的每个像素都保留了重要的深度信息,因而弥补了传统方法的数据冗余问题。另外,本文使用一种多通道深度表达方式,在一个像素 RGB 值的 3 条通道中分别存储不同的数据(投影线与模型表面交点集的最远点距离、最近点距离及最远点处投影线与模型表面法向夹角),从而保存模型的部分内部信息,弥补了传统方法的数据丢失问题。

1 相关工作

三维模型识别过程一般可分为 2 步:1) 使用一种模型表示方法,将三维模型转换成识别系统可理解的数据表示;2) 使用特定的识别架构对表示数据进行识别。

1.1 模型表示方法

为了对模型进行特征提取,需要将模型表示为适当的形式。三维模型的表示方法主要分为基于三维形状的表示方法和基于二维视图的表示方法。

基于三维形状的表示方法的主要思想是基于模型的表面或者包围空间的几何性质,使用特定的数据格式记录模型信息。已有的表示方法包括根据模型表面法向和曲率采用直方图或者特征集表示的方法^[6-7]、对模型表面进行密集采样并计算形状直径的方法^[8],以及以形状直径函数进行稀疏

表示的方法^[9],或是像文献^[10-11]一样用多边形网格的热核信号表示网格模型。还有 Saupe 等^[12]提出的基于球面坐标的表示方法以及三维体素表示等方法^[1-13]。

基于二维视图的表示方法虽然丢失了部分三维信息,但也有诸多优点:1) 由于其数据维度较低,因而在相同的数据量下二维数据较三维数据能够存储更多的信息,从而保证了较高的分析识别效率;2) 二维视图能够较好地表示模型的任何几何性质,如孔洞或者噪面等;3) 视图表示方法在其他领域也有较好的应用,对三维模型的视图表示可以用来与其他图像或者草图作比较。

二维表示方法中较经典的有 Murase 等^[14]提出的方法,该方法在多个相机姿态下使用多种不同的光照对三维模型进行绘制,从而对每个模型渲染出多个视图数据。LightField 方法^[15]在计算机图形学中较为普及,其从多个不同视角绘制了模型的轮廓图。另外,Macrini 等^[16]将模型的轮廓拆分成多个部分并以有向无环图表示,Cyr 等^[17]使用曲率匹配和视图相似群组来定义形状相似度量,Eitz 等^[18]使用局部 Gabor 滤波器从多个角度生成三维模型线条画,还有 Schneider 等^[19]使用基于 SIFT 特征空间^[20]的 Fisher 向量^[21]来表示模型的人工手绘图等。

1.2 基于深度学习的形状识别

近年来,随着深度学习技术的兴起,越来越多的三维模型识别方法开始使用深度神经网络。Wang 等^[3]提出一种基于草图的模型检索方法,其通过草图对类内每个模型寻找一个最接近的视图,再使用视图对卷积神经网络(convolutional neural network, CNN)进行训练。Wu 等^[1]将三维体素数据输入到三维卷积滤波器中进行训练。Shi 等^[4]对模型的柱面全景图进行训练。Su 等^[5]从多个角度对模型生成多幅视图,将视图投入到一个并行 CNN 中进行训练。网络中设定了一个 View-pooling 层以合并不同视图的卷积结果。

在前沿的深度识别方法中,多视图方法^[5]是一种很有效的方法,其利用人眼对于模型识别过程中的多视角特点,使用一种并行多分支的深度网络对多视角视图进行识别,从而达到了很好的效果。本文使用这种并行深度网络,并以多角度全景图为输入,从而达到更好的识别效果。

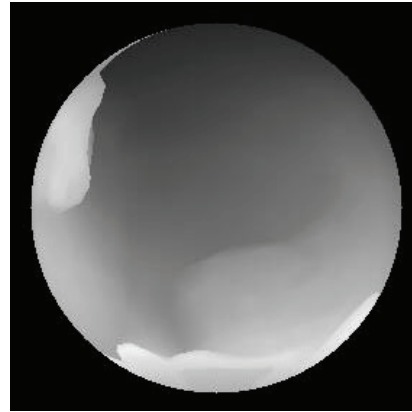
2 全景图绘制

DeepPano^[4]中使用柱面全景图作为 CNN 的输入。

相对于传统基于视图的方法,这种方法的输入数据更为稠密和连续,从而提高了识别效率.然而,基于柱面投影的全景图仍有诸多缺点:1)圆柱形投影面对模型主轴的选取非常敏感,限制了全景图识别方法的能力;2)圆柱形全景图没有表示出圆柱体顶面和底面的信息,有可能导致特征丢失.



a. bunny 模型和球面包围盒



b. bunny 的球面投影

图1 球面投影方法

在三维空间坐标系下,三角网格模型 M 表面上的每个点 (x, y, z) 都将转换成球坐标 (r, θ, φ) 的形式;其中, $\theta \in [0, \pi]$, $\varphi \in [0, 2\pi]$, $r \in [0, S_{\max}]$, S_{\max} 表示模型表面距中心点最远处的距离.后面的坐标表示将全部使用球坐标的形式.

给出一个单位球面 S , 其球心设为模型 M 的几何中心(这里通过将每个三角面片中心以面片面积为权重加权求平均以得到模型几何中心).将模型 M 从中心点向外投影到球面上,于是对于球面上任意一点 $s=(\theta_s, \varphi_s)$, 都可以定义一个投影射线 l , 该射线与模型 M 表面的交点集合定义为 $P_s=\{p_i\}$, $i=1,2,\dots,k$. 之后,利用投影线交点集合 P_s 的数据对单位球面 S 上每一点进行绘制,并将绘制好的球面 S 展开为平面图像.

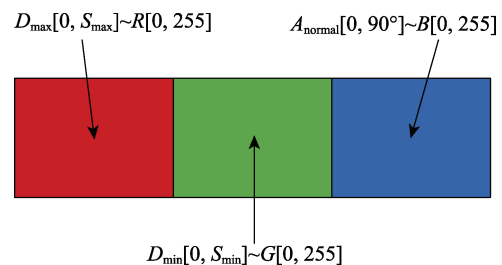
对于球面 S 上的每个点 s , 本文赋予一个 RGB 颜色值 $c_s=(r_s, g_s, b_s)$, 从而绘制出一幅球面图像.第2.1节将介绍 c_s 的计算方法.

2.1 球面投影绘制

DeepPano 方法^[4]将交点集合中距中心最远点的距离转换为灰度值储存在对应像素中,该方法容易丢失部分内部信息,且精度较低.本文对一个着色点 c_s 中的3条通道存储不同的信息,如图2所示,3条通道中分别存储最远交点距离、最近交点距离及最近交点处投影线与面片法向夹角(这里没有区分面片的正反方向).对于 R, G 通道,将最远

为了弥补柱面全景图的局限性,本文对投影方法做出了改进:使用球面深度投影方法绘制模型全景图.如图1所示,图1a所示为将 bunny 模型投影到球面上,投影点为模型中心点;图1b所示为球面上的颜色灰度值表示投影线与模型交点到投影中心的距离.

交点距离和最近交点从 $[0, S_{\max}]$ 线性映射到 $RGB[0, 255]$ 中;对于 B 通道,将法向夹角从 $[0, 90^\circ]$ 线性映射到 $[0, 255]$ 中.通过这种绘制方法,三维模型内部的部分信息得以保存,从而在某种程度上弥补了传统方法中无法表示模型内部的缺陷.

图2 每个像素的 R, G, B 3条通道中分别存储不同数据

2.2 球面平铺

对球面完成着色后,需要将球面图像展开平铺为平面图像,这样才能够将二维的图像数据传输入到 CNN 中进行训练.图3所示为一种球面展开方法,其中将 $\pi \times 2\pi$ 球面坐标 (θ, φ) 直接映射到 112×224 的像素空间中.

通过这种方法得到的全景图像中,其每个像素都表示重要的深度数据和夹角数据,因而弥补了传统方法中像素冗余的缺陷.这种映射方法的缺点是形状畸变较大,全景图中靠近两侧的部分被大幅度地拉长.为了降低畸变所导致的识别

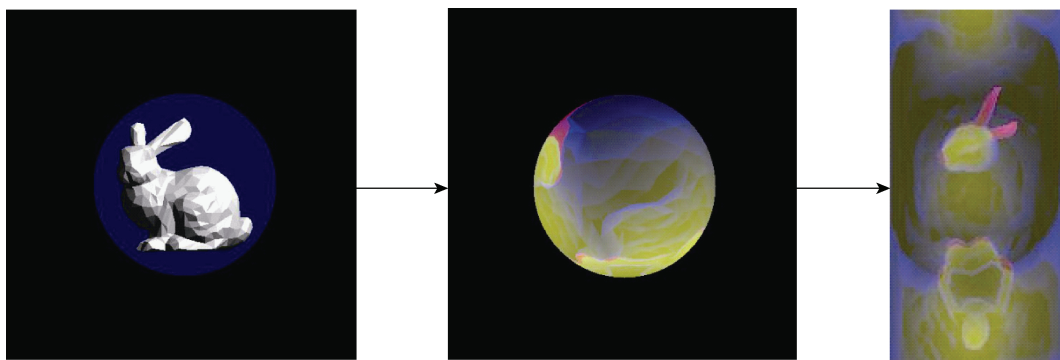


图3 三维模型投影到三维球面上再展开为平面全景图

误差, 本文从多个方向构建不同的球面坐标系, 从而生成多幅不同的全景图像. 多视图方法^[5]中绕主轴选取 12 个方向生成多个视图, 这种方向选取方法对于球面来说有很大的局限性, 因此本文方法中选取了张开的 8 个方向, 生成 8 个全景图, 如图 4 所示; 这 8 种展开方法对应的球面坐标系分别是将原坐标系原点方向 $(0^\circ, 0^\circ)$ 旋转到 $(45^\circ, 0^\circ)$, $(45^\circ, 90^\circ)$, $(45^\circ, 180^\circ)$, $(45^\circ, 270^\circ)$, $(135^\circ, 0^\circ)$, $(135^\circ, 90^\circ)$, $(135^\circ, 180^\circ)$ 和 $(135^\circ, 270^\circ)$ 这 8 个方向. 由于球面展开的接缝不同, 这些图像中畸变的部分各不相同, 从而实现了整体数据的畸变均匀化. 之后, 8 幅全景图共同输入到并行网络中进行整合训练, 从而得到高精度的识别网络.

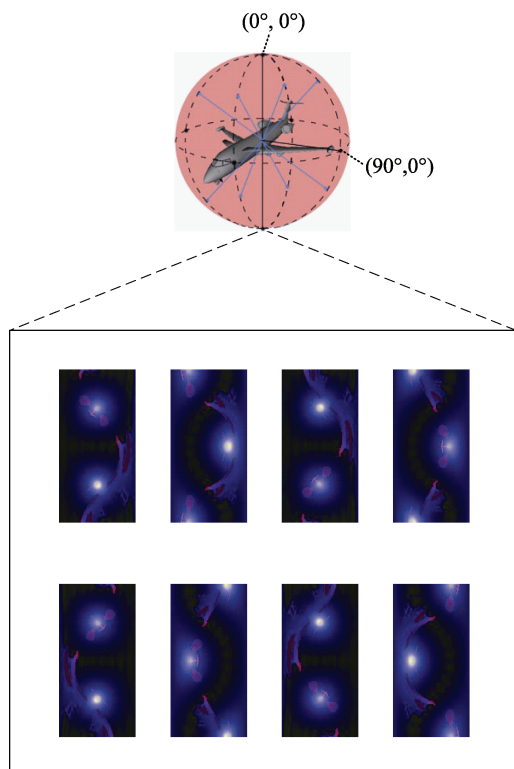


图4 将一个飞机模型经过投影再展开得到 8 幅不同的全景图

3 全景图识别网络

第 2.2 节中对每个模型生成了 8 幅全景图, 将这些全景图分别输入到 CNN 中可以得到 8 个分类结果; 之后将其合并, 即可得到整个模型的识别结果. 这样做的问题在于, 由于三维模型形状的不规则性, 这些全景图中可能存在变形较大的图片, 从而影响整体的识别效果. 因此, 本文提出一种并行的网络结构对这些全景图进行整合识别.

网络结构如图 5 所示, 对一个模型 M , 首先将每幅全景图分别传入到网络 1 的不同分支中, 得到 8 个全景图描述符 $d_1 \sim d_8$; 再将这些描述符传入

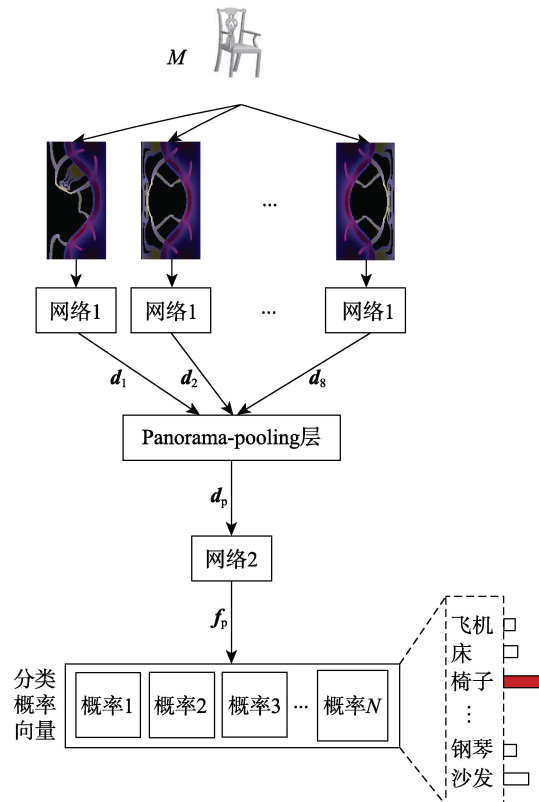


图5 多视图网络结构

Panorama-pooling 层,得到一个整合描述符 d_p ;最后将整合描述符传入网络 2 中,得到模型 M 的分类概率向量 f_p .

CNN 中使用 CNN 激活特征函数^[22]生成特征描述符向量,网络中包含 5 层卷积层和 3 层全连接层,本文将表示为($conv_1, conv_2, \dots, conv_5$ 共 5 层卷积层和 fc_6, fc_7, fc_8 共 3 层全连接层). 一个 Panorama-pooling 层将网络分割成了网络 1 和网络 2 两部分. 网络 1 复制出了 8 个分支,它们共享相同的参数,每个分支对其中一幅全景图进行特征提取;之后通过 Panorama-pooling 层整合成为模型的一个整体描述符.

Panorama-pooling 层通过对多个视图进行特征选拔,可以有效地降低图片变形对识别特征的影响;其可以设置为 max-pooling 层,也可以设置为 maxout 层^[23],并且可以安置在 CNN 中间的任意一层. 多视图方法^[5]中实验表明,将 Panorama-pooling 层安置在 fc_6 层之后或者 fc_7 层之后可以得到最好的识别效果. 本文将安置在 fc_6 层之后.

网络的输入是 8 个 112×224 的 RGB 数据,经过网络 1 和 Panorama-pooling 层后可以得到一个 4096 维的描述符向量,再经过网络 2 后最终可以得到一个分类概率向量,向量中的每个数值表示模型属于一个类的概率.

4 实验及结果分析

相对于基于柱面深度全景图的 DeepPano 方法^[4],基于球面深度全景图的识别方法对模型主轴不敏感,且可以表示各个方向的数据,不会造成 DeepPano 中的数据丢失. 但即便如此,球面全景图在某些方面仍具有局限性. 例如,将全景图从球面投影到平面的过程中,全景图上不同的位置会产生不同程度的畸变,这干扰了特征识别的效果. 为了弥补这个缺陷,本文借用 MVCNN^[5]的多视图思想,从多个角度对球面全景图进行展开和平面投影,从而得到多幅不同畸变的全景图. 通过对这些全景图进行整合识别,可以得到一个更高的识别精度.

本文使用 Princeton 大学的 ModelNet 数据库进行实验,该数据库中含有 127915 个三维模型,662 种形状分类. 本文中用的是 ModelNet 的子集: ModelNet40 和 ModelNet10. ModelNet40 中包含有 12311 个模型(测试模型 2468 个,训练模型 9843 个). ModelNet10 中包含有 4899 个模型(测试模型 908 个,

训练模型 3991 个).

4.1 模型检索

本文使用训练出的网络对模型进行检索. 多视图方法^[5]中使用多个描述符之间的最短距离,即对于有 n_x 个全景图描述符的模型 x 和有 n_y 个全景图描述符的模型 y ,它们之间的距离表示为

$$d(x, y) = \frac{1}{2} \frac{\sum_{j=1}^{n_y} \min_i \|x_i - y_j\|_2}{n_y} + \frac{1}{2} \frac{\sum_{i=1}^{n_x} \min_j \|x_i - y_j\|_2}{n_x}.$$

其中, $\|x_i - y_j\|_2$ 表示模型 x 的第 i 个全景图描述符向量与模型 y 的第 j 个全景图描述符向量之间的 L_2 距离.

本文使用这种距离度量进行检索实验,另外也尝试对每个测试模型使用整合的描述符直接计算欧氏距离并进行了检索实验. 将 2 种方法的平均检索精度(mean average precision, mAP)与其他方法进行比较,包括比较经典的 SPH 方法^[2]和 LFD 方法^[15],以及基于三维体素的 ShapeNets 方法^[1]和基于二维视图的 MVCNN 方法^[5],结果如表 1 所示. 可以看出,基于整合描述符的检索方法的 mAP 整体高于基于多描述符的方法,并且明显超过了其他前沿方法;基于多描述符的方法在 ModelNet40 数据库上的检索精度也超过了其他方法.

表 1 5 种方法的检索 mAP 比较 %

方法	mAP	
	ModelNet10	ModelNet40
SPH	44.05	33.26
LFD	49.82	40.91
3D ShapeNets	68.26	49.23
MVCNN	58.13	50.80
本文(多描述符)	65.51	54.29
本文(整合描述符)	76.67	58.64

4.2 模型分类

有 2 种对模型的分类方法:一种是基于网络输出的分类概率向量对模型分类;另一种是使用支持向量机(support vector machine, SVM)对全景图描述符进行分类.

在第 1 种方法中,对于训练集有 N 个形状分类的网络,输入一个模型的 8 个全景图数据,其输出会是一个 N 维的分类概率向量,本文取向量中概率最大的项所对应的类作为该模型估计所属类.

第 2 种方法按照文献[5]中的 SVM 分类方法, 首先使用训练集的描述符一对多 SVM 进行训练(一个模型的 8 个全景图描述符看作 8 个不同的样本); 之后对于测试模型, 将 8 个全景图描述符输入到 SVM 得到 8 个分类概率向量, 对这 8 个分类概率向量进行加和; 最后取最大项对应的类作为该模型估计所属类。

本文对这 2 种方法进行实验, 并且将分类结果与一些前沿方法做了比较, 结果如表 2 所示。可以看出, 除了 ModelNet10 数据集上基于 SVM 的分类方法效果略差于 DeepPano^[4]以外, 本文方法分类精度大体上超过了其他前沿方法。

表 2 5 种方法的分类精度比较 %

方法	分类精度	
	ModelNet10	ModelNet40
SPH	79.79	63.59
LFD	79.87	75.47
3D ShapeNets	83.54	77.32
MVCNN	88.66	77.63
本文(多描述符)	88.44	83.87
本文(整合描述符)	89.98	86.18

5 结 语

本文提出一个卷积神经网络下基于多视角全景图的模型识别方法, 较好地弥补了传统视图识别方法中特征丢失和数据冗余等缺陷, 并且其识别精度超过了前沿的方法, 得到了更好的检索和分类结果。另外, 由于本文方法是基于多空间角度的全景图方法, 其对模型的位置和角度变化不敏感, 具有姿态不变性。

未来这个方向仍有许多可供发展的空间, 如可以使用其他映射方法将球面图像映射为平面图像, 或者对不同的视角选取方法做进一步的实验等。另外, 除了柱面全景图和球面全景图以外, 还可以寻找更好的全景图绘制方法。

参考文献(References):

- [1] Wu Z R, Song S R, Khosla A, *et al.* 3D shapenets: a deep representation for volumetric shapes[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 1912-1920
- [2] Kazhdan M, Funkhouser T, Rusinkiewicz S. Rotation invariant spherical harmonic representation of 3D shape descriptors[C] //Proceedings of International Conference on Computer Graphics and Interactive Techniques. Aire-la-Ville: Eurographics Association Press, 2003: 156-164
- [3] Wang F, Kang L, Li Y. Sketch-based 3d shape retrieval using convolutional neural networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 1875-1883
- [4] Shi B G, Bai S, Zhou Z C, *et al.* DeepPano: deep panoramic representation for 3-D shape recognition[J]. IEEE Signal Processing Letters, 2015, 22(12): 2339-2343
- [5] Su H, Maji S, Kalogerakis E, *et al.* Multi-view convolutional neural networks for 3D shape recognition[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 945-953
- [6] Liu X G, Su R J, Kang S B, *et al.* Directional histogram model for three-dimensional shape similarity[C] //Proceedings of Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2003: 813-820
- [7] Horn B K P. Extended gaussian images[J]. Proceedings of the IEEE, 1984, 72(12): 1671-1686
- [8] Chaudhuri S, Koltun V. Data-driven suggestions for creativity support in 3D modeling[J]. ACM Transactions on Graphics, 2010, 29(6): Article No.183
- [9] Shu Zhenyu, Wang Pengfei, Yu Xin, *et al.* 3D model recognition via local sparse representation[J]. Journal of Computer-Aided Design & Computer Graphics, 2014, 26(11): 1938-1947 (in Chinese)
(舒振宇, 王鹏飞, 于 欣, 等. 基于局部稀疏表示的三维模型识别算法[J]. 计算机辅助设计与图形学学报, 2014, 26(11): 1938-1947)
- [10] Bronstein A M, Bronstein M M, Guibas L J, *et al.* Shape Google: geometric words and expressions for invariant shape retrieval[J]. ACM Transactions on Graphics, 2011, 30(1): Article No.1
- [11] Kokkinos I, Bronstein M M, Litman R, *et al.* Intrinsic shape context descriptors for deformable shapes[C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE Computer Society Press, 2012: 159-166
- [12] Saupe D, Vranić D V. 3D model retrieval with spherical harmonics and moments[M] //Lecture Notes in Computer Science. Heidelberg: Springer, 2001, 2191: 392-397
- [13] Knopp J, Prasad M, Willems G, *et al.* Hough transform and 3D SURF for robust three dimensional classification[M] //Lecture Notes in Computer Science. Heidelberg: Springer, 2010, 6316: 589-602
- [14] Murase H, Nayar S K. Visual learning and recognition of 3-D objects from appearance[J]. International Journal of Computer Vision, 1995, 14(1): 5-24
- [15] Chen D Y, Tian X P, Shen Y T, *et al.* On visual similarity based 3D model retrieval[J]. Computer Graphics Forum, 2003, 22(3): 223-232
- [16] Macrini D, Shokoufandeh A, Dickinson S, *et al.* View-based 3-D object recognition using shock graphs[C] //Proceedings of International Conference on Pattern Recognition. Los Alamitos:

- IEEE Computer Society Press, 2002, 3: 24-28
- [17] Cyr C M, Kimia B B. A similarity-based aspect-graph approach to 3D object recognition[J]. *International Journal of Computer Vision*, 2004, 57(1): 5-22
- [18] Eitz M, Richter R, Boubekeur T, *et al.* Sketch-based shape retrieval[J]. *ACM Transactions on Graphics*, 2012, 31(4): Article No.31
- [19] Schneider R G, Tuytelaars T. Sketch classification and classification-driven analysis using fisher vectors[J]. *ACM Transactions on Graphics*, 2014, 33(6): Article No.174
- [20] Lowe D G. Object recognition from local scale-invariant features[C] //Proceedings of the 7th IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 1999, 2: 1150-1157
- [21] Perronnin F, Sánchez J, Mensink T. Improving the fisher kernel for large-scale image classification[C] //Proceedings of the 11th European Conference on Computer Vision. Heidelberg: Springer, 2010: 143-156
- [22] Donahue J, Jia Y Q, Vinyals O, *et al.* Decaf: a deep convolutional activation feature for generic visual recognition[C] //Proceedings of the 31st International Conference on Machine Learning. New York: ACM Press, 2014, 32: I-647-I-655
- [23] Goodfellow I J, Warde-Farley D, Mirza M, *et al.* Maxout networks[C] //Proceedings of the 30th International Conference on Machine Learning. New York: ACM Press, 2013, 28: III-1319-III-1327