



# 基于 2D-3D 语义传递的室内三维点云模型语义分割

熊汉江<sup>1</sup> 郑先伟<sup>1</sup> 丁友丽<sup>1</sup> 张 艺<sup>2</sup> 吴秀杰<sup>2</sup> 周 妍<sup>1</sup>

1 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

2 武汉大学数学与统计学院,湖北 武汉,430079

**摘 要:**针对现有三维点云模型重建对象化和结构化信息缺失的问题,提出一种基于图模型的二维图像语义到三维点云语义传递的算法。该算法利用扩展全卷积神经网络提取 2D 图像的室内空间布局 and 对象语义,基于以 2D 图像超像素和 3D 点云为结点构建融合图像间一致性和图像内一致性的图模型,实现 2D 语义到 3D 语义的传递。基于点云分类实验的结果表明,该方法能够得到精度较高的室内三维点云语义分类结果,点云分类的精度可达到 73.875 2%,且分类效果较好。

**关键词:**语义三维点云模型;语义传递;语义标记;点云分类

中图分类号:P208

文献标志码:A

近年来,随着室外三维点云模型的大规模应用<sup>[1-2]</sup>,室内三维点云模型的研究成为了众多学者研究的热点。然而,由于室内空间的封闭性和复杂性<sup>[3]</sup>,且构建 3D 特征描述算子困难,因此室内三维点云模型的语义提取和对象分类也非常困难,从而造成大多数几何精确而语义未知的“盲点云”,难以与实际应用对接。相比于构建几何结构完整的点云模型<sup>[4-6]</sup>,语义三维点云模型是在获得点云模型的基础上,根据对象信息将无意义的“盲点云”逐点标记为具有实际意义的语义对象。室内语义三维模型不仅需要表达对象信息,还需表达结构信息。这种描述点云对象及其空间结构关系的语义三维模型,在语义感知的虚拟现实交互、3D 对象检索和提取等方面有着重要的价值,因此成为了众多学者研究的热点。

目前,语义三维点云建模研究的普遍做法是直接借鉴二维图像语义分类的研究思路,即基于 3D 几何特征进行逐点的特征分类器训练<sup>[7-11]</sup>,并通过空间上下文约束进行优化,得到语义分类结果<sup>[12-13]</sup>。随着深度学习在二维图像分类上的发展,该方法逐渐扩展到三维点云分类研究中,成为

三维点云语义标注的重要算法<sup>[14]</sup>。Wu 等<sup>[15]</sup>构建 3D ShapeNet 研究通过构造 3D 卷积网络,从对象体素中学习形状描述算子。然而,由于缺乏类似于 2D SIFT(scale-invariant feature transform)特征算子这样具有旋转、平移、缩放、光照不变性的 3D 特征算子,使得构造有效的 3D 特征算子成为研究的难点。此外,缺乏足够标记点云训练数据也成为影响分类精度的另一大关键问题。为此,Koppula 等<sup>[7]</sup>提出利用多维特征和上下文关系构建图模型进行语义分类的方法,结合图像局部纹理特征、形状先验特征、对象关系特征提高 3D 特征算子的鲁棒性;袁理等<sup>[16]</sup>提出了一种针对待建图像动态选择 2D-3D 特征变换子空间的策略,能显著提高重建精度;针对室内空间由于对象遮挡、重叠而产生的复杂场景,Munoz 等<sup>[12]</sup>提出一种基于搜索的分类方法,该方法递归地进行分割和分类,不断添加同一类别的邻域块,实现对象拟合。这些研究表明,使用联合视觉特征和 3D 形状特征可以提高在遮挡的现实环境中语义对象的提取,然而这些方法依然无法摆脱对 3D 训练数据的依赖。

在二维图像语义分类研究中,构建大规模 2D

收稿日期:2018-05-17

项目资助:国家重点研发计划(2018YFB0505401);国家自然科学基金(41871361,41701445);测绘遥感信息工程国家重点实验室自主科研项目。

第一作者:熊汉江,博士,教授,主要从事地理空间数据可视化、语义三维模型重建等研究。xionghanjiang@whu.edu.cn

通讯作者:郑先伟,副教授。zhengxw@whu.edu.cn

标记训练数据集已经有了数十年的研究历史<sup>[17-19]</sup>,形成了大量样本丰富、分类完整的标准数据集,如 LabelMe<sup>[17]</sup>、ImageNet<sup>[18]</sup>、ImageNet-segment<sup>[19]</sup>等。Boulch 等<sup>[14]</sup>利用卷积神经网络图像分类方法对三维点云模型实现逐点云的语义类别标注<sup>[14]</sup>。相比于直接对 3D 点云数据进行分割和语义标注,该方法避免了构建 3D 特征算子和训练数据集,实现了三维点云的语义对象获取。然而,该方法没有考虑图像之间的空间上下文一致性。

基于上述问题,本文提出了一种基于图模型的二维图像语义到三维点云语义的传递算法,通过将实景图像作为语义分类的信息源,然后基于 FCN(full convolutional network)算法对室内图像对象语义分类和空间布局提取,探索融合图像间一致性和图像内一致性的图模型,以此实现从 2D 图像到 3D 点云的语义传递和分割。实验表明,本文方法能得到精度较高的室内三维点云的精细结构化语义分割模型,并且能够有效地规避缺乏 3D 特征描述算子和训练数据集的问题。

## 1 基于 FCN 的室内场景语义分类和空间布局提取

全卷积神经网络(FCN)精细场景语义分类算法从 2015 年一经提出就成为对象分类的主流算法<sup>[20]</sup>。随着室内服务和应用需求的提高,基于 FCN 的分类方法也逐渐扩展到室内对象语义的分类中。然而,单纯的室内对象语义不足以描述室内复杂的空间结构信息,而室内空间普遍呈现为横平竖直的规整结构,因此,室内空间布局作为描述其结构特征的重要手段。当前不少学者将 FCN 算法应用于室内空间布局的估计中,通过同时进行语义标签和空间布局的训练,相互辅助,得到优化的布局估计和语义分类结果。本文采用文献<sup>[21]</sup>提出的基于 FCN-VGG 16 的室内场景估计和语义分类方法,将两个任务同时融入 FCN 进行训练,得出场景的对象语义分类和空间布局估计结果,其流程如图 1 所示。

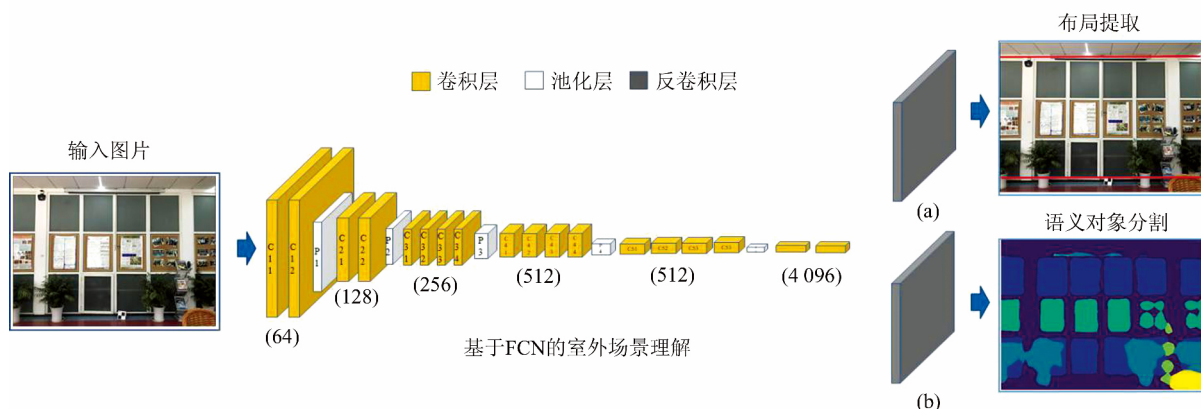


图 1 基于 FCN-VGG 16 的室内场景理解与布局估计

Fig.1 Indoor Scene Understanding and Layout Estimation Based on FCN-VGG 16

该网络的第一部分输出结果为室内对象语义分类。为了减少训练的难度,本文使用已经在室内图像数据集上预训练好的网络作为初始网络。该网络为基于 Caffe 的 FCN-VGG 16 框架<sup>[22]</sup>,在 NYUDv2 RGBD 室内数据集上进行训练,数据集包含室内场景中常见的对象类<sup>[23]</sup>。原始网络中<sup>[20]</sup>包含两个输入项:RGB 图像和深度图。在本文研究中,只选择其中的 RGB 图像作为输入项。为了保证分类的精度,本文额外采集了 84 张室内场景的实景图像进行标记并输入 FCN 网络。当对网络进行微调后,得到高精度的室内对象语义分类结果。在采集的图像中,考虑到待重建室内环境中的对象语义,在训练数据的标记中将室内对象划分为墙壁、地板、天花板、门、窗户、板子、

画板、电视、空调、椅子、桌子、装饰画 12 个类别。

该网络的第二部分输出结果为室内空间布局估计。布局估计的目的是识别室内规则的三维结构信息,可以理解为将室内对象划分为墙面、地板和天花板 3 个类别,并提取边界。由于室内环境存在大量遮挡,单纯的基于线特征的方法不能完整地提取出空间布局的结构线。而 FCN 在语义分类和边界提取方面的高效性能使它在室内遮挡严重的环境依然可以显著识别出空间布局的结构线。首先,通过联合训练布局结构线和表面语义标签,剔除遮挡对象,使得语义表面的边界和布局结构线经过训练后可互相匹配,得出在遮挡情况下的布局粗估计。然后通过提取直线和灭点用于补齐粗估计中不完整的结构线,得到一系列候选

的空间布局。最后对这些候选集进行排序,选择得分最高的作为布局估计结果。

## 2 基于图模型的 2D-3D 语义传递

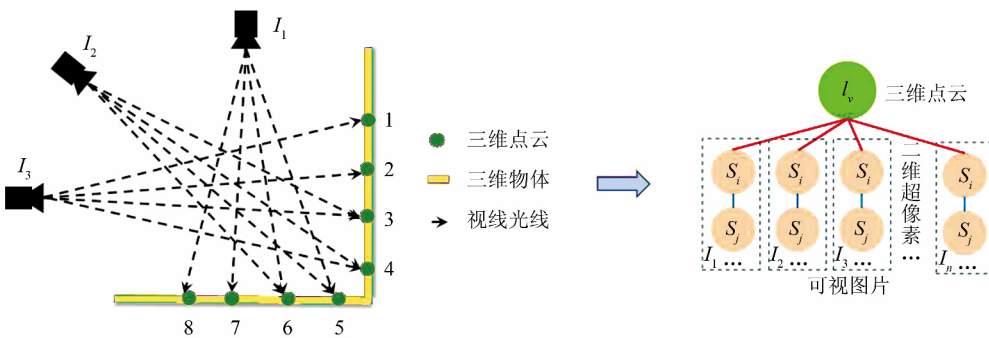
### 2.1 超像素标记池构建

假设 3D 点云模型集合定义为  $P = \{p_i\}$ , 每个点用三维坐标和 RGB 颜色进行描述  $\{x_i, y_i, z_i, R_i, G_i, B_i\}$ 。点云模型使用分层式运动恢复结构(structure from motion, SfM)和多视角密集匹配(patch-based multi-view stereo, PMVS) (SfM-PMVS) 方法从  $R$  张实景图像  $I_R = \{I_r\}_{r=1}^R$  重建得到。为了对 SfM-PMVS 构建的稠密点云进行语义标注, 本文使用 SLIC (simple line interface method) 方法<sup>[24]</sup>对图像进行超像素分割, 构建特征点的区域块, 并以超像素中特征点的类别表示该超像素的类别, 然后通过 SfM 的可视图模型将超像素反投影到三维空间, 实现对稠密点云的语义标注。语义标记池可以表示为  $S = \{S_i, l_{S_i}\}$ ,  $l_{S_i} \in L$  表示超像素的语义标记,  $i \in N$  表示超像素编号 ( $N$  为超像素个数)。给定输入图像, 结合语义分割和特征点, 得到特征点的语义标记值, 然后以特征点所在的超像素作为缓冲区, 将超像素内特征点的标记值作为该超像素的标记

值。除了可以实现稠密点云的语义标记, 以超像素为单位进行操作还可以为 3D 点对应的 2D 特征点建立缓冲区, 避免由于投影误差引起的偏移问题。2D-3D 语义传递的目标是依据超像素池中的像素标记, 给点云中的每一个点  $p_i$  赋予对应语义标记  $l_i$ 。

### 2.2 图模型构建

为了实现 2D 超像素语义到 3D 点云语义的传递, 本文把 3D 点云  $P = \{p_i\}$  和超像素  $S = \{S_i\}$  作为结点  $V = \{p_i\} \cup \{S_i\}$ , 连接线作为边  $\epsilon$ , 构建基于 MRF (Markov random field) 的图模型  $G = \{v, \epsilon\}$ 。本文通过 SfM 点云重建过程中编码的可视图模型构建 3D 点云和超像素的连接关系, 进而为 3D 点云和其对应的 2D 超像素之间建立语义传递通道。可视图模型连接相机和点云, 能够表示三维点与二维图像的对应关系, 从图 2(a) 可以看出 3 号点云的可视图图像为图像 2 和图像 3。3D 点云通过可视图相机参数  $\{M_i\}$  投影到对应的图像空间, 并将投影点所在的超像素空间和该 3D 点连接, 2D 超像素结点之间的连接通过计算与其相邻的超像素实现 (如图 2(b) 所示), 其中, 红色线段表示 2D 超像素和 3D 点的连接边, 蓝色线段表示 2D 超像素之间的连接边。



(a) 可视图模型中三维点与二维图像的对应关系

(b) 二维超像素和三维点云为结点构建的图模型中二维超像素之间及其与三维点云之间的连接关系

图 2 可视图到图模型的构建

Fig.2 Construction from Visibility Model to Graph Model

### 2.3 多类别标记能量函数构建

本文的目标是为三维点云中的每个点赋予相应的语义标签, 因此, 需要解决每个三维点与哪些二维超像素进行语义标签的传递的问题。对于 SfM 系统生成的影像点云, 其三维点与二维图像分割的超像素的对应关系可能存在以下两种情况: ①三维点由二维图像的 SIFT 特征点通过相机参数反投影到三维空间中得到, 通过超像素标记池获取 SIFT 特征点所在超像素的语义标签可将其传递给对应的三维点; ②三维点没有与之对应的 SIFT 特征点, 但是可以根据相机参数将三维点投影到二维图像上得到其对应的二维超像

素。因此, 可以利用图模型得到每个三维点和对应超像素之间的连接关系。然而, 在 2D 到 3D 对应结点间语义标记传递的过程中, 3D 点对应的多个 2D 超像素可能具有不同的标记值, 为此, 本文探索图像内一致性和图像间一致性进行语义标记预测, 抑制标记误差的影响, 提高 3D 点云语义标记的精度。以上问题可以表达为多类别标记问题, 以结点  $v \in V$  的标记  $l(v)$  作为变量, 构建势能函数:

$$E = \sum_{v \in V} \sum_{l(\cdot) \in L} \phi_d(l(v)) + \sum_{(v_1, v_2) \in \epsilon} \sum_{l(\cdot) \in L} \phi_s(l(v_1), l(v_2)) \quad (1)$$

式中,  $L$  是标记集合;  $V$  为结点; 势能函数包括数



据项  $\psi_d$  和平滑项  $\psi_s$ , 数据项用于约束语义的正确性, 平滑项保证邻域结点具有相似的语义, 通过图像间一致性和图像内一致性实现。

在正确传递的情况下, 2D 超像素类别标记与语义传递后的 3D 点云的类别标记结果是一致的, 若超像素的标签分布函数为  $p_{S_i}(\cdot)$ , 则数据项可以定义为:

$$\psi_d(l(S_i)) = p_{S_i}(l(S_i)) \quad (2)$$

相邻超像素结点之间的平滑项通过图像内一致性构建, 约束邻域超像素之间具有相似的标记, 表示为:

$$\psi_{s,2D}(l(S_i), l(S_j)) = p(l(S_i), l(S_j)) \quad (3)$$

式中,  $p(l(S_i), l(S_j))$  为从超像素  $S_i$  的标签  $l(S_i)$  和相邻超像素  $S_j$  的标签  $l(S_j)$  在邻域  $\epsilon$  内与其他相邻超像素标签相似的概率。表示为:

$$\frac{p(l(S_i), l(S_j)) = \sum_{(S_i, S_k) \in \epsilon_1} N_{(l(S_i)=l(S_k))} + \sum_{(S_j, S_m) \in \epsilon_2} N_{(l(S_j)=l(S_m))}}{N_{S_i} + N_{S_j}} \quad (4)$$

其中,  $S_k$  表示超像素  $S_i$  所在邻域  $\epsilon_1$  内的其他超像素;  $S_m$  表示超像素  $S_j$  所在邻域  $\epsilon_2$  内的其他超像素;  $N_{S_*}$  表示超像素  $S_*$  出现的次数。

相连的 3D 点和 2D 超像素之间的平滑项通过图像间一致性构建, 表示为:

$$\psi_{s,3D}(l(S_i), l(p_j)) = \begin{cases} 1, l(S_i) = l(p_j) \\ c, l(S_i) \neq l(p_j) \end{cases} \quad (5)$$

式中,  $c > 1$  为一个常数。因此, 能量函数可以定义如下:

$$E = \sum_{S_q \in I_R} \sum_{l(\cdot) \in L} \psi_d(l(S_q)) + \lambda_1 \cdot \sum_{(S_i, S_j) \in \epsilon} \sum_{l(\cdot) \in L} \log \psi_{s,2D}(l(S_i), l(S_j)) + \lambda_2 \cdot \sum_{(S_i, p_j) \in \epsilon} \sum_{l(\cdot) \in L} \log \psi_{s,3D}(l(S_i), l(p_j)) \quad (6)$$

其中,  $\lambda_1$  和  $\lambda_2$  为拉格朗日乘子, 表示不同势能的权重。

最后, 3D 点云  $\{p_i\}$  的语义标注  $L$  通过最小化势能函数得到, 本文使用图割算法求得最优解:

$$\psi(L) = \arg \min_{l \in L} \sum_{S_q \in I_R} \psi_d(l(S_q)) + \lambda_1 \cdot \sum_{(S_i, S_j) \in \epsilon} \psi_{s,2D}(l(S_i), l(S_j)) + \lambda_2 \cdot \sum_{(S_i, p_j) \in \epsilon} \psi_{s,3D}(l(S_i), l(p_j)) \quad (7)$$

### 3 实验与结果分析

#### 3.1 实验数据

为了减少训练的难度, 本文使用已经在室内

图像数据集上预训练好的网络作为初始网络。全卷积网络为基于 Caffe 的 FCN-VGG 16 框架<sup>[22]</sup>, 训练数据为 NYUDv2 RGBD 室内数据集的 RGB 图像, 包含常见的室内场景, 如卧室、餐厅、教室、办公室、客厅、洗手间等, 其中标记图像包括 40 个常见的对象类, 共 1 499 张标记图像和 407 024 张未标记图像<sup>[23]</sup>。为了保证分类的精度, 本文额外采集了 64 张室内场景的实景图像, 使用语义标记工具 LabelMe<sup>[17]</sup> 按照对应的类别将室内对象划分为墙壁、地板、天花板、门、窗户、板子、画板、电视、空调、椅子、桌子和装饰画这 12 个类别。

此外, 为了利用 FCN 进行室内三维布局的估计, 需要对训练图像进行布局结构线标记。忽略室内空间的所有遮挡, 提取并标记出墙、天花板、地板之间完整的三维结构线。即使在实际场景中存在遮挡, 训练标记中的所有像素都被认为是室内粗略三维布局, 其他像素都作为背景。

测试数据包括 261 张会议室图像和由此生成的三维点云数据, 每一个点由三维坐标 XYZ 和颜色 RGB 组成 (如图 3 所示)。本文对 261 张会



(a) 测试数据集中的会议室图像



(b) 测试数据集中与会议室图像对应的三维点云数据

图 3 会议室数据集

Fig.3 Dataset of Meeting Room

会议室图像利用 SLIC(simple linear iterative clustering)进行超像素分割,分割区域大小设置为 10,正则项系数设为 1,分割结果如图 4 所示。



(a) 测试数据集中会议室图像



(b) 超像素分割结果

图 4 超像素分割结果

Fig.4 Result of Superpixel Segmentation

### 3.2 FCN 分类实验结果和分析

本文使用 FCN-VGG 16 卷积神经网络进行室内场景估计和语义分类,将两个任务同时融入 FCN 网络进行训练,得出场景语义分类和布局估计结果。由图 5 和图 6 可见,FCN 可以得到精确的室内对象语义,且边界划分清晰,分类精度为 96%。在布局结果提取方面,对于存在遮挡的区域,该算法也能较为精确地提取室内空间布局,并以完整的结构线进行区分。因此,可以为后续的语义传递提供可靠的基础。

### 3.3 点云分类实验结果和分析

分类实验利用 C++ 语言基于多类别分类优化程序 gco-v3.0<sup>[18]</sup> 得到,优化方法采用图割  $\alpha$ -expansion 算法。分类实验利用平均分类精度,分类精度为 73.875 2%。分类后的结果如图 7 所示。可见,三维点云可以按照图像的语义分类得到正确的结果,且分类效果较好。

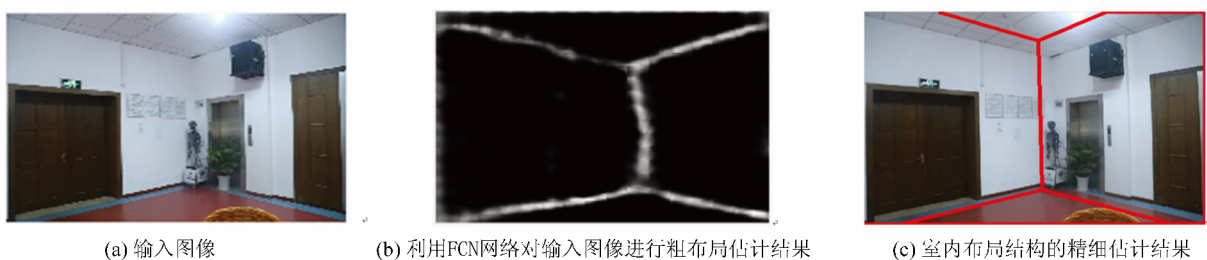
## 4 结 语

本文针对三维点云模型,提出了一种基于图



图 5 语义分类结果

Fig.5 Results of Semantic Classification



(a) 输入图像

(b) 利用FCN网络对输入图像进行粗布局估计结果

(c) 室内布局结构的精细估计结果

图 6 室内空间布局估计结果

Fig.6 Results of Layout Estimation

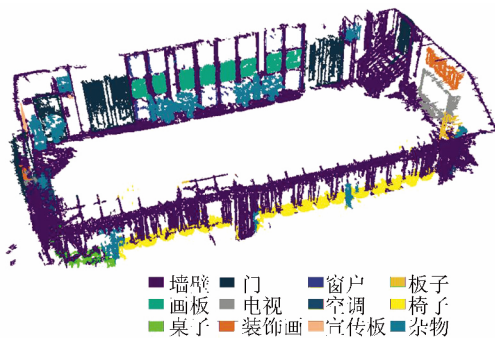


图7 点云语义标记结果

Fig.7 Semantic Label Result of Point Cloud

模型的 2D-3D 语义传递算法。该算法利用二维图像分类方法中取得的显著成绩和大量可用的 2D 标记训练数据,基于扩展 FCN 对图像进行语义分割和布局提取,融合图像间一致性和图像内一致性,在不依赖大量训练数据集的同时,进行精确的语义传递,得到室内三维点云精细的结构化语义分割模型。

### 参 考 文 献

- [1] Lu Xueliang, Tong Xiaochong, Zhang Yongsheng, et al. An Improved Region-Growing Surface Triangulation Algorithm for Urban Dense Point Cloud [J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(6): 832-837 (卢学良, 童晓冲, 张永生, 等. 城市密集点云的区域生长表面构网改进算法[J]. 武汉大学学报·信息科学版, 2016, 41(6): 832-837)
- [2] Liu Rufe, Lu Xiushan, Yue Guowei, et al. An Automatic Extraction Method of Road from Vehicle-Borne Laser Scanning Point Clouds [J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(2): 250-256 (刘如飞, 卢秀山, 岳国伟, 等. 一种车载激光点云数据中道路自动提取方法[J]. 武汉大学学报·信息科学版, 2017, 42(2): 250-256)
- [3] Tang Shengjun. Multi-view Image Enhancement of RGB-D Indoor High-Precision 3D Mapping Method [D]. Wuhan: Wuhan University, 2017 (汤圣君. 多视图图像增强的 RGB-D 室内高精度三维测图方法[D]. 武汉: 武汉大学, 2017)
- [4] Newcombe R A, Izadi S, Hilliges O, et al. Kinect Fusion: Real-time Dense Surface Mapping and Tracking [C]. *IEEE International Symposium on Mixed and Augmented Reality*, Austin, TX, USA, 2011
- [5] Furukawa Y, Ponce J. Accurate, Dense, and Robust Multiview Stereopsis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(8): 1 362-1 376
- [6] Wu C. VisualSFM: A Visual Structure from Motion System [OL]. <http://ccwu.me/vsfm/>, 2011
- [7] Koppula H S, Anand A, Joachims T, et al. Semantic Labeling of 3D Point Clouds for Indoor Scenes [C]. *Advances in Neural Information Processing Systems*, Granada, Spain, 2011
- [8] Anand A, Koppula H S, Joachims T, et al. Contextually Guided Semantic Labeling and Search for Three-dimensional Point Clouds [J]. *The International Journal of Robotics Research*, 2013, 32(1): 19-34
- [9] Xiong X, Munoz D, Bagnell J A, et al. 3-D Scene Analysis via Sequenced Predictions over Points and Regions [C]. *IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011
- [10] Kalogerakis E, Hertzmann A, Singh K. Learning 3D Mesh Segmentation and Labeling [J]. *ACM Transactions on Graphics*, 2010, 29(4): 1-12
- [11] Lai K, Fox D. Object Recognition in 3D Point Clouds Using Web Data and Domain Adaptation [J]. *The International Journal of Robotics Research*, 2010, 29(8): 1 019-1 037
- [12] Munoz D, Bagnell J A, Hebert M. Stacked Hierarchical Labeling [C] // *European Conference on Computer Vision*. Berlin, Heidelberg: Springer, 2010
- [13] Murphy K P, Torralba A, Freeman W T. Using the Forest to See the Trees: A Graphical Model Relating Features, Objects, and Scenes [C]. *Advances in Neural Information Processing Systems*, Vancouver and Whistler, British Columbia, Canada, 2004
- [14] Boulch A, Guerry J, Saux B L, et al. SnapNet: 3D Point Cloud Semantic Labeling with 2D Deep Segmentation Networks [J]. *Computers & Graphics*, 2017, 71: 189-198
- [15] Wu Z, Song S, Khosla A, et al. 3D ShapeNets: A Deep Representation for Volumetric Shapes [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015
- [16] Yuan Li, Chen Qinghu, Liao Haibin, et al. Rapid Three-Dimensional Reconstruction of Face with Single Vision [J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(4): 487-491 (袁理, 陈庆虎, 廖海斌, 等. 单视影像下的人脸快速三维重建[J]. 武汉大学学报·信息科学版, 2012, 37(4): 487-491)
- [17] Russell B C, Torralba A, Murphy K P, et al. LabelMe: A Database and Web-Based Tool for Image Annotation [J]. *International Journal of Computer Vision*, 2008, 77(1-3): 157-173

- [18] Deng J, Dong W, Socher R, et al. ImageNet: A Large-Scale Hierarchical Image Database[C]. IEEE Conference on CVPR, Vancouver, BC, Canada, 2009
- [19] Kuettel D, Guillaumin M, Ferrari V. Segmentation Propagation in ImageNet[C]. European Conference on Computer Vision, Florence, Italy, 2012
- [20] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015
- [21] Mallya A, Lazebnik S. Learning Informative Edge Maps for Indoor Scene Layout Prediction[C]. IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015
- [22] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[C]. The 22nd ACM International Conference on Multimedia, Orlando, Florida, USA, 2014
- [23] Gupta S, Arbelaez P, Malik J. Perceptual Organization and Recognition of Indoor Scenes from RGB-D Images[C]. IEEE Conference on CVPR, Portland, Oregon, 2013
- [24] Achanta R, Shaji A, Smith K, et al. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34 (11): 2 274-2 282

## Semantic Segmentation of Indoor 3D Point Cloud Model Based on 2D-3D Semantic Transfer

XIONG Hanjiang<sup>1</sup> ZHENG Xianwei<sup>1</sup> DING Youli<sup>1</sup> ZHANG Yi<sup>2</sup> WU Xiujie<sup>2</sup> ZHOU Yan<sup>1</sup>

<sup>1</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,  
Wuhan University, Wuhan 430079, China

<sup>2</sup> School of Mathematics and Statistics, Wuhan University, Wuhan 430079, China

**Abstract:** In this paper, we propose an effective algorithm based on graph model for semantic transfer from 2D images to 3D point clouds, which can effectively solve the problem of objectification and lack of structured information of 3D point cloud model. Our proposed method uses the extended full convolutional neural network to extract the indoor space layout and object semantics of 2D images, and then implements the transfer of 2D semantics to 3D semantics based on the 2D image superpixels and 3D point clouds as nodes to construct a graph model of consistency between images and intra-image consistency. The experiment from 3D point cloud shows that the proposed method can obtain accurate indoor 3D point cloud semantic classification results. The accuracy of point cloud classification can reach 73.875 2%, and the classification effect is better.

**Key words:** semantic 3D point cloud model; semantic transfer; semantic mark; point cloud classification

**First author:** XIONG Hanjiang, PhD, professor, specializes in the theories and methods of geospatial data information visualization, massive spatial data network services and fast transmission, and rapid reconstruction of 3D models. E-mail: xionghanjiang@whu.edu.cn

**Corresponding author:** ZHENG Xianwei, associate professor. E-mail: zhengxw@whu.edu.cn

**Foundation support:** The National Key Research and Development Program of China, No. 2018YFB0505401; the National Natural Science Foundation of China, Nos. 41871361, 41701445; the LIESMARS (State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing) Special Research Funding.