

DOI:10.19651/j.cnki.emt.2106027

# 基于卷积神经网络和 NBV 的三维重建方法\*

李爱军

(天津职业大学 电子信息工程学院 天津 300410)

**摘要:** 为了使三维(3D)重建适用于多种目标形状,并提高处理速度,提出一种基于机器学习的3D重建方法,重点解决“下一个最优视点”(NBV)规划问题。首先,给出NBV的定义和计算,建立离散NBV搜索空间;然后,生成NBV,同时对该空间进行迭代式重建。此外,为了处理NBV的学习问题,提出一个基于3D卷积神经网络的分类方法,将可能的传感器位姿考虑为一个分类问题。实验结果表明,所提方法的重建精度优于VoxNet网络方法,能较好地满足约束条件;与高精度信息增益方法相比,所提方法也取得了较优和接近的重建覆盖率,对于不同形状,基本上4次扫描就可以达到较高的覆盖率,且重建速度快约90倍。

**关键词:** 三维重建;下一个最优视点;分类问题;卷积神经网络;处理速度

**中图分类号:** TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6020

## 3D reconstruction method based on convolution neural network & NBV

Li Aijun

(School of Electronic Information Engineering, Tianjin Vocational University, Tianjin 300410, China)

**Abstract:** To make 3D reconstruction suitable for various target shapes and improve the processing speed, a 3D reconstruction method based on machine learning is proposed, which focuses on solving the "next best viewpoint" (NBV) planning problem. Firstly, the definition and calculation of NBV are given, and the discrete NBV search space is established. Then, NBV is generated, and the space is reconstructed iteratively. In addition, in order to deal with the learning problem of NBV, a classification method based on 3D convolution neural network is proposed, which considers the possible position and pose of sensor as a classification problem. The experimental results show that the reconstruction accuracy of the proposed method is better than that of the VoxNet network method, which can meet the constraints better. Compared with the high-precision information gain method, the proposed method also achieves better and close reconstruction coverage. Basically, it can achieve high coverage in 4 scans for different shapes, and the reconstruction speed is about 90 times faster.

**Keywords:** 3D reconstruction; next best viewpoint; classification problem; convolution neural network; processing speed

## 0 引言

3D目标重建<sup>[1]</sup>是一种广泛应用的计算机视觉任务,其应用领域广泛,如航天设备、电影动画、医学影像、目标探测和测量等。3D目标重建是从多个视点感测目标表面,生成对象的3D模型,当目标为未知时,一般通过传感器定位、感测、配准,以及下一个传感器位置的规划<sup>[2]</sup>。其中,规划步骤的难点主要是如何寻找能够最大化目标表面的传感器位置和朝向,尤其重要,称之为“下一个最优视点”(NBV)问题。

迄今,大部分用于NBV规划的先进方法都需要基于重建场景的表征、需求和约束来手动设计。其中,合成方法依赖目标表面相关信息<sup>[3]</sup>。合成方法的优势是速度快,但在处理遮挡的目标时性能通常会下降。基于搜索的方法则定义一个效用函数,在可行传感器位置上执行搜索,以优化效用函数。如文献[4]利用概率均匀网格进行信息表征,并将前沿单元格定义为相关特征,从在能够观察到最大数量的前沿单元格的传感器位置上执行搜索。文献[5-6]提出的模型使用了基于信息增益的效用函数。其目标旨在寻找能够观察到熵值较高的单元格(立体像素)的传感器位姿。

收稿日期:2021-03-17

\* 基金项目:天津市教育科学“十三五”规划课题(VESP3011)、2018年全国高等院校计算机基础教育研究会教学改革重点项目(2018-AFCEC-035)、2021年全国高等院校计算机基础教育研究会教学改革项目(2021-AFCEC-429)资助

基于搜索的方法的劣势是较为耗时,因为其需要对效用函数进行多次评估。文献[7]基于广义立体像和基础地理信息数据进行精确三维重建,并将其引用到遥感图像的阴影检测中。

另一方面,随着并行处理硬件技术、机器学习技术的快速发展,三维重建技术获得更多突破口。如文献[8]提出基于深度学习的单幅图像三维重建,将 2D 信息微调 and 复制组合成 3D 信息。文献[9]尝试学习效用函数来进行 3D 重建,但未完成整体 NBV 预测。

目前很多三维重建方法针对特定的形状,如果采用高精度基础数据,则需要花费更多处理时间,为了使 3D 目标重建适用于更多形状,并降低重建的时间复杂度,本文提出了基于监督式深度学习的 NBV 规划方法,该方法可自动生成数据集,通过 3D 卷积神经网络(3D-CNN),完成训练和测试,解决了 NBV 问题。实验比较了机器学习方法和非机器学习方法,实验结果表明,提出的网络架构的精度优于 VoxNet 网络和信息增益法,能够在多个未知目标的重建过程中预测 NBV,并取得较高的重建覆盖率,且处理速度非常快。

## 1 NBV 问题

NBV 的概念最早由 Connolly 提出。在此类任务中,将一个传感器以多个位姿放置在感兴趣目标周围,以得到其 3D 形状。由于目标形状的相关信息是有限的,因此通过定位、感知、配准和规划 NBV 的步骤来完成重建,在这个过程中,深度相机<sup>[10]</sup>扮演重要角色,很多原始精确数据出自深度相机。在下一迭代中,使用上一次迭代计算出的 NBV 做为当前传感器位姿。在定位过程中,将传感器放置在给定位姿。传感器的位姿(视点)定义了其位置和朝向,即向量  $\mathbf{v} = (x, y, z, \alpha, \beta, \gamma)^T$ , 其中,  $\alpha$  表示围绕  $x$  轴旋转,  $\beta$  表示围绕  $y$  轴旋转,  $\gamma$  表示围绕  $z$  轴旋转。在感知步骤,对目标表面进行测量,并得到目标形状的点云( $z$ )。感知步骤后,将得到的点云配准至单个模型。随着重建的开展,将采集到的信息存储到局部模型中。本文研究中,使用均匀概率占用网格  $M$ , 其中每个立体像素均有一个关联概率,表示目标的部分或全部表面位于该立体像素的容量内的可能性。

当前,研究人员通常将 NBV 计算作为搜索问题来处理,旨在寻找能够实现度量最大化的视点。不同于传统方法,本文方法基于局部模型提供的信息,直接预测 NBV。该预测必须基于从过去的重建中得到的知识。为此,本文将 NBV 学习问题定义为函数学习任务:

$$f(M); R^n \rightarrow R^3 \times SO(3) \quad (1)$$

以使得在  $\mathbf{v} = f(M)$  处的感知能够增加局部模型中包含的目标相关的知识。 $f$  的输入为立体像素网格,表示为式(1)中的  $R^n$ , 其中  $n$  为网格中的立体像素量。如果仅考虑占用概率,则  $f$  的域为  $[0, 1]^n$ 。输出为应进行感知的传

感器位姿。

提出的数据驱动方法将 NBV 计算分解为两个步骤。第 1 个步骤为  $f$  的离线学习过程,并利用过去重建中的大量样本;第 2 个步骤使用  $f$  对 NBV 进行估计,该步骤在重建过程中在线执行。总之,  $f$  的学习意味着求解回归问题。

## 2 提出的 3D 重建方法

本文提出了利用目标相关的完整信息和离散搜索空间,得出真实 NBV,所提方法的工作原理如下:给定一个示例对象(目标整个表面为已知),首先,建立离散 NBV 搜索空间;然后,生成 NBV,同时对该空间进行迭代式重建。下文将解释离散搜索空间,给出 NBV 的定义和计算,构建数据集的迭代程序。

### 2.1 生成搜索空间

该步骤中,针对所有示例目标构建视点(搜索空间)的离散集合。这种视点为放置传感器的唯一可用位姿。因此,感知结果和 NBV 均受该集合约束。形式上,离散搜索空间表示为  $V_d, V_d \subset V = R^3 \times SO(3)$ 。相关研究表明<sup>[11]</sup>,通过一定的标准化,可以得到目标周围的视图范围和每个视点的距离图像集合。具体来说,将目标放置在一个全局参考坐标系的中心。然后,通过旋转极坐标的点,创建出传感器的可能位姿集合  $V_d$ 。极坐标向量的大小是固定的,其提供了从目标中心到传感器位置的距离。接下来对角度进行均匀移动,以得到传感器位姿。参考坐标系和传感器位姿样例如图 1 所示。

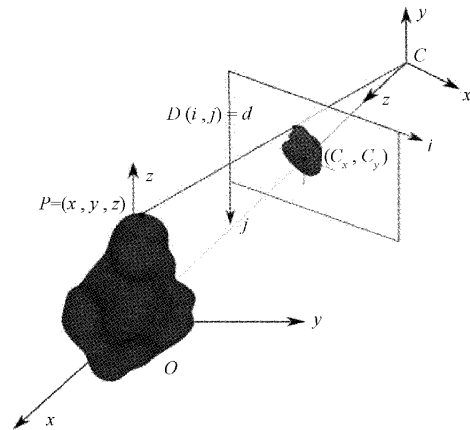


图 1 参考坐标系和传感器位姿

一旦建立起视点,则针对每个实例目标,生成与  $V_d$  中每个视点相对应的感知。该步骤对于提高性能很有必要,因为在后续的训练和验证节点中,需要从选定视点来观察目标。因此,需要每个视点的距离图像( $D$ )。首先根据相机参考坐标系,然后根据全局参考坐标系,将距离图像转换为点云。该步骤结束时,对于每个视点和每个目标,可得到一个与全局参考坐标系相关的点云。为表示方便,下文将参考坐标系中的每个点云称为“感知”,表示为  $z$ 。从某个

视点生成的每个感知表示为  $z(v)$ 。可能感知的集合表示为  $Z_d$ 。

## 2.2 真实 NBV

理论上, NBV 被定义为能够提高感兴趣目标的相关信息量的视点,但在实践中, NBV 通常表示能够优化效用函数的视点。该效用函数间接地测量信息增量。例如,统计未知立体像素数量的函数<sup>[12]</sup>(概率为 0.5 的立体像素)假定隐藏目标表面与未知像素之间存在正相关的关系。信息增益方法<sup>[5-6]</sup>也应用了同样的推理。此类方法有助于给出视点优度的近似。但其不能提供真实 NBV。为此,本节将介绍对于给定的概率网格,计算真实 NBV 的方法。提出的方法基于 NBV 必须最大化扫描表面的事实,且考虑到现实重建中所需的重叠。

假设传感器被放置的位姿集合表示为  $S = \{s_0, \dots, s_n\}$ ,  $S \subseteq V_d$ , 示例目标表示为  $W_{obj}$ , 为一个点云。目标点云应足够密集,以使得点之间的距离小于立体像素分辨率。同时,将目标的累积点云表示为  $P_{acu}$ 。值得一提,由于重建是感知的迭代整合,  $P_{acu}$  是到迭代  $i$  为止的传感器感知的整合结果,即  $P_{acu} = \bigcup_{i=0, n} z(s_i)$ 。利用函数  $Coverage(A, W_{obj})$ , 计算出点云  $A$  相对于真实  $W_{obj}$  的覆盖百分比。

现在,将真实 NBV( $v^*$ )定义为能够增加累积点云的视点:

$$v^* = \underset{v}{\operatorname{argmax}} Coverage(z(v) \cup P_{acu}, W_{obj}) \quad (2)$$

式(2)遵循如下约束:

- 1)  $v^*$  必须无冲突,即,其必须位于自由空间中。
- 2) 感知  $z(v^*)$  与  $P_{acu}$  的重叠必须高于一个阈值,  $overlap(z(v^*), P_{acu}) > thresh_1$ 。

3) 表面之间的公共区域必须至少包含 3 个 3D 特征 ( $thresh_2 \geq 3$ ),因为在重叠量之外,还需要 3D 特征来完成在 3 个轴上的 3D 配准。

根据经验,设  $thresh_1 = 50\%$ 。为计算重叠,定义半径  $gap$ ,以比较两个点云  $P_{acu}$  和  $z$ 。考虑到这些点云中包含的数据量,需要利用 KdTree 算法,对  $z$  中半径  $gap$  内的  $P_{acu}$  的所有近邻进行高速搜索。使用 NARF 点<sup>[13]</sup>作为 3D 特征。

## 2.3 迭代生成样本

本文提出如下算法,以生成数据集。该算法在给定一个已知目标表面的情况下,输出多个“下一个最优视点”。

该算法利用不同的初始化传感器位姿对示例目标进行多次重建,然后保存计算出的 NBV。在重建过程中,将当前网格( $M$ )、累积点云( $P_{acu}$ )和计算出的 NBV( $v^*$ )保存为一个样本。即使对于同一个目标,不同的初始视点也会生成不同的局部模型和不同的视点序列。

为生成数据集,需要输入的实例目标( $W_{obj}$ ),对应点间隙( $gap \leftarrow 0.005$ ),终止重建标准( $S_{cov}$ ),最大迭代次数( $\max_{iter}$ ),重叠百分比( $thresh_1$ ),视点集合( $V_d = \{v_1, v_2, \dots, v_n\}$ ),以及感知集合( $Z = \{z_1, z_2, \dots, z_n\}$ )。目标完

## 算法 1 数据集样本生成算法

```

1. for  $i \leftarrow 1; n$  do
2.    $v^* \leftarrow v_i$ 
3.    $iter \leftarrow 0$ 
4.    $P_{acu} \leftarrow \emptyset$ 
5.   Initialize( $M$ )
6.   while  $Coverage(P_{acu}, W_{obj}) < S_{cov}$ ,  $iter < \max_{iter}$  do
7.      $P_{acu} \leftarrow P_{acu} \cup z(v^*)$ 
8.      $P_{acu} \leftarrow DownsizeFiltering(P_{acu})$  // 缩减过滤
9.      $M \leftarrow UpdateGrid(M, z(v^*))$  // 更新网格
10.     $\Delta_{max} \leftarrow 0$ 
11.    for  $j \leftarrow 1; n$  do
12.       $z \leftarrow Perception(v_j)$ 
13.      if  $overlap(z', P_{acu}, gap) > thresh_1$  then
14.        if  $|NARF(z' \cap P_{acu})| > thresh_2$  then
15.           $\Delta \leftarrow Coverage(z', P_{acu}, W_{obj}) - Coverage(P_{acu}, W_{obj})$ 
16.          if  $\Delta > \Delta_{max}$  then
17.             $v^* \leftarrow v_j$ 
18.             $\Delta_{max} \leftarrow \Delta$ 
19.          end if
20.        end if
21.      end if
22.    end for
23.     $iter++$ 
24.    保存到数据集( $P_{acu}, M, v^*$ )
25.  end while
26. end for

```

成重建的条件为  $P_{acu}$  的覆盖百分比达到  $S_{cov}$ ,或迭代次数达到  $\max_{iter}$ ,见算法的第 6 行。通过取每个  $v \in V_d$ ,从不同的初始视点  $v_i$ (第 2 行)开始每次重建。读取在位姿  $v^*$  处感知到的每个点云  $z$ ,并将其添加到当前的  $P_{acu}$ (第 7 行);其后,对  $P_{acu}$  执行过滤操作,以保持密度均匀(第 8 行)。然后,在算法第 9 行,根据最近一次的感知  $z$ ,对概率网格进行更新。从第 11~22 行,在所有可能的视点上执行穷举搜索,以找到其感知  $z$  能够最大化当前  $P_{acu}$  的增量视点。为避免与当前感知  $z$  发生冲突,本文将  $z'$  定义为评估中的感知(第 12 行)。 $z'$  必须满足最小重叠百分比(第 13 行),且必须包含被计算为 NARF 点的  $thresh_2$  个 3D 特征。那么,  $v^*$  为能够最大化增量和满足约束的 NBV。最后,算法保存包含  $P_{acu}, M$  和  $v^*$  的三元组(第 24 行)。

## 2.4 利用深度学习进行 NBV 分类预测

为处理 NBV 学习问题,本文提出一个基于深度学习的分类方法,将可能的传感器位姿考虑为一个分类。深度

学习方法可表示为输入数据和输出数据两部分。第 1 部分由隐藏层或深度网络结构所组成,负责进行特征提取。第 2 部分包含对分类标签进行预测的网络层。

一般来说,传感器位姿被定义为一个连续空间。但为了使用分类方法,必须得到可能分类的离散集合。为此,本文将位姿聚集为一个集合。每个可能的位姿将被称为一个分类。采用的架构是 3D 卷积神经网络<sup>[14-15]</sup>(3D-CNN),3D-CNN 是 CNN 的一个特殊案例。迄今为止,很少有研究将 3D-CNN 用于 NBV 预测。为此,本文提出了基于问题本质进行 NBV 分类预测的架构。

为简化符号,使用了如下缩写:1)  $C(f, k, s)$ :3D 卷积层,其中包含  $f$  个特征,核大小为  $k \times k \times k$ ,步长为  $s \times s \times s$ 。2)  $P(s)$ :步长  $s \times s \times s$  的最大池化操作。3)  $FC(n)$ :全连接层,将  $n$  个参数作为输出。提出的架构的连接结构如图 2 所示,其中  $C$  表示分类数量。除池化层之外,在每层之后,使用 ReLU 激活函数。应用 softmax 函数与交叉熵以提供单独编码。提出的网络架构将大小为  $32 \times 32 \times 32$  的卷馈入到网络中,并给出 14 个可能的分类作为输出。

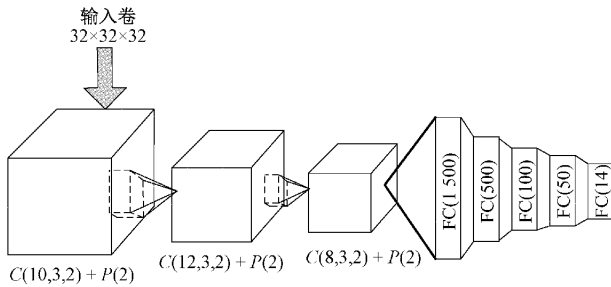


图 2 提出的网络架构

### 3 实验结果与分析

#### 3.1 数据集与参数设置

实验中使用的数据集的每个样本均是概率网格及其相应的 NBV 的元组。概率网格中包含  $32 \times 32 \times 32$  个立体像素,且包含整个目标。向每个网格分配的 NBV 为集合  $V_d$  中的一个视点。对于该数据集,围绕目标创建 20 个传感器位姿,但仅  $C=14$  为有用位姿,因为剩余 6 个位姿在地面以下,不能提供目标相关信息。

生成整个数据集的处理时间约为 200 h。硬件平台配置了主频 2.8 GHz Intel® Core™ i5 和 8G RAM。使用 12 个示例目标创建出样本。从 263 个初始视点每个目标进行重建,选择围绕视图范围均匀分布的视点<sup>[16]</sup>。考虑到初始位姿之间的微小差异不会提供重要信息,本文未使用整个  $V_d$  集合作为初始视点。将覆盖停止标准  $S_{cov}$  设为 80%,由此实现在覆盖率和迭代次数之间的合理权衡。最大扫描次数  $\max_{iter}$  被设为 10。对应点距  $gap$  被设为 0.005 m。在重建过程中,由于目标形状和初始视点的不同,不同目标的重建达到停止标准(覆盖率或最大迭代次

数)的迭代次数也不同。不同目标重建的迭代次数从 3~9 次不等,大部分在 3~5 次之间。数据集被限制为 14 个分类。

#### 3.2 不同网络的测试

使用所描述的数据集对本文网络进行训练。此外,本文还实施并训练了其他两个网络,即简单全连接网络<sup>[17]</sup>和 VoxNet 网络<sup>[18]</sup>,其中,全连接网络<sup>[17]</sup>简称 FC 网络,以比较不同网络架构的精度。FC 网络分为 4 层,其架构为 FC(1500)-FC(750)-FC(100)-FC(50)。本文将全连接网络的实验结果作为基线。虽然 VoxNet<sup>[18]</sup>是针对目标识别而设计的网络,但该网络设计与提出的架构较为相似,因此在本文实验中被纳入比较。

利用 Adam 随机优化方法执行网络训练。学习率为 0.001,批大小为 200。在 FC 层和最后一个卷积层之后,添加 0.7 的 Dropout。利用来自截断正态分布的数值对卷积层进行初始化,其中  $\sigma = 0.1, \mu = 0$ 。本文实验在 Tensorflow 中实施。利用相同的超参数完成对本文架构、FC 和 VoxNet 的训练。训练代数设为 500 代。对数据集进行分割,80%用于训练,20%用于测试。根据测试集的学习,训练在 500 代后终止,该代之后,测试精度在  $\pm 0.02$  的范围内上下浮动。图 3 所示为每个网络达到的测试精度。从图中可发现,本文架构的测试精度优于 VoxNet 和 FC 网络。虽然测试精度未达到 1,但得到了令人满意的结果。由于 FC 网络表现出较差的验证精度,后续实验中将不对该网络进行测试。

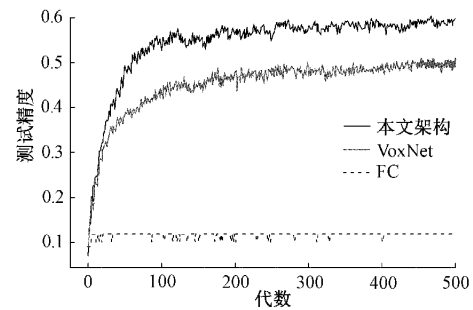


图 3 不同网络的测试精度比较

#### 3.3 与机器学习方法比较

该节的目标在于证明,本文网络架构计算出的传感器位姿能够提高目标表面覆盖率。实验流程如下:1)将传感器放置在随机初始位置;2)获取一个感知;3)对概率网格进行更新;4)将网格馈入每个受训网络,并向前传播,此时每个网络将提供一个分类(NBV);5)将传感器放置在给定分类,并获取一个新的感知。重复该过程,直到覆盖率不再增加,或网络提供了之前已访问过的传感器位姿。

针对 3 个新对象执行重建,这些对象包括电话、熨斗和瓶子。每个网络对每个目标进行 10 次重建。对于两个网络,使用相同的随机初始位姿。表 1 所示为从随机初始位置中进行 10 次重建的平均覆盖率( $S$ )、标准偏差( $\sigma$ )和迭代次



数( $I$ )结果。由表 1 结果可知,提出的网络架构实现了比其他网络更高的覆盖率,且更好地满足了重叠约束(3.2 节),这是因为覆盖表面的百分比增量比 VoxNet 更加平缓。

表 1 10 次重建的比较结果

目标	本文网络			VoxNet <sup>[18]</sup>		
	$S_{cov}/\%$	$\sigma_s$	$I$	$S_{cov}/\%$	$\sigma_s$	$I$
瓶子	93.01	3.44	4.1	83.11	11.77	2.1
熨斗	80.65	1.35	4.7	77.02	1.70	2.7
电话	84.01	3.25	3.6	79.12	5.35	1.9

图 4 所示为使用本文 NBV 迭代的重建过程,其中,黑色立体像素表示占用空间,浅灰色立体像素表示未知空间。对于图 4(a)“瓶子”的重建,初始的覆盖率为 54.21%,第 1 次迭代后的覆盖率为 80.92%;第 2 次迭代后,覆盖率为 90.53%;第 3 次迭代后,覆盖率为 91.55%;第 4 次迭代后,覆盖率为 93.01%。可以看出经过 4 次迭代后,已经非常接近全覆盖。对于图 4(b)“熨斗”的重建,原始的覆盖率为 42.10%,经过 4 次迭代后,覆盖率为 80.65%,由于熨斗特殊的形状,此覆盖率是后续迭代的稳态值。对于图 4(c)“手机”的重建,只需要 2 次迭代达到稳态值,84.01%,这可能是手机的造型较为简单,3D-CNN 可以快速学习到手机的外观特征。这几个实例充分说明本文方法可以快速达到重建效果,而且保持了较好的覆盖率。

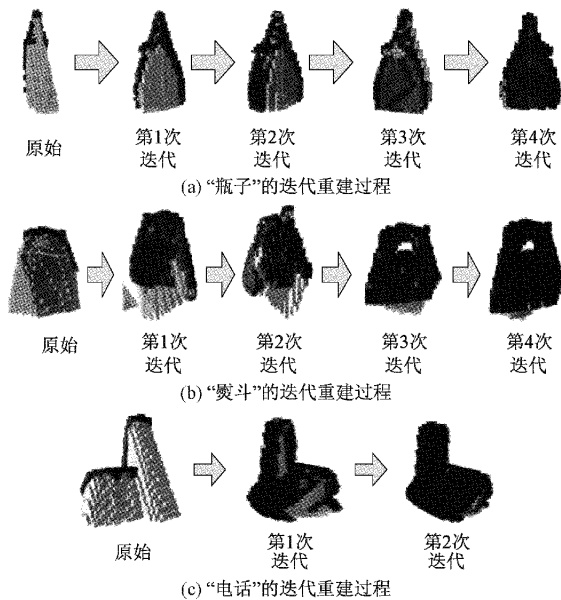


图 4 对目标进行重建的应用实例

### 3.4 与非机器学习方法比较

本节将训练后的所提方法与信息增益方法<sup>[5]</sup>(非机器学习方法)进行比较。所有方法均需要一组候选视点,以便根据其自身的度量对其进行评价。为进行公平比较,所有方法均使用围绕目标周围的 20 个候选视点。实验流程如

下:从一个固定的初始视点进行一次初始扫描,其后每个方法迭代地计算 NBV 并更新模型,直至 10 次扫描为止。重建的对象如图 5 所示,包括“球体”、“茶壶”、“兔子”和“龙”。重建的结果如图 6 所示,其中,黑色立体像素表示占用空间,浅灰色立体像素表示未知空间。第 1 行是本文重建结果,第 2 行是信息增益方法的重建结果。可以看出,两种方法获得的结果都有一定的表明缺陷,前 3 个对象的重建结果相差无几,可能本文方法会更好些,因此对覆盖面积进行计算,获得覆盖百分比/扫描次数如表 2 所示,可以看出本文方法较优,但在“龙”的重建中表现不佳,这可能是由于:1)数据集中未包含类似形状,完成重建的样本数量不足;2)与以往方法相比,本文方法省去了传感器建模过程。

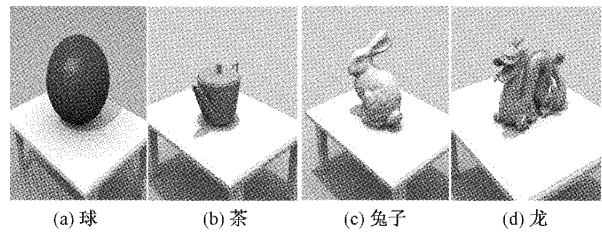


图 5 用于比较的对象

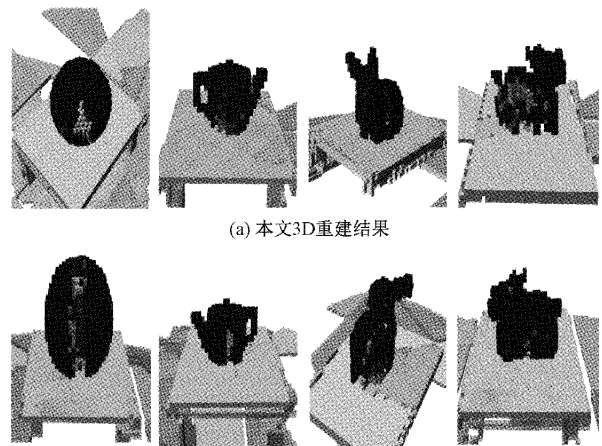


图 6 重建结果比较

表 2 两种方法的比较结果(覆盖百分比/扫描次数)

	本文方法	信息增益方法 <sup>[5]</sup>	VoxNet <sup>[18]</sup>
球体	96.7%/9	96.2%/7	95.0%/6
茶壶	94.9%/5	92.5%/7	93.7%/6
兔	95.1%/3	93.2%/9	94.4%/9
龙	89.3%/8	90.4%/10	88.4%/10

最后,所提方法在早期迭代中效率很高,在极短时间内即达到了很好的覆盖率,但在后期迭代中表现稍差。所提方法的下一个最优视点的计算时间是 0.12 s,而信息增益方法需要 10.83 s,所提方法快 90 倍。所提方法表现出的快速计算优势是因为信息增益方法在候选视点评估过程中

对传感器进行显式建模,需要应用极为耗时的光线追踪。表 3 所示为 4 次扫描后的覆盖率,从表中可发现,本文方法利用极短的处理时间,在早期迭代即达到了较高的覆盖率。

表 3 4 次扫描后的比较结果(覆盖率) %

	本文方法	信息增益方法 <sup>[5]</sup>	VoxNet <sup>[18]</sup>
球体	93.5	92.2	91.9
茶壶	89.9	85.0	88.4
兔	94.0	90.1	89.8
龙	71.3	85.0	82.8

## 4 结 论

本文基于监督式深度学习的 NBV 计算方法用于 3D 重建。该方法包含数据集自动生成算法,以及一个 3D 卷积神经网络。通过对所提网络进行训练来辅助 NBV 的计算。实验结果表明,所提方法能够预测出覆盖大部分目标表面的 NBV。此外,所提方法不需要进行极为耗时的光线追踪,处理时间很少。未来,本文会尝试将 NBV 学习问题作为回归问题来处理,以避免分类集合的限制。

### 参考文献

- [1] DELMERICO J, ISLER S, SABZEVARI R, et al. A comparison of volumetric information gain metrics for active 3D object reconstruction [J]. *Autonomous Robots*, 2018, 54(12): 97-108.
- [2] 叶春凯, 万旺根. 针对重复性结构下的三维场景重建方法研究[J]. *电子测量技术*, 2019, 42(16): 134-138.
- [3] DING L, DEAN-BEN X L, RAZANSKY D. Efficient 3-D model-based reconstruction scheme for arbitrary optoacoustic acquisition geometries [J]. *IEEE Transactions on Medical Imaging*, 2017, 36(9): 1858-1867.
- [4] BIRCHER A, KAMEL M, ALEZIS K, et al. Receding horizon “next-best-view” planner for 3d exploration [C]. *IEEE International Conference on Robotics and Automation(ICRA)*, Montreal, Canada: IEEE Press, 2016: 1462-1468.
- [5] 孙瀚. 模型驱动的自主重建研究[D]. 合肥: 中国科学技术大学, 2018.
- [6] ISLER S, SABZEVARI R, DELMERICO J, et al. An information gain formulation for active volumetric 3d reconstruction[C]. *IEEE International Conference on Robotics and Automation*, Stockholm, Sweden: IEEE Press, 2016: 3477-3484.
- [7] 董胜光, 秦建新, 郭云开. 基于广义立体像对的高分辨率遥感影像阴影检测[J]. *电子测量与仪器学报*, 2019, 33(3): 111-117.
- [8] 郑恩, 成耀天, 林靖宇. 采用去抖动模糊算法的稠密三维重建[J]. *计算机工程与应用*, 2018, 54(1): 217-223.
- [9] HEPP B, DEY D, SINHA S N, et al. Learn-to-score: Efficient 3d scene exploration by predicting view utility[C]. *The European Conference on Computer Vision (ECCV)*, Munich, Germany: IEEE, 2018: 437-452.
- [10] 耿璇, 王召巴, 金永, 等. 基于单个深度相机的人体三维重建[J]. *国外电子测量技术*, 2020, 39(9): 143-146.
- [11] 石广升. 基于 Kinect 的物体三维模型构建和姿态估计方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2015.
- [12] VASQUEZ-GOMEZ J I, SUCAR L E, MURRIETA C R. View/state planning for three-dimensional object reconstruction under uncertainty [J]. *Autonomous Robots*, 2017, 41(9): 89-109.
- [13] 胡修祥, 张良. 结合 NARF 特征的改进型 3D-NDT 多视点云配准[J]. *信号处理*, 2015(12): 132-137.
- [14] 段萌, 王功鹏, 牛常勇. 基于卷积神经网络的小样本图像识别方法[J]. *计算机工程与设计*, 2018, 39(1): 224-229.
- [15] 胡海洋, 丁佳民, 胡华, 等. 基于三维卷积神经网络的工作流识别方法[J]. *计算机集成制造系统*, 2018, 24(7): 163-173.
- [16] MATURANA D, SCHERER S. Voxnet: A 3D convolutional neural network for real-time object recognition [C]. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Hamburg, Germany: IEEE, 2015: 922-928.
- [17] 孟录, 高恒上, 张含光, 等. 基于全连接神经网络的三维人体姿态估计[J]. *仪器仪表学报*, 2020, 41(10): 167-179.
- [18] DEWEI Z, QI C, ZILONG Z, et al. An improved method for model-based training, detection and pose estimation of texture-less 3D objects in occlusion scenes[J]. *Procedia Cirp*, 2019, 83(12): 541-546.

### 作者简介

李爱军, 硕士, 讲师, 主要研究方向为三维重建、神经网络等。  
E-mail: aij\_li@126.com