

# 透明刚体非侵入式三维重建技术研究进展

潘智辉<sup>†</sup>, 陈睿<sup>†</sup>, 胡昌平, 徐静\*

清华大学机械工程系, 北京 100083

**摘要** 三维重建在自动驾驶、元宇宙等工业领域有着重要的应用前景。其中,透明刚体的三维重建由于其复杂的成像规律在重建过程中存在一定挑战。本综述回顾了透明刚体的三维重建技术,重点介绍了基于优化可视外壳的方法和基于深度学习的方法,讨论了基于可视外壳重建的原理和各类优化方法,还讨论了基于深度学习的 RGB-D 深度补全和多视图方法,并介绍了目前透明物体的数据集,最后对可能的发展方向进行了讨论和展望。

**关键词** 三维重建; 透明物体; 可视外壳; 深度学习

中图分类号 O439

文献标志码 A

DOI: 10.3788/LOP223415

## Research Progress in Non-Intrusive Three-Dimensional Reconstruction of Transparent Rigid Bodies

Pan Chifai<sup>†</sup>, Chen Rui<sup>†</sup>, Hu Changping, Xu Jing\*

Department of Mechanical Engineering, Tsinghua University, Beijing 100083, China

**Abstract** The three-dimensional (3D) reconstruction technology has considerable application prospects in industrial fields such as autopilot and metaverse. The 3D reconstruction of transparent rigid bodies exhibits some challenges owing to its complex imaging laws. This review examines the 3D reconstruction technology of transparent rigid bodies, focuses on the methods based on optimized visual hull and deep learning, discusses the principle and different optimization methods based on visual hull reconstruction, describes the RGB-D depth completion and multiview reconstruction based on deep learning, presents the current dataset of transparent objects, and finally discusses the possible development direction of the 3D reconstruction technology.

**Key words** three-dimensional reconstruction; transparent object; visual hull; deep learning

## 1 引言

随着工业机器人、自动驾驶等技术的发展,获取周边物体信息的感知技术得到越来越多的重视并迅速发展。其中,物体的三维信息(如点云)对于机器人的识别、抓取等任务起到非常关键的作用。机器人能够通过视觉信息在非结构化的环境中进行规划与决策,最后利用执行机构完成设定的目标。同时,在元宇宙应用方面,物体的数字化身是元宇宙的基础单元,而三维重建是数字化身的基础方法之一。相比于使用专业三维软件对物体进行建模,自动三维重建可以为元宇宙工业减少成本并提供便利。随着自动驾驶、元宇宙等领域的发展,三维重建技术有着巨大的应用潜力和前景。

过去十几年内,针对物体的三维重建已发展了很

多不同的方法,如全自动扫描<sup>[1]</sup>、结构光<sup>[2-4]</sup>、多视角立体重建<sup>[5-6]</sup>等主动或被动的重建方法。这些方法能够很好地应用于非透明物体的三维重建。但是透明物体与光线存在相对复杂的交互:除了反射以外,光线还会在不同材质的界面相交处发生折射并穿过透明物体的表面。因此,传统的重建方法在透明体上的表现欠佳,具体表现为特征匹配困难,或是在利用结构光解码时存在信息缺失的问题。

为解决透明体重建问题,各种针对性的方法相继被提出。从对象看可以大致分为刚体对象以及流体对象,如气体<sup>[7-8]</sup>、火焰<sup>[9-10]</sup>等。从方法上看可以大致分为非侵入式和侵入式。典型的侵入式方法将物体浸泡在液体里,通过改变光路来辅助重建<sup>[11-12]</sup>,但侵入式方法可能对物体表面产生破坏。在机器人、自动驾驶、元宇

收稿日期: 2022-12-28; 修回日期: 2023-02-16; 录用日期: 2023-03-01; 网络首发日期: 2023-03-11

通信作者: jingxu@tsinghua.edu.cn

<sup>†</sup> 共同第一作者

宙等运用到视觉感知的领域中,透明刚体的非侵入式三维重建方法受到更多关注。

在透明刚体的重建框架中(图 1),除使用偏振信息<sup>[13-14]</sup>、结构光信息<sup>[15-16]</sup>恢复表面的方法外,优化可视外壳(visual hull)<sup>[17]</sup>和深度学习也是常用的方法。优化可视外壳方法先获取重建物体的粗糙表面轮廓,再使用三角测量法等方式进行优化。深度学习方法在不

需要对透明物体复杂的成像规律进行建模的前提下,通过监督学习可以很好地从数据集中学习到图像信息与物体三维信息之间对应关系的概率分布,并使用其对透明物体进行重建。本文对非侵入式透明刚体的重建方法中的基于优化可视外壳和基于深度学习的重建方法进行重点介绍。

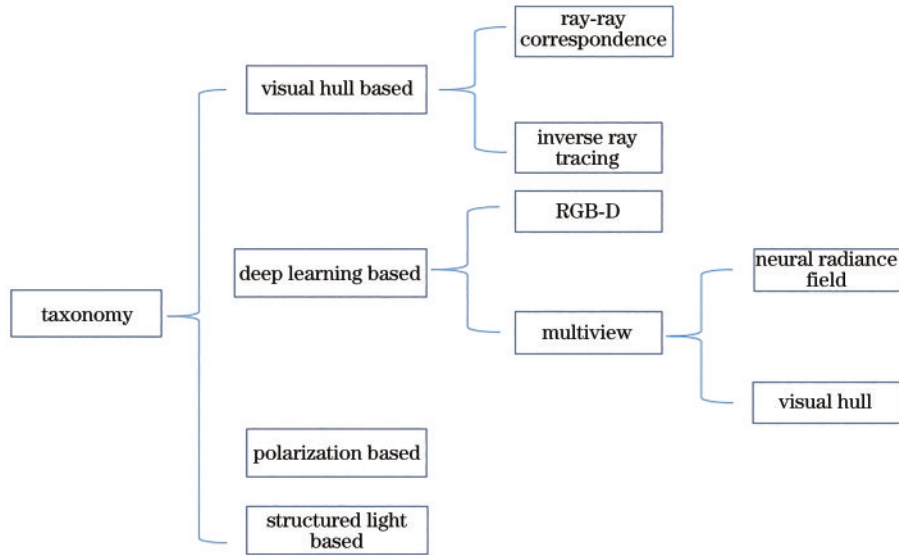


图 1 非侵入式透明刚体三维重建方法分类

Fig. 1 General taxonomy of 3D reconstruction of a transparent rigid body

## 2 基于优化可视外壳的重建方法

可视外壳是利用物体剪影进行三维重建得到的一种几何实体,如图 2 所示,一般获取可视外壳的方法是在物体四周采集多张图片,通过阈值化处理将物体与背景分离得到物体的剪影轮廓。在三维空间中,剪影轮廓和视点会形成一个锥体,最后利用足够多视角下锥体的交集来对物体进行三维重建。该三维重建方法最早由 Laurentini<sup>[17]</sup>提出,并在后来运用于透明物体的三维重

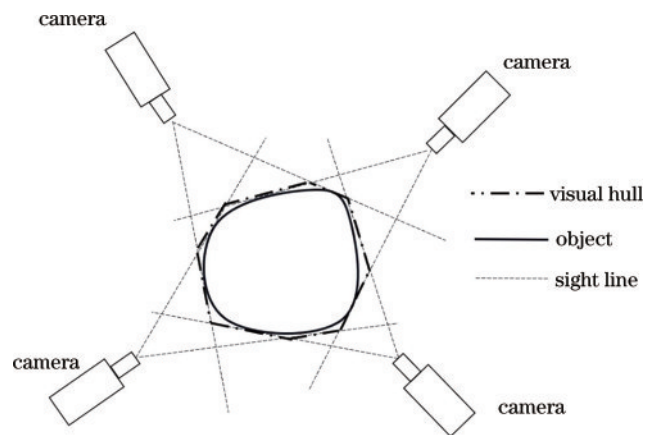


图 2 利用剪影重建可视外壳的原理(2D 的情况)

Fig. 2 Principle of using silhouette to reconstruct the visual hull (two-dimensional case)

建当中。由于透明物体的边界相比于非透明物体更难以识别,透明物体的可视外壳并不能满足精度需求,一般需要通过优化方法来提高可视外壳的精度,按照优化的方法可以分为光线对应和反向光线追踪的方法。

光线对应基于“光线在透明物体里发生不超过两次折射”的假设,通过入射光线和出射光线的对应关系来优化物体轮廓,其基本思想是确认入射光线的方向或使用三角测量法等方法来确认光线与透明物体的交点来对透明物体的表面进行重建或优化。Ben-Ezra 等<sup>[18]</sup>最早基于扭曲恢复透明物体表面形状,通过先验知识假设一个初始轮廓,然后使用相机内参从相机视点出发计算光线终点为远处特征的人射光线,最后基于“远处特征的人射光线是相互平行”的这一假设来优化初始轮廓。其优化初始轮廓的思想对后续的透明物体重建框架产生了深远的影响。Kutulakos 等<sup>[19]</sup>对使用三角测量法来恢复物体表面的重建框架进行了理论分析,指出当光线发生多于两次的折射或反射时,无论添加多少个相机视角,也无法通过对光线的回溯来重建物体表面。这为基于三角测量的重建方法提供了理论基础,发生全发射的区域由于发生了多于两次的折射或反射,理论上将无法使用三角测量法进行重建。Qian 等<sup>[20]</sup>使用光线对应的方法来优化物体轮廓,在透明物体的正面和背面分别放置显示器来确定进入和离开透明物体的光线。基于“光线在透明物体里发生不

超过两次折射”的假设,使用三角测量法确认透明物体一侧的深度,再使用“香农定律得到的法向量和物体表面法向量一致”这一条件来优化另一侧的深度。这个方法并没有使用可视外壳的方法来初始化轮廓,所以重建得到的是放置了显示屏那两侧的表面。Wu 等<sup>[21]</sup>则在此基础上使用可视外壳的方法进行透明物体轮廓的初始化,使用与 Qian 等<sup>[20]</sup>相似的优化方法,通过约束香农定律的法向量和物体轮廓的法向量来优化可视外壳。该方法相比 Qian 等<sup>[20]</sup>的方法在设备布置上更加方便,使用转台来得到物体的剪影并且不需要手动移动显示屏和相机,而且得到的是透明物体的整体轮廓而非正面和背面的一对表面信息。

基于反向光线追踪的优化是另一种常见的优化方式,其基本思想是减少反向光线追踪得到的光线和真实光线之间的偏转误差来优化可视外壳。Miyazaki 等<sup>[22]</sup>使用光线追踪的方法,通过优化测量和渲染出的偏振信息之间的误差来优化透明物体的表面。虽然该方法使用了光线追踪的方法来优化表面,但需要对透明物体有一定的先验知识来假设初始轮廓。Kim 等<sup>[23]</sup>则对对称的透明酒杯类物体进行了三维重建。该方法使用剪影来生成初始化的可视外壳,然后对物体只发生两次折射的区域(玻璃杯杯柄)使用反向光线追踪的方法,减少渲染出的光线偏转的误差来估计折射率,再对发生了四次折射的区域(玻璃杯杯腔)使用相同的方法优化内壁轮廓。Lü 等<sup>[24]</sup>也使用与 Kim 类似的方法优化物体表面,但是其方法可以应用于任意形状的透明物体。该方法根据反向光线追踪的误差设计了损失函数来优化初始化的可视外壳,可以很好地使用可微框架优化。

综上所述,基于优化可视外壳的方法已经比较成熟,主要的创新点在于优化方法的不同。除了重建精度的差异,不同的优化方法对所需测量设备的数量和

操作复杂度的要求也不同。同时,相比于对透明物体轮廓进行简单先验假设,可视外壳对于各类优化方法可提供更好的初始化信息,也为之后的研究提供了很好的重建框架。

### 3 基于深度学习的重建方法

传统的三维重建方法如光流法<sup>[25]</sup>、三角测量法<sup>[11]</sup>,需要对重建过程中的物理系统进行详尽建模,而透明物体的成像规律相比于非透明物体更加复杂,会出现如全反射等复杂的光学现象。所以在传统的透明物体重建过程中,为了简化物理模型,大多数文章使用“光线在透明体内部只发生不超过两次折射”的假设。相比于直接对物理系统建模,人们可以使用深度神经网络在不需要对系统有一个详尽理解的前提下学习到观测图像与物体三维信息之间对应关系的概率分布,并将其运用于表面重建。由于透明物体复杂的光学效应,一般 RGB-D 相机得到的深度信息是有精度损失的(图 3),基于学习的方法可以使用预测出的法向量、边缘等信息对损失的深度进行补充优化。Chen 等<sup>[26]</sup>最早把透明物体的抠图问题视为折射流问题,并使用深度神经网络对折射流进行学习,但该方法仅限于不带颜色的透明物体。该方法对于之后使用神经网络处理透明物体重建的问题产生了深远影响。Stets 等<sup>[27]</sup>使用虚拟引擎渲染出透明物体在随机环境下的 RGB 图像、分割掩膜(segmentation mask)、深度图(depth map)、法向图(normal map)并使用神经网络进行监督学习,输入单张 RGB 图像可生成相对深度图。但多视图信息能够提供更多的信息进行重建,因此更为常用。接下来将基于学习的透明物体三维重建方法按照 RGB-D 深度补充类方法和多视图类方法进行阐述。

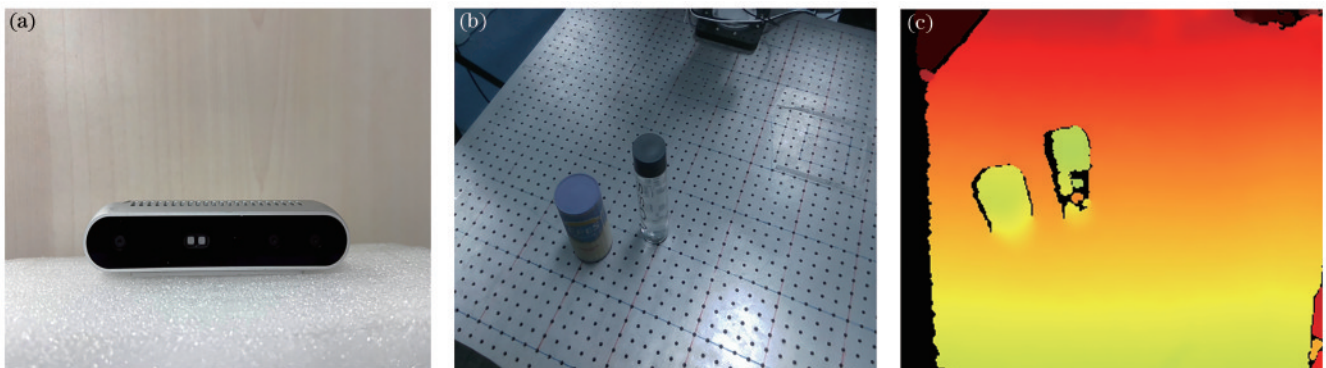


图 3 RGB-D 相机与其拍摄的 RGB 图像和深度信息。(a) RealSense D415 RGB-D 深度相机;(b)使用 RealSense D415 拍摄的包含透明物体和非透明物体的 RGB 图片;(c)使用 RealSense D415 获取的深度信息,其中透明物体的深度是缺失的

Fig. 3 RGB-D camera and the RGB image and depth map from it. (a) RealSense D415 RGB-D camera; (b) RGB image containing opaque and transparent objects from RealSense D415; (c) the depth information obtained by RealSense D415, in which the depth of transparent object is missing

#### 3.1 基于 RGB-D 深度补充的方法

基于 RGB-D 深度补充的方法的基本思想是使用

图像中跟深度信息相关联的信息来优化和补充缺失的深度。Sajjan 等<sup>[28]</sup>借鉴 Zhang 等<sup>[29]</sup>关于修补深度信息

的框架并将其运用于透明物体的深度补全,认为相比于深度信息,透明物体的 RGB 图像里展示出了更多关于物体表面曲率的信息,并提出以下方法:先使用 RGB 图像推理出透明物体的边缘信息和法向信息,再将这两个信息结合进行全局优化,对 RGB-D 相机得到的初始深度进行补全。由于透明物体真实的法向信息和深度信息难以从传感器直接获得,Sajjan 等<sup>[28]</sup>还制作了仿真数据集和真实数据集来给神经网络进行训练和测试:使用渲染引擎 blender<sup>[30]</sup>和光线追踪渲染引擎 blender cycle<sup>[31]</sup>生成的仿真数据集包含 50000 张 RGB-D 图像信息;使用同等大小的非透明物体替换透明物体的方法采集的真实数据集包含 286 张 RGB-D 图像信息。

相比使用透明物体的边缘信息和法向信息对深度进行补全,Xu 等<sup>[32]</sup>利用 RGB-D 相机采集的原始深度信息来补全缺失的深度。该方法采用一个新的网络结构将原始深度信息逆投影生成粗糙点云,并将粗糙点云放入点云修复模块修复,然后把修补好的点云再投影回深度信息,最后把 RGB 图像和修补好的深度信息放入深度补全模块来预测最终深度。

Dai 等<sup>[33]</sup>为减小仿真数据集和真实数据的差异,提出一种基于域随机化增强的深度相机仿真方法,在 blender<sup>[30]</sup>渲染引擎里仿真模拟 RealSense D415 相机,并将其得到的左右红外图像进行立体匹配,生成带有真实传感器噪声的深度图像,同时对物体类别、材质、位置等参数进行域随机化,最终生成的仿真数据集具有很好的泛化性和 sim-to-real 迁移效果。

除了使用如法向、边缘等特征信息来优化缺失深度外,基于任务的目标优化也被引入透明体重建中。Jiang 等<sup>[34]</sup>首次将机器人抓取任务中的 affordance map 引入透明物体重建的框架中,affordance map 表示每一个抓取点的置信度信息。该方法结合 affordance map 和法向、边缘信息一起进行全局优化,并且采取多步优化策略,对深度不连续的部分(如玻璃杯边缘和内腔之间)和深度连续的部分(如玻璃杯底与桌面连接的部

分)采取了不同的优化方法来补全深度。

在网络结构和优化方法的创新方面,Chen 等<sup>[35]</sup>将 Transformer 网络<sup>[36]</sup>引入透明物体的三维重建当中,基于 Swin Transformer 网络<sup>[37]</sup>构建编码器来提取上下文信息,并使用编码器-特征融合-解码器的网络结构。Dai 等<sup>[33]</sup>则基于 Swin Transformer 网络从 RGB 图像和初始深度提取出多尺度特征信息,并使用交叉注意力 Transformer 网络来融合特征,最后基于置信图插值完成深度的预测。Tang 等<sup>[38]</sup>则将生成对抗网络(GANs)<sup>[39]</sup>引入透明物体重建的框架当中,在 Sajjan 等<sup>[28]</sup>的网络框架的基础上用生成对抗网络来代替全局优化,使用生成器来预测补全深度并使用判别器来将其与真实深度博弈。Fang 等<sup>[40]</sup>基于密集连接卷积网络模块(dense block)<sup>[41]</sup>构建类似于编码器-解码器的网络结构 DFNet,由于密集连接卷积网络模块的参数数量少,DFNet 在 GPU 内存占用和推理速度上相比 Sajjan 等<sup>[28]</sup>的模型更有优势。

相比使用点云、深度等信息的重建方法,利用隐式表示(implicit representation)的重建方法由于其出色的重建效果在三维重建领域受到越来越多的关注。Zhu 等<sup>[42]</sup>受神经辐射场<sup>[43-45]</sup>的启发,先生成包含点云的体素网格,再将 RGB 信息、体素信息、光线信息输入隐式表示函数里,生成光线在体素里终止的概率和位置,最后沿光线通过最大池化的方法预测深度。得益于包含点云的体素网格的轻便性,该方法相比于传统神经辐射场类渲染方法<sup>[43]</sup>有更快的渲染速度,并且比 Sajjan 等<sup>[28]</sup>的方法在推理速度上有了更大的提升。

表 1 比较了以上提及的 RGB-D 深度补全的方法,主要涉及深度补全的效果、推理速度和 GPU(graphics processing unit)内存占有率表现这 3 个角度。其中,>A 表示该方法的表现(深度补全的效果/推理速度/GPU 内存占有率表现)好于标签为 A 的方法(更好的补全效果/更快的推理速度/更少的 GPU 内存占有率)。

表 1 RGB-D 深度补全方法的比较  
Table 1 Comparison of RGB-D depth completion methods

Label	Method	Depth completion performance	Inference speed	GPU memory occupation performance
A	ClearGrasp <sup>[28]</sup>			>B
B	LIDF <sup>[42]</sup>	>A (better than ClearGrasp <sup>[28]</sup> )	>C	
C	TransparentNet <sup>[32]</sup>	>A	>A	>A
D	DepthGrasp <sup>[38]</sup>	>A		
E	DFNet <sup>[40]</sup>	>A	>B	>C
F	A4T <sup>[34]</sup>	>A		
G	TODE-Trans <sup>[35]</sup>	>B, E		
H	SwinDRNet <sup>[33]</sup>	>B		

目前基于 RGB-D 补全的方法的基本思路一般是将更多的与深度相关的信息引入神经网络中,人们希

望神经网络能够从这些信息中提炼出深度信息从而恢复损失的深度。目前,法向、边缘、点云等信息已经被

很好地用于恢复损失深度。由于图像中的信息有限,想要从图像中提取出新的信息引入神经网络已经比较困难。Jiang 等<sup>[34]</sup>将 affordance map 用于深度修补的思路是值得借鉴的,相比于直接从图像中再去获取新的和深度相关的信息,将与深度有关的“任务信息”等非图像类信息引入深度补全是值得研究的方向。这类研究可以让深度补全任务与和三维重建不相关的学科交叉,比如还可以将触觉、听觉等非图像类信息引入深度补全任务当中,目前这方面还没有很多的研究,是一个值得探索的方向。同时,由于大部分 RGB-D 深度补全的方法都是基于神经网络的方式,当深度学习领域里出现了性能相比之前更好的网络结构时,这些网络结构都能在调整后应用于透明物体的三维重建当中。这些网络可以带来好的重建效果,但是网络本身对于透明物体的三维重建任务的适应性和针对性还有待观察。从透明物体重建原理的角度来设计网络是一个值得研究的方向,可以从成像原理、光学原理等角度设计网络模块、损失函数等,这将会对该重建任务更有针对性和专一性。

### 3.2 基于多视图的方法

基于学习的多视图方法主要分为使用神经辐射场(neural radiance field)的方法或是可视外壳(visual hull)的方法。使用神经辐射场的方法使用体渲染(volume rendering)沿相机光线上对神经辐射场中的颜色、透明度等信息进行积分,从而渲染出相机光线经过的像素的颜色(图 4)。从不同相机投射光线,可以得到物体不同视角的图片。

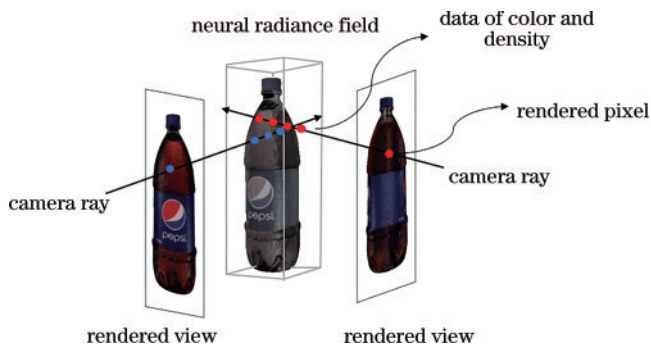


图 4 神经辐射场的渲染原理

Fig. 4 Principle of rendering for the neural radiance field

对于使用神经辐射场的方法,Ichnowski 等<sup>[46]</sup>将神经辐射场引入透明物体的三维重建中并用其处理抓取任务。相比于传统的神经辐射场方法,该方法在估计深度时不采取累计体素密度(volume density)的方式,而是沿光线上选取第一个达到预设体素密度阈值的深度。Dai 等<sup>[47]</sup>也将神经辐射场运用于透明物体的抓取之中。相比于传统的神经辐射场方法,他们先预测截断符号距离场(truncated signed distance field prediction),再将其转换为体素密度。Dai 等<sup>[47]</sup>的重建

方法在抓取速度上有很大的优势,并且仅需 6 张图片就能重建并进行实时的抓取工作。

对于使用可视外壳的方法,相比于第 2 节传统的可视外壳优化方法,深度学习给出了新的优化方法。该方法具有更好的重建精度,如表 2 所示。传统可视外壳优化方法受限于“光线在透明物体里不多于两次折射”的假设,在发生全反射或多于两次折射的区域将产生误差。Li 等<sup>[48]</sup>则提出了基于物理模型的神经网络,采用的可微渲染层能够根据透明物体的前景和后景的法向量和环境图生成 RGB 图像和全反射图,再基于代价聚合的思想通过采样生成透明物体前景和后景的法向量的聚合体,使用代价聚合体和可微渲染层的输入输出来预测和优化前后景法向量,最后使用优化后的前后景法向量来预测和优化可视外壳。该方法在重建的速度上要明显优于 Wu 等<sup>[21]</sup>的方法。设备和操作方面虽然仅需使用手持相机对透明物体进行拍摄,但还需要获取透明物体周围的环境图。

表 2 传统的可视外壳优化方法与基于深度学习的优化方法的重建精度比较

Table 2 Comparison of reconstruction accuracy between the traditional visual-hull-based optimization method and deep-learning-based optimization method

Object	Chamfer distance / mm	
	Traditional visual hull based optimization	Deep learning based optimization
Mouse	0.804	0.535
Pig	0.558	0.487
Dog	0.225	0.186

### 3.3 透明物体数据集

对于基于深度学习的重建方法来说,数据集和网络结构一样重要。很多工作在提出新的网络结构的同时也给出了新的数据集。表 3 将包含有透明物体深度信息的数据集进行了汇总。其中,原始深度(raw depth)表示直接从传感器获取的深度信息,由于透明物体复杂的成像规律,其原始深度往往都是不准确或是部分缺失的。为了获取透明物体的准确深度(ground truth depth),可以使用相同尺寸的非透明物体对透明物体进行替换然后再采取其深度信息,或是在仿真环境下获取仿真物体的深度信息。在现有的透明物体真实数据集中,ClearPose<sup>[49]</sup>的数据集数据量和物体数量最多。而在透明物体仿真数据集中,DREDS<sup>[33]</sup>的数据量和物体数量最多,并且由于使用了仿真相机进行深度匹配,该数据集还有带有真实噪声的原始深度,这在仿真数据集中比较特殊。TransCG<sup>[40]</sup>和 ClearPose<sup>[49]</sup>数据集除了透明物体和不透明物体,还包含半透明物体的数据。而 ClearGrasp<sup>[28]</sup>、TOD<sup>[50]</sup>、TODD<sup>[32]</sup>数据集只包含透明物体的数据。

表 3 透明物体数据集列表  
Table 3 List for datasets of transparent object

Name of the dataset	Number of frames	Type of data	Number of objects	Type of material
TOD <sup>[50]</sup>	40000(real)	RGB image, raw depth, GT(ground truth) depth, pose, and segmentation mask	15	Transparent
ClearGrasp <sup>[28]</sup>	50000(synthetic) +286(real)	RGB image, GT depth, surface normal, pose, and semantic segmentation	9(synthetic) +10(real)	Transparent
Omniverse <sup>[42]</sup>	60000(synthetic)	RGB image, GT depth, 2D/3D bounding box, pose, and segmentation mask	Unknown	Transparent, opaque
TODD <sup>[32]</sup>	15000(real)	RGB image, raw depth, GT depth, pose, and segmentation mask	6	Transparent
TransCG <sup>[40]</sup>	58000(real)	RGB image, GT depth, surface normal, pose, and segmentation mask	51	Transparent, translucent, and opaque
ClearPose <sup>[49]</sup>	350000(real)	RGB image, GT depth, surface normal, pose, and segmentation mask	63	Transparent, translucent, and opaque
DREDS <sup>[33]</sup>	130000(synthetic)	RGB image, raw depth, and GT depth	1861	Transparent, opaque
STD <sup>[33]</sup>	27000(real)	RGB image, raw depth, GT depth, segmentation mask, normalized object coordinate space map	50	Transparent, opaque

随着时间的推移,透明物体的数据集的规模已经越来越大,如 ClearPose<sup>[49]</sup> 包含 350000 张透明物体真实照片,对于透明物体重建这一个具体的任务来说,已经初具规模。但现在大部分的真实数据集都将物体类别集中在杯子、容器等透明物体上,类别数量十分有限,目前物体类别数量最多的真实数据集 ClearPose<sup>[49]</sup> 包含 63 种不同类别的透明物体(大部分为不同类别的玻璃容器),这对于泛化的透明体重建任务来说并不是一个完全可靠的数量。更具有泛化性的真实物体数据集将是一个值得发展和探索的领域,数据集中需要加入更多非容器类的透明物体以及非均匀的、透明和不透明混合的物体。由于透明物体会把周围环境也同时显现出来,关于玻璃物体周围环境的泛化也是值得探索的,也可以针对特定的工业应用场景进行环境的布置,如生产线的装配环境或是户外环境等。此外,仿真现实迁移技术也值得关注。在仿真环境中,对透明物体以及场景进行泛化和随机化将更加容易,并且相比于真实数据集,仿真数据集的成本更低并且生产效率更高。但是,由于仿真数据集和真实数据集存在差异(gap),使用网络从仿真数据集中学习再迁移到真实透

明物体的测试集上的效果还很有限,减少真实数据集和仿真数据集的差异将是未来数据集的一个发展方向。同时,减少仿真与现实差异的网络结构模块或是损失函数,也是透明物体重建任务网络结构的发展方向。

## 4 方法比较

为了充分阐明不同重建方法的优越性和局限性,表 4 从重建质量、重建速度和应用价值几个方面对基于优化可视外壳的方法和基于深度学习的方法进行了比较。基于优化可视外壳的方法具有较高的精度,重建误差为 0.10 mm,但“光线在透明物体里发生不超过两次折射”的假设限制了该类方法的重建精度,在发生多于两次折射或是全反射的区域存在点云缺失的问题。在时间开销方面,对于每个重建物体都需要 30 min~1 h 的图片采集过程以及 10 min~1 h 不等的可视外壳生成和优化过程。在硬件方面,该类方法需要相机、显示器、旋转台等设备采集图片,硬件搭建成本较高。综上所述,基于优化可视外壳的方法具有精度高、重建时间长、所需设备多的特点,更适用于对于

表 4 不同重建方法的比较  
Table 4 Comparison between different reconstruction methods

Method	Error /mm	Time	Strength	Shortcoming	Application
Visual hull based	0.10	Longer than one hour	High-accuracy reconstruction	Slow reconstruction speed and complicated setup	Measurement
Deep learning based	10	Real time	Fast reconstruction speed, simple setup	Low-accuracy reconstruction, high-cost dataset	Robot manipulation and pose estimation

精度要求高的测量任务。基于深度学习的方法的重建精度较低,重建误差为 10 mm。但该类方法具有很好的实时性,且该类方法对于设备的要求较低。其中,基于 RGB-D 深度补全的方法仅需一台深度相机就可以完成图片的采集。因此,基于深度学习的方法可以适用于实时性要求较高或硬件使用受限的任务,比如机器人操作、物体姿态估计等。此外,基于深度学习的方法的重建效果依赖于数据集的质量,多样性好的数据集会让模型拥有更好的泛化性。面对获取大规模真实透明物体的数据集成本较高的问题,一个有效的解决方案是使用仿真数据集以降低成本,但仿真数据集与真实数据集之间的迁移仍是一个待研究与解决的问题。

## 5 总结和展望

随着自动驾驶、机器人等工业领域的快速发展,物体的三维重建得到越来越多的重视。由于透明物体成像时复杂的光学规律,透明物体的三维重建存在一定的挑战。本文针对透明刚体的三维重建方法进行了原理的阐述,详细讨论了基于优化可视外壳的方法和基于深度学习的方法。由于要使用到显示屏等设备来确认光线,基于可视外壳的方法更偏向于测量等应用场景。而基于学习的方法在训练完网络后大多可以在相机上实时应用,所以更偏向于环境感知等应用场景。

伴随深度学习的迅猛发展,网络结构的创新和数据集的完善将给基于深度学习的重建方法带来新的变革。其中,神经辐射场是一个值得关注的方向,其由于出色的重建效果在近年来得到广泛关注。如何减少神经辐射场重建方法所需的训练时间和照片数量将是未来值得研究的方向。目前关于重建的研究方向主要集中在提高重建速度和精度上,而如何使用简单的设备(如不带深度传感器的单目相机)对透明物体进行实时重建也是一个值得探索的研究方向。

未来工业上的对于透明物体的感知将会越来越重要,无论是使用机器人对玻璃类物体进行抓取,还是对透明物体进行数字化重建,都需要获取精确的点云信息。相信随着越来越多人研究透明物体的重建,新的重建方式或是优化方法将会给透明物体重建领域带来变革性的发展,也将会给工业以及人类生活带来便利。

## 参 考 文 献

- [1] Wu S H, Sun W, Long P X, et al. Quality-driven Poisson-guided autoscanning[J]. *ACM Transactions on Graphics*, 2014, 33(6): 1-12.
- [2] 苏显渝, 张启灿, 陈文静. 结构光三维成像技术[J]. *中国激光*, 2014, 41(2): 0209001.  
Su X Y, Zhang Q C, Chen W J. Three-dimensional imaging based on structured illumination[J]. *Chinese Journal of Lasers*, 2014, 41(2): 0209001.
- [3] 张启灿, 吴周杰. 基于格雷码图案投影的结构光三维成像技术[J]. *红外与激光工程*, 2020, 49(3): 0303004.  
Zhang Q C, Wu Z J. Three-dimensional imaging technique based on Gray-coded structured illumination[J]. *Infrared and Laser Engineering*, 2020, 49(3): 0303004.
- [4] Li B W, An Y T, Cappelleri D, et al. High-accuracy, high-speed 3D structured light imaging techniques, and potential applications in intelligent robotics[J]. *International Journal of Intelligent Robotics and Applications*, 2017, 1(1): 86-103.
- [5] 吴铮铮, 寇展. 基于单目多视角影像的场景三维重建[J]. *光学与光电技术*, 2020, 18(5): 51-56.  
Wu Z, Kou Z. 3D reconstruction of a scene based on monocular multi-view image[J]. *Optics & Optoelectronic Technology*, 2020, 18(5): 51-56.
- [6] Chen R, Han S F, Xu J, et al. Point-based multi-view stereo network[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 1538-1547.
- [7] Ji Y, Ye J W, Yu J Y. Reconstructing gas flows using light-path approximation[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 2507-2514.
- [8] Atcheson B, Ihrke I, Heidrich W, et al. Time-resolved 3D capture of non-stationary gas flows[C]//SIGGRAPH Asia '08: ACM SIGGRAPH Asia 2008 papers, December 10-13, 2008, Singapore. New York: ACM Press, 2008: 1-9.
- [9] Ihrke I, Magnor M. Image-based tomographic reconstruction of flames[C]//Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation, August 27, Goslar, Germany. 2004, New York: ACM Press, 2004: 365-373.
- [10] Wu Z H, Zhou Z, Tian D L, et al. The reconstruction of three-dimensional flame with color temperature[J]. *The Visual Computer*, 2015, 31(5): 613-625.
- [11] Han K, Wong K Y K, Liu M M. A fixed viewpoint approach for dense reconstruction of transparent objects [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 4001-4008.
- [12] Han K, Wong K Y K, Liu M M. Dense reconstruction of transparent objects by altering incident light paths through refraction[J]. *International Journal of Computer Vision*, 2018, 126(5): 460-475.
- [13] Sun Z, Qiao Y, Jiang Z G, et al. An accurate Fourier-based method for three-dimensional reconstruction of transparent surfaces in the shape-from-polarization method [J]. *IEEE Access*, 2020, 8: 42097-42110.
- [14] Drouet F, Stolz C, Laligant O, et al. 3D reconstruction of external and internal surfaces of transparent objects from the polarization state of highlights[J]. *Optics Letters*, 2014, 39(10): 2955-2958.
- [15] Guo H Y, Zhou H W, Banerjee P P. Surface shape reconstruction of transparent objects using structured light [C]//OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP), July 19-23, 2021,

- Washington, DC. Washington, D.C.: Optica Publishing Group, 2021: DTth5C. 4.
- [16] He K J, Sui C Y, Huang T Y, et al. 3D surface reconstruction of transparent objects using laser scanning with a four-layers refinement process[J]. *Optics Express*, 2022, 30(6): 8571-8591.
- [17] Laurentini A. The visual hull concept for silhouette-based image understanding[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, 16(2): 150-162.
- [18] Ben-Ezra, Nayar. What does the motion reveal about transparency? [C]//*Proceedings Ninth IEEE International Conference on Computer Vision*, October 13-16, 2003, Nice, France. New York: IEEE Press, 2008: 1025-1032.
- [19] Kutulakos K N, Steger E. A theory of refractive and specular 3D shape by light-path triangulation[J]. *International Journal of Computer Vision*, 2008, 76(1): 13-29.
- [20] Qian Y M, Gong M L, Yang Y H. 3D reconstruction of transparent objects with position-normal consistency[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4369-4377.
- [21] Wu B J, Zhou Y, Qian Y M, et al. Full 3D reconstruction of transparent objects[J]. *ACM Transactions on Graphics*, 2018, 37(4): 1-11.
- [22] Miyazaki D, Ikeuchi K. The shape estimation of transparent objects by using the inverse polarization ray tracing[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(11): 2018-2029.
- [23] Kim J, Reshetouski I, Ghosh A. Acquiring axially-symmetric transparent objects using single-view transmission imaging[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1484-1492.
- [24] Lü J H, Wu B J, Lischinski D, et al. Differentiable refraction-tracing for mesh reconstruction of transparent objects[J]. *ACM Transactions on Graphics*, 2020, 39(6): 1-13.
- [25] Bruss A R, Horn B K P. Passive navigation[J]. *Computer Vision, Graphics, and Image Processing*, 1983, 21(1): 3-20.
- [26] Chen G Y, Han K, Wong K Y K. TOM-net: learning transparent object matting from a single image[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 9233-9241.
- [27] Stets J, Li Z Q, Frisvad J R, et al. Single-shot analysis of refractive shape using convolutional neural networks [C]//*2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, January 7-11, 2019, Waikoloa, HI, USA. New York: IEEE Press, 2019: 995-1003.
- [28] Sajjan S, Moore M, Pan M K, et al. Clear grasp: 3D shape estimation of transparent objects for manipulation [C]//*2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 31 - August 31, 2020, Paris, France. New York: IEEE Press, 2020: 3634-3642.
- [29] Zhang Y D, Funkhouser T. Deep depth completion of a single RGB-D image[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 175-185.
- [30] Blender. Blender physics engine[EB/OL]. [2022-10-08]. <https://docs.blender.org/manual/en/latest/physics/index.html>. 3.
- [31] Blender. Blender cycles[EB/OL]. [2022-10-08]. <https://docs.blender.org/manual/en/2.80/render/cycles/introduction.html>. 3.
- [32] Xu H P, Wang Y R, Eppel S, et al. Seeing glass: joint point cloud and depth completion for transparent objects [EB/OL]. (2021-09-30)[2022-12-01]. <https://arxiv.org/abs/2110.00087>.
- [33] Dai Q Y, Zhang J Y, Li Q W, et al. Domain randomization-enhanced depth simulation and restoration for perceiving and grasping specular and transparent objects[M]//Avidan S, Brostow G, Cissé M, et al. *Computer vision-ECCV 2022. Lecture notes in computer science*. Cham: Springer, 2022, 13699: 374-391.
- [34] Jiang J Q, Cao G Q, Do T, et al. A4T: hierarchical affordance detection for transparent objects depth reconstruction and manipulation[J]. *IEEE Robotics and Automation Letters*, 2022, 7(4): 9826-9833.
- [35] Chen K, Wang S C, Xia B H, et al. TODE-trans: transparent object depth estimation with a transformer [EB/OL]. (2022-09-18)[2022-10-09]. <https://arxiv.org/abs/2209.08455>.
- [36] Vaswani A, Shazeer N, Parmar N, et al. Attention is all You need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, December 4-9, 2017, Long Beach, California, USA. New York: ACM, 2017: 6000-6010.
- [37] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//*2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 9992-10002.
- [38] Tang Y J, Chen J H, Yang Z G, et al. DepthGrasp: depth completion of transparent objects using self-attentive adversarial network with spectral residual for grasping[C]//*2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, September 27-October 1, 2021, Prague, Czech Republic. New York: IEEE Press, 2021: 5710-5716.
- [39] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [40] Fang H J, Fang H S, Xu S, et al. TransCG: a large-scale real-world dataset for transparent object depth completion and a grasping baseline[EB/OL]. (2022-02-17)[2022-10-09]. <https://arxiv.org/abs/2202.08471>.
- [41] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 21-26, 2017, Honolulu, HI, USA. New



- York: IEEE Press, 2017: 2261-2269.
- [42] Zhu L Y, Mousavian A, Xiang Y, et al. RGB-D local implicit function for depth completion of transparent objects[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 4647-4656.
- [43] Mildenhall B, Srinivasan P, Tancik M, et al. NeRF[J]. *Communications of the ACM*, 2022, 65(1): 99-106.
- [44] Liu L, Gu J, Zaw Lin K, et al. Neural sparse voxel fields [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 15651-15663.
- [45] Schwarz K, Liao Y, Niemeyer M, et al. GRAF: generative radiance fields for 3D-aware image synthesis [C]//*Proceedings of the 34th International Conference on Neural Information Processing Systems*, December 6 - 12, 2020, Vancouver, BC, Canada. New York: ACM, 2020: 20154-20166.
- [46] Ichnowski J, Avigal Y, Kerr J, et al. Dex-NeRF: using a neural radiance field to grasp transparent objects[EB/OL]. (2021-10-27)[2022-10-09]. <https://arxiv.org/abs/2110.14217>.
- [47] Dai Q Y, Zhu Y, Geng Y R, et al. GraspNeRF: multiview-based 6-DoF grasp detection for transparent and specular objects using generalizable NeRF[EB/OL]. (2022-10-12)[2022-11-23]. <https://arxiv.org/abs/2210.06575>.
- [48] Li Z Q, Yeh Y, Chandraker M. Through the looking glass: neural 3D reconstruction of transparent shapes[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1259-1268.
- [49] Chen X T, Zhang H J, Yu Z R, et al. ClearPose: large-scale transparent object dataset and benchmark[M]//Avidan S, Brostow G, Cissé M, et al. *Computer vision-ECCV 2022. Lecture notes in computer science*. Cham: Springer, 2022, 13668: 381-396.
- [50] Liu X Y, Jonschkowski R, Angelova A, et al. KeyPose: multi-view 3D labeling and keypoint estimation for transparent objects[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11599-11607.