



图学学报  
*Journal of Graphics*  
ISSN 2095-302X, CN 10-1034/T

## 《图学学报》网络首发论文

题目: 基于投票决策的实时遮挡处理技术  
作者: 蒋素琴, 张梦骏, 李蔚清, 苏智勇  
收稿日期: 2020-09-24  
网络首发日期: 2021-02-07  
引用格式: 蒋素琴, 张梦骏, 李蔚清, 苏智勇. 基于投票决策的实时遮挡处理技术. 图学学报. <https://kns.cnki.net/kcms/detail/10.1034.T.20210205.1743.002.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于投票决策的实时遮挡处理技术

蒋素琴<sup>1</sup>, 张梦骏<sup>2</sup>, 李蔚清<sup>3</sup>, 苏智勇<sup>1</sup>

(1. 南京理工大学自动化学院, 江苏南京 210094; 2. 上海机电工程研究所, 上海 201109; 3. 南京理工大学计算机科学与工程学院, 江苏南京 210094)

**摘 要:** 针对当前基于深度信息的虚实遮挡处理技术面临的实时性差和精度低的问题, 提出一种基于局部区域深度估计和基于 patch 相似性噪声点投票融合的实时虚实遮挡处理算法。算法将真实场景视频序列作为输入, 首先利用局部区域深度估计算法通过稀疏重建估算出稀疏关键点的深度信息, 对稀疏深度施加目标区域的约束限制深度向周围像素的传播, 从而快速恢复出目标区域的相对深度图; 然后, 噪声点投票融合算法利用深度比较确定虚实物体的前后位置关系, 基于 patch 相似性和投票决策的方法对区域内的像素进行投票和融合绘制; 最后输出具有真实遮挡关系的融合效果。实验结果表明, 该算法不仅可以提高虚实遮挡的实时性, 还能够获得真实和虚拟场景不同空间关系下的良好融合效果。

**关 键 词:** 虚实遮挡; 深度估计; 区域约束; 投票决策; 图像块相似性

**中图分类号:** TP 391

**文献标识码:** A

**文章编号:** 2095-302X(0000)00-0000-00

## Real-time virtual and real occlusion processing technology based on voting decision

JIANG Su-qin<sup>1</sup>, ZHANG Meng-jun<sup>2</sup>, LI Wei-qing<sup>3</sup>, SU Zhi-yong<sup>1</sup>

(1. School of Automation, Nanjing University of Science and Technology, Nanjing Jiangsu 210094, China;

2. Shanghai Institute of Mechanical and Electrical Engineering, Shanghai 201109, China;

3. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing Jiangsu 210094, China)

**Abstract:** Aiming at the problems of poor real-time performance and low accuracy faced by the current virtual and real occlusion processing technology based on depth information, a real-time virtual and real occlusion processing algorithm based on local area depth estimation and fusion of noise points based on patch similarity is proposed. Algorithm takes real scene video sequence as input. Firstly, the local area depth through sparse reconstruction, and the sparse depth is imposed on the target area to limit the depth propagation to the surrounding pixels, so as to quickly restore the relative depth map of the target area. Then, the noise point voting fusion algorithm uses depth comparison to determine the front and back position relationship of the virtual and real objects, and votes and merges the pixels in

收稿日期: 2020-09-24; 定稿日期: 2021-01-31

**Received:** 24 September, 2020; **Finalized:** 31 January, 2021

**基金项目:** “十三五”装备预研项目(61409230104, 1017, 315100104); 中央高校基本科研业务费专项(30918012203); 上海航天科技创新基金(SAST2019009); 上海市青年科技英才扬帆计划资助(18YF1410200)

**Foundation items:** "13th Five-Year" Equipment Pre-research Project (61409230104, 1017, 315100104); Special Funds for Fundamental Scientific Research of Central Universities(30918012203); Shanghai Aerospace Science and Technology Innovation Fund (SAST2019009); Supported by Yang Fan Program for Young Scientists in Shanghai(18YF1410200)

**第一作者:** 蒋素琴(1995-), 女, 江苏盐城人, 硕士研究生。主要研究方向为计算机视觉和增强现实。E-mail: Jiangsq0@163.com

**First author:** JIANG Su-qin (1995-), female, master student. Her main research interests cover computer vision and augmented reality.

E-mail: Jiangsq0@163.com

**通信作者:** 苏智勇(1981-), 男, 江苏泰州人, 副教授, 博士。主要研究方向为计算机图形学、计算机视觉。E-mail: su@njust.edu.cn

**Corresponding author:** SU Zhi-yong (1981-), male, associate professor, Ph.D. His main research interests cover computer graphics, computer vision, etc.

E-mail: su@njust.edu.cn

the area based on the patch similarity and voting decision method. Finally, the fusion effect with real occlusion relationship is output. The experimental results show that the proposed algorithm can not only improve the real-time performance of virtual and real occlusion, but also obtain a good fusion effect under different spatial relationships between real and virtual scenes.

**Keywords:** occlusion; depth estimation; regional constraints; voting decision; patch similarity

增强现实(Augmented Reality, AR)技术<sup>[1]</sup>是一种将虚拟信息与真实世界巧妙融合的技术,能够有效增强真实环境和虚拟环境之间的无缝融合,最终达到自然、逼真、和谐的人机交互。目前大多数的AR系统只是简单地将真实环境和虚拟对象进行叠加,没有考虑两者的空间位置关系和虚拟对象的信息延迟问题。错误的空间遮挡关系,会导致用户在感官方向上的迷失和空间位置上的错乱;而信息的延迟也不利于增强现实的沉浸式体验。所以只有高效、实时地解决两者在空间中的一致性问题,才能得到具有强烈真实感的虚实融合效果。

目前对于虚实遮挡的研究主要分为基于模型的方法、基于对象的方法和基于深度的方法。基于模型<sup>[2-3]</sup>的方法需要事先对会发生遮挡的真实物体进行建模并将其作为掩膜精准覆盖,利用深度测试确定空间关系。基于对象<sup>[4-5]</sup>的方法中,真实物体和虚拟对象的相互遮挡关系是固定不变的,从而将问题转换为虚拟对象的跟踪注册。基于深度<sup>[6-7]</sup>的方法利用深度传感器捕获深度或者利用深度估计算法得到场景中每个物体到视点的距离,通过比较真实场景与虚拟对象距离相机的远近关系,进行相应的遮挡渲染。其中,基于深度的虚实遮挡方法由于不需要预先建立真实场景的三维模型,不受复杂环境制约,更加适用于动态场景,而成为当前研究的主流方法。

当前,基于深度的虚实遮挡处理技术在实际应用中主要存在两个问题。一个是深度数据的精度问题。不管是利用深度传感器获取的深度图,还是基于图像/视频通过稀疏重建得到的深度图,都可能存在噪声和孔洞问题。另一个问题是虚实遮挡处理的实效性。由于虚实遮挡处理需要计算场景深度图,逐一比较场景中每个像素的深度信息,因此往往比较耗时,使得整个处理过程很难达到实时性的要求。

针对如何获得场景中各像素的深度信息的问题, Kim等<sup>[8]</sup>提出一种基于立体匹配的算法获得真实场景的深度信息,用于真实场景的重建和虚实遮挡处理。Valentin等<sup>[9]</sup>利用关键帧图像匹配技术获取场景的深度信息,实现虚拟物体的跟踪和实时的虚实遮挡处理。以上算法基于稠密匹配的思想,都存

在计算量大,实时性差的问题。Valentini等<sup>[10]</sup>利用深度传感器获取场景的三维几何结构并对用户进行有效的跟踪,实现动态环境的遮挡。Du等<sup>[11]</sup>提出基于边缘捕捉的深度图增强方法,解决深度图的边界噪声和虚实遮挡的边缘视觉伪影问题。吕浩等<sup>[12]</sup>提出联合双边滤波器的方法,填充像素缺失区域并保持锐利的深度图像边缘特征,提升了深度图的质量。以上处理都是利用深度传感器获取深度图,但是图像质量易受噪声和设备的影响,适用于理想的室内环境。基于稀疏重建的方法<sup>[13-15]</sup>可以提高运行效率。Stühmer等<sup>[16]</sup>利用单个手持摄像机采集多幅连续图像,实时估计多幅图像的深度图,进行密集的几何图像重建。Holynski等<sup>[17]</sup>提出利用同步跟踪与建模(simultaneous localization and mapping, SLAM)技术快速深度感知的方法,利用图像序列和稀疏重建的3D点信息,快速恢复出场景中每个像素点的深度值,从而解决虚实遮挡问题。Weerasekera等<sup>[18]</sup>根据稀疏地图和稠密深度预测的点间置信度,利用置信权值和概率数据来融合深度图,使得解决方案对稀疏的错误深度的存在不敏感。以上算法都是利用稀疏重建得到场景的稠密深度信息,对稀疏点的可靠性要求高,没有考虑错误稀疏点对深度恢复的影响。精度问题关键是得到准确的稀疏数据或者解决错误数据融合导致的孔洞问题,从而实现逼真的虚实遮挡效果。

针对如何提高虚实融合实时性问题, Roxas等<sup>[19]</sup>提出实时分割前景物体的方法,利用前景物体轮廓的深度信息与虚拟物体进行遮挡处理。上述遮挡实现虽然简化了计算量,但是在复杂背景和强烈遮挡的动态环境下,无法完美分割出前景物体的轮廓信息。Hebborn等<sup>[20]</sup>在融合过程中,根据传感器捕获的真实场景深度和渲染虚拟对象得到的区域,指定深度图像的前景、背景、无效和未知区域,实时进行有效区域的动态遮挡处理。实时性问题重点关注虚实融合的有效区域,从而加快动态场景的遮挡处理。

针对上述问题,本文提出通过施加区域约束提高深度图计算效率,通过融合局部错误深度值改善虚实遮挡效果。在提高深度图效率方面,提出增加

区域约束从而限制稀疏3D点深度传播的改进方法。因为虚实遮挡只出现在虚拟物体绘制的区域，所以只在虚拟区域传播深度到周围像素，从而快速恢复出深度图。在噪声数据和局部错误深度值导致的孔洞问题方面，提出基于patch相似性和投票决策的方法。考虑单个像素比较结果的偶然性，基于patch进行像素处理，对patch内部像素进行投票决策和区域融合渲染绘制。

## 1 算法基本原理

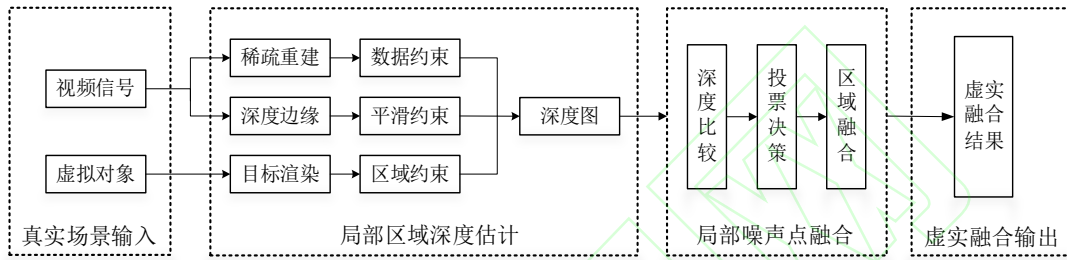


图 1 算法实现流程图

Fig. 1 Algorithm implementation flow chart

### 1.2 基于局部区域的深度估计

本文基于 Holynski<sup>[17]</sup>的快速深度致密化算法实现局部区域的深度数据恢复。原始算法要恢复出完整的深度图信息，计算量大效率低，无法做到实时融合。因此，本文提出基于局部区域的深度估计模块来提高实时性。模块的总体设计思路如图 2 所示。模块设计思路主要分为三个步骤：

1) 输入准备：将视频信号和虚拟对象作为输入。

2) 局部区域深度估计：通过给稀疏 3D 点施加三种不同种类的约束，限制深度数据的传播范围，快速恢复出局部区域的深度图。其中数据约束是利用 DSO(Direct Sparse Odometry)技术跟踪并稀疏重建的 3D 点信息；平滑约束是通过图像对齐并定位的深度边缘信息；区域约束是通过渲染虚拟目标生成特定区域的二值图像信息。

3) 输出结果：输出场景几何中特定目标区域的深度图。

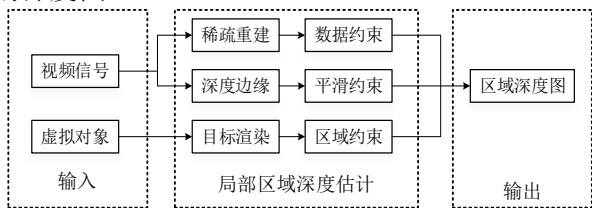


图 2 局部区域深度估计模块设计图

Fig. 2 Design drawing of local area depth estimation module

### 1.1 算法流程

算法实现的具体流程如图 1 所示。真实场景视频信息和虚拟对象作为输入，局部区域深度估计模块通过稀疏重建得到稀疏 3D 点的深度信息；再叠加不同的约束共同限制深度的传播，从而快速恢复出虚拟对象区域的深度图。局部噪声点融合模块通过深度比较确定虚实物体的前后位置关系，基于 patch 进行投票决策和区域融合，最终输出具有真实遮挡关系的融合结果。

局部区域深度估计模块通过建立损失函数来描述步骤 2) 中施加的不同约束，计算最小损失函数得到深度传播的最优解，从而将局部区域深度估计任务简单描述成一个二次优化的问题进行求解。各项施加的约束和激励的任务如图 3 所示。其中数据项公式(1)(2)和平滑项公式(3)皆是 Holynski<sup>[17]</sup>方法中的公式。

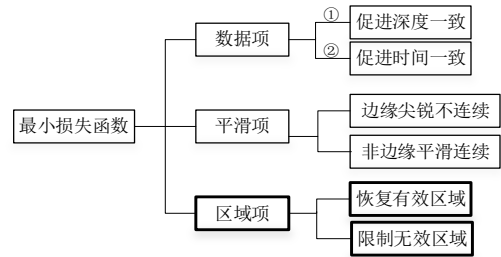


图 3 最小损失函数任务示意图

Fig. 3 Minimal loss function task diagram

图 3 中数据项①促进深度数据的一致：

$$E_{data}(p) = w_{sparse} \|D(p) - D_{sparse}(q)\|_2^2 \quad (1)$$

其中， $E_{data}$  是数据项①的损失函数， $D_{sparse}$  是跟踪的稀疏点深度。在传播过程中，如果待传播的像素  $D(p)$  与稀疏点重合则  $w_{sparse}$  设置为 1，否则为 0。

图 3 中数据项②促进时间的一致：

$$E_{temp}(p) = w_{temp}(p) \|D(p) - D_{temp}(p)\|_2^2 \quad (2)$$



其中,  $E_{temp}$  是数据项②的损失函数,  $D_{temp}$  是重投影上一帧的密集像素而获得的深度图。对于重叠重投影点的所有像素  $D(p)$ ,  $w_{temp}$  为 1, 其他地方为 0。

平滑项促使深度图在边缘地区有尖锐且明显的不连续性, 而其他地方整体平滑且连续:

$$E_{smooth}(p) = w_{pq} \|D(p) - D(q)\|_2^2 \quad (3)$$

其中,  $E_{smooth}$  是平滑项的损失函数,  $w_{pq}$  是平滑项权重。如果  $D(p)$  是边缘像素, 将权重  $w_{pq}$  设置为零, 允许边缘像素通过均值漂移不受惩罚从而形成清晰的不连续性。而其他区域像素  $D(p)$ ,  $w_{pq}$  强制执行高平滑度, 保证深度的平滑与连续。

区域项限制深度值只在有效区域内传播:

$$E_{area}(p) = w_{area} \|D(p) - D_{area}(q)\|_2^2 \quad (4)$$

其中,  $E_{area}$  是区域项的损失函数,  $D_{area}$  是虚拟物体渲染的目标区域二值图。对于重叠有效区域的所有像素点  $D(p)$ ,  $w_{area}$  是 1, 其他为 0。

得到二次优化函数为:

$$E = \arg \min_D \lambda_d \sum p E_{data}(p) + \lambda_t \sum p E_{temp}(p) + \lambda_s \sum_{(p,q) \in N} E_{smooth}(p,q) + \lambda_a \sum p E_{area}(p) \quad (5)$$

其中  $N$  是水平和垂直相邻像素的集合, 数据项、平滑项和区域项对应的平衡系数分别取 1、0.01、1 和 1。公式(5)是一个标准的泊松问题, 可以通过快速优化解算器进行求解。

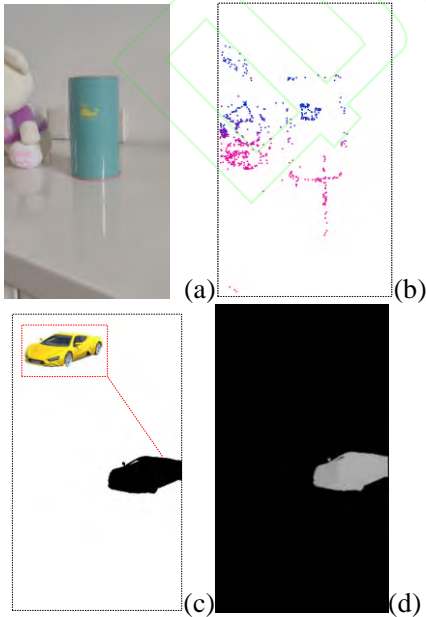


图 4 目标区域深度恢复示意图((a) 原图; (b) 稀疏 3D 点; (c) 区域二值图像; (d) 目标区域深度图)

Fig. 4 Schematic diagram of target area depth recovery((a)

Original image;(b) Sparse 3D points;(c) Regional binary image;(d) Target area depth map)

图 4 所示为区域深度估计流程示意图, 将图 4(c)中虚拟黄色小车与图 4(a)中真实存钱罐进行虚实融合处理。图 4(a)是采集的真实场景视频输入的某一帧图像; 图 4(b)是对图 4(a)进行稀疏重建得到的稀疏 3D 信息; 图 4(c)是渲染虚拟对象区域生成的二值图像, 其中黑色区域是渲染的位置信息, 空白的地方是无关区域; 图 4(d)是将图 4(c)中黑色区域部分作为致密化处理的区域约束, 从而恢复出的局部区域深度图。

### 1.3 基于投票决策的噪声点融合

基于投票决策的虚实遮挡融合处理模块的总体设计思路如图 5 所示。输入有局部错误的深度数据; 基于噪声点的投票决策融合模块通过深度比较确定区域内每个像素点虚拟物体和真实物体前后位置关系, 再利用投票决策的方法, 进行区域的融合处理; 输出有真实遮挡关系的融合图像。

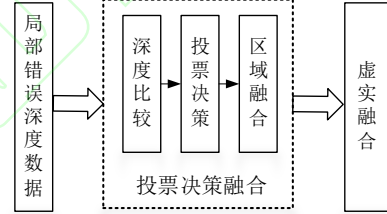
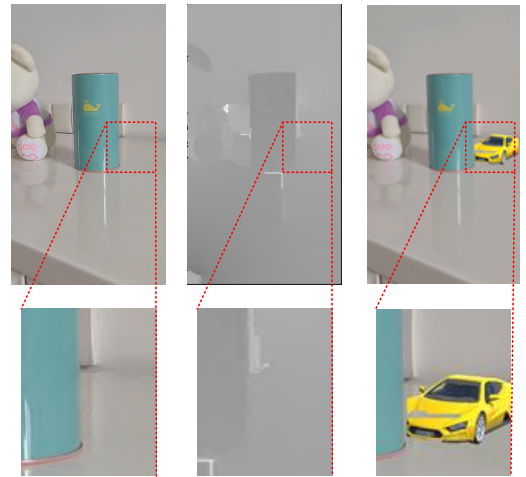


图 5 基于投票决策融合的方案示意图

Fig. 5 Scheme diagram based on voting decision fusion

#### 1.3.1 融合问题

基于 Holynski 的快速深度致密化算法在处理无纹理区域、非朗伯表面、反射和半透明表面<sup>[17]</sup>时, 会生成错误的稀疏点, 从而导致虚实融合中的孔洞和噪声现象, 如图 6 所示。



(a) 原始深度图 (b) 局部错误点 (c) 孔洞现象  
(a) Original depth map (b) Local error (c) Hole phenomenon

图6 局部错误点的融合示意图

Fig. 6 Schematic diagram of the fusion of local error points

图6(a)中插座区域存在倒影和反射,导致恢复出的深度图存在不平滑、不连续和残缺的问题,如图6(b)所示。错误的深度值会对虚实物体之间的相互遮挡产生不利的影响。如图6(c)所示,黄色小车能够被真实存钱罐挡住车头部分。但是在出现局部错误深度值的地方,融合过程中虚拟小车的车盖部分出现孔洞问题。

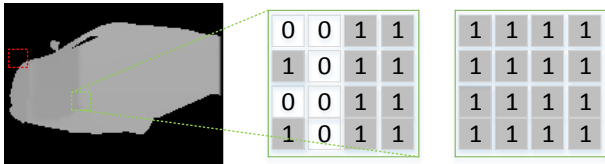
针对上述问题,本文提出一种基于 patch 的投票决策和融合的思想。由于深度图整体上是平滑连续的,错误的像素点较少且分布较稀疏。因此,本文首先对整幅图像进行分块处理,形成一系列  $n \times n$  大小的 patch; 然后对每个 patch 区域内部像素的深度值进行统计分析; 最后,采取少数服从多数的投票策略,解决局部错误点导致的融合孔洞问题,提升融合效率。

### 1.3.2 投票决策融合

投票决策融合模块主要分为三个步骤,具体实现如图5所示。

- 1) 比较每个像素点对应的真实场景和虚拟场景的深度值;
- 2) 对区域内的所有像素进行投票决策,少数像素服从多数像素;
- 3) 区域融合绘制。

比较像素点的真实深度和虚拟深度大小来描述两者的前后关系。若像素点的真实深度大于虚拟对象深度,则该像素点用 1 标志,表示绘制虚拟物体像素。否则该像素点用 0 标志,表示绘制真实场景像素。针对目标区域出现局部的错误深度信息导致的孔洞问题,融入投票决策的思想。将注意力从单个像素点转换到整个 patch,而将问题从逐点像素的比较与绘制转变为区域内所有像素的投票决策和区域绘制。



(a) 不规则区域两种情况 (b) 区域像素绘制 (c) 区域融合

(a) Two cases of irregular areas (b) Area pixel drawing (c)

Regional integration

图7 错误点投票决策融合

Fig. 7 Voting decision fusion at wrong points

基于局部区域的深度估计恢复出的深度图是不规则的,如图7(a)所示,需要执行中心规则正方形区域和边缘零散像素两种不同处理。对于规则 patch 区域,以绿色区域为例,根据深度比较关系确定每个像素的标志如图7(b)所示,利用投票决策的思想判断该区域是否绘制,因为大多数像素满足标志 1,所以该区域内的像素都标志为 1,执行都绘制虚拟对象像素的条件,区域融合结果如图7(c)所示。针对图7(a)边缘零散像素,以红色区域为例,由于边缘像素分散且稀疏,投票决策的意义不大,所以强制执行边缘像素满足最近的 patch 内像素的绘制条件,从而边缘能得到比较平滑的融合效果。综上所述,基于投票决策的噪声点融合模块能够针对具有局部错误深度数据的深度图做到良好的虚实融合效果,解决像素融合过程中的孔洞问题。

## 2 实验

### 2.1 环境与数据

在 Holynski 的算法思想基础上,本文在 Ubuntu16.04 系统平台上,结合 OpenCV 4.0 函数库开发实现区域深度数据的恢复,利用 Unity3d 2019.2.16f1 游戏引擎实现虚实融合效果。本文在光照充足的条件下采集实验数据。数据来源于非理想化的日常场景,包括透明物体、稀疏纹理区域等。

### 2.2 实验过程

实验采用普通手机摄像头(分辨率  $1920 \times 1080$ ),拍摄连续视频作为实验数据。系统视频运行速度一般保持在 30fps,视频比较流畅。利用 SLAM 跟踪场景 3D 关键点,针对快速深度致密化算法和本文算法估计出深度信息的时间进行实验对比。本文针对真实物体和虚拟物体不同空间位置关系进行噪声点数据融合,利用 Unity3d 进行虚实融合实验。

### 2.3 实验结果

#### 2.3.1 时间结果比较

表1给出了完整场景和局部目标区域深度估计运行时间比较结果。分别测试3组数据,不同视频的时长、对应的图像序列数和目标区域大小各不相同。一般场景越复杂,特定模板区域范围越大,所需要的恢复时间越长。

表1 深度估计每帧时间对比

Table 1 Depth estimation time comparison per frame

数据	快速深度致密化算法[17]	本文算法
1	12 ms	7.6 ms
2	9.6 ms	5.2 ms

从每组的完整场景和局部目标区域处理的时间可知,在特定目标区域进行深度恢复,更有利于虚实遮挡的处理。

表 2 投票决策遮挡时间对比

Table 2 Voting decision occlusion time comparison

数据	原始深度测试遮挡处理	本文投票决策算法
1	8.3 ms	8.7 ms
2	7.7 ms	8.1 ms
3	6.0 ms	6.5 ms

从每组的原始深度测试和本文投票决策虚实遮挡处理的单帧时间对比可知,利用投票决策算法时间会增加,但是差距很少,基本可以忽略不计。

2.3.2 融合效果比较

实验对比了不同大小的 patch 对虚实遮挡融合结果的影响。实验数据以单帧为例,为了准确地观察实验结果和对比,图 8(虚拟物体遮挡真实场景)和图 9(虚拟物体被遮挡)展示了完整的实验效果和对应虚拟区域的放大视图。

在虚拟对象遮挡真实场景的实验中,虚拟物体是三阶魔方模型,快速深度致密化算法得到的融合

效果如图 8(a)所示。由于真实书籍的边角处的深度值不精确,虚实融合中魔方中心区域出现孔洞问题。在虚拟物体被真实物体遮挡的实验中,虚拟物体是黄色的卡通汽车,快速深度致密化算法得到的虚实融合效果如图 9(a)所示。由于真实场景桌子的反光现象,该部分深度数据不准确,从而出现卡通汽车车盖部分虚拟像素缺失的问题。图 8 和图 9(b)-(f)分别是本文算法采用不同大小的 patch 融合得到的虚实遮挡结果。

从虚拟物体和真实场景不同的空间关系(遮挡前后)的虚实融合对比结果可知,随着选取的 patch 的增大,原始算法融合过程的孔洞越来越小。边缘的细微残缺像素也能实现平滑的绘制,与真实场景的物体更加融洽。实验表明基于 patch 的像素点投票决策融合的方法,能够较好地解决孔洞和噪声点的问题。在 patch 大小为  $6 \times 6$  时,融合效果最佳,对于真实环境和虚拟对象不同的空间关系都适用。但是 patch 过小时会导致孔洞现象加剧,一般 patch 的选取不要小于  $3 \times 3$ 。

(a) (b) (c) (d) (e) (f)

图 8 虚拟物体遮挡真实场景融合图(a)是原始快速深度致密化算法[17]的实验结果, (b)-(f)是不同 patch 融合的实验结果, 其对应 patch 大小分别为 (b)  $2 \times 2$  (c)  $3 \times 3$  (d)  $4 \times 4$  (e)  $5 \times 5$  (f)  $6 \times 6$

Fig. 8 Virtual object occludes real scene fusion map (a) is the experimental result of the original fast deep densification algorithm [17], (b)-(f) are the experimental results of the fusion of different patches, and the corresponding patch sizes are

(b)  $2 \times 2$  (c)  $3 \times 3$  (d)  $4 \times 4$  (e)  $5 \times 5$  (f)  $6 \times 6$

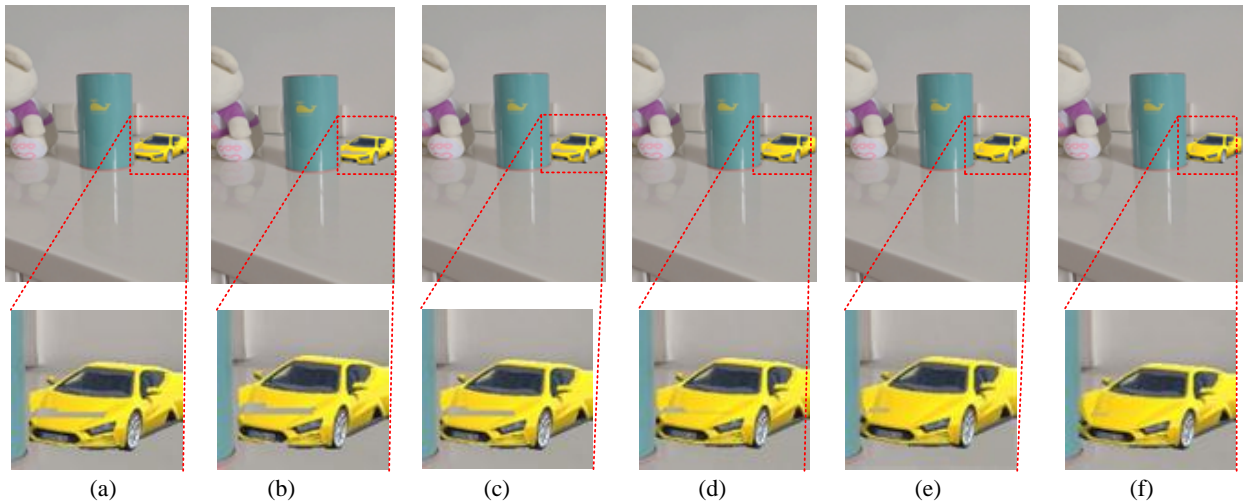




图9 虚拟物体被遮挡融合图(a)是原始快速深度致密化算法[17]的实验结果, (b)-(f)是不同 patch 融合的实验结果, 其对应 patch 大小分别为 (b)  $2 \times 2$  (c)  $3 \times 3$  (d)  $4 \times 4$  (e)  $5 \times 5$  (f)  $6 \times 6$

**Fig. 9** Fusion image of virtual objects being occluded (a) is the experimental result of the original fast deep densification algorithm [17], (b)-(f) are the experimental results of the fusion of different patches, and the corresponding patch sizes are (b)  $2 \times 2$  (c)  $3 \times 3$  (d)  $4 \times 4$  (e)  $5 \times 5$  (f)  $6 \times 6$

### 3结论

本文提出一种基于 patch 投票决策的实时遮挡处理技术, 解决遮挡处理的实时性和深度数据精确性两个问题。算法能够做到实时的遮挡融合处理, 同时解决不精确的深度数据导致的噪声点和孔洞问题。

本文算法是建立在深度估计算法所得深度图中错误像素点较少且分布较稀疏的基础上。如果出现大范围的错误深度数据, 就无法正确处理融合中的孔洞问题。此外, 针对错误的稀疏深度值, 研究如何从根源出发纠正稀疏重建的错误深度信息, 是本文继续要研究的问题。

#### 参考文献(References)

- [1] 朱淼良, 姚远, 蒋云良. 增强现实综述[J]. 中国图象图形学报, 2004, 9(007):767-774.  
ZHU M L, YAO Y, JIANG Y L. Augmented reality review[J]. Chinese Journal of Image and Graphics, 2004, 9(007):767-774 (in Chinese).
- [2] Whitaker R T, Crampton C, Breen D E, et al. Object calibration for augmented reality[C]//Computer Graphics Forum. Edinburgh, UK: Blackwell Science Ltd, 1995, 14(3): 15-27.
- [3] 李腾飞. 增强现实虚实交互技术的研究与实现[D]. 辽宁: 沈阳工业大学, 2019.  
LI T F. Research and implementation of augmented reality interactive technology[D]. Liaoning: Shenyang University of Technology, 2019 (in Chinese).
- [4] Lepetit V, Berger M O. A semi-automatic method for resolving occlusion in augmented reality[C]//Proceedings IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2000, 2: 225-230.
- [5] 张岩, 何晓波, 管涛. 基于多视图几何的位置与遮挡一致性恢复方法研究[J]. 图学学报, 2011(1):111-118.  
ZHANG Y, HE X B, GUAN T. Research on the restoration method of position and occlusion consistency based on multi-view geometry[J]. Journal of Graphics, 2011(1):111-118 (in Chinese).
- [6] Breen D E, Whitaker R T, Rose E, et al. Interactive occlusion and automatic object placement for augmented reality [J]. Computer Graphics Forum (S0167-7055), 1996, 15(3): C-11-C-22.
- [7] 冒佳仪. 基于深度图像的虚实遮挡处理系统[D].湖北: 华中科技大学, 2019.  
MAO J Y. Virtual and real occlusion processing system based on depth image[D]. Hubei: Huazhong University of Science and Technology, 2019 (in Chinese).
- [8] Kim H, Yang S J, Sohn K. 3D reconstruction of stereo images for interaction between real and virtual worlds[C]//Proceedings of IEEE & ACM International Symposium on Mixed & Augmented Reality, Tokyo: IEEE, 2003:169-176.
- [9] Valentin J, Kowdle A, Barron J T, et al. Depth from motion for smartphone AR[J]. ACM Transactions on Graphics, 2018, 37(6): 1-19.
- [10] Valentini P P. Natural interface in augmented reality interactive simulations[J]. Virtual and Physical Prototyping, 2012, 7(2): 137-151.
- [11] Du C, Chen Y L, Ye M, et al. Edge snapping-based depth enhancement for dynamic occlusion handling in augmented reality[C]//2016 IEEE International Symposium on Mixed and Augmented Reality. Mexico: IEEE, 2016: 54-62.
- [12] 吕浩, 陈世峰. 基于多方向滤波的强边缘深度图像补全方法[J]. 集成技术, 2016, 5(6):36-44.  
LV H, CHEN S F. A strong edge depth image complement method based on multi-directional filtering[J]. Integrated Technology, 2016, 5(6):36-44 (in Chinese).
- [13] Mal F, Karaman S. Sparse-to-dense: Depth prediction from sparse depth samples and a single image[C]//2018 IEEE International Conference on Robotics and Automation.Brisbane: IEEE, 2018: 1-8.
- [14] Chen Z, Badrinarayanan V, Drozdov G, et al. Estimating depth from rgb and sparse sensing[C]//Proceedings of the European Conference on Computer Vision.Berlin: Springer, 2018: 167-182.
- [15] Xu Y, Zhu X, Shi J, et al. Depth completion from sparse lidar data with depth-normal constraints[C]//Proceedings of the IEEE International Conference on Computer Vision.Seoul: IEEE, 2019: 2811-2820.
- [16] Stühmer J, Gumhold S, Cremers D. Real-time dense geometry from a handheld camera[C]//Joint Pattern Recognition Symposium. Springer, Berlin, Heidelberg, 2010: 11-20.



- 
- [17] Holynski A, Kopf J. Fast depth densification for occlusion-aware augmented reality[J]. *ACM Transactions on Graphics*, 2018, 37(6): 1-11.
- [18] Weerasekera C S, Dharmasiri T, Garg R, et al. Just-in-time reconstruction: Inpainting sparse maps using single view depth predictors as priors[C]//2018 IEEE International Conference on Robotics and Automation. Brisbane: IEEE, 2018: 1-9.
- [19] Roxas M, Hori T, Fukiage T, et al. Occlusion handling using semantic segmentation and visibility-based rendering for mixed reality[C]//Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology. New York: ACM Press, 2018: 1-8.
- [20] Hebborn A K, Hühner N, Müller S. Occlusion matting: realistic occlusion handling for augmented reality applications[C]//2017 IEEE International Symposium on Mixed and Augmented Reality. Nantes: IEEE, 2017: 62-67.