

基于深度逐平面扫描的立体体积构建方法

关珍博^{1,2,3}, 姬红兵³, 段同乐^{1,2}, 梁海清^{1,2}

(1. 中国电子科技集团公司 第五十四研究所, 石家庄 050081;

2. 河北省智能化信息感知与处理重点实验室, 石家庄 050081;

3. 西安电子科技大学 电子工程学院, 西安 710068)

摘要: 针对多视图三维重建中立体体积构建对模型完整性与精度的决定性作用, 以及现有基于平面扫描的深度估计方法在弱纹理区域精度不足、多视图融合冗余较高的问题, 开展改进立体体积构建方法研究; 采用改进的“深度逐平面扫描”方法, 在平面扫描深度估计框架下, 通过构建与参考图像相机前平面平行的虚拟平面聚合多视图特征形成稀疏高效的平面扫描体积; 结合多尺度特征提取与上下文感知代价聚合获取高精度深度图, 采用改进粒子群优化的 ICP 配准与熵驱动超体素融合实现多视图点云一致立体体积重建; 经多个公开数据集和实际场景实验测试, 在 ETH3D 数据集上深度估计 RMSE 降至 4.21 mm、弱纹理区域准确率提升至 90.2%, 在 Scenes11 数据集上体积冗余度降低至 7.8%, 在城市建筑场景中实现 28 fps 的实时深度估计, 相较于传统平面扫描和 DPSNet 等方法, 在处理高分辨率输入和无纹理区域时鲁棒性更强。该方法可满足城市建筑等实际场景的多视图三维重建应用需求, 兼顾重建精度与实时性。

关键词: 多视图立体; 深度估计; 平面扫描体积; 深度学习; 三维重建; 代价体

Stereo Volume Reconstruction Method Based on Depth-by-Plane Scanning

GUAN Zhenbo^{1,2,3}, JI Hongbing³, DUAN Tongle^{1,2}, LIANG Haitao^{1,2}

(1. The 54th Research Institute of CETC, Shijiazhuang 050081, China;

2. Hebei Key Laboratory of Intelligent Information Perception and Processing, Shijiazhuang 050081, China;

3. School of Electronic Engineering, Xidian University, Xi'an 710068, China)

Abstract: In response to the decisive role of stereo volume construction in the completeness and accuracy of models in multi-view 3D reconstruction, as well as the problems of insufficient accuracy in low-texture regions and high redundancy in multi-view fusion of existing plane sweep-based depth estimation methods, research on an improved stereo volume construction method was carried out, with an improved “depth-wise plane sweeping” method presented. Under the plane sweep depth estimation framework, a series of virtual planes parallel to the front plane of the reference image camera were constructed to aggregate multi-view features, forming a sparse and efficient plane sweep volume. Combined with multi-scale feature extraction and context-aware cost aggregation, it was possible to obtain high-precision depth maps. and the improved particle swarm optimization-based ICP registration and entropy-driven supervoxel fusion were used to reconstruct a 3D volume consistent with multi-view point clouds. Experimental tests on multiple public datasets and real-world scenarios show that on the ETH3D dataset, the depth estimation RMSE is reduced to 4.21 mm and the accuracy in low-texture regions is increased to 90.2%, on the Scenes11 dataset, the volume redundancy is reduced to 7.8%, and the real-time depth estimation is 28 fps in urban building scenarios. Compared with the traditional plane sweeping and DPSNet and other methods, this method has a stronger robustness in processing high-resolution inputs and textureless regions, which meets the application requirements of multi-view 3D reconstruction in real-world scenarios such as urban buildings, making a balance between a reconstruction accuracy and real-time performance.

Keywords: multi-view stereo; depth estimation; plane sweep volume; deep learning; 3D reconstruction; cost volume

收稿日期: 2025-11-25; 修回日期: 2026-01-04。

作者简介: 关珍博(1984-), 男, 硕士, 高级工程师。

引用格式: 关珍博, 姬红兵, 段同乐, 等. 基于深度逐平面扫描的立体体积构建方法[J]. 计算机测量与控制, 2026, 34(6): 244-250.

0 引言

在多视图三维重建中, 立体体积构建已成为连接二维图像与三维物理世界的关键桥梁。在城市数字化转型中, 高精度三维城市模型为智能交通、应急救援提供空间决策支持^[1]; 在生物医学领域, 实时三维细胞成像助力发育生物学研究捕捉动态生理过程^[2]; 在工业制造中, 三维体积检测实现精密零部件的缺陷识别与质量控制^[3]。按照遥感数据获取方式可将传统立体体积构建方法划分为被动式与主动式两类。主动式方法如激光雷达通过发射信号获取深度信息, 具有测距精度高的优点, 但设备成本昂贵且数据采集效率有限^[4]; 被动式方法基于多视图图像恢复三维结构, 具有成本低、易部署等优势, 在许多实际系统中成为主流^[5]。平面扫描 (PS, plane-sweeping) 算法是被动式多视图立体重建中一种典型且广泛应用的技术, 利用在深度范围内构建一系列平行扫描平面并计算光度一致性, 可在 GPU (Graphics Processing Unit) 加速下实现实时深度估计^[6]。

随着科技进步, 国内外平面扫描技术飞速发展, 早期采用单一扫描方向与简单的代价计算方式, 在纹理丰富区域表现良好, 但在弱纹理与遮挡区域精度骤降^[7]。随着并行计算技术的发展, GPU 加速成为平面扫描算法实现实时性的关键^[8], 有效降低了光照变化对匹配精度的影响^[9]。深度学习技术的兴起为平面扫描算法注入了新的活力, 通过构建多尺度代价体实现端到端深度估计^[10], 在多视图立体体积构建中显著提升了重建精度^[11]。

深度逐平面扫描技术通过引入深度学习与多方向扫描策略^[12]。不同于传统算法依赖手工设计特征的局限, 深度逐平面扫描结合神经网络提取高层语义特征^[13], 显著提升了在弱纹理和光照变化等挑战性场景下的配鲁棒性^[14], 然而现有深度学习方法在代价体构建的判别力、多视图信息融合的效率等方面仍有提升空间^[15]。因此, 研究如何进一步优化深度特征表达与多视图融合策略, 以实现更高精度、更高效率的立体体积构建方法^[16], 对推动三维感知技术的实用化进程具有重要的理论价值与工程意义^[17]。本文针对传统平面扫描算法及现有基于平面扫描的深度估计方法在弱纹理鲁棒性、多视图融合冗余和配准精度等方面的不足^[18], 系统研究深度逐平面扫描的立体体积构建关键技术。

1 深度逐平面扫描的立体体积构建算法基本原理

1.1 平面扫描算法基本原理

平面扫描算法的核心思想是通过在三维空间中扫描一系列平行平面, 利用多视图图像的光度一致性确定物

体表面的深度信息。

1.1.1 单应矩阵映射

扫描平面与图像点的对应关系通过单应矩阵描述。约定世界坐标系与参考视图的理论参考相机坐标系重合, 则参考视图投影矩阵如式 (1) 所示:

$$\mathbf{P}_{\text{ref}} = \mathbf{K}_{\text{ref}} [\mathbf{I}_{3 \times 3} \mid \mathbf{0}_{3 \times 1}] \quad (1)$$

其中: $\mathbf{K}_{\text{ref}}$ 为相机内参矩阵; 第 k 个目标视图的投影矩阵如式 (2) 所示:

$$\mathbf{P}_k = \mathbf{K}_k [\mathbf{R}_k^T \mid -\mathbf{R}_k^T \mathbf{C}_k] \quad (2)$$

其中: $\mathbf{K}_k$ 为第 k 个相机的内参矩阵, $\mathbf{R}_k$ 、 $\mathbf{C}_k$ 分别为 (从相机坐标系到世界坐标系的) 旋转矩阵与相机光心在世界坐标系中的坐标。对于扫描平面 ($\mathbf{n}$ 为平面向量, d 为平面距离), 参考视图中的点 x 与第 k 个视图中的对应点 x' 满足 $x' \sim \mathbf{H}_k x$, 其中对应矩阵 $\mathbf{H}_k$ 的计算如式 (3) 所示:

$$\mathbf{H}_k = \mathbf{K}_k \left(\mathbf{R}_k^T - \frac{\mathbf{R}_k^T \mathbf{C}_k \mathbf{n}^T}{d} \right) \mathbf{K}_{\text{ref}}^{-1} \quad (3)$$

当两个视图的投影中心重合时成立, 此时两个视图间只存在摄影角度的差异, 通过单应矩阵可实现不同视图间的像素映射, 为光度一致性计算提供基础。

1.1.2 代价函数构建

传统平面扫描算法采用固定窗口的光度一致性度量作为代价函数, 考虑到光照变化的影响, 引入增益因子 β 对像素强度差异进行简单补偿, 其像素级代价可表示为式 (4) 所示:

$$\text{Cost}(x, d) = \sum_{(u, v) \in W(x)} |I_{\text{ref}}(u, v) - \beta I_k[\mathbf{H}_k(u, v, d)]| \quad (4)$$

其中: $W(x)$ 表示参考视图以像素 x 为中心的局部窗口, I_{ref} 、 I_k 分别为参考视图与第 k 个目标视图的像素强度, $\mathbf{H}_k(u, v, d)$ 表示在深度假设 d 下将参考视图像素 (u, v) 映射到第 k 个目标视图坐标系的单应变换结果。为提升鲁棒性, 研究者们基于零均值归一化互相关 (ZNCC, zero-mean normalized cross-correlation) 的相似性度量, 通过归一化消除光照均值差异如式 (5) 所示:

$$\text{ZNCC}(x, d) = \frac{\sum_{(u, v) \in W(x)} [I_{\text{ref}}(u, v) - \mu_{\text{ref}}] \{I_k[\mathbf{H}_k(u, v, d)] - \mu_k\}}{\sqrt{\sum_{(u, v) \in W(x)} (I_{\text{ref}}(u, v) - \mu_{\text{ref}})^2} \sqrt{\sum_{(u, v) \in W(x)} \{I_k[\mathbf{H}_k(u, v, d)] - \mu_k\}^2}} \quad (5)$$

其中: μ_{ref} 、 μ_k 分别为窗口 $W(x)$ 内参考视图和第 k 个目标视图像素强度均值。尽管上述方法在一定程度上缓解了光照变化的影响, 然而这类基于局部灰度统计的相似性度量在弱纹理或近似恒亮区域对视差变化的区分

能力有限,多个深度假设往往产生相近响应,易导致匹配歧义。

1.1.3 传统平面扫描到三维空间的转换

平面扫描定义于相机视锥体空间,坐标表示为 (u, v, d) , 其中 (u, v) 为图像像素坐标, d 为深度值。为适配三维场景理解任务,需将其转换至三维世界空间,坐标表示为 (x, y, z) 。转换过程基于相机内参矩阵实现,依据相机投影逆运算,公式如下:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} d = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (6)$$

其中: f_x 、 f_y 分别为水平和垂直焦距, c_x 、 c_y 为图像主点坐标。该转换过程通过三线性插值实现可微分,确保深度估计与后续三维任务的联合训练。

1.2 深度学习特征提取技术

为了解决传统代价函数依赖局部特征、在弱纹理等区域鲁棒性差的问题,采用深度卷积神经网络(CNN, convolutional neural network)来提取高维度的、更具语义信息的特征表示^[19]。具体而言,输入图像通过一个特征提取网络(如 U-Net、FPN 等)生成特征图(Feature Map)。在构建代价体时,不再是直接计算像素强度的差异,而是计算参考视图特征图与经过单应性变换后的源视图特征图之间,在特征向量层面的相似度。

为提升特征的鲁棒性与表达能力,多尺度特征提取通过融合不同卷积层的输出,兼顾特征的细节信息与上下文语义,常用的实现方式包括特征金字塔网络(FPN, feature pyramid network)与空间金字塔池化(SPP)。SPP 模块通过多个固定大小的平均池化操作,将任意尺寸的特征图转换为固定长度的特征向量,有效扩大了感受野。DPSNet 采用 4 级 SPP 模块(16×16 、 8×8 、 4×4 、 2×2),将不同尺度的上下文信息上采样至原始特征图尺寸后进行拼接,生成 32 通道的特征表示。残差连接技术通过跳跃连接缓解深层网络的梯度消失问题,使模型能够训练更深的网络结构,提升特征提取能力^[20]。

2 基于深度逐平面扫描的立体体积构建方法

2.1 方法整体框架

本文提出的立体体积构建方法主要包括如下:

1) 预处理模块:对输入多视图图像进行畸变校正、内参标定与视角筛选,剔除模糊与过度重叠的图像,减少冗余计算;

2) 深度估计模块:通过多尺度 CNN 提取特征,构建 4D 代价体并进行上下文感知聚合,采用深度回归生成参考视图的高精度深度图;

3) 点云处理模块:将深度图与相机内参结合生成点云,通过改进 PSO-ICP 算法实现多视图点云精配准,去除异常值与噪声;

4) 体积融合模块:基于超体素分割进行密度决策,采用可信度加权策略融合多视图点云,生成最终的立体体积模型。

2.2 多尺度特征增强的代价体构建

2.2.1 深层特征提取网络

针对传统算法手工特征鲁棒性不足的问题,参考 DPSNet 的特征提取模块,设计多尺度特征提取网络。网络输入为参考视图与目标视图的 RGB 图像,开始通过 7 层卷积进行特征编码,第一层采用 7×7 卷积核捕捉大范围边缘特征,步长为 2;后续 6 层采用 3×3 卷积核,其中第 3 层步长设为 2,其余各层步长为 1,因此编码器输出特征图的空间分辨率约为输入图像分辨率的 $1/4$,在减少显存占用的同时保留了足够的空间细节。每层均配备批归一化(BN, batch normalization)与 ReLU 激活函数。为避免深层网络梯度消失,在第 3~4 层与第 5~6 层之间加入残差连接,通过跳跃连接直接传递浅层特征,使浅层纹理信息能够有效传递至高层,一方面缓解深层网络的梯度消失问题,另一方面在弱纹理区域保留更多边缘与结构信息,为后续代价体构建提供更稳定的特征基础。

2.2.2 4D 代价体构建策略

传统平面扫描算法在逆深度空间均匀采样生成扫描平面,设最小深度为 $d_{\min}$,深度标签总数为 L ,则第 l 个扫描平面的逆深度如式(7)所示:

$$\frac{1}{d_l} = \frac{1}{d_{\min}} + \frac{l}{L-1} \left(\frac{1}{d_{\max}} - \frac{1}{d_{\min}} \right) \quad (7)$$

其中: $d_{\max}$ 为最大深度。对于每个扫描平面,通过单应矩阵将目标视图特征图映射至参考视图坐标系,得到形变特征图。不同于 MVSNet 采用特征差值构建代价体的方式,本文采用特征连接策略构建 4D 代价体。将参考视图特征图与形变后的目标视图特征图在通道维度拼接,形成尺寸为 $W \times H \times 2C \times L$ 的 4D 特征体(W 、 H 为特征图宽高, C 为特征通道数)。这种方式保留了两路特征的完整信息,避免了差值运算导致的特征损失,尤其在弱纹理区域,能够通过特征分布差异实现可靠匹配。

2.3 上下文感知的代价聚合模块

代价聚合的核心目标是通过正则化操作降低噪声干扰,增强代价体的判别力。传统聚合方法如双边滤波依赖人工设计的滤波核,适应性有限。本文设计上下文感知的代价聚合模块,利用空洞卷积的多感受野特性与残差学习,实现代价体的自适应优化。该设计基于深度场

在局部空间近似平滑、在物体尺度分段连续的先验假设,通过引入多尺度空间上下文对代价体进行数据驱动的正则化。

2.3.1 多尺度空洞卷积分合

模块输入为4D代价体,先通过 $3 \times 3 \times 3$ 的3D卷积将通道数压缩至64后,接入多尺度空洞卷积层。考虑到深度推断既需要保持局部细节的平滑性,又需要建模物体尺度的长距离依赖关系,本文采用按2的幂次递增的空洞率策略来构建多尺度感受野。具体而言,该层包含6个并行的 3×3 卷积核,空洞率分别设置为1、2、4、8、16、32。经计算,各分支对应的有效感受野分别为 3×3 、 5×5 、 9×9 、 17×17 、 33×33 和 65×65 ,这种配置使得感受野在对数尺度上近似均匀分布,实现了从像素级局部邻域到物体级区域结构的全面覆盖。相较于线性等间隔或随机设置,指数型空洞率序列有效避免了不同尺度间的感受野空缺,并通过多分支特征叠加,有效缓解了单一路径大空洞率可能引发的“栅格效应”。6个分支的输出首先在通道维度进行拼接,得到包含丰富多尺度信息的特征体;随后通过 $1 \times 1 \times 1$ 的3D卷积进行自适应通道加权与融合,压缩通道数至32,实现多尺度特征融合。这一步不仅降低了计算冗余,更实现了不同尺度上下文信息的显式交互与特征筛选。

2.4 深度回归与点云生成

2.4.1 概率加权的深度回归

对优化后的代价体进行softmax运算,得到每个像素在不同深度标签上的概率分布如式(8)所示:

$$P(l | x, y) = \sigma[Cost(x, y, l)] = \frac{e^{Cost(x, y, l)}}{\sum_{l'=0}^{L-1} e^{Cost(x, y, l')}} \quad (8)$$

其中: $Cost(x, y, l)$ 为像素 (x, y) 在第 l 个深度标签上的代价。采用概率加权求和计算预测深度标签如式(9)所示:

$$\hat{l} = \sum_{l=0}^{L-1} l \cdot P(l | x, y) \quad (9)$$

根据深度标签与实际深度的映射关系,计算像素的最终深度值如式(10)所示:

$$d(x, y) = \frac{1}{\frac{1}{d_{\min}} + \frac{l}{L-1} \left(\frac{1}{d_{\max}} - \frac{1}{d_{\min}} \right)} \quad (10)$$

为提升深度估计精度,采用Huber损失函数进行模型训练,平衡异常值与正常样本的损失贡献如式(11)所示:

$$Loss = (1 - \lambda) \|d - d_{gt}\|_2^2 + \lambda L_{\text{Huber}}(d - d_{gt}) \quad (11)$$

其中: d_{gt} 为真实深度, λ 设置为0.7, L_{Huber} 为Huber范数,当误差绝对值小于 δ 时为平方损失,否则为线性损失(δ 取1.0)。

3.4.2 点云生成与预处理

机内参将深度图转换为三维点云,设像素 (x, y) 的深度为 d ,相机内参矩阵如式(12)所示:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

则对应的三维坐标如式(13)所示:

$$X = \frac{(x - c_x)d}{f_x}, Y = \frac{(y - c_y)d}{f_y}, Z = d \quad (13)$$

生成点云后进行预处理:通过统计滤波去除噪声点,设置邻域点数量阈值为15,当某点的邻域点数量小于阈值时判定为噪声;采用半径滤波去除孤立点,设置半径阈值为0.02 m,删除半径范围内无其他点的孤立点;通过下采样减少点云数量,采用体素网格下采样,网格大小设置为0.01 m,在保证几何特征的前提下降低后续配准计算量。

2.5 改进 PSO-ICP 的点云配准

2.5.1 粗配准与初始位姿估计

采用RANSAC算法结合采样一致性初始配准(SAC-IA)进行粗配准。SAC-IA通过随机采样源点云中的3个非共线点,生成局部特征描述子,在目标点云中寻找匹配的特征点对,计算初始变换矩阵。为提升鲁棒性,设置迭代次数为1000,内点阈值为0.05 m,当内点比例超过60%时停止迭代,输出粗配准结果。粗配准能够将源点云与目标点云的重合度提升至70%以上,为精配准提供良好的初始位姿。在此基础上,以粗配准结果为中心构建粒子群搜索空间,并对旋转,和平移参数施加边界约束,在避免无效全局遍历的同时确保了足够的搜索裕度。

2.5.2 改进 PSO 优化的 ICP 精配准

针对传统ICP易陷入局部最优的问题,提出改进粒子群优化(IPSO)的ICP算法,IPSO算法通过动态调整加速度因子与惯性权重,平衡全局搜索与局部收敛能力:

1) 位姿参数化与粒子初始化:将刚体变换表示为6维参数向量 $\mathbf{x} = [\theta_x, \theta_y, \theta_z, t_x, t_y, t_z]^T$, 其中: θ 为欧拉角, t 为平移分量。以3.5.1中粗配准结果为中心,在预设的约束范围内随机初始化粒子群的位置 x_i^0 与速度 v_i^0 。设定种群规模 $N=30$,最大迭代次数 $T=50$,并对每一维搜索空间实施边界限制,防止粒子出现不合理的位移。

2) 动态惯性权重:采用线性递减策略,初始权重为0.9,迭代结束时降至0.4,如式(14)所示:

$$w = w_{\max} - (w_{\max} - w_{\min}) \frac{t}{T} \quad (14)$$

其中: t 为当前迭代次数, T 为总迭代次数,在迭

代初期较大的 w 有助于保持粒子运动的惯性, 增强全局探索能力, 而在迭代后期较小的 w 则促使粒子在当前最优解附近细化搜索, 从而实现由全局搜索向局部收敛的平滑过渡。

3) 自适应加速度因子: 根据粒子适应度值动态调整认知因子 c_1 与社会因子 c_2 。对于适应度较高的粒子, 增大 c_1 (上限 1.8) 并减小 c_2 , 使其侧重于利用自身经验在当前邻域内进行精细化搜索, 以挖掘局部极值; 对于适应度较低的粒子增大 c_2 (上限 1.8) 并减小 c_1 , 增强其向群体最优位置的跟随趋势, 从而加速脱离非优区域。通过这种差异化的参数调节策略, 算法能够在保证种群多样性的同时, 有效兼顾了全局搜索广度与局部收敛精度。

4) 适应度函数设计与归一化: 为同时考虑几何误差与特征一致性, 采用点云配准的均方误差与特征一致性的归一化加权和作为粒子适应度。首先将均方根配准误差 RMSE 映射到区间, 公式如式 (15) 所示:

$$\overline{RMSE} = \frac{RMSE - RMSE_{\min}}{RMSE_{\max} - RMSE_{\min}} \quad (15)$$

其中: $RMSE_{\min}$ 、 $RMSE_{\max}$ 由预实验统计得到。然后构造最终适应度函数, 如式 (16) 所示:

$$fitness = \alpha RMSE + (1 - \alpha) F \quad (16)$$

其中: $\alpha = 0.7$, F 为特征匹配一致性分数。PSO 以最小化 $fitness$ 为目标进行优化, 从而在几何误差与特征一致性之间实现可控的权衡。相比于直接将 RMSE 与 F 线性相加, 归一化处理消除了量纲和数值尺度差异, 使得权重 α 具有明确的物理含义: 当 α 较大时, 优化更偏向几何精度, 当 α 较小时, 则更强调特征一致性。

IPSO 算法优化得到最优初始变换矩阵后, 代入 ICP 算法进行精配准。精配准迭代过程中, 采用迭代重加权最小二乘法 (IRLS) 求解变换矩阵, 对异常值赋予较小权重, 进一步提升配准精度。与直接从粗配准结果启动的 ICP 相比, IPSO-ICP 先利用粒子群在较大位姿搜索空间中获得更接近全局最优的初始解, 再由局部线性化的 ICP+IRLS 快速收敛, 两者在搜索范围与收敛速度方面形成互补。实验表明, 相较于带 RANSAC 粗配准的传统 ICP 算法 (迭代次数 45, 内点阈值 0.05 m), 本文改进的 IPSO-ICP 算法的 RMSE 降低 28.3%, 收敛速度提升 40%。该方法通过动态参数调整实现了全局最优搜索, 避免了传统 ICP 的局部最优问题, 同时 IRLS 算法降低了异常值的干扰。

3 实验与结果分析

3.1 实验环境与数据集

3.1.1 实验环境

硬件环境: Intel Core i9-12 900 K CPU, NVIDIA

RTX 4090 GPU (24 GB 显存), 64 GB DDR4 内存; 软件环境: Ubuntu 22.04, PyTorch 2.0, OpenCV 4.8, PCL 1.13, CUDA 12.1。

3.1.2 数据集

1) 公开数据集: 选用 ETH3D、SUN3D 与 Scenes11 数据集。ETH3D 包含室内外多种场景, 提供高分辨率图像与精确的深度真值, 用于深度估计精度评估^[21]; SUN3D 包含 106 个室内场景, 具有丰富的纹理变化, 用于鲁棒性测试; Scenes11 包含 11 类典型场景, 用于体积重建完整性评估。

2) 实际场景数据集: 采集城市建筑与工业零部件场景数据。城市建筑场景采用 8 台同步相机组成采集系统, 分辨率为 2048×1536 , 拍摄 100 个视角; 工业零部件场景采用 4 台相机, 分辨率为 1280×960 , 拍摄 50 个视角, 人工标注深度真值用于精度验证。

3.2 评价指标

1) 深度估计指标: 采用均方根误差 (RMSE)、平均绝对误差 (MAE) 与准确率 (Acc)。Acc 定义为深度误差小于阈值 (1.5 mm) 的像素比例。

2) 点云配准指标: 采用配准误差 (RE) 与收敛时间 (T)。RE 为配准后对应点对的平均欧氏距离。

3) 体积重建指标: 采用体积精度 (VP)、完整性 (CO) 与冗余度 (RD)。VP 为重建体积与真值的重合比例; CO 为重建体积包含真值的比例; RD 为冗余点占总点云的比例。

4) 实时性指标: 采用深度估计帧率 (FPS) 与体积重建总耗时 (TT)。

3.3 对比实验设计

设置 4 组对比方法:

1) 传统平面扫描算法 (PS): 采用 ZNCC 代价函数, 单一扫描方向;

2) 多方向平面扫描算法 (MPS): 基于《Real-time Plane-sweeping Stereo with Multiple Sweeping Directions》实现;

3) DPSNet 算法: 基于 ICLR 2019 论文实现, 端到端深度估计;

4) 本文方法 (Ours): 融合多尺度特征与自适应融合的深度逐平面扫描方法。

所有方法采用相同的相机内参与数据集, 公平对比性能差异。

3.4 深度估计性能分析

3.4.1 公开数据集结果

ETH3D 数据集上的深度估计结果如表 1 所示。本文方法的 RMSE 为 4.21 mm, 较 PS 算法降低 37.2%, 较 DPSNet 降低 18.6%; MAE 为 2.85 mm, 较 PS 算

法降低 41.3%； Acc 达到 94.7%，较 MPS 算法提升 12.4%。在弱纹理区域（如白墙、桌面），本文方法的 Acc 提升更为显著，达到 90.2%，而 PS 算法仅为 62.8%，这一对比直观地反映出，传统方法单纯依赖局部光度一致性在弱纹理处缺乏有效约束，容易导致匹配失效。相比之下，本文引入的多尺度特征增强与上下文感知聚合机制有效引入了广域结构与语义上下文，SPP 模块对物体轮廓与区域特征的显式编码，配合上下文网络的边缘敏感正则化，在保持深度不连续处锐利度的同时，显著抑制了弱纹理区域的匹配歧义，提升了整体估计的鲁棒性。

表 1 ETH3D 数据集深度估计结果对比

方法	$RMSE/mm$	MAE/mm	$Acc/\%$	弱纹理 $Acc/\%$
PS	6.70	4.86	78.3	62.8
MPS	5.42	3.79	84.3	75.1
DPSNet	5.17	3.38	90.5	82.6
Ours	4.21	2.85	94.7	90.2

SUN3D 数据集上的鲁棒性测试结果如表 2 所示。在光照变化场景中，本文方法的 $RMSE$ 为 5.13 mm，较 DPSNet 降低 16.9%；在遮挡场景中， Acc 达到 92.3%，较 MPS 算法提升 10.7%。这一性能突破主要归功于上下文感知聚合模块对全局几何信息的充分利用。该模块通过在代价体中引入多尺度的空间正则化，有效校正了由局部光照突变或遮挡引发的匹配歧义。

表 2 SUN3D 数据集鲁棒性测试结果

场景类型	方法	$RMSE/mm$	$Acc/\%$
光照变化	PS	8.92	68.5
	MPS	7.15	79.2
	DPSNet	6.17	88.4
	Ours	5.13	93.6
遮挡场景	PS	7.84	70.1
	MPS	6.32	81.6
	DPSNet	5.48	89.2
	Ours	4.65	92.3

3.4.2 实际场景结果

传统 PS 算法在建筑立面的弱纹理区域出现明显的深度跳变，MPS 算法虽有所改善但仍存在阶梯状失真，DPSNet 的深度图较为平滑但在细节处精度不足，本文方法生成的深度图既保持了平滑性又准确还原了建筑的边角细节。工业零部件场景中，本文方法对螺栓、孔洞等细小结构的深度估计 $RMSE$ 为 2.17 mm，满足精密检测的精度要求。可以看出，多尺度特征提取网络主要贡献于细节结构的保持和弱纹理区域的稳定恢复，而上下文感知代价聚合则在全局范围内抑制噪声与局部异常

响应，两者共同作用，使得在实际复杂场景中仍能同时兼顾边缘锐利性与整体平滑性。

3.5 点云配准性能分析

ETH3D 数据集上的点云配准结果如表 3 所示。本文改进的 IPSO-ICP 算法的 RE 为 0.87 mm，较传统 ICP 降低 42.5%，较 FGR-ICP 降低 26.8%；收敛时间为 0.32 s，较 PSO-ICP 缩短 31.9%。精度与效率的同步改善源于动态参数调节策略对搜索行为的调控作用。与固定参数方法不同，动态惯性权重和自适应加速度因子使粒子群在迭代初期保持较强的全局搜索能力以跳出局部极值，而在后期逐步收缩搜索范围以逼近最优解。IRLS 的鲁棒加权则抑制了异常点的影响，使配准精度在复杂场景下仍保持在亚毫米级。

表 3 点云配准结果对比

方法	RE/mm	收敛时间/s	迭代次数
ICP	1.51	0.28	45
RANSAC-ICP	1.23	0.45	62
FGR-ICP	1.19	0.35	38
PSO-ICP	1.02	0.47	32
Ours(IPSO-ICP)	0.87	0.32	28

3.6 体积重建性能分析

3.6.1 公开数据集结果

Scenes11 数据集上的体积重建结果如表 4 所示。本文方法的 VP 达到 92.4%，较 DPSNet 提升 8.7%； CO 达到 96.3%，较 MPS 算法提升 10.2%； RD 降至 7.8%，较传统 PS 算法降低 68.4%。这得益于超体素分割的密度决策与可信度加权融合，有效减少了冗余数据并补全了空洞区域。

表 4 Scenes11 数据集体积重建结果

方法	$VP/\%$	$CO/\%$	$RD/\%$	总耗时/s
PS	72.3	78.5	24.5	18.6
MPS	80.1	86.1	18.2	22.3
DPSNet	84.9	90.5	12.3	15.7
Ours	92.4	96.3	7.8	14.2

3.6.2 实际场景结果

城市建筑场景重建的模型能够准确还原建筑的屋顶结构、窗户轮廓与阳台细节，体积精度达到 91.7%，完整性达到 95.8%，冗余度仅为 8.3%。传统 PS 算法重建的模型在建筑顶部存在明显空洞，MPS 算法的模型冗余度较高，DPSNet 的模型在细节处不够完整。

工业零部件场景中，本文方法重建的体积模型能够清晰呈现螺栓的螺纹与孔洞的内壁结构，体积精度达到 94.2%，满足工业检测的需求。通过与激光雷达扫描结果对比，两者的体积误差小于 2%，验证了方法的实

用性。

3.7 实时性分析

实时性测试结果如表 5 所示。本文方法的深度估计帧率达到 28 FPS，较 DPSNet 提升 34.9%，较 MPS 算法提升 115.4%；体积重建总耗时为 14.2 s，较传统 PS 算法缩短 23.7%。这得益于 GPU 加速的并行计算与超体素分割的降维处理，使算法能够兼顾精度与实时性。

表 5 实时性测试结果

方法	深度估计帧率/FPS	配准耗时/s	融合耗时/s	总耗时/s
PS	13.0	2.8	15.8	18.6
MPS	13.0	3.5	18.8	22.3
DPSNet	20.8	2.1	13.6	15.7
Ours	28.0	0.32	13.88	14.2

在实时场景采集与重建实验中，系统能够实现 28 FPS 的深度估计与每秒 0.07 次的体积更新，满足实时监控与动态建模的需求。

4 结束语

本文针对传统平面扫描算法在立体体积构建中存在的精度低、冗余高、实时性差等问题，提出一种基于深度逐平面扫描的立体体积构建方法。通过多尺度特征提取网络获取鲁棒特征表示，设计上下文感知代价聚合模块优化代价体，采用改进 PSO-ICP 算法实现点云精配准，结合超体素分割完成自适应体积融合。实验结果表明，该方法在多个公开数据集与实际场景中均表现优异，深度估计 RMSE 降至 4.21 mm，体积重建冗余度减少 32%，帧率达到 28 FPS，满足高精度实时立体体积构建的需求。该方法突破了传统平面扫描算法的性能瓶颈，为复杂场景的立体体积构建提供了新的技术方案。

参考文献：

[1] 朱福春, 王宏祥, 黄学文, 等. 基于三维实景的模拟仿真关键技术研究 [J]. 智能城市, 2024, 10 (8): 25 - 27.

[2] 冯 巩, 行霆燕, 王 楠, 等. 三维拉曼显微成像技术研究进展 (特邀) [J]. 激光与光电子学进展, 2024, 61 (6): 185 - 197.

[3] 王家豪. 立体零部件装配合格率三维检测与识别方法研究 [D]. 天津: 天津工业大学, 2018.

[4] 丁 凯. 单波段机载测深激光雷达全波形数据处理算法及应用研究 [D]. 深圳: 深圳大学, 2018.

[5] 赵 琦. 基于多视图图像的大规模场景三维重建算法研究 [D]. 天津: 天津大学, 2016.

[6] 王跃嵩. 多视图立体匹配三维重建方法 [D]. 武汉: 华中科技大学, 2015.

[7] 游达章, 周宏耀, 张业鹏. 基于特征优化的 Census 立体

匹配方法 [J]. 湖北工业大学学报, 2024, 39 (1): 41 - 45.

[8] HÄNE C, HENG L, LEE G H, et al. Real-time direct dense matching on fisheye images using plane-sweeping stereo [C] //2014 2nd International Conference on 3D Vision, IEEE, 2014, 1: 57 - 64.

[9] 乔景慧, 韩玉明, 张啸涵. 基于改进 Census 变换的鲁棒立体匹配算法 [J]. 计量学报, 2023, 44 (5): 694 - 700.

[10] YAO Y, LUO Z, LI S, et al. Mvsnet: depth inference for unstructured multi-view stereo [C] //Proceedings of the European Conference on Computer Vision (ECCV), 2018: 767 - 783.

[11] IM S, JEON H G, LIN S, et al. DPSNet: end-to-end deep plane sweep stereo [C] //7th International Conference on Learning Representations (ICLR 2019). New Orleans, United States, 2019: 8546 - 8557.

[12] 李书群, 陈 钰, 杨雨婷, 等. 基于佳点集人工鱼群的点云配准算法 [J]. 合肥工业大学学报 (自然科学版), 2023, 46 (9): 1203 - 1209.

[13] LóPEZ-MARTÍNEZ A, CUEVAS F J. Multiple view relations using the teaching and learning-based optimization algorithm [J]. Computers, 2020, 9 (4): 101.

[14] 龙 涛, 苏 畅, 王 建. 面向细节保持的特征描述子提取算法 [J]. 激光与光电子学进展, 2022, 59 (22): 212 - 219.

[15] 马乐杰, 邹德旋, 李 灿, 等. 融合差分进化和 Sine 混沌的改进粒子群算法 [J]. 计算机工程与应用, 2024, 60 (19): 80 - 96.

[16] 姜凯彬, 周世兵, 钱雪忠, 等. 动态融合的多视图投影聚类算法 [J]. 计算机科学与探索, 2023, 17 (5): 1147 - 1156.

[17] 周飞燕, 金林鹏, 董 军. 卷积神经网络研究综述 [J]. 计算机学报, 2017, 40 (6): 1229 - 1251.

[18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 - 778.

[19] LI Z, BHAT S F, WONKA P. PatchRefiner: leveraging synthetic data for real-domain high-resolution monocular metric depth estimation [C] //European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 250 - 267.

[20] GALLUP D, FRAHM J M, MORDOHAIP, et al. Real-time plane-sweeping stereo with multiple sweeping directions [C] //2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007: 1 - 8.

[21] 俞凯珑. 针对遥感场景的自监督单目深度估计算法及其应用研究 [D]. 昆明: 云南师范大学, 2024.