# 基于曲率与极线迭代的多视图三维重建方法

王若蓝

（太原学院 智能与信息工程系，山西 太原 030032）

**摘　要：**在多视图立体视觉领域，针对弱纹理、遮挡及复杂几何场景下存在的匹配精度不足的问题，提出一种融合曲率调制卷积与极线迭代优化的强匹配深度学习方法SM-MVS方法。首先，设计多尺度曲率感知网络，利用曲率先验动态调制卷积核增强边缘特征并抑制噪声；其次，提出极线几何迭代模块，在低分辨率深度图的基础上引入对极几何约束，采用极线动态采样与跨视图注意力机制逐步逼近真实深度；最后，通过门控循环单元（GRU）网络迭代优化直接得到高精度深度图，提升匹配精度、减少显存压力。在DTU公开数据集上的实验表明，该方法在完整性指标上相较于经典MVSNet提升22.1%，总体评分提升11.5%，适用于复杂场景下的物体三维重建。

**关键词：**深度学习；极线几何；多视立体视觉；三维重建

**DOI：**10.11907/rjdk.251642　　扫描二维码阅读全文：

**中图分类号：**TP391　　**文献标识码：**A　　**文章编号：**1672-7800(2026)007-0186-06

# Multi-view 3D Reconstruction Method Based on Curvature and Epipolar Iteration

WANG Ruolan

(*Department of Intelligence and Information Engineering*, *Taiyuan University*, *Taiyuan* 030032, *China*)

**Abstract:** Learning-based multi-view stereo (MVS) has achieved notable progress, yet remains challenged by textureless regions, severe occlusions, and complex geometric structures. To tackle these limitations, this paper introduces SM-MVS, a novel framework that synergistically integrates curvature-guided feature enhancement with epipolar geometry-based iterative refinement. The proposed approach consists of two key components: a multi-scale curvature-aware network that adaptively modulates convolutional kernels using geometric priors to reinforce edge features while suppressing noise, and an epipolar optimization module that progressively refines depth estimation through dynamic sampling and cross-view attention. Comprehensive evaluations on the DTU dataset demonstrate that SM-MVS achieves a 22.1% improvement in completeness and an 11.5% increase in overall score compared to MVSNet, while maintaining competitive computational efficiency. These results validate the framework's robustness and practical utility for high-fidelity 3D reconstruction in complex scenarios.

**Key Words:** deep learning; epipolar geometry; multi-view stereo; 3D reconstruction

## 0　引言

三维重建是计算机视觉中的重要研究方向，在自动驾驶、虚拟现实和文物保护等领域都有广泛应用[1-3]。在众多方法中，多视图立体视觉（Multi-View Stereo，MVS）对不同视角下的图像进行特征提取和匹配，恢复场景的稠密三维结构，基本原理是利用视差估计并结合代价优化生成点云，因此被视为三维重建的经典方法。

然而，传统MVS方法多依赖人工特征SIFT（Scale-Invariant Feature Transform）、SURF（Speeded Up Robust Features）进行匹配，在弱纹理、重复结构、遮挡和光照变化等复杂条件下表现不佳，重建结果常出现孔洞、噪声及细节缺失，这些问题使其在真实场景中的适用性受到限制，阻碍了模型在工程系统中的大规模推广[4]。

随着深度学习技术迅速发展，基于深度学习技术的MVS方法应运而生，逐渐成为了主流。MVSNet（Multi-View Stereo Network）作为开创性工作，提出可微分代价体的观点，使用3D卷积神经网络（Convolutional Neural Network，CNN）对代价体进行正则化，构建了端到端的深度估计框架，成为了基于深度学习技术的MVS方法的代表性方法，虽然重建精度较高但消耗显存较大[5]。TransMVSNet

**收稿日期：**2025-09-17

**基金项目：**山西省基础研究计划项目（202203021212015，202103021223005）；山西省大学生创新创业训练计划项目（20251675）；太原学院2024年青年科研项目（24TYQN16）

**作者简介：**王若蓝（1999-），女，太原学院智能与信息工程系助教，研究方向为人工智能与计算机视觉。

(Transformer-based Multi-view Stereo Network)方法提出一个基于2D CNN的特征匹配转换器,利用注意力聚合图像内和图像之间的远程上下文信息,改善了三维重建的效果[6]。CasMVSNet(Cascade Multi-View Stereo Network)融合多种不同尺度下的信息,引入级联方法与方差差异性度量实现了效率与精度的平衡,但显存需求增加,影响了重建效率[7]。PatchmatchNet(Patchmatch Network)等方法针对代价体进行迭代优化,依赖多阶段局部代价体假设评估深度值,限制了模型效率[8]。

除了在代价体优化方面做的工作,许多基于深度学习的主流MVS方法积极提升匹配代价或深度图预测精度。例如,AA-RMVSNet(Adaptive Aggregation Recurrent Multi-View Stereo Networ)引入递归机制和自适应聚合模块,有效提升了代价体的鲁棒性与匹配精度[9]。GAC-Net(Geometric and Attention-based Network)融合多尺度特征及通道注意力机制,显著提升了稀疏深度图的细化效果[10]。MVSTER(Multi-View Stereo Transformer)将Transformer网络引入MVS方法,跨视图执行特征聚合任务,显著提升了深度细化阶段的几何一致性[11]。朱代先等[12]引入渲染参考损失和深度一致性损失,将神经渲染与多视图立体技术有效结合,缓解了噪声问题。刘韵婷等[13]引入ECA(Efficient Channel Attention)注意力机制增强特征提取,采用卷积增强的门控循环单元GRU(Gated Recurrent Unit)模块对代价体进行正则化,但依赖三维概率中间体,在去噪的同时高分辨率细节严重丢失。DSC-MVSNet(Depthwise Separable Convolution based Multi-View Stereo Network)首次将3D深度可分离卷积引入代价体正则化,计算量降至普通3D卷积的1/14,并结合3D注意力机制缓解了多视图中相似特征误匹配的难题[14]。Min等[15]提出自适应Wasserstein损失和偏移模块,解决深度预测的分布差异和离散化问题,但计算开销显著增加且依赖初始深度估计的准确性。

虽然,这些方法在不同方面提升了MVS方法的性能,但对图像中的几何结构先验利用不足,在复杂几何形态或弱纹理区域的特征区分能力受限,且依赖构建全局代价体及使用3DCNN(3D Convolutional Neural Network)正则化,因此难以在大规模或高分辨率场景中高效应用。

综上,本文提出一种显式融合几何数据达到强匹配的深度学习框架SM-MVS,从3个维度进行创新:①多尺度曲率感知网络(CAFEN):提出曲率调制卷积,将局部曲率作为动态权重嵌入卷积核,结合多尺度特征金字塔,使用曲率先验知识提取低级与高级语义信息,以提升弱纹理与遮挡下的特征鲁棒性;②极线几何迭代模块:设计基于极线采样的优化机制,显式引入极线,使用极线约束匹配代价提升匹配和深度估计的精度,减少噪声;③优化策略:采用粗略估计—极线迭代策略,使用GRU网络预测深度估计,以直接优化深度图,降低计算量,缓解显存压力。

## 1　网络架构

为了提升深度估计的精度与细节表征能力,在多尺度级联特性突出的CVP-MVSNet(Cost Volume Pyramid based Multi-View Stereo Network)基础上,提出针对性的改进方案。具体为,整体框架采用由粗到细的深度重建思路,在低分辨率下进行初始估计,中分辨率阶段实现极线约束优化,高分辨率阶段通过残差细化恢复细节。整体网络架构如图1所示。

本文方法输入一个参考图像、$N-1$个源图像及其内参和相对相机姿态。首先,设定参考图像为$I_0$、相邻图像为$\{I_i\}_{i=1}^{N}$、图像大小为$H\times W\times 3$;其次,引入曲率感知模块,在特征层面嵌入边缘与纹理结构信息,提升深度估计的结构敏感性与区域判别力;再次,结合相机姿态建立跨视图极线关联,实现多分辨率下的精准对齐与一致性优化,以提升匹配稳健性与准确度;最后,利用多轮GRU更

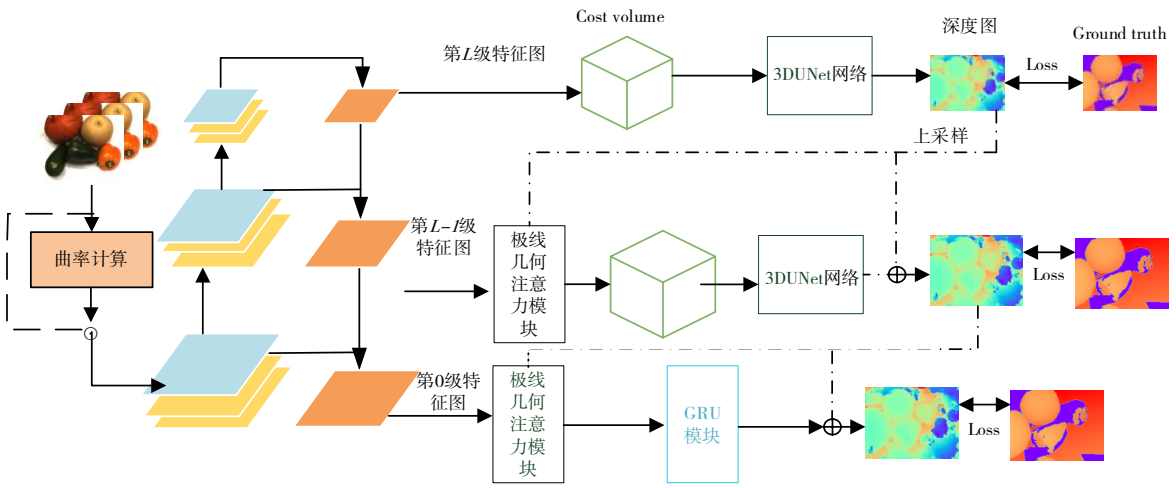



**Fig. 1　Network architecture**

**图1　网络架构**

新机制逐步细化深度结果，在保证高精度的同时显著降低显存占用，实现高分辨率深度的高效重建。

本文方法在特征层面融合曲率先验，在几何层面强化极线约束，并在优化层面采用轻量迭代结构，兼顾了结构细节恢复与计算效率，为多视图高精度深度估计提供了一种高效可靠的方法。

### 1.1 多尺度曲率感知网络

在复杂图像任务中，传统固定网格卷积对物体形变、旋转、遮挡等几何变化适应能力有限。为此，本文提出多尺度曲率感知网络（CAFEN），设计曲率调制卷积，将局部平均曲率作为动态权重作用于卷积核响应，并采用多尺度曲率金字塔结构增强弱纹理与遮挡边界的特征判别性，如图 2 所示。

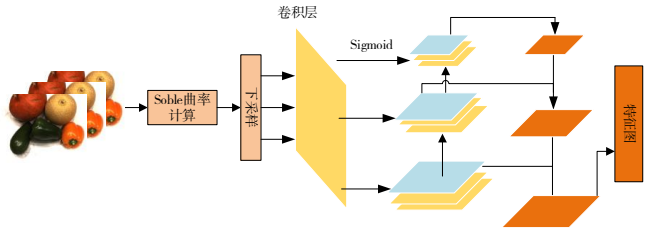



**Fig. 2 Multi-scale curvature-aware network**

**图 2 多尺度曲率感知网络**

曲率调制卷积引入图像曲率作为几何先验，以实现卷积核的自适应调节。该机制依据局部曲率特征进行差异化处理，在平坦区域抑制噪声，在边缘与角点等高曲率区域增强特征表达，从而有效缓解传统卷积的边缘模糊问题，提升细节重建质量。其中，曲率计算的准确性直接影响调制效果，该计算过程基于图像梯度，对噪声敏感且涉及邻域操作。

为此，本文采用镜像填充处理边界问题，通过反射边界像素保持梯度连续性，避免零填充导致的人工突变。同时，引入高斯平滑（3×3 核，σ=1.5）抑制噪声，该参数经实验验证，在实验场景中表现最好。

首先，使用（3×3）Sobel 算子计算灰度图像的曲率。

$$\kappa = \frac{|Gxx + Gyy|}{(1 + Gx^2 + Gy^2)^{1.5} + \epsilon} \tag{1}$$

式中：$Gx$ 为图像的 $x$ 方向的一阶偏导数；$Gy$ 为 $y$ 方向一阶偏导数；$Gxx$ 为 $x$ 方向二阶偏导数；$Gyy$ 为 $y$ 方向二阶偏导数。

在此基础上，对原始曲率图 $k$ 执行下采样操作，生成与 FPN 各层级分辨率一致的适配曲率图。同时，适配曲率图经 3×3 卷积层及 Sigmoid 函数激活生成权重 $\gamma_L$，将可学习通道缩放因子 $\alpha_L$ 与该层级特征图 $f_L$ 逐元素相乘得到调制后的特征图，如图 2 所示。

$$f_L^{\text{mod}} = f_L \odot (1 + \alpha_L \cdot \gamma_L) \tag{2}$$

### 1.2 极线几何迭代模块

现有 MVS 方法通常依赖可微单应性变换得到代价体，但单应性假设场景在局部近似为平面，面对复杂的几何特性和遮挡会导致错误对齐，产生噪声。为此，本文提出极线跨视图注意力机制，引入严格的对极几何约束，将匹配限制在极线邻域内，从而在提升深度估计精度的同时降低计算冗余与噪声干扰，如图 1 所示。

由于传统方法通常为所有像素设置一个固定的深度采样范围，忽略了场景深度的不确定性分布。为此，本文提出一种基于不确定性的自适应采样策略。首先，在初始深度图 $D^L$ 的基础上，结合极线几何注意力在局部范围内构建小规模代价体 $C^{L-1}$、输出深度图 $D^{L-1}$，以增强跨视图匹配的鲁棒性；其次，根据第一阶段得到的置信度 $\sigma(p)$ 为每个像素动态计算其深度采样半径：

$$\Delta(p) = \Delta_{\min} + (\Delta_{\max} - \Delta_{\min}) \cdot \frac{\sigma(p) - \sigma_{\min}}{\sigma_{\max} - \sigma_{\min}} \tag{3}$$

在每个像素 $p$ 的深度采样半径内均匀采样 $M$=32 个候选深度 $d_k$，并引入极线跨视图注意力机制，结合对极几何约束，仅在极线邻域内搜索匹配，通过极线几何注意力（Epipolar Geometry Attention，EGA）模块对特征进行聚合。

将源视图特征图映射到参考视图，通过方差聚合得到初始代价体 $C^{L-1}$，如图 3 所示。由此可见，针对参考视图 $I_0$ 中的像素 $p_0$，利用基本矩阵 $F$ 可计算其在源视图中的极线，如式（4）所示。同时，以当前投影点 $p_{s(t)}$ 为中心，在极线范围内采样 $j$=5 个点 $\left\{q_{i,j}\right\}_{i=0}^{N}$。

$$I_s = FP_0 \tag{4}$$

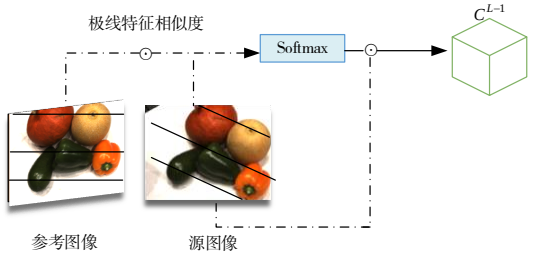



**Fig. 3 Epipolar geometry attention module**

**图 3 极线几何注意力模块**

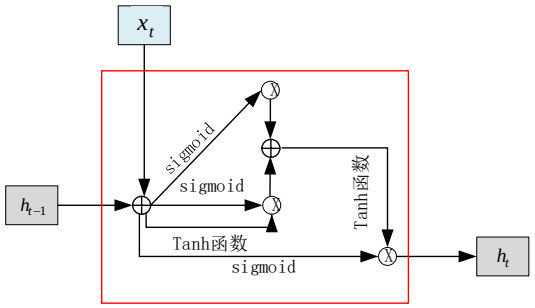
在此基础上，通过跨视图注意力计算参考特征与每个采样点特征的相似度，并基于 softmax 归一化得到注意力权重值 $\left\{w_{i,j}\right\}_{i=0}^{N}$。最后，将其与采样点特征相乘得到新的特征图，即代价体，如式（5）所示。

$$\tilde{F}(p, d_k) = \sum_{i,j} w_{i,j} F_s^i(q_{i,j}) \tag{5}$$

此外，语义信息也可引导特征匹配，即利用像素级语义信息对极线采样与匹配评分进行加权和约束，提升匹配可靠性。本文模型尚未引入语义模块，但为后续改进提供了可行方向。

### 1.3 GRU 残差优化网络

为了避免高分辨率下全局代价体的巨大显存消耗，该阶段不生成全局代价体。具体为，针对每个像素单独计算一个长度为 $M$ 的局部候选向量，并利用 GRU 直接预测深度残差得到最终深度图 $D^0$。GRU 单元结构如图 4 所示。




GRU

**Fig. 4　GRU unit**

**图4　GRU单元**

GRU相较于长短期记忆网络(Long Short Term Memory Network, LSTM)依赖分离的遗忘门与输入门的不同之处在于:GRU通过单一的更新门(update gate)实现了对历史状态衰减与当前候选状态激活的协同调控,在保证模型表达能力的条件下,显著降低了参数复杂性与计算开销。

首先,利用GRU叠加多个GRU层,基于GRU深度在每个时间步上产生的隐藏状态,利用极线注意力模块提取每个像素的局部匹配特征,以构造一维匹配向量;其次,将其输入到一个两层GRU网络中,通过4次迭代逐步预测深度残差。具体计算过程为:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{6}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{7}$$

$$\tilde{ht} = \tanh(W_h \cdot [r_t \odot ht - 1, x_t] + b_h) \tag{8}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{9}$$

综上,本文利用GRU的记忆功能,将稀疏深度图进一步深度估计聚合成深度图,使其既具备稀疏深度图的信息,又有精细深度值估计的优点,避免了普通全局代价体的显存占用,提升了整体场景空间的重建效果。

## 2　实验结果与分析

本文在DTU数据集上对所提出方法进行系统性训练与评估[16]。训练阶段,所有输入图像均被下采样至640×512分辨率,在计算效率与特征保持之间达到平衡。每次训练迭代随机选取5张图像构成一个样本组,以增强模型对不同视角组合的适应能力。深度搜索范围根据数据集特性设定为[425-935] mm,以覆盖场景中可能的主要深度分布。网络基于PyTorch框架实现,采用Adam优化器进行端到端训练,共16个epoch,初始学习率设为1×10-3,每轮衰减为0.9,批量大小为4。

测试阶段,在原始高分辨率图像(1 600×1 184)上进行,使用7个输入视图,沿用MVSNe的训练损失函数进行监督学习。为了进行泛化实验,在Tanks and Temples基准的中间集上进行评估,并与其他方法进行比较[17]。

### 2.1　实验环境

本文实验在Ubuntu 20.04系统下完成,硬件配置为Intel Core i9-9900k CPU与两块NVIDIA RTX TITAN GPU(24 GB与24.2 GB),详细信息如表1所示。

**Table 1　Experimental environment**

**表1　实验环境**

| 类别 | 类型或版本 |
|---|---|
| 操作系统 | Ubuntu 20.04 |
| CPU | Intel Core i9-9900k |
| GPU | NVIDIA RTX TITAN |
| GPU0内存/M | 24 212 |
| GPU1内存/M | 24 220 |

### 2.2　评价标准

在三维重建领域中,广泛使用准确度(Accuracy)、完整度(Completeness)、F-score三项指标评判三维重建的效果。假设$G$为真实点云集,$R$为预测点云集,预测点云$r \in R$,三项指标的定义及计算公式为:

(1)准确度(Accuracy)表示预测点云中每个点到真实点云的最近距离的平均值,值越小表示重建结果越精确。

$$Acc = \frac{1}{|R|}\sum_{r \in R} \min_{g \in G} \| r - g \|_2 \tag{10}$$

(2)完整度(Completeness)表示真实点云中每个点到预测点云的最近距离的平均值,值越小表示重建越完整。

$$Comp = \frac{1}{|G|}\sum_{g \in G} \min_{r \in R} \| g - r \|_2 \tag{11}$$

(3)F-score是准确度与完整度的调和平均数,综合反映整体重建质量。在DTU数据集中,F-score为完整度与准确度两者的平均数,值越高表明该方法的性能越优秀。

### 2.3　实验结果

表2为本文方法在DTU数据集的定量结果。由此可知,相较于早期MVSNet方法,本文方法在F-score上提升31.4%;相较于DSC-MVSNet等最新方法,本文方法在F-score上稳定领先9%~12%。实验证明,本文方法在完整度上表现突出,在弱纹理区域具有更强的重建能力。

**Table 2　Experimental results of different models on the DTU dataset**

**表2　不同模型在DTU数据集上的实验结果**

| 方法 | Acc | Comp | F-score |
|---|---|---|---|
| MVSNet[5] | 0.396 | 0.527 | 0.462 |
| CasMVSNe[7] | 0.325 | 0.385 | 0.355 |
| PVA-MVSNet[6] | 0.379 | 0.336 | 0.357 |
| CVP-MVSNet[18] | 0.296 | 0.406 | 0.351 |
| EGF-MVSNet[13] | 0.343 | 0.351 | 0.347 |
| 文献[12] | 0.363 | 0.289 | 0.326 |
| DSC-MVSNet[14] | 0.316 | 0.372 | 0.344 |
| 文献[15] | 0.364 | 0.354 | 0.359 |
| 文献[19] | 0.355 | 0.301 | 0.328 |
| 文献[20] | 0.351 | 0.287 | 0.319 |
| Ours | 0.324 | 0.310 | 0.317 |

在DTU场景12上,本文方法与CVA-MVSNet的点云重建定性结果如图5所示。由此可见,本文方法重建的点云整体更清晰、结构更完整,能有效恢复物体轮廓并减少点云断裂。尤其在低纹理区域与深度突变边界,本文方法的点分布更均匀、细节更丰富,展现出了更强的鲁棒性。

为了验证本文方法在不同复杂场景下的表现，在Scan13、Scan34上进行比较，如图6所示。由此可见，Scan13纹理丰富、结构精细，在保留细节的同时有效抑制了噪声；Scan34空间范围大、包含远距离物体，在结构连贯性、整体完整性与远景清晰度方面表现更优，进一步验证了该方法的鲁棒性更强、精度更高。

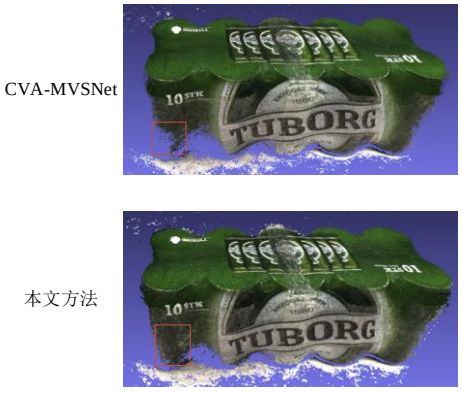



**Fig. 5 Comparison of point cloud results on DTU scenario 12**

**图5 DTU场景12上的点云结果比较**

为了进行泛化评估，将在DTU数据集上训练得到的模型迁移至Tanks and Temples基准的Intermediate子集进行测试，如表3所示。由此可知，在与多种深度学习方法比较下，本文方法在Francis等多个场景上均取得了更具竞争力的重建质量，跨数据集适应能力良好。图7为Tanks and Temples数据集中经典场景的重建效果。由此可见，所提模型能较好地恢复场景的整体结构与细节层次，点云分布密集且连续，在复杂环境下仍具有较强的重建能力。虽然，局部区域存在少量稀疏点或轻微噪声，但不影响整体的几何一致性。

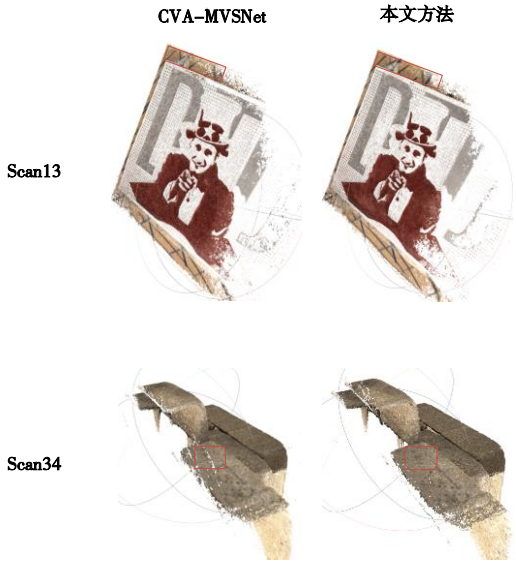



**Fig. 6 Comparison of point cloud results on DTU scans 13 and 34**

**图6 DTU场景13、34上的点云结果比较**

**Table 3 Result of Tanks and Temples dataset index**

**表3 Tanks and Temples数据集指标结果**

| 方法 | Mean | Family | Francis | Horse | Light-house | M60 | Panther | Play-ground | Train |
|---|---|---|---|---|---|---|---|---|---|
| MVSNet[5] | 43.48 | 55.99 | 28.55 | 25.07 | 50.79 | 53.96 | 50.86 | 47.90 | 34.69 |
| CasMVSNet[7] | 56.42 | 76.36 | 58.45 | 46.20 | 55.53 | 56.11 | 54.02 | 58.17 | 46.56 |
| AA-RMVSNet[9] | 61.51 | 77.77 | 59.53 | 51.53 | 64.02 | 64.05 | 59.47 | 60.85 | 54.90 |
| 文献[12] | 60.31 | 79.68 | 62.33 | 53.60 | 56.86 | 61.06 | 57.66 | 56.48 | 54.85 |
| 文献[15] | 60.53 | 77.18 | 57.96 | 52.55 | 60.03 | 60.73 | 61.23 | **60.28** | 54.26 |
| Vis-MVSNet[21] | 60.03 | 77.40 | 60.23 | 47.07 | 63.44 | 62.21 | 57.28 | **60.54** | 52.07 |
| 本文方法 | **62.17** | 79.01 | **63.42** | **51.70** | 60.75 | 59.62 | **57.27** | 57.36 | **54.76** |

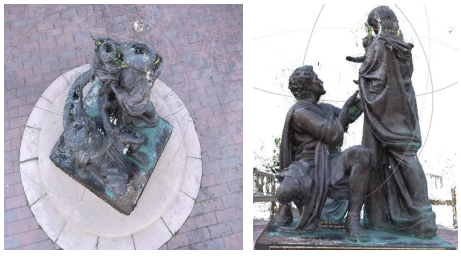


**Fig. 7 Family scene visualization**

**图7 Family场景可视化**

## 2.4 消融实验

本文通过消融实验验证曲率计算中关键参数的选择合理性，以避免全量消融所带来的巨大计算开销。具体为，选取4个具有代表性的训练场景（scan1、scan4、scan9、scan10）和2个测试场景（scan11、scan12）进行消融实验。表4为不同边界填充方案对曲率估计质量的影响。表5为高斯平滑参数对最终深度图重建精度的作用效果，结果以F-score进行展示。由此可知，在边界处理方面，镜像填充方法重建效果相较于零填充、重复填充方法的效果更好；在高斯参数选择方面，3×3卷积核配合σ=1.5参数能取得最佳的深度重建精度，而增大核尺寸或调整标准差均会导致性能下降。

**Table 4 Performance comparison of boundary filling schemes**

**表4 边界填充方案性能比较**

| 填充方法 | Acc | Comp | F-score |
|---|---|---|---|
| 零填充 | 0.347 | 0.335 | 0.341 |
| 重复填充 | 0.339 | 0.327 | 0.333 |
| 镜像填充(本文) | 0.336 | 0.319 | 0.328 |

**Table 5 Performance comparison of Gaussian smoothing parameters**

**表5 高斯平滑参数性能比较**

| 核尺寸 | σ=1.0 | σ=1.5 |
|---|---|---|
| 3×3 | 0.336 | 0.328 |
| 5×5 | 0.345 | 0.349 |

在深度图优化阶段，本文使用了基于GRU的迭代更新策略。为了分析迭代次数对精度与效率的影响，在2、4、6

次迭代后进行比较实验,如表6所示。由此可知,2次迭代的深度估计精度明显偏低;4次迭代在保证精度的同时,计算效率与显存消耗均保持在合理范围内;6次迭代的精度仅有轻微提升,但推理时间与显存占用显著增加。综上,4次迭代在精度与效率之间取得了最佳平衡,因此采用4次GRU迭代进行后续实验。

**Table 6　Performance comparison of GRU iteration times**

**表6　GRU的迭代次数性能比较**

| 迭代次数 | Acc | Comp | F-score |
|---|---|---|---|
| 2次 | 0.345 | 0.329 | 0.337 |
| 4次 | 0.336 | 0.319 | 0.328 |
| 6次 | 0.330 | 0.322 | 0.326 |

此外,通过消融实验定量分析所提方法中关键组件的优势和有效性,采用精度与完整性作为评价指标,以比较多尺度曲率感知网络、极线几何迭代模块的有效性,如表7所示。由此可知,多尺度曲率感知网络、极线几何迭代模块均能显著提升重建标准,互补可达到最佳效果。

**Table 7　Results of ablation experiment on DTU dataset**

**表7　在DTU数据集上的消融实验结果**

| Multi-Scale Curvature-Aware Network | Epipolar Geometry Attention Module | Acc | Comp | F-score |
|---|---|---|---|---|
| √ | × | 0.312 | **0.347** | **0.330** |
| × | √ | **0.323** | 0.321 | 0.322 |
| √ | √ | 0.310 | 0.369 | 0.345 |

## 3　结语

本文提出SM-MVS多视图立体视觉方法,解决了弱纹理、遮挡及复杂几何场景下匹配精度不足的问题。该方法包括3个方面的创新:①构建多尺度曲率感知网络,利用曲率信息动态调制卷积核强化边缘特征、抑制噪声;②设计极线几何迭代优化模块,结合对极约束与跨视图注意力实现自适应深度采样,减少匹配误差;③采用基于GRU的轻量迭代优化更新深度图残差,在保证精度的同时降低显存占用。

实验表明,SM-MVS在三维重建完整度与细节恢复方面均优于多数主流方法,尤其在弱纹理与复杂结构场景中表现突出,为复杂环境下高精度三维重建提供了新思路。未来,将探索语义信息在匹配过程中的辅助作用,有望与现有的几何约束机制结合,利用像素级语义信息对极线采样与匹配评分进行加权和约束,形成鲁棒性更强的语义—几何协同匹配框架。

**参考文献:**


[1] SUN X M, REN L. A survey of 3D reconstruction with RGB-D cameras [J]. Software Guide, 2021, 20(5): 249-252.
孙晓明,任磊. RGB-D相机的三维重建综述[J]. 软件导刊,2021,20(5):249-252.

[2] REMONDINO F, EL-HAKIM S. Image-based 3D modelling: a review [J]. The Photogrammetric Record, 2006, 21(115): 269-291.

[3] SCARAMUZZA D, FRAUNDORFER F. Visual odometry [J]. IEEE Robotics & Automation Magazine, 2011, 18(4): 80-92.

[4] GOESELE M, CURLESS B, SEITZ S M. Multi-view stereo revisited[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006: 2402-2409.

[5] YAO Y, LUO Z X, LI S W, et al. MVSNet: depth inference for unstructured multi-view st-ereo[C]// European Conference on Computer Vision, 2018:767-783.

[6] DING Y, YUAN W, ZHU Q, et al. Transmvsnet: global context-aware multi-view stereo network with transformers [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2022: 8585-8594.

[7] GU X, FAN Z, ZHU S, et al. Cascade cost volume for high-resolution multi-view stereo and stereo matching[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2020: 2495-2504.

[8] BLEYER M, RHEMANN C, ROTHER C. PatchMatch stereo-stereo matching with slanted support windows[EB/OL]. https://www.microsoft.com/en-us/research/wp-content/uploads/2011/01/PatchMatchStereo_BM-VC2011_6MB.pdf.

[9] WEI Z, ZHU Q, MIN C, et al. AA-RMVSNet: adaptive aggregation recurrent multi-view stereo network[C]// IEEE International Conference on Computer Vision, 2021: 6187-6196.

[10] ZHU K, GAN X, SUN M. GAC-Net_geometric and attention-based network for depth completion [DB/OL]. https://arxiv.org/abs/2501.07988.

[11] WANG X, ZHU Z, QIN F, et al. MVSTER: epipolar transformer for efficient multi-view stereo[C]// European Conference on Computer Vision, 2022: 573-591.

[12] ZHU D X, KONG H R, QIU Q, et al. Multi-view 3D reconstruction algorithm integrating attention mechanism and neural rendering[J]. Electronic Measurement Technology, 2024, 47(5): 158-166.
朱代先,孔浩然,秋强,等. 注意力机制与神经渲染的多视图三维重建算法[J]. 电子测量技术,2024,47(5):158-166.

[13] LIU Y T, GAO Y, DAI J L, et al. Multi-view 3D reconstruction integrating ECA attention layer and lightweight regularization [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(7): 179-186.
刘韵婷,高宇,戴佳霖,等. 融合ECA注意力层和轻量正则化的多视图三维重建[J]. 电子测量与仪器学报,2024,38(7):179-186.

[14] ZHANG S, WEI Z, XU W, et al. DSC-MVSNet: attention aware cost volume regularization based on depthwise separable convolution for multi-view stereo[J]. Complex & Intelligent Systems, 2023, 9(6): 6953-6969.

[15] MIN Q, ZHAO J, ZHANG Z, et al. Adaptive learning for multi-view stereo reconstruction[DB/OL]. https://arxiv.org/abs/2404.05181.

[16] AANS H, JENSEN R R, DAHL A B. Large-scale data for multiple-view stereopsis[J]. International Journal Computer Vision, 2016, 120(2):153-168.

[17] KNAPITSCH A, PARK J, ZHOU Q Y, et al. Tanks and temples: benchmarking large-scale scene reconstruction[J]. ACM Transactions on Graphics, 2017, 36(4): 1-13.

[18] YANG J, MAO W, ALVAREZ J M, et al. Cost volume pyramid based depth inference for multi-view stereo[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2020: 4877-4886.

[19] ZHU G Z, WEI B, YANG A F, et al. Multi-view 3D reconstruction method based on self-attention mechanism[J]. Laser & Optoelectronics Progress, 2023, 60(16): 323-330.
朱光照,韦博,杨阿峰,等. 基于自注意力机制的多视图三维重建方法[J]. 激光与光电子学进展,2023,60(16):323-330.

[20] SONG S, TRUONG K G, KIM D, et al. Prior depth-based multi-view stereo network for online 3D model reconstruction[J]. Pattern Recognition, 2023, 136: 109198.

[21] ZHANG J, LI S, LUO Z, et al. Vis-mvsnet: visibility-aware multi-view stereo network[J]. International Journal of Computer Vision, 2023, 131(1): 199-214.


(责任编辑:刘嘉文)