

基于改进型运动恢复结构和 3DGS 的红外图像三维重建

李 游¹,黄泽润¹,蔡 礼²,何慧雯³,曾 晗¹

(1. 华东交通大学 电气与自动化工程学院,江西 南昌 330013;2. 国网江西省电力有限公司电力科学研究院,江西 南昌 330096;
3. 中国电力科学研究院有限公司,北京 100192)

摘 要:面向电力设备的三维重建技术在状态感知、故障诊断、数字孪生和智能巡检等工作中具有重要意义。针对热红外图像纹理弱、对比度低、细节不足,导致传统多视图重建中特征匹配困难、位姿估计不稳定和模型不完整等问题,文中提出一种基于视觉几何组运动恢复结构(visual geometry grounded deep structure from motion, VGGSFm)与改进三维高斯溅射(3d gaussian splatting, 3DGS)的三维重建方法。首先,利用 VGGSFm 对多视角热红外图像进行相机位姿估计和稀疏结构恢复,为后续重建提供几何先验;随后,以相机参数和稀疏点云初始化 3DGS 显式场景表示,实现热红外场景重建与新视角表达;最后,引入深度法向正则约束,增强高斯元分布的几何稳定性和结构一致性。

关键词:红外图像;电力设备;三维重建;VGGSFm;三维高斯溅射;深度正则化

中图分类号:TP 206 **文献标志码:**B **文章编号:**1006-348X(2026)03-0070-06

0 引言

近年来,变电站巡检人员已开始利用热红外相机采集变电站设备的热红外图像。当前,针对电力设备进行温度监测已有较多的研究,胡钰泊^[1]等提出了基于改进 yolov8 的电力设备红外图像检测方法,提升了复杂场景下红外设备目标检测的性能;Chi Hao^[2]等人提出了一种面向变电站设备的轻量融合网络,在突出热目标的同时保留背景纹理,并兼顾了移动端部署需求;张秋铭^[3]等针对在线监测系统中红外图像分割效果差、速度慢等问题,提出了一种改进的 Chan - Vese 模型分割算法,提升了红外图像分割的精度与效率,并在 Dice 重合率和错误分割率等指标上取得了较好效果。然而,热红外相机提供的照片都是二维形态,无法在三维空间里展示其分布和趋势,难以应用于工程实践。因此,有必要研究一种把热成像图应用到变电站设备 3D 模型的方法,来增强立体空间信息的直观性和可视化能力,并辅助运维人员在工作中做

出准确分析。

与文中研究密切相关的重建方法可以分为两类,即基于多模态融合的图像三维重建与基于单模态的三维重建。多模态融合的重建方法常使用可见光图像与红外图像,可见光图像纹理丰富、结构信息完整,但在夜间、雨雾等复杂环境下稳定性较差;红外图像能够反映设备热状态、抗光照干扰能力较强,但存在纹理模糊、细节不足的问题,因此有研究人员将二者结合,同时保留结构纹理与温度信息。代表性的研究有:Skorokhodov^[4]等人提出的在预训练视觉几何 Transformer^[5]进行轻量化微调的基础上,面向可见光-热红外跨模态三维重建,以实现两种模态下相机位姿与场景结构的联合恢复。但该方法需要额外的多模态微调数据支持,因此,可见光与热红外图像的融合方法在实际场景中并没有得到广泛应用。

由于热红外图普遍存在特征不显著、对比度低等问题,仅依靠热红外成像设备和图像进行三维重建的研究较少。面对这个问题,研究人员提出了许多方案,其中代表性有:Sentenac^[6]等人提出了基于红外双

收稿日期:2026-02-10

基金项目:国家自然科学基金地区基金项目(52567009);国家自然科学基金面上项目(52277148);江西省自然科学基金项目(20252BAC240319)。

作者简介:李游(2001),女,硕士研究生,从事电气设备多模态检测方面的研究。

目相机的自动热三维重建方法,实现了仅依赖红外成像的自动化热三维重建。因此,目前大部分红外三维重建方法都需要激光雷达、深度相机等专用设备,工程门槛较高;此外,红外三维重建通常对建模能力要求较高,训练和渲染开销较大。

鉴于此,文中提出了一套基于深度正则化三维高斯溅射(3d gaussian splatting, 3DGS)^[7]和基于学习的运动恢复结构技术,具体为:为解决红外图像特征稀疏易导致重建失败的问题,引入视觉几何组运动恢复结构(visual geometry grounded deep structure from motion, VGGSFm)方法^[8],恢复各视角相机位姿和稀疏点云结构,为后续三维高斯表示提供可靠的几何先验;利用3DGS,构建电力设备的三维场景表示,并在基础框架中进一步引入深度正则项,提升重建结果的几何稳定性与结构一致性。

1 改进型运动恢复结构技术

运动恢复结构(structure from motion, SFM)是一种基于多视图几何的三维重建方法,可以由多张图像同时恢复场景的三维结构和各视角相机位姿。典型的SFM系统一般由特征提取与匹配、相机位姿初始化与稀疏点云恢复、更新特征轨迹与图像集合、束调整优化环节四个部分组成。

传统SFM在弱纹理场景中存在特征点提取与匹配困难的问题。为了解决这一问题,文中采用VGGSFm作为位姿估计前端,利用深度点跟踪,直接从图像中提取跨视角2D轨迹,并通过可学习三角化和可微分Bundle Adjustment模块,对相机参数和三维点进行联合优化。

与普通SFM依赖局部两两匹配和增量式相机注册不同,VGGSFm采用深度前馈式点跟踪替代成对匹配链式拼接,直接从输入图像集合中输出跨视图的2D轨迹。设输入图像集合为 $I = \{I_i\}_{i=1}^{N_t}$,对于参考图像中的查询点,跟踪器输出其在各视图中的轨迹:

$$T^j = \left((y_i^j, m_i^j) \cdots, (y_{N_t}^j, m_{N_t}^j) \right) \quad (1)$$

式中: m_i^j 、 y_i^j 分别表示第 j 个点在第 i 个摄像机中的可见性标记与二维坐标。VGGSFm先从图像集合中选取查询帧,并使用SuperPoint和SIFT从该查询帧中提

取出 N_t 个关键点,并从特征图中提取出查询点对应的查询描述符。这些点对应的查询描述符与所有 N_t 帧在不同空间分辨率下的特征图进行相关性计算,构造出代价体积分金字塔,并将其展平为 $V \in \mathbb{R}^{N_t \times N_t \times C}$ 输入Transformer。因Transformer网络需要接收一维向量序列,故将 L 层代价体积分金字塔逐层展平并拼接为长度为 C 的向量:

$$C = \sum_{l=1}^L H_l W_l \quad (2)$$

式中: C 为拼接后向量长度; H_l 和 W_l 分别为第 l 层代价体积的高度和宽度, L 为金字塔层数。Transformer模块先把该张量线性投影到固定的隐藏维度大小,再经过自注意力机制,提取跨帧几何共视关系,最后,前馈神经网络对数据进行非线性变换,进一步提取深层几何特征。经过多层串联处理后,将该富含上下文信息的张量送进并行的解码器中,最终输出粗级轨迹、预测置信度、轨迹特征描述符与可见性指示符。

为满足SFM对亚像素精度的要求,进一步采用由粗到细的跟踪策略:先在全图范围内粗定位,再在局部图像块中细化,最终得到亚像素级精细对应关系。这样可以减少传统轨迹串联中的漂移问题,并提升弱纹理区域的对应稳定性。

可学习相机初始化模块直接对所有输入图像进行联合相机恢复。具体而言,设第 i 张图像的全局特征为 $\phi(I_i)$,轨迹描述子为 $d^p(y_i^j)$,利用 $\phi(I_i)$ 作为query、 $d^p(y_i^j)$ 作为key-value做交叉注意力,再将其与由八点法估计得到的初始相机信息拼接,输入Transformer主干网络。网络接收这些融合了追踪轨迹几何与语义信息的数据后,最终输出所有相机的初始姿态 P 。相机参数采用旋转四元数 q_i 、平移 t_i 以及焦距对数 $\log f_i$ 等8自由度形式表示。

在得到初始相机参数 P 与多视图轨迹 y_i^j 后,该步骤进一步利用可学习三角化模块恢复初始稀疏点云。首先,根据精细轨迹集合和相机初始位姿,通过多视图直接线性变换进行空间相交计算,得到一个初始三维点云 X_a ,求解公式如下:

$$X_a = \arg \min_{\|x\|} \|AX\| \quad (3)$$

式中: A 为由精细轨迹集合中的多视图二维观点和对应相机投影矩阵构成的线性方程矩阵; X 为待求解的齐次三维点变量; X_a 为线性三角化得到的初始齐次

三维点云。

随后,计算其到各相机视线的距离以及该视线上最近点位置,并将这些几何信息与轨迹特征、点的谐波位置编码一同输入 Transformer,得到深度特征矫正的初始3D点云 X_c 。

当初始3D点云和相机初始位姿都得到后,使用可微光束平差法(BA)同时优化三维点位置和相机参数:

$$(X_d, P_b) = BA(T, X_c, P) = \arg \min_{X, P} L_{BA} \quad (4)$$

$$L_{BA} = \sum_{i=1}^{N_l} \sum_{j=1}^{N_r} m_i^j \left\| \pi(P_i, X_i) - y_i^j \right\|_2^2 \quad (5)$$

式中: X_c 为三角化器校正后的初始点云; P 为初始相机参数集合; L_{BA} 为光束平差损失函数; X_d 和 P_b 分别为优化后的三维点云和相机参数; $\pi(P_i, X_i)$ 表示三维点 X_i 通过相机参数 P_i 投影到图像平面的结果。

由于该模块是可微的,因此,误差可反向传播至整个架构的可学习参数中,使得预测误差逐渐减少。整体重建流程如图1所示。

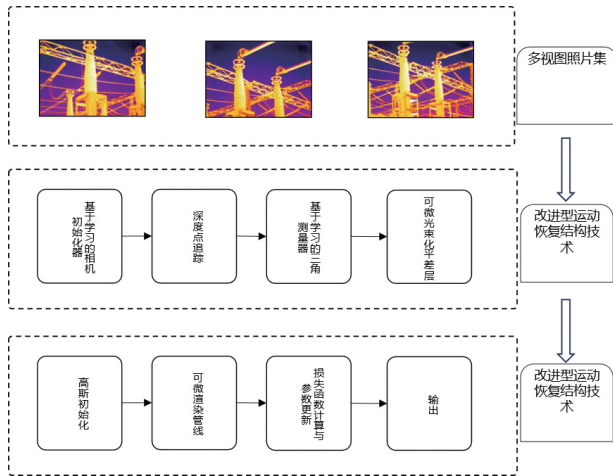


图1 红外图像三维重建流程

2 改进3DGS建模及渲染方法

2.1 3DGS建模

2.1.1 3DGS的二维表达

3DGS是一种基于显式场景表示的三维重建与渲染方法。与一般显式表示如MVS多视图立体几何重建方法、深度图重建方法不同,3DGS将场景表示为大量可优化的三维高斯元。

每个高斯元包含中心位置、尺度以及颜色等属

性,其中颜色通常由球谐函数进行建模。空间中以点 μ 为原点的三维高斯的基本形式是:

$$G(x) = \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (6)$$

式中: μ 是三维高斯中心,也就是它在三维空间中的位置; Σ 是协方差矩阵,决定它的大小、长短轴和朝向。

将三维高斯椭球渲染成图像需要对中心点进行投影,再在二维图像平面对齐其缩放及尺度。公式表达为:

$$\lambda p = K[R|t]\mu \quad (7)$$

式中: λ 是齐次坐标的尺度因子; p 是图像平面上的像素点; R 、 t 分别为旋转矩阵和平移向量; K 是相机内参矩阵。

其次,三维高斯投影的形状由二维协方差 Σ' 决定。公式为:

$$\Sigma' = JR_c \Sigma R_c^T J^T \quad (8)$$

式中: R_c 是界坐标到相机坐标的线性部分; J 是透视投影在当前点附近的Jacobian矩阵。

当三维高斯椭球投影到图像上后,每个高斯对当前像素的贡献表示为一个二维函数。其二维高斯权重公式如下:

$$g_i(v) = \exp\left[-\frac{1}{2}(v - p_i)^T (\Sigma'_i)^{-1}(v - p_i)\right] \quad (9)$$

式中: v 是二维图像坐标; p_i 为第 i 个高斯投影后的二维中心坐标; Σ'_i 为第 i 个高斯投影后的协方差矩阵。再与权重和基础不透明度参数相乘,就可以得到随坐标而变的不透明度 α_i ,最后根据alpha合成计算出像素颜色:

$$C(v) = \sum_{i=1}^N c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \alpha_i = o_i g_i(v) \quad (10)$$

式中: $C(v)$ 表示二维图像位置 v 处的渲染颜色; c_i 为第 i 个高斯的颜色, o_i 表示第 i 个高斯的不透明度。

2.1.2 自适应密度调节与损失函数设计

3DGS在参数更新过程中穿插执行自适应密度控制。当某一区域的高斯因位置梯度较大而表明当前表示不足或过于粗糙时,需要进行增密;而当某些高斯长期不透明度很低,则应进行剪枝。具体方法是当它们对渲染贡献极弱,或在图像平面上投影足迹过大造成场景过度平滑、覆盖不准时,可视为无效高斯并删除之。

3DGS 的监督主要来自图像重建误差。其常见形式可写为:

$$L_{\text{rgb}} = (1 - \lambda_{\text{ssim}}) L_1 + \lambda_{\text{ssim}} L_{\text{D-SSIM}} \quad (11)$$

式中: L_{rgb} 表示图像重建损失; L_1 表示渲染图像与真实图像之间的损失; $L_{\text{D-SSIM}}$ 表示基于结构相似性的差异项; λ_{ssim} 为结构相似性损失项的平衡系数, 一般取 0.2。

2.2 基于深度-法向联合的正则 3DGS 改进策略

由于 3DGS 是一种全新的三维重建方法, 算法本身存在很多有待改进的方向。文中在不改变原始 3DGS 基本表示与渲染框架的前提下, 引入深度-法向几何先验正则化机制。对于深度约束, 先由当前高斯场渲染得到深度 $D_{\text{render}}(p)$, 与颜色合成类似:

$$D_{\text{render}}(p) = \sum_{i \in N(p)} z_i \alpha_i(p) \prod_{j=1}^{i-1} (1 - \alpha_j(p)) \quad (12)$$

式中: z_i 表示第 i 个高斯中心在相机坐标系下的深度值; $N(p)$ 表示参与像素 p 渲染的高斯集合。再与参考深度 $D^*(p)$ 构成深度监督项:

$$L_{\text{depth}} = \frac{1}{|\Omega|} \sum_{p \in \Omega} \exp(-\eta |\nabla I(p)|) |D_{\text{render}}(p) - D^*(p)| \quad (13)$$

式中: η 表示边缘权重调节系数; $\nabla I(p)$ 为图像梯度, 当像素处是强边缘时, $|\nabla I|$ 大, 权重变小, 说明不要过度相信该处的深度; L_{depth} 表示深度正则损失; Ω 表示参与损失计算的像素区域。

对于法向约束, 同样和深度正则约束类似:

$$L_{\text{normal}} = \frac{1}{|\Omega|} \sum_{p \in \Omega} \|N_{\text{render}}(p) - N^*(p)\|_1 \quad (14)$$

式中: L_{normal} 表示法向正则损失; $N_{\text{render}}(p)$ 表示由渲染深度估计得到的像素 p 处的法向量; $N^*(p)$ 表示参考法向量。

最终得到增加深度-法向联合正则化后的损失函数表示为:

$$L_{\text{total}} = (1 - \lambda_{\text{ssim}}) L_1 + \lambda_{\text{ssim}} L_{\text{D-SSIM}} + \lambda_{\text{depth}} L_{\text{depth}} + \lambda_{\text{normal}} L_{\text{normal}} \quad (15)$$

式中: L_{total} 表示总损失函数; λ_{depth} 、 λ_{normal} 分别表示深度正则损失与法向正则损失的权重系数。该改进方案易于与现有 3DGS 框架集成, 适合热红外图像三维重建任务。

3 实验与结果分析

文中实验均在 Ubuntu 22.04.1 LTS 环境下完成,

实验硬件平台主要由 Ultra 7 265KF 处理器、NVIDIA GeForce RTX 5080 显卡以及 64 GB 系统内存组成, 可为多视图图像特征匹配、相机位姿恢复及三维高斯溅射优化提供较为充足的计算资源, 保证了模型训练与渲染过程具有较高的计算效率和稳定性。

3.1 实验评价指标

为了全面评价文中方法在变电站电力设备热红外场景三维重建任务中的重建质量与渲染效果, 主要从图像保真性、结构一致性以及感知相似性三个角度选取评价指标, 对不同方法在测试集上的新视角合成结果进行定量分析。

1) 峰值信噪比 Q_{PSNR}

$$Q_{\text{PSNR}} = 10 \log_{10} \left(\frac{(2^2 - 1)^2}{E_{\text{MSE}}} \right) \quad (16)$$

式中: n 为每个采样值的比特数; E_{MSE} 为原图像三维重建后图像之间均方误差。

2) 结构相似性指数 SSIM

PSNR 从像素差值角度衡量图像之间的差异, 而 SSIM 是一种从视觉感知角度出发设计的图像质量评价指标, 能够综合比较两幅图像在亮度、对比度以及结构三方面的相似程度, 定义为:

$$Q_{\text{SSIM}} = \frac{(2\mu_x \mu_y + \varepsilon_1)(2\sigma_{xy} + \varepsilon_2)}{(\mu_x^2 + \mu_y^2 + \varepsilon_1)(\sigma_x^2 + \sigma_y^2 + \varepsilon_2)} \quad (17)$$

式中: μ_x 和 μ_y 分别表示图像块 x 与 y 的均值; σ_x^2 和 σ_y^2 分别表示图像 x 和 y 的方差; σ_{xy} 表示二者协方差; ε_1 和 ε_2 为防止分母为零的稳定常数。

3) 学习感知图像块相似度 LPIPS

为了进一步从感知层面衡量重建图像与真实图像之间的差异, 文中选取学习感知图像块相似度 LPIPS 作为补充指标。LPIPS 属于负指标, 其核心思想是先将图像输入预训练深度网络, 提取多层特征表示, 再计算其在特征空间中的距离:

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (y'_{hw} - y'_{0hw})\|_2^2 \quad (18)$$

式中: $d(x, x_0)$ 表示渲染图像 x 与真实图像 x_0 之间的感知距离; H 和 W 分别表示图像或特征的高度和宽度; y'_{hw} 表示渲染图像 x 在第 l 层特征图位置 (h, w) 处的特征值; y'_{0hw} 表示真实图像 x_0 在对应位置处的特征值, h 和 w 分别表示当前参与计算的空间位置索引, l 表示预训练网络的第 l 个特征层, w_l 为对应层的可学习

权重。

3.2 图像匹配实验

文中对改进的特征提取与匹配算法的性能进行了定量和定性评估,并与其它先进的特征提取与匹配算法进行了比较。评估主要包括匹配有效性以及匹配可视化效果两个方面:

1) 匹配有效性实验

为评价不同图像配对方法在热红外弱纹理场景下的匹配性能,采用 Precision、Recall 和 F1-score 对不同图像配对方法的性能进行量化评价。表 1 给出 ORB+FLANN、SIFT+FLANN 以及 VGGSFM 三种方法的实验结果。

表 1 三种不同匹配算法的重建指标对比

方法	Precision / %	Recall / %	F1-score / %	平均有效匹配点 对数	平均匹配时间 / s
ORB+FLANN	79.4	73.1	76.1	112	0.018
SIFT+FLANN	88.7	85.3	87.0	179	0.073
VGGSFM	94.9	92.6	93.7	269	0.214

由表 1 可知,VGGSFM 的 Precision、Recall 和 F1-score 分别达到 94.9%、92.6% 和 93.7%,均优于 SIFT+FLANN 和 ORB+FLANN,说明该方法在热红外弱纹理场景下,能够建立更加充分和准确的匹配关系。SIFT+FLANN 的整体性能次之,而 ORB+FLANN 的匹配精度和匹配完整性均相对较低,表明传统二值局部特征在低对比度、纹理稀疏的热红外外场景中更容易受到限制。虽然 ORB+FLANN 在平均匹配时间上具有一定优势,但其匹配质量明显不足。综合考虑匹配精度、匹配完整性、有效匹配点数量以及运行效率,VGGSFM 更适合作为文中热红外图像三维重建任务中的图像配对前端。

2) 可视化评价

图 2 展示了三种图像配对方法在热红外变电站设备场景中的匹配效果对比。从图中可以直观看出,ORB 方法得到的匹配点对数量较少,且存在一定误匹配现象;SIFT 方法相较于 ORB 在匹配数量和稳定性方面有所提升,但在设备主体中部等纹理稀疏区域中,其匹配点仍不够密集且存在漏匹配问题;相比之下,文中方法不仅提升了匹配点数量,而且显著增强了弱纹理区域的匹配完整性,使得设备关键结构区域都能够获得较为连续和准确的对应关系。因此,文中最终选用 VGGSFM 作为图像配对与位姿恢复的前端方法。

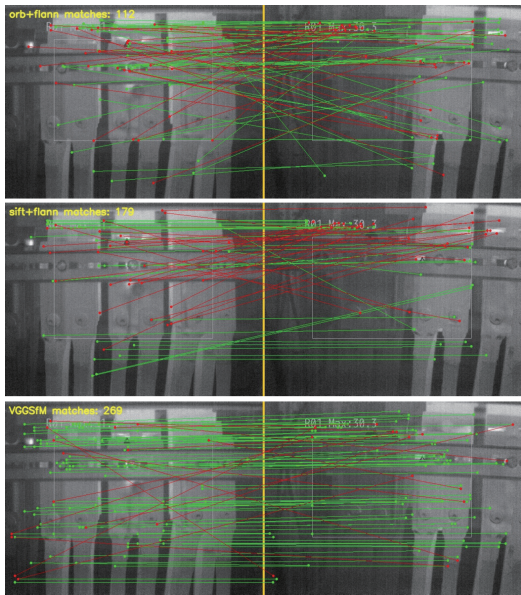


图 2 特征点匹配可视化

3.3 图像三维重建实验

为验证文中所提方法在变电站电力设备热红外场景三维重建任务中的有效性,选取 COLMAP、NERF 以及文中基于 VGGSFM 初始化并结合改进 3DGS 的方法进行对比实验。为进一步量化差异,采用 PSNR、SSIM、LPIPS 以及重建时间作为评价指标。结果如下:

从表 2 的 PSNR 指标可以看出,改进 3DGS 的数值达到 28.91 dB,明显高于 COLMAP 的 18.74 dB 和 NERF 的 24.63 dB,说明其在热红外电力设备场景中能够更准确地恢复图像的灰度分布和局部细节;从 SSIM 指标来看,改进 3DGS 同样优于 NERF 和 COLMAP;从 LPIPS 指标来看,改进 3DGS 的结果为 0.128,低于 NERF 和 COLMAP。这些数据表明该方法所重建出的模型在主观视觉效果上更接近真实观测结果,如图 3 所示。

为了进一步评估不同方法的工程可用性,表 2 给出了三种方法的重建时间对比结果。可以看出,NERF 的重建时间最长,达到 15 h 48 min;COLMAP 的重建时间为 1 h 12 min,明显快于 NERF;相比之下,改进 3DGS 的重建时间仅为 26 min,在三者中时间最短。该结果说明,3DGS 显著降低优化与渲染过程中的时间开销,已经具备了搭建数字孪生可视化平台要求的实时性能力。

表 2 重建性能指标对比

算法	PSNR/dB	SSIM	LPIPS	重建速度
COLMAP	18.74	0.548	0.386	1 h 12 min
NERF	24.63	0.816	0.241	15 h 48 min
改进 3DGS	28.91	0.921	0.128	26 min



图3 电力设备及母线构架的3DGS重建结果

综上所述,基于VGGSfM初始化并结合改进3DGS的方法在保真度、结构一致性、感知质量和重建效率等方面均表现最佳,是更适合该类场景的三维重建方案。

4 结语

针对热红外图像弱纹理、低对比度条件下电力设备三维重建精度不足、重建结果不完整以及重建效率受限等问题,文中提出了一种基于VGGSfM与改进3DGS的热红外电力设备三维重建方法,并完成了相关实验验证,结论如下:

采用VGGSfM完成多视角热红外图像的相机位姿恢复与初始稀疏点云重建,较好解决了传统方法在弱纹理热红外场景下特征提取困难、匹配不稳定的问题,为后续三维重建提供了可靠的几何基础和经验。实验结果表明,VGGSfM的综合性能优于SIFT和ORB,更适合用于热红外电力设备三维重建,而改进3DGS在PSNR、SSIM、LPIPS以及重建速度等指标上均表现更优,说明该

方法更适合热红外电力设备场景下的三维建模任务。

参考文献:

- [1] 胡钰泊,李厚君,游桂杨.基于YOLOv8的电力设备红外图像检测[J].电气技术与经济,2026(01):222-226.
- [2] CHI H, LUO D L, WANG S. LMD Fusion: A lightweight infrared and visible image fusion network for substation equipment based on mask and residual dense connection[J]. Infrared Physics & Technology, 2024, 138: 105218.
- [3] 张秋铭,李云红,罗雪敏,等.改进Chan-Vese模型的电力设备红外图像分割算法[J].红外技术,2023,45(02):129-136.
- [4] Skorokhodov V, Xu C, Sun S, et al. SEAR: Simple and efficient adaptation of visual geometric transformers for RGB+Thermal 3D reconstruction[EB/OL]. arXiv, 2026: arXiv: 2603.18774.
- [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates, Inc., 2017, 30: 5998-6008.
- [6] SENTENAC T, BUGARIN F, DUCAROUGE B, et al. Automated thermal 3D reconstruction based on a robot equipped with uncalibrated infrared stereovision cameras[J]. Advanced Engineering Informatics, 2018, 38: 203-215.
- [7] KERBL B, KOPANAS G, LEIMKÜHLER T, et al. 3D gaussian splatting for Real-Time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(04): 1-14.
- [8] WANG J Y, KARAEV N, RUPPRECHT C, et al. VGGSfM: Visual geometry grounded deep structure from motion[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA: IEEE, 2024: 21686-21697.