

内窥镜视频关键帧获取及三维重建方法研究

薛怀琦,王双园,张玉荣,姚志远

上海理工大学 机械工程学院,上海 200093

摘要:目的 鉴于内窥镜在获取人体内部图像时所面临的数量和质量限制,提出一种新的高清晰度内窥镜视频帧抽取策略三维重建方法。方法 首先,采用基于结构相似性指数(SSIM)和 Laplace 算子复合方法筛选和提取具有高清晰度和低噪声的关键帧;然后,利用基于 Neighbor2Neighbor 深度学习方法对初步的关键帧队列进行处理,减少图像中的噪声并增强其质量,并利用 COLMAP 增量式 SFM 技术将优质关键帧转化为稀疏点云数据;最后,运用 OpenMVS 深度图融合技术进行稠密重建和曲面拟合,从而获得高精度的三维模型。结果 实验结果表明:所提方法不仅显著增强了图像特征匹配度,而且大大提升了三维重建的准确性和精度。结论 该方法不仅显著提高了图像特征匹配的精确度,而且提升了三维重建的准确性和精度,预计将为临床医生提供更直观的诊断依据。

关键词:内窥镜图像;Neighbor2Neighbor;增量式 SFM;深度图融合算法

中图分类号:TP391 文献标识码:A doi:10.16055/j.issn.1672-058X.2026.0001.007

Research on Keyframe Acquisition and 3D Reconstruction Methods for Endoscopic Videos

XUE Huaiqi, WANG Shuangyuan, ZHANG Yurong, YAO Zhiyuan

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract: Objective In view of the limitations in quantity and quality when endoscopes acquire internal human body images, this study proposes a new high-definition endoscope video frame extraction strategy for 3D reconstruction.

Methods First, a composite method based on the Structural Similarity Index Measure (SSIM) and the Laplace operator was used to screen and extract key frames with high definition and low noise. Then, the Neighbor2Neighbor deep-learning method was used to process the preliminary key-frame queue to reduce noise in the images and enhance their quality. The COLMAP incremental Structure-from-Motion (SFM) technique was employed to convert high-quality key frames into sparse point-cloud data. Finally, the OpenMVS depth-map fusion technology was used for dense reconstruction and surface fitting to obtain a high-precision three-dimensional model. **Results** The experimental results showed that the proposed method not only significantly enhanced the image feature matching degree but also greatly improved the accuracy and precision of three-dimensional reconstruction. **Conclusion** This method not only significantly improves the accuracy of image feature matching but also enhances the accuracy and precision of three-dimensional reconstruction. It is expected to provide clinicians with a more intuitive diagnostic basis.

Keywords: endoscopic image; Neighbor2Neighbor; incremental SFM; depth-map fusion algorithm

收稿日期:2024-01-23 修回日期:2024-04-09 文章编号:1672-058X(2026)01-0055-09

基金项目:海洋智能装备与系统教育部重点实验室开放基金(MIES-2020-05)项目资助。

作者简介:薛怀琦(1999—),男,江苏靖江人,硕士研究生,从事机器视觉研究。

通信作者:王双园(1986—),男,安徽合肥人,博士,从事机器学习研究。Email:wsyusst@163.com。

引用格式:薛怀琦,王双园,张玉荣,等.内窥镜视频关键帧获取及三维重建方法研究[J].重庆工商大学学报(自然科学版),2026,43(1):55-63.

XUE Huaiqi, WANG Shuangyuan, ZHANG Yurong, et al. Research on keyframe acquisition and 3D reconstruction methods for endoscopic videos[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(1): 55-63.

内窥镜为一种能够透过狭小路径深入到身体各个组织进行检查的设备,它的诞生与进步打破了视觉界限,给身体健康检查带来了许多方便。尽管内窥镜可以捕捉图像,但是由于探头的大小限制,它的视线范围极为狭窄,仅能看见局部的二维图像。这对于需要识别复杂结构的重要疾病区域来说,并不便于进行深入解读和评估。因此,在复杂胃肠道环境中,运用内窥镜视频帧抽取和三维重建技术,能够更有效地揭示病灶的具体位置、大小等关键信息,从而提升疾病诊断与治疗精准度。目前,光学三维成像技术主要包括距离选通法^[1]、结构光法^[2]、激光雷达法^[3]、立体视觉法^[4]和运动中恢复结构法^[5]等。使用单目相机内窥镜获取人体内部的光学影像信息是目前的主流方式。相较于双目内窥镜,单目内窥镜在实际探测过程中成本较低、易于使用,且不需要特别校准。此外,相较于其他方法,从运动中恢复结构的算法(Structure From Motion, SFM)可以从运动中恢复空间位置关系,能够适应更多的环境和场景,使用 SFM 方法可以更好地对人体内复杂结构进行精准重建。

SFM 方法已经广泛应用于相机运动跟踪和三维重建研究中^[6]。在医疗领域内,使用内窥镜的环境独特之处:一是病灶处的三维重建需要大量不同角度的图像,但通常内窥镜不能长时间在人体内停留;二是内窥镜自带的光源容易出现照明不足或者过强的情况^[7];三是人体器官中的体液粘在镜头表面时,会使拍摄得到的图像畸变较高。Chen 等^[8]使用 3D AlexNet 模型对前列腺图像进行预处理以提取图像特征,在分割图像后实现了对前列腺肿瘤的重建,但是简单地将 2D 分割结果拼接成 3D 分割结果可能会导致锯齿状和错误;Zhuang 等^[9]使用动态规划算法对上呼吸道进行自动分割,应用解剖弯曲方法进行三维重建,但是传统的数字处理(Digital Processing, DP)通常会将受斑点噪声或过饱和 A 线污染的图像区域误识别为正常气道腔结构的可能区域,从而导致异常的分割结果;刘瑶等^[10]提出一种新的卷积神经网络模型 RetKIND,用于低光图像的增强,但是在部分数据集上效果不佳,在 SSIM 指标等方面仍有改进空间。以上工作的缺点:目前大规模使用的单目内窥镜无法加装额外的传感器,无法消除内窥镜采集到的图像噪声和畸变。

考虑以上问题,本文提出一种创新的解决方案,针对单目内窥镜医学影像,旨在进行二维影像到三维结构的完整转换方法。本文所进行的工作如下:

(1) 使用内窥镜录制人体内部视频,利用基于拉

普拉斯算子的图像清晰性度量算法剔除模糊视频帧,再使用结构相似性指数(SSIM)来滤去相似帧,得到用于三维重建的初始关键帧序列。

(2) 利用一种 Neighbor2Neighbor 无监督训练图像处理方法,针对初始关键帧序列训练特定的降噪模型,提高图像特征点匹配结果的准确性和匹配数量,降噪后得到的关键帧序列作为三维重建的关键帧数据集。

(3) 使用 COLMAP 的增量式 SFM 技术生成关键帧稀疏点云,采用基于深度图融合算法的 OpenMVS 平台将多个视角下的深度图进行融合,经过稠密重建和曲面拟合得到三维模型。

1 内窥镜图像三维重建方法研究

1.1 内窥镜三维重建方法

本文提出的内窥镜三维重建方法分为如下步骤:

(1) 内窥镜视频获取。利用内窥镜设备获取目标器官的视频序列。

(2) 高清晰度关键帧提取。自适应帧采样是根据帧间的结构相似性指数(SSIM)自适应调整帧采样间隔,避免重复获取相似帧。剔除模糊帧是利用拉普拉斯算子计算图像清晰度,剔除模糊帧,得到高质量的关键帧序列。

(3) 图像特征点匹配优化。利用 Neighbor2Neighbor 深度学习模型,从单张含噪图像中生成相似噪声图像,用于训练降噪模型。通过正则化损失函数训练模型,实现对原始关键帧的降噪和增强。

(4) 关键帧三维重建。利用 COLMAP 进行增量式 SFM,根据关键帧间的特征匹配关系,逐步重建稀疏点云。利用 OpenMVS 进行稠密重建和曲面拟合,生成高质量的三维模型。流程如图 1 所示。

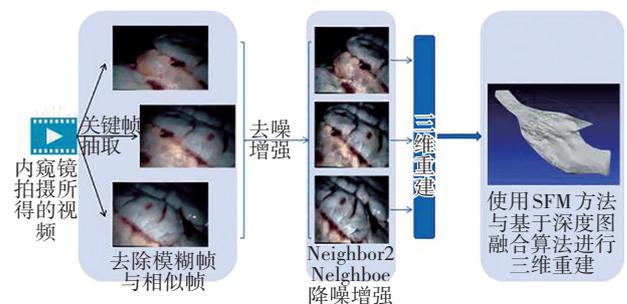


图 1 内窥镜图像三维重建方法

Fig. 1 3D reconstruction method of endoscopic images

1.2 高清晰度关键帧提取方法

使用 SFM 方法进行三维重建需要获得被重建物体不同角度的照片。当前医疗机构普遍使用的内窥

镜都不能长时间停留在人体内,短时间内获得不同角度的大量图片较为困难。常用的解决方法是利用内窥镜视频流进行抽帧获得图像帧序列,由于视频编码过程中会采用各种算法来减少数据量,在低比特率的情况下,为了保持流畅度,编码器会牺牲部分帧的质量,导致某些帧模糊;编码器会利用前后帧的相似性来进行预测编码,降低数据量,导致某些帧相似^[11]。用包含模糊帧和相似帧的图像序列进行三维重建时会导致精度下降、噪声增加、性能降低和完整性不足。本文提出一种高清晰度关键帧获取方法,利用内窥镜视频进行抽帧并剔除模糊帧与相似帧后获得初始关键帧列表。

图 2 展示了整体流程。流程中, C 代表图像的清晰度, C_1 是清晰度的临界值; S 用于衡量图像间的相似度,其中 S_l 和 S_h 分别是相似度的下限和上限阈值;step 是指定视频帧抓取的间隔;而 K 代表选定的关键帧。整个处理流程主要分为两个阶段:

(1) 自适应间隔采样。根据连续帧之间的相似度来自动调整采样间隔,这种方法能最小化无谓的帧分析和计算工作,同时确保以最快的速度捕获到视频中的高清帧。假设相机匀速运动,按固定步长提取关键帧,一旦发现图像相似度增大则需增加步长,反之减小步长;同时采用结构相似性指数 SSIM (Structural Similarity Index),来度量相似性^[12],SSIM 取值范围为 -1 到 1,1 表示完全相同,0 表示没有相似性,-1 通常表示负相关。实际应用中重点关注正值区间,因为它们表示图像之间的结构和纹理相似度。为了提高关键帧的视觉质量和有效利用计算资源,本文选择一个经验性的 SSIM 阈值为 0.95,这个阈值是根据大量实验和经验得出的,可以较好地平衡视觉质量与计算资源。太低的阈值会导致提取过多关键帧,增加计算量;太高的阈值会遗漏重要视觉信息,影响重建质量。将低于 SSIM 阈值的视频帧生成初始关键帧队列。若是整个帧中都没有达到 0.95,则采用运动检测和场景深度判断,运动变化与深度变化较大的帧会被提取出来作为关键帧。

(2) 剔除模糊视频帧。使用基于拉普拉斯梯度函数的图像清晰度度量算法来保证图像的清晰度。本文采用一个简化的模糊图像检测方法,将灰度图像的方差阈值设定为 100 作为判断依据。若图像的方差低于预设阈值,则判定该图像为模糊;反之,则认为图像清晰。图像模糊时,则跳过当前图像,将清晰的图像生

成用于三维重建的关键帧列表。

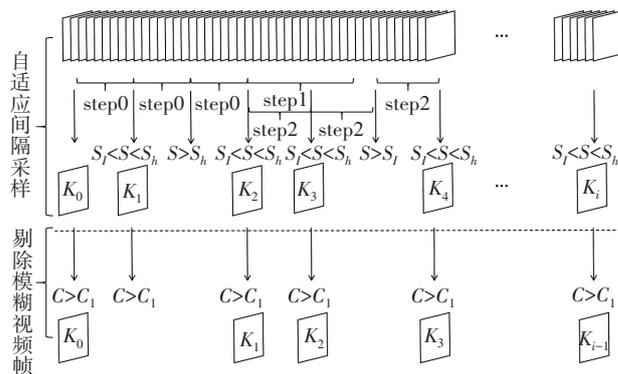


图 2 自适应帧采样流程

Fig. 2 Adaptive frame sampling process

自适应间隔采样过程中实时监测相邻关键帧之间的 SSIM 值变化,SSIM 主要考量图片的 3 个关键特征:亮度 (Luminance)、对比度 (Contrast)、结构 (Structure):

$$S(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

式(1)中, μ_x 和 μ_y 代表图像 x 和图像 y 的平均亮度,通过平均所有像素的值得到; σ_x^2 和 σ_y^2 代表图像 x 和图像 y 的亮度方差; σ_{xy} 是图像 x 和图像 y 的亮度协方差; c_1 和 c_2 是小的常数,用于避免方差除法运算中出现数值不稳定的情况;结构对比比较的是经过归一化后 $(x - \mu_x) / \sigma_x$ 与 $(x - \mu_y) / \sigma_y$ 的比值。

这种自适应调整步长的机制满足 3 个原则:一是当图像内容变化较大时,相似度减小,通过减小步长为帧率的一半来更频繁地提取关键帧,以确保不遗漏重要视觉信息;二是当图像内容变化较小时,相似度增大,通过增大步长为帧率的一半来减少关键帧的提取,从而节省计算资源和存储空间;三是设置一个步长的限制,即步长的最小值为帧率,确保至少每帧都会被检查一次,将满足相似性指数要求的视频帧加入初始的关键帧队列。

接着处理得到的初始关键帧队列。利用 Laplace 算子计算图像各像素点的梯度,对图像进行卷积操作,剔除模糊图像,增强图像的边缘和细节信息,得到最终的关键帧队列。Laplace 算子计算如式(2):

$$\Delta\sigma_{ij} = \frac{\partial^2\sigma_{ij}}{\partial u^2} + \frac{\partial^2\sigma_{ij}}{\partial v^2} \quad (2)$$

式(2)中, $\Delta\sigma_{ij}$ 表示图像中某个像素点的梯度; σ_{ij} 代表图像中某个像素点的灰度值; u 和 v 是图像平面上的坐标,其本质就是在 (u, v) 点处求二阶导数;本文将 K_{size} 设为 1,即代表在计算相邻像素的差异时,所有像素都被均等

地对待。鉴于此,使用式(3)中的 kernel 进行滤波:

$$k = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (3)$$

这个卷积核是基于拉普拉斯算子的概念和图像处理中的经验设计的,它考虑了像素在水平和垂直方向上的二阶导数。中心像素取-4的原因是因为它需要反映周围像素对该像素的影响,确保当图像在中心像素处有强烈的边缘时,拉普拉斯值会有较大的正值或负值,从而在边缘检测时突出这些区域。

将梯度强度低于 100 的像素视为模糊区域。对于每一个像素,如果其梯度强度低于阈值,则将其置为 0,否则保留原始灰度值。若整个图像 70% 的区域都被标记为模糊,可以认为这张图像就是模糊的,需要被剔除。

1.3 图像特征点匹配优化方法

本文用 Neighbor2Neighbor 模型消除关键帧中的干扰物反光和内窥镜光源直射产生的亮块,减小对于特征提取以及匹配准确性的影响。传统基于监督学习的方法需要用无噪声(Noisy-Clean)图像训练,但在医学真实场景中,比较难以获取 Noisy-Clean 的图像对。本文采用的 Neighbor2Neighbor 使用与 ground truth 相似的图片进行训练,可以在无需 Noisy-Clean 图像的条件下进行内窥镜图像的去噪和增强,不同深度学习模型可以针对不同的胃部环境进行训练。相较于传统利用高斯滤波和直方图均衡化处理的图像而言,利用深度学习的图像处理方法在处理胃部环境图像时更具有针对性。

Neighbor2Neighbor^[14]是一种无需干净图像即可训练降噪网络的方案,通过对 Wang 等^[15]提出的 Noise2Noise 方法进行推广,实现了仅用单张含噪图像及对应的相似含噪图像就能进行有效训练。在 Noise2Noise 框架中,通过比较同一对象或场景的不同噪声观测来学习降噪模型。然而,Neighbor2Neighbor 进一步创新了这个理念,通过特定采样器,从单幅含噪图像中生成具有相似噪声特征的对图像。这种方法允许在没有大量配对噪声数据的情况下,有效地训练和优化降噪网络,从而在各种图像降噪任务中实现高性能的表现。通过引入正则项,克服了因为相似含噪图像的采样位置不同,从而导致得到的图像过于平滑的问题。Neighbor2Neighbor 的训练和推理流程如图 3 所示。

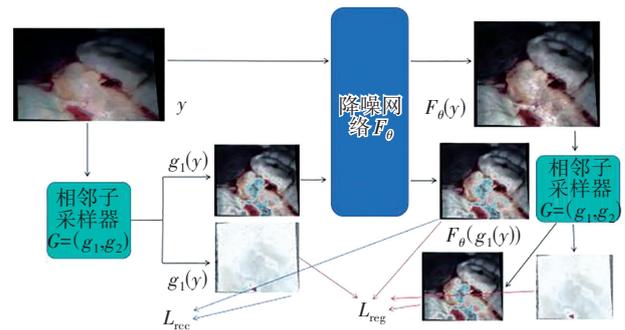


图 3 Neighbor2Neighbor 方法训练流程

Fig. 3 Training process of Neighbor2Neighbor method

相邻子采样器是一种图像处理采样技术,它主要用于从原始图像中提取或生成相邻的、大小或分辨率不同的子图像。带有相邻子采样器 $G=(g_1, g_2)$ 的噪声图像 y 生成一对子采样图像 $(g_1(y), g_2(y))$ 。去噪网络分别以 $g_1(y)$ 作为输入,并以 $g_2(y)$ 作为目标进行训练。训练过程采用一种正则化损失 L ,包括以下两项:左侧为网络输出与带噪声目标之间的重构项 L_{rec} ,这是网络输出与带噪声目标之间的差异度量。网络的目标是尽可能地还原输入图像的无噪声版本,因此需要最小化网络输出与目标 $g_2(y)$ 之间的差异。在右侧,考虑由同一相邻子采样器 G 生成的子采样噪声图像与 $(g_1(y), g_2(y))$ 之间的真实像素值可能存在本质差异,进一步添加了正则化项 L_{reg} 来约束这种差异。出现两次的相邻子采样器 G (蓝色)代表同一个相邻子采样器。

$$L = L_{rec} + \gamma \cdot L_{reg} = f_{\theta}(g_1(y)) - g_2(y) + \gamma \cdot (f_{\theta}(g_1(y)) - g_2(y) - (g_1(f_{\theta}(y)) - g_2(f_{\theta}(y))))^2 \quad (4)$$

式(4)中, L_{rec} 是网络输出和噪声目标之间计算的重构项; L_{reg} 是正则项; γ 是控制正则项强度的超参数。在训练开始时,可以将 γ 初始化为一个较小的正值,本文设为 0.1,使训练初期正则项的影响相对较小。为了稳定学习,停止 $g_1(f_{\theta}(y))$ 和 $g_2(f_{\theta}(y))$ 的梯度,并在训练过程中逐渐增加 γ 到指定值。总的训练轮数被设置为 300,正则化项参数的增长比例设为 2.0,正则化项参数动态调整,遵循式(5):

$$L = e/300 * 2.0 \quad (5)$$

式(5)中, L 为正则化项参数, e 为当前训练轮数。

采样器 G 使用近邻采样,图像被分为多个 2×2 的像素区块。从每个这样区块的 4 个像素中随机选取两个相邻像素分配至两个不同的子图中,从而构建出一对“在细节上有所差异但整体保持相似性”的子图,将这两张子图定义为“Neighbor”,如图 4 所示。

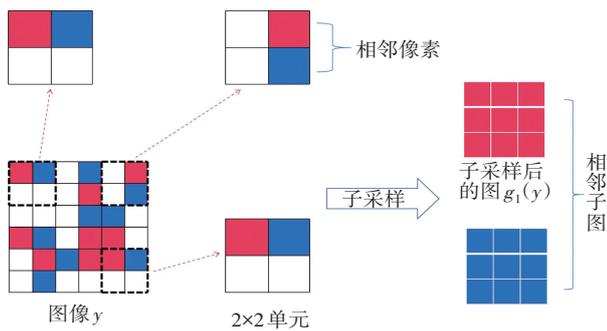


图 4 生成具有相邻子采样器的图像对示例

Fig. 3 An example of generating image pairs with adjacent sub-samplers

因为在每个 2×2 的区块中随机选择像素,所以每个子图都会丢失一些像素信息,这些丢失的信息保证了两个子图在细节上的差异性;由于采样是在整个图像上均匀进行的,因此两个子图在整体上仍然保持了较高的相似性;随机采样能增加训练数据的多样性。

1.4 关键帧序列三维重建方法

1.4.1 利用增量式 SFM 技术进行稀疏重建

稀疏重建的主要实现手段是 SFM,基于 SFM 重建的基本原理是通过采集设备的位置移动,对静态场景拍摄不同视角的图像序列,求解相机的位置和姿态,进而恢复场景的三维信息^[16]。SFM 通常被划分成增量式、全局式以及混合式。本文采用的是增量式 SFM 方法,该方法有能处理较大视差变化和丰富特征匹配点的优势。

增量式 SFM 工作流程:首先,在两个关键帧上执行特征检测和匹配,以确立它们之间的几何关系;然后,选择特征匹配质量较高的两帧作为起始点,采取随机采样一致性(RANSAC)方法剔除不准确的匹配点;此后,在数据集基础上逐步引入新图像,每次迭代都会计算新增图像的相机位置与姿态,通过三角化算法更新三维点云;每次添加新图像并重建后均会执行光束平差(Bundle Adjustment, BA)优化步骤,以减少累积误差,增强整体解决方案的鲁棒性;最终所有图像都被完全添加,实现了对整个场景的完整恢复。其技术流程如图 5 所示。

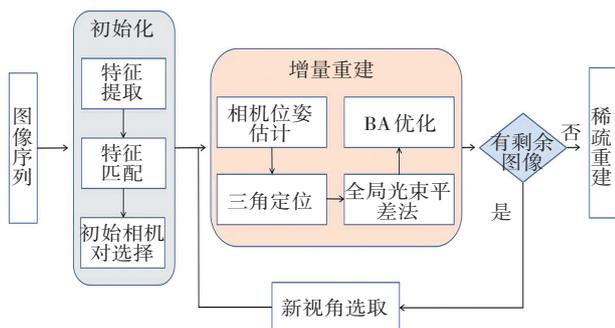


图 5 稀疏点云的生成流程

Fig. 5 The generation process of sparse point cloud

通常采用 Levenberg-Marquardt^[17] 算法进行全局光束平差法优化,以减少误差。光束平差法是为多个视角的三维空间点以及相机位姿找到全局视觉一致解决方案的过程。光束平差法通过调整所有相机投影矩阵和三维点来最小化给定像素坐标点和重投影点之间的整体误差。以下是等效的光束平差法最小化方程:

$$\min \sum_{i,j} d(x'_{ij}, x_{ij})^2 = \min \sum_{i,j} d(\mathbf{P}'_j \cdot \mathbf{X}'_i, x_{ij})^2 \quad (6)$$

式(6)中, x_{ij} 为第 j 幅图像的第 i 个二维点,利用内窥镜的摄像机矩阵,求得最大似然相机投影矩阵 \mathbf{P}'_j 和最大似然三维点 \mathbf{X}'_i ,使重投影图像点 x'_{ij} 与给定图像点 x_{ij} 尽可能接近, $d(a, b)$ 为点 a 与 b 之间的图像距离。

1.4.2 利用深度图融合算法生成稠密点云

多视图立体重建(Multi-View Stereo, MVS)是一种计算机视觉和计算机图形学领域的技术,用于从多个视角图像中还原出三维场景的形状和结构。相比于其他 MVS 算法,基于深度图融合的 MVS 算法通过融合多个图像的深度信息,可以生成非常稠密的点云,从而提供更详细的几何形状;并且通过融合多个视角信息,该方法可以减少单一视角引起的误差,提高最终重建的准确性。

所有的深度图生成方式都存在原始深度图上出现空洞、物体边缘处深度信息缺失和噪声影响的情况,这样的情况会损坏深度图的质量,在点云中产生较多错误点,所以本文首先通过深度图滤波方式优化初始深度信息^[18]。在将深度信息转换为三维空间点云时,多个图像点可能被映射到不同的空间位置,导致点云数据的冗余。本文采用如下方法来提高点云的准确性并减少冗余:

(1) 分析不同相机视角间的几何关系,为成对的二维图像建立多组单应性矩阵,用于描述两个视角之间的投影变换。

(2) 利用单应性矩阵识别出图像对之间的对应点,同时分析对应点之间的视差误差分布。

(3) 基于误差统计特性设计目标函数,最小化这些误差,并使用最大似然估计方法(MLE)优化目标函数,从而得到误差分布的最佳参数。

(4) 利用从 MLE 获得的参数来构建一个误差概率置信模型,用于评估每个三维点的可信度,并且在处理点云时过滤掉冗余数据,从而提高点云的质量^[19]。

文中采用的基于深度图融合技术的多视图立体匹配(MVS)算法,基本流程如图 6 所示。

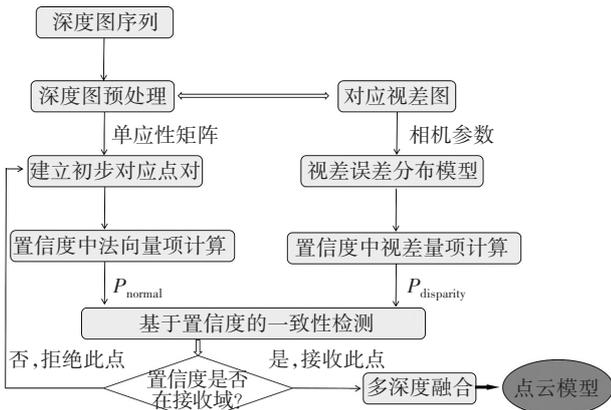


图 6 基于深度图融合的 MVS 流程

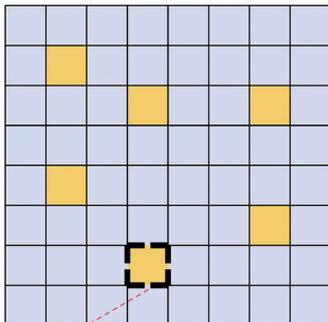
Fig. 6 Process of MVS based on depth map fusion

用联合双边滤波方法对深度图进行滤波预处理, 采用图像纹理图的信息 I 来增强深度图信息 D , 输出高质量的深度图 D_p , 具体过程如下式:

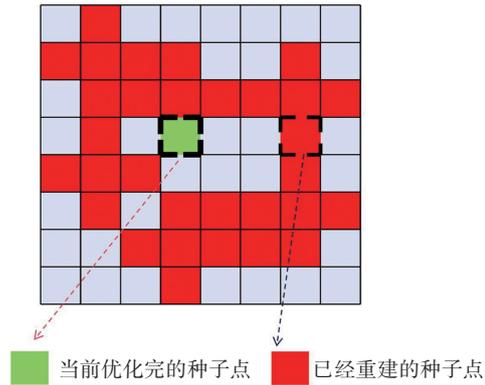
$$D_p = k_p \sum_{q' \in \Omega} \omega_s(p', q') \omega_r(I_p, I_q) D_{q'} \quad (7)$$

式(7)中, k_p 代表标准化系数, 用于平衡不同因素的影响; Ω 代表插值选取窗口的大小; p' 和 q' 是低质量深度图里像素的位置, p 和 q 则是对应 p' 和 q' 的高质量深度图里像素的位置; ω_s 和 ω_r 代表像素空间距离项和颜色空间距离项的权重, 用于在计算过程中调整这两个因素对最终结果的影响。

计算种子点优化后结果的置信度, 按照置信度高 低建立队列, 对队列中的点进行优化, 并将这些点的周围点加入队列。如图 7(a) 所示, 把黄色种子点深度和法向量值赋给相邻的上下左右 4 个像素点, 作为它们的深度和法向量初值, 并将这些点加入队列进行优化。如图 7(b) 所示, 当前优化完的种子点 (绿色) 四周有已经被优化过的点 (红色) 和未被优化过的点 (蓝色), 比较绿色点与红色点的置信度大小, 如果绿色点的置信度高, 则把绿色点的信息赋给红色点。图 7(b) 中, 两类领域点将被选为种子点: 没有没重建过的点和置信度高于已经重建的点。



(a) 初始阶段的种子点选取



(b) 种子点优化和重建过程

图 7 像素点深度与法向量信息恢复

Fig. 7 Recovering pixel depth and normal vector information

2 三维重建实验及结果分析

本文通过实验验证了提出的三维重建方法。利用医用内窥镜对具备已知参数的 3D 打印胃部模型实施重建, 便于验证重建结果的准确性。内窥镜为单目内窥镜, 采用奥林巴斯 CV-290 电子胃肠镜系统。实验的软件环境包括使用 Pycharm 2022. 3. 1 版本, 在 64 位的 Windows 11 操作系统上进行, 编程语言为 Python, 硬件采用 AMD(R) Core(TM) R7-5800H CPU @ 3. 20 GHz, 32. 00GB 内存。此外, 实验还利用基于 RTX-3060 GPU 的 Nvidia CUDA 11. 0 并行计算工具包来实现算法的加速。实验设备如图 8 所示。

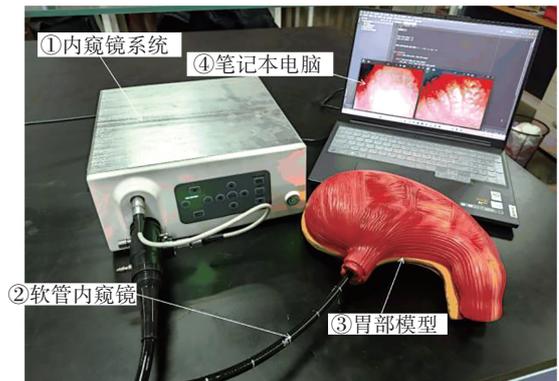


图 8 内窥镜视频流采集设备

Fig. 8 Endoscopic video stream acquisition device

2. 1 利用内窥镜的胃部模型重建实验

本文使用 3D 打印技术仿造人体胃部设计了一个 1 : 1 大小的胃部模型, 总长为 30 cm(从贲门到幽门的距离), 胃体部分最大宽度为 15 cm(胃在中等程度充盈时的横截面尺寸), 胃壁平均厚度约为 3 mm。在胃壁上喷涂医用级硅胶模拟胃壁的皱襞、血管分布和不同层的厚度差异, 使用奥林巴斯 CV-290 内窥镜对胃部模型进行拍摄, 得到分辨率为 720 * 480 pixel 的图像序列。

本文算法获得初始关键帧序列后,用 Neighbor2Neighbor 方法对初始高清晰度关键帧序列进行去噪和增强。实验选取 516 张关键帧作为深度学习的样本,将图像随机裁剪为 256×256 大小的子图作为模型的输入。采用 3e-4 的初始学习率,并采用学习率衰减策略,每训练 20 个 epoch 后学习率衰减为原来的 0.5 倍;模型的批量大小设置为 4;损失函数中的权重参数分别设置为 $\lambda_1 = 1.0, \lambda_2 = 1.0$;损失函数中的增加比例参数设为 2.0;模型总共训练了 300 个 epoch,以确保网络充分收敛。

得到三维重建所用的最终高清晰度关键帧序列后进行三维重建,重建流程和结果见图 9 所示。

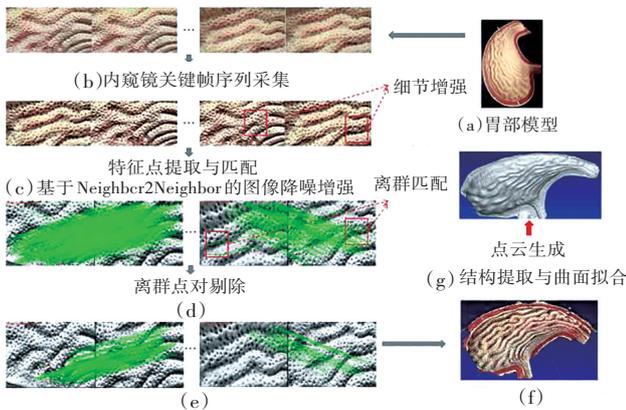
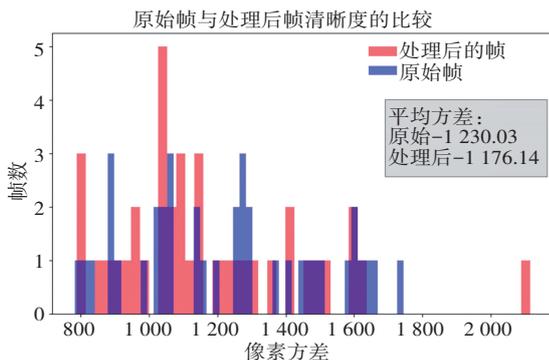


图 9 胃模型重建过程与结果

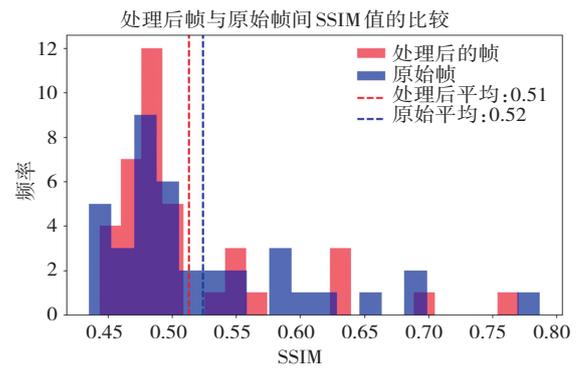
Fig. 9 Reconstruction process and results of stomach model

2.1.1 高清晰度关键帧获取

实验设置 SSIM 阈值为 0.95,方差阈值为 100。本文将采用高清晰度关键帧获取方法所得的关键帧序列与采用普通方法抽帧所得的图像序列进行对比,图 10 展示了两个图像序列之间的相似度和清晰度对比结果。图 10 中可以明显看出:经过本文高清晰度关键帧获取算法处理后的关键帧序列,相比于普通抽帧方法处理得到的图像集有更低的像素平均方差和 SSIM 平均值,这代表得到的关键帧序列整体清晰度更高,相似度更低。



(a) 原始帧与处理后帧清晰度的比较柱状图



(b) 处理后帧与原始帧间 SSIM 的比较

图 10 清晰度与相似度的比较

Fig. 10 Comparison of clarity and similarity

2.1.2 图像特征点匹配优化

本文利用基于 Noise2Noise 方法改进而来的 Neighbor2Neighbor 方法,通过训练一个专门针对当前数据集的降噪模型从而实现对不同病灶视频流抽帧所得数据集的精准降噪。通过比较图 9(b)和图 9(c),可以明显看出:在使用本文提出方法对图像进行增强处理后,纹理特征有了显著的改善,许多细小的纹理细节都得到了清晰地展现。图 11 展示了在未经处理的原始图像上进行特征匹配的结果。与图 9(d)所示的图像进行对比,可以明显发现:在使用本文提出的图像增强和降噪算法进行处理后,提取到的特征点数量相比于未处理的原始图像有了显著的增长。

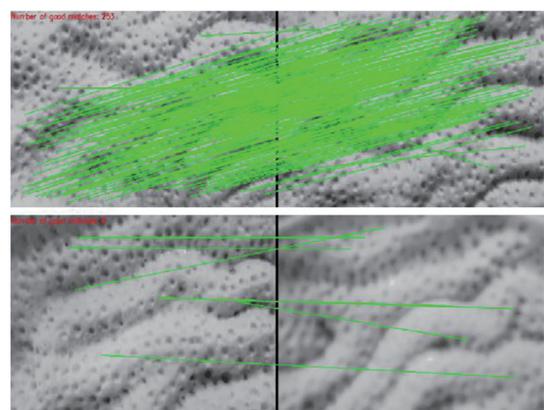


图 11 比较特征点匹配效果

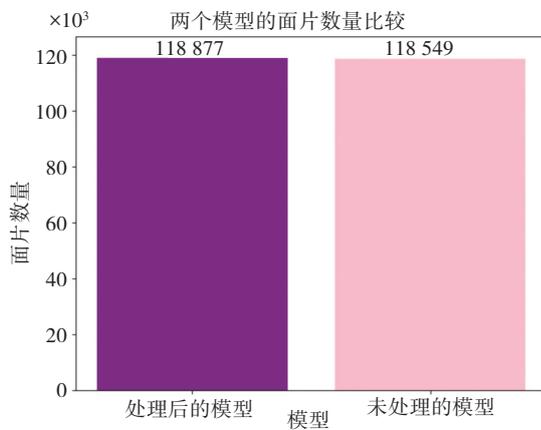
Fig. 11 Comparison of feature point matching results

2.2 结果分析

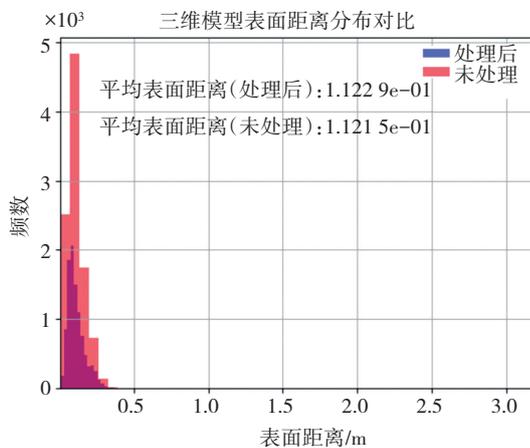
为了验证本文所提出方法在提升三维重建精度方面的优越性,分别利用经过本文方法处理得到的高清晰度关键帧序列和普通方法处理得到的图集对胃部模型进行了三维重建,重建得到的三维模型局部如图 12 所示。图中从定量的角度对该结论进行了验证,对两种重建模型计算面片数量 (Polygon Count, PC) 及表面平均距离 (Average Surface Distance, ASD) 进行了对比。

结果显示:应用本文方法重构的三维胃部模型在面片使用效率和表面拟合精确度上均显著优于常规方法。

结合表 1 可见:相比于普通方法,本文所提出的算法在略微增加时间消耗的情况下,重建后的胃部模型面片数有了明显增长,这意味着模型能更好地表现复杂几何结构;同时,表面平均距离也得到了显著提高,这直接反映了模型在整体形态和局部特征上与原始解剖结构的贴合程度更高,从而有效提升了重建模型的精确性。实验结果表明本文方案对于单目内窥镜适用。



(a) 两个模型的面片数量比较



(b) 三维模型表面距离分布对比

图 12 胃模型三维重建结果对比

Fig. 12 Comparison of 3D reconstruction results of stomach models

表 1 胃模型结构提取算法的时间效率对比

Table 1 Comparison of time efficiency of the structure extraction algorithms for stomach models

模型	用时 /s	模型面片数量	平均表面距离/m
高清晰度关键帧三维重建	1 561	118 877	1.1229e-01
普通视频帧三维重建	1 223	118 549	1.1215e-01

3 结论

考虑现有单目视觉三维重建方法在没有额外辅助设备的情况下很难适用于内窥镜环境,本研究提出一种全新的解决方案,该方案利用单目内窥镜对医学影像进行三维重建。这种方法具有以下几个特点:

(1) 利用基于拉普拉斯算子的图像清晰度度量算法剔除模糊视频帧,再使用结构相似性指数(SSIM)来滤去相似帧,克服了传统方法抽帧得到的图集中掺杂低质量图片的缺陷。

(2) 引入一种基于深度学习无监督训练的图像处理方法 Neighbor2Neighbor,对获得的预处理数据集进行降噪和增强,以减小杂质反光和内窥镜自带光源直射造成图像中高亮区域对特征提取和匹配的干扰。

(3) 最后,在使用 COLMAP 提供的增量式 SFM 技术生成稀疏点云后,采用基于深度图融合算法的 OpenMVS 平台进行稠密重建和曲面拟合生成,得到最终的三维模型。

实验结果表明:与利用普通帧进行三维重建相比,本文所提出的方法取得了较为明显的精确性提升,总体性能更优秀。

本文提出的方法在提高内窥镜三维重建精度方面取得了一定效果,但也存在一些不足之处。在关键帧提取阶段,仅利用图像的清晰度和结构相似性,未能充分考虑图像内容信息;在图像增强阶段,采用的 Neighbor2Neighbor 方法需要大量训练样本,样本不足时效果可能受限;在三维重建阶段,COLMAP 和 OpenMVS 算法对图像质量和特征匹配效果要求较高,本文方法在这 3 方面仍有提升空间。

参考文献(References):

- [1] 王新伟,孙亮,王敏敏,等.水下二维及三维距离选通成像去噪技术研究[J].红外与激光工程,2020,49(2):27-37.
WANG Xin-wei, SUN Liang, WANG Min-min, et al. Deblurring methods for underwater 2D and 3D range-gated imaging [J]. Infrared and Laser Engineering, 2020, 49(2): 27-37.
- [2] 范承成,德晓薇,郭金家,等.基于三角位移法姿态矫正的激光线扫描海底地形三维测绘[J].光学精密工程,2022,30(10):1170-1180.
FAN Cheng-cheng, DE Xiao-wei, GUO Jin-jia, et al. 3D mapping of submarine topography by laser line scanning based on pose correction by triangular displacement method[J].

- Optics and Precision Engineering, 2022, 30(10): 1170–1180.
- [3] 徐国权, 李广英, 万建伟, 等. 脉冲调制激光雷达水下成像系统[J]. 红外与激光工程, 2022, 51(3): 274–281.
XU Guo-quan, LI Guang-ying, WAN Jian-wei, et al. Underwater imaging system of pulse modulated lidar[J]. Infrared and Laser Engineering, 2022, 51(3): 274–281.
- [4] 谢亮亮, 屠大维, 张旭, 等. 深海原位激光扫描双目立体视觉成像系统[J]. 仪器仪表学报, 2020, 41(6): 106–114.
XIE Liang-liang, TU Da-wei, ZHANG Xu, et al. Deep sea in situ binocular stereo vision imaging system with laser scanning[J]. Chinese Journal of Scientific Instrument, 2020, 41(6): 106–114.
- [5] 梁庆娟. 基于增量式 SFM 的影像密集匹配点云生成及其与激光点云的融合方法研究[D]. 徐州: 中国矿业大学, 2023: 5–10.
LIANG Qing-juan. Research on the generation of image dense matching point cloud based on incremental SFM and their fusion with laser point cloud[D]. Xuzhou: China University of Mining and Technology, 2023: 5–10.
- [6] 肖腾. 基于影像关联图优化的鲁棒 SfM 方法研究[J]. 测绘学报, 2023, 52(6): 1042–1043.
XIAO Teng. Research on robust structure from motion via optimizing view-graph[J]. Acta Geodaetica et Cartographica Sinica, 2023, 52(6): 1042–1043.
- [7] 刘诗源. 基于内窥镜的图像增强及色彩校正方法研究[D]. 保定: 河北大学, 2017: 10–15.
LIU Shi-yuan. Study on enhancement and color correction method of endoscopic images[D]. Baoding: Hebei University, 2017: 10–15.
- [8] CHEN J, WAN Z, ZHANG J, et al. Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet[J]. Computer Methods and Programs in Biomedicine, 2021, 10(16): 200–216.
- [9] ZHUANG Z, CHEN D, LIANG Z, et al. Automatic 3D reconstruction of an anatomically correct upper airway from endoscopic long range OCT images[J]. Biomedical Optics Express, 2023, 14(9): 4594–4608.
- [10] 刘瑶, 贾晓芬. 融合残差结构与注意力机制的暗光图像增强算法[J]. 重庆工商大学学报(自然科学版), 2024, 41(4): 86–96.
LIU Yao, JIA Xiao-fen. Low-light image enhancement algorithm integrating residual structure and attention mechanism[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(4): 86–96.
- [11] LIU D, YU R, CHEN W. Compression artifact reduction of low bit-rate videos via deep neural networks using self-similarity prior[J]. IET Image Processing, 2023, 17(2): 480–491.
- [12] OSORIO F, VALLEJOS R, BARRAZA W, et al. Statistical estimation of the structural similarity index for image quality assessment[J]. Signal, Image and Video Processing, 2022, 16(4): 1035–1042.
- [13] JIA D, ZHANG C, WU N, et al. Autofocus algorithm using optimized Laplace evaluation function and enhanced mountain climbing search algorithm[J]. Multimedia Tools and Applications, 2022, 81(7): 10299–10311.
- [14] HUANG T, LI S, JIA X, et al. Neighbor2Neighbor: A self-supervised framework for deep image denoising[J]. IEEE Transactions on Image Processing, 2022, 31: 4023–4038.
- [15] WANG T, HUANG X, DING J, et al. THz imaging enhancement based on Noise2Noise algorithm[J]. Electronics Letters, 2023, 59(5): 1007–1023.
- [16] 毕海芸, 郑文俊, 曾江源, 等. SFM 摄影测量方法在活动构造定量研究中的应用[J]. 地震地质, 2017, 39(4): 656–674.
BI Hai-yun, ZHENG Wen-jun, ZENG Jiang-yuan, et al. Application of SFM photogrammetry method to the quantitative study of active tectonics[J]. Seismology and Geology, 2017, 39(4): 656–674.
- [17] GUO C, SHI Y, WU H, et al. A combination of library search and Levenberg-Marquardt algorithm in optical scatterometry[J]. Thin Solid Films, 2023, 14(7): 767–780.
- [18] 张耀, 吴一全, 陈慧娟. 基于深度学习的视觉同时定位与建图研究进展[J]. 仪器仪表学报, 2023, 44(7): 214–241.
ZHANG Yao, WU Yi-quan, CHEN Hui-xian. Research progress of visual simultaneous localization and mapping based on deep learning[J]. Chinese Journal of Scientific Instrument, 2023, 44(7): 214–241.
- [19] 曾光. 基于深度图的点云融合算法与系统研究[D]. 武汉: 华中科技大学, 2018: 15–20.
ZENG Guang. Researches on depth map based point cloud fusion methods and system[D]. Wuhan: Huazhong University of Science and Technology, 2018: 15–20.

责任编辑:李翠薇