Journal of Image and Signal Processing 图像与信号处理, 2024, 13(2), 107-116
Published Online April 2024 in Hans. https://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2024.132010

Hans汉斯

基于时空分析的手术场景三维重建方法研究

王晓雨，孟晓亮*，张立晔，宋　政，韩储屹

山东理工大学计算机科学与技术学院，山东　淄博

收稿日期：2024年1月29日；录用日期：2024年2月29日；发布日期：2024年3月12日

摘　要

内窥镜图像的深度估计与手术场景三维重建是微创手术中提高外科医师手术效率的关键因素。本文提出一种基于时空分析的手术场景三维重建方法，深度估计网络采用编码器–解码器结构，编码器使用ResNet34模块、改进的SAB注意力机制、改进的FPN模块以及特征增强模块；解码器通过上采样获取图像的深度信息和位姿信息，实现内窥镜图像的准确深度估计。在跟踪重建方面，通过时空跟踪优化相机位姿，将空间维度的深度信息与时间维度相结合，通过时空分析与融合，还原手术场景的三维结构。评估采用Hamlyn公共数据集，实验结果表明本文所提方法可有效提高内窥镜图像深度估计的准确性，同时通过与时间维度的深度信息融合，可准确还原手术场景的三维信息，进一步辅助外科医师实现术中精准导航。

关键词

内窥镜图像，深度估计，时空分析，位姿优化，三维重建

Research on Three-Dimensional Reconstruction Method of Surgical Scene Based on Spatiotemporal Analysis

Xiaoyu Wang, Xiaoliang Meng*, Liye Zhang, Zheng Song, Chuqi Han

School of Computer Science and Technology, Shandong University of Technology, Zibo Shandong

Received: Jan. 29th, 2024; accepted: Feb. 29th, 2024; published: Mar. 12th, 2024

Abstract

Depth estimation of endoscopic image and 3D reconstruction of surgical scene are key factors to

*通讯作者。

文章引用: 王晓雨, 孟晓亮, 张立晔, 宋政, 韩储屹. 基于时空分析的手术场景三维重建方法研究[J]. 图像与信号处理, 2024, 13(2): 107-116. DOI: 10.12677/jisp.2024.132010

improve surgical efficiency of surgeons in minimally invasive surgery. In this paper, a 3D reconstruction method of surgical scene based on spatiotemporal analysis is proposed, and the proposed network is designed as an encoder-decoder structure. The encoder uses ResNet34 module, improved SAB attention mechanism, improved FPN module and feature enhancement module, and the decoder obtains the depth information and pose information of the image through up-sampling, so as to realize accurate depth estimation of the endoscope image. In terms of tracking and reconstruction, the camera pose is optimized through spatiotemporal tracking, the depth information of the spatial dimension is combined with the time dimension, and the three-dimensional structure of the surgical scene is restored through spatiotemporal analysis and fusion. Hamlyn public dataset was used for evaluation, and experimental results show that the method proposed in this paper could effectively improve the accuracy of depth estimation of endoscopic images. At the same time, three-dimensional information of surgical scene could be accurately restored through fusion with depth information and time dimension to assist surgeons to achieve accurate intraoperative navigation.

Keywords

Endoscopic Image, Depth Estimation, Spatiotemporal Analysis, Pose Optimization, Three-Dimensional Reconstruction

1. 引言

从内窥镜图像序列中还原手术场景的三维信息具有非常重要的研究意义[1]，通过深度学习网络对内窥镜图像进行深度估计，然后对不同图像帧进行跟踪和重建，将空间深度信息与时间维度相结合，最终实现手术场景的三维重建。

图像深度估计与手术场景三维重建是手术场景还原过程中的两个必要步骤。深度估计的准确度直接影响到后续跟踪和三维重建的准确度。针对深度估计问题，Chen 等人[2]提出一种通过学习相对深度以预测像素级深度的多尺度网络。利用相对深度损失函数对网络进行训练，并在无约束环境下实现了单目图像的深度估计。2018 年，Li 等人[3]运用分类回归损失函数，结合扩张卷积(Dilated convolution)，将图像的多尺度深度信息进行融合，进而实现更加准确的深度图估计。Dong 等人[4]提出概率符号距离函数(PSDF)，设计了一种包含体素、网格和网格的混合数据结构，使用贝叶斯更新逐步细化，实时以更多的细节和更少的噪声重建高质量的表面。2019 年，Xie 等人[5]使用胃肠道内窥镜数据集，在位姿优化和空间点定位方面引入局部姿态优化算法和最小几何距离法，但该方法在低帧频时容易出现跟踪丢失的现象。2020 年，孙蕴瀚等人[6]提出一种基于自监督卷积网络的单图像深度估计方法，设计了编解码结构的 CNN，没有全连接层，应用了残差结构、密集连接结构和跳跃连接，通过预测视差的方式学习图像场景内的深度信息，实现了端到端的单幅图像深度估计。2021 年，Sucar 等人[7]跟踪进程中使用 NeRF (Neural Radiance Fields) [8]中的网络架构，优化当前帧相对于固定网络的姿态，重建进程判断关键帧，进行联合优化网络参数和关键帧相机姿势。同年，David Recasens [9]等人提出使用自监督深度神经网络和光度残差跟踪相机位姿[10]，以实现准确和稠密的人体内腔三维重建。并通过实验证明所提方法在图像缩放后性能不会下降，但预测深度值与真实深度值之间的误差还有改进的空间。

本文针对内窥镜手术视频连续且相邻图像帧之间场景重合度较大的特点，提出基于深度学习网络的

内窥镜图像深度估计方法和基于时空分析的手术场景三维重建方法。首先，设计深度学习网络，以实现内窥镜图像深度的准确估计；其次，基于不同图像帧得到的深度图，通过优化相机位姿矩阵，实现手术场景的准确三维重建，从而辅助外科医师进一步提高手术效率。

2. 相关工作

深度学习网络以其强大的特征学习能力在各个领域的应用日趋广泛与成熟。基于单目深度估计[11]的深度学习网络是一种通过深度神经网络从单幅彩色图像中获取深度图的过程。与传统的图像处理方式进行单目深度估计的方法相比，深度学习方法通过构建多层神经网络学习图像特征，从而获取准确度更高的深度图[12]。

深度学习方法可分为有监督学习方法和无监督学习方法两类。自监督学习属于无监督学习方法的一种，可直接从几何约束中学习深度信息[13]。自监督单目深度估计方法通常使用立体图像对或单目图像序列进行训练。本文基于深度学习的单目深度估计采用编码器－解码器网络，RGB 图像作为输入，深度图作为输出。编码器网络由卷积层和池化层组成，用于捕获深度特征；解码器网络通过反卷积层，输出图像尺寸与输入图像尺寸相同。

目标跟踪分为单目标跟踪和多目标跟踪[14]。单目标跟踪是对连续图像帧中单个目标进行跟踪；多目标跟踪是对连续视频画面中多个目标进行跟踪，因此多目标跟踪算法是多变量估计问题，跟踪过程中需要考虑及解决跟踪目标的遮挡、快速移动、旋转等问题[15]。

针对场景的三维重建，早期研究主要侧重于对显式表示的重建研究，如体素[16]、点云[17]或网格[18]。近年来，隐式表示变得更加流行[19] [20]，自从隐式表示在 KinectFusion [21]中首次出现以来，GPU 加速的 TSDF [22] (截断地带符号距离函数)算法被公认为场景重建的标准，利用 GPU 的大规模并行处理单元实现良好的实时性能，得到隐式表达式。TSDF 可以将多个视角的深度图像融合成一个三维模型。对于每个视角，算法会将深度图像转换成点云，再将点云投影到网格上，计算每个网格内的距离值，然后将多个视角的距离值进行融合，得到一个完整的三维模型。通过移动立方体算法构建三角网格模型，恢复场景的立体表面，最终得到手术场景三维重建的结果。

3. 方法

本文基于时空分析的手术场景三维重建过程由深度估计、时空跟踪和三维重建构成，具体结构如图 1 所示。

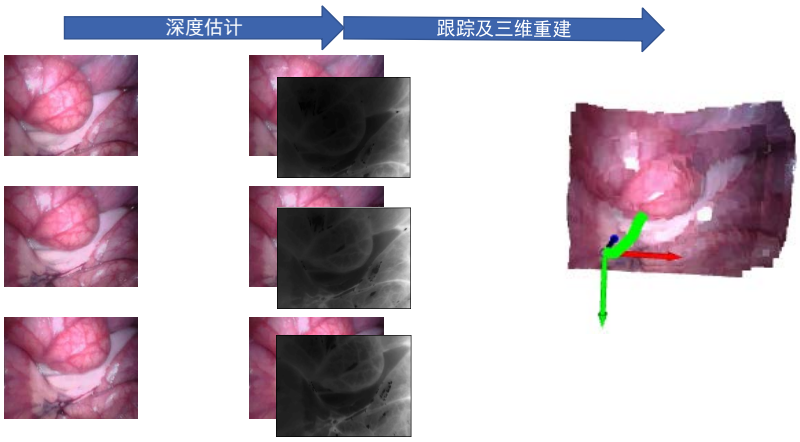

Figure 1. The overall process of three-dimensional reconstruction of surgical scene
图 1. 手术场景三维重建整体流程

首先，深度估计采用 U-Net 网络[23]的编码器-解码器架构，编码网络使用 ResNet34 模块[24]，并通过改进的 SAB 注意力机制、改进的 FPN 模块以及特征增强模块提高深度估计的准确度。解码网络在输出位置使用 Sigmoid 激活函数和 ReLU 非线性激活函数。最终可得到不同图像帧之间的位姿关系以及每一帧图像的深度信息。

其次，针对跟踪问题，本文使用光度跟踪[9]，将反投影光度误差作为依据，使用李代数，将相机的旋转和平移作为参数，使用牛顿高斯法进行迭代，过程中加入由粗到细的金字塔模块以提高收敛速度，以实现相机位姿变换矩阵的优化。

最后，进行三维重建，对每幅关键图像帧中的像素预测深度，通过优化后的相机位姿矩阵还原手术场景的三维信息。计算体素和权重并进行平均以更新 TSDF 值，计算隐式面后使用移动立方体算法[25]构建三角网格，实现手术场景的三维重建。

3.1. 深度估计

本文使用内窥镜图像序列进行图像深度估计，采用 U-Net 网络结构，该网络由编码器和解码器两部分组成。深度估计整体流程图如图 2 所示。首先，输入被测图像依次经过 ResNet34 模块、改进的 SAB 注意力机制、改进的 FPN 模块以及特征增强模块，再通过上采样、卷积操作以及 Sigmoid 激活函数解码，通过多尺度估计，输出不同维度的特征图。其次，被测图像与相邻的下一帧图像一同作为输入进入编码器提取特征信息，使用位姿估计解码器，对特征信息进行卷积操作，使用 ReLU 非线性激活函数进行解码，得到相邻两帧图像之间的位姿信息。

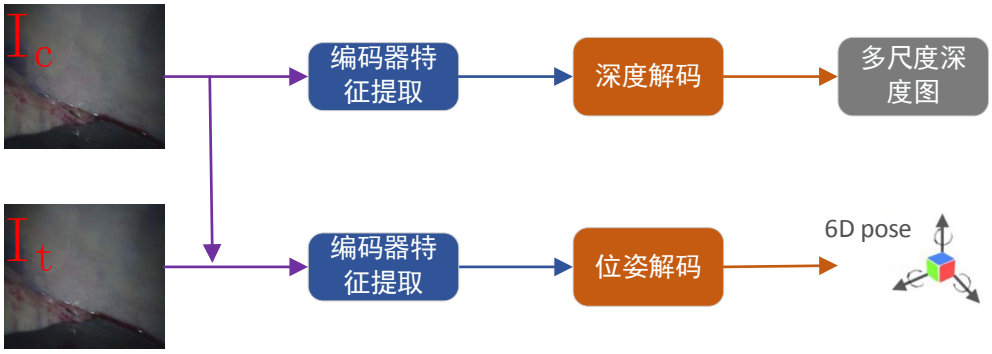

Figure 2. Overall flow chart of depth estimation
图 2. 深度估计整体流程图

深度估计的编码器和解码器的框架图如图 3 所示，在编码器部分，被测图像I_c进入编码器后，首先，通过 ResNet34 模块，并在其中加入改进的 SAB 注意力机制。改进的 SAB 注意力机制如图 4 所示，该注意力机制可使该模块根据当前图像帧的信息，触发对过去相关状态的回忆，以更加专注于图像中重要信息的提取。其次，通过改进的 FPN 模块，将原始 FPN 模块统一固定特征维度(通道数)改进为逐层改变通道数，对传入特征图进行上采样后改变通道数，再与 ResNet34 模块输出中相同通道数的特征图结合，并在融合的结果中加入一个 3 × 3 的卷积块。最后，接入特征增强模块，在特征图上附加一个步长为 2 的 3 × 3 卷积块，并进行下采样及改变通道，然后将其与 FPN 输出中相同通道的特征图相结合，使每个结果都经过一个 3 × 3 的卷积层以生成新的特征。

在解码器部分，将经过编码器后新特征的结果使用 3 × 3 的卷积、ELU()非线性激活函数后，再进行上采样，与特征增强模块中的特征图相融合，最后利用 Sigmoid 激活函数输出多尺度深度图。

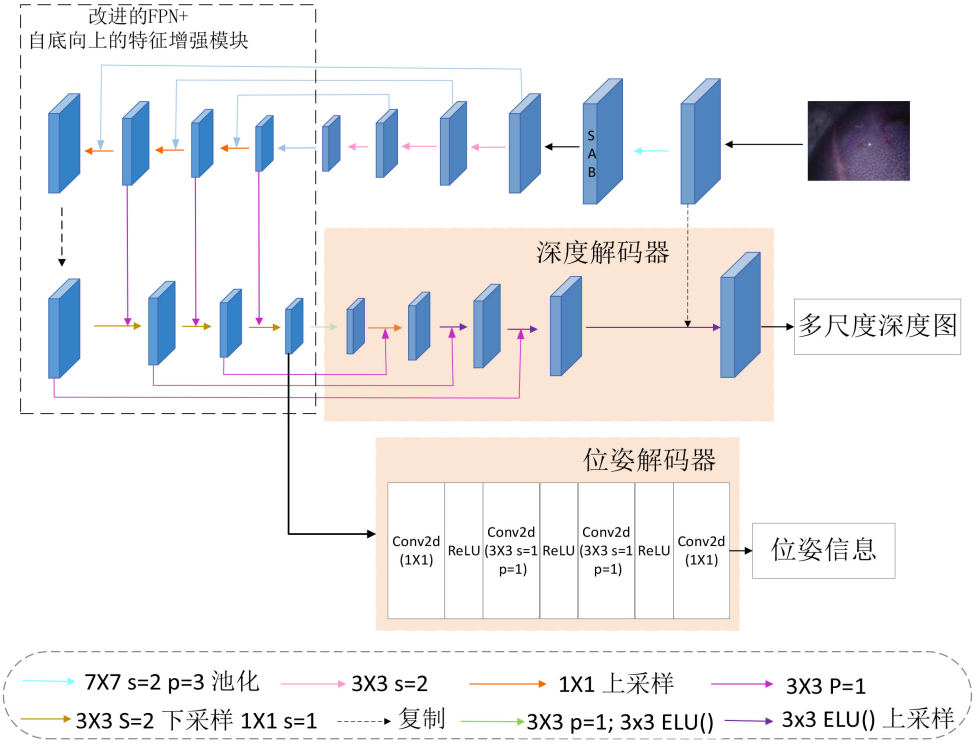

Figure 3. Depth estimation network framework diagram
图 3. 深度估计网络框架图

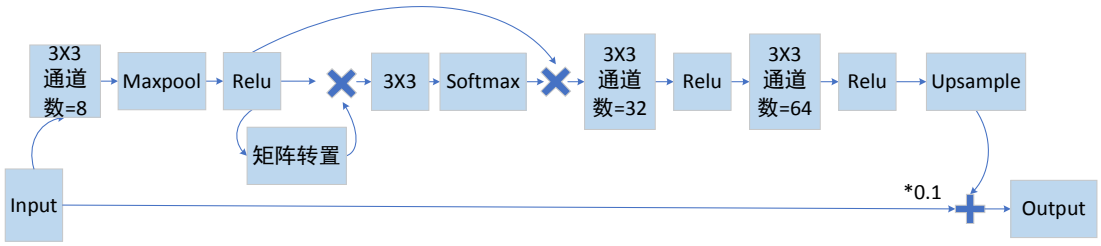

Figure 4. Improved SAB attention mechanism frame diagram
图 4. 改进的 SAB 注意力机制框架图

在位姿信息估计部分，由于位姿估计需要将被测图像 I_c 与下一帧图像 I_t 两帧图像作为输入，使用两帧图像的三通道矩阵进行计算时，将两个矩阵按照通道维度进行拼接。采用两帧图像拼接后的矩阵进入编码器，将尺寸最小的最深层特征图结果作为位姿解码器的输入进行解码，得到位姿信息。

3.2. 时空追踪

用基于关键帧的光度方法进行相机位姿的跟踪，跟踪帧对于被测帧运动的位姿估计通过最小化光度重投影误差进行确定。利用重投影误差的时间几何一致性，将时间维度与空间维度的深度图结合，得到优化后的位姿矩阵。位姿优化流程如图 5 所示。

图 5 中被测图像的下一帧图像，本文称为跟踪帧，根据读取的跟踪帧与被测图像的深度图及转换后的灰度图，计算两者之间的光度重投影误差，并通过计算第 k 次迭代优化后与前一次误差的差值 $l_k[p]$，优化变换矩阵 T_{ct}，得到优化后的位姿矩阵。

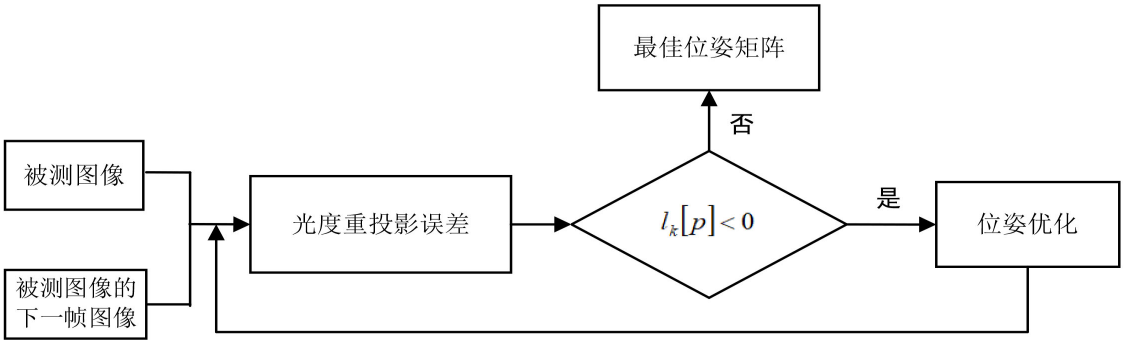

Figure 5. Depth estimation network structure diagram
图 5. 位姿优化流程图

3.2.1. 光度重投影误差

在进行图像跟踪过程中，跟踪帧与被测图像的像素匹配至关重要，本文使用光度重投影误差 l_{ct} 表示两帧图像之间的匹配重合度，同时也作为优化相机变换矩阵的依据，其公式如下：

$$l_{ct}=\sum_{p\in\Omega_t}\left(I_{c\to t}[p]-I_t[p],\alpha\right), \tag{1}$$

其中，$p\in\Omega_t$ 为跟踪帧中对应的像素，$I_{c\to t}$ 为被测图像到跟踪帧的反投影，α 为阈值。

3.2.2. 变换矩阵的优化

本文使用李代数 $F\in se(3)$，将相机的旋转和平移作为参数，通过非线性最小二乘法计算最小化误差以寻找最优函数。并使用牛顿高斯法进行迭代，根据优化前与优化后的误差值，计算得到最优变换矩阵。牛顿高斯法公式如下：

$$\Delta T=\min_{p\in\Omega_t}F\left(T_k[p],l_k[p]\right), \tag{2}$$

$$l_k[p]=\sum_{p\in\Omega_t}\left\{\left(I_{c_{T_k}\to t}[p]-I_t[p],\alpha\right),\left(I_{c_{T_{k-1}}\to t}[p]-I_t[p],\alpha\right)\right\}, \tag{3}$$

其中，T_k 为第 k 次迭代时的矩阵，ΔT 为优化前与优化后的误差值，$I_{c_{T_k}\to t}$ 为使用第 k 次迭代优化的变换矩阵的跟踪帧。

为加速收敛，我们使用由粗到细的金字塔模型，设置金字塔的粗尺度，可以跟踪速度较快的目标，但对于运动缓慢的目标，变化较小，无法准确跟踪到，与实际不相符，因此我们让其尺度逐渐变细，使包含的像素更少，使小运动目标的跟踪更加精确，最终得到优化后的位姿变换矩阵。

3.3. 三维重建

图像三维重建阶段，将深度估计获得的 RGBD 关键帧融合到隐式表面表示中。首先使用基于 TSDF(截断地带符号距离函数)值计算隐式面，将场景划分为体素，并且为每个体素存储一个累积的带符号距离函数，该函数表示到最近表面的距离，若超过阈值的深度值则被截断。可以通过对关键帧中每个像素预测的深度值计算体素和权重并进行平均，并更新 TSDF。最终使用移动立方体算法从隐式表示中构建得到三角网格模型。

4. 实验

4.1. 数据集和训练环境

本文数据集采用 Hamlyn 公共数据集，数据集中内窥镜图像分辨率为 384 × 288，训练环境使用

NVIDIA TESLA T4 16G GPU 服务器以及 Windows 10 操作系统。

4.2. 性能评估指标

4.2.1. 深度估计

为评估本文深度估计方法的准确性，使用绝对相对误差(AbsRel)、相对误差(SqRel)、均方根误差(RMSE)和均方根对数误差(logRMSE)对深度估计进行性能评价。各评价指标的具体公式如下：

$$\mathrm{AbsRel}=\frac{1}{n}\sum_{i}^{n}\frac{\left|d_i-d_i^*\right|}{d_i}, \tag{4}$$

$$\mathrm{SqRel}=\frac{1}{n}\sum_{i}^{n}\frac{\left|d_i-d_i^*\right|^2}{d_i}, \tag{5}$$

$$\mathrm{RMSE}\left(\log(\,)\sqrt{\frac{1}{n}\sum_{i}^{n}\left|\log d_i-\log d_i^*\right|^2}\right), \tag{6}$$

其中，n 为图像总像素数，d_i 为真实的深度值，d_i^* 为预测的深度值。

4.2.2. 跟踪与重建

本文对跟踪与重建结果的评估采用平均重投影误差和三角化点数。平均重投影误差即每个被测图像的三维点，通过相机位姿投影到跟踪帧二维坐标的位置，并与跟踪帧实际点的平均距离误差。平均重投影误差越小，说明整体的准确度越高。重建中三角化的成功率依赖于相机位姿和二维点的匹配精度，因此三角化的点数越多，说明相机的位姿和二维点的匹配度越高。

4.3. 实验结果

4.3.1. 深度估计结果分析

本文深度估计实验使用 Hamlyn 内窥镜图像数据集进行模型训练，然后选取数据集中的场景，并选取其中一部分图像的深度估计结果进行呈现，如图 6 所示。

原图

结果

Figure 6. The depth estimation result
图 6. 深度估计结果

由图 6 可以看出，本文所提深度估计网络可还原出内窥镜图像的深度信息，但是无法从图中直接判断其深度估计的准确性。

为进一步评价本文的深度估计方法，选取部分内窥镜图像进行比较，采用前述深度估计评价指标计算不同方法得到的结果，如表 1 所示。

Table 1. Comparison of performance evaluation indexes of depth estimation
表 1. 深度估计性能评价指标比较

方法	AbsRel	SqRel	logRMSE
Ref. [9]	0.50	13.17	0.60
本文方法	0.38	8.68	0.43

从表 1 可以看出，本文所提深度估计方法的各项性能指标均优于其它的方法，进一步验证了本文方法的有效性。

4.3.2. 跟踪与重建结果分析

将不同时刻的内窥镜深度图及原图进行跟踪与三维重建，可还原整个手术场景的三维立体信息，跟踪及摄像机轨迹如图 7 所示。

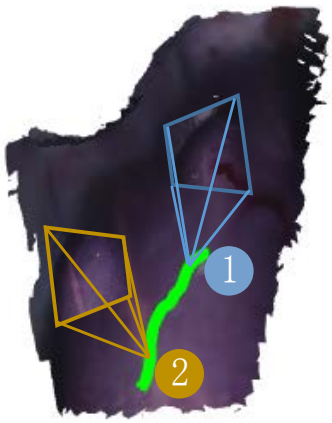

Figure 7. Tracking and camera trajectory
图 7. 跟踪及摄像机轨迹

图 7 中绿色轨迹为摄像机运行的轨迹，得到深度信息更准确、更完整的三维重建结果。将该场景内窥镜图像的跟踪及重建结果与文献[9]方法进行计算，本文方法的平均重投影误差降低约 3%、三角化点数增加约 1600 个像素，这表明整体重建的结果更优且准确度更高，进一步验证了本文所提方法的有效性。

5. 结论

本文提出一种基于时空分析的手术场景三维重建方法。深度估计采用基于改进的 Monodepth2 方法实现了内窥镜图像的准确深度估计。手术场景的跟踪重建，利用深度估计得到的深度图，采用光度误差优化相机位姿，并使用 TSDF 算法及移动立方体算法构建模型完成重建。实验结果表明本文所提方法获取的深度值信息更接近于真实深度值，通过跟踪及三维重建，可得到手术场景的三维形貌，进一步辅助外科医师提高手术效率，实现术中精准导航。

基金项目

国家自然科学基金(No. 62001272)。

参考文献

[1] Isachsen, T.M.E. (2021) Fast and Accurate GPU-Accelerated, High-Resolution 3D Registration for the Robotic 3D Reconstruction of Compliant Food Objects. *Computers and Electronics in Agriculture*, **180**, Article ID: 105929. https://doi.org/10.1016/j.compag.2020.105929

[2] Chen, W., Fu, Z., Yang, D., *et al.* (2016) Single-Image Depth Perception in the Wild. *Advances in Neural Information Processing Systems*, **29**, 730-738.

[3] Li, B., Dai, Y.C. and He, M.Y. (2018) Monocular Depth Estimation with Hierarchical Fusion of Dilated CNNs and Soft-Weighted-Sum Inference. *Pattern Recognition*, **83**, 328-339. https://doi.org/10.1016/j.patcog.2018.05.029

[4] Dong, W., Wang, Q., Wang, X. and Zhang, H.B. (2018) PSDF Fusion: Probabilistic Signed Distance Function for On-The-Fly 3D Data Fusion and Scene Reconstruction. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *ECCV* 2018: *Computer Vision—ECCV* 2018, Springer, Cham, 714-730. https://doi.org/10.1007/978-3-030-01240-3_43

[5] Xie, C., Yao, T., Wang, J., *et al.* (2020) Endoscope Localization and Gastrointestinal Feature Map Construction Based on Monocular Slam Technology. *Journal of Infection and Public Health*, **13**, 1314-1321. https://doi.org/10.1016/j.jiph.2019.06.028

[6] 孙蕴瀚, 史金龙, 孙正兴. 利用自监督卷积网络估计单图像深度信息[J]. 计算机辅助设计与图形学学报, 2020, **32**, 643-651. https://www.jcad.cn/cn/article/doi/10.3724/SP.J.1089.2020.1782

[7] Sucar, E., Liu, S., Ortiz, J., *et al.* (2021) IMAP: Implicit Mapping and Positioning in Real-Time. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 11-17 October 2021, 6229-6238. https://doi.org/10.1109/ICCV48922.2021.00617

[8] Mildenhall, B., Srinivasan, P.P., Tancik, M., *et al.* (2021) Nerf: Representing Scenes as Neural Radiance Fields for View Synthesis. *Communications of the ACM*, **65**, 99-106. https://doi.org/10.1145/3503250

[9] Recasens, D., Lamarca, J., Fácil, J.M., *et al.* (2021) Endo-Depth-And-Motion: Reconstruction and Tracking in Endoscopic Videos Using Depth Networks and Photometric Constraints. *IEEE Robotics and Automation Letters*, **6**, 7225-7232. https://doi.org/10.1109/LRA.2021.3095528

[10] Shen, M., Gu, Y., Liu, N., *et al.* (2019) Context-Aware Depth and Pose Estimation for Bronchoscopic Navigation. *IEEE Robotics and Automation Letters*, **4**, 732-739. https://doi.org/10.1109/LRA.2019.2893419

[11] 陈苑锋. 视觉深度估计与点云建图研究进展[J]. 液晶与显示, 2021, **36**, 896-911. https://doi.org/10.37188/cjlcd.2020-0047

[12] Ming, Y, Meng, X, Fam, C, *et al.* (2021) Deep Learning for Monocular Depth Estimation: A Review. *Neurocomputing*, **438**, 14-33. https://doi.org/10.1016/j.neucom.2020.12.089

[13] Jing, L. and Tian, Y. (2020) Self-Supervised Visual Feature Learning with Deep Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 4037-4058. https://doi.org/10.1109/TPAMI.2020.2992393

[14] Luo, W., Xing, J., Milan, A., *et al.* (2021) Multiple Object Tracking: A Literature Review. *Artificial Intelligence*, **293**, Article ID: 103448. https://doi.org/10.1016/j.artint.2020.103448

[15] Cai, B., Xu, X., Xing, X., *et al.* (2016) BIT: Biologically Inspired Tracker. *IEEE Transactions on Image Processing*, **25**, 1327-1339. https://doi.org/10.1109/TIP.2016.2520358

[16] Xie, H, Yao, H, Zhang, S, *et al.* (2020) Pix2Vox++: Multi-Scale Context-Aware 3D Object Reconstruction from Single and Multiple Images. arXiv: 2006.12250.

[17] Yang, G., Huang, X., Hao, Z., *et al.* (2019) PointFlow: 3D Point Cloud Generation with Continuous Normalizing Flows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 4540-4549. https://doi.org/10.1109/ICCV.2019.00464

[18] Wang, N., Zhang, Y., Li, Z., *et al.* (2018) Pixel2mesh: Generating 3D Mesh Models from Single Rgb Images. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *ECCV* 2018: *Computer Vision—ECCV* 2018, Springer, Cham, 55-71. https://doi.org/10.1007/978-3-030-01252-6_4

[19] Huang, Z., Stojanov, S., Thai, A., *et al.* (2022) Planes vs. Chairs: Category-Guided 3D Shape Learning without Any 3D Cues. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M. and Hassner, T., Eds., *ECCV* 2022: *Computer Vision—ECCV* 2022, Springer, Cham, 727-744. https://doi.org/10.1007/978-3-031-19769-7_42

[20] Xu, Q., Wang, W., Ceylan, D., *et al.* (2019) DISN: Deep Implicit Surface Network for High-Quality Single-View 3D Reconstruction. arXiv: 1905.10711.

[21] Huang, K. and Hao, Q. (2021) Joint Multi-Object Detection and Tracking with Camera-LiDAR Fusion for Autonomous Driving. 2021 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Prague, 27 Septem-

ber-1 October 2021, 6983-6989. https://doi.org/10.1109/IROS51168.2021.9636311

[22] Reijgwart, V., Millane, A., Oleynikova, H., *et al.* (2019) Voxgraph: Globally Consistent, Volumetric Mapping Using Signed Distance Function Submaps. *IEEE Robotics and Automation Letters*, **5**, 227-234. https://doi.org/10.1109/LRA.2019.2953859

[23] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *MICCAI* 2015: *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[24] Zhang, Z., Xu, C., Yang, J., *et al.* (2018) Progressive Hard-Mining Network for Monocular Depth Estimation. *IEEE Transactions on Image Processing*, **27**, 3691-3702. https://doi.org/10.1109/TIP.2018.2821979

[25] Lorensen, W.E. and Cline, H.E. (1987) Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *ACM SIGGRAPH Computer Graphics*, **21**, 163-169. https://doi.org/10.1145/37402.37422