

基于特征点的双目内窥图像三维拼接

摘要: 针对内窥镜常规三维重建算法精度低, 视场小, 尺度信息丢失等缺点, 结合胃器官具体场景, 提出基于特征点的双目内窥图像三维拼接方法。移动内窥镜获取左右图, 通过双目匹配转化成点云, 并进行预处理弥补光照不均、弱纹理等场景特点造成的误差; 通过相邻左视图的特征点检测与匹配估计相机位姿变化; 最后基于计算所得的变换矩阵, 通过迭代最近点(ICP)进行多点云的配准, 实现对整个胃器官的稠密三维重建。拼接后的均方根误差为 2.07, 内窥镜视场扩大2.20倍, 增加了观察范围; 与常规方法对比, 不仅保留了器官尺度信息, 还使稠密度增加了12倍, 方便医生对病灶等目标区域进行三维测量, 有助于提升诊断准确率和效率。

关键词: 内窥镜; 双目匹配; 三维拼接; 配准; 特征点

中图分类号: TP 29 **文献标识码:** A **文章编号:**

3D mosaic of binocular endoscopic images based on feature points

Abstract: There are shortcomings of binocular endoscope 3D reconstruction in the conventional algorithm, such as low accuracy, small field of view and loss of scale information. Combining with the specific scenes of stomach organs, a method of 3D endoscopic image mosaic based on feature points was proposed. The left and right images were obtained by moving endoscope, which were converted into point cloud by binocular matching. Preprocessed to make up for the errors caused by the scene features such as uneven illumination and weak texture. The camera pose changes were estimated by detecting and matching the feature points of adjacent left images. Finally, based on the calculated transformation matrix, point cloud registration was carried out by the iterative closest point (ICP) algorithm, and the three-dimensional dense reconstruction of the whole gastric organ was realized. The results show that the root mean square error was 2.07, and the endoscopic field of view was expanded by 2.20 times, increasing the observation range. Compared with conventional methods, the scale information of organs is not only retained, but also the density is increased by 12 times, which is convenient for doctors to carry out three-dimensional measurement of target areas such as lesions. These improvements are helpful to improve the accuracy and efficiency of diagnosis.

Key words: endoscope; binocular matching; 3D mosaic; registration; feature points

1 引言

内窥镜检查具有分辨率高、创伤小等优点，在临床诊断和治疗上有着广泛的应用。与传统内窥镜相比，双目内窥镜具备三维成像功能，能够提供手术深度信息，帮助医生精准、高效、安全地操作内窥镜，同时方便对人体器官进行三维重建^[1-2]。器官三维重建，特别是实时的稠密三维重建，使画面更加符合真实场景，能够帮助医生对重要的解剖结构及其空间位置进行准确判断，显著提高手术的速度和安全性，减轻患者痛苦，在临床手术中具有重要意义^[3-4]。

现阶段双目内窥镜主要通过双目匹配算法实现三维重建^[5]，其基本原理是模仿人眼的双目立体视觉，通过特征匹配的方式获取物点在左右视图中的视差值，并根据三角测量法结合双目相机参数进行建模。如周嘉悦等^[6]基于双目视觉提取眼底图像的血管进行立体匹配，实现视网膜血管图像的三维重建。然而它们都是基于单视角进行三维重建，视野的局限性会影响医生诊断和手术操作的准确性和安全性。

完整场景的三维重建主要依靠三维拼接技术，常见的方法有多视图几何和点云配准等^[7-8]。当前已有多个研究团队基于多视图几何研发出内窥镜图像三维重建系统，如 Mahmoud 等^[9]使用同时定位与地图构建（simultaneous localization and mapping, SLAM）方法对腹腔单目图像进行动态建图；Widya 等^[10]基于胃镜视频使用运动恢复结构方法（structure from motion, SFM）重建胃器官。该方法一般是稀疏重建，精度低且会丢失尺度信息，不利于后期对重建结果的观察和测量。而且该方法需要通过运动相机观察不同视角，在人体器官这类狭小空间操作难度大。在点云配准方面，随机一致性采样（random sample consensus, RANSAC）是根据提取点云三维特征进行匹配的算法^[11]，胃肠道等器官在正常情况下表面较为光滑，其纹理特征不足以用 RANSAC 进行特征匹配。而作为目前应用最为广泛的点云配准方法——迭代最近点算法（iterative closest point, ICP），通过不断地重复搜索以获取最优的变换矩阵，虽然其容易理解、拼接效果理想，但十分依赖初始矩阵，不容易陷入局部最优解，且需要巨大的计算资源^[12]。

针对上述问题，本文提出了基于特征点的三维拼接方法，通过对内窥镜图像特征点的检测匹配和多点云配准实现了完整的胃模型稠密三维重建，扩大

了医生观测范围，协助手术操作，同时可为后期手术导航打下基础。

2 基于特征点的内窥三维拼接

本文提出的基于特征点的内窥图像三维拼接方法具体流程如 1 所示。操作双目内窥镜获取不同位点下胃模型的左右图序列，通过半全局块匹配算法（semi-global Matching, SGBM）算法进行双目匹配生成各位点下的点云，并进行离群点剔除和降采样等预处理。同时对相邻的左视图进行特征提取与匹配，获取相邻两位点在 X 和 Y 方向上的偏移量，计算对应的初始矩阵。最后通过改进的 ICP 算法对点云进行配准和拼接，从而完成对整个胃器官的三维重建。

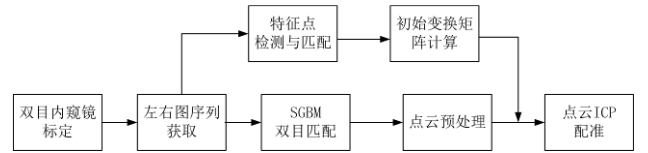


图 1 双目内窥镜图像三维拼接流程图

Fig 1. Workflow of 3D mosaic with binocular endoscopic images

2.1 双目图像获取与匹配

首先对双目相机进行标定，采用张正友棋盘格标定法^[13]计算双目相机内部参数和外部参数。使用黑白棋盘格作为标定板，在双目内窥镜工作距离的范围内拍摄不同位姿的标定板双目图像，利用 OpenCV 库对拍摄的双目图像实现标定。用双目系统的重投影矩阵 Q 来表示双目相机的内外参数信息，其中重投影矩阵 Q 的定义是：

$$Q \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} \quad (1)$$

d 是双目匹配得到的视差， x 、 y 是对应相机的图像坐标系的二维点， X 、 Y 、 Z 和 W 是对应的未归一化的三维坐标和归一化系数 W 。将当前的双目相机等效为了两个相机的主点均为 c_x 、 c_y ，焦距为 f ，且两个相机之间的 X 方向平移向量是 t_x 的对称系统，将对双目深度的求解进行简化。

完成标定后，操作双目内窥镜拍摄胃模型获取左右图序列。移动镜头得到胃模型上下两层共 16

对的左右视图，其中上下各 8 对，图像分辨率为 1280*800，所拍图像已能覆盖胃模型的主要区域。通过双目匹配得到视差图，从而生成相应的点云。在众多的匹配算法中，SGBM 改进了原本半全局匹配（SGM）的代价计算，增加了后处理部分，相比于全局匹配方法（GC）速度更快；相比于局部匹配算法（SAD）精度更高，针对体内组织弱纹理甚至无纹理等场景更为适用，因此选择 SGBM 算法进行双目匹配^[14]。

2.2 点云预处理

由于内窥镜图像存在缺少纹理、细节较弱等困难区域，加上视线遮挡，光照不均匀等原因，使用 SGBM 匹配算法得到的视差图仍会出现误匹配问题，使得产生的点云存在空洞和离群点^[15]。空洞和离群点不仅影响可视化结果，在点云配准过程中也会造成干扰，因此需要对点云进行空洞填充和离群点剔除。

使用半径滤波剔除离群点，该方法假定原始点云中每个点在指定的半径邻域中至少包含一定数量的近邻点。原始点云中符合假设条件的点被视为正常点进行保留，反之，则视为噪声点并进行去除，因此对离群点具有很好的去除效果。

因光照不均匀等原因生成的大片空洞，视差为固定值，转化成点云后集中在特定区域。通过点云分割截取 Z 值在正确深度范围内的点云区域，剔除空洞，得到用于拼接的点云。

实验所使用的内窥镜相机的分辨率是 1280*800，因此由视差图得到的点云数达到百万量级，为了在保持点云形状特征的同时提高配准的速度，对点云进行降采样。采用体素滤波降采样，设置 voxel 参数，即用在设定边长的栅格中所有点的重心表示当前栅格以此减小拼接计算量。

2.3 特征检测与匹配

特征点法是相机运动估计的常用方法。特征点由关键点和描述子两部分组成，关键点指该特征点在图像中的位置，有些还具有朝向、大小等信息；描述子通常是一个向量，描述关键点周围像素的信息。常用的特征检测算法有 SURF、SIFT、ORB 等^[16]。ORB 不具备尺度变换性，只能适用于正对拍摄的场景。SIFT 算法的优点是特征稳定，对旋转、尺度变换、亮度保持不变性，对视角变换、噪声也有一定程度的稳定性；缺点是实时性不高，并且对于边缘光滑目标的特征点提取能力较弱。SURF 是 SIFT 的一种改进，其用一种更为高效的方式完成特

征的提取和描述，提升了计算速度和鲁棒性^[17]。

各位点的点云是根据与左视图对应的视差图生成的，因此在相邻的两帧内窥图像，通过左视图进行特征检测与匹配，从而估计相机的移动。综合比较精度、鲁棒性和计算速度，以及内窥镜的具体场景，本研究采用 SURF 特征法，具体流程如下：

（1）构建黑塞矩阵，生成所有的兴趣点，用于特征的提取。

（2）构建尺度空间。在图像金字塔中生成 0 组 L 层不同尺度的图像，两个量（0，L）构成高斯金字塔的尺度空间，给定一组坐标（0，L）唯一确定高斯金字塔中的一幅图像，以解决图像在所有尺度下描述的问题。

（3）特征点的定位。将经过黑塞矩阵处理的每个像素点与二维图像空间和尺度空间邻域内的 26 个点进行比较，初步定位出关键点，再经过滤除能量比较弱的关键点以及错误定位的关键点，筛选出最终的稳定的特征点。

（4）特征点主方向分配。在特征点的圆形邻域内，统计 60 度扇形内所有点的水平、垂直 haar 小波特征总和，然后扇形以 0.2 弧度大小的间隔进行旋转并再次统计该区域内 haar 小波特征值之后，最后将值最大的那个扇形的方向作为该特征点的主方向。

（5）生成特征点描述子。在沿着特征点主方向的周围取一个 4*4 的矩形区域块，每个区域块统计 25 个像素的水平方向和垂直方向的 haar 小波特征，每个 haar 小波特征有 4 个特征向量，所以最后得到 64 维的特征描述子。

（6）特征点匹配。通过计算两个特征点间的欧式距离来确定匹配度，欧氏距离越短，代表两个特征点的匹配度越好。

2.4 多点云配准

本研究采用多点云配准方法，对点云序列进行配准实现三维拼接。将相邻两位点的前一点云作为源点云，后一点云作为目标点云，使用 ICP 进行配准。ICP 算法的基础思想是近似将距离最近的两点认为是同一点。对于给定的两片点云 P 和 Q，其初始位置关系是 $[R_0|t_0]$ 。选取在初始位姿关系中点云 P 的任意一点 pi 和点云 Q 中的最近点 qi 作为匹配点对，从而建立误差函数，通过使误差函数达到最小值从而求得 R 和 t 的最佳值，该过程即为迭代中的一轮。通过求得的 R 和 t 组合成新的位置关系 $[R_1|t_1]$ ，更新点云的位置关系，继续重复上述过程最后实现误差函数收敛或达到迭代次数的上限

这两个条件之一。对于第 i 对点其误差可表示为:

$$ei = pi - (Rpi' + t) \quad (2)$$

因此为了利用最小二乘法, 误差函数可以表示为:

$$\min_{R, t} J = \frac{1}{2} \sum \| (pi - (Rpi' + t)) \|^2_2 \quad (3)$$

初始变换矩阵会大大影响 ICP 配准的速度和准确性。一般使用对极几何方法, 通过二维图像点的对应关系, 恢复出在两帧之间摄像机的运动, 而在此过程中会对 t 归一化, 直接导致度不确定性, 使得最终的重建结果无法进行三维测量。本研究通过特征匹配的结果生成初始变换矩阵。根据匹配点在相邻两个视场左视图的位置变化, 计算 X 和 Y 方向上偏移量的平均值, 作为三维变换矩阵在平移矩阵部分的初值。由于是相邻两位点, 点云在 X 和 Y 方向上的变换大小远大于 Z 方向和旋转矩阵变化量。

对于每一对点云, 设置大小两个阈值, 使其进行粗、精两次配准, 保证点云有较好配准结果的前提下提高计算效率。变换矩阵中的旋转部分是非线性的, 导致 ICP 求解实质上是一个非线性最小二乘问题。考虑到在短时间内的运动非常小, 旋转角度约为 0, 将 ICP 求解近似地转换为一个线性最小二乘问题。

得到姿态图, 即点云节点和包含点云配准的变换矩阵的边后, 通过 Bundle Adjustment 算法^[18]进行图优化, 以减小由于配准过程中的位姿估计误差累积。最终完成内窥图像的三维拼接, 实现对整个胃模型的三维重建。

3 实验结果与分析

3.1 系统搭建与标定结果

本研究搭建了双目内窥镜三维拼接实验系统, 系统结构如图 2a 所示。样机系统主要包括: 双目胃镜镜体, LED 光源, 信号获取电路, 图像处理工作站, 其中镜体探头(图 2c)包括双目相机系统, 送水送气通道, 手术器械通道, 以及照明系统。相机参数为: 焦距为 1059.6 像素, 基线距离为 5.9mm, CMOS 尺寸为 $1.75 \mu m \times 1.75 \mu m$, 左相机主点坐标 c_x 和 c_y 分别为 633.6 像素和 367.1 像素。

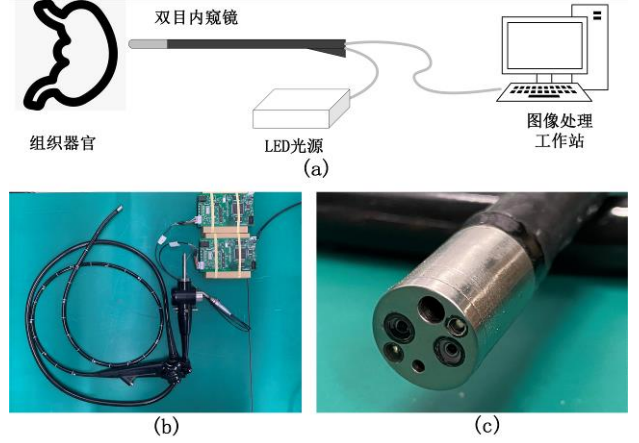


图 2 双目内窥镜三维拼接系统 (a) 系统示意图; (b) 双目内窥镜实物图; (c) 双目内窥镜镜体探头

Fig 2. 3D mosaic system for binocular endoscope (a) System diagram; (b) Binocular endoscope; (c) The tip of binocular endoscope

在相机标定时, 使用棋盘格为 9×6 、方格边长为 12mm 的黑白棋盘格作为标定板, 在双目内窥镜工作距离的范围内拍摄了 15 组不同位姿的标定板双目图像。最终, 得到的重投影矩阵 Q

$$Q = \begin{bmatrix} 1 & 0 & 0 & -633.6 \\ 0 & 1 & 0 & -367.1 \\ 0 & 0 & 0 & 1059.6 \\ 0 & 0 & 0.17 & 0 \end{bmatrix}$$

3.2 点云预处理结果

在半径滤波离群点剔除处理中, 根据当前图像大小, 将半径 15 以内的球形中点数小于 16 的点定义为离群点, 进行剔除。如图 3 显示了点云经过离群点剔除预处理后的结果, 其中 3a 为原始点云, 3b 是经过半径离群点处理的点云, 3c 是在 3b 基础上进行点云分割, 以去除深度值在一定范围内的空洞部分。由图 3 可知, 经过离群点剔除预处理之后, 大量的空洞和误匹配点被剔除, 处理后的点云能反映内窥图像的三维结构。

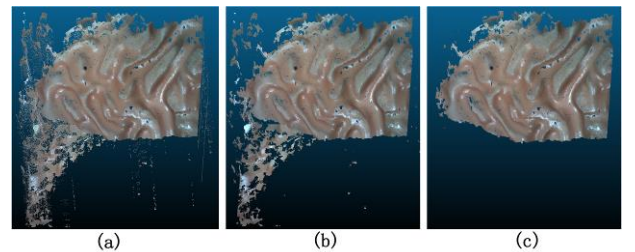


图 3 点云离群点剔除效果 (a) 原始点云; (b) 半径离群点剔除; (c) 点云分割结果

Fig 3. Result of point cloud outlier removal (a) Original point cloud; (b) Radius outlier removal; (c) Result of point cloud segmentation

在体素化降采样中, 设定 voxel 参数为 5, 平均一个点云的大小从百万减小到 10 万。表 1 比较了预处理前后相邻两个点云配准的速度。可知经去空洞、降采样等预处理操作之后, 特别是降采样之后, 显著降低了点云配准所用的时间, 提高了拼接的实时性。图 4 展示了点云预处理前后相邻点云配准结果, a、b、c 分别对应原始点云、去离群点及去离群点的降采样后点云的配准结果。红色区域存在明显的错误匹配, 剔除误配点和空洞, 配准的准确率得到提升。

表 1 点云预处理前后配准时间比较

Tab.1 Comparison of registration time before and after point cloud preprocessing

	配准时间 (s)	读取时间 (s)
原始点云	72.33	1.02
去离群点	63.61	0.84
去离群点&降采样	1.03	0.17

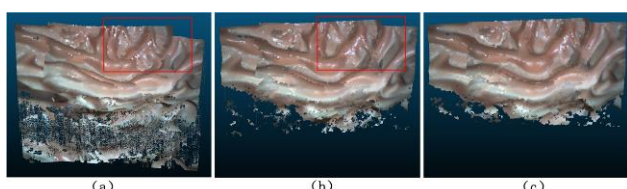


图 4 点云预处理前后配准准确率比较 (a) 原始点云; (b) 去离群点; (c) 去离群点&降采样

Fig 4. Comparison of registration time before and after point cloud preprocessing (a) Original point cloud; (b) Outlier removal; (c) Outlier removal & downsampling

3.3 特征检测与匹配结果

图 5 是相邻两位点左视图特征匹配的结果, 其中红点为检测到的特征点, 蓝线表示匹配到的特征点对。在图中, 蓝线多为等长的平行线, 说明在相邻两帧图像上主要表现为 X 和 Y 方向上的移动。

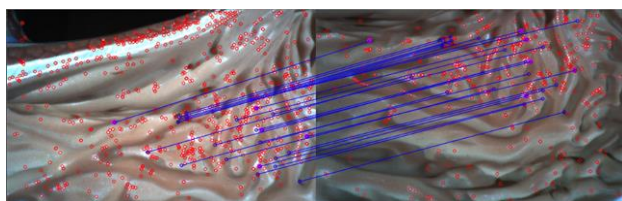


图 5 左视图特征检测与匹配结果

Fig 5. Result of left images feature detection and matching

3.4 三维拼接结果

经过对点云序列进行配准, 完整胃模型的双目内窥图像三维重建结果如图 6 所示。图 6a 为胃模型

的平面图, 图 6b 为胃模型点云三维拼接的结果。在图 b 中, 三维拼接的结果已经能展示胃模型主要区域的特征, 且模型上各种纹理如胃壁褶皱能清晰的显现, 有利于医生的诊断。图 6c 是视场变化的结果, 其中红色框为单幅图像的视场大小, 为 29.05° , 经过 16 幅图的拼接, 视场是原来的 2.20 倍, 视场大小为 63.91° , 大大增加了观测范围, 方便医生手术操作。

为了评价点云拼接的精度, 将结果与胃模型的扫描点云进行配准。图 6d 白色部分是模型经扫描仪扫描得到的点云, 可看作真值。结果显示二者很好地进行匹配, 均方根误差 (RMS) 为 2.07, 6e 是两者配准的可视化结果, 颜色表示拼接点与扫描点距离的绝对值。大部分为蓝色区域, 说明双目点云在拼接后较好还原了模型的三维特征。

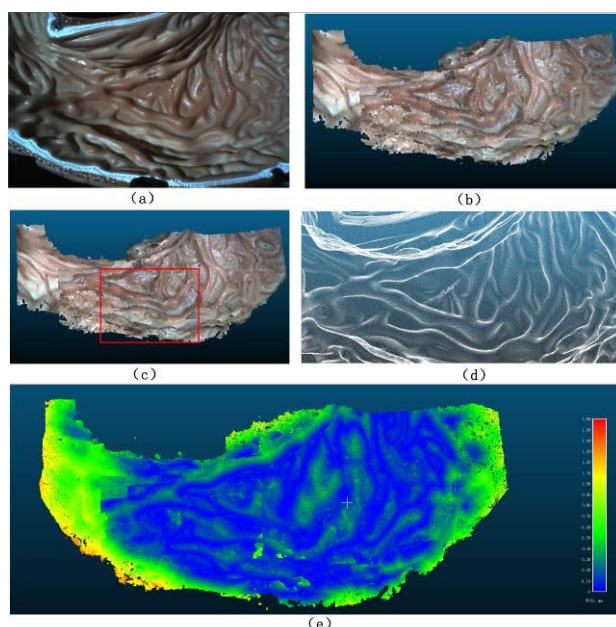


图 6 胃双目内窥三维拼接结果及评价 (a) 胃模型平面图; (b) 本研究三维拼接结果; (c) 视场变化情况; (d) 扫描点云; (e) 与扫描点云配准结果;

Fig 6. Results and evaluation of 3D mosaic with gastric binocular endoscopic (a) Plane figure of the model; (b) Result of 3D mosaic; (c) Change of the field of view; (d) Scanned point cloud; (e) Registration result with scanned point cloud

为进一步评价本研究方法, 使用 SFM 方法对相同的内窥图像序列进行三维重建, 得到的结果比较如图 7 所示。SFM 结果有 27568 个点, 而本方法结果点云有 344201 个点, 提升了约 12 倍; 对器官模型的某一区域进行放大, 本方法仍能反映其三维特征, 且保留了尺度信息, 允许进行三维测量。SFM 方法

不仅失去尺度信息，还因太稀疏而使器官特征难以辨认。结果表明，本方法精度高，方便医生在使用内窥镜时对疑难部位进行观察和测量。

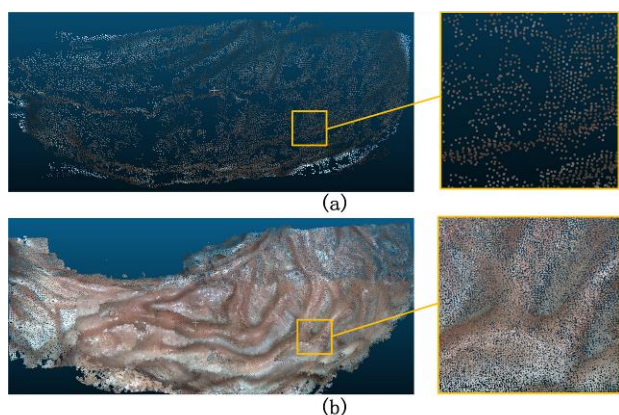


图 7 与其他方法的比较 (a) 本研究方法结果; (b) SFM 方法重建结果

Fig 7. Comparison with other methods (a) Result of our method; (b) Result of SFM

4 结语

本研究提出了一种基于特征点的双目内窥镜图像三维拼接方法，通过 SGBM 双目匹配获取不同位点的点云，通过相邻位点特征检测与匹配计算变换矩阵，为多点云 ICP 配准减小计算量，提升准确率。三维拼接后的内窥镜图像，不仅扩大了视场，且保留了尺度信息，有利于反映患病部位真实场景、方便医生进行手术操作，也为后续的手术导航提供支持。

参考文献:

- [1] SORENSEN S M, SAVRAN M M, KONGE L, et al. Three-dimensional versus two-dimensional vision in laparoscopy: a systematic review [J]. **Surgical Endoscopy**, 2016, 30(1): 11-23.
- [2] AREZZO A, VETTORETTO N, FRANCIS N K, et al. The use of 3D laparoscopic imaging systems in surgery: EAES consensus development conference 2018 [J]. **Surgical Endoscopy**, 2019, 33: 3251-3274.
- [3] LIAO H, TSUZUKI, MASAYOSHI, et al. Fast image mapping of endoscopic image mosaics with three-dimensional ultrasound image for intrauterine fetal surgery [J]. **Minimally Invasive Therapy & Allied Technologies**, 2009, 18(6): 332-340.
- [4] CAI H, WANG R, LI Y, et al. Role of 3D reconstruction in the evaluation of patients with lower segment oesophageal cancer [J]. **Journal of thoracic disease**, 2018, 10(7): 3940-3947.
- [5] SUI C, WU J, WANG Z, et al. A Real-Time 3D Laparoscopic Imaging System: Design, Method, and Validation [J]. **IEEE Transactions on Biomedical Engineering**, 2020, PP(99): 1-.
- [6] 周嘉悦, 韩少峰, 郑昱, 等. 基于双目视觉的视网膜血管三维重建[J]. **中国医疗器械杂志**, 2020, 44(1):13-19.
ZHOU J, HAN S, ZHENG Y, et al. Three-dimensional Reconstruction of Retinal Vessels Based on Binocular Vision [J]. **Chinese Journal of Medical Instrumentation**, 2020, 44(1): 13-19.
- [7] CRANDALL D, OWENS A, SNAVELY N, et al. SfM with MRFs: discrete-continuous optimization for large-scale structure from motion [J]. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. 2013, 35(12):2841-53
- [8] POMERLEAU F O, COLAS F, SIEGWART R. A Review of Point Cloud Registration Algorithms for Mobile Robotics [J]. **Foundations & Trends in Robotics**, 2015, 4(1): 1-104.
- [9] MAHMMOUD N, COLLINS T, HOSTETTLE A, et al. Live Tracking and Dense Reconstruction for Handheld Monocular Endoscopy [J]. **IEEE Transactions on Medical Imaging**, 2019, 38(1): 79-89.
- [10] WIDYA A R, MONNO Y, OKUTOMI M, et al. Whole Stomach 3D Reconstruction and Frame Localization From Monocular Endoscope Video [J]. **IEEE Journal of Translational Engineering in Health and Medicine**, 2019, 7: 3300310.
- [11] TSAI C Y, WANG C W, WANG W Y. Design and implementation of a RANSAC RGB-D mapping algorithm for multi-view point cloud registration [C]// **CACS International Automatic Control Conference**. Taiwan: CACS, 2013, pp. 367-370.
- [12] SHARP G C, LEE S W, WEHE DK, ICP Registration

-
- Using Invariant Features [J]. **IEEE Transactions on Pattern Analysis & Machine Intelligence**, 2002, 24(1): 90-102.
- [13] ZHANG Z. A Flexible New Technique for Camera Calibration [J]. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2000, 22(11): 1330-4.
- [14] 张欢, 安利, 张强, 等. SGBM 算法与 BM 算法分析研究 [J]. **测绘与空间地理信息**, 2016, 000(010):214-216.
- HUAN Z, LI A N, QIANG Z, et al. SGBM Algorithm and BM Algorithm Analysis and Research [J]. **Geomatics & Spatial Information Technology**, 2016, 000(010):214-216.
- [15] WANG C, ODA M, HAYASHI Y, et al. A visual SLAM-based bronchoscope tracking scheme for bronchoscopic navigation [J]. **International Journal of Computer Assisted Radiology and Surgery**, 2020,15: 1619–1630.
- [16] KARAMI E, PRASAD S, SHEHATA M, Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images [C]// **Electrical and Computer Engineering Conference**. Canada: IEEE, 2015.
- [17] BAY H, TUYTELAARS T, GOOL L V, SURF: Speeded up robust features[C]// **European conference on Computer Vision**. Austria, 2006, pp 404-417.
- [18] GRANSHAW S I. Bundle Adjustment Methods in Engineering Photogrammetry [J]. **The Photogrammetric Record**, 2006, 10(56):181-207.