

摘要

特征点是多视图三维恢复的基础,可以通过图像的特征点提取与匹配来实现不同视角下的图像对应和几何信息获取。因此,通过提取两幅图像中具有显著特征的关键点或兴趣点来进行配对,能够在两幅图像之间建立对应关系,其提取和匹配的精度和速度直接影响到后续任务的准确性和实时性。然而由于低纹理和光照等特殊环境的影响,提取的特征点数量较少且会产生大量错误匹配。当使用 RANSAC (Random Sample Consensus) 剔除错误匹配的过程中需要随机选择所有样本去迭代求解,极大影响了匹配速度,并且大量错误匹配存在于初始匹配中,会不利于多级匹配的匹配精度。为此,该论文对特征提取和匹配方法进行了更进一步的研究,取得的主要成果可概括为:

第一,针对使用固定阈值的 ORB (Oriented FAST and Rotated BRIEF)算法在胶囊内窥镜图像变化不显著的区域难以提取特征点,显著区域出现特征点冗余的问题,提出了一种改进的 ORB 算法来提取胶囊内镜图像的特征点。该方法首先在特征点提取阶段使用变异系数计算自适应阈值,提高算法在同质区域特征点的提取能力。然后,使用四叉树方法对特征点进行过滤,以消除过度集中和重叠的特征点。最后,使用 BEBLID (Boosted Efficient Binary Local Image Descriptor)二进制描述子来构建具有强描述的二进制描述符,并使用 RANSAC 来避免不匹配。

第二,针对由于需要随机选择所有的样本去进行迭代求解 RANSAC 的最优模型从而导致剔除错误匹配耗时较多的问题,提出了一种便于视觉 SLAM (Simultaneous Localization and Mapping)增速的 GMS (Grid-based Motion Statistics)与 RANSAC 融合错误匹配剔除方法。该方法首先基于 GMS 算法获得匹配对邻域内匹配对数的数量,并以此为可信度对匹配对进行排序分组,进而采用 RANSAC 算法进行进一步剔除误匹配时,将其随机选取所有匹配对所带来耗费大量时间的问题,转换为优先从可信度高的匹配对中选取样本的问题,进而实施最优模型的迭代求解。

第三,针对多级匹配方法中使用基于重采样的方法来剔除错误匹配时,基于重采样的方法受初始匹配准确性低的影响,造成匹配精度降低的问题。建立了一种基于多级精细匹配策略的新特征匹配方法即称为 KTGP-ORB 方法。该方法利用特征描述符在汉明空间中局部外观的相似性生成初始对应关系,并结合局部图像运动平滑性约束,利用 GMS 算法提高初始匹配的准确性;最后采用 PROSAC (Progressive Sampling Consensus)算法优化匹配,得到基于欧式空间全局灰度信息上的精准匹配。

本论文在不同场景下针对特征点提取与匹配的不同问题进行研究,改善了复杂场景中特征点提取困难、匹配速度慢以及匹配精度低的问题,支撑了 SLAM 和 SFM (Structure from Motion)技术在更广泛领域的应用,丰富了多视图几何学的方法和技术体系。

感谢国家自然科学基金(No. 62172190)、江苏省"双创人才"计划(证书号:JSSCRC2021532)和无锡市"太湖人才-创新领军人才"计划(证书日期:202110)的支持!

关键词: 特征提取; 特征匹配; 随机采样一致性; ORB 算法; GMS 算法

Abstract

Feature points are the basis of multi-view 3D recovery, and can be extracted and matched to achieve image correspondence and geometric information from different views. Therefore, by extracting key points or points of interest with significant features in two images for matching, a correspondence can be established between the two images, and the accuracy and speed of their extraction and matching directly affect the accuracy and real-time performance of subsequent tasks. However, the number of extracted feature points is low and a large number of false matches can be generated due to specific environments such as low texture and illumination. When using RANSAC (Random Sample Consensus), all samples need to be randomly selected for iterative solving, which greatly affects the matching speed, and the presence of a large number of false matches in the initial matching can be detrimental to the matching accuracy of multi-level matching. To this end, this thesis has further investigated the feature extraction and matching methods, and the main results achieved can be summarized as follows:

Firstly, an improved ORB (Oriented FAST and Rotated BRIEF) algorithm is proposed to extract feature points from capsule endoscopy images for the problem that it is difficult to extract feature points in regions with insignificant changes in capsule endoscopy images using a fixed threshold, and that redundancy of feature points occurs in significant regions. The method first uses variance coefficients to calculate adaptive thresholds in the feature point extraction stage to improve the algorithm's ability to extract feature points in homogeneous regions. Then, the feature points are filtered using a quadtree approach to eliminate over-concentrated and overlapping feature points. Finally, BEBLID (Boosted Efficient Binary Local Image Descriptor) binary descriptors are used to construct binary descriptors with strong descriptions, and RANSAC is used to avoid mismatches.

Secondly, to address the problem of the need to randomly select all the samples to iteratively solve the optimal model of RANSAC, which leads to a time-consuming method of rejecting false matches, a fusion of GMS (Grid-based Motion Statistics) and RANSAC is proposed to facilitate the speedup of visual SLAM (Simultaneous Localization and Mapping). The method first obtains the number of matching pairs in the neighborhood of a matching pair based on the GMS algorithm, and uses this as a confidence level to rank and group the matching pairs, and then uses the RANSAC algorithm to further reject false matches, converting the time-consuming problem of randomly selecting all matching pairs into a problem of prioritizing the selection of samples from pairs with high confidence levels, and then implementing the iterative solution of the optimal model.

Thirdly, to address the problem that when resampling-based methods are used in multi-level matching methods to reject false matches, the resampling-based methods are affected by the low initial matching accuracy, resulting in reduced matching accuracy. A new feature matching method based on a multi-stage fine matching strategy is established, known as the KTGP-ORB method. The method uses the similarity of local appearance of feature descriptors in Hamming space to generate the initial correspondence, and combines the local image motion smoothness

constraint with the GMS algorithm to improve the initial matching accuracy; finally, the PROSAC (Progressive Sampling Consensus) algorithm is used to optimize the matching and obtain an accurate matching based on the global grey-scale information in Euclidean space.

This thesis investigates the different problems of feature extraction and matching in different scenes, improving the problems of difficult feature extraction, slow matching and low matching accuracy in complex scenes, supporting the application of SLAM and SFM (Structure from Motion) techniques in a wider range of fields, and enriching the methodology and technical system of multi-view geometry.

Thanks to the National Natural Science Foundation of China (No. 62172190), Jiangsu Province "Innovative and Entrepreneurial Talent" Program (Certificate No. JSSCRC2021532) and Wuxi "Taihu Talents - Innovation Leaders" Program (Certificate Date: 202110) for their support!

Keywords: feature extraction; feature matching; random sample consensus; ORB algorithm; GMS algorithm

目 录

第一章 绪论.....	1
1.1 课题研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 特征点提取.....	2
1.2.2 误匹配剔除.....	3
1.2.3 多级匹配.....	4
1.3 论文主要研究内容.....	5
1.4 论文组织结构.....	6
第二章 用于胶囊内窥镜定位的特征点提取	7
2.1 引言.....	7
2.2 方法概述.....	8
2.3 方法细述.....	8
2.3.1 自适应阈值.....	8
2.3.2 特征点均匀化.....	9
2.3.3 结合 BEBLID 算子的特征点描述	10
2.4 实验与分析.....	11
2.4.1 数据集选择.....	11
2.4.2 评价指标选择.....	11
2.4.3 实验结果.....	11
2.4.4 方法对比分析.....	13
2.5 本章小结.....	15
第三章 面向视觉 SLAM 提速的错误匹配剔除	17
3.1 引言.....	17
3.2 方法概述.....	17
3.3 方法细述.....	18
3.3.1 GMS-RANSAC 方法分析	18
3.3.2 方法改进.....	20
3.4 实验与分析.....	20
3.4.1 数据集选择.....	20
3.4.2 评价指标选择.....	20
3.4.3 实验结果.....	21
3.4.4 参数变化实验.....	22
3.4.5 方法对比分析.....	24
3.5 本章小结.....	27

第四章 基于多级细化策略的特征匹配	29
4.1 引言	29
4.2 方法概述	30
4.3 方法细述	30
4.3.1 初始关系构建	30
4.3.2 假匹配剔除	31
4.3.3 匹配优化	32
4.4 实验与分析	33
4.4.1 数据集选择	33
4.4.2 评价指标选择	34
4.4.3 实验结果	34
4.4.4 参数变化实验	36
4.4.5 方法对比分析	39
4.5 本章小结	42
第五章 总结与展望	43
5.1 主要结论	43
5.2 展望	43
参考文献	45

第一章 绪论

本章对阐述了该论文的研究背景和意义，并对特征点提取和匹配的研究现状进行了概述，总结出了目前存在的问题。进而给出了本文研究的内容、解决的关键技术，并设计了本文研究的整体架构、以及各章节的具体组成。

1.1 课题研究背景与意义

工具的使用是人类文明发展的重要组成部分，人类文明的发展离不开对工具的使用，从石器时代到青铜器时代再到铁器时代，人类的工具不断进化，这些工具的进步推动了农业和手工业的发展，极大促进了人类文明的进步。再之后的工业革命更是推动了人类社会文明的进步和生产力的提高，从蒸汽时代到电气时代再到信息时代，人类已经发展出了各种各样的机器和工具，这些机器和工具已经成为了人类生产和生活中不可或缺的一部分。近年来，随着科技水平的不断提高，又一次工业革命正在兴起，步入全球自动化的时代已经成为一种新的趋势。在这一背景下，人工智能技术迅猛发展，其主要目的是让机器模拟人类的智能，实现智能化的自动化过程。在人工智能的热潮中，多视图三维恢复技术在计算机视觉领域中获得了大量研究，在航空航天、国防军工、医疗影像、农业生产、安防监控等诸多领域已有广泛应用。其利用计算机算法和技术来推断物体的三维结构，对于实现计算机视觉中的各种任务有着至关重要的作用，推动着计算机视觉技术的迅猛发展。

在多视图三维恢复中，需要首先获取物体的多个视角的二维图像。这些图像可以通过不同的相机拍摄角度来获取，也可以是通过运动相机等设备采集。然后，通过图像处理技术，如图像匹配、立体匹配等算法，将这些二维图像进行对应和匹配，以获取不同视角下物体的几何信息。最后，通过三维重建算法将这些几何信息转化为物体的三维模型。因此，图像匹配是多视图三维恢复中非常重要的一个环节。图像匹配是通过选用合适的方法去利用图像本身各种信息，在两幅或多幅图像之间建立联系的过程。该技术在三维重建^[1]、目标识别与跟踪^[2]、图像检索^[3]、即时定位与地图构建(Simultaneous Localization and Mapping, SLAM)^[4]等诸多领域中都具有着重要的作用。图像特征提取与匹配作为一项基础而关键的技术一直是研究热点和研究难点。其中，基于局部特征点的图像匹配是目前图像匹配中的主流方式之一，这类方法通过提取图像中的角点、交叉点、闭合区域中心等具有一定物理结构的点来实现图像间的匹配。但是在现实场景的图像中，由于光照变化、遮挡、图像旋转、噪声、模糊等复杂干扰因素的存在，图像的质量和特征点的提取会收到极大的影响，这些因素对特征点匹配的准确性和可靠性有着很大程度的影响^[5]。这些影响会产生大量匹配误差，并累积到匹配之后的处理环节从而严重影响最终的视觉任务的效果，例如，在基于特征点的视觉 SLAM 任务中，大量的错误匹配会导致较大的错误相机位姿估计，使得最终的定位和地图构建结果与事实严重不符^[4]。此外，特征匹配的耗时也是需要考虑的因素，否则难以满足像 SLAM 这样具有实时性要求的视觉任务。为了克服上述干扰因素的影响以获取更好的匹配效果，研究人员一直致力于研究更加鲁棒性的特征与更加高效的匹配方法。

尽管有越来越多的研究者提出更加鲁棒的特征点提取和匹配方法，不断提升了匹配的效果，但是由于图像匹配所面临的场景和问题是多样化的，有些问题需要在具体的应用场景中进行综合考虑并采用相应的处理方法来解决。比如在胶囊内窥镜这种特殊场景下，由于胶囊内窥镜设备的限制，内置的摄像头和日常生活中使用的摄像头有很大差异，其拍摄的图像分辨率相对较低，帧率也相对较低。此外，由于消化道内部环境的复杂性，胶囊内窥镜的拍摄视角受到限制，只能观察到胶囊进入的部位，无法灵活调整视角，同时还会受到液体、气体等物质的干扰。并且胶囊内窥镜只能靠自身配置的光源进行照明，在消化道内移动时其所拍摄的照片的光照变化比较剧烈，对图像质量影响十分巨大。这些缺点对于现有的特征提取和匹配方法形成了很大的挑战。

由于图像匹配所面临的干扰因素较多且多样化，因此没有一种特征匹配算法能够在所有情况下都具有足够的鲁棒性和实时性^[6]。综上所述，特征点作为多视图三维恢复中进行图像匹配的关键，在提取和匹配过程中仍然存在多方面的挑战和难题，需要进一步深入的研究，以便能够满足不同视觉任务的不同需求。

1.2 国内外研究现状

图像匹配作为多视图三维恢复中的关键技术早在上世纪 70 年代就被提出来了，其核心思想是通过两幅图像中具有相同或相似的内容或结构构建两幅图像之间的联系^[7-8]。其中，基于特征点的匹配方法是图像匹配中的一个最为基本的方法，其是通过提取两幅图像中具有物理意义的显著关键点或兴趣点来进行图像之间的匹配^[9]。

1.2.1 特征点提取

在众多的特征点提取的经典方法中，Moravec^[10]在 1981 年首次提出兴趣点的概念，将图像中不同于其它像素点的点，称为兴趣点，也叫做特征点。但只考虑了邻域的灰度差异，判断方法简单，因此鲁棒性较差。1988 年，Harris 在 Moravec 算法的基础上提出了 Harris 算子^[11]保证了方向的不变性，降低了检测时间，但对尺度变化比较敏感，抗噪能力差。尺度不变特征转换(Scale Invariant Feature Transform, SIFT)算法^[12]是 Lowe 在 1999 年提出的一种基于尺度空间的特征点检测方法，并在 2004 年进行了完善总结。该算法充分考虑了变换过程中图像的光照、比例和旋转的变化，但是计算复杂，无法在普通计算机的 CPU 上实时计算 SIFT 特征并实现定位。在 SIFT 算法的基础上，Bay 于 2006 年提出了加速鲁棒特征(Speeded Up Robust Features, SURF)算法^[13]，该算法采用了 Hessian 矩阵和降维特征描述符，在一定程度上提高了 SIFT 算法的执行效率，为算法在计算机上实时运行提供了可能。在处理复杂场景的图像时，这些算法依旧耗时很长无法满足实时性的需求。此外，还存在其它特征，通过适当降低精度和鲁棒性，以提高计算速度，特别是 Rosten 等人在 2006 年至 2010 年期间提出的加速分割检测(Features from Accelerated Segment Test, FAST)算法^[14]，该算法以其快速著称，它利用某一场域中每个像素的灰度值和中心点的灰度值来判断中心点是否为一个角点，由于该算法只检测角点，因此不具有尺度不变性和方向性；但 FSAT 算法只能用于特征点的检测，不能详细描述特征点。2010 年提出的二进制鲁棒独立元素特征(Binary Robust Independent Elementary

Features, BRIEF)算法^[15]通过在特征点的图像窗口中随机选择像素点来比较灰度值,生成二进制编码的描述子来对检测到的特征点进行描述。但在比较像素灰度值的过程中,只使用像素点灰度值之间的关系。未能充分利用像素点周边的灰度值信息,会造成图像信息的丢失。Rublee 等人在 2011 年结合 FSAT 算法和 BRIEF 提出了一种具有旋转不变性的(Oriented FAST and Rotated BRIEF, ORB)算法^[16]。ORB 在精度和速度之间进行很好的折衷,所提取的特征对于大多数场景表现良好,但没有解决尺度变化的问题^[17]。不过在同年提出的二进制尺度不变特征点(Binary Robust Invariant Scalable Keypoints, BRISK)算法^[18]中解决了 ORB 中存在的尺度问题。Alahi 等人在 2012 年在 BRIEF 算法的基础上,在选择像素点时仿照了人眼视网膜的采样模式,提出了快速视网膜关键点(Fast Retina Keypoints, FREAK)算法^[19]。Suárez 等人在 2019 年提出一种高效率的图像描述符(Boosted Efficient Local Image Descriptor, BELID)^[20],其特征使用积分图像来有效计算一对图像正方形区域的平均灰度值之差,并使用 BoostedSSC 算法^[21]选择一组特征组合形成强描述。次年, Suáñez 等人对 BELID 的特征提取过程进行改进,提出一种高效的二值局部图像描述符(Boosted Efficient Binary Local Image Descriptor, BEBLID)^[22],解决了不对称引起的匹配问题,并将描述符二值化。2015 年 Spyrou 等人^[23]对多种特征提取方法在胶囊内窥镜图像上进行了比较评估,ORB 特征点相比其它算法在速度方面有很大的优势,但是在进行特征点检测时,能提取到的 ORB 特征点数量是很稀少的,大部分特征点会出现在纹理更丰富的图像区域甚至导致特征点冗余,而纹理较弱或缺乏的区域几乎没有特征点。针对这一问题, Mur-Arta 在 ORB_SLAM3 中采用双阈值和四叉树来优化特征点提取^[24],但是该算法仍然使用人为设置的阈值,没有考虑像素邻域相关性这一条件,不能真正实现自适应提取。Ma 等人^[25]利用动态局部阈值来代替固定阈值,基于邻域图像块的灰度值计算局部阈值,能够提取更多的特征点。但是该方法只简单考虑了平均灰度值,这导致在该方法在灰度值变化较小的图像上效果大大降低。

1.2.2 误匹配剔除

特征匹配作为一项基础而关键的技术在视觉 SLAM 系统中有着重要地位^[26-28],特征匹配的准确与否直接影响到视觉 SLAM 系统的准确性和鲁棒性。但由于重复纹理、光照条件变化和相机成像变换等,在匹配过程中,不可避免的会产生大量错误匹配和虚假匹配。这些错误匹配会带来时间上的额外开销,加之匹配筛选阶段需要从特征匹配产生的匹配点对中筛除这些错误匹配,又会占用额外的时间资源。Fischler 等人早在 1982 年提出随机采样一致性算法^[29],这是错误匹配筛选阶段被人们一直以来公认的最普适的方法,其是通过从输入样本中随机选择样本,通过计算得到最佳模型,满足该模型的样本称为内点,也就是正确匹配点。近年来,为了提高 RANSAC 算法效率的精度,一些研究人员提出了许多改进算法。Torr 等人在 2000 年提出一种随机抽样最大似然估计(Maximum Likelihood Estimation Sample and Consensus, MLESAC)算法^[30],它使用和 RANSAC 相同的采样策略来生成假定的解决方案,但选择的是相似的最大值而不是内点数。Matas 等人在 2004 年提出一种随机 RANSAC (Randomized Random Sample Consensus, R-RANSAC)算法^[31],首先用 d 个随机选取的样本点来验证模型,一旦得以验

证,再用全部 N 个数据来验证,如果样本中的数据与模型不拟合,则直接放弃,进而减少了计算量以便提高运算效率。Chum 等人在 2005 年提出渐进样本一致性(Progressive Sampling Consensus, PROSAC)算法^[32],优化了 RANSAC 算法的采样策略,它不是从整个数据集中均匀采样,而是从逐渐增大的最佳对应点集合中进行采样,可以节省计算量提高运行速度。其它类似方法还有空间一致的随机采样一致性(Spatially Consistent Random Sample Consensus, SCRAMSAC)算法^[33]、全局 RANSAC(Universal-RANSAC, USAC)算法^[34]等,这类方法统称为基于重采样的方法,严重依赖于采样的准确性,当初始匹配中的错误匹配数量太多时,需要的采样次数大幅增加,效率大打折扣。Xu 等人^[35]在 2012 年根据统计思想选取输入数据的随机子集,并从每个子集中选取样本的所有选择点来拟合最佳模型,使得所用点的样本数量呈指数级下降,进而提高了算法的效率。Ci 等人^[36]等人在 2022 年提出利用所有匹配点对的前 20%和 80%组成新的样本集,用前 20%的匹配点对计算矩阵模型,并验证计算出的模型,利用变换模型验证剩余 80%是否为内点,提高了内点在样本点集中的比例,从而有效提高了图像匹配的精度和速度,但其二八比例分界点的固定性导致了其适应性的不灵活。Bian 等人在 2017 年提出一种基于网格的运动统计的特征匹配(Grid-based Motion Statistics, GMS)算法^[37],利用统计学的方法,在局部网格区域内统计匹配的数量,根据数量的多少来判定该网格区域内的所有匹配是否为正确匹配。Chen 等人^[38]在 2018 年对 GMS 进行了优化,将九格网格统计优化为五格网格统计,将 7 次旋转矩阵的计算改为 3 次旋转矩阵的计算,但降低了算法的旋转不变性。Zhu 等人^[39]在 2019 年提出了基于改进网格运动统计特征的 GMS-RANSAC,引入了距离一致性原理,利用距离之间的相似性来剔除异常值,提高了精度但运行时间增加了。Lan 等人^[40]在 2020 年提出一种改进的 GMS-RANSAC,使用双向暴力匹配和 GMS 算法后,进而基于 RANSAC 进一步剔除异常值,最后使用 GMS 算法验证所有特征对是否完全匹配。Zhang 等人^[41]在 2022 年提出将 GMS-RANSAC 运用到 ORB-SLAM2 中,也是目前在精度上最有效的剔除异常值的方法,虽然可以通过设置阈值来减少迭代次数以降低运行时间,但牺牲精度的代价太高,特别是大场景数据下精度和速度的矛盾问题无法调和。

1.2.3 多级匹配

在处理复杂场景时,单一级别的特征匹配方法往往不能满足实际需求,尤其在视觉 SLAM 系统中设计一种快速精确的误匹配提出策略仍是一个难点。Ye 等人为了提高匹配的精度,将 K 最近邻(K Nearest Neighbor, KNN)、近邻比、双向匹配、余弦相似度(Cosine Similarity, CS)和 PROSAC 结合起来,提出了 MP-ORB 匹配方法^[42]。Sun 等人^[43]为了解决在非均匀照明和无纹理随机出现而导致的亮度和对比度的不规则动态变化对匹配的影响,提出由 KNN、阈值滤波(Threshold Filtering, TF)、特征向量范数(Eigenvector Norm, EN)和 RANSAC 组成的多阶段匹配模块来消除不匹配。Sun 等人根据特征描述符在汉明空间中基于局部外观的相似性、局部图像结构的相似性和特征对在欧氏空间中的全局灰度信息来进行匹配,提出一种由 KNN、阈值滤波、双向滤波、特征向量范数和 RANSAC 组成的 KTBER 多阶段匹配技术^[44]。之后 Sun 等人^[45]引入 RTC (Ratio-test Criterion)技术

[46]来比较最近邻距离和次近邻之比与阈值的大小,提出由 KNN、阈值滤波、RTC、余弦相似度和 RANSAC 组成的 KRCR 多级匹配策略。

1.3 论文主要研究内容

本文的研究内容主要包括三方面的关键技术,一是特征点的提取,二是特征点错误匹配剔除,三是特征点的精细匹配。本文研究的总体框架如图 1-1 所示,主要内容概括如下:

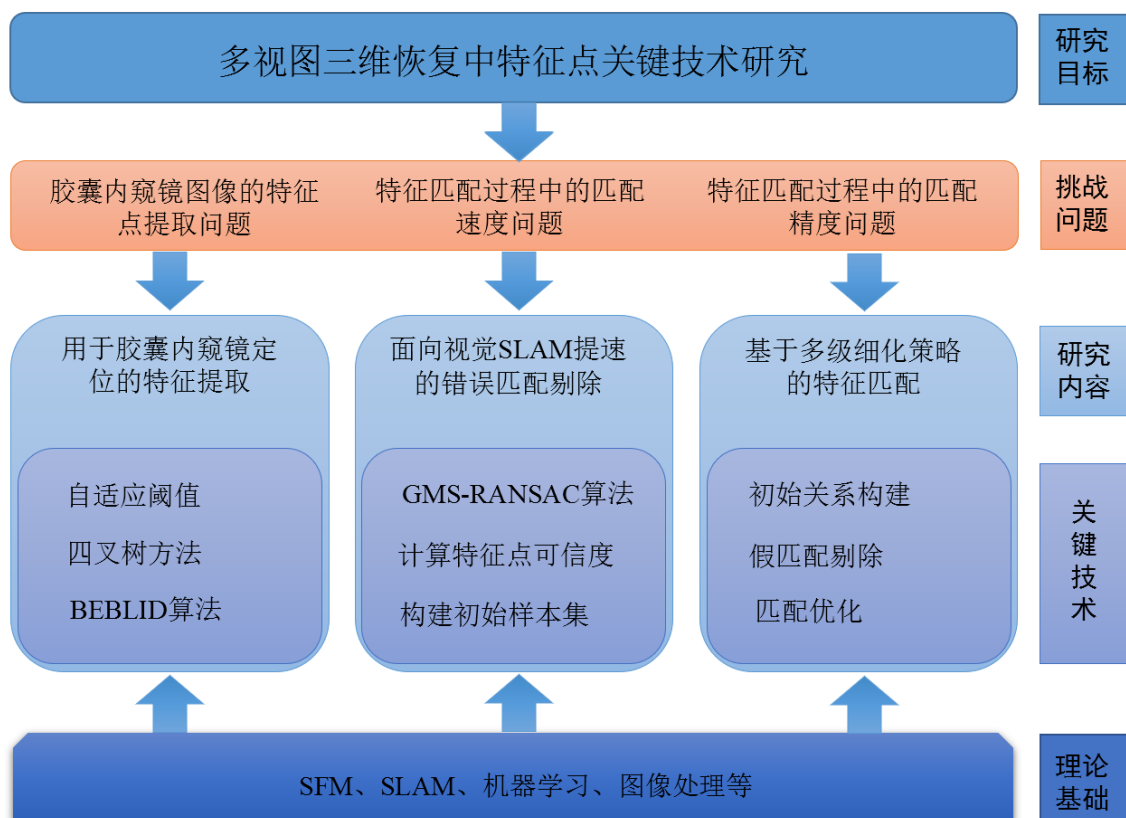


图 1-1 本文总体研究框架图

(1) 用于胶囊内窥镜定位的特征点提取

针对使用固定阈值的 ORB 算法在胶囊内窥镜图像变化不显著的区域难以提取特征点,显著区域出现特征点冗余的问题,拟设计一种改进的 ORB 算法来提取胶囊内窥镜图像的特征点。预通过分析固定阈值在图像局部灰度值之间的差异性较小场景下的不足,拟构造自适应阈值达到在同质区域稳定提取特征点的目的。然后研究基于四叉树方法的特征点平均分发机制,并分析 ORB 算法原有描述符的缺点,拟通过 BEBLID 算法来代替,以便实现数量充足的高质量特征均匀分布在胶囊内窥镜图像上的目的。

(2) 面向视觉 SLAM 提速的错误匹配剔除

针对由于需要随机选择所有的样本去进行迭代求解 RANSAC 的最优模型从而导致剔除错误匹配耗时较多的问题,拟设计一种面向视觉 SLAM 提速的 GMS 与 RANSAC 融合错误匹配剔除方法。分析和透视 GMS 算法和 RANSAC 算法的基本原理,并对融合使用时的不足进行分析,探究其中匹配对邻域内匹配对数的数量与 RANSAC 算法剔除错误匹配的关联,设计 RANSAC 算法的输入样本的选取方法,进而设计一种错误匹配

剔除方法，拟达到提高 GMS 和 RANSAC 融合的特征匹配方法剔除错误匹配的速度的目的。

(3) 基于多级细化策略的特征匹配

针对多级匹配方法中使用基于重采样的方法来剔除错误匹配时，基于重采样的方法受初始匹配准确性低的影响，造成匹配精度降低的问题，拟基于多级精细匹配策略，建立一种新的特征匹配方法。预通过深入分析多级匹配各阶段之间的关联，并对剔除错误匹配阶段的不足进行分析，探究初始匹配的准确性对基于重采样的方法的影响，拟加入运动的平滑性约束对初始匹配进行预处理，以便达到提高匹配精度的目的，进一步提升在复杂场景下有效剔除错误匹配和低质量匹配的能力。

1.4 论文组织结构

本文一共包括五个章节，各章节的具体内容如下所示：

第一章：绪论。本章首先重点介绍了特征点在多视图三维恢复中的地位以及关于特征点的提取与匹配的研究背景和研究意义，随后本文介绍目前国内外对于特征点提取与匹配的研究现状。最后介绍本文的主要研究内容和论文的整体结构。

第二章：用于胶囊内窥镜定位的特征点提取方法。本章分析了使用固定阈值的 ORB 算法在胶囊内窥镜图像上难以提取特征点的原因，以及目前解决方法的不足之处，针对这些不足提出了一种用于胶囊内窥镜定位的特征点提取方法。具体介绍了本文方法的实现流程，对其中的核心方法即自适应阈值、特征点均匀化和特征点描述进行详述。最后，在胶囊内窥镜图像上通过实验结果体现本章方法的有效性。

第三章：面向视觉 SLAM 提速的错误匹配剔除方法。本章分析了 GMS 和 RANSAC 融合算法剔除错误匹配速度偏慢的原因，并对 GMS-RANSAC 算法的基本原理进行剖析，提出了一种便于视觉 SLAM 增速的 GMS 与 RANSAC 融合错误匹配剔除方法。具体介绍了本章方法的实现流程，对其中的核心改进进行详述。最后，以在不同数据集上的实验结果体现本章方法较原 GMS-RANSAC 算法的速度提升。

第四章：基于多级细化策略的特征匹配。本章分析了多级匹配中导致匹配精度降低的原因，及目前多级匹配方法的不足之处，提出了一种基于多级细化策略的特征匹配方法。具体介绍了本章方法的实现流程，对其中的核心方法，即初始关系构建、假匹配剔除和匹配优化进行详述。最后，以在不同光照变化和模糊的图像上的实验结果，体现本章方法的有效性及精度优势。

第五章：总结与展望。本章对全文进行了总结，概述了当前提出的方法优势和以解决的问题，并提出了方法的局限性与未来工作的方向，为后续研究提供一些建议和思路。

第二章 用于胶囊内窥镜定位的特征点提取

虽然目前视觉 SLAM 取得了长足的进展并积累了丰富的研究成果,但相关的技术在一些特殊的环境诸如胶囊内窥镜胃肠环境中,其受光照、胃肠道壁表面纹理低和高光反射等特殊环境的影响,现有的使用固定阈值 ORB 算法难以提取到足够的特征点,且在变化大的区域提取的特征点会过度集中甚至重叠,导致其基于 SLAM 的三维恢复和定位等工作无法顺利开展。因此,本章节将通过研究图像局部灰度值之间的差异性对提取特征点的影响,对特征点的提取阈值进行研究,并探究特征点的均匀化方法和描述符,提出一种新的提取方法,拟达到在胶囊内窥镜图像上能够快速稳定提取到充足的分布均匀的特征点的目的。

本章的创新点和成果概括为:(1)提出了一种基于动态局部阈值的 ORB 算法,确保在胶囊内窥镜图像能够提取充足的特征点。(2)采用基于四叉树的方法对特征点进行管理和优化,达到特征点在胶囊内窥镜图像均匀分布的目的。(3)在特征点描述阶段,提出利用 BEBLID 算法生成二进制特征描述符,提高特征点的质量和准确性

2.1 引言

与传统的有线内窥镜相比,无线胶囊内窥镜(Wireless Capsule Endoscope, WCE)是用一个小型、可吞咽的内窥镜工作的,内窥镜安装在一个胶囊中,作为人类消化系统的无创无痛、无交叉感染的临床成像和诊断工具已被普遍使用。但是,无论采集路径规划还是病灶诊断^[47],都离不开 WCE 在腔内的具体位置^[48-50]。目前市售的 WCE 主要采用磁力定位技术^[51-53]进行定位,这种技术需要昂贵的设备同时,也可能会对人产生不良影响,而且不能直接提供胶囊在胃肠道内移动的距离信息,因此用于胶囊内窥镜定位的视觉里程计方法应运而生。然而,目前能够实现良好相机定位的 SLAM 方法^[4,54]和用于三维结构恢复的 SFM(Structure from Motion)方法^[55]对内腔视频几乎是失效的,主要原因是胃肠道内表面的高光性和视频帧的低频率导致无法提取和跟踪所致。在本章中,提出了一种基于 WCE 图像的特征提取方法,提升了完全基于视觉特征的位移技术来进行 WCE 胶囊定位的可能性。

针对 WCE 的定位的需求更加关注特征点的有效提取,目前主要有基于尺度空间的 SIFT 算法^[12]、SURF 算法^[13]、FAST 算法^[14]和 ORB 算法^[16]等。在这些算法中,ORB 是最有效的并得到人们的青睐,但通过实验发现它在提取 WCE 的特征点方面几乎是失效的。首先,在特征点的检测方面,ORB 算法采用 FAST 点提取算法。FAST 通过比较图像中较突出的像素与周围像素的灰度值,选择图像中比较突出的像素作为特征点。而 FAST 算法使用任意设置的全局固定阈值,其影响是检测到的特征点过于集中甚至是重叠。在 WCE 图像没有明显变化的区域,只能提取到少数甚至没有特征点。其次,在特征点描述阶段,ORB 算法采用增强的 BRIEF^[15]描述符对特征点进行描述。BRIEF 算法通过在特征点的图像窗口中随机选择的像素来比较灰度值,生成二进制的特征描述。然而,在比较像素灰度值的过程中,只使用像素点的灰度值之间的关系,未能充分利用像

素点周围的灰度值信息，会导致图像信息的丢失。

为此，本章提出了一种改进的 ORB 方法。在特征点的检测阶段，基于邻域图像块的灰度利用变异系数^[56]计算出动态的局部阈值进行改进，然后利用四叉树方法对特征点进行管理和优化。在特征描述阶段，再使用增强有效二进制局部图像描述符(BEBLID)来描述特征点，这样在 WCE 图像上也可以提取到有效特征点。实验表明，该方法对 WCE 视频特征点的提取有更好的适应性，特别是在几乎不损失速度的前提下，对变化较小的胶囊内窥镜图像区域也能提取到足够多的特征，为 WCE 无线胶囊定位提供了良好的支撑。

2.2 方法概述

总体来看，本章方法首先采用基于变异系数计算局部自适应阈值的特征点提取方法，然后利用四叉树方法对特征点进行平均和分发，在特征描述阶段，采用准确率更高的 BEBLID 描述子来对特征点进行描述。本章方法的技术路线如图 2-1 所示。

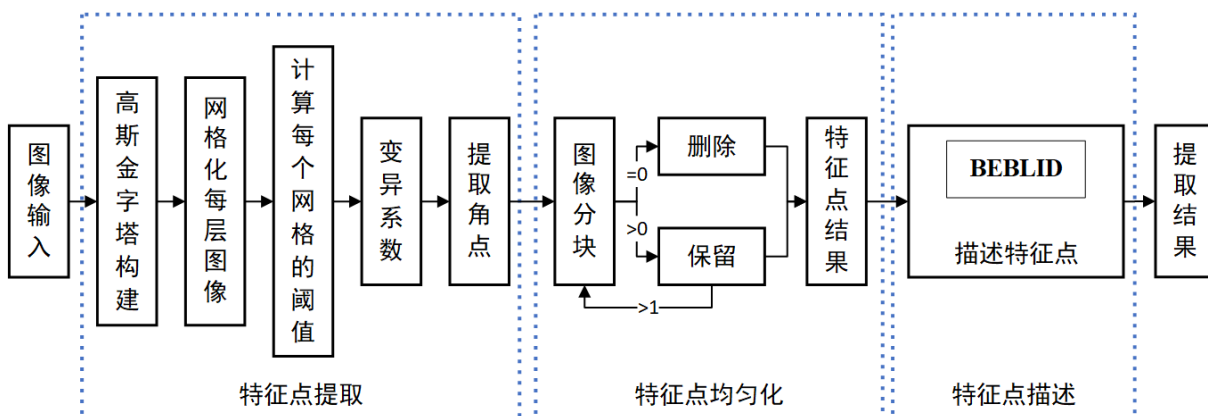


图 2-1 本章方法技术路线图

在 FAST 算法提取角点的过程中，阈值的选择决定了提取的结果，由于胶囊内窥镜所拍摄的腔内场景复杂多变，FAST 固定的阈值选取并不能适应腔内场景的特征提取需求。为此，本章首先将图像按块进行划分，每个块算作一个独立区域，每个像素点的灰度值算作一个样本值，然后计算像素点灰度值之间的变异系数当作当前图像块的阈值。

此外，为了解决特征点集中甚至冗余的问题，本章利用四叉树方法对提取的特征点进行筛选和均匀化。其具体过程为将一张图像看作一个节点，如果该节点中的特征点数量大于 1，那就把该节点分为 4 个节点，如果特征点数量为 0，则删掉该节点；然后，往复进行这个操作，直到节点数量大于需要的特征点的数量或者无法再分裂为止。

最后，进一步考虑实时性的要求，本章对提取的特征点的描述采用了 BEBLID 方法。BEBLID 描述子是根据像素对的比较来呈现特征的，其速度比较快的关键是比较次数的有限性，其本质也是为了追求速度而牺牲了部分准确性的折衷。

2.3 方法细述

2.3.1 自适应阈值

ORB 算法通过构造高斯图像金字塔，并采用 FAST 算法来完成特征点的选取。FAST 算法^[14]是通过比较一个候选像素点周围的像素灰度来判定是否为特征点。以候选像素点

p 为中心, 计算半径为 3 的圆的 16 个像素点的灰度值并与 p 点比较, 将这些像素点按顺时针方向编号为 1 到 16。如果这 16 个像素点中有 N 个连续的像素点与中心像素点 p 的灰度差的绝对值大于等于 t (t 为阈值), 则判定中心像素点 p 为特征点。在实际应用中, 为了加快判定过程, 可以计算 p 点与像素 1、5、9、13 之间的灰度差, 如果有三个大于或小于 t 的差异, 则进一步与其他像素进行比较。如果不满足条件, 则直接排除点 p 。

固定阈值的方法会使计算简单, 减少计算时间, 但是这样的方法使用在高光和变化较少的胶囊内窥镜图像上并不能满足特征点提取的要求, 为此本章采用变异系数计算阈值 t 来改进这一问题并达到自适应性的需要。

变异系数是概率分布离散程度的一个归一化量度, 可以用来衡量各像素点的变异程度。当数据之间的测量尺度相差太大时, 可以利用变异系数来消除测量尺度的影响。变异系数的计算如式(2-1)所示。

$$c_v = \frac{\sigma}{\mu} \quad (2-1)$$

其中 c_v 是变异系数, σ 为样本的标准差, μ 为样本的平均值。

利用变异系数计算自适应阈值时, 首先根据设定的网格大小对整个图像进行网格划分, 每个网格作为一个独立的图像区域。在每个网格中阈值 t 的计算如式(2-2)所示。

$$t = \frac{D_i}{M_i} * \alpha \quad (i=1, 2, \dots) \quad (2-2)$$

其中, D_i 表示网格 i 中像素点的灰度值的标准差, M_i 表示网格 i 中像素点的灰度值的平均值, α 的计算如式(2-3) -(2-4)所示。

$$\alpha = \frac{M_i}{M'_i} \quad (2-3)$$

$$M'_i = \frac{\sum_{j=1}^{n_i} f_i(x, y) - f_i(x, y)_{\max} - f_i(x, y)_{\min}}{n_i - 2} \quad (2-4)$$

其中, M'_i 为网格 i 中像素点的灰度值的截断平均值, n_i 表示网格 i 中像素点的个数, $f_i(x, y)$ 表示网格 i 中 (x, y) 处的灰度值, $f_i(x, y)_{\max}$ 表示网格 i 中灰度值的最大值, $f_i(x, y)_{\min}$ 表示网格 i 中灰度值的最小值。

2.3.2 特征点均匀化

原始 ORB 算法^[16]在提取特征点时, 由于在纹理丰富的区域提取到大量特征点而在纹理较少或缺乏纹理的区域提取到较少的特征点, 因此提取的特征点会出现聚集和冗余的情况。为了使胶囊内窥镜图像中的特征点分布更加均匀, 本章采用了四叉树方法对 2.3.1 节中提取到的特征点采用了均匀化的策略。

首先使用 2.3.1 节中提出的基于动态局部阈值的 FAST 算法来完成特征点的提取。为了进一步均匀化图像中的特征点, 利用四叉树方法对图像的二维空间进行分块, 然后对每个块中的特征点进行处理。具体步骤如下:

步骤 1、将整个图像作为四叉树的初始节点, 得到初始的四叉树结构。

步骤 2、对图像中的每个节点，判断其中特征点的数量。如果特征点个数为 0，则删除该节点，如果特征点个数为 1，则该节点不再被细分。如果节点中的特征点数大于 1，则该节点继续细分成为 4 个子节点。

步骤 3、重复步骤 2，当节点数达到所需的特征点数，或者无法再进行分裂，则四叉树的拆分结束。

步骤 4、选择每个节点包含的特征点。如果特征点数量大于 1，则选择响应值最大的特征点来表示该节点。

通过上述方法，特征点可以均匀地分布在胶囊内窥镜图像区域内。

2.3.3 结合 BEBLID 算子的特征点描述

BEBLID 算法是在 BELID 基础上改进的一种二进制描述符，其使用 AdaBoost 算法^[57]在不均衡的数据集进行训练，从而帮助选择弱学习器(weak-learner, WL)。BEBLID 和 BRIEF 类似，都是二进制的描述符，也是基于灰度值的差异，但是 BEBLID 计算的是一对图像正方形区域的平均灰度值的差，并将它们组合成一种强描述，用于表示中心区域接近高斯分布的局部描述信息。在特征点周围选择固定数量的不同大小的方形区域，且每对方形区域大小相同，其通过特征提取函数计算每对方形块中像素点的平均灰度值的差异来描述特征点。BEBLID 描述符提取示意图如图 2-2 所示，红色和蓝色的框表示以特征点 p 为中心，大小为 s 的方形区域， $f(x)$ 为特征提取函数， T 为阈值， $h(x)$ 代表特征。BEBLID 特征描述符的提取函数 $f(x)$ 定义如式(2-5)所示。

$$f(x; p_1, p_2, s) = \frac{1}{s^2} \left(\sum_{q \in R(p_1, s)} I(q) - \sum_{r \in R(p_2, s)} I(r) \right) \quad (2-5)$$

其中 $I(q)$ 、 $I(r)$ 是像素点 q 和 r 的灰度值， $R(p_1, s)$ ， $R(p_2, s)$ 是中心位于 p_1 、 p_2 的方形边框，大小是 s 。

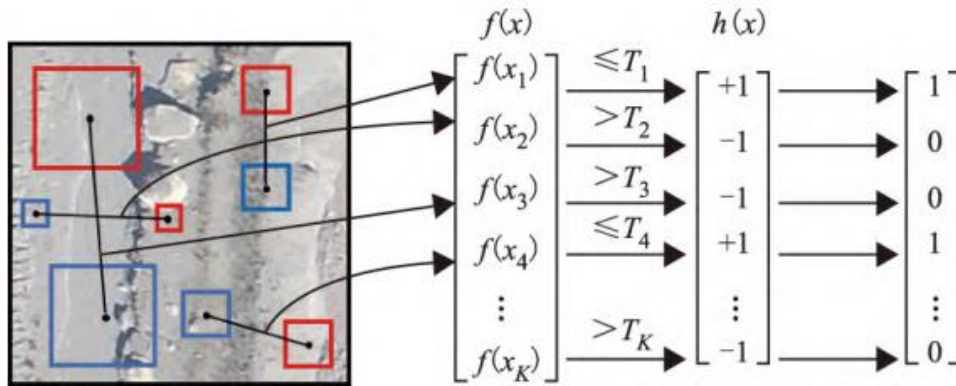


图 2-2 BEBLID 描述符提取示意图

将特征提取函数得到的结果与阈值 T 进行比较，如果大于阈值 T ，则记为-1，反之则记为+1。为了最终的结果的为二进制的形式，将-1 转化为 0，+1 转化为 1，从而得到 BEBLID 描述符。判断公式如式(2-6)所示。

$$h_x(x; f, T) = \begin{cases} +1 & \text{if } f(x) \leq T \\ -1 & \text{if } f(x) > T \end{cases} \quad (2-6)$$

2.4 实验与分析

本节通过实验来验证本章提出的方法在胶囊内窥镜图像上的可行性以及与其它方法之间的对比分析，主要内容包括评价指标选择、数据集选择、实验结果以及与现有方法的实验对比分析。本节的实验在基于 Intel i7-4710MQ 处理器、8GB 内存、500GB 硬盘容量、Ubuntu18.04 系统的笔记本电脑上进行。

2.4.1 数据集选择

本章在 PillCam 采集的一组 120 个胶囊内窥镜视频数据集上进行评估，分辨率为 576*576 像素，每个视频持续时间大约为 2.5 个小时，共 45000 帧，包含完整的使用胶囊内窥镜进行拍摄的全过程视频。这些视频是在医院利用 PillCam 胶囊内窥镜采集而来的真实的数据集。该数据集总共包括大量胃、结肠和小肠三个部位的视频，本章只使用小肠部位的视频中图像帧来进行实验评估，将胃和结肠部位的视频片段排除在本研究之外。

2.4.2 评价指标选择

本章的实验中采用了四种评价指标。第一个标准是特征点的数量，数量是最直观反应特征点提取效果的指标之一。

第二个指标是匹配精度，计算公式如式(2-7)所示：

$$precision = \frac{m_c}{m} * 100\% \quad (2-7)$$

其中， m_c 表示正确匹配的匹配数量， m 表示没有消除错误匹配的匹配数量。

第三个指标是时间，本章方法的运算时间包括特征点提取时间、描述时间和匹配时间。不考虑输入处理时间和指标计算时间。

第四个指标是方向误差，其含义是通过匹配好的特征点之间的对应关系，恢复出在两帧之间摄像机的运动，但是往往因为误匹配的存在，所以求出的摄像机的运动与真实的运动之间是存在的偏差的。

2.4.3 实验结果

在本部分将验证本章的方法在胶囊内窥镜图像上的可行性。由于目前不存在胶囊内窥镜的真实值的数据集，因此通过使用采样视频帧的旋转变换来评估提出的方法。本章采用了和 Spyrou 等人^[58]相同的评估方法，具体来说，为了从胶囊内窥镜的视频中获得一组具有代表性的帧，以 5fps 的帧速率从数据集中选定视频帧进行旋转变换和/或缩放变换，然后，从两帧图像中提取特征点，并使用 RANSAC 确定原始视频帧和旋转视频帧之间的内部对应关系^[59]。基于这些对应关系，估计一个变换矩阵，从中提取相应的旋转和/或缩放非常简单。最后，通过将这些值与相应的已知人工值进行比较，从而估计了旋转(方向)和/或位移(缩放)的误差。

为了估计两幅图像之间的转换，本章使用了 Opencv 的内置 RANSAC 算法来实现。为了确定临时对应集，必定会导致更多的随机实验，因此本章设定了 30 次实验，取 30 次实验的平均值作为平均误差。



图 2-3 小肠部位的胶囊内窥镜图像

图 2-3 是真实的胶囊内窥镜图像，所需要匹配的对象是在真实胶囊内窥镜图像的基础上通过旋转和缩放得到的。在实验中，待匹配得图像有两种情况：一种是只有旋转没有进行缩放，共测试了 8 个不同的旋转角度，从 5 度到 45 度，旋转步长为 5 度。另一种是在 8 个不同的旋转角度的基础上，将图像按 0.9 的比例进行缩放。匹配结果如图 2-4、图 2-5 所示，每组图像显示正确的匹配结果和错误的匹配结果。绿线表示匹配正确，红线表示匹配错误。

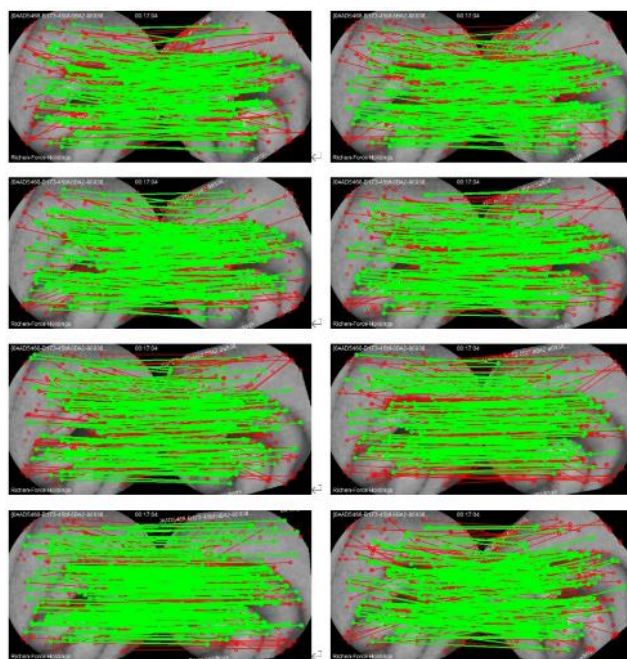


图 2-4 旋转 8 个角度的匹配结果

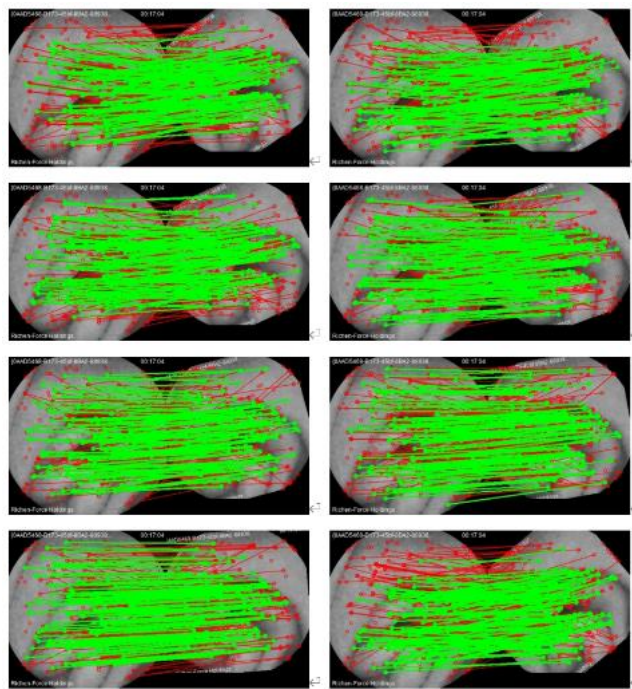


图 2-5 旋转 8 个角度并按 0.9 比例缩放的匹配结果

图 2-4、图 2-5 分别为两种情况的结果。图 2-4 显示了将胶囊内窥镜图像旋转 8 个角度进行测试的算法的图像匹配效果，图 2-5 显示了将胶囊内窥镜图像旋转 8 个角度并按 0.9 比例缩放进行测试的算法的图像匹配效果，图中的结果为从左到右，自上而下角度依次增大。从图 2-4 和图 2-5 中可以看出，本章的方法在缩放和进行不同角度旋转的胶囊内窥镜图像上，都能提取充足的特征点，并能完成匹配。但是匹配结果比较相似，难以直观区分差异。为了更详细地分析算法地性能差异，对选取的 8 组图像组的平均绝对方向误差进行统计，如表 2-1 所示。相较于无缩放的误差，有缩放的误差会更小。不进行缩放时的误差在在 5 度到 20 度之间是比较大的，而进行缩放后只有在 5 度时是比较高的。

表 2-1 原始比例和 0.9 比例下每个旋转角度的估计误差

实际旋转角度	无缩放时的平均绝对方向误差	有缩放时的平均绝对方向误差
5	21.4035	13.3594
10	4.3211	0.2672
15	36.0287	6.7480
20	35.1488	2.0812
25	2.0323	0.6260
30	3.8051	0.3977
35	4.7720	5.5424
40	3.7651	4.5578

2.4.4 方法对比分析

本部分将展示本章的提出的方法和其他类似方法在胶囊内窥镜图像上的结果对比分析，以便更清晰的体现本章提出的方法的优势。在实验中，分别采用 ORB、当前最好的基于视觉特征的位移技术 ORB_SLAM3 中的特征提取方法^[24]和本章提出的改进方法在 5.3 节中的数据集上进行实验得到比较结果。

第一个评价指标是数量，本章设置目标特征点的数目为 500，表 2-2 展现了 ORB、ORB_SLAM3 中的特征提取方法和本章的方法提取的特征点数量，其中平均值是 8 组实验的平均数量。两种情况下，ORB 提取的特征点平均数量分别为 187 和 159 个，ORB_SLAM3 中的特征提取方法提取的特征点平均数量分别为 478 和 476，远远达不到提取要求，本章改进的方法在不同的角度变换以及有无缩放的场景下提取数目都能达到本章设定的目标数目。由此实验可以看出，本章改进后 ORB 算法在胶囊内窥镜图像上的提取效果都优于 ORB 和 ORB_SLAM3 中的方法。

表 2-2 ORB、ORB_SLAM3 的特征提取方法和本章方法提取的特征点数量

实际旋转角度	ORB		ORB_SLAM3		本章方法	本章方法
	无缩放	有缩放	无缩放	有缩放	无缩放	有缩放
5	148	136	464	459	507	507
10	171	139	468	480	507	505
15	170	143	482	474	507	505
20	172	143	482	483	507	506
25	199	169	482	482	507	506
30	208	175	486	482	507	507
35	218	177	484	481	507	503
40	210	191	477	474	504	503
平均值	187	159	478	476	507	505

第二个评价指标是精确率，相比于 ORB_SLAM3 中的方法和本章的方法，ORB 算法在胶囊内窥镜图像上提取的特征点数量少了很多，因为本章方法在意的是提取足够的特征点的同时保证匹配的精确率，所以将 ORB 方法排除在外，只对比 ORB_SLAM3 中的方法和本章提出的方法。实验结果如表 3 所示，在无缩放的情况下，ORB_SLAM3 中的特征提取方法的平均精确率为 51.7150%，本章的方法为 54.7703%。在有缩放的情况下，本章的方法的平均精确率比 ORB_SLAM3 中的特征提取方法低了 1.8157%，但在角度较小的情况下，本章提出的方法是较优的，在 5-15 度之间平均精确率提高了 0.9233%。

表 2-3 ORB、ORB_SLAM3 的特征提取方法和本章方法的匹配精确率

实际旋转角度	ORB_SLAM3(%)	ORB_SLAM3(%)	本章方法(%)	本章方法(%)
	无缩放	有缩放	无缩放	有缩放
5	62.5430	52.8986	61.4458	54.2587
10	50.1832	45.3571	52.7508	45.5975
15	48.9437	54.0146	52.4845	55.1839
20	51.2027	58.5185	55.3459	52.9032
25	49.1525	54.6387	57.4675	55.1155
30	54.5139	53.9855	51.7886	51.3245
35	49.3243	56.8905	53.3762	47.5570
40	47.8571	53.0466	53.5032	52.8846
平均值	51.7150	53.6688	54.7703	51.8531

第三个标准是时间，实验结果如图 2-6 所示，本章改进的 ORB 方法和 ORB_SLAM3 中的方法的计算时间都高于 ORB 方法。但考虑到 ORB 提取的特征点数量少很多，所以可以排除在外。虽然优化后的 ORB 算法的提取速度比优化前要慢，但在效果更好的情况下可以和 ORB_SLAM3 中的方法保持几乎相同的计算时间。可以满足系统的实时性要求。

第四个指标是方向误差，表 2-4 显示了 ORB、ORB_SLAM3 中的特征提取方法和本章的方法在有无缩放情况下 8 个角度上的方向误差。其中，平均值表示 8 个角度的平均绝对方向误差的平均值。从表 2-4 可以看出，在没有缩放的情况下，ORB 的平均误差为 28.2001 度，ORB_SLAM3 中的特征提取方法为 15.4883 度，本章的方法为 13.9095 度。从表 2-5 可以看出，在有缩放的情况下，ORB 的平均误差为 14.5736 度，ORB_SLAM3 中的特征提取方法为 21.2225 度，本章的方法只有 4.1975 度。实验结果表明，本章的方法优于 ORB 和 ORB_SLAM3 中的方法，在胶囊内窥镜图像上具有一定的鲁棒性。

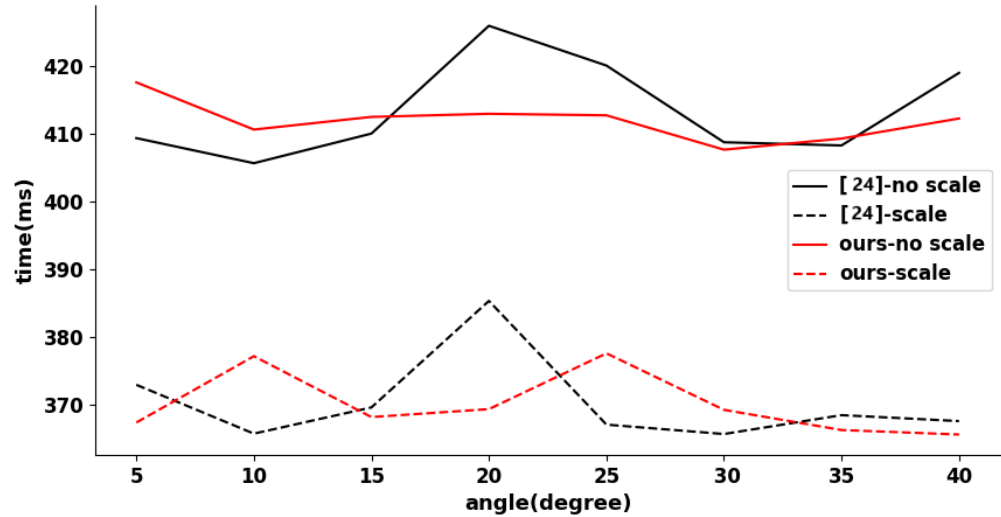


图 2-6 三种方法在有无缩放情况下的平均的提取时间

表 2-4 三种方法在无缩放情况下的平均绝对方向误差

实际旋转角度	ORB	ORB_SLAM3	本章方法
5	26.6887	20.1484	21.4035
10	34.4837	11.3446	4.3211
15	80.6692	40.2004	36.0287
20	6.6569	7.6963	35.1488
25	24.9077	13.7508	2.0323
30	23.3149	9.3273	3.8051
35	28.7053	19.8742	4.7720
40	0.1743	1.5647	3.7651
平均值	28.2001	15.4883	13.9095

表 2-5 三种方法在缩放情况下的平均绝对方向误差

实际旋转角度	ORB	ORB_SLAM3	本章方法
5	0.3835	37.5493	13.3594
10	16.9175	11.0954	0.2672
15	2.6067	17.0243	6.7480
20	35.5334	23.4020	2.0812
25	5.1912	45.4308	0.6260
30	11.3170	2.8833	0.3977
35	26.4908	31.7732	5.5424
40	18.1486	0.6219	4.5578
平均值	14.5736	21.2225	4.1975

2.5 本章小结

采用完全基于视觉特征的位移技术来进行无线胶囊内窥镜（WCE）定位时，合适的视觉特征提取技术十分重要。本章提出了一种基于局部自适应阈值、二叉树方法和

BEBLID 描述子的改进 ORB 算法,以达到在胶囊内窥镜图像上稳定提取特征点的目的。本章的方法基于 ORB 算法,重点在特征点提取阶段和特征点描述阶段进行研究和改进。首先,采用基于变异系数计算局部自适应阈值的特征点提取方法,该方法可以稳定地提取胶囊内窥镜图像上的特征点。同时,利用四叉树方法实现对特征点的管理和优化,避免了特征点的过度集中和重叠。在特征点描述阶段,采用 BEBLID 描述子来进行特征描述,保证算法的正确率和速度。通过对胶囊内窥镜图像数据集的验证,本章提出的改进 ORB 算法在胶囊内窥镜图像有较好的表现,相较于 ORB_SLAM3 中的方法在有缩放时旋转角度较小的情况下的平均匹配精度提高了 0.9233%,无缩放情况下的平均匹配精度提高了 3.0553%,有效提高了特征点提取的数量,特征点能够均匀分布在图像内,对胶囊内窥镜图像这种纹理较弱的区域有良好的适应性。

第三章 面向视觉SLAM提速的错误匹配剔除

在基于视觉的三维恢复和地图构建中, 基于 GMS 和 RANSAC 融合的特征匹配方法在求解 RANSAC 的最优模型的过程中, 依然需要随机选择所有的样本去进行迭代求解, 未能充分利用 GMS 算法得到的结果, 严重影响了匹配速度, 妨碍了视觉 SLAM 实现的实时性。因此, 本章在 GMS-RANSAC 算法的基础上, 分析 GMS 算法中得到的匹配对的邻域内匹配对的数量与后续算法的关联性, 以及 RANSAC 算法的输入样本集对算法的影响, 设计一种错误匹配删除的方法, 拟达到提高 GMS 和 RANSAC 融合的特征匹配方法剔除错误匹配的速度的目的。

本章的创新点和成果概括为: (1) 提出了一种基于可信度的 GMS 结果优化的方法, 确保了提供给 RANSAC 样本的准确性。(2) 实现了一种减少 RANSAC 输入样本的策略, 缩短了 GMS-RANSAC 算法剔除错误匹配的时间。

3.1 引言

实时性是 SLAM 的核心特性, 在视觉 SLAM 中更加备受人们的关注^[60]。而特征点提取与匹配是视觉 SLAM^[61-62]中的最关键和最耗时步骤, 其是指在不同时间、不同视角或不同传感器拍摄的同一场景的两幅图像之间建立可靠的特征对应关系, 以便获得图像与图像、或图像与地图之间的关系描述^[63], 并为 SLAM 后续的相机姿态估计、优化操作, 以及各类任务的完成等提供有效支撑。加之图像特征的局部性, 错配现象普遍存在, 虽然通过迭代优化等一系列附加过程的策略, 在一定程度上降低了错误率^[64], 但错误加入的匹配和附加的过程都会带来时间上的额外开销, 加重了对视觉 SLAM 实时性能的提升, 并已成为 SLAM 发展的瓶颈问题之一。

近年来, 在视觉 SLAM 中, 更倾向于使用随机采样一致性技术^[29]来解决错误匹配剔除问题, 且是大家认可的一种通用方法。但当数据量较大时, 为了达到相应的精度要求, 则必须增加迭代的次数, 其匹配时间会迅速增大, 这对必须保持实时性的 SLAM 来说是不可容忍的^[6]; 但是, 如果减少匹配次数必将牺牲系统的精度, 进而降低了 SLAM 的适应范围。通常, SLAM 往往面对的是大场景下的大数据量, 则很难同时确保匹配时间短且精度高。为此, 张等人最近提出了 GMS-RANSAC 方法并将其运用到 SLAM 系统中来来平衡这一问题的解决^[41], 即首先通过 GMS 算法筛选一遍, 然后再利用 RANSAC 进一步消除错误匹配。虽然 GMS-RANSAC 是目前视觉 SLAM 在复杂环境下精度较高的方法, 但依旧在好多场合满足不了实时性的要求。

为此, 本章受到 GMS-RANSAC 的启发, 提出一种改进的基于网格运动估计(GMS)和 RANSAC 相融合的剔除错误匹配的方法, 其通过优化 RANSAC 的输入样本, 在确保精度的同时可大幅度减少迭代次数, 缩短匹配时间。通过实验比较, 时间平均节省了 23.59%。

3.2 方法概述

本章方法的技术路线如图 3-1 所示。首先通过 ORB 算法分别提取两张图像的特征

点,然后基于描述子求取汉明距离对 ORB 特征进行暴力匹配。紧接着用 GMS 算法对已经得到的匹配结果进行粗略筛选,以便减少匹配次数。GMS 算法利用特征点周边邻域的匹配对数量来判断当前特征对是否是正确匹配,正确匹配的体现为特征附近的匹配对数量充足,而错误匹配的则体现为特征附近的匹配对几乎为零。为此,本章将特征点周边邻域的匹配对数量记为该特征点的可信度,匹配对数量越多,可信度越高,当前特征点为正确匹配点的可能性越高,反之亦然。简而言之就是,依据 GMS 算法所得的可信度对匹配对进行排序分组,而将 RANSAC 算法原来的随机选取,改变为优先从可信度较高的匹配对中进行,直到达到精度要求,这样可以降低迭代次数以便换回运行时间效率。

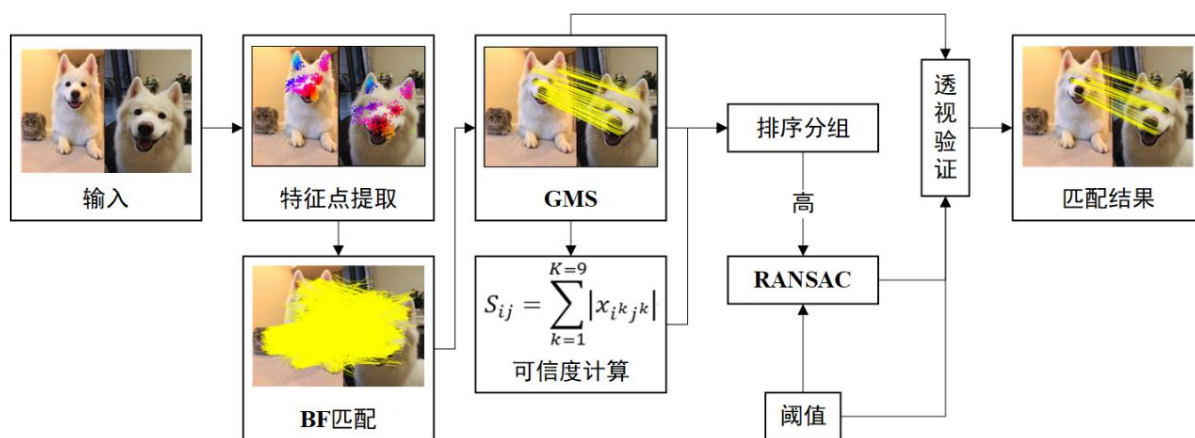


图 3-1 本章方法技术路线图

3.3 方法细述

3.3.1 GMS-RANSAC 方法分析

GMS^[37]被称作基于网格运动统计的快速鲁棒的特征匹配方法,其基于运动的平滑性会导致匹配的特征点周围区域的匹配点较多这一事实,则通过计数周围区域的匹配点个数多少来判断一个匹配是否正确。正确匹配的特征点附近的匹配对数量充足,而错误匹配的特征点附近的匹配对几乎为零。

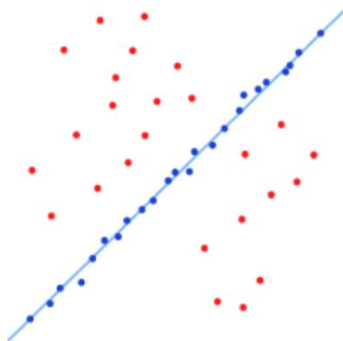


图 3-2 RANSAC 示意图

RANSAC 也称作随即抽样一致方法^[29],通过迭代随机抽样和拟合模型的方法从一组包含离群数据的观测数据中估算出数学模型的参数。RANSAC 算法假设数据中包含异常数据,这些异常数据被记作外点,正确数据被记作内点。图 3-2 为 RANSAC 的示意图,其中蓝色的点为内点,红的点为外点,而直线为 RANSAC 找到的直线。

RANSAC 算法的实现如算法 3-1 所示。

算法 3-1 RANSAC 算法

输入：匹配结果

输出：筛选结果图像

- 1: 设匹配点为样本集。
 - 2: 从样本集中随机选取 4 对点作为假设内点, 根据它们之间的变换关系求解变换模型。
 - 3: 计算模型中所有匹配点的投影误差, 将误差在阈值范围内的点作为模型适用的内点, 计算内点集合中匹配特征对的个数。
 - 4: 如果当前内点个数大于最优集合中内点个数时, 更新最优集合, 迭代次数增加 1 次。
 - 5: 如果迭代次数等于预设次数, 则将最终的内点点集记录为当前最优集合, 否则将内点点集设置为新的输入样本集, 然后再执行 1。
 - 6: 利用当前最优集合对模型进行重新计算, 得到最终的数学模型。
 - 7: 利用所求的模型对匹配结果进行筛选, 去除异常值。
-

根据 RANSAC 算法, 其迭代次数估算如下:

假设内点在数据集中的概率为 t , 如式(3-1)所示。

$$t = \frac{n_{inliers}}{n_{inliers} + n_{outliers}} \quad (3-1)$$

其中, $n_{inliers}$ 是内点数, $n_{outliers}$ 是外点。

那么每次计算模型都使用 n 个点的情况下, n 个点都是内点的概率为 t^n , 则 n 个点中至少有一个是外点的概率(采样失败的概率)为 $1-t^n$ 。也就是说, 在迭代 k 次的情况下, $(1-t^n)^k$ 表示 k 次随机抽样都没有找到一次全是内点的情况。因此, 能采样到正确的 n 个点去计算正确模型的概率 p 如式(3-2)所示。

$$p = 1 - (1 - t^n)^k \quad (3-2)$$

由此可得迭代次数 k 如式(3-3)所示。

$$k = \frac{\log(1-p)}{\log(1-t^n)} \quad (3-3)$$

最优模型受迭代次数的限制, 迭代次数的上限与获得最佳模型的概率有很强的一致性, 即迭代次数的上限越大, 获得最佳模型的概率越大。但是增加迭代次数的同时, 降低了算法的速度。

由此可见, 在 ORB-SLAM3 系统中, RANSAC 粗略地迭代更少的次数, 以确保响应时间满足实时性要求, 同时肯定降低了算法的准确度。而 GMS-RANSAC 算法是先利用 GMS 算法筛选之后, 再将筛选所得结果利用 RANSAC 算法进一步处理, 这样在求解 RANSAC 的最优模型的过程中, 能够减少求得最优模型所需要的迭代次数, 从而提升算法的速度。但是该方法依然需要随机选择所有的样本去进行迭代求解, 同样会浪费大量

时间。

3.3.2 方法改进

GMS-RANSAC 算法在对待所有 GMS 算法得到的匹配对时，由于不加区分的随机选择了初始集，显然忽略了特征点之间的差异，从而导致了不必要的迭代次数。由于在 GMS 算法中，特征点的邻域得分与特征点是正确匹配的概率之间是正相关的，因此可以考虑根据特征点的邻域得分来选择初始匹配对集。即首先根据邻域得分的高低将匹配对进行排序和分组，然后利用邻域得分越高的，优先作为初始集来求解最优模型，以便降低迭代次数节约处理时间。改进后的 GMS-RANSAC 算法实现如算法 3-2 所示。

算法 3-2 改进的 GMS-RANSAC 算法

输入：两张图像

输出：匹配结果图像

- 1: 提取两张图片的 ORB 特征点。
 - 2: 计算描述子的汉明距离，对 ORB 特征点进行暴力匹配。
 - 3: 用 GMS 算法对前一步得到的匹配结果进行粗略的筛选，剔除错误匹配。
 - 4: 记录 GMS 算法求解过程中计算的特征点邻域范围内的匹配对数量作为该特征点的可信度。
 - 5: 对 GMS 算法求解的结果根据对应的可信度进行排序，并根据可信度的高低将结果按 1/2 的比例分为两组。
 - 6: 只用可信度高的一组作为 RANSAC 求解最优模型的输入样本，并根据求得的最优模型对全部结果进一步去除异常值。
-

3.4 实验与分析

本节旨在通过实验来验证本章方法的可行性，并与其它方法进行对比分析，主要包括评价指标选择、数据集选择、实验结果、参数变化实验、以及与现有方法的实验对比分析。实验在基于 Intel i7-4710MQ 处理器、8GB 内存、500GB 硬盘容量、Ubuntu18.04 系统的笔记本电脑上进行。

3.4.1 数据集选择

本章采用 GMS 中使用的图像数据来展现实验结果，在实验对比分析小节选用 TUM^[65]和 KITTI^[66]两个数据集来进行实验。TUM 和 KITTI 两个数据集提供了大量室内和室外场景，这些数据集对特征提取和匹配都提出了巨大的挑战，且更能验证本章方法的有效性。

3.4.2 评价指标选择

本章主要采用精确率、召回率、以及算法运行时间三个标准指标来评价和比较本章方法。

第一个标准指标是精确率 *Precision* 如式(4)所示。

$$Precision = \frac{Num_{correct_matches}}{Num_{correct_matches} + Num_{false_matches}} \times 100\% \quad (3-4)$$

其中, $Num_{correct_matches}$ 为正确匹配次数, $Num_{false_matches}$ 为错误匹配对的个数。

第二个标准指标是召回率 *Recall* 如式(5)所示。

$$Recall = \frac{Num_{correct_matches}}{Num_{correct_matches} + Num_{true_matches}} \times 100\% \quad (3-5)$$

其中, $Num_{correct_matches}$ 为正确匹配次数, $Num_{true_matches}$ 为实际为真实匹配点对但是没有匹配上的点的个数。

第三个标准指标是运行时间, 本章算法的运算时间包括特征点提取时间、描述时间和匹配时间, 但不包括输入处理时间和指标计算时间。

3.4.3 实验结果

本部分将展现本章改进的 GMS-RANSAC 算法的实验结果。在算法流程中, 对输入的两张图片提取相同数量(预设数量为 3000)的 ORB 特征点之后, 对两张图片的特征点进行暴力匹配, 示例结果如图 3-3 所示。

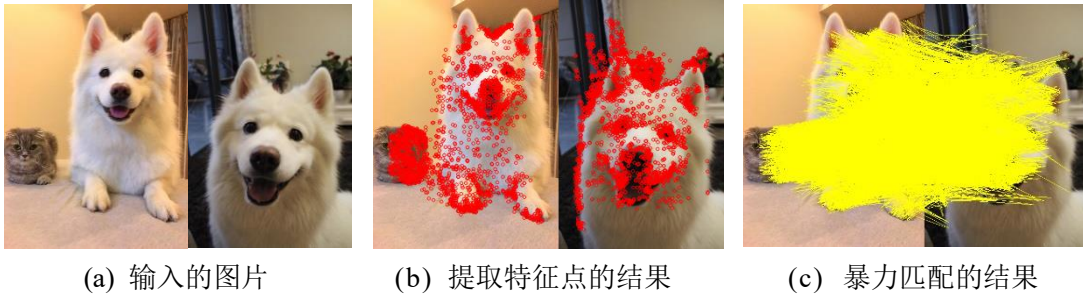


图 3-3 暴力匹配结果示例

进而利用 GMS 算法进行筛选, 剔除错误匹配, 减少输入给 RANSAC 错误的样本数量。GMS 筛选的结果如图 3-4(a)所示。然后对所得结果进行 RANSAC 筛选, 进一步剔除错误匹配, 得到为本章方法的结果, 如图 3-4(b)所示。

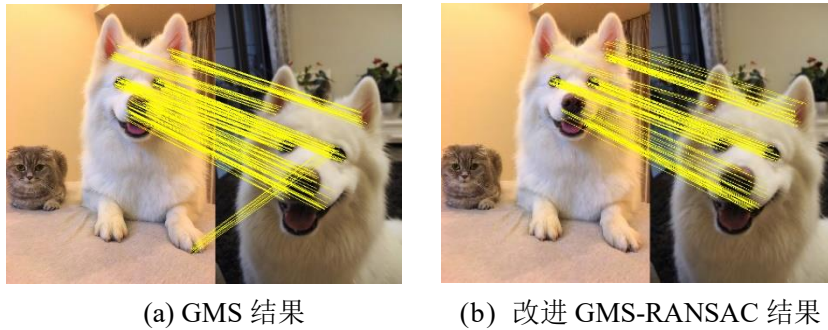


图 3-4 暴力匹配之后不同方法的筛选结果

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/316135021154010053>