

运动推断结构技术中的特征跟踪方法综述

曹明伟 李书杰 贾 伟 刘晓平

(合肥工业大学计算机与信息学院 合肥 230009)

摘 要 运动推断结构(Structure From Motion, SFM)是一种从图像或视频序列生成三维点云模型的技术. 由于 SFM 能够产生与输入图像或视频场景具有较高几何一致性的三维模型, 因此, 吸引了国内外学者的广泛关注, 并将其应用于计算机视觉和图形学领域生成场景的三维模型, 例如一些经典的应用有三维重建、增强现实、虚拟现实和无人驾驶等. 特征跟踪作为运动推断结构的关键基础技术, 可用来在图像和视频序列中寻找对应的匹配点, 这些匹配点的质量直接影响着 SFM 生成的三维点云模型的几何形状. 因此, 为了提高 SFM 生成的点云模型与真实场景在几何上的一致性, 研究者们提出了大量的特征跟踪方法, 这些方法主要分为两类: 处理无序图像的特征跟踪方法和处理视频序列的特征跟踪方法. 然而, 在现有的特征跟踪方法中, 研究者们主要集中于在特定应用环境下如何提高特征跟踪的精度与时间效率, 而忽略了特征跟踪领域尚未解决的一些问题. 为了促进特征跟踪技术的发展和提高 SFM 生成的点云模型的质量, 该论文综述了现有的特征跟踪方法. 具体地说, 该论文首先详细地分析了处理无序图像集合和视频序列的特征跟踪方法, 阐述每个特征跟踪方法的核心思想、优点与不足. 其次, 总结了特征跟踪领域一些可用的算法, 如特征检测子、特征描述子和特征匹配方法, 以及一些标准的测试数据集. 第三, 在不同类型的数据集上评估了一些经典的特征跟踪方法, 以便为选择合适的特征跟踪方法提供参考. 第四, 总结了特征跟踪领域一些迫切需要解决的问题, 讨论了影响特征跟踪方法时间效率和精度的主要因素, 以便促进特征跟踪技术的发展. 最后, 探讨了特征跟踪技术的发展趋势, 为未来研究指明方向.

关键词 运动推断结构; 特征跟踪; 三维重建; 特征检测子; 特征描述子

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2018.02536

A Survey on Feature Tracking Methods for SFM

CAO Ming-Wei LI Shu-Jie JIA Wei LIU Xiao-Ping

(School of Computer and Information, Hefei University of Technology, Hefei 230009)

Abstract Structure from Motion (SFM) is a technique which takes image sequences or video sequences as input, then produces 3D point-cloud model. SFM technique has received a significant amount of attention in the past years both in computer vision and computer graphics communities motivated by applications in 3D reconstruction, augmented reality, virtual reality, map navigation, 3D change detection, scene completion, 3D salient object detection, and driverless car, due to the model produced by SFM technique has high geometric consistency with the scenes of images and video sequences. Generally speaking, feature tracking is a key fundamental component for SFM technique, which is used to produce feature matches from consecutive two image sequences or video frames, and the quality of the produced feature matches has a significant effect to the shape of the 3D model produced by SFM technique. Thus, in order to improve the geometric consistency of the 3D model with real scenes, a large number of feature tracking methods have been proposed,

收稿日期: 2016-11-08; 在线出版日期: 2017-07-21. 本课题得到国家重点研发计划(2016YFC0800100)、国家自然科学基金(61370167, 61305093, 61673157, 61802103, 61877016)和安徽省自然科学基金(JZ2015AKZR0664)资助. 曹明伟, 男, 1986年生, 博士研究生, 中国计算机学会(CCF)会员, 主要研究方向为大规模三维重建、计算机视觉、增强现实. E-mail: cmwqq2008@163.com. 李书杰, 女, 1983年生, 博士, 讲师, 硕士生导师, 主要研究方向为模式识别、智能系统. 贾伟, 男, 1978年生, 博士, 副研究员, 硕士生导师, 中国计算机学会(CCF)会员, 主要研究方向为计算机视觉、模式识别. 刘晓平(通信作者), 男, 1964年生, 博士, 教授, 中国计算机学会(CCF)杰出会员, 中国计算机学会(CCF)常务理事, 主要研究领域为3D建模、计算机动画、协同计算与仿真. E-mail: lxp@hfut.edu.cn.

according to the different inputs, these methods can be roughly divided into two categories: the first type is used to handle unordered images where the resolutions are often not same to each other, and another one is used to process video sequences where each frame has the same resolution. However, among these feature tracking methods, most of them mainly focus on the problem that how to improve the accuracy and efficiency for specific applications such as indoor scene reconstruction and outdoor reconstruction where many repetitive features are existed, thus the essential drawbacks about the feature tracking problem has received few attentions. Although, some researchers have been proposed many state-of-the-art feature tracking methods in recent years, but it is very difficult to select a good one from the existing feature tracking methods for some special applications. In other words, we urgently need some guides to help us to make decision when selecting a suitable feature tracking method. To advance the research progress of feature tracking method, and improve the quality of the 3D model produced by SFM technique, in this paper we have made a comprehensive investigation for feature tracking methods. In detail, we first analyze some state-of-the-art feature tracking methods, and elaborate the core idea, advantages and disadvantages behind them. Second, we collect a significant amount of available resources such as feature detectors, feature descriptors and feature matching methods as well as benchmark datasets; these resources are very valuable both in theoretical research and practical applications. Third, we have made a comprehensive assessment for the state-of-the-art methods on various datasets such as 3D reconstruction dataset with scale variance and lighting variance, based on this evaluation we presented many valuable experiences which can be used as a guide to select a suitable feature tracking method from the existing methods for the specific applications. Fourth, we concluded some hard problems which should be solved in the future, and discussed what could affect the efficiency and accuracy of the feature tracking methods, this discussion can further push the research progress of feature tracking. Finally, we discussed the development trends of feature tracking, so as to point out the directions for feature research.

Keywords structure from motion; feature tracking; 3D reconstruction; feature detector; feature descriptor

1 引言

随着计算机视觉与图形学技术的发展,产生了丰富的信息表达方式,这些表达方式多数是建立在三维模型基础之上,因此,三维模型在信息的呈现上起到至关重要的作用.例如,虚拟现实、增强现实、无人驾驶和无人机导航等技术,均建立在三维场景模型基础之上.运动推断结构(Structure From Motion, SFM)^[1-8]是从图像或视频序列中恢复三维模型的一种技术,相比传统的以 3D MAX、Sketch Up、PhotoScan 等专业的 3D 建模软件创建三维模型, SFM 技术不需要专业的建模技术基础,具有操作便利和成本低廉等优点,因此吸引了国内外学者的广泛研究并取得了较大研究进展,相关成果已经被应用于计算机视觉的多种任务中,例如,大规模三

维重建^[1,2,4,6,9]、增强现实^[10-11]、三维地图构建^[12]、三维人脸重建^[13]、变化检测^[14]、动作识别^[15-16]、自动驾驶^[17]、场景填充^[18]和全景图生成^[19]等.目前,已有的比较知名 SFM 系统有: Bundler^[20]、ETH-3D^[21]、ACTS^[22]、LS-ACTS^[23]、Visual SFM (VSFM)^[24]、Hyper SFM^[25]、Dynamic SFM^[26]、1DSFM^[27]和 LS-SFM^[28]等.在 SFM 系统中,特征跟踪是重要的组成部分,也是 SFM 系统生成三维模型的第一步,是 SFM 研究的重点内容之一.

特征跟踪作为运动推断结构的关键基础技术,可用在在图像和视频序列中寻找对应的匹配点,从而保证生成的点云与真实场景在几何上的一致性.根据 SFM 系统的不同输入对象,可以将特征跟踪方法分为两类:(1)处理视频序列的特征跟踪方法;(2)处理无序图像序列的特征跟踪方法,这些无序的图像数据主要来自于一些图像共享网络,例如

Flickr 和 Facebook 等图片分享网站. 针对视频序列,特征跟踪过程具有时序性,即按照时间顺序在相邻两帧之间进行特征匹配. KLT(Kanade Lucas Tomasi)是经典的视频序列下的特征跟踪方法^[29-31],此方法通过在连续的视频之间寻找目标之间的偏移量实现特征跟踪. 针对无序图像集合,不存在时序性,需要在任意两幅图像之间进行特征匹配. 基于特征检测与匹配框架(Detection and Matching Framework,DMF)是处理无序图像集合的经典特征跟踪方法. DMF 类方法主要包含 3 个关键步骤:(1)特征点检测;(2)计算特征描述子;(3)特征匹配. 例如 Bundler、ETH-3D、ACTS、VSFM 等 SFM 系统的特征跟踪方法均建立在 DMF 框架之上. 在这些 SFM 系统中,主要采用局部特征检测算法检测特征点,并采用强力匹配(Brute Force Matching, BFM)方法进行特征点匹配,最后采用 RANSAC 算法去除错误匹配. 由于 DMF 是一个特征跟踪框架,因此对每一个组成部分进行等价替代或改进,即可获得新的特征跟踪方法. 随着众多特征检测和匹配方法被相继提出,研究者们提出大量的基于 DMF 框架之上的特征跟踪方法,从而进一步促进了 SFM 术的研究和发展.

国外的很多研究机构,如华盛顿大学 Grails 实验室、CMU 机器人研究所、德国慕尼黑工业大学、北卡大学、MIT Media Lab、美国康奈尔大学、斯坦福大学人工智能实验室、微软 Redmond 研究室、Google 研究院和 Acute3D 公司等,已经对特征跟踪和 SFM 技术进行了多年研究,并以多项重要的研究成果产业化. 例如微软公司的 PhotoSynth 系统以及 Google 公司的 MapGL 地图导航系统,均建立在以特征跟踪为基础的 SFM 技术之上. 与之相比,国内的学术机构也早已开展了相关研究工作,主要研

究机构包括浙江大学 CAD&CG 国家重点实验室、北京航空航天大学虚拟现实国家重点实验室、清华大学、中国科学院自动化研究所和计算技术研究所、中国科学院深圳先进技术研究院、北京大学、北京理工大学、上海交通大学、山东大学、上海科技大学、西安交通大学和东北大学人工智能与机器人研究所等,探索出一系列具有价值的研究成果,相关研究成果已经在国际著名期刊和顶级会议上进行了报道,其代表作如文献[5,22-23,32-58].

由于研究者们提出的特征跟踪方法有各自的优点与不足,对于 SFM 技术的初学者和一些企业的工程技术人员难以从已有的大量方法中选择合适的特征跟踪方法. 为了便于为 SFM 系统选取和设计合适的特征跟踪方法,本文对现有的特征跟踪方法进行综述,并对一些经典的特征跟踪方法进行详细分析和对比,同时总结现有方法存在的问题以及未来研究方向,为设计鲁棒的 SFM 系统奠定基础.

本文第 2 节全面综述现有特征跟踪方法,并汇总可用的资源;第 3 节在公共数据集上对现有经典特征跟踪方法进行评估,并在 SFM 系统中测试不同特征跟踪方法对三维重建的影响;第 4 节总结特征跟踪方法亟待解决的问题以及可选的解决方案;第 5 节给出特征跟踪的发展趋势和展望;最后是本文的总结.

2 特征跟踪方法分析

在过去 30 年中,大量的特征跟踪方法被提出,图 1 为特征跟踪方法的发展历史进程. 除 KLT 算法外,均采用作者加日期的方式表达具体的特征跟踪方法,同一年提出的方法,按不同的月份进行排序. 下面将对每类特征跟踪方法进行分析.

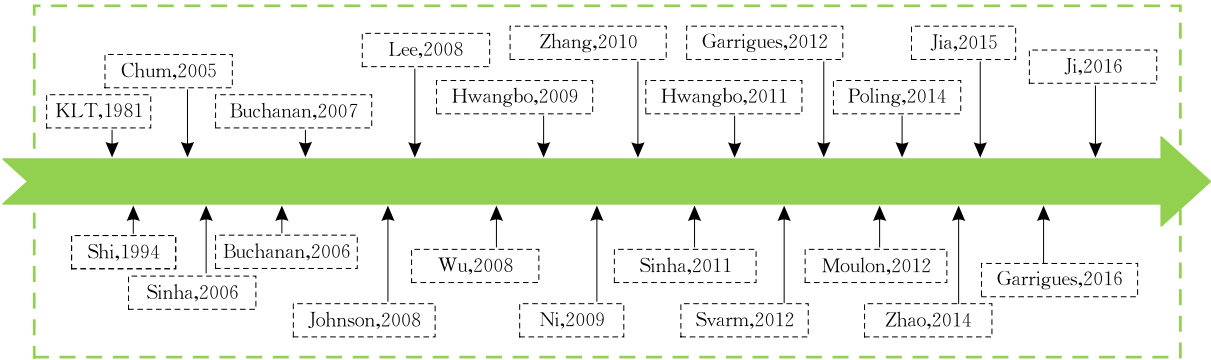


图 1 特征跟踪方法进程

2.1 KLT 类方法

KLT^[29-31]特征跟踪方法自 1981 年被提出以

来,被广泛用于视频序列的特征跟踪. KLT 方法根据相邻两幅图像之间的位置偏移计算特征点,并采

用光流对特征点进行跟踪. 由于 KLT 方法需要大量的计算才能确定特征点的位置, 因此, 在一些计算能力有限的设备上难以有效运行. 随着图形处理器 (Graphics Process Units, GPU) 的出现以及在并行计算方面表现出的高效性, Sinha 等人^[59-60]在 GPU 上采用图形编程语言 OpenGL 和 CG 实现了 KLT, 即 KLTGPU 方法. Sinha 等人是首次在 GPU 上实现 KLT 算法, 为后续算法的改进提供了帮助, 在特征跟踪方法的并行化道路上具有重要的里程碑意义.

图 2 为 KLTGPU 方法的流程, 可以看出 KLT-GPU 方法有 3 个组成部分: (1) 构造金字塔, 根据文献[61]可知, 图像金字塔的创建极其耗时, 也是导致 KLT 方法效率低下的主要原因; (2) 特征点再次选择; (3) 特征跟踪. 在实现过程中这 3 部分分别由不同的 GPU 内核代码完成, 且每个模块的内核代码分布在不同的 GPU 处理器上执行, 实现了指令与数据的同时并行. 根据文献[59]的测试结果, 在相同环境下 KLTGPU 方法的效率是 KLT 的 20 倍.

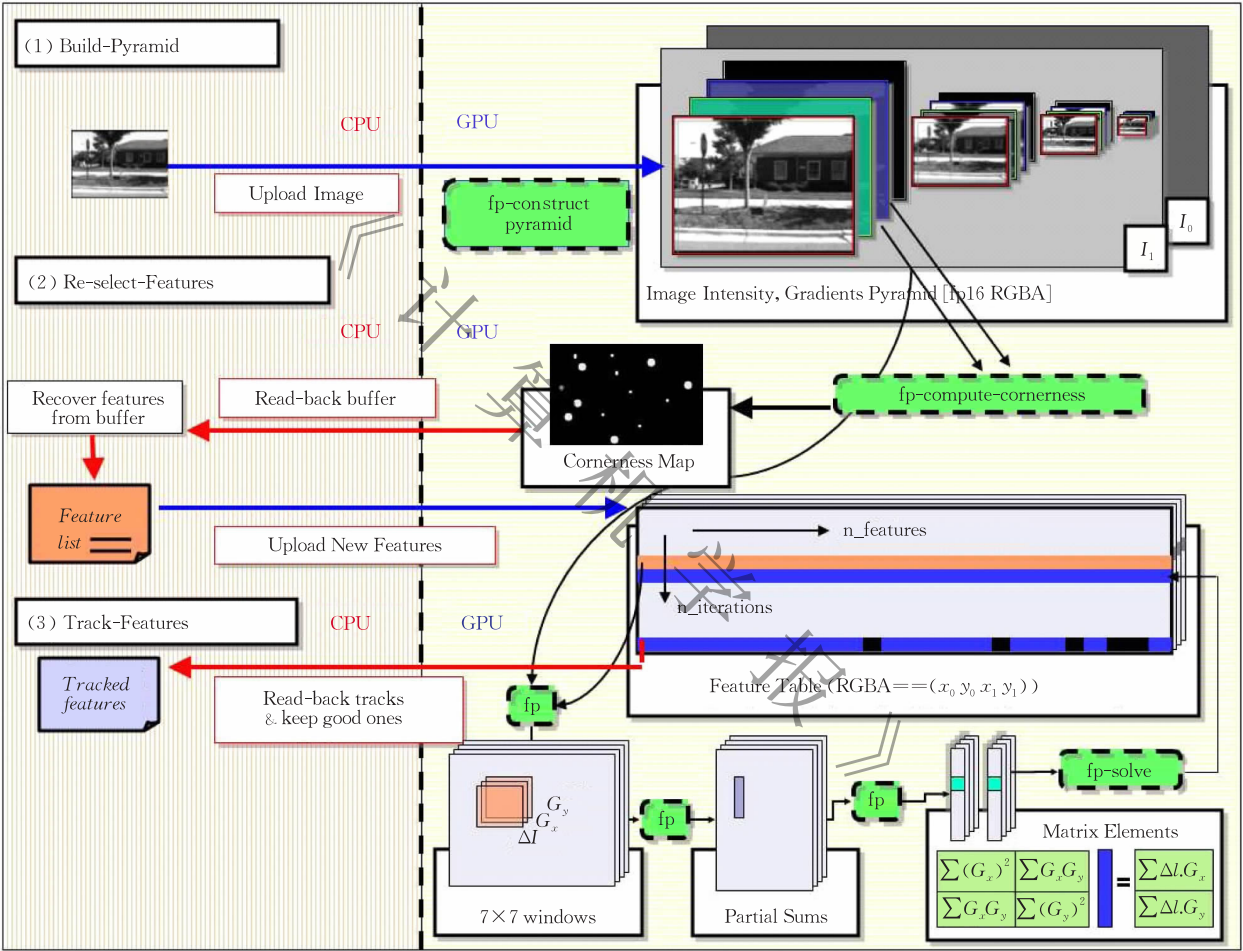


图 2 KLTGPU 方法流程^[59]

Zach 等人^[62]认为, 随着 GPU 硬件体系结构的变化, KLTGPU 方法没有充分利用新一代 GPU 架构的硬件资源. 此外, KLTGPU 方法在采样频率较高的视频序列上难以实时获取每个 GPU 处理器发送的数据流, 即多路数据的同步存在问题, 容易导致特征跟踪丢失. 为了弥补 KLTGPU 方法存在的不足, Zach 等人在新一代体系结构的 GPU 上采用 CUDA 实现了 KLT 方法, 并采用文献[63]提出的实时立体流数据获取技术较好地解决了数据并行的同步问题. Buchanan 等人^[64]在 KLT 基础之上提出

一种结合局部与全局运动的特征跟踪方法, 首先检测特征点和拟合运动模型, 然后以单个特征点作为 KLT 的输入反向跟踪 $M-1$ 帧, 最后将两次都匹配的特征点组合为特征跟踪链. 根据文献[64]所述, 两遍跟踪策略较好地弥补了 KLT 算法在目标遮挡和图像噪声等方面的不足.

随着移动设备 (如智能手机) 的广泛使用, Hwangbo 等人^[65-66]将 KLT 方法应用于移动设备的特征跟踪. 由于移动设备的自我运动 (Ego-Motion) 造成视频序列出现较大的抖动, 从而导致 KLT 方

法在跟踪特征点时丢失(Drift)频率较高. 为了弥补特征丢失问题,Hwangbo 等人从两个方面对 KLT 方法进行改进:(1)为 KLT 方法选择好的初始化条件;(2)采用惯性传感器(Inertial Measurement Unit, IMU)辅助 KLT 方法进行特征跟踪. 此外,为了提高时间效率,Hwangbo 等人在实现 KLT 方法时充分利用了移动设备的 GPU 资源.

图 3 展示了文献[65]使用的 IMU 辅助的 KLT 方法原理,可以看出,对于新的视频帧 I_{t+1} 采用 p_{t+1}^0 代替不在收敛区域内的特征点 p_t , 此外还利用摄像

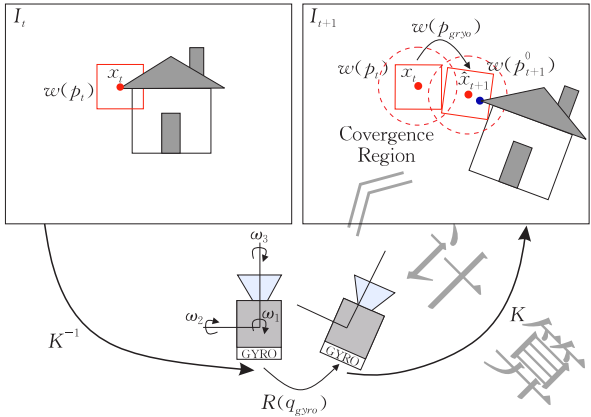


图 3 IMU 辅助的 KLT 方法原理^[65]

机的旋转信息 $p_t^{t+1}(gyro)$ 计算两幅图像之间的单应性矩阵(Homography Matrix)以便为 KLT 方法选择好的初始化条件. 实验结果表明,Hwangbo 等人在移动设备上实现的 KLT 算法充分利用了 IMU 信息,较好地解决了较长视频序列的特征跟踪丢失问题.

2.2 DMF 类方法

如前所述,DMF 是基于特征检测与匹配的适用于无序图像集合的特征跟踪框架,此特征跟踪框架包含以下 3 个组成部分:(1)特征检测;(2)计算特征描述子;(3)特征匹配. DMF 的流程如图 4 所示,对于给定的无序图像集合,首先检测图像上的特征点(图 4(b)上的黄色圆形表示 SIFT 特征点),然后计算每个特征点的描述子(图 4(c)上的绿色矩形是 SIFT 特征描述子的可视化表示);其次,根据描述子之间的相似性对特征点进行匹配(图 4(d)表示两幅图像之间的 SIFT 特征匹配结果),重复这一过程直到输入图像集合处理完毕,最后生成特征跟踪链(如图 4(e)所示). 由于在 DMF 类方法的每个部分都有多种算法可供选择,为了便于分析具体的特征跟踪算法,在 2.3 节我们将对现有的特征检测子、特征描述子、匹配算法和可用测试数据集进行总结.

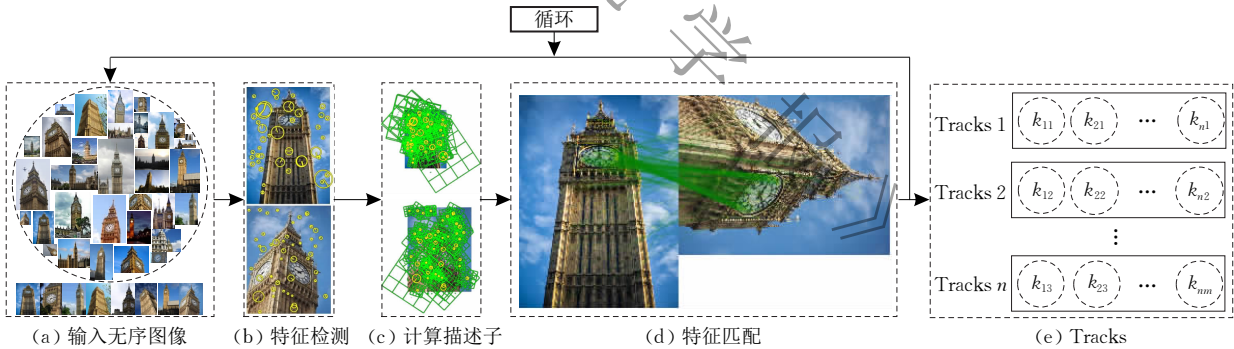


图 4 DMF 类方法流程

DMF 思想首次被应用于 Bundler 系统^[20],并取得了较大成功,自此,DMF 思想在 SFM 领域被广泛采用,后继的研究主要集中于效率和精度的提升. 具体地说,Bundler 系统采用 SIFT 算法检测特征点和计算对应的特征描述子,采用 BFM 算法对描述子进行强力匹配,从而确定不同图像上特征集合之间的对应关系. 由于 SIFT 和 BFM 算法的计算效率较低,难以满足系统的时间效率需求. 为了提高时间效率,研究者们提出了多种不同的策略. 例如 ETH-3D 系统采用比 SIFT 更加快速的 SURF 算法检测特征点和计算特征描述子,从而提高了 ETH-3D

系统的时间效率. 根据文献[24]所述,在处理分辨率较大的图像时,SURF 的时间效率仍然需要进一步提高. 此外,还发现在相同数据下,SURF 检测出的特征点数远少于 SIFT,容易加剧特征跟踪丢失. 为了提高时间效率,Wu 等人^[67]采用 OpenGL 和 CUDASDK 在 GPU 设备上实现了 SIFT 特征,即 SIFTGPU 特征,此特征被成功用于 VSFM^[24]和 LS-ACTS^[23]等系统. 此外,一些研究者在特征匹配方法上做出大量努力进一步提高特征跟踪效率,如 Garcia 等人^[68]在 GPU 上实现 KNN 算法——gKNN,用于加速特征匹配. 为了避免对 GPU 硬件

资源的依赖, Muja 等人^[69]在层次化聚类之上提出一种快速的二进制特征匹配方法, 采用 KD-Tree 加速特征匹配. 后来 Muja 等人^[70]将文献[69]的方法做进一步拓展, 提出一种适用于高维特征的快速近似最近邻方法(Fast Approximate Nearest Neighbor, FANN), 此方法发展为快速特征匹配的工具包——FLANN(Fast Library for Approximate Nearest Neighbors)^[71]. 受 Muja 等人思想的启发, Buchanan 等人^[72]提出一种可交互的特征跟踪(Interactive Feature Tacking, IFT)方法, IFT 方法采用 KD-Tree 和动态规划(Dynamic Programming, DP)技术加速特征匹配, 同时此方法可以根据用户输入, 选择性地跟踪一些特征点.

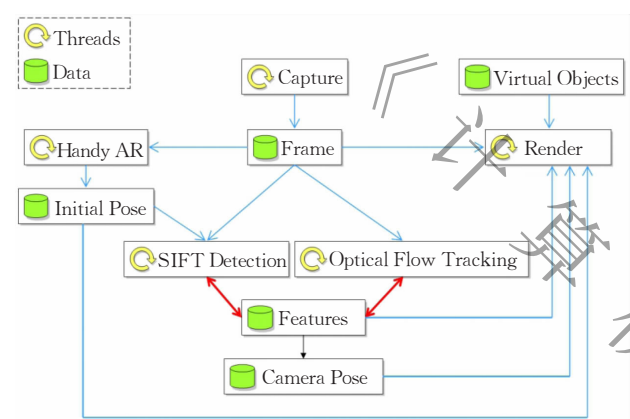


图 5 HFT 方法结构^[72]

SFM 系统所追求的另一目标是快速生成精确的三维点云模型, 而特征跟踪方法的鲁棒性和精确性对点云模型的质量影响较大. 因此, 在 DMF 框架下, 一些研究者侧重于提高特征跟踪方法的精度. 对于最为常见的特征跟踪丢失问题, Zhang 等人^[33]发现在极线附近的特征点容易出现丢失, 为了解决此问题, Zhang 等人提出一种基于两遍匹配策略的特征跟踪算法, 即第一遍匹配采用传统的 SIFT 描述子的相似性作为度量依据, 第二次根据对极几何原理, 采用平面的运动分割(Planar Motion Segmentation, PMS)作为匹配约束, 此算法被成功应用于 ACTS 系统中. 后来, Zhang 等人将此方法进一步扩展, 提出一种适用于非连续图像集合的 ENFT(Efficient Non-Consecutive Feature Tacking)^[23]方法, 已被成功用于 LS-ACTS 系统. 由于 ENFT 方法的使用, 为 SFM 系统的鲁棒性奠定很好基础, 实验结果表明 LS-ACTS 系统产生的点云模型在精确性方面远好于 VSFM 系统. Wu 等人^[73]发现视点变

化对特征匹配的精确性有较大影响, 因此, 提出一种视点不变块(Viewpoint-Invariant Patches, VIP)特征用于高分辨率的图像匹配, 但是 VIP 特征的计算很耗时, 在一些计算性能较低的移动设备上难以有效运行. 为了提高特征跟踪准确性, Chum 等人^[74]采用 RANSAC^[75]算法去除错误的特征匹配, 通过随机采样策略将满足矩阵约束的最大数目的特征集合作为特征匹配结果. 后来 Ni 等人^[76]提出一种 GroupSAC 算法用于去除错误的特征匹配, 在高分辨率图像下的测试结果表明了此算法的有效性. 为了满足手持增强现实设备的需求, Lee 等人^[77]提出一种混合特征跟踪(Hybrid Feature Tracking, HFT)方法, HFT 方法是基于特征检测、描述子匹配、光流跟踪、RANSAC 和多线程技术的结合, 在时间效率和精确性方面均有提高, HFT 方法的结构如图 5 所示. 为了弥补特征的歧义性问题, Svarm 等人^[78]提出一种基于图理论的特征跟踪方法(Graph-Theoretical Approach to Point Tracking), 此方法根据逐个特征匹配建立点图(Point-Graph, PG), 其中每个节点表示被检测的特征点、边表示特征点之间的匹配关系, 通过求解 PG 的每个连通路径创建特征跟踪链, 最后采用 Gomory-Hu^[79]算法去除错误匹配. 为使 SFM 技术能够产生较稠密的点云模型, Garrigues 等人^[80]提出一种实时的半稠密特征点跟踪(Real-Time Semi-Dense Point Tracking)方法, 稠密性主要体现在采用的 FAST 特征点检测算法具有较高的重复性(Repeatability), 此方法已被用于 Video Extruder^[81]系统, 在移动设备上能够实时提取场景目标的运动轨迹. 图 6 展示了 Video Extruder 系统在移动设备上的运行实例, 可以清楚看见行人的稠密运动轨迹, 足以证明此特征跟踪方法的有效性.



图 6 Video Extruder 系统运行实例^[81]

随着子空间学习(Subspace Learning, SSL)理论的发展, 为特征跟踪方法提供了新的研究途径.

例如 Poling 等人^[82]提出子空间约束 (Subspace Constraints) 的特征跟踪方法, 此方法能够同时跟踪一系列的特征集合, 允许在不同图像之间共享信息, 因此, 能够有效地适应目标的形变问题. Jia 等人^[32]提出一种新的基于 SSL 的特征跟踪框架——ROML, 并采用 ADMM 算法对 ROML 进行求解. ROML 框架的提出具有里程碑意义, 采用不同的求

解算法, 可以获得不同的特征跟踪方法, 为后继研究者指明方向. 最近, Zhao 等人^[34]将度量学习 (Metric Learning) 应用于特征跟踪, 进一步促进了机器学习算法在特征跟踪领域的研究; 图 7 为基于度量学习的特征跟踪方法流程, 可以看出每个任务 (Task) 负责一条特征跟踪链的学习, 经过多任务的结构化学习和特征映射, 从而获得鲁棒的特征匹配 (如图 7 的右下角).

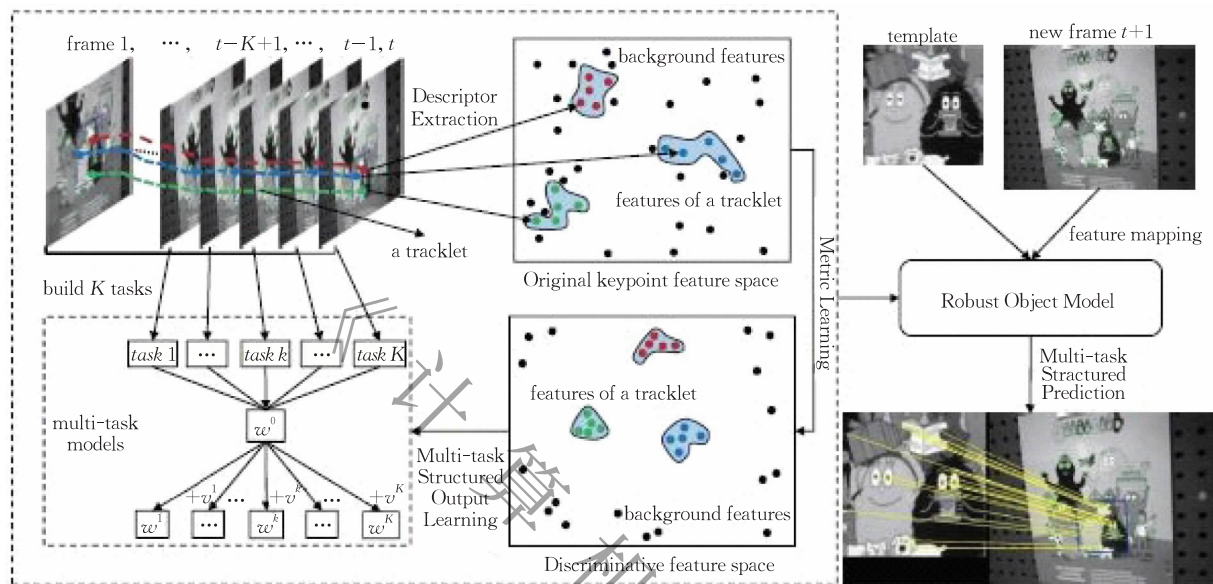


图 7 基于度量学习的特征跟踪方法流程^[34]

通过以上分析可知, KLT 方法与 DMF 类方法在设计原理上存在较大差异, 为了更好地理解两类方法的不同, 我们对两类方法的各自特点进行总结, 如下: (1) 对于 KLT 类方法不需要特征检测子与描述子, 采用光流进行特征跟踪, 适用于处理视频序列和分辨率相同的图像序列; 由于计算光流的时间复杂度较高, 因此, KLT 算法的时间效率较低; (2) 对于 DMF 类方法需要特征检测子、特征描述子和特征匹配方法 (如强力匹配和 KNN 等), 此类方法既适用于处理视频场景又适用于处理图像序列, 且对图像的分辨率没有要求, 因此, DMF 方法的适用性更强, 也是目前研究特征跟踪方法的主流思想. 然而, DMF 类方法也存在不足: (1) 时间效率取决于特征检测子和特征匹配所消耗的时间代价, 如 Bundler 系统使用了具有较高时间代价的 SIFT 特征检测子, 从而导致 Bundler 系统在特征跟踪阶段需要消耗大量的时间; (2) 算法的精度在很大程度上取决于特征描述子的表达性, 例如使用二进制描述子的特征跟踪方法的精度通常低于基于统计直方图的特征描述子.

2.3 可用算法与数据集

随着特征检测子、描述子以及特征匹配等技术的发展, 导致每个 SFM 系统所采用的 DMF 类方法有所不同, 因此, 对于相同的输入数据每个算法产生的结果也相差甚远. 为了便于分析问题, 对现有 SFM 系统进行调研, 总结现有 SFM 系统中常用的特征检测算法、描述子生成算法、特征匹配算法和测试数据集.

2.3.1 特征检测算法

在 SFM 系统中常用的特征检测算法如表 1 所示 (×表示不具有某种特性, √表示具有某种特性, √数量越多表示某种特性越好), 从早期的 Harris 特征到经典的 SIFT 特征, 甚至最近较为知名的 ORB 特征. 根据 SFM 系统的结果^[3, 27, 28, 83-85], 表明处理网络大规模数据集时, 具有多尺度和主方向的特征点产生的点云模型在精度方面高于其他特征检测算法结果. 文献^[85]在目前最大的 YFCC 100M 数据上测试, 采用 SIFT 算法检测特征点, 实验结果表明由于 SIFT 具有主方向和多尺度特性, 能够有效处理光照、尺度和图像扭曲等因素的影响, 使得

SFM 系统能够产生精确的点云模型. 但是, 由于 SIFT 算法的计算量较大, 导致 SFM 系统的效率较为低下, 文献[85]使用 43 台计算机组成的分布式集群花费 6 天时间重建出 YFCC 100M 数据集的点云模型. 最近 ORB^[86] 和 AGAST^[87] 算法在计算效率上表现出较强优势, 但是不具有多尺度特性. 每种算法具有各自的优点和不足, 因此, 可以根据具体问题选择合适的特征检测算法.

表 1 常用的特征检测算法

名称	类型		效率	重复性	文献
	主方向	多尺度			
Harris	×	×	✓✓	✓	[88]
MSER	✓	✓	✓✓	✓	[89]
SIFT	✓✓	✓✓	✓	✓✓✓	[61]
SURF	✓	✓	✓	✓✓	[90]
FAST	×	×	✓✓✓	✓✓	[91]
AGAST	×	×	✓✓✓	✓✓✓	[87]
FRIF	✓	✓	✓✓	✓✓	[92]
BRISK	✓	✓	✓✓	✓✓	[93]
ORB	✓	×	✓✓✓	✓✓	[86]
KAZE	✓✓	✓✓	✓✓	✓✓✓	[94]
DART	✓	×	✓✓	✓✓	[95]

2.3.2 特征描述子

特征描述子是对特征点的表示, 也是判断特征点相似性的依据. 表 2 给出了现有 SFM 系统常用的特征描述子(×表示不具有某种特性, ✓表示具有某种特性, ✓数量越多表示某种特性越好), 从早期的基于统计直方图的数值型描述子, 如 SIFT、SURF, 到目前较为受关注的二进制类型的描述子, 如 BRIEF^[96]、LDB^[101] 等. 每种类型的特征描述子都具有各自的优势与不足, 数值型的特征描述子精度较高, 计算效率较低; 二进制类型的特征描述子计算效率较高, 但是精度低于数值性描述子. 在具体的应用中, 我们可以根据实际问题选择合适的描述子生成

表 2 常用的特征描述子算法

名称	类型			效率	区分性	文献
	数值	二进制	描述子长度/bit			
SIFT	✓	×	128	✓	✓✓✓	[61]
SURF	✓	×	64	✓	✓✓	[90]
BRIEF	×	✓	128,256,512	✓✓✓	✓	[96]
BRISK	×	✓	512	✓✓✓	✓	[93]
ORB	×	✓	256	✓✓✓	✓✓	[86]
KAZE	✓	×	64	✓✓	✓✓✓	[94]
FRIT	×	✓	512	✓✓	✓✓	[92]
LIOP	✓	×	144	✓✓	✓✓✓	[97]
MROGH	✓	×	256	✓✓	✓✓✓	[98]
FREAK	×	✓	512	✓✓	✓✓	[99]
LATCH	×	✓	256	✓✓	✓✓✓	[100]
LDB	×	✓	256	✓✓✓	✓✓✓	[101]
DART	✓	×	8	✓✓	✓✓	[95]

算法, 例如, 当输入的数据量较大时, 可以选择二进制描述子, 如 LDB、ORB 和 BRIEF 等; 当输入的数据量较小且对算法精度要求较高时, 可以选择数值型描述子, 如 SIFT、SURF 和 LIOP 等, 以便产生更精确的点云模型.

2.3.3 特征匹配算法

特征匹配的目的是寻找相似的特征点, 表 3 给出了现有 SFM 系统常用的特征匹配方法(×表示不具有某种特性, ✓表示具有某种特性). 在早期的 SFM 系统中, 大多采用强力匹配算法对两个特征集合进行逐个比较, 取相似度最高的作为最终匹配结果, 此方法的最大不足是时间效率较低. 对于分别包含 N 和 M 个特征点的两个特征集合进行匹配, 时间复杂度为 $O(N \times M)$. 为了提高匹配效率和精度, 研究者们提出了 KNN 匹配算法, 选取相似度最高的 K 个匹配点作为候选结果, 然后根据候选特征点之间的距离关系决定最终匹配目标. 后来, 为满足大规模特征匹配需求, Garcia 等人^[68] 在 GPU 上实现了 KNN 算法, 利用并行技术提高特征匹配效率. 随着高维二进制描述子的出现(如 BRIEF、BRISK 和 FRIF 为 512 维), Muja 等人^[70] 提出一种适用于高维特征的可伸缩最近邻(Scalable Nearest Neighbors, SNN)算法, 此算法采用聚类思想为特征集合建立一棵层次化的聚类树, 在树形结构下查找候选特征点, 从而节省大量时间. 最近, Cheng 等人^[44] 提出一

表 3 常用的特征匹配算法

名称	特点		文献
	效率	准确性	
BF	✓	✓	[102]
KNN	✓✓	✓✓	[69]
gKNN	✓✓✓	✓✓	[68]
SNN	✓✓	✓✓	[70]
ANN	✓✓	✓✓	[103]
LSH	✓✓	✓✓	[104]
MIH	✓✓	✓✓	[105]
CasHashing	✓✓✓	✓✓✓	[44]
HKM	✓✓	✓✓	[106]
OKD-trees	✓✓	✓✓✓	[107]
JII	✓✓	✓✓	[108]
FGM	✓✓	✓✓	[109]
FeatMat	✓✓	✓✓✓	[110]
RPM	✓	✓✓	[111,112]
APM	✓✓	✓✓✓	[113]
KGT	✓✓	✓✓✓	[114]
Hungarian-BP	✓	✓✓	[115]
HLSS	✓✓	✓✓	[116]
Nv-Tree	✓✓	✓✓	[117]
PSM	✓	✓✓	[118]
FANNG	✓✓✓	✓✓✓	[119]
PQ-Tree	✓✓✓	✓✓✓	[120]

种基于局部感知哈希 (Locality Sensitive Hashing, LSH) 的层次哈希 (Cascade Hashing, CH) 算法, CH 算法能够快速匹配数值型和二进制类型的描述子, 实验结果表明 CH 是目前效率最高 (没有使用 GPU 加速) 的特征匹配算法.

2.3.4 特征跟踪数据集

数据是验证特征跟踪算法有效性的重要依据, 一些研究者创建了大量的测试数据集, 用于验证特征跟踪算法的性能. 表 4 和表 5 总结出一些较为常用的经典数据集 (×表示不具有某种特性, ✓表示具有某种特性), 这些数据集在作者的主页上均可下载

获得, 数据集的详细信息和详细链接地址可参见对应文献. 需要说明的是: Castle^[121] 和 Palace^[122] 是常用的小样本数据集; PoseNet^[123] 是 CVPR2015 发布的较大数据集, 采用车载相机拍摄的街道场景. 1DSFM^[27] 是 ECCV2014 发布的测试 SFM 系统的无序图像集合, 包含 20 种测试集合, 每类集合包含大量的分辨率不一致的图像序列. Disambig dataset^[124]、Duplicate dataset^[125] 和 ArtsQuad^[28] 数据集是从网络搜索而得, 包含大量的重复特征 (Repeating Feature), 对描述子的区分性要求较高, 是目前测试特征跟踪方法鲁棒性的常用数据. KITTI suite^[126]

表 4 可用的数据集

名称	类型		分辨率	焦距		来源	数量/体积	文献
	视频	图像序列		恒定	变化			
Castle	×	✓	2832×2128	✓	×	拍摄	11 幅	[121]
Palace	×	✓	1200×800	✓	×	拍摄	68 幅	[122]
Tsinghua	×	✓	4368×2912	×	✓	拍摄	193 幅	[44]
Shengkelou	×	✓	4368×2912	✓	×	拍摄	102 幅	[44]
PoseNet dataset	✓	×	2048×1024	✓	×	拍摄	6 类序列	[123]
1DSFM dataset	×	✓	大小不一	×	✓	网络	20 类序列	[27]
NotreDame	×	✓	大小不一	×	✓	网络	8.0 GB	[124]
Louvre	×	✓	大小不一	×	✓	网络	3.5 GB	[124]
SacreCoeur	×	✓	大小不一	×	✓	网络	3.0 GB	[124]
Seville		✓	大小不一	×	✓	网络	1.5 GB	[124]
Acts dataset	×	✓	680×480	✓	×	拍摄	6 类序列	[22]
Plant	×	✓	960×540	×	✓	拍摄	583 幅	[23]
Gangwan	×	✓	960×540	×	✓	拍摄	2 类序列	[23]
Garden	✓	×	320×240	×	✓	拍摄	N/A	[22]
Building	✓	×	320×240	×	×	拍摄	N/A	[22]
Strecha dataset	×	✓	大小不一	×	✓	拍摄	6 类序列	[127]
barcelona	×	✓	2272×1704	✓	×	拍摄	191 幅	[128]
CAB	×	✓	1696×1132	✓	×	拍摄	312 幅	[128]
capitole	×	✓	3292×2264	✓	×	拍摄	99	[128]

表 5 可用的数据集

名称	类型		分辨率	焦距		来源	数量/体积	文献
	视频	图像序列		恒定	变化			
redmond	×	✓	3968×2232	✓	×	拍摄	261	[128]
Urban Facades	×	✓	2880×2160	×	✓	拍摄	9 类序列	[129]
Cathedral	×	✓	大小不一	×	✓	网络	448	[125]
Big Ben	×	✓	大小不一	×	✓	网络	390	[125]
Cereal	×	✓	2736×3648	×	✓	拍摄	25	[125]
Church	×	✓	大小不一	×	✓	网络	277	[125]
Street	×	✓	1536×2048	×	✓	拍摄	19	[125]
SfMwO	×	✓	620×480	✓	×	拍摄	258	[130]
Ford Car	×	✓	781×601	✓	×	拍摄	344	[131]
Zurich	×	✓	640×480	✓	×	拍摄	1005	[132]
New College	×	✓	384×512	✓	×	拍摄	25 GB	[133]
Rome16K	×	✓	大小不一	×	✓	网络	7.0 GB	[134]
Dubrovnik6K	×	✓	大小不一	×	✓	网络	3.5 GB	[134]
SUN3D	×	✓	640×480	✓	×	拍摄	100 GB	[84]
ArtsQuad	×	✓	大小不一	×	✓	网络	9 GB	[28]
KITTI suite	×	✓	1224×370	×	✓	拍摄	428 GB	[126]
YFCC 100M	×	✓	大小不一	×	✓	网络	10 TB	[135]

是采用车载系统拍摄的城市场景数据集,用于测试 SFM 系统在大规模室外环境下的性能. YFCC 100M^[135]是由 Yahoo! 公司发布的大规模数据集,这些数据来自于世界各地的摄影师分享在 Flickr 上的图像集合,也是目前最大最全的数据集,据统计约有 10 TB(其中一些序列只提供链接地址,需要自己下载).

3 实验评估

为了更好地了解现有特征跟踪方法的精度和时间性能,在 Tsinghua、Notre Dame、Rushmore^[136]、Yosemite^[136]和 Church^[125]数据集上对 KLT、Bundler、ETH-3D 和 VSFM 的特征跟踪方法进行测试.

表 6 特征跟踪算法详细信息

算法	检测子	描述子	匹配方法	源代码地址
KLT	无	无	光流	http://www.ces.clemson.edu/~stb/klt
Bundler	SIFT	SIFT	强力匹配	http://www.cs.cornell.edu/~snaveily/bundler/
ETH-3D	SURF	SURF	KNN	http://cvg.ethz.ch/research/chzach/opensource.html
VSFM	SIFTGPU	SIFTGPU	强力匹配	http://ccwu.me/vsfm/

3.1 特征跟踪方法测试

在不同数据集上特征跟踪方法的可视化结果如图 8~图 12 所示. 为了便于观察匹配结果,在实验结果中仅显示了采用单应性矩阵(Homography Matrix)验证后的特征.

3.1.1 定性分析

(1) 在 Notre Dame 数据集上测试分析

图 8 展示了在 Notre Dame 数据集上每个特征跟踪方法的结果. 可以看出,KLT 方法跟踪到的特征点数较少,且存在较多错误,原因在于 KLT 方法采用运动光流跟踪特征点,对摄像机的垂直波动较为敏感. 由图 8 可知,由于相机的垂直波动导致 Notre Dame 数据集中每幅图像中的建筑物的顶部不在同一水平线上,从而造成 KLT 方法不能有效地跟踪特征点(此图为黑白印刷). 虽然 Bundler 方法能够跟踪到较多的特征点,但是在第一组结果中,特征点的匹配主要集中在建筑物的上部. 经过分析发现导致这问题的主要原因是:相机运动导致图像的尺度变化超越了 SIFT 的尺度空间所能容忍的范围. ETH-3D 和 VSFM 均能够跟踪到较多的特征点,且正确率较高;相对而言,VSFM 方法的结果更精确一些,且特征点的分布很均匀.

(2) 在 Tsinghua 数据集上测试分析

图 9 展示了在 Tsinghua 序列上每个特征跟踪方法的结果(此图为黑白印刷),可以看出 KLT 方

Tsinghua 序列是一组具有对称结构的经典数据; Notre Dame 具有对称结构、少量重复特征和光照变化;Rushmore 和 Yosemite 是测试特征跟踪方法的 benchmark 数据,具有大量重复特征和光照变化; Church 是一组具有光照变化、尺度变化和重复特征的标准数据集. KLT 是特征跟踪对比的基准算法; Bundler 和 ETH-3D 是具有较高知名度的开源 SFM 系统;VSFM 是带有用户界面的著名 SFM 系统;表 6 给出了这些特征跟踪方法所使用的特征检测子、特征描述子、特征匹配方法和源代码地址信息. 为了客观地评价每一个算法,在实验过程中采用算法默认参数,所有实验结果均在配有 Intel i5 CPU 和 8.0 GB RAM 的 PC 机上运行获得.

法的特征匹配结果较为清晰,没有交叉匹配现象;但是 KLT 方法检测的特征点数较少且分布在建筑物的下部,特征跟踪丢失较为严重. Bundler 算法检测出较多的特征点,且特征点的分布较为均匀,但是从图中可以看出 Bundler 方法存在较多的交叉匹配. ETH-3D 方法检测出的特征点数多于 Bundler 方法且交叉匹配少于 Bundler 方法. 比较而言,VSFM 方法能够检测出较多的特征点,且错误匹配少于 Bundler 和 ETH-3D 方法. 经过分析发现,Bundler 方法存在较多错误匹配的原因是:当物体颜色与背景颜色相似时,SIFT 描述子的区分性较弱,从而导致 Bundler 方法出现较多的错误匹配. 因此,根据在 Tsinghua 序列上的测试结果我们可以得出如下结论:当目标物体与场景颜色相似时,基于统计直方图的特征描述子的区分性较弱.

(3) 在 Rushmore 数据集上测试分析

图 10 展示了在 Rushmore 数据集上每个特征跟踪方法的测试结果(此图为黑白印刷). 由于摄像机的垂直波动导致 Rushmore 中每幅图像上的物体不在同一水平线上,造成 KLT 方法跟踪到较少的特征点. Bundler 方法跟踪到的特征点数多于 KLT 方法,但是,由于光照变化较大导致 Bundler 方法出现较多的错误匹配. ETH-3D 采用 SURF 特征,它是 SIFT 特征的改进,比 SIFT 具有更强的光照不变性,因此 ETH-3D 方法能够跟踪到较多的正确的

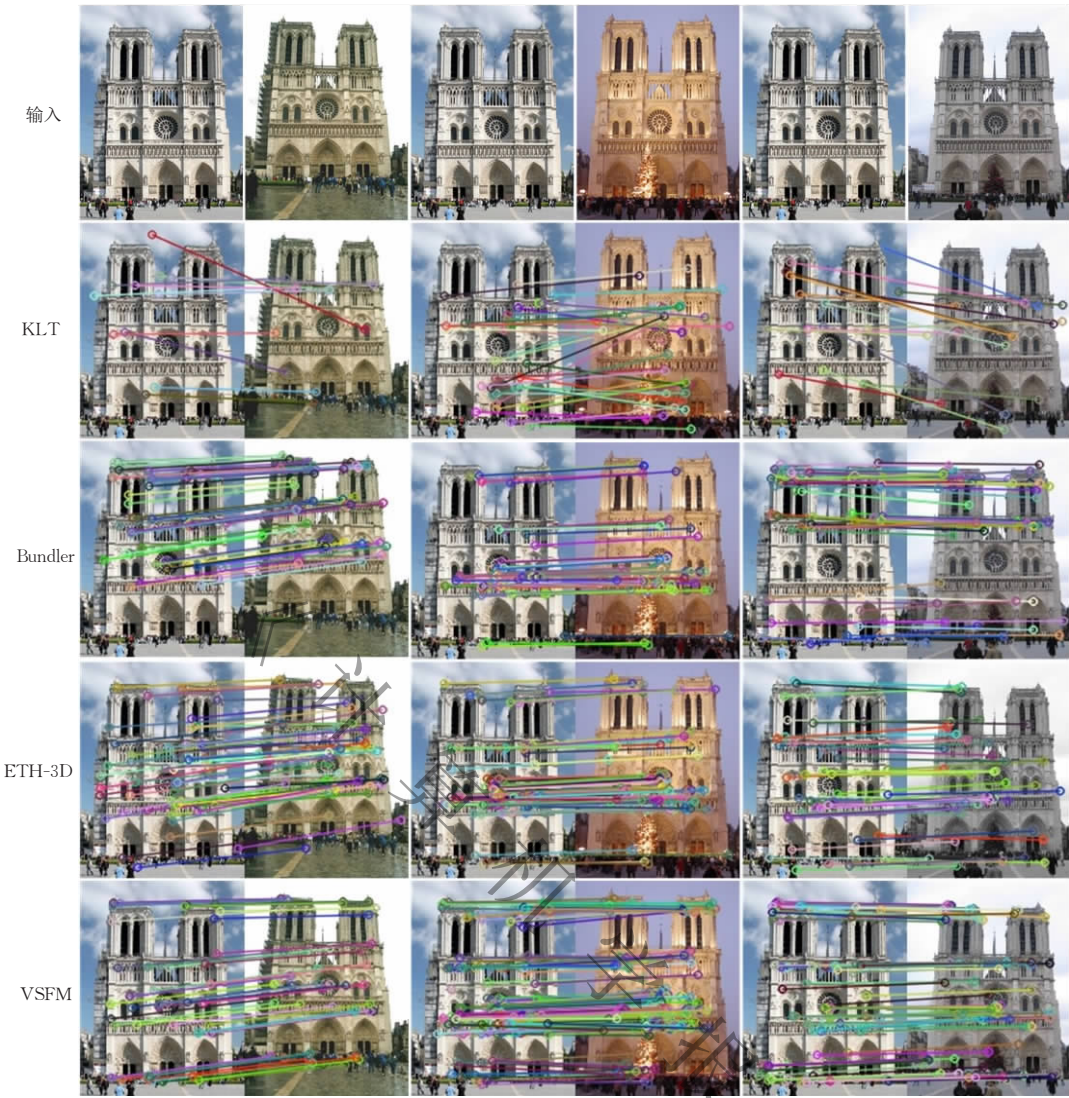


图 8 在 Notre Dame 数据集上测试结果



图 9 在 Tsinghua 数据集上测试结果

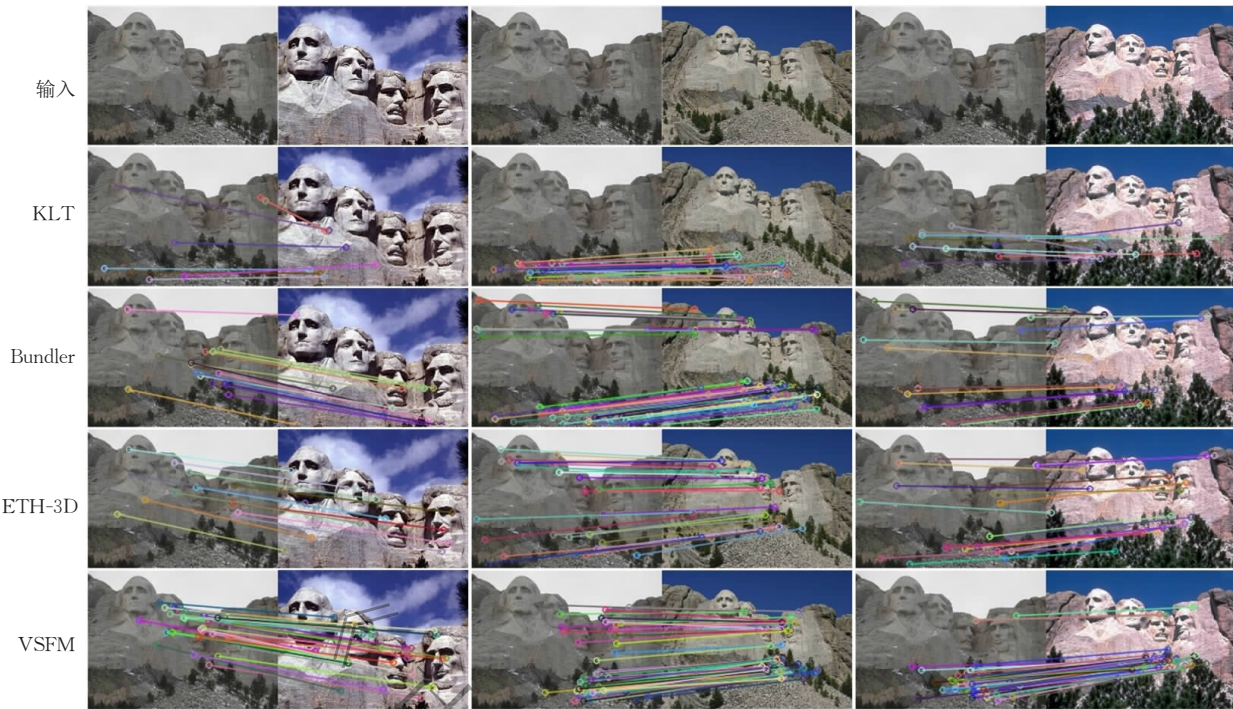


图 10 在 Rushmore 数据集上测试结果

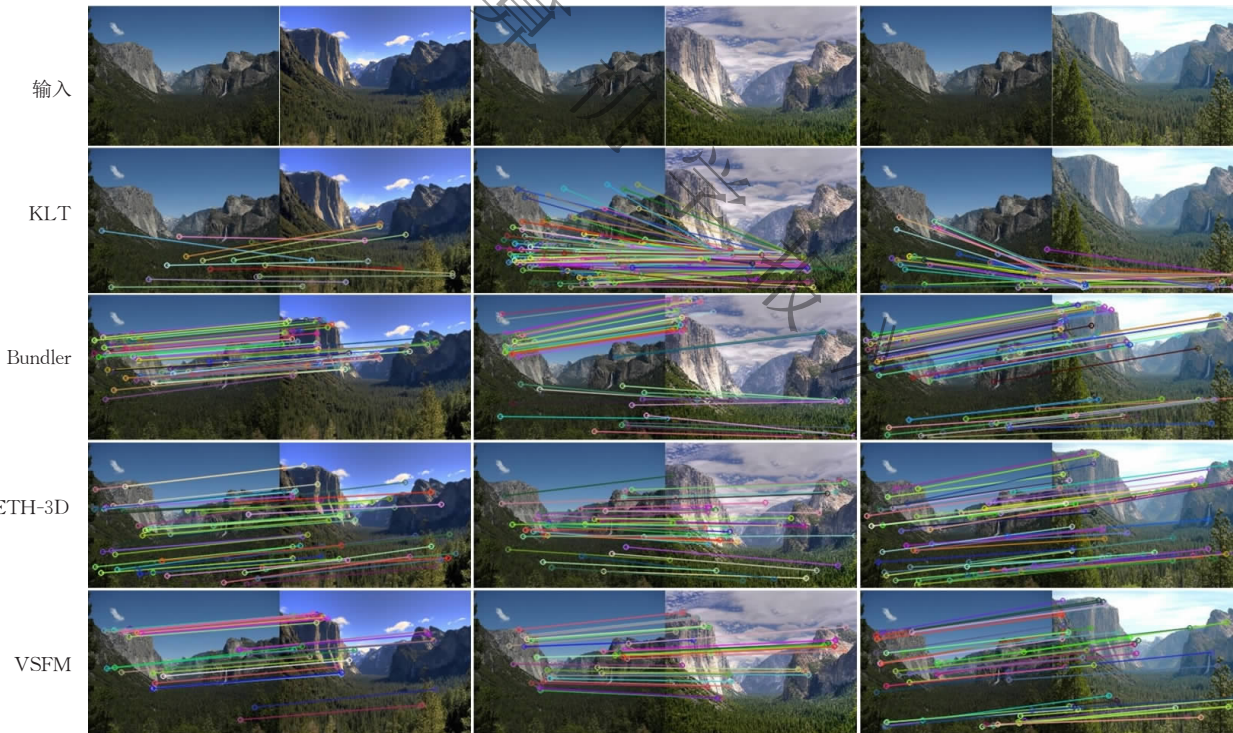


图 11 在 Yosemite 数据集上测试结果

特征点. VSFM 方法采用的 SIFTGPU 特征是 SIFT 特征在图形处理器上的并行实现,受图形处理器显示分辨率的限制,需要对图像进行降采样,导致 VSFM 方法跟踪到的特征点出现聚簇现象,但是正确率仍然高于 ETH-3D 方法. 因此,根据在 Rushmore 数据集上的测试结果可知,图像的分辨率能够影响

SIFT 特征检测子的质量.

(4) 在 Yosemite 数据集上测试分析

图 11 展示了特征跟踪方法在 Yosemite 数据集上的测试结果(此图为黑白印刷). 此数据集中每幅图像之间垂直波动较小,因此, KLT 能够跟踪到较多的特征点. 但是,跟踪结果存在较多的匹配错误,



图 12 在 Church 数据集上测试结果

且特征点分布较为聚簇. Bundler 方法跟踪到的特征点数多于 KLT 方法,特征点分布较为均匀.在第二组实验中,由于左侧山峰的颜色与右侧天空类似,因此 Bundler 方法错误地将左侧山峰上的特征点匹配到右侧天空上的特征点. ETH-3D 方法跟踪到的特征点数最多,且特征点分布很均匀,错误匹配较少.受 VSFM 方法中 SIFTGPU 对图像降采样影响,导致 VSFM 方法跟踪到的特征点数少于 ETH-3D 方法.此外,在第一组和第二组测试结果中,VSFM 方法跟踪到的特征点出现聚簇.

(5) 在 Church 数据集上测试分析

图 12 展示了在 Church 数据集上每个特征跟踪方法的测试结果(此图为黑白印刷). Church 数据集

中的每幅图像含有大量的重复结构,是测试特征跟踪方法处理重复特征能力的最具挑战性的测试数据.从测试结果可以看出,KLT 方法存在大量的错误匹配,在第三组测试结果中 KLT 只跟踪到很少的特征点,且全为错误匹配. Bundler 和 VSFM 能够跟踪到较多的特征点,但是存在较多错误匹配,经过分析发现建筑物表面的重复特征是导致错误匹配的主要因素. ETH-3D 方法跟踪到的特征点数少于 Bundler 和 VSFM 方法,原因在于 ETH-3D 方法使用的 SURF 特征采用滤波器替代 SIFT 的高斯差分空间,在重复特征较多的环境下,降低了描述子的区分性.此外,从测试结果可以看出,在重复特征较多的环境下,采用单应性矩阵难以去除错误匹配的特

征点,因此,如何提高特征描述子的区分性是特征跟踪领域迫切需要解决的问题.

3.1.2 定量分析

为了公正地评估每个特征跟踪方法的性能,我们统计出每个方法在不同数据集上的匹配精度(Precision).

$$Precision = \frac{N_c}{N} \tag{1}$$

其中, N 表示强力匹配数目; N_c 表示采用RANSAC^[137]矫正后的正确匹配数.

(1) 精度分析

图 13 展示了在 Tsinghua 数据集上每个特征跟踪方法的精度,可以看出 KLT 的精度最低,为 25.34%;ETH-3D 方法的精度最高,为 82.75%;VSFM 特征跟踪方法的精度为 78.73%;Bundler 方法的精度为 72.56%. 在 Notre Dame 数据集上每个特征跟踪方法的精度如图 14 所示,可以看出 KLT 方法的精度最低,为 41.26%;VSFM 方法的精度最高,为 82.75%;Bundler 和 ETH-3D 方法的精度分别为 72.35 和 79.83.

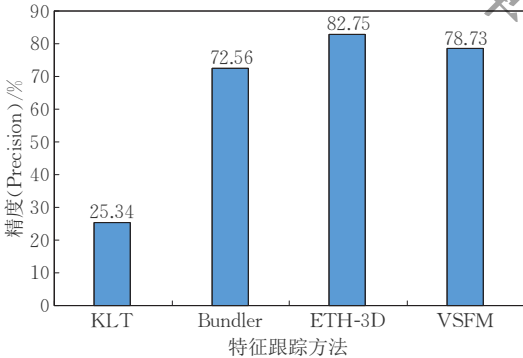


图 13 在 Tsinghua 数据集上每个特征跟踪方法的精度

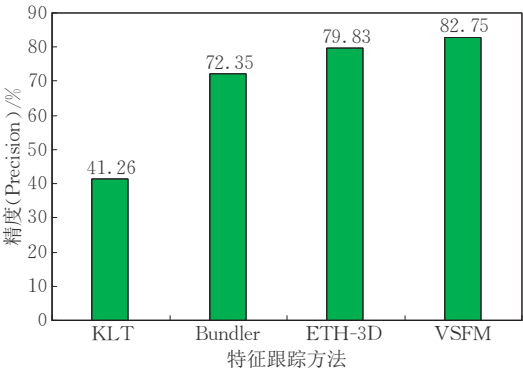


图 14 在 Notre Dame 数据集上每个特征跟踪方法的精度

图 15 展示每个特征跟踪方法在 Rushmore 数据集上精度,效果最好的是 VSFM 方法,精度为 75.64%;KLT 方法的精度最低,为 32.16%;Bundler 和 ETH-3D 方法的精度分别为 62.47 和 73.86. 在 Yosemite

数据集上每个特征跟踪方法的精度如图 16 所示,KLT 方法的精度最低,为 58.63%;精度最高的是 ETH-3D 方法,为 78.54%;VSFM 方法的精度比 Bundler 方法稍高一些,分别为 71.34%和 74.26%. 图 17 展示了在 Church 数据集上每个特征跟踪方法的精度,可以看出 Bundler 方法的精度最高为 51.53%,KLT 方法的精度最低,为 25.24%;ETH-3D 和 VSFM 方法的精度分别为 40.62%和 47.84%.

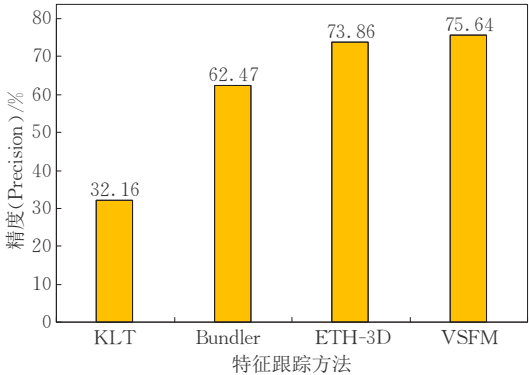


图 15 在 Rushmore 数据集上每个特征跟踪方法的精度

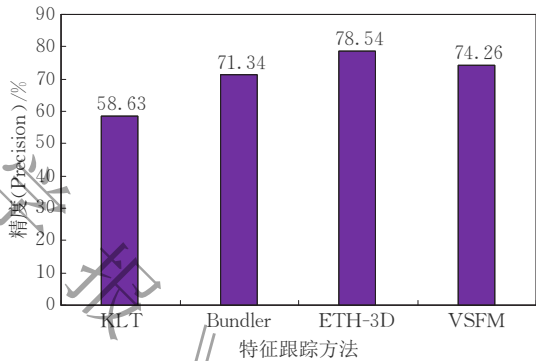


图 16 在 Yosemite 数据集上每个特征跟踪方法的精度

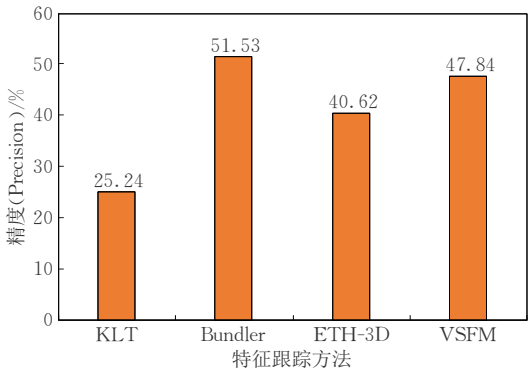


图 17 在 Church 数据集上每个特征跟踪方法的精度

通过以上分析,可以发现在所有数据集上 KLT 方法的精度最低,基于 DMF 框架的特征跟踪方法精度排名交替出现,主要原因在 DFM 类特征跟踪方法采用的局部特征容易受到图像质量的影响.具体地说,影响因素包括输入数据的分辨率、光照变化

和重复特征结构等. 因此, DMF 类特征跟踪方法的精度与图像的分辨率、光照变化、几何结构和重复特征有较大联系. 此外, 从实验结果中我们还可以发现, DMF 类特征跟踪方法的有效性还取决于特征检测子的质量, 例如在 Church 序列上, Bundler 算法使用的 SIFT 能够检测出较多的特征点, 这也进一步提高了特征匹配的精度.

(2) 时间效率分析

我们统计出每个特征跟踪方法分别在 Tsinghua、Notre Dame、Rushmore、Yosemite 和 Church 数据集上的时间效率(包括特征检测、计算描述子和匹配所需时间之和), 结果如图 18~图 22 所示. 图 18 展示了在 Tsinghua 数据集上, 每个特征跟踪方法的平均时间效率, VSFM 方法的效率最高, 处理两幅图像的特征检测、计算描述子和匹配所用平均时间之和为 18.57 s; Bundler 方法的时间效率最低, 平均处理时间为 70.65 s; ETH-3D 的平均处理时间为 52.46 s; KLT 效率排在第二位, 为 29.53 s. 在 Notre Dame 数据集上每个特征跟踪方法的时间效率如图 19 所示, 效率最低的是 Bundler 方法, 平均处理时间为 23.52 s; 效率最高的是 VSFM 方法, 平均处理时间为 10.85 s; ETH-3D 和 KLT 方法的平均处理时间分别为 20.36 s 和 15.34 s.

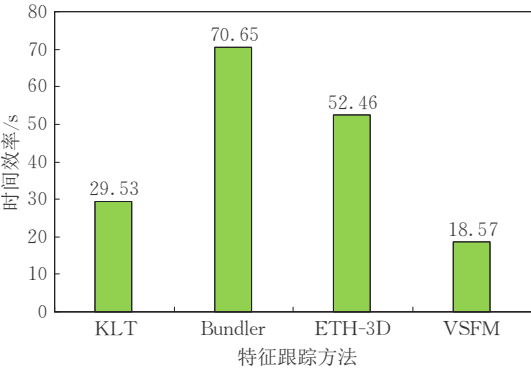


图 18 在 Tsinghua 数据集上每个特征跟踪方法的时间效率

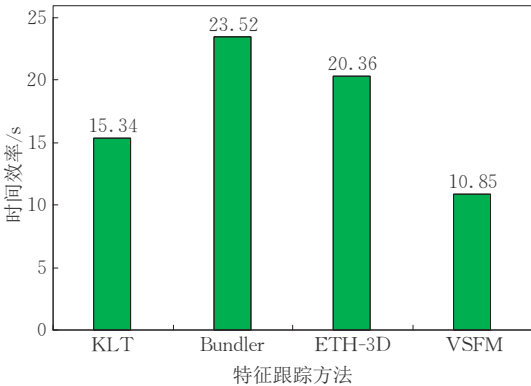


图 19 在 Notre Dame 数据集上每个特征跟踪方法的时间效率

图 20 展示了在 Rushmore 数据集上每个特征跟踪方法的时间效率, Bundler 方法的时间效率最低, 平均处理时间为 57.67 s; ETH-3D 方法的平均处理时间为 48.45 s; VSFM 方法的时间效率最高, 平均处理时间为 31.84 s. 图 21 展示了在 Yosemite 数据集上每个特征跟踪方法的时间效率, 可以看出效率最高的 VSFM 方法, 平均处理时间为 32.64 s; 效率最低的是 Bundler 方法, 平均处理时间为 63.47 s; ETH-3D 采用效率较高的 SURF 特征, 因此时间效率比 Bundler 高一些, 平均处理时间为 51.26 s.

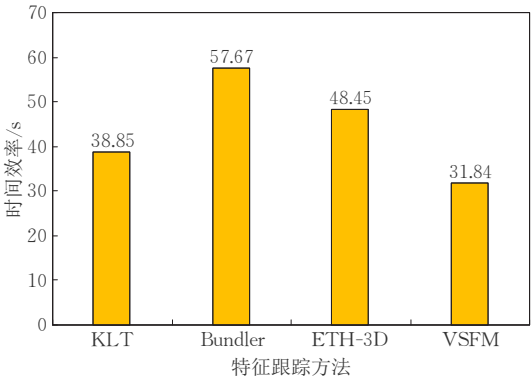


图 20 在 Rushmore 数据集上每个特征跟踪方法的时间效率

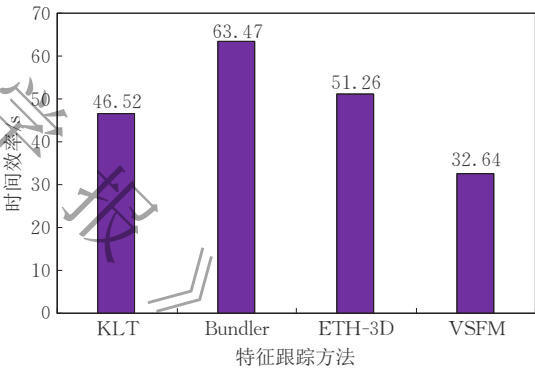


图 21 在 Yosemite 数据集上每个特征跟踪方法的时间效率

图 22 展示了每个特征跟踪方法在 Church 数据集上的时间效率, 可以看出效率最低的是 Bundler 方法, 平均处理时间为 78.84 s; 效率最高的是 VSFM 方法, 平均处理时间为 36.28 s; ETH-3D 和 KLT 方法的平均处理时间分别为 50.32 s 和 50.26 s. 通过以上分析可知, 在所有测试数据集上 Bundler 方法的时间效率最低, 因为此方法采用 SIFT 算法检测特征点和计算描述子, 从而导致特征跟踪方法的时间效率低下. 由于 VSFM 方法采用 SIFTGPU 算法检测特征点和计算描述子, 它是 SIFT 算法在图形处理器上的并行实现, 因此 VSFM 方法具有较高的实时性. ETH-3D 采用 SURF 算法检测特征点和计算描述子, SURF 是 SIFT 算法在时间效率上的改进版本, 因此 ETH-3D 方法的时间效率高于 Bundler

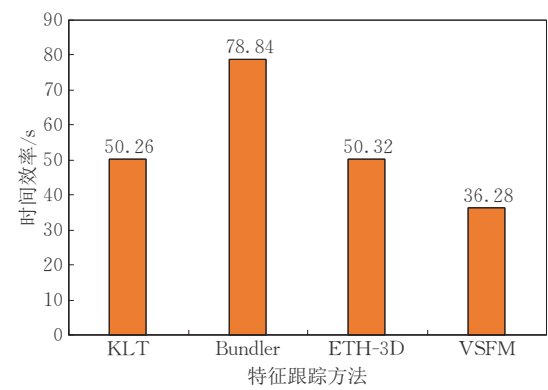


图 22 在 Church 数据集上每个特征跟踪方法的时间效率

方法. 根据实验得出结论: 特征跟踪方法的时间效率在很大程度上取决于特征检测子与特征描述子的时间效率.

3.2 SFM 重建结果

为了测试在 DMF 框架下, 不同特征跟踪方法的性能, 我们将 Bundler、ETH-3D 和 VSFM 的特征跟踪方法应用于类似与 Bundler 的增量式 SFM 系统中, 通过比较产生的点云模型的精度来评估每个特征跟踪方法的有效性. 在 Tsinghua^[44]、Hotel、Burgos、Castle^[121]、Kler 和 Lion 数据集上对上述特征跟踪方法进行测试, 结果如图 23 所示(此图为黑白印刷).

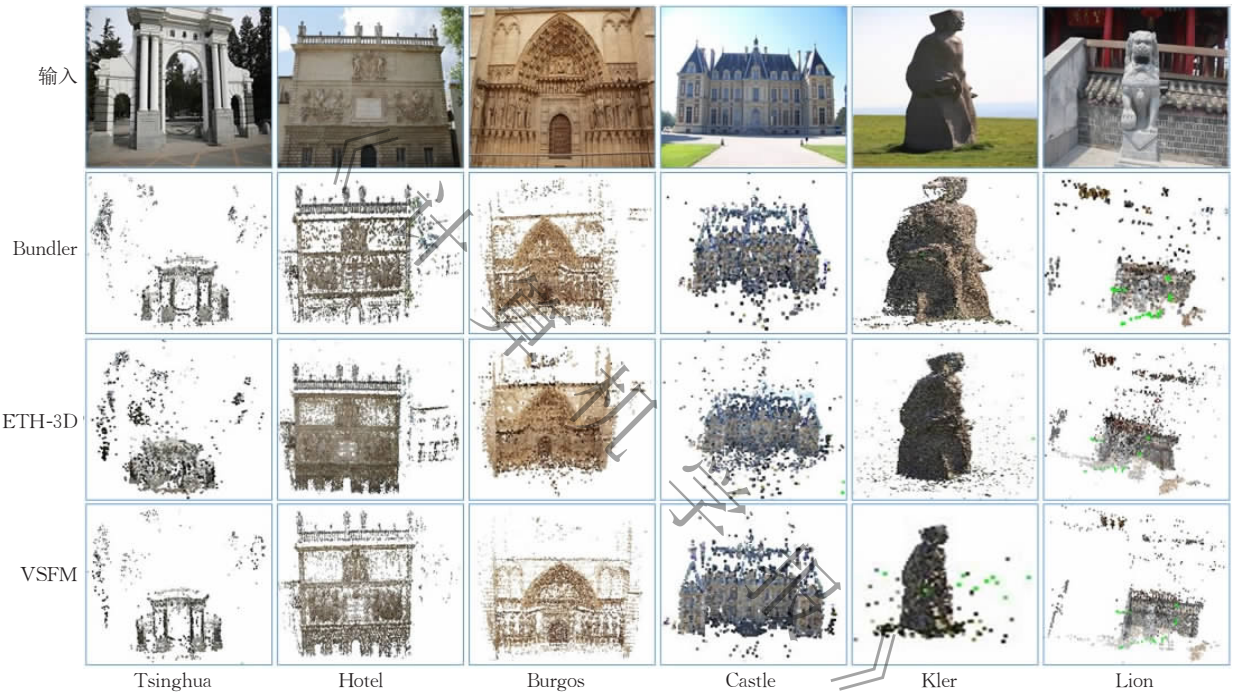


图 23 特征跟踪方法对 SFM 重建结果的影响

在 Tsinghua 数据集上, Bundler 和 ETH-3D 特征跟踪方出现较多的错误匹配, 导致重建出的点云模型存在大量的噪声(outliers); VSFM 方法产生的噪声较少, 且点云模型与真实场景具有较好的几何一致性. 在 Hotel 数据集上, ETH-3D 方法产生的建筑物表面的点云模型最精确, 但是由于建筑物的表面存在较多的重复结构, 导致在点云模型的右侧出现较多的歧义点云; Bundler 和 VSFM 方法产生的点云模型较为稀疏, 且在建筑物的表面存在噪声.

在 Burgos 数据集上, VSFM 方法生成的点云模型较为精确; Bundler 和 ETH-3D 方法产生的建筑物点云模型出现扭曲, 且存在大量噪声, 这也进一步证明了 Bundler 和 ETH-3D 方法容易受到图像

质量的影响. 对于 Castle 数据集, 每个方法产生的点云模型均存在空洞和噪声, 相比较而言, VSFM 方法产生的点云模型较为精确一些. 在 Kler 序列上, ETH-3D 方法产生的点云模型最完整; Bundler 方法产生的点云模型右上角存在残缺; VSFM 方法生成的点云模型存在大量噪声. 在 Lion 序列上, 每个方法生成的点云模型均存在大量的噪声, 相比较而言, VSFM 方法产生的噪声稍少一些; Bundler 方法产生的噪声最多, ETH-3D 方法产生的噪声少于 Bundler 方法. 从测试结果可知, 每个特征跟踪方法具有各自特点, 没有一种方法能够同时满足所有需求.

为了更精确地度量每个特征跟踪方法的性能, 我们统计出每个特征跟踪方法所产生的点云模型的

误差信息,其中,摄像机位置误差(Cerr)表示 SFM 系统估计的摄像机位置与 groundtruth 之间的平均误差,旋转角度误差(Rerr)表示 SFM 系统估计出的摄像机方向与 groundtruth 之间的平均误差,结果如表 7 所示. 对于三维点云误差,我们统计出每个

三维点云模型与 groundtruth 之间的误差之和,结果如图 24 所示,可以看出每个算法的性能随着输入数据的变化而改变,例如在 Tsinghua 数据集上 VSFM 产生的点云模型的误差大于 ETH-3D,然而在 Hotel 数据集上,VSFM 的精度高于 ETH-3D.

表 7 摄像机位置(像素)与方向误差(角度)

方法	数据集											
	Tsinghua		Hotel		Burgos		Castle		Kler		Lion	
	Cerr	Rerr	Cerr	Rerr	Cerr	Rerr	Cerr	Rerr	Cerr	Rerr	Cerr	Rerr
Bundler	0.35	0.34	0.41	0.28	0.33	0.26	0.23	0.31	0.45	0.39	0.54	0.48
ETH-3D	0.26	0.31	0.38	0.34	0.25	0.24	0.25	0.35	0.42	0.36	0.56	0.45
VSFM	0.28	0.26	0.32	0.31	0.31	0.30	0.28	0.31	0.46	0.40	0.53	0.42

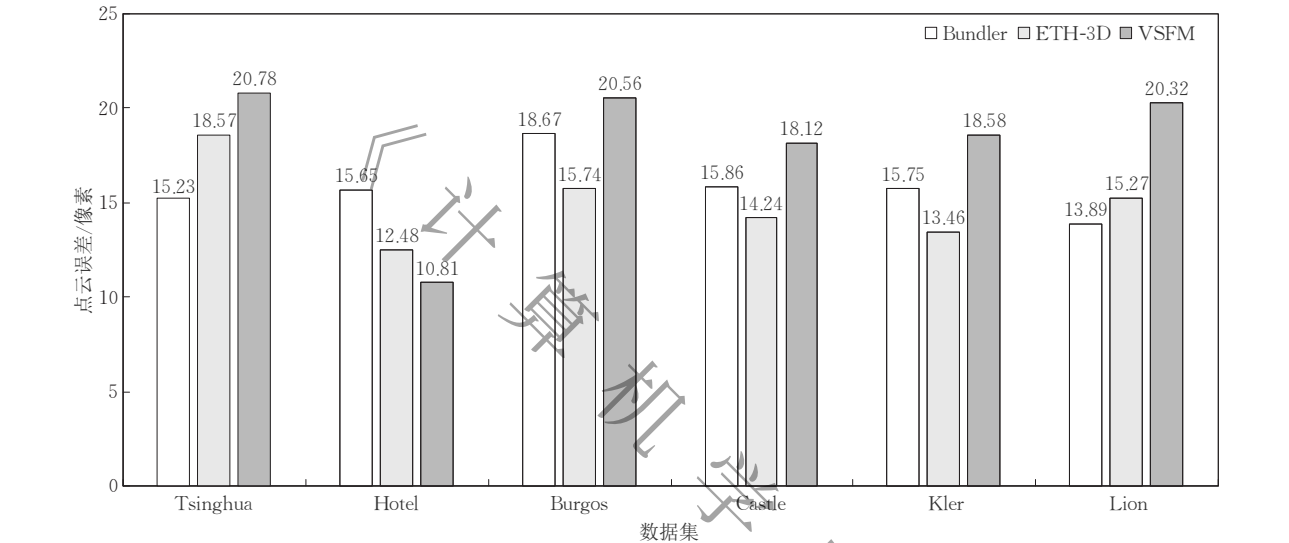


图 24 三维点云位置误差

4 待解决问题

众所周知,SFM 研究工作是一项具有挑战性的任务,特征跟踪作为 SFM 的重要组成部分,仍然存在一些亟待解决的问题,尤其当一些网络大数据^[123,133,135]的发布使得特征跟踪方法存在的问题表现得更加突出. 本文对现有 SFM 系统使用的特征跟踪方法进行总结和实验测试,发现迫切需要解决的问题主要反映在以下五个方面:(1) 时间效率;(2) 特征匹配精度;(3) 特征消歧;(4) 结构化特征选取;(5) 非线性相机影响. 为了便于进一步提升特征跟踪方法的性能,下面我们将对特征跟踪方法存在的问题进行详细分析.

4.1 特征跟踪效率

高效地生成三维模型是每个三维重建系统追求的目标之一,特征跟踪是影响 SFM 系统效率的主要因素之一,在 DMF 框架下影响特征跟踪方法效

率的因素主要体现在以下方面:

- (1)特征检测算法. 高效的特征检测算法能够在很大程度上提高特征跟踪方法的效率,尤其在一些分辨率较大的图像序列上特征检测算法的效率显得更为重要. 例如文献[85]在 YFCC 100M 数据集上采用 SIFT 算法检测特征点,其特征检测部分所花费时间约占整个三维重建过程的 30%. 因此,为了提高时间效率,应该采用高效的特征检测算法,经典的高效特征检测算法有 FAST、AGAST 和 ORB,但是,这些特征存在各自不足,需要进一步改进才能在特征跟踪和 SFM 系统中体现出优势.
- (2)特征描述子. 通常计算特征描述子所花费时间大于特征检测时间,例如 SIFT 特征描述子的 128 向量所需计算时间将会更长. 为了提高时间效率,可以采用低维的 PCA-SIFT 和 SURF 特征或者二进制描述子如 BRIEF、LDB 和 BRISK 等.
- (3)特征匹配方法. 高效的特征匹配方法能够提高特征跟踪的时间效率,如 Bundler 的特征跟

踪方法采用强力匹配方法对特征点进行匹配,时间复杂度为 $O(MN^2)$,其中 M 表示图像数, N 表示特征点数. 可以改用效率较高的特征匹配方法,如 FLANN^[71]、Cascade Hashing^[44] 和 FANNG^[119] 等.

(4) 可以充分利用 GPU 等硬件资源对现有算法进行并行化,以便提高特征跟踪方法的时间效率. 例如 VSFM 系统采用的 SIFTGPU 算法是 SIFT 算法的 GPU 实现,效率提高约 10 倍. 但是, SIFTGPU 在算法实现上存在一定问题,对于同样的输入图像, SIFTGPU 检测出的特征点数远少于原始 SIFT 算法,容易导致特征跟踪丢失,最终造成 SFM 系统重建出的点云模型存在大量空洞.

4.2 特征跟踪的精度

特征跟踪算法的精确性直接影响 SFM 系统生成的三维模型质量. 在 DMF 特征跟踪框架下,影响特征跟踪精度的因素主要体现在以下几个方面:

(1) 特征检测子的重复性(Repeatability). 特征检测子的重复性反映的是图像在经过一些变换后,特征点数的变化,当图像经过某种变化(仿射变换、光照变化等)后同样能够检测出较多的特征点,说明特征检测子的重复性较高. 在特征跟踪过程中,特征点数目越多,特征跟踪方法保留下的特征点数相对较多. 因此,重复性高的特征检测算法能够提高特征跟踪的精度. 重复性较高的经典特征检测算法有 SIFT、FAST、AGAST 和 KAZE,其中, KAZE 和 SIFT 特征的计算效率较低.

(2) 描述子的区分性(Discriminative)或表达性(Representative). 特征描述子之间的相似性通常采用 Hamming 和 L2 距离度量,对于区分性较弱的描述子容易造成错误匹配,从而导致特征跟踪错误,关于此问题文献[101]做出了详细分析. SIFT 和 SURF 这样的基于统计直方图的特征描述子,通常具有较高的区分性. 以 BRIEF、ORB 和 BRISK 为代表的二进制描述子的区分性稍低一些,如图 25 所示,对于不同的图像区域, BRIEF 生成了相同的描述子.

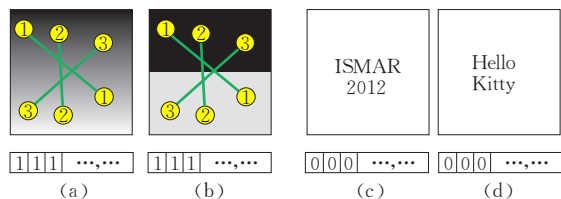


图 25 BRIEF 描述子表达性低的示例^[101]

(3) 外点(outlier)去除算法. 采用描述子匹配方法创建的特征跟踪链通常包含一些错误匹配,为了提高匹配精度,可以采用特定算法进行优化. 例如,

RANSAC^[138]算法是去除特征匹配错误最常用的方法之一,但是此算法较为耗时. 文献[23]采用两次匹配策略去除错误匹配,同时解决特征的跟踪丢失问题,然而两次匹配同样降低了特征跟踪的时间效率. 因此,寻找 RANSAC 以外的有效外点去除算法还需进一步探索.

4.3 特征消歧

特征消歧(Disambiguation)是特征跟踪中最具挑战性的任务,此问题产生于输入图像序列上存在相同的目标对象,去除重复特征造成的错误匹配的过程称之为特征消歧. 如图 26 所示(此图为黑白印刷), A 和 C 是正确的匹配结果, B 是错误匹配;由于输入序列上存在相同的建筑,无论采用何种描述子算法在不同位置都能产生相同的描述子,即存在大量的重复特征(Repeating Feature),从而造成错误匹配.

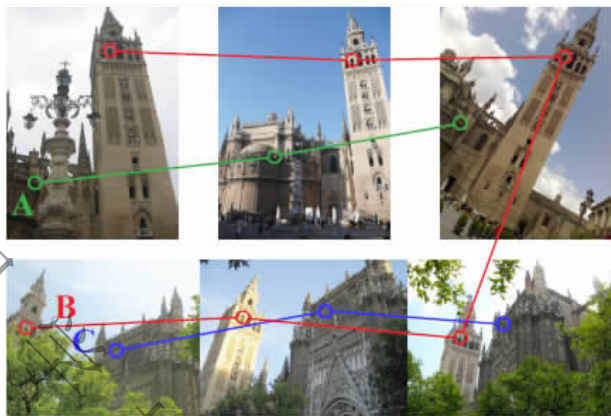
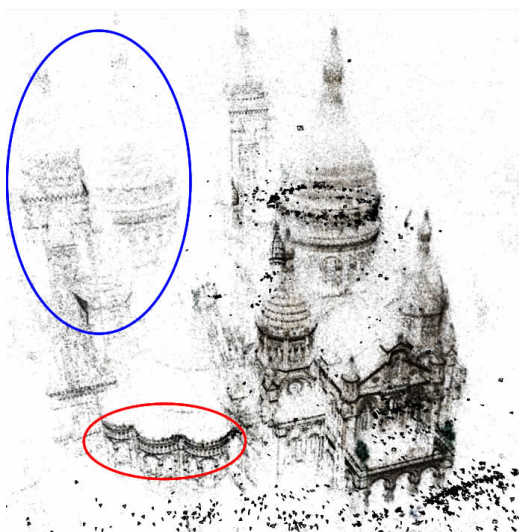


图 26 特征的匹配的歧义性^[139]

图 27 展示了重复特征对重建点云模型的影响,其中图 27(a)是建筑物的参考模型,图 27(b)是 SFM 系统生成的点云模型,红色和蓝色圆形所标注的是受重复特征影响导致 SFM 系统生成虚假的点云模型. 目前,针对特征消歧问题有 4 个代表性的研究工作^[125,129,139-140]. 其中文献[140]根据摄像机姿态排除错误的特征匹配;文献[139]根据特征与图像之间的拓扑结构对特征跟踪链进行优化,以便去除错误匹配;文献[129]根据输入图像序列的几何对称性排除错误匹配,但是此方法只适用于表面结构对称的建筑;文献[125]的方法作为 SFM 系统的后处理步骤,首先对输入的点云模型的摄像机姿态进行分割,然后根据摄像机姿态和图像的超像素分割去除错误匹配,最后将保留的摄像机姿态对应的点云模型进行合并,从而获得较为精确的点云模型. 但是,此方法作为后处理步骤,具有较大的时间开销;此外,此方法的有效性严重依赖摄像机姿态的准确性. 迄今,没有一种通用的特征消歧算法,需要根据输入



(a) 参考视图



(b) SFM重建的点云模型

图 27 重复特征对 SFM 重建结果的影响^[139]

数据集的特点设计合适的算法. 文献^[139]还指出特征消歧容易导致模型的过分割, 从而导致新问题的出现. 因此, 针对特征消歧问题的研究, 任重而道远.

4.4 结构化特征选取

结构化特征选取的目的是: 从特征跟踪方法所产生的特征集合中选取有效的特征进行重建点云模型, 避免外点(outliers)对三维重建精度的影响. 目前存在的结构化特征选取方法有两种:

(1) 对于特征跟踪方法生成的原始特征集合, 采用 RANSAC^[75] 算法去除可能的的外点(outliers), 每次选取特征点数目最多的特征跟踪链作为结构化特征, 用于重建点云模型. 此方法的缺点是: 当外点数目较多时, RANSAC 算法的迭代过程很耗时, 难

以满足 SFM 系统的效率需求.

(2) 根据特定的标准每次选取有效的结构化特征. Beder 等人^[141]最先研究自动的结构化特征选取问题, 提出场景点的置信椭圆(C Confidence Ellipsoids)概念, 选取置信椭圆空间内误差最小的特征跟踪链作为结构化特征参与三维点云重建. Dunn 等人^[142]根据初始模型结构和相机特性提出一种 NBV(Next Nest View)算法, 为 SFM 系统的下一次增量重建选择最佳的结构化特征. 后来, Haner 等人^[143]对 NBV 算法进行改进, 增加了不确定迭代(uncertainty propagation)机制——UNBV 算法, 提高了 NBV 算法对噪声的鲁棒性. 不确定迭代机制的加入提高了三维点云模型的质量, 图 28 为 NBV 与 UNBV 算法



(a) NBV算法结果



(b) UNBV算法结果

图 28 结构化特征选取算法对比结果^[143]

的对比结果(此图为黑白印刷),可以看出 UNBV 算法产生的三维点云与摄像机位置较为精确. Liu 等人^[144]通过大量实验,对特征匹配数目与三维点云的质量进行统计,建立一种经验关系,通过最小化重投影误差,选取误差最小的特征跟踪链作为结构化特征,有效提高了三维点云模型的质量.然而,此方法的缺点是:重投影优化是全局过程,每次优化需要计算所有特征点之间的单应性矩阵,很耗时.最近,Shen 等人^[145]将结构化特征选取过程看作联合优化问题,使用量子启发(quantum inspired)评估算法寻找最优解,从而选取最佳的结构化特征,提高了点云模型的质量.关于结构化特征选取问题,Shen 等人提出的算法在精度和实时性方面均具有较大优势,具有重要的研究和参考价值.

4.5 非线性相机影响

由于非线性相机产生的图像具有大视角和低分辨率特性^[146],近来受到研究者的关注.现有的特征跟踪算法主要用于处理透视相机(Perspective Camera)采集的图像序列.然而,对于非线性相机,如全方位相机(Omnidirectional Camera),鱼眼相机(Fisheye Camera)等,现有的特征跟踪方法容易产生错误匹配. Sturm 等人^[146-147]最先开展关于全方位相机的特征跟踪方法研究,他们采用 SIFT 算法检测特征点,然后根据特征点的尺度空间大小计算尺度因子,将透视图像缩小到一定比例后与全方位图像进行匹配,从而提高了特征匹配精度.但是,计算尺度因子很耗时,为了提高特征跟踪效率,Sturm 等人将全方位图像展开后与透视图像进行匹配,进一步提高了特征跟踪的精度.图 29 展示了非线性相机对特征匹配的影响(此图为黑白印刷),第 1 行为透视图像与全方位图像匹配结果(绿色线表示正确匹配,红色线表示错误匹配),可以看出存在大量的错误匹配;第 2 行是透视图像降采样后与全方位图像匹配结果,由于降采样后尺度空间的一致性提高,从而提高了特征匹配的精度;第 3 行是透视图像与展开的全方位图像匹配结果,可以看出特征的匹配错误被进一步降低.根据文献^[146-147]所述,Sturm 等人提出的特征跟踪方法依赖于 SIFT 特征点的尺度空间特性,因此,他们提出的特征跟踪方法难以适用于不具备多尺度性质的特征点,如 FAST 和 ORB 等特征.最近 Ventura 等人^[148]设计一种适用于非线性相机的 SFM 系统,但是他们仍然采用传统的特征跟踪方法,因此,产生的点云模型存在大量的外点(outliers).对于非线性相机模型下特征跟踪方法的研究正处于探索阶段,缺乏有效的理论支持和通用算法,因此,

未来还需研究者进行深入研究和探索.

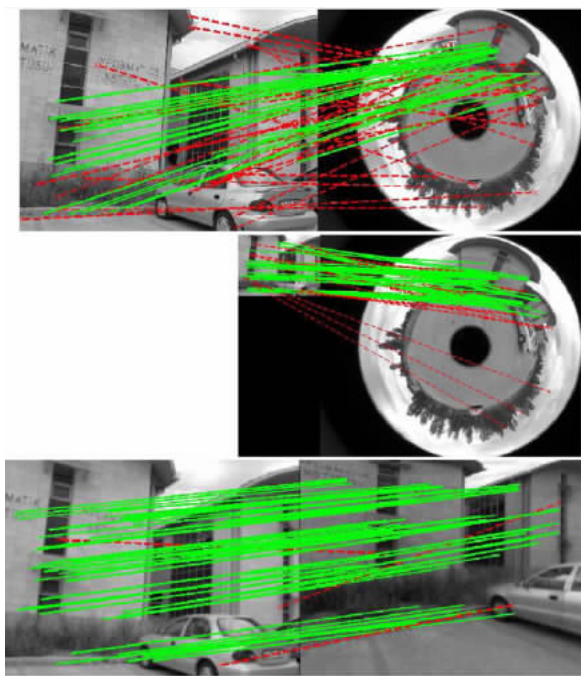


图 29 非线性相机对特征跟踪的影响^[146]

通过以上分析可知,特征跟踪方法还存在一些挑战性的问题需要解决,本文对此进行归纳和总结便于研究者集中于问题解决方案的寻找,而不是寻找问题的本身.

5 发展趋势与展望

随着网络大数据以及深度学习技术的出现,面向 SFM 的特征跟踪方法将面临一些新的具有挑战性的问题.在大数据时代,这些问题也将引领特征跟踪方法的未来研究方向.展望未来,面对网络大数据,以下几个方面将是特征跟踪方法的未来研究方向.

5.1 特征跟踪算法的并行化

大数据的出现有效地提高了 SFM 系统生成的点云模型质量,但是也带来一些新的挑战,其中最为突出的是大数据下的特征跟踪效率.以 SIFT 特征为例,文献^[85]在 YFCC 100M 数据集上计算特征跟踪所耗时间约为整个 SFM 重建工作的 30%,所耗时间之长令人难以接受.因此,为了提高效率,对特征跟踪方法的并行化是未来的发展方向之一.如今,并行化的处理方式有多种:(1)采用 MPI 技术;(2)采用 GPU 硬件;(3)采用大数据专用处理平台,如 Hadoop 和 Spark 等;这三种并行化的理念各不相同,无论选择何种解决方案,均需要坚实的技术基础.

5.2 缓解手工特征依赖

鲁棒的局部特征是 SFM 系统的重要基础,现有的 SFM 系统采用一些手工设计(Hand Crafted)的局部特征,如 SIFT、SURF、ORB、FAST 和 KAZE 等.采用不同的特征对 SFM 系统生成的点云模型质量有较大影响,关于此问题,在 3.2 节的实验中已经得到很好的验证.因此,如何避免手工局部特征对 SFM 的影响是未来的研究方向之一.最近,一些基于传统机器学习与深度学习的局部特征的提出,为解决此问题提供了可能的新途径.由于此类特征集中于最近 2~3 年内被提出,截止到目前,还没有发现被 SFM 系统所采用,有效性还需进一步验证.为了便于读者研究此问题,我们对基于学习的特征生成算法进行总结,详细信息如表 8 所示;其中,文献[149]首次提出采用传统机器学习方法学习图像特征描述子;文献[163]首次采用深度学习技术学习局部特征描述子,采用监督学习技术训练 3 层的神经网络,在 Oxford 数据上测试结果优越于经典的 SIFT 特征.

表 8 基于学习的局部特征

名称	学习方式		文献	备注
	传统学习	深度学习		
LLID	✓	×	[149]	特征描述子
IPSD	✓	×	[150]	特征描述子
BGM	✓	×	[151]	特征描述子
BinBoost	✓	×	[152,153]	特征描述子
CNN	×	✓	[154]	特征描述子
TNet	×	✓	[155]	特征描述子
BOLD	✓	×	[156]	特征描述子
Deepdesc	×	✓	[157]	特征描述子
PN-Net	×	✓	[158]	特征描述子
Hybrid++	×	✓	[159]	特征匹配
MatchNet	×	✓	[160]	特征匹配
DeepCompare	×	✓	[161]	特征匹配
NO-IMI/GNO-IMI	×	✓	[162]	特征匹配
CNN Matching	×	✓	[163]	特征匹配
TILDE	✓	×	[164]	特征检测子
DetNet-L,DetNet-S	×	✓	[165]	特征检测子
DeepBit	×	✓	[166]	特征描述子
Kwang's	×	✓	[167]	特征点方向
LIFT	×	✓	[168]	特征检测子/ 描述子

对于特征跟踪问题,国内外研究者主要从特征检测子,特征描述子和特征匹配三方面研究深度学习技术的应用:

(1)对于描述子,TNet^[154]采用 Siamese 网络学习出一种新的二进制局部描述子,在 PatchData^[149]数据集上 TNet 表现出比传统的 SIFT 和 SURF 具有较好的鲁棒性.但是,由于 TNet 网络结构的加深,使得特征匹配效率有所降低.为了提高效率,文献^[157]对 Siamese 网络进行简化,训练出一种具有较好表达性和较高计算效率的局部描述子——Deepdesc.文献^[158]采用共联的 Triple 网络——PN-Net 训练出一种具有旋转不变性的局部描述子,有效避免了图像旋转和缩放所造成的错误匹配.以上介绍的基于深度学习技术的图像描述子在训练时采用有监督学习技术,即训练数据需要带有特定的标签,因此,训练之前需要对数据进行标注,需要大量时间.为了克服有监督学习的不足,Lin 等人^[166]采用无监督学习技术和 VGG-Net 训练出一种具有较高鲁棒性的二进制描述子——DeepBit,实验结果表明,DeepBit 描述子的表达性较强.

(2)对于特征检测子,文献^[165]首次将深度学习技术应用于特征点检测,作者将特征检测问题看作回归问题(Regression Problem),采用深度神经网络训练出一种带有主方向的局部特征检测子——DetNet,与传统的 SIFT 相比,DetNet 具有较高的重复性(Repeatability).此外,对于 DetNet 和 SIFT 特征检测子,均采用 SIFT 描述子对其进行表示时,DetNet 的匹配精度远高于 SIFT,原因在于:基于深度学习技术的 DetNet 特征点的方向比 SIFT 的主方向稳定,为特征匹配奠定基础.

(3)对于特征匹配,DeepCompare^[161]首次将深度学习技术应用于图像匹配,求解深度图像估计问题,DeepCompare 比传统的 SIFT 方法具有较大优势.受 DeepCompare 思想的启发,文献^[159,162]将深度学习技术应用于航空图像匹配,实验结果表明,与传统的强力匹配方法相比,深度学习技术具有较大优势.最近,关注度较高的是 LIFT(Learned Invariant Feature Transform)^[168]特征,图 30 是 LIFT 特征的流程,可以看出 LIFT 是集特征检测子与描述子于一体的局部特征,与传统的 SIFT 特征相比,在多个实验中 LIFT 均表现出较高的性能.图 31~图 33(此图为黑白印刷)展示了在Radcliffe数据集



图 30 LIFT 特征的流程^[168]

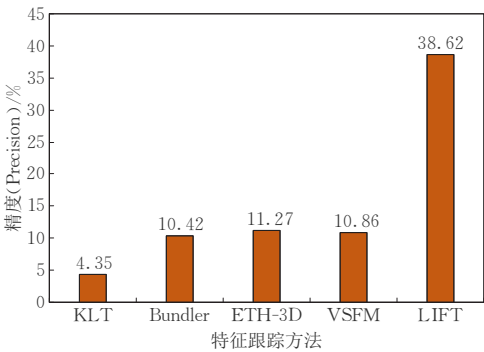


图 31 在 Radcliffe 数据集上每个特征跟踪方法的精度

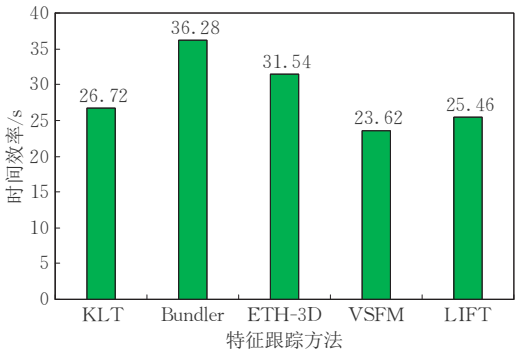


图 32 在 Radcliffe 数据集上每个特征跟踪方法的时间效率

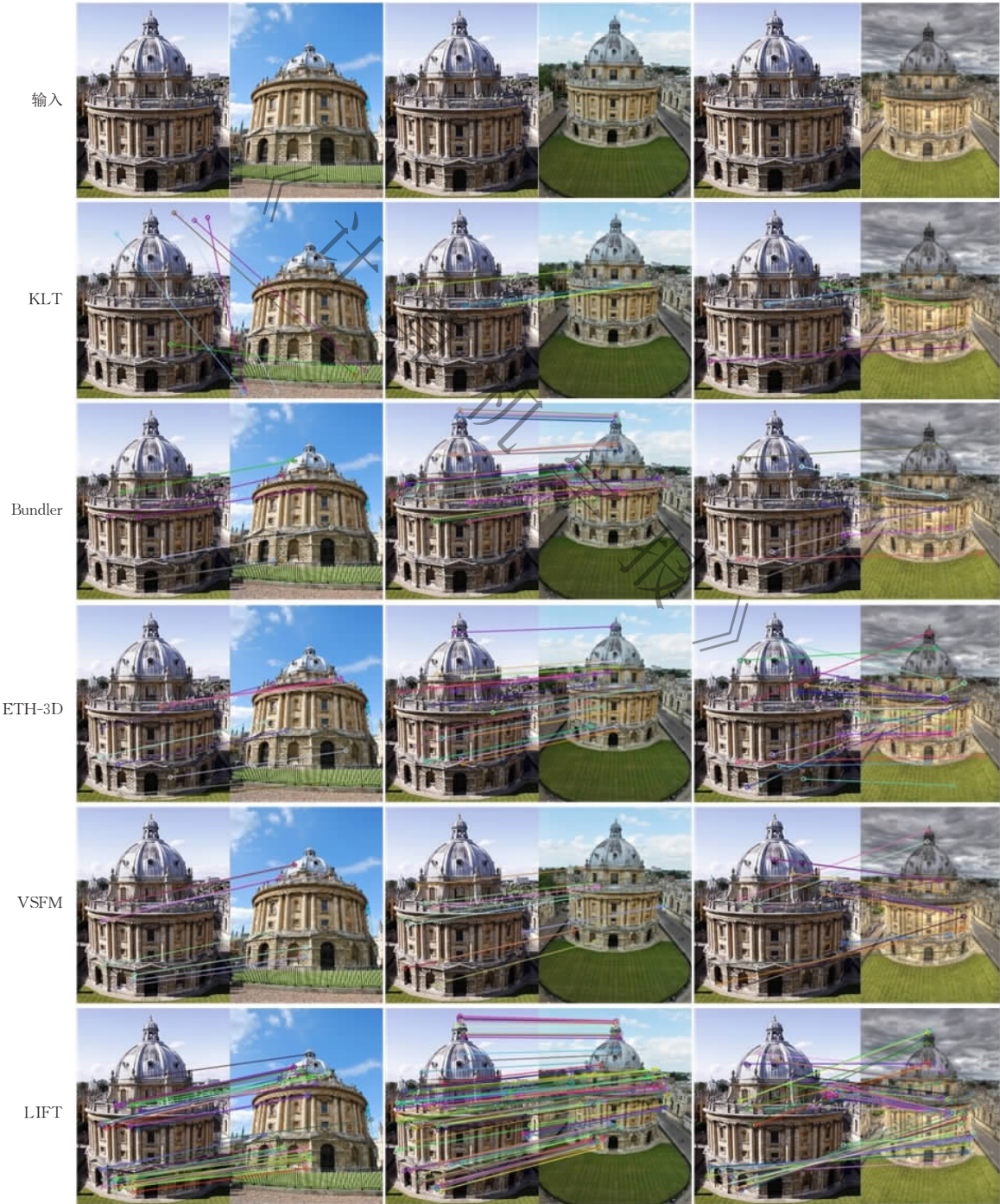


图 33 在 Radcliffe 数据集上的测试结果

上与 KLT 等特征跟踪方法的对比结果,可以看出 LIFT 在时间效率和精度方面都具有较大优势。

深度学习技术在特征跟踪领域的应用与研究才刚刚开始,现有的研究作为后继研究者提供了一定的基础,相信在不久的将来,基于深度学习技术的特征跟踪方法一定能够取得较大的进步,也是未来的重点研究方向。

5.3 移动端的特征跟踪

随着移动设备的普及(如智能手机、平板电脑等),给三维重建工作带来了新的挑战,相应的特征跟踪方法同样需要做出改进。由于移动设备在采集图像过程中处于不稳定状态,常常造成图像的抖动和模糊。这些问题,容易造成特征跟踪丢失,从而导致 SFM 系统重建的点云模型较为稀疏,甚至重建失败。解决这些问题对移动端三维重建具有重要的理论和实用价值,因此,移动端的特征跟踪是未来重点研究方向之一。目前解决图像抖动采用的是视频稳定(Video Stabilization)技术^[169],如微软公司开发的 HyperLapse^[170],虽然此应用没有开放源代码,但是理论基础为研究移动端的视频稳定技术提供重要的指导。图像的去模糊^[171-174]一直是计算机视觉研究的热点,但是这些方法的计算复杂度较高,难以在移动设备上有效运行。因此,在移动设备上实现快速鲁棒的特征跟踪是一项艰巨任务,未来还需要研究者们付出大量的心血。

展望未来,面向移动端的快速鲁棒的特征跟踪和以网络大数据为基础、深度学习为工具的 PC 端的快速特征跟踪是今后研究面向 SFM 的特征跟踪方法的必然趋势。

6 总 结

特征跟踪作为 SFM 的重要组成部分之一,一直是热点研究问题,从以 KLT 为基础的视频序列的特征跟踪方法到以 DMF 框架为基础的无序图像序列的特征跟踪,可以说,特征跟踪在以 SFM 为技术手段的三维重建领域是一个非常活跃的研究方向。从整体上讲,尚未建立一套完整的理论体系能够同时满足移动端和 PC 端的快速鲁棒的特征跟踪方法。

为了便于设计和选择合适的特征跟踪方法,本文全面综述了现有的特征跟踪方法。主要贡献如下:(1)全面综述和分析了现有经典的特征跟踪方法,总结了现有可用资源(特征检测子、特征描述子、特征匹配方法和可用数据集);(2)对经典的特征跟踪

方法进行全面的实验评估,总结和分析了现有特征跟踪方法的优点与缺点;(3)对影响特征跟踪方法效率和精度的因素进行分析,为设计快速鲁棒的特征跟踪方法奠定基础;本文还探讨了一些亟待解决的问题,并对未来发展进行了展望。

参 考 文 献

- [1] Wang Jun, Zhu Li. 3D building façade reconstruction based on image matching-point cloud fusing. Chinese Journal of Computers, 2012, 35(10): 2072-2079(in Chinese)
(王俊,朱利.基于图像匹配-点云融合的建筑立面三维重建.计算机学报,2012,35(10):2072-2079)
- [2] Wang Yu-Xi, Liu Yue. Robust structure from motion approach based on multi cameras. Chinese Journal of Scientific Instrument, 2015, 36(1): 230-238(in Chinese)
(王宇希,刘越.一种基于多摄像机的鲁棒运动结构重建方法.仪器仪表学报,2015,36(1):230-238)
- [3] Schönberger J L, Frahm J-M. Structure-from-motion revisited //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 600-611
- [4] Zhang Feng, Shi Li-Min, Sun Feng-Mei, et al. An image based 3D reconstruction system for large indoor scenes. Acta Automatica Sinica, 2010, 36(5): 625-633(in Chinese)
(张峰,史利民,孙凤梅等.一种基于图像的室内大场景自动三维重建系统.自动化学报,2010,36(5):625-633)
- [5] Zhang Guo-Feng, Qin Xue-Ying, Dong Zi-Long, et al. Camera tracking based on structure and motion recovery for augmented video. Chinese Journal of Computers, 2006, 29(12): 2104-2111(in Chinese)
(章国锋,秦学英,董子龙等.面向增强视频的基于结构和运动恢复的摄像机定标.计算机学报,2006,29(12):2104-2111)
- [6] Chen Hui, Ma Shi-Wei, Andreas Nuechter. Non-synchronous point cloud algorithm for 3D reconstruction based on laser scanning and SFM. Chinese Journal of Scientific Instrument, 2016, 35(6): 1148-1157(in Chinese)
(陈辉,马世伟,Andreas Nuechter.基于激光扫描和SFM的非同步点云三维重构方法.仪器仪表学报,2016,35(6):1148-1157)
- [7] Kong C, Lucey S. Prior-less compressible structure from motion//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 460-471
- [8] Zheng E, Wu C. Structure from motion using structure-less resection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 2075-2083
- [9] Zhang Cong-Xuan, Chen Zhen, Li Ming, et al. 3D motion detection and dense structure reconstruction based on variational optical flow. Chinese Journal of Scientific Instrument, 2012, 33(6): 1315-1323(in Chinese)

- (张聪炫, 陈震, 黎明等. 基于变分光流的三维运动检测与稠密结构重建. 仪器仪表学报, 2012, 33(6): 1315-1323)
- [10] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces//Proceedings of the IEEE International Symposium on Mixed and Augmented Reality. Nara, Japan, 2007: 225-234
- [11] Li Z, Wang Y, Guo J, et al. Diminished reality using appearance and 3D geometry of internet photo collections//Proceedings of the IEEE International Symposium on Mixed and Augmented Reality. Adelaide, Australia, 2013: 11-19
- [12] Colbert M, Bouguet J-Y, Beis J, et al. Building indoor multi-panorama experiences at scale//Proceedings of the ACM SIGGRAPH 2012 Talks. Los Angeles, USA, 2012, Article No. 24
- [13] Wang Kun, Zheng Nan-Ning. 3D face modelling based on SFM algorithm. Chinese Journal of Computers, 2005, 28(6): 1048-1053(in Chinese)
(王琨, 郑南宁. 基于 SFM 算法的三维人脸模型重建. 计算机学报, 2005, 28(6): 1048-1053)
- [14] Sakurada K, Okatani T, Deguchi K. Detecting changes in 3D structure of a scene from multi-view images captured by a vehicle-mounted camera//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 137-144
- [15] Qin Lei, Hu Qiong, Huang Qing-Ming, et al. Action recognition using trajectories of spatio-temporal feature points. Chinese Journal of Computers, 2014, 37(6): 1281-1288(in Chinese)
(秦磊, 胡琼, 黄庆明等. 基于特征点轨迹的动作识别. 计算机学报, 2014, 37(6): 1281-1288)
- [16] Guo Zi-Xin, Yi Yang, Li Han-Ju. Recognizing actions from videos in the wild via adaptive feature fusion. Chinese Journal of Computers, 2013, 36(11): 2330-2339(in Chinese)
(郭梓鑫, 衣杨, 李汉巨. 基于自适应特征融合的自然环境视频行为识别. 计算机学报, 2013, 36(11): 2330-2339)
- [17] Song S, Chandraker M. Robust scale estimation in real-time monocular SFM for autonomous driving//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1566-1573
- [18] Mansfield A, Kobyshev N, Riemenschneider H, et al. Frankenhorse: Automatic completion of articulating objects from image-based reconstruction//Proceedings of the British Machine Vision Conference. Nottingham, UK, 2014: 1-12
- [19] Zhang G, He Y, Chen W, et al. Multi-viewpoint panorama construction with wide-baseline images. IEEE Transactions on Image Processing, 2016, 25(7): 3099-3111
- [20] Snavely N, Seitz S M, Szeliski R. Photo tourism: exploring photo collections in 3D. ACM Transactions on Graphics, 2006, 25(3): 835-846
- [21] Zach C. Robust bundle adjustment revisited//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 772-787
- [22] Dong Z, Zhang G, Jia J, et al. Keyframe-based real-time camera tracking//Proceedings of the IEEE International Conference on Computer Vision. Kyoto, Japan, 2009: 1538-1545
- [23] Zhang G, Liu H, Dong Z, et al. ENFT: Efficient non-consecutive feature tracking for robust structure-from-motion. arXiv preprint arXiv:151008012, 2015
- [24] Wu C. Towards linear-time incremental structure from motion//Proceedings of the IEEE International Conference on 3D Vision-3DV. Seattle, USA, 2013: 127-134
- [25] Ni K, Dellaert F. HyperSfM//Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT). Zurich, Switzerland, 2012: 144-151
- [26] Wang T Y, Kohli P, Mitra N J. Dynamic SFM: Detecting scene changes from image pairs. Computer Graphics Forum, 2015, 34(5): 177-189
- [27] Wilson K, Snavely N. Robust global translations with 1DSfM//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 61-75
- [28] Crandall D, Owens A, Snavely N, et al. Discrete-continuous optimization for large-scale structure from motion//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 3001-3008
- [29] Dorini L, Goldenstein S. Unscented KLT: Nonlinear feature and uncertainty tracking//Proceedings of the 19th Brazilian Symposium on Computer Graphics and Image Processing. Brasilia, Brazilian, 2006: 187-193
- [30] Zhang G, Vela P. Good feature to track for visual SLAM//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1373-1382
- [31] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision//Proceedings of the International Joint Conference on Artificial Intelligence. Vancouver, Canada, 1981: 674-679
- [32] Jia K, Chan T-H, Zeng Z, et al. ROML: A robust feature correspondence approach for matching objects in a set of images. International Journal of Computer Vision, 2015, 117(2): 1-25
- [33] Zhang G, Dong Z, Jia J, et al. Efficient non-consecutive feature tracking for structure-from-motion//Proceedings of the European Conference on Computer Vision. Hersonissos, Greece, 2010: 422-435
- [34] Zhao L, Li X, Xiao J, et al. Metric learning driven multi-task structured output optimization for robust keypoint tracking//Proceedings of the AAAI Conference on Artificial Intelligence. Austin, USA, 2015: 3864-3870
- [35] Jiang H, Liu H, Tan P, et al. 3D reconstruction of dynamic scenes with multiple handheld cameras//Proceedings of the European Conference on Computer Vision. Florence, Italy, 2012: 601-615
- [36] Cao Ming-Wei, Li Shu-Jie, Jia Wei, et al. Fast and robust feature tracking algorithm for large-scale SFM. Chinese Journal of Scientific Instrument, 2017, 38(1): 252-260(in Chinese)
(曹明伟, 李书杰, 贾伟等. 面向大规模 SFM 的快速鲁棒特征跟踪算法. 仪器仪表学报, 2017, 38(1): 252-260)

- [37] Wang K, Zhang G, Bao H. Robust 3D reconstruction with an RGB-D camera. *IEEE Transactions on Image Processing*, 2014, 23(11): 4893-4906
- [38] Song Y, Chen X, Wang X, et al. Fast estimation of relative poses for 6-DOF image localization//*Proceedings of the IEEE International Conference on Multimedia Big Data*. Beijing, China, 2015: 156-163
- [39] Cui H, Shen S, Gao W, et al. Efficient large-scale structure from motion by fusing auxiliary imaging information. *IEEE Transactions on Image Processing*, 2015, 24(11): 3561-3573
- [40] Fan B, Wu F, Hu, Z. Towards reliable matching of images containing repetitive patterns. *Pattern Recognition Letters*, 2011, 32(14): 1851-1859
- [41] Wang Z, Fan B, Wang G, et al. Exploring local and overall ordinal information for robust feature description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(11): 2198-2211
- [42] Fan B, Wu F, Hu Z. Rotationally invariant descriptors using intensity order pooling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(10): 2031-2045
- [43] Wu F C, Wang Z H, Wang X G. Feature vector field and feature matching. *Pattern Recognition*, 2010, 43(10): 3273-3281
- [44] Cheng J, Leng C, Wu J, et al. Fast and accurate image matching with cascade hashing for 3D reconstruction//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014: 1-8
- [45] Zhu H, Liu Y, Fan J, et al. Video-based outdoor human reconstruction. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016, 27(4): 760-770
- [46] Liu Y, Cao X, Dai Q, et al. Continuous depth estimation for multi-view stereo//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Miami, USA, 2009: 2121-2128
- [47] Gao X, Zhang T. Loop closure detection for visual slam systems using deep neural networks//*Proceedings of the Chinese Control Conference*. Hangzhou, China, 2015: 5851-5856
- [48] Gurdjos P, Sturm P, Wu Y. Euclidean Structure from $N \geq 2$ Parallel Circles: Theory and Algorithms//*Proceedings of the European Conference on Computer Vision*. Graz, Austria, 2006: 238-252
- [49] Feng Y, Fan L, Wu Y. Fast Localization in large-scale environments using supervised indexing of binary features. *IEEE Transactions on Image Processing*, 2016, 25(1): 343-358
- [50] Feng Y, Wu Y, Fan L. Online learning of binary feature indexing for real-time SLAM relocalization//*Proceedings of the Asian Conference on Computer Vision (Workshops)*. Singapore, 2014: 206-217
- [51] Zhang Guofeng, He Yi, Chen Weifeng, et al. Multi-viewpoint panorama construction with wide-baseline images. *IEEE Transactions on Image Processing*, 2016, 25(7): 3099-3111
- [52] Shi L, Guo F, Gao W, et al. 3D reconstruction via semantic interaction. *Journal of Computer-Aided Design & Computer Graphics*, 2011, 23(5): 839-848
- [53] Shen S. Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes. *IEEE Transactions on Image Processing*, 2013, 22(5): 1901-1914
- [54] Hainan C, Shuhan S, Zhanyi H. Global fusion of generalized camera model for efficient large-scale structure from motion. *Science China Information Sciences*, 2016, 59(1): 1-3
- [55] Liu Hao-Min, Zhang Guo-Feng, Bao Hu-Jun. A survey of monocular simultaneous localization and mapping. *Journal of Computer-Aided Design and Computer Graphics*, 2016, 28(6): 855-868(in Chinese)
(刘浩敏, 章国锋, 鲍虎军. 基于单目视觉的同时定位与地图构建方法综述. *计算机辅助设计与图形学学报*, 2016, 28(6): 855-868)
- [56] Sun Feng-Mei, Wang Bo. A note on the roots distribution and stability of the PnP problem. *Acta Automatica Sinica*, 2010, 36(9): 1213-1219(in Chinese)
(孙凤梅, 王波. 关于 PnP 问题多解的分布与解的稳定性的讨论. *自动化学报*, 2010, 36(9): 1213-1219)
- [57] Guo Yang, Xu Xin-He. An analytic solution for the P5P problem with an un-calibrated camera. *Chinese Journal of Computers*, 2007, 30(7): 1195-1200(in Chinese)
(郭阳, 徐心和. 未标定摄像机 P5P 问题的一种解析解. *计算机学报*, 2007, 30(7): 1195-1200)
- [58] Wang Y, Ji R, Chang S-F. Label propagation from ImageNet to 3D point clouds//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA, 2013: 3135-3142
- [59] Sinha S N, Frahm J-M, Pollefeys M, et al. GPU-based video feature tracking and matching//*Proceedings of the EDGE, Workshop on Edge Computing Using New Commodity Architectures*. Boston, USA, 2006: 189-196
- [60] Sinha S N, Frahm J-M, Pollefeys M, et al. Feature tracking and matching in video using programmable graphics hardware. *Machine Vision and Applications*, 2011, 22(1): 207-217
- [61] Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91-110
- [62] Zach C, Gallup D, Frahm J-M. Fast gain-adaptive KLT tracking on the GPU//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage, USA, 2008: 1-7
- [63] Kim S, Gallup D, Frahm J, et al. Gain adaptive real-time stereo streaming//*Proceedings of the International Conference on Vision Systems*. Las Vegas, USA, 2007: 163-179
- [64] Buchanan A, Fitzgibbon A. Combining local and global motion models for feature point tracking//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, USA, 2007: 1-8
- [65] Hwangbo M, Kim J-S, Kanade T. Inertial-aided KLT feature tracking for a moving camera//*Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. Louis, USA, 2009: 1909-1916
- [66] Hwangbo M, Kim J-S, Kanade T. Gyro-aided feature tracking for a moving camera: fusion, auto-calibration and GPU implementation. *International Journal of Robotics Research*, 2011, 30(14): 1755-1774

- [67] Wu C, Agarwal S, et al. Schematic surface reconstruction//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012; 1498-1505
- [68] Garcia V, Debreuve E, Nielsen F, et al. K-nearest neighbor search: Fast GPU-based implementations and application to high-dimensional feature matching//Proceedings of the IEEE International Conference on Image Processing. Hong Kong, China, 2010; 3757-3760
- [69] Muja M, Lowe D G. Fast matching of binary features//Proceedings of the IEEE Conference on Computer and Robot Vision. Toronto, Canada, 2012; 404-410
- [70] Muja M, Lowe D G. Scalable nearest neighbor algorithms for high dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(11): 2227-2240
- [71] Muja M, Lowe D G. Flann, fast library for approximate nearest neighbors with automatic configuration//Proceedings of the International Conference on Computer Vision Theory and Applications. Lisboa, Portugal, 2009; 331-340
- [72] Buchanan A, Fitzgibbon A. Interactive feature tracking using KD trees and dynamic programming//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA, 2006; 626-633
- [73] Wu C, Clipp B, Li X, et al. 3D model matching with viewpoint-invariant patches (VIP)//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008; 1-8
- [74] Chum O, Matas J. Matching with PROSAC-progressive sample consensus//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA, 2005; 220-226
- [75] Civera J, Grasa O G, Davison A J, et al. 1-point RANSAC for extended Kalman filtering: Application to real-time structure from motion and visual odometry. *Journal of Field Robotics*, 2010, 27(27): 609-631
- [76] Ni K, Jin H, Dellaert F. GroupSAC: Efficient consensus in the presence of groupings//Proceedings of the IEEE International Conference on Computer Vision. Kyoto, Japan, 2009; 2193-2200
- [77] Lee T, Hollerer T. Hybrid feature tracking and user interaction for markerless augmented reality//Proceedings of the IEEE Virtual Reality. Reno, USA, 2008; 145-152
- [78] Svärm L, Simayijiang Z, Enqvist O, et al. Point track creation in unordered image collections using Gomory-Hu trees//Proceedings of the IEEE International Conference on Pattern Recognition. Tsukuba, Japan, 2012; 2116-2119
- [79] Gomory R E, Hu T C. Multi-terminal network flows. *Journal of the Society for Industrial and Applied Mathematics*, 1961, 9(4): 551-570
- [80] Garrigues M, Manzanera A. Real time semi-dense point tracking//Proceedings of the International Conference Image Analysis and Recognition. Montreal, Canada, 2012; 245-252
- [81] Garrigues M, Manzanera A, Bernard T M. Video extruder: A semi-dense point tracker for extracting beams of trajectories in real time. *Journal of Real-Time Image Processing*, 2016, 11(4): 785-798
- [82] Poling B, Lerman G, Szlam A. Better feature tracking through subspace constraints//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Stockholm, Sweden, 2014; 3454-3461
- [83] Agarwal S, Snavely N, Simon I, et al. Building Rome in a day//Proceedings of the IEEE International Conference on Computer Vision. Kyoto, Japan, 2009; 72-79
- [84] Xiao J, Owens A, Torralba A. SUN3D: A database of big spaces reconstructed using SfM and object labels//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013; 1625-1632
- [85] Heinly J, Schonberger J L, Dunn E, et al. Reconstructing the World* in Six Days* (as captured by the Yahoo 100 million image dataset)//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015; 3287-3295
- [86] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain, 2011; 2564-2571
- [87] Mair E, Hager G D, Burschka D, et al. Adaptive and generic corner detection based on the accelerated segment test//Proceedings of the European Conference on Computer Vision. Hersonissos, Greece, 2010; 183-196
- [88] Harris C, Stephens M. A combined corner and edge detector //Proceedings of the Alvey Vision Conference. Boston, USA, 1988; 32-40
- [89] Forssén P-E, Lowe D G. Shape descriptors for maximally stable extremal regions//Proceedings of the IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil, 2007; 1-8
- [90] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded up robust features//Proceedings of the European Conference on Computer Vision. Graz, Austria, 2006; 404-417
- [91] Rosten E, Drummond T. Machine learning for high-speed corner detection//Proceedings of the European Conference on Computer Vision. Graz, Austria, 2006; 430-443
- [92] Wang Z, Fan B, Wu F. FRIF: Fast robust invariant feature//Proceedings of the British Machine Vision Conference. Bristol, UK, 2013; 16.1-16.11
- [93] Leutenegger S, Chli M, Siegwart R Y. BRISK: Binary robust invariant scalable keypoints//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain, 2011; 2548-2555
- [94] Alcantarilla P F, Bartoli A, Davison A J. KAZE features//Proceedings of the European Conference on Computer Vision. Firenze, Italy, 2012; 214-227
- [95] Marimon D, Bonnin A, Adamek T, et al. DARTs: Efficient scale-space extraction of DAISY keypoints//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010; 2416-2423
- [96] Calonder M, Lepetit V, Strecha C, et al. BRIEF: Binary robust independent elementary features//Proceedings of the European Conference on Computer Vision. Hersonissos, Greece, 2010; 778-792

- [97] Wang Z, Fan B, Wu F. Local intensity order pattern for feature description//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain, 2011: 603-610
- [98] Fan B, Wu F, Hu Z. Aggregating gradient distributions into intensity orders: A novel local image descriptor//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 2377-2384
- [99] Alahi A, Ortiz R, Vanderghenst P. FREAK: Fast retina keypoint//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 510-517
- [100] Levi G, Hassner T. LATCH: Learned arrangements of three patch codes//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. New York, USA, 2016: 1-9
- [101] Yang X, Cheng K-T. LDB: An ultra-fast feature for scalable augmented reality on mobile devices//Proceedings of the IEEE International Symposium on Mixed and Augmented Reality. Atlanta, USA, 2012: 49-57
- [102] Heinly J, Dunn E, Frahm J-M. Comparative evaluation of binary features//Proceedings of the European Conference on Computer Vision. Firenze, Italy, 2012: 759-773
- [103] Hwang Y, Han B, Ahn H-K. A fast nearest neighbor search algorithm by nonlinear embedding//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 3053-3060
- [104] Andoni A, Indyk P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions//Proceedings of the IEEE Symposium on Foundations of Computer Science. Washington, USA, 2006: 459-468
- [105] Norouzi M, Punjani A, Fleet D J. Fast exact search in hamming space with multi-index hashing. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(6): 1107-1119
- [106] Nister D, Stewenius H. Scalable recognition with a vocabulary tree//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA, 2006: 2161-2168
- [107] Silpa-Anan C, Hartley R. Optimised KD-trees for fast image descriptor matching//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8
- [108] Xia Y, He K, Wen F, et al. Joint inverted indexing//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 3416-3423
- [109] Zhou F, De La Torre F. Deformable graph matching//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 2922-2929
- [110] Hu Y-T, Lin Y-Y. Progressive feature matching with alternate descriptor selection and correspondence enrichment //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 346-354
- [111] Chui H, Rangarajan A. A new algorithm for non-rigid point matching//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head, USA, 2000: 44-51
- [112] Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding, 2003, 89(2): 114-141
- [113] Lian W, Zhang L, Yang M-H. An efficient globally optimal algorithm for asymmetric point matching//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 504-515
- [114] Talker L, Moses Y, Shimshoni I. Using spatial order to boost the elimination of incorrect feature matches//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 1809-1817
- [115] Zhang Z, Shi Q, Mcauley J, et al. Pairwise matching through max-weight bipartite belief propagation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 1202-1210
- [116] Ong E-J, Bober M. Improved hamming distance search using variable length substrings//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2000-2008
- [117] Lejsek H, Jónsson B, Amsaleg L. NV-Tree: Nearest neighbors at the billion scale//Proceedings of the Proceedings of the ACM International Conference on Multimedia Retrieval. Trento, Italy, 2011: 528-530
- [118] Naidan B, Boytsov L, Nyberg E. Permutation search methods are efficient, yet faster search is possible. Proceedings of the VLDB Endowment, 2015, 8(12): 1618-1629
- [119] Harwood B, Drummond T. FANNG: Fast approximate nearest neighbour graphs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 5713-5722
- [120] Wieschollek P, Wang O, Sorkine-Hornung A, et al. Efficient large-scale approximate nearest neighbor search on the GPU//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2027-2035
- [121] Moulon P, Monasse P, Marlet R. Global fusion of relative motions for robust, accurate and scalable structure from motion//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 3248-3255
- [122] Jiang N, Cui Z, Tan P. A global linear method for camera pose registration//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 481-488
- [123] Kendall A, Grimes M, Cipolla R. PoseNet: A convolutional network for real-time 6-DOF camera relocalization//Proceedings of the IEEE International Conference on Computer Vision. Santiago, USA, 2015: 2938-2946
- [124] Wilson K, Snavely N. Network principles for SfM: Disambiguating repeated structures with local context//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 513-520

- [125] Heinly J, Dunn E, Frahm J M. Correcting for duplicate scene structure in sparse 3D reconstruction//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 780-795
- [126] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the KITTI vision benchmark suite//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 3354-3361
- [127] Strecha C, Hansen W V, Gool L V, et al. On benchmarking camera calibration and multi-view stereo for high resolution imagery//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8
- [128] Cohen A, Zach C, Sinha S N, et al. Discovering and exploiting 3D symmetries in structure from motion//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 1514-1521
- [129] Ceylan D, Mitra N J, Zheng Y, et al. Coupled structure-from-motion and 3D symmetry detection for urban facades. *ACM Transactions on Graphics*, 2014, 33(1): 57-76
- [130] Crocco M, Rubino C, Del Bue A. Structure from motion with objects//Proceedings of the European Conference on Computer Vision. Amsterdam, Netherlands, 2016: 100-112
- [131] Bao S Y, Savarese S. Semantic structure from motion//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 2025-2032
- [132] Shao H, Tuytelaars T, Van Gool L. HPAT indexing for fast object/scene recognition based on local appearance//Proceedings of the Image and Video Retrieval, Second International Conference. Urbana-Champaign, USA, 2003: 307-312
- [133] Smith M, Baldwin I, Churchill W, et al. The new college vision and laser data set. *International Journal of Robotics Research*, 2009, 28(5): 595-599
- [134] Li Y, Snavely N, Huttenlocher D P. Location recognition using prioritized feature matching//Proceedings of the European Conference on Computer Vision. Hersonissos, Greece, 2010: 791-804
- [135] Thomee B, Shamma D A, Friedland G, et al. YFCC100M: The new data in multimedia research. *Communications of the ACM*, 2016, 59(2): 64-73
- [136] Zitnick C, Ramnath K. Edge foci interest points//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain, 2011: 359-366
- [137] Moisan L, Stival B. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal of Computer Vision*, 2004, 57(3): 201-218
- [138] Moisan L, Moulon P, Monasse P. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing on Line*, 2012, 2: 56-73
- [139] Wilson K, Snavely N. Network principles for SfM: Disambiguating repeated structures with local context//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 513-520
- [140] Roberts R, Sinha S N, Szeliski R, et al. Structure from motion for scenes with large duplicate structures//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 3137-3144
- [141] Beder C, Steffen R. Determining an initial image pair for fixing the scale of a 3D reconstruction from an image sequence//Proceedings of the 28th Annual Pattern Recognition Symposium of the German Association for Pattern Recognition. Berlin, Germany, 2006: 12-22
- [142] Dunn E, Frahm J-M. Next best view planning for active model improvement//Proceedings of the British Machine Vision Conference. London, UK, 2009: 1-11
- [143] Haner S, Heyden A. Covariance propagation and next best view planning for 3D reconstruction//Proceedings of the European Conference on Computer Vision. Firenze, Italy, 2012: 545-556
- [144] Liu Z, Monasse P, Marlet R. Match selection and refinement for highly accurate two-view structure from motion//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 818-833
- [145] Shen S, Hu Z. How to select good neighboring images in depth-map merging based 3D modeling. *IEEE Transactions on Image Processing*, 2014, 23(1): 308-318
- [146] Bastanlar Y, Temizel A, Yardimci Y, Sturm P. Multi-view structure-from-motion for hybrid camera scenarios. *Image and Vision Computing*, 2012, 30(8): 557-572
- [147] Puig L, Sturm P, Guerrero J. Hybrid homographies and fundamental matrices mixing uncalibrated omnidirectional and conventional cameras. *Machine Vision and Applications*, 2013, 24(4): 721-738
- [148] Ventura J. Structure from Motion on a Sphere//Proceedings of the European Conference on Computer Vision. Amsterdam, Netherlands, 2016: 53-68
- [149] Winder S A, Brown M. Learning local image descriptors//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA, 2007: 1-8
- [150] Kavukcuoglu K, Fergus R, Lecun Y. Learning invariant features through topographic filter maps//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 1605-1612
- [151] Trzcinski T, Christoudias M, Lepetit V, et al. Learning image descriptors with the boosting-trick//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2012: 269-277
- [152] Trzcinski T, Christoudias M, Fua P, et al. Boosting binary keypoint descriptors//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 2874-2881

- [153] Trzcinski T, Christoudias M, Lepetit V. Learning image descriptors with boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 597-610
- [154] Simo-Serra E, Trulls E, Ferraz L, et al. Fracking deep convolutional image descriptors. *Computer Science*, 2015, 20(3): 52-61
- [155] Kumar B, Carneiro G, Reid I. Learning local image descriptors with deep siamese and triplet convolutional networks by minimising global loss functions//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015: 501-512
- [156] Balntas V, Tang L, Mikolajczyk K. BOLD-binary online learned descriptor for efficient image matching//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015: 2367-2375
- [157] Simo-Serra E, Trulls E, Ferraz L, et al. Discriminative learning of deep convolutional feature point descriptors//*Proceedings of the IEEE International Conference on Computer Vision*. Santiago, Chile, 2015: 118-126
- [158] Balntas V, Johns E, Tang L, et al. PN-Net: Conjoined triple deep network for learning local image descriptors//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 55-66
- [159] Altwaijry H, Trulls E, Hays J, et al. Learning to match aerial images with deep attentive architectures//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 3539-3547
- [160] Han X, Leung T, Jia Y, et al. MatchNet: Unifying feature and metric learning for patch-based matching//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015: 3279-3286
- [161] Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015: 4353-4361
- [162] Babenko A, Lempitsky V. Efficient indexing of billion-scale datasets of deep descriptors//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 2055-2063
- [163] Fischer P, Dosovitskiy A, Brox T. Descriptor matching with convolutional neural networks: A comparison to sift. *Computer Science*, 2015, 20(8): 20-32
- [164] Verdie Y, Yi K M, Fua P, et al. TILDE: A temporally invariant learned DETector//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015: 5279-5288
- [165] Lenc K, Vedaldi A. Learning covariant feature detectors//*Proceedings of the ECCV Workshop on Geometry Meets Deep Learning*. Amsterdam, Netherlands, 2016: 381-398
- [166] Lin K, Lu J, Chen C-S, et al. Learning compact binary descriptors with unsupervised deep neural networks//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 1183-1192
- [167] Yi K M, Verdie Y, Fua P, et al. Learning to assign orientations to feature points//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 107-116
- [168] Yi K M, Trulls E, Lepetit V, et al. LIFT: Learned invariant feature transform//*Proceedings of the European Conference on Computer Vision*. San Diego, USA, 2016: 467-483
- [169] Liu S, Yuan L, Tan P, et al. Bundled camera paths for video stabilization. *ACM Transactions on Graphics*, 2013, 32(4): 78-96
- [170] Kopf J, Cohen M F, Szeliski R. First-person hyper-lapse videos. *ACM Transactions on Graphics*, 2014, 33(4): 1-10
- [171] Gregson J, Heide F, Hullin M B, et al. Stochastic deconvolution//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA, 2013: 1043-1050
- [172] Fergus R, Singh B, Hertzmann A, et al. Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 2006, 25(3): 787-794
- [173] Levin A, Weiss Y, Durand F, et al. Efficient marginal likelihood optimization in blind deconvolution//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Colorado, USA, 2011: 2657-2664
- [174] Hacoen Y, Shechtman E, Lischinski D. Deblurring by example using dense correspondence//*Proceedings of the IEEE International Conference on Computer Vision*. Sydney, Australia, 2013: 2384-2391



CAO Ming-Wei, born in 1986, Ph.D. candidate. His research interests include large-scale semantic 3D reconstruction based on structure from motion and simultaneous localization and mapping, 3D object detection and augmented reality.

LI Shu-Jie, born in 1983, Ph.D. Her research interests include pattern recognition and intelligence system.

JIA Wei, born in 1978, Ph.D., associate professor. His research interests include computer vision and pattern recognition.

LIU Xiao-Ping, born in 1964, Ph.D., professor. His research interests include 3D reconstruction, computer animation, cooperative computing and simulation.

Background

In recent years, 3D reconstruction has received wide attentions from academia and industry due to its great application prospects. It is well known that Structure from Motion (SFM) is one of important techniques for 3D reconstruction, which has been successfully exploited in many computer vision tasks such as augmented reality, 3D-map, 3D face reconstruction, change detection, action recognition, driverless and panoramic, etc. Generally, a typical SFM system consists of feature tracking, camera pose estimation, triangulation, and bundle adjustment. Among them, feature tracking is a key component of SFM, whose goal is to accurately locate correspondences between different images. In the past ten years, the research on feature tracking has made a great progress. So far, however, there lacks a thorough and detailed analysis on current research progress and development trends of feature tracking. To solve this problem, in this paper, we make a comprehensive review of feature tracking technique. In detail, we firstly introduce the goal of feature tracking for video sequence and un-order image sequence, and analyze the basic theory of feature tracking technique. Then, the research status and progress of feature tracking are summarized. Moreover, we collected many available resources about feature

tracking including feature detectors, feature descriptors, matching strategies and testing datasets. We also evaluate the state-of-the-art feature tracking methods, and discuss what influences the performances of feature tracking. Finally, we conclude the problems that should be solved in the future, and present the future research directions.

We have conducted a series of studies on SFM and feature tracking from the point of efficiency and quality. To this end, we have developed a novel SFM system which outperforms the state-of-the-art systems including Bundler^[20], ETH-V3D^[21] and Visual SFM^[24]. And we have also acquired the warranty of two related Chinese patents. Based on above research work and related work of exports and scholars at home and abroad, we analyzed and summarized the latest research results, prospect a further research on feature tracking in the field of SFM.

This research was supported by the National Key Research and Development Program of China (No. 2016YFC0800100), the National Natural Science Foundations of China (61370167, 61305093, 61673157, 61802103, 61877016), and the Natural Science Foundations of Anhui Province (JZ2015AKZR0664).