

基于移动增强现实的智慧城市导览

张运超 陈靖 王涌天 刘越

(北京理工大学计算机学院 北京 100081)

(zhangyunchao163@163.com)

Smart City Guide Using Mobile Augmented Reality

Zhang Yunchao, Chen Jing, Wang Yongtian, and Liu Yue

(School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081)

Abstract A new technique for smart city guide using mobile augmented reality is proposed, which satisfies the personalized, multi-scale, comprehensive needs of users and presents active interface with virtual-real fusion. Mobile side is limited by computing power and resource storage capacity. However, mobile devices usually integrate multiple inertial sensors, which are portable and easy to display. Server side is used for city-scale location recognition based on vocabulary tree method. Dynamic partition method with GPS information reduces the range of image retrieval. Hierarchical k -means clustering on BRISK feature with binary descriptors improves the real-time performance of vocabulary tree. Hybrid features based on BRISK and optical flow are executed in parallel for real-time and robust tracking. Regular re-initialization with BRISK feature is used for reducing errors generated by optical flow. Matching point sets mapping is applied for eliminating drift of feature points during initialization of BRISK feature. Sequence frames and keyframe information are used for reducing jitter with pose estimation. Experimental results on UKbench and real environment demonstrate the advantage of virtual-real fusion for city-scale smart guide. Users can easily interact with surrounding real environment. The prototype system has been successfully applied to smart guide system of Shanghai Telecom Experience Venue and other such guide systems.

Key words smart city guide; mobile augmented reality; dynamic partition; hierarchical clustering; hybrid feature tracking

摘要 提出一种采用移动增强现实技术实现智慧城市导览的方法,满足用户个性化、多尺度、按需推送的智能导览需求,呈现用户虚实融合的周边环境。移动终端计算性能以及资源存储能力有限,但集成多种传感器,方便携带,易于显示。利用服务器实现基于词汇树的海量场景识别定位系统。依据地理位置信息动态划分分区缩减了场景检索范围,基于二进制鲁棒尺度不变特征(binary robust invariant scalable keypoints, BRISK)进行层级式聚类提高了识别算法的实时性。移动终端利用服务器返回的识别结果进行BRISK特征与光流算法结合的混合特征跟踪注册方法,并通过点集映射消除特征点漂移。利用前后帧信息以及关键帧信息减少跟踪抖动。UKbench标准图像库以及真实环境下的实验结果表明,虚实融合的智能导览效果良好。该原型系统已成功应用于上海电信体验馆等展馆智能导览系统。

关键词 智慧城市导览;移动增强现实;动态区域划分;层级式聚类;混合特征跟踪

中图法分类号 TP391

增强现实技术是计算机视觉研究的热点问题,它通过将计算机生成的虚拟信息精准地叠加在真实环境中,达到虚实融合的效果.在智慧城市的概念下,移动增强现实研究用户位置精确定位技术、智能移动终端姿态跟踪技术以及位置相关内容推送等关键技术,实现城市真实环境的时空导览以及信息多尺度展示.智慧城市导览融合移动互联网、物联网以及计算机视觉等多项技术,通过智能手机实现城市环境真实与虚拟的结合,如图1所示:



Fig. 1 Smart city guide using augmented reality.

图1 基于增强现实的智慧城市导览

Layar, Wikitude, CityLens 等流行应用丰富了用户通过增强的视角实现城市周边环境导览以及实景导航.然而单纯依赖 GPS,电子罗盘等纯硬件的位置定位技术并不能满足智慧导览的精度需求.基于视觉的增强现实技术精度能够达到“亚像素”级别,众多研究人员展开了相关技术研究.希腊研究人员^[1]开发了基于增强现实的文化古迹导游系统“Archeoguide”,用于雅典古神庙的漫游导览.北京理工大学开发了基于增强现实的圆明园数字重建系统^[2]以及博物馆移动智能导览系统^[3].由于户外环境复杂,智能移动终端性能有限,相关系统仅限于某一特定工作范围.

实现城市级智慧城市导览,需要解决海量场景识别定位以及实时鲁棒跟踪注册问题.移动终端计算性能以及资源存储能力有限,需要借助服务器实现海量图像检索以及增强数据处理.采用客户端/服务器(client/server, C/S)架构的智慧城市导览系统,选择合理的图像识别算法和控制网络传输的文件大小对系统流畅度以及精度影响较大.

主流图像识别算法包含基于线性搜索的特征匹配式识别算法^[4]以及基于机器学习的特征分类识别算法^[5],其中特征分类算法采用机器学习的方法通过有监督学习将图像识别算法转换为图像特征分类

问题,适合海量场景识别,例如 SVM^[6]和 Random Ferns^[7].线性搜索策略计算代价较大,不适合海量识别系统.Nister 等人^[8]提出的词汇树算法(vocabulary tree, VT)也是一种基于特征分类的场景识别算法,借助文本检索的思想将提取的图像尺度不变特征(scale invariant feature transform, SIFT)局部特征进行层级式聚类完成特征分类,具有较高的识别速度和识别精度.

本文采用的 C/S 架构对于网络传输提出比较高的要求.传统方法直接发送 JPG 格式视频帧进行服务器识别,特征提取在服务器端进行. JPG 图像一般较大,连续发送 JPG 图像的方式比较耗时,同时也浪费用户的流量.文献^[9]在移动端提取加速鲁棒特征(speed up robust features, SURF)^[10],将经编码的描述符传输至服务器端进行量化识别.实时 SURF 特征提取对于移动终端性能要求较高,实时性较差.近年来提出的二进制局部特征提取方法相比 SURF 特征具有较高的实时性,然而二进制局部特征并不适合采用层级式聚类的词汇树等海量图像检索系统.文献^[11]已经证明二进制向量在低维汉明空间存在边缘密集效应,另一方面二进制向量聚类也无法求出理想的聚类中心.

移动终端的实时跟踪注册一直是计算机视觉的研究热点问题.由于网络传输时延,无法通过服务器实时计算移动终端当前姿态.文献^[12]提出的基于即时定位与地图构建(simultaneous localization and mapping, SLAM)^[13]的即时地图构建与定位算法 PTAM 以及文献^[14]提出的多目标跟踪方法能够适用于环境可控的室内,在户外环境变化较大的情况下容易失去鲁棒性.与我们工作最为接近的是文献^[15-16]提出的 SIFT 与光流结合的方法,该方法通过 SIFT 特征提取初始化,后续帧连续进行光流计算估计摄像机姿态,直至光流算法失效才重新开始初始化.虽然采用并行执行的策略,由于 SIFT 特征速度较慢以及光流算法控制初始化的方法本身存在较大的累积误差,上述系统容易发生抖动.在复杂的室外环境下进行实时跟踪注册需要更加鲁棒的跟踪注册算法.

针对上述算法存在的问题,本文选择较高维的二进制鲁棒尺度不变特征(binary robust invariant scalable keypoints, BRISK)^[17]特征在服务器端进行层级式聚类,以汉明空间概率中心向量作为聚类中心,采用扩展的汉明距离衡量特征差异.移动终端采用 BRISK 特征提取与光流相结合的并行执行方法

进行混合跟踪注册,系统整体执行流程如图 2 所示,主要工作如下:

1) 设计了一种对二进制特征进行层级式聚类(B-HKM)的方法,并对词汇树(VT)检索结果进行快速几何变化相似度打分(GV)重排,提高了海量识别系统的实时性。

2) 采用 BRISK 特征提取与光流算法并行执行的方法进行实时鲁棒跟踪注册,利用 BRISK 特征提

取定期重新初始化以及点集映射消除跟踪漂移误差,利用前后帧信息与关键帧信息减少跟踪过程中的抖动。

3) 我们在重庆、南京、北京等地采集了大量 GPS 标注的建筑数据(POI)供研究,采用动态分区的方法缩小图像检索范围。另外采集了部分校园建筑图像以及 GPS 数据,在真实环境下进行场景识别以及跟踪注册算法测试。

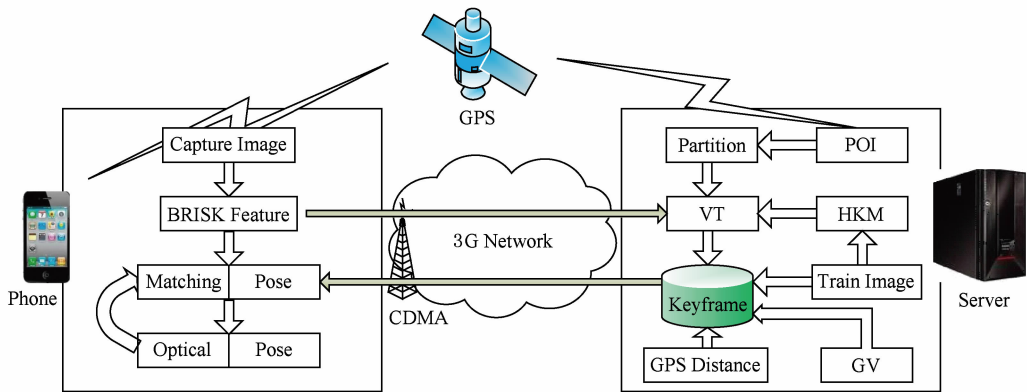


Fig. 2 The framework of mobile augmented reality.

图 2 移动增强现实系统框图

1 动态区域划分

进行城市级智能导航,单个城市待识别场景图像数量可达百万级以上。Nokia^[18]采用无区域划分的城市导航策略,在词汇树检索结果重排阶段剔除 GPS 距离较远的识别结果。然而国内建筑较国外相似度极高,无区域划分的方式需要返回大量的 Top-K 检索结果。同时随着检索结果的增多,检索结果重排阶段的资源占用以及时间代价增大。部分研究人员提出动态加载当前 GPS 位置周边区域建筑特征的策略,当前帧图像与加载图像特征进行实时特征匹配。该方法相对于词汇树检索算法能够大量减少单用户图像检索范围。然而随用户数目增多,服务器内存占用急剧上升,不适合高并发访问系统。本文提出一种区域划分的策略,分区训练词汇树能够缩小检索范围并对用户并发访问具有较好的效果。热点信息(POI)在分区结构中存储形式如图 3 所示。区域划分策略需要考虑热点信息(POI)分布不均、分区边缘设置等问题。POI 分布不均,单纯通过地理位置网格均匀划分的方法导致部分分区的 POI 数量偏少,对于整体词汇树相关阈值的确定以及资源分配效果不好。

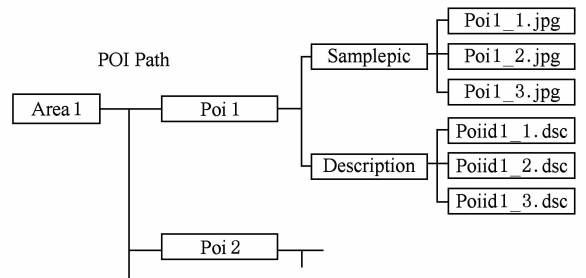


Fig. 3 The structure of area partition.

图 3 区域划分结构

本文提出一种动态分区策略,设定分区分配表,在均匀地理位置区域划分的基础上分区边界向 POI 密度较低的区域偏移,直至分区密度大致趋近于整体均值。由于 GPS 精度有限以及观察者位置与建筑实际位置的差异,导致分区边缘容易发生分区错误问题,如图 4 所示。基于用户手机 GPS 位置测定应属于分区 2,而此时位于分区 1 的热点 2 很容易被误认为属于分区 2。本系统采用的解决策略是将边缘区域的特征分别加入两边的词汇树进行查询。当用户位置处在分区边缘时,在线查询比较两分区的词汇树打分结果,将两边结果加入 Top-K 重排。如果分区错误,则通过词汇树打分阈值以及重排几何约束剔除。

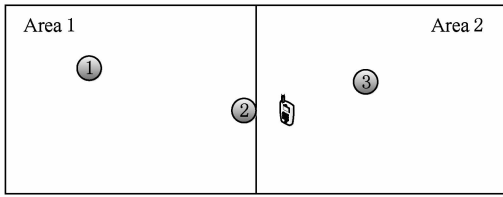


Fig. 4 Error of partition edge.

图4 分区边缘错误分区情况

词汇树采用 TF-IDF 模型量化训练图像各特征在词汇树中分布情况. 词汇树各节点权值计算如式(1)所示:

$$t_{i,j} = m_{i,j} \times \lg \frac{N}{n_i}, \quad (1)$$

$m_{i,j}$ 为视觉单词 i 在图像 j 中出现次数, N 为训练图像总数, n_i 为包含视觉单词 i 的训练图像数目.

训练图像 $\mathbf{T}_j = [t_{0,j}, \dots, t_{i,j}, \dots]$ 与查询图像 $\mathbf{Q} = [q_0, \dots, q_i, \dots]$ 经词汇树量化, 其相似度测量作为词汇树打分, 如式(2)所示:

$$\text{Score}[j] = \sum (t_{i,j} - q_i)^p. \quad (2)$$

热点 2 在分区 2 中无训练图像, 故其相比分区 1 具有较低的相似度, 该结果可以被剔除.

2 基于 BRISK 特征的层级式聚类

BRISK 是一种二进制局部特征提取算法, 对外光照、尺度以及视角变化具有一定的鲁棒性. 相比传统的 SIFT^[19], SURF 特征提取算法具有较高的实时性, 适合在资源有限的移动终端进行快速特征提取. 移动终端将提取的 BRISK 特征传输至服务器端, 服务器采用基于层级式聚类的词汇树算法进行快速识别.

2.1 汉明空间中心向量

K 均值聚类要求把一堆二进制特征向量分 K 份, 每份用一个均值向量表示, 也就是所谓的聚类中心, 而且要保证各向量具有良好的区分性. 传统实值向量如 SIFT, SURF 描述的差异性采用欧氏距离衡量, 如式(3)所示:

$$\min \sum_{i=1}^N d_{\text{euc}}(\bar{\mathbf{X}}, \mathbf{X}_i) = \min \sum_{i=1}^N \sum_{j=1}^L (\bar{x}_j - x_{i,j})^2, \quad (3)$$

其中, L 为向量维数, N 为向量总数. 均值向量到各向量距离和最小, 如式(4)所示:

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{i,j}, \quad j = 1, 2, \dots, L. \quad (4)$$

汉明空间一般采用汉明距离衡量二进制特征差异, 如式(5)所示. 若保证中心向量到各向量汉明距离和最小, 则由可知中心向量由各维统计结果确定,

如式(6)和式(7)所示.

$$\min \sum_{i=1}^N d_{\text{ham}}(\bar{\mathbf{X}}, \mathbf{X}_i) = \min \sum_{i=1}^N \sum_{j=1}^L (\bar{x}_j \oplus x_{i,j}); \quad (5)$$

$$N_j^1 = \sum_{i=1}^N x_{i,j}, \quad N_j^0 = \sum_{i=1}^N (\sim x_{i,j}); \quad (6)$$

$$\bar{x}_j = \begin{cases} 0, & N_j^0 > N_j^1; \\ 0, 1, & N_j^0 = N_j^1, \quad j = 1, 2, \dots, L; \\ 1, & N_j^0 < N_j^1. \end{cases} \quad (7)$$

文献[11]已经证明汉明空间存在密集边缘特性, 若空间维数为 L , 汉明空间两二进制向量汉明距离为 D , 则到两向量距离相等的向量至少包含 $C_D^{D/2} \times 2^{L-D}$ 个, 占空间总比例如式(8)所示:

$$\frac{C_D^{D/2} \times 2^{L-D}}{2^L} = \frac{C_D^{D/2}}{2^D}. \quad (8)$$

式(8)趋近于 $\sqrt{\frac{2}{\pi D}}$, 曲线如图 5 所示, 当 $D=2$ 时, 50% 的空间向量到距离为 D 的两随机向量距离相等. 因此要保证聚类中心的区分性, 汉明距离不能过低.

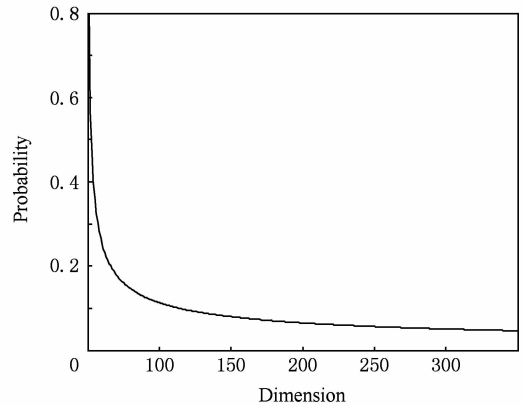


Fig. 5 Proportion of equal distance vector in Hamming space.

图5 等距向量占汉明空间比例

2.2 扩展汉明距离

通过汉明距离中心无法精确衡量该维特征分布情况, 特别是当 N_j^1, N_j^0 比较接近时, 汉明距离中心也不具有明显的区分性, 直接进行层级式聚类效果不好. 我们采用概率中心与词汇树节点进行关联, 替代原始二进制中心向量, 词汇树聚类中心如式(9)所示:

$$(P(\bar{x}_1), P(\bar{x}_2), \dots, P(\bar{x}_j), \dots, P(\bar{x}_L)), \quad (9)$$

其中:

$$P(\bar{x}_j=0) = N_j^0/N, \quad (10)$$

$$P(\bar{x}_j=1) = N_j^1/N. \quad (11)$$

则扩展的汉明距离量化特征的方法如式(12)所示:

$$d_{\text{e-ham}}(\bar{\mathbf{X}}, \mathbf{X}_i) = \sum_{j=1}^L [P(\bar{x}_j=1)(x_{i,j} \oplus 0) + P(\bar{x}_j=0)(x_{i,j} \oplus 1)]. \quad (12)$$

通过概率中心能够有效地量化特征分布情况, 由于采用 256 维的 BRISK 特征, 也能够有效避免边缘密集特性。

2.3 快速几何一致性重排

场景图像存在相似结构以及冗余信息都会导致误匹配, 需要对词汇树检索结果进行几何校验。通常的几何一致校验方法, 例如 RANSAC, PROSAC^[20] 常用于特征匹配后剔除野点。在词汇树检索系统里, 采用的策略是对候选识别结果进行基于最近邻搜索的特征匹配。然后通过 RANSAC 或者 PROSAC 进行几何一致性校验。结合基于线性搜索的特征匹配和几何一致性校验的方法时间消耗比较长, 因此需要限制返回检索结果数目, 否则影响系统实时性。

我们利用 GPS 信息剔除地理位置过远的候选图像, 然后借鉴文献[21]的思想, 采用几何变换相似度打分来代替 RANSAC, 并改进了文中重排直方图权值的计算方法, 取得较好的时间性能以及重排结果。通过两张 640×480 标准图像获取 500 个初始匹配点集, 分别采用 RANSAC, PROSAC 以及本文提出的 GV Score 进行几何一致校验, Intel Xeon 3.6 GHz 4 核处理器下几何一致性打分耗费时间对比如图 6 所示:

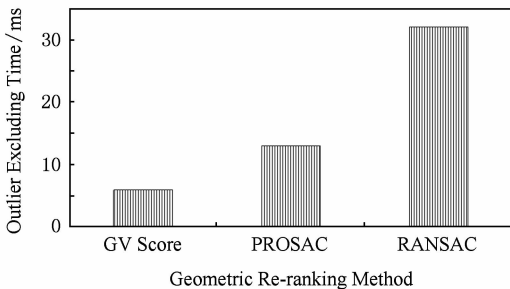


Fig. 6 The time cost of fast geometric re-ranking.

图 6 快速几何重排打分时间效率

几何变换相似度打分方法通过词汇树量化过程寻找查询图像与候选图像的匹配点对。依据是相似的特征描述在词汇树中具有相同的搜索路径。若某一节点有且仅有唯一的查询图像描述以及候选图像描述, 则把它们作为匹配点对。若传感器数据有效, 也可以用文献[22]的方法剔除部分具有相似对称结构的匹配点对。打分过程如图 7 所示, (A, a) , (B, b) 分别是查询图像与候选图像的两对匹配点对。分别计算查询图像以及候选图像上属于匹配点的两点位置坐标的欧氏距离。若匹配点对有 n 个, 则两图像各有 C_n^2 个点坐标距离。计算查询图像与候选图像对应点坐标距离的比值, 如式(13)所示:

$$\lambda = \lg \left(\frac{\sqrt{\|x_A - x_B\|^2 + \|y_A - y_B\|^2}}{\sqrt{\|x_a - x_b\|^2 + \|y_a - y_b\|^2}} \right). \quad (13)$$

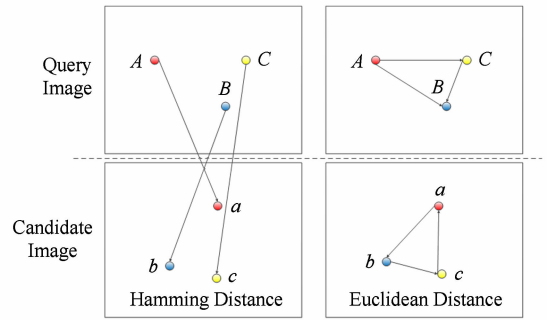


Fig. 7 The process of fast geometric re-ranking.

图 7 快速几何重排打分过程

我们统计具有相似比值的特征点, 如式(14)所示:

$$C(\alpha) = \sum I\left(\frac{\alpha}{c} \leq \lambda < \frac{\alpha+1}{c}\right), \quad (14)$$

为简化计算, 作者采用直方图统计距离比值 λ 。若比值落在 $\left[\frac{\alpha}{c}, \frac{\alpha+1}{c}\right)$ 的匹配点对数目最多, 则将其结果作为查询图像与候选图像几何相似度分数, 如式(15)所示:

$$Score = \max C(\alpha). \quad (15)$$

文献[22]中的方法按照均匀的点坐标距离比值间隔做直方图统计, 每一满足比例间隔约束的匹配点对在统计直方图中权值都是 1, 没有考虑匹配点对的描述差异误差。我们将满足比例间隔约束的点对权值设为式(14), 则统计直方图结果如式(16)所示:

$$C(\alpha) = \sum w \times I\left(\frac{\alpha}{c} \leq \lambda < \frac{\alpha+1}{c}\right), \quad (16)$$

$$w = \frac{k}{k + d_{\text{ham}}(A, a) + d_{\text{ham}}(B, b)}, \quad (17)$$

其中 $k > 0$ 为经验常数, $d_{\text{ham}}(A, a)$, $d_{\text{ham}}(B, b)$ 分别为匹配点对描述符的汉明距离。则汉明距离较远的匹配点对所对应的权值较小, 当且仅当两匹配点对汉明距离都为 0 时权值最大, 为 1, 也即此时匹配点对完全一致。

3 混合特征跟踪注册

复杂自然场景的六自由度跟踪注册是增强现实的核心关键问题。六自由度也即是摄像机坐标系相对于真实世界坐标系的旋转平移变化。移动端获取到精确的关键帧后需要实时计算摄像头相对于当前场景的六自由度姿态, 姿态计算算法实时性以及鲁棒性直接增强信息叠加流畅性以及准确性。

BRISK 特征提取算法具有较高的实时性。然而在低功耗的移动设备下, 采用 BRISK 特征进行特征

匹配以及姿态计算,实时性仍然难以满足用户需求.本文采用 BRISK 特征提取与光流跟踪并行执行的混合特征跟踪注册方法.

光流是空间运动物体在成像平面上像素运动的瞬时速度,光流计算利用图像序列中像素在时间域以及前后帧之间的对应关系计算出相邻帧之间物体的运动信息.假设图像相邻帧之间亮度守恒.光流算法计算速度较快,能够快速估计姿态,但是存在累积误差.需要通过 BRISK 特征提取的方式定期初始化,消除累积误差.然而,BRISK 特征提取需要消耗一定时间.当 BRISK 特征提取线程开启至特征提取完毕期间,在线程开启帧提取的特征点位置已经发生偏移,需要将特征点映射至 BRISK 特征提取结束帧.映射方法如式(18)所示:

$$m_{\text{stop}} = \mathbf{H}m_{\text{start}}, \quad (18)$$

其中 \mathbf{H} 为线程起始帧至结束帧的单应矩阵,可由光流算法求得, $\mathbf{H}_{\text{start}}$, \mathbf{H}_{stop} 分别为通过光流算法估计的两帧相对于关键帧的姿态,则 \mathbf{H} 可以通过式(19)精确求解:

$$\mathbf{H} = \mathbf{H}_{\text{stop}} \mathbf{H}_{\text{start}}^{-1}. \quad (19)$$

在光流跟踪过程中,每一帧均直接估计当前帧相对于关键帧的姿态,姿态估计稳定性较差,容易发生抖动.而通过前后帧求解姿态的方法存在累积误差,容易发生漂移.本文采用综合利用前后帧信息以及关键帧的方法,姿态估计包含前后帧信息 \mathbf{H}_i^{i-1} ,也包含前一帧相对于关键帧的预测位姿信息 \mathbf{H}_i^{i-1} ,则当前帧相对于关键帧的精确位姿信息 \mathbf{H}_i^w 可以由式(20)求解:

$$\mathbf{H}_i^w = \mathbf{H}_i^{i-1} \mathbf{H}_i^{i-1}, \quad (20)$$

同时计算当前帧相对于关键帧的预测位姿信息 \mathbf{H}_i^i ,为下一帧计算作准备.通过综合前后帧信息以及关键帧的方法能够有效保证跟踪注册算法的鲁棒性.

4 实验与分析

实验配置为服务器采用 Intel Xeon 3.6 GHz 4 核处理器,移动终端为双核 1.5 GHz Android 智能手机.户外实验采用 3G 网络传输图像数据.实验所用数据集为在重庆、南京、北京以及校园等地采集的经 GPS 标注的图像数据,包含 783 个不同场景,每个场景包含 50~100 张不同视角、光照等变化的图像,采用 640×480 图像分辨率以 JPG 格式存储.如图 8 所示.实验主要测试区域划分对于识别性能影响,基于概率中心的二进制层级式聚类(B-HKM)相对于传统欧氏距离中心层级式聚类(V-HKM)的时

间比较以及性能比较,测试本文识别系统与其他识别系统精度比较,最后对于跟踪注册算法鲁棒性进行测定.



Fig. 8 The GPS tagging image data sets.

图 8 GPS 标注的场景数据库

4.1 区域划分

我们对采集的数据通过 GPS 信息进行区域划分.测试所划分区域范围大小对各场景识别率的影响.测试过程中使用的是分支 $K=10$,深度为 $L=6$ 的词汇树算法.每个场景选择 20 帧的图像样本进行训练.对于待测试的场景,每个场景选择另外 20 张图像作测试.测试图像包含户外光照、遮挡、旋转、尺度等各种变化,我们计算 Top-10 识别率.测试结果分别如图 9 所示.我们将场景分成 1,2,4,8,16 个分区,岁分区数目增长,识别精度有明显提升,然而精度提升到一定程度时上升趋势不再明显.

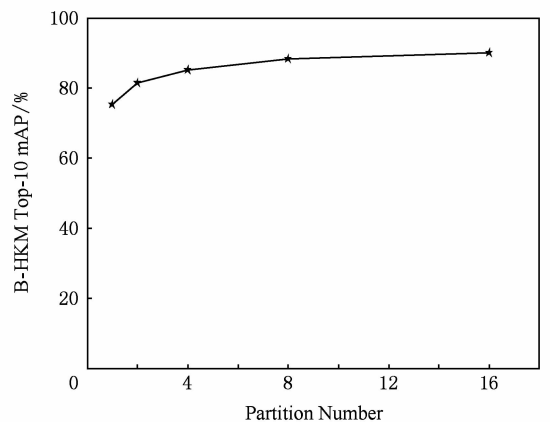


Fig. 9 The effect of partition numbers for Top-10 recognition rate.

图 9 分区数目对 Top-10 识别率影响

4.2 识别性能

采用 UKbench 标准数据库对本文提出的基于概率中心的二进制特征层级式聚类(B-HKM)方法进行有效性测试. UKbench 标准图像库包含2 550个不同场景,每个场景包含旋转、视角等不同变化. 主要比较基于概率中心的 B-HKM 方法与传统欧氏中心向量层级式聚类(V-HKM)的时间性能以及识别性能. 其中 B-HKM 采用 BRISK 特征提取的方法, V-HKM 采用 SURF 特征提取的方法. B-HKM 与 V-HKM 算法的时间性能对比如图 10 所示(查询图像均提取 500 个特征点,返回 Top-10 结果),测试在不同词汇树规模下 B-HKM 与 V-HKM 算法识别结果比较,如图 11 所示.

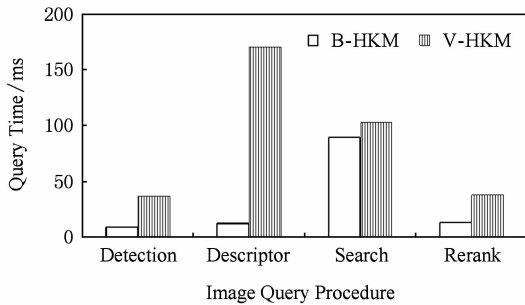


Fig. 10 The comparison of time cost between B-HKM and V-HKM.

图 10 B-HKM 与 V-HKM 各阶段时间对比

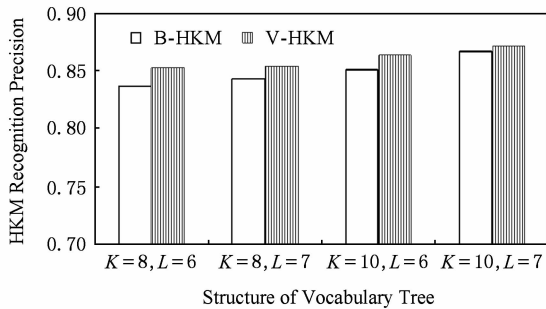


Fig. 11 The comparison of t precision between B-HKM and V-HKM.

图 11 B-HKM 与 V-HKM 不同词汇树规模下精度对比

由图 10 可见, B-HKM 的方法在时间性能上远胜于 V-HKM 的方法, 虽然图 11 可以看出 B-HKM 识别性能稍差, 但在较高的词汇树分支以及深度条件下性能差距并不明显.

与传统图像检索不同的是, 增强现实系统需要获取唯一且精确的关键帧信息. 本文结合了 BRISK 特征层级式聚类以及快速几何一致性打分, 在 UKbench 标准数据库以及我们自己在各地搜集的

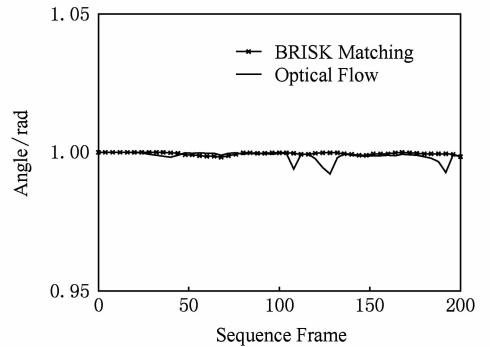
场景数据库测试综合识别精度(统计 Top-1 识别结果), 同最近主流算法对比结果如表 1 所示:

Table 1 The Recognition Rate of Recent Method

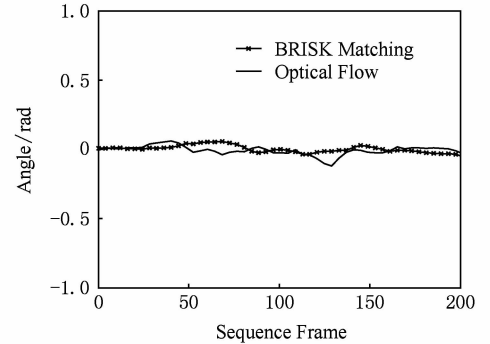
表 1 主流方法的识别率		%
Method	UKbench	Our Data Sets
Nister ^[8]	76.0	81.3
Jegou ^[23]	83.2	86.9
He ^[9]	85.9	89.7
Proposed	87.2	93.2

4.3 跟踪注册稳定性

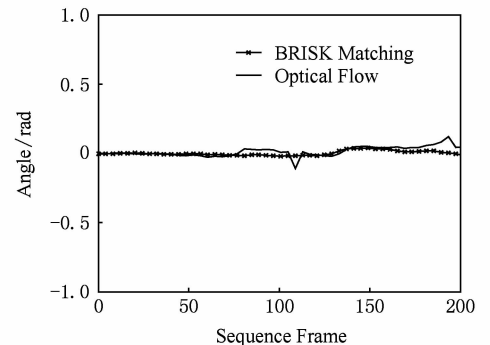
我们在校园环境结合 3G 网络实现混合特征跟踪注册, 跟踪算法主要测试在未开启初始化条件



(a) Rotation Matrix-x



(b) Rotation Matrix-y



(c) Rotation Matrix-z

Fig. 12 The robustness of rotation matrix.

图 12 旋转矩阵稳定性

下光流跟踪算法估计摄像机姿态的稳定性. 我们选择一段序列帧, 直接在线计算光流跟踪算法估计的旋转矩阵以及平移矩阵, 同时以离线计算相同帧相对于关键帧的姿态作为参考基准, 比较旋转矩阵变化如图 12 所示, 以及平移矩阵变化如图 13 所示. 通过实验结果可知, 光流跟踪算法在一定帧序列内能够保持较好的姿态计算正确率, 超过一定帧率会发生姿态漂移.

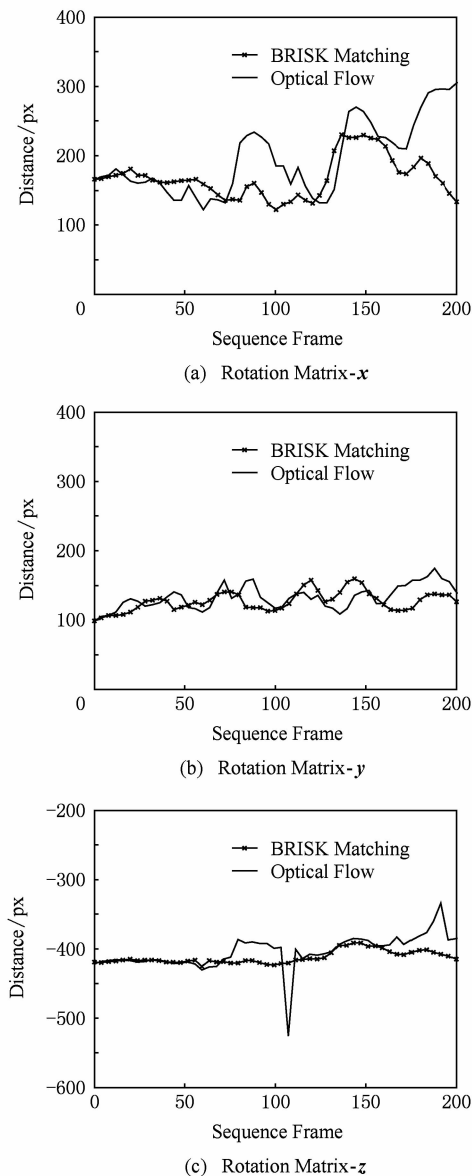


Fig. 13 The robustness of translation matrix.

图 13 平移矩阵稳定性

5 结 论

移动增强现实通过智能手机以虚实融合的形式实现智慧城市导览. 采用动态区域划分以及 B-HKM

聚类的大量场景识别算法能够实现用户位置精确定位. BRISK 与光流结合的混合特征跟踪注册方法提供给用户实时动态交互的全新体验, 位置相关信息以虚实融合的方式推送给用户, 满足用户多尺度、多时空的信息需求. 现阶段在城市环境开展大规模的智慧导览还面临许多问题, 高精度的数据采集以及图像后处理是亟待解决的问题. 另外现实环境的诸多变化, 类似场景遮挡、光照等也对移动增强现实系统提出了很大的挑战.

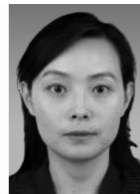
参 考 文 献

- [1] Vlahakis V, Ioannidis M, Karigiannis J, et al. Archeoguide: An augmented reality guide for archaeological sites [J]. Computer Graphics and Applications, 2002, 22(5): 52-60
- [2] Chen Jing, Wang Yongtian, Guo Junwei, et al. Augmented reality registration algorithm based on nature feature recognition [J]. Scientia Sinica Informationis, 2010, 40(11): 1437-1449 (in Chinese)
(陈靖, 王涌天, 郭俊伟, 等. 基于特征识别的增强现实跟踪定位算法[J]. 中国科学: 信息科学, 2010, 40(11): 1437-1449)
- [3] Zhong Zhipeng, Wang Yongtian, Chen Jing, et al. A museum guide system based on mobile visual search [J]. Journal of Computer-Aided Design & Computer Graphics, 2012, 24(4): 555-562 (in Chinese)
(钟志鹏, 王涌天, 陈靖, 等. 一个基于移动视觉搜索技术的博物馆导览系统[J]. 计算机辅助设计与图形学学报, 2012, 24(4): 555-562)
- [4] Fan Zhiqiang, Zhao Qiping. A data-clustering based robust SIFT feature matching method [J]. Journal of Computer Research and Development, 2012, 49(5): 1123-1129 (in Chinese)
(范志强, 赵沁平. 一种基于数据聚类的鲁棒 SIFT 特征匹配方法[J]. 计算机研究与发展, 2012, 49(5): 1123-1129)
- [5] Qin Lei, Gao Wen. Scene image categorization based on content correlation [J]. Journal of Computer Research and Development, 2009, 46(7): 1198-1205 (in Chinese)
(秦磊, 高文. 基于内容相关性的场景图像分类方法[J]. 计算机研究与发展, 2009, 46(7): 1198-1205)
- [6] Berg A, Zhang H, Maire M, et al. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition [C] //Proc of the 2006 IEEE Int Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2006: 2126-2136
- [7] Calonder M, Ozuysal M, Lepetit V, et al. Fast keypoint recognition using random ferns [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2010, 32(3): 448-461
- [8] Nister D, Stewenius H. Scalable recognition with a vocabulary tree [C] //Proc of the 2006 IEEE Int Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2006: 2161-2168

- [9] He Junfeng, Feng Jinyuan, Liu Xianglong, et al. Mobile product search with bag of hash bits [C] //Proc of the 19th ACM Int Conf on Multimedia. New York: ACM, 2011; 839-840
- [10] Bay H, Ess A, Tuytelaars T, et al. SURF: Speeded up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359
- [11] Trzcinski T, Lepetit V, Fua P. Thick boundaries in binary space and their influence on nearest-neighbor search [J]. Pattern Recognition Letters, 2012, 33(16): 2173-2180
- [12] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces [C] //Proc of the 6th IEEE and ACM Int Symp on Mixed and Augmented Reality. Piscataway, NJ: IEEE, 2007; 225-234
- [13] Davison J, Reid D, Molton D, et al. MonoSLAM: Real-time single camera SLAM [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067
- [14] Wagner D, Schmalstieg D, Bischof H. Multiple target detection and tracking with guaranteed framerates on mobile phones [C] //Proc of the 8th IEEE and ACM Int Symp on mixed and augmented reality. Piscataway, NJ: IEEE, 2009; 57-64
- [15] Lee T, Hoellerer T. Multithreaded hybrid feature tracking for markerless augmented reality [J]. IEEE Trans on Visualization and Computer Graphics, 2009, 15(3): 355-368
- [16] Taehee L, Hollerer T. Hybrid feature tracking and user interaction for markerless augmented reality [C] //Proc of Virtual Reality Conf. Piscataway, NJ: IEEE, 2008; 145-152
- [17] Leutenegger S, Chli M, Siegwart R. BRISK: Binary robust invariant scalable keypoints [C] //Proc of the 2011 IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2011; 2548-2555
- [18] Schindler G, Brown M, Szeliski R. City-scale location recognition [C] //Proc of the 2007 IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2007; 1-7
- [19] Lowe D. Distinctive image features from scale-invariant keypoints [J]. Int Journal of Computer Vision, 2004, 60(2): 91-110
- [20] Chum O, Matas J. Matching with PROSAC-progressive sample consensus [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2005; 220-226
- [21] Tsai S, Chen D, Takacs G, et al. Fast geometric re-ranking for image-based retrieval [C] //Proc of the 17th IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2010; 1029-1032
- [22] Kurz D, Himane B. Inertial sensor-aligned visual feature descriptors [C] //Proc of the 2011 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2011; 161-166
- [23] Jegou H, Douze M, Schmid C. Improving bag-of-feature for large scale image search [J]. Int Journal of Computer Vision, 2010, 87(3): 316-336



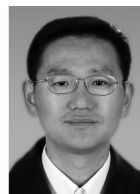
Zhang Yunchao, born in 1987. PhD candidate in Beijing Engineering Research Center for Mixed Reality and Novel Display Technology, School of Optics and Electronics, Beijing Institute of Technology. Student member of China Computer Federation. His research interests include augmented reality and virtual reality.



Chen Jing, born in 1974. PhD and associate professor in Beijing Institute of Technology. Member of China Computer Federation. Her main research interests include computer vision and augmented reality.



Wang Yongtian, born in 1957. Professor and PhD supervisor in Beijing Institute of Technology. His main research interests include new 3D display, virtual reality and augmented reality.



Liu Yue, born in 1968. Professor and PhD supervisor in Beijing Institute of Technology. His main research interests include medical image processing and augmented reality.