

基于城市级位置服务的移动增强现实

张运超¹, 陈 靖², 王涌天^{1,2}

(1.北京理工大学计算机学院 北京 100081; 2.北京理工大学光电学院 北京 100081)

摘 要: 基于位置服务的移动增强现实以虚实融合的方式呈现用户周围环境. 本文采用 C/S 架构实现基于视觉的移动增强现实系统, 解决手机处理速度以及存储空间不足的问题. 服务器端采用基于分布式聚类的词汇树算法进行城市级海量场景识别; 手机端采用 BRISK 特征提取与光流算法相结合的 6 自由度跟踪注册方式. 本文在 UKbench 以及 SUN397 数据库进行实验测试, 实验结果表明, 文中提出的分布式聚类算法能快速构建词汇树并有效解决单 PC 计算内存瓶颈; 移动终端通过 BRISK 特征提取以及并行流水线设计, 系统跟踪实时性良好.

关键词: 增强现实; 词汇树; 分布式聚类; 移动终端; 光流跟踪

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2014)08-1503-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.08.007

City-Scale Location Services Based on Mobile Augmented Reality

ZHANG Yun-chao¹, CHEN Jing², WANG Yong-tian^{1,2}

(1. School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China;

2. School of Optics and Electronics, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Mobile Augmented Reality system presents surroundings in a way of virtual-real fusion for location-based services. This paper employs C/S architecture for vision-based mobile AR system, which solves the problem of processing speed and storage space on mobile phone. Vocabulary tree algorithm which is based on distributed clustering algorithm is used for city-scale scene recognition on the server. BRISK feature and optical flow algorithm is used in combination for visual tracking on mobile devices. Experimental results on UKbench and SUN397 show that the proposed distributed clustering algorithm can quickly build vocabulary tree and solve the limitations of memory. Mobile terminals keep real-time tracking through extraction of BRISK feature and design of parallel pipeline.

Key words: Augmented Reality; vocabulary tree; distributed clustering; mobile terminals; optical tracking

1 引言

增强现实技术(Augmented Reality, 简称 AR)是计算机视觉领域研究的热点问题. 它通过将计算机生成的虚拟信息精准叠加在观察者周围真实环境中, 从而实现虚拟场景与用户所处真实环境完美融合. 近年来随着智能手机的普及, 基于地理位置服务(Location Based Service, 简称 LBS)相关应用层出不穷. 而结合移动 AR 的 LBS 应用以其虚实结合的自然呈现方式吸引了大量用户. 相比较 GPS, 电子罗盘等硬件方式, 计算机视觉方式实现 AR 场景识别和跟踪注册, 精度可达亚像素级别. 然而手机端资源存储以及计算性能不足, 场景识别精度和跟踪注册实时性面临极大挑战.

针对户外海量场景识别精度问题, 相关研究人员相继提出了 SVM, PCA 及 Ferns 等场景识别算法^[1~3]. 这些算法通常采用构建分类器的方法对特征进行有监督学习和分类. SVM 和 PCA 运行速度较慢, 难以满足实时性较高的 AR 系统. 而 Ferns 内存占用过高, 无法适用于海量场景识别. Nister^[4]借助文本检索思想将词汇树算法(Vocabulary Tree, 简称 VT)用于图像识别. 通过特征聚类的方式构建 VT 并利用倒排文件统计视觉单词在每张训练图像中出现频率. 在线阶段通过查询倒排文件能够快速精确识别图像.

为解决 AR 跟踪注册的实时性问题, Klein^[5]采用 Fast 特征快速构建地图实时跟踪注册. Wagner^[6]利用 Fast 以及 Modified SIFT 算法进行小范围多目标跟踪, 并

选择 PROSAC^[7] 剔除野点. 然而 Fast 特征不具有尺度和旋转不变性. Leutenegger 在 Fast 的改进算法 AGAST^[8] 基础上结合 BRIEF^[9] 特征描述提出了 BRISK^[10] 特征提取算法. BRISK 特征通过构建尺度金字塔和非最大值抑制实现尺度空间上特征点检测, 适用于户外大尺度变化环境, 并且相比传统的 SIFT^[11] 以及 SURF^[12] 等鲁棒特征提取算法具有良好的实时性.

国内, 哈工大^[13] 采用 SIFT 特征与隐马尔科夫模型结合的方式实现高效移动识别定位. 浙大研究团队^[14] 提出 SURF 特征与 KLT 跟踪并行执行的 AR 跟踪注册方法. 传统 VT 算法离线训练需要耗费大量时间和内存, 而上述移动 AR 跟踪注册研究工作大多集中于实验室环境, 适用于小范围可控环境. 户外场景复杂, 图像数据量大, 上述算法无法适用于城市级 LBS 的实时跟踪注册. 针对城市级 AR 系统存在的问题, 本文提出一个适用于海量户外场景的移动 AR 系统. 系统采用基于分布式聚类的 VT 算法在服务器端进行海量场景识别, 手机端采用 BRISK 特征提取与光流跟踪并行执行的跟踪注册算法, 同时利用汉明距离对初始跟踪点集进行筛选, 解决姿态计算中内点率不高造成的计算浪费.

2 系统设计与框架

移动设备数据存储量小, 处理能力弱, 但内置多种传感器, 易于携带, 方便显示, PC 工作站数据存储量大, 计算能力强, 借助分布式处理系统能够支持城市级 LBS 的场景识别需求. 综合两端各自的处理优势, 本文提出一种基于分布式 C/S 架构的移动 AR 系统实现 LBS. 借助 PC 工作站以及分布式处理系统作为后台云服务处理器, 通过基于分布式聚类的 VT 算法进行快速精确户外场景识别, 获取跟踪目标关键帧. 手机端采用 BRISK 特征提取与光流跟踪相结合的算法进行精准高效的六自由度跟踪, 该方法能够有效解决两者之间需求矛盾,

形成互补. 整体算法流程如图 1, 具体设计步骤如下:

服务器识别: 离线阶段, 训练图像通过初始聚类中心选择以及分布式聚类构建 VT 倒排文件, 另外存储 BRISK 特征关键帧. 在线阶段, 服务器接收移动端发送的视频帧, 通过 VT 算法识别之后返回识别结果 ID 以及对应的关键帧. **移动端跟踪:** 手机端接收来自服务器的关键帧, 通过 BRISK 特征进行跟踪初始化, 并采用光流算法加速计算姿态.

3 算法实现

3.1 基于分布式聚类的 VT 场景识别算法

VT 算法识别精度高, 速度快, 目前被广泛用于图像识别领域. 针对数据量较大的 LBS 户外场景识别, 本文采用分布式聚类对传统 VT 算法进行改进, 加快训练速度, 也解决了单 PC 内存受限问题. 对图像特征进行分布式 k-means 聚类, 其核心包括初始聚类中心选择以及特征点与聚类中心分布式计算.

VT 算法进行海量场景识别, 离线阶段, 首先通过图像特征聚类构建 VT, 训练图像通过其特征在 VT 中分布进行量化, 假设聚类分支数为 k , 树深度为 L , 则

$$d_j = [w_{1,j}, w_{2,j}, \dots, w_{t,j}] \quad (1)$$

d_j 表示第 j 个训练图像, t 为 VT 单词总数目 $\sum_{i=1}^L k^i$, $w_{i,j}$ 为第 j 个训练图像的特征在 VT 第 i 个节点上的权值, $w_{i,j}$ 采用 TF-IDF 模型计算

$$w_{i,j} = m_{i,j} \times \lg \frac{N}{n_i} \quad (2)$$

$m_{i,j}$ 为单词 i 在图像 j 中出现次数, N 为训练图像总数, n_i 为包含单词 i 的训练图像数目. 在线识别阶段查询图像也采用训练图像相同的量化策略, 其相似度测量为:

$$S(d_j, q) = \|d_j - q\|_p \quad (3)$$

3.1.1 初始聚类中心选择

聚类算法受初始值影响较大, 传统 k-means 聚类算法随机选择初始聚类中心, 容易造成聚类算法迭代次数过多, 难以收敛, 影响算法时间效率和性能. 为保证快速收敛, 初始聚类中心应具有较好区分性, 也即筛选出距离较远特征描述作为聚类中心. 记包含 i ($i = 0, 1, 2, \dots, k$) 个聚类中心的集合为 K_i , 其余特征 x 到聚类中心集合 K_i 的距离 $d(x, K_i)$ 为:

$$d(x, K_i) = \min_{y \in K_i} d(x, y), x \notin K_i, y \in K_i \quad (4)$$

特征 x 被选为初始聚类中心的概率为:

$$P(x \in K_i) \sim d^2(x, K_{i-1}) \quad (5)$$

即特征 x 到已存在聚类中心的最近距离越大, 被选择为聚类中心概率越大. 除第 1 个聚类中心随机选取外, 其余聚类中心由特征到已存在聚类中心集合的距离确定. 这样的选择策略能够保证聚类中心各特征之

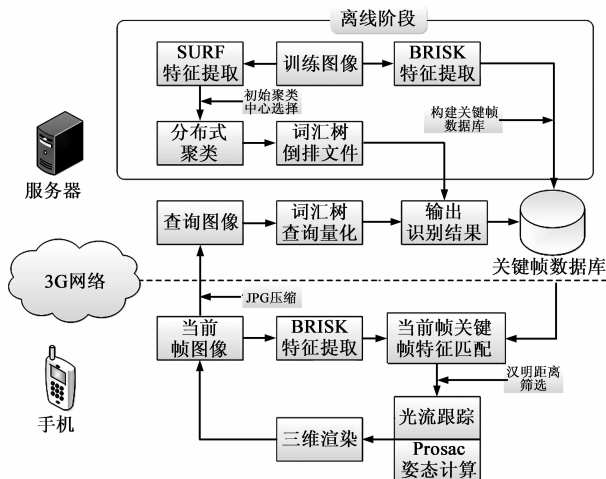


图1 AR导航整体流程图

间区分性较高,减少迭代次数,聚类效果良好.

3.1.2 分布式 k-means 聚类

训练图像特征分布式 k-means 聚类采用 Map/Reduce 思想^[15,16].由配置服务器(Configuration Server,简称 CS)根据映射规则划分训练图像数据,经由中心调度节点(Master)将计算任务分配给各分布式计算节点(Compute Unit,简称 CU),如图 2.

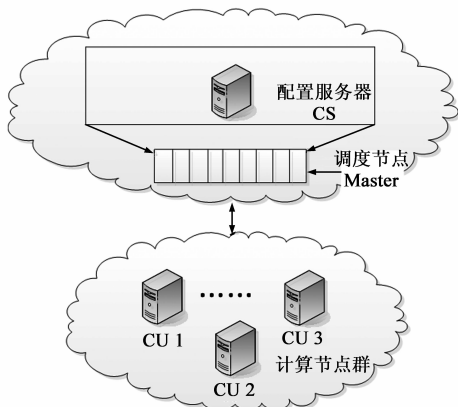


图2 分布式系统

VT 的构造采用深度优先遍历串行生成策略,即先沿某一分支构造 VT,待子树构造完毕再构造兄弟子树,如图 3.树的构造在专门的规约节点-中心计算节点(CenterCU)上进行.每个树节点只存储自身聚类中心,节点编号采用广度优先方式,有 *id* 字段存储自身的编号.

VT 的生成采用边生成,边构造,边序列化的方式.由于采用深度优先串行聚类策略,所以构造 VT 的过程中,CenterCU 在内存中分配能够容纳 60 个树节点($K = 10, L = 6$)的数组,用以保存聚类过程树节点的中间结果.每次迭代,聚类都将节点所在数组下标一并发送给 CU,每个 CU 进行分布式聚类计算,获取每个 CU 各自的聚类中心,结果以各 CU 均值的形式返回 CenterCU 时,更新对应数组下标的值,直到算法收敛,此时聚类成功,对应数组下标值不再改变,即最终生成了树的节点.

3.2 基于 BRISK 特征的跟踪注册算法

考虑到进行 LBS 跟踪注册时,观察者一般离建筑等户外场景距离较远,可将场景近似为平面.通过当前帧与关键帧的特征匹配点对集合即可建立两者之间单

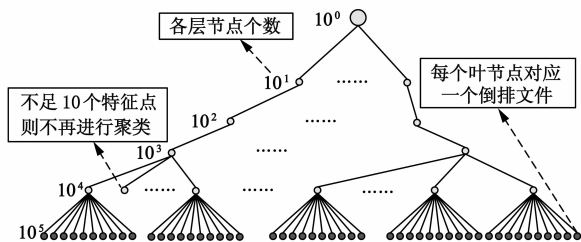


图3 VT层级式聚类过程

应关系,如式(6),通过单应关系可计算跟踪过程中摄像机的姿态变换:

$$Sm_c = H_k^c m_k \quad (6)$$

(m_c, m_k)为一对 2D 特征匹配点对,分别位于当前帧和关键帧上.随机选择四对不共线匹配点对即可求解初始单应矩阵 H_k^c ,通过初始单应矩阵求解所有匹配点对的反投影误差 d

$$d = \|m_c - H_k^c m_k\| \quad (7)$$

选择最多个反投影误差 d 小于某个给定阈值的初始单应矩阵 H_k^c 作为当前帧与关键帧的单应矩阵,并通过 L-M 优化算法求解最终的单应矩阵.

$$\min \|m_c - H_k^c m_k\| \quad (8)$$

初始匹配点对的确定对于跟踪注册算法至关重要,是求解当前帧与关键帧单应关系的基础.本文采用 BRISK 特征提取算法进行跟踪注册,并采用光流算法加速计算.算法主要包括跟踪初始化以及光流跟踪两部分,并采用并行处理策略保证系统流畅.

3.2.1 基于汉明距离筛选的跟踪初始化

BRISK 特征采用二进制描述,特征之间区分性采用汉明距离衡量,即两特征描述分别为 $x = (x_1, x_2, \dots, x_k), y = (y_1, y_2, \dots, y_k), x_i, y_i \in \{0, 1\}$ 则其汉明距离为:

$$d(x, y) = \sum_{i=1}^k x_i \oplus y_i \quad (9)$$

当前帧与关键帧进行特征匹配,通常能够获取大量匹配点对.检测两幅图像特征匹配后匹配点对的汉明距离 $d(x, y)$,图 4(a)为 PROSAC 几何一致校验后内点对的汉明距离分布统计直方图,图 4(b)为剩余匹配点对的汉明距离分布统计直方图.对比可知,汉明距离与内点密切相关.

3.2.2 光流跟踪

本文利用光流算法预测跟踪点集在序列帧中位置.光流场的计算基于图像像素守恒的基本假设.假定 t 时刻图像坐标 (x, y) 处灰度值为 $I(x, y, t)$, $t + \Delta t$ 时刻该点运动至点 $(x + \Delta x, y + \Delta y, t + \Delta t)$ 处,灰度值变为 $I(x + \Delta x, y + \Delta y, t + \Delta t)$,由图像像素守恒的基本假

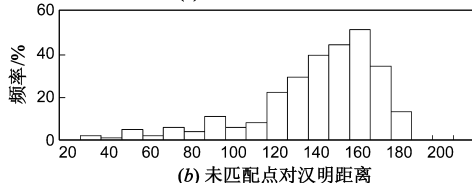
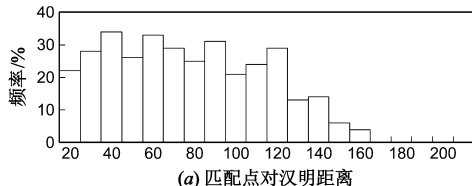


图4 匹配点对与未匹配点对汉明距离对比

设,知

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) \quad (10)$$

由忽略高次项的泰勒展开即可得光流约束方程

$$\frac{\partial I}{\partial x} \frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y} \frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t} = 0 \quad (11)$$

通过对约束方程的求解即可预测跟踪点集的序列帧位置坐标.

3.2.3 基于点集映射的并行流水线

为保证跟踪算法流畅进行,设计并行流水线,重新初始化与光流跟踪并行执行,如图 5 所示:

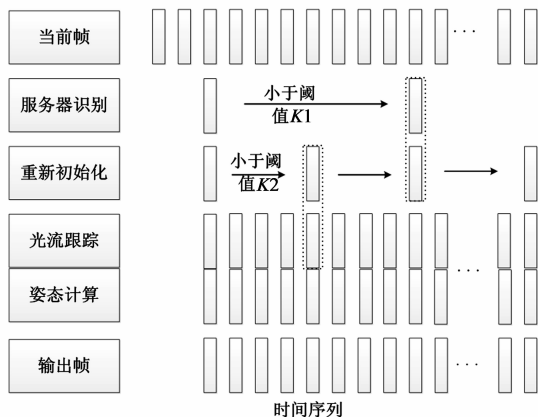


图5 跟踪并行流水线

初始阶段,当前帧捕获图像发送至服务器进行场景识别,继而当前帧与返回的识别关键帧进行跟踪初始化过程,通过特征匹配确定初始跟踪点集并计算姿态,后续帧采用光流算法预测跟踪点集位置并计算姿态.通过特征匹配进行初始化需要消耗一定时间,特征匹配完成后,初始帧提取的特征点位置已经发生偏移,需要通过点集映射将初始帧提取的特征点映射至匹配完成帧,如式(12).

$$m_{over} = H m_{begin} \quad (12)$$

点对 (m_{begin}, m_{over}) 之间的映射关系通过单应矩阵 H 确定.初始帧以及完成帧的姿态可以通过光流算法估计, H_{begin} 以及 H_{over} 分别是两帧图像相对于关键帧的单应矩阵.则由式(13)即可求解点集映射单应矩阵.

$$H = H_{over} H_{begin}^{-1} \quad (13)$$

4 实验结果及分析

实验配置为服务器采用 Intel Xeon E3 处理器,内存 16GB.手机端为双核 Android 智能手机,配有 3G 网卡.采用 15 个计算节点分布式计算平台.本实验将从识别算法性能以及跟踪算法实时性方面进行实验验证.

4.1 识别性能

实验主要测试初始聚类中心选择算法以及分布式 K-means 聚类算法的时间效率和内存占用.

4.1.1 识别精度

本文采用 Ukbench 图像库测试海量场景识别精度,分别测试传统聚类,初始聚类中心选择聚类以及分布式聚类等方法的时间和识别率.图 6 为不同分支 K 及树深 L 下层级式聚类花费时间,图 7 为 VT 算法识别返回 Top5 识别率.实验结果表明,初始聚类中心选择算法以及分布式聚类算法能够在保证图像识别精度的前提下大幅减少聚类花费时间.

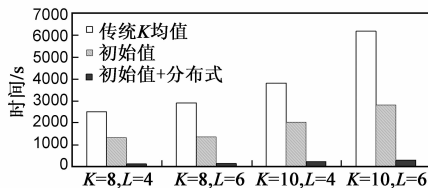


图6 不同聚类算法聚类时间对比

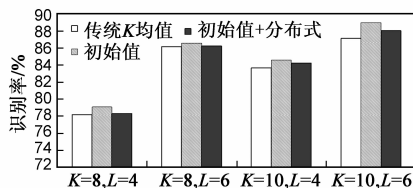


图7 不同聚类算法的识别率对比

4.1.2 聚类算法内存和时间开销

采用 SUN397 数据库^[17]测试分布式 k-means 算法时间效率和内存占用.图 8 为采用 SUN397 数据库测试单 PC 聚类以及 15PC 分布式聚类在不同场景规模下所消耗时间,图 9 为单 PC 聚类内存占用情况.

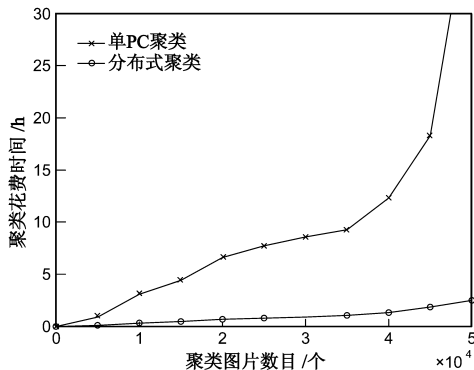


图8 分布式与单PC聚类时间对比

当训练图像数目超过一定数量以后,参与聚类图像占用内存已超过单 PC 最大内存,需要不停进行 IO 操作,耗费大量时间.通过分布式聚类算法能够提升训练速度,同时也能够有效地解决单 PC 聚类大量图像内存不足的问题.

4.2 跟踪性测试

特征跟踪测试采用牛津大学图像库^[9].跟踪初始化过程中初始跟踪点集内点率低,并且影响姿态计算

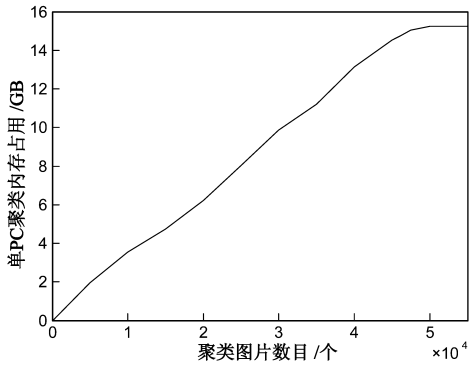


图9 单PC聚类内存占用

时间效率.通过对匹配点对的汉明距离进行筛选,测试在不同汉明距离误差约束条件下图像匹配点数以及内点数变化情况.图 10 可以看出,随着匹配点对的汉明距离误差约束增大,匹配点对增多,而内点数在增大到一定程度后迅速下降.结果表明,汉明距离筛选能够保证内点率,限定汉明距离误差对内点数目无影响,并且有助于加速姿态计算.

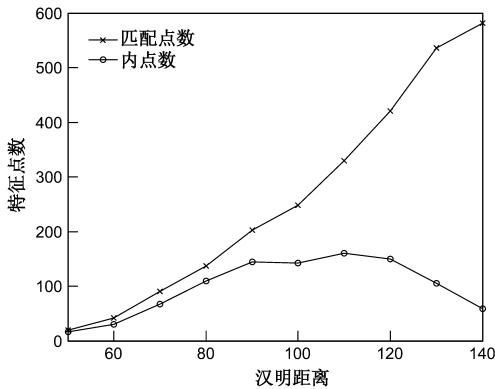


图10 匹配点数以及内点数随汉明距离筛选变化

选择 SIFT, SURF 以及 BRISK 等尺度不变特征对跟踪算法实时性进行对比分析.为验证点集筛选的有效性,同时对非筛选点集表 1 以及筛选点集表 2 两种情况进行实验.结果表明,点集筛选对于系统实时性提升明显, BRISK 特征在移动端进行跟踪初始化时间效率相比 SIFT, SURF 等特征提升 10 倍左右.

表 1 不进行点集筛选的初始化/ms

提点	描述	匹配	PROSAC	总时间
SIFT	8925	268.1	129.17	10859
SURF	501.1	181.3	73.54	850.3
BRISK	30.31	35.12	27.61	107.7

4.3 应用

校内环境下构建基于 AR 的 LBS 导览系统.在校园周边采集 113 个不同建筑视频序列.视频包含不同视

表 2 进行点集筛选的初始化/ms

提点	描述	匹配	PROSAC	总时间	
SIFT	1537	3512	169.7	61.23	5279
SURF	94.31	201.3	120.7	41.17	457.4
BRISK	14.67	15.1	20.3	11.61	61.68

角,尺度以及旋转变化.每个建筑截取 10 帧图像做训练,20 帧图像做查询测试,通过分布式聚类算法进行离线训练.智能手机在校园环境通过 3G 信号与分布式服务器进行通信,完成基于 AR 的 LBS 导览,效果如图 11.包含传输时间在内的在线识别时延低于 800ms,跟踪注册帧率可达 21fps.跟踪过程中基本无较大抖动.



图11 校园环境移动AR系统

5 结论

本文采用基于分布式聚类的 VT 算法实现户外海量场景识别,在保证识别精度的前提下实现了 VT 的快速训练构建,大幅提高了训练时间效率,并解决了聚类海量图像过程中的单 PC 内存受限问题.采用 BRISK 特征提取算法结合光流跟踪实现手机端在户外场景下实时的六自由度跟踪,通过点集映射消除初始化过程中跟踪点集的漂移,跟踪过程基本无抖动,本文通过并流水线的设计保证了系统运行的实时性.

参考文献

- [1] Berg A, Zhang H, Maire M, et al. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition [A]. IEEE International Conference on Computer Vision and Pattern Recognition Proceedings [C]. San Diego: IEEE, 2006. 2126 - 2136.
- [2] Zisserman A, Bosch A, Munoz X. Representing shape with a spatial pyramid kernel [A]. ACM International Conference on

- Image and Video Retrieval Proceedings [C]. Amsterdam: Springer, 2007. 401 – 408.
- [3] Calonder M, Ozuysal M, Lepetit V, et al. Fast keypoint recognition using random ferns [J]. IEEE Trans on PAMI, 2010, 32 (3): 448 – 461.
- [4] Nister D, Stewenius H. Scalable recognition with a vocabulary tree [A]. IEEE International Conference on Computer Vision and Pattern Recognition Proceedings [C]. New York: IEEE, 2006. 2161 – 2168.
- [5] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces [A]. 6th IEEE and ACM International symposium on mixed and augmented reality Proceedings [C]. Nara: IEEE, 2007. 225 – 234.
- [6] Wagner D, et al. Multiple target detection and tracking with guaranteed framerates on mobile phones [A]. 8th IEEE International symposium on mixed and augmented reality Proceedings [C]. Orlando: IEEE, 2009. 57 – 64.
- [7] Chum O, Matas J. Matching with PROSAC—progressive sample consensus [A]. IEEE conference on computer vision and pattern recognition Proceedings [C]. San Diego: IEEE, 2005. 220 – 226.
- [8] Mair E, Hager G, Burschka D, et al. Adaptive and generic corner detection based on the accelerated segment test [A]. European Conference on Computer Vision Proceedings [C]. Heraklion: Springer, 2010. 183 – 196.
- [9] Michael Calonder, Vincent Lepetit, Christoph Strecha, et al. BRIEF: binary robust independent elementary features [A]. European Conference on Computer Vision Proceedings [C]. Heraklion: Springer, 2010. 778 – 792.
- [10] Leutenegger S, Chli M, Siegwart R. BRISK: binary robust invariant scalable keypoints [A]. IEEE International Conference on Computer Vision Proceedings [C]. Barcelona: IEEE, 2011. 2548 – 2555.
- [11] D. G. Lowe. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.
- [12] Bay H, Ess A, Tuytelaars T, et al. SURF: speeded up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346 – 359.
- [13] 陈凤东, 洪炳荣. 基于特征地图的机器人全局定位与自主泊位方法 [J]. 电子学报, 2010, 38(6): 1256 – 1261. Chen Fengdong, et al. A global localization and self-docking method for mobile robot based on feature map [J]. Acta Electronica Sinica, 2010, 38(6): 1256 – 1261 (in Chinese)
- [14] 李扬, 孙超, 张明敏, 张文杰, 潘志庚. 跟踪与匹配并行的增强现实注册方法 [J]. 计算机辅助设计与图形学学报, 2011, 16(4): 680 – 685.
- Li Yang, Sun Chao, et al, Pan Zhigeng. Parallel tracking and matching for natural feature based AR registration [J]. Journal of Computer-Aided Design & Computer Graphics, 2011, 16 (4): 680 – 685. (in Chinese)
- [15] 黄龙涛, 等. 基于 MapReduce 的并行 Web 服务自动组合 [J]. 电子学报, 2012, 40(7): 1397 – 1403. Huang Longtao, et al. Automatic service composition in parallel with MapReduce [J]. Acta Electronica Sinica, 2012, 40 (7): 1397 – 1403. (in Chinese)
- [16] 杨际祥, 等. 一种大规模分布式计算负载均衡策略 [J]. 电子学报, 2012, 40(11): 2226 – 2231. Yang Jixiang, et al. A load balancing strategy for large-scale distributed computing [J]. Acta Electronica Sinica, 2012, 40 (11): 2226 – 2231. (in Chinese)
- [17] Xiao J, et al. SUN database: Large-scale scene recognition from Abbey to Zoo [A]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings [C]. San Francisco: IEEE, 2010. 3485 – 3492.

作者简介



张运超 男, 1987 年出生, 博士, CCF 学生会员, 主要研究方向为增强现实、模式识别.

E-mail: zhangyunchao163@163.com



陈靖(通信作者) 女, 1974 年出生, 副研究员, 主要研究方向为增强现实与虚拟现实、计算机视觉、模式识别.

E-mail: chen74jing29@bit.edu.cn



王涌天 男, 1957 年出生, 教授, 主要研究方向包括光学系统设计和 CAD、新型光学元件、虚拟现实和增强现实技术.