

一种基于视觉即时定位与地图构建的改进增强现实注册方法

郑国强^{1*}, 周治平^{1,2}

¹江南大学物联网工程学院, 江苏 无锡 214122;

²江南大学物联网技术应用教育部工程研究中心, 江苏 无锡 214122

摘要 在无标记增强现实中, 基于视觉即时定位与地图构建, 结合 ORB 特征提取算法和顺序抽样一致性, 提出了一种改进的注册方法。利用相似函数衡量匹配样本点对之间的差异性, 从具有最高品质函数的数据子集中抽取样本, 得到较优的匹配点对, 完成了三维地图重建过程中的平面检测。通过奇异值分解, 实现了虚拟物体的方向和位置控制; 结合相机的位姿估计, 完成了虚拟物体的注册。实验结果表明, 所提方法的平均注册效率提高了 34.5%。

关键词 机器视觉; 增强现实; 注册; 视觉即时定位与地图构建; 顺序抽样; ORB 算法

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.061501

Improved Augmented Reality Registration Method Based on VSLAM

Zheng Guoqiang^{1*}, Zhou Zhiping^{1,2}

¹ School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China;

² Engineering Research Center of Internet of Things Technology Applications, Ministry of Education, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract In the markerless augmented reality, an improved registration method which combines oriented FAST and rotated BRIEF (ORB) feature and sequential sampling consistency is proposed based on visual simultaneous localization and mapping (VSLAM). The difference between pairs of matching samples is measured through the similarity function, and the samples are extracted from the data subset of the highest quality function. So that the better matching point pairs are obtained and the plane detection is completed in the three-dimensional map reconstruction process. The direction and position of the virtual object is represented by SVD. The registration of the virtual object is completed with the use of pose estimation of cameras. The experimental results show that the proposed method makes the average registration efficiency increases by 34.5%.

Key words machine vision; augmented reality; registration; visual simultaneous localization and mapping; sequential sampling; ORB algorithm

OCIS codes 150.0155; 100.3010; 100.3020; 100.4999

1 引言

增强现实(AR)是一种结合真实场景和虚拟场景的技术,该技术通过计算平台将虚拟信息融合到真实世界的方式来实现增强,从而提升用户对周围信息的感知。根据 Carmigniani 等^[1]的定义,它是虚拟世界与现实世界的一体化,通过三维(3D)注册实现实时交互。具有虚拟与真实的无缝结合,三维

注册和实时交互 3 个特点。随着软硬件的发展,增强现实技术正成为一种极为重要的关键技术,被应用到如服务机器人、自动驾驶、场景三维重建等不同的应用场景中。

Davison 等^[2]实现了基于视觉即时定位与地图构建(SLAM)的实时运行的突破性成果,研发了可穿戴主动视觉系统,可通过实时捕捉到的图像完成自身运动的估测。袁梦等^[3]提出一种点线融合特征

收稿日期: 2018-08-06; 修回日期: 2018-09-10; 录用日期: 2018-10-10

* E-mail: 6161913037@vip.jiangnan.edu.cn

和半直接单目视觉里程计相结合的方法,用于获取更为稳健的相机姿态估计。张剑华等^[4]提出结合视觉 SLAM 的惯性传感器融合的算法,通过惯性矫正相机位姿取得了很好的跟踪效果。之后, Mur-Artal 等^[5]提出一种单目视觉 SLAM,多线程的应用较为理想地实现了跟踪和建图功能。该算法成功建立了点云重建和关键帧选择的最佳适应策略,使得系统具有出色的稳健性,是目前最为优秀的 SLAM 算法之一。研究者们开始将性能较好的 SLAM 用于解决增强现实领域中的三维注册问题。Polvi 等^[6]提出 SlidAR 方法,利用 3D 射线投射和极线几何进行虚拟对象定位,通过交互提高了实用性。但是,该系统需要人为介入,不能完全自主化实现全部过程。以上结合视觉 SLAM 算法中,AR 系统通过计算系统中的单应性矩阵完成三维注册,该方法通过平面检测过程中选择的 4 个点对的坐标确定相机的位姿和世界坐标,并通过视觉 SLAM 算法生成的三维地图信息提高配准精度。但是检测算法在注册速度较低时,系统在三维注册过程中容易出现虚拟对象错位,导致虚拟对象不稳定或者视觉闪烁等结果。Chum 等^[7]提出的 PROSAC 算法可以有效剔除误匹配点对的出现,用以提升平面检测过程中的效率。因此,基于无标记的增强现实系统如何实现更高效

的注册成为一个亟待解决的问题。

针对以上问题,在视觉 SLAM 的基础上提出一种改进的虚拟对象注册方法。首先为了保证 AR 系统的快速性,选择使用 ORB 特征。接着通过 PROSAC 方法优化平面检测进而求解单应性矩阵,克服了由于场景三维信息干扰点多和点云误差大的缺陷,计算出平面法向量并通过奇异值分解求解旋转矩阵。最后在摄像机视角下的渲染坐标系中根据相机的位姿估计完成注册。

2 系统框架

所提的增强现实的注册方法的系统由视觉 SLAM 和注册两个模块组成,系统框架如图 1 所示。其中视觉 SLAM 模块通过相机完成 ORB 特征提取的过程。在当前帧的基础上开始地图构建的初始化,通过处理实时图像的帧间匹配获取相机的位姿评估,同时插入关键帧,接着通过 Urban 等^[8]提出的光束平差法(BA)优化相机位姿的求解。当建图初始化完成后,三维地图开始调用注册模块,随之确定平面并完成注册。最后将由视觉 SLAM 获得的摄像机姿态转换为 OpenGL 的模型视图矩阵,以实现虚拟增强。

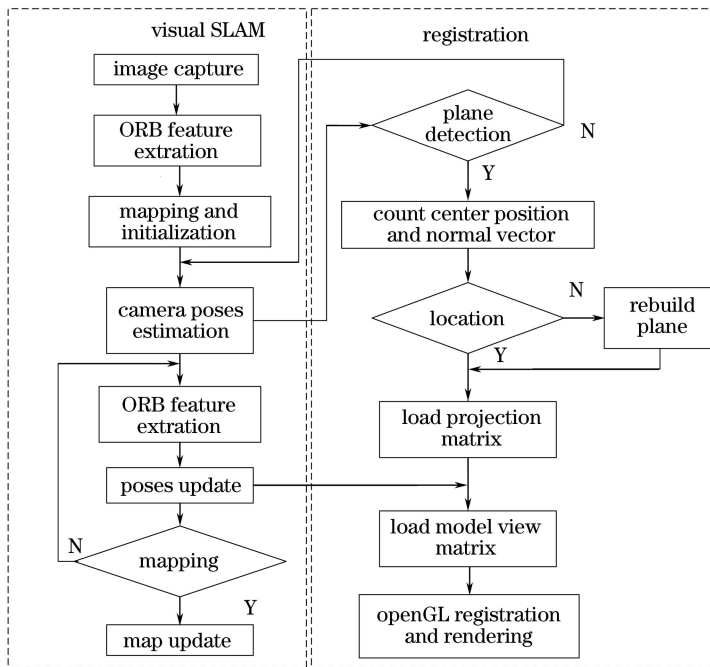


图 1 系统框架

Fig. 1 System framework

2.1 特征提取与平面检测

Rublee 等^[9]将 ORB 特征点的提取通过使用

FAST(Features from accelerated segment test)算法完成基于序列图像的特征点检测。通过使用分段

检测的方法对相邻区域内的像素点进行比对,选出与周围像素具有显著差异的点即为检测点。oFAST(Oriented FAST)算法检测是通过灰度质心方法给 FAST 的主角点增加一个主方向。然后采用 BRIEF(Binary Robust Independent Elementary Features)作为特征描述子,以 FAST 特征点为中心,符合中心附近两元素高斯分布的二进制串即为得到的 BRIEF 特征描述子。接着使用特征点的质心方向,将方向信息添加到 BRIEF 描述子中。第 i 对测试点对为 x_i, y_i , 则 n 对对可以定义一个 $2 \times n$ 的矩阵 \mathbf{Q} , 利用矩阵 \mathbf{Q} 得到 θ 以及与其相应的旋转矩阵 \mathbf{B}_θ , 可由此构建矫正矩阵 $\mathbf{Q}_\theta = \mathbf{B}_\theta \mathbf{Q}$ 。最终得到的具有旋转不变性的 rBRIEF(Rotated BRIEF)描述子为: $g_n(I, \theta) = f_n(I) | x_i, y_i \in \mathbf{Q}_\theta$, 其中 $f_n(I)$ 为第 n 对点描述子 I 的二进制串函数, $g_n(I, \theta)$ 为第 n 对点对应的 rBRIEF 描述子在特征提取过程中的表达函数。为提高描述子的抗噪声能力,使用 5×5 的滑动窗口的灰度像素均值表示该处灰度值。由于基于中间点对所生成的 rBRIEF 描述子向量之间具有较大的相关性,增加了点对匹配过程的复杂度,使用逐步搜索的贪婪算法,剔除高相关性因子,最终得到符合条件的低相关性的描述子,有效解决了算法的噪声敏感性问题。

相机位姿的追踪由关键帧提取的匹配点对求得,由于在匹配点对的求取过程中通常会包含异常值,假设相似度量度的方法比仅仅只是随机猜测选取样本的方法可以更好地对准确性进行预测,所提系统框架采用顺序采样评估 PROSAC 算法快速剔除误匹配点。使用文献[10]中的马氏距离作为品质函数对原始数据集中的样本进行评估,并根据品质高低对数据集中的样本进行降序排列。抽取具有最高品质函数的样本进行迭代,通过连续假设和验证的方式在迭代中获取最优解。最后,根据选取的最优点集完成优势平面的检测与选取。

2.2 虚拟物体的跟踪与注册

在视觉 SLAM 的基础上,通过使用 ORB 特征计算相机位姿并完成跟踪注册。但是,跟踪过程中 ORB 算法提取到的特征点的数目分布可能过于集中,不能很好地反映整张图片的整体特征,同时,过度提取部分区域的特征点致使计算过程十分耗时,因此,特征点分布的影响也是不可忽略的。将图像均匀分割成 $a \times b$ 的模块,候选特征点随机分布在不同的栅格中,对栅格按照从左至右由上至下的顺

序排列,栅格模块标记为 $\mathbf{W} = \{\omega_{11}, \omega_{12}, \dots, \omega_{ab}\}$ 。若子模块中没有检测到候选特征点则设为不感兴趣子模块栅格,且在后面的计算中不予考虑;若检测到候选特征点则设为感兴趣子模块栅格。设对应感兴趣子模块栅格检测到的候选特征点数为 h_j , 根据所提的特征点总数,求得单位模块子栅格中的均值,根据所求栅格均值设立动态阈值为 j , j 满足 $j = \sum h_j / (a \times b)$ 。判断 h_j 与 j 的大小,若满足 $h_j \leq j$, 则候选特征点即为 ORB 提取的特征点;若满足 $h_j > j$, 则根据角点响应函数选择出最好的 j 个特征点。计算该模块内 Harris 的角点,有

$$\mathbf{E} = G(x, y, \sigma) \otimes \begin{bmatrix} D_x^2 & D_x D_y \\ D_y D_x & D_y^2 \end{bmatrix}, \quad (1)$$

式中:矩阵 \mathbf{E} 为特征提取的 Harris 算子; $G(x, y, \sigma)$ 为高斯函数, σ 为高斯函数的标准差; D_x, D_y 分别为 x 和 y 方向上的一阶导数,对应的角点响应函数为

$$F_{\text{Fre}} = \text{Det}(\mathbf{E}) - k \cdot [\text{Tr}(\mathbf{E})]^2, \quad (2)$$

式中: $\text{Det}(\mathbf{E})$ 为矩阵 \mathbf{E} 对应的行列式; $\text{Tr}(\mathbf{E})$ 为求取矩阵 \mathbf{E} 的迹; k 为经验系数。根据文献[11],经验系数取值范围 $k \in [0.04, 0.06]$, 那么如果当该检测带点的 F_{Fre} 值超过设定阈值 T_{Fre} 时,可以判定该检测点为角点。根据响应函数的大小对感兴趣子模块栅格中的候选特征点进行排序,再根据排序选择出排序最好的 j 个特征点。

根据提取的 ORB 特征点计算基于平面的单应性矩阵和基于非平面的基础矩阵。评价公式为

$$S_L = \sum_i \{ \rho_L [d_{\text{cr-L}}^2(x_c^{(i)}, x_r^{(i)})] + \rho_L [d_{\text{rc-L}}^2(x_c^{(i)}, x_r^{(i)})] \}, \quad (3)$$

$$\rho_L(d^2) = \begin{cases} \Gamma - d^2, & d^2 < T_L \\ 0, & d^2 \geq T_L \end{cases}$$

式中: S_L 为计算的模型,下标 L 为单应性矩阵 \mathbf{H} 或者基础矩阵 \mathbf{F} 的统一形式; d_{rc} 和 d_{cr} 是帧间对称传递误差,其中 r 表示参考帧, c 表示当前帧; T_L 为离群点判定阈值,符合 χ^2 分布(根据文献[5]取阈值为 0.95, $T_H = 5.99$, $T_F = 3.84$); $\Gamma = T_L$; x_c, x_r 分别表示基于当前帧和参考帧所提取的 ORB 特征点。该评价公式可以分别计算两种情况下对应的模型矩阵得分,再选择出得分较高模型,接着使用 BA 优化相机位姿,其中相机位姿表示为

$$\{\mathbf{R}, \mathbf{t}\} = \arg \min_{\mathbf{R}, \mathbf{t}} \min_{m \in \mathcal{X}} \text{cov} \rho \cdot \left(\sum_m \|\mathbf{x}^{(m)}(\cdot) - \pi_{(\cdot)}(\mathbf{R}\mathbf{P}^{(m)} + \mathbf{t})\|_2 \right), \quad (4)$$

式中: $\mathbf{R} \in \text{SO}(3)$, $\text{SO}(3)$ 为特殊欧氏群; \mathbf{R} 为旋转矩阵; \mathbf{t} 为平移矩阵; cov 表示关键点的协方差; $\pi(\cdot)$ 表示投影函数; $\mathbf{x}^{(m)}(\cdot)$ 为关键点; $\mathbf{P}^{(i)}$ 为世界坐标系下对应的匹配点; ρ 为 Huber 代价函数; m 为匹配点集中元素。在获取相机位置估计的同时, 通过相机姿势对关键帧进行三角测量获得三维地图点云, 完成地图点云的构建。

PROSAC 算法平面检测确定了平面的位置, 通过使用文献[12]中的方法求解可得 \mathbf{H} 和 \mathbf{F} 阵。但是虚拟物体的注册必须经过从世界坐标系到相机坐标系再到屏幕坐标系的转换, 再用文献[13]中的奇异值分解(SVD)求得变换矩阵 $\mathbf{T}_{4 \times 4}$, 变换矩阵可以描绘成旋转矩阵 $\mathbf{R}_{3 \times 3}$ 及平移矩阵 $\mathbf{t}_{3 \times 1}$ 的组合形式。假设现实空间中任意一点坐标为 (X, Y, Z) , 对应方向上焦距大小为 (f_x, f_y) , 屏幕坐标为 (u, v) , 像素坐标下原点平移为 (c_x, c_y) , 则虚拟物体在屏幕上注册到真实空间坐标满足

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{3 \times 3} & 3 \times 1 \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (5)$$

3 实验与分析

实验验证是基于 Ubuntu 16.0 的系统平台实现

表 2 算法匹配率的对比

Table 2 Matching rate comparison among different algorithms

Test scene	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Scene 6	Scene 7	Scene 8	Scene 9	Scene 10	Average
SIFT	92	91	88	91	89	86	90	85	88	89	88.9
SURF	95	92	94	90	95	96	95	94	92	90	93.3
ORB	93	92	89	93	90	94	90	92	96	93	92.2

所提的改进注册方法是基于 PROSAC 实现的, 相比与传统的 RANSAC 算法, PROSAC 具有更高的计算效率。PROSAC 算法首先以输入数据集中的较高相似度的顺序生成要采样的集合, 以减少异常值的数量, 有效防止了因随机提取的异常值而

表 3 算法采样情况以及总体时间的对比

Table 3 Sampling and total time comparison between two algorithms

Algorithm	Average number of samples / (100)	Maximum number of samples	Minimum number of samples	Average matching points	Total time of two methods / ms
RANSAC	1698.2	2000	58	46.20	62.45
PROSAC	239.54	2000	9	53.65	40.89

此外, 改进前后的两种注册方法在 TUM RGB-D 标准库上的整体运行效率对比如表 4 所示。结果

的。其中使用的 CMOS 摄像头分辨率为 $640 \text{ pixel} \times 480 \text{ pixel}$ 。硬件配置为 AMD A8-4500M APU with Radeon (tm) HD Graphics 1.9 GHz 四核处理器, 内存 10 GB。实验对 3 种算法进行比较, 其中特征点数提取阈值设置为 1000, 实验给出了 3 种特征下算法的运行时间, 计算 300 帧的平均运行时间。算法运行时间的对比如表 1 所示, 由表可以看出, 所选择的 ORB 算法在基于顺序采样评估的视觉 SLAM 增强现实注册方法下实时运行有着明显的速度优势。

表 1 算法运行时间的对比

Table 1 Running time comparison among different algorithms

Algorithm	ORB (oFAST+rBRIEF)	SURF	SIFT
Run time of average frame / ms	23.5	319.6	5809.1

实验测试了 10 个场景(来自 ISPRS 图像数据)下, 3 种算法的匹配率对比如表 2 所示。表中依次对 3 种算法的匹配率进行了对比, 其中黑色加粗表明该算法下的最佳匹配率, 从表中可以看出, ORB 算法平均匹配率略低于 SURF 算法, 但对顺序采样评估的视觉 SLAM 增强现实注册的准确度与速度进行整体评估, ORB 算法为最佳选择, 能够兼顾效率和精度。

导致的验证次数的增加。此外, 通过设置, 保证非随机性的最小值, 减少重复次数并缩短处理时间, 在完成足够的验证过程时可以减少重复验证的次数, 从而提高了算法的计算效率。算法采样情况以及总体时间的对比如表 3 所示。

表明, 改进前后, 所提的基于视觉 SLAM 的增强现实注册方法运行效率提升了 $30.9\% \sim 33.5\%$, 实时

性得到很大程度的提高。

表4 整体运行效率的对比

Table 4 Total running efficiency comparison between two methods

Data sets	Original /ms	Improved /ms
fr1/xyz	43.15	28.95
fr1/rpy	49.04	33.89
fr1/desk	24.69	16.94
fr1/room	38.53	25.62
fr1/360	51.74	34.92

4 结 论

针对增强现实领域中的三维注册的低效率问题,提出了基于 PROSAC 算法的一种改进注册方法,通过品质函数的使用,有效去除了误匹配点对,减少了在剔除误匹配过程中算法的迭代次数,提高了注册方法实时性。实验结果表明,所提方法具有较高的注册效率,大幅提升了系统的速度。但是,所提注册方法是基于三维地图的稀疏点云,只能访问有限的空间配置信息。未来的工作将结合密集点云对所提方法进行探究。

参 考 文 献

- [1] Carmigniani J, Furht B, Anisetti M, *et al.* Augmented reality technologies, systems and applications[J]. *Multimedia Tools and Applications*, 2011, 51(1): 341-377.
- [2] Davison A J, Reid I D, Molton N D, *et al.* MonoSLAM: Real-time single camera SLAM [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(6): 1052-1067.
- [3] Yuan M, Li A H, Zheng Y, *et al.* Point-line feature fusion in monocular visual odometry [J]. *Laser & Optoelectronics Progress*, 2018, 55(2): 021501.
袁梦, 李艾华, 郑勇, 等. 点线特征融合的单目视觉里程计 [J]. *激光与光电子学进展*, 2018, 55(2): 021501.
- [4] Zhang J H, Wang Y Y, Wang Z Y, *et al.* Recovery of lost map for monocular simultaneous localization and mapping [J]. *Journal of Image and Graphics*, 2018, 23(3): 372-383.
张剑华, 王燕燕, 王曾媛, 等. 单目同时定位与建图中的地图恢复融合技术 [J]. *中国图象图形学报*,

2018, 23(3): 372-383.

- [5] Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: A versatile and accurate monocular SLAM system [J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1147-1163.
- [6] Polvi J, Taketomi T, Yamamoto G, *et al.* SlidAR: A 3D positioning method for SLAM-based handheld augmented reality [J]. *Computers & Graphics*, 2016, 55: 33-43.
- [7] Chum O, Matas J. Matching with PROSAC - progressive sample consensus [C] // *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 20-25, San Diego, CA, USA. New York: IEEE, 2005: 220-226.
- [8] Urban S, Wursthorn S, Leitloff J, *et al.* MultiCol bundle adjustment: A generic method for pose estimation, simultaneous self-calibration and reconstruction for arbitrary multi-camera systems [J]. *International Journal of Computer Vision*, 2017, 121(2): 234-252.
- [9] Rublee E, Rabaud V, Konolige K, *et al.* ORB: An efficient alternative to SIFT or SURF [C] // *2011 International Conference on Computer Vision*, November 6-13, Barcelona, Spain. New York: IEEE, 2011: 2564-2571.
- [10] Tuytelaars T. Wide baseline stereo matching based on local, affinely invariant regions [C] // *British Machine Vision Conference*, September 11-14, Bristol. British: BMVA Press, 2000: 412-425.
- [11] Zhang Y S, Zou Z R. Automatic registration method for remote sensing images based on improved ORB algorithm [J]. *Remote Sensing for Land & Resources*, 2013, 25(3): 20-24.
张云生, 邹峥嵘. 基于改进 ORB 算法的遥感图像自动配准方法 [J]. *国土资源遥感*, 2013, 25(3): 20-24.
- [12] Hartley R I. In defense of the eight-point algorithm [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(6): 580-593.
- [13] Faugeras O D, Lustman F. Motion and structure from motion in a piecewise planar environment [J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 1988, 2(3): 485-508.