中国科学技术大学

国创成果报告

这里是标题
标题只能有三行
不能多

姓　　名：wy wfd zsq
院　　系：少年班学院
导　　师：黄章进 副教授
完成时间：二〇一七年五月

致　谢

感谢原本科模板的作者 XPS、硕博模板的作者刘青松以及它们的维护者的辛勤工作！

感谢大家对本模板更新工作的支持！

本模板以及本示例文档还存在许多不足之处，欢迎大家测试并及时提供反馈。

ywg@USTC

在中国科技大学完成本科和硕博连读学业的九年里，我所从事的学习和研究工作，都是在导师以及系里其他老师和同学的指导和帮助下进行的。在完成论文之际，请容许我对他们表达诚挚的谢意。

首先感谢导师 XXX 教授和 XXX 副教授多年的指导和教诲，是他们把我带到了计算机视觉的研究领域。X 老师严谨的研究态度及忘我的工作精神，X 老师认真细致的治学态度及宽广的胸怀，都将使我受益终身。

感谢班主任 XXX 老师和 XX 老师多年的关怀。感谢 XXX、XX、XX 等老师，他们本科及研究生阶段的指导给我研究生阶段的研究工作打下了基础。

感谢 XX、XXX、XXX、XX、XXX、XXX、XXX、XX 等师兄师姐们的指点和照顾；感谢 XXX、XX、XXX 等几位同班同学，与你们的讨论使我受益良多；感谢 XXX、XX、XXX、XX、XXX 等师弟师妹，我们在 XXX 实验室共同学习共同生活，一起走过了这段愉快而难忘的岁月。

感谢科大，感谢一路走过来的兄弟姐妹们，在最宝贵年华里，是你们伴随着我的成长。

最后，感谢我家人一贯的鼓励和支持，你们是我追求学业的坚强后盾。

赵钱孙

2017 年 5 月 8 日

目 录

致 谢 I
目 录 III
摘 要 V
ABSTRACT VII
第一章 Introduction 1
1.1 增强现实技术 1
1.2 增强现实的定位方式 1
1.3 V-SLAM 技术 1
1.3.1 V-SLAM 的基本原理 1
1.3.2 基于关键帧 BA 的单目 V-SLAM 系统 2
第二章 基于单目相机的 3 维实时重建的 PTAM 算法介绍 3
2.1 PTAM 算法的介绍 3
2.2 PTAM 算法工作的基本流程 3
2.2.1 PTAM 管线之一: Tracking 3
2.2.2 PTAM 管线之二: Mapping 5
2.3 PTAM 算法的实际效果 6
2.3.1 在一般电脑上运行 6
2.3.2 在智能手机上运行 6
2.3.3 PTAM 算法的评价 6
参考文献 9

摘　要

本文是中国科学技术大学本硕博毕业论文模板示例文件。本模板由 ywg@USTC 创建，适用于撰写学士、硕士和博士学位论文，本模板由原来的本科模板和硕博模板整合优化而来。本示例文件除了介绍本模板的基础用法外，本文还是一个简要的学位论文写作指南。

关键词： 中国科学技术大学 学位论文 LaTeX 通用模板 学士 硕士 博士

ABSTRACT

This is USTC thesis template for bachelor, master and doctor user's guide. The template is created by ywg@USTC and a derivative of USTC Bachelor and Master-PhD templates. Besides that the usage of the template, a brief guideline for writing thesis is also provided.

Keywords: University of Science and Technology of China (USTC), Thesis, Universal LaTeX Template, Bachelor, Master, PhD

第一章 Introduction

1.1 增强现实技术

增强现实技术 (augmented reality) 是一种将真实世界信息和虚拟世界信息“无缝”集成的新技术，是把原本在现实世界的一定时间空间范围内很难体验到的实体信息 (视觉信息, 声音, 味道, 触觉等), 通过电脑等科学技术，模拟仿真后再叠加，将虚拟的信息应用到真实世界，被人类感官所感知，从而达到超越现实的感官体验。比起传统方式来说，它更加的直观，更加的高效，因此也有着更加广阔的应用前景。近年来，增强现实技术已在军事，生活，游戏等众多领域运营并取得了成功。例如，宜家家居公司已经开发了一个 APP 使得用户可以使用智能手机观察不同的家具在自己房间的摆放效果；而任天堂公司也开发了 Pokemon-Go 游戏，使得玩家可以通过智能手机在现实世界里发现精灵。

1.2 增强现实的定位方式

增强现实需要实时定位设备在环境中的方位，定位的方案虽然有许多种，但多数方案都存在局限或者代价太高难以普及，例如 GPS 无法在室内及遮挡严重的环境里使用，且精度较低，而基于无线信号的定位方案则需要事先布置场景。基于视觉的同时定位与地图构建技术（visual simultaneous localization and mapping V-SLAM）以其成本低廉、小场景精度较高、无需预先布置场景等优势成为比较常采用的定位方案。

1.3 V-SLAM 技术

V-SLAM 技术指的是使用图像作为外部信息的唯一来源，来定位一个机器人、一辆车或者一个移动的相机在整个场景中的位置，同时，重建环境的三维结构。

1.3.1 V-SLAM 的基本原理

V-SLAM 技术根据拍摄的视频、图像信息推断摄像头在环境的方位，同时构建环境地图，其原理为多视图几何原理 (Multiple view geometry theroy) V-SLAM 的目标为同时恢复出每帧图像对应的相机运动参数 C_1,C_2 $\cdots C_m$ 以及场景三维结构 X_1,X_2 $\cdots X_n$, 每个相机运动参数 C_i 包含了相机的位置和朝向信息，通常表达为一个 3×3 的旋转矩阵 R_i 和一个三维位置变量 p_i。R_i 与 p_i 将一个世界坐标系下的三维点 X_j 变换至 C_i 的局部坐标系

$$(X_{ij}, Y_{ij}, Z_{ij})^{\mathrm{T}} = R_i(X_j - p_i) \tag{1.1}$$

进而投影至图像中

$$h_{ij} = \left(f_x X_{ij}/Z_{ij} + c_x . f_y Y_{ij}/Z_{ij} + c_y\right)^{\mathrm{T}} \tag{1.2}$$

其中，f_x,f_y 分别为沿图像 x,y 轴的图像焦距,(c_x, c_y) 为镜头光心在图像中的位置，通常假设这些参数已实现标定且保持不变，由式 (1.1) (1.2)，三维点在图像中的投影位置 h_{ij} 可表示为一个关于 C_i 和 X_j 的函数，记为

$$h_{ij} = h(C_i, X_j) \tag{1.3}$$

V-SLAM 算法需要将、对不同图像中对应于相同场景的图像点进行匹配，而这个过程是通过求解如下目标函数

$$\underset{C_1, C_m, X_1, X_n}{\arg\min} \sum_{i=1}^{m} \sum_{j=1}^{n} \left\| h(C_i, X_j) - \tilde{x_{ij}} \right\|_{\Sigma_{ij}} \tag{1.4}$$

得到一组最优的 C_1，$C_2 \cdots C_m$,X_1,X_2 $\cdots X_n$，使得所有 X_j 在 C_i 图像中的投影位置 h_{ij} 与观测到的图像点位置 x_{ij} 尽可能靠近，这里假设图像观测点符合高斯分布 $x_{ij} \sim \mathrm{N}\ (\tilde{x_{ij}}, \Sigma_{ij})$，$\| e \| = e^{\mathrm{T}}\ \Sigma^{-1}\ e$ 求解目标函数(1.4)的过程也成为集束调整（bundle adjustment,BA）, 该最优化问题可利用线性方程的稀疏结构高效求解。

1.3.2 基于关键帧 BA 的单目 V-SLAM 系统

由于现阶段大多数 AR 产品都以智能手机以及平板电脑作为载体，而智能手机的摄像头大多以单目为主，双目、三目摄像头甚至深度摄像头都未得到普及，因此本文主要讨论基于单目视觉的同时定位与地图构建方法。目前，主流的 V-SLAM 方法主要为：基于滤波器、基于关键帧 BA 和基于直接跟踪，我们先来看看这三种方法。比较并分析其优劣，而后详细介绍基于关键帧 BA 的 V-SLAM 方法。其中比较具有代表性的有 MonoSLAM 以及 MSCKF

基于滤波器的 V-SLAM 的方法将系统每一时刻的状态 t 用一个高斯概率模型表达，$x_t \sim \mathrm{N}\ (\tilde{x_t}, \mathrm{P}_{ij})$, 其中 $\tilde{x_t}$ 为当前时刻系统状态估计值，P_t 为该估计值误差的协方差矩阵，系统状态由滤波器不断更新。

而基于关键帧 BA 的 V-SLAM 方法是近年来最流行的方法之一，他的主要思想是将相机跟踪（Tracking）和地图构建（Mapping）作为两个独立的任务在两个线程并行执行，而 Mapping 线程仅维护视频流中抽取的关键帧。PTAM 是最著名的基于关键帧 BA 的方法之一，也是我们介绍的重点

基于直接跟踪的 V-SLAM 方法则是直接通过比较像素颜色来求解相机运动，具有代表性的算法有 DTAM 以及 LSD-SLAM。

第二章 基于单目相机的 3 维实时重建的 PTAM 算法介绍

2.1 PTAM 算法的介绍

关键词，双管线系统，为后面的算法奠定了基础

我们首先简述 PTAM 实现单目相机 SLAM 的原理。单目相机模型不同于双目相机模型，实时追踪时相机视界中的点不能和其它相机视界中的点进行匹配，只能和自己的关键帧匹配，从而加大了 3D 重建中定位的难度。PTAM 算法提出利用单目相机实时追踪特征点的可行性，实现三维重建。

PTAM 算法主要思想是将 Tracking 和 Mapping 两个过程放在不同的管线 (进程) 中进行: Tracking 进程专门实现相机位置的估计，Mapping 进程则用于进行关键帧之间的误差消除。

2.2 PTAM 算法工作的基本流程

如果记 W 为真实世界的坐标系,PTAM 算法将维护一个关键帧集合: $Img = \{I_1, I_2, \ldots, I_m\}$，这 m 个关键帧分别对应 m 个相机坐标系 K_i，以及一个三维重建的结果。我们用 E_{K_iW} 表示从世界坐标系到相机坐标系的仿射变换 (Affine Transformation)。

2.2.1 PTAM 管线之一: Tracking

追踪进程需要解决如下的问题：

当读入了新的关键帧之后，原来算法在重建过程中提取的 3 维空间中的特征点现在在照片中的坐标是什么？现在的相机姿态应该怎么估计？

我们假定程序可以从映射进程得到一个关键帧集合 Img 以及 3 维重建的特征点的坐标集合 (相对于世界坐标系)$P = P_W = \{\mathbf{p}_{1W}, \ldots, \mathbf{p}_{sW}\}$，为了统一形式，将第 j 个点坐标记为 $\mathbf{p}_{jW} = (p_{jx}, p_{jy}, p_{jz}, 1)$。

根据已知结论，仿射坐标系变换对应公式为

$$\mathbf{p}_{jK_{t+1}} = E_{K_{t+1}W}\mathbf{p}_{jW} \tag{2.1}$$

为了把三维空间中的视界投影到二维空间，算法遵循 [1] 中的 FOV 相机模型，构建一个 $\mathbb{R}^3$ 到 $\mathbb{R}^2$ 的映射为:

$$f(x, y, z, 1) = (u_0, v_0) + (x/z, y/z)\begin{bmatrix} f_u & 0 \\ 0 & f_v \end{bmatrix}\frac{r'}{r} \tag{2.2}$$

$$r = \sqrt{\frac{x^2 + y^2}{z^2}} \tag{2.3}$$

$$r' = \frac{1}{\omega} arctan(2rtan\frac{\omega}{2}) \tag{2.4}$$

其中我们假定焦距 f_u, f_v，主点位置 (u_0, v_0) 和畸变系数 ω 已知。这时对于实时相机姿态的更新，相当于对于 式 2.2 求微分，在假定线性运动的情况下更新相机姿态的变换仿射矩阵。若我们设从上一关键帧到下一个关键帧的仿射变换矩阵为 T，则有以下的关系成立:

$$E_{K_{t+1}W} = TE_{K_tW} = exp(\mu)E_{K_tW} \tag{2.5}$$

其中 μ 为 6 维向量，代表矩阵 T 的 6 个自由度。所以问题转化为: 根据图像中特征点 (u, v) 的变化和在每个特征点三维空间中的预估位置，求解 μ, T 的取值。

我们假定点 p 为我们需要定位的特征点，则我们首先需要在当前的关键帧图像中找到该点的新投影坐标 $(\hat{u}, \hat{v})$。为此，我们构造图像的尺度空间，利用的 FAST-10[2] 计算角点的方法提取可能的特征点。在假定相机移动很慢的情况下，特征点匹配算法从上一帧的位置开始，在一定的半径阈值内，在对应的视界空间上，依据设计好的评价函数 (scoring function) 进行搜索。算法会在搜索过程中得到一个坐标 $(\hat{u}, \hat{v})$，以及 $\sigma = 2^l$ 作为视界空间金字塔中搜索对应的层数。

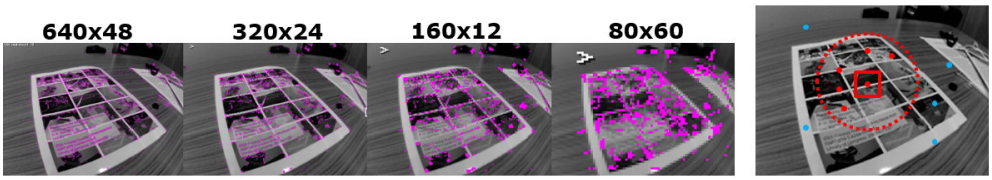

图 2.1 视界空间中利用 FAST 提取特征点和搜索对应特征点示意图，左图表示视界空间的 down sampling，以及基于 FAST 方法的特征点选取，右图为在对应的尺度空间中搜索对应的特征点

在得到所有特征点的信息之后，算法求解以下的优化问题计算相机姿态的更新 (设 P_{t+1} 为当前特征点集合):

$$\underset{\mu}{argmin} \sum_{j \in P_{t+1}} \psi(\frac{\|\mathbf{e}_j\|}{\sigma_j}, \sigma_T) \tag{2.6}$$

$$\mathbf{e}_j = (\hat{u}_j, \hat{v}_j) - f(exp(\mu)E_{K_tW}\mathbf{p_j}) \tag{2.7}$$

这里 ψ 为 Tukey loss:

$$\psi(x, c) = \begin{cases} x(1 - \frac{x^2}{c^2}) & \text{for } |x| < c \\ 0 & \text{for } |x| > c \end{cases} \tag{2.8}$$

为了保证算法的稳定性，追踪进程会进行两次：第一次会从三维模型中抽取 50 个特征点投影匹配，第二次则抽取 1000 个特征点进行匹配。同时，如果在特征点投影搜索配对的过程中，如果特征点搜索失败的比率大于某一个阈值的话，算法将会判定这一帧失效，并且自动舍弃（这可能是由于抖动，位移量过大特征点不在视界中等多种原因造成）。

2.2.2 PTAM 管线之二: Mapping

在追踪进程持续运行的同时，映射进程会根据追踪进程估计的相机姿态，完成关键帧的选取及三维重建的主要任务。为此，算法首先在初始化阶段，需要用户指定视频流中的两帧 (第一帧一般为起始帧)，单独进行第一次特征点配对。这样做的目的有两个，一方面是为了利用五点法求解相机模型中的固有参数，另一方面是为了先定位特征点到三维空间中作为初始信息。

2.2.2.1 关键帧的选择和插入

在程序进程中，当视频流被读入之后，对于每一帧图像，算法需要判断是否需要把该帧图像加入关键帧集合。判断主要依据是距离上一关键帧的时间和预估的相机距离，以及图像的质量。在选定关键帧之后，根据追踪进程返回的信息，结合上一关键帧的对应特征点信息，定位深度，从而把特征点加入到三维重建模型中。

2.2.2.2 利用 BA(Bundle Adjustment) 校正

为了消除帧与帧之间的积累误差，映射进程还需要对整体进行一个调整，即所谓的 BA。算法对于所有关键帧求解一次优化问题，也会对局部的几个特征点求解局部的优化问题，它们有以下形式:

$$\underset{\{\mu\},\{\mathbf{p}\}}{argmin} = \sum_{i=1}^{N}\sum_{j\in P_i}\psi\left(\frac{\|\mathbf{e}_j\|}{\sigma_j},\sigma_T\right) \tag{2.9}$$

这里的计算将会采用 Levenberg-Marquardt BA 算法 (详见 [3])。在完成校正之后，新的特征点的三维信息和关键帧信息将作为追踪进程的输入，传入 Tracking 进程，两个进程在工作的时候将会保持相互的通信，这样最大程度上保证算法运行的流畅度。上述算法的核心过程，在下方图 2.2中进行了简要的总结。

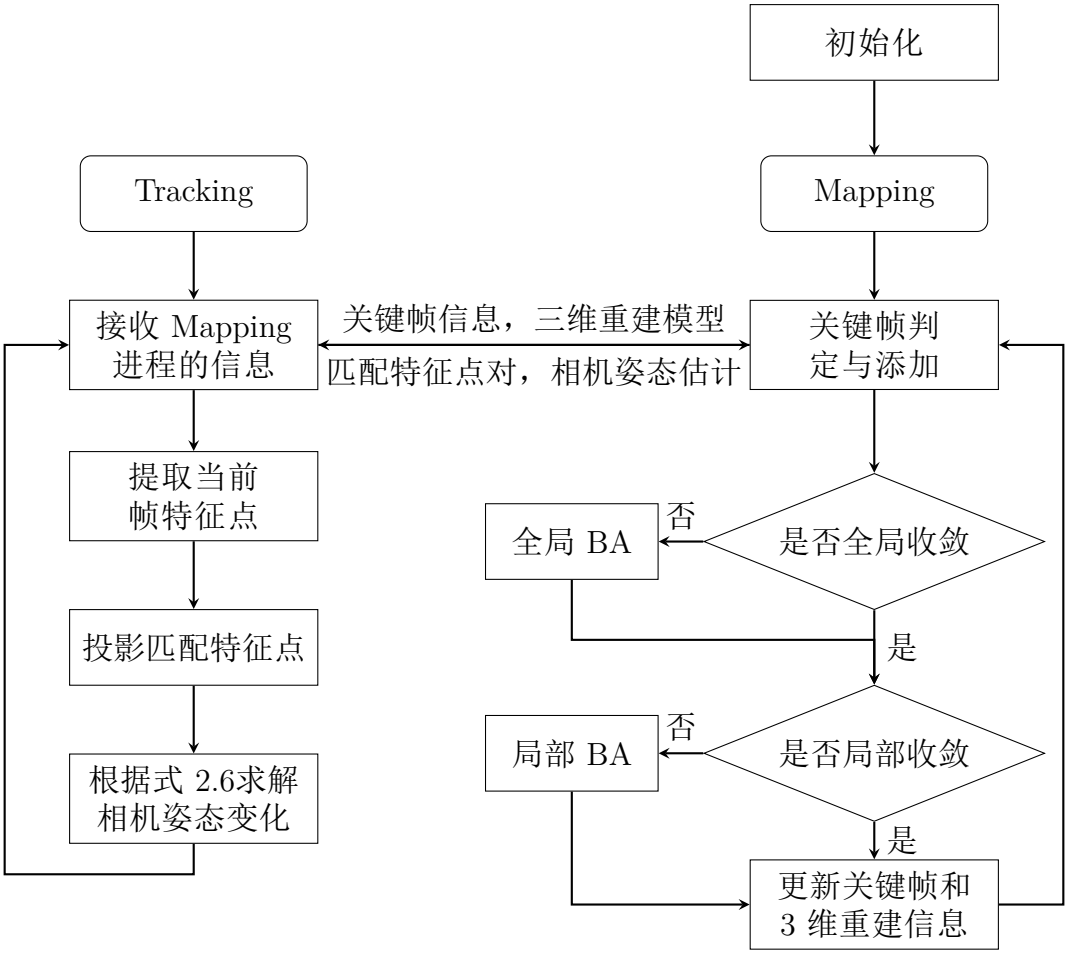

图 2.2 基于 BA 的单目经典算法 PTAM 的流程示意图

2.3 PTAM 算法的实际效果

本节重点比较和讨论对于 PTAM 算法运行效果的评价。

2.3.1 在一般电脑上运行

在一般的笔记本电脑上运行的结果，如下图所示。对于开放的场景 (如寝室内部)，实际测试中帧率几乎达不到 30 帧/秒，能够在 20 帧每秒的情况也比较少见。对于特定的平面场景 (如桌面)，实际测试结果。

2.3.2 在智能手机上运行

智能手机的性能会是制约此类算法在手机上表现的一个重要因素，[4] 中尝试把 PTAM 的系统移植到 Iphone 中，得到了图 2.3所示的结果。

2.3.3 PTAM 算法的评价

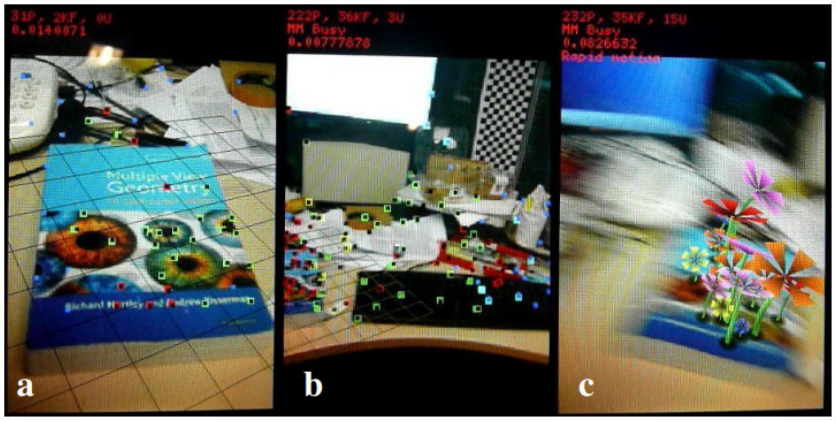

图 2.3 在智能手机上运行的结果，(a) 图为起始状态，(b) 图表现了特征点的追踪，(c) 图表示了在运动过程中的平面捕捉与 AR 渲染。由于手机运算能力限制，特征点数量较 PC 端大幅减少

参考文献

[1] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces. 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR, 2007..

[2] Machine learning for high-speed corner detection. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2006, 3951 LNCS:430–443.

[3] Hartley R I, Zisserman A. Multiple View Geometry in Computer Vision. Second ed., Cambridge University Press, ISBN: 0521540518, 2004.

[4] Klein G, Murray D. Parallel tracking and mapping on a camera phone. Proceedings of Science and Technology Proceedings - IEEE 2009 International Symposium on Mixed and Augmented Reality, ISMAR 2009, 2009. 83–86.