



计算机工程

Computer Engineering

ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目：动态场景下基于视觉特征的 SLAM 方法
作者：张金凤，石朝侠，王燕清
网络首发日期：2019-10-25
引用格式：张金凤，石朝侠，王燕清. 动态场景下基于视觉特征的 SLAM 方法. 计算机工程. <http://kns.cnki.net/kcms/detail/31.1289.tp.20191025.1559.006.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

动态场景下基于视觉特征的 SLAM 方法

张金凤¹, 石朝侠¹, 王燕清²

(1. 南京理工大学, 计算机科学与工程学院, 南京 210094; 2. 南京晓庄学院, 信息工程学院, 南京 211171)

摘 要: 同时定位与地图构建(SLAM)作为机器人领域的研究热点, 近年来取得了长足进步, 但很少有 SLAM 算法考虑到场景中的动态或可移动目标。针对场景中动态、可移动性目标的问题, 提出了一种适用于动态场景的 SLAM 方法, 将基于深度学习的目标检测算法引入到经典 ORB_SLAM2 方法中, 将特征点分为潜在动态特征和非潜在动态特征, 基于非潜在动态特征点计算出运动模型, 筛选出场景中的静态特征点进行位姿跟踪, 利用非潜在动态特征中的静态特征点进行建图。在 KITTI 和 TUM 数据集上和 ORB_SLAM2 进行对比, 提高了跟踪轨迹精度, 在保障地图精度的同时, 改善了地图的适用性。

关键词: 同时定位与地图构建; 局部特征; 动态场景; 深度学习; 目标检测。



SLAM method based on visual features in dynamic scene

Jinfeng Zhang¹, Chaoxia Shi¹, Yanqing Wang²

(1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; 2. School of Information Engineering, Nanjing Xiaozhuang University, Nanjing 211171, China)

【Abstract】 As a research hotspot in the field of robotics, Simultaneous localization and mapping(SLAM) has made great progress in recent years, but few SLAM algorithms take dynamic or movable targets in the scene into account. To handle the problem of scene with mobile or movable targets, a SLAM method which introduces deep learning based object detection into classic ORB_SLAM2 is proposed to make it more suitable for dynamic scene. The features extracted are divided into dynamic features and potential dynamic features. We calculate the motion model with potential dynamic features, and then use the model to select the static features in potential dynamic features for global map building. Compared with ORB_SLAM2 on KITTI and TUM datasets, The tracking accuracy as well as the application performance of the map are improved.

【Key words】 Simultaneous localization and mapping; Local features; Dynamic scenario; Deep learning; object detection.

0 概述

同时定位和地图构建(Simultaneous localization and mapping, SLAM)问题在机器人和计算机视觉领域有着悠久的历史。过去已经探索了不同的传感器模式, 包括 2D 激光扫描仪、3D 扫描仪、单目相机、双目相机和 RGB-D 传感器。现有的 SLAM 方法, 从所处理的输入数据类型的角度, 可分为基于深度相机的方法和基于单目相机的方法。从采用方法的角度来看, 也可以分为依靠特征点进行匹配并构建稀疏地图的特征点法, 和最小化光度误差并构建稠密地图的直接法。本文选用基于 ORB(Oriented Brief)特征点的 SLAM 方法^{[1][2]}。ORB 特征点^[3]可以在 CPU

上实时计算; 相比 Harris 角点等简单角点特征, 又具有良好的旋转和缩放不变性。并且, ORB 定义的描述子, 在运动范围较大时也可以实现良好的回环检测和重定位效果。对于如何消除场景中移动目标和潜在移动目标的影响, 我们引入基于深度学习的对象检测算法来检测场景中的移动目标和潜在移动目标。基于非潜在动态目标的运动模型对特征点全集进行筛选, 利用其中的静态特征点进行位姿跟踪, 利用静态特征点中的非潜在动态特征点进行地图构建。

1 相关工作

基金项目: 国家自然科学基金面上项目(61371040)

作者简介: 张金凤 (1994-), 女, 硕士研究生, 主研方向为视觉 SLAM; 石朝侠, 副教授, 博士; 王燕清, 副教授, 博士。

E-mail: 2543771964@qq.com scx@njjust.edu.cn

SLAM 技术涉及的技术范围很广,也是机器人和计算机视觉领域的一个热门研究方向。目前大多数 SLAM 方法的一个基本假设是环境是静态的。然而,像人类这样的活动对象存在于许多真实场景中。因此,大多数最初设计用于在静态环境中执行 SLAM 的方法都无法处理复杂的动态场景。为了解决这个问题,我们需要从环境中识别出具有移动性的目标,然后在进行姿态估计之前丢弃它们。在稠密 SLAM 中,许多运动目标检测方法往往基于光流方法。如果图像中存在运动目标,就会产生光流,通过计算光流的不一致性,可以区分静态背景和运动对象。例如,Fang^[4]等人使用最优估计和均匀采样的方法来检测动态对象。与其他导数光流方法相比,该方法具有较强的时效性,但精度较低,计算量较大,尤其是在大图像采样量较大的情况下。Wang^[5]等人基于点轨迹对图像进行聚类,将动态对象排除在最小化能量函数之外。他们的方法是健壮的,但不是实时的。因此,这一领域的研究还需要进一步完善。近两年来,在动态场景中出现了许多 SLAM 算法。它们大多使用语义分割或目标检测算法来识别环境中的动态因素。

在 dynaSLAM^[6]中,作者提出了动态对象检测系统。使用 MASK-RCNN^[7]进行实例分割,并对具有移动性的物体进行分割,比如事先没有在 MASK-RCNN 中标记的行人手中的书籍。将之前的 20 个关键帧的 RGB 信息和深度信息映射投影到当前帧上,在没有动态对象的情况下完成背景修复。该工作实现了对动态场景的良好处理,但无法实现实时操作,ORB_SLAM2 项目的实时性能不容忽视。在 DS-SLAM^[8],通过将语义分割网络与光流法相结合,提供八叉树地图的语义表示,重点降低基于视觉的 SLAM 中动态对象的影响。VSO^[9]的主要思想是使用语义作为不变的场景表示。基于的原理是,视角、尺度、光照等方面的变化只影响物体的低层外观,而不影响其语义。基于这一思想,作者提出了一种将语义约束集成到姿态优化和地图优化中的视觉语义测距方法。

近年来,驾驶环境中的目标检测问题越来越受到研究领域和工业界的关注。自动检测道路上的车辆和行人等物体,使驾驶员能够了解道路状况、交通信息等。因此,目标检测可以用于各种应用,如自动车辆^[10]和自动监视系统^[11]。然而,在开发可靠的驾驶环境下的目标检测方法中存在着许多挑战:

目标遮挡问题^[12]、尺度^[13]的大方差等。特别是,由停放的汽车、过往车辆和行人引起的阻塞是道路上最关键的因素之一。这种遮挡会使道路上的目标检测变得困难。已有一些研究成果^{[14][15][16]}中。Dalal 提出了一种基于 HOG 和线性 SVM 的行人检测方法^[14]。其主要思想是利用 HOG 特征和线性 SVM 学习正、负样本模板。Felzenszwalb 提出了一种基于多尺度变形分量检测模型的可变形零件模型^[17]。DPM 是深度学习之外的最佳对象检测模型。DPM 通常采用滑动窗口检测方法,通过构造一个尺度金字塔对每个尺度进行搜索。区域卷积神经网络(RCNN:Regions with CNN features)^{[18][19][20]}是最早的卷积神经网络在检测模型中应用网络进行目标检测。RCNN 在传统的选择性搜索的基础上选择候选集,建立 CNN 特征提取网络的检测思想来检测目标。YOLO 是 Joseph Redmon 和 Ali Farhadi 等人在 2015 年提出的一种基于神经网络的目标检测系统^[21],是第一个基于 CNN 的目标检测算法检测端到端模型,其将整个图像作为网络模型的输入,将图像划分为 $S \times S$ 网格。如果一个物体的中心落在网格上,网络负责检测这个物体,然后输出该区域的物体和该区域的置信度。

2 体系架构

在本节中,我们将介绍这个系统的细节。首先是整个系统的框架,接着简要介绍了基于深度学习的目标检测方法,实时检测出场景中具有潜在移动性的目标,根据检测结果对特征点进行分类。然后介绍了对位姿跟踪和建图模块输入特征点的筛选。最后显示了在 KITTI 数据集和 TUM 数据集上实时运行的结果。

2.1 主要框架

在现有面向动态环境的 SLAM 算法改进方法中,实时性是不可忽视的重要方面。首先,我们考虑了基于像素的语义分割方法与 SLAM 的结合,但实验发现基于像素的语义分割方法的运行速度和准确率成反比。所以综合考虑,引入基于深度学习的目标检测方法检测环境中潜在的动态对象。其良好的运行速度,满足了 SLAM 系统的实时性要求。如图 1 所示,该方法在基于特征点的 SLAM 方法中增加了基于深度学习的目标检测方法,将输入图像中提取出的特征点分为两类,一类是潜在动态特征

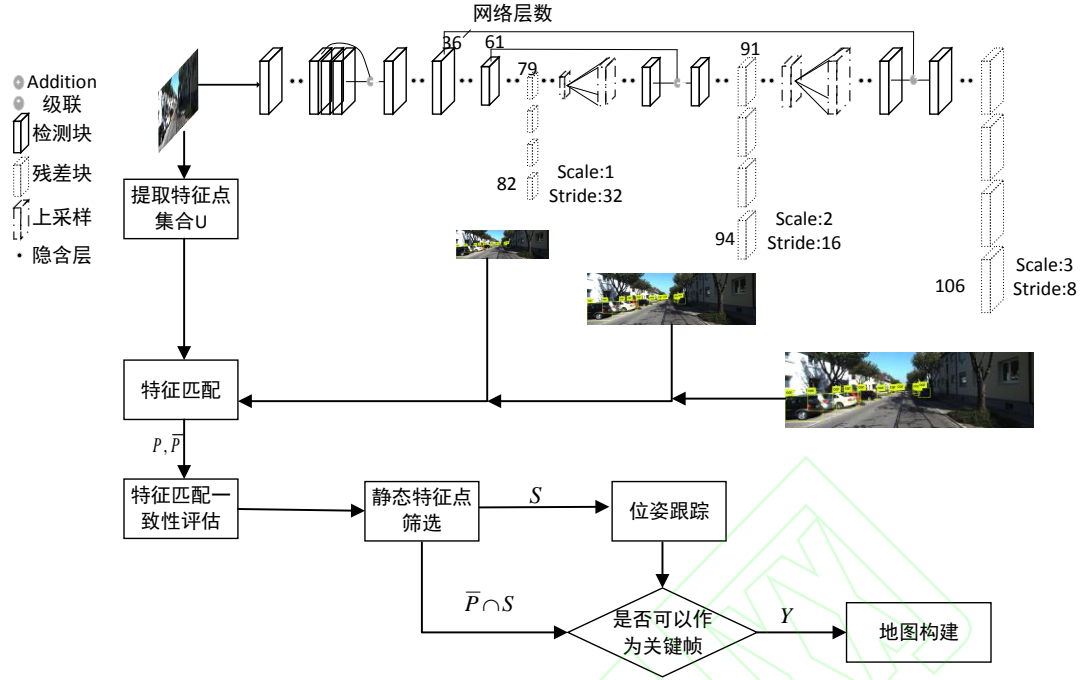


图1 主要框架

点,这一类特征点具有可移动性,在场景中并不是长时间固定存在的(如车、人等目标上提取的特征点),在重定位和闭环检测时可能会导致场景匹配失败;另一类是非潜在动态特征点,这一类特征点在场景中往往不能移动或者在足够长的时间内不会移动(如建筑物、树木等目标上提取的特征点)。

我们首先利用基于深度学习的目标检测将特征点集合 U 分为潜在动态特征点 P (人、车等目标) 和非潜在动态特征点 \bar{P} (建筑、道路、植被等)。然后使用特征匹配一致性评估跟踪的车辆位姿运动模型将特征点集合 U 分为动态特征点集合 D (场景中实际移动了的特征点,如正在移动的车和人上提取的特征点) 和静态特征点集合 S (场景中未移动的特征点,如建筑物、停在路边的车上提取的特征点)。上述集合之间的关系为:

$$U = D \cup S = P \cup \bar{P} \quad (1)$$

在位姿跟踪模块,只基于场景中的静态特征点进行跟踪,排除动态特征点的影响,可以有效提高跟踪精度。在建图模块,只基于场景中的非潜在动态特征点中的静态特征点 $S \cap \bar{P}$ 进行建图,改善地图的适用性,防止在重定位和闭环检测时受到场景中动态目标的干扰。

2.2 ORB 特征点提取

ORB 特征点由两部分组成:关键点和描述符。提取 ORB 特征主要分为两个步骤:

1) FAST 角点提取:找到图像中的角点,计算

特征点的主方向,并为后续的简要描述符添加旋转不变特征;

2) BRIEF 描述子:描述上一步提取的特征点周围的图像区域。

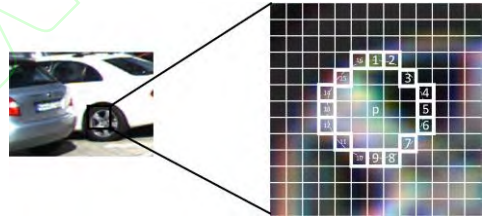


图2 FAST 角点

FAST 角点,如图2所示。其主要依据是,如果一个像素与其周围像素显著不同(太亮或太暗),那么它可能是角点。ORB 添加了尺度和旋转的描述。对于任意一个特征点 p 来说,定义 p 的邻域像素的矩为:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \quad (2)$$

其中 $I(x,y)$ 为点 (x,y) 处的灰度值。那么我们可以得到图像的质心为:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (3)$$

那么特征点与质心的夹角定义为 FAST 特征点的方向:

$$\theta = \arctan(m_{01}, m_{10}) \quad (4)$$

为了提高方法的旋转不变性,需要确保 x 和 y

在半径为 r 的圆形区域内, 即 $x, y \in [-r, r]$, r 等于邻域半径。在提取有向 FAST 关键点后, 计算每个点的描述符。ORB 选择了 BRIEF 作为特征描述方法, 该算法采用随机选取点的方法, 选择特征点周围 $S \times S$ 大小的像素块, 随机选取 n 对像素点, 定义:

$$\tau(p; x, y) := \begin{cases} 1 & \text{if } p(x) < p(y) \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$p(x)$ 是点 x 处的灰度值, 则特征点 p 的描述子定义为:

$$f_n(p) := \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x_i, y_i) \quad (6)$$

2.3 基于深度学习的目标检测算法

随着深度学习技术的快速发展, 目标检测算法从基于手工特征的传统算法转向了基于深度神经网络的检测技术。本文采用了由残差块构成的全卷积网络作为网络主体, YOLOv3 中将其命名为 Darknet-53 网络, 其细节图见图 1。在不同尺度上进行特征提取, 获得最后目标检测结果。如图 3 所示, 定义检测框集合为 $R = \{r_1, r_2, r_3, \dots\}$, 单个检测框定义为 $r(x, y, w, h) \in R$, 其中 (x, y) 为检测框左上角在帧中的坐标, (w, h) 为检测框的宽和高, 对特征点集合 U 中的每个特征点 p 判断:

$$\begin{cases} p(u, v) \\ r(x, y, w, h) \end{cases} \Rightarrow \begin{cases} x \leq u \leq x+w \text{ 且 } y \leq v \leq y+h \\ u < x \text{ 或 } u > x+w \text{ 或 } v < y \text{ 或 } v > y+h \end{cases} \begin{matrix} p \in \bar{P} \\ p \in P \end{matrix} \quad (7)$$

经过上述操作, 特征点全集 U 被分为潜在动态特征点集合 P 和非潜在动态特征点集合 \bar{P}



图 3 特征点分类示意图

2.4 特征点匹配

在提取特征点后和参考帧中特征点进行粗匹配, 由前面的定义可知, \bar{P} 中的特征点都是非潜在动态特征点, 因此大多数都是静态特征点, 或者运动范围非常小的特征点。我们基于属于集合 \bar{P} 的特征点对, 采用改进的^[22]算法, 获得能够符合场景中静态特征点的最优运动模型, 算法步骤如图 4 所示: 具体操作分为如下 4 步:

1) 首先根据特征点的描述子对 U 中特征点进行粗匹配, 比较特征点描述向量之间的 Hamming 距离, 距离越小表明两个特征点之间相似度越高, 若 Hamming 距离小于一定的阈值, 则表示两个特征点

匹配成功。我们记匹配点对集合为 U_m :

$$U_m = \{ \{u_c, u_r\}, \sum_{1 \leq i \leq n} f_n^i(u_c) \oplus f_n^i(u_r) < \delta \} \quad (8)$$

其中 u_c 为在当前帧中的一个特征点, u_r 为在参考帧中的匹配特征点, δ 为定义的距离阈值。定义 $U_{\bar{P}}$:

$$U_{\bar{P}} = \{ \{u_c, u_r\}, \{u_c, u_r\} \in U_m \text{ 且 } u_c \in \bar{P} \} \quad (9)$$

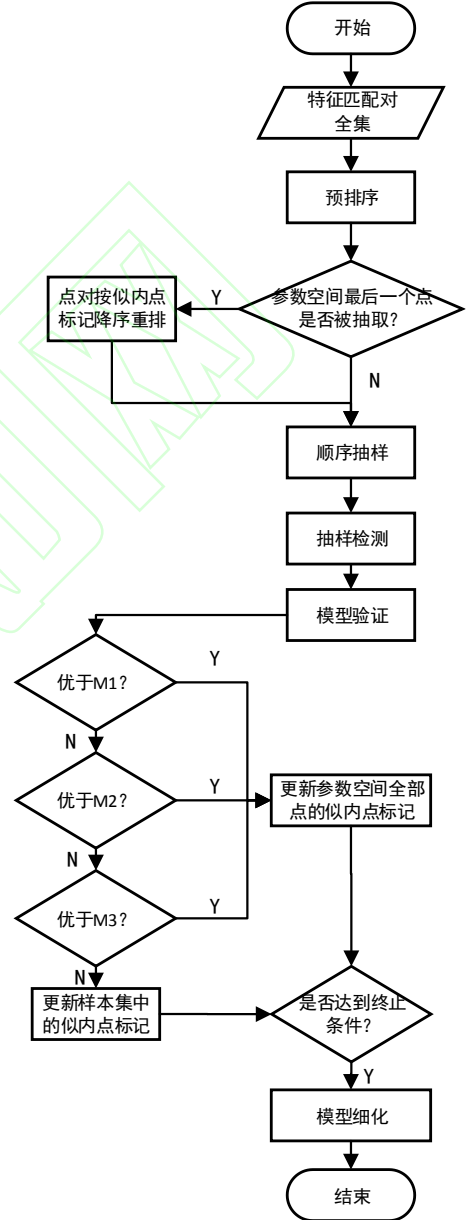


图 4 特征匹配一致性检测

2) 对匹配点对集合 $U_{\bar{P}}$ 进行顺序抽样, 并标记每对点对为内点的概率 η , 通过验证假设模型来更新其 η 值, 假设对于当前抽取的样本集 I^* 的假设模型为 M^* , 当前前三个最优模型为 M_1 、 M_2 和 M_3 , 对应样本集为 I_1 、 I_2 和 I_3 , 若 $I_k \in I^*$ ($k \in \{1, 2, 3\}$),

则更新概率 η_i 为:

$$\eta_i = \begin{cases} \eta_i + b_1, & u_i \in I^* \text{ 且 } u_i \in I \\ \eta_i + b_2, & u_i \notin I^* \text{ 且 } u_i \in I \\ \eta_i - b_1, & u_i \in I^* \text{ 且 } u_i \notin I \\ \eta_i, & u_i \notin I^* \text{ 且 } u_i \notin I \end{cases} \quad \left(I = \begin{cases} I_1 & I_1 \in I^* \\ I_2 & I_2 \in I^* \\ I_3 & I_3 \in I^* \end{cases} \right) \quad (10)$$

若更新了模型 M_1 , 则 M_3 被删除, 模型 M_1, M_2 变成新的 M_2 和 M_3 。对 M_2 和 M_3 的更新策略可依此类推。

3) 顺序取完匹配点对中的所有点对并更新 η 值后, 根据各个点对的 η 值进行重新排序, 重复执行2)中操作, 直到在某次操作后对匹配点对重排序时, 并未改变匹配点对的顺序。根据 M_1, M_2 和 M_3 分别计算其相对的运动模型:

$$f(M_k) = (q_k, t_k) \quad (k = \{1, 2, 3\}) \quad (11)$$

其中 q_k, t_k 为模型 M_k 对应的旋转四元数和平移距离, 定义:

$$\arg \min_k (diff_k) = \lambda_1 (q_k^{-1} \cdot q) + \lambda_2 (t_k - t) \quad (k = \{1, 2, 3\}) \quad (12)$$

其中 q, t 为上一帧的旋转四元数和距离, λ_1, λ_2 为常数。选择 $diff$ 值最小的模型作为最优模型 M 输出, 对应的匹配点对集合作为 I 输出。

4) 根据最优模型 M , 计算出相机运动的本质矩阵, 设匹配点对 $\{i_c, i_r\}$ 于集合 I , 其在当前帧中的特征点为 i_c 其在参考帧中的匹配特征点为 i_r 。 i_c 和 i_r 的归一化坐标为

$$\begin{aligned} i_c &= (a_c, b_c, 1) \\ i_r &= (a_r, b_r, 1) \end{aligned} \quad \{i_c, i_r\} \in S \quad (13)$$

根据对极约束:

$$(a_r, b_r, 1) \begin{pmatrix} e1 & e2 & e3 \\ e4 & e5 & e6 \\ e7 & e8 & e9 \end{pmatrix} \begin{pmatrix} a_c \\ b_c \\ 1 \end{pmatrix} = 0 \quad (14)$$

求得本质矩阵 $E = \begin{pmatrix} e1 & e2 & e3 \\ e4 & e5 & e6 \\ e7 & e8 & e9 \end{pmatrix}$ 。根据 E 可以计

算出两帧之间的旋转矩阵 R 和平移向量 t 。

$$E = t^{\wedge} R \quad (15)$$

其中 t^{\wedge} 是 t 的反对称矩阵。

2.5 位姿跟踪和建图

对于集合 U 中的特征点筛选分为以下几步:

1) 如图5, 对于当前帧中提取出的潜在动态特征点 $p_c (p_c \in U)$, 在参考帧中的对应匹配点为

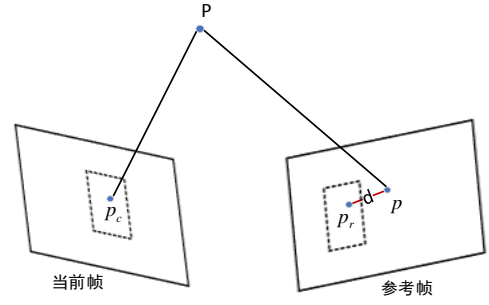


图5 示意图

p_r , 根据上文计算出的本质矩阵 E , 可以获得空间点 $P(X, Y, Z)$ 在参考帧中的投影点为 p , p 和 p_r 地齐次坐标为:

$$\begin{aligned} p &= (u, v, 1) \\ p_r &= (u_r, v_r, 1) \end{aligned} \quad (16)$$

满足:

$$\begin{cases} \sqrt{(u_r - u)^2 + (v_r - v)^2} \leq d \\ \sum_{1 \leq i \leq n} f_n^i(p) \oplus f_n^i(p_r) < \delta / 2 \end{cases} \quad (17)$$

(d 为设置的距离阈值)则保留特征点 p 加入到集合 S 中, 否则丢弃。

2) 对集合 S 中的特征点进行跟踪;

对于建图模块的操作:

1) 对跟踪模块输入的关键帧中的特征点集合进行筛选, 基于特征点集合 L 进行建图:

$$L = \{l, l \in S \text{ 且 } l \notin \bar{P}\} \quad (18)$$

2) 插入该关键帧到地图中, 具体操作不做赘述。

如图6, (a)中为输入图像; (b)为ORB_SLAM系统提取特征点后的图像帧; (c)为本方法对特征点进行筛选后传入跟踪模块的图像帧。



(a)输入图片



(b)ORB_SLAM2



(c)本方法

图6 本方法和ORB_SLAM2在动态场景下提取特征点情况对比

表 1 绝对路径误差(ATE)对比

data	ORB_SLAM2.ATE(m)			OUR_SYSTEM.ATE(m)			IMPROVEMENT(%)		
	rmse	mean	std	rmse	mean	std	rmse	mean	std
walking_static	0.4249	0.4083	0.1175	0.0088	0.0079	0.0040	97.93	98.06	96.59
walking_rpy	0.5614	0.4875	0.2816	0.0595	0.0472	0.0362	89.40	90.28	87.48
walking_halfsphere	0.5548	0.4423	0.3351	0.0394	0.0326	0.0220	92.89	92.62	93.43

表 2 相对位移误差(RPE)对比

Translational	ORB_SLAM2			OUR_SYSTEM			IMPROVEMENT(%)		
	rmse	mean	std	rmse	mean	std	rmse	mean	std
walking_xyz	0.3812	0.2709	0.2683	0.0275	0.0236	0.0141	92.80	94.30	94.76
walking_static	0.2359	0.1100	0.2086	0.0117	0.0104	0.0054	95.05	90.55	97.43
walking_rpy	0.3201	0.2154	0.2368	0.0821	0.0617	0.0541	74.37	71.36	77.15
walking_halfsphere	0.2752	0.1566	0.2263	0.0412	0.0344	0.0225	85.05	78.03	90.04

表 3 相对旋转误差(RPE)对比

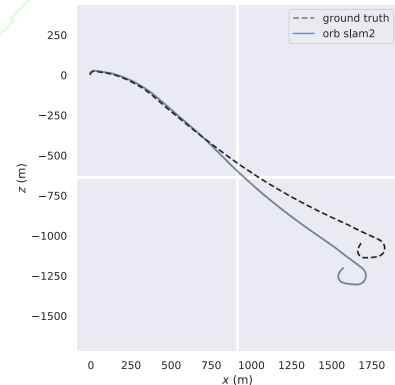
Rotational	ORB_SLAM2			OUR_SYSTEM			IMPROVEMENT(%)		
	rmse	mean	std	rmse	mean	std	rmse	mean	std
walking_xyz	7.7616	5.4991	5.4775	1.1595	0.9820	0.6165	85.06	82.14	88.74
walking_static	4.2410	2.0340	3.7215	0.4413	0.3913	0.2041	89.59	80.76	94.51
walking_rpy	6.3291	4.2921	4.6513	2.2104	1.8188	1.2561	65.07	57.62	72.99
walking_halfsphere	5.4652	3.2209	4.4152	1.7349	1.4942	0.8817	68.25	53.61	80.03

3 实验结果

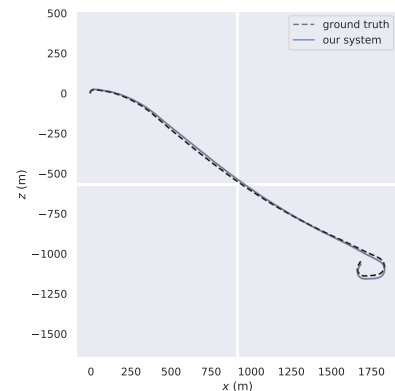
在本节中, 我们使用 TUM RGB-D 和 KITTI 数据集对系统进行了评估。我们将系统与原始 ORB_SLAM2 进行了比较, 将重点放在具有更多动态对象的序列的结果上。系统在动态环境中的性能是评价的重要组成部分。

3.1 KITTI

在 KITTI 数据集中, 一些移动的车辆和行人对跟踪和定位精度有影响。该系统将环境中的车辆和行人作为动态对象进行检测和处理。由于数据集中动态对象较少, 跟踪模块性能提升不明显, 但是可以有效减小没有闭环情况下的累积误差。从图 7 可以看出, 在没有闭环的情况下, 因为在地图中滤除了所有具有移动性的物体, 本方法的轨迹误差明显较小。



(a)ORB_SLAM2



(b)本方法

图 7 (a)为 ORB_SLAM2 绘制路径与实际路径对比图; (b)

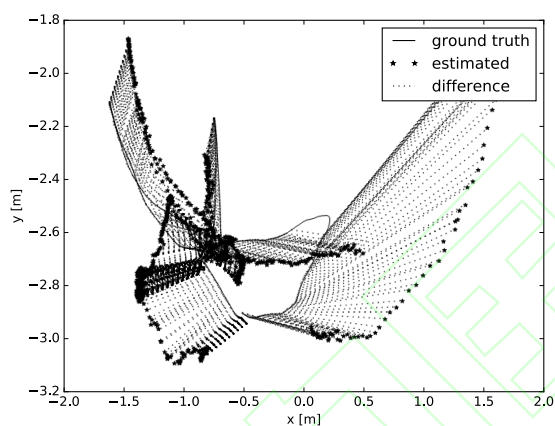
为本方法绘制路径与实际路径对比图。

3.2 TUM

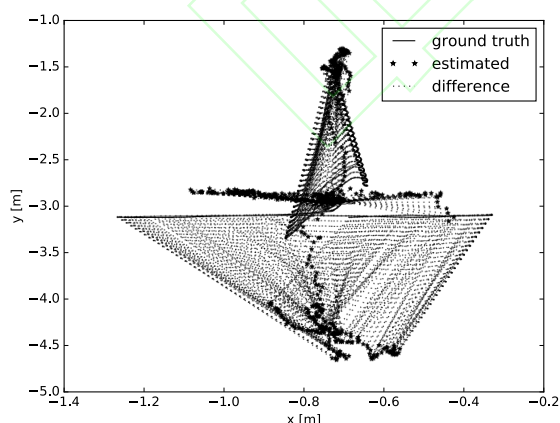
在 TUM 数据集上, 这个系统的性能明显优于 ORB_SLAM2。如图 8 所示, ORB_SLAM2 在序列 walking_xyz 和 walking_halfsphere 上的系统的轨迹跟踪误差较高。由于该方法主要针对场景中的动态对象。因此, 列出了在 TUM 数据集中动态目标较多的数据集上的运行效果, 表 1 中列出了本系统和 ORB_SLAM2 在绝对路径误差(ATE)方面的结果对比, 表 2 和表 3 分别列出了相对位姿误差(RPE)在平移和旋转方面的结果提升。其中 improvement 定义为:

$$\eta = \frac{our - ori}{ori} \times 100\% \quad (19)$$

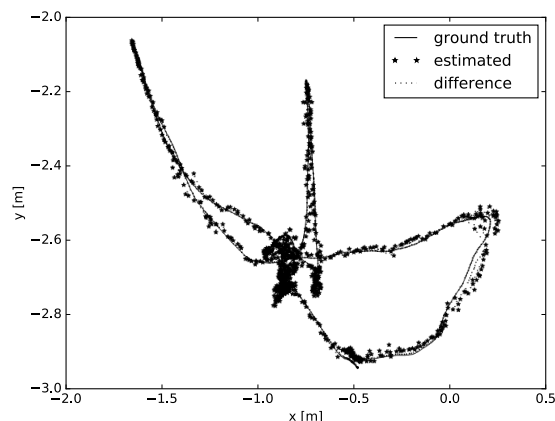
其中 *our* 为本文提出方法的运行结果, *ori* 为 ORB_SLAM2 运行结果。可以看出提升效果是显著的。



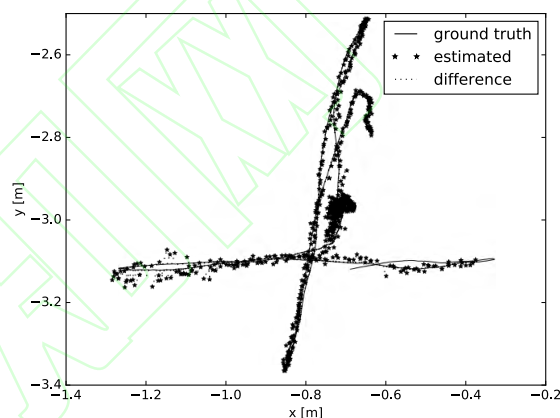
(a)ORB_SLAM2(walking_halfsphere)



(b)ORB_SLAM2(walking_xyz)



(c)本方法(walking_halfsphere)



(d)本方法(walking_xyz)

图 8 (a)、(b)为 ORB_SLAM2 绝对路径误差图, (c)、(d)为本方法绝对路径误差图。

4 结束语

为了减少动态对象对跟踪定位的影响, 提出了一种将基于深度学习的目标检测引入到基于视觉特征的实时 SLAM 系统。将特征点进行分类处理, 减少动态特征点给位姿跟踪和建图造成的误差, 与 ORB_SLAM2 相比, 系统的性能有了明显的提高, 并且系统的运行速度可以满足实时性的要求。通过 ATE 和 RPE 的比较, 该系统在大多数情况下都能达到较高的精度。在 TUM 动态对象数据集中, 该系统相对于 ORB_SLAM2 系统具有明显的性能提升。在 KITTI 数据集中, 对于那些动态对象较多的场景, 该系统比单目和双目 SLAM 系统更为精确。在未来的工作中, 我们将重点研究如何更详细的处理动态对象区域, 考虑到使用像素级语义分割算法对图像进行处理, 同时保证其实时性。

参考文献

- [1] Mur-Artal, Raul et al. "ORB-SLAM: A Versatile and Accurate Monocular SLAM System." *IEEE Transactions on Robotics* 31 (2015): 1147-1163.
- [2] Mur-Artal, Raul and Juan D. Tardós. "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras." *IEEE Transactions on Robotics* 33 (2017): 1255-1262.
- [3] Rublee, Ethan et al. "ORB: An efficient alternative to SIFT or SURF." *2011 International Conference on Computer Vision* (2011): 2564-2571.
- [4] Fang, Yuqiang and Bin Dai. "An improved moving target detecting and tracking based on Optical Flow technique and Kalman filter." *2009 4th International Conference on Computer Science & Education* (2009): 1197-1202.
- [5] Wang, Youbing and Shoudong Huang. "Towards dense moving object segmentation based robust dense RGB-D SLAM in dynamic scenarios." *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)* (2014): 1841-1846.
- [6] Bescós, Berta et al. "DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes." *IEEE Robotics and Automation Letters* 3 (2018): 4076-4083.
- [7] He, Kaiming et al. "Mask R-CNN." *IEEE transactions on pattern analysis and machine intelligence* (2018): n. pag.
- [8] Yu, Chao et al. "DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments." *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2018): 1168-1174.
- [9] Lianos, Konstantinos-Nektarios et al. "VSO: Visual Semantic Odometry." *ECCV* (2018).
- [10] Szegedy, Christian et al. "Scalable, High-Quality Object Detection." *ArXiv abs/1412.1441* (2015): n. pag.
- [11] Wen, Xuezhi et al. "Efficient Feature Selection and Classification for Vehicle Detection." *IEEE Transactions on Circuits and Systems for Video Technology* 25 (2015): 508-517.
- [12] Kotus, Józef et al. "Detection and localization of selected acoustic events in acoustic field for smart surveillance applications." *Multimedia Tools and Applications* 68 (2012): 5-21.
- [13] Wu, Tianfu et al. "Learning And-Or Model to Represent Context and Occlusion for Car Detection and Viewpoint Estimation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (2015): 1829-1843.
- [14] Dalal, Navneet and Bill Triggs. "Histograms of oriented gradients for human detection." *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* 1 (2005): 886-893 vol. 1.
- [15] 夏胡云,叶学义,罗霄晗,王鹏. 多尺度空间金字塔池化 PCANet 的行人检测[J]. *计算机工程*, 2019, 45(2): 270-277.
- [16] 张祯伟,石朝侠.改进视觉词袋模型的快速图像检索方法[J].*计算机系统应用*,2016,25(12):126-131.
- [17] Felzenszwalb, Pedro F. et al. "Object Detection with Discriminatively Trained Part Based Models." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (2009): 1627-1645.
- [18] Ren, Shaoqing et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2015): 1137-1149.
- [19] Girshick, Ross B. et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." *2014 IEEE Conference on Computer Vision and Pattern Recognition* (2013): 580-587.
- [20] Girshick, Ross B.. "Fast R-CNN." *2015 IEEE International Conference on Computer Vision (ICCV)* (2015): 1440-1448.
- [21] Redmon, Joseph and Ali Farhadi. "YOLOv3: An Incremental Improvement." *ArXiv abs/1804.02767* (2018): n. pag.
- [22] 贺黎,周传伟,张坤,石朝侠.基于顺序采样评估的二进制特征匹配方法[J].*计算机工程*,2017.