

# 面向三维多目标追踪的运动补偿优化方法

王顺洪, 张 昱\*, 沈江楠, 吉建民, 张燕咏

(中国科学技术大学计算机科学与技术学院, 安徽合肥 230027)

**摘要:** 三维多目标追踪是自动驾驶系统中的关键模块之一, 其结果的优劣主要取决于追踪模块中数据关联过程的准确度. 现有的追踪方法大多从外观特征或运动特征出发计算两帧之间物体的相似度, 而基于运动特征的方法通常根据当前帧和历史帧三维包围框之间的交并比 (Intersection over Union, IoU) 进行关联, 然而这种方式在观测点物体自身运动时存在严重缺陷. 在观测点物体自身进行运动时, 观测到的两帧数据将处于不同的局部坐标系, 导致无法使用运动模型准确预测已追踪物体在下一帧中的位置. 本文针对上述问题, 通过引入观测点自身的惯性测量单元 (Inertial Measurement Unit, IMU) 或全球定位系统 (Global Positioning System, GPS) 数据, 在一帧数据到达之后计算当前帧局部坐标系与上一帧局部坐标系之间的旋转和平移关系, 并对已追踪的物体状态按得到的坐标变换关系进行运动补偿, 使其抵消因观测点自身运动造成的偏移量. 这种运动补偿增强了追踪模块的数据关联环节, 提高追踪时三维包围框的关联成功率, 降低误关联数量, 改善三维多目标追踪的精度. 在相关追踪框架及 KITTI 数据集上的原型验证表明, 所提的运动补偿优化方法实现了 1% 左右的精度提升.

**关键词:** 自动驾驶; 运动补偿; 三维多目标追踪; 运动特征; 全球定位系统; 惯性测量单元

**基金项目:** 科技创新 2030—“新一代人工智能”重大项目 (No. 2018AAA0100500); 国家自然科学基金 (No. 62272434); 安徽省重点研究与开发计划标准化专项 (No. 202104h04020039)

**中图分类号:** TP399

**文献标识码:** A

**文章编号:** 0372-2112(2024)02-0528-12

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20220104

## Motion Compensation Optimization Method for 3D Multi-Object Tracking

WANG Shun-hong, ZHANG Yu\*, SHEN Jiang-nan, JI Jian-min, ZHANG Yan-yong

(School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China)

**Abstract:** Three-dimensional (3D) multi-object tracking is a key module in the autonomous driving system, and the quality of the tracking results mainly depends on the accuracy of the data association process in the tracking module. Existing tracking methods mostly calculate the similarity of objects between two frames from appearance characteristics or motion characteristics, while methods based on motion characteristics usually associate the current frame with the historical frame by using the intersection over union (IoU) of three-dimensional bounding box. However, this method has serious drawbacks when the observation point is moving. When the observation point is moving, the data observed in two frames would lie in different local coordinate systems, making it impossible to use the motion model to accurately predict the position of the tracked objects in the next frame. To solve the above problems, this paper introduces the inertial measurement unit (IMU) or the global positioning system (GPS) data of the observation point itself, and calculates the relationship of rotation and translation between local coordinate systems of the current and the previous frames after each frame data arrives then the state of the tracked object is compensated according to the obtained coordinate transformation relationship, making it counteract the offset caused by the movement of the observation point itself. This motion compensation enhances the data association process in the tracking module, improving the correlation success rate of the 3D bounding boxes, reducing the number of false correlations, and improving the accuracy of 3D multi-object tracking. The prototype verification on related

tracking frameworks and the KITTI dataset shows the proposed motion compensation optimization method achieves an accuracy improvement of about 1%.

**Key words:** autonomous driving; motion compensation; three-dimensional multi-object tracking; motion features; global positioning system(GPS); inertial measurement unit (IMU)

**Foundation Item(s):** Science and Technology Innovation 2030- Major Project (No.2018AAA0100500); National Natural Science Foundation of China (No.62272434); Anhui Provincial Key Research and Development Program Standardization Project (No.202104h04020039)

## 1 引言

随着商业界和学术界对自动驾驶技术的深入研究,自动驾驶相关的感知、决策等技术成果出现井喷式增长,而三维多目标追踪任务就属于感知任务中的关键一环.三维多目标追踪需要将当前帧中的物体与历史帧中的物体关联起来,同一物体在每帧数据中出现的位置构成该物体的运动轨迹.这对车辆行为预测以及行驶路径规划都有重要意义.

三维多目标追踪任务在对两帧之间的物体进行关联时,通常采用运动模型搭配递归滤波器实现运动状态的精确估计<sup>[1,2]</sup>.使用根据历史帧信息估计出的物体速度等状态信息,预测物体在当前帧的位置,并使用预测出的物体位置信息与当前帧检测到的物体进行距离、包围框相交程度等方面的相似性度量,以此为依据关联识别同一物体.然而与路侧固定传感器不同,对于自动驾驶系统而言,观测点自身往往存在不规则位移,这就导致系统感知到的当前场景中的物体在不同时刻下实际处于不同的局部坐标系.一方面,这会导致两帧之间同一物体的距离或方位产生很大差别,使距离、包围框相交程度等度量方式均会因为观测点自身的不规则位移而失去区分度,对帧之间的数据关联造成很大影响;另一方面,现有的方法<sup>[1]</sup>在使用滤波器估计物体运动的速度等相关信息时,实际估计出的是观测点与物体之间的相对运动,这与使用运动模型来预测物体自身运动的初衷不符.

本文针对以上问题提出了一种运动补偿优化方法,通过使用车载 GPS/IMU 传感器数据获得观测点自身的运动情况,在一帧数据到达之后计算当前帧局部坐标系与上一帧局部坐标系之间的旋转和平移关系,并对已追踪的物体状态按得到的坐标变换关系进行运动补偿,使关联时两帧数据中的物体处于同一局部坐标系下,从而优化追踪模块中的数据关联效果,提升追踪精度.

相关研究表明<sup>[3]</sup>,自动驾驶系统实时感知和处理的速度至少为 30 FPS(Frame Per Second),其中检测追踪最坏执行时间希望控制在 45 ms 以内<sup>[4]</sup>.然而, KITTI<sup>[5]</sup>等数据集<sup>[5-7]</sup>给出的数据频率是 10 FPS,即两帧之间时

间间隔最多为 100 ms.而主流三维多目标检测排行榜上的很多检测器都达不到这样的响应时延要求,这对追踪过程消耗的时间更提出了巨大的挑战.因此,本文提出的运动补偿优化方法将遵循快速轻量的原则,在有极低时间开销的代价下得到追踪框架的明显精度提升.

需要强调的是,本文提出的“运动补偿”概念与传统意义上应用于视频领域的“运动补偿”<sup>[8,9]</sup>不同.前者指在局部坐标系下获得的物体信息上补偿观测点的运动量,以获得在新的局部坐标系下对应的真实物体信息;后者指一种描述相邻视频帧(“相邻”在这里表示在编码关系上相邻,而在播放顺序上两帧未必相邻)差别的方法.这个区别需要在此重点强调,以免造成误解.

本文的主要贡献包括以下 4 个方面:

(1)提出了一种应用于三维多目标追踪领域的运动补偿优化方法,该方法能够对观测到的物体位置使用观测点自身的“运动”进行“补偿”,进而抵消观测点自身运动以提升关联准确度.

(2)针对提出的运动补偿优化方法,从 IMU 和 GPS 两种硬件设备出发,分别提出了基于 IMU 局部信息进行运动补偿和基于 GPS 全局信息进行运动补偿的方法;并对不同方法之间的优劣进行分析.

(3)提出了将运动补偿优化方法应用于传统三维多目标追踪框架中的模式,即在每一帧数据到达时,对之前保存的物体状态进行运动补偿.然后以 AB3DMOT 为例,展示了增加运动补偿步骤之后的数据流图,如图 3 所示.

(4)实验结果表明,提出的运动补偿优化方法能够在有极低时间开销的代价下,改善基于运动特征的三维多目标追踪方法的关联成功率,有效降低追踪精度评估指标中的 ID-switch 和 Fragmentations,实现追踪精度的有效提升.使用相关追踪框架在 KITTI 数据集上实现了 1% 左右的精度提升.

## 2 相关工作

三维多目标追踪在当前仍然是一个热门的研究课题.在现有的多个自动驾驶开源数据集上,如 KITTI<sup>[5]</sup>,

MOTChallenge<sup>[6]</sup>, nuScenes<sup>[7]</sup>等,每隔一段时间就会有新的方法占领榜首.许多追踪方法都尝试从一些新的角度出发来改善追踪精度,其中不乏有使用各种类型的神经网络结构进行关联的方法<sup>[10]</sup>.本文仅针对使用了物体运动特征的追踪方法,如AB3DMOT<sup>[1]</sup>等.

本文提到的运动补偿对多种使用物体运动特征进行数据关联的系统有效.观测点的运动导致不同帧之间的观测数据处于不同的局部坐标系,这对需要通过连续数据来感知场景的系统而言是一个很明显的障碍,而这样的系统在生活中十分常见.本文将运动补偿技术应用到三维多目标追踪系统中.

下面就本文的相关工作进行介绍.

## 2.1 三维多目标追踪

现有的三维多目标追踪方法大多是基于“检测-追踪”范式来实现的,这使目标检测和目标追踪变成了两个阶段的任务.对于每一帧数据,首先使用目标检测器检测出感兴趣的类别的物体,随后在帧之间关联检测到的物体,完成追踪任务.对于追踪任务,通常分为对物体位置的预测以及检测结果与预测结果的数据关联两个阶段.围绕这两个基本阶段,现有的工作尝试使用新的关联方法或者学习网络来提高多目标追踪的性能.AB3DMOT<sup>[1]</sup>使用卡尔曼滤波器进行运动预测,并使用匈牙利算法进行数据关联,实现实时三维多目标追踪.JRMOT<sup>[11]</sup>将二维RGB图像和三维点云的信息进行最优融合形成多模态特征,再通过行人重识别(Re-Identification)技术进行数据关联.FanTrack<sup>[12]</sup>提出了一种使用CNN进行数据处理的在线多目标追踪方法进行关联.GNN3DMOT<sup>[10]</sup>通过在新的特征交互机制中引入图神经网络,提高了数据关联的性能.这些都是基于本文所关注的“检测-追踪”框架的.不过本文提出的方法仅针对使用了运动特征的方法有效.

## 2.2 运动预测

运动预测环节对于追踪环节至关重要,直接影响到数据关联的准确度.现有的工作主要尝试根据不同的物体运动模型搭配滤波器建立起对单个物体运动的建模和预测过程.

### 2.2.1 运动模型

Schubert等<sup>[13]</sup>将几种常用的运动模型,如常速度运动模型(Constant Velocity, CV)、常加速度运动模型(Constant Acceleration, CA)、常转动速率运动模型(Constant Turn Rate and Velocity, CTRV),用于车辆轨迹预测并进行了评估,而这些模型同样被用于多目标追踪工作中.AB3DMOT使用了常速度运动模型,并使用卡尔曼滤波器预测出物体在相应的坐标系下的三个坐标轴上的速度.在物体运动较慢且为直线运动时,常速度运动模型往往能够较为准确地预测出物体在下一个时间

间隔可能出现的位置,然而实际情况中物体的运动可能伴随有比较强烈的加减速操作.PC3T<sup>[2]</sup>使用了常加速度运动模型,经试验发现,与常速度运动模型相比,也可以更好地预测被追踪物体的位置.

### 2.2.2 运动预测存在的问题

在为物体建立这些运动模型时,常常会忽略观测点自身也在运动的问题,这样估计出来的运动状态将是物体与观测点自身的相对运动,而观测点与被观测物体二者自身的运动会导致相对运动更加不规律.解决这个问题的关键在于如何消除观测点自身的运动.PC3T将所有检测到的物体坐标都通过观测点的GPS坐标转换到全局坐标系下,以此来消除观测点运动的影响,但是这种方法对GPS的精度有很高要求.为增强方法的普适性,本文除基于GPS的全局信息来消除观测点运动的影响,还进一步利用IMU得到的信息,尝试从局部的角度抵消不同帧的局部坐标系之间因观测点自身运动造成的偏差,以更准确地建立物体的运动模型.

### 2.2.3 对观测点运动预测的方法

在涉及多帧物体运动如物体检测或者光流等的任务中,为了消除观测点运动对数据造成的影响,现有的方法常常会使用神经网络对观测点的运动进行预测.如文献[14,15]中,使用神经网络估计出两帧数据之间观测点的运动,并将其作用于估计出的数据.此类方法不需要基于诸如IMU之类的具体硬件设备,然而,这类基于神经网络学习的方法在推理精度上受限,相较于通过IMU/GPS这些硬件设备进行定位的方法,对于观测点运动的估计存在相对更大的误差.同时,模型推理会增加很多时间和资源开销,这对运行在端侧的自动驾驶系统是难以接受的.而本文提出的融合IMU/GPS的方法快速轻量,能够在极低的时延下提供较为精确的观测点运动估计.

## 2.3 基于IMU的融合方法

利用IMU数据在帧间更新位姿的方法已经在SLAM和视觉-惯导里程计等领域中广泛应用<sup>[16,17]</sup>.利用车载IMU的实时数据能够较为精确地估计出两帧之间的位姿差距,进而构建出物体的完整运行轨迹.Bloesch等<sup>[16]</sup>利用扩展的Kalman滤波(Extended Kalman Filter, EKF)将视觉和IMU数据进行融合,构建了视觉-惯导里程计.Sahoo等<sup>[17]</sup>利用单目摄像头和IMU数据融合,构建了具有鲁棒性的SLAM系统.本文提出的方法在初衷上与这些方法一致,即通过传感器数据减小观测点在帧间运动时对观测数据造成的影响.不过,本文提出的方法相较于这些领域的方法有如下的差异和/或优势:

(1)本文提出的方法应用于基于激光雷达的3D多

目标追踪领域,而现有方法均基于视觉传感器;在应用于基于激光雷达的3D多目标追踪领域时主要针对被检测到的物体进行坐标系上的变换,而现有基于视觉传感器的方法通常为了估计观测点自身的整体运动.

(2)现有方法均使用IMU构建连续的轨迹或测量整体的过程,这会导致过程中存在累积误差.而本文提出的方法对IMU数据不进行连续更新,不存在累积误差.

(3)现有方法大多紧耦合地与传感器数据融合,而本文采用松耦合的方式,方便拓广到其他基于“检测-追踪”的框架进行增量式改进.

(4)本文提出的方法不仅可使用IMU数据实现,也可使用GPS数据或两者数据结合实现.在实现方法上更具灵活性.

### 3 运动补偿优化方法

#### 3.1 运动补偿

本文提出的“运动补偿”概念,是指在三维多目标追踪领域,对观测到的物体位置使用观测点自身的“运动”进行“补偿”。“运动补偿”操作使前后两帧的数据处

于同一个局部坐标系,消除了观测点自身的转向、加减速等运动的影响,对物体运动状态的预测和建模会更加准确.

如图1所示,假设某个在第 $t$ 时刻和第 $t+1$ 时刻被观测到的物体都在全局范围内没有移动,设该物体处于 $t$ 时刻、第 $t$ 帧局部坐标系下的物体状态为 $S_t^i$ (上角标 $t$ 表示处于第 $t$ 帧局部坐标系,下角标 $i$ 表示描述的是第 $t$ 时刻的信息,后同).如图1(a)所示,其被观测到的坐标为 $(x_t^i, y_t^i)$ .由于观测点自身运动,在第 $t+1$ 时刻局部坐标系发生了平移和旋转,则在第 $t+1$ 帧局部坐标系下,如图1(b)所示,如果按先前记录的物体坐标 $(x_t^i, y_t^i)$ ,记录的状态将远离物体的实际所在的位置 $(x_{t+1}^i, y_{t+1}^i)$ .包围框出现了与物体运动不符的偏移,这将给多目标追踪任务之后需要进行的运动预测和数据关联步骤带来困难.尤其是当观测点高速运动或进行剧烈转向时,这种现象尤为明显.传统的基于运动特征的方法往往会因此而发生无法关联甚至误关联的情况.显然,如果按照预定的假设,即该物体在 $t$ 和 $t+1$ 时刻没有移动,图1所示的未进行运动补偿的包围框位置将无法与新检测到的物体包围框位置进行关联.

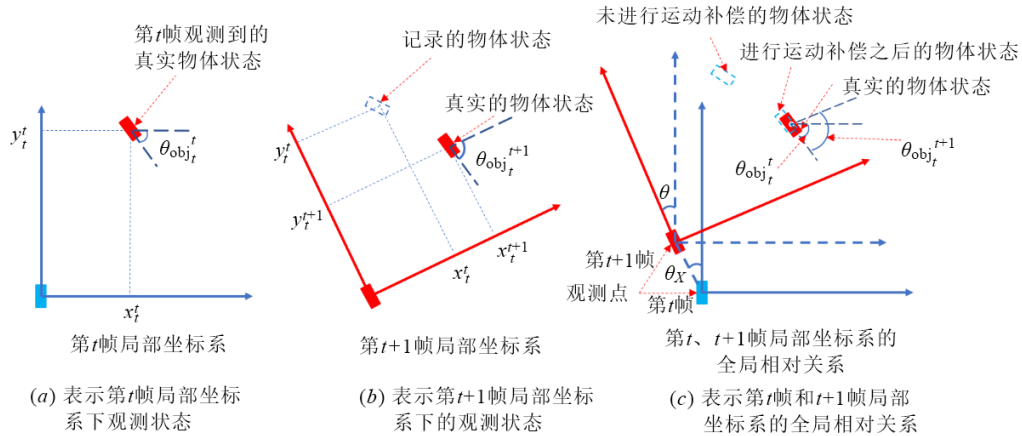


图1 运动补偿示意图

为了解决这个问题,在此提出运动补偿优化方法,通过观测点配备的GPS/IMU设备提供的运动信息,获得前后两帧局部坐标系之间的旋转矩阵 $R$ 和平移关系 $\Delta X$ .将第 $t$ 帧记录的每个物体位置 $X_t^i$ 都按照上述得出的旋转和平移关系进行坐标系变换得到第 $t+1$ 帧局部坐标系中的位置 $X_{t+1}^i$ :

$$X_{t+1}^i = R(X_t^i + \Delta X) \quad (1)$$

而估计得到的每个物体运动速度 $V_t^i$ 都按照上述得出的旋转关系进行坐标系变换,得到第 $t+1$ 帧局部坐标系中的速度 $V_{t+1}^i$ :

$$V_{t+1}^i = R V_t^i \quad (2)$$

另外,由于观测点与被观测物体之间相对位置的变化,需要对被观测物体的偏航角 $\theta_{obj_t}^i$ 做相应的调整,得到第 $t+1$ 帧局部坐标系中的偏航角 $\theta_{obj_t}^{t+1}$ :

$$\theta_{obj_t}^{t+1} = \theta_{obj_t}^i \pm \theta \quad (3)$$

其中, $\theta$ 为两个局部坐标系之间的旋转角.具体的加减情况需要依据角度定义确定.

上述三个过程即在进行所谓的“运动补偿”过程,得到的物体状态集合将变为 $S_{t+1}^i$ ,如图1(c)所示,它是仅与真实的物体状态 $S_{t+1}^i$ 有少许偏差的物体状态.这里的偏差主要取决于计算得出的坐标系变换关系的准确性.

当 $t$ 时刻保存的物体状态与 $t+1$ 时刻检测到的物

体状态处于同一局部坐标系下之后,通过滤波器等预测得到的物体运动特征就与观测点自身运动无关. 这使运动预测精度提高,进一步提升数据关联的成功率,减少追踪丢失的情况.

下面将从使用IMU局部信息进行运动补偿和根据GPS全局信息进行运动补偿两种途径详细描述运动补偿步骤所需的信息和相应的计算.

### 3.2 根据IMU局部信息进行运动补偿

惯性测量单元(Inertial Measurement Unit, IMU)能够提供速度、加速度等局部运动信息,被广泛应用于各个领域. 由于其价格低廉,一般的车载系统中均有配备,这为本文提出的运动补偿优化方法提供了硬件基础. 本节将详细描述如何利用IMU提供的局部运动信息计算两个时刻之间观测点(配备了IMU)的局部坐标系变换关系.

#### 3.2.1 IMU局部运动信息

通常而言,IMU能够提供以其自身为中心的三维直角坐标系的三个坐标轴方向上的速度、加速度和角速度信息. 由于这些信息无法提供观测点在全局范围内的运动信息,故将其称之为局部运动信息. 本文主要关注IMU垂直向上的轴上的角速度 $w$ ,以及水平面上的前向速度 $v_f$ 和左向速度 $v_l$ .

#### 3.2.2 旋转关系的计算

对两帧之间两个局部坐标系内同一个物体状态偏移影响最大的因素就是坐标系间的旋转关系(在很小的时间间隔内位移一般很小). IMU提供了角速度信息,由于只考虑局部坐标系在水平面上的旋转,故选取垂直向上的轴上的角速度 $w$ ,而在两帧之间能获取的信息只有一个初角速度和一个末角速度,因而可将观测点转向的过程视为一个匀角加速度旋转过程,于是第 $t$ 帧与第 $t+1$ 帧局部坐标系之间的旋转角 $\theta$ 为

$$\theta = \frac{w_t + w_{t+1}}{2} \Delta t \quad (4)$$

其中, $w_t$ 和 $w_{t+1}$ 分别表示第 $t$ 帧和第 $t+1$ 帧的角速度, $\Delta t$ 表示第 $t$ 帧与第 $t+1$ 帧之间的时间间隔.

得到旋转角之后,两帧之间的局部坐标系旋转矩阵即可依照特定的角速度方向以及坐标轴定义进行设置.

#### 3.2.3 平移关系的计算

虽然在时间间隔很短的情况下观测点的位移可能很小,但是如果不进行位移补偿,在某些情况下仍然会造成一定的关联失败问题,这一点将在第4节实验部分进行验证讨论. 为了计算局部坐标系之间的平移关系,我们选取IMU提供的前向速度 $v_f$ 和左向速度 $v_l$ . 与计算旋转角时的问题一致,由于在两帧之间能获取的信息只有初速度和末速度,因此可以将观测点位移的过程

视为一个匀加速运动过程,于是第 $t$ 帧与第 $t+1$ 帧局部坐标系之间的平移量 $\Delta X$ 为

$$\Delta X = \frac{v_{ft} + v_{ft+1}}{2} \Delta t + \frac{v_{lt} + v_{lt+1}}{2} \Delta t \quad (5)$$

其中, $v_{ft}$ 和 $v_{ft+1}$ 分别表示第 $t$ 帧和第 $t+1$ 帧的前向速度, $v_{lt}$ 和 $v_{lt+1}$ 分别表示第 $t$ 帧和第 $t+1$ 帧的左向速度, $\Delta t$ 表示第 $t$ 帧与第 $t+1$ 帧之间的时间间隔. 式(5)中向量指代的方向由旋转角 $\theta$ 与实际坐标轴定义确定. 在这里需要说明的一点是,计算位移时需要使用两帧局部坐标系原点连线与上一帧局部坐标系正前方的夹角作为位移方向的依据,然而在仅利用IMU局部信息时无法通过可用信息计算得到这个角度,因而使用两个局部坐标系之间的旋转角近似替代这个角,当然只能在两帧之间时间间隔很小时这样做.

### 3.3 根据GPS全局信息进行运动补偿

全球定位系统能够提供观测点在全局范围内的经纬度、海拔、偏航角等信息,是导航技术的基础. 本节将详细描述如何通过利用GPS提供的全局运动信息计算两个时刻之间观测点(配备了GPS)的局部坐标系变换关系.

#### 3.3.1 GPS全局信息

通常而言,GPS能够提供观测点在全局范围的经纬度、海拔、偏航角、横摆角、俯仰角信息. 由于这些信息指代观测点在全局范围内的姿态和位置关系,故将其称为全局运动信息. 本文主要关注GPS提供的经纬度坐标(lat和lon)以及偏航角 $\theta_{yaw}$ 信息,其中lat、lon分别表示纬度坐标和经度坐标.

#### 3.3.2 旋转关系的计算

由于GPS直接提供了车辆的偏航角 $\theta_{yaw}$ 信息,可以直接通过计算两帧之间偏航角的差得到局部坐标系之间的旋转角 $\theta$ ,即

$$\theta = \theta_{yaw\ t+1} - \theta_{yaw\ t} \quad (6)$$

其中, $\theta_{yaw\ t}$ 和 $\theta_{yaw\ t+1}$ 分别表示第 $t$ 帧和第 $t+1$ 帧的偏航角.

同样,在计算旋转矩阵时,需要根据 $\theta_{yaw}$ 的方向定义以及局部坐标系的坐标轴定义进行设置.

#### 3.3.3 平移关系的计算

使用全局信息计算两帧局部坐标系之间的平移关系时,需要采用不同于用IMU速度信息进行计算的方式. 如图2所示,计算时使用GPS提供的经纬度坐标(lat, lon). 由此根据大圆距离公式<sup>[19]</sup>可以得到观测点在两帧之间的位移距离:

$$\Delta d = R_{\text{earth}} \times \arccos \left( \sin(\text{lat}_t) \times \sin(\text{lat}_{t+1}) + \cos(\text{lat}_t) \times \cos(\text{lat}_{t+1}) \times \cos(\text{lon}_t - \text{lon}_{t+1}) \right) \quad (7)$$

其中, $R_{\text{earth}}$ 表示地球半径,lat<sub>*t*</sub>和lat<sub>*t+1*</sub>分别表示第 $t$ 帧和

第  $t+1$  帧观测点的纬度,  $\text{lon}_t$  和  $\text{lon}_{t+1}$  分别表示第  $t$  帧和第  $t+1$  帧观测点的经度.

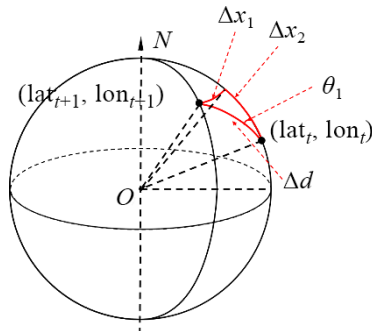


图2 大圆距离公式计算位移示意图

计算得到局部坐标系之间的位移距离之后,需要确定位移方向. 前文已提及,通过IMU局部信息进行运动补偿的方式无法通过可用信息计算得到位移方向的角度,而只能使用旋转角近似替代. 不过在使用GPS全局信息时,由于已知观测点在两帧中的全局坐标点,可以很容易获得相应的位移方向. 如假设俯航角是以逆时针为正方向、观测点正前方与正东方向的夹角,需要求得两帧局部坐标系原点连线与上一帧局部坐标系正前方的夹角  $\theta_x$ . 对于该夹角,可以考虑如图2所示由  $(\text{lat}_t, \text{lon}_t)$ ,  $(\text{lat}_{t+1}, \text{lon}_{t+1})$ ,  $(\text{lat}_t, \text{lon}_{t+1})$  三个点构成的三角形,再次利用大圆距离公式,算得各点间距离  $\Delta d$ ,  $\Delta x_1, \Delta x_2$  为

$$\Delta x_1 = R_{\text{earth}} \times \arccos\left(\sin^2(\text{lat}_{t+1}) + \cos^2(\text{lat}_{t+1}) \times \cos(\text{lon}_t - \text{lon}_{t+1})\right) \quad (8)$$

$$\Delta x_2 = R_{\text{earth}} \times \arccos(\cos(\text{lat}_t - \text{lat}_{t+1})) \quad (9)$$

由此可以求出两帧坐标系原点连线与南北方向的夹角  $\theta_1$  ( $\theta_1$  范围  $0 \sim \frac{\pi}{2}$ ), 即

$$\theta_1 = \arctan\left(\frac{\Delta x_1}{\Delta x_2}\right) \quad (10)$$

此时再通过前后两帧的经纬度判断第  $t+1$  帧局部坐标系在前一帧坐标系的具体方位,即处于第  $t$  帧坐标系的第几象限,用位移方向对  $\theta_1$  进行转换,求出前后两帧坐标系原点连线与正北方向的夹角  $\theta_2$  (角度范围为  $-\pi \sim \pi$ ), 则两帧局部坐标系原点连线与上一帧局部坐标系正前方的夹角  $\theta_x$  为

$$\theta_x = \frac{\pi}{2} + \theta_2 - \theta_{\text{yaw } t} \quad (11)$$

## 4 运动补偿优化方法在 MOT 的应用

### 4.1 对 AB3DMOT 的运动补偿

#### 4.1.1 原始追踪框架

如图3所示,原始 AB3DMOT 的追踪框架大致可以

分为运动预测、数据关联、状态更新等步骤. 未进行运动补偿时的数据流如图中红框箭头所示. 首先检测器给出  $t+1$  帧的检测结果  $D_{t+1}^{t+1}$ . 追踪器从之前存储的状态中取出第  $t$  帧的状态  $S_t^t$ , 经过运动预测步骤之后,获得先前状态在第  $t$  帧坐标系下的  $t+1$  时刻预测状态  $P_{t+1}^t$ . 随后对这两个集合进行数据关联,获得的关联结果  $A_{t+1}$  (包含关联成功的状态、未关联成功的状态、关联成功的物体、新出现的物体) 由状态更新步骤对原有的状态进行更新.

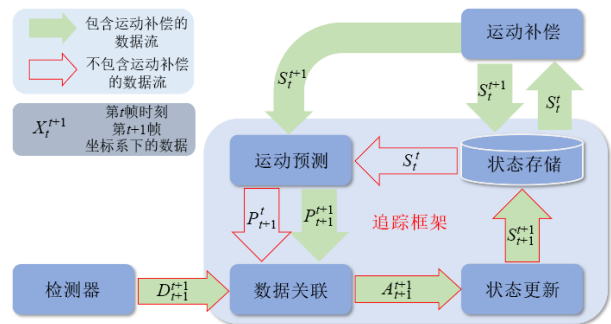


图3 AB3DMOT增加运动补偿步骤结构图

在这个过程中主要存在两个问题. 第一,在数据关联过程中,由于检测器给出的结果  $D_{t+1}^{t+1}$  与运动预测给出的预测状态  $P_{t+1}^t$  处于不同的局部坐标系,同一物体的预测状态和新检测到的状态之间可能存在特别大的差距,这种差距在观测点转向、高速运动等情况下尤其明显. 这样的差距会导致数据关联步骤几近失效,而只有在观测点低速直线运动时才有较小影响. 第二,在状态更新过程中,由于之前存储的状态为  $S_t^t$ , 而即将更新的状态为  $S_{t+1}^{t+1}$ , 二者处在不同的局部坐标系下,这会导致在更新物体运动状态时根据滤波器等方法估计出的运动为观测点和物体之间的相对运动,这与采用运动模型估计物体自身运动的初衷不一致.

#### 4.1.2 增加运动补偿步骤后的框架

在图3中,增加了运动补偿步骤之后的数据流以绿色箭头表示. 首先检测器给出第  $t+1$  帧的检测结果  $D_{t+1}^{t+1}$ , 追踪器从之前存储的状态中取出第  $t$  帧的状态  $S_t^t$ . 这个状态经过运动补偿之后得到第  $t+1$  帧局部坐标系下的第  $t$  帧时刻物体状态  $S_t^{t+1}$ , 一方面直接传递给运动预测模块,另一方面覆盖之前存储的  $S_t^t$  (需要注意这里的直接覆盖原有状态的行为与状态更新中的更新行为不同,这里不会更新对物体运动的估计). 经过运动预测步骤之后,获得先前状态在第  $t+1$  帧坐标系下的  $t+1$  时刻预测状态  $P_{t+1}^{t+1}$ . 随后对检测得到的结果  $D_{t+1}^{t+1}$  和预测状态  $P_{t+1}^{t+1}$  两个集合进行数据关联. 获得的关联结果  $A_{t+1}$  由状态更新步骤进行更新.

增加了运动补偿步骤之后的追踪框架解决了之前存在的两个问题. 第一,在数据关联过程中,检测器给

出的结果  $D_{t+1}^{t+1}$  与经过运动补偿之后运动预测给出的预测状态  $P_{t+1}^{t+1}$  处于相同的局部坐标系。这使观测点自身的运动被消除, 预测出的位置能更接近物体在  $t+1$  帧局部坐标系下的真实位置; 原先因为观测点转向、高速运动等情况造成无法关联的帧现在能够关联成功。第二, 在状态更新过程中, 由于之前存储的处于第  $t$  帧局部坐标系的状态  $S_t^t$  在经过运动补偿过程之后已经被覆盖为处于第  $t+1$  帧局部坐标系的状态  $S_t^{t+1}$ , 使用关联后的新状态  $S_{t+1}^{t+1}$  对  $S_t^{t+1}$  进行更新时, 二者处于同一局部坐标系, 使用滤波器等方法更新能够得到对物体真实运动更好的估计。而且这样的运动估计消除了观测点自身运动的影响, 与采用的运动模型估计物体运动的初衷一致。

#### 4.2 对 ThunderMOT 的运动补偿

根据 Bateni 等人所做的分析<sup>[4]</sup>, 自动驾驶系统中多目标追踪作为感知任务端到端最长时延不得超过 45 ms。然而, 根据对 KITTI 多目标追踪任务排行榜上开源工作的实际测量, 基于检测追踪的工作端到端响应时延普遍在 100 ms 以上, 远不能达到真实系统的实时性要求。由此本研究组提出一种面向自动驾驶的端到端实时三维多物体追踪框架 ThunderMOT, 实现在 10 Hz 数据采样频率下追踪响应延迟控制在 45 ms 以内。ThunderMOT 引入快速预测模块, 它根据当前被追踪物体的运动状态直接预测物体位置, 再将预测结果直接作为每帧的追踪结果进行输出; 而传统检测追踪中延迟不满足实时性要求的检测和关联模块会串行形成 ThunderMOT 的慢速更新模块, 在后台更新被追踪物体的运动状态。ThunderMOT 框架不是本文研究问题的重点, 其相关结构将在其他论文中详细阐述。在本文中仅以该框架为例, 对追求快速处理单帧数据而会因此损失一定精度的结构进行运动补偿优化分析。

由于其有快慢双分支的结构, 在其原有结构上做运动补偿也需要根据快慢分支的特点分别进行。对于快速分支而言, 原有框架通过在第  $t$  帧保存的状态上进行预测, 仅能得到在第  $t$  帧局部坐标系下的  $t+1$  时刻预测结果  $P_{t+1}^t$ 。而由于观测点自身无法预知的运动, 这个预测结果很有可能远离真实的物体位置。为此增加运动补偿步骤之后, 输出结果补偿为  $P_{t+1}^{t+1}$ , 在消除了观测点运动的影响之后, 同一局部坐标系下的预测结果将更加贴近真实的物体状态。对于慢速分支而言, 其整个过程与传统的追踪框架类似, 因而采用与第 3.1 节中所述的补偿模式即可。

#### 4.3 适用性分析

针对上述提出的几种运动补偿方案, 有必要在这里对这些方案在实际使用时的精度与适用性进行分析, 并讨论在现实情况中可能存在的问题和可能对精

度造成影响的干扰因素。

##### 4.3.1 误差累积问题

对于一个迭代更新的系统而言, 误差累积问题是需要进行分析的。如果误差在迭代更新过程中得不到消除, 那么累计误差对系统精度造成的影响将会越来越大, 甚至会导致系统出现错误, 这显然是不希望看到的。不过, 误差累积问题在本文提出的运动补偿优化方法中并不完全存在。虽然上述进行的运动补偿操作表面上看是一个迭代更新的过程, 但是对于关联成功的物体状态而言, 实际上每一次运动补偿作用的结果只会影响之后的一次数据关联过程, 关联成功后最新时刻的物体状态会以更新的形式覆盖掉经过运动补偿得到的上个时刻的物体状态。也就是说, 对于关联成功的物体状态而言, 运动补偿操作并没有“迭代”的过程。而对于未成功关联的物体状态而言, 由于一些追踪器有追踪状态生命周期管理参数的设置(如 AB3DMOT 中使用  $\max_{\text{ages}}$  表示被追踪物体在未被关联的情况下保留的最大帧数), 这些状态不会被马上删除, 而是会继续保存一定时间。那么在数据关联过程之后, 会使用预测出的状态进行更新, 这会导致下一次进行运动补偿时依赖的状态即是上一次运动补偿得到的状态预测得出的结果, 即产生了“迭代”的过程, 这里存在误差累积的情况, 但是这种情况无法被避免。因为对于一个未检测到物体而言, 无法拿到任何有关它的信息来消除误差。不过在现有的追踪器中, 物体关联失败后保存的帧数一般都很少(一般为 2~3 帧), 这样少量的累计误差是可以接受的。故而误差累积问题对本文提出的运动补偿过程影响很小。

##### 4.3.2 旋转角的计算误差

根据 IMU 提供的角速度信息计算旋转角, 本质上是一个积分过程, 但由于只能获得初速度和末速度, 故假设该过程为匀加速旋转, 这样计算得到的角度结果在计算的层面上就存在一定的误差。另外, 由于 IMU 本身存在一定的精度问题, 尤其是会因为剧烈加减速而产生大幅抖动, 这样的局部计算更有可能因此受到影响。一个比较好的解决方案是采用一个滤波器, 这样能够在一定程度上抑制 IMU 数据大幅抖动的情况, 而且对真实的 IMU 数据有一个比较好的估计, 这一点在后续的工作中是可以进行改进的。

而根据 GPS 提供的全局方位角计算观测点在两帧之间的旋转角, 这种方法直接计算两帧的方位角差即可, 故误差完全来源于 GPS 的精度。在 GPS 信号良好的区域内理论上来说误差比较小, 但是在隧道、山区等场景下可能无法得到很好的结果。坐标系的旋转对物体的坐标变化影响特别大。如果旋转角的计算出现很大误差的话, 可能会得到比不使用运动预测更差的结果。

这在实际使用时是一个无法避开的问题。考虑到这个问题,仍然可以采用一个滤波器对观测点的偏航角进行一个平稳估计,这在一定程度上能够降低因短时误差造成追踪系统失效的风险。

#### 4.3.3 位移的计算误差

与上述分析一致,使用IMU信息计算位移时,匀加速运动的假设在计算上就存在一定的误差。此外,由于局部信息的不完整性,本文提出的方法使用两个局部坐标系之间的旋转角来近似两帧局部坐标系原点的连线与上一帧局部坐标系正前方之间的夹角,这两个角实际上并不相等。再加上IMU的精度问题,实际位移的精度将受多个因素影响。同样,GPS的坐标精度直接影响了使用全局信息的位移计算,这取决于GPS的实际精度与稳定性。不过由于两帧数据之间的时间差一般很短(如引言所述,最多为100 ms),观测点在这个很小的时间间隔内的位移也相应的很小,利用局部信息计算时提到的角度近似也基本成立,故而位移的小范围的误差是可以接受的。

#### 4.3.4 多硬件时间同步问题

需要承认运动补偿优化方法在实际实施时存在一定的局限性。传感器的精度造成的误差对于此类使用多硬件设备的系统而言只是众多问题中的一个。另一个更重要的问题在于如何实现多硬件设备之间的时间同步,这是多硬件设备协同工作的首要条件。这个问题不在本文讨论的范畴,在此仅作为局限性提出。

## 5 实验评估

针对上述提出的运动补偿优化方法的实现,本节将着重设计实验评估运动补偿的效果。

首先,需要检验和评估运动补偿优化方法对改善追踪精度的效果。我们设计了一组对照试验,分别在AB3DMOT和ThunderMOT为追踪器实验对象,比较在其原有流程上添加运动补偿步骤与否的精度差别。由于运动补偿优化方法主要能够改善的是追踪的关联成功度,因此我们更加关注评估指标中的追踪中断次数(Fragmentations)以及追踪编号置换(ID-switch)的次数。

其次,对比使用IMU和GPS信息<sup>\*</sup>分别进行位移和角度差的计算,进而检验运动补偿的效果。我们设计的对照试验中,对运动补偿步骤有四种不同的实现方法,并通过四种实现方法的对比,分析实现方法对结果优劣的影响。

最后,为了证实本文提出的运动补偿优化方法轻量可行,不会在原有的追踪过程基础上增加过多时间损耗,我们统计了运动补偿过程消耗的时间,并对其在整个追踪过程中的时间占比进行分析和评估。

## 5.1 实验设置

实验环境为一台配置2个Intel Xeon CPU E5-2690 v3(每个CPU含有12物理核心,开启超线程),4个Geforce RTX 2080 Ti GPU(每个4352核心,12GB显存,实际测试时仅使用一个),256GB内存的服务器。服务器软件配置情况:操作系统为Ubuntu 18.04,Python版本为3.7.7,CUDA版本为10.2。

第一组实验中,采用SA-SSD作为检测器,AB3DMOT作为追踪器的结构进行实验。其中,AB3DMOT的参数设置为 $\min_{\text{hits}}=1$ ,  $\max_{\text{ages}}=2$ 。

第二组实验中,以ThunderMOT三维多目标追踪框架,采用SA-SSD作为检测器,物体运动模型采用常速度运动模型,状态管理参数同样设置为 $\min_{\text{hits}}=1$ ,  $\max_{\text{ages}}=2$ 。

追踪精度评估采用KITTI多目标追踪数据集,依照惯例,选择训练数据中的11个序列作为验证数据集,数据到达时间间隔 $\Delta t$ 为100 ms,共3908帧,387个物体,物体类型包括机动车、自行车和行人。在此,我们只评估在所有类型中数量最多的机动车类型,共217个物体。

追踪精度评估使用CLEAR MOT Metrics<sup>[18]</sup>中的MOTA, MOTP, Fragmentations, ID-switch指标,指标计算在KITTI官方提供的信息的基础上将二维包围框交并比更改为三维包围框交并比计算。

MOTA表示多目标追踪精度,计算方式如下:

$$\text{MOTA} = 1 - \frac{\text{FN} + \text{FP} + \text{ID}_{\text{switch}}}{\text{GT}} \quad (12)$$

其中, FN表示总的漏报数量, FP表示总的误报数量, ID-switch表示所有追踪目标身份交换的次数。

MOTP表示多目标追踪准确度,计算方式如下:

$$\text{MOTP} = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \quad (13)$$

其中,  $t$ 表示第 $t$ 帧,  $c_t$ 表示第 $t$ 帧中预测轨迹与真实轨迹成功匹配上的数目,  $d_{t,i}$ 表示第 $t$ 帧中第 $i$ 个匹配对之间的距离,本文使用三维边框的交并比来度量。

真实追踪轨迹由于某种原因发生中断并在后来又继续被准确追踪上的情况被称为追踪碎片。Fragmentations表示整个序列中碎片的总数。

被追踪目标的身份发生被错误更改时,称为身份跳变。在对序列的追踪过程当中, ID-switch表示所有追踪目标身份交换的次数。

## 5.2 针对KITTI数据集的计算配置

在KITTI数据集中,检测得出的物体坐标处于修正

<sup>\*</sup>数据集中的GPS和IMU数据已经完成同步,由于使用的是集成GPS和IMU的OXTS RT 3003设备,IMU和GPS的同步直接由硬件完成。在数据集中,GPS和IMU数据按照帧号给出,无需考虑同步问题。

相机坐标系中,因而我们的运动补偿过程也将基于修正相机坐标系展开。

KITTI数据集修正相机坐标系是以观测点相机为中心、正右方为 $x$ 轴正方向、正下方为 $y$ 轴正方向、正前方为 $z$ 轴正方向的右手系。数据集提供的GPS/IMU信息中:垂直向上方向上的角速度 $w_u$ 取逆时针为正;全局偏航角 $\theta_{yaw}$ 为观测点正前方与正东方的夹角,逆时针为正。

根据以上坐标系轴及观测量的方向定义,可以确定计算时需要采用的旋转矩阵和位移计算方式。

使用IMU的局部信息计算时,旋转角 $\theta$ 为

$$\theta = \frac{w_{ut} + w_{ut+1}}{2} \Delta t \quad (14)$$

旋转矩阵 $R$ 为

$$R = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \quad (15)$$

位移 $\Delta X$ 为

$$\Delta X = \begin{bmatrix} -\frac{v_f^t + v_f^{t+1}}{2} \Delta t \sin\theta - \frac{v_l^t + v_l^{t+1}}{2} \Delta t \cos\theta \\ 0 \\ \frac{v_f^t + v_f^{t+1}}{2} \Delta t \cos\theta - \frac{v_l^t + v_l^{t+1}}{2} \Delta t \sin\theta \end{bmatrix} \quad (16)$$

使用GPS的全局信息进行计算时,旋转角 $\theta$ 为

$$\theta = \theta_{yaw_{t+1}} - \theta_{yaw_t} \quad (17)$$

旋转矩阵 $R$ 同上。位移 $\Delta X$ 为

$$\Delta X = \begin{bmatrix} -\Delta d \sin\theta_x \\ 0 \\ \Delta d \cos\theta_x \end{bmatrix} \quad (18)$$

其中,取 $R_{earth} = 6371393$  m,则位移距离 $\Delta d$ 为

$$\Delta d = R_{earth} \arccos(\sin(\text{lat}_t) \sin(\text{lat}_{t+1}) + \cos(\text{lat}_t) \cos(\text{lat}_{t+1}) \cos(\text{lon}_t - \text{lon}_{t+1})) \quad (19)$$

另外,每个被观测物体的偏航角在KITTI数据集中被定义为 $\text{rotation}_y$ ,这个角为物体朝向与观测点修正相机坐标系 $x$ 轴正方向的夹角,顺时针为正。故当旋转角为 $\theta$ 时,物体的偏航角应更新为

$$\text{rotation}_{y_{t+1}} = \text{rotation}_{y_t} + \theta \quad (20)$$

### 5.3 精度评估

根据以上配置和计算方式,我们分别以AB3DMOT和ThunderMOT为实验对象,在KITTI数据集的11个序列的训练数据上进行测试。分别测试了以下五种配置的情况:未进行运动补偿的原始AB3DMOT运行结果(None Motion Compensation, NMC);使用IMU局部信息计算旋转角、使用IMU局部信息计算位移进行运动补偿的结果(Motion Compensation.IMU+IMU, MC.II);使用IMU局部信息计算旋转角、使用GPS全局信息计算位移

进行运动补偿的结果(Motion Compensation.IMU+GPS, MC.IG);使用GPS全局信息计算旋转角、使用GPS全局信息计算位移进行运动补偿的结果(Motion Compensation.GPS+GPS, MC.GG);使用GPS全局信息计算旋转角、使用IMU局部信息计算位移进行运动补偿的结果(Motion Compensation.GPS+IMU, MC.GI)。

使用AB3DMOT提供的三维多目标追踪评估工具以及KITTI官方提供的二维多目标追踪评估工具对实验结果进行评估,评估结果(表1)及其分析如下。

表1 用AB3DMOT提供的三维多目标追踪评测工具  
评测AB3DMOT的结果

配置/指标	MOTA	MOTP	ID-switch	Fragmentations
NMC	0.854 2	0.792 7	6	39
MC.II	0.855 0	0.804 7	0	31
MC.IG	0.855 7	0.804 5	8	39
MC.GG	0.855 7	0.804 5	8	39
MC.GI	0.856 9	0.804 8	0	31

#### 5.3.1 AB3DMOT三维多目标追踪评估结果

由于AB3DMOT提供的KITTI三维多目标追踪评估工具会在评估时根据检测器给出的物体置信度使用不同的阈值进行进一步过滤,最后选择结果最好的一个阈值给出评测结果,因此结果并不代表真实的检测及追踪结果,而只具有一定的参考意义。从表1可以看出,SA-SSD+AB3DMOT的“检测-追踪”框架在增加了运动补偿步骤之后(对比NMC与其他组),追踪精度MOTA和MOTP均有小幅提升。这说明本文提出的运动补偿优化方法在一定程度上能够提升三维多目标追踪的精度。

对比采用运动补偿的四种配置组别可以发现:使用IMU局部信息或GPS全局信息计算旋转角对结果精度的影响不大,这说明两种方法均有效。而从精度上看,使用GPS全局信息计算旋转角的方式(MC.GG和MC.GI)略微占优。从之前对方法的分析中我们也能体会到这一点,利用IMU提供的角速度信息计算旋转角本质上是一个积分过程,而对于只能获取始末数据的积分过程而言,存在误差是必然的。另外,使用IMU局部信息计算位移的方式(MC.II和MC.GI)会比使用GPS全局信息计算位移的方式(MC.IG和MC.GG)略优,这一点可能是由于GPS坐标的抖动或者是由于计算时全局到局部的转换之间存在误差。

#### 5.3.2 AB3DMOT二维多目标追踪评估结果

KITTI官方提供的评测脚本不存在AB3DMOT提供的评测脚本中的最优阈值选取的过程,而是直接使用原始输出结果进行评估,这是对数据的真实评估。从表2可以看出,在增加了运动补偿步骤之后(对比NMC与其他组),追踪精度均有小幅提升,这与三维多目标

表 2 KITTI 官方提供的二维多目标追踪评测工具评测结果

配置/指标	MOTA	MOTP	ID-switch	Fragmentations
NMC	0.776 7	0.865 9	15	91
MC.II	0.780 3	0.871 9	1	76
MC.IG	0.780 0	0.872 0	9	84
MC.GG	0.780 0	0.872 0	9	84
MC.GI	0.780 6	0.872 0	1	76

追踪评估结果一致;可以看到我们所关注的与关联成功度最相关的两个指标 ID-switch, Fragmentations 在增加了运动补偿步骤之后均得到了更好的结果,说明本文提出的运动补偿优化方法能够提高关联成功度以提升追踪精度。

由于 KITTI 数据集体量较小,而其中能够体现出观测点局部坐标系剧烈变化的情况更少. 为了更加明显地体现出本文提出的方法对性能的改善作用,特对 KITTI 数据集中包含转弯情况的 0014 号场景的精度结果进行进一步分析. 精度结果如表 3 所示. 在增加了运动补偿步骤之后,追踪精度有 2% 以上的提升,而这样的提升得益于大幅减少了观测点转弯过程中局部坐标系剧烈变化导致的数据关联失败,从 ID-switch, Fragmentations 两个指标的降低上可以明显发现这一点。

表 3 KITTI 官方提供的二维多目标追踪评测工具评测 AB3DMOT 的结果(仅 0014 序列)

配置/指标	MOTA	MOTP	ID-switch	Fragmentations
NMC	0.656 9	0.794 9	6	12
MC.II	0.681 3	0.851 2	0	6
MC.IG	0.681 3	0.851 1	0	6
MC.GG	0.678 8	0.850 9	0	6
MC.GI	0.678 8	0.850 8	0	6

### 5.3.3 ThunderMOT 三维多目标追踪评估结果

表 4 对比了运动补偿的四种配置组别,可以发现,精度规律与三维多目标追踪结果中分析的基本一致。

表 4 用 AB3DMOT 提供的三维多目标追踪评测工具评测 ThunderMOT 的结果

配置/指标	MOTA	MOTP	ID-switch	Fragmentations
NMC	0.836 7	0.757 4	0	45
MC.II	0.847 7	0.758 7	0	40
MC.IG	0.847 8	0.756 1	0	38
MC.GG	0.847 6	0.756 3	0	38
MC.GI	0.847 7	0.758 5	0	40

从表 4 可以看出,对于需要进行快速追踪任务而丢失一定精度的框架而言,在增加了运动补偿步骤之后(对比 NMC 与其它组),追踪精度有较为明显的提高。

对比运动补偿的四种配置组别,可以发现:,使用

IMU 局部信息计算位移的方式(MC.II 和 MC.GI)会比使用 GPS 全局信息计算位移的方式(MC.IG 和 MC.GG)在 MOTP 指标上略优,这是由于该框架直接使用预测得出的包围框作为结果,运动补偿的结果好坏在某种程度上直接影响了 MOTP 指标的优劣。

### 5.4 时间评估

针对上述对 AB3DMOT 和 ThunderMOT 分别进行的五组配置下的实验,为了证实本文提出的运动补偿优化方法轻量可行,我们分别统计了处理一帧数据的检测过程、追踪过程、总过程以及单个物体进行运动补偿消耗的平均时间\*,用以分析增加运动补偿步骤与否对时间消耗的影响。

AB3DMOT 上施加运动补偿的统计结果如表 5 所示. 从结果可以看出,本文提出的运动补偿优化方法并不会对原始的追踪过程造成很大的时间消耗,单个物体进行运动补偿时消耗的时间不超过 0.25 ms,对于 KITTI 数据集而言,系统中同时出现的物体不会超过 20 个,则单帧进行运动补偿所消耗的时间也仅是毫秒级. 从追踪时间统计可以看出,进行了运动补偿之后的单帧处理时间仅有 1~2 ms 的提升. 这证明了本文提出的运动补偿优化方法的轻量性。

对比运动补偿的四种配置组别,可以发现,使用 GPS 全局信息计算位移的方式(MC.IG 和 MC.GG)消耗时间会比使用 IMU 局部信息计算位移的方式(MC.II 和 MC.GI)略长. 从算法角度分析,对于单个物体的位移计算两种方式的时间复杂度均为  $O(1)$ . 不过使用 IMU 局部信息计算位移的方式(MC.II 和 MC.GI)仅需分别计算两个轴上的位移并进行矢量求和即可(式(5)),而使用 GPS 全局信息计算位移的方式(MC.IG 和 MC.GG)需要对 GPS 坐标进行多次三角函数计算(式(7)),且计算位移方向时同样需要类似计算(式(8)和式(9)). 这使使用 GPS 全局信息计算位移的方式消耗时间会是使用 IMU 局部信息计算位移的方式的常数倍。

ThunderMOT 上施加运动补偿的统计结果如表 6 所示. 由于对单个物体的运动补偿过程与 AB3DMOT 中一致,在此不再重复统计该项时间. 表中的慢速分支时间指单帧图像从输入检测器开始到完成单帧追踪任务的总时间,即“检测时间+追踪时间”,慢速分支时间可以对应于 AB3DMOT 实验中的总时间. 而快速分支时间为单帧响应时间,通过运动预测可以快速得到当前帧的近似状态,这个时间包含运动预测和运动补偿时间. 从结果可以看出,在增加了运动补偿步骤之后

\*对不同过程的时间测量,采用 Python 中的修饰器对相应的函数进行插桩,并使用 Python 自带的时间库 time 中的 perf\_counter() 接口进行时间测量。

表 5 AB3DMOT 上施加运动补偿消耗的时间统计

单位: ms

配置	检测时间	追踪时间	总时间	单个物体运动补偿时间
NMC	101.53	19.40	121.00	—
MC.II	95.47	20.65	116.19	0.1
MC.IG	99.65	22.00	121.73	0.22
MC.GG	97.11	21.69	118.87	0.21
MC.GI	98.05	20.86	118.98	0.11

表 6 ThunderMOT 上施加运动补偿消耗时间统计 单位: ms

配置	慢速分支时间	快速分支时间
NMC	113.54	1.79
MC.II	112.05	2.13
MC.IG	113.95	2.15
MC.GG	113.68	2.21
MC.GI	113.15	2.13

(对比 NMC 与其他组), 快速分支时间仅增加了不到 1 ms, 这说明本文提出的运动补偿优化方法十分轻量, 对于此类极致追求速度的追踪框架也不会 在时间上造成较大影响. 结合第 4.3 节中的精度提升, 可以看出本文提出的运动补偿优化方法是轻量且有效的.

## 6 总结与展望

本文主要从三维多目标追踪过程中因观测点自身运动导致追踪关联失败的问题展开, 提出了一种轻量的运动补偿优化方法: 通过利用 GPS/IMU 提供的观测点运动信息, 在两帧数据之间对观测到的物体进行运动补偿, 以此抵消观测点运动, 从而提升关联成功率, 进而提升三维多目标追踪精度. 实验及分析证明, 本文提出的运动补偿优化方法能够有效提升关联成功率, 并因此提高追踪关联精度, 而且耗时很短, 单帧仅需 1~2 ms 即可完成运动补偿步骤, 能够高效地完成优化任务. 对于一些极致追求性能的框架同样能够在有极低时间开销的代价下得到明显的性能提升, 不影响框架原先所追求的极致性能. 未来的工作主要包括: (1) 考虑使用滤波器对 GPS/IMU 得到的数据进行进一步滤波, 降低因硬件数据波动造成的误差影响; (2) 在进行运动补偿时使用更为精细化的计算方式, 减少因近似造成的不必要的误差.

### 参考文献

[1] WENG X S, WANG J R, HELD D, et al. 3D multi-object tracking: A baseline and new evaluation metrics[C]// ZHANG H. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE, 2020:10359-10366.

[2] WU H, HAN W, WEN C, et al. 3D multi-object tracking

in point clouds based on prediction confidence-guided data association[J]. IEEE Transactions on Intelligent Transportation Systems, 2021: 1-10.

- [3] DA SILVA CARVALHO M D, KOARK F, RHEINLÄNDER C, et al. Real-time image recognition system based on an embedded heterogeneous computer and deep convolutional neural networks for deployment in constrained environments[C]//SAE Technical Paper Series. Warrendale: SAE International, 2019. DOI: 10.4271/2019-01-1045.
- [4] BATENI S, LIU C. Predictable data-driven resource management: an implementation using autoware on autonomous platforms[C]//CHEN M. 2019 IEEE Real-Time Systems Symposium (RTSS). Piscataway: IEEE, 2019: 339-352.
- [5] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 3354-3361.
- [6] LEAL-TAIXÉ L, MILAN A, REID I, et al. Motchallenge 2015: Towards a benchmark for multi-target tracking[EB/OL]. (2015)[2022]. <https://arxiv.org/abs/1504.01942>.
- [7] CAESAR H, BANKITI V, LANG A H, et al. nuscenes: A multimodal dataset for autonomous driving[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020:11621-11631.
- [8] 钟平, 冯进良, 于前洋, 等. 动态图像序列帧间运动补偿方法探讨[J]. 光学技术, 2003, 29(4): 441-444.
- ZHONG P, FENG J L, YU Q Y, et al. Discussion on motion compensation method between frames of dynamic image sequence[J]. Optical Technique, 2003, 29(4): 441-444. (in Chinese)
- [9] NAKAYA Y, HARASHIMA H. Motion compensation based on spatial transformations[J]. IEEE Transactions on circuits and systems for video technology, 1994, 4(3): 339-356.
- [10] WENG X, WANG Y, MAN Y, et al. Gnn3dmot: Graph neural network for 3d multi-object tracking with 2d-3d

multi-feature learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 6499-6508.

- [11] SHENOI A, PATEL M, GWAK J Y, et al. Jrmot: A real-time 3d multi-object tracker and a new large-scale dataset [C]//ZHANG H. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE, 2020: 10335-10342.
- [12] BASER E, BALASUBRAMANIAN V, BHATTACHARYYA P, et al. Fantrack: 3D multi-object tracking with feature association network[C]//Arnaud de la Fortelle. 2019 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE, 2019: 1426-1433.
- [13] SCHUBERT R, RICHTER E, WANIELIK G. Comparison and evaluation of advanced motion models for vehicle tracking[C]//Jürgen Grosche. 2008 11th International Conference on Information Fusion. Piscataway: IEEE, 2008: 1-6.
- [14] BRAZIL G, PONS-MOLL G, LIU X, et al. Kinematic 3d object detection in monocular video[C]//Vittorio Ferrari. European Conference on Computer Vision. GLASGOW: Springer, 2020: 135-152.
- [15] JIA S, PEI X, JING X, et al. Self-supervised 3d reconstruction and ego-motion estimation via on-board monocular video[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 7557-7569.
- [16] BLOESCH M, BURRI M, OMARI S, et al. Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback[J]. The International Journal of Robotics Research, 2017, 36(10): 1053-1072.
- [17] SAHOO B, BIGLARBEKIAN M, MELEK W. Monocular visual inertial direct SLAM with robust scale estimation for ground robots/vehicles[J]. Robotics, 2021, 10(1): 23.
- [18] BERNARDIN K, STIEFELHAGEN R. Evaluating multiple object tracking performance: the clear mot metrics[J]. EURASIP Journal on Image and Video Processing, 2008, 2008: 1-10.
- [19] ROBUSTO C C. The cosine-haversine formula[J]. The American Mathematical Monthly, 1957, 64(1): 38-40.

### 作者简介



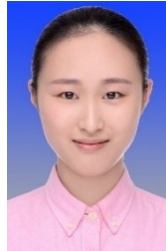
王顺洪 男,1998年生. 中国科学技术大学硕士生. 主要研究方向为边缘计算、自动驾驶系统的感知等.

E-mail: wangshunhong@mail.ustc.edu.cn



张昱 女,1972年生. 博士,教授,CCF杰出会员. 主要研究方向为面向新兴领域的编程系统与优化、如智能无人系统、量子计算等.

E-mail: yuzhang@ustc.edu.cn



沈江楠 女,2000年生. 中国科学技术大学本科生. 主要研究方向为自动驾驶系统的感知、机器学习等.

E-mail: jnshen.ustc@gmail.com



吉建民 男,1984年生. 博士,副教授,CCF会员. 主要研究方向为移动机器人、深度强化学习等.

E-mail: jianmin@ustc.edu.cn



张燕咏 女,1975年生. 博士,教授,CCF会员. 主要研究方向为边缘计算、人工智能物联网、无人系统的感知等.

E-mail: yanyongzhang\_ustc@ustc.edu.cn