

doi:10.3969/j.issn.1003-3106.2025.11.001

引用格式:刘明,方勇军,吴函,等.毫米波雷达与视觉一体化多目标协同感知与跟踪方法[J].无线电工程,2025,55(11):2131-2141. [LIU Ming, FANG Yongjun, WU Han, et al. A Collaborative Perception and Tracking Method Based on Radar and Camera [J]. Radio Engineering, 2025, 55(11): 2131-2141. ]

## 毫米波雷达与视觉一体化多目标协同感知与跟踪方法

刘明<sup>1,2,3</sup>, 方勇军<sup>2,3</sup>, 吴函<sup>2</sup>, 李乾坤<sup>2</sup>, 李冬冬<sup>2</sup>, 张朝阳<sup>1,3</sup>

(1. 浙江大学, 浙江 杭州 310058;

2. 浙江大华技术股份有限公司, 浙江 杭州 310053;

3. 全省视觉物联融合技术重点实验室, 浙江 杭州 310053)

**摘要:**在交通监控领域,雷视一体机对目标协同感知与跟踪时,对目标定位和测速有较高的跟踪精度要求。因透视成像原理,距离越远的目标对应的图像区域越小,且视觉目标检测框抖动显著。当雷视标定存在适量误差、遮挡或视觉目标框抖动时,远处视觉目标位置信息从图像坐标系映射到雷达坐标系时会存在较大的偏差,从而影响远处多目标跟踪的准确率。特别是多个传感器协同感知和跟踪目标时,进一步增加了目标跟踪的难度。针对上述问题,提出一种基于二阶段匹配和自适应卡尔曼滤波的多传感器多目标协同感知与跟踪方法。该方法在鸟瞰图(Bird's-Eye View, BEV)平面关联前后帧数据之后,增加图像视角(Perspective View, PV)平面的匹配过程,通过提升关联准确率,有效解决跟踪(较远)目标跟踪精度低的问题。基于图像点与距离位置抖动关系模型,提出自适应多传感器多目标跟踪方法,利用图像点与距离关系模型更新卡尔曼滤波器参数,根据目标传感器数据源,自适应选择合适的观测矩阵和测量协方差矩阵,对目标位置速度参数进行估计,有效提高对目标空间位置和速度的实时预测精度,进而提高 BEV 平面的目标关联准确率。实验结果表明,所提方法相较于未添加二阶段匹配策略且仅使用普通卡尔曼滤波器时多目标跟踪准确率(Multiple Object Tracking Accuracy, MOTA)指标提升 16.3%,显著提高了交通场景毫米波雷达和视觉一体机进行目标感知和跟踪的准确率。

**关键词:**多目标跟踪;卡尔曼滤波;多传感器跟踪;自适应卡尔曼;路侧场景

中图分类号:TP273.5

文献标志码:A

开放科学(资源服务)标识码(OSID):



文章编号:1003-3106(2025)11-2131-11

## A Collaborative Perception and Tracking Method Based on Radar and Camera

LIU Ming<sup>1,2,3</sup>, FANG Yongjun<sup>2,3</sup>, WU Han<sup>2</sup>, LI Qiankun<sup>2</sup>, LI Dongdong<sup>2</sup>, ZHANG Zhaoyang<sup>1,3</sup>

(1. Zhejiang University, Hangzhou 310058, China;

2. Zhejiang Dahua Technology Co., Ltd., Hangzhou 310053, China;

3. Zhejiang Provincial Key Laboratory of Harmonized Technology of Vision & Transmission, Hangzhou 310053, China)

**Abstract:** In traffic surveillance systems, radar-camera devices are used to collaboratively perceive and monitor the roadside environment. Due to the principles of perspective imaging, the greater the distance to a target, the smaller its corresponding pixel area in the image. Furthermore, the bounding boxes generated by visual detection exhibit significant jitter. If calibration errors or visual occlusion exist, or if the detection boxes shake, a significant error will be introduced when the target's position is mapped from the image coordinate system to the radar coordinate system, affecting tracking accuracy. This is especially true for collaborative target sensing and tracking with multiple sensors, which further increases the difficulty. To address these challenges, a multi-sensor, multi-target collaborative perception and tracking method is proposed, leveraging a two-stage matching strategy and an adaptive Kalman filter. This method improves association precision by adding a secondary matching strategy of Perspective View (PV) plane after the Bird's-

收稿日期:2025-07-06

基金项目:浙江省重点研发计划(2023C01021, 2025C01065);国家自然科学基金(62394292);中央高校基础研究基金(226-2024-00069);

Foundation Item: Zhejiang Provincial Key Research and Development Program(2023C01021, 2025C01065); National Natural Science Foundation of China(62394292); Fundamental Research Funds for Central Universities(226-2024-00069)

Eye View (BEV) plane is associated with the data of frame before and after. This effectively solves the problem of low tracking accuracy for distant targets caused by significant mapping errors. Based on the relationship model between image points and range-position jitter, an adaptive multi-sensor multi-target tracking method is proposed. By using the relationship model to update the parameters of the Kalman filter, and adaptively selecting the appropriate observation matrix and measurement covariance matrix according to the target sensor data source, the position and velocity parameters of the target are estimated. This effectively improves the real-time prediction accuracy of the target's spatial position and velocity, and further enhances the accuracy of target association in the BEV plane. Experimental results show that the proposed method improves the Multiple Object Tracking Accuracy (MOTA) index by 16.3% compared to the method without the two-stage matching strategy and only using the ordinary Kalman filter, significantly improving the accuracy of target perception and tracking in traffic scenes using millimeter-wave radar and vision integrated systems.

**Keywords:** multi-target tracking; Kalman filter; multi-sensor tracking; adaptive Kalman; roadside scene

## 0 引言

在智能交通领域,越来越多的相机或雷达代替人工检测违法违章危险行为。智能设备不仅能提高工作效率,还能降低成本<sup>[1]</sup>。相机可获取目标丰富的特征信息,如车牌、车型、车轮廓等信息,但是受光照和距离影响,在远处和夜间无法保持稳定有效的检测精度;雷达则可获取目标准确的位置、速度信息,且光照不影响其检测跟踪效果,在交通领域,较远距离的检测跟踪效果不会显著下降。但雷达无法获取目标的车牌、车型等视觉细节特征。雷达视觉一体机因同时具有二者的优势,尤其是在测距测速领域,目前已大量替代传统的单视觉相机。

雷视一体机协同感知和跟踪,可克服单一传感器的弊端,增强对目标的综合感知能力。首先是提升感知和跟踪范围。毫米波雷达和视觉一体机的有效检测范围可达 500 m,而视觉在 200 m 处检测性能已显著下降。其次,互相取长补短,克服单个传感器的弊端,提升整体感知能力。目标可被多个传感器感知和跟踪,对于近距离横向运动目标,雷达传感器感知能力弱,但视觉对横向目标跟踪能力强,因此对于横向运动目标,雷达权重会下降,视觉权重会增强。再如,夜间雷达目标跟踪能力强,视觉感知能力下降,此时雷达权重上升,视觉权重下降。此外,雷视多传感器协同感知和跟踪也面临诸多问题亟需解决,如快速准确的标定、高准确率高召回率的融合、实时且平滑的目标轨迹。标定是多传感器协同感知和跟踪的必要前提。在实际产品部署应用中,标定或多或少存在误差。以实际检测场景为例,目标在 200 m 位置时,视觉框抖动 1 pixel,映射到 BEV 坐标系之后的距离抖动可超过 2 m,此 PV 坐标系映射误差严重影响感知跟踪算法对目标关联的准确率。

针对雷视多传感器协同感知和跟踪的重要价值和应用中的种种问题,国内外学者提出多种方法。

在多传感器目标跟踪方面,主要有融合后跟踪、融合前跟踪、端到端的检测和跟踪,其中融合后跟踪为主流的跟踪方式,此类方法通常先将多个传感器信息融合为一个目标,取各自传感器的优势属性作为这个目标的输出属性,然后使用长短期记忆(Long Short Term Memory, LSTM)或卡尔曼滤波等预测方法获得目标信息下一帧位置,最后使用交并比(Intersection over Union, IoU)<sup>[2]</sup>、图结构<sup>[3]</sup>、外观特征<sup>[4]</sup>等方式对比预测值和量测值的特征相似度进行关联,完成对目标协同感知跟踪任务。相比其他跟踪方式,此类跟踪方法耗时少,但易产生更多的虚假跟踪,需要根据具体场景适配特定参数或差异化策略/方法才能取得较好的跟踪效果<sup>[5]</sup>。融合前跟踪方式则是先对每个传感器的信息进行跟踪,然后对各自传感器的轨迹进行关联,文献[6-8]均从不同角度对多传感器轨迹关联方法进行了改进,但此类依赖轨迹的方法需要目标有丰富的轨迹数据才能有较好的效果<sup>[9]</sup>,对于交通路侧监控这种实时性要求较高的场景并不适用。端到端检测和跟踪方法目前也取得了一定成果,Zeng 等<sup>[10]</sup>首次提出完全基于端到端的多目标检测跟踪模型,实现帧与帧之间的数据传输和信息更新,即无缝检测和跟踪,业界在此基础上对端到端检测跟踪模型展开了深入研究<sup>[11-13]</sup>。虽然端到端方法可以减少数据传递误差,有更好的研究前景,但从无人驾驶行业最新研究和应用可以看出,端到端方法需继续发展才能达到传统方法的性能,使用卡尔曼预测然后根据目标特征进行关联的方式<sup>[14]</sup>仍为现阶段效果最好的跟踪方式,不仅耗时较少,同时在多个数据集下多项指标测试下有综合优势。但无人驾驶行业的数据与路侧场景存在差异,方法无法直接复用。

本文在广大学者研究的基础上,针对路侧监控场景下的跟踪问题,提出了一种基于不同坐标系下的二阶段匹配方法和适用于多传感器感知跟踪的自

适应卡尔曼估计方法。该方法首先利用目标在不同坐标系下的属性准确率差异,解决(远距离)目标标定误差相对较大导致的跟踪准确率下降的问题;其次在状态预测环节,使用自适应卡尔曼估计方法,与普通自适应卡尔曼估计不同的是,所提方法不仅能根据传感器自适应地切换卡尔曼初始化时的先验估计协方差矩阵和量测更新时的观测矩阵,还能将各传感器的量测误差模型嵌入到卡尔曼估计方法中,实现对多传感器信息更充分的利用,最终达到提升跟踪准确率的目的。

## 1 多传感器多目标跟踪方法

多目标跟踪方法已被 Pang 等<sup>[15]</sup>总结并分解为预处理、数据关联、运动模型/更新、生命周期管理 4 个模块,目前主流跟踪方法均按此通用范式对各模块进行改进。本文以此通用范式为基础,提出针对交通路侧监控场景,多传感器密集多目标协同感知与跟踪的改进方法。

多传感器密集多目标跟踪的难点在于正确关联量测值和稳定更新滤波器。本文在交通场景雷达视觉协同感知和跟踪多目标时,在“关联模块”和“更新模块”提出创新性方法,有效地解决实际问题。下文基于关键点和创新点进行论述。

### 1.1 目标状态向量

目标跟踪方法对雷视数据帧实时流水式处理,其输入为单帧雷达检测结果和视频检测结果关联之后的融合目标,即每个输入目标同时具备视觉检测信息(目标框)和雷达检测信息(速度、位置)。本文方法也可兼容输入为未关联的视觉检测或雷达检测结果,只需按照单帧雷达视频目标检测数据关联方法对 2 个传感器检测结果关联为融合目标以及按照融合目标属性估计方法对单传感器检测结果补充为融合目标,再按融合目标的处理方式调用跟踪流程即可。一个融合目标的完整属性信息如下所示:

$$S_F = \{x, y, v_x, v_y, l, u, r, d, source\}, \quad (1)$$

式中: $x, y, v_x, v_y$  表示在雷达坐标系(世界坐标系)下的位置和速度, $l, u, r, d$  表示视觉目标框左上角和右下角的坐标, $source$  表示此融合目标属性信息的来源( $source = 1$  表示雷达, $source = 2$  表示视觉, $source = 3$  表示雷视)。当目标仅被视觉检测到时,融合目标属性中的  $x, y, v_x, v_y$  分别表示根据视觉检测框和标定信息估计的位置和速度;当目标仅被雷达检测到时,融合目标属性中的  $l, u, r, d$  表示根据雷达位置和标定信息估计的视觉框(左上角和右下角坐标)。

### 1.2 单帧雷达视频目标检测数据关联

如图 1 所示,车道 1 中的黑色车辆(最左侧车道号是 1),在单视觉传感器中检测结果  $ID = 1 (S_F^{V1} = \{*, *, *, *, l, u, r, d, 2\})$ ,在单雷达传感器中检测结果是  $ID = 1 (S_F^{R1} = \{x, y, v_x, v_y, *, *, *, *, 1\})$ 。对于跟踪器而言,若雷达视觉传感器均获得目标检测结果时,期望输入是完整属性  $S_F^{R1V1} = \{x, y, v_x, v_y, l, u, r, d, 3\}$ 。但是环境中还有其他目标(如 2 车道银色车辆和 3 车道白色车辆),需将单帧多个单传感器检测结果正确匹配为完整的融合检测结果(即理想完整正确的检测结果  $S_F^{R1V1}, S_F^{R2V2}, S_F^{R3V3}$ )。

匹配方法可采用 BEV 下满足预设阈值的最近邻匹配方法。但此匹配方法效果受标定、视觉目标检测、雷达目标检测效果影响。标定误差、视觉检测框抖动、视觉目标半遮挡、雷达目标位置波动,通常导致一个目标的视觉检测结果和雷达检测结果无法匹配。此时真实目标就有 2 个检测结果,即  $S_F^{R*}, S_F^{V*}$ 。因此,多传感器进行多目标协同感知与跟踪时,需考虑多传感器目标无法匹配的情况。

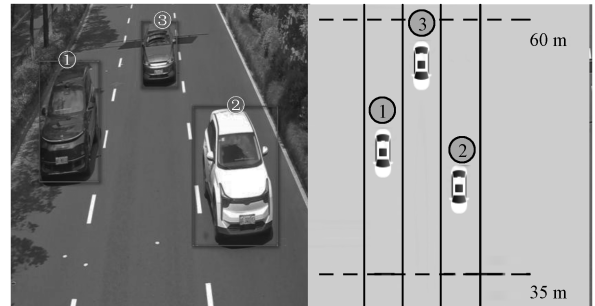


图 1 单帧雷达视觉传感器目标检测

Fig.1 Single-frame detection result of radar and camera sensor

再如因遮挡、雨雾、光照、目标较远或视野盲区时,视觉目标检测能力下降,此时一个车辆只有雷达目标检测结果,即仅有  $S_F^{R*}$ 。本文采用的是毫米波雷达,当目标静止时,其检测能力显著下降或存在视野盲区时,此时一个车辆只有视觉目标检测结果,即仅有  $S_F^{V*}$ 。因此,多传感器进行多目标协同感知与跟踪时,需考虑部分传感器失效的情况。

具体示例如图 1 中黑车和白车,视觉检测结果是  $S_F^{V1}$  和  $S_F^{V2}$ ;雷达检测结果是  $S_F^{R1}$  和  $S_F^{R2}$ 。按照上文分析,理想匹配结果是  $S_F^{R1V1}$  和  $S_F^{R2V2}$ 。差一点的情形有  $S_F^{V1}S_F^{R1}$  和  $S_F^{R2V2}$  或  $S_F^{R1V1}$  和  $S_F^{V2}S_F^{R2}$ ;再差一点的情形有  $S_F^{V1}S_F^{R1}S_F^{V2}S_F^{R2}$ 。若传感器“失效”则存在如下恶劣情况: $S_F^{V1}S_F^{R2}$  或  $S_F^{V2}S_F^{R1}$  等。

对于多传感器多目标协同感知与跟踪而言,

$S_F^{V1} S_F^{R1} S_F^{V2} S_F^{R2}$  是普遍现象。

### 1.3 融合目标属性估计

对于跟踪器而言,其观测量的维度和属性是确定的。对于多传感器协同感知与跟踪而言, $S_F^{R1V1}$  和  $S_F^{R2V2}$  是一种理想常见情形,但  $S_F^{V1} S_F^{R1} S_F^{V2} S_F^{R2}$  也是常见现象。对于跟踪器而言,无法直接处理  $S_F^{V1} S_F^{R1} S_F^{V2} S_F^{R2}$  格式的数据,因此需使输入数据送入跟踪系统前的目标属性完整,即需要针对只有单传感器检测结果的目标进行属性估计/补全。对于某一目标,若只被视觉传感器检测到( $S_F^{V*}$ ),则将目标框的下边缘中心点坐标作为这个目标在图像坐标系下的位置点,调用标定映射函数<sup>[16]</sup>投影到 BEV 坐标系下即可。

对于目标只被雷达传感器检测的情况( $S_F^{R*}$ ),因雷达检测信息稀疏,无法直接运用标定映射等方式计算目标框的大小。但考虑到视觉目标透视成像原理(目标距离越远目标检测框越小),可按式(2)、式(3)分别构建目标距离与目标宽、高的线性映射模型:

$$w = k_w y - 1 + b_w, \quad (2)$$

$$h = k_h y - 1 + b_h. \quad (3)$$

式(2)、式(3)中参数求解过程是在线采集数据、在线拟合、在线应用。在线采集数据是在设备运行期间采集适量的不同距离下关联成功的融合目标数据(分别取视觉框的宽高  $w$ 、 $h$  和雷达检测的距离  $y$ ),如表 1 所示。

表 1 在线数据集  
Tab. 1 Online data set

序号	距离值	图像目标宽高
1	$y_1$	$w_1, h_1$
2	$y_2$	$w_2, h_2$
⋮	⋮	⋮
$n$	$y_n$	$w_n, h_n$

在线拟合是将视觉框的宽高和雷达检测的距离  $y$ , 带入到式(2)、式(3)中,求得线性模型的参数  $k_w$ 、 $b_w$ 、 $k_h$ 、 $b_h$ 。在线应用是求得参数后,带入任意雷达距离,即可计算对应位置视觉目标的宽高信息,再将目标的雷达位置调用标定映射函数投影到图像坐标系下,作为目标的下边缘中心点,从而实现根据目标雷达距离计算目标框的目的。计算公式为:

$$(u_i^c, v_i^c) = F(x_i, y_i), \quad (4)$$

$$l = u_i^c - w/2, \quad (5)$$

$$r = u_i^c + w/2, \quad (6)$$

$$u = v_i^c - h, \quad (7)$$

$$d = v_i^c \quad (8)$$

通过上述处理流程, $S_F^{R*}$ 、 $S_F^{V*}$  均转换为  $S_F^{RV}$  (完整的位置、速度、目标框、信息源信息)。为后续跟踪流程提供更准确更全面的数据。其中  $(u_i^c, v_i^c)$  表示第  $i$  个视觉点坐标,  $(x_i, y_i)$  表示第  $i$  个雷达坐标点,  $F(\cdot)$  表示标定函数。

### 1.4 关联和跟踪原理

在路侧监控场景下,因传感器具备高帧率(20 Hz 以上)的检测效率,此帧率下机动车非机动车目标的运动可近似认为是常速度或常加速度模型,且监控设备的实时性要求计算量尽可能小,因此本文与大多数研究一样,使用卡尔曼常速度模型对运动系统建模。卡尔曼滤波器的更新方程如式(9)~式(15)所示:

$$\mathbf{T}_{\text{bev}} = \{x, y, v_x, v_y\}, \quad (9)$$

$$\mathbf{T}'_{\text{bev}} = \mathbf{A} \mathbf{T}_{\text{bev}}, \quad (10)$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (11)$$

$$\mathbf{P}' = \mathbf{A} \mathbf{P} \mathbf{A}^T + \mathbf{Q}, \quad (12)$$

$$\mathbf{K} = \mathbf{P}' \mathbf{H}^T (\mathbf{H} \mathbf{P}' \mathbf{H}^T + \mathbf{R})^{-1}, \quad (13)$$

$$\mathbf{T}_{\text{bev}} = \mathbf{T}'_{\text{bev}} + \mathbf{K} (\mathbf{D}_{\text{bev}} - \mathbf{H} \mathbf{T}'_{\text{bev}}), \quad (14)$$

$$\mathbf{P} = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}', \quad (15)$$

式中:  $\mathbf{T}_{\text{bev}}$  为世界坐标系下的目标位置和速度,  $\mathbf{A}$  为状态转移矩阵,  $dt$  为 2 帧的时间间隔,  $\mathbf{P}$  为上一时刻的最优估计协方差矩阵,  $\mathbf{P}'$  为当前的先验估计协方差矩阵,  $\mathbf{Q}$  为过程噪声协方差矩阵,  $\mathbf{K}$  为卡尔曼增益矩阵。从卡尔曼的更新式中可以看出,卡尔曼滤波器的先验估计协方差矩阵  $\mathbf{P}$ 、将实际测量值与状态量对齐的观测矩阵  $\mathbf{H}$ 、传感器测量噪声  $\mathbf{R}$  都会对滤波器的更新造成较大影响,且这些参数也与具体的传感器存在相关性。

基于跟踪原理,对于跟踪器而言,影响跟踪效果因素有如下 5 点:① 初始值因素;② 预测方向因素;③ 量测因素;④ 更新因素-量测;⑤ 更新因素-传感器精度。跟踪器收敛发散图如图 2 所示,绿色实线是真值,红色虚线是可接收的收敛区域(目标轨迹收敛在红色区间内,即视作跟踪器是稳定且跟踪精度满足要求)。轨迹 1 是发散的,即目标跟踪失败。轨迹 2 亦是发散,尽管初期有收敛趋势,但后期发散。轨迹 3 和轨迹 4 均是收敛轨迹(正确稳定跟踪),但由于轨迹 3 的初始状态好于轨迹 4,因此更快速地收敛。

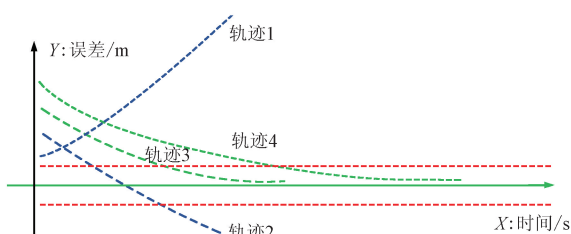


图 2 跟踪器收敛发散图

Fig. 2 Convergence and divergence of tracker

关联错误是导致轨迹发散(长期)或抖动(短期)的重要原因。如图 3 所示,红色轨迹点是当前帧目标预测值,蓝色点是量测点,真实的匹配对是轨迹 1-量测 1,轨迹 2-量测 2。若关联错误,如量测 2 更新轨迹 1,则会导致轨迹 1 的运动向下(量测点 2 的方向偏离),进而导致短期内抖动;若轨迹 1 后续未能被正确量测更新,则会导致发散(目标跟踪错误/失败)。

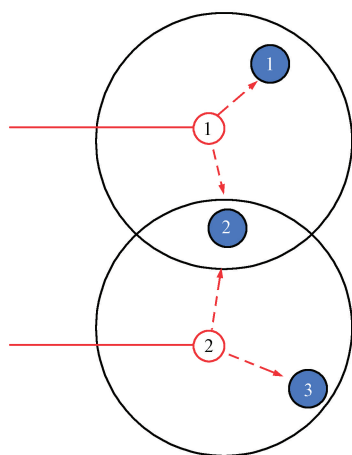


图 3 量测更新示意

Fig. 3 Schematic diagram of updating by measurement

### 1.5 雷达视频多传感器多目标跟踪中问题

在雷达视觉协同感知和跟踪目标过程中,复杂场景下,雷视一体机面临遮挡、目标拥挤等问题。如图 4 所示,如果目标拥挤,行人、非机动车、机动车均存在,导致目标跟踪错误的概率大大增加,进而触发违法违规误报事件。但对于传感器特性而言,近距离视觉检测效果好(目标类型、定位精度),基于此在近距离优先采用视觉信息更新滤波器;远处雷达检测效果好,基于此在远距离优先采用雷达信息更新滤波器。因此多传感器协同感知与跟踪,需针对目标的传感器信息合理选用准确数据才能充分发挥多传感器协同感知跟踪的优势。

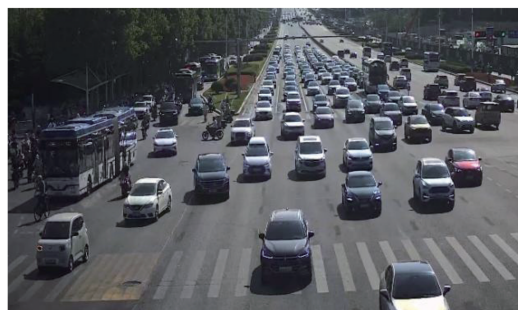


图 4 交通复杂场景

Fig. 4 Traffic congestion scene

在密集目标场景或远距离区域,跟踪失败/轨迹发散的主要原因是关联和更新。关联是指采用正确的量测数据更新对应目标轨迹;更新是指采用雷达位置和速度信息还是视觉目标框信息更新对应目标轨迹(如何充分利用多个传感器信息)。

本文针对交通密集场景或远距离场景,毫米波与视觉一体机协同感知与跟踪目标中的关联难题,提出两阶段数据关联方法,有效提升关联准确率,进而提高目标稳定准确跟踪的概率。针对如何充分利用多个传感器信息的难题,提出自适应更新方法,提高目标稳定跟踪的概率。

### 1.6 两阶段数据关联

在数据关联环节,主要是将当前的检测结果与上一帧已有的轨迹进行关联。由于 2 帧数据存在时间差,需要对上一帧轨迹执行预测步骤,预测的当前帧位置再与当前检测结果进行关联操作。在图像领域,一般是预测目标框的位置,然后使用 IoU、改进 IoU<sup>[17]</sup>等方式进行数据关联。但是在多传感器场景,雷达只能检测位置信息,即便通过预处理环节的目标框拟合方式构建虚拟框,也会因定位或标定误差原因造成较大的关联误差,因此,优先选择在 BEV 下使用距离方式进行关联更能准确利用不同传感器的信息。而在远处,视觉点映射到 BEV 坐标系之后,其位置误差或许已严重超出正常的阈值范围,单纯加大关联阈值的方式又会引入更多的误关

联问题。则此时在图像坐标系下进行关联能获得更准确的关联结果。

相关研究显示<sup>[18-19]</sup>,在多个模态的数据关联时,使用两阶段数据关联方法是一种有效的关联方法,本文借鉴两阶段关联方法,具体的流程如图5所示,但与之前的研究不同,本文两阶段匹配不是在同一视角内调整阈值,而是在不同的视角下根据传感器特性先后完成关联。在本文提出的关联算法的第一阶段,使用欧氏距离作为2个目标的代价函数,能将位置映射准确的目标优先关联,第一阶段关联之后,剩余未关联的目标均因位置映射误差超出阈值而无法关联。从后文图6中可以看出,距离越远每个点代表的距离就会越大,而视觉检测时目标框跳

动几个点是正常的检测误差,但其映射到 BEV 坐标系之后,误差就会急剧增大,针对这种现象,本文在第二阶段的关联在图像坐标系下进行,雷达映射到图像坐标系之后,其测量误差与图像一致,即使偏离几个 pixel,2 传感器产生的目标框仍为重叠状态,此时使用 IoU 距离作为 2 个目标的代价函数进行关联,可有效解决第一阶段遗留的误差问题。

在图5所示的流程中,输入的  $T_{bev}$  和  $D_{bev}$  分别代表 BEV 下目标上一帧的轨迹信息的最优估计值和目标当前帧的量测信息。轨迹信息与量测信息的具体内容与式(1)结构相同。 $T_{bev}$  需要先使用卡尔曼预测方程预测轨迹在当前的位置  $T'_{bev}$ ,计算方法见式(10)、式(11)。

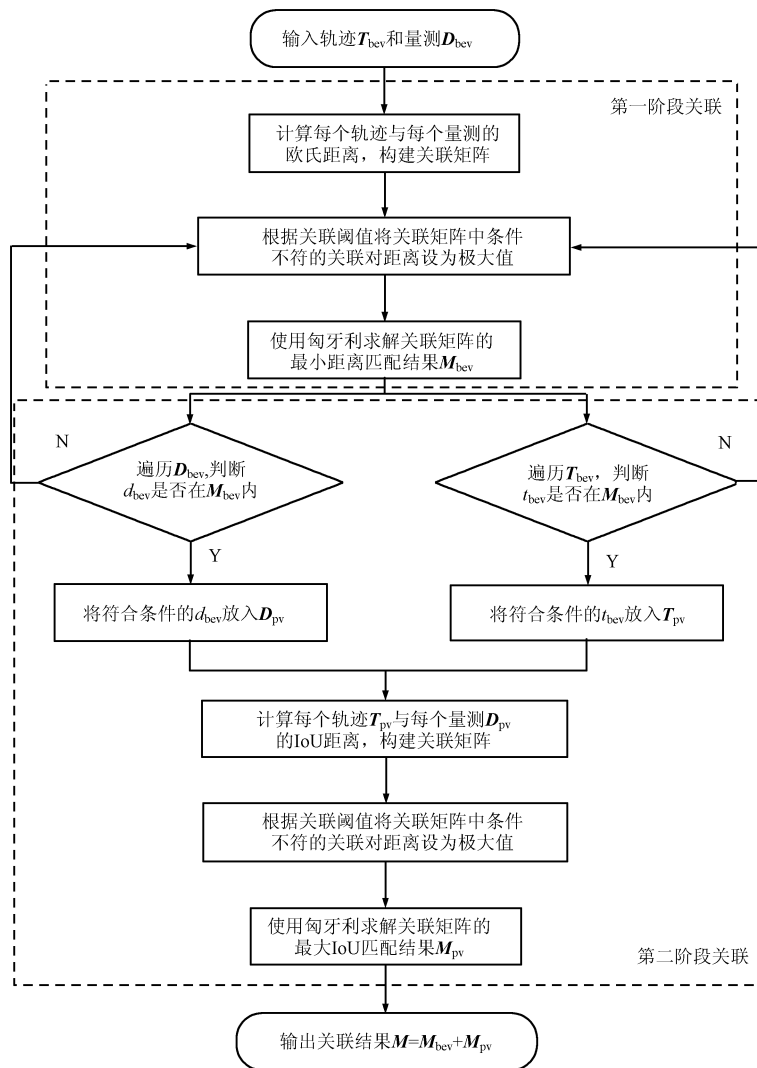


图5 两阶段关联流程

Fig. 5 Flowchart of two-stage association

### 1.7 自适应更新方法

由于路侧监控设备独特的视角和多传感器的协同感知与跟踪原因,一般的卡尔曼滤波器参数定义方式很难充分利用多个传感器信息,且这些参数也与具体的传感器存在相关性,例如雷达在远处测量精度更高,视觉近处测量精度更高。因此,需基于传感器信息,合理自适应选用准确的数据才能充分发挥多个传感器协同感知与跟踪的优势。

为了提高雷达和视觉在不同距离下的位置估计精度,需要对目标距离和传感器定位误差进行建模。然后根据此定位误差模型构建卡尔曼自适应参数。首先,建立雷达传感器误差模型,雷达传感器的检测方式是主动发射毫米波,通过物体反射的方式完成测量,而反射过程中,目标越大,则反射点越难确定这个目标的精确位置,因此,雷达目标的传感器误差为目标实际长度,考虑到监控场景下目标均为来向或去向运动,其反射点通常为目标下半部分的某一点,可将传感器误差缩小为目标长度的一半。以机动车为例,雷达传感器误差  $E_R$  如下:

$$E_R = \begin{bmatrix} w_T/2 \\ l_T/2 \end{bmatrix}, \quad (16)$$

式中:  $l_T$ 、 $w_T$  分别表示小型车辆的长和宽。

对于视觉传感器来说,其检测的特点是距离越远,目标图像点越少,检测的距离误差越大,图 6 为视觉点根据标定关系映射到 BEV 坐标系之后的效果图,从图中可以证实,近处的视觉定位误差远低于雷达传感器,远处的定位误差则远高于雷达。根据此检测特性,构建的视觉传感器测量误差模型  $E_V$ ,如式 (17) 所示。



图 6 图像点映射 BEV 示意  
Fig. 6 Pixel to BEV mapping

$$E_V(Y) = YM - (Y + \alpha)M, \quad (17)$$

$$Y = \begin{pmatrix} u \\ v \end{pmatrix}. \quad (18)$$

视觉误差模型的原理是计算 2 个图像点在 BEV 坐标系下的距离,其中  $Y$  表示图像坐标, $M$  表示图像坐标系到 BEV 坐标系的标定映射矩阵,考虑到视觉框检测时会存在几个 pixel 的抖动,视觉误差模型计算误差时还需要加上目标检测时横向纵向的抖动范围  $\alpha$ 。

确定 2 个传感器的测量误差模型后,若给定任一位置,均可获得传感器的定位误差。在此前提下,构建卡尔曼自适应参数。根据式 (9) ~ 式 (15) 的卡尔曼更新方程,需要初始化的参数有先验估计协方差矩阵  $P$ 、观测矩阵  $H$ 、测量噪声矩阵  $R$ 、过程噪声协方差矩阵  $Q$ 。 $P$  参数对滤波器收敛有重要影响,不符合实际的  $P$  参数会让模型始终无法收敛或轨迹波动大,为保持  $P$  参数在合理的范围内更新,根据目标初始位置初始化  $P$  参数,如式 (19) 所示形式:

$$P = \begin{bmatrix} [1 \ 0]E & 0 & 0 & 0 \\ 0 & [0 \ 1]E & 0 & 0 \\ 0 & 0 & [1 \ 0]E_v^T & 0 \\ 0 & 0 & 0 & [0 \ 1]E_v^T \end{bmatrix}, \quad (19)$$

$$E = \begin{cases} E_R, E_R < E_V(Y) \\ E_V, \text{other} \end{cases}, \quad (20)$$

$$E_v = \begin{cases} [1 \ 1], \text{source} = 1 \\ [\max \ \max], \text{other} \end{cases}, \quad (21)$$

式中:  $E$  表示在指定位置  $Y$  时的 2 传感器最小测量误差,  $E_v$  表示速度的测量误差,当目标源为雷达时,其速度测量误差为 1,当目标源为视频时,速度测量误差为最大值,表示对速度源的不可信状态。

对于观测矩阵,其意义是将预测量转换到观测量的同一个向量空间内,从而使用观测量对预测量进行修正,但在多传感器场景,雷达传感器与视觉传感器不仅检测精度存在差异,观测量也存在差异。因此观测矩阵  $H$  设定为式 (22) 所示形式,在更新过程中,会根据传感器来源不同使用不同的观测矩阵。

$$H = \begin{cases} \begin{bmatrix} x & 0 & 0 & 0 \\ 0 & y & 0 & 0 \\ 0 & 0 & v_x & 0 \\ 0 & 0 & 0 & v_y \end{bmatrix}, \text{source} = 1 \\ \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}, \text{other} \end{cases}. \quad (22)$$

同样,不同传感器的测量噪声也存在差异,矩阵

$R$  设定为式(23)所示形式,更新过程中根据传感器来源不同调整测量噪声。

$$R = \begin{cases} \begin{bmatrix} [1 \ 0]E & 0 & 0 & 0 \\ 0 & [0 \ 1]E & 0 & 0 \\ 0 & 0 & [1 \ 0]E_v^T & 0 \\ 0 & 0 & 0 & [0 \ 1]E_v^T \end{bmatrix}, s_{source}=1 \\ \begin{bmatrix} [1 \ 0]E_v(Y) & 0 \\ 0 & [0 \ 1]E_v(Y) \end{bmatrix}, other \end{cases} \quad (23)$$

最后,过程噪声矩阵  $Q$  代表整个系统的噪声,与单个传感器无关,因此矩阵  $Q$  按照通用方法设置即可。

### 1.8 生命周期管理

借鉴 Pang 等<sup>[15]</sup>提出的方法,本文使用该方法管理对象轨迹及其生命周期。如果一个轨迹在指定的时间内没有被任何检测结果更新,则该轨迹被丢弃。由于雷达和视觉存在辅助验证目标是否虚检的作用,当一个目标同时被雷达和视觉检测到,则可认为此目标真实存在,若此类型的检测结果没有与任何轨迹关联,则此检测结果可直接创建一个新的轨迹,而目标若只被单个雷达或单个视觉检测到,则仍需经过多帧的确认步骤,才可创建一个新的轨迹。

## 2 实验结果分析

### 2.1 数据集和指标

因目前尚无路侧端的多传感器数据集,无法充分验证研究的有效性,因此除了车侧的 nuScenes 数

据集外,还针对路侧场景采集了私有数据集,该数据集由交通雷达摄像头一体机采集。摄像头包括 400 万像素和 800 万像素;雷达包括 250 m 雷达、350 m 雷达和 500 m 雷达 3 种型号。满足大部分流量监控场景的需求。数据采集和标注方法主要参考 RoScenes<sup>[20]</sup>,在数据采集和标注时加入雷达标注。跟踪度量方面则使用业界常用的目标身份切换次数 (ID Switch, IDS)、最大跟踪轨迹 (Mostly Tracked trajectories, MT)、最大丢失轨迹 (Mostly Lost trajectories, ML)、真实目标总数 (Ground Truth Trajectory, GT)、MOTA 等指标。

### 2.2 定量分析

为展示本文提出的传感器检测误差模型的有效性,挑选一个目标的轨迹直观展示跟踪效果。如图 7 所示,同样未经过任何平滑处理的情况下,雷达检测的目标轨迹始终在真值附近 ( $x$  方向  $\pm 1$  m,  $y$  方向  $\pm 2.5$  m 以内波动)。视觉检测的目标轨迹则在近处目标位置与真值基本一致,而在远处, PV 下虽轨迹较为稳定,但 BEV 视角下的位置出现剧烈抖动,  $y$  方向的误差可达  $\pm 10$  m。使用本文提出的传感器误差模型选择误差最小的位置作为此融合目标的轨迹点之后,近处的位置误差基本接近真值,远处的位置误差也不超过雷达的检测误差。此例证明了传感器检测误差模型的有效性。

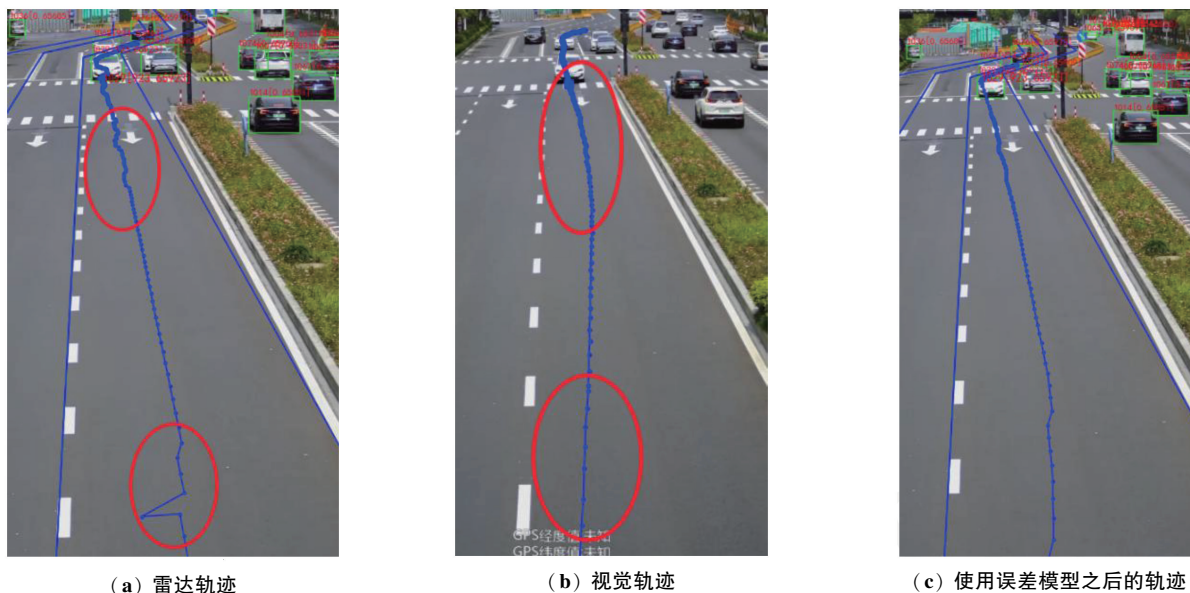


图 7 误差模型效果

Fig.7 Error model effect

为了验证本文提出的多目标跟踪方法的整体有效性,与未添加二阶段匹配策略且仅使用普通卡尔曼的行业主流方法<sup>[18,21]</sup>分别在自定义数据集和 nuScenes 数据集进行对比,自定义数据集对比结果如表 2 所示,从表中可以看出,本文提出的方法 MOTA 指标提升了 16.3 个百分点,其中 IDS 指标下降了 44 次。通过分析数据,指标得到提升的原因集中于图 8 所示的情况,因距离较远导致的视觉检测效果差时,下一帧只能被雷达检测到,此时从视觉位置切换到雷达位置,2 个传感器的标定误差和测量误

差稍大就会造成跟踪错误。因遮挡导致的视觉检测效果较差时,即便使用了本文提出的检测误差模型,因视觉框跳动范围超出正常范围,也会造成前后 2 帧的位置误差超过阈值,而此时 PV 视角下的 IoU 比对仍会生效,如图中 BEV 视角下红色箭头指向的 2 个目标对,其虽为同一目标,但 BEV 下的距离误差达 20 m,而 PV 视角下的误差不超过 10 pixel。这些指标提升场景也符合本文针对路侧场景下的优化预期。最终多种方法共同生效,保障多目标跟踪指标的有效提升。

表 2 跟踪指标

Tab. 2 Tracking index

方法	检测算法	MOTA/%	GT	MT 数量	ML 数量	IDS
deepsort <sup>[21]</sup>	LargeKernel3D	69.4	5 899	306	117	63
plop-mot <sup>[18]</sup>	LargeKernel3D	73.6	5 899	320	93	47
本文方法	LargeKernel3D	85.7	5 899	374	80	19

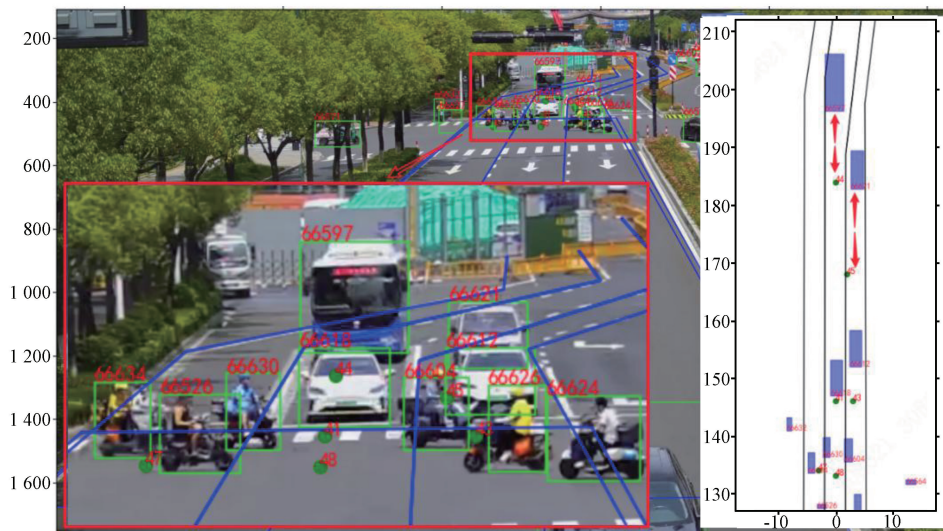


图 8 优化场景示例

Fig. 8 Example of optimization scener

图 9 展示了在实际应用场景下,路侧端恶劣场景下的跟踪结果,图中左侧上方图片为实际进行目标检测的图片,并将检测结果叠加到图片中进行了展示,左侧下方图片为远处区域进行放大之后的结果,方便在调试或结果展示过程中,对目标真实情况进行观察。右侧则展示了 BEV 下各个环节的结果,红色圆圈代表雷达检测结果,黄色矩形则代表视觉

检测结果,黑白相间的车辆图形则为融合跟踪结果。从图中可以看出,在极远处这种恶劣场景下,视觉虽然能检测到远处的小目标,但是在 BEV 下其定位误差极大,图中雷达点与其对应的 ID 为 2 790 的视觉目标定位误差约为 20 m,此情况下仍能正确进行关联,并只输出一个融合结果。展示了本文所提方法在恶劣场景下的有效性。

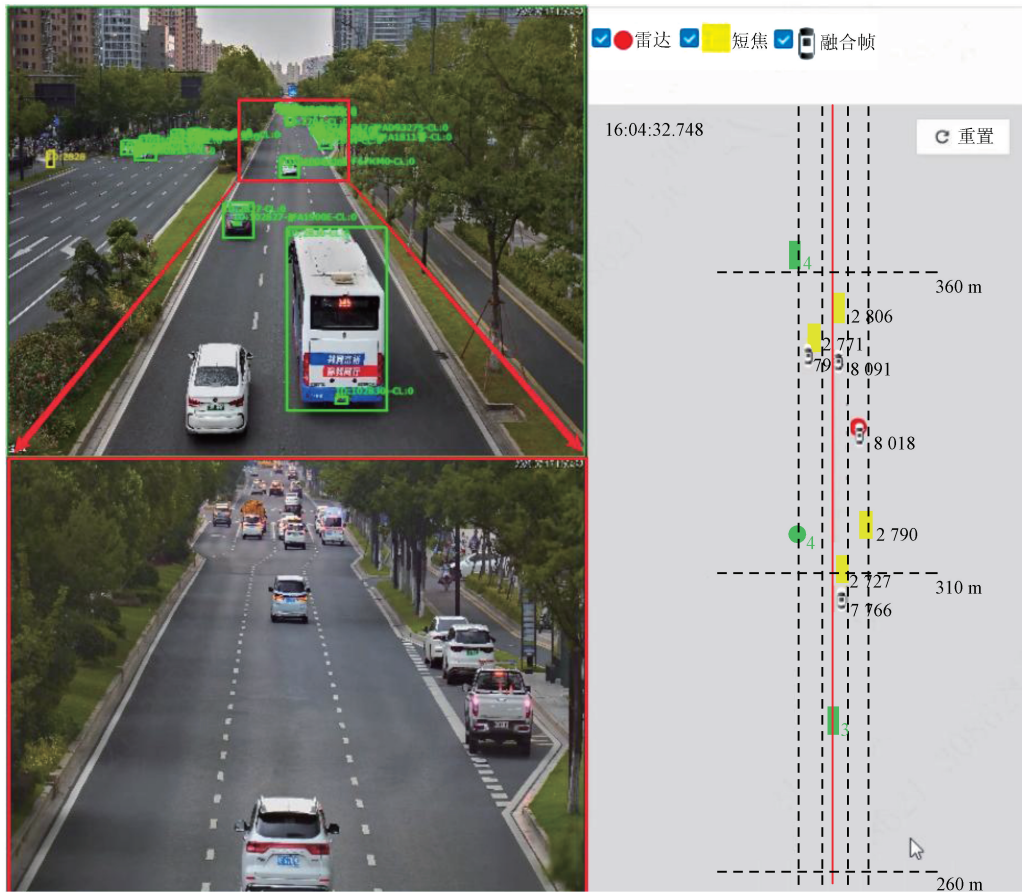


图9 恶劣场景跟踪效果  
Fig. 9 Tracking effect of adverse scenes

表3展示了在 nuScenes 数据集下的指标对比结果,在 nuScenes 数据下,本文新方法的指标提升幅度不大,主要因为 nuScenes 数据集是自动驾驶相关数据集,其检测范围比路侧场景近很多,在近距离时,视觉检测的准确率很高,基本不会面临路侧场景下视觉定位剧烈抖动的情况。因此本文关于视觉定位方面的优化效果有限,仅新提出的自适应卡尔曼滤波器能有效提升指标。

表3 nuScenes 数据集跟踪指标  
Tab. 3 Tracking index in nuScenes dataset

方法	检测算法	MOTA/%	GT	MT 数量	ML 数量	IDS
Simpletrack <sup>[15]</sup>	CenterPoint	56.6	17 081	5 476	1 780	575
ploy-mot <sup>[18]</sup>	LargeKernel3D	62.1	17 081	5 946	1 649	292
本文方法	LargeKernel3D	62.9	17 081	6 098	1 462	253

### 3 结束语

本文针对交通雷视一体机协同感知和跟踪目标,高精度估计目标位置速度信息需求和处理远距

离目标跟踪精度低的问题,提出一种基于二阶段关联和卡尔曼滤波的多传感器多目标跟踪方法。该方法首先通过拟合方式将雷达无法检测的视觉框信息补充完整,使送入跟踪系统的每个目标都具有完整一致的属性信息。其次,提出一种二阶段关联方法,在距离关联的基础上,补充 PV 下的 IoU 关联方法,解决远处测量误差或标定误差较大导致关联准确率低的问题。最后,提出一种自适应跟踪滤波器的方法,对雷达传感器和视觉传感器的测量误差建模,将测量误差模型引入到卡尔曼滤波器中,使卡尔曼滤波器的更新过程处于收敛状态(避免发散),其更适用于路侧的多传感器跟踪场景。试验结果表明,本文所提出的传感器误差模型可有效度量雷达和视觉传感器在同一位置对目标估计的准确性。所提出的二阶段关联方法和自适应卡尔曼方法也能在真实应用场景中有效提升跟踪指标,为多传感器跟踪的行业应用提供一种有效方法。在后续的研究中,将着重探索目标轨迹行为推理和对目标视觉信息关键点

方面的应用,以期能解决监控场景下车辆遮挡导致的感知跟踪问题。



### 参考文献

- [1] 杨立峰. 人工智能时代我国城市智能道路交通发展研究[C]//绿色数智 提质增效——2024年中国城市交通规划年会. 绿色数智 提质增效——2024年中国城市交通规划年会论文集. 昆明: 中国建筑工业出版社, 2024: 1880-1890.
- [2] KARUNASEKERA H, WANG H, ZHANG H. Multiple Object Tracking with Attention to Appearance, Structure, Motion and Size[J]. *IEEE Access*, 2019, 7: 104423-104434.
- [3] FENGW T, LI B P, OUYANG W L. Multi-object Tracking with Multiple Cues and Switcher-aware Classification[C]//2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA). Sydney: IEEE, 2022: 1-10.
- [4] 苏佳, 冯康康, 孟俊彤, 等. 基于改进 CStrack 关联策略的多目标跟踪算法[J]. *无线电工程*, 2024, 54(3): 597-606.
- [5] 刘敬一, 高晓倩, 郭琦, 等. 基于运动属性约束的船舶 AIS 航迹去纠缠方法[J]. *无线电工程*, 2023, 53(3): 678-685.
- [6] 王志伟, 刘永祥, 杨威, 等. 基于航迹概率假设密度的多传感器多目标跟踪[J]. *系统工程与电子技术*, 2024, 46(2): 526-533.
- [7] 齐滨, 田金, 邱龙皓, 等. 检测前跟踪中的目标交叉航迹标签分层管理[J]. *声学学报*, 2023, 48(6): 1164-1173.
- [8] 王登辉. 多传感器多目标航迹关联与融合算法研究[D]. 辽宁: 沈阳理工大学, 2023.
- [9] 朱俊杰. 三维多目标跟踪航迹片段关联方法[D]. 浙江: 杭州电子科技大学, 2023.
- [10] ZENG F G, DONG B, ZHANG Y, et al. MOTR: End-to-End Multiple-object Tracking with Transformer[C]//European Conference on Computer Vision. Tel Aviv: Springer, 2022: 659-675.
- [11] ZHANG Y, WANG T C, ZHANG X Y. MOTRv2: Bootstrapping End-to-End Multi-object Tracking by Pretrained Object Detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 22056-22065.
- [12] 欧群雍, 谭同德, 袁红斌. 结合 CNN 和 Bi-LSTM 的多行人目标检测跟踪方法[J]. *无线电工程*, 2022, 52(9): 1633-1641.
- [13] GAOR P, WANG L M. MeMOTR: Long-term Memory-augmented Transformer for Multi-object Tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 9867-9876.
- [14] WANGX Y, QI S Z, ZHAO J Y, et al. MCTrack: A Unified 3D Multi-object Tracking Framework for Autonomous Driving[EB/OL]. (2024-09-23) [2025-05-06]. <http://arxiv.org/abs/2409.16149>.
- [15] PANG Z Q, LI Z C, WANG N Y. Simpletrack: Understanding and Rethinking 3d Multi-object Tracking[C]//European Conference on Computer Vision. Tel Aviv: Springer, 2022: 680-696.
- [16] 袁野. 基于毫米波雷达与视觉传感器融合的防侵入预警系统的研究与实现[D]. 江苏: 扬州大学, 2023.
- [17] 许华杰, 郑力文. 基于自适应卡尔曼滤波的视觉多目标跟踪[J]. *计算机工程与应用*, 2025, 61(5): 200-210.
- [18] LIX Y, XIE T, LIU D D, et al. Poly-mot: A Polyhedral Framework for 3D Multi-object Tracking[C]//2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Detroit: IEEE, 2023: 9391-9398.
- [19] ZHANG Y F, WANG X H, YE X Q, et al. ByteTrackV2: 2D and 3D Multi-object Tracking by Associating Every Detection Box[EB/OL]. (2023-03-27) [2025-05-06]. <https://arxiv.org/abs/2303.15334>.
- [20] ZHUX S, SHENG H L, CAI S J, et al. RoScenes: A Large-scale Multi-view 3d Dataset for Roadside Perception[C]//European Conference on Computer Vision. Milan: Springer, 2024: 331-347.
- [21] PUJARAA, BHAMARE M. DeepSORT: Real Time & Multi-object Detection and Tracking with YOLO and TensorFlow[C]//2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS). Trichy: IEEE, 2022: 456-460.

### 作者简介

刘明 男, (1981—), 硕士, 正高级工程师。主要研究方向: 以视频雷达为核心的多维感知融合技术。

方勇军 男, (1985—), 正高级工程师。主要研究方向: 新一代感知技术。

吴函 男, (1996—), 硕士, 工程师。主要研究方向: 人工智能。

李乾坤 男, (1988—), 硕士, 正高级工程师。主要研究方向: 人工智能。

李冬冬 男, (1989—), 博士, 工程师。主要研究方向: 人工智能。

张朝阳 男, (1973—), 博士, 教授。主要研究方向: 无线人工智能与网络智能、智能感知通信计算一体化。