

一种基于 2D 和 3D 联合信息的改进 MDP 跟踪算法

王正宁¹ 周 阳¹ 吕 侠¹ 曾凡伟¹ 张 翔¹ 张锋军²

(电子科技大学信息与通信工程学院 成都 611731)¹

(中国电子科技集团公司第三十研究所 成都 610041)²

摘 要 在线多目标跟踪算法是自动驾驶和辅助驾驶系统的重要组成部分。目前,大部分多目标跟踪方法集中于图像域跟踪。虽然通过建立自适应在线模型或最小化能量函数可以解决大多数跟踪问题,但是如何处理复杂交通场景下目标的相互遮挡仍是研究者们面临的难题。文中基于 2D 和 3D 联合信息提出了一种改进的基于马尔科夫决策过程(MDP)的跟踪算法,通过将原始 MDP 跟踪算法的相似性特征由图像域拓展到空间域,使用一种新的光流特征描述子即多图像前后向跟踪误差(Multi-image FB error)来代替原算法的多区域前后向跟踪误差(Multi-aspect FB error),取得了良好的跟踪效果。最后,采用 KITTI 数据库对本文算法进行测试,结果显示其综合性能相较于原算法有显著提升。

关键词 多目标跟踪,马尔科夫决策过程,多图像光流,2D-3D 联合特征

中图分类号 TP181 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.03.013

Improved MDP Tracking Method by Combining 2D and 3D Information

WANG Zheng-ning¹ ZHOU Yang¹ LV Xia¹ ZENG Fan-wei¹ ZHANG Xiang¹ ZHANG Feng-jun²

(School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)¹

(The 30th Research Institute of CETC, Chengdu 610041, China)²

Abstract Online multi-object tracking (MOT) plays an important role in autonomous driving and ADAS system. Most of recent MOT methods concentrate on tracking in image domain. Although they can solve most of problems by building adaptive online models or optimizing energy functions, it's still an obstacle for researchers to handle mutual occlusion in complex traffic scenes. In this paper, an improved tracking method was proposed by introducing 3D information to the Markov decision processes (MDP) tracker. The original MDP similarity feature was extended from image domain to spatial domain with 2D-3D combined feature, and a new optical flow descriptor, called multi-image FB error, was addressed to replace the original multi-aspect FB error. This method was tested on KITTI benchmark and the results verified that the comprehensive performance of the proposed method is refined significantly in comprehensive performance compared with the original method.

Keywords Multi-object tracking, Markov decision process, Multi-image optical flow, 2D-3D combined feature

1 引言

多目标跟踪(MOT)是 ADAS 和自动驾驶系统的关键技术之一,其主要任务是提取环境中目标随时间推移的运动轨迹。不同于单目标跟踪算法在整个视频序列中保持对同一目标的跟踪,多目标跟踪算法所跟踪的目标集合随着时间推移是不断变化的,这要求跟踪系统具备不断加入新目标的能力。因此,大多数多目标算法都采用边检测边跟踪的框架:使用检测算法^[1-2]获取当前帧中所有的目标,并将其分别与已知目标连接,最终延长原有目标轨迹或者产生新目标。

根据目标连接时所使用的数据范围,可将多目标跟踪算法划分为全局方法和在线方法。最近的研究^[3-5]为了实现目

标连接的高可靠性,大部分都采用全局框架,同时利用历史和未来信息,通过能量函数最小化模式实现目标的精准匹配。但是这种非因果系统并不适用于无人驾驶等技术的应用场景,后者要求跟踪具备实时性。在线跟踪方法^[6-8]则采用严格的因果系统处理目标连接问题,当前帧的轨迹只与历史轨迹和当前检测目标有关。虽然该方法具备更好的实时性,但交通场景的复杂性和检测噪声等因素,使得不具备全局优化能力的在线方法进行目标连接的出错概率更高。最近,有学者试图将两种框架结合起来^[9],并取得了较好的跟踪效果。

不论在线方法还是离线方法,多目标跟踪的核心都在于其目标连接算法。目标连接的基本思路在于构造相似性特征,并用相应的度量函数进行相似性度量。一些方法^[11-13]将

到稿日期:2018-07-02 返修日期:2018-09-10 本文受四川科技厅项目(2018GZ0071)资助。

王正宁(1979—),男,博士,副教授,主要研究领域为图像及视频处理、智能交通系统、多媒体通讯系统及应用,E-mail:zhengning.wang@uestc.edu.cn;周 阳(1991—),男,硕士生,主要研究领域为基于视觉的目标检测与跟踪算法;吕 侠(1994—),男,硕士生,主要研究领域为基于视觉的目标检测与跟踪算法;曾凡伟(1994—),男,硕士生,主要研究领域为图像压缩与拼接;张 翔(1995—),男,硕士生,主要研究领域为图像压缩与拼接;张锋军(1975—),男,研究员级高工,主要研究领域为软件工程和网络安全。

这一过程视为学习和分类过程,并引入在线或离线学习模型加以实现。在文献[12]中,多目标跟踪任务被描述为一个马尔科夫决策过程(MDP),并且其每次决策均包含一次基于离线训练模型的SVM二分类过程。影响目标连接的另一关键因素是相似性特征的构造,大部分方法^[3,10]在提取图像域特征(如HOG、颜色空间、光流等)时会忽略空间信息的利用,从而导致算法无力应对复杂拥挤的交通环境。

近年来,不少学者将相关滤波和目标跟踪相结合,且取得了不错的跟踪效果,其中比较经典的是由Henriques等提出的CSK^[14]算法和KCF^[15]算法。CSK使用图像灰度特征,引进循环矩阵生成样本,利用相关滤波器进行跟踪。KCF采用HOG特征,将特征由单通道转换成了多通道,通过结合岭回归和核技巧,得到了更好的跟踪效果。

本文提出了一种基于2D-3D联合特征的改进MDP多目标跟踪算法,该算法将原始MDP跟踪算法^[16]的相似性特征由图像域拓展到空间域,并使用一种新的光流特征描述子即多图像前后向跟踪误差来代替原算法的多区域前后向跟踪误差,取得了良好的跟踪效果。具体来说,本文主要贡献如下:

(1)将原MDP算法基于图像域的2D目标跟踪拓展到空间域中的3D目标跟踪,采用2D-3D融合目标集合代替原有的2D检测目标集合。同时,从2D检测目标和3D包围盒中提取目标外观特征和空间特征,提高了目标连接环节的判别能力和匹配准确率。

(2)提出了一种新的光流特征描述子,即多图像前后向跟踪误差,来代替原MDP算法的多区域前后向跟踪误差。本文引入双目相机系统提取的左右图像,用于在相邻两帧的4张图像之间进行光流匹配,以此来代替原算法在一张图像的不同区域中提取光流信息。

本文第2节简要介绍Yu等提出的基于MDP的多目标跟踪算法,以明确算法的主要框架;第3节详细阐述本文基于原MDP算法的改进框架;第4节重点介绍改进的MDP算法基于KITTI^[20]数据库的测试结果及其分析;最后总结全文。

2 MDP多目标跟踪算法

2.1 基于MDP的多目标跟踪模型

MDP跟踪模型由Xiang等^[12]提出,它首先定义了一个包含4个状态的状态空间,用于描述被跟踪目标的4种可能状态:Active, Tracked, Inactive, Lost。其中,Active表示激活状态,代表目标被检出;Inactive与Active相对,代表目标生存期的结束;Tracked和Lost分别代表跟踪状态和丢失状态。4种状态之间可以相互转移,其转移关系如图1所示。

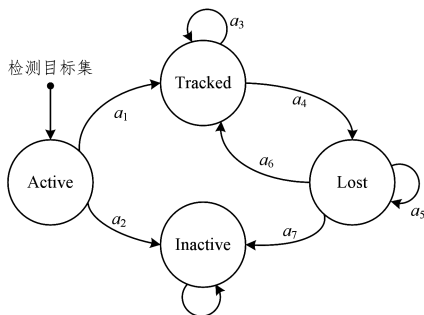


图1 基于马尔科夫决策模型的跟踪过程

Fig.1 Tracking process based on Markov decision model

a_i 代表不同的决策,特定决策产生唯一状态;每个决策带来的收益由奖励函数 $R(\cdot)$ 描述,对于不同状态下的不同决策,奖励函数用来评估其好坏。跟踪任务在该模型下被推导为最佳决策的选择过程,此即基于MDP框架的多目标跟踪机制。

2.2 MDP决策机制

MDP针对不同的状态转移过程,分别采用不同的奖励函数。

(1) S_{active} 状态下,观测目标可能进入 S_{tracked} 状态或者 S_{inactive} 状态。该过程可视为跟踪前的预处理,主要目的是滤除检测环节的噪声。此时的奖励函数为:

$$R_{\text{active}}(s, a) = y(a)(\omega_{\text{active}}^T \phi_{\text{active}}(s) + b_{\text{active}}) \quad (1)$$

其中, ϕ_{active} 为激活特征,由目标的位置、尺度构成; ω 和 b 为SVM权重及偏移; $y(a)$ 为决策标签, $y(a_1) = 1, y(a_2) = -1$ 。

(2) S_{tracked} 状态下,目标保持 S_{tracked} 状态或转移至 S_{lost} 状态。此时的决策选择依赖于跟踪结果,由前后向LK匹配误差(FB error)^[11]和预测包围盒与检测框IOU共同决定。

$$R_{\text{tracked}}(s, a) = \begin{cases} y(a), & \text{if } e_{\text{medFB}} < e_0 \text{ and } iou > o_0 \\ -y(a), & \text{otherwise} \end{cases} \quad (2)$$

其中, e_0 和 o_0 为固定阈值; $y(a_3) = 1, y(a_4) = -1$ 。

(3) S_{tracked} 状态下目标可能维持 S_{lost} 状态或转移至 S_{tracked} 状态,亦或进入 S_{inactive} 状态终止轨迹。若目标处于 S_{lost} 状态的时长超过时间阈值,则转入 S_{inactive} ;若目标丢失时间在时间阈值内,则将该目标依次与检测器输出的观测目标连接,若成功匹配某个观测目标,则转移至 S_{tracked} 状态,否则维持 S_{lost} 。此时的奖励函数为:

$$R_{\text{lost}}(s, a) = y(a)(\max_{k=1}^M (\omega_k^T \phi_{\text{lost}}(t, d_k) + b)) \quad (3)$$

其中, $\phi_{\text{lost}}(\cdot)$ 表示相似性特征; (t, d_k) 代表跟踪目标-观测目标对,每对 (t, d_k) 生成一个相似性特征;决策标签 $y(a_5) = 1, y(a_6) = -1$ 。

2.3 连接特征的构造及在线模型的更新

MDP采用人工设计的相似性特征,其结构如表1所列。其中前5维由多区域FB error组成:一个整体FB error和上、下、左、右4个半区的区域FB error。其提取方法如图2所示。

表1 相似性特征的构造

Table 1 Construction of similarity features

类型	标记	特征含义描述
FB error	ϕ_1, \dots, ϕ_5	FB errors, 占据5个分量, 包括帧间FB errors一对, 交叉FB errors一对及循环FB errors
NCC	ϕ_6	光流匹配邻域间的相关系数
	ϕ_7	跟踪目标与检测目标间的相关系数
Height ratio	ϕ_8	跟踪目标框与光流预测框的高度之比
	ϕ_9	跟踪目标框与检测目标框的高度之比
Overlap	ϕ_{10}	检测目标框与光流预测框IOU
Score	ϕ_{11}	归一化检测分数
Distance	ϕ_{12}	预测目标与观测目标间的欧氏距离

相似性特征分类器依旧采用二分类SVM,使用增强学习框架进行线下训练:首先初始化SVM分类器的权重为 (ω_0, b_0) ;输入训练集,使用MDP框架进行目标跟踪,根据Ground Truth判断决策是否正确,若决策错误,将对应的相似性特征 $\phi(t_{ij}^l, d_k)$ 加入训练集,使用训练集 S 训练SVM分类器,多次迭代,直至训练集中所有目标都能够被正确跟踪,则MDP决

策框架的参数训练完成。

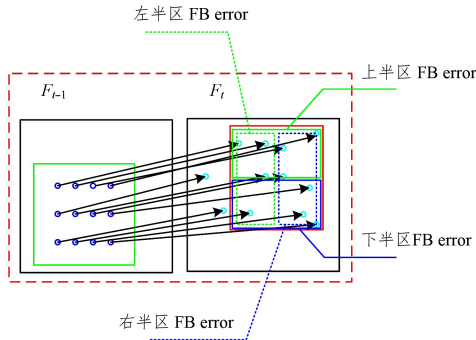


图2 多区域 FB error

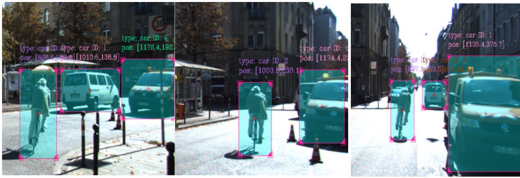
Fig. 2 Multi-zone FB error

数据连接环节观测目标集的连接对象是 MDP 为每个目标建立的在线模型。采用与 TLD^[17] 相似的图像块模板形式,不过 MDP 的模板数固定为 K 个。在模型更新模式上,MDP 由于采用了离线训练模式,无法直接套用 TLD 的 P-N learning 框架,但作者也没有完全采用常规的逐帧更新,而是提出了一种保守更新方式:每个目标在特定时刻只与 K 个模板中的一个绑定,且只要目标处于 S_{tracked} 状态则绑定不变,当且仅当目标跟踪失败时才更改目标绑定的模板。这种更新方式在一定程度上减小了在线模型随时间推移积累的误差,进而降低了跟踪漂移的风险。

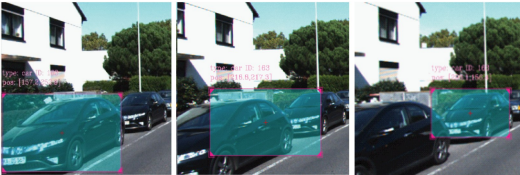
3 MDP 跟踪算法的改进

3.1 MDP 算法的缺陷

MDP 多目标跟踪算法的跟踪性能虽然较好,但仍存在提升空间。图3展示了 MDP 算法处理效果较差的场景:图3(a)中由于最右侧汽车的遮挡,跟踪器将其错检为目标1,而目标1的白色汽车被错检为目标5;图3(b)中3台轿车的外观和高度十分相似且并排放置,导致跟踪器出现漂移。



(a) 目标相互遮挡



(b) 目标外观相似

图3 MDP 算法跟踪过程的不足表现

Fig. 3 Insufficient performance of tracking process of MDP algorithm

上述问题的根源是 MDP 对图像域信息的过度依赖:用于目标连接的相似性特征主要基于 FB error、NCC 和图像距离等 2D 信息,显然很难排除相互遮挡或外观相似的目标干扰。

3.2 基于 2D 和 3D 信息的 MDP 跟踪方法的改进

针对 MDP 特征结构单一、对复杂场景处理能力较差的缺陷,本节提出一种联合空间和图像信息的改进 MDP 跟踪

算法,以提高跟踪性能。

3.2.1 联合图像和空间特征的 MDP 框架

改进后的 MDP 算法的框架延续了原算法的 4 种状态及其转移关系(见图 1),但评估决策选择的奖励函数有所变更。

(1) S_{active} 状态下,检测目标集合变为 2D-3D 融合目标集,因此判别过程也演变为 2D-3D 综合判别。其中,2D 特征 $\phi_{\text{active}}^{2d}$ 继承了原始版本;3D 特征 $\phi_{\text{active}}^{3d}$ 则由 3D 框中心坐标、三维尺度、检测分数等构成 7 维向量,此时奖励函数为:

$$R_{\text{active}}(s, a) = y(a) \cdot (\eta \cdot (\omega_{\text{active}, 2d}^T \phi_{\text{active}}^{2d}(s) + b_{\text{active}}^{2d}) + (1 - \eta) \cdot (\omega_{\text{active}, 3d}^T \phi_{\text{active}}^{3d}(s) + b_{\text{active}}^{3d})) \quad (4)$$

其中, $\omega_{\text{active}, 3d}^T$ 与 $\omega_{\text{active}, 2d}^T$ 均通过数据集训练得到,常量参数 η 表示 2D 框判别分数所占的权重。

(2) S_{tracked} 状态下,相比于原 MDP 算法,改进算法一方面采用 KF(Kalman Filter)同时对目标的空间运动状态和尺度变化进行估计,来代替原有的 2D 运动估计;另一方面使用 3D 空间距离取代 iou 作为判决指标之一。

$$d_{3d} = \|C_{\text{pre}} - C_{\text{det}}\|_2 \quad (5)$$

其中, C_{pre} 代表 KF 预测的 3D 包围盒中心坐标, C_{det} 表示被匹配的 2D-3D 目标的 3D 包围盒中心坐标。此时的奖励函数为:

$$R_{\text{tracked}}(s, a) = \begin{cases} y(a), & \text{if } e_{\text{medFB}} < e_0 \ \& \ d_{3d} > T_{d, 3d} \\ -y(a), & \text{otherwise} \end{cases} \quad (6)$$

(3) S_{lost} 状态下的目标连接决策,综合考虑了 2D 特征和 3D 特征。其中,2D 特征 ϕ_{lost}^{2d} 继承了原算法的结构,但对 FB error 做了改进;3D 特征 ϕ_{lost}^{3d} 主要融合了空间信息。改进后 S_{lost} 状态下的决策机制与 S_{active} 相似,由 2D 和 3D 两部分加权组成:

$$R_{\text{lost}}^{3d}(s, a) = y(a) \cdot (\max_{k=1}^M (\tau \cdot (\omega_{2d}^T \phi_{\text{lost}}^{2d}(t, d_k) + b_{2d}) + (1 - \tau) \cdot (\omega_{3d}^T \phi_{\text{lost}}^{3d}(t, d_k) + b_{3d}))) \quad (7)$$

$$\tau = \frac{s_{2d}}{s_{2d} + s_{3d}} \quad (8)$$

式(7)中, τ 代表 2D 特征分类结果所占的权重, s_{2d} 和 s_{3d} 分别指式(4)中 2D 和 3D 框的形态得分。在 S_{lost} 阶段,图像域特征和空间域特征所占的权重由 S_{active} 阶段框的质量分数决定。

3.2.2 2D-3D 相似性特征构造

改进后 MDP 算法的相似性特征分为 2D 和 3D 两部分。其中,2D 特征部分继承于原算法(见表 1);新增的 3D 特征部分如表 2 所列,为 6 维归一化向量,主要包括距离、尺度、速度方向、速度幅度(大小)、IOU 和框置信度 6 个部分。

表 2 3D 相似性特征结构

Table 2 Structures of 3D similarity features

类型	标记	特征含义的描述
Center distance	ϕ_1^{3d}	检测目标与 KF 预测目标的空间距离
Size ratio	ϕ_2^{3d}	目标与模板 3D 包围盒高度、宽度、长度比值的加权线性组合
Cos Δ alpha	ϕ_3^{3d}	帧间速度方向角度差的余弦值
Velocity ratio	ϕ_4^{3d}	帧间速度幅度差的比值
Overlap _{3d}	ϕ_5^{3d}	检测目标与预测目标 3D 框空间 IOU
Score _{3d}	ϕ_6^{3d}	目标 3D 包围盒的检测分数

Center distance 采用欧氏距离并结合距离阈值 $T_{d, 3d}$ 对幅值进行约束:

$$d_{3d} = \frac{\|C_{\text{pre}} - C_{\text{det}}\|_2}{T_{d, 3d}} \quad (9)$$

尺度比 Size ratio 为加权比值:

$$r_{3d} = \frac{1}{K} \sum_i^K \xi_1 \cdot \min\left(\frac{h_i}{h_{t,k}}, \frac{h_{t,k}}{h_i}\right) + \xi_2 \cdot \min\left(\frac{w_i}{w_{t,k}}, \frac{w_{t,k}}{w_i}\right) + \xi_3 \cdot \min\left(\frac{l_{i,k}}{l_t}, \frac{l_t}{l_{i,k}}\right) \quad (10)$$

$$\xi_j = \frac{\sigma_j^{-1}}{\sigma_1^{-1} + \sigma_2^{-1} + \sigma_3^{-1}}, j=1, 2, 3 \quad (11)$$

式(10)中的 σ_j 代表第 j 维标准差,即认为任一目标 t_k 的各维尺度服从正态分布 $N_k \sim (\mu_{k,j}, \sigma_{k,j})$ 。虽然理想情况下同一目标的空间尺度在整个生存期内不会发生变化,但检测模块引入的误差会导致某些维度比值的浮动较大,因此采用正态分布建模。以训练集中的 3D 框为训练数据,通过最大似然估计可求得任一维的标准差 σ_i 。式(11)中, h_i, w_i, l_i 分别代表第 i 个模型的 3D 尺度, $h_{t,k}, w_{t,k}, l_{t,k}$ 表示检测目标 t_k 的 3D 尺度。

速度比 Velocity ratio 代表速度相似性,由方向一致性和幅度一致性两方面衡量。其中,方向一致性的计算公式为:

$$\mu_{\cos} = \begin{cases} \cos(\alpha_t - \alpha_{t-1}), & \text{if } |\alpha_t - \alpha_{t-1}| < \frac{\pi}{2} \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

幅度一致性的计算公式为:

$$\mu_{\parallel v} = \frac{\|v_t - v_{t-1}\|}{\max(\|v_t\|, \|v_{t-1}\|)} \quad (13)$$

速度特征的构造原理基于车辆运动的渐变假设,交通环境下的车辆运动一般遵循匀变速原则,同一车辆的相邻帧的运动规律基本保持稳定。

3D 交并比 Overlap_3d 是二维 IOU 的三维扩展,采用 3D 框交集的体积比表示。

2D 特征继承了除 FB error 外的其他特征分量,并采用如图 4 所示的多图像 FB error 气流描述子来替代原算法的多区域 FB error 气流描述子。

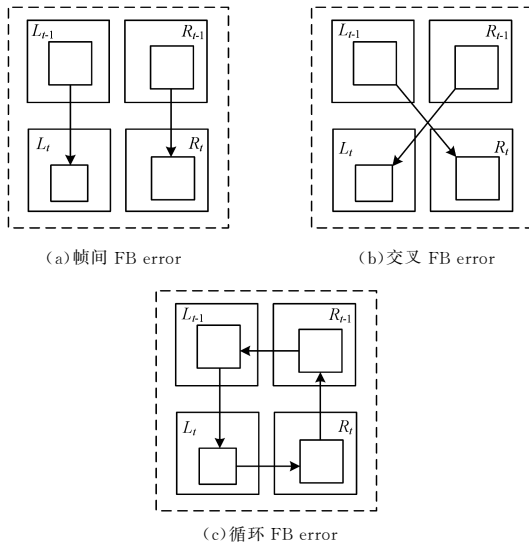


图 4 多图像 FB error

Fig. 4 Multi-image FB error

依据文献[17],多图像 FB error 的每一个分量均取 LK 气流匹配后 FB error 总集的中值。其中,循环 FB error 的获取要相对复杂一些:从 $t-1$ 时刻左图像 2D 框 $B_{L,t-1}$ 出发,对 $t-1$ 时刻的右图像进行 LK 气流匹配,以 FB error 中值为阈

值,选取匹配点并求出匹配后的 2D 框 $B_{R,t-1}$;以 $B_{R,t-1}$ 为基础,在 $t-1$ 时刻右图像和 t 时刻右图像之间匹配,得到框 $B_{R,t}$;然后从 t 时刻右图像到 t 时刻左图像得到 $B_{L,t}$;最后由 t 时刻左图像匹配回 $t-1$ 时刻左图像,此时将匹配误差总集的中值作为循环 FB error。

MDP 算法中的多区域 FB error 将整个匹配区域分成上、下、左、右 4 个半区(见图 2),试图同时描述目标的整体相似性和局部相似性。但是用于描述 FB error 的中值误差 e_{medFB} 是统计意义上的误差大小,取中值的目的是避免误差较大区域的干扰,若强行对框内匹配点进行分区统计,则容易出现匹配误差分布不均的情况。由于原始连接特征 ϕ_{lost} 前 5 个分量均由 FB error 构成,除描述整体误差的分量外,其他局部误差分量是不稳定的;4 个分区的匹配效果取决于目标运动方向、相机角度以及检测模块给出的检测框的精确度,而离线训练的 SVM 固定权重无法自适应这些变动,最终起决定性作用的依然是目标整体匹配误差。

相较于多区域 FB error,本节提出的多图像 FB error 具有更强的相似性描述能力。首先,构成 FB error 特征的 5 个分量均使用全局匹配误差中值,而不像多区域 FB error 进行区域划分,这在一定程度上保证了 FB error 中值作为统计量的可靠性;其次,多图像 FB error 充分利用了左右相机采集的丰富图像资源,使多视角图像信息相互佐证,且循环匹配使得提取特征的稳定性更高,判别能力更强。

4 实验及结果分析

本节主要通过综合实验客观评估基于空间和图像信息的 MDP 多目标跟踪算法的性能。

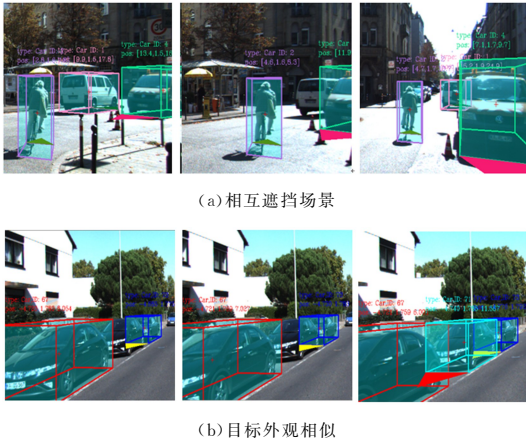
数据集方面,我们使用 KITTI 数据库跟踪数据集。该数据集被分为训练集和测试集两部分,本文仅使用训练集进行训练和验证实验。训练集由 21 个序列组成,编号为 0 至 20,其中序列 12,13,15,16,17 和 19 的主要内容为行人,不符合本算法对车辆目标跟踪的要求,因此被剔除。剩下的 15 个序列被分为训练集和验证集两部分,其中训练集包括序列 0,1,2,3,5,7,9,11,18;验证集包括 4,6,8,10,14,20。

检测目标集合方面,由于改进后的 MDP 方法由 2D 图像域拓展到 3D 空间域,其用于跟踪和连接的检测目标集合也由原始的 2D 集合变为了 2D-3D 融合目标集。为减小检测环节对跟踪环节的影响,我们采用 KITTI 数据集自带的 2D 和 3D 标签生成 2D-3D 检测集合。具体来说,通过标签中的目标类别标签、截断指标和遮挡指标对标签集合进行筛选:首先删去所有标签为“Don't care”和“Pedestrian”的目标;其次忽略所有尺寸(长或宽)小于 25 像素的目标;最后舍去遮挡等级大于 2 或截断等级大于 1 的目标。将筛选后的集合同时作为检测集合和 Ground Truth,其中检测集合在 2D 框和 3D 框的尺度、中心等参数中加入了高斯噪声,以区分 Ground Truth。

评价指标方面,本文采用文献[19]提出的多目标跟踪精度(MOTP)和多目标跟踪准确率(MOTA)、文献[18]提出的最多跟踪目标数(MT)和最多丢失目标数(ML),以及常用的准确率(Recall)、召回率(Precision)和测量值(F1)。

图 5 展示了改进后 MDP 算法在遮挡和目标外观相似两

种情况下的跟踪效果,为了方便比较,选取了与图2相同的场景。通过对比可知,原算法中出现的误跟踪和目标漂移现象均得以避免,且输出为能体现空间尺度的3D框。



(a) 相互遮挡场景

(b) 目标外观相似

图5 改进的MDP算法在特定场景中的跟踪效果

Fig. 5 Tracking effect of improved MDP algorithm in specific scene

为了更为客观地对比算法改进前后的性能,表3、表4给出了验证序列的各项测试结果。从表3的准确率和召回率指标来看,改进的MDP算法相较于原算法有较大提升;表4给出的各项指标中,新算法的MT,ML,MTOA均明显优于原算法,MOTP基本与原算法持平。需要注意的是:在序列14和序列20中,原始MDP和改进MDP的多项指标都出现了明显差异:前者明显低于后者。其主要原因是,序列14中存在大量外观相似车辆并排停放场景(见图6(a)),而序列20中交通比较拥挤,遮挡场景频繁(见图6(b)),原MDP单纯使用图像信息进行目标连接对于这种情况效果较差,而本文提出的基于空间信息改进的MDP算法通过联合空间和图像信息,使复杂交通状况下的目标连接更加稳健,因而效果更好。

表3 准确率-召回率指标的实验结果

Table 3 Experimental results of recall rate and accuracy

(单位:%)

序列	MDP/Ours		
	Recall	Precision	F1
4	85.4/89.8	83.5/83.7	84.4/86.7
6	86.1/94.7	86.3/84.3	86.2/89.2
8	91.0/93.3	59.6/59.9	72.0/72.9
10	75.4/86.0	76.1/75.8	75.7/80.6
14	52.6/74.9	63.3/65.6	57.4/70.0
20	61.1/90.0	78.8/80.1	68.8/84.7
Avg	68.3/89.4	76.5/77.6	74.1/80.7

表4 CLEARMOT-MT^[18-19]测量指标的实验结果

Table 4 Experimental results of CLEARMOT-MT measurement index

序列	MDP/Ours			
	MT	ML	MOTA/%	MOTP/%
4	60.0/67.5	20.0/15.0	68.3/71.9	80.4/81.2
6	66.7/93.3	6.7/6.7	72.1/76.5	81.0/80.5
8	79.2/83.3	12.5/4.2	28.4/30.7	73.5/73.5
10	37.0/55.6	29.6/14.8	51.2/56.2	81.2/80.9
14	33.3/60.0	33.3/0.0	21.4/33.8	68.0/66.4
20	21.7/81.7	32.5/8.3	44.4/67.0	76.3/75.9
Avg	49.8/73.7	22.4/8.2	47.0/62.9	77.1/76.6



(a) 序列 14



(b) 序列 20

图6 部分测试序列样张

Fig. 6 Part of test sequence proofs

为了进一步突出改进算法的优越性,表5将本文算法与当前主流的多目标跟踪算法进行了比较。通过对比可知,本文提出的基于空间和图像信息的MDP跟踪方法的综合性能明显强于基于离线全局方法的DCO、TBD算法以及基于半在线模型的NOMT算法,其中对比DCO和TBD的优势尤为明显;本文算法相比于原MDP算法也有明显优势,4个指标均有提升;与同样使用3D信息的CWIT算法相比,MT和ML较为接近,但MOTA和MOTP大幅超出。总体来说,基于2D-3D联合特征的MDP算法改进,实现了对原算法性能的超越,并且对比其他主流跟踪算法,其性能更优。

表5 本文算法与主流多目标跟踪算法的性能比较

Table 5 Comparison of our algorithm and mainstream multi-target tracking algorithms

Algorithm	Tracking Model	MT	ML	MOTA/%	MOTP/%
CIWT ^[21]	Online	49.85	10.31	75.39	79.25
NOMT ^[22]	Near-Online	41.08	25.23	66.60	78.17
DCO ^[23]	Offline	15.14	30.92	37.28	74.36
TBD ^[24-25]	Offline	20.46	32.62	55.07	55.07
MDP ^[26]	Online	47.73	14.32	74.23	83.62
Ours	Online	50.43	10.23	83.50	86.00

结束语 本文提出了一种基于2D-3D联合信息的改进

MDP多目标跟踪算法,通过从2D-3D检测目标集中提取2D和3D信息并构造相应特征,将跟踪2D信息拓展到3D信息。此外,本文还引入了左右图像信息,构造了一种基于多图像前后向跟踪误差的光流特征描述子。通过基于KITTI数据集的综合实验,最终证明了本文方法相较于原MDP算法有显著的性能提升。

参考文献

- [1] XIANG Y, CHOI W, LIN Y, et al. Subcategory-Aware Convolutional Neural Networks for Object Proposals and Detection[C]// Applications of Computer Vision. IEEE, 2017: 924-933.
- [2] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016: 770-778.
- [3] ROTH S, ANDRIYENKO A, SCHINDLER K. Discrete-continuous optimization for multi-target tracking[C]// Computer Vision and Pattern Recognition. IEEE, 2012: 1926-1933.
- [4] MILAN A, ROTH S, SCHINDLER K. Continuous optimization for multi-target tracking[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(1): 58-72.

- [5] ZHANG L, LI Y, NEVATIA R. Global data association for multi-object tracking using network flows[C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008; 1-8.
- [6] SONG X, CUI J, ZHA H, et al. Vision-Based Multiple Interacting Targets Tracking via On-Line Supervised Learning[C] // European Conference on Computer Vision. Springer-Verlag, 2008; 642-655.
- [7] CHOI W, PANTOFARU C, SAVARESE S. A general framework for tracking multiple people from a moving camera[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(7): 1577-1591.
- [8] JU H Y, LEE C R, YANG M H, et al. Online Multi-object Tracking via Structural Constraint Event Aggregation[C] // Computer Vision and Pattern Recognition. IEEE, 2016; 1392-1400.
- [9] CHOI W. Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor[C] // IEEE International Conference on Computer Vision. IEEE, 2015; 3029-3037.
- [10] BERCLAZ J, FLEURET F, TURETKEN E, et al. Multiple object tracking using k-shortest paths optimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(9): 1806-1819.
- [11] KALAL Z, MIKOLAJCZYK K, MATAS J. Forward-backward error: Automatic detection of tracking failures[C] // 2010 20th International Conference on Pattern recognition (ICPR). IEEE, 2010; 2756-2759.
- [12] XIANG Y, ALAHI A, SAVARESE S. Learning to track: Online multi-object tracking by decision making[C] // 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2015; 4705-4713.
- [13] LI Y, HUANG C, NEVATIA R. Learning to associate: Hybrid-boosted multi-target tracker for crowded scene[C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009; 2953-2960.
- [14] HENRIQUES J F, RUI C, MARTINS P, et al. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels[C] // European Conference on Computer Vision. Springer-Verlag, 2012; 702-715.
- [15] HENRIQUES J F, RUI C, MARTINS P, et al. High-Speed Tracking with Kernelized Correlation Filters[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(3): 583-596.
- [16] BELLMAN R. A Markovian decision process[J]. Journal of Mathematics and Mechanics, 1957, 6(5): 679-684.
- [17] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [18] BERNARDIN K, STIEFELHAGEN R. Evaluating multiple object tracking performance: the CLEAR MOT metrics[OL]. <http://doi.org/10.1155/2008/246309>.
- [19] LI Y, HUANG C, NEVATIA R. Learning to associate: Hybrid-boosted multi-target tracker for crowded scene[C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009; 2953-2960.
- [20] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012; 3354-3361.
- [21] OŠEP A, MEHNER W, MATHIAS M, et al. Combined image- and world-space tracking in traffic scenes[C] // IEEE International Conference on Robotics and Automation. 2017; 1988-1995.
- [22] CHOI W. Near-online multi-target tracking with aggregated local flow descriptor[C] // Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2015; 3029-3037.
- [23] ANDRIYENKO A, SCHINDLER K, ROTH S. Discrete-continuous optimization for multi-target tracking[C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012; 1926-1933.
- [24] GEIGER A, LAUER M, WOJEK C, et al. 3D Traffic Scene Understanding From Movable Platforms[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(5): 1012-1025.
- [25] ZHANG H, GEIGER A, URTASUN R. Understanding High-Level Semantics by Modeling Traffic Patterns[C] // IEEE International Conference on Computer Vision. IEEE, 2013; 3056-3063.
- [26] XIANG Y, ALAHI A, SAVARESE S. Learning to track: Online multi-object tracking by decision making[C] // Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2015; 4705-4713.