

DOI: 10.19650/j.cnki.cjsi.J2209848

基于一致性点漂移的智能车视觉目标跟踪方法*

朱 浩, 李 鑫

(重庆邮电大学自动化学院 重庆 400065)

摘 要:针对智能车未知运动下的多目标跟踪问题,提出一种基于一致性点漂移的视觉多目标跟踪方法。首先利用一致性点漂移算法构建智能车未知运动模型,得到局部目标状态变换关系;其次建立一种基于外观相似性和运动一致性的自适应特征融合函数;最后通过匈牙利算法求解轨迹与检测的对应关系,以实现面向智能车的鲁棒数据关联。实验结果表明,与现有的 5 种主流目标跟踪方法对比,所提方法在多个指标方面具有更好的效果,相较于结构约束(SCEA)算法,在 KITTI 数据集中较大运动场景下,所提方法多目标跟踪准确率提高了 6.3%,在实拍实验数据下,所提方法多目标跟踪准确率提高了 7.3%,证明该算法能在智能车未知运动下有效的进行多目标跟踪。

关键词:智能车;一致性点漂移;多目标跟踪;数据关联

中图分类号: TP391.4 TH89

文献标识码: A

国家标准学科分类代码: 510.40

Visual multi-object tracking method for intelligent vehicle based on coherent point drift

Zhu Hao, Li Xin

(College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract:To address the problem of multi-object tracking under the unknown motion of the intelligent vehicle, a visual multi-object tracking method is proposed, which is based on the coherent point drift. First, the unknown motion model of the intelligent vehicle is formulated by the coherent point drift algorithm. The local object state transformation relationship is achieved. Secondly, an adaptive feature fusion function is constructed, which is based on appearance similarity and motion consistency. Therefore, the Hungarian algorithm is utilized to solve the correspondence between the track and the detection. Finally, the robust data association for the intelligent vehicle is realized. Compared with the current five mainstream multi-object tracking methods, results show that the proposed algorithm has better results in multiple indicators. Compared with the SCEA algorithm, the multi-object tracking accuracy of the proposed method is increased by 6.3% in the large motion scene of the KITTI dataset. Under the real-shot experimental data, the multi-object tracking accuracy of the proposed method is increased by 7.3%, which can effectively perform multi-object tracking under the unknown motion of the intelligent vehicle.

Keywords: intelligent vehicle; coherent point drift; multi-object tracking; data association

0 引 言

多目标跟踪(multi-object tracking, MOT)技术是计算机视觉领域的一个重要研究内容,在智能监控^[1]、自动驾驶^[2]、机器人定位导航^[3]等领域均有广泛的应用。而自动驾驶任务中一个最主要的需求就是能够实时感知并理解周围的环境^[4]。多目标跟踪的任务就是为了解析跟踪对象的行为,在目标检测的

基础上获取目标的运动特性并预测目标在下一时刻的运动趋势^[5],以便智能车后续的决策与控制做出正确的判断。但对于智能车的目标跟踪而言,由于智能车行驶会造成前后两帧的目标位置不在同一车体坐标系下,传统的目标建模方式无法有效刻画面向智能车的目标运动,因此,针对如何在智能车未知运动下设计一个有效的多目标跟踪算法具有重大意义。近年来,多目标跟踪问题受到了学术界的广泛关注,并获得了大量的研究成果。

收稿日期:2022-05-24 Received Date: 2022-05-24

* 基金项目:国家自然科学基金(62073052)、重庆市自然科学基金(cstc2021jcyj-msxmX0373)、重庆市教委重点研究项目(KJZD-K202200603)资助

根据如何利用图像序列中目标的检测信息可以将 MOT 分为在线方法和离线方法。离线方法利用整个图像序列中所有可用的检测,并将跟踪作为一个全局优化问题^[6],通常比在线方法有更高的准确性,但离线方法只能应用于整个图像序列存在的情况^[7]。相比之下,在线方法只依赖于历史帧及当前帧的信息,这使得在线方法适合于智能车实时感知场景。大多数在线 MOT 利用对象的特征表达计算数据关联的概率^[8],目前主流的跟踪器通常使用卡尔曼滤波器^[9-10]、检测框交并比(intersection-over-union, IOU)^[11]进行数据关联。Bewley 等^[9]提出了一种简单实时目标跟踪方法(simple online and realtime tracking, SORT)考虑检测框的位置和尺寸信息,使用卡尔曼滤波和匈牙利算法解决运动预测和数据关联问题。但在复杂场景下,由于忽略了目标的外观信息而导致跟踪准确度下降,不能满足智能车高精度感知的要求,Luo 等^[12]提出了基于超像素特征和可变形的组件模型(deformable part model, DPM)特征构成目标的外观信息表示。Bao 等^[13]提出了一种多尺度选择性跟踪内核模块,并采用注意力机制来捕获目标更有效的外观信息。Bae 等^[14]提出了集成学习算法在线学习轨迹外观模型并基于轨迹置信度进行目标跟踪(confidence multi-object tracking, CMOT)。除此之外,在计算不同帧间检测的相似性时,目标的运动也是比较常见的特征信息。Sadeghian 等^[15]提出了一种递归神经网络(recurrent neural networks, RNN)结构,该结构在时间窗口上对目标的外观、运动和交互信息进行联合数据关联,最终实现多目标跟踪。You 等^[16]提出了一种新的基于时空拓扑的检测器(spatial-temporal topology-based detector, STTD)算法,该算法将临近目标视为一个簇,并引入拓扑结构来描述属于同一簇的运动目标的动力学模型,最后结合外观特征和 IOU 计算成本矩阵。当存在运动噪声或长期遮挡的情况时,数据关联可能会失败,Tian 等^[17]提出了一种新的跟踪方法,通过 3 个数据关联阶段将目标间的结构域和时域信息结合来解决此问题。Bae 等^[18]在文献[14]的基础上结合外观、形状和运动特征定义小轨迹亲和力和模型,利用小轨迹的可检测性和连续性定义小轨迹置信度,然后根据其置信度以不同方式关联轨迹和检测来解决在线多目标跟踪问题。尽管上述基于特征融合的跟踪方法可以产生准确的跟踪结果,但上述方法^[15-18]的特征融合率通常由主观设定,并不能适应不同场景下的特征表征强度不同问题。对于目标的运动特征,在未知的车体运动下,大多数目标运动模型都会失效,为了处理运动干扰下目标跟踪问题,郑浦等^[19]提出了一种基于光流法与三帧差分法的运动目标检测算法,并设计一种改进的核相关滤波算法进行目标跟踪。针对跟踪场景中的背景复杂、平面内旋转等问题,张伟俊等^[20]利用目标

与相机运动的连续性来计算目标和背景的运动模式,并以此为基础建立基于运动估计的图像观测模型。Yoon 等^[21]通过考虑来自多个目标的运动关系构建相对运动网络目标跟踪算法(relative motion network multi-object tracking, RMOT)来解决相机意外运动的影响。然而,当车体运动较大时,会导致对象运动预测失败,目标间的相对运动网络就不再可靠。因此在此基础上又提出了一种新的数据关联方法^[22],该方法将对象之间的相互关系称为结构约束(structural constraint event aggregation, SCEA)来克服剧烈的相机运动,并利用结构约束来跟踪丢失的目标。但在目标数量较多的情况下,该方法的计算成本会显著增加,无法满足实时性的要求。进一步,针对目标位置漂移和状态预测失效等问题,余仁伟等^[23]提出一种基于薄板样条函数的无人机多目标跟踪方法。利用薄板样条函数刻画无人机运动,但该方法并未考虑目标自身运动影响,导致对无人机运动的刻画不准确。

基于上述研究,本文考虑到用点集配准方法可以很好地刻画智能车的旋转、平移运动,提出一种基于一致性点漂移(coherent point drift, CPD)^[24]的视觉多目标跟踪算法。首先提取图像的加速鲁棒特征(speeded up robust features, SURF),利用 CPD 算法得到前后两帧图像的空间变换关系;其次由目标状态转移计算目标状态框内的特征点并进行剔除;进一步,对剔除后的特征点再次进行空间变换来准确刻画智能车运动,并在智能车运动的基础上构建目标状态运动模型;进而计算当前帧的检测与轨迹之间的颜色、结构、运动及尺寸相似度,考虑到不同环境下的特征权重不同,构建外观相似性和运动一致性的自适应特征融合函数,以提高数据关联的准确度;最后利用匈牙利算法对检测和轨迹关系进行分配。对比不同跟踪算法的评价指标,验证本文算法的有效性。相较于现有的多目标跟踪算法,本文主要面向智能车下的多目标跟踪。由于智能车前后两个时刻未知的运动关系,会导致传统目标运动模型的假设不再成立,本文利用一致性点漂移算法计算前后两帧的空间变换关系来刻画智能车运动,构建智能车运动下的目标状态空间模型,解决由于智能车未知运动而带来的数据关联失效问题。本文基于不同场景中目标密集程度,建立外观和运动的自适应相似性度量函数。解决场景变化对特征权重的影响,以提高数据关联的准确度。

1 多目标跟踪模型框架

本文提出的视觉目标跟踪算法主要包括 3 个主要部分:1)CPD 空间变换刻画智能车运动;2)轨迹与检测的特征相似度计算及自适应融合;3)轨迹管理。本文所提

基于一致性点漂移的智能车视觉目标跟踪方法流程如图 1 所示。

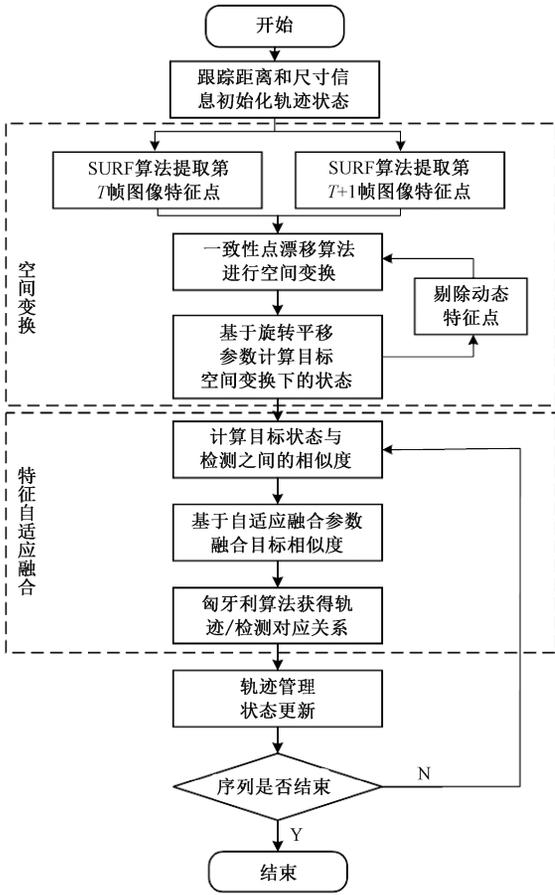


图 1 基于一致性点漂移的多目标跟踪方法流程

Fig. 1 Flowchart of the multi-object tracking method based on coherent point drift

1.1 空间变换模型

本文轨迹 i 在 t 时刻的状态定义为 $x_t^i = [u_t^i, v_t^i, w_t^i, h_t^i, \dot{u}_t^i, \dot{v}_t^i, \dot{w}_t^i, \dot{h}_t^i]^T \in X_t$, 检测 j 在 t 时刻的状态定义为 $d_t^j = [u_t^j, v_t^j, w_t^j, h_t^j]^T \in D_t$, 其中 u_t, v_t, w_t, h_t 代表边界框中心点位置及宽高信息, \dot{u}_t, \dot{v}_t 为速度, \dot{w}_t, \dot{h}_t 代表对象边框的宽度和高度的变化率。

为刻画智能车未知运动, 需获取图像序列前后两帧之间的变换关系。本文通过提取图像 SURF 特征, 再根据特征点的描述向量进行特征匹配, 得到匹配度最大的 n 个特征点。

定义第 $t-1$ 帧特征点集为 $A_{N \times D} = (a_1, \dots, a_N)^T$, 第 t 帧特征点集为 $B_{M \times D} = (b_1, \dots, b_M)^T$, D 为点集的维度, 目标函数由 CPD 算法^[24] 给出:

$$Q(\theta, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M P^{\text{old}}(m | a_n) \| a_n - \Phi(b_m, \theta) \|^2 +$$

$$\frac{D}{2} \log \sigma^2 \sum_{n=1}^N \sum_{m=1}^M P^{\text{old}}(m | a_n) \quad (1)$$

式中: $P^{\text{old}}(m | a_n)$ 为后验概率分布; $\Phi(b_m, \theta)$ 为两特征点集的 CPD 空间变换函数。

$$\Phi(b_m, \theta) = \Phi(b_m; R, t) = R b_m + t \quad (2)$$

使用期望最大化解法 (expectation-maximization, EM) 估计 θ, σ^2 , EM 算法通过在 E 步和 M 步之间交替计算直到获取参数 R_t^g, t_t^g 的最优解。

由于特征点集可能包含车辆、行人等冗余的移动目标特征点, 会导致空间变换不准确。本文通过状态转移和当前空间变换参数得到当前时刻的目标预测状态为:

$$\bar{x}_t^i = F(x_{t-1}, R_t^g + t_t^g) + w_t \quad (3)$$

式中: \bar{x}_t^i 表示当前时刻的目标预测状态; x_{t-1} 表示上一时刻的目标状态; R_t^g, t_t^g 表示所有特征点计算得到的旋转平移参数。得到目标预测状态后, 对目标状态框内的特征点集进行剔除, 保留背景特征点, 再次进行一致性点漂移变换, 对变换参数进行迭代优化, 得到最终的空间变换 R_t, t_t 。具体步骤如图 2 所示。轨迹状态的最优预测如下:

$$x_t = F(x_{t-1}, R_t + t_t) + w_t \quad (4)$$

式中: x_t 为轨迹状态的最优预测; w_t 为过程噪声; F 是基于匀速运动模型的转移矩阵。

1.2 相似度刻画模型

对于本文所提出的跟踪方法, 每一帧的对象由一组边界框组成, 定义 I 个目标状态预测集和 J 个观测值集表示为 $X_t = \{x_t^1, \dots, x_t^I\}$ 和 $D_t = \{d_t^1, \dots, d_t^J\}$ 。在计算目标相似度时, 采用 4 个特征来衡量目标间的相似性, 其中, 结构相似度由 P_h 表示, 颜色相似度由 P_a 表示, 尺寸相似度由 P_s 表示, 运动相似度由 P_e 表示。

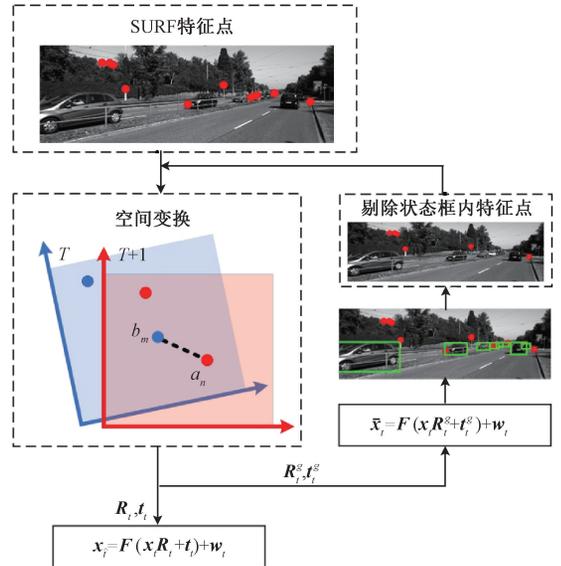


图 2 空间变换示意图

Fig. 2 Schematic diagram of space transformation

1) 结构相似度

结构特征是在图像相似度衡量上应用广泛的特征信息,其中低频信息可以很好地表示图像的结构特征,也是一种对光照变化鲁棒的特征信息。本文利用感知哈希算法对图像低频信息生成一串“指纹”到每个检测上,通过比较“指纹”的相似度来确定轨迹与检测的相似度。

首先将图片缩放到 8×8 pixels 图像,通过忽略图片大小和比例的不同,留下结构、底纹等基本信息;计算灰度图像的离散余弦变换(discrete cosine transform, DCT);根据 DCT 矩阵,构成图像的 64 bit 的哈希指纹整数;最后利用汉明距离来度量哈希指纹之间的相似度。

$$P_h^{i,j}(\mathbf{x}_i^i, \mathbf{d}_i^j) = 1 - d_h(\text{hash}(\mathbf{x}_i^i), \text{hash}(\mathbf{d}_i^j)) \quad (5)$$

式中: $P_h^{i,j}$ 表示第 i 个轨迹和第 j 个检测之间的结构相似度; $\text{hash}(\cdot)$ 表示计算目标低频特征得到的哈希指纹; d_h 表示两哈希指纹之间的汉明距离。最后将距离进行归一化处理,即可得到目标间的结构相似度。

2) 颜色相似度

颜色特征是在目标跟踪中应用最为广泛的视觉特征。传统的红、绿、蓝(RGB)颜色空间中,3种颜色分量的取值与所生成的颜色之间的联系并不符合人们对颜色相似性的主观判断。相对于 RGB 颜色空间,色相、饱和度、亮度(HSV)空间对光照变化的敏感度较低,能够较好地反映图片的颜色分布。本文利用 HSV 空间下的颜色直方图来表示目标的颜色特征,用 Bhattacharyya 距离计算检测与轨迹之间的外观相似度,即:

$$P_a^{i,j}(\mathbf{x}_i^i, \mathbf{d}_i^j) = \sum_{k=1}^K \sqrt{\mathcal{H}^k(\mathbf{x}_i^i) \mathcal{H}^k(\mathbf{d}_i^j)} \quad (6)$$

式中: $P_a^{i,j}$ 表示第 i 个轨迹和第 j 个检测之间的颜色相似度; $\mathcal{H}^k(\mathbf{x}_i^i)$ 和 $\mathcal{H}^k(\mathbf{d}_i^j)$ 表示第 i 条轨迹和第 j 个检测的直方图; K 表示直方图 bin 的数量。

3) 尺寸相似度

基于大小的相似度是根据检测的目标包围框与轨迹的预测包围框进行计算的。考虑第 j 个检测包围框的 w_j, h_j 与第 i 个轨迹预测包围框的 w_i, h_i 之间的差值,利用宽高信息对相似度进行归一化处理。具体定义如下:

$$P_s^{i,j}(\mathbf{x}_i^i, \mathbf{d}_i^j) = 1 - \frac{1}{2} \left(\frac{|h_i - h_j|}{(h_i + h_j)} + \frac{|w_i - w_j|}{(w_i + w_j)} \right) \quad (7)$$

式中: $P_s^{i,j}$ 表示第 i 个轨迹和第 j 个检测之间的尺寸相似度。

4) 运动相似度

当在相邻帧中跟踪同一个目标时,其运动信息有助于预测该目标在下一帧中的位置。本文的目标运动相似度是考虑第 j 个检测的包围框与第 i 个轨迹的预测包围框之间的 CIOU^[25] (complete IOU, CIOU) 来衡量。在智能车剧烈运动时,前后两帧之间的同一个检测可能出现

距离过大的情况,传统的 IOU 在框 A 与框 B 不重合的时候 IOU 值为 0,不能正确反映二者的距离大小。而 CIOU 是基于中心点的距离和对角线距离的比值计算,并且引入框的宽高比信息,能够更好的衡量检测框之间的距离损失。具体定义为:

$$P_c^{i,j}(\mathbf{x}_i^i, \mathbf{d}_i^j) = 1 - IOU + \alpha v + \frac{\rho^2(\text{bbox}(\mathbf{x}_i^i), \text{bbox}(\mathbf{d}_i^j))}{c^2} \quad (8)$$

$$IOU = \frac{\text{bbox}(\mathbf{x}_i^i) \cap \text{bbox}(\mathbf{d}_i^j)}{\text{bbox}(\mathbf{x}_i^i) \cup \text{bbox}(\mathbf{d}_i^j)} \quad (9)$$

$$\alpha = \frac{v}{(1 - IOU) + v} \quad (10)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w_i}{h_i} - \arctan \frac{w_j}{h_j} \right)^2 \quad (11)$$

式中: $P_c^{i,j}$ 表示第 i 个轨迹和第 j 个检测之间的运动相似度; bbox 为目标检测框状态信息; c 是包围它们的最小方框的对角线距离; ρ 为框中心点坐标的欧氏距离; α, v 用来衡量宽高比的相似性。

本文将颜色相似度和结构相似度线性融合定义为外观相似度,运动一致性则由目标的尺寸相似度及运动相似度线性融合定义。数据关联的自适应相似性度量函数可表示为:

$$\Omega(\mathbf{x}_i^i, \mathbf{d}_i^j) = (1 - \lambda_t) \left(\frac{P_a^{i,j}(\mathbf{x}_i^i, \mathbf{d}_i^j) + P_s^{i,j}(\mathbf{x}_i^i, \mathbf{d}_i^j)}{2} \right) + \lambda_t \left(\frac{P_h^{i,j}(\mathbf{x}_i^i, \mathbf{d}_i^j) + P_c^{i,j}(\mathbf{x}_i^i, \mathbf{d}_i^j)}{2} \right) \quad (12)$$

式中: λ_t 为自适应融合参数。对于场景中的每个目标,其外观或运动的区分能力是不同的。当环境中目标较为密集时,目标间的检测框位置及大小一般情况下更为相近,此时的颜色和结构相似度的权重就会增大。本文根据当前帧的所有目标与轨迹之间的重叠率生成权重 λ_t 。 λ_t 定义为:

$$\lambda_t = \frac{\text{sum}(P_c(\mathbf{x}_i, \mathbf{d}_i) > 0)}{N \times M} \quad (13)$$

式中: N, M 分别当前时刻轨迹和检测的数量; $\text{sum}(P_c(\mathbf{x}_i, \mathbf{d}_i) > 0)$ 表示 P_c 中大于 0 的元素个数,即相互重叠的目标个数。基于式(12)中的相似函数,得到第 i 个轨迹与第 j 个检测之间的代价函数 $C_i^{i,j} = -\ln \Omega(\mathbf{x}_i^i, \mathbf{d}_i^j)$ 。计算分配矩阵 $\mathbf{A} = [\alpha^{i,j}]_{N \times M}$ 以使成本最小化:

$$\begin{aligned} & \text{argmin} \sum_{i,j} C_i^{i,j} \alpha^{i,j}, \\ & \text{s. t.} \quad \sum_i \alpha^{i,j} = 1, \sum_j \alpha^{i,j} = 1 \forall i, j \end{aligned} \quad (14)$$

利用匈牙利分配算法对代价函数求解,从而实现数据关联。

1.3 状态管理

在当前帧的目标关联完成时,如果还存在未匹配的轨迹和未匹配的检测,则进行二次数据关联。通过当前帧轨迹的一步预测值和前 3 帧平均状态转移的预测值与未关联的检测进行相似度计算,得到新的分配矩阵。更新对应轨迹的颜色统计直方图、哈希指纹及状态信息。根据卡尔曼滤波方程^[26-27],在状态空间变换的基础上,获得协方差矩阵和轨迹的预测值:

$$\mathbf{x}_i = \mathbf{F}(\mathbf{x}_{i-1}\mathbf{R}_i + \mathbf{t}_i) + \mathbf{w}_i \quad (15)$$

$$\mathbf{P}_i = \mathbf{F}\mathbf{P}_{i-1}\mathbf{F}^T + \mathbf{Q} \quad (16)$$

其中:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (17)$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

基于 t 时刻关联到的检测,更新与其关联的轨迹的状态,得到一个更精确的结果:

$$\mathbf{K}_i = \mathbf{P}_i\mathbf{H}^T(\mathbf{H}\mathbf{P}_i\mathbf{H}^T + \mathbf{M})^{-1} \quad (18)$$

$$\mathbf{x}_i = \mathbf{x}_i + \mathbf{K}_i(\mathbf{d}_i - \mathbf{H}\mathbf{x}_i) \quad (19)$$

$$\mathbf{P}_i = \mathbf{P}_i - \mathbf{K}_i\mathbf{H}\mathbf{P}_i \quad (20)$$

式中: \mathbf{P} 为协方差; \mathbf{Q} 为系统的噪声,代表整个系统的可靠程度; \mathbf{M} 为测量协方差; \mathbf{F} 是基于匀速运动模型的转移矩阵; \mathbf{H} 是观测矩阵。

在状态更新之后,对于未匹配且尚未超出图像边界的轨迹,对其历史状态进行曲线拟合,并利用曲线的预测状态继续对目标进行跟踪。终止连续 5 帧没有被任何检测关联的轨迹,并将无关联的检测初始化为新的轨迹。如果连续 6 帧检测的距离和尺寸相似度大于设定的阈值,则创建一个新的轨迹。

2 实验及分析

本文跟踪算法使用 MATLAB 编程实现,在处理器为 AMD Ryzen 7 4800 H,内存为 16 GB 的计算机上完成。本文将所提的基于一致性点漂移的目标跟踪方法与 5 种算法包括在线方法 (SORT^[9]、CMOT^[13]、RMOT^[21]、SCEA^[22]) 和离线方法 TBD^[7]) 进行评估比较。

2.1 数据集

本文使用 KITTI Object Tracking Evaluation (2D bounding-boxes) 的 training 数据集,并且只考虑“Car”类。训练数据集由 21 个序列组成(序列 0017 仅考虑“Pedestrian”),分辨率平均为 1 242×375,采用 KITTI 提供的 Regionet 检测器的检测结果进行实验。通过选择不同场景下的 KITTI 数据集序列进行对比实验。

2.2 评价指标

本文采用基于 CLEARMOT 的评价指标,包括身份切换总数 (IDS ↓); 多目标跟踪准确度 (MOTA ↑); 多目标跟踪精度 (MOTP ↑),为度量匹配对在所有帧中的估计位置的总误差; 精确度 (Prcn ↑),为度量在序列的所有跟踪结果中正确跟踪对象的比率; 召回率 (Recall ↑),为度量序列正确跟踪对象占真实值的比率; MT ↑,表示目标 80% 被跟踪到的轨迹数量; ML ↓,表示目标 80% 丢失的轨迹数量; FM ↓,表示所有轨迹被打断的总次数。其中,多目标跟踪准确度考虑所有帧的漏报 (FN ↓)、误报 (FP ↓) 和 ID 切换次数 (IDS ↓),综合显示了 MOT 的性能。

2.3 数据集结果分析

本文对考虑“Car”类的序列都进行了跟踪实验,20 个序列下总体跟踪结果如表 1 所示。本文所提算法在整体序列下达到了最高的跟踪准确度及目标身份的一致性。除此以外,本文还选取了其中具有代表性特殊场景的序列显示多目标跟踪结果。可以看出,本文所提算法能较好地处理智能车未知运动带来的数据关联失效问题,具有更强的适应性,表现出了更加具有竞争力的跟踪效果。

表 1 各跟踪算法结果对比

Table 1 Comparison of tracking results of each algorithm

算法	MOTA ↑	MOTP ↑	Recall ↑	Prcn ↑	MT ↑	ML ↓	FP ↓	FN ↓	IDS ↓	FM ↓
TBD ^[7] (离线)	60.2	81.6	69.3	89.7	218	100	2 524	9 760	373	354
SORT ^[9]	43.0	74.4	59.2	79.5	135	149	4 849	12 965	307	546
CMOT ^[13]	45.0	79.5	64.0	78.4	147	135	5 593	11 452	428	489
RMOT ^[21]	52.2	75.7	67.7	82.5	185	96	4 566	10 268	358	519
SCEA ^[22]	65.9	80.8	74.2	90.6	318	85	2 452	8 210	176	223
本文	67.1	81.5	74.2	91.7	343	102	2 121	8 208	130	198

1) 智能车旋转、平移问题

当智能车出现颠簸、急转弯、快速运动等情况时,传统的目标运动模型在图像坐标中的假设不再成立,从而造成数据关联的失败。现有目标跟踪算法对于此类情况的适应性并不强,为此本文选取 KITTI 数据集中智能车旋转平移较大的 0004 和 0014 序列,各算法评价指标定量比较如表 2 所示。本文利用一致性点漂移算法的空间变换关系来刻画智能车的运动,减少了车体运动对数据

关联的影响。而在目标密集场景下,由于距离特征不明显,极易产生关联错误。本文考虑不同场景下的特征权重问题,通过外观和运动的自适应特征融合函数,减少场景变化对特征权重的影响,提高数据关联的准确度。如图 3 所示,可以看出在智能车大幅度旋转运动下,传统算法 SCEA(虚线框)无法对上下帧目标进行数据关联,造成大量漏报,本文所提算法(实线框)则展示出更强的适应性,跟踪效果提升明显。

表 2 各跟踪算法在智能车剧烈运动场景下定量比较
Table 2 Quantitative comparison under severe camera motion

序列	算法	MOTA ↑	MOTP ↑	Recall ↑	Prcn ↑	MT ↑	ML ↓	FP ↓	FN ↓	IDS ↓	FM ↓
0004	TBD(离线)	61.8	81.8	71.4	89.3	7	6	80	268	10	3
	SORT	11.1	67.1	50.7	56.7	2	11	363	462	8	21
	CMOT	34.2	78.9	57.0	72.2	2	10	206	403	8	11
	RMOT	42.9	75.5	68.9	73.4	3	5	234	291	10	16
	SCEA	73.7	79.6	82.7	90.6	15	5	80	162	4	4
	本文	82.3	80.9	89.0	93.3	21	4	60	103	3	4
0014	TBD(离线)	54.8	76.7	56.9	98.0	2	2	6	227	5	6
	SORT	26.4	70.9	42.3	74.6	0	5	76	304	8	13
	CMOT	23.0	73.4	39.5	73.2	1	5	76	319	11	11
	RMOT	29.2	70.2	45.7	75.8	2	3	77	286	10	13
	SCEA	57.5	78.5	59.8	97.5	4	3	8	212	4	6
	本文	59.8	79.1	62.0	97.9	5	3	7	200	5	8



图 3 智能车剧烈运动场景下定性比较

Fig. 3 Qualitative comparison under severe camera motion

2) 目标遮挡问题

目标遮挡是相机目标跟踪过程中的一大难题,当目标被遮挡时,其特征就会变得不可靠甚至丢掉特征,导致数据关联失败。为此,本文选取 KITTI 数据集中存在大量遮挡场景的 0001 序列,各算法评价指标定量比较如表 3 所示。本文采用拟合预测的方式继续对遮挡的目标进行跟踪,利用轨迹的前几帧状态预测目标被遮挡时的状态,可以有效对漏报轨迹进行恢复,从而保持了轨迹 ID 的一致性。具体跟踪效果如图 4 所示,其中,SCEA 算法(虚线框)当目标出现遮挡情况时,目标轨迹被立刻终

止,而目标再次出现后,ID 已经发生切换;而本文所提算法(实线框)在不同遮挡情况下,利用对目标状态的预测,图 4 中第 380 帧,可以有效的对遮挡目标进行恢复,并保持 ID 的一致性。

除此上述特殊场景之外,本文还评估了所提方法在 KITTI 的其他非特殊序列的跟踪性能。不同 MOT 方法的重要指标定量比较结果如图 5 所示,结果显示本文所提算法在不同序列的不同指标下均展示出非常具有竞争力的结果。算法在提高多目标跟踪准确度(MOTA ↑)的同时精确度(Prcn ↑)也维持在很高的水平,且在目标误报、目标

表 3 各跟踪算法在目标遮挡场景下定量比较

Table 3 Quantitative comparison under object occlusion scenarios

序列	算法	MOTA ↑	MOTP ↑	Recall ↑	Prcn ↑	MT ↑	ML ↓	FP ↓	FN ↓	IDS ↓	FM ↓
0001	TBD(离线)	57.4	81.4	70.7	85.4	39	15	350	848	36	27
	SORT	40.5	69.4	62.7	75.1	29	20	604	1 080	39	61
	CMOT	24.8	78.1	60.3	65.3	19	17	930	1 150	98	71
	RMOT	37.0	71.6	65.0	70.8	24	13	777	1 014	35	48
	SCEA	64.4	80.9	78.8	85.0	55	13	402	615	15	11
	本文	67.4	82.3	79.3	87.1	57	18	341	599	6	8

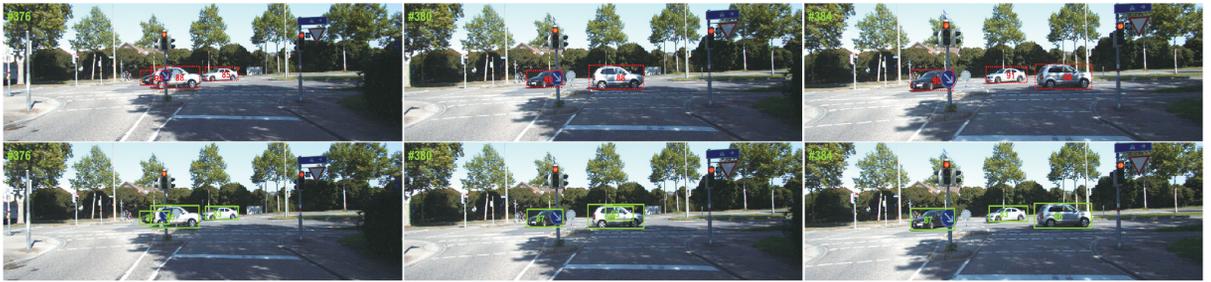


图 4 目标遮挡场景下定性比较

Fig. 4 Qualitative comparison under object occlusion scenarios

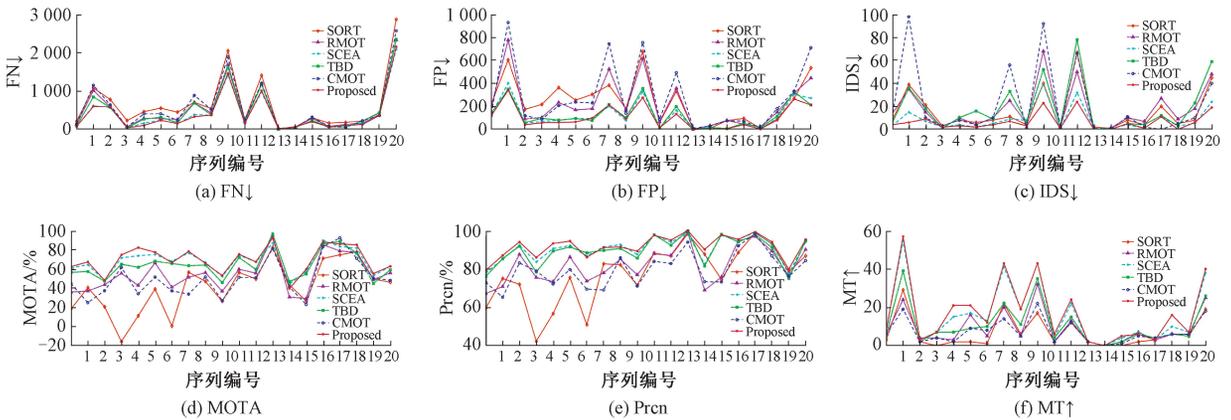


图 5 重要指标定量比较曲线结果

Fig. 5 Quantitative comparison curve results of important indicators

漏报和 ID 切换次数上, 本文所提算法明显优于其他算法。

在算法实时性方面, 各算法在不同平台处理一帧的速度对比如表 4 所示。由于本文算法加入了图像序列特征提取步骤, 导致运行速度略有下降, 但对于数据的处理

速率依旧大于相机的采样频率 (10 Hz)。并且本文算法相对于对比算法在多目标跟踪准确度、多目标跟踪精确度、目标误报、目标漏报和 ID 切换次数等指标上都有明显提升, 具有最好的跟踪结果。

表 4 不同算法运行时间比较

Table 4 Comparison of running time of different algorithms

算法	SORT	CMOT	RMOT	SCEA	本文
运行环境	RTX 2060	Ryzen 2.90 GHz	Ryzen 2.90 GHz	Ryzen 2.90 GHz	Ryzen 2.90 GHz
语言	Python	MATLAB	MATLAB	MATLAB	MATLAB
每帧用时/s	0.002 0	0.047 6	0.021 2	0.052 6	0.058 8

2.4 实拍实验结果分析

本文所提算法在智能车平台上进行了测试,实拍数据由3个序列组成,包含颠簸、急弯和一般道路场景。图像序列分辨率平均为 $1\ 280\times 960$,采样频率为10 Hz。采用Fast-RCNN检测算法,并选择检测阈值为0.8得到的检测结果进行跟踪实验。

为了验证本文算法的有效性,将本文算法与SCEA算法进行对比。场景跟踪效果如图6~8所示。

由图6可以看出,由于出现颠簸情况,SCEA算法

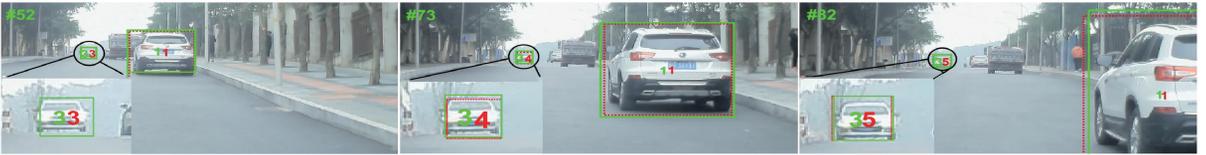


图6 颠簸情况下的跟踪结果对比

Fig. 6 Comparison of tracking results under bumpy conditions

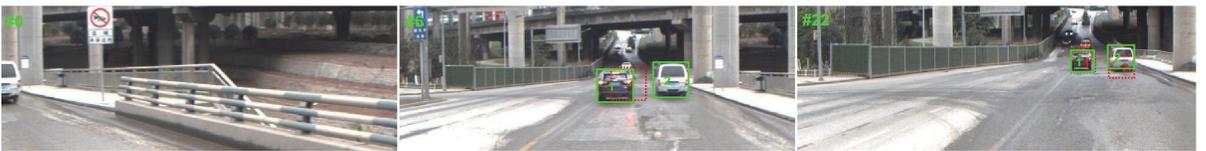


图7 急弯及颠簸情况下的跟踪结果对比

Fig. 7 Comparison of tracking results in sharp bends and bumps



图8 一般道路下的跟踪结果对比

Fig. 8 Comparison of tracking results under general roads

表5 实拍实验下定量比较

Table 5 Quantitative comparison under real shooting experiments

算法	MOTA \uparrow	MOTP \uparrow	FP \downarrow	FN \downarrow	IDS \downarrow
SCEA	73.3	74.9	456	280	19
本文	80.6	75.6	324	210	14

3 结论

本文提出了一种基于CPD的视觉多目标跟踪方法。针对智能车未知运动下的多目标跟踪问题,利用一致性点漂移算法刻画智能车前后时刻的未知运动,其次,构建外观相似性和运动一致性的自适应特征融合函数,提高

(虚线框)对同一个目标的连续跟踪中出现多次ID切换。由图7可以看出,由第0帧~22帧出现急转弯情况,且路面较为颠簸,SCEA算法跟踪目标出现位置漂移和跟踪失败的情况。而本文所提算法(实线框)在上述情况下均可以有效维持目标ID的一致性,保持跟踪的鲁棒性。在图8的一般道路下,本文算法也能保持较好的跟踪效果。此外本文算法与对比算法在实拍序列下的重要参数定量比较结果如表5所示,结果表明所提算法具有更好的跟踪性能。

数据关联的准确性,以实现面向智能车的鲁棒目标跟踪。本文在KITTI公开数据集和实拍实验数据上进行实验评估,结果表明本文所提的算法相比于现有多目标跟踪算法具有更好的性能。其中,在KITTI数据集智能车剧烈运动场景下,所提算法多目标跟踪准确率相比于SCEA算法提高了6.3%。

参考文献

- [1] MABROUK A B, ZAGROUBA E. Abnormal behavior recognition for intelligent video surveillance systems: A review[J]. Expert Systems with Applications, 2018, 91: 480-491.
- [2] LI P, QIN T. Stereo vision-based semantic 3D object and ego-motion tracking for autonomous driving [C].

- Proceedings of the European Conference on Computer Vision(ECCV), 2018: 646-661.
- [3] ZHU H, ZHANG G, LI Y, et al. A novel robust Kalman filter with unknown non-stationary heavy-tailed noise[J]. Automatica, 2021, 127: 109511.
- [4] ZHU H, YUEN K V, MIHAYLOVA L, et al. Overview of environment perception for intelligent vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(10): 2584-2601.
- [5] LUO W, XING J, MILAN A, et al. Multiple object tracking a literature review [J]. Artificial Intelligence, 2021, 293: 103448.
- [6] ZHU H, MI J, LI Y, et al. VB-Kalman based localization for connected vehicles with delayed and lost measurements: Theory and experiments [J]. IEEE/ASME Transactions on Mechatronics, 2022, 27(3): 1370-1378.
- [7] GEIGER A, LAUER M, WOJEK C, et al. 3D traffic scene understanding from movable platforms [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(5): 1012-1025.
- [8] 洪锋, 鲁昌华, 蒋薇薇, 等. 基于时空一致性约束视频目标车辆的检测与跟踪算法研究 [J]. 电子测量与仪器学报, 2022, 36(3): 105-112.
- HONG F, LU CH H, JIANG W W, et al. Research on vehicle detection& tracking algorithm based on spatio-temporal consistent dual-stream network video target [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 105-112.
- [9] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking [C]. IEEE International Conference on Image Processing, 2016: 3464-3468.
- [10] 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, 23(6): 5668-5677.
- [11] WENG X, WANG J, HELD D, et al. 3D multi-object tracking: A baseline and new evaluation metrics [C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, 2020: 10359-10366.
- [12] LUO W, STENGER B, ZHAO X, et al. Trajectories as topics: Multi-object tracking by topic discovery [J]. IEEE Transactions on Image Processing, 2018, 28(1): 240-252.
- [13] BAO F, CAO Y, ZHANG S, et al. Using segmentation with multi-scale selective kernel for visual object tracking [J]. IEEE Signal Processing Letters, 2022, 29: 553-557.
- [14] BAE S H, YOON K J. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 1218-1225.
- [15] SADEGHIAN A, ALAHI A, SAVARESE S. Tracking the untrackable: Learning to track multiple cues with long-term dependencies [C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017: 300-311.
- [16] YOU S, YAO H, XU C. Multi-object tracking with spatial-temporal topology-based detector [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(5): 3023-3035.
- [17] TIAN W, LAUER M, CHEN L. Online multi-object tracking using joint domain information in traffic scenarios [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(1): 374-384.
- [18] BAE S H, YOON K J. Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(3): 595-610.
- [19] 郑浦, 白宏阳, 李政茂, 等. 抖动干扰下运动目标精准检测与跟踪算法设计 [J]. 仪器仪表学报, 2019, 40(11): 90-98.
- ZHENG P, BAI H Y, LI ZH M, et al. Design of accurate detection and tracking algorithm for moving target under jitter interference [J]. Chinese Journal of Scientific Instrument, 2019, 40(11): 90-98.
- [20] 张伟俊, 钟胜, 徐文辉. 融合显著性与运动信息的相关滤波跟踪算法 [J]. 自动化学报, 2021, 47(7): 1572-1588.
- ZHANG W J, ZHONG SH, XU W H. Correlation filter based visual tracking integrating saliency and motion cues [J]. Acta Automatica Sinica, 2021, 47(7): 1572-1588.

- [21] YOON J H, YANG M H, LIM J, et al. Bayesian multi-object tracking using motion context from multiple objects[J]. IEEE Winter Conference on Applications of Computer Vision, 2015: 33-40.
- [22] YOON J H, LEE C R, YANG M H, et al. Structural constraint data association for online multi-object tracking[J]. International Journal of Computer Vision, 2019, 127(1): 1-21.
- [23] 余仁伟, 朱浩, 蔡昌恺. 基于薄板样条函数的无人机多目标跟踪算法[J]. 仪器仪表学报, 2021, 42(3): 168-176.
- YU R W, ZHU H, CAI CH K. Multi-object tracking algorithm for UAV based on the thin plate spline function[J]. Chinese Journal of Scientific Instrument, 2021, 42(3): 168-176.
- [24] MYRONENKO A, SONG X. Point set registration: Coherent point drift[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(12): 2262-2275.
- [25] ZHENG Z, WANG P, LIU W, et al. Distance-IOU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [26] ZHU H, ZHANG G, LI Y, et al. An adaptive Kalman filter with inaccurate noise covariances in the presence of outliers[J]. IEEE Transactions on Automatic Control, 2022, 67(1): 374-381.

- [27] ZHU H, WANG Q, LI Y F, et al. Variational Bayesian based localization for intelligent vehicle using lidar and GPS data fusion: Algorithm and experiments[J]. IEEE/ASME Transactions on Mechatronics, 2022, DOI: 10.1109/TMECH.2022.3187975.

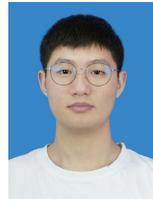
作者简介



朱浩 (通信作者), 2005 年于大连海事大学获得学士学位, 分别在 2007 年于重庆大学获得硕士学位, 2012 年于重庆大学获得博士学位, 现为重庆邮电大学教授, 主要研究方向为信号处理、信息融合、多目标跟踪。

E-mail: sandwish@163.com

Zhu Hao (Corresponding author) received his B. Sc. degree from Dalian Maritime University in 2005, received M. Sc. degree from Chongqing University in 2007 and Ph. D. degree from Chongqing University in 2012. He is currently a professor at Chongqing University of Posts and Telecommunications. His main research interests include single processing, information fusion and multi-object tracking.



李鑫, 2020 年于成都信息工程大学获得学士学位, 现于重庆邮电大学硕士研究生, 主要研究方向为智能汽车, 目标跟踪。

E-mail: lixinfirst163@163.com。

Li Xin received his B. Sc. degree from Chengdu University of Information Technology in 2020. He is currently M. Sc. candidate at Chongqing University of Posts and Telecommunications. His main research interests include intelligent vehicle and multi-object tracking.