DOI: 10.16452/j.cnki.sdkjzk.2024.05.011 文章编号:1672-3767(2024)05-0110-10

面向自助收银台场景下的遮挡目标全关联跟踪模型

王金富,赵建立,房 胜,李 哲
(山东科技大学 计算机科学与工程学院,山东 青岛 266590)

摘 要:自助收银台监控视频中商品目标之间频繁发生的遮挡现象会导致目标外观信息缺失,同时静止状态商品的运动信息无法为目标关联提供有价值的跟踪线索,导致自助收银台场景下目标跟踪困难。本研究提出一种遮挡目标全关联跟踪模型,将自助收银台场景下商品目标逐帧运动造成的遮挡现象理解为一个渐变的推理过程。模型使用基于检测的跟踪范式在目标外观信息缺失和运动信息为零时,提出遮挡率和层次信息作为遮挡目标关联的辅助信息,并借助卡尔曼滤波算法对遮挡现象推理过程中多个目标和不同轨迹之间进行关联。实验结果表明,本研究方法能够提高自助收银台场景下的商品目标跟踪精度,且遮挡率和层次信息能够有效减少目标轨迹碎片的数量,多目标跟踪精度和身份识别 F1 分数分别达到 80.7%和 80.4%,比 ByteTrack 模型分别提升 10.6%和 9.8%。

关键词:多目标跟踪;目标检测;全关联;遮挡;自助收银台
中图分类号:TP391 **文献标志码**:A

Fully-associated tracking model for occluded objects inself-service cashier scenarios

WANG Jinfu, ZHAO Jianli, FANG Sheng, LI Zhe
(College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: The frequent occlusion between commodity objects in self-service cashier surveillance videos can lead to the problem of missing appearance information of the objects, and the motion information of stationary products cannot provide valuable tracking clues for object association, leading to the difficulty of object tracking in self-service cashier scenarios. In this study, a fully associated tracking model for occluded objects was proposed and the occlusion phenomena caused by the frame-by-frame movement of commodity objects in the self-service cashier scenarios were interpreted as a gradual inference process. The tracking by detection (TBD) paradigm was used to propose occlusion rate and hierarchical information as auxiliary information for occlusion object association when the appearance information of the object was missing and the motion information was zero. The Kalman filtering algorithm was used to complete the association between multiple objects and different trajectories in the inference process of occlusion phenomena. The experimental results show that the proposed method can improve the tracking accuracy of commodity objects in self-service cashier scenarios and the occlusion rate and hierarchical information can effectively reduce the number of target trajectory fragments, with mutiple object tracking accuracy and identification F1 score reaching 80.7% and 80.4% respectively, 10.6% and 9.8% higher than the ByteTrack model.

Key words: multiple object tracking; object detection; fully-association; occlusion; self-service cashier

计算机视觉领域中,多目标跟踪任务与越来越多的场景[1-3]相结合表现出了强大活力和独特价值。自助收银台作为商超自助结算设备,在提高结算效率、降低人工成本、提升消费体验等方面有重要作用。但自助

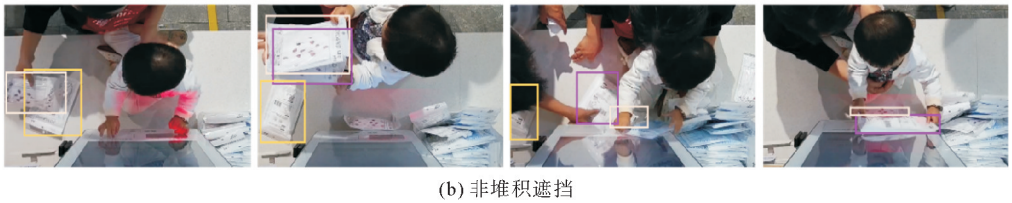
收稿日期:2023-06-22
基金项目:山东省自然科学基金项目(ZR2022MF325);山东省科技型中小企业创新能力提升工程项目(2021TSGC1072)
作者简介:王金富(1997—),男,山东潍坊人,硕士研究生,主要从事视觉目标跟踪相关研究.
赵建立(1977—),男,山东青岛人,教授,博士,主要从事大数据和人工智能领域的研究,本文通信作者.
E-mail:jlzhao@sdust.edu.cn

收银台无法判断商品扫码、付款的状态。本研究通过跟踪商品目标获取商品运动轨迹,为判断商品扫码、付款状态提供必要信息,进而减少因自助收银而带来的货损问题。

目标跟踪问题分为目标检测和目标关联两个子问题。目标检测旨在从帧序列中提取目标信息,目标关联旨在将同一目标在帧序列中的位置串联起来。现有跟踪模型中,检测和跟踪联合(joint detection and tracking,JDT)范式将目标检测和目标关联两问题联合起来,通过信息交流提升精度。文献[4]充分挖掘目标运动信息,设计了一个不依赖外观的跟踪模型。文献[5]通过目标外观信息完成目标跟踪。文献[6-7]利用 Transformer 模型实现良好的跟踪精度,但对遮挡现象导致的视觉信息缺失问题的跟踪效果不好。基于检测的跟踪(tracking by detection,TBD)范式较为灵活,能够根据场景特点在两个子问题上分别选择合适模型。近年来,YOLO 系列模型[8-9]为 TBD 范式提供了可靠的目标检测支撑。Deep Sort[10]、OC-Sort[11]都是以卡尔曼滤波为基础的目标关联算法,虽然能够取得良好的跟踪效果,但为保证模型精准度放弃了遮挡现象造成的低阈值目标。ByteTrack[12]首次使用二次关联算法,但只采用运动信息对低阈值目标进行关联,无法有效跟踪静止状态的目标。针对遮挡现象,文献[13-14]选择依赖外观特征的算法来判断遮挡目标是否再次出现,却忽略了多个目标外观相似的情况。文献[15]将遮挡目标轨迹分成各个时期片段,通过较大的相似度计算量将碎片合并成一条轨迹。Stadler 等[16]设计了一种轨迹回归方法来解决遮挡造成的轨迹碎片问题,但仅适用于动态目标跟踪。

自助收银台场景如图 1 所示,不同颜色框代表不同的商品目标。该场景中频繁发生的遮挡现象会导致目标同时存在水平位置相近、运动信息相同、外观相似或缺失等问题,而上述研究大多使用运动信息或外观信息来解决目标跟踪问题。虽然它们难以完成自助收银台场景中遮挡目标的跟踪任务,却充分证明了 TBD 范式的灵活性以及模型与场景相匹配的重要性。

(a) 堆积遮挡



(b) 非堆积遮挡

图 1　自助收银台场景下遮挡现象示意图

Fig. 1　Schematic diagram of occlusion phenomenon in self-service cashier scenarios

针对上述问题,本研究在 TBD 范式基础上提出以遮挡率和层次信息为辅助信息的遮挡目标二次处理全关联模型。根据遮挡的渐变过程,在运动信息判别的基础上对遮挡目标分类处理。引入遮挡率和目标层次信息对动态目标进行逐帧推理,结合目标外观信息和运动信息完成对该过程中多个目标和不同轨迹之间的关联。静态目标在遮挡率变化预测的基础上,通过目标层次信息对比和继承判断的方式,保证静态目标被遮挡前后的轨迹连贯性,进而减少轨迹碎片。

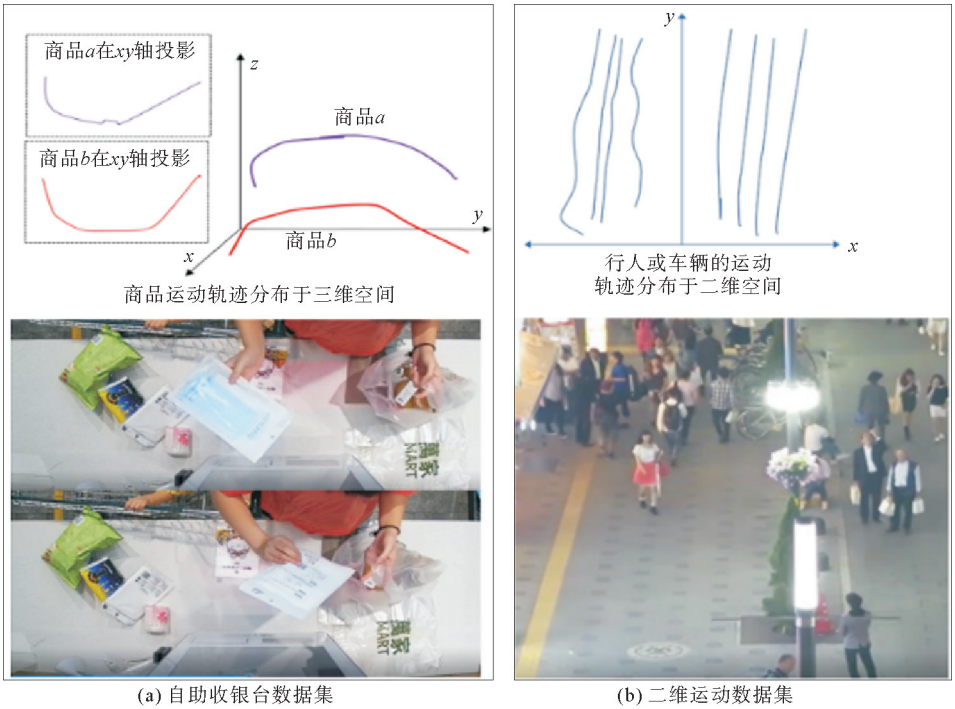
1　自助收银台场景分析

将自助收银台场景的监控视频作为多目标跟踪数据集。通过大量观察、统计分组、随机抽样、对比推断、相关分析等方法,完成数据集遮挡现象分析并给出遮挡率和层次信息的定义。

1.1　自助收银台数据集分析

自助收银台因为设备条件、环境和成本等因素,视频数据仅提供自上至下俯视角度的二维水平信息。针

对目标在高度上的遮挡、覆盖等现象,模型难以获取可用的三维信息。利用自助收银台数据集的二维信息跟踪三维空间内高度信息变化频繁的商品目标,不仅增加了遮挡现象产生的频率,也对模型提出了更高的要求。对比很多数据集的大场景、小目标,自助收银台数据集的场景范围较小,目标尺度相对较大。具体情形如图 2,其中图 2(a)为自助收银台数据集,图 2(b)为二维运动数据集。



(a) 自助收银台数据集 (b) 二维运动数据集

图 2 数据集高度信息比较示意图

Fig. 2 Schematic diagram for comparing height information of datasets

1.2 遮挡现象分析

由 1.1 节可知,商品目标仅含有二维水平位置信息,但商品轨迹属于三维空间轨迹。当多个目标的第三维度信息相近时,产生的遮挡现象可进行逐帧递推。初始时自助收银台不存在商品目标,目标从下到上依次增加产生遮挡现象,然后目标从上到下依次被拿走,遮挡消失。其中,遮挡率逐帧递增形成普通遮挡和严重遮挡;目标间层次信息依次叠加形成堆积遮挡。因为数据集中很多商品目标是静止状态,而遮挡产生是动态目标形成的,故遮挡现象中目标跟踪的推理可行性较高。另外,新出现目标以运动状态从边缘区域进入,不会出现遮挡现象。

设某场景下目标 a 的三维位置集合 Q_a:$\{(x_{at_i},y_{at_i},z_{at_i}),(x_{at_j},y_{at_j},z_{at_j}),\cdots\}$。其中,$(x_{at_i},y_{at_i})$ 为 a 在 t_i 帧的水平面坐标,z_{at_i} 为 a 在 t_i 帧的高度信息。很多数据集中多目标间的位置关系:$(x_{ct_j},y_{ct_j})\neq(x_{dt_j},y_{dt_j})$且 $z_{ct_j}\approx z_{dt_j}$。故此类目标很少会产生因水平位置相近、高度位置不同的遮挡现象。自助收银台数据集中高度信息的变化带来多个目标在某一时刻:$(x_{et_k},y_{et_k})\approx(x_{ft_k},y_{ft_k})$,但 $z_{et_k}\neq z_{ft_k}$,即目标 e 和 f 在 t_k 时刻会产生水平位置相近、高度位置不同的非堆积遮挡现象。当多于两个目标水平位置相近且高度位置不同时则会产生堆积遮挡现象。

一般而言,商品借助人手才会具备运动能力,动态商品目标相对较少。当多个目标在面积有限的自助收银台上发生遮挡时,静态目标的运动信息对于水平位置的预测会因为目标间水平位置的重叠而难以区分。若遮挡的是同种商品(同种包装、同款商品、外观相似商品等,如图 1(b)所示)还会导致目标的外观特征无法为目标关联提供有价值的线索。针对上述问题,模型在遮挡现象分析的基础上提出遮挡率和层次信息作为辅助变量进行目标跟踪的逐帧推理。

1) 目标遮挡率。将某个目标及其所有相近目标的位置重叠关系量化为目标遮挡率,计算式为:

$$C_{\mathrm{box}_a}^{t}=\frac{s\{(x,y,\gamma,h)_{\mathrm{box}_a}^{t}\cap[(x,y,\gamma,h)_{\mathrm{box}_1}^{t}\cup\cdots\cup(x,y,\gamma,h)_{\mathrm{box}_M}^{t})]\}}{s\{(x,y,\gamma,h)_{\mathrm{box}_a}^{t}\}}。\tag{1}$$

式中:$C_{\mathrm{box}_a}^{t}$ 为 a 在 t 帧的遮挡率,box_a 为目标 a 的矩形检测框;$(x,y,\gamma,h)_{\mathrm{box}_a}^{t}$ 为 t 帧中 a 的矩形检测框在图像坐标系的位置信息,(x,y)为 a 的中心坐标,γ 为 a 的长宽比,h 为 a 的高;$s\{(x,y,\gamma,h)_{\mathrm{box}_a}^{t}\}$为根据 a 的坐标信息得到的面积,M 为与 a 相邻的目标总数。当两个目标的遮挡率大于等于 80%时,目标存在难以检测的严重遮挡关系,故不考虑此类情况。另外,遮挡率受目标的大小、长宽比影响较大。

2) 层次信息。多件商品叠放一起时,引入层次信息作为辅助变量,强调堆积遮挡时目标间的上下层关系。以两个目标间的遮挡率为基础对层次信息进行更新,将 a 对其相邻目标 b 的遮挡率量化为:

$$C_{(\mathrm{box}_b,\mathrm{box}_a)}^{t}=\frac{s\{(x,y,\gamma,h)_{\mathrm{box}_a}^{t}\cap(x,y,\gamma,h)_{\mathrm{box}_b}^{t}\}}{s\{(x,y,\gamma,h)_{\mathrm{box}_a}^{t}\}}。\tag{2}$$

式中,$C_{(\mathrm{box}_b,\mathrm{box}_a)}^{t}$为 a 对 b 造成的遮挡率,取值范围为 $[0,1]$。通过式(2)更新目标的层次信息,每个商品的层次信息初始值为 1,当 a 与其他目标存在遮挡关系,即 $1\geqslant C_{(\mathrm{box}_b,\mathrm{box}_a)}^{t}>0(b=1,2,\cdots,M)$时,$a$ 的层次信息需要叠加相应的层次信息,计算式为:

$$T^{a}=\sum_{b=1}^{M}T^{a'}+T^{b}。\tag{3}$$

式中:b 为目标 a 遮挡的某个目标;T^{b} 为 b 的层次信息;M 为被 a 遮挡的目标总数;$T^{a'}$ 为 a 当前层次信息,T^{a} 为更新后 a 的层次信息。层次信息用于比较和区分上下层目标,特别是外观相似的商品目标。

当 a 不再遮挡 b 目标,即 $C_{(\mathrm{box}_b,\mathrm{box}_a)}^{t}=0(b=1,2,\cdots,M)$时,$a$ 的层次信息需要减去相应目标的层次信息,计算式为:

$$T^{a}=\sum_{b=1}^{M}T^{a'}-T^{b}。\tag{4}$$

另外,计算目标遮挡率和层次信息时,通过目标动静状态的判别分类处理。若两个遮挡目标有一个处于静态,则动态目标层次信息增加,静态目标层次信息不变。若两个目标同时处于静态或动态,仅计算遮挡率,不需要层次信息来区分上下层关系。因为同为静态目标已经完成层次信息计算和推理,而动态目标间层次信息会因为目标的运动而不断变化。

3) 遮挡现象分类。为了更好地推理和解决遮挡目标的跟踪问题,在式(2)遮挡率的基础上将两件商品之间的非堆积遮挡分为普通遮挡和严重遮挡。当 $C_{(\mathrm{box}_b,\mathrm{box}_a)}^{t}<80\%$时,认为两个目标之间存在普通遮挡关系;当 $C_{(\mathrm{box}_b,\mathrm{box}_a)}^{t}\geqslant 80\%$时,则认为两个目标之间存在严重遮挡关系。

在式(3)、式(4)层次信息的基础上定义非堆积遮挡和堆积遮挡。非堆积遮挡是指两个商品之间形成上下两层的遮挡关系,堆积遮挡是指至少三件商品在高度上存在遮挡关系。

2 遮挡目标全关联跟踪模型

针对自助收银台场景下遮挡现象导致的目标难以关联和轨迹碎片问题,在二次关联算法[12]的基础上提出一种遮挡目标全关联跟踪模型,模型整体结构如图 3 所示。首先使用目标检测器从帧序列中提取 t 帧的目标信息,然后通过运动信息和外观信息进行目标关联,并将关联成功的目标和轨迹更新至 T_{rack}^{t} 中。值得注意的是,一次关联根据阈值范围和通过卡尔曼滤波得到的预测值,将低阈值和低预测值目标舍弃,在许多通用数据集中有效保证了目标关联的准确性。但自助收银台场景复杂多变的遮挡现象使目标舍弃变得频繁,多目标跟踪精度难以保证。因此,针对一次关联失败的目标集合 T_{arget}^{t} 和未更新的轨迹集合 T_{race}^{t},结合遮挡率和层次信息进行二次关联以减少遮挡目标的舍弃,其结果更新至 T_{rack}^{t},进入下一帧。

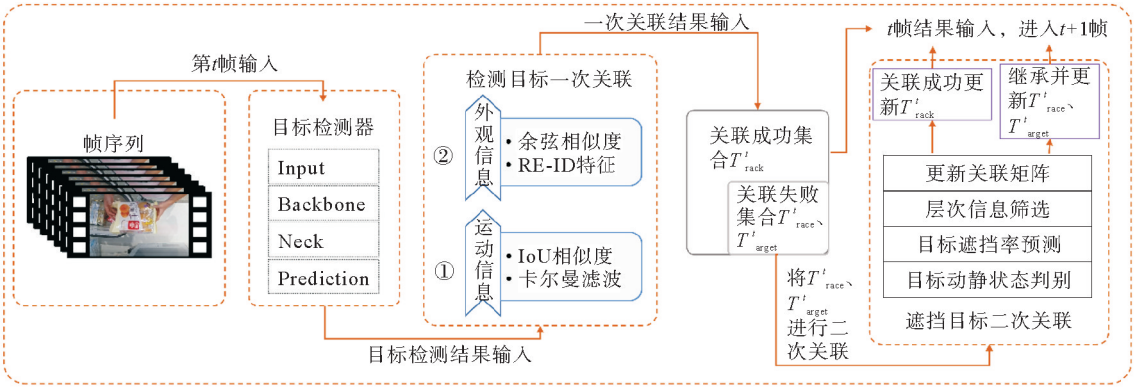


图 3 全关联模型整体架构图

Fig. 3 Overall architecture of the fully-associated model

2.1 目标一次关联

选用 YOLOv5 作为目标检测器，其结构包括 Input、Backbone、Neck、Prediction 四部分。将自助收银台数据集目标设置为：手、商品、手机、其他共 4 类。为减少高阈值目标对遮挡目标造成的干扰和计算量，首先利用卡尔曼滤波算法完成对运动的信息处理，并结合外观信息进行一次关联，其步骤如下。

1) 卡尔曼滤波算法。对目标检测器的输出结果进行卡尔曼滤波预测：

$$(x,y,\gamma,h)\rightarrow(x,y,\gamma,h,\dot{x},\dot{y},\dot{\gamma},\dot{h})。\tag{5}$$

式中：(x,y)为目标中心坐标；γ 为目标长宽比；h 为目标高度；$(\dot{x},\dot{y},\dot{\gamma},\dot{h})$为假设目标匀速直线运动通过卡尔曼滤波算法预测的结果。

2) 运动信息。利用对轨迹的预测位置和当前帧的检测位置进行马氏距离计算：

$$d^{(1)}(i,j)=(\boldsymbol{d}_j-\boldsymbol{y}_i)^{\mathrm{T}}\boldsymbol{S}_i^{-1}(\boldsymbol{d}_j-\boldsymbol{y}_i)。\tag{6}$$

式中：$d^{(1)}(i,j)$为马氏距离，$\boldsymbol{d}_j$ 为第 j 个检测框位置，$\boldsymbol{y}_i$ 为第 i 条轨迹的目标预测位置，$\boldsymbol{S}_i$ 为预测位置与平均追踪位置之间的协方差矩阵。根据马氏距离计算交并比(intersection of union，IoU)相似度，得到运动信息的关联结果。

3) 外观信息。利用外观特征提取网络得到的重识别(re-identifaction，RE-ID)特征，计算上一帧目标轨迹 T^{t-1}_{rack} 与当前帧目标 j 的外观特征相似度：

$$d^{(2)}(i,j)=\min\{1-\boldsymbol{r}_j^{\mathrm{T}}\boldsymbol{r}_k^{(i)}\mid\boldsymbol{r}_k^{(i)}\in\mathbf{R}_i\}。\tag{7}$$

式中：$\boldsymbol{r}_j^{\mathrm{T}}$、$\boldsymbol{r}_k^{(i)}$ 为归一化后第 j 个检测目标与第 i 个轨迹特征向量之间的余弦距离，$d^{(2)}(i,j)$为第 j 个检测目标与第 i 个轨迹中 N 个特征向量的余弦距离最小值，$\mathbf{R}_i$ 表示一条轨迹中外观特征向量的集合。

4) 一次关联。结合运动信息和外观信息，目标一次关联矩阵权重为：

$$c_{i,j}=\lambda d^{(1)}(i,j)+(1-\lambda)d^{(2)}(i,j)。\tag{8}$$

式中：$c_{i,j}$ 为关联权重，λ 为关联系数。因为商品目标间的 $d^{(2)}(i,j)$大小接近，故外观信息权重较低，λ 设为 0.8。最后，利用匈牙利算法和级联匹配完成目标关联并更新至 T^t_{rack}。根据一次关联结果，针对遮挡现象造成的低阈值目标进行二次关联。

2.2 目标的二次关联

结合自助收银台场景总结和遮挡现象分析，本小节在遮挡率和层次信息的基础上，针对一次关联中舍弃的遮挡目标进行二次关联。

2.2.1 目标动静状态判别

通过运动信息判别遮挡目标的动静状态进行分类处理，单目标状态判别式为：

$$s_{\mathrm{tate}}=\begin{cases}0, & v=0\ \text{且}\ \dot{v}=0;\\ 1, & \text{其他}。\end{cases}\tag{9}$$

式中：v 为目标运动速度，$\dot{v}$ 为目标预测速度。当目标的 v 和 $\dot{v}$ 同时为 0 时，状态值为 0，表示目标为静止状

态,其他情况下目标状态值为 1,表示目标为运动状态。静态目标遮挡率虽然较高,但其视觉信息不变,故优先考虑继承的处理方式。动态目标位于目标上层会改变视觉信息,优先进行目标关联。

2.2.2 动态目标关联

动态目标关联重点在于目标遮挡的过程性变化。本数据集中动态目标数量较少且层次信息较大,一次关联失败的主要原因在于非堆积遮挡。通过轨迹信息进行遮挡率分析,对动态目标优先进行关联处理。对未更新轨迹集合 T^t_{race}(含继承轨迹)、未关联目标集合 T^t_{arget},假设目标遮挡率线性变化,用拉格朗日公式对未更新轨迹当前帧的遮挡率进行预测:

$$\begin{cases}\Delta C^t_{\text{box}_a}=C^{t-1}_{\text{box}_a}-C^{t-2}_{\text{box}_a},\\ \hat{C}^t_{\text{box}_a}=C^{t-1}_{\text{box}_a}+\Delta C^t_{\text{box}_a}。\end{cases}\tag{10}$$

式中:$\Delta C^t_{\text{box}_a}$ 为目标 a 的遮挡率变化,$C^{t-1}_{\text{box}_a}$ 为目标 a 在 $t-1$ 帧的遮挡率,$t-1$ 为帧序号,$\hat{C}^t_{\text{box}_a}$ 为目标 a 在 t 帧的遮挡率预测值。通过 $\Delta C^t_{\text{box}_a}$ 的正负,判断目标遮挡变化方向是否一致。利用式(10)对动态目标的轨迹进行遮挡率预测:若预测值大于 80%,即当前帧无法检测到该目标,保留其轨迹信息继承到下一帧;若预测值小于 80%,则对该轨迹进行适配目标预测。利用式(11)进行阈值筛选:

$$|\hat{C}^t_{\text{box}_a}-C^t_{\text{box}_N}|<t^{(1)}。\tag{11}$$

式中:$\hat{C}^t_{\text{box}_a}$ 为 T^t_{rack} 中某一目标轨迹的遮挡率预测值,$C^t_{\text{box}_N}$ 为当前帧 t 中 T^t_{arget} 中相邻目标遮挡率,$t^{(1)}$ 为预测阈值。关联权重更新算式为:

$$c_{i,j}=\frac{1-|\hat{C}^t_{\text{box}_a}-C^t_{\text{box}_a}|}{2}+\frac{\lambda d^{(1)}(i,j)+(1-\lambda)d^{(2)}(i,j)}{2}。\tag{12}$$

最后,通过匈牙利算法和卡尔曼滤波更新算法完成动态目标关联。

2.2.3 静态目标处理

静态目标的遮挡中,对 T^t_{race} 内轨迹的层次信息较高或遮挡率较小的目标优先进行关联,并根据关联结果进行下一帧的继承判断。当目标遮挡关系保持相对不变时,若其层次信息较低或遮挡率较大,优先考虑继承处理。当某一静态目标的被遮挡关系消失时,模型会在目标层次信息对比和遮挡率预测的基础上进行数据关联。因为静态目标遮挡率较大,故放弃外观特征计算,利用式(13)进行目标二次关联:

$$c_{i,j}=\frac{1-|\hat{C}^t_{\text{box}_a}-C^t_{\text{box}_a}|}{2}+\frac{d^{(1)}(i,j)}{2}。\tag{13}$$

最后,通过匈牙利算法和卡尔曼滤波更新算法完成对当前帧的目标二次关联和确认。遮挡目标全关联跟踪算法伪代码如算法 1 所示。需要说明的是,静态目标遮挡现象存在极端情形:大商品始终严重遮挡小商品。这是一项艰难的目标跟踪挑战,本研究不予考虑。

算法 1 遮挡目标全关联跟踪算法

输入:YOLO 目标检测器结果逐帧输入(每一帧目标检测框坐标、目标检测结果置信度、目标类别、目标外观特征信息)

输出:目标跟踪结果(在目标检测器结果的基础上对每一帧的目标赋予不同的 ID,并将视频中目标运动的轨迹逐帧连接并显示出来)

1) 开始进行目标检测

2) **循环**:每当第 t 帧处理结束,进入循环

3) 获取当前帧的检测结果集合 T^t_{arget}

4) 根据目标检测结果通过卡尔曼滤波获取目标的动静状态值 s_{tate}

5) 根据目标检测框坐标计算检测目标的遮挡率 $C^t_{\text{box}_a}$

6) 第一次目标关联:

7) 卡尔曼滤波预测

8) 使用式(6)计算马氏距离 $d^{(1)}(i,j)$

9) 使用式(7)计算目标的外观特征相似度 $d^{(2)}(i,j)$

10) 根据式(8)的结果,利用匈牙利算法和卡尔曼预测算法完成目标第一次关联

11) 结束

12) 当前帧第一次关联后更新的轨迹集合 T^t_{rack} 和未更新轨迹集合 T^t_{race}
13) 当前帧未与轨迹关联的目标集合 T^t_{arget}
14) 第二次目标关联:
15) 根据目标动静状态值进行分类处理
16) 通过式(2)计算目标遮挡率 $C^t_{(\text{box}_b,\text{box}_a)}$,更新目标的层次关系 T^a
17) 将层次信息较低和遮挡率较高的静态目标和静态轨迹保持原状态继承到下一帧
18) 根据卡尔曼滤波算法判断动态轨迹是否确认为静态轨迹,判断静态目标是否确认为动态轨迹
19) 使用式(6)计算马氏距离 $d^{(1)}(i,j)$
20) 使用式(7)计算目标的外观特征相似度 $d^{(2)}(i,j)$
21) 目标和轨迹的层次信息相近优先使用式(10)获得 T^t_{race} 中遮挡率的预测值 $\widehat{C}^t_{\text{box}_a}$
22) 根据式(12)和式(13)的结果,利用匈牙利算法和卡尔曼算法预测完成目标第二次关联
23) 当前帧二次关联后更新的轨迹集合 T^t_{rack} 和当前帧的轨迹集合 T^t_{race}
24) 当前帧未与轨迹关联的目标集合 T^t_{arget},可能存在边缘区域新出现的目标
25) 结束
26) 判断是否存在下一帧,如果存在下一帧则继续循环
27) **否则,结束循环**
28) t 帧的检测目标集合 T^t_{arget} 根据目标关联的结果不断减少,直到集合内的目标无法与轨迹中的目标相关联
29) 根据跟踪结果和目标 ID,将视频中目标运动的轨迹逐帧连接并显示出来
30) $C^t_{\text{box}_a}$ 通过式(1)计算得到
31) T^a 通过式(3)、式(4)计算得到
32) s_{tate} 通过卡尔曼滤波算法和式(9)计算得到
33) 结束

3 实验与分析

3.1 数据集

自助收银台数据集在综合考虑节假日、工作日等多种自助收银台使用情况的基础上收集而成,共有 516 个视频,每个视频至少 30 s,且视频中自助收银台至少使用一次,共得到 5 000 张照片。其中,训练集、验证集、测试集按照 6∶2∶2 的比例划分。数据集根据遮挡现象分为非堆积遮挡类和堆积遮挡类。以一个完整视频为基本单位进行分类,当视频中没有遮挡现象或出现非堆积遮挡现象时,该视频归为非堆积遮挡类;当视频中出现堆积遮挡现象时,该视频归为堆积遮挡类。数据集内容如图 4 所示,每一列代表一位用户的扫码付款行为,同一列从上到下目标数量或遮挡次数增加。图 4(a)中,非堆积遮挡类视频中商品目标数量较少或目标堆叠次数低于商品个数。该类目标跟踪推理过程中的干扰因素较少,目标跟踪精度能够满足自助收银台防货损需求。图 4(b)中,自助收银台的有限空间会导致用户购买商品数量较多或目标面积较大时产生堆积遮挡现象。商品目标跟踪的推理过程中,除目标遮挡造成的视觉信息缺失问题外,静态目标的运动信息和相似外观使得两件商品难以区分。该类多目标跟踪任务不仅要求目标检测结果的精度高,对目标跟踪算法也提出更高的要求。另外,用户对于目标运动的方式也会有不同的影响,例如手持多件袋装商品、多个刚性商品自始至终叠放在一起等,这都会增加目标跟踪的难度。

(a) 非堆积遮挡类 (b) 堆积遮挡类

图 4 自助收银台数据集

Fig. 4 Self-service cashier dataset

3.2 目标检测器及参数设置

YOLOv5 为目标检测器。训练网络输入图像大小为 608×608,batch size 为 32,初始学习率为 0.01,循环学习率为 0.1,学习率动量为 0.973,权重衰减为 0.000 5,交并比损失系数为 0.05。本研究模型综合准确率可达到 92.1%,召回率为 79%,当 IoU 阈值设为 0.5 时,平均精度为 0.856。

3.3 遮挡率分析

遮挡率分布比如图 5 所示。非堆积遮挡类很多目标间不存在遮挡关系,故遮挡率在 0～20%占比最高,在 20%～40%占比最低。堆积遮挡类遮挡率在 60%～80%占比最高,在 20%～40%占比最低。一方面遮挡率分布跟运动导致的遮挡率均匀变化有关,另一方面跟商品数量有较大的关系。总体来说,遮挡率分布于较高或较低的值。因为自助收银台数据集中形状规则的刚体目标受力面积较大,遮挡率较低的情况存在于非刚性商品的遮挡现象。通过调优,式(11)的 $t^{(1)}$ 设置为 0.1,式(12)的 λ 设为 0.2,视频帧率根据数据集设为 10。

将所有遮挡率的变化大小取为正值,增量分布如图 6 所示。可以看出,遮挡率增量普遍低于 30%,大部分的增量区间在 15%～20%之间。实验结果证明,自助收银台数据集目标的遮挡率变化是可以捕捉和预测的,无法关联的遮挡目标与遮挡率直接相关且正相关。

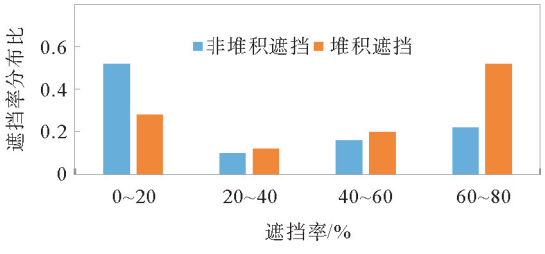


图 5 目标遮挡率分布

Fig. 5 Distribution of target occlusion rate

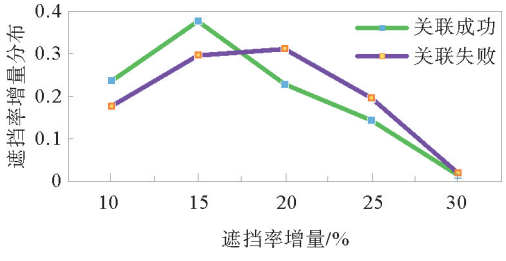


图 6 遮挡率增量分布

Fig. 6 Distribution of incremental occlusion rate

3.4 对比实验

为验证本模型跟踪性能,选择 Motr[6]、TransMOT[7]、ByteTrack[12]、FairMOT[17]、CenterTrack[18]、JDE[19]模型分别对两类视频进行对比,实验结果如表 1、表 2 所示,虽然手和其他类目标会对商品目标跟踪造成干扰,但能满足商品防货损的需求。对比 ByteTrack,本模型的多目标跟踪精度(multiple object tracking accuracy, MOTA)提升 0.3%,身份识别分数(identification F1 score, IDF1)提升 0.5%,目标身份变换(Identity switch,IDs)减少 5 789 次。表 2 中跟踪结果差的主要原因是左右手间 IDs 和商品外观相似的干扰。对堆积遮挡来说,二次关联针对性处理后跟踪精度更高。实验结果表明,针对自助收银台场景下的遮挡现象,对比 ByteTrack,本研究模型的 MOTA 提升 10.6%,IDF1 提升 9.8%,跟踪效果明显提高。

3.5 消融实验

随机选取 20 个视频,以 Deep Sort[10]为基线对一次关联和二次关联进行消融实验,结果如表 3 所示。由表 3 可以看出,在一次关联的基础上增加遮挡率预测和层次信息作为辅助变量,堆积遮挡类目标的 MOTA 下降 3.1%,IDF1 下降 3%。主要原因在于运动信息和外观信息适用于非遮挡目标的轨迹推理,遮挡率预测和层次信息适用于遮挡目标的轨迹推理,两个部分同时进行会造成一定的推理干扰。二次关联中,非堆积遮挡类使用遮挡率预测时 MOTA 提升 0.8%,IDF1 提升 0.5%,IDs 减少 221 次。堆积遮挡类在不考虑层次信息时,使用遮挡率预测会因为上下层目标难以区分而降低跟踪精度。静态遮挡目标通过层次信息优先级对比后,跟踪结果的 MOTA 提升 12.5%,IDF1 提升 15.6%,IDs 次数减少 523 次。

表 1 非堆积遮挡类对比实验

Table 1 Comparison experiment of non-stacked occlusion

模型	MOTA/%	IDF1/%	IDs/次
JDE[19]	66.3	65.9	1 952
CenterTrack[18]	68.5	69.4	1 993
FairMOT[17]	70.5	69.9	2 186
TransMOT[7]	76.5	74.9	1 965
Motr[6]	76.9	76.5	1 962
ByteTrack[12]	80.4	79.9	1 761
本研究	**80.7**	**80.4**	1 865

表 2 堆积遮挡类对比实验

Table 2 Comparativeexperiments of stacked occlusion

模型	MOTA/%	IDF1/%	IDs/次
JDE[19]	29.4	31.9	2 354
CenterTrack[18]	35.1	34.2	2 761
FairMOT[17]	36.6	37.2	2 687
TransMOT[7]	39.4	40.1	2 346
Motr[6]	42.6	40.3	2 156
ByteTrack[12]	48.9	50.3	2 241
本研究	**59.5**	**60.1**	**1 985**

表 3 消融实验

Table 3 Ablation experiment

类组	模型	MOTA/%	IDF1/%	IDs/次
非堆积遮挡类	Deep Sort	79.3	78.7	963
	Deep Sort+遮挡率(一次关联)	79.3	78.7	1 067
	Deep Sort+遮挡率+层次信息(一次关联)	79.5	78.8	986
	Deep Sort+遮挡率(二次关联)	79.9	78.6	1 095
	Deep Sort+遮挡率+层次信息(二次关联)	**80.1**	**79.2**	**742**
堆积遮挡类	Deep Sort	40.2	35.9	2 374
	Deep Sort+遮挡率(一次关联)	36.6	33.8	2 830
	Deep Sort+遮挡率+层次信息(一次关联)	37.1	32.9	2 587
	Deep Sort+遮挡率(二次关联)	37.8	34.7	2 861
	Deep Sort+遮挡率+层次信息(二次关联)	**52.7**	**51.5**	**1 851**

实验发现,当出现抽取底层静态遮挡目标的小概率事件时,模型对遮挡目标跟踪的推理结果会变得不可靠。首先,因为遮挡目标全关联模型假设目标层次信息从下到上增加产生遮挡、层次信息从上到下减小遮挡消失。其次,遮挡率预测使模型在 TBD 范式基础上加深了对目标检测器的依赖,当检测结果不准确或难以检测时,跟踪性能也会受到很大干扰。另外,商品目标运动因为用户的自主操作而复杂多变,当商品发生完全覆盖、翻转和变形等情况时,算法难以奏效。

4 总结与展望

将自助收银台的监控视频作为多目标跟踪数据集,针对非堆积遮挡、堆积遮挡等现象造成的目标关联难题,在 TBD 范式的基础上提出二次处理的全关联模型。模型根据一次关联结果,在目标动静状态判别的基础上,通过遮挡率、层次信息、运动信息等内容实现对遮挡目标关联。针对自助收银台场景下遮挡现象引起的轨迹碎片问题,模型通过遮挡率预测和目标层次信息对比的方法,将遮挡前后的目标轨迹衔接,保证商品目标的轨迹连贯性。下一步工作重点将放在两个方面:一是如何解决遮挡现象中商品目标外观相似导致的跟踪难点,二是如何提高目标检测结果的精度,特别是针对严重遮挡的目标。

参考文献:

[1] HAN T,BAI L,GAO J Y,et al. DR. VIC:Decomposition and reasoning for video individual counting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans,Jun. 18-24,2022:3083-3092.

[2] MUELLER M,SMITH N,GHANEM B. A benchmark and simulator for UAV tracking[C]//14th European Conference on

Computer Vision (ECCV 2016). Amsterdam, Oct. 11-14, 2016: 445-461.

[3] CIOPPA A, GIANCOLA S, DELIEGE A, et al. SoccerNet-tracking: Multiple object tracking dataset and benchmark in soccer videos[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, Jun. 18-24, 2022: 3491-3502.

[4] WANG G A, GU R S, LIU Z Z, et al. Track without appearance: Learn box and tracklet embedding with local and global motion patterns for vehicle tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Oct. 10-17, 2021: 9876-9886.

[5] KIM C, FUXIN L, ALOTAIBI M, et al. Discriminative appearance modeling with multi-track pooling for real-time multi-object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, Jun. 20-25, 2021: 9553-9562.

[6] ZENG F T, DONG B, ZHANG Y A, et al. MOTR: End-to-end multiple-object tracking with transformer[C]//Proceedings of European Conference on Computer Vision. Cham: Springer Nature Switzerland. Tel Aviv, Oct. 23-27, 2022: 659-675.

[7] CHU P, WANG J, YOU Q Z, et al. TransMOT: Spatial-temporal graph transformer for multiple object tracking[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, Jan. 3-7, 2023: 4870-4880.

[8] 徐岩,李永泉,郭晓燕,等. 基于 YOLOv3-tiny 的火焰目标检测算法[J]. 山东科技大学学报(自然科学版),2022,41(6):95-103.
XU Yan, LI Yongquan, GUO Xiaoyan, et al. Flame object detection algorithm based on YOLOv3-tiny[J]. Journal of Shandong University of Science and Technology(Natural Science), 2022, 41(6): 95-103.

[9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Jun. 27-30, 2016: 779-788.

[10] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on IMAGE PROCESsing (ICIP). Beijing, Sep. 17-20, 2017: 3645-3649.

[11] CAO J K, PANG J M, WENG X S, et al. Observation-centric sort: Rethinking sort for robust multi-object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Jun. 17-24, 2023: 9686-9696.

[12] ZHANG Y F, SUN P Z, JIANG Y, et al. Bytetrack: Multi-object tracking by associating every detection box[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, Tel Aviv, Oct. 23-27, 2022: 1-21.

[13] AL-SHAKARJI N M, BUNYAK F, SEETHARAMAN G, et al. Multi-object tracking cascade with multi-step data association and occlusion handling[C]//2018 the 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, Auckland, Nov. 27-30, 2018: 1-6.

[14] 朱松豪,吕址函,宋杰. 基于特征学习的双路径红外-可见光行人重识别算法[J]. 山东科技大学学报(自然科学版),2022,41(5):82-90.
ZHU Songhao, LÜ Zhihan, SONG Jie. Dual-path infrared-visible person re-identification algorithm based on feature learning[J]. Journal of Shandong University of Science and Technology(Natural Science), 2022, 41(5): 82-90.

[15] SONG Y M, YOON K, YOON Y C, et al. Online multi-object tracking with GMPHD filter and occlusion group management[J]. IEEE Access, 2019, 7: 165103-165121.

[16] STADLER D, BEYERER J. Improving multiple pedestrian tracking by track management and occlusion handling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, Jun. 20-25, 2021: 10958-10967.

[17] ZHANG Y F, WANG C Y, WANG X G, et al. FairMOT: On the fairness of detection and re-identification in multiple object tracking[J]. International Journal of Computer Vision, 2021, 129: 3069-3087.

[18] ZHOU X Y, KOLTUN V, KRÄHENBÜHL P. Tracking objects as points[C]//European Conference on Computer Vision. Cham: Springer International Publishing, Glasgow, Aug. 23-28, 2020: 474-490.

[19] WANG Z D, ZHENG L, LIU Y X, et al. Towards real-time multi-object tracking[C]//European Conference on Computer Vision. Cham: Springer International Publishing, Glasgow, Aug. 23-28, 2020: 107-122.

(责任编辑:傅 游)