

中图法分类号: TP39 文献标识码: A 文章编号: 1006-8961(2024)07-1984-14

论文引用格式: Wang K, Dai F, Guo W Y, Wang J F and Wang X X. 2024. Integrating similarity and interaction force between objects for multiple object tracking. Journal of Image and Graphics, 29(07):1984-1997(王凯, 戴芳, 郭文艳, 王军锋, 王小侠. 2024. 融合目标相似性和作用力的多目标跟踪. 中国图象图形学报, 29(07):1984-1997)[DOI:10.11834/jig.230340]

融合目标相似性和作用力的多目标跟踪

王凯, 戴芳*, 郭文艳, 王军锋, 王小侠

西安理工大学理学院, 西安 710076

摘要: 目的 多目标跟踪是计算机视觉一个重要的研究方向,为了解决多目标跟踪中错跟和漏跟导致跟踪精度低的问题,提出一种融合目标相似性和作用力的多目标跟踪算法。**方法** 首先将多目标跟踪问题转化为一个最大后验概率问题,其次将最大后验概率问题映射到网络流中,利用最小代价流寻找最优路径,这样获得的最优路径就是目标轨迹。为了计算网络流中目标节点之间的代价,从以下两方面考虑:1)将目标的外观、运动和位置信息三者结合,计算目标间的相似度;2)考虑目标与目标的相互影响,参考社会力模型中个体之间的吸引力来计算目标节点之间的作用力。**结果** 在MOT15、MOT16和MOT17共3个公开数据集进行实验评估并与12种方法进行比较,实验结果表明,本文算法在MOTA (multiple object tracking accuracy)、MT (mostly tracked tracklets)、ML (mostly lost tracklets)、FP (false positives)、FN (false negatives)等指标上明显优于OACDASM (online association by continuous-discrete appearance similarity measurement)、STURE (spatial-temporal mutual representation learning)、IQHMOT (identity-quantity harmonic multi-object tracking)和GCNNMatch (graph convolutional neural network match)等典型算法。在MOT15数据集中选取ETH-Bahnhof、TUD-Stadtmitte与PETS09-S2L1 3个视频序列进行消融实验,验证增加目标作用力之后的数据关联结果,消融实验结果表明,增加目标作用力之后可以改善目标跟踪的精度和其他指标,尤其在遮挡不明显的视频序列中。**结论** 本文在目标多特征的基础之上增加目标节点间作用力,加强了目标间的数据关联,减少错跟的目标数量,有效地提高了目标跟踪的精度。

关键词: 多目标跟踪(MOT);最小代价流;目标作用力;目标相似性;社会力模型

Integrating similarity and interaction force between objects for multiple object tracking

Wang Kai, Dai Fang*, Guo Wenyan, Wang Junfeng, Wang Xiaoxia

School of Science, Xi'an University of Technology, Xi'an 710076, China

Abstract: Objective In the field of computer vision, object tracking is a critical task. Currently, many different types of multi-object tracking algorithms have been proposed, which usually include the following steps: object detection, feature extraction, similarity calculation, data association, and ID assignment. In this process, the object in the video sequence is first detected and a rectangular box is drawn to label the specific object detected. Then, the features of each object are extracted, such as location and appearance features. Then, the similarity of the object is determined by calculating the probability that the object in the adjacent video frames is the same object. Finally, through data association, the objects

收稿日期:2023-06-20;修回日期:2023-09-06;预印本日期:2023-09-13

* 通信作者:戴芳 daifang@xaut.edu.cn

基金项目:国家自然科学基金项目(61976176)

Supported by: National Natural Science Foundation of China (61976176)

belonging to the same object in adjacent frames are associated and an ID is assigned to each object precisely. This paper mainly focuses on the feature extraction and data association stage of the object, using combined features to represent the characteristics of the object and then increasing the interaction force between the objects for enhanced data association to address the problem of mistracking in object tracking and thus improve the accuracy of object tracking. **Method** First, the multi-object tracking problem is transformed into a maximum a posteriori probability problem. Second, the maximum a posteriori probability problem is mapped to the network flow and the minimum cost flow is used to find the optimal path. To calculate the cost between the object nodes in the network flow, we consider two aspects. First, we calculate the similarity between the objects by combining the appearance, motion, and position information of the objects. Second, we consider the interaction between objects and objects, referring to the attraction between individuals in the social force model to calculate the force between object nodes. **Result** The experimental evaluation on three public datasets MOT15, MOT16, and MOT17 and a comparison with the latest 12 methods show that the proposed algorithm performs well in multiple object tracking accuracy, mostly tracked tracklets, mostly lost tracklets, false positives, false negatives, and other indicators; these indicators are significantly better than those of online association by continuous-discrete appearance similarity measurement, spatial-temporal mutual representation learning, identity-quantity harmonic multi-object tracking, graph convolutional neural network match (GCNNMatch), and other typical algorithms. Ablation experiments were carried out on three video sequences of TUD-Stadtmitte, ETH-Bahnhof, and PETS09-S2L1 in the MOT15 dataset to verify the data association results after increasing the object force. The ablation experimental results show that the object tracking accuracy and other indicators can be improved after increasing the object force, especially in video sequences where occlusion is not obvious. **Conclusion** In this paper, the force between the target nodes is added based on the target multi-feature, which strengthens the data association between the targets, reduces the number of misfollowed targets, and effectively improves the accuracy of target tracking.

Key words: multi-object tracking(MOT); minimum cost flow; object force; object similarity; social force model

0 引言

多目标跟踪(multiple object tracking, MOT)是一项基础的计算机视觉任务,主要应用于自动驾驶、机器人导航和视频分析等(Law 和 Deng, 2020),其重要性已经得到了学术界和工业界的重视(Ciaparrone 等, 2020; Sun 等, 2021)。与单目标跟踪相比, MOT 需要解决许多复杂的问题,如目标的出现或消失、目标的遮挡、目标尺寸变化以及背景干扰等。研究者

们提出了许多 MOT 方法,但仍然不能达到理想的跟踪结果,特别是在拥挤和混乱的场景中,如图 1 所示,每个矩形框代表一个目标,矩形框上的数字表示目标的 ID(identification)。因此, MOT 依然是一个值得继续研究且具有挑战性的问题(Abdulghafoor 和 Abdullah, 2022; Sun 等, 2021)。

依赖于滤波技术的多目标跟踪,如 Kalman 滤波(Li 等, 2010)和粒子滤波(Okuma 等, 2004),可以应用于实时跟踪任务,但它们通常容易出现跟踪不准确等问题。近年来,随着目标检测性能的显著提升,

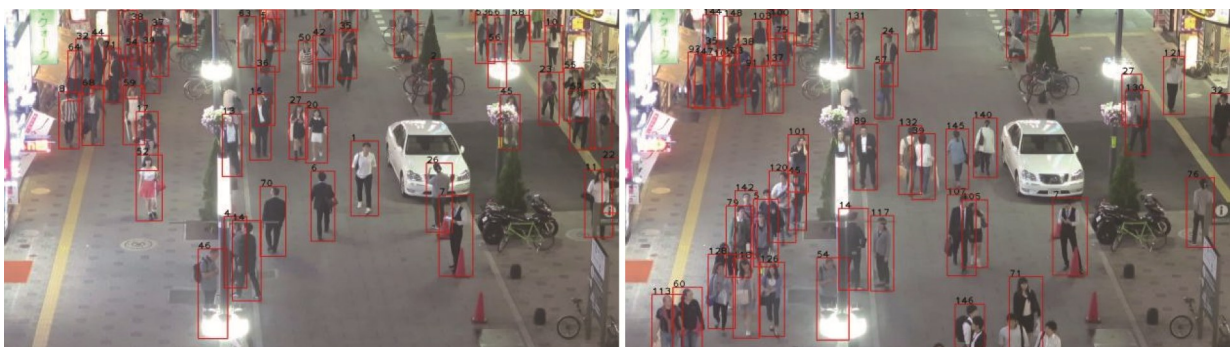


图1 拥挤场景中的 MOT

Fig. 1 MOT in a crowded scene

越来越多的研究人员考虑利用 TBD (tracking-by-detection) 方法对多个目标进行跟踪。TBD 方法是通过数据关联将独立的目标连接起来形成轨迹,所以目标跟踪问题被表述为目标之间的关联,也就是说只需要将检测产生的目标通过关联算法连接起来。由于检测结果作为数据关联的输入,对关联结果和最终的跟踪结果有很大影响,因此 TBD 方法的关键在于目标检测结果和数据关联的准确性。交并比 (intersection over union, IoU) 算法 (Bochini 等, 2017) 是一种离线的基于检测的 MOT 算法。该算法指出,当检测的精度很高,视频的帧率也很高时,可以通过计算目标与目标之间的 IoU,再通过设定阈值来判断目标与背景即可完成目标跟踪任务。但由于未引入任何视频帧间信息、运动模型和外观模型,导致漏跟和错跟问题难以解决。若出现频繁遮挡、目标形变的情况,会导致 ID 频繁切换,且极其依赖目标检测算法的性能。

机器学习技术的迅速发展,使得目标检测有了很大的进步,但数据关联仍然是 MOT 的一个挑战。Bewley 等人 (2016) 使用 Faster R-CNN (region convolutional neural network) 对目标进行检测,利用 Kalman 滤波处理视频帧间的关联性,最后用 Hungarian 算法 (Kuhn, 1955) 进行目标匹配,使目标跟踪性能提升了几十倍。然而此算法的 ID 切换问题比较明显,不适用目标遮挡以及运动不稳定的跟踪任务。随后又结合动作和外观信息 (Wojke 等, 2017), 使用卷积神经网络 (convolutional neural network, CNN) 提取目标特征,增加了对缺失目标和遮挡目标的鲁棒性,且算法快速高效,易于实现。ByteTrack 算法 (Zhang 等, 2022) 是一种基于 YOLOX (you only look once) 目标检测的跟踪算法,与其他非 ReID (re-identification) 算法一样,仅仅使用目标检测所得到的检测框进行跟踪,跟踪方法与 DeepSort (Wojke 等, 2017) 相同,均使用了 Kalman 滤波预测检测框,使用 Hungarian 算法进行目标和轨迹间的匹配。ByteTrack 算法的最大创新在于使用了低分框对跟踪算法进行了二次匹配,有效优化了跟踪过程中因为遮挡造成的 ID 切换问题。

将数据关联视为全局优化问题的基于检测的 MOT 方法,注重设计优化算法,如能量最小化 (Milan 等, 2014, 2016)、条件随机场 (Bombardelli 等, 2019)、最小代价网络流 (Ren 等, 2021; Berclaz 等, 2011; Han

等, 2017) 等。条件随机场多用于目标跟踪的建模问题,而能量函数最小化和最小代价网络流作为一种优化方法实现 MOT。Pirsiavash 等人 (2011) 提出一种近似贪婪的算法解决网络流的问题。该算法能够保证较高的跟踪精度,但对于目标的 IDs (identification_switch) 处理不理想。Zhang 等人 (2008) 通过最小代价流算法找到全局最优关联轨迹,为了解决目标遮挡问题,构建了显式遮挡模型。虽然该算法具有较高的时间复杂度,但是在实验中发现,对目标跟踪是非常高效的。Liu 等人 (2017) 提出一种虚拟缺失检测和遮挡模型,对于遮挡和遗漏的目标,建立虚拟节点,通过反嵌入式迭代最短路径算法,解决网络流优化问题。这种方法虽然优于很多跟踪器,但是在有些场景中出现了很高的 IDs。Hu 等人 (2020) 所提出的跟踪器成功地使用了一个单目标跟踪器来处理不可靠的检测,并同时改进了数据关联。此外,他们结合了 ReID 网络和单目标跟踪器分别学习外观和运动模型,最后利用最大权重团图算法在线模式对目标进行最优分配。此方法产生了许多冗余的矩形框,而置信度过滤器和非最大抑制算法不能有效地过滤掉不合格的边界框,导致了过多的 IDs。乐应英等人 (2013) 针对低视点多目标跟踪场景的遮挡问题,提出一种能够遮挡自适应感知的多目标跟踪算法。该算法在低视点跟踪场景下,能够减弱部分遮挡、短时全遮挡和长时全遮挡对跟踪性能的影响,跟踪鲁棒性得到了提升。方岚和于凤芹 (2020) 利用轨迹置信度将多目标跟踪分为局部关联和全局关联两个层次。在局部关联中,置信度高的可靠轨迹利用外观、位置大小相似度与当前帧检测点进行关联;在全局关联中,置信度低的不可靠轨迹引入运动模型和有效关联范围进一步关联分段的轨迹,在提取目标外观特征时引入增量线性可判别分析方法以解决身份转换问题,依据新增样本与目标样本均值的外观特征差异自适应地更新目标外观模型。结果表明引入的在线线性可判别外观学习对遮挡产生的身份转换具有良好的解决效果。

上述跟踪器在跟踪目标时忽略了场景环境对目标运动的影响,没有反映目标之间的制约关系,导致目标跟踪的 ID 频繁更换。社会力模型 (social force model) (Helbing 和 Molnár, 1995) 通过考虑目标的排斥力、静态障碍物对目标的排斥力以及引导目标向某一方向前进的驱动力来预测每个个体目标的运

动,提高了目标跟踪的精度。

早期,社会力模型主要用于研究行人的异常检测(Mehran等,2009),常常与概率假设密度(Krishanth等,2017;Feng等,2017)滤波器和粒子滤波器(Ata-ur-Rehman等,2014)结合,用来预测目标在下一帧的位置,实现目标跟踪。Ren等人(2018)将目标分为真实目标和候选目标,利用Hungarian算法进行匹配,最后利用社会力模型加速度项去预测目标在下一帧中的位置,但跟踪精度比较低。Mazzon和Cavallaro(2013)利用改进的社会力模型模拟非观测区域的运动,使用来自一个观察区域的信息初始化社会力模型,然后让模型基于一组兴趣点(目标)和障碍物在非观察区域内传播路径,最后根据目标再次出现的下一个观察区域的可用信息选择最佳预测路径,但是此算法仅对特定的场景具有很好的跟踪效果,适用范围受到了限制。本文中使用最小代价流解决多目标跟踪问题,与其他文献不同的是,在Leal-Taixé等人(2011)的方法中,通过对相同时刻,即同一帧中目标之间的作用力表现社会力模型,此外,还加入了群体模式,作者认为如果两个目标在一段合理的时间彼此靠近,它们很可能属于同一个群体,因此具有相同的速度,所以使用这个群体中所有其他目标的平均速度来预测目标的下一个位置。Zhang等人(2008)的目的是解决目标遮挡引起的跟踪困难,但是,Zhang等人(2008)和Leal-Taixé等人(2011)均没有用到目标的特征信息。本文基于目标检测,只考虑目标之间的关联性,旨在提高目标跟踪

的精度,因此将Leal-Taixé等人(2011)方法中的计算方式引入本文中,考虑计算相邻帧中目标的相互作用力,而不是同一帧中的目标作用力,此外还增加了目标的外观、运动和位置信息,加强数据关联的效果。

本文提出一种融合目标相似性和作用力的多目标跟踪算法。主要研究思路是,将目标跟踪定义为一个最大后验概率问题,基于最小代价网络流算法求解最大后验概率问题的最优解,即寻找目标轨迹。为了精确计算目标与目标之间的相似性,结合了目标的外观、IoU和位置信息,使用全局特征计算目标相似性。实际场景中,每一个目标的运动都会受到其他目标的影响,一个目标的运动速度会受到它的相邻目标的影响。为了减少目标之间的IDs,考虑目标及其邻近目标之间的相互作用力。特别地,对于同一视频帧,如果两个目标距离越近,它们之间的斥力就越大。相邻视频帧中的两个目标的吸引力越大,它们是同一个目标的可能性就越大,引入了社会力模型来计算目标之间的作用力。具体方法就是利用检测到的目标建立一个具有始发点和结束点的网络,其中每一个目标作为网络中的一个节点。然后,将数据关联问题转化为一个最大后验概率问题,再将该问题映射到网络流中,利用最小代价流(minimum cost flow, MCF)求解网络流问题,找到最优轨迹集。社会力模型在不改变问题的复杂度的前提下,以一种有效的方式引入到网络流中,实现了全局最优。本文方法称为NF_SFM算法,图2展示了NF_SFM算法的整个研究框架。

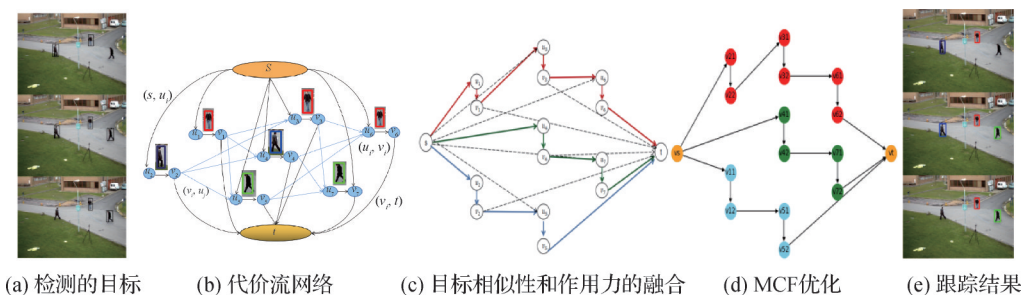


图2 NF_SFM的目标跟踪框架

Fig. 2 Object tracking framework of NF_SFM ((a) detected objects; (b) cost flow network; (c) fusion of object similarity and force; (d) MCF optimization; (e) tracking results)

1 MOT问题

设 $O = \{o_i\}$ 是检测的目标集,第 i 个目标 $o_i =$

(x_i, s_i, a_i, t_i) , 其中, x_i 表示目标位置, s_i 表示目标大小, a_i 表示目标的外观, t_i 表示目标所在的视频帧。定义一个目标的轨迹为一个有序的目标集合 $T_k = \{o_{k_1}, o_{k_2}, \dots, o_{k_n}\}$, 多目标跟踪的任务是找到 $T = \{T_k\}$

这样的集合。数据关联的目的就是在给定检测集 $\mathbf{O}$ 的情况下,最大化 $\mathbf{T}$ 的后验概率。假设检测的目标是独立的,目标跟踪可以表示为一个最大后验概率问题,即

$$\begin{aligned} \mathbf{T}^* &= \arg \max_{\mathbf{T}} P(\mathbf{T}|\mathbf{O}) = \\ &= \arg \max_{\mathbf{T}} P(\mathbf{O}|\mathbf{T})P(\mathbf{T}) = \\ &= \arg \max_{\mathbf{T}} \prod_i P(o_i|\mathbf{T})P(\mathbf{T}) \end{aligned} \quad (1)$$

由于轨迹 $\mathbf{T}$ 的空间巨大,直接优化式(1)变得十分困难。为了减少搜索空间,假设每一个目标只属于一条轨迹,即所有的轨迹不相交,这就转化为 $\mathbf{T}_k \in \mathbf{T}$ 不能有重叠的。进一步假设每个目标的运动是相互独立的,此时,式(1)可以被分解为

$$\begin{aligned} \mathbf{T}^* &= \arg \max_{\mathbf{T}} \prod_i P(o_i|\mathbf{T}) \prod_{\mathbf{T}_k \in \mathbf{T}} P(\mathbf{T}_k) \\ \text{s.t. } &\mathbf{T}_k \cap \mathbf{T}_l = \emptyset, \quad \forall k \neq l \end{aligned} \quad (2)$$

式中, $\mathbf{T}_k$ 和 $\mathbf{T}_l$ 分别表示两条目标轨迹,将每个轨迹建模为一个 Markov 链,即

$$P(o_i|\mathbf{T}) = \begin{cases} 1 - \beta_i & \mathbf{T}_k \in \mathbf{T}, o_i \in \mathbf{T}_k \\ \beta_i & \text{其他} \end{cases} \quad (3)$$

$$P(\mathbf{T}_k) = P_{\text{enter}}(o_{k_0})P_{\text{link}}(o_{k_{i+1}}|o_{k_i})P_{\text{exit}}(o_{k_N}) \quad (4)$$

式中, $P(o_i|\mathbf{T})$ 服从 Bernoulli 分布,用来描述一个目标是真实的还是虚假的目标的概率。 $P(\mathbf{T}_k)$ 是一个 Markov 链,包括初始概率 P_{enter} 、转移概率 P_{link} 和终止概率 P_{exit} 。在这里 $P(o_i|\mathbf{T})$ 不仅可以建模为与 $\mathbf{T}$ 相关的检测,即真实的检测,还可以建模为与 $\mathbf{T}$ 无关的观测,即虚假的目标。这使得该方法可以选择观测值,而不是假设所有的输入都是真实的目标,无需在关联后进行额外的处理以消除错误轨迹。

2 解决 MOT 问题的最小代价流方法

在这一节中,将 MOT 问题表述为最小代价流的优化问题。最小代价流是一种有效的解决整数规划问题的方法。使用最小代价流解决目标跟踪问题具有以下优点:1)为了保证一开始就要选定哪些节点有可能成为起点、哪些节点有可能成为终点,将网络设定为具有源节点 s 和汇聚节点 t 的有向图;2)所有弧的容量都是 1,可以有效防止轨迹重叠;3)为了保证每条弧最多被选择一次,引入过渡弧,这样每个目标用一对节点表示,即一个前节点和一个后节点;4)对于每个目标,前节点通过有向内弧线与 s 相连,

后节点通过有向外弧线与 t 相连,这样的弧线允许每一条流都是轨迹的开始或结束;5)定义每条包含源节点和汇聚节点的弧的代价都是比较大的正值,过渡弧的代价是负值,这样可以避免轨迹过早被终止。设 $G = (N, E)$ 为一有向网络,其节点 v_i 和 u_j 间的代价 $C_{i,j}$ 与每条弧 $E(i, j) \in E$ 相关。视频中连续 3 帧目标构建的网络示意图如图 3 所示,图 3(a)表示连续 3 帧视频序列,图 3(b)表示视频帧中的目标构建的网络示意图,它包含两个特殊的节点,源节点 s 和汇聚点 t ,所有经过图的流量都必须从 s 点出发,到 t 点结束,因此每一个流表示成一条轨迹 $\mathbf{T}_k$ 。给定检测的目标集 $\mathbf{O}$,对于每一个目标 $o_i \in \mathbf{O}$,创建两个节点 u_i, v_i ,由节点 u_i 和 v_i 连接组成的弧 $c(u_i, v_i)$ 的代价为 C_i ,流 $f(u_i, v_i)$ 的流量为 f_i ;由节点 s 和节点 u_i 连接组成的弧 $c(s, u_i)$ 的代价为 $C_{\text{en},i}$,流 $f(s, u_i)$ 的流量为 $f_{\text{en},i}$;由节点 v_i 和 t 连接组成的弧 $c(v_i, t)$ 的代价为 $C_{\text{ex},i}$,流 $f(v_i, t)$ 的流量为 $f_{\text{ex},i}$ 。对每个目标的转移概率 $P_{\text{link}}(o_j|o_i) \neq 0$,由节点 v_i 和 u_j 连接组成的弧 $c(v_i, u_j)$,其代价为 $C_{i,j}$,流 $f(v_i, u_j)$ 的流量为 $f_{i,j}$ 。

在最小代价网络流问题中,目标就是找到使网络上所有流的总代价最小的变量的值,为此,最大后验概率问题可以转换为以下的整数规划问题,即

$$\min \sum_i C_{\text{en},i} f_{\text{en},i} + \sum_{i,j} C_{i,j} f_{i,j} + \sum_i C_{\text{ex},i} f_{\text{ex},i} + \sum_i C_i f_i \quad (5)$$

$$\text{s.t. } \begin{cases} f_{\text{en},i} + \sum_j f_{j,i} = f_i = f_{\text{ex},i} + \sum_j f_{i,j}, \forall i \\ f_{\text{en},i}, f_{\text{ex},i}, f_{i,j}, f_i \in \{0,1\} \end{cases} \quad (6)$$

式中,

$$\begin{cases} C_{\text{en},i} = -\log P_{\text{enter}}(o_i) \\ C_{\text{ex},i} = -\log P_{\text{exit}}(o_i) \\ C_{i,j} = -\log P_{\text{link}}(o_j|o_i) \\ C_i = \log \frac{\beta_i}{1 - \beta_i} \end{cases} \quad (7)$$

式(6)为流量守恒约束。寻找最优关联等价于将流量从源节点 s 发送到汇聚节点 t ,使得总代价最小的每一条流表示一条运动轨迹,从而形成多个目标的运动轨迹,式(7)中的 $C_{\text{en},i} = C_{\text{ex},i} = 0.1$,然而,对于任何更有可能是真实的检测 ($\beta_i < 0.5$),由于检测并不能将所有目标都检测出来,所以在本文中,对任意 $i, \beta_i = 0.4$ 。

最小代价网络流方法是多目标跟踪的一种直观表示。每一条路径都可以解释为一条目标轨迹,从

s 点发出到 t 点接收的流的数量等于目标轨迹的数量,网络 G 中所有流的总代价对应于关联假设的对数似然。流量守恒约束保证了没有流共用一条边,因此轨迹不会重叠。与传统的网络流算法相似,考虑到目标的外观相似度和路径的平滑度的约束,根

据连接两个节点的代价为每条弧 $E(i,j)$ 分配代价 $C_{i,j}$,因此,本文提出一种新的计算代价的方式,即

$$C_{ij} = C_{\text{Fea},ij} + C_{\text{Sfm},ij} \quad (8)$$

式中, $C_{\text{Fea},ij}$ 和 $C_{\text{Sfm},ij}$ 分别表示目标 o_i 和 o_j 间的相似性和相互作用力。

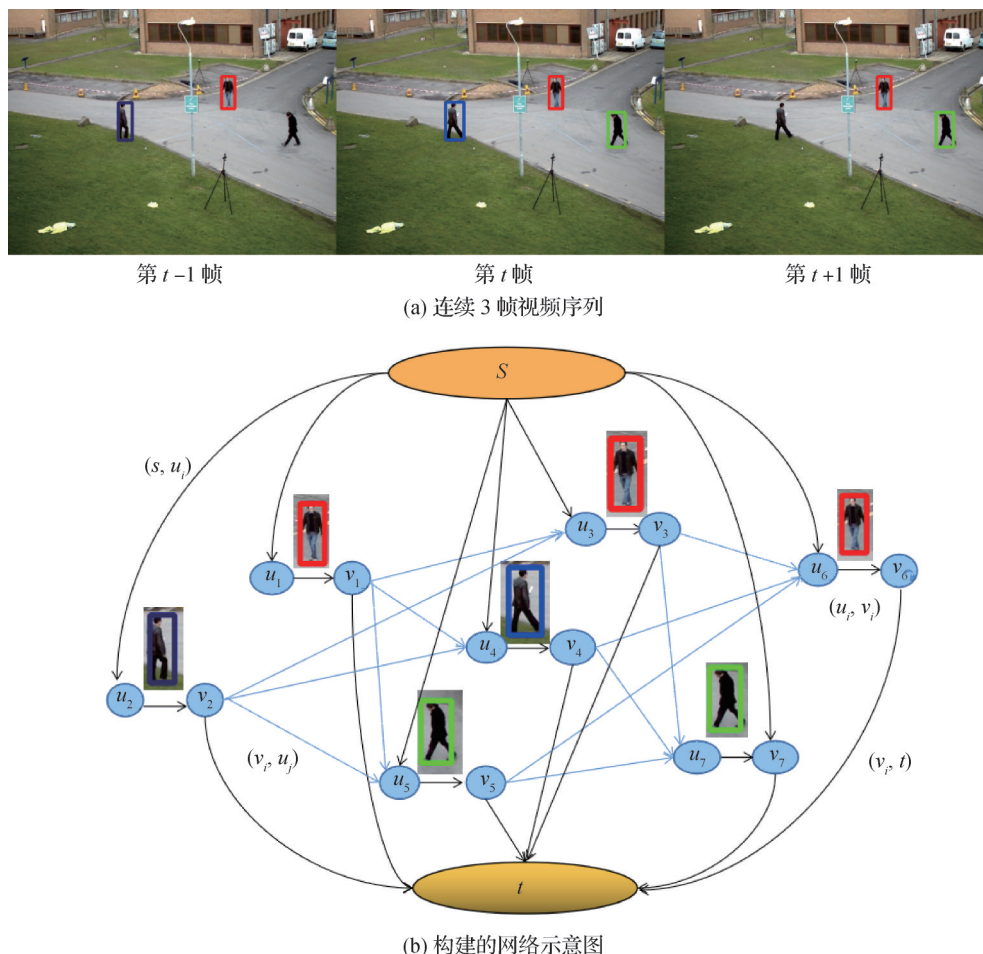


图3 连续3帧视频构建的网络示意图(Leal-Taixé等,2011)

Fig. 3 Schematic diagram of the network constructed by three consecutive frames of video (Leal-Taixé et al., 2011)

((a) video sequence with 3 consecutive frames; (b) schematic of the constructed network)

2.1 目标相似性

在MOT中仅仅使用目标的外观信息也可以达到对目标的关联,但是当目标的外观极其相似的时候,就不能区分是否为同一目标。因此,为了解决这一问题,引入IoU,通过对特定区域的目标计算外观特征,可以解决外观相似目标的IDs,当不同区域的两个目标的矩形框相差较小时,IoU值较为接近,此时,目标的位置坐标就表现得尤为重要,坐标位置可以对目标定位,这样可以更好地关联目标。为了提高复杂场景下的目标跟踪精度,提出了组合目标特征来计算目标间的相似性。对于两个目标 o_i, o_j ,根

据外观、运动和位置信息对跟踪结果的影响,对跟踪影响最大的外观特征设置最大的权重,其次是位置和外观,将3种特征进行加法融合,计算它们之间的相似性 $C_{\text{Fea},ij}$,具体为

$$C_{\text{Fea},ij} = \mu \text{Cor}_{ij} + \nu \text{IOU}_{ij} + (1 - \mu - \nu) \text{Posi}_{ij} \quad (9)$$

式中, Cor_{ij} 表示目标之间的外观相似度, IOU_{ij} 表示目标的运动特征, Posi_{ij} 表示目标的位置信息。由于将以上3种信息分别作为目标相似性度量时,外观相似性表现出比较好的效果,所以在式(9)中取参数 $\mu = 0.5, \nu = 0.2$ 。

在目标跟踪中,最常用的颜色模型是基于RGB

空间的,但这并不符合人们对颜色相似度的视觉判断,所以使用 HSV 空间的色直方图,它更接近人类对颜色的体验和感知。外观代价 $Cor_{i,j}$ 为

$$Cor_{i,j} = \frac{\sum_I (H_i(I) - \bar{H}_i)(H_j(I) - \bar{H}_j)}{\sqrt{\sum_I (H_i(I) - \bar{H}_i)^2 \sum_I (H_j(I) - \bar{H}_j)^2}} \quad (10)$$

式中, H_i, H_j 分别表示目标 o_i, o_j 在 HSV 颜色空间下的颜色直方图, $\bar{H}_i = \frac{1}{N} \sum_J H_i(J)$, N 是直方图的 bin 数, I 表示通道数。

运动信息是由视频连续帧中目标位置的变化给出的。利用目标检测的矩形框,定义运动信息的代价 $IOU_{i,j}$ 为

$$IOU_{i,j} = \frac{a(o_i^t) \cap a(o_j^{t+1})}{a(o_i^t) \cup a(o_j^{t+1})} \quad (11)$$

式中, $a(o_i^t)$ 表示在视频帧 t 中目标 o_i 的矩形框。

考虑到目标的运动,将目标放在一个二维空间的坐标系中。当前帧的目标与下一帧目标之间的位置代价定义 (Boragule 和 Jeon, 2017) 为

$$Posi_{i,j} = 1 - \frac{|x_i - x_j|}{2(w_i + w_j)} - \frac{|y_i - y_j|}{2(h_i + h_j)} \quad (12)$$

式中, x_i, y_i, w_i, h_i 分别表示目标 o_i 的检测框的左上角位置的横坐标、纵坐标、宽度和高度。

2.2 目标作用力

虽然只使用目标的外观、运动和位置信息就可以计算相似性,但是考虑到真实的运动环境,如果目标在运动过程中没有遇到任何障碍,大多数情况下都会沿着一条直线行走。但是,环境变得复杂,运动目标变得拥挤,当空间变得越来越狭小,目标将不再沿着直线行走,目标之间的相互影响在拥挤的环境中表现得尤为重要。为了进一步增强目标之间的关联性,将目标之间的相互作用力也加入到目标相似性的计算范围内。

假设每个目标试图保持一定的速度和方向,在 $t + \Delta t$ 中与在 t 中具有相同的速度,用当前帧目标与下一帧目标之间的距离计算目标之间的相互作用力 (Leal-Taixé 等, 2011), 具体为

$$C_{Sfm,i,j} = -\log F\left(\frac{\|o_j - o_i\|}{\Delta t}, V_{\max}\right) \quad (13)$$

为了能够在任何帧率下工作,观测之间的时间间隔也被考虑在内,因此使用速度测量而不是距离。

假设目标运动速度不能超过最大速度 $V_{\max}$, 将速度映射到概率,并使用高斯误差函数,得到

$$F(V_i, V_{\max}) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{-V_i + \frac{V_{\max}}{2}}{\frac{V_{\max}}{4}}\right) \quad (14)$$

式中, erf 为高斯误差函数,其表达式为

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \quad (15)$$

为了解决优化问题式 (5) 和式 (6), 在求解最小代价流问题时使用动态规划算法计算最优路径。可以通过复杂度 $O(KN \log N)$ 得到全局最优解, K 为最优轨迹数, N 为视频序列的长度,动态规划算法具有更低的计算复杂度。

3 算法性能测试与分析

3.1 数据集

在标准的 MOTChallenge 数据集 MOT15、MOT16 和 MOT17 上对本文方法进行评估,这些数据集包含多个具有挑战性的行人序列:目标之间的频繁遮挡、拥挤的运动场景、不同的视角和相机运动,以及不同的目标大小和帧率。

MOT15 数据集由 11 个训练视频序列和 11 个测试视频序列组成, MOT17 和 MOT16 包含的视频序列是一样的,两个数据集都包含相同的 7 个训练视频序列和 7 个测试视频序列, MOT16 只提供 DPM (deformable parts models) (Felzenszwalb 等, 2008) 产生的检测结果,而 MOT17 不仅提供了基于 DPM 的检测结果,而且提供了两种更好的检测器 Faster R-CNN (Ren 等, 2017) 和 SDP (scale dependent pooling) (Yang 等, 2016) 的检测结果。

3.2 评价指标

本文采用以下 7 项指标评价目标跟踪性能:

1) MOTA (multiple object tracking accuracy), 多目标跟踪精度,用数量描述目标跟踪的精度; 2) MT (mostly tracked tracklets), 大部分跟踪轨迹的数量,包含真实轨迹 80% 以上的节点; 3) ML (mostly lost tracklets), 大部分丢失轨迹的数量,其中包括真实轨迹中小于 20% 的节点; 4) FP (false positives), 假阳性,跟踪模型预测它是真实的,但实际值是假的; 5) FN (false negatives), 假阴性,跟踪模型预测它

应该是假的,但实际值是真的;6)IDs,表示轨迹改变其匹配ID的次数;7)IDF1(idenlity F1 score):正确识别的检测数与真实检测数和计算检测数的平均值之比。

3.3 实验结果

首先,选取 MOT15 数据集的 3 个视频序列, TUD-Stadtmitte 视频序列包含 175 帧图像,这是一个侧视图的过马路场景,视频中目标较少且单个目标很大,但是频繁出现目标互相遮挡。ETH-Bahnhof 视频序列包含 1 000 帧,是用移动摄像机拍摄的行人密集的户外镜头,在户外,光照也给目标跟踪带来了很大的困难,移动的相机加上低照度造成了大量的运动模糊,使这个序列极具挑战性。PETS-2009 视频序列包含 795 帧视频,是一个俯视角度拍摄的静态视频,视频中选择场景的行人密度是 5.6,目标稀疏,运动较快,目标相对较小,同时也出现了频繁的目标遮挡,是一个长时间的视频序列。

选取 3 个不同角度和不同密度的视频序列验证本文中增加目标与目标节点之间的作用力之后,能否改善目标跟踪的精度和其他指标,如表 1 所示,增加目标节点之间的作用力对于目标跟踪的各项指标

都达到了很好的效果。

表 1 显示了目标节点作用力对目标跟踪结果的影响。本节选取 MOT15 数据集上 3 个视频序列,计算目标节点作用力后,在 TUD-Stadtmitte 视频序列中,尽管 MOTA 的值降低了 0.3%,但是 IDF1 的值增加了,并且 IDs 也减少了;在 ETH-Bahnhof 视频序列中,MOTA 提高了 1%,IDs 由 125 降到了 69,其他 5 个指标接近;在 PETS09-S2L1 视频序列中,MOTA 有明显的提高,IDs 也发生了巨大的变化,由原来的 986 缩小到 55,除 ML 之外,其他 4 个指标也优于仅使用目标相似性的值。

由表 1 可知,ETH-Bahnhof、TUD-Stadtmitte 视频序列与 PETS09-S2L1 视频序列相比,FP 和 FN 这两个指标在融合算法后数值变化很小,是因为这两个值受目标检测的影响比较大,所以当检测结果一定时,它们的值没有发生太大的变化。PETS09-S2L1 视频序列上 MOTA 指标的提升幅度远远超过其他两个序列,是因为此视频帧率为所选视频序列中的最低帧率,相邻两帧目标的移动距离小,相互作用力表现得更加明显,且目标稀少,目标与目标之间的距离较大,更加容易区分。

表 1 目标节点作用力对跟踪能力的影响

Table 1 Impact of object node force on tracking ability

视频序列	代价	MOTA/(% (↑))	IDF1/(% (↑))	MT/(↑)	ML/(↑)	FP/(↓)	FN/(↓)	IDs/(↓)
TUD-Stadtmitte	目标相似性	75.9	69.5	6	0	0	273	6
	目标相似性+目标作用力	75.6	73.9	6	0	2	275	5
ETH-Bahnhof	目标相似性	45.4	58.7	99	39	1 939	895	125
	目标相似性+目标作用力	46.2	57.6	99	39	1 940	894	78
PETS09-S2L1	目标相似性	54.8	34.1	13	0	312	724	986
	目标相似性+目标作用力	76.6	48.1	14	0	290	702	56

注:加粗字体表示各序列各列最优结果。“↑”表示值越高越好,“↓”表示值越低越好。

表 1 说明利用目标节点之间的作用力去关联目标之间的相似程度,是一种有效的数据关联方法,对目标跟踪的各项指标有显著的改善。图 4(a)为增加目标作用力之前连续 3 帧的跟踪结果,图 4(b)为增加目标作用力之后 3 帧的跟踪结果,由于在 PETS09-S2L1 视频序列上的效果较明显,所以选取 PETS09-S2L1 视频序列上连续 3 帧跟踪结果进行分析。第 294 帧中,图 4(a)中 ID 为 3 的目标,在跟踪过程中发生了两次 IDs,分别变成了第 295 帧中的 6 和 296 帧中的 8,ID 为 5 的目标在第 296 帧中变为 3,其他目标

也有这种情况发生,这是因为目标之间的关联不紧密,在增加了目标之间的作用力之后,由图 4(b)可知,所有目标的 ID 均没有发生变化,说明增加目标之间的作用力之后对目标之间的关联起到了非常大的作用,减少 IDs,从而提高了目标跟踪的精度。

本文是基于目标检测的跟踪方法,为了提高跟踪的性能,在 MOT15 数据集上采用的检测方法是 DG-net (<https://github.com/TimoK93/ApLift>),由于 MOT16 和 MOT17 包含相同的视频序列,所以对 MOT16 和 MOT17 分别采用了 FasterCNN (Ren 等,

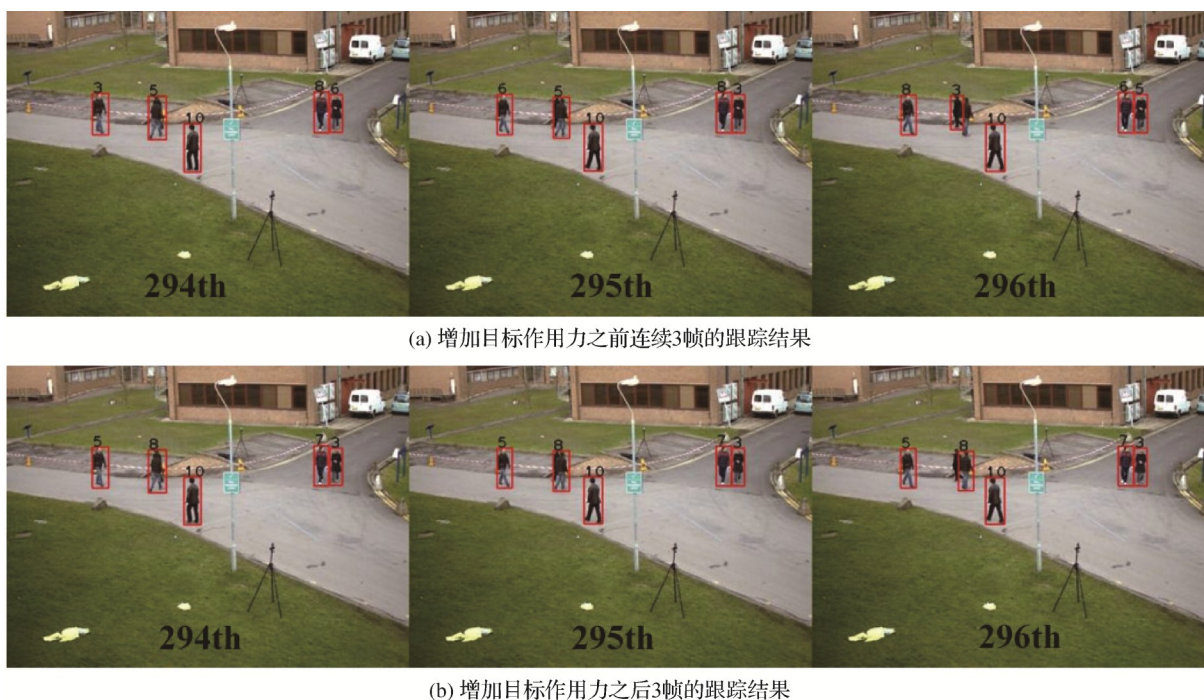


图4 在MOT15数据集上的跟踪结果

Fig. 4 Tracking results on the MOT15 dataset ((a) tracking results for three consecutive frames before increasing target force ; (b) tracking results for 3 frames after increasing target force)

2017)和DG-net,将NF_SFM算法跟踪结果与近几年的MOT算法进行了比较,结果如表2—表4所示。需要将跟踪结果上传至MOTchallenge官网上得到目标跟踪的指标集,由于上传时只需要上传NF_SFM算法的跟踪结果,所以在官网上得不到算法的跟踪时间,这与算法的运行设备有关,例如,PETS-2009视频序列,共795帧,4476个目标,NF_SFM算法在2.40 GHz的设备上在1 min之内找到最优解,即跟踪完成。

表2显示了NF_SFM算法与MOT15排行榜上表现最好的方法的性能比较。Trctrd15 (Bergmann等, 2019)、HSAN (hard-soft attention network) (Liu等, 2021)和OACDASM (online association by continuous-discrete appearance similarity measurement) (Li等, 2022)使用了目标重识别。此外Trctrd15 (Bergmann等, 2019)使用Faster R-CNN来检测目标,GCNNMatch (graph convolutional neural network match) (Papakis等, 2020)是一种深度神经网络的在线MOT方法,将NF_SFM算法与这些方法进行比较,在MOT15数据集上,NF_SFM算法获得了最高的MOTA,为50.78%。可以看出,NF_SFM算法的MOTA明显优于其他几种方法,MT也是这几种方法中的最优值,ML则比最低的175高出6,排在所有方法的第2位,FP和FN也在所有方法中是最优的,但

是NF_SFM算法的IDF1和IDs在所有方法中显示出比较差的结果。

表3显示了NF_SFM算法与MOT16排行榜上表现最好方法的性能比较。IQHMOT (identity-quantity harmonic multi-object tracking) (He等, 2022)使用了Faster R-CNN的检测方法,STURE (spatial-temporal mutual representation learning) (Wang等, 2022)利用在目标关联过程中的各种设计损失,利用神经网络提取出更具区别性的检测和序列表示。该方法通过相互学习时空特征来强化现有检测特征,从而消除特征差异。MPTC (multi-pedestrian tracking with clusters) (Stadler和Beyerer, 2021)是一种基于检测和数据关联相结合的方法,用FPN (feature pyramid network) (Lin等, 2017)和ResNet-101 (He等, 2016)在MOT数据集上训练Faster R-CNN,MPNS (message passing networks) (Brasó和Leal-Taixé, 2020)是以检测为网络节点,构建一个网络,依赖一个CNN网络,提取目标特征;Aplift (Hornakova等, 2021)则使用了行人重识别的方法。最后将NF_SFM算法的跟踪结果与上述多目标跟踪方法进行对比,NF_SFM算法的MOTA比其他的方法都高,尽管MOTA比MPTC (Stadler和Beyerer, 2021)高出了0.3%,并且FN要比MPTC小得多,但是其他的指标还是略显不足,ML

在所有算法中达到了最优值。NF_SFM算法是基于检测的跟踪,由于MOT16和MOT17具有相同的视频序列,在MOT16数据集上选取了与MOT17数据集不同的检测方法,检测结果中目标包含了视频中的大多数目标,同时也包含了MOT16数据集Ground Truth之外的目标,将NF_SFM算法跟踪结果上传官网进行评价时会与官方给定的目标不一致,导致在MOT16数据集上FP出现较差的结果。

表4显示了NF_SFM算法与MOT17排行榜上表

现最好的方法的性能比较。DeepMOT(Xu等,2020)是一种端对端的多目标跟踪方法的训练框架,JBNOT(Henschel等,2019)分析了联合检测对多目标跟踪的适应性,引入了两种检测类型之间不同的亲和力。本文提出的方法在所选的7个评价指标中有5个排名第1,如表4所示,在跟踪任务中,NF_SFM算法优于其他方法。

图5展示了MOT15和MOT16数据集部分跟踪结果。

表 2 在 MOT15 数据集上与其他方法比较

Table 2 Comparison with other methods on the MOT15 dataset

方法	评价指标						
	MOTA/(% (↑))	IDF1/(% (↑))	MT(↑)	ML(↓)	FP(↓)	FN(↓)	IDs(↓)
Trectrd15 (Bergmann 等, 2019)	44.1	46.7	130	189	6 477	26 577	1 318
OACDASM (Li 等, 2022)	30.9	46.3	137	258	14 613	27 312	541
HSAN (Liu 等, 2021)	38.0	48.3	168	175	12 245	24 521	1 310
AMIR15 (Sadeghian 等, 2017)	37.6	46.0	114	193	7 933	29 397	1 026
GCNNMatch (Papakis 等, 2020)	46.7	43.2	157	203	6 643	25 311	820
NF_SFM(本文)	50.58	39.3	191	181	5 610	23 298	1 453

注:加粗字体表示各列最优结果。

表 3 在 MOT16 数据集上与其他方法比较

Table 3 Comparison with other methods on the MOT16 dataset

方法	评价指标						
	MOTA/(% (↑))	IDF1/(% (↑))	MT(↑)	ML(↓)	FP(↓)	FN(↓)	IDs(↓)
OACDASM (Li 等, 2022)	49.2	49.2	140/-	277/-	9 832	82 018	736
HSAN (Liu 等, 2021)	53.0	54.7	154/-	271/-	5 259	79 671	779
IQHMOT (He 等, 2022)	58.6	62.4	-/26.1	-/36.5	4 074	71 206	370
Aplift (Hornakova 等, 2021)	61.7	66.1	260/-	237/-	9 168	60 180	495
GCNNMatch (Papakis 等, 2020)	57.2	55.0	174/-	258/-	3 905	73 494	559
STURE (Wang 等, 2022)	44.6	39.9	-/11.7	-/47.69	3 543	-/-	801
MPTC (Stadler 和 Beyerer, 2021)	62.9	65.1	195/-	240/-	3 361	63 565	685
MPNS (Brasó 和 Leal-Taixé, 2020)	58.6	61.7	-/27.3	-/34.0	4 949	70 252	354
NF_SFM (本文)	63.2	40.5	235/31.0	161/21.2	8 760	55 721	2 693

注:加粗字体表示各列最优结果。“/”前后的数字表示同一个评价指标的两种表达方式,“/”后数值的单位为%,-表示论文未提供相应实验结果。

表 4 在 MOT17 数据集上与其他方法比较

Table 4 Comparison with other methods on the MOT17 dataset

方法	评价指标						
	MOTA(↑)	IDF1(↑)	MT(↑)	ML(↓)	FP(↓)	FN(↓)	IDs(↓)
Trackror++ (Bergmann 等, 2019)	53.5	53.3	-/19.5	-/36.6	12 201	248 047	2 072
OACDASM (Li 等, 2022)	51.7	50.5	516/-	823/-	28 484	241 734	2 456
HSAN (Liu 等, 2021)	53.7	54.5	528/-	804/-	20 393	238 002	2 639
DeepMOT (Xu 等, 2020)	53.7	53.8	-/19.4	-/36.6	11 731	247 444	1 947
JBNOT (Henschel 等, 2019)	52.6	50.8	465/-	844/-	31 572	232 659	3 050
NF_SFM (本文)	57.4	45.5	562/23.9	796/33.8	10 284	225 672	4 614

注:加粗字体表示各列最优结果。“/”前后的数字表示同一个评价指标的两种表达方式,“/”后数值的单位为%,-表示论文未提供相应实验结果。



图5 跟踪结果示意图(不同颜色的矩形框表示不同的目标)

Fig. 5 Schematic diagram of tracking results , with rectangular boxes of different colors representing different objects

4 结 论

本文将目标跟踪定义为一个最大后验概率问题,通过目标的颜色、外观和位置提取目标特征,计算目标相似性,并且引入社会力模型增强这种相似性,然后基于最小代价网络流算法求解最优解,获得目标轨迹。与现有的MOT方法通过最大目标后验来跟踪多个目标不同,基于目标的运动,本文充分运用了场景信息,增加了目标节点之间的相互作用力,提高了数据关联的能力。从消融实验可以看出,增加目标之间的作用力,可以有效地抑制目标的IDs。在公开的3个数据集MOT15、MOT16和MOT17上进行了实验,结果表明,NF_SFM算法的跟踪结果在MOTA、FP和FN上都优于其他对比的MOT方法,达到了很好的跟踪精度。NF_SFM算法是基于检测的多目标跟踪算法,目标检测的精度对跟踪的结果有很大的影响,在视频序列中,由于目标遮挡、环境阴暗等问题,导致目标检测阶段存在目标的漏检,当目标不能被检测器检测到时,目标也就不能被映射到网络中,因而当目标出现、消失和遮挡时都会赋予一个新的ID,所以引起较高的IDs。由IDF1的定义可知,该指标的前提是在ID相同的情况下跟踪的准确率与召回率的衡量指标,当目标出现、消失和遮挡时都会赋予一个新的ID,所以这些目标的ID在视频序列中很难保持不变,这样就引起较差的IDF1。未来可以考虑在进行目标检测的过程中就进行Re-ID特征的提取,增加目标检测的精度,从而降低IDs。

参考文献(References)

- Abdulghafoor N H and Abdullah H N. 2022. A novel real-time multiple objects detection and tracking framework for different challenges. *Alexandria Engineering Journal*, 61(12): 9637-9647 [DOI: 10.1016/j.aej.2022.02.068]
- Ata-ur-Rehman, Naqvi S M, Mihaylova L and Chambers J A. 2014. Multi-target tracking by using particle filtering and a social force model//*Proceedings of the 17th International Conference on Information Fusion (FUSION)*. Salamanca, Spain: IEEE: 1-6
- Berclaz J, Fleuret F, Türetken E and Fua P. 2011. Multiple object tracking using K-shortest paths optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9): 1806-1819 [DOI: 10.1109/TPAMI.2011.21]
- Bergmann P, Meinhardt T and Leal-Taixé L. 2019. Tracking without bells and whistles//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 941-951 [DOI: 10.1109/ICCV.2019.00103]
- Bewley A, Ge Z Y, Ott L, Ramos F and Upcroft B. 2016. Simple online and realtime tracking//*Proceedings of 2016 IEEE International Conference on Image Processing (ICIP)*. Phoenix, USA: IEEE: 3464-3468 [DOI: 10.1109/ICIP.2016.7533003]
- Bochinski E, Eiselein V and Sikora T. 2017. High-speed tracking-by-detection without using image information//*Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. Lecce, Italy: IEEE: 1-6 [DOI: 10.1109/AVSS.2017.8078516]
- Bombardelli F, Gül S and Hellge C. 2019. Compressed-domain video object tracking using markov random fields with graph cuts optimization//*Proceedings of the 40th German Conference on Pattern Recognition*. Stuttgart, Germany: Springer: 127-139 [DOI: 10.1007/978-3-030-12939-2_10]
- Boragule A and Jeon M. 2017. Joint cost minimization for multi-object tracking//*Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. Lecce, Italy: IEEE: 1-6 [DOI: 10.1109/AVSS.2017.8078481]
- Brasó G and Leal-Taixé L. 2020. Learning a neural solver for multiple object tracking//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 6246-6256 [DOI: 10.1109/CVPR42600.2020.00628]
- Ciaparrone G, Sánchez F L, Tabik S, Troiano L, Tagliaferri R and Herrera F. 2020. Deep learning in video multi-object tracking: a survey. *Neurocomputing*, 381: 61-88 [DOI: 10.1016/j.neucom.2019.11.023]
- Fang L and Yu F Q. 2020. Multi-object tracking based on adaptive online discriminative appearance learning and hierarchical association. *Journal of Image and Graphics*, 25(4): 708-720 (方岚, 于凤芹. 2020. 自适应在线判别外观学习的分层关联多目标跟踪. *中国图象图形学报*, 25(4): 708-720) [DOI: 10.11834/jig.190320]
- Felzenszwalb P, McAllester D and Ramanan D. 2008. A discriminatively trained, multiscale, deformable part model//*Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage, USA: IEEE: 1-8 [DOI: 10.1109/CVPR.2008.4587597]
- Feng P M, Wang W W, Dlay S, Naqvi S M and Chambers J. 2017. Social force model-based MCMC-OCSVM particle PHD filter for multiple human tracking. *IEEE Transactions on Multimedia*, 19(4): 725-739 [DOI: 10.1109/TMM.2016.2638206]
- Han G, Yu X Y and Liu L. 2017. Robust multi-object tracking based on higher-order graph and min-cost flow network//*Proceedings of the 4th International Conference on Systems and Informatics (ICSAI)*. Hangzhou, China: IEEE: 484-490 [DOI: 10.1109/ICSAI.2017.8248341]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning

- for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- He Y H, Wei X, Hong X P, Ke W and Gong Y H. 2022. Identity-quantity harmonic multi-object tracking. *IEEE Transactions on Image Processing*, 31: 2201-2215 [DOI: 10.1109/TIP.2022.3154286]
- Helbing D and Molnár P. 1995. Social force model for pedestrian dynamics. *Physical Review E*, 51(5): 4282-4286 [DOI: 10.1103/PhysRevE.51.4282]
- Henschel R, Zou Y Z and Rosenhahn B. 2019. Multiple people tracking using body and joint detections//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach, USA: IEEE: 770-779 [DOI: 10.1109/CVPRW.2019.00105]
- Hornakova A, Kaiser T, Swoboda P, Rolinek M, Rosenhahn B and Henschel R. 2021. Making higher order mot scalable: an efficient approximate solver for lifted disjoint paths//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 6310-6320 [DOI: 10.1109/ICCV48922.2021.00627]
- Hu Y J, Zhang X, Li Y X and Tian R. 2020. Online multiple object tracking using single object tracker and maximum weight clique graph//2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP). Tampere, Finland: IEEE: 1-6 [DOI: 10.1109/MMSP48831.2020.9287090]
- Krishanth K, Chen X, Tharmarasa R, Kirubarajan T and McDonald M. 2017. The social force PHD filter for tracking pedestrians. *IEEE Transactions on Aerospace and Electronic Systems*, 53(4): 2045-2059 [DOI: 10.1109/TAES.2017.2680718]
- Kuhn H W. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1/2): 83-97 [DOI: 10.1002/nav.3800020109]
- Law H and Deng J. 2020. CornerNet: detecting objects as paired keypoints. *International Journal of Computer Vision*, 128(3): 642-656 [DOI: 10.1007/s11263-019-01204-1]
- Leal-Taixé L, Pons-Moll G and Rosenhahn B. 2011. Everybody needs somebody: modeling social and grouping behavior on a linear programming multiple people tracker//Proceedings of 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). Barcelona, Spain: IEEE: 120-127 [DOI: 10.1109/ICCVW.2011.6130233]
- Li H L, Dong Y S and Li X L. 2022. Online association by continuous-discrete appearance similarity measurement for multi-object tracking. *Neurocomputing*, 487: 86-98 [DOI: 10.1016/j.neucom.2022.02.055]
- Li X, Wang K J, Wang W and Li Y. 2010. A multiple object tracking method using Kalman filter//Proceedings of 2010 IEEE International Conference on Information and Automation. Harbin, China: IEEE: 1862-1866 [DOI: 10.1109/ICINFA.2010.5512258]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Liu P X, Li X F, Feng H Y and Fu Z Z. 2017. Multi-object tracking by virtual nodes added min-cost network flow//Proceedings of 2017 IEEE International Conference on Image Processing (ICIP). Beijing, China: IEEE: 2577-2581 [DOI: 10.1109/ICIP.2017.8296748]
- Liu Y T, Li X S, Bai T X, Wang K F and Wang F Y. 2021. Multi-object tracking with hard-soft attention network and group-based cost minimization. *Neurocomputing*, 447: 80-91 [DOI: 10.1016/j.neucom.2021.02.084]
- Mazzon R and Cavallaro A. 2013. Multi-camera tracking using a multi-goal social force model. *Neurocomputing*, 100: 41-50 [DOI: 10.1016/j.neucom.2011.09.038]
- Mehran R, Oyama A and Shah M. 2009. Abnormal crowd behavior detection using social force model//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 935-942 [DOI: 10.1109/CVPR.2009.5206641]
- Milan A, Roth S and Schindler K. 2014. Continuous energy minimization for multitarget tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1): 58-72 [DOI: 10.1109/TPAMI.2013.103]
- Milan A, Schindler K and Roth S. 2016. Multi-target tracking by discrete-continuous energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10): 2054-2068 [DOI: 10.1109/TPAMI.2015.2505309]
- Okuma K, Taleghani A, De Freitas N, Little J J and Lowe D G. 2004. A boosted particle filter: multitarget detection and tracking//Proceedings of the 8th European Conference on Computer Vision. Prague, Czech Republic: Springer: 28-39 [DOI: 10.1007/978-3-540-24670-1_3]
- Papakis I, Sarkar A and Karpatne A. 2020. GCNNMatch: graph convolutional neural networks for multi-object tracking via sinkhorn normalization [EB/OL]. [2023-06-05]. <https://arxiv.org/pdf/2010.00067.pdf>
- Pirsiavash H, Ramanan D and Fowlkes C C. 2011. Globally-optimal greedy algorithms for tracking a variable number of objects//Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, USA: IEEE: 1201-1208 [DOI: 10.1109/CVPR.2011.5995604]
- Ren H L, Xu F, Zou F S, Jia K, Di P and Kang J. 2018. Multi-pedestrian tracking based on social forces//Proceedings of 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR). Shenyang, China: IEEE: 527-532 [DOI: 10.1109/IISR.2018.8535956]

- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Ren W H, Wang X C, Tian J D, Tang Y D and Chan A B. 2021. Tracking-by-counting: using network flows on crowd density maps for tracking multiple targets. *IEEE Transactions on Image Processing*, 30: 1439-1452 [DOI: 10.1109/TIP.2020.3044219]
- Sadeghian A, Alahi A and Savarese S. 2017. Tracking the untrackable: learning to track multiple cues with long-term dependencies//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 300-311 [DOI: 10.1109/ICCV.2017.41]
- Stadler D and Beyerer J. 2021. Multi-pedestrian tracking with clusters//*Proceedings of the 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. Washington, USA: IEEE: 1-10 [DOI: 10.1109/AVSS52988.2021.9663829]
- Sun Z H, Chen J, Chao L, Ruan W J and Mukherjee M. 2021. A survey of multiple pedestrian tracking based on tracking-by-detection framework. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(5): 1819-1833 [DOI: 10.1109/TCSVT.2020.3009717]
- Sun Z H, Chen J, Mukherjee M, Liang C, Ruan W J and Pan Z G. 2022. Online multiple object tracking based on fusing global and partial features. *Neurocomputing*, 470: 190-203 [DOI: 10.1016/j.neucom.2021.10.107]
- Wang H D, Li Z Y, Li Y P, Nai K and Wen M. 2022. STURE: Spatial-temporal mutual representation learning for robust data association in online multi-object tracking. *Computer Vision and Image Understanding*, 220: #103433 [DOI: 10.1016/j.cviu.2022.103433]
- Wojke N, Bewley A and Paulus D. 2017. Simple online and realtime tracking with a deep association metric//*Proceedings of 2017 IEEE International Conference on Image Processing*. Beijing, China: IEEE: 3645-3649 [DOI: 10.1109/ICIP.2017.8296962]
- Xu Y H, Šep A, Ban Y T, Horaud R, Leal-Taixé L and Alameda-Pineda, X. 2020. How to train your deep multi-object tracker//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 6786-6795 [DOI: 10.1109/CVPR42600.2020.00682]
- Yang F, Choi W and Lin Y Q. 2016. Exploit all the layers: fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 2129-2137 [DOI: 10.1109/CVPR.2016.234]
- Yue Y Y, Xu D, He K J and Zhang H. 2023. An adaptive occlusion-aware multiple targets tracking algorithm for low viewpoint. *Journal of Image and Graphics*, 28(2): 441-457 (乐应英, 徐丹, 贺康建, 张浩. 2023. 低视点下遮挡自适应感知的多目标跟踪算法. *中国图象图形学报*, 28(2): 441-457) [DOI: 10.11834/jig.210853]
- Zhang L, Li Y and Nevatia R. 2008. Global data association for multi-object tracking using network flows//*Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage, USA: IEEE: 1-8 [DOI: 10.1109/CVPR.2008.4587584]
- Zhang Y F, Sun P Z, Jiang Y, Yu D D, Weng F C, Yuan Z H, Luo P, Liu W Y and Wang X G. 2022. ByteTrack: multi-object tracking by associating every detection box//*Proceedings of the 17th European Conference on Computer Vision*. Tel Aviv, Israel: Springer: 1-21 [DOI: 10.1007/978-3-031-20047-2_1]

作者简介

王凯,男,硕士研究生,主要研究方向为多目标跟踪。

E-mail: 977247348@qq.com

戴芳,通信作者,女,教授,主要研究方向为图像处理与计算机视觉。E-mail: daifang@xaut.edu.cn

郭文艳,女,副教授,主要研究方向为智能优化和图像处理。

E-mail: wyguo@xaut.edu.cn

王军锋,男,副教授,主要研究方向为图形图像处理与分析。

E-mail: jfwang2000@126.com

王小侠,女,副教授,主要研究方向为图像处理和流体问题数值模拟。E-mail: wangxiaoxia@xaut.edu.cn