

基于深度学习的多目标跟踪算法研究

Multiple Object Tracking Algorithm Based on Deep Learning

陆平/LU Ping
邓硕/DENG Shuo
李伟华/LI Weihua

(中兴通讯股份有限公司, 广东 深圳 518052)
(ZTE Corporation, Shenzhen 518052, China)

自从深度学习技术出现以来, 计算机视觉领域得到了快速发展, 深度学习技术最先用于图像分类问题。近年来, 基于深度学习的多目标跟踪算法也取得了一定的突破。多目标跟踪是计算机视觉领域一个非常具有挑战性的研究方向, 且有着十分广泛的现实应用场景, 例如: 智能视频监控控制、异常行为分析、移动机器人研究等。传统的多目标跟踪算法往往由于目标检测效果较差, 导致跟踪效果不佳; 而基于深度学习的检测器可以获得较好的目标检测效果, 进而提高目标跟踪的准确度。

因此, 文章中我们着重研究了基于深度学习的多目标跟踪算法。首先通过 GoogLeNet^[1]+长短期记忆网络(LSTM)^[2]模型进行目标检测, 以获得准确的目标检测结果。在此基础上, 提出了直接根据目标检测的特征图对检测目标进行深度特征的提取的方法, 深度特征相比于尺度不变特征变换(SIFT)^[3]等传统特征可以更准确地反应检测目标的外观特征, 因此可以提高目标跟踪算法的准确性。此外, 还在基于马尔科夫蒙特卡罗

收稿日期: 2017-05-20
网络出版日期: 2017-07-06

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 04-0014-006

摘要: 提出了一种基于深度学习的多目标跟踪算法。首先, 通过 GoogLeNet+长短期记忆网络(LSTM)模型进行目标检测, 以获得准确的目标检测结果; 其次, 直接根据目标检测的特征图对检测目标进行深度特征的提取, 深度特征相比于传统特征可以更准确地反映检测目标的外观特征, 因此可以有效提高跟踪的准确性。此外, 还在传统数据驱动马尔科夫蒙特卡罗(DDMCMC)算法的基础上, 提出了层次的数据驱动马尔科夫蒙特卡罗(HDDMCMC)算法, 可以进一步提高多目标跟踪的准确性。实验结果证明了所提出算法的有效性。

关键词: 多目标跟踪; 深度学习; 目标检测; MCMC算法

Abstract: In this paper, a multi-target tracking algorithm based on deep learning is proposed. Firstly, GoogLeNet + long short-term memory (LSTM) model is used to obtain accurate object detection results. Secondly, the feature map of object detection is directly used to extract the deep feature for tracking. Compared with the traditional feature, the deep feature can reflect the appearance of objects more accurately, which could improve the tracking accuracy effectively. What's more, based on the traditional Data Driven Markov Chain Monte Carlo (DDMCMC) algorithm, the Hierarchical Data Driven Markov Chain Monte Carlo (HDDMCMC) algorithm is proposed to further improve the tracking accuracy. The experiment results prove the effectiveness of our algorithm.

Keywords: multiple object tracking; deep learning; object detection; MCMC algorithm

(MCMC)算法的多目标跟踪算法的基础上, 提出了层次的数据驱动马尔科夫蒙特卡罗(HDDMCMC)算法。

1 基于深度学习的多目标跟踪算法的研究现状

1.1 传统多目标跟踪算法

MCMC算法是一种经典的多目标跟踪算法, Yu等人在MCMC算法的基础上提出了数据驱动的马尔科夫蒙特卡罗(DDMCMC)算法^[4]。在使用该算法获得检测数据后, 我们按照传统的MCMC算法, 对当前轨迹中的检测数据进行位置和长相特征的衡量, 来

计算当前的后验概率大小, 然后在迭代过程中不断进行不同状态之间的转移来进行寻找全局最优结果。

Tang等人提出了一种基于图分割的多目标跟踪算法^[5], 通过在时间和空间上对边界框进行聚类来进行轨迹匹配。Tang等人在用该算法解决轨迹匹配这一最优化问题时, 提出了一种基于KL(Kernighan-Lin)算法的近似解法, 运用该算法求得的结果与剪枝法求得的结果相比, 准确率略微下降, 但是运算速度有较大提高。

1.2 基于深度学习的多目标跟踪算法

Bing等人提出了一种联合学习

卷积神经网络(CNN)特征和时域约束度量的模型,通过这个训练模型结合相应的损失函数可以构建出轨迹亲和力模型,再通过传统的图匹配方法将所有的轨迹进行联合,利用 softassign 算法求出轨迹匹配最优解,得到最终结果。

Fengwei 等人提出了一种基于深度学习的目标检测外观特征,将这种特征应用到多目标跟踪后,获得了较高的跟踪准确度,但需要大量的训练数据和时间,且无法达到实时跟踪的效果。

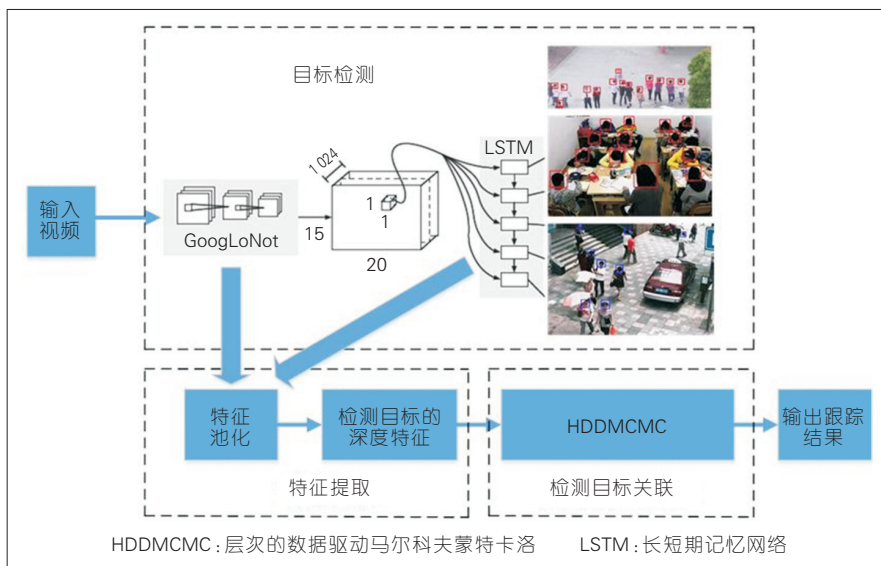
基于深度学习的多目标跟踪算法的跟踪准确度相对于传统多目标跟踪算法有了较大提升,但仍需要大量的训练数据和时间,且实时性不足。针对这一问题,我们在基于深度学习的目标检测算法的基础上,提出一种可直接利用目标检测的特征图得到深度特征的多目标跟踪算法,并基于 HDDMCMC 算法得到跟踪结果。

2 基于目标检测特征图的多目标跟踪算法框架

我们提出的算法框架如图 1 所示。算法整体上可以分为目标检测、特征提取、检测目标关联这 3 个模块。对于输入视频,首先通过 GoogLeNet + LSTM 模型进行目标检测,根据检测模块中的特征图,对检测目标提取深度特征,利用深度特征可以计算外观相似度,进而通过 HDDMCMC 算法对检测目标进行匹配,形成完整的轨迹,得到跟踪结果。基于目标检测特征图的特征提取模块和基于 HDDMCMC 算法的检测目标关联模块是该算法的主要创新点。

3 基于 GoogLeNet+LSTM 的目标检测

目标检测是基于数据联合的多目标跟踪算法的基础,针对在复杂场景下目标过小和目标遮挡等问题,我们利用 GoogLeNet + LSTM 这一框架进行目标检测。对于输入视频,我们首



▲ 图 1 基于深度学习的多目标跟踪算法框架

先使用 GoogLeNet 进行卷积,在最后一层得到 $1 \times 1 \times 024 \times 15 \times 20$ 的特征图阵列,把它进行转置成为 $300 \times 1 \times 024$ 的特征图阵列。每个 1 024 维的向量对应原图中 139×139 的区域。

在利用 GoogleNet 进行卷积以后可以得到 $300 \times 1 \times 024$ 的特征图阵列,然后通过 LSTM 子模块并行处理每一个 1 024 维向量。对于每一个输出的隐状态,经过两个不同的全连接层:一个直接输出框的位置和宽高,一个再经过 softmax 层输出这个框的置信度。LSTM 子模块共有 5 个这样的 LSTM 单元,即对于每个输入预测 5 个可能的框和对应的置信度,在训练中使得框的位置集中在感知区域中心的 64×64 的位置,置信度是从高到低排的。

每个 1 024 维的向量经过 LSTM 阵列处理以后,可以得到对应原图中 64×64 小块的 5 个检测框以及对应的置信度。因为 LSTM 是并行处理的,所以生成的对 64×64 区域的检测结果是重叠的。最后框处理子模块需要对整个视频帧的所有检测框进行筛选,然后再通过给定一个阈值去掉置信度低的框,并给出最终检测结果。具体流程是:对于已经确定的框,如果一个待选的框与它有相交,

则去除这个框,限制一个已经确定的框至多去除一个待选框。在上述匹配的时候代价用 (m, d) 衡量, m 表示评估的二者是否相交,取值 $\{0,1\}$, d 衡量评估的两个框之间曼哈顿距离。 m 的重要性大于 d ,即对两种匹配方案得到的结果,先比较 m 大小,如果不能得出结论,再比较 d 的大小。我们使用匈牙利算法来寻找代价最小的匹配方式,假设筛选的置信度阈值为 0.5,那么则去掉那些置信度低于 0.5 的框。

为有效对目标检测模型进行训练,我们采用了如下训练方法:首先在 LSTM 子模块得到较多待选框,但存在检测错误或误差。错误或误差的情况有 3 种:

- (1) 把不是人头尖的地方框出;
- (2) 预测框位置和真值框位置的差异;
- (3) 对同一个目标产生了多个预测框。

对于情况 (1),可通过赋予待选框较低的置信度来杜绝;对于情况 (2),需要修正待选框和与之匹配的真值框之间的误差;对于情况 (3),可通过给同一个目标后生成的预测框赋予较低的置信度来杜绝。

模型训练时的损失函数如公式

(1)所示:

$$L(G, C, f) = \alpha \sum_{i=1}^{|C|} l_{pos}(\hat{b}_{pos}^i, \hat{b}_{pos}^{f(i)}) + \sum_{j=1}^C l_c(\hat{b}_{pos}^j, \{f^{-1}(j)\}) \quad (1)$$

其中, G 代表框的真值, C 代表待选框, f 代表匹配算法, b_{pos}^i 代表真值框里面的第 i 个框, \hat{b}_{pos}^j 代表待选框里面第 j 个框, l_{pos} 代表两者之间的曼哈顿距离, l_c 是交叉熵损失, 也就是对应网络里面的 softmax 损失。这个损失函数的前一项代表待选框和与之匹配的真值框的位置误差, 后一项代表待选框的置信度, α 调整这两种损失之间的平衡。

匹配算法为匈牙利算法, 其中用的比较函数如公式(2)所示:

$$\Delta(b_i, \hat{b}_j) = (o_{ij}, r_j, d_{ij}) \quad (2)$$

其中, o_{ij} 取值为 $\{0, 1\}$, 若待选框的中心落在真值框中, 则为 0, 否则为 1; r_j 为该待选框生成的序号, 目的是在匹配的时候, 偏好先生成置信度比较高的框, 故匹配同一个目标时, 排序越靠前代价越低; d_{ij} 是两个框之间的距离, 即距离误差。

4 基于目标检测特征图的深度特征提取

在检测模块中得到对应框的位置时, 得到的是人头的位置, 按一定比例放大可以覆盖全身。得到行人的框位置和大小后, 我们进一步利用 GoogLeNet 的最后一层卷积层得到的特征图阵列去提取特征。我们使用的方法是快速-区域 CNN (RCNN)^[6] 中提到的感兴趣区域 (ROI) 池化, 即根据特征图阵列相对于输入图片的缩小比例, 把原图中的感兴趣区域在特征图阵列中对应的感兴趣区域池化成一个 1 024 维的向量。其中缩小比例为 32 倍, 即得到框的位置按 32 倍比例缩小, 并且为了减少背景影响, 最大化池化这一池化类型更加有效。通过将检测模块中的特征图进行池化, 可得到对检测模块中检测到

的每个目标的深度特征, 因为每个特征都是高度抽象的, 可以很好地表征目标的外观特征。

文中, 我们所提出算法的重要特点就是利用目标检测中的特征图进行池化来得到目标跟踪所需要的深度特征, 而不需要重新进行训练, 因此可以在不牺牲目标跟踪算法的实时性的前提下, 提高跟踪准确性。

5 基于 HDDMCMC 算法的目标检测关联

5.1 传统 MCMC 算法

为了对检测数据进行最优关联, 可利用 MCMC 算法进行建模。传统 MCMC 算法的计算过程为: 在迭代过程中, 均匀随机选取一个转移动作 (包括产生、消失、融合、分裂、扩展、收缩和交换等), 把当前的行人轨迹按照此转移动作进行转换, 即从状态 w 转移到状态 w' 。在此情况下, 可以计算得到转移前后的后验概率 $\pi(w)$ 和 $\pi(w')$, 以及转移概率 $q(w'|w, D)$ 和 $q(w|w', D)$, 从而可以计算得出此时的接收概率 $A(w', w)$ 。将此接收概率与从均匀 $0 \sim 1$ 分布中随机抽取的值 u 进行比较, 如果出现 $A(w', w) > u$, 则接收当前这个转移, 使得当前状态转移为 w' , 即 $w = w'$, 否则不接受此转移。为了得到最大后验概率, 我们将转移后的状态与当前的最优状态 w^* 进行比较, 如果转移后的状态优于当前的最优状态, 则对最优状态进行更新, 即 $w^* = w$ 。

在传统 MCMC 算法的基础上, 我们提出了 HDDMCMC 算法, 即把传统的单层 MCMC 算法分为段内 MCMC 算法和段间 MCMC 算法。如图 2 所示, 假设当前处理的时间段是 $[t+2T, t+3T-1]$, 我们首先对此段执行段内 MCMC 算法并得到相应的段内行人轨迹。在此之前, $[t, t+T-1]$ 段和 $[t+T, t+2T-1]$ 段内的 MCMC 算法已执行完成, 并执行了两者之间的段间 MCMC 算法, 从而得到对应时间

帧内的行人轨迹; 然后把其中已走出当前区域的行人轨迹选择出来, 再对 $[t, t+2T-1]$ 段和 $[t+2T, t+3T-1]$ 段再执行段间 MCMC 算法, 得到当前的行人轨迹。HDDMCMC 算法即按照此种方式不断执行, 直到视频序列结束。

5.2 段内 MCMC 算法

在多目标跟踪算法中, 考虑到运动的稳定性和连续性 (即同一个目标在前后帧视频数据中, 外观特征不会发生剧烈变化), 在段内 MCMC 算法中, 使用第 4 节中的深度特征对目标轨迹的相似度进行度量。

可以将每个检测目标看成一个节点, 以段内时间 $[t, t+T]$ 来进行叙述。假设视频帧 t 内的节点集合为 $N_t = \{N^t(1), N^t(2), N^t(3), \dots, N^t(i), \dots, N^t(N_t-1), N^t(N_t)\}$, 后验概率 $P(w|D)$ 的设定如公式(3)所示:

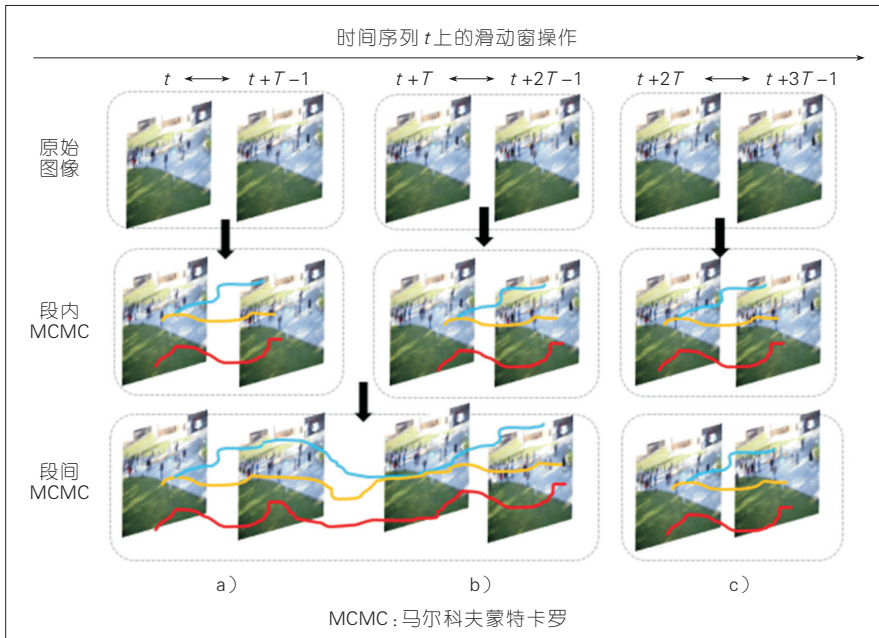
$$P(w|D) = P(\{\tau_1, \tau_2, \tau_3, \dots, \tau_k\} | D) = \left(\prod_{k=1}^K \prod_{n=1}^{N_k-1} P(N^{n+1}(k) | N^n(k)) \right) \times \left(\prod_{k=1}^K e^{-\frac{1}{l_k}} \right) \times (e^{-|\tau_0|}) \quad (3)$$

其中, $N^n(k)$ 和 $N^{n+1}(k)$ 分别代表第 k 条行人轨迹中第 $n+1$ 个和第 n 个节点, $P(N^{n+1}(k) | N^n(k))$ 即为两节点的相似度, 可用两节点深度特征的夹角余弦值来进行计算。公式第 2 项中的 l_k 代表的是不同轨迹的长度, 保证行人轨迹的完整; 公式第 3 项中的 $|\tau_0|$ 代表的是虚警的个数, 保证虚警率较低。

图 3 为 MCMC 算法中 7 种操作: 产生和消失这两个转移动作如图 3a) 所示, 其中, 左边黑色的点代表虚警集合中的点, 经过产生操作后转化为右边红色点所代表的行人轨迹, 消失操作为反向顺序; 融合和分裂这两个转移动作如图 3b) 所示; 扩展和收缩这两个转移动作如图 3c) 所示, 交换和转移动作如图 3d) 所示, 此操作前后互为逆操作。

5.3 段间 MCMC 算法

段间 MCMC 算法使用的数据主



▲图2 HDDMCMC 算法

要是段内 MCMC 算法生成的目标轨迹。在段间 MCMC 算法中,主要采取的转移动作包括融合和分裂操作以及交换操作。因为,在经过段内 MCMC 算法之后,生成许多较为可靠的目标轨迹。此时若存在同一目标轨迹断裂的情况,就是由于检测数据不稳定,对应目标碰撞或者遮挡导致对应的漏检帧数过多等原因造成的。因此,段间 MCMC 的目的是将两个时间段的目标轨迹数据做进一步的数据联合。当前状态下,后验概率项更新为公式(4)所示。

$$P(w|D) = P(\{\tau_1, \tau_2, \tau_3, \dots, \tau_K\} | D) = \prod_{k=1}^K \prod_{n=1}^{N_k-1} P(N^{n+1}(k) | N^n(k)) \times \prod_{k=1}^K e^{-\frac{1}{l_k}} \quad (4)$$

在公式(4)中,我们不再考虑虚警因素,因为此处主要是对目标轨迹的划分操作。由于已有之前较为可靠的两个时间段内的目标轨迹数据作为辅助,因此可以在更加宽松的条件下进行抽样工作。对于融合操作,此时允许的时间间隔设定为 $t_{gap} = 10$,且两个目标的轨迹段与段之间连接处的帧差不能超过6。对于概率中的标准差的设定,此时为

$3\sigma = size(d_i^t)$,即允许在理想位置周围的变化幅度是一个当前目标的大小。对于分裂操作和交换操作,抽样的时间节点在段与段连接处的时间

帧为左右各6帧时间以内。通过这种方式,可使得不同状态之间进行转移的单位是之前已生成的较为完整的目标轨迹片段。

在当前的段间 MCMC 算法结合之后,把已走出视频场景的目标移除当前数据集,此时的当前数据集假设为 τ' 。当下一个段内 MCMC 得到轨迹之后,通过段间 MCMC 算法与 τ' 进行匹配,即继续在之前目标数据的基础之上,结合当前的目标数据,做进一步的数据联合来优化。整个算法按照这样的滑动方式不断进行。

6 实验结果及分析

6.1 MOT2015 训练数据库的实验结果

本小节中,我们介绍的是 MOT2015 训练数据库^[8]中的实验结果,此数据库总共包括11个视频,有固定视角、运动视角等场景,同时又有俯视视角、平时视角等场景,由于场景变化多样,不同目标之间的接触

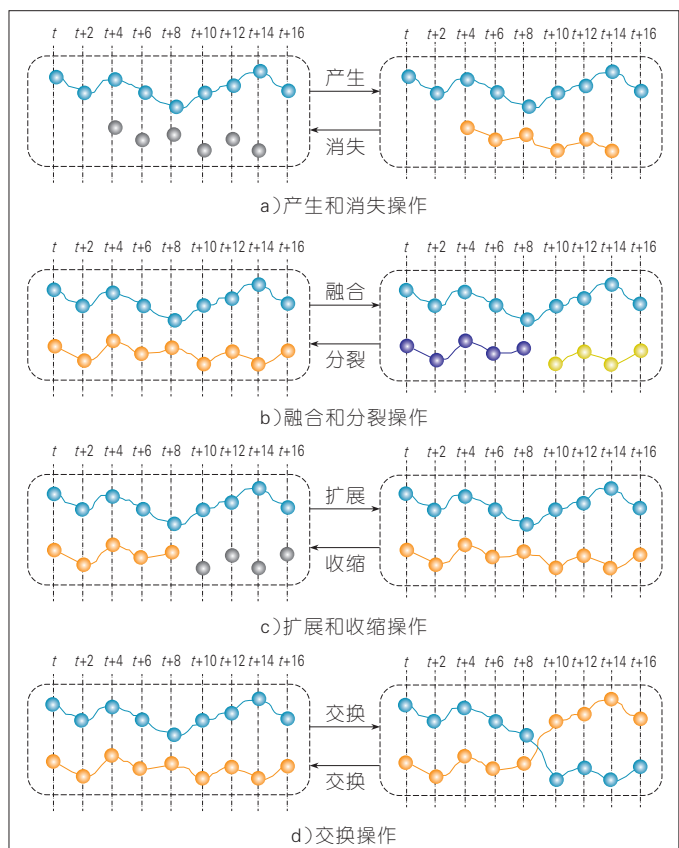


图3 ▶ 7种转移操作示意

和碰撞较多,挑战性较大。

我们使用的衡量指标^[9]包括多目标跟踪的准确度(MOTA)、多目标跟踪的精确度(MOTP)、漏检数目(FN)、虚警数目(FP)等,其中MOTA指标是用来衡量整体跟踪准确度,最具代表性。实验结果如表1所示,其中MOTA和MOTP是11个数据库的平均值,FN、FP和ID Sw是11个数据库的总和。

文章中,我们所提出算法的主要创新点为基于目标检测特征图的特征提取模块和基于层次的数据驱动马尔科夫蒙特卡洛算法的检测目标关联模块,为了验证这两个创新点的有效性,我们分别用4种方法进行了实验:

(1) SIFT+MCMC,即用SIFT特征结合传统MCMC算法。

(2) SIFT+HDDMCMC,即用传统SIFT特征结合文中所提出的HDDMCMC算法。

(3) CNN+MCMC,即用文中提出的通过目标检测特征图提取的深度特征结合传统MCMC算法。

(4) CNN+HDDMCMC,即用通过目标检测特征图提取的深度特征结合HDDMCMC算法,也即文中我们提出算法的完整版。

实验结果如表1所示,对比SIFT+MCMC和SIFT+HDDMCMC可知:我们提出的基于层次的HDDMCMC算法相比于传统的MCMC算法可以在一定程度上提升跟踪效果;对比SIFT+MCMC和CNN+MCMC可知:通过用目标检测特征图提取的深度特征来代替传统的SIFT特征,可以在很大程度上提升跟踪效果;CNN+HDDMCMC算法,即文中我们提出算法的完整版的各项检测指标都优于其他方法,更进一步证明了所提出算法的有效性。

图4和图5是两张跟踪效果的对比图(为便于观察,只显示了要说明的区域)。图4中第1行是使用SIFT+HDDMCMC的实验结果,第2行是使用CNN+HDDMCMC的实验结果,每1

行的3张图片代表视频中的连续3帧(前后2帧之间间隔1帧)。第1行的第3张图片在红色箭头标识处出现了跟踪错误,这是由于目标之间的遮挡且用SIFT特征求目标之间的外观相似度时并不十分准确而产生的;第2行图片中,由于深度特征能更准确地反应检测目标的外观特征,没有出现错误。

图5中第1行是使用CNN+MCMC的实验结果,第2行是使用CNN+HDDMCMC的实验结果。每1行的第2张图片由于目标之间的遮挡发生了漏检,导致了轨迹断裂,使得第1行中的第3张图片在红色箭头标识处出现了错误;而在第2行图片中,由于此处使用了HDDMCMC算法,通过段间MCMC算法对两段轨迹进行了匹配,从而避免了错误。

基于此数据库的对比实验可知:文中所提出的算法对于生成稳定的目标轨迹具有十分重要的作用。

6.2 MOT2015 测试数据库实验结果及其对比

本小节中,我们介绍的是MOT2015测试数据库中的实验结果,此数据库同样包括11个视频,我们将文中所提出算法的实验结果与其他算法进行了对比,比较方法包括如下3种:

(1) 结构化支持向量机算法(LP_SVM)^[10],此算法是Shaofei等人提出的,主要思想是使用最小网络流模式,此模型的参数通过结构化支持向量机训练得到。

(2) 近时多目标跟踪算法(NOMT)^[11],此算法是由Wongun等人

▼表1 MOT2015 训练数据库实验结果

	MOTA	MOTP	FN	FP
SIFT+MCMC	36.2	73.6	27 872	14 967
SIFT+HDDMCMC	39.8	73.6	25 688	13 576
CNN+MCMC	45.3	73.7	20 854	11 378
CNN+HDDMCMC	47.6	73.8	19 638	8 895

CNN:卷积神经网络
FN:漏检数目
FP:虚警数目

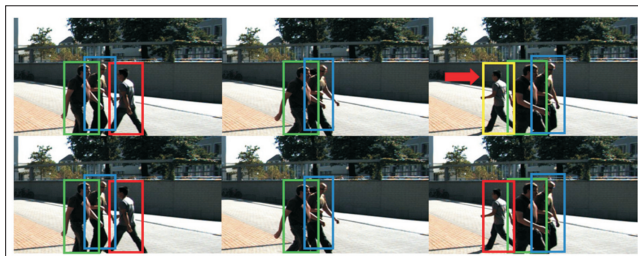
HDDMCMC:层次的数据驱动马尔科夫蒙特卡洛
MCMC:马尔科夫蒙特卡洛

MOTA:多目标跟踪的准确度
MOTP:多目标跟踪的精确度
SIFT:尺度不变特征变换

图4▶
是否使用深度特征的跟踪效果对比



图5▶
是否使用HDDMCMC的跟踪效果对比



提出的,主要思想是通过建立一个聚合本地光流描述子来计算检测目标之间的相似度,实现了接近实时的多目标跟踪。

(3) 子卷积神经网络算法(MDP_SubCNN)^[12],此算法是由Yu等人提出的,主要思想是通过学习的策略来建立检测目标之间的相似度矩阵,并通过马尔科夫决策过程建模。

实验结果如表2所示,其中Our表示的是文中我们提出的算法,与其他法对比可知:本算法的MOTA指标最高,整体效果最好;FN值优于其他算法;FP值相对于其他算法较低;MOTP值只略逊于MDP_SubCNN算法,总的来说,我们提出算法的效果较为理想。

7 结束语

多目标跟踪作为计算机视觉领域中的一个经典问题,有着十分广泛的应用前景。近年来,随着深度学习技术的发展,基于深度学习的多目标跟踪算法取得了一定的突破,获得了高于传统多目标跟踪算法的跟踪准确度。在文章中,我们提出了一种新的基于深度学习的多目标跟踪算法,通过用GoogLeNet+LSTM框架进行目标检测的特征图对检测目标深度特征的提取,以及对传统MCMC算法的改进,有效地提高目标跟踪算法的准确性和实时性。最后对文中所提出的算法进行了实验验证,并与一些相关的多目标跟踪算法进行了比较,得到了较为理想的结果。

致谢

本研究得到了上海交通大学彭瑾龙、林巍晓老师的指导与帮助,对他们谨致谢意!

参考文献

- [1] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA:IEEE, 2015: 1-9. DOI: 10.1109/CVPR.2015.7298594

▼表2 MOT2015 测试数据库实验结果

	MOTA	MOTP	FN	FP
LP_S SVM	25.2	71.7	36 932	8 369
NOMT	33.7	71.9	32 547	7 762
MDP_SubCNN	47.5	74.2	22 969	8 631
Our	48.2	74.0	21 318	9 385

CNN: 卷积神经网络 FN: 漏检数目 FP: 虚警数目	LP_S SVM: 结构化支持向量机算法 MCMC: 马尔科夫蒙特卡罗	MOTA: 多目标跟踪的准确度 MOTP: 多目标跟踪的精确度 SIFT: 尺度不变特征变换
HDDMCMC: 层次的数据驱动马尔科夫蒙特卡罗	MDP_SubCNN: 子卷积神经网络算法	

- [2] HOCHREITER S, SCHMIDUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780
- [3] LOWER D G. Distinctive Image Features from Scale-Invariant Key Points[J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [4] LOWE G D. Object Recognition from Local Scale-Invariant Features[C]// ICCV '99 Proceedings of the International Conference on Computer Vision. USA:IEEE, 1999
- [5] OH S, RUSSELL S, SASTRY S. Markov Chain Monte Carlo Data Association for General Multiple-Target Tracking Problems[C]// Decision and Control, 2004. CDC. 43rd IEEE Conference on. USA:IEEE, 2004
- [6] GIRSHICK R. Fast R-CNN[EB/OL].(2015-09-27)[2017-06-25]. <https://arxiv.org/abs/1504.08083>
- [7] YU Q, MEDIONI G, COHEN I. Multiple Target Tracking Using Spatio-Temporal Markov Chain Monte Carlo data association [C]// Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007. USA:IEEE, 2007. DOI: 10.1109/CVPR.2007.382991
- [8] Multiple Object Tracking Benchmark[EB/OL]. [2017-06-25]. <https://motchallenge.net/>
- [9] BERNARDIN K, STIEFELHAGEN R. Evaluating Multiple Object Tracking Performance: the CLEAR MOT Metrics[J]. EURASIP Journal on Image and Video Processing, 2008, 2008(1): 1-10
- [10] WANG S F, FOWLKES C C. Learning Optimal Parameters for Multi-target Tracking[EB/OL].(2016-10-05)[2017-06-25]. <https://arxiv.org/abs/1610.01394>
- [11] CHOI W. Near-Online Multi-Target Tracking with Aggregated Local Flow Descriptor[C]//Proceedings of the IEEE International Conference on Computer Vision. USA:IEEE, 2015: 3029-3037
- [12] 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. USA: IEEE, 2015: 4705-4713. DOI: 10.1109/ICCV.2015.534
- [13] MILAN A, LEAL T L, SCHINDLER K, et al. Joint Tracking and Segmentation of Multiple Targets[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA:IEEE, 2015. DOI: 10.1109/CVPR.2015.7299178
- [14] YOON J H, YANG M H, LIM JONGWOO, et al. Bayesian Multi-Object tracking Using Motion Context from Multiple Objects[C]// Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. USA: IEEE, 2015. DOI: 10.1109/WACV.2015.12
- [15] MCLAUGHLIN N, RINCON J M D, MILLER P. Enhancing Linear Programming with Motion Modeling for Multi-target Tracking [C]// Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. USA:IEEE, 2015
- [16] XIANG Y, ALAHI A, SAVARESE S, et al. Learning to Track: Online Multi-Object Tracking by Decision Making[C]// Computer Vision (ICCV), 2015 IEEE Conference on. USA:IEEE, 2015. DOI: 10.1109/ICCV.2015.534
- [17] BEWLEY A, GE Z, OTT L, et al. Simple Online and Realtime Tracking[C]//Image Processing (ICIP), 2016 IEEE International Conference on. USA: IEEE, 2016: 3464-3468. DOI: 10.1109/ICIP.2016.7533003

作者简介



陆平,中兴通讯股份有限公司云计算及政企业务产品总经理,北京邮电大学和南京邮电大学兼职教授;主要从事云计算、大数据、增强现实、基于多媒体服务技术等方面的研究;主持国家、省部级基金项目10余项;获得省部级科技进步奖多项;拥有20余项发明专利,发表论文15篇,出版专著2部。



邓硕,中兴通讯股份有限公司高级工程师;研究方向为图像处理、视频智能分析等技术,主要从事基于深度学习的图像智能分析技术预研与开发;参与多项中兴通讯产学研合作项目及国家、省部级基金项目。



李伟华,中兴通讯股份有限公司高级工程师;主要研究领域包括人工智能、图像及视频分析等方面;拥有20余项发明专利,承担标准Editor席位并输出多篇国际和国内标准。