



(12) 发明专利

(10) 授权公告号 CN 108875588 B

(45) 授权公告日 2022. 04. 15

(21) 申请号 201810512107.2

G06V 10/774 (2022.01)

(22) 申请日 2018.05.25

G06V 10/80 (2022.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 108875588 A

(56) 对比文件

CN 106097391 A, 2016.11.09

CN 107992819 A, 2018.05.04

CN 106709568 A, 2017.05.24

CN 102436662 A, 2012.05.02

US 2012000987 A1, 2012.01.05

US 2017116753 A1, 2017.04.27

CN 104318588 A, 2015.01.28

(43) 申请公布日 2018.11.23

(73) 专利权人 武汉大学

地址 430072 湖北省武汉市武昌区珞珈山  
武汉大学

张瑞林等. 基于改进YOLOv2网络的遗留物检测算法.《浙江理工大学学报(自然科学版)》.2017,第39卷(第03期),第325-332页.

卢湖川等. 目标跟踪算法综述.《模式识别与人工智能》.2018,第31卷(第1期),第61-76页.

辛鹏等. 全卷积网络多层特征融合的飞机快速检测.《光学学报》.2018,第38卷(第3期),第1-7页.

(72) 发明人 陈丽琼 田胜 邹炼 范赐恩

杨烨 胡雨涵

(74) 专利代理机构 武汉科皓知识产权代理事务  
所(特殊普通合伙) 42222

代理人 严彦

审查员 代冯楠

(51) Int. Cl.

G06V 40/20 (2022.01)

G06V 20/52 (2022.01)

G06K 9/62 (2022.01)

G06N 3/04 (2006.01)

权利要求书3页 说明书10页 附图3页

(54) 发明名称

基于深度学习的跨摄像头行人检测跟踪方法

人重识别有机地融合起来,实现了大范围场景下精确的跨摄像头行人检测和跟踪。

(57) 摘要

本发明公开了一种基于深度学习的跨摄像头行人检测跟踪方法,包括通过训练行人检测网络,对输入的监控视频序列进行行人检测;以行人检测得到的目标框初始化跟踪目标,提取行人检测网络中候选框所对应区域的浅层特征和深层特征,实现跟踪;有目标消失时行人重识别,包括得到目标消失信息之后,利用行人重识别算法,从行人检测网络得到的候选图像中找到与消失目标匹配度最高的图像,继续跟踪;跟踪结束时输出行人目标在多个摄像头下的运动轨迹。本发明提取的特征能克服光照变化和视角变化的影响,而且跟踪和行人重识别部分均从行人检测网络中提取特征,将行人检测、多目标跟踪和行人



1. 一种基于深度学习的跨摄像头行人检测跟踪方法,其特征在于,卷积神经网络的浅层卷积层包含适用于目标定位的高分辨率特征,深层卷积层包含易于区分的视觉特征,利用浅层特征与深层特征结合的方式提取行人在不同摄像头下不变的本质特征,克服光照变化和视角变化的影响,并且跟踪和行人重识别部分用到的特征也从检测网络中间层提取,将行人检测、关联跟踪和行人重识别有机地融合起来,并通过各模块间的信息交互,有效地降低了跟踪过程中跨摄像头的噪声干扰,提高了跨摄像头行人检测和跟踪的准确性和实时性,

实现方式包括设置三个模块,行人检测模块、多目标跟踪模块和行人重识别模块,输入为多个摄像头下的视频;

行人检测模块用于利用深度卷积神经网络,检测监控区域所有摄像头拍摄的视频序列中的行人目标,得到每个目标的置信度和边框坐标,利用改进的Faster RCNN网络进行行人检测,根据输入的不同摄像头下的视频序列,计算出行人目标在图像中的位置相应边框坐标信息以及置信度得分;

多目标跟踪模块用于将检测结果作为跟踪的初始化目标,提取检测网络的中间层特征对检测出来的行人目标进行特征建模,实现构建行人外观特征模型,利用相关滤波器预测计算目标的运动轨迹,实现多目标跟踪;

行人重识别模块用于解决跨摄像头情况下的目标消失问题,当跟踪过程中目标离开当前摄像头视野时,根据跟踪模块传来的消失目标信息,从行人检测模块检测出来的候选目标图像库中,通过求取全局特征和局部特征,找到相似度最高的匹配目标,返回给多目标跟踪模块继续跟踪;

以上模块结合实现的整个系统在主动结束跟踪或者所有行人目标离开监控区域摄像头的视野时,最后输出不同行人在多个摄像头下的运动轨迹以及运动信息;

系统执行包括以下步骤:

步骤S1,通过训练行人检测网络,对输入的监控视频序列进行行人检测;

步骤S2,以行人检测得到的目标框初始化跟踪目标,实现跟踪,包括提取行人检测网络中候选框所对应区域的浅层特征和深层特征,将浅层特征和深层特征结合起来作为物体的外观特征模型;

步骤S3,持续进行新目标行人检测,并判断是否有目标消失,如果有则进入步骤S4进行行人重识别;实现方式如下,

1) 由于新的目标会不时进入摄像头视野,采用非固定性间隔多帧的方式,每隔F帧调用一次检测模块进行行人检测,为新出现的目标分配新的ID;

2) 如果跟踪框的置信度下降或者离开视频区域,多目标跟踪模块会把该目标最后一帧的跟踪信息传给行人重识别模块,并且调用行人检测模块对其他摄像头下的当前帧进行检测,得到的所有目标检测框也传给行人重识别模块,作为行人重识别的图像库;

跨摄像头进行目标重识别实现方式为,当摄像头中的某个目标消失在它的视野时,多目标跟踪模块会调用行人检测模块和行人重识别模块,根据特征匹配的信息,在其他摄像头中找到该目标,保持ID不变继续跟踪;

步骤S4,行人重识别,包括得到目标消失信息之后,从行人检测网络得到的候选图像中找到与消失目标匹配度最高的图像,继续跟踪,跟踪结束时输出行人目标在多个摄像头下

的运动轨迹。

2. 根据权利要求1所述基于深度学习的跨摄像头行人检测跟踪方法,其特征在于:所述步骤S1包括以下子步骤,

步骤S11,构建行人检测网络的训练集;

步骤S12,训练行人检测网络;

步骤S13,对于待测试的视频序列,首先将每一帧分离出来,把图片送入行人检测网络中进行行人检测,得到所有行人目标的置信度score和边框信息,当置信度score大于设定阈值时,认为是行人目标,保留目标框;

步骤S14,基于步骤S13初步检测所得的准目标,利用非极大值抑制算法去除冗余框,得到最终的检测结果。

3. 根据权利要求1所述基于深度学习的跨摄像头行人检测跟踪方法,其特征在于:所述步骤S2包括以下子步骤,

步骤S21,根据行人检测得到的行人目标候选框,提取行人检测网络中候选框所对应区域的浅层特征和深层特征,将浅层特征和深层特征结合起来作为物体的外观特征模型;

步骤S22,使用相关滤波器计算跟踪目标在各个位置的置信度得分,得分最高的区域为物体的预测位置区域,从而得到各个目标的预测位置集合;

步骤S23,更新各个目标在当前帧的位置,再通过物体的外观特征模型进行比较,得出各个目标的外观置信度分数,根据置信度分数降低的程度判断当前跟踪的目标是否被遮挡,不断更新各个目标的跟踪状态,从而实现目标的自跟踪;

实现方式为,行人检测模块检测到各个摄像头新出现的目标为激活态,当目标检测框满足激活跟踪条件,即置信度高于设定阈值时,将该目标再转换为跟踪态;当目标检测框不满足激活跟踪条件,即目标检测框置信度太低或者和已经处于跟踪态的目标高度重叠,将该目标转为闲置态;

跟踪态的目标是当前正在各个摄像头图像序列中跟踪的目标,在非检测帧根据运动预测模型对每一个目标采取自跟踪策略;在检测帧用外观模型特征和运动预测模型特征匹配目标检测框,以此达到修正跟踪目标位置和目标尺度变换的目的;其中,非检测帧是指不需要检测算法提供检测框的视频帧;检测帧是指需要检测算法提供检测框作为初始跟踪框的视频帧;

丢失态为可能被遮挡的目标或者离开某一摄像头的目标,暂时未能在图像数据中显示出来的目标,目标处于丢失态时,根据行人重识别算法,检测到丢失目标后,保持该目标原始ID不变,状态恢复为跟踪态;若丢失目标长时间处于丢失态,未能通过目标重识别模块转为正常的跟踪态,则将该目标转为闲置态,不再对该目标进行任何后续跟踪计算,以减少计算耗时;

步骤S2由多目标跟踪模块实现,把检测到的行人结果传给跟踪模块,作为初始化跟踪目标,给每个目标一个不同的ID,并记录他们的轨迹。

4. 根据权利要求1所述基于深度学习的跨摄像头行人检测跟踪方法,其特征在于:所述步骤S4包括以下子步骤,

步骤S41,将目标跟踪所得最后一帧消失目标的图像,以及行人检测所得的多张候选图像作为输入,对于每一张图像首先通过行人检测网络的低层得到浅层特征图,再利用显著

度检测算法提取目标的显著度剔除背景中的冗余信息,然后送入深层卷积层,输出深层特征图;融合浅层特征图和深层特征图,对融合后特征图进行全局池化得到全局特征向量,进行水平池化得到局部特征向量;

步骤S42,根据步骤S41求出的全局特征向量,计算消失目标与候选目标全局特征向量之间的L2距离,然后通过目标对齐的方法计算出局部特征向量间的距离;

步骤S43,计算消失目标图像与候选对象图像的最终距离,最终距离等于全局特征向量距离加上局部特征向量距离,根据最终距离的大小得到候选对象图像与消失目标图像的初始相似度排序;

步骤S44,通过重排序算法得到最终的相似性排序结果,选择相似性最高的图像继续跟踪。

## 基于深度学习的跨摄像头行人检测跟踪方法

### 技术领域

[0001] 本发明属于计算机视觉、视频分析技术领域，具体地涉及一种基于深度学习的跨摄像头行人检测跟踪方法。

### 背景技术

[0002] 随着人们对公共安全问题的重视以及监控摄像头数量和覆盖程度的快速增加，智能化的多摄像头监控发挥着越来越重要的作用。行人作为监控的主体，不仅具有一般目标的通性，还具有类内多样性，这正是行人检测与跟踪的难点。跨摄像头行人检测与跟踪是指检测并跟踪多个摄像头下的行人目标，当有目标离开当前摄像头的视野时能迅速在邻近摄像头区域重新找回该目标，进行持续的全方位跟踪，最终得到有效的行人轨迹和运动信息。多摄像头协同跟踪可以弥补单摄像头系统视野范围的局限性，能够实现目标在更大场景下的长时间跟踪，可以广泛应用于智能安防、机器人交互、公安刑侦等领域。但跨摄像头行人检测与跟踪也远比单摄像头跟踪难度大，具体表现在以下方面：1) 实际监控场景中背景复杂，行人目标高度密集、互相遮挡等导致目标检测存在漏检和误检问题；2) 跟踪过程中行人之间的频繁遮挡、随时都会出现新目标或者有旧目标离开当前摄像头的视野，给多目标跟踪增加难度；3) 跨摄像头的光照变化、视角变化和行人的姿态变化导致同一个行人在不同摄像头下外貌特征发生变化，难以精确地在其他摄像头重新找回消失的目标。

[0003] 传统的行人检测与跟踪系统中，检测部分主要是用人工提取的形状特征或纹理特征，如HOG、LBP、Haar特征等，再加上SVM或Adaboost分类器进行行人检测。跟踪部分根据如何初始化目标可以分为两类：一类是基于检测的跟踪 (Detection-Based Tracking, DBT)，给定一个视频序列，在每帧中进行行人检测，然后进行顺序或批量跟踪；另一类是无需检测的跟踪 (Detection-Free tracking, DFT)，这种方法需要在第一帧手动初始化一定数量的目标，然后在后续帧定位这些物体。相对来说，基于检测的跟踪方法应用更为广泛，因为它可以自动发现新目标、自动终止消失的目标，本发明的跟踪系统属于第一种方法。传统目标重识别算法主要集中在提取行人的外貌特征、形状特征和局部特征等浅层特征，然后在图像库中找出匹配度最高的行人目标。但是人工提取的检测特征、重识别特征难以克服目标遮挡、光照变化和视角变化带来的消极影响，从而限制了精度的提升，而精准的行人检测和行人重识别是实现跨摄像头行人跟踪的前提条件，因此寻找高鲁棒性的特征并且构建一个统一的行人检测与跟踪系统是关键。

[0004] 现在已有的一些关于行人检测与跟踪的专利 (包括发明专利和发明专利公布专利) 如下：

[0005] 1) 申请号为：CN201410266099.X的中国发明专利《视频行人检测与跟踪方法》，此方法根据人头为人体的重要组成部分且具有刚体不变形，通过对不同人头目标的样本学习与训练对视频进行行人检测，确定人头目标，并基于人头色彩特征的差异性，采用粒子滤波对人头进行跟踪。此方法简便灵活，但在实际监控场景中行人可能十分密集或者存在遮挡，人头区域较小且差异性不大，仅仅对人头进行检测和跟踪会导致漏检率增加，影响行人检

测与跟踪的精度。

[0006] 2) 申请号为:CN201611027589.X的中国发明专利《一种基于混合特征的带缓存机制的实时行人检测与跟踪方法》,此方法的思想是在监控视频帧通过滑动窗口的方式,提取每一滑动窗口内图像的混合特征(包括图像的颜色特征、边缘特征、纹理特征等),并将其输入分类器进行检测,采用前后双向跟踪实现实时跟踪。所提供的方法虽然有很好的硬件可实现性,但是人工提取的特征难以克服光照变化以及遮挡问题,限制了精度的提升。

[0007] 3) 申请号为:CN201710709817.X的中国发明专利《基于深度学习的多目标行人检测与跟踪方法》,此方法的思想是利用行人检测卷积神经网络进行多目标行人检测和关键点检测,再对检测到的行人进行上半身姿态特征和颜色直方图特征提取,得到该行人的总特征,与上一个检测到的行人匹配,关联跟踪轨迹。此方法在行人跟踪阶段需要进一步计算姿态特征以及颜色直方图特征等,计算量需要考虑,且这些特征在光照变化、视角变化的时候不具有鲁棒性,会影响系统的稳定性。此方法也没有考虑多摄像头系统下的行人检测与跟踪的情况。

[0008] 4) 申请号为:CN201611122720.0的中国发明专利《面向轨道交通全景监控的跨摄像头可疑行人目标跟踪系统》,此方法的思想是先用混合高斯背景建模法检测出正在运动的可疑行人,再用Haar\_like特征的特征值作为运动目标跟踪过程中的特征描述算子,然后用朴素贝叶斯对特征进行处理。对于多摄像头之间的目标匹配模块,采用的是Surf特征点匹配和色彩直方图匹配相融合的方法,选用最佳匹配目标。所提供的方法实现了对可疑行人的跨摄像头跟踪,但是用的传统匹配方法限制了多摄像头间行人重识别的匹配精度。

[0009] 因此,探寻一种高精度、高鲁棒性、实时的跨摄像头行人检测与跟踪方法已经成为计算机视觉领域亟待解决的问题。

## 发明内容

[0010] 本发明的目的在于克服目标遮挡以及跨摄像头带来的光照变化、视角变化、尺度变化等问题,提出一种基于深度学习的跨摄像头行人检测与跟踪方法。能够在目标离开当前摄像头视野时在邻近摄像头重新找回该目标,实现大面积场景下多行人目标的持续跟踪,并且能获取行人的运动轨迹及运动信息,进行后续分析。

[0011] 为实现上述目的,本发明的技术方案包括一种基于深度学习的跨摄像头行人检测跟踪方法,包括以下步骤:

[0012] 步骤S1,通过训练行人检测网络,对输入的监控视频序列进行行人检测;

[0013] 步骤S2,以行人检测得到的目标框初始化跟踪目标,实现跟踪,包括提取行人检测网络中候选框所对应区域的浅层特征和深层特征,将浅层特征和深层特征结合起来作为物体的外观特征模型;

[0014] 步骤S3,持续进行新目标行人检测,并判断是否有目标消失,如果有则进入步骤S4进行行人重识别;

[0015] 步骤S4,行人重识别,包括得到目标消失信息之后,从行人检测网络得到的候选图像中找到与消失目标匹配度最高的图像,继续跟踪,跟踪结束时输出行人目标在多个摄像头下的运动轨迹。

[0016] 而且,所述步骤S1包括以下子步骤,

- [0017] 步骤S11,构建行人检测网络的训练集;
- [0018] 步骤S12,训练行人检测网络;
- [0019] 步骤S13,对于待测试的视频序列,首先将每一帧分离出来,把图片送入行人检测网络中进行行人检测,得到所有行人目标的置信度score和边框信息,当置信度score大于设定阈值时,认为是行人目标,保留目标框;
- [0020] 步骤S14,基于步骤S13初步检测所得的准目标,利用非极大值抑制算法去除冗余框,得到最终的检测结果。
- [0021] 而且,所述步骤S2包括以下子步骤,
- [0022] 步骤S21,根据行人检测得到的行人目标候选框,提取行人检测网络中候选框所对应区域的浅层特征和深层特征,将浅层特征和深层特征结合起来作为物体的外观特征模型;
- [0023] 步骤S22,使用相关滤波器计算跟踪目标在各个位置的置信度得分,得分最高的区域为物体的预测位置区域,从而得到各个目标的预测位置集合;
- [0024] 步骤S23,更新各个目标在当前帧的位置,再通过物体的外观特征模型进行比较,得出各个目标的外观置信度分数,根据置信度分数降低的程度判断当前跟踪的目标是否被遮挡,不断更新各个目标的跟踪状态,从而实现目标的自跟踪。
- [0025] 而且,所述步骤S4包括以下子步骤,
- [0026] 步骤S41,将目标跟踪所得最后一帧消失目标的图像,以及行人检测所得的多张候选图像作为输入,对于每一张图像首先通过行人检测网络的低层得到浅层特征图,再利用显著度检测算法提取目标的显著度剔除背景中的冗余信息,然后送入深层卷积层,输出深层特征图;融合浅层特征图和深层特征图,对融合后特征图进行全局池化得到全局特征向量,进行水平池化得到局部特征向量;
- [0027] 步骤S42,根据步骤S41求出的全局特征向量,计算消失目标与候选目标全局特征向量之间的L2距离,然后通过目标对齐的方法计算出局部特征向量间的距离;
- [0028] 步骤S43,计算消失目标图像与候选对象图像的最终距离,最终距离等于全局特征向量距离加上局部特征向量距离,根据最终距离的大小得到候选对象图像与消失目标图像的初始相似度排序;
- [0029] 步骤S44,通过重排序算法得到最终的相似性排序结果,选择相似性最高的图像继续跟踪。
- [0030] 与现有跨摄像头行人检测与跟踪技术相比,本发明具有以下优点和有益效果:
- [0031] (1) 本发明可以自动检测监控视频中的行人位置并持续跟踪,利用卷积神经网络在行人数据集上学习到的特征克服行人遮挡、光照变化、视角变化等干扰,提高了跟踪精度;
- [0032] (2) 本发明中目标跟踪与行人重识别模块共用检测网络中特征,通过浅层特征与深层特征结合的方式将行人检测、目标跟踪、行人重识别有机地融合起来,相比于三部分单独的网络结构,本发明的方法可以提高行人检测与跟踪的速度,满足实时监控的需求;
- [0033] (3) 本发明针对跨摄像头情况下出现的目标消失问题,采用行人重识别算法在其他摄像头下找回消失的目标,实现了大面积、长时间的跨摄像头行人跟踪。

## 附图说明

- [0034] 图1为本发明实施例的总体技术框图。
- [0035] 图2为本发明实施例的跨摄像头下重新找回消失目标的示意图。
- [0036] 图3为本发明实施例的跟踪目标状态图。
- [0037] 图4为本发明实施例的重排序算法的示意性流程图。

## 具体实施方式

[0038] 为了更加清楚地阐述本发明的目的、技术方案和有益效果,以下结合附图和具体实施例进行详细描述。应理解本发明的具体实施方式不限于这里的示例描述。

[0039] 近年来,深度学习在许多计算机视觉任务,如图像识别、目标检测、目标跟踪等任务中都取得了十分不错的表现。卷积神经网络的浅层卷积层包含适用于目标定位的高分辨率特征,深层卷积层包含易于区分的视觉特征。本发明的创新点在于利用浅层特征与深层特征结合的方式提取行人在不同摄像头下不变的本质特征,克服光照变化和视角变化的影响,并且跟踪和行人重识别部分用到的特征也从检测网络中间层提取,将行人检测、关联跟踪和行人重识别有机地融合起来,并通过各模块间的信息交互,有效地降低了跟踪过程中跨摄像头的噪声干扰,提高了跨摄像头行人检测和跟踪的准确性和实时性。

[0040] 图1是本发明技术方案的总体框图,具体实施时,本发明方法也可采用模块化方式实现,提供基于深度学习的跨摄像头行人检测与跟踪系统,除去输入和输出部分,包含三个模块:行人检测模块、多目标跟踪模块和行人重识别模块。

[0041] 输入为多个摄像头下的视频;

[0042] 行人检测模块的主要功能是利用深度卷积神经网络,检测监控区域所有摄像头拍摄的视频序列中的行人目标,得到每个目标的置信度和边框坐标,实施例中利用改进的Faster RCNN网络进行行人检测,根据输入的不同摄像头下的视频序列,计算出行人目标在图像中的位置(即边框坐标信息)以及置信度得分;

[0043] 多目标跟踪模块主要负责将检测结果作为跟踪的初始化目标,提取检测网络的中间层特征对检测出来的行人目标进行特征建模,实现构建行人外观特征模型,利用相关滤波算法(滤波器)预测计算目标的运动轨迹,实现多目标跟踪;

[0044] 行人重识别模块主要解决跨摄像头情况下的目标消失问题,当跟踪过程中目标离开当前摄像头视野时,根据跟踪模块传来的消失目标信息,从行人检测模块检测出来的候选目标图像库中,通过求取全局特征和局部特征,找到相似度最高的匹配目标,返回给多目标跟踪模块继续跟踪。

[0045] 整个系统可在主动结束跟踪或者所有行人目标离开监控区域摄像头的视野时,最后输出不同行人在多个摄像头下的运动轨迹以及运动信息。

[0046] 本发明实施例基于深度学习的跨摄像头行人检测与跟踪方法,参见图2,图中的S表示某监控区域中每一个摄像头的标号,S-1和S+1表示与摄像头S邻近的摄像头标号,具体包括以下步骤:

[0047] 步骤S1,对输入的监控视频序列进行行人检测,具体实施时可预先搭建好监控区域的摄像头网络,对拍摄的视频序列中所有的行人目标进行检测。本步骤可采用检测模块实现。



[0048] 具体包括以下步骤:

[0049] 步骤S11,构建行人检测的训练集,具体实施时可在准备阶段,将目前公共的行人检测专用数据集Caltech数据集、INRIA Person数据集、ETH数据集等制作成PASCAL VOC2007数据集的格式,作为行人检测网络的训练集,具体包括数据集的图片以及标注文件(xml格式),标注文件中存有每个目标的边框信息( $x^*, y^*, w^*, h^*$ )和类别,其中( $x^*, y^*$ )表示边框左上角的坐标, $w^*$ 和 $h^*$ 分别表示边框的宽和高。

[0050] 步骤S12,行人检测网络训练阶段,采用卷积神经网络VGG16作为检测网络的基网络,VGG16的具体网络结构如表1所示,由13个卷积层和3个全连接层组成。具体实施时,也可采用ZFNet、ResNet等其他网络结构。

[0051] 表1

卷积神经网络 VGG16 (13 个卷积层+3 个全连接层)			
层数	卷积层	卷积核大小	特征通道数
第一层	Conv1_1	3×3	64
	Conv1_2	3×3	64
第二层	Conv2_1	3×3	128
	Conv2_2	3×3	128
第三层	Conv3_1	3×3	256
	Conv3_2	3×3	256
	Conv3_3	1×1	256
第四层	Conv4_1	3×3	512
	Conv4_2	3×3	512
	Conv4_3	1×1	512
第五层	Conv5_1	3×3	512
	Conv5_2	3×3	512
	Conv5_3	1×1	512
全连接层	FC1		4096
全连接层	FC2		4096
全连接层	FC3		2

[0054] 本发明实施例用在ImageNet上预先训练的VGG16模型进行参数的初始化,使用近似联合训练(端到端)的方式训练改进的Faster RCNN网络,具体改进如下:

[0055] ①输入图像尺寸统一将短边缩放到800,用来提高对小物体的检测率;

[0056] ②Faster RCNN网络中的区域生成网络(Region Proposal Networks,RPN)中的滑

动窗采用长宽比1:0.41 (这是经过统计之后的行人目标的平均长宽比),这样可以在保证精度的前提下提高检测速度;

[0057] ③由于监控视频中拍摄距离远近的关系,行人的大小差异十分巨大,将滑动窗的尺寸增设为9种,分别为 $\{32 \times 32, 64 \times 64, 96 \times 96, 128 \times 128, 160 \times 160, 192 \times 192, 224 \times 224, 256 \times 256, 288 \times 288\}$ 。采用长宽比1:0.41后,例如 $32 \times 32$ 滑动窗的实际大小为 $32 \times 13$ ,以此类推。

[0058] 本发明检测网络的损失函数由两部分组成,包括分类损失和边框回归损失,定义如式(1):

$$[0059] \quad L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

[0060] 式中下标*i*表示滑动窗的标号,取值范围为 $1, 2, \dots, N_{cls}$ ,  $N_{cls}$ 表示一次批量训练使用的滑动窗数量,  $N_{reg}$ 表示滑动窗坐标位置的数量;  $\lambda$ 是调控分类损失和边框回归损失的因子(具体实施时在本发明实施例中设 $\lambda=10$ );  $p_i$ 表示该窗口预测为行人的概率,  $p_i^*$ 表示真实概率(窗口为正样本  $p_i^*=1$ , 负样本  $p_i^*=0$ )。

[0061] 已知训练样本中每个目标的真实边框坐标信息为 $(x^*, y^*, w^*, h^*)$ ; RPN网络中滑动窗的边框信息为 $(x_a, y_a, w_a, h_a)$ , 其中 $(x_a, y_a)$ 表示滑动窗的左上角坐标,  $w_a, h_a$ 分别表示滑动窗的宽和高; RPN网络预测的边框信息为 $(x, y, w, h)$ , 其中 $(x, y)$ 表示预测边框左上角的坐标,  $w$ 和 $h$ 分别表示预测边框的宽和高。  $t_i$ 表示预测边框与滑动窗坐标关系的4维参数化向量 $(t_x, t_y, t_w, t_h)$ ,  $t_i^*$ 表示真实边框与滑动窗坐标关系的4维参数化向量 $(t_x^*, t_y^*, t_w^*, t_h^*)$ , 具体计算公式如下:

$$[0062] \quad \begin{aligned} t_x &= (x - x_a) / w_a, t_y = (y - y_a) / h_a, \\ t_w &= \log(w / w_a), t_h = \log(h / h_a), \\ t_x^* &= (x^* - x_a) / w_a, t_y^* = (y^* - y_a) / h_a, \\ t_w^* &= \log(w^* / w_a), t_h^* = \log(h^* / h_a), \end{aligned} \quad (2)$$

[0063] 公式(1)中分类损失 $L_{cls}$ 采用交叉熵损失, 边框回归损失 $L_{reg}$ 采用SmoothL1损失, 具体定义如式(3) - (5)所示:

$$[0064] \quad L_{cls}(p_i, p_i^*) = -\log(p_i \times p_i^*) \quad (3)$$

$$[0065] \quad L_{reg}(t_i, t_i^*) = smooth_{L1}(t_i - t_i^*) \quad (4)$$

$$[0066] \quad smooth_{L1}(\Delta t) = \begin{cases} 0.5|\Delta t|^2, & \text{if } |\Delta t| < 1 \\ |\Delta t| - 0.5, & \text{otherwise} \end{cases} \quad (5)$$

[0067] 式(5)中 $\Delta t = t_i - t_i^*$ , 表示变量之间的差值。通过最小化损失函数 $L(\{p_i\}, \{t_i\})$ 来进行网络的训练和调节, 行人检测网络采用随机梯度下降法(Stochastic Gradient Descent, SGD)进行反向传播, 迭代80000次, 学习率在前30000次迭代过程中设置为0.001, 后面每20000次下降为原来的0.1。

[0068] 步骤S13,对于待测试的视频序列,首先将每一帧分离出来,利用S12训练所得改进的Faster RCNN网络对多个摄像头下的视频序列进行行人检测,得到所有行人目标的置信度score和预测边框信息(x,y,w,h)。每个预测边框的置信度由softmax分类器计算得到,当置信度score大于设定阈值(具体实施时,可根据精度需要预设相应阈值,一般设阈值T为0.7以上,本发明实施例中设阈值T=0.8)时,认为是行人目标,保留目标框。

[0069] 步骤S14,由步骤S13已经初步检测到了多个准目标,多个准目标往往只对应一个行人目标,利用非极大值抑制(Non-maximum Suppression,NMS)算法对这些候选框去除冗余框,得到最终的检测结果,实现方式如下:

[0070] 首先对目标框的得分score排序,得分最高的目标框记为 $\text{Boxes}_{\text{MaxScore}}$ ,遍历排序后的目标框集合,分别与 $\text{Boxes}_{\text{MaxScore}}$ 进行比较,计算重叠面积与两者间最小面积的比例,保留比例小于设定IoU阈值(具体实施时,可根据精度需要预设相应阈值,本发明实施例设置为0.5)的包围框,若超过则抑制掉该目标框,经过遍历迭代得到最终目标框 $\text{Boxes}_{\text{best}}$ 。

[0071] 步骤S2,以行人检测得到的目标框初始化跟踪目标,实现跟踪,进一步包括以下步骤:

[0072] 步骤S21,提取目标的浅层特征和深层特征,二者结合起来构建物体的外观特征模型。根据步骤S1提供的行人目标候选框,从行人检测网络的第一层Conv1\_2提取浅层特征,卷积神经网络第一层通常提取目标的形状、边缘等全局特征,称之为浅层特征;从行人检测网络的第五层Conv5\_3提取深层特征,卷积层层次越深,提取的高级语义信息越丰富,具有某种程度上的平移和尺度不变性,这种不变性特征称之为深层特征。先对深层特征图上采样到与浅层特征图大小一致,然后将浅层特征与放大后的深层特征连接起来,得到融合后的特征,这种融合后的特征既可以获得目标的全局信息,又可以保留细节信息,有助于提高跟踪的精度。

[0073] 设对于某个目标Z提取到的深层特征和浅层特征,共D个特征通道,分别记为 $Z^1, \dots, Z^d, \dots, Z^D$ ,  $N_d$ 表示第d个通道 $Z^d$ 中的采样点数。使用连续域的特征表示可以获得物体更高的定位精度,为了处理连续域的学习问题,引入一个内插模型,将跟踪框转化为一维区间 $[0, T) \in \mathbb{R}$ ,  $\mathbb{R}$ 表示实数集。对每一个特征通道 $Z^d$ ,定义一个插值模型如式(6):

$$[0074] \quad J_d\{Z^d\}(t) = \sum_{n=0}^{N_d-1} Z^d[n] b_d(t - \frac{T}{N_d}n) \quad (6)$$

[0075] 其中,标量T表示支持区间的大小,t为区间 $[0, T)$ 中的取值, $b_d$ 表示一个三次立方插值函数,插值模型 $J_d\{Z^d\}(t)$ 实际上可以视为插值函数 $b_d$ 平移之后叠加形成的函数,特征通道d中第n维的特征值 $Z^d[n]$ 可以视为相应的位移函数的权值, $n \in \{0, \dots, N^d-1\}$ 。得到在位置区间 $[0, T)$ 的物体连续外观特征值 $J_1\{Z^1\}, \dots, J_d\{Z^d\}, \dots, J_D\{Z^D\}$ ,  $\Sigma J_d\{Z^d\}$ 表示物体Z的外观特征模型,简化表示为 $J\{Z\}$ 。

[0076] 步骤S22,通过训练样本训练滤波器,对于目标Z在步骤S21中其外观模型所构造的位置区间 $[0, T)$ 内,使用相关滤波器计算跟踪目标在各个位置的置信度得分,得分最高的区域为物体的预测位置区域,得到各个目标的预测位置集合,从而构建物体的运动预测模型。然后移动目标跟踪框以预测位置区域为中心,更新物体的外观特征模型;

[0077] 实施例中,使用一个线性卷积算子,将步骤S21所得样本空间的目标Z的外观特征模型 $J\{Z\}$ ,映射到目标置信函数 $s(t) = S_f(Z)$ ,  $s(t)$ 即为目标在位置t的置信度得分,因此s

(t) 的最大值所对应的位置 $t$ 就是目标所移动到的新位置。

[0078] 算子 $S_f$ 的推导方程如式(7)：

$$[0079] \quad S_f\{Z\} = \sum_{d=1}^D f^d * J_d\{Z^d\} \quad (7)$$

[0080] 其中,\*表示卷积操作, $f^d$ 是特征通道 $d$ 的连续滤波器, $D$ 表示总的特征通道数。例如VGG16的第一层通道数为64,第五层通道数为512,那么 $D=64+512=576$ 。 $d$ 取值为 $1 \sim D$ 。训练连续卷积滤波器 $f$ ,再通过求解上式,由 $P\{x\} = \max(s(t))$ 获得目标 $x$ 的移动新中心点,记为 $P\{x\}$ ,即为物体的运动预测模型。

[0081] 步骤S23,更新各个目标在当前帧的位置,再通过物体的外观特征模型进行比较,得出各个目标的外观置信度分数,判断当前跟踪的目标是否被遮挡,更新各个目标的跟踪状态,从而实现目标的自跟踪;

[0082] 实施例中,根据马尔科夫决策过程,定义跟踪目标的状态为 $s \in S$ ,其中 $S$ 由激活态 $S_{Active}$ 、跟踪态 $S_{Tracked}$ 、丢失态 $S_{Lost}$ 和闲置态 $S_{Inactive}$ 组成,即 $S = S_{Active} \cup S_{Tracked} \cup S_{Lost} \cup S_{Inactive}$ 。

[0083] 如图3所示跟踪过程中目标的状态图,行人检测模块检测到各个摄像头新出现的目标为激活态 $s = S_{Active}$ ,当目标检测框满足激活跟踪条件,即置信度高于设定阈值时,将该目标再转换为跟踪态;当目标检测框不满足激活跟踪条件,即目标检测框置信度太低或者和已经处于跟踪态的目标高度重叠,将该目标转为闲置态。跟踪态的目标是当前正在各个摄像头图像序列中跟踪的目标,在非检测帧根据运动预测模型对每一个目标采取自跟踪策略;在检测帧用外观模型特征和运动预测模型特征匹配目标检测框,以此达到修正跟踪目标位置和目标尺度变换的目的。其中,非检测帧是指不需要检测算法提供检测框的视频帧;检测帧是指需要检测算法提供检测框作为初始跟踪框的视频帧。丢失态为可能被遮挡的目标或者离开某一摄像头的目标,暂时未能在图像数据中显示出来的目标,目标处于丢失态时,根据行人重识别算法,检测到丢失目标后,保持该目标原始ID不变,状态恢复为跟踪态。若丢失目标长时间处于丢失态,未能通过目标重识别模块转为正常的跟踪态,则将该目标转为闲置态,不再对该目标进行任何后续跟踪计算,这样可以减少算法的计算耗时。

[0084] 本步骤可由多目标跟踪模块实现,把检测到的行人结果传给跟踪模块,作为初始化跟踪目标,给每个目标一个不同的ID,并记录他们的轨迹。

[0085] 步骤S3,持续进行新目标行人检测,并判断是否有目标消失,如果有则进入步骤S4进行行人重识别。

[0086] 实施例中,考虑到跟踪过程中会遇到以下两种情况:

[0087] 1) 由于新的目标会不时进入摄像头视野,采用非固定性间隔多帧的方式,每隔 $F$ 帧(在本发明实施例中 $F=5$ )调用一次检测模块进行行人检测,为新出现的目标分配新的ID。

[0088] 2) 如果跟踪框的置信度下降或者离开视频区域,多目标跟踪模块会把该目标最后一帧的跟踪信息传给行人重识别模块,并且调用行人检测模块对其他摄像头下的当前帧进行检测,得到的所有目标检测框也传给行人重识别模块,作为行人重识别的图像库。跨摄像头进行目标重识别的具体实现如图2所示,当摄像头 $S$ 中的某个目标消失在它的视野时,多目标跟踪模块会调用行人检测模块和行人重识别模块,根据特征匹配的信息,在其他摄像头中找到该目标,保持ID不变继续跟踪。

[0089] 步骤S4,行人重识别,包括得到目标消失信息之后,利用行人重识别算法,从行人检测网络得到的候选图像中找到与消失目标匹配度最高的图像,继续跟踪,跟踪结束时输出行人目标在多个摄像头下的运动轨迹。

[0090] 本步骤可采用行人重识别模块实现。

[0091] 行人重识别模块接收到跟踪模块传来的消失目标信息之后,会在检测到的匹配图像库中找到与消失目标匹配度最高的图像,传回给跟踪模块继续跟踪,此时传回的图像保持跟消失目标一样的ID,在跟踪结束时输出不同行人目标在多个摄像头下的运动轨迹。

[0092] 实施例具体包括以下步骤:

[0093] 步骤S41,设对某消失目标,通过行人检测得到N-1张候选图像,行人重识别模块的输入为目标跟踪模块传入的一张消失目标的图像以及行人检测模块传入的N-1张候选图像,对于每一张图像首先通过行人检测网络的第一层(低层)得到浅层特征图,然后利用显著度检测算法提取目标的显著度从而剔除背景中的冗余信息后送入深层卷积层,在第五层(高层)输出得到深层特征图。对浅层特征图与深层特征图进行融合,可先对深层特征图上采样到与浅层特征图大小一致,然后连接到一起,所以通道数是相加的。

[0094] 在全局特征分支对浅层特征图与深层特征图融合之后的特征图进行全局池化,得到一维的全局特征向量。在局部特征分支对融合后的特征图做水平池化,也就是将图片分块为H行(具体实施时,用户可以预设取值,本发明实施例中设H=7),对每一行上都提取全局特征,再使用 $1 \times 1$ 卷积核进行降维,这样就得到了H维局部特征向量并且每一个向量都对应表征图像的一个横块,因此每一张输入图像可以被表征为一维全局特征(大小为 $1 \times 2048$ )和H维局部特征(大小为 $7 \times 128$ )。

[0095] 步骤S42,根据步骤S41求出的全局特征向量,计算消失目标与N-1个候选目标全局特征向量之间的L2距离,然后通过目标对齐的方法计算出局部特征向量间的距离;

[0096] 由于全局特征向量是一维向量,因此目标全局特征向量与N-1个候选对象全局特征向量之间的距离可以直接计算向量间的欧氏距离得到。对于同一个目标的两张图像,一张图像中目标的某块部分的局部特征总是与另一图像中目标的相对应部分的局部特征有着很高的相似度,例如不同图像下的同一个行人的头部部分的局部特征相似度很高,而不同的目标之间的局部特征则没有这一特性。基于这样一个特点,就可以将候选对象中的目标与非目标对象区分开来。将局部特征向量距离定义为从图像的顶部到底部各个局部部分进行动态匹配实现局部特征对齐的最小总距离。

[0097] 假设两张图像的局部特征向量分别为 $F = \{f_1, \dots, f_H\}$ 和 $G = \{g_1, \dots, g_H\}$ ,首先将距离归一化到 $[0, 1)$ 区间,归一化操作如式(8):

$$[0098] \quad d_{i_h, j_h} = \frac{e^{\|f_{i_h} - g_{j_h}\|_2} - 1}{e^{\|f_{i_h} - g_{j_h}\|_2} + 1} \quad i_h, j_h \in 1, 2, 3, \dots, H \quad (8)$$

[0099] 式中,e为自然对数的底数, $d_{i_h, j_h}$ 为第一张图像由上到下第 $i_h$ 部分特征向量与第二张图像由上到下第 $j_h$ 部分特征向量的距离,基于上式可定义一个矩阵D,其中下标 $(i_h, j_h)$ 对应的元素值 $d_{i_h, j_h}$ 。因此给出局部特征向量距离的定义:矩阵D中从下标(1,1)到(H,H)的最短路径的距离。这个最短路径可以用以下动态规划方程式(9)描述:

$$[0100] \quad L_{i_h, j_h} = \begin{cases} d_{i_h, j_h} & i_h = 1, j_h = 1 \\ L_{i_h-1, j_h} + d_{i_h, j_h} & i_h \neq 1, j_h = 1 \\ L_{i_h, j_h-1} + d_{i_h, j_h} & i_h = 1, j_h \neq 1 \\ \min(L_{i_h-1, j_h}, L_{i_h, j_h-1}) + d_{i_h, j_h} & i_h \neq 1, j_h \neq 1 \end{cases} \quad (9)$$

[0101] 其中  $L_{i_h, j_h}$  为下标 (1,1) 到  $(i_h, j_h)$  的最短路径的总距离,  $L_{H,H}$  则表示两张图像的局部特征向量距离。

[0102] 步骤S43, 计算消失目标图像与候选对象图像的最终距离, 等于步骤S42计算出的全局特征向量距离加上局部特征向量距离, 通过最终距离的大小给出候选对象图像与目标图像的初始相似度排序, 距离越小, 排名越靠前。

[0103] 步骤S44, 通过重排序算法得到最终的相似性排序结果, 选择相似性最高的图像传回给多目标跟踪模块, 实现对消失目标的持续跟踪。重排序算法的具体实现如图4所示, 首先根据步骤S43得到的初始相似度排序, 选出排名靠前的k张候选图像, 称之为top-k图像; 加上消失目标图像一共k+1张图像, 把这个k+1张图像的特征向量相加, 然后除以 (k+1) 得到平均特征向量; 以求出的平均特征向量作为消失目标的特征, 然后重新计算步骤S42中与N-1个候选目标图像的距离; 最后根据计算得到的距离重新排序, 排名第一的候选目标图像即为与消失目标匹配度最高的候选图像。

[0104] 以上就是本发明的详细步骤, 应当理解的是本说明书未详细阐述的部分均属于现有技术。具体实施时, 可采用计算机软件技术实现以上流程的自动运行。

[0105] 本发明的输入为多个摄像头下的视频序列, 行人检测模块以及行人重识别模块都是服务于更加精确的多目标跟踪, 最后输出不同行人目标在多摄像头下的运动轨迹和运动信息。通过在街道、火车站、地铁站、图书馆等人群密集场所进行监控实验, 本发明能够进行精准的行人检测和跟踪, 在跨摄像头之间目标交互或者有目标消失时能成功找回目标并继续跟踪, 本发明提出的基于深度学习的浅层特征与深层特征结合的方式可以很好地克服目标遮挡、光照变化以及尺度变化等因素影响, 实现对行人目标精确的实时跟踪。

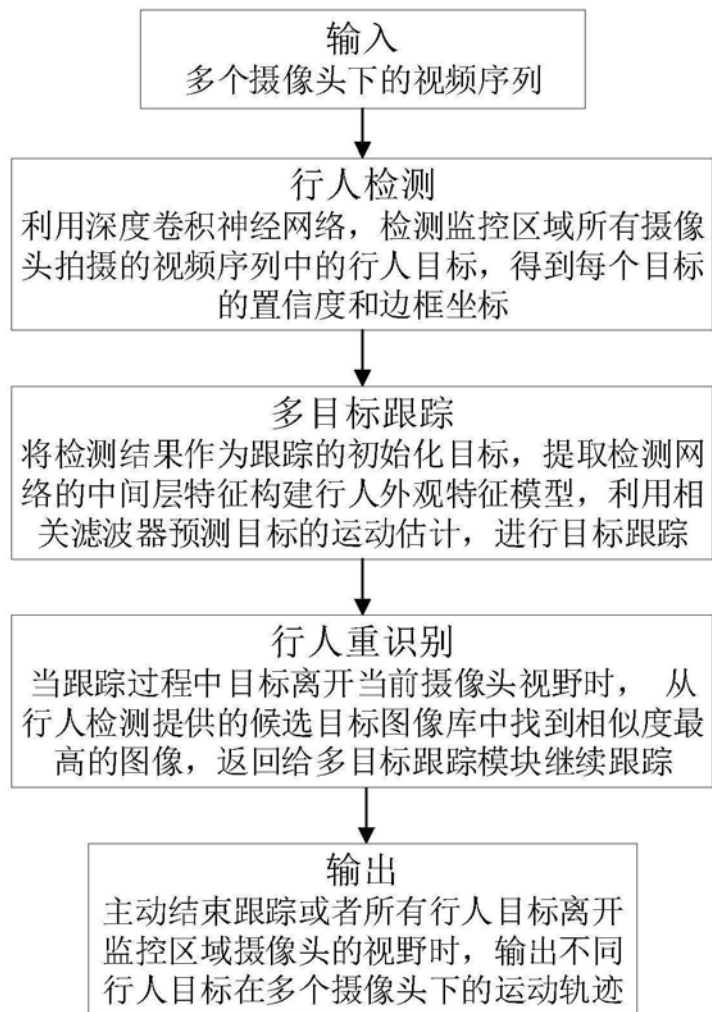


图1

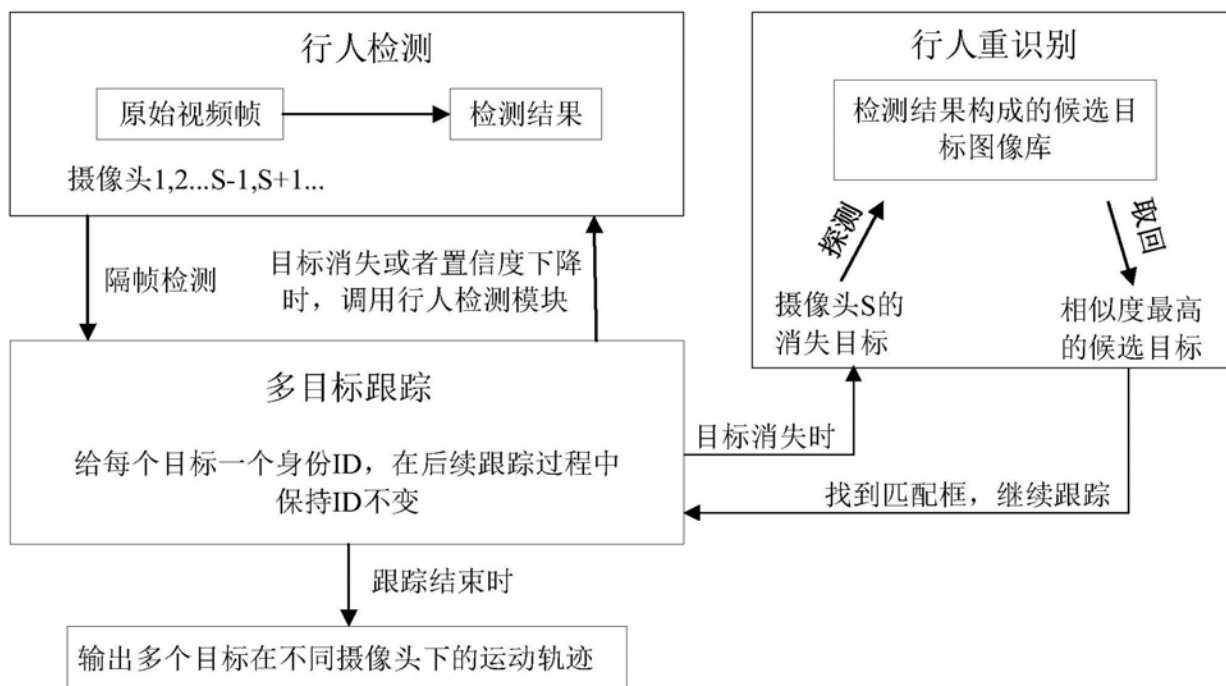


图2

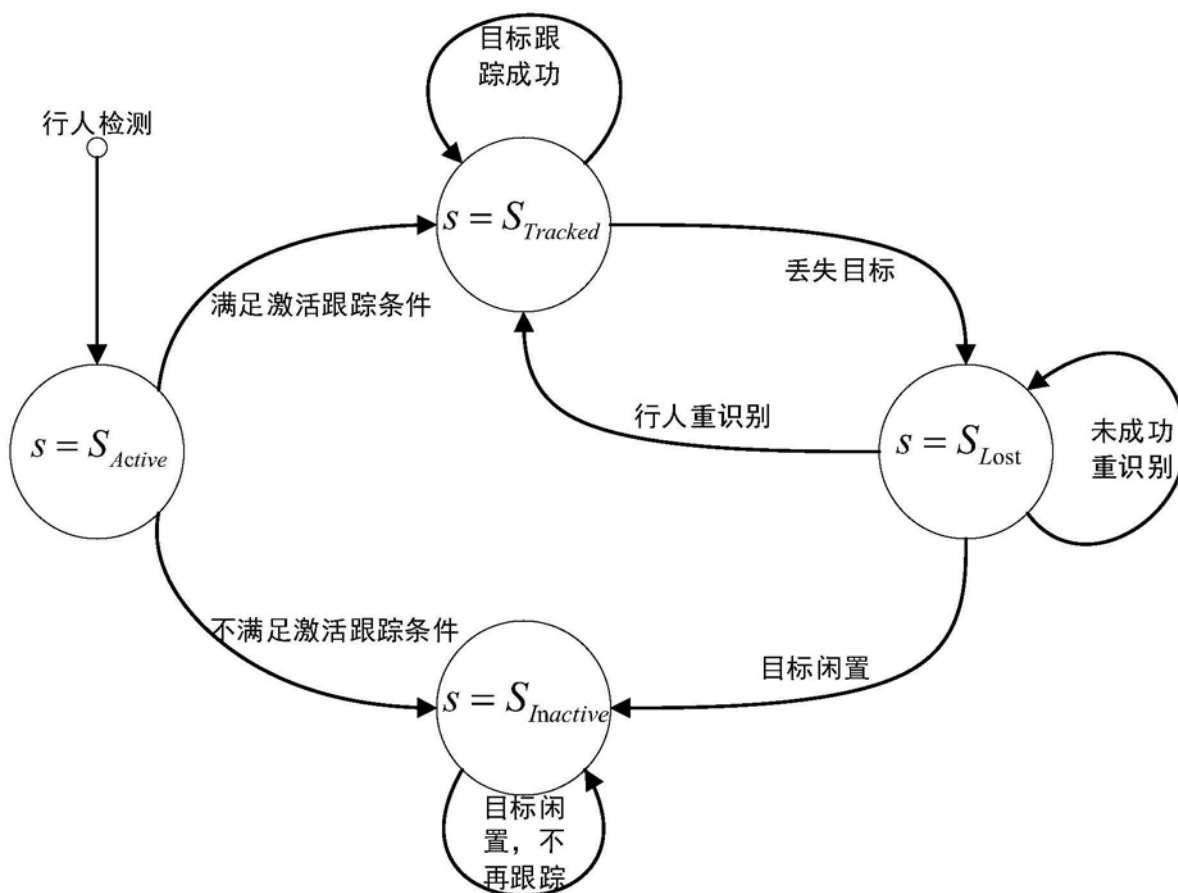


图3



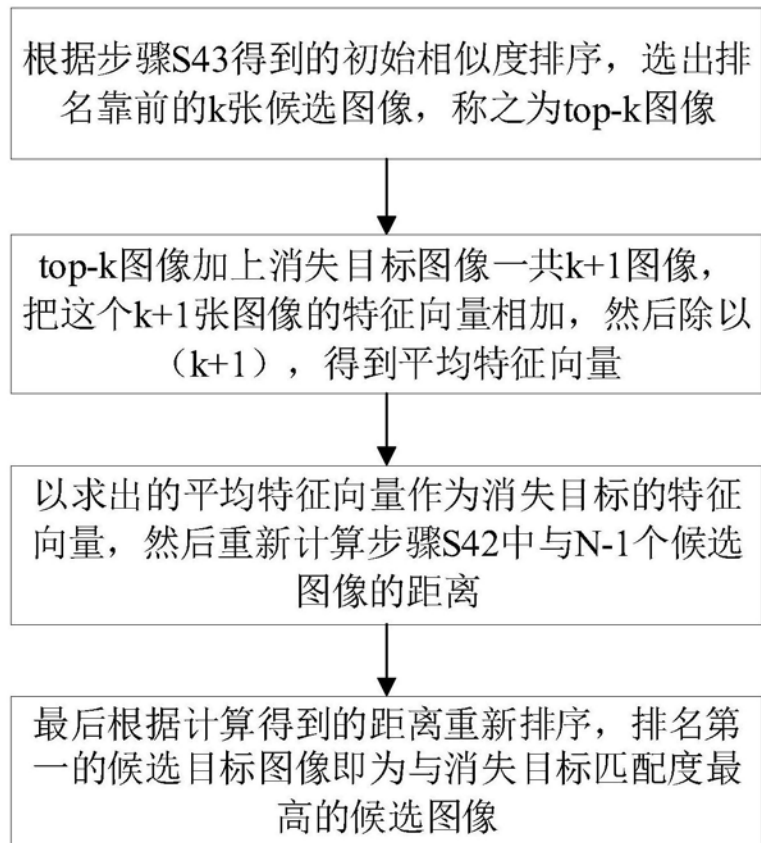


图4