



(12) 发明专利

(10) 授权公告号 CN 108229352 B

(45) 授权公告日 2021.09.07

(21) 申请号 201711397963.X

H04N 7/18 (2006.01)

(22) 申请日 2017.12.21

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 108229352 A

CN 106803913 A, 2017.06.06

CN 106931968 A, 2017.07.07

(43) 申请公布日 2018.06.29

US 2016092733 A1, 2016.03.31

US 2016104385 A1, 2016.04.14

(73) 专利权人 上海交通大学
地址 200240 上海市闵行区东川路800号

CN 107146177 A, 2017.09.08

CN 106941602 A, 2017.07.11

(72) 发明人 邵奔驰 姜飞 申瑞民

Jifeng Dai等.R-FCN: Object Detection via Region-based Fully Convolutional Networks.《30th Conference on Neural Information Processing Systems》.2016,

(74) 专利代理机构 上海科盛知识产权代理有限公司 31225

王忠民等.一种基于卷积神经网络深度学习的人体行为识别方法.《计算机科学》.2016,第43卷(第11A期),

代理人 翁惠瑜

审查员 郭弘倩

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/32 (2006.01)

H04N 5/14 (2006.01)

H04N 5/76 (2006.01)

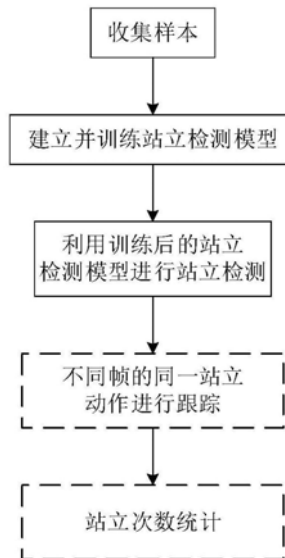
权利要求书1页 说明书4页 附图4页

(54) 发明名称

一种基于深度学习的站立检测方法

(57) 摘要

本发明涉及一种基于深度学习的站立检测方法,该方法包括:1)收集样本,各所述样本包括样本图片和对应的标注文件;2)建立站立检测模型,该站立检测模型基于卷积神经网络结构,并基于所述样本以R-FCN目标检测算法进行训练,所述站立检测模型包括高年级站立检测模型和低年级站立检测模型;3)利用训练后的站立检测模型对待测视频进行站立检测。与现有技术相比,本发明具有检全率、准确率高,适用于复杂教室环境等优点。



1. 一种基于深度学习的站立检测方法,其特征在于,该方法包括:

1) 收集样本,各所述样本包括样本图片和对应的标注文件,所述标注文件的信息包括站立人员类型,所述站立人员类型包括高年级学生、低年级学生和教师,一至三年级为低年级,四年级及以上的年级划分为高年级;

2) 建立站立检测模型,该站立检测模型基于卷积神经网络结构,并基于所述样本以R-FCN目标检测算法进行训练,所述站立检测模型包括高年级站立检测模型和低年级站立检测模型;

3) 利用训练后的站立检测模型对待测视频进行站立检测;

所述建立站立检测模型具体为:

201) 利用所有样本训练一基础站立模型;

202) 利用带有高年级学生标注的样本和带有低年级学生标注的样本分别在所述基础站立模型的基础上进行进一步训练,获得高年级站立检测模型和低年级站立检测模型。

2. 根据权利要求1所述的基于深度学习的站立检测方法,其特征在于,该方法还包括步骤:

4) 根据前一帧的站立检测结果和当前的站立检测结果进行站立的跟踪。

3. 根据权利要求2所述的基于深度学习的站立检测方法,其特征在于,所述跟踪具体为:

401) 获取第一个图像帧及检测到的站立框坐标,各站立框对应建立有一tracklet数组,且状态初始化为ALIVE;

402) 获取下一个图像帧,判断镜头是否发生变换,若是,则将所有tracklet数组的状态改为DEAD,重新建立新的tracklet数组,返回步骤402),若否,则执行步骤403);

403) 遍历当前图像帧检测到的所有站立框,利用跟踪算法为每一站立框选择最佳匹配的一个tracklet数组;

404) 对于在当前图像帧下未被匹配的tracklet数组,判断其状态是否为ALIVE,若是,则状态修改为WAIT,若否,则状态修改为DEAD,返回步骤402),直至处理完成所有图像帧。

4. 根据权利要求3所述的基于深度学习的站立检测方法,其特征在于,所述判断镜头是否发生变换具体为:

获取相邻两个图像帧,判断两个图像帧之间灰度值的差值是否大于第一阈值,若是,则判定两个图像帧的像素点发生变化;

判断像素点发生变化的两个图像帧间变化的像素点个数占总像素点的比例是否大于第二阈值,若是,则判定为发生镜头视角变换,若否,则未发生镜头视角变换。

5. 根据权利要求3所述的基于深度学习的站立检测方法,其特征在于,所述选择最佳匹配的一个tracklet数组具体为:

选择与站立框最近的一个tracklet,计算该tracklet的边框与所述站立框的之间宽度差与长度差之和,当和小于站立框宽度的三分之一,且tracklet的边框与所述站立框重合度大于0.3时,判断所述站立框与该tracklet最佳匹配。

6. 根据权利要求2所述的基于深度学习的站立检测方法,其特征在于,该方法还包括步骤:

5) 对跟踪获得的站立次数进行计数。

一种基于深度学习的站立检测方法

技术领域

[0001] 本发明涉及一种图像处理技术,尤其是涉及一种基于深度学习的站立检测方法。

背景技术

[0002] 在传统教学的录播教室以及类似的室内监控中,需要一套能够自动检测站立行为的系统,以便能够评估整体氛围和参与者的参与程度。但由于站立行为的特征又和个体的身高有着密切的关系,而同一场景中个体的身高有着差别,而不同场景下的身高分布也是不一样的,使得在传统的教室环境中检测人的站立行为仍是一件艰巨的任务。

[0003] 现有一种基于主从摄像机的学生跟踪定位方法,该方法使用两个从摄像机和一个带云台装置的主摄像机。从摄像机可自动或手动生成感兴趣区域,使用背景差分法检测有无学生进入或离开感兴趣区域,且向主摄像机发送检测结果;主摄像机根据从摄像机传递过来的信息判断站起的学生数目,并根据站起的学生数目选择全景录制模式或定位录制模式,其中定位录制模式中使用帧间差分法检测所有运动目标的轮廓,判断轮廓中心点最高的目标为站起的学生,其流程图如图1所示。该方法虽然具有一定定位精度,但还存在以下不足:

[0004] 1、该方法将摄像头安置于黑板两侧,安装高度为学生坐下后与头顶齐平处,使得摄像头基本处于学生视线的前方,较容易对学生产生心理压力。而且在此高度下,学生可能在有意或无意的情况下触碰到摄像机,造成结果的偏差。

[0005] 2、需要一主一从两个摄像头去完成站立检测。

[0006] 3、因为低年级的小学生站立高度和坐着时的高度相差不大,所以此方法在低年级的学生上的效果较差。

发明内容

[0007] 本发明就是为了克服上述现有技术存在的缺陷而提供一种基于深度学习的站立检测方法。

[0008] 本发明的目的之一是只需一个摄像头即可实现站立检测。

[0009] 本发明的目的之二是着重提高在低年级学生上的检测效果。

[0010] 本发明的目的之三是将不同帧中的同一站立行为串联起来,避免同一站立行为在不同帧的重复计数。

[0011] 本发明的目的可以通过以下技术方案来实现:

[0012] 一种基于深度学习的站立检测方法,该方法包括:

[0013] 1) 收集样本,各所述样本包括样本图片和对应的标注文件;

[0014] 2) 建立站立检测模型,该站立检测模型基于卷积神经网络结构,并基于所述样本以R-FCN目标检测算法进行训练,所述站立检测模型包括高年级站立检测模型和低年级站立检测模型;

[0015] 3) 利用训练后的站立检测模型对待测视频进行站立检测。

- [0016] 所述标注文件的信息包括站立人员类型。
- [0017] 所述站立人员类型包括高年级学生、低年级学生和老师。
- [0018] 所述建立站立检测模型具体为：
- [0019] 201) 利用所有样本训练一基础站立模型；
- [0020] 202) 利用带有高年级学生标注的样本和带有低年级学生标注的样本分别在所述基础站立模型的基础上进行进一步训练,获得高年级站立检测模型和低年级站立检测模型。
- [0021] 该方法还包括步骤：
- [0022] 4) 根据前一帧的站立检测结果和当前的站立检测结果进行站立的跟踪。
- [0023] 所述跟踪具体为：
- [0024] 401) 获取第一个图像帧及检测到的站立框坐标,各站立框对应建立有一tracklet数组,且状态初始化为ALIVE；
- [0025] 402) 获取下一个图像帧,判断镜头是否发生变换,若是,则将所有tracklet数组的状态改为DEAD,重新建立新的tracklet数组,返回步骤402),若否,则执行步骤403)；
- [0026] 403) 遍历当前图像帧检测到的所有站立框,利用跟踪算法为每一站立框选择最佳匹配的一个tracklet数组；
- [0027] 404) 对于在当前图像帧下未被匹配的tracklet数组,判断其状态是否为ALIVE,若是,则状态修改为WAIT,若否,则状态修改为DEAD,返回步骤402),直至处理完成所有图像帧。
- [0028] 所述判断镜头是否发生变换具体为：
- [0029] 获取相邻两个图像帧,判断两个图像帧之间灰度值的差值是否大于第一阈值,若是,则判定两个图像帧的像素点发生变化；
- [0030] 判断像素点发生变化的两个图像帧间变化的像素点个数占总像素点的比例是否大于第二阈值,若是,则判定为发生镜头视角变换,若否,则未发生镜头视角变换。
- [0031] 所述选择最佳匹配的一个tracklet数组具体为：
- [0032] 选择与站立框最近的一个tracklet,计算该tracklet的边框与所述站立框宽度差与长度差之和,当和小于站立框宽度的三分之一,且tracklet的边框与所述站立框重合度大于0.3时,判断所述站立框与该tracklet最佳匹配。
- [0033] 该方法还包括步骤：
- [0034] 5) 对跟踪获得的站立次数进行计数。
- [0035] 与现有技术相比,本发明具有以下有益效果：
- [0036] 1) 本发明对采集的样本进行标注,将站立行为细分为老师和学生、低年级和高年级,有效提高了模型的可靠性。
- [0037] 2) 本发明采用基于深度学习的目标检测算法,从课堂的真实视频中抽取大量的数据(如两万样本),能够在复杂的教室环境中取得较好的检测结果。
- [0038] 3) 本发明对高年级和低年级样本进行分别训练,获得高年级站立检测模型和低年级站立检测模型,有效地解决了高年级站立与低年级站立之间存在较大差别的问题。经过测试,在高年级和低年级上都能够达到检全率90%,准确率80%。
- [0039] 4) 本发明采用tracking算法对站立动作进行跟踪,能串联不同帧的同一站立行

为,因此能得到真实站立次数的数据,为进一步分析评估提供依据。

附图说明

- [0040] 图1为现有的一种学生跟踪定位方法的流程示意图;
- [0041] 图2为本发明的流程示意图;
- [0042] 图3为低年级站立行为示意图;
- [0043] 图4为高年级站立行为示意图;
- [0044] 图5为本发明站立动作的跟踪流程示意图。

具体实施方式

[0045] 下面结合附图和具体实施例对本发明进行详细说明。本实施例以本发明技术方案为前提进行实施,给出了详细的实施方式和具体的操作过程,但本发明的保护范围不限于下述的实施例。

[0046] 如图2所示,本发明提供一种基于深度学习的站立检测方法,该方法包括以下步骤:

[0047] 1) 收集样本

[0048] 样本按照PASCAL VOC数据集的格式制作,PASCAL VOC为图像识别和分类提供了一整套标准化的优秀的数据集,使用的标注工具为LabelImg。本发明方法主要关注的是学生的站立,而老师在教室中的走动也会被判定为站立,因此本发明将教室中的站立行为细分为学生的站立和老师的站立,在制作样本的时候也是分别标注出了学生的站立和老师的站立。

[0049] 本实施例中,采用的样本个数为2万个。各样本包括样本图片和对应的标注文件,标注文件的信息包括站立人员类型,从视频中截取的包含站立的帧存放在JPEGImages文件夹下,而对应的标注文件则存放在Annotations下。所述站立人员类型包括高年级学生、低年级学生和老师。高低年级划分的标准是1-3年级为低年级,四年级及以上的年级划分为高年级。

[0050] 2) 建立站立检测模型

[0051] 如图3和图4所示分别为低年级和高年级学生的站立行为示意图,由图中可知,低年级学生的站立与坐着时候的相差不大,而高年级学生站立与坐的时候相差很大的。因此如果将两种情况并在一个模型中会出现冲突,若要在低年级上把站立基本都检测出来,在高年级的视频上就会增加许多的误检。本发明将这两种情况进行划分并分别处理。

[0052] 建立站立检测模型具体为:

[0053] 201) 使用基于ResNet-101的R-FCN模型,利用所有将高低年级的样本一同训练一基础站立模型;

[0054] 202) 利用带有高年级学生标注的样本和带有低年级学生标注的样本分别在所述基础站立模型的基础上进行进一步训练,获得高年级站立检测模型和低年级站立检测模型。本实施例中,利用fine-tuning来得到针对高年级的模型和针对低年级的模型。

[0055] 训练所用的框架是caffe,训练时迭代次数为30000次,训练参数如下:

[0056] base_lr:0.001

[0057] lr_policy:"step"

[0058] gamma:0.1

[0059] stepsize:10000

[0060] display:20

[0061] momentum:0.9

[0062] weight_decay:0.0005

[0063] 利用训练后的站立检测模型即对待测视频进行站立检测。

[0064] 在某些实施例中,本方法还包括步骤:

[0065] 4) 根据前一帧的站立检测结果和当前的站立检测结果进行站立的跟踪。

[0066] 而在进行站立跟踪时还需要进行镜头变化的判断。这是因为在整个教室的监控视频中,镜头会出现转动以及拉近拉远等操作,会使得不同帧中的同一个站立的位置发生很大的变化,从而无法有效的进行匹配。

[0067] 判断是否发生镜头的变化的算法具体:将当前帧和前一帧转化为灰度图,如果两帧之间灰度值的差值大于第一阈值thres0,判定此像素点发生变化,当发生变化的像素点占总像素点的比例超出第二阈值thres1,判断镜头发生了变化。否则判断镜头没有发生变化。本实施例中所使用的thres0为20,thres1为0.2(20%)。

[0068] 如图5所示,站立跟踪的具体流程为:

[0069] 401) 获取第一个图像帧及检测到的站立框坐标,各站立框对应建立有一tracklet数组,且状态初始化为ALIVE,tracklet用来记录跟踪的信息。

[0070] 402) 获取下一个图像帧,判断镜头是否发生变换,若是,则将所有tracklet数组的状态改为DEAD,重新建立新的tracklet数组,返回步骤402),若否,则执行步骤403)。

[0071] 403) 遍历当前图像帧检测到的所有站立框,利用跟踪算法为每一站立框选择最佳匹配的一个tracklet数组。

[0072] 选择最佳匹配的一个tracklet数组的具体过程为:

[0073] 首先找到与每个站立框最近的tracklet,并计算出每个站立框和最近的tracklet边框之间宽度差与长度差之和,当和小于站立框宽度的三分之一,且tracklet的边框与此站立框重合度大于0.3时,判断站立框与当前的tracklet之间匹配,否则此站立框没有相匹配的tracklet。若此tracklet已经被匹配了,则更新站立框的最近tracklet,并标记此站立框仍需判断是否与最近的tracklet匹配。

[0074] 404) 对于在当前图像帧下未被匹配的tracklet数组,判断其状态是否为ALIVE,若是,则状态修改为WAIT,因为有一帧跟踪信息消失了有可能是这一帧的站立行为没有检测到,若否,则状态修改为DEAD,标志一个站立行为跟踪的结束,返回步骤402),直至处理完成所有图像帧。

[0075] 以上详细描述了本发明的较佳具体实施例。应当理解,本领域的普通技术人员无需创造性劳动就可以根据本发明的构思作出诸多修改和变化。因此,凡本技术领域中技术人员依本发明的构思在现有技术的基础上通过逻辑分析、推理或者有限的实验可以得到的技术方案,皆应在由权利要求书所确定的保护范围内。

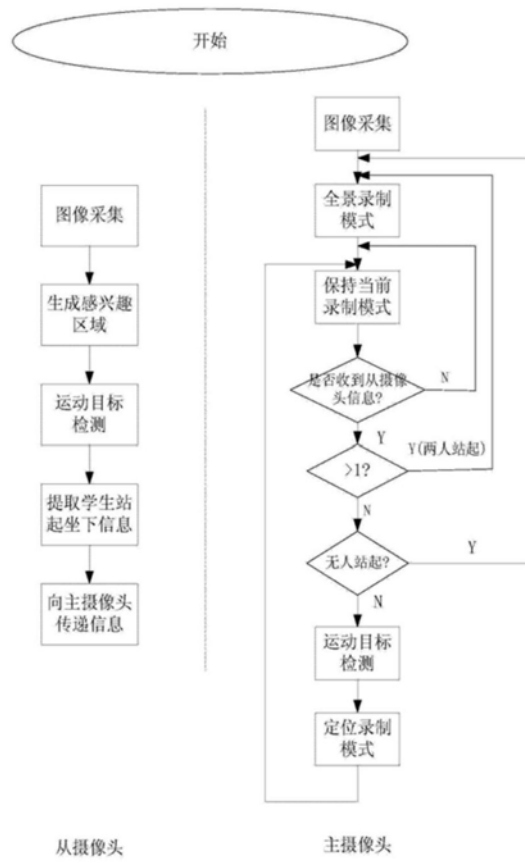


图1

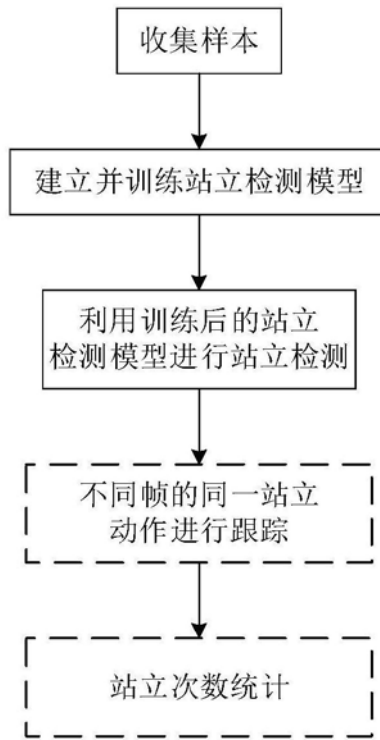


图2



图3



图4

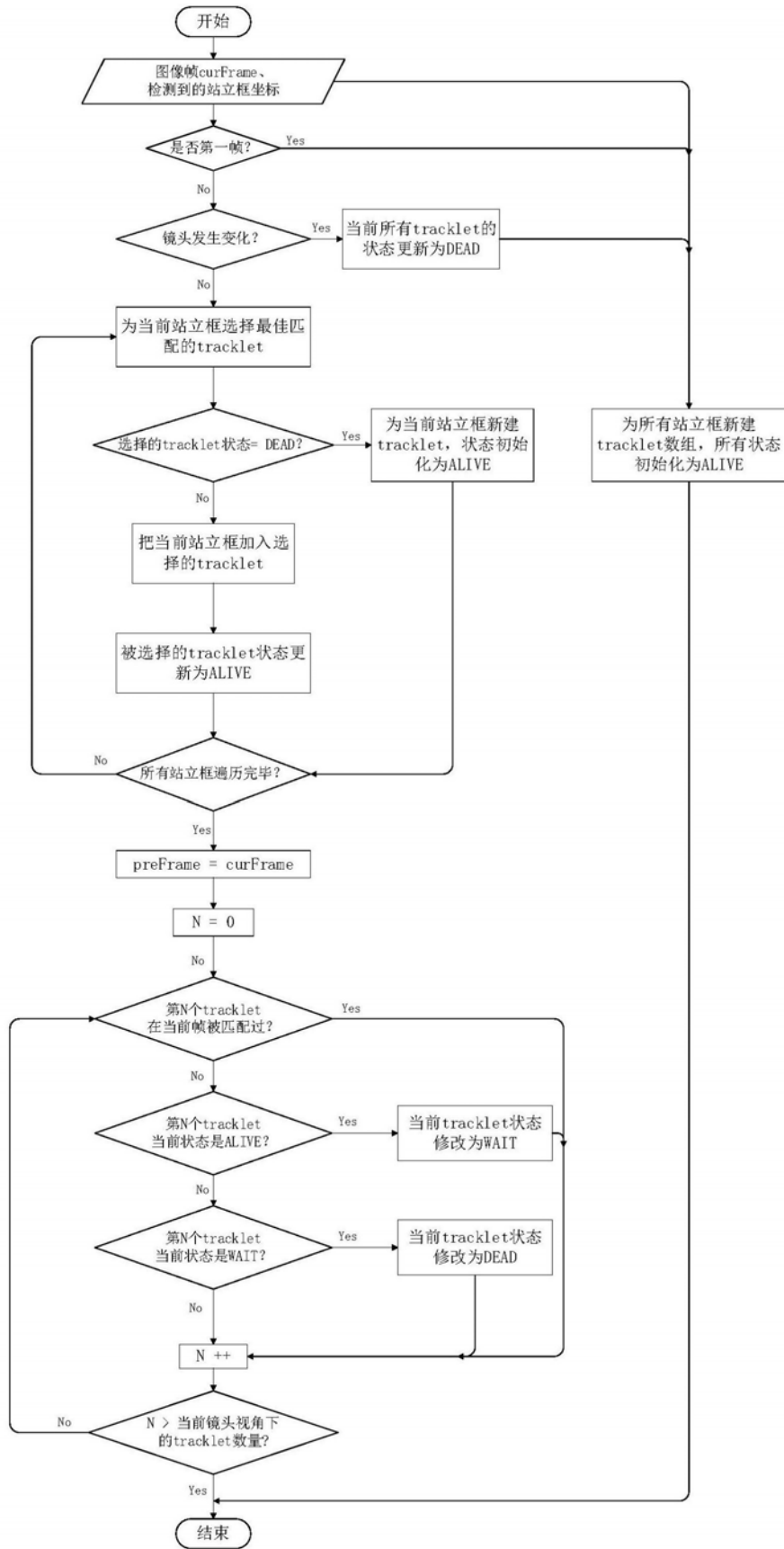


图5