



(12) 发明专利

(10) 授权公告号 CN 111507283 B

(45) 授权公告日 2021. 11. 30

(21) 申请号 202010317276.8

CN 108629766 A, 2018.10.09

(22) 申请日 2020.04.21

CN 109740446 A, 2019.05.10

(65) 同一申请的已公布的文献号

CN 108090857 A, 2018.05.29

申请公布号 CN 111507283 A

CN 110287792 A, 2019.09.27

(43) 申请公布日 2020.08.07

US 2018060649 A1, 2018.03.01

(73) 专利权人 浙江蓝鸽科技有限公司

CN 110781843 A, 2020.02.11

地址 314006 浙江省嘉兴市南湖区南溪东路1968号

左国才 等. 基于 CNN 人脸识别模型的大学生课堂行为分析研究.《智能计算机与应用》. 2019, 第9卷(第6期),

(72) 发明人 张新华 李琳璐 司俊 陈诚

Samuel Lukas 等. Student Attendance System in Classroom Using Face

(74) 专利代理机构 上海一平知识产权代理有限公司 31266

Recognition Technique.《2016 International Conference on Information and Communication Technology Convergence (ICTC)》. 2016,

代理人 成春荣 竺云

贾鹏宇 等. 基于人工智能视频处理的课堂学生状态分析.《现代教育技术》. 2019, 第29卷(第12期),

(51) Int. Cl.

G06K 9/00 (2006.01)

(56) 对比文件

CN 110781843 A, 2020.02.11

CN 107609517 A, 2018.01.19

审查员 张露

权利要求书3页 说明书10页 附图2页

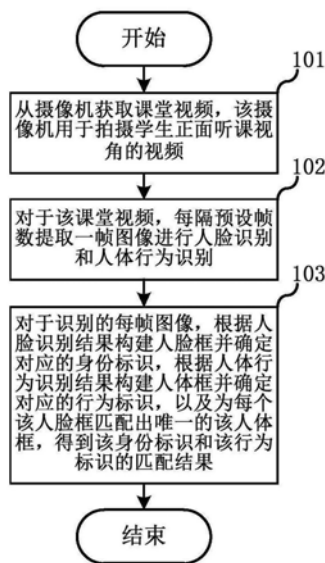
(54) 发明名称

基于课堂场景的学生行为识别方法及系统

(57) 摘要

本申请涉及教学领域,公开了一种基于课堂场景的学生行为识别方法及系统。该方法包括:从摄像机获取课堂视频,该摄像机用于拍摄学生正面听课视角的视频;对于该课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别;对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个该人脸框匹配出唯一的该人体框,得到该身份标识和该行为标识的匹配结果。本申请的实施方式可以准确地得到课堂上每个学生的身份标识与人体行为标识的匹配结果,实现监控课堂上每个学生的实时行为的效果。

CN 111507283 B



1. 一种基于课堂场景的学生行为识别方法,其特征在于,包括:

从摄像机获取课堂视频,所述摄像机用于拍摄学生正面听课视角的视频;

对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别;对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个所述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果;

所述根据人体行为识别结果构建人体框,进一步包括子步骤:根据人体动作识别结果中的对应每个人体的头顶、左肩、右肩、左脚、右脚的坐标信息,以构建对应每个人体的所述人体框;

其中,所述为每个所述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果,进一步包括子步骤:如果某个人体框内有多个人脸框,计算该人体框的上边沿的中点与该多个人脸框中每个人脸框的上边沿的中点之间的距离,选择距离最短的人脸框为该人体框的唯一匹配人脸框,得到对应该唯一匹配人脸框的身份标识和对应该人体框的行为标识的匹配结果。

2. 如权利要求1所述的基于课堂场景的学生行为识别方法,其特征在于,所述对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别,进一步包括:

对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别得到对应所述每帧图像的所述人脸识别结果,所述人脸识别结果包括对应每张人脸的身份标识和眼睛、嘴的坐标信息;

对于所述课堂视频,每间隔预设帧数提取一帧图像进行人体行为识别得到对应所述每帧图像的所述人体行为识别结果,所述人体行为识别结果包括对应每个人体的行为标识和头顶、左肩、右肩、左脚、右脚的坐标信息。

3. 如权利要求2所述的基于课堂场景的学生行为识别方法,其特征在于,所述对于所述课堂视频,每间隔预设帧数提取一帧图像进行人体行为识别得到对应所述每帧图像的所述人体行为识别结果,进一步包括:

建立包含学生行为标签的课堂视频的训练样本集,所述学生行为标签包括坐、站立、举手和趴桌子;

用所述训练样本集对预设检测模型进行训练,该预设检测模型为OpenPose、Detectron或AlphaPose模型;

使用训练后的预设检测模型对所述每帧图像进行人体行为识别得到对应每个人体的行为标识和头顶、左肩、右肩、左脚、右脚的坐标信息。

4. 如权利要求2所述的基于课堂场景的学生行为识别方法,其特征在于,所述根据人脸识别结果构建人脸框并确定对应的身份标识,进一步包括:

根据所述人脸识别结果中的对应每张人脸的身份标识,为每个人脸框确定对应的身份标识;

确定人体框对应的行为标识,进一步包括:

根据所述人体动作识别结果中的对应每个人体的行为标识,为每个人体框确定对应的行为标识。

5. 如权利要求4所述的基于课堂场景的学生行为识别方法,其特征在于,所述为每个所

述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果,进一步还包括:

如果某个人体框中只有一个人脸框,则得到对应该人脸框的身份标识和对应该人体框的行为标识的匹配结果。

6.如权利要求1所述的基于课堂场景的学生行为识别方法,其特征在于,所述对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个所述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果之后,还包括:

周期性地获取预设时间段内的多帧图像的多组匹配结果;

根据所述多组匹配结果,为每个身份标识确定占比最高的行为标识作为该身份标识在该预设时间段的行为标识。

7.如权利要求1-6中任意一项所述的基于课堂场景的学生行为识别方法,其特征在于,所述对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别之前,还包括:

对于所述课堂视频中的每帧图像,将该图像灰度化;

采用3x3的拉普拉斯算子对灰度化后的图像进行处理;

计算处理后图像输出的方差,如果该方差小于预设阈值,则判定该图像为模糊图像并将其从所述课堂视频中删除以更新所述课堂视频;

所述从摄像机获取课堂视频之前,还包括:

将所述摄像机布置在教室讲台侧的中间位置,并调整摄像机与距其最远的学生的距离以满足该学生的人脸图像像素大于 $m*m$,其中 $m*m$ 为所述人脸识别的识别精度。

8.一种基于课堂场景的学生行为识别系统,其特征在于,包括:

获取模块,用于从摄像机获取课堂视频,所述摄像机用于拍摄学生正面听课视角的视频;

识别模块,用于对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别,对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个所述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果,其中,所述识别模块还用于如果某个人体框内有多个人脸框,计算该人体框的上边沿的中点与该多个人脸框中每个人脸框的上边沿的中点之间的距离,选择距离最短的人脸框为该人体框的唯一匹配人脸框,得到对应该唯一匹配人脸框的身份标识和对应该人体框的行为标识的匹配结果;

所述识别模块还用于根据人体动作识别结果中的对应每个人体的头顶、左肩、右肩、左脚、右脚的坐标信息,以构建对应每个人体的所述人体框。

9.如权利要求8所述的基于课堂场景的学生行为识别系统,其特征在于,所述识别模块还用于如果某个人体框中只有一个人脸框,则得到对应该人脸框的身份标识和对应该人体框的行为标识的匹配结果。

10.一种基于课堂场景的学生行为识别系统,其特征在于,包括:

存储器,用于存储计算机可执行指令;以及,

处理器,用于在执行所述计算机可执行指令时实现如权利要求1至7中任意一项所述的

方法中的步骤。

11. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有计算机可执行指令,所述计算机可执行指令被处理器执行时实现如权利要求1至7中任意一项所述的方法中的步骤。

基于课堂场景的学生行为识别方法及系统

技术领域

[0001] 本申请涉及教学领域,特别涉及基于课堂场景的学生行为识别技术。

背景技术

[0002] 人体行为识别技术可以充分利用设备资源,从已有的课堂录播视频中提取学生听课状况相关的数据信息,辅助老师进行教学管理、教学决策。

[0003] 但是,由于课堂场景学生行为识别属于复杂场景下多人姿态行为识别问题,目前通用的识别算法准确率低,使得人体行为识别结果与身份识别结果匹配准确率较低,大都只能实现整个班级学生行为数据整体粗略地统计汇总,而不能准确地将课堂人体行为细化精确到对应的每一个身份明确的学生。

发明内容

[0004] 本申请的目的在于提供一种基于课堂场景的学生行为识别方法及系统,可以准确地得到课堂上每个学生的身份标识与人体行为标识的匹配结果,实现监控课堂上每个学生的实时行为的效果。

[0005] 本申请公开了一种基于课堂场景的学生行为识别方法,包括:

[0006] 从摄像机获取课堂视频,所述摄像机用于拍摄学生正面听课视角的视频;

[0007] 对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别;

[0008] 对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个所述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果。

[0009] 在一个优选例中,所述对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别,进一步包括:

[0010] 对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别得到对应所述每帧图像的所述人脸识别结果,所述人脸识别结果包括对应每张人脸的身份标识和眼睛、嘴的坐标信息;

[0011] 对于所述课堂视频,每间隔预设帧数提取一帧图像进行人体行为识别得到对应所述每帧图像的所述人体行为识别结果,所述人体行为识别结果包括对应每个人体的行为标识和头顶、左肩、右肩、左脚、右脚的坐标信息。

[0012] 在一个优选例中,所述对于所述课堂视频,每间隔预设帧数提取一帧图像进行人体行为识别得到对应所述每帧图像的所述人体行为识别结果,进一步包括:

[0013] 建立包含学生行为标签的课堂视频的训练样本集,所述学生行为标签包括坐、站立、举手和趴桌子;

[0014] 用所述训练样本集对预设检测模型进行训练,该预设检测模型为OpenPose、Detectron或AlphaPose模型;

[0015] 使用训练后的预设检测模型对所述每帧图像进行人体行为识别得到对应每个人体的行为标识和头顶、左肩、右肩、左脚、右脚的坐标信息。

[0016] 在一个优选例中,所述根据人脸识别结果构建人脸框并确定对应的身份标识,进一步包括:

[0017] 根据所述人脸识别结果中的对应每张人脸的眼睛、嘴的坐标信息计算人脸框的上、下、左、右边界的坐标信息,以构建对应每张人脸的所述人脸框;

[0018] 根据所述人脸识别结果中的对应每张人脸的身份标识,为每个人脸框确定对应的身份标识;

[0019] 所述根据人体行为识别结果构建人体框并确定对应的行为标识,进一步包括:

[0020] 根据所述人体动作识别结果中的对应每个人体的头顶、左肩、右肩、左脚、右脚的坐标信息,以构建对应每个人体的所述人体框;

[0021] 根据所述人体动作识别结果中的对应每个人体的行为标识,为每个人体框确定对应的行为标识。

[0022] 在一个优选例中,所述为每个所述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果,进一步包括:

[0023] 如果某个人体框中只有一个人脸框,则得到对应该人脸框的身份标识和对应该人体框的行为标识的匹配结果;

[0024] 如果某个人体框内有多个人脸框,计算该人体框的上边沿的中点与该多个人脸框中每个人脸框的上边沿的中点之间的距离,选择距离最短的人脸框为该人体框的唯一匹配人脸框,得到对应该唯一匹配人脸框的身份标识和对应该人体框的行为标识的匹配结果。

[0025] 在一个优选例中,所述对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个所述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果之后,还包括:

[0026] 周期性地获取预设时间段内的多帧图像的多组匹配结果;

[0027] 根据所述多组匹配结果,为每个身份标识确定占比最高的行为标识作为该身份标识在该预设时间段的行为标识。

[0028] 在一个优选例中,所述对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别之前,还包括:

[0029] 对于所述课堂视频中的每帧图像,将该图像灰度化;

[0030] 采用3x3的拉普拉斯算子对灰度化后的图像进行处理;

[0031] 计算处理后图像输出的方差,如果该方差小于预设阈值,则判定该图像为模糊图像并将其从所述课堂视频中删除以更新所述课堂视频;

[0032] 在一个优选例中,所述从摄像机获取课堂视频之前,还包括:

[0033] 将所述摄像机布置在教室讲台侧的中间位置,并调整摄像机与距其最远的学生的距离以满足该学生的人脸图像像素大于 $m*m$,其中 $m*m$ 为所述人脸识别的识别精度。

[0034] 本申请还公开了一种基于课堂场景的学生行为识别系统包括:

[0035] 获取模块,用于从摄像机获取课堂视频,所述摄像机用于拍摄学生正面听课视角的视频;

[0036] 识别模块,用于对于所述课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别,对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个所述人脸框匹配出唯一的所述人体框,得到所述身份标识和所述行为标识的匹配结果。

[0037] 本申请还公开了一种基于课堂场景的学生行为识别系统包括:

[0038] 存储器,用于存储计算机可执行指令;以及,

[0039] 处理器,用于在执行所述计算机可执行指令时实现如前文描述的方法中的步骤。

[0040] 本申请还公开了一种计算机可读存储介质所述计算机可读存储介质中存储有计算机可执行指令,所述计算机可执行指令被处理器执行时实现如前文描述的方法中的步骤。

[0041] 本申请实施方式中,与现有技术相比,至少包括以下优点和效果:

[0042] 通过获取学生正面听课视角的课堂视频,对该课堂视频逐帧或间隔提取帧进行人脸识别和人体行为识别,并且对于识别的每帧图像分别根据人脸识别结果、人体行为识别结果构建对应的人脸框及相应的身份标识、人体框及相应的行为标识,并为每个所述人脸框匹配出唯一的所述人体框,最终得到的人体行为标识与身份标识精准地一一对应的匹配结果,进而可以监控课堂上每个学生的实时行为,以便更好地辅助授课教师进行课堂管理和教学决策。例如,本申请的实施方式可以应用到课堂考勤,能够准确地判断学生缺席、迟到、早退以实现无感考勤,与学生签到或老师点名的考勤模式相比,节省教学时间。

[0043] 进一步地,考虑到人脸框和人体框可能存在多种位置关系的情况,在“为每个人脸框匹配出唯一的人体框”时,如果某个人体框中只有一个人脸框,则该人体框就是该人脸框的唯一匹配人体框,如果某个人体框内有多个人脸框,计算该人体框的上边沿的中点与该多个人脸框中每个人脸框的上边沿的中点之间的距离,选择距离最短的人脸框为该人体框的唯一匹配人脸框,匹配准确度较高,进一步提高了人体行为标识与身份标识的匹配结果的准确性。

[0044] 进一步地,在对课堂视频进行人脸识别和人体行为识别之前,对课堂视频的每帧图像进行预处理,将场景杂乱、运动模糊的图像帧删除,提高了后续识别、匹配过程的效率和识别、匹配结果的准确性。

[0045] 进一步地,在对于每帧图像得到对应的人体行为标识与身份标识精准地一一对应的匹配结果之后,周期性地获取预设时间段内连续多帧图像的多组匹配结果,根据该多组匹配结果,为每个身份标识确定占比最高的行为标识作为该身份标识在该预设时间段的的行为标识,有效地避免了单帧图像因噪声或局部遮挡等可能会造成的识别结果误差,进一步提高了人体行为标识与身份标识的匹配结果的准确性。

[0046] 进一步地,预先设置摄像机的安装位置及角度以保证尽可能地拍摄到学生正脸,以及调整摄像机与最远端学生人脸的距离以保证所有学生人脸图像的像素满足后续识别精度要求,进一步提高人体行为标识与身份标识的匹配结果的准确性。

[0047] 进一步地,在人体行为识别以及根据人体行为识别结果构建人体框时,设置自定义的学生行为标签,基于带有标签的训练样本集训练模型,并且根据训练好的模型进行每帧图像中人体行为的识别和关键点坐标信息的确认,进而可以基于得到的关键点坐标信息构建人体框以及对应该人体框的行为标识,为后续的识别过程和匹配过程提供有效基础。

[0048] 本申请的说明书中记载了大量的技术特征,分布在各个技术方案中,如果要罗列出本申请所有可能的技术特征的组合(即技术方案)的话,会使得说明书过于冗长。为了避免这个问题,本申请上述发明内容中公开的各个技术特征、在下文各个实施方式和例子中公开的各项技术特征、以及附图中公开的各个技术特征,都可以自由地互相组合,从而构成各种新的技术方案(这些技术方案均因视为在本说明书中已经记载),除非这种技术特征的组合在技术上是不可行的。例如,在一个例子中公开了特征A+B+C,在另一个例子中公开了特征A+B+D+E,而特征C和D是起到相同作用的等同技术手段,技术上只要择一使用即可,不可能同时采用,特征E技术上可以与特征C相组合,则,A+B+C+D的方案因技术不可行而应当不被视为已经记载,而A+B+C+E的方案应当视为已经被记载。

附图说明

[0049] 图1是根据本申请第一实施方式的基于课堂场景的学生行为识别方法流程图;

[0050] 图2是本申请的摄像机布置的一个示例示意图;

[0051] 图3是根据本申请的人脸框和人体框示例位置关系示意图;

[0052] 图4是根据本申请第二实施方式的基于课堂场景的学生行为识别系统结构示意图。

具体实施方式

[0053] 在以下的叙述中,为了使读者更好地理解本申请而提出了许多技术细节。但是,本领域的普通技术人员可以理解,即使没有这些技术细节和基于以下各实施方式的种种变化和修改,也可以实现本申请所要求保护的技术方案。

[0054] 为使本申请的目的、技术方案和优点更加清楚,下面将结合附图对本申请的实施方式作进一步地详细描述。

[0055] 本申请的第一实施方式涉及一种基于课堂场景的学生行为识别方法,其流程如图1所示,该方法包括以下步骤:

[0056] 在步骤101中,从摄像机获取课堂视频,该摄像机用于拍摄学生正面听课视角的视频。

[0057] 可选地,该步骤101之前,还可以包括以下步骤:

[0058] 将该摄像机布置在教室讲台侧位置,并调整摄像机与距其最远的学生的距离以满足该学生的人脸图像像素大于 $m*m$,其中 $m*m$ 为该人脸识别的识别精度。

[0059] 如图2所示为该摄像机布置的一个示例示意图,图中摄像机离两侧墙的水平距离分别为 a 和 b 。优选地,设定 $a=b$ 以保证摄像机尽可能地拍摄到学生的正脸,保证识别结果的准确性。可选地, a 也可以不等于 b ,只要能够拍摄到学生正面听课视角的设置都在本申请的保护范围内。

[0060] 可选地,步骤101之前,还包括以下步骤:

[0061] 预先构建课程表库和学生档案库。

[0062] 其中,该课程表库中课程表的属性信息包括教室信息(例如教室的唯一编码)、上课时间、下课时间、上课班级信息(例如班级的唯一编码)等。其中,学生档案库中的学生档

案属性信息包括学生信息(例如包括学生姓名及其对应的学籍和人脸图像等)、所属班级信息等。

[0063] 可选地,步骤101之前,还包括以下步骤:

[0064] 首先,响应于某一教室的上课时间,从该课程表库中获取该教室的该上课时间对应的课程表;然后根据所获取的课程表确定当前上课班级信息,并根据当前上课班级信息从所述学生档案库中获取对应的学生信息(包括学生姓名及其对应的学籍和人脸图像等)。这为后续的学生人脸识别和学生行为识别提供基础。

[0065] 之后,进入步骤102,对于该课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别。

[0066] 其中,步骤102中的预设帧数可以为零或非零整数。在一个实施例中,该预设帧数为0,则该步骤102可以实现为:对于该课堂视频逐帧进行人脸识别和人体行为识别。

[0067] 在另一个实施例中,该预设帧数为非零整数,优选地,该预设帧数小于该摄像机的帧率。例如,对于目前通用摄像机拍摄的课堂视频,帧率为30帧/秒(每秒钟30帧图像),当该预设帧数为1时,即每间隔1帧提取一帧图像进行人脸识别和人体行为识别,换言之,从每秒钟内的30帧图像中提取15帧图像进行人脸识别和人体行为识别,这可以在保证准确率的前提下,节省系统资源。

[0068] 可选地,该步骤102之前,还可以包括以下步骤A和B:

[0069] 在步骤A中,对于该课堂视频中的每帧图像,将该图像灰度化;之后采用3x3的拉普拉斯算子对灰度化后的图像进行处理;之后执行步骤B,计算处理后图像输出的方差,并且如果该方差小于预设阈值,则判定该图像为模糊图像并将其从该课堂视频中删除以更新该课堂视频。因为考虑到教师或学生在摄像机在学生区域走动时,可能会产生因人体运动而造成的部分图像模糊,该步骤A和B可以一定程度上检测并剔除模糊图像范畴的图像,可以有效提高后续学生行为识别的效率。其中该预设阈值可以根据需要进行设置。

[0070] 可选地,该步骤102可以进一步包括以下步骤①和②:

[0071] 在步骤①中,对于该课堂视频,每间隔预设帧数提取一帧图像进行人脸识别得到对应该每帧图像的该人脸识别结果,该人脸识别结果包括对应每张人脸的身份标识和眼睛、嘴的坐标信息;在步骤②中,对于该课堂视频,每间隔预设帧数提取一帧图像进行人体行为识别得到对应该每帧图像的该人体行为识别结果,该人体行为识别结果包括对应每个人体的行为标识和头顶、左肩、右肩、左脚、右脚的坐标信息。其中,步骤①和②可以按照先后顺序依次执行或者二者并行执行。

[0072] 需要说明的是,人脸识别算法为通用算法,本申请不对其进行详细的说明及限定。

[0073] 可选地,该步骤②还可以进一步包括以下步骤:

[0074] 首先,建立包含学生行为标签的课堂视频的训练样本集,该学生行为标签包括坐、站立、举手和趴桌子;之后,用该训练样本集对预设检测模型进行训练,该预设检测模型为OpenPose、Detectron或AlphaPose模型(此三种模型均为开源的算法模型);之后,使用训练后的预设检测模型对该每帧图像进行人体行为识别得到对应每个人体的行为标识和头顶、左肩、右肩、左脚、右脚的坐标信息。其中,优选地,该训练样本集中各标签的数据量不少于两万个。

[0075] 如下表1为各学生行为标签的自定义。例如,可以但不限于使用coco官方工具、

labelling、CVAT等打标签工具对该课堂视频进行人工打标签操作,再使用视频切片工具,把所打好标签的课堂视频裁剪成对应的每帧图像。

[0076] 表1

学生行为标签	定义
坐	1. 人坐在座位上; 2. 手不能举起,可放桌子上或自然下垂。
站立	1. 人站立起来; 2. 弯腰、半站立等不计入。
举手	1. 人坐在座位上; 2. 举起左、右、左右手。
趴桌子	1. 坐在座位上; 2. 单手、双手、或无手在桌子上; 3. 头枕在手上或头枕在桌子书本上等。

[0077] 之后,进入步骤103,对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个该人脸框匹配出唯一的该人体框,得到该身份标识和该行为标识的匹配结果。

[0079] 可选地,步骤103中“根据人脸识别结果构建人脸框并确定对应的身份标识”可以进一步实现为:首先根据该人脸识别结果中的对应每张人脸的眼睛、嘴的坐标信息计算人脸框的上、下、左、右边界的坐标信息,以构建对应每张人脸的该人脸框;然后根据该人脸识别结果中的对应每张人脸的身份标识,为每个人脸框确定对应的身份标识。在一个实施例中,如果根据该人脸识别结果中的对应每张人脸的左眼、右眼、嘴的中心坐标分别为 (x_1, y_1) 、 (x_2, y_2) 、 (x_3, y_3) ,则可以根据“左边界为 $x_1 - (x_1 + x_2) * \xi_1$ 、右边界为 $x_2 + (x_1 + x_2) * \xi_1$ 、上边界为 $(y_1 + y_2) / 2 - [y_3 - (y_1 + y_2) / 2] * \xi_2$ 、下边界为 $y_3 + [y_3 - (y_1 + y_2) / 2] * \xi_2$ ”来构建人体框。其中, ξ_1 、 ξ_2 为可调节的经验参数。

[0080] 可选地,该步骤103中“该根据人体行为识别结果构建人体框并确定对应的行为标识”可以进一步实现为:首先根据该人体动作识别结果中的对应每个人体的头顶、左肩、右肩、左脚、右脚的坐标信息,以构建对应每个人体的该人体框;然后根据该人体动作识别结果中的对应每个人体的行为标识,为每个人体框确定对应的行为标识。在一个实施例,如果根据该人体行为识别结果中识别出头顶、左肩、右肩、左脚、右脚的坐标信息,则可以根据“以头顶的上边沿作为人体框水平方向的上边沿、以脚的下边沿为人体框的下边沿、分别以左右肩的位置为人体框水平方向的左右边沿”来构建人体框。

[0081] 需要说明的是,如果人体的左脚、右脚被遮挡,可以根据头顶、左肩、右肩三点的坐标位置的比例关系来估算下边界的位置来构建人体框。

[0082] 考虑到在进行步骤103的人脸框和人体框匹配时人脸框和人体框可能存在多种位置关系,如图3所示,可能存在一个人体框内只包含一个人脸框,也可能存在一个人体框内包含多个人体框(全部或部分)。可选地,该步骤103中“为每个该人脸框匹配出唯一的该人体框,得到该身份标识和该行为标识的匹配结果”进一步实现为:如果某个人体框中只有一个人脸框,则得到对应该人脸框的身份标识和对应该人体框的行为标识的匹配结果;如果

某个人体框内有多个人脸框,计算该人体框的上边沿的中点M与该多个人脸框中每个人脸框的上边沿的中点N之间的距离,选择距离最短的人脸框为该人体框的唯一匹配人脸框,得到对应该唯一匹配人脸框的身份标识和对应该人体框的行为标识的匹配结果。

[0083] 可选地,步骤102时,还包括以下步骤:

[0084] 为提取的每帧图像从该课堂视频中获取对应的时间点。

[0085] 可选地,该步骤103之后,还包括以下步骤:

[0086] 周期性地从步骤103得到的匹配结果中获取预设时间段t内的多帧图像的多组匹配结果,并且根据该多组匹配结果,为每个身份标识确定占比最高的行为标识作为该身份标识在该预设时间段t的行为标识。这一可选步骤是考虑到单帧图像因噪声或局部遮挡等可能会造成的识别结果误差,取预设时间段内对连续多帧图像的判断结果进行分析,进而从多帧图像输出一个可能性最大的结果,从而达到剔除错误数据和冗余数据的目的。

[0087] 例如,该课堂视频的帧率为25帧/秒,并且在对于该课堂视频逐帧进行人脸识别和人体行为识别时,取 $t=2$ 秒,即获取2秒内的连续50帧图像的各学生的课堂行为的识别结果,将每个学生行为数量统计如下表2,那么例如,在该t秒内张三的动作行为标识为起立,因此剔除了“坐下”、“举手”的误判结果,同时也将50组结果数据仅保留一组数据进行存储,剔除了其余的冗余数据。

[0088] 表2

学生姓名 \ 动作	起立	坐下	举手	趴桌	输出动作
	帧数	帧数	帧数	子帧数	行为结果
[0089] 张三	43	4	1	0	起立
李四	2	48	0	0	坐下
王五	0	5	10	35	趴桌子

[0090] 本申请的第二实施方式涉及一种基于课堂场景的学生行为识别系统,其结构如图4所示,该基于课堂场景的学生行为识别系统包括获取模块和识别模块。

[0091] 具体的,该获取模块用于从摄像机获取课堂视频,该摄像机用于拍摄学生正面听课视角的视频。

[0092] 在实际应用中,本实施方式的学生行为识别系统可以利用教室中已安装的摄像机进行连接实施,也可以将摄像机作为该系统的一个组成部分同时实施。如果是后者,该摄像机在安装时,需要被布置在教室讲台侧位置,并调整摄像机与距其最远的学生的距离以满足该学生的人脸图像像素大于 $m*m$,其中 $m*m$ 为该人脸识别的识别精度。

[0093] 如图2所示为该摄像机布置的一个示例示意图,图中摄像机离两侧墙的水平距离分别为a和b。优选地,设定 $a=b$ 以保证摄像机尽可能地拍摄到学生的正脸,保证识别结果的准确性。可选地,a也可以不等于b,只要能够拍摄到学生正面听课视角的设置都在本申请的保护范围内。

[0094] 可选地,该基于课堂场景的学生行为识别系统还包括构建模块和存储模块,该构建模块用于预先构建课程表库和学生档案库,该存储模块用于存储该课程表库和该学生档

案库。其中,该课程表库中课程表的属性信息包括教室信息(例如教室的唯一编码)、上课时间、下课时间、上课班级信息(例如班级的唯一编码)等。其中,学生档案库中的学生档案属性信息包括学生信息(例如包括学生姓名及其对应的学籍和人脸图像等)、所属班级信息等。

[0095] 可选地,该获取模块还用于响应于某一教室的上课时间,从该课程表库中所获取的该教室的该上课时间对应的课程表,并且根据获取课程表确定当前上课班级信息,并根据当前上课班级信息从所述学生档案库中获取对应的学生信息(包括姓名及其对应的学籍和人脸图像等)。这为后续的学生人脸识别和学生行为识别提供基础。

[0096] 该识别模块用于对于该课堂视频,每间隔预设帧数提取一帧图像进行人脸识别和人体行为识别,对于识别的每帧图像,根据人脸识别结果构建人脸框并确定对应的身份标识,根据人体行为识别结果构建人体框并确定对应的行为标识,以及为每个该人脸框匹配出唯一的该人体框,得到该身份标识和该行为标识的匹配结果。

[0097] 其中,该预设帧数可以为零或非零整数。在一个实施例中,该预设帧数为0,则该步骤102可以实现为:对于该课堂视频逐帧进行人脸识别和人体行为识别。

[0098] 在另一个实施例中,该预设帧数为非零整数,优选地,该预设帧数小于该摄像机的帧率。例如,对于目前通用摄像机拍摄的课堂视频,帧率为30帧/秒(每秒钟30帧图像),当该预设帧数为1时,即该识别模块每间隔1帧提取一帧图像进行人脸识别和人体行为识别,换言之,该识别模块从每秒钟内的30帧图像中提取15帧图像进行人脸识别和人体行为识别,这可以在保证准确率的前提下,节省系统资源。可选地,该基于课堂场景的学生行为识别系统还包括预处理模块,该预处理模块用于对于该课堂视频中的每帧图像,将该图像灰度化,采用3x3的拉普拉斯算子对灰度化后的图像进行处理,计算处理后图像输出的方差,并且如果该方差小于预设阈值,则判定该图像为模糊图像并将其从该课堂视频中删除以更新该课堂视频。

[0099] 该识别模块还用于对于该课堂视频,每间隔预设帧数提取一帧图像进行人脸识别得到对应该每帧图像的该人脸识别结果,该人脸识别结果包括对应每张人脸的身份标识和眼睛、嘴的坐标信息,以及对于该课堂视频,每间隔预设帧数提取一帧图像进行人体行为识别得到对应该每帧图像的该人体行为识别结果,该人体行为识别结果包括对应每个人体的行为标识和头顶、左肩、右肩、左脚、右脚的坐标信息。

[0100] 可选地,该识别模块还用于建立包含学生行为标签的课堂视频的训练样本集,该学生行为标签包括坐、站立、举手和趴桌子,用该训练样本集对预设检测模型进行训练,该预设检测模型为OpenPose、Detectron或AlphaPose模型,以及使用训练后的预设检测模型对该每帧图像进行人体行为识别得到对应每个人体的行为标识和头顶、左肩、右肩、左脚、右脚的坐标信息。其中,优选地,该训练样本集中各标签的数据量不少于两万个。

[0101] 如上表1为各学生行为标签的自定义。例如,可以但不限于使用coco官方工具、labelling、CVAT等打标签工具对该课堂视频进行人工打标签操作,再使用视频切片工具,把所打好标签的课堂视频裁剪成对应的每帧图像。

[0102] 可选地,该识别模块还用于根据该人脸识别结果中的对应每张人脸的眼睛、嘴的坐标信息计算人脸框的上、下、左、右边界的坐标信息,以构建对应每张人脸的该人脸框,并且根据该人脸识别结果中的对应每张人脸的身份标识,为每个人脸框确定对应的身份标

识。在一个实施例中,该识别模块还用于如果根据该人脸识别结果中的对应每张人脸的左眼、右眼、嘴的中心坐标分别为 (x_1, y_1) 、 (x_2, y_2) 、 (x_3, y_3) ,则可以根据“左边界为 $x_1 - (x_1 + x_2) * \xi_1$ 、右边界为 $x_2 + (x_1 + x_2) * \xi_1$ 、上边界为 $(y_1 + y_2) / 2 - [y_3 - (y_1 + y_2) / 2] * \xi_2$ 、下边界为 $y_3 + [y_3 - (y_1 + y_2) / 2] * \xi_2$ ”来构建人体框。其中, ξ_1 、 ξ_2 为可调节的经验参数。

[0103] 可选地,该识别模块还用于根据该人体动作识别结果中的对应每个人体的头顶、左肩、右肩、左脚、右脚的坐标信息,以构建对应每个人体的该人体框,并且根据该人体动作识别结果中的对应每个人体的行为标识,为每个人体框确定对应的行为标识。在一个实施例,该识别模块还用于如果根据该人体行为识别结果中识别出头顶、左肩、右肩、左脚、右脚的坐标位置,则可以根据“以头顶的上边沿作为人体框水平方向的上边沿、以脚的下边沿作为人体框的下边沿、分别以左右肩的位置为人体框水平方向的左右边沿”来构建人体框。

[0104] 需要说明的是,如果人体的左脚、右脚被遮挡,可以根据头顶、左肩、右肩三点的坐标位置的比例关系来估算下边界的位置来构建人体框。

[0105] 考虑到人脸框和人体框匹配时,人脸框和人体框可能存在多种位置关系,如图3所示,可能存在一个人体框内只包含一个人脸框,也可能存在一个人体框内包含多个人体框(全部或部分)。可选地,该识别模块还用于如果某个人体框中只有一个人脸框,则得到对应该人脸框的身份标识和对应该人体框的行为标识的匹配结果,以及如果某个人体框内有多个人脸框,计算该人体框的上边沿的中点M与该多个人脸框中每个人脸框的上边沿的中点N之间的距离,选择距离最短的人脸框为该人体框的唯一匹配人脸框,得到对应该唯一匹配人脸框的身份标识和对应该人体框的行为标识的匹配结果。

[0106] 可选地,该识别模块还用于为提取的每帧图像从该课堂视频中获取对应的时间点。

[0107] 可选地,该基于课堂场景的学生行为识别系统还包括后处理模块,该后处理模块用于周期性地从识别模块识别中获取预设时间段内的多帧图像的多组匹配结果,并且根据该多组匹配结果,为每个身份标识确定占比最高的行为标识作为该身份标识在该预设时间段的标识。

[0108] 例如,该课堂视频的帧率为25帧/秒,并且在对于该课堂视频逐帧进行人脸识别和人体行为识别时,取 $t=2$ 秒,即后处理模块获取2秒内的连续50帧图像的各学生的课堂行为的识别结果,将每个学生行为数量统计如上表2,那么例如,在该 t 秒内张三的动作行为标识为起立,因此剔除了“坐下”、“举手”的误判结果,同时也将50组结果数据仅保留一组数据进行存储,剔除了其余的冗余数据。

[0109] 第一实施方式是与本实施方式相对应的方法实施方式,第一实施方式中的技术细节可以应用于本实施方式,本实施方式中的技术细节也可以应用于第一实施方式。

[0110] 需要说明的是,本领域技术人员应当理解,上述基于课堂场景的学生行为识别系统的实施方式中所示的各模块的实现功能可参照前述基于课堂场景的学生行为识别方法的相关描述而理解。上述基于课堂场景的学生行为识别系统的实施方式中所示的各模块的功能可通过运行于处理器上的程序(可执行指令)而实现,也可通过具体的逻辑电路而实现。本申请实施例上述基于课堂场景的学生行为识别系统如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请实施例的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产

品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机、服务器、或者网络设备等)执行本申请各个实施例所述方法的全部或部分。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read Only Memory)、磁碟或者光盘等各种可以存储程序代码的介质。这样,本申请实施例不限制于任何特定的硬件和软件结合。

[0111] 相应地,本申请实施方式还提供一种计算机可读存储介质,其中存储有计算机可执行指令,该计算机可执行指令被处理器执行时实现本申请的各方法实施方式。计算机可读存储介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括但不限于,相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读存储介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0112] 此外,本申请实施方式还提供一种基于课堂场景的学生行为识别系统,其中包括用于存储计算机可执行指令的存储器,以及,处理器;该处理器用于在执行该存储器中的计算机可执行指令时实现上述各方法实施方式中的步骤。其中,该处理器可以是中央处理单元(Central Processing Unit,简称“CPU”),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,简称“DSP”)、专用集成电路(Application Specific Integrated Circuit,简称“ASIC”)等。前述的存储器可以是只读存储器(read-only memory,简称“ROM”)、随机存取存储器(random access memory,简称“RAM”)、快闪存储器(Flash)、硬盘或者固态硬盘等。本发明各实施方式所公开的方法的步骤可以直接体现为硬件处理器执行完成,或者用处理器中的硬件及软件模块组合执行完成。

[0113] 需要说明的是,在本专利的申请文件中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。本专利的申请文件中,如果提到根据某要素执行某行为,则是指至少根据该要素执行该行为的意思,其中包括了两种情况:仅根据该要素执行该行为、和根据该要素和其它要素执行该行为。多个、多次、多种等表达包括2个、2次、2种以及2个以上、2次以上、2种以上。

[0114] 在本申请提及的所有文献都被认为是整体性地包括在本申请的公开内容中,以便在必要时可以作为修改的依据。此外应理解,以上所述仅为本说明书的较佳实施例而已,并非用于限定本说明书的保护范围。凡在本说明书一个或多个实施例的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本说明书一个或多个实施例的保护范围之内。

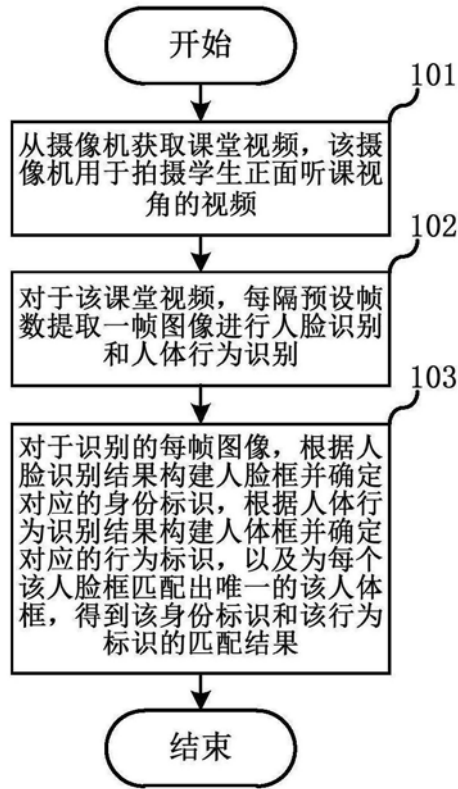


图1

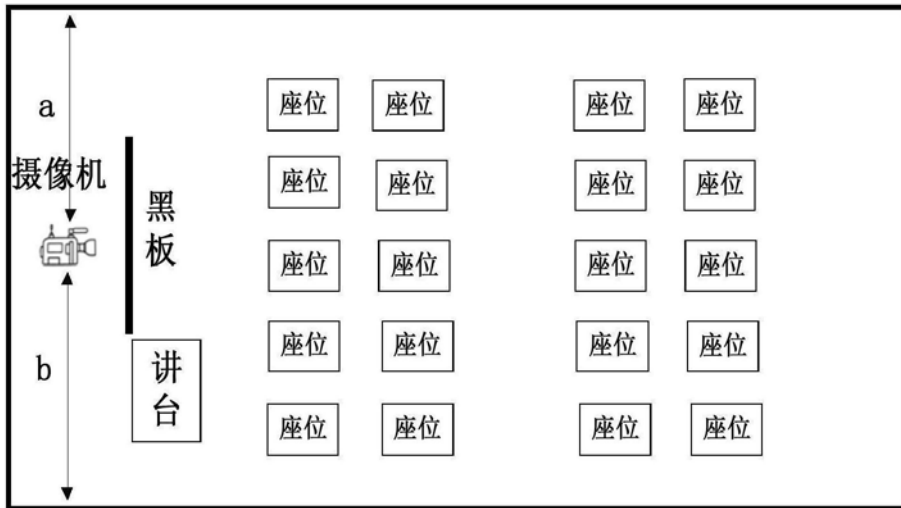


图2

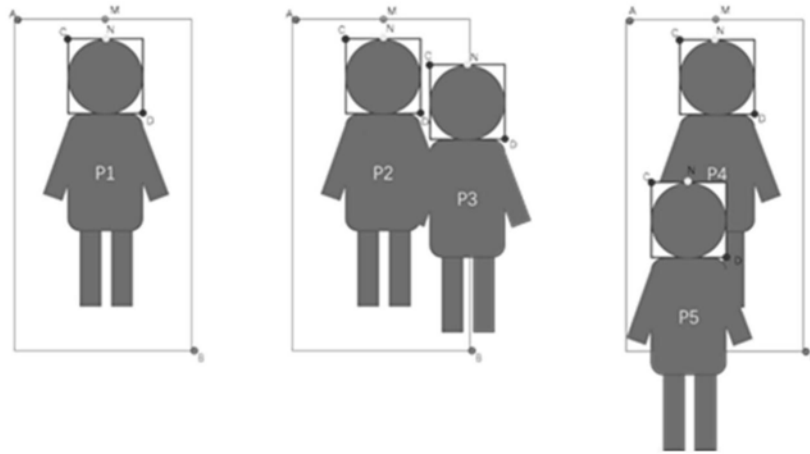


图3



图4