



(12) 发明专利

(10) 授权公告号 CN 109344746 B

(45) 授权公告日 2022. 02. 01

(21) 申请号 201811081090.6

G06V 10/74 (2022.01)

(22) 申请日 2018.09.17

G06V 10/774 (2022.01)

(65) 同一申请的已公布的文献号
申请公布号 CN 109344746 A

(56) 对比文件

CN 104376575 A, 2015.02.25

CN 104376575 A, 2015.02.25

(43) 申请公布日 2019.02.15

CN 106778502 A, 2017.05.31

(73) 专利权人 曜科智能科技(上海)有限公司
地址 201203 上海市浦东新区中国(上海)
自由贸易试验区张江路665号3层

CN 104902258 A, 2015.09.09

CN 106096553 A, 2016.11.09

审查员 李浩

(72) 发明人 曹志杰 吴旻烨

(74) 专利代理机构 上海光华专利事务所(普通
合伙) 31219

代理人 高彦

(51) Int. Cl.

G06V 40/10 (2022.01)

G06V 20/52 (2022.01)

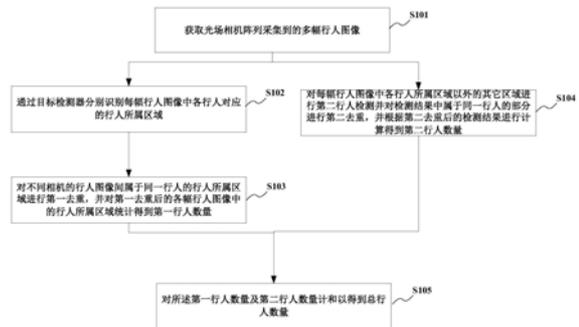
权利要求书4页 说明书12页 附图5页

(54) 发明名称

行人计数方法、系统、计算机设备和存储介质

(57) 摘要

本发明的行人计数方法、系统、计算机设备和存储介质,通过获取光场相机阵列采集到的多幅行人图像;通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域;对不同相机的行人图像间属于同一行人的行人所属区域进行第一去重,并对去重的行人所属区域统计得到第一行人数量;对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对进行第二去重,根据第二去重后的检测结果进行计算得到第二行人数量;对第一行人数量及第二行人数量计和以得到总行人数量;本发明基于光场相机阵列的优势出发,结合深度学习高性能的目标检测方法,避免漏检问题,亦配合深度学习的行人去重方式,得到更为精准的行人计数结果。



1. 一种行人计数方法,其特征在于,包括:

获取光场相机阵列采集到的多幅行人图像;其中,所述多幅行人图像包括:在所述光场相机阵列中每个相机的视野区域下采集到的一幅行人图像;

通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域;

将各幅所述行人图像的行人所属区域中的像素点的二维图像坐标转换为三维世界坐标;对不同相机的行人图像间属于同一行人的行人所属区域进行第一去重,并对第一去重后的各幅行人图像中的行人所属区域统计得到第一行人数量;

将各幅所述行人图像的行人所属区域以外的其它区域中的像素点的二维图像坐标转换为三维世界坐标;对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对检测结果中属于同一行人的部分进行第二去重,并根据第二去重后的检测结果进行计算得到第二行人数量;

所述对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测,包括:对应每幅行人图像,利用用于回归人头分布的补偿检测模型处理所述行人图像中的其它区域,得到第一人头分布密度图像;

将所述三维世界坐标转换为对应的二维地面坐标,以得到二维地面坐标对应的第二映射点,每幅行人图像所对应的各所述第二映射点构成采集该行人图像的相机的部分空间覆盖区域;通过容斥定理对光场相机阵列中各不同相机的部分空间覆盖区域间的重复覆盖区域加以消除,以保留各个相机的单独覆盖区域,并获取对应各单独覆盖区域的去重第二人头分布密度图像;

其中,所述第一人头分布密度图像和第二人头分布密度图像用于供积分以计算所述第二行人数量;

对所述第一行人数量及第二行人数量计和以得到总行人数量。

2. 根据权利要求1所述的行人计数方法,其特征在于,

所述目标检测器通过多层全卷积神经网络模型实现,其生成用于框选所述行人图像中每个行人的目标定界框作为所述行人所属区域;

所述目标检测器通过迁移学习训练为令其目标定界框仅用于框选识别为行人分类的部分图像;

所述迁移学习训练是以只有行人定界框的数据集对在多目标检测的数据集上训练好的所述多层全卷积神经网络模型进行的训练。

3. 根据权利要求2所述的行人计数方法,其特征在于,

所述迁移学习训练的具体方法基于多层全卷积神经网络模型,其网络结构不变,仅将多种类的数据集变化为单种类,使得只需判断一个目标定界框中内容是人还是背景;提高分类精确度的分类损失函数为:

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \quad \text{where} \quad \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$$

其中, c_i^p 代表对这个定界框进行对类别p的打分,打分越高,说明目标定界框内容越可能属于类别p。

4. 根据权利要求1所述的行人计数方法,其特征在于,所述第一去重,包括:

将所述三维世界坐标转换为对应的二维地面坐标,以得到二维地面坐标对应的第一映射点;

以分别获取自两个行人图像中各一行人所属区域的第一映射点成对,判断每对第一映射点所对应的一对像素点分别所在的行人所属区域是否是在对应的两个相机的公共视野区域下采集到的;

若否,则将该对第一映射点分别归类于不同行人;

若是,则计算该对第一映射点之间的相似程度;

判断所述相似程度是否大于一预设阈值;

若是,则判断该对第一映射点归类于不同行人;

若否,则判断该对第一映射点疑似属于同一行人,并进行第三去重。

5. 根据权利要求4所述的行人计数方法,其特征在于,所述第三去重,包括:

将疑似属于同一行人的每对第一映射点对应的一对像素点分别所在的行人所属区域分别输入一孪生神经网络模型的两个神经网络中;

计算分别在所述两个神经网络最后一个全连接层输出的特征向量间的相似程度;

判断所述特征向量间的相似程度是否大于一预设阈值;

若是,则判断该两个行人所属区域属于同一行人,进而去重;

若否,则判断该两个行人所属区域属于不同行人。

6. 根据权利要求4或5所述的行人计数方法,其特征在于,所述相似程度的计算方式包括:余弦距离、欧氏距离、标准化欧氏距离、马氏距离、汉明距离、及曼哈顿距离中的任意一种。

7. 一种行人计数系统,其特征在于,包括:

通信单元,用于获取光场相机阵列采集到的多幅行人图像;其中,所述多幅行人图像包括:在所述光场相机阵列中每个相机的视野区域下采集到的一幅行人图像;

处理单元,用于通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域;将各幅所述行人图像的行人所属区域中的像素点的二维图像坐标转换为三维世界坐标;对不同相机的行人图像间属于同一行人的行人所属区域进行第一去重,并对第一去重后的各幅行人图像中的行人所属区域统计得到第一行人数量;

处理单元,还用于将各幅所述行人图像的行人所属区域以外的其它区域中的像素点的二维图像坐标转换为三维世界坐标;对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对检测结果中属于同一行人的部分进行第二去重,并根据第二去重后的检测结果进行计算得到第二行人数量;

所述对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测,包括:对应每幅行人图像,利用用于回归人头分布的补偿检测模型处理所述行人图像中的其它区域,得到第一人分布密度图像;

将所述三维世界坐标转换为对应的二维地面坐标,以得到二维地面坐标对应的第二映射点,每幅行人图像所对应的各所述第二映射点构成采集该行人图像的相机的部分空间覆盖区域;通过容斥定理对光场相机阵列中各不同相机的部分空间覆盖区域间的重复覆盖区域加以消除,以保留各个相机的单独覆盖区域,并获取对应各单独覆盖区域的去重第二人头分布密度图像;

其中,所述第一人分布密度图像和第一人分布密度图像用于积分以计算所述第二行人数量;

对所述第一行人数量及第二行人数量计和以得到总行人数量。

8. 根据权利要求7所述的行人计数系统,其特征在于,

所述目标检测器通过多层全卷积神经网络模型实现,其生成用于框选所述行人图像中每个行人的目标定界框作为所述行人所属区域;

所述目标检测器通过迁移学习训练为令其目标定界框仅用于框选识别为行人分类的部分图像;

所述迁移学习训练是以只有行人定界框的数据集对在多目标检测的数据集上训练好的所述多层全卷积神经网络模型进行的训练。

9. 根据权利要求8所述的行人计数系统,其特征在于,

所述迁移学习训练的具体方法基于多层全卷积神经网络模型,其网络结构不变,仅将多种类的数据集变化为单种类,使得只需判断一个目标定界框中内容是人还是背景;提高分类精确度的分类损失函数为:

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \quad \text{where} \quad \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$$

其中, c_i^p 代表对这个定界框进行对类别p的打分,打分越高,说明目标定界框内容越可能属于类别p。

10. 根据权利要求7所述的行人计数系统,其特征在于,所述第一去重,包括:

将所述三维世界坐标转换为对应的二维地面坐标,以得到二维地面坐标对应的第一映射点;

以分别获取自两个行人图像中各一行人所属区域的第一映射点成对,判断每对第一映射点所对应的一对像素点分别所在的行人所属区域是否是在对应的两个相机的公共视野区域下采集到的;

若否,则将该对第一映射点分别归类于不同行人;

若是,则计算该对第一映射点之间的相似程度;

判断所述相似程度是否大于一预设阈值;

若是,则判断该对第一映射点归类于不同行人;

若否,则判断该对第一映射点疑似属于同一行人,并进行第三去重。

11. 根据权利要求10所述的行人计数系统,其特征在于,所述第三去重,包括:

将疑似属于同一行人的每对第一映射点对应的一对像素点分别所在的行人所属区域分别输入一孪生神经网络模型的两个神经网络中;

计算分别在所述两个神经网络最后一个全连接层输出的特征向量间的相似程度;

判断所述特征向量间的相似程度是否大于一预设阈值;

若是,则判断该两个行人所属区域属于同一行人,进而去重;

若否,则判断该两个行人所属区域属于不同行人。

12. 根据权利要求10或11所述的行人计数系统,其特征在于,所述相似程度的计算方式包括:余弦距离、欧氏距离、标准化欧氏距离、马氏距离、汉明距离、及曼哈顿距离中的任意

一种。

13. 一种计算机设备,其特征在于,所述计算机设备包括:通信器、处理器及存储器;

所属通信器,通信连接光场相机阵列;

所述存储器,存储有计算机程序;

所述处理器,用于运行所述计算机程序,以实现如权利要求1至6中任一项所述的行人计数方法。

14. 根据权利要求13所述的计算机设备,其特征在于,所述计算机设备与光场相机阵列之间通过通信网络连接。

15. 一种计算机存储介质,其特征在于,存储有计算机程序;所述计算机程序被运行时实现如权利要求1至6中任一项所述的行人计数方法。

行人计数方法、系统、计算机设备和存储介质

技术领域

[0001] 本发明涉及图像视觉技术领域,尤其涉及行人计数方法、系统、计算机设备和存储介质。

背景技术

[0002] 行人计数是计算机视觉的常用应用之一,通过计算机视觉的相关算法配合相关场景的相机,可以统计出该场景下的实时的人数状况,这可以为商场,景区等场所提供客流量信息来做进一步分析,也可以为街道提供行人数量,来分析安全隐患,或者为地铁、机场等提供可靠的客流量分布信息,以便有关部门合理配置交通工具的班次等等。

[0003] 近年来,关于行人计数的相关技术,有用深度学习的方法的,也有用传统的方法的,用深度学习的方法的,可以分为两种方式,一种是检测方式,通过检测出来场景中的行人,来统计行人数量,缺点在于遮挡会导致很多人检测不出来。另一种深度学习的方式是回归方式,通过神经网络大致回归行人的密度分布,可以有效克服遮挡情况,但是容易受到场景,着装,光照等影响,换一个场景可能需要通过进一步的训练,通用性较差。用传统的方法,主要优点在于速度快,但是无论从检测的精度角度,还是应对遮挡的问题,都不如深度学习的方法。同时也有人用多相机,有专利尝试用多相机多视角去弥补单相机有些人检测不到的问题,但是大多采用传统的检测方法,没有结合深度学习的优势,效果也不是很好。

发明内容

[0004] 鉴于以上所述现有技术的缺点,本发明的目的在于提供行人计数方法、系统、计算机设备和存储介质,解决现有技术的问题。

[0005] 为实现上述目标及其他相关目标,本发明提供一种行人计数方法,包括:获取光场相机阵列采集到的多幅行人图像;其中,所述多幅行人图像包括:在所述光场相机阵列中每个相机的视野区域下采集到的一幅行人图像;通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域;对不同相机的行人图像间属于同一行人的行人所属区域进行第一去重,并对第一去重后的各幅行人图像中的行人所属区域统计得到第一行人数量;对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对检测结果中属于同一行人的部分进行第二去重,并根据第二去重后的检测结果进行计算得到第二行人数量;对所述第一行人数量及第二行人数量计和以得到总行人数量。

[0006] 于本发明的一实施例中,所述目标检测器通过多层全卷积神经网络模型实现,其生成用于框选所述行人图像中每个行人的目标定界框作为所述行人所属区域。

[0007] 于本发明的一实施例中,所述目标检测器被训练为令其目标定界框仅用于框选识别为行人分类的部分图像。

[0008] 于本发明的一实施例中,所述第一去重,包括:将各幅所述行人图像的行人所属区域中的像素点的二维图像坐标转换为三维世界坐标,并将所述三维世界坐标转换为对应的二维地面坐标,以得到二维地面坐标对应的第一映射点;以分别获取自两个行人图像中各

一行人所属区域的第一映射点成对,判断每对第一映射点所对应的一对像素点分别所在的行人所属区域是否是在对应的两个相机的公共视野区域下采集到的;若否,则将该对第一映射点分别归类于不同行人;若是,则计算该对第一映射点之间的相似程度;判断所述相似程度是否大于一预设阈值;若是,则判断该对第一映射点归类于不同行人;若否,则判断该对第一映射点疑似属于同一行人,并进行第三去重。

[0009] 于本发明的一实施例中,所述第三去重,包括:将疑似属于同一行人的每对第一映射点对应的一对像素点分别所在的行人所属区域分别输入一孪生神经网络模型的两个神经网络中;计算分别在所述两个神经网络最后一个全连接层输出的特征向量间的相似程度;判断所述特征向量间的相似程度是否大于一预设阈值;若是,则判断该两个行人所属区域属于同一行人,进而去重;若否,则判断该两个行人所属区域属于不同行人。

[0010] 于本发明的一实施例中,所述相似程度的计算方式包括:余弦距离、欧氏距离、标准化欧氏距离、马氏距离、汉明距离、及曼哈顿距离中的任意一种。

[0011] 于本发明的一实施例中,所述对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测,包括:对应每幅行人图像,利用用于回归人头分布的补偿检测模型处理所述行人图像中的其它区域,得到对应的人头分布密度图像;其中,所述人头分布密度图像用于供积分计算所述第二行人数量。

[0012] 于本发明的一实施例中,所述第二去重,包括:将各幅所述行人图像的行人所属区域以外的其它区域中的像素点的二维图像坐标转换为三维世界坐标,以得到对应的第二映射点,每幅行人图像所对应的各所述第二映射点构成采集该行人图像的相机的部分空间覆盖区域;通过容斥定理对光场相机阵列中各不同相机的部分空间覆盖区域间的重复覆盖区域加以消除,以保留各个相机的单独覆盖区域,并获取对应各单独覆盖区域的去重人头分布密度图像;其中,所述部分人头分布密度图像用于供积分以计算所述第二行人数量。

[0013] 为实现上述目标及其他相关目标,本发明提供一种行人计数系统,包括:通信单元,用于获取光场相机阵列采集到的多幅行人图像;其中,所述多幅行人图像包括:在所述光场相机阵列中每个相机的视野区域下采集到的一幅行人图像;处理单元,用于通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域;对不同相机的行人图像间属于同一行人的行人所属区域进行第一去重,并对第一去重后的各幅行人图像中的行人所属区域统计得到第一行人数量;对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对检测结果中属于同一行人的部分进行第二去重,并根据第二去重后的检测结果进行计算得到第二行人数量;对所述第一行人数量及第二行人数量计和以得到总行人数量。

[0014] 于本发明的一实施例中,所述目标检测器通过多层全卷积神经网络模型实现,其生成用于框选所述行人图像中每个行人的目标定界框作为所述行人所属区域。

[0015] 于本发明的一实施例中,所述目标检测器被训练为令其目标定界框仅用于框选识别为行人分类的部分图像。

[0016] 于本发明的一实施例中,所述第一去重,包括:将各幅所述行人图像的行人所属区域中的像素点的二维图像坐标转换为三维世界坐标,并将所述三维世界坐标转换为对应的二维地面坐标,以得到二维地面坐标对应的第一映射点;以分别获取自两个行人图像中各一行人所属区域的第一映射点成对,判断每对第一映射点所对应的一对像素点分别所在的

行人所属区域是否是在对应的两个相机的公共视野区域下采集到的;若否,则将该对第一映射点分别归类于不同行人;若是,则计算该对第一映射点之间的相似程度;判断所述相似程度是否大于一预设阈值;若是,则判断该对第一映射点归类于不同行人;若否,则判断该对第一映射点疑似属于同一行人,并进行第三去重。

[0017] 于本发明的一实施例中,所述第三去重,包括:将疑似属于同一行人的每对第一映射点对应的一对像素点分别所在的行人所属区域分别输入一孪生神经网络模型的两个神经网络中;计算分别在所述两个神经网络最后一个全连接层输出的特征向量间的相似程度;判断所述特征向量间的相似程度是否大于一预设阈值;若是,则判断该两个行人所属区域属于同一行人,进而去重;若否,则判断该两个行人所属区域属于不同行人。

[0018] 于本发明的一实施例中,所述相似程度的计算方式包括:余弦距离、欧氏距离、标准化欧氏距离、马氏距离、汉明距离、及曼哈顿距离中的任意一种。

[0019] 于本发明的一实施例中,所述对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测,包括:对应每幅行人图像,利用用于回归人头分布的补偿检测模型处理所述行人图像中的其它区域,得到对应的人头分布密度图像;其中,所述人头分布密度图像用于供积分计算所述第二行人数量。

[0020] 于本发明的一实施例中,所述第二去重,包括:将各幅所述行人图像的行人所属区域以外的其它区域中的像素点的二维图像坐标转换为三维世界坐标,以得到对应的第二映射点,每幅行人图像所对应的各所述第二映射点构成采集该行人图像的相机的部分空间覆盖区域;通过容斥定理对光场相机阵列中各不同相机的部分空间覆盖区域间的重复覆盖区域加以消除,以保留各个相机的单独覆盖区域,并获取对应各单独覆盖区域的去重人头分布密度图像;其中,所述部分人头分布密度图像用于供积分以计算所述第二行人数量。

[0021] 为实现上述目标及其他相关目标,本发明提供一种计算机设备,所述计算机设备包括:通信器、处理器及存储器;所述通信器,通信连接光场相机阵列;所述存储器,存储有计算机程序;所述处理器,用于运行所述计算机程序,以实现所述的行人计数方法。

[0022] 于本发明的一实施例中,所述计算机设备与光场相机阵列之间通过通信网络连接。

[0023] 为实现上述目标及其他相关目标,本发明提供一种计算机存储介质,存储有计算机程序;所述计算机程序被运行时实现所述的行人计数方法。

[0024] 如上所述,本发明的行人计数方法、系统、计算机设备和存储介质,通过获取光场相机阵列采集到的多幅行人图像;其中,所述多幅行人图像包括:在所述光场相机阵列中每个相机的视野区域下采集到的一幅行人图像;通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域;对不同相机的行人图像间属于同一行人的行人所属区域进行第一去重,并对第一去重后的各幅行人图像中的行人所属区域统计得到第一行人数量;对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对检测结果中属于同一行人的部分进行第二去重,并根据第二去重后的检测结果进行计算得到第二行人数量;对所述第一行人数量及第二行人数量计和以得到总行人数量;本发明基于光场相机阵列的多视角、计算方便等优势出发,结合深度学习高性能的目标检测方法,避免漏检问题,亦配合深度学习的行人去重方式,得到更为精准的行人计数结果。

附图说明

- [0025] 图1显示为本发明实施例中行人计数方法的流程示意图。
- [0026] 图2显示为本发明实施例中目标检测器检测效果的示意图。
- [0027] 图3显示为本发明实施例中第一去重的流程示意图。
- [0028] 图4显示为本发明实施例中第三去重的流程示意图。
- [0029] 图5显示为本发明实施例中第三去重使用回归模型的架构示意图。
- [0030] 图6A显示为本发明实施例中回归模型处理得到人头分布密度图的过程示意图。
- [0031] 图6B显示为本发明实施例中经掩盖行人所属区域之后的人头分布密度图。
- [0032] 图7显示为本发明实施例中第二去重的流程示意图。
- [0033] 图8显示为本发明实施例中行人计数系统的功能模块图。
- [0034] 图9显示为本发明实施例中实现行人计数的计算机设备及其设备连接方式的系统架构图。

具体实施方式

[0035] 以下通过特定的具体实例说明本发明的实施方式,本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用,本说明书中的各项细节也可以基于不同观点与应用,在没有背离本发明的精神下进行各种修饰或改变。需说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。

[0036] 本发明的技术方案,是关于行人计数方案的改进,针对现有技术中相机采集图像数据的局限,以及深度学习算法易受干扰的问题,取两者之长加以结合并进行优化,提升行人计数的精准性。

[0037] 如图1所示,展示本发明实施例中行人计数方法的流程示意图。

[0038] 所述方法具体包括:

[0039] 步骤S101:获取光场相机阵列采集到的多幅行人图像;其中,所述多幅行人图像包括:在所述光场相机阵列中每个相机的视野区域下采集到的一幅行人图像。

[0040] 光场相机阵列包含多个相机,每个相机从不同角度进行行人图像的采集,进而可以拼合成一张带有深度的总图像;而在本发明实施例中,是利用各幅行人图像进行行人检测及去重以完成行人数量统计。

[0041] 步骤S102:通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域。

[0042] 目前深度学习的行人检测方法大多依赖于通用的目标检测方法,这些方法可以检测很多类别的对象(每个对象对应一个分类,例如宠物、植物、行人、建筑物等等),但是本发明的技术方案只需要检测行人,所以采用对通用目标检测模型的调用,并通过在行人检测数据集上的迁移学习,训练得到专用于检测行人的模型。同时,为了考虑实时视频检测的需要,需要快速的行人检测方法,而常见的目标检测算法,如Faster R-CNN,存在着速度慢的缺点。

[0043] 故优选的,本发明的一实施例中采用SSD模型制作所述目标检测器。

[0044] SSD全称是Single Shot MultiBox Detector,其出自Jisoo Jeong,Hyojin Park,

Nojun Kwak等人的论文《Enhancement of SSD by concatenating feature maps for object detection》,其本身是一个多层的全卷积网络,该模型的每个用于特征提取的卷积层都会产生多个目标定界框,以此实现从不同卷积层产生多种尺度的目标定界框,所述目标定界框用于框选出图像中的目标,例如行人、植物、宠物、建筑物等。

[0045] 需说明的是,在本申请各实施例中,将框选出图像中的行人的目标界定框所占据的行人图像中的区域称之为行人所属区域。

[0046] 如图2所示,展示本实施例中通过目标检测器201来实现框选出一幅行人图像202中的各个行人所属区域203的处理过程。

[0047] 需说明的是,本发明通过SSD模型这样一个多层全卷积网络来作为目标检测器201实现了目标定界框产生,仅对行人进行框选,抛弃对应其他多余分类的定界框,形成了高效率的目标检测器,十分适合视频流的行人目标检测。

[0048] 具体的,由于SSD模型本身是用于多目标检测,但是对于行人计数来说,不需要检测其他种类的目标,如上图中的手提包,或者背包,因此以只有行人定界框的数据集对在多目标检测的数据集上训练好的SSD模型进行迁移学习,即进行网络模型的微调。

[0049] 模型微调时,网络结构不变,变化的是原来数据集是多种类的,现在的数据集是单种类,即只需判断一个目标定界框中内容是人还是背景,用以下公式说明:

$$[0050] \quad L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \quad \text{where} \quad \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$$

[0051] 上面的公式是SSD中用来提高分类精确度的分类损失函数, \hat{c}_i^p 代表对这个定界框进行对类别p的打分,打分越高,说明目标定界框内容越可能属于类别p。

[0052] 整个分类损失函数的计算是对目标定界框计算是否是背景的打分值,即 $i \in Neg$ 的部分,再对打分值进行 $-\log$ 处理,然后要判断SSD检测出来的这个目标定界框与数据集中人为标定好的目标定界框进行重合度计算,重合度超过一定阈值, x_{ij}^p 就等于1,表示SSD检测出来的第i个目标定界框与标定好的第j个目标定界框(该框中物体类别为p)重合度高,否则 x_{ij}^p 就等于0。与检测出来的目标定界框重合度高的那些标定好的定界框属于什么类别,那检测出来的定界框就需要对这个类别进行打分,再取 $-\log$, 并累加起来,即公式中 $i \in Pos$ 的部分。

[0053] 以上是对SSD多目标检测时的分类损失函数的解释,而本发明的实施例中,利用行人数据集训练SSD实现微调时,由于数据集的标定好的目标定界框都是只针对行人这一类别,所以对于SSD检测出来的每个目标定界框,除了对背景进行打分之外,如果存在与它重合度较高的标定好的目标定界框,那么就只需要对是否为行人打分即可,而不存在对其他类别的打分,这是模型微调时的不同。

[0054] 通过模型微调,训练出来的SSD可以达到更好的行人检测效果,同时排除掉其他类别的检测目标(如宠物、植物、建筑物等)的干扰。

[0055] 步骤S103:对不同相机的行人图像间属于同一行人的行人所属区域进行第一去重,并对第一去重后的各幅行人图像中的行人所属区域统计得到第一行人数量。

[0056] 具体的,通过图3实施例展示该第一去重的具体流程。

[0057] 所述第一去重,包括:

[0058] 步骤S301:将各幅所述行人图像的行人所属区域中的像素点的二维图像坐标转换为三维世界坐标,并将所述三维世界坐标转换为对应的二维地面坐标,以得到二维地面坐标对应的第一映射点。

[0059] 具体的,先将像素点在二维图像坐标系中的坐标利用相机内参转换为相机坐标系下的坐标,进而利用相机外参再转换为世界三维坐标,之后为排除行人身高影响,再将世界三维坐标转换为二维的地面坐标系中的地面坐标。

[0060] 举例来讲,假设某个相机所采集行人图像中的目标定界框中心的图像坐标为(x,y),从2D的图像坐标(x,y)计算出相机坐标下的空间坐标(X,Y,Z):

$$[0061] \quad \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = d * K^{-1} * \begin{bmatrix} x \\ y \\ 1 \end{bmatrix};$$

[0062] 其中,d代表深度信息,K代表该相机的内参矩阵,这都是标定好的光场相机阵列提供的。

[0063] 进而,从相机坐标(X,Y,Z),计算出相机阵列的公共世界坐标(X',Y',Z'):

$$[0064] \quad \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ W \end{bmatrix} = T^{-1} * \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$[0065] \quad \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \frac{1}{W} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix};$$

[0066] 其中,T是相机的外参数矩阵,W是齐次坐标的系数。

[0067] 如此,从每个相机采集的行人图像检测出来的行人上的各像素点,都可以映射到公共世界坐标系当中的三维世界坐标。

[0068] 由于上述实施例中取的像素点取决于目标定界框的尺寸,即从中心图像坐标出发,因此如果行人的身高不同,会导致两个人位置很相似的人,计算出来的世界三维坐标距离却比较大,因此,为了排除身高的影响,可以将世界坐标系中的点投影到二维地面坐标系中的点,计算过程如下:

$$[0069] \quad \begin{bmatrix} X'' \\ Y'' \end{bmatrix} = P * R * \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix};$$

$$[0070] \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix};$$

[0071] 上式中,R代表旋转矩阵,将世界坐标系中的点的坐标转换为一个Z轴向上的标准笛卡尔坐标系,然后P是一个投影矩阵,去掉获取笛卡尔坐标系中的高度坐标,保留平面坐标。

[0072] 步骤S302:以分别获取自两个行人图像中各一行人所属区域的第一映射点成对,判断每对第一映射点所对应的一对像素点分别所在的行人所属区域是否是在对应的两个相机的公共视野区域下采集到的。

[0073] 步骤S303:若否,则将该对第一映射点分别归类于不同行人;

[0074] 步骤S304:若是,则计算该对第一映射点之间的相似程度;进而进至步骤S305;

[0075] 步骤S305:判断所述相似程度是否大于一预设阈值;

[0076] 步骤S306:若是,则判断该对第一映射点归类于不同行人;

[0077] 步骤S307:若否,则判断该对第一映射点疑似属于同一行人,并进行第三去重。

[0078] 举例来说,在得到了二维地面坐标系中的每个行人的第二映射点坐标(X'' , Y'')之后,假如光场相机阵列中的A,B两个相机,且A相机中的某个行人映射出来二维地面坐标系中的一个点P1的空间坐标为(X''_1 , Y''_1),B相机中的某个行人映射出来的二维地面坐标系中的一个点P2空间坐标为(X''_2 , Y''_2),以P1及P2成对,采用几何去重方式判断两者是否属于重复的同一行人,以进行去重处理:

[0079] 判断P1和P2是否在两个相机的公共视野区域。

[0080] 如果不在公共视野区域,即在各自独有的视野区域,那么肯定不会重复,所以直接计数;如果P1和P2都在公共区域,那么进行下一步操作,计算P1和P2的相似程度,所述相似程度的计算方式包括:余弦距离、欧氏距离、标准化欧氏距离、马氏距离、汉明距离、及曼哈顿距离中的任意一种。

[0081] 在本实施例中采用欧氏距离的计算方式,即用如下公式表示:

$$[0082] \quad distance(P1, P2) = \sqrt{(X''_1 - X''_2)^2 + (Y''_1 - Y''_2)^2}$$

[0083] 进而,设定一相似程度阈值 α ,如果P1和P2的 $distance > \alpha$,则认为是两个不同的人,属于两个相机分别检测到了对方没有检测到的人,因此是两个人,都计入第一行人数量。

[0084] 如果两个点的 $distance < \alpha$,则认为很可能是同一个人的重复观测,当然也可能由于检测误差,以及两个不同的人站的很近,导致了空间坐标很近,就需要进行进一步的第三去重。

[0085] 于本发明的一实施例中,在本发明的一实施例中,可以通过提取特征,用行人重识别(ReID)的方法去进行所述第三去重的优化。

[0086] 如图4所示,展示实施例中所述第三去重的流程,其包括:

[0087] 步骤S401:将疑似属于同一行人的每对第一映射点对应的一对像素点分别所在的行人所属区域分别输入一孪生神经网络模型的两个神经网络中。

[0088] 于本发明的一实施例中,所述孪生神经网络模型,即Siamese neural network,其结构即如图5所示,可以将 $distance < \alpha$ 的P1和P2对应的一对像素点分别所在的行人所属区域分别输入孪生神经网络的两个神经网络中,P1所对应的像素点归类于行人1,P2对应的像素点归类于行人2,归类于行人1的像素点集合和归类于行人2的像素集合即为通过目标定界框在各自所在行人图像中框选出的行人所属区域1和行人所属区域2的部分图像,进而也就是如图5所示,分别将两个框选出的行人所属区域分别输入孪生神经网络模型的两个神经网络以判断两个部分图像所分别对应的行人1和行人2是否为同一行人。

[0089] 步骤S402:计算分别在所述两个神经网络最后一个全连接层输出的特征向量间的相似程度;

[0090] 步骤S403:判断所述特征向量间的相似程度是否大于一预设阈值;

[0091] 步骤S404:若是,则判断该两个行人所属区域属于同一行人,进而去重;

[0092] 步骤S405:若否,则判断该两个行人所属区域属于不同行人。

[0093] 举例来说,孪生神经网络模型最后的L2Norm层是用来计算欧氏距离损失函数的,即 Euclidean loss,是训练中用的,训练完成后,在衡量两个目标定界框中行人的部分图像(各个像素点)的相似程度时,只需要去掉L2Norm层,直接获得模型最后的全连接层FC7层输出的256维的特征向量即可。

[0094] 对于需要判断是否为同一个人的两个目标定界框的部分图像,我们利用上面的网络结构的获取256维的特征向量,然后需要比较这两个特征向量的相似度,可以采用欧式距离(也可以是余弦距离、标准化欧氏距离、马氏距离、汉明距离、及曼哈顿距离中的任意一种)来计算两个特征向量的距离作为相似度,距离越近,说明相似度越高,则两个检测框中的人越可能是同一个人,具体我们用一个阈值 β 去衡量,具体如下所示:

$$[0095] \quad \text{identical}(b1, b2) = \begin{cases} \text{true} & \text{if } \text{distance}(v1, v2) > \beta \\ \text{false} & \text{if } \text{distance}(v1, v2) \leq \beta \end{cases};$$

[0096] 如果 $\text{identical}(b1, b2)$ 等于true表示 $b1, b2$ 两个目标定界框中的人是同一个人,反之则为不同的人。 $\text{distance}(v1, v2)$ 表示从两个目标定界框 $v1, v2$ 中提取出来的两个特征向量的欧式距离,如果这个欧式距离大于 β ,则判断为同一个人,反之就认为是两个不同的人。 β 的值视实际使用效果而定。

[0097] 同时,为防止遗漏对行人图像中行人所属区域以外的其它区域中的行人的计数,因此需要对该些其它区域进行计数。

[0098] 具体的,如图2所示,虽然通过SSD模型检测之后已框选出大部分行人,但由于全身大部分区域被遮挡,右上区域的一些人未被检测出来,故需要对这些区域进行再次检测

[0099] 步骤S104:对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对检测结果中属于同一行人的部分进行第二去重,并根据第二去重后的检测结果进行计算得到第二行人数量。

[0100] 具体的,对于没有检测到行人的区域,我们进一步可以采用回归的方式进行补偿,即将画面输入到一个回归模型,可以是一个用于回归人头分布的全卷积神经网络。

[0101] 如图6A所示,展示根据目标定界框标定行人所属区域后的行人图像601经回归模型602进行回归得到人头分布密度图603(可以通过例如高斯函数作为卷积核)的过程。密度图603中高亮部分表示该处存在人头的概率较大,背景表示存在人头概率为0,对整张图进行积分(即像素值的累加),即可获得这样图中的人头数,也就是行人数量。

[0102] 由于行人图像601中还包含了行人所属区域,因此需要对该些区域进行剔除后再进行行人计数。具体的,如图6B所示,可以通过与目标定界框尺寸对应的掩膜604(mask)遮蔽各个行人所属区域对应在人头密度分布图603中的部分,再进行积分,可以得到该行人的数量601在第一行人数量以外的行人数量。

[0103] 但是,由于光场相机阵列的多个相机采集图像间可能会有重叠区域,位于其它区域的重叠区域中的行人可能会被重复计数,因此,需要进行第二去重。

[0104] 如图7所示,展示实施例中所所述第二去重的具体流程,其包括:

[0105] 步骤S701:将各幅所述行人图像的行人所属区域以外的其它区域中的像素点的二维图像坐标转换为三维世界坐标,以得到对应的第二映射点,每幅行人图像所对应的各所述第二映射点构成采集该行人图像的相机的部分空间覆盖区域。

[0106] 在本实施例中,进行坐标转换的方式与前述实施例中的S301类似,将行人图像中剔除行人所属区域之后的其它区域中的各像素点从图像坐标系映射到世界坐标系,构成部分空间覆盖区域,不同相机所拍摄的图像中的不同点可能在世界坐标系中对应的是同一点,从而就可以据以去重。

[0107] 步骤S702:通过容斥定理对光场相机阵列中各不同相机的部分空间覆盖区域间的重复覆盖区域加以消除,以保留各个相机的单独覆盖区域,并获取对应各单独覆盖区域的去重人头分布密度图像;其中,所述部分人头分布密度图像用于供积分以计算所述第二行人数量。

[0108] 具体的,采用容斥定理对多个相机的第二映射点构成的部分空间覆盖区域中各个三维世界坐标相同的点进行去重。

[0109] 举例来说,假设相机1的部分空间覆盖区域为A,相机2的部分空间覆盖区域为B,相机3的部分空间覆盖区域为C

[0110] 则根据容斥定理去剔除A、B、C间的重叠区域:

[0111] $S(A \cup B \cup C) = S(A) + S(B) + S(C) - S(A \cap B) - S(B \cap C) - S(C \cap A) + S(A \cap B \cap C)$;

[0112] 可以计算出各个相机部分空间覆盖区域中的独立覆盖部分的集合。

[0113] 进而,根据光场相机阵列中各相机的独立覆盖部分的集合所对应的部分分布密度图像进行积分,即对各部分分布密度图像的每个点对应的概率密度值进行求和,即可计算出回归的第二行人数量。

[0114] 步骤S105:对所述第一行人数量及第二行人数量计和以得到总行人数量。

[0115] 第一行人数量代表目标检测器检测到的各幅行人图像中行人所属区域对应的行人数量,第二行人数量代表行人所属区域以外其它区域中的行人数量,相加即获得光场相机阵列拍摄行人图像统计得到的总行人数量。

[0116] 如图8所示,展示本发明实施例中提供的一种行人计数系统800,由于该行人计数系统800的原理与前述方法实施例大致相同,因此方法实施例中的各种技术细节均能应用于本实施例中,故不作重复赘述。

[0117] 所述系统800包括:

[0118] 通信单元801,用于获取光场相机阵列采集到的多幅行人图像;其中,所述多幅行人图像包括:在所述光场相机阵列中每个相机的视野区域下采集到的一幅行人图像;

[0119] 处理单元802,用于通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域;对不同相机的行人图像间属于同一行人的行人所属区域进行第一去重,并对第一去重后的各幅行人图像中的行人所属区域统计得到第一行人数量;对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对检测结果中属于同一行人的部分进行第二去重,并根据第二去重后的检测结果进行计算得到第二行人数量;对所述第一行人数量及第二行人数量计和以得到总行人数量。

[0120] 于本发明的一实施例中,所述目标检测器通过多层全卷积神经网络模型实现,其生成用于框选所述行人图像中每个行人的目标定界框作为所述行人所属区域。

[0121] 于本发明的一实施例中,所述目标检测器被训练为令其目标定界框仅用于框选识别为行人分类的部分图像。

[0122] 于本发明的一实施例中,所述第一去重,包括:将各幅所述行人图像的行人所属区

域中的像素点的二维图像坐标转换为三维世界坐标,并将所述三维世界坐标转换为对应的二维地面坐标,以得到二维地面坐标对应的第一映射点;以分别获取自两个行人图像中各一行人所属区域的第一映射点成对,判断每对第一映射点所对应的一对像素点分别所在的行人所属区域是否是在对应的两个相机的公共视野区域下采集到的;若否,则将该对第一映射点分别归类于不同行人;若是,则计算该对第一映射点之间的相似程度;判断所述相似程度是否大于一预设阈值;若是,则判断该对第一映射点归类于不同行人;若否,则判断该对第一映射点疑似属于同一行人,并进行第三去重。

[0123] 于本发明的一实施例中,所述第三去重,包括:将疑似属于同一行人的每对第一映射点对应的一对像素点分别所在的行人所属区域分别输入一孪生神经网络模型的两个神经网络中;计算分别在所述两个神经网络最后一个全连接层输出的特征向量间的相似程度;判断所述特征向量间的相似程度是否大于一预设阈值;若是,则判断该两个行人所属区域属于同一行人,进而去重;若否,则判断该两个行人所属区域属于不同行人。

[0124] 于本发明的一实施例中,所述相似程度的计算方式包括:余弦距离、欧氏距离、标准化欧氏距离、马氏距离、汉明距离、及曼哈顿距离中的任意一种。

[0125] 于本发明的一实施例中,所述对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测,包括:对应每幅行人图像,利用用于回归人头分布的补偿检测模型处理所述行人图像中的其它区域,得到对应的人头分布密度图像;其中,所述人头分布密度图像用于供积分计算所述第二行人数量。

[0126] 于本发明的一实施例中,所述第二去重,包括:将各幅所述行人图像的行人所属区域以外的其它区域中的像素点的二维图像坐标转换为三维世界坐标,以得到对应的第二映射点,每幅行人图像所对应的各所述第二映射点构成采集该行人图像的相机的部分空间覆盖区域;通过容斥定理对光场相机阵列中各不同相机的部分空间覆盖区域间的重复覆盖区域加以消除,以保留各个相机的单独覆盖区域,并获取对应各单独覆盖区域的去重人头分布密度图像;其中,所述部分人头分布密度图像用于供积分以计算所述第二行人数量。

[0127] 需要说明的是,应理解以上装置的各个单元的划分仅仅是一种逻辑功能的划分,实际实现时可以全部或部分集成到一个物理实体上,也可以物理上分开。且这些单元可以全部以软件通过处理元件调用的形式实现;也可以全部以硬件的形式实现;还可以部分单元通过处理元件调用软件的形式实现,部分单元通过硬件的形式实现。例如,处理单元可以为单独设立的处理元件,也可以集成在上述装置的某一个芯片中实现,此外,也可以以程序代码的形式存储于上述装置的存储器中,由上述装置的某一个处理元件调用并执行以上处理单元的功能。其它单元的实现与之类似。此外这些单元全部或部分可以集成在一起,也可以独立实现。这里所述的处理元件可以是一种集成电路,具有信号的处理能力。在实现过程中,上述方法的各步骤或以上各个单元可以通过处理器元件中的硬件的集成逻辑电路或者软件形式的指令完成。

[0128] 例如,以上这些单元可以是配置成实施以上方法的一个或多个集成电路,例如:一个或多个特定集成电路(Application Specific Integrated Circuit,简称ASIC),或,一个或多个微处理器(digital signal processor,简称DSP),或,一个或者多个现场可编程门阵列(Field Programmable Gate Array,简称FPGA)等。再如,当以上某个单元通过处理元件调用程序代码的形式实现时,该处理元件可以是通用处理器,例如中央处理器

(CentralProcessingUnit,简称CPU)或其它可以调用程序代码的处理器。再如,这些单元可以集成在一起,以片上系统(system-on-a-chip,简称SOC)的形式实现。

[0129] 如图9所示,展示本发明实施例中提供的计算机设备900,所述计算机设备900包括:通信器901、处理器902及存储器903;所述通信器901,通信连接光场相机阵列904;所述存储器903,存储有计算机程序;所述处理器902,用于运行所述计算机程序,以实现前述实施例中的行人计数方法。

[0130] 可选的,通信器901、处理器902及存储器903之间可以通过系统总线连接,即如图中粗线所示,系统总线可以是外设部件互连标准(PeripheralPomponentInterconnect,简称PCI)总线或扩展工业标准结构(ExtendedIndustryStandardArchitecture,简称EISA)总线等。该系统总线可以分为地址总线、数据总线、控制总线等。需说明的是,系统总线在图中虽然仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。通信接口用于实现数据库访问装置与其他设备(例如客户端、读写库和只读库)之间的通信。

[0131] 上述的处理器902可以是通用处理器,包括中央处理器(CentralProcessingUnit,简称CPU)、网络处理器(NetworkProcessor,简称NP)等;还可以是数字信号处理器(DigitalSignalProcessing,简称DSP)、专用集成电路(ApplicationSpecificIntegratedCircuit,简称ASIC)、现场可编程门阵列(Field-ProgrammableGateArray,简称FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

[0132] 存储器903可能包含随机存取存储器(RandomAccessMemory,简称RAM),也可能还包括非易失性存储器(non-volatilememory),例如至少一个磁盘存储器。

[0133] 优选的,在本实施例中,所述计算机设备900与光场相机阵列904之间通过通信网络905连接,所述通信网络905连接可以是一个或多个有线或无线网络的任何合适的组合形式;所述通信网络可以包括物联网、互联网、内联网、广域网(WAN)、局域网(LAN)、无线网络、数字用户线(DSL)网络、帧中继网络、异步传输模式(ATM)网络、虚拟专用网络(VPN)和/或任何其它合适的通信网络中的任何一个或多个,则所述通信器901是符合该些网络通信协议要求的通信电路实现。

[0134] 光场相机阵列904可以集成或连接相应的通信器,从而能接入通信网络905以收发数据。

[0135] 虽然在本实施例中,所述计算机设备900可以是中心化网络架构中的服务器实现,但是在其它实施例中,它也可以是去中心化网络架构中的分布式网络设备实现。

[0136] 具体的,可以通过安装在某场景下的光场相机阵列获取视频流,视频流通过通信网络传到控制中心的服务器上,服务器上执行人数统计的程序,然后将人数输出到屏幕,供有关人员使用。

[0137] 为实现上述目标及其他相关目标,本发明提供一种计算机存储介质,存储有计算机程序;所述计算机程序被运行时实现所述的行人计数方法。所述存储介质包括:ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0138] 综上所述,本发明的行人计数方法、系统、计算机设备和存储介质,通过获取光场相机阵列采集到的多幅行人图像;其中,所述多幅行人图像包括:在所述光场相机阵列中每个相机的视野区域下采集到的一幅行人图像;通过目标检测器分别识别每幅行人图像中各行人对应的行人所属区域;对不同相机的行人图像间属于同一行人的行人所属区域进行第

一去重,并对第一去重后的各幅行人图像中的行人所属区域统计得到第一行人数量;对每幅行人图像中各行人所属区域以外的其它区域进行第二行人检测并对检测结果中属于同一行人的部分进行第二去重,并根据第二去重后的检测结果进行计算得到第二行人数量;对所述第一行人数量及第二行人数量计和以得到总行人数量;本发明基于光场相机阵列的多视角、计算方便等优势出发,结合深度学习高性能的目标检测方法,避免漏检问题,亦配合深度学习的行人去重方式,得到更为精准的行人计数结果。

[0139] 上述实施例仅例示性说明本发明的原理及其功效,而非用于限制本发明。任何熟悉此技术的人士皆可在不违背本发明的精神及范畴下,对上述实施例进行修饰或改变。因此,举凡所属技术领域中具有通常知识者在未脱离本发明所揭示的精神与技术思想下所完成的一切等效修饰或改变,仍应由本发明的权利要求所涵盖。

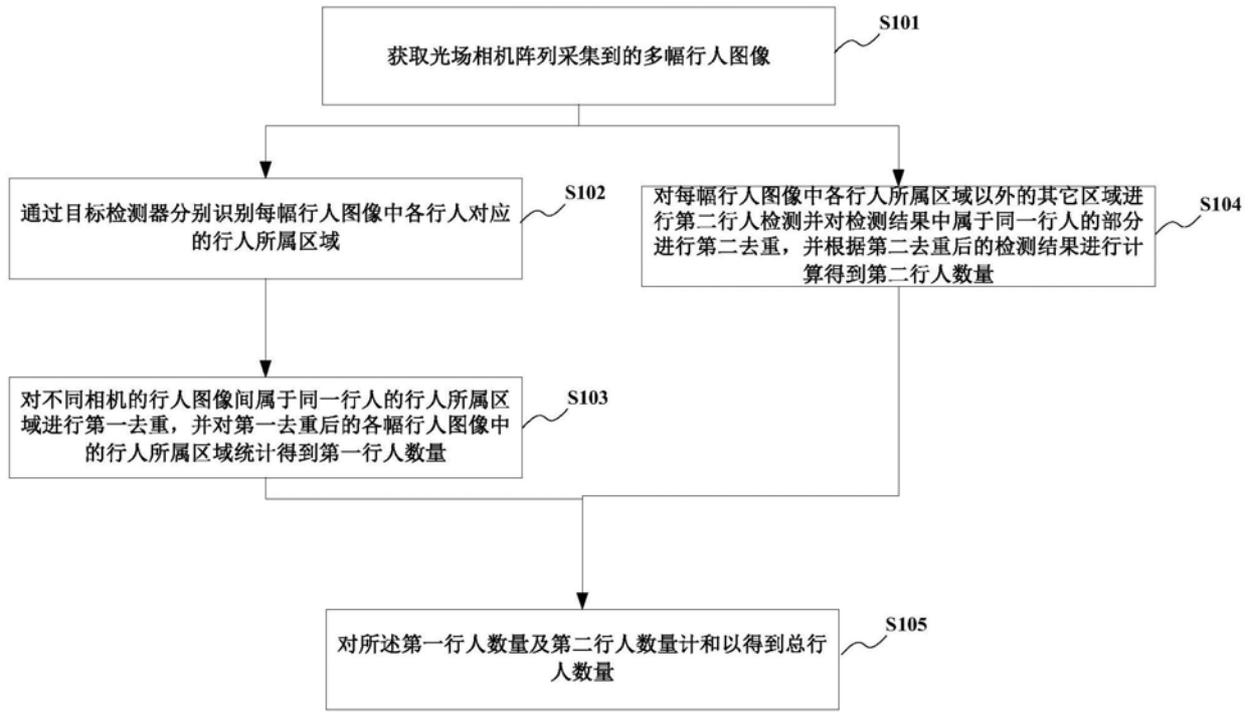


图1

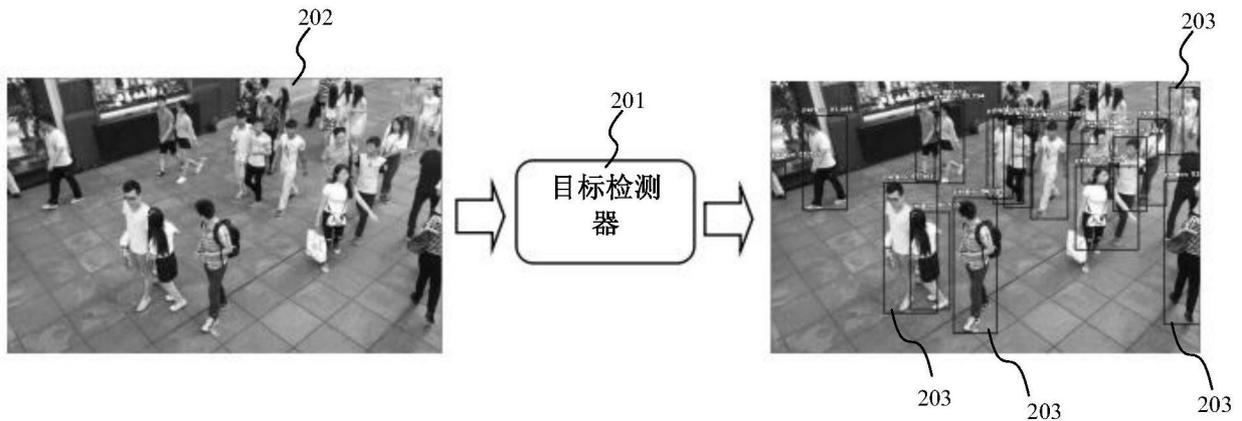


图2

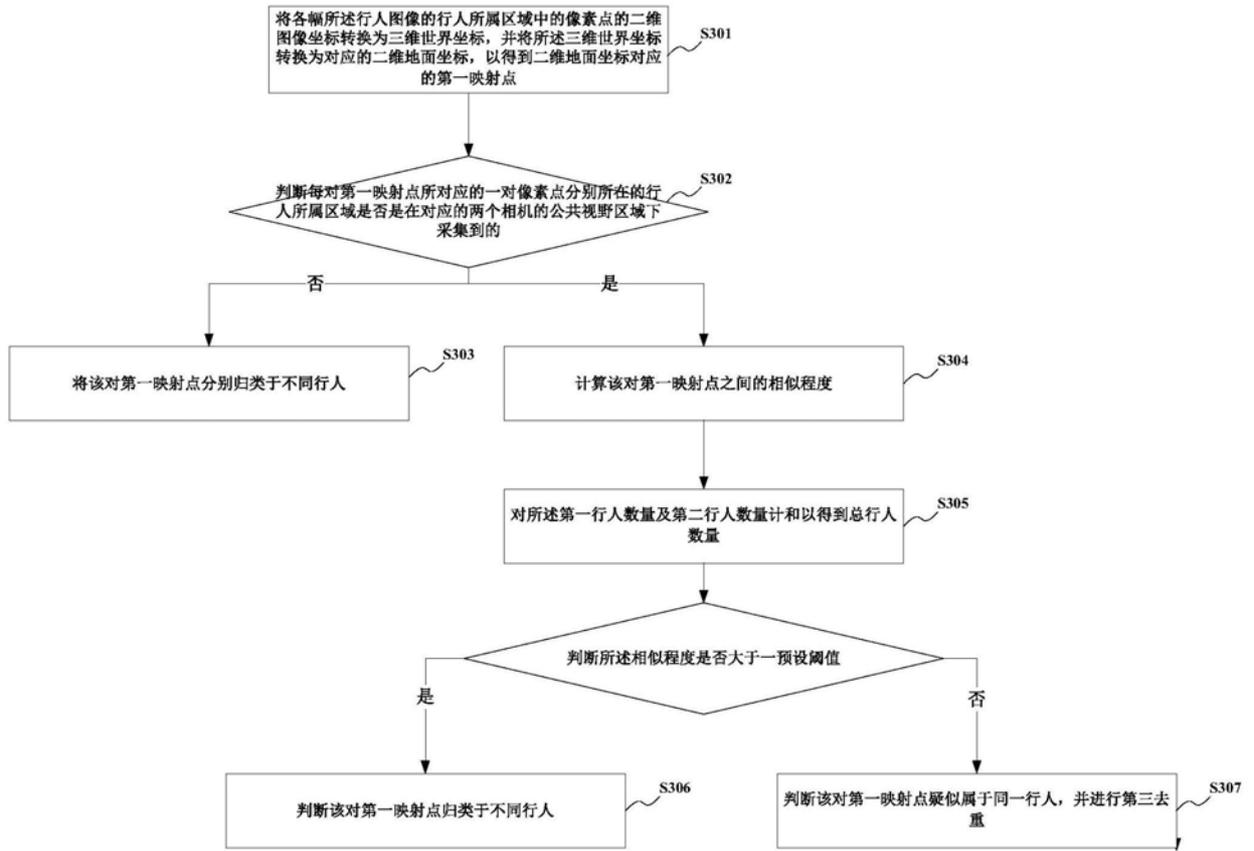


图3

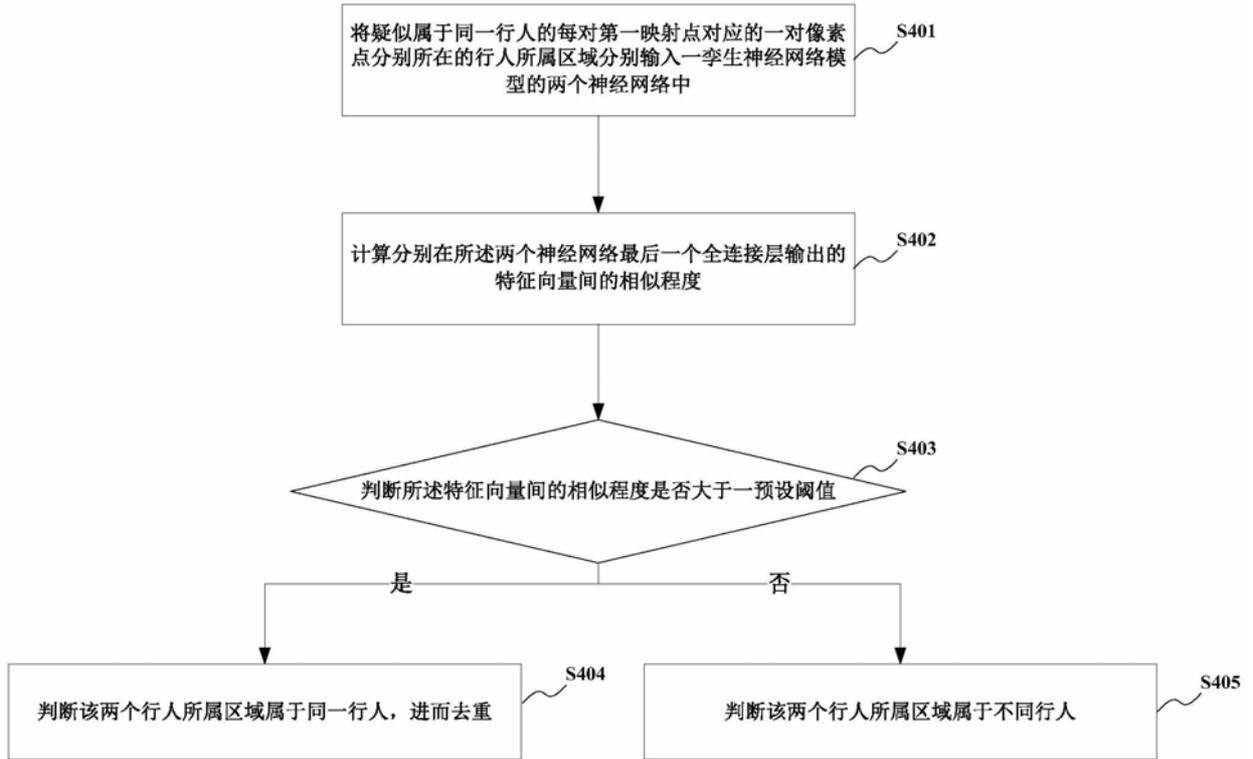


图4

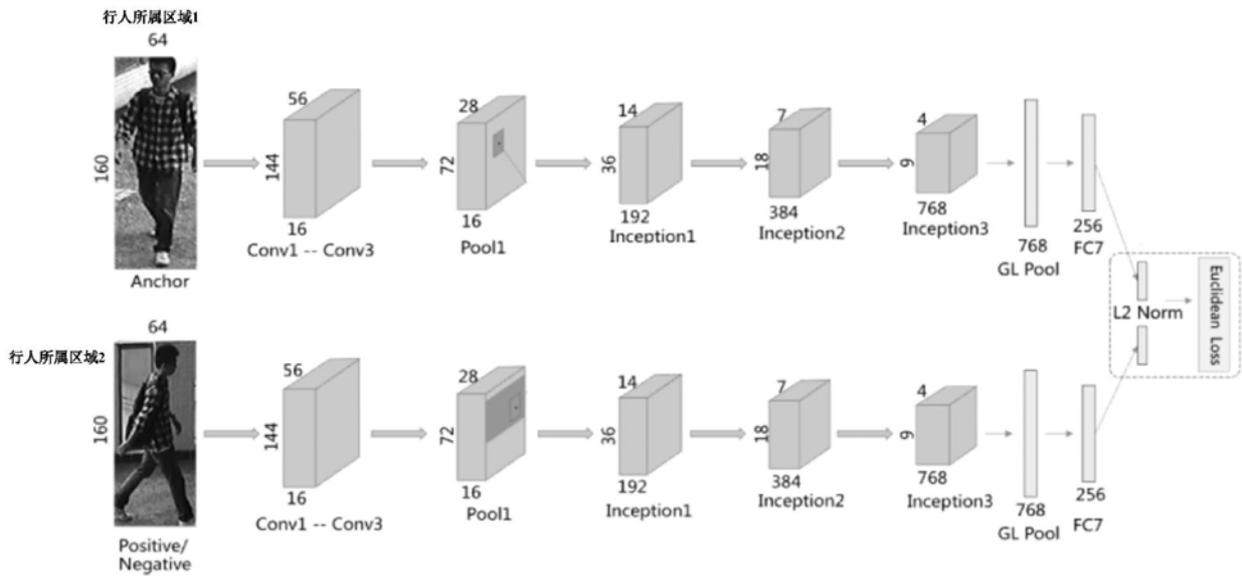


图5

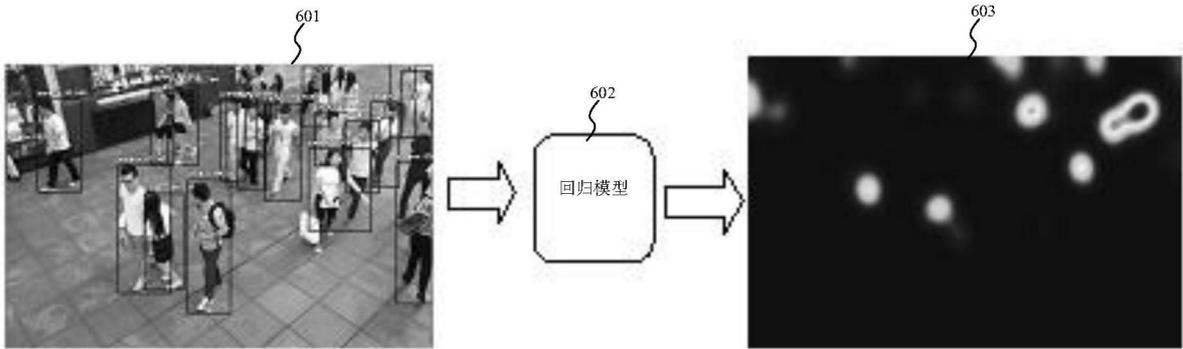


图6A



图6B

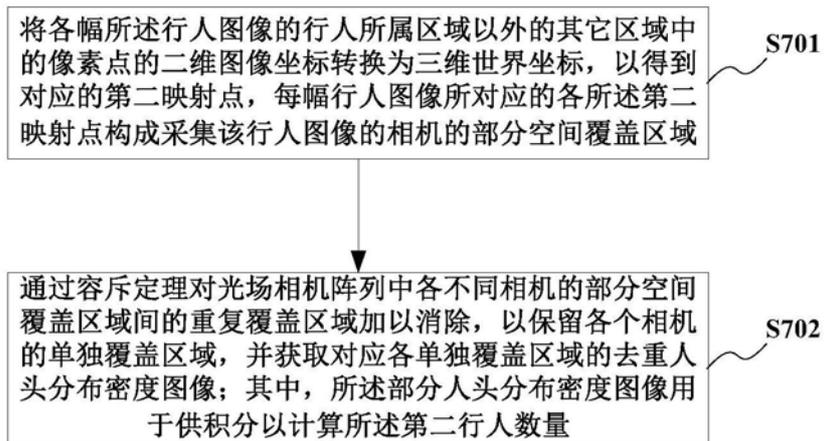


图7

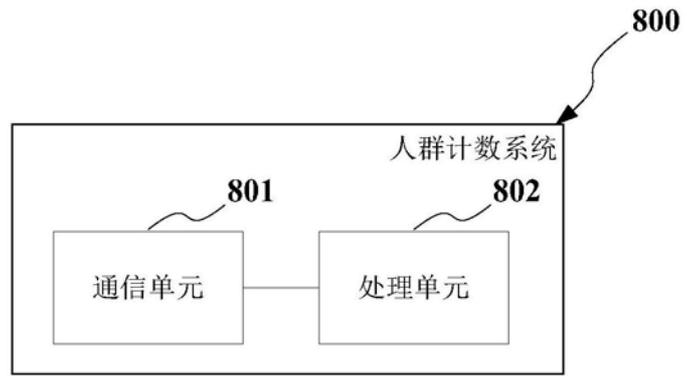


图8

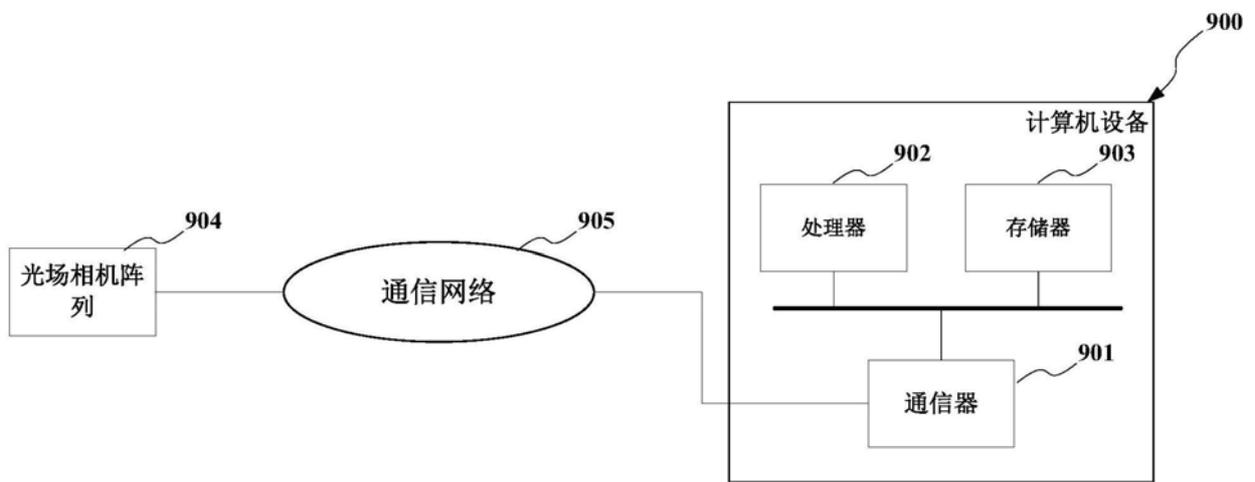


图9