



(12) 发明专利申请

(10) 申请公布号 CN 114694230 A

(43) 申请公布日 2022. 07. 01

(21) 申请号 202210409189.4

(22) 申请日 2022.04.19

(71) 申请人 山东师范大学

地址 250014 山东省济南市历下区文化东路88号

(72) 发明人 张桂娟 赵传淼 陈伟达 宋婷
陆佃杰

(74) 专利代理机构 济南圣达知识产权代理有限公司 37221

专利代理师 闫圣娟

(51) Int. Cl.

G06V 40/16 (2022.01)

G06V 20/40 (2022.01)

权利要求书2页 说明书11页 附图2页

(54) 发明名称

基于时空一致性的群体表情情绪识别方法及系统

(57) 摘要

本公开提出了基于时空一致性的群体表情情绪识别方法及系统,包括如下步骤:对获取的待识别的人群运动场景中的视频图像,进行人脸检测识别得到每帧图像中的每个个体的人脸表情;根据每个个体的人脸表情识别个体的情绪是否为恐慌情绪;基于时间一致性计算有部分帧图像中被遮挡个体是否为恐慌情绪,基于空间一致性计算所有帧图像中被遮挡个体是否为恐慌情绪;根据所有个体中为恐慌情绪个体的比例,得到群体恐慌度。本公开基于提出时间一致性的个体情绪计算方法和空间一致性的个体情绪计算方法,解决了人群中的遮挡导致面部表情信息采集不完整的情绪识别问题,再通过群体恐慌度,解决了人群运动场景中群体情绪识别的问题。



1. 基于时空一致性的群体表情情绪识别方法, 其特征在于, 包括如下步骤:

对获取的待识别的人群运动场景中的视频图像, 进行人脸检测识别得到每帧图像中的每个个体的人脸表情;

根据每个个体的人脸表情识别个体的情绪是否为恐慌情绪; 基于时间一致性计算有部分帧图像中被遮挡个体是否为恐慌情绪, 基于空间一致性计算所有帧图像中都被遮挡个体是否为恐慌情绪;

根据所有个体中为恐慌情绪个体的比例, 得到群体恐慌度。

2. 如权利要求1所述的基于时空一致性的群体表情情绪识别方法, 其特征在于: 对获取的待识别的人群运动场景中的视频图像, 进行人脸检测识别得到每帧图像中的每个个体的人脸表情的方法, 包括如下步骤:

输入待识别的人群运动场景中的视频;

把视频进行分帧处理成一帧一帧的图片;

针对分帧处理的每帧图像进行人脸检测;

对每张图片检测到的人脸进行裁剪, 得到只包含人脸的图片;

将裁剪后的人脸图片输入到深度残差网络中得到人脸表情识别结果。

3. 如权利要求2所述的基于时空一致性的群体表情情绪识别方法, 其特征在于:

针对分帧处理的每帧图像进行人脸检测采用基于haar特征提取算法实现;

或者,

对检测到的人脸进行裁剪, 得到只包含人脸的图片的步骤如下:

对进行人脸检测后每张图片, 设置与人脸大小相近的矩形框对检测到的人脸进行框选;

把每张图片里框选出来的人脸裁剪出来并标号, 按照标号的顺序输出一张图片中每一个人脸。

4. 如权利要求2所述的基于时空一致性的群体表情情绪识别方法, 其特征在于: 对于深度残差网络中得到人脸表情识别结果, 采用矩阵的形式, 得到情绪识别结果矩阵;

或者, 深度残差网络为Resnet34网络;

Resnet34网络结构包括卷积层、残差模块、池化层、全连接层以及Softmax分类器, 所述卷积层、残差模块、池化层、全连接层以及Softmax分类器依次连接。

5. 如权利要求1所述的基于时空一致性的群体表情情绪识别方法, 其特征在于: 基于时间一致性计算有部分帧图像中被遮挡个体a是否为恐慌情绪的方法, 具体为: 在能够采集到个体a完整面部表情信息的所有帧图片中, 识别为恐慌情绪的帧数的比例大于设定的比例阈值, 即判定该个体a的情绪在被识别的人群运动的视频中为恐慌情绪, 否则为非恐慌情绪。

6. 如权利要求1所述的基于时空一致性的群体表情情绪识别方法, 其特征在于: 基于空间一致性计算所有帧图像中被遮挡个体j是否为恐慌情绪, 具体为: 将每一帧图片转化为深度图; 计算每一帧图像中所有已经通过计算得到情绪结果的个体与被遮挡个体j的距离, 根据距离得到个体j的情绪受群体中其他个体的影响值之和, 根据影响值确定是否为恐慌情绪, 其他个体对个体j的恐慌情绪的影响为负值, 非恐慌情绪的影响为正值。

7. 如权利要求1所述的基于时空一致性的群体表情情绪识别方法, 其特征在于: 群体恐

慌度包括无恐慌、轻度恐慌、中度恐慌和重度恐慌。

8. 基于时空一致性的群体表情情绪识别系统,其特征不在于,包括:

人脸识别模块:被配置为用于对获取的待识别的人群运动场景中的视频图像,进行人脸检测识别得到每帧图像中的每个个体的人脸表情;

个体恐慌判断模块:被配置为用于根据每个个体的人脸表情识别个体的情绪是否为恐慌情绪;基于时间一致性计算有部分帧图像中被遮挡个体是否为恐慌情绪,基于空间一致性计算所有帧图像中被遮挡个体是否为恐慌情绪;

群体恐慌度判断模块:被配置为用于根据所有个体中为恐慌情绪个体的比例,得到群体恐慌度。

9. 一种电子设备,其特征不在于,包括存储器和处理器以及存储在存储器上并在处理器上运行的计算机指令,所述计算机指令被处理器运行时,完成权利要求1-7任一项方法所述的步骤。

10. 一种计算机可读存储介质,其特征不在于,用于存储计算机指令,所述计算机指令被处理器执行时,完成权利要求1-7任一项方法所述的步骤。

基于时空一致性的群体表情情绪识别方法及系统

技术领域

[0001] 本公开涉及深度学习、计算机视觉技术相关技术领域，具体的说，是涉及基于时空一致性的群体表情情绪识别方法及系统。

背景技术

[0002] 本部分的陈述仅仅是提供了与本公开相关的背景技术信息，并不必然构成在先技术。

[0003] 近年来，人群运动仿真研究得到了重视。随着社会城市化的飞速发展和人们生活水平的提高，人们出入公共场所的次数越来越多，致使公共场所中人群密度和流动性非常大，例如车站、学校和大型娱乐购物广场等。一旦发生危险情况，如火灾、地震等，可能导致人群中产生恐慌情绪，从而发生拥堵延误、踩踏死亡和其他安全事故。识别人群的真实群体情绪，然后采用相应的情绪感染和调控措施，减少群众恐慌，是确保疏散安全的有效途径。

[0004] 在人群运动的场景中通过人脸表情识别得到每个个体的情绪，再根据统计学的办法定义群体的恐慌程度，进一步得到人群的群体情绪是要解决的重要问题。在人群运动的场景中，由于人群的流动性大会出现遮挡现象，导致人脸表情信息采集不完整。因此，导致识别信息不完整的被遮挡的个体的情绪成为难点。发明人发现，传统的群体情绪识别方法不能识别在人群运动场景中的人员情绪，也没有解决人脸遮挡导致信息不完整的情况下的人脸表情识别的问题。

发明内容

[0005] 本公开为了解决上述问题，提出了基于时空一致性的群体表情情绪识别方法及系统，基于提出时间一致性的个体情绪计算方法和空间一致性的个体情绪计算方法，解决了人群中的遮挡导致面部表情信息采集不完整的情绪识别问题，再通过群体恐慌度，解决了人群运动场景中群体情绪识别的问题；通过该方法可以识别群体的恐慌程度，从而为人群运动的研究提供真实的数据。

[0006] 为了实现上述目的，本公开采用如下技术方案：

[0007] 一个或多个实施例提供了基于时空一致性的群体表情情绪识别方法，包括如下步骤：

[0008] 对获取的待识别的人群运动场景中的视频图像，进行人脸检测识别得到每帧图像中的每个个体的人脸表情；

[0009] 根据每个个体的人脸表情识别个体的情绪是否为恐慌情绪；基于时间一致性计算有部分帧图像中被遮挡个体是否为恐慌情绪，基于空间一致性计算所有帧图像中被遮挡个体是否为恐慌情绪；

[0010] 根据所有个体中为恐慌情绪个体的比例，得到群体恐慌度。

[0011] 一个或多个实施例提供了基于时空一致性的群体表情情绪识别系统，包括：

[0012] 人脸识别模块：被配置为用于对获取的待识别的人群运动场景中的视频图像，进

行人脸检测识别得到每帧图像中的每个个体的人脸表情；

[0013] 个体恐慌判断模块：被配置为用于根据每个个体的人脸表情识别个体的情绪是否为恐慌情绪；基于时间一致性计算有部分帧图像中被遮挡个体是否为恐慌情绪，基于空间一致性计算所有帧图像中被遮挡个体是否为恐慌情绪；

[0014] 群体恐慌度判断模块：被配置为用于根据所有个体中为恐慌情绪个体的比例，得到群体恐慌度。

[0015] 一种电子设备，包括存储器和处理器以及存储在存储器上并在处理器上运行的计算机指令，所述计算机指令被处理器运行时，完成上述方法所述的步骤。

[0016] 一种计算机可读存储介质，用于存储计算机指令，所述计算机指令被处理器执行时，完成上述方法所述的步骤。

[0017] 与现有技术相比，本公开的有益效果为：

[0018] 基于时间一致性的个体情绪计算和空间一致性的个体情绪计算来解决人群中的遮挡导致面部表情信息不完整的问题的方法，能够识别出待测视频中所有个体是否处于恐慌情绪，同时恐慌度定量了整个群体处于何种恐慌程度。该方法能够在人群遮挡导致面部表情信息采集不完整的情况下较为准确的得到群体的情绪，识别群体的恐慌程度，从而为人群运动的研究如人群疏散提供了准确的数据支撑，具有简化程度高，模块化程度高，模型精度高，准确契合场景等特点。

[0019] 本公开的优点以及附加方面的优点将在下面的具体实施例中进行详细说明。

附图说明

[0020] 构成本公开的一部分的说明书附图用来提供对本公开的进一步理解，本公开的示意性实施例及其说明用于解释本公开，并不构成对本公开的限定。

[0021] 图1是本公开实施例1的方法流程图；

[0022] 图2是本公开实施例1的残差模块结构示意图。

具体实施方式

[0023] 下面结合附图与实施例对本公开作进一步说明。

[0024] 应该指出，以下详细说明都是示例性的，旨在对本公开提供进一步的说明。除非另有指明，本文使用的所有技术和科学术语具有与本公开所属技术领域的普通技术人员通常理解相同含义。

[0025] 需要注意的是，这里所使用的术语仅是为了描述具体实施方式，而非意图限制根据本公开的示例性实施方式。如在这里所使用的，除非上下文另外明确指出，否则单数形式也意图包括复数形式，此外，还应当理解的是，当在本说明书中使用术语“包含”和/或“包括”时，其指明存在特征、步骤、操作、器件、组件和/或它们的组合。需要说明的是，在不冲突的情况下，本公开中的各个实施例及实施例中的特征可以相互组合。下面将结合附图对实施例进行详细描述。

[0026] 实施例1

[0027] 在一个或多个实施方式公开的技术方案中，如图1所示，基于时空一致性的群体表情情绪识别方法，包括如下步骤：

[0028] 步骤1、对获取的待识别的人群运动场景中的视频图像,进行人脸检测识别得到每帧图像中的每个个体的人脸表情;

[0029] 步骤2、根据每个个体的人脸表情识别个体的情绪是否为恐慌情绪;基于时间一致性计算有部分帧图像中被遮挡个体是否为恐慌情绪,基于空间一致性计算所有帧图像中被遮挡个体是否为恐慌情绪;

[0030] 根据每个个体的人脸表情识别个体的情绪是否为恐慌情绪,可选的,把每帧图像中的每个个体的人脸表情分别输入到训练好的Resnet34网络中,输出得到没有被遮挡的个体的情绪是否为恐慌情绪;

[0031] 步骤3、根据所有个体中为恐慌情绪个体的比例,得到群体恐慌度。根据群体恐慌度可以生成最优的人群疏散方案。

[0032] 人群疏散方案包括疏散路径规划、人群疏散出口选择等。

[0033] 本实施例中,基于时间一致性的个体情绪计算和空间一致性的个体情绪计算来解决人群中的遮挡导致面部表情信息不完整的问题的方法,能够识别出待测视频中所有个体是否处于恐慌情绪,同时恐慌度定量了整个群体处于何种恐慌程度。该方法能够在人群遮挡导致面部表情信息采集不完整的情况下较为准确的得到群体的情绪,识别群体的恐慌程度,从而为人群运动的研究如人群疏散提供了准确的数据支撑,具有简化程度高,模块化程度高,模型精度高,准确契合场景等特点。

[0034] 步骤1中,对获取的待识别的人群运动场景中的视频图像,进行人脸检测识别得到每帧图像中的每个个体的人脸表情的方法,包括如下步骤:

[0035] 步骤(1)输入待识别的人群运动场景中的RGB视频。

[0036] 步骤(2)把视频进行分帧处理成一帧一帧的图片并保存。

[0037] 步骤(3)针对分帧处理的每帧图像进行人脸检测。

[0038] 步骤(4)对每张图片检测到的人脸进行裁剪并保存。

[0039] 步骤(5)输入裁剪后的人脸图片到深度残差网络(Resnet34)结构中得到人脸表情识别结果。

[0040] 所述步骤(1)中,获取待识别的人群运动场景中的RGB视频;

[0041] 可选的,可以采用`cap=cv.VideoCapture()`函数输入待识别的视频,该待识别的视频可以为一段人群运动的场景中能够捕捉到人群面部表情的RGB视频。

[0042] 所述步骤(2)中,把视频进行分帧处理,得到一帧一帧的图片并保存;

[0043] 具体的,可以采用`face_detect_demo(frame)`函数把输入的视频进行分帧处理,得到每帧图片后保存在文件夹里。

[0044] 所述步骤(3)中,针对分帧处理的每帧图像进行人脸检测;

[0045] 可选的,可以调用OpenCV的API接口进行图片中的人脸检测,具体的,采用`face_detect=cv.CascadeClassifier`函数调用OpenCV内部的“haarcascade_frontalface_default.xml”人脸检测模块。

[0046] 可选的,人脸检测算法可以采用基于haar特征提取算法实现。

[0047] haar算法的计算特征是用一张图片中一块区域内黑色的值减去白色的值。本实施例中,可以采用积分图作为一种加速计算的方法,坐标A(x,y)的积分图是其左上角的所有像素之和,积分图的定义如下:

[0048]
$$I(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x', y') \quad (1)$$

[0049] 其中 $I(x, y)$ 表示积分图, $i(x, y)$ 表示原始图像。积分图中每一点 (x, y) 都是这个点对应左上角区域所有值的和,并且只需要遍历一遍图像就可得到积分图,对于任意一点 (x, y) ,积分图的计算公式为:

[0050]
$$I(x, y) = G(x, y) + I(x-1, y) + I(x, y-1) - I(x-1, y-1) \quad (2)$$

[0051] 其中, I 表示点 (x, y) 的积分值, G 表示点 (x, y) 的灰度值。

[0052] 本实施例中,利用积分图可以很高效的计算出图像中的特征区域。

[0053] 所述步骤(4)中,分别对每张图片检测到的人脸进行裁剪,得到只包含人脸的图片,将裁减后的图片保存。

[0054] 对检测到的人脸进行裁剪,得到只包含人脸的图片,可选的,可以采用`cv.rectangle()`函数设置矩形框对检测到的人脸进行框选操作。采用`cv.imwrite()`函数对框选出来的人脸进行裁剪操作。

[0055] 对检测到的人脸进行裁剪,得到只包含人脸的图片的步骤可以如下:

[0056] (4-1)对进行人脸检测后每张图片,设置与人脸大小相近的矩形框对检测到的人脸进行框选。

[0057] 具体的,将进行人脸检测后图片转为灰度,将灰度化后的图片进行直方均匀化处理,在得到图片矩阵中检测脸的区域,然后把脸部区域用矩形框框出来。

[0058] (4-2)把每张图片里框选出来的人脸裁剪出来并标号,按照标号的顺序输出一张图片中每一个人脸。

[0059] 所述步骤(5)中,输入裁剪后的人脸图片到深度残差网络结构中得到人脸表情的识别结果。

[0060] 其中,深度残差网络可以为Resnet34网络。

[0061] 本实施例中,采用训练好的Resnet34网络结构,可以准确地识别人脸表情的情绪结果。

[0062] 可选的,本实施例的Resnet34网络结构包括卷积层、残差模块、池化层、全连接层以及Softmax分类器,所述卷积层、残差模块、池化层、全连接层以及Softmax分类器依次连接。

[0063] 具体的,选中一张人脸的图片作为输入,放到经过训练好的Resnet34网络结构中,经过卷积层、池化层和全连接层最终输出这个个体情绪的量化的值和标签。

[0064] 卷积层用于提取图像特征及输入图片中的信息,

[0065] 其中,图像特征是由图像中的每个像素通过组合或者独立的方式所体现,比如图片的纹理特征,颜色特征。

[0066] 池化层用于对卷积层中提取的特征进行挑选。可选的,池化操作可以为最大池化和平均池化,池化层是由 $n*n$ 大小的矩阵窗口滑动来进行计算的,求 $n*n$ 大小的矩阵中的最大值和平均值。

[0067] 全连接层用于将池化层的所有特征矩阵转化成一维的特征大向量,全连接层放在Resnet34网络结构中的最后,用于对图片进行分类到了全连接层,然后输出结果。

[0068] 针对人脸表情识别,传统方法主要依赖人工提取特征的优劣,算法的鲁棒性较差,

而传统卷积神经网络无法提取到更深层次的图像特征,因此本实施例将采用深度残差网络进行人脸表情识别。

[0069] ResNet网络主要由残差模块组成,将残差模块的输出送入到全连接层进行特征的融合,最后由Softmax分类器进行分类。

[0070] 深度卷积网络自然的整合了低中高不同层次的特征,特征的层次可以靠加深网络的层次来丰富。从而,在构建卷积网络时,网络的深度越高,可抽取的特征层次就越丰富,所以一般会倾向于使用更深层次的网络结构,以便取得更高层次的特征。但是在使用深层次的神经网络结构时会遇到以下问题,如梯度消失,梯度爆炸问题和网络退化的问题。深度残差网络与传统卷积神经网络相比,在网络中引入残差模块,该模块的引入有效地缓解了网络模型训练时反向传播的梯度消失问题,进而解决了深层网络难以训练和性能退化的问题。

[0071] Softmax分类器常常用于多分类识别中,它的输入是一个任意实数的向量,输出是一个向量,其中每个元素的取值在(0,1)之间,且其和为1。假设有一个数组,其Softmax表达式为:

$$[0072] \quad S_i = \frac{e^i}{\sum_j e^j} \quad (3)$$

[0073] 其中, S_i 是表示Softmax的取值, e^i 表示第i个元素, $\sum_j e^j$ 表示所有元素之和。

[0074] 图2是残差网络的基本单元即为残差模块的结构,可以看到x是这一层残差块的输入,也称F(x)为残差,x为输入值,F(x)是经过第一层线性变化并激活后的输出,该图表示在残差网络中,第二层进行线性变化之后激活之前,F(x)加入了这一层输入值x,然后再进行激活后输出。在第二层输出值激活前加入x。刚开始输入的x,按照常规的神经网络进行权值叠加然后通过激活函数,在次权值叠加后,此时再把输入的信号和此时的输出叠加然后通过激活函数,最具代表性的是一个直接到达输出前的连线,这条路径称作shortcut连接,也称为捷径。在线性拟合中的残差说的是数据点距离拟合直线的函数值的差,这里的x就是我们的拟合的函数,而H(x)的就是具体的数据点,那么通过训练使得拟合的值加上F(x)的就得到具体数据点的值,因此F(x)代表残差。

[0075] 由于是群体的视频经过分帧操作生成的图片,一张图片中会包含多个个体的人脸,按照之前的标号的顺序依次输入到Resnet34网络结构中,分别得到一张图片中每个个体的人脸表情所反映的真实情绪。

[0076] 进一步地,还包括网络训练的步骤,如下:

[0077] (5-1) 获取人脸表情数据集作为训练集和测试集;

[0078] 本实施例中,Resnet34网络结构使用RAFDB人脸表情数据集进行训练和测试。该数据集分为原始图像和处理后的对齐的面部图像。对齐后的图像包含12271个训练样本和3068个对齐后的测试样本。使用该数据集进行预训练,训练了1500个epoch得到预训练模型,使用该模型在测试集上达到了74%的准确率。

[0079] (5-2) 将数据集中的图像作为输入,以个体情绪为输出,输入至Resnet34网络结构进行训练得到训练后的Resnet34网络结构。

[0080] 进一步地,根据训练后的Resnet34网络结构,对待检测的裁剪后的人脸图片进行识别,并采用矩阵的形式表示个体情绪识别的结果,得到情绪识别结果矩阵R。

[0081] 可选的,用矩阵的形式来表示以上的个体情绪识别的结果,假设把一个视频分为m帧,该视频中的群体中有n个个体,得到的情绪识别结果矩阵R如下:

$$[0082] \quad R = \begin{pmatrix} r_{11} & \cdots & r_{1j} & \cdots & r_{1n} \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ r_{i1} & \cdots & r_{ij} & \cdots & r_{in} \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ r_{m1} & \cdots & r_{mj} & \cdots & r_{mn} \end{pmatrix} \quad (4)$$

$$[0083] \quad [r_{i1} \cdots r_{ij} \cdots r_{in}] \quad (5)$$

$$[0084] \quad \begin{bmatrix} r_{1j} \\ \vdots \\ r_{ij} \\ \vdots \\ r_{mj} \end{bmatrix} \quad (6)$$

[0085] 其中,公式(5)表示该群体的每个个体在第i帧这一时刻的情绪识别结果。公式(6)表示个体j在该视频所有帧中的情绪识别结果。 r_{ij} 表示个体j在i帧这一时刻的情绪识别结果,定义 r_{ij} 可能的取值如下:

$$[0086] \quad r_{ij} = \begin{cases} 1 & no \text{ panic} \\ -1 & panic \\ 0 & null \end{cases} \quad (7)$$

[0087] 其中,个体j在i帧这一时刻的情绪识别结果是非恐慌情绪,则记录为 $r_{ij}=1$;个体j在i帧这一时刻的情绪识别结果是恐慌情绪,则记录为 $r_{ij}=-1$;如果个体j在i帧这一时刻的没有检测到面部表情,则记录为 $r_{ij}=0$ 。

[0088] 步骤2中,没有被遮挡的个体的情绪是否为恐慌情绪也可以采用Resnet34网络识别,把每帧图像中的每个个体的人脸表情分别输入到训练好的Resnet34网络中,输出得到没有被遮挡的个体的情绪是否为恐慌情绪;

[0089] 训练过程包括:构建包括人脸表情以及对应情绪的训练集,以每个个体的人脸表情作为输入,个体情绪作为输出训练,得到训练后的Resnet34网络。

[0090] 步骤2中,基于时间一致性的个体情绪计算和空间一致性的个体情绪,计算得到被遮挡个体的情绪。

[0091] 人群运动场景中,恐慌的消极情绪是最具指导性意义的情绪,所以对恐慌情绪的识别也就具有重要的意义。在人群运动的真实场景中,由于人群的运动属性和密集性,部分个体会在某个或某几个不同的时间段处于面部表情被遮挡的状态中,导致无法采集到该个体某一帧或某几帧的面部表情,或者部分个体会在整个时间段全部处于面部表情被遮挡的状态中,导致无法采集到该个体任何一帧的面部表情,为此,本实施例引入了时间一致性个体情绪的计算和空间一致性个体情绪的计算的概念,在面部信息采集不完整或者未采集到的情况下辅助得出一个个体的情绪。

[0092] 基于时间一致性的个体情绪计算即为在不同帧下面部表情所反映的情绪会有所

不同,以及由于遮挡导致某个个体A的面部表情在某帧或某几帧时采集不到的情况下解决个体情绪识别的问题。

[0093] 在一些实施例中,可选的,基于时间一致性计算有部分帧图像中被遮挡个体是否为恐慌情绪的方法,具体为:在能够采集到个体a完整面部表情信息的所有帧图片中,识别为恐慌情绪的帧数的比例大于设定的比例阈值,即判定该个体a的情绪在被识别的人群运动的视频中为恐慌情绪,否则为非恐慌情绪。

[0094] 可选的,设定的比例阈值不小于50%。

[0095] 具体的,本实施例中,个体a可以采集到完整面部表情信息的所有帧数中识别为恐慌情绪的帧数占该个体a可以采集到完整面部表情信息的所有帧数的一半以上,这时即可判定该个体在人群运动的视频中处于恐慌情绪。

[0096] 本实施例中,基于时间一致性的对个体面部表情情绪进行识别,解决了某个个体在某个或某几个不同的时间段处于面部表情被遮挡的状态中,导致无法采集到该个体某一帧或某几帧的面部表情的问题。

[0097] 可选的,本实施例中,基于时间一致性的个体情绪计算方法可以通过定义恐慌指标实现。

[0098] 首先是基于时间一致性的个体面部表情情绪识别,若某个个体a在某个或某几个不同的时间段处于面部表情被遮挡的状态中,导致无法采集到该个体某一帧或某几帧的面部表情。

[0099] 为了解决这个问题,定义一个个体是否判定为恐慌情绪的指标P,对于第j个个体,它的恐慌情绪的指标 P_j 的计算如下:

$$[0100] \quad P_j = \frac{\sum_{i=1}^m |r_{ij}|}{\sum_{i=1}^m |r_{ij}|} * 100\% \quad (8)$$

[0101] 其中,m是该视频被分成的所有帧数,分母代表个体j可以采集到完整面部表情信息即可以准确识别出情绪结果的人脸图片的所有帧数,分子代表个体j人脸照片识别为恐慌情绪的帧数。

[0102] 某个个体的最终情绪为r,对于个体j,有个体j的最终情绪为 r_j , r_j 的表达式如下:

$$[0103] \quad r_j = \begin{cases} -1 & P \geq 50\% \\ 1 & P < 50\% \end{cases} \quad (9)$$

[0104] 本实施中,设定 $P \geq 50\%$ 即 $r_j = -1$ 时,表示一个个体可以采集到完整面部表情信息的所有帧数中识别为恐慌情绪的帧数占该个体可以采集到完整面部表情信息的所有帧数的一半以上,这时即可定义该个体在人群运动的视频中处于恐慌情绪;如果 $P < 50\%$ 即 $r_j = 1$ 时,表示一个个体可以采集到完整面部表情信息的所有帧数中识别为恐慌情绪的帧数占该个体可以采集到完整面部表情信息的所有帧数的一半以下,这时即可定义该个体在人群运动的视频中处于非恐慌情绪。

[0105] 在一些实施例中,基于空间一致性计算所有帧图像中被遮挡个体是否为恐慌情绪

的方法,可以为:将每一帧图片转化为深度图;计算每一帧图像中所有已经通过计算得到情绪结果的个体与被遮挡个体j的距离,根据距离得到个体j的情绪受群体中其他个体的影响值之和,根据影响值确定是否为恐慌情绪,其他个体对个体j的恐慌情绪的影响为负值,非恐慌情绪的影响为正值。

[0106] 当某个个体j在所有帧都处于被遮挡的状态中时,通过把每一帧图片转化为深度图,确定帧图像中所有已经通过计算得到情绪结果的个体与个体j的距离,根据个体之间的距离得到周围个体对被遮挡个体j的恐慌影响,被遮挡个体的情绪是其他所有个体对遮挡个体恐慌影响的加权和。

[0107] 个体j的情绪受群体中其他个体的影响,且这个影响与二者的距离有关。其中其他个体对个体j的恐慌情绪的影响为负值,非恐慌情绪的影响为正值,最终周围个体对个体j的情绪的影响为各个个体的影响求和。

[0108] 深度图像(depth image)也被称为距离影像(range image),是指将从图像采集器到场景中各点的距离(深度)作为像素值的图像,它直接反映了景物可见表面的几何形状。深度图像经过坐标转换可以计算为点云数据,有规则及必要信息的点云数据也可以反算为深度图像数据。深度图是一种图像或图像通道,其中包含与场景对象的表面到视点的距离有关的信息,用于模拟的3D形状或重建它们。

[0109] 深度图像可采用伪灰度、点云或者网格表示。局部特征提取作为三维模型建模和三维目标识别的关键环节,包含特征检测和特征描述两个部分。本实例中,深度图可以得到个体之间的纵向距离,在由可以直接得到的横向距离根据勾股定理即可计算出任意两个个体的距离。

[0110] 深度图是由相机拍摄的,其每个像素值代表的是物体到相机xy平面的距离。单位为mm。通过机器学习的方法,将大量训练图像和对应的深度图像输入到定义好的模型中进行有监督的学习,模型训练完成后,便可将实际待测试的图像输入到训练好的模型中进行深度的计算。

[0111] 深度数据流所提供的图像帧中,每一个像素点代表的是在深度感应器的视野中,该特定的(x,y,z)坐标处物体到离摄像头空间最近的物体的距离(以毫米为单位)。

[0112] 深度图中,个体s与个体j之间的距离 d_{sj} ,如下:

$$[0113] \quad d_{sj} = \sqrt{(x_s - x_j)^2 + (y_s - y_j)^2 + (z_s - z_j)^2} \quad (10)$$

[0114] 其中,j为面部表情在整个时间段全部处于被遮挡的状态中的个体,j的位置坐标为 (x_j, y_j, z_j) 。s为整个群体中可以通过人脸表情识别出情绪的任意一个个体。定义s是j的邻居,s的位置坐标为 (x_s, y_s, z_s) 。

[0115] 其中, r_s 表示个体s的最终情绪,j为面部表情在整个时间段全部处于被遮挡的状态中的个体,s为整个群体中可以通过人脸表情识别出情绪的个体。 w_{sj} 是个体s对个体j情绪影响的权重,这个权重与两个个体之间的距离 d_{sj} 成反比,两个个体之间的距离越大,这个影响的权重就越小;两个个体之间的距离越小,这个影响的权重就越大。 w_{sj} 的表示如下:

$$[0116] \quad w_{sj} \propto \frac{1}{d_{sj}} \quad (11)$$

[0117] 如果一个个体的面部表情在整个时间段全部处于被遮挡的状态中,那么该个体的情绪受整个群体中所有已识别出情绪的个体的影响。定义一个面部表情在整个时间段全部处于被遮挡的状态中的个体的情绪是否判定为恐慌情绪的指标V,对于个体j,它的恐慌情绪的指标V的计算如下:

$$[0118] \quad V = \sum_{s \in neighbor(j)} w_{sj} r_s \quad (12)$$

[0119] 其中,个体s是可以通过人脸面部表情直接识别出情绪的个体和通过时间一致性计算得出最终情绪个体。 $r_s = 1$,表示该个体s的情绪是非恐慌情绪,则个体s对个体j的非恐慌情绪影响的值为正值; $r_s = -1$,表示该个体s的情绪是恐慌情绪,则个体s对个体j的恐慌情绪影响的值为负值。

[0120] 本实施例中,定义了当 $V \leq 0$ 时,个体j的情绪判定为恐慌情绪;当 $V > 0$ 时,个体j的情绪判定为非恐慌情绪。

[0121] 基于时间一致性的个体情绪计算和空间一致性的个体情绪计算在公式(4)-公式(14)和矩阵R上进行计算,通过该方法可以准确地得出在由于有遮挡导致人脸信息采集不完整甚至采集不到的情况下的个体的最终情绪。

[0122] 本实施例中,基于空间一致性的个体面部表情情绪识别,解决了某个个体在整个时间段全部处于面部表情被遮挡的状态中,导致无法采集到该个体任何一帧的面部表情的问题。

[0123] 步骤3中,通过定义群体恐慌度的概念计算群体人脸表情情绪。得到个体的最终情绪后,根据个体的恐慌情绪个数占群体的百分比定义了群体恐慌度,来描述一个群体的恐慌情绪的指标,最终得到群体的真实情绪,群体恐慌度具体如下:

[0124] 对于待检测的群体,群体恐慌度包括无恐慌、轻度恐慌、中度恐慌和重度恐慌,通过描述人群恐慌程度指标S判定,其中,人群恐慌程度指标S为:

$$[0125] \quad S = \frac{n_p}{n} * 100\% \quad (15)$$

[0126] 其中,n表示该视频中人群的总数, n_p 表示该视频中人脸表情情绪识别为恐慌情绪的人数。

[0127] 可设置的,如表1所示,为本实施例中人群恐慌度的定义以及判定标准,无恐慌指待检测的人群中没有个体带有恐慌情绪,即 $S = 0$;轻度恐慌指待检测的人群中是恐慌情绪的人数占人群的总人数的0-30%,即 $S \in (0, 0.30]$;中度恐慌指待检测的人群中是恐慌情绪的人数占人群的总人数的30%-60%,即 $S \in (0.30, 0.60]$;重度恐慌指待检测的人群中是恐慌情绪的人数占人群的总人数的60%-100%,即 $S \in (0.60, 1]$ 。

[0128] 表1

[0129]	恐慌度	无恐慌	轻度恐慌	中度恐慌	重度恐慌
	表达式	$S = 0$	$S \in (0, 0.30]$	$S \in (0.30, 0.60]$	$S \in (0.60, 1]$

[0130] 通过上述方法可以确定划分为一个群体的人群的整体情绪处在哪一种恐慌程度中,根据恐慌程度可以进行疏散方案的部署,能够最大限度的保证人群的安全。

[0131] 实施例2

[0132] 基于实施例1,本实施例提供基于时空一致性的群体表情情绪识别系统,包括:

[0133] 人脸识别模块:被配置为用于对获取的待识别的人群运动场景中的视频图像,进行人脸检测识别得到每帧图像中的每个个体的人脸表情;

[0134] 个体恐慌判断模块:被配置为根据每个个体的人脸表情识别个体的情绪是否为恐慌情绪;基于时间一致性计算有部分帧图像中被遮挡个体是否为恐慌情绪,基于空间一致性计算所有帧图像中被遮挡个体是否为恐慌情绪;

[0135] 群体恐慌度判断模块:被配置为用于根据所有个体中为恐慌情绪个体的比例,得到群体恐慌度。

[0136] 实施例3

[0137] 本实施例提供一种电子设备,包括存储器和处理器以及存储在存储器上并在处理器上运行的计算机指令,所述计算机指令被处理器运行时,完成实施例1的方法所述的步骤。

[0138] 实施例4

[0139] 本实施例提供一种计算机可读存储介质,用于存储计算机指令,所述计算机指令被处理器执行时,完成实施例1的方法所述的步骤。

[0140] 本公开所提出的电子设备可以是移动终端以及非移动终端,非移动终端包括台式计算机,移动终端包括智能手机(Smart Phone,如Android手机、IOS手机等)、智能眼镜、智能手表、智能手环、平板电脑、笔记本电脑、个人数字助理等可以进行无线通信的移动互联网设备。

[0141] 应理解,在本公开中,该处理器可以是中央处理单元CPU,该处理器还可以是其他通用处理器、数字信号处理器DSP、专用集成电路ASIC,现成可编程门阵列FPGA或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0142] 该存储器可以包括只读存储器和随机存取存储器,并向处理器提供指令和数据、存储器的一部分还可以包括非易失性随机存储器。例如,存储器还可以存储设备类型的信息。

[0143] 在实现过程中,上述方法的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。结合本公开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。软件模块可以位于随机存储器、闪存、只读存储器、可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器,处理器读取存储器中的信息,结合其硬件完成上述方法的步骤。为避免重复,这里不再详细描述。本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元即算法步骤,能够以电子硬件或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本公开的范围。

[0144] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0145] 在本公开所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其他方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能的划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另外一点,所显示或讨论的相互之间的耦合或者直接耦合或者通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性、机械或其它的形式。

[0146] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读存储介质中。基于这样的理解,本公开的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案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)执行本公开各个实施例所述方法的全部或部分步骤。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)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0147] 以上所述仅为本公开的优选实施例而已,并不用于限制本公开,对于本领域的技术人员来说,本公开可以有各种更改和变化。凡在本公开的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本公开的保护范围之内。

[0148] 上述虽然结合附图对本公开的具体实施方式进行了描述,但并非对本公开保护范围的限制,所属领域技术人员应该明白,在本公开的技术方案的基础上,本领域技术人员不需要付出创造性劳动即可做出的各种修改或变形仍在本公开的保护范围以内。



图1

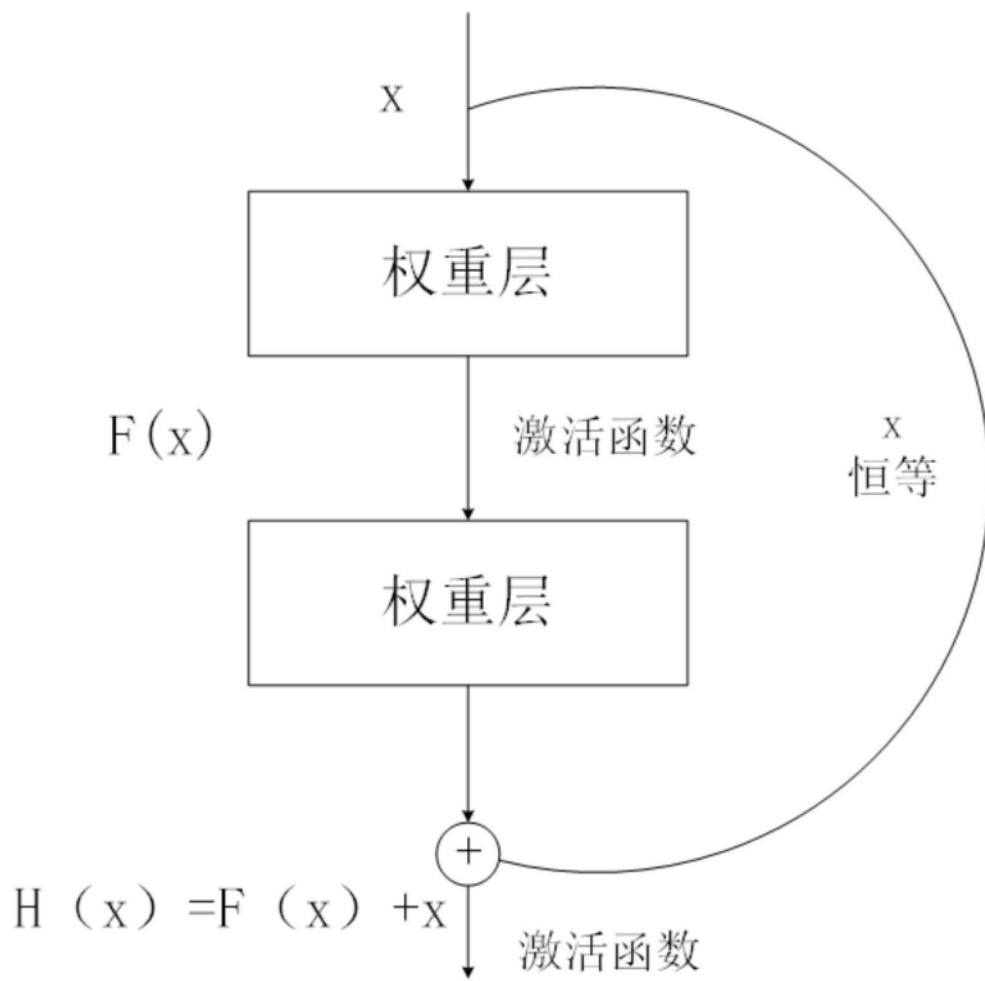


图2