



(12) 发明专利

(10) 授权公告号 CN 111722717 B

(45) 授权公告日 2024.03.15

(21) 申请号 202010564421.2

(22) 申请日 2020.06.18

(65) 同一申请的已公布的文献号  
申请公布号 CN 111722717 A

(43) 申请公布日 2020.09.29

(73) 专利权人 歌尔科技有限公司  
地址 266104 山东省青岛市崂山区北宅街道投资服务中心308室

(72) 发明人 李晶晶

(74) 专利代理机构 深圳市世纪恒程知识产权代理事务所 44287  
专利代理师 王韬

(51) Int. Cl.  
G06F 3/01 (2006.01)  
G06V 40/10 (2022.01)

G06V 10/764 (2022.01)  
G06V 10/774 (2022.01)  
G06V 10/82 (2022.01)  
G06N 3/0464 (2023.01)  
G06N 3/08 (2023.01)

(56) 对比文件

CN 108596079 A, 2018.09.28  
CN 108898045 A, 2018.11.27  
CN 108932500 A, 2018.12.04  
CN 109033978 A, 2018.12.18  
CN 110245593 A, 2019.09.17  
CN 111027487 A, 2020.04.17  
CN 111160114 A, 2020.05.15  
JP 2010271944 A, 2010.12.02

审查员 陈莉

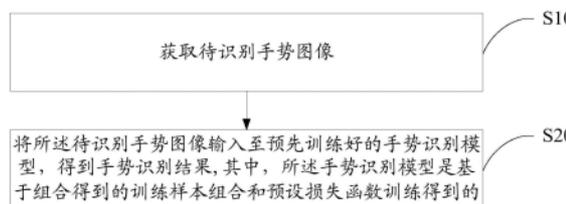
权利要求书3页 说明书12页 附图2页

(54) 发明名称

手势识别方法、装置及计算机可读存储介质

(57) 摘要

本发明公开了一种手势识别方法。该手势识别方法包括：获取待识别手势图像；将所述待识别手势图像输入至预先训练好的手势识别模型，得到手势识别结果；其中，所述手势识别模型是基于组合得到的训练样本组合和预设损失函数训练得到的。本发明能够实现对图像差异度极小的手势的精确识别，提高手势识别模型的鲁棒性和手势识别结果的准确性。



1. 一种手势识别方法,其特征在于,所述手势识别方法包括以下步骤:  
 获取待识别手势图像;  
 将所述待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果;  
 其中,所述手势识别模型是基于组合得到的训练样本组合和预设损失函数训练得到的;

其中,所述获取待识别手势图像的步骤之前,还包括:  
 获取训练样本数据,所述训练样本数据包括训练样本图像及其类别标签;  
 根据所述类别标签对所述训练样本图像进行组合,得到训练样本组合;  
 利用所述训练样本组合和预设损失函数对预设手势识别模型进行训练,得到所述手势识别模型;

所述利用所述训练样本组合和预设损失函数对预设手势识别模型进行训练,得到所述手势识别模型的步骤包括:

将所述训练样本组合输入至特征提取层进行特征提取,得到各训练样本图像对应的特征向量;

根据所述特征向量计算各训练样本组合中同类样本特征之间的第一余弦角度和不同类样本特征之间的第二余弦角度;

将所述第一余弦角度、所述第二余弦角度代入预设损失函数,计算得到损失值;

根据所述损失值进行特征提取层的后向传播和参数更新,直至训练完成,得到训练好的特征提取层;

利用所述训练好的特征提取层和所述训练样本数据对分类器进行训练,得到训练好的分类器,其中,所述手势识别模型包括所述训练好的特征提取层和所述训练好的分类器;

所述预设损失函数如下:

$$J = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\cos(\text{angle\_same})-0.1}}{e^{\cos(\text{angle\_same})-0.1} + \sum_{\text{angle\_diff}} e^{\cos(\text{angle\_diff})}};$$

其中, $J$ 为损失值, $N$ 表示批处理数据的个数, $\text{angle\_same} = 0.9\theta_{\text{same}} + 0.3$ ,  
 $\text{angle\_diff} = \theta_{\text{diff}} + 0.2$ , $\theta_{\text{same}}$ 为第一余弦角度, $\theta_{\text{diff}}$ 为第二余弦角度。

2. 如权利要求1所述的手势识别方法,其特征在于,所述根据所述类别标签对所述训练样本图像进行组合,得到训练样本组合的步骤包括:

根据所述类别标签对相同类别的训练样本图像进行两两组合;

向各个组合中随机加入一个另一类别的训练样本图像,得到训练样本组合。

3. 如权利要求1所述的手势识别方法,其特征在于,所述根据所述损失值进行特征提取层的后向传播和参数更新,直至训练完成,得到训练好的特征提取层的步骤包括:

根据所述损失值进行所述特征提取层的后向传播和参数更新,直至达到预设训练停止条件时,得到初步训练好的特征提取层;

对所述训练样本数据进行困难样本挖掘,基于挖掘得到的困难样本对所述初步训练好的特征提取层进行再次训练,得到训练好的特征提取层。

4. 如权利要求1所述的手势识别方法,其特征在于,所述获取待识别手势图像的步骤包

括：

获取实时图像，通过预先训练好的手势唤醒模型检测所述实时图像中是否存在用户区域；

若所述实时图像中存在用户区域，则对所述实时图像中的用户区域进行裁剪，得到用户区域图像；

通过所述手势唤醒模型检测所述用户区域图像中是否存在手部区域；

若所述用户区域图像中存在手部区域，则对所述用户区域图像中的手部区域进行裁剪，得到待识别手势图像。

5. 如权利要求4所述的手势识别方法，其特征在于，所述对所述用户区域图像中的手部区域进行裁剪，得到待识别手势图像的步骤包括：

获取所述用户区域图像中手部区域的第一属性数据；

根据所述第一属性数据对所述手部区域进行筛选，得到筛选后的手部区域；

对所述筛选后的手部区域进行裁剪，得到待识别手势图像。

6. 如权利要求1至5中任一项所述的手势识别方法，其特征在于，所述将所述待识别手势图像输入至预先训练好的手势识别模型，得到手势识别结果的步骤包括：

判断所述待识别手势图像的数量为一个还是为至少两个；

若所述待识别手势图像的数量为一个，则将所述待识别手势图像输入至预先训练好的手势识别模型，得到手势识别结果；

若所述待识别手势图像的数量为至少两个，则基于预设规则和预先训练好的手势识别模型对所述待识别手势图像进行识别，得到手势识别结果。

7. 如权利要求6所述的手势识别方法，其特征在于，所述基于预设规则和预先训练好的手势识别模型对所述待识别手势图像进行识别，得到手势识别结果的步骤包括：

将各待识别手势图像分别输入至预先训练好的手势识别模型，得到对应的手势类别及其分类概率；

获取各待识别手势图像的第二属性数据，所述第二属性数据包括手部检测概率、像素信息和清晰度信息；

基于所述手势类别的分类概率和所述第二属性数据对各待识别手势图像进行评分，得到各待识别手势图像的综合评分；

根据所述综合评分和各待识别手势图像的手势类别，确定手势识别结果。

8. 如权利要求1至5中任一项所述的手势识别方法，其特征在于，所述手势识别方法还包括：

获取当前应用程序，并获取所述当前应用程序对应的手势类别与操作指令之间的映射关系；

根据所述手势识别结果和所述映射关系，确定目标操作指令；

基于所述目标操作指令控制所述当前应用程序执行对应的操作。

9. 一种手势识别装置，其特征在于，所述手势识别装置包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的手势识别程序，所述手势识别程序被所述处理器执行时实现如权利要求1至8中任一项所述的手势识别方法的步骤。

10. 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上存储有手势识

别程序,所述手势识别程序被处理器执行时实现如权利要求1至8中任一项所述的手势识别方法的步骤。

## 手势识别方法、装置及计算机可读存储介质

### 技术领域

[0001] 本发明涉及人工智能领域,尤其涉及一种手势识别方法、装置及计算机可读存储介质。

### 背景技术

[0002] 随着计算机在社会中的普及,促进人机交互(HCI,Human-Computer Interaction)技术的发展将对计算机的使用产生积极影响,因此,人们越来越重视研究用于跨域人机障碍的新技术。长期以来,手势一直被认为是一种可以提供更自然、更有创意和更直观的与计算机进行通信的交互技术,所以在人机交互中加入手势是一个重要的研究领域。

[0003] 在人机交互领域,手势通常被定义为:人类通过手掌和手指的不同姿势组合形成的具有特定含义信息的集合体。可通过对用户的手势进行识别,最终实现根据手势的类型对计算机或其他电子设备执行对应的控制操作。然而,目前对手势图像进行识别时,往往对差异度较大的手势识别效果较好,而对比于差异度较小的手势识别效果较差,手势识别结果的准确性较差。

### 发明内容

[0004] 本发明的主要目的在于提供一种手势识别方法、装置及计算机可读存储介质,旨在实现对图像差异度极小的手势的精确识别,提高手势识别模型的鲁棒性和手势识别结果的准确性。

[0005] 为实现上述目的,本发明提供一种手势识别方法,所述手势识别方法包括以下步骤:

[0006] 获取待识别手势图像;

[0007] 将所述待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果;

[0008] 其中,所述手势识别模型是基于组合得到的训练样本组合和预设损失函数训练得到的。

[0009] 进一步地,所述获取待识别手势图像的步骤之前,还包括:

[0010] 获取训练样本数据,所述训练样本数据包括训练样本图像及其类别标签;

[0011] 根据所述类别标签对所述训练样本图像进行组合,得到训练样本组合;

[0012] 利用所述训练样本组合和预设损失函数对预设手势识别模型进行训练,得到所述手势识别模型。

[0013] 可选地,所述根据所述类别标签对所述训练样本图像进行组合,得到训练样本组合的步骤包括:

[0014] 根据所述类别标签对相同类别的训练样本图像进行两两组合;

[0015] 向各个组合中随机加入一个另一类别的训练样本图像,得到训练样本组合。

[0016] 可选地,所述预设手势识别模型包括特征提取层和分类器,所述利用所述训练样本组合和预设损失函数对预设手势识别模型进行训练,得到所述手势识别模型的步骤包

括：

[0017] 将所述训练样本组合输入至所述特征提取层进行特征提取,得到各训练样本图像对应的特征向量;

[0018] 根据所述特征向量计算各训练样本组合中同类样本特征之间的第一余弦角度和不同类样本特征之间的第二余弦角度;

[0019] 将所述第一余弦角度、所述第二余弦角度代入预设损失函数,计算得到损失值;

[0020] 根据所述损失值进行特征提取层的后向传播和参数更新,直至训练完成,得到训练好的特征提取层;

[0021] 利用所述训练好的特征提取层和所述训练样本数据对所述分类器进行训练,得到训练好的分类器;其中,所述手势识别模型包括所述训练好的特征提取层和所述训练好的分类器。

[0022] 可选地,所述根据所述损失值进行特征提取层的后向传播和参数更新,直至训练完成,得到训练好的特征提取层的步骤包括:

[0023] 根据所述损失值进行所述特征提取层的后向传播和参数更新,直至达到预设训练停止条件时,得到初步训练好的特征提取层;

[0024] 对所述训练样本数据进行困难样本挖掘,基于挖掘得到的困难样本对所述初步训练好的特征提取层进行再次训练,得到训练好的特征提取层。

[0025] 可选地,所述获取待识别手势图像的步骤包括:

[0026] 获取实时图像,通过预先训练好的手势唤醒模型检测所述实时图像中是否存在用户区域;

[0027] 若所述实时图像中存在用户区域,则对所述实时图像中的用户区域进行裁剪,得到用户区域图像;

[0028] 通过所述手势唤醒模型检测所述用户区域图像中是否存在手部区域;

[0029] 若所述用户区域图像中存在手部区域,则对所述用户区域图像中的手部区域进行裁剪,得到待识别手势图像。

[0030] 可选地,所述对所述用户区域图像中的手部区域进行裁剪,得到待识别手势图像的步骤包括:

[0031] 获取所述用户区域图像中手部区域的第一属性数据;

[0032] 根据所述第一属性数据对所述手部区域进行筛选,得到筛选后的手部区域;

[0033] 对所述筛选后的手部区域进行裁剪,得到待识别手势图像。

[0034] 可选地,所述将所述待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果的步骤包括:

[0035] 判断所述待识别手势图像的数量为一个还是为至少两个;

[0036] 若所述待识别手势图像的数量为一个,则将所述待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果;

[0037] 若所述待识别手势图像的数量为至少两个,则基于预设规则和预先训练好的手势识别模型对所述待识别手势图像进行识别,得到手势识别结果。

[0038] 可选地,所述基于预设规则和预先训练好的手势识别模型对所述待识别手势图像进行识别,得到手势识别结果的步骤包括:

- [0039] 将各待识别手势图像分别输入至预先训练好的手势识别模型,得到对应的手势类别及其分类概率;
- [0040] 获取各待识别手势图像的第二属性数据,所述第二属性数据包括手部检测概率、像素信息和清晰度信息;
- [0041] 基于所述手势类别的分类概率和所述第二属性数据对各待识别手势图像进行评分,得到各待识别手势图像的综合评分;
- [0042] 根据所述综合评分和各待识别手势图像的手势类别,确定手势识别结果。
- [0043] 可选地,所述手势识别方法还包括:
- [0044] 获取当前应用程序,并获取所述当前应用程序对应的手势类别与操作指令之间的映射关系;
- [0045] 根据所述手势识别结果和所述映射关系,确定目标操作指令;
- [0046] 基于所述目标操作指令控制所述当前应用程序执行对应的操作。
- [0047] 此外,为实现上述目的,本发明还提供一种手势识别装置,所述手势识别装置包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的手势识别程序,所述手势识别程序被所述处理器执行时实现如上所述的手势识别方法的步骤。
- [0048] 此外,为实现上述目的,本发明还提供一种计算机可读存储介质,所述计算机可读存储介质上存储有手势识别程序,所述手势识别程序被处理器执行时实现如上所述的手势识别方法的步骤。
- [0049] 本发明提供一种手势识别方法、装置及计算机可读存储介质,通过获取待识别手势图像;然后,将所述待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果;其中,所述手势识别模型是基于组合得到的训练样本组合和预设损失函数训练得到的。本发明中,该预先训练好的手势识别模型是通过对训练样本图像进行组合,进而基于组合得到的训练样本组合对预设手势识别模型进行训练、同时采用特定的预设损失函数来完成手势特征学习,最终训练得到的。对训练样本进行组合和采用特定的预设损失函数的目的均在于最小化类内(即同类之间)的特征距离和最大化类间(即不同类之间)的特征距离,从而可使得训练得到的手势识别模型可实现对图像差异度极小的手势的精确识别,因此,通过上述方式训练得到的手势识别模型进行手势识别时,可大大提高手势识别模型的鲁棒性和手势识别结果的准确性。

## 附图说明

- [0050] 图1为本发明实施例方案涉及的硬件运行环境的终端结构示意图;
- [0051] 图2为本发明手势识别方法第一实施例的流程示意图;
- [0052] 图3为本发明手势识别方法第二实施例的流程示意图。
- [0053] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

## 具体实施方式

- [0054] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。
- [0055] 参照图1,图1为本发明实施例方案涉及的硬件运行环境的终端结构示意图。
- [0056] 本发明实施例终端可以是PC(personal computer,个人计算机),也可以是智能手

机、平板电脑、便携计算机、电子书阅读器、MP3 (Moving Picture Experts Group Audio Layer III, 动态影像专家压缩标准音频层面3) 播放器、MP4 (Moving Picture Experts Group Audio Layer IV, 动态影像专家压缩标准音频层面3) 播放器和智能音箱等终端设备。

[0057] 如图1所示,该终端可以包括:处理器1001,例如CPU(Central Processing Unit, 中央处理器),通信总线1002,用户接口1003,网络接口1004,存储器1005。其中,通信总线1002用于实现这些组件之间的连接通信。用户接口1003可以包括显示屏(Display)、输入单元比如键盘(Keyboard),可选用户接口1003还可以包括标准的有线接口、无线接口。网络接口1004可选的可以包括标准的有线接口、无线接口(如无线保真Wireless-Fidelity,Wi-Fi接口)。存储器1005可以是高速RAM存储器,也可以是稳定的存储器(non-volatile memory),例如磁盘存储器。存储器1005可选的还可以是独立于前述处理器1001的存储装置。

[0058] 本领域技术人员可以理解,图1中示出的终端结构并不构成对终端的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0059] 如图1所示,作为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块以及手势识别程序。

[0060] 在图1所示的终端中,网络接口1004主要用于连接后台服务器,与后台服务器进行数据通信;用户接口1003主要用于连接客户端,与客户端进行数据通信;而处理器1001可以用于调用存储器1005中存储的手势识别程序,并执行以下手势识别方法的各个步骤。

[0061] 基于上述硬件结构,提出本发明手势识别方法各个实施例。

[0062] 本发明提供一种手势识别方法。

[0063] 参照图2,图2为本发明手势识别方法第一实施例的流程示意图。

[0064] 在本实施例中,该手势识别方法包括:

[0065] 步骤S10,获取待识别手势图像;

[0066] 在本实施例中,本发明实施例的终端可以是智能手机、PC、平板电脑、便携计算机、电子书阅读器、MP3播放器、MP4播放器和智能音箱等终端设备。为便于说明,以智能音箱为例进行说明。

[0067] 在本实施例中,先获取待识别手势图像。待识别手势图像的获取方式可以包括但不限于:1)在一定条件下,智能音箱调用摄像头拍摄得到待识别手势图像;2)在一定条件下,智能音箱调用摄像头拍摄得到当前图像,进而对该当前图像进行识别处理,并裁剪得到手部区域的图像(即待识别手势图像)。

[0068] 步骤S20,将所述待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果;其中,所述手势识别模型是基于组合得到的训练样本组合和预设损失函数训练得到的。

[0069] 然后,待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果;其中,手势识别模型是基于组合得到的训练样本组合和预设损失函数训练得到的。

[0070] 作为其中一种识别方式,可以直接将待识别手势图像输入至手势识别模型中,得到手势识别结果,该手势识别结果可以为该待识别手势图像中的手势属于各手势类别的分类概率,当然也可以为该待识别手势图像中的手势的所属的手势类别(即分类概率最大值

所对应的手势类别)。

[0071] 当然,在实际应用场景中,待识别手势图像可能包括多个,而最终只需确定得到一个手势类别,因此,作为另一种识别方式,可基于手势识别模型得到的结果及其他因素进行综合评价,得到最终的手势识别结果。

[0072] 需要说明的是,本发明实施例中的手势识别模型是基于组合得到的训练样本组合和预设损失函数训练得到的。具体的,在手势识别模型的训练过程中,先对训练样本图像进行组合,使得组合得到的训练样本组合中既包括相同类别的训练样本图像,也包括不同类别的训练样本图像,进而基于训练样本组合对预设手势模型进行训练,同时采用特定的预设损失函数来完成手势特征学习,以最小化类内(即同类之间)的特征距离、同时最大化类间(即不同类之间)的特征距离,从而使得训练得到的手势识别模型可实现对图像差异度极小的手势的精确识别,即实现对手势的细粒度分类。

[0073] 本发明实施例提供一种手势识别方法,通过获取待识别手势图像;然后,将待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果;其中,手势识别模型是基于组合得到的训练样本组合和预设损失函数训练得到的。本发明实施例中,该预先训练好的手势识别模型是通过对训练样本图像进行组合,进而基于组合得到的训练样本组合对预设手势识别模型进行训练、同时采用特定的预设损失函数来完成手势特征学习,最终训练得到的。对训练样本进行组合和采用特定的预设损失函数的目的均在于最小化类内(即同类之间)的特征距离和最大化类间(即不同类之间)的特征距离,从而可使得训练得到的手势识别模型可实现对图像差异度极小的手势的精确识别,因此,通过上述方式训练得到的手势识别模型进行手势识别时,可大大提高手势识别模型的鲁棒性和手势识别结果的准确性。

[0074] 进一步地,基于上述第一实施例,提出本发明手势识别方法的第二实施例。

[0075] 在本实施例中,在上述步骤S10之前,该手势识别方法还包括:

[0076] 步骤S30,获取训练样本数据,所述训练样本数据包括训练样本图像及其类别标签;

[0077] 在本实施例中,介绍了手势识别模型的训练过程,具体如下:先获取训练样本数据,训练样本数据包括训练样本图像及其类别标签。其中,训练样本数据可以从ImageNet(一个用于视觉对象识别软件研究的大型可视化数据库)数据库中获取,类别标签用于表征手势的类别,可以为向量的形式,例如,包括手势1-3,可以分别用向量(1,0,0)、(0,1,0)和(0,0,1)来表示。

[0078] 需要说明的是,在获取训练样本数据时,为避免训练样本数据过少,可对训练样本数据进行数据扩增处理,例如,可采用光照变换、色度与饱和度变换等数据扩增手段,具体的扩增方法可参照现有技术。

[0079] 步骤S40,根据所述类别标签对所述训练样本图像进行组合,得到训练样本组合;

[0080] 然后,根据类别标签对训练样本图像进行组合,得到训练样本组合。训练样本组合中既包括相同类别的训练样本图像,也包括不同分类的训练样本图像,以便于在基于训练样本组合训练手势识别模型时,通过最小化类内(即同类之间)的特征距离、最大化类间(即不同类之间)的特征距离,最终使得训练好的手势识别模型可对图像差异度极小的手势进行精确识别,实现对手势的细粒度分类。

[0081] 作为其中一种组合方式,步骤S40包括:

[0082] 步骤a41,根据所述类别标签对相同类别的训练样本图像进行两两组合;

[0083] 步骤a42,向各个组合中随机加入一个另一类别的训练样本图像,得到训练样本组合。

[0084] 作为其中一种组合方式,可先根据类别标签对相同类别的训练样本图像进行两两组合,然后,向各个组合中随机加入一个另一类别的训练样本图像,得到训练样本组合。

[0085] 当然,在具体实施时,还可以通过其他组合方式进行组合,例如,先根据类别标签对相同类别的训练样本图像进行两两组合,然后根据各个组合的类别标签对不同分类的组合进行两两组合,得到训练样本组合。

[0086] 步骤S50,利用所述训练样本组合和预设损失函数对预设手势识别模型进行训练,得到所述手势识别模型;

[0087] 在对训练样本图像进行组合得到训练样本组合之后,利用训练样本组合和预设损失函数对预设手势识别模型进行训练,得到手势识别模型。其中,该预设损失函数与同类手势的特征向量和不同类手势的特征向量相关,以用于最小化类内(即同类之间)的特征距离、同时最大化类间(即不同类之间)的特征距离,从而使得训练得到的手势识别模型可实现对图像差异度极小的手势的精确识别,即实现对手势的细粒度分类。

[0088] 其中,预设手势识别模型实质上是一个精简的低计算量浅层卷积神经网络,可基于现有的MobileNetV1模型裁剪得到。MobileNetV1,是针对移动端以及嵌入式视觉的应用提出的一类模型,是基于一种流线型结构使用深度可分离卷积来构造轻量级权重深度神经网络。预设手势识别模型包括特征提取层和分类器。其中,特征提取层用于对待识别手势图像进行特征提取,分类器用于基于特征提取层的输出对待识别手势图像进行分类。

[0089] 具体的,步骤S50包括:

[0090] 步骤a51,将所述训练样本组合输入至所述特征提取层进行特征提取,得到各训练样本图像对应的特征向量;

[0091] 本实施例介绍了手势识别模型的训练过程,具体如下:

[0092] 将训练样本组合输入至特征提取层进行特征提取,得到各训练样本图像对应的特征向量。以训练样本组合包括两个相同类别和一个不同类别的训练样本图像为例进行说明,将两个相同类别的训练样本图像中的其中一个训练样本图像记为original,另一图像记为same,然后将另一类别的训练样本图像记为diff,在经过特征提取后,得到3个特征向量,分别为 $f_{original}$ 、 $f_{same}$ 和 $f_{diff}$ 。

[0093] 步骤a52,根据所述特征向量计算各训练样本组合中同类样本特征之间的第一余弦角度和不同类样本特征之间的第二余弦角度;

[0094] 然后,根据特征向量计算各训练样本组合中同类样本特征之间的第一余弦角度和不同类样本特征之间的第二余弦角度。其中,第一余弦角度 $\theta_{same}$ 的计算公式如下述公式(1),第二余弦角度 $\theta_{diff}$ 的计算公式如下述公式(2):

$$[0095] \quad \theta_{same} = \arccos \frac{f_{original} \cdot f_{same}}{|f_{original}| |f_{same}|} \quad \text{公式 (1);}$$

$$[0096] \quad \theta_{diff} = \arccos \frac{f_{original} \cdot f_{diff}}{|f_{original}| |f_{diff}|} \text{公式 (2)}。$$

[0097] 需要说明的是,当训练样本组合包括两个相同类别和一个不同类别的训练样本图像时,由于该训练样本组合中实际上包括2组不同的训练样本,因此,第二余弦角度对应的应当有2个。因此,在具体实施时,可分别将两个训练样本图像作为original,然后分别计算两个original与另一类别的图像diff之间的余弦角度,可得到2个余弦角度,再取2个余弦角度的平均值作为第二余弦角度。当然,可以随机选取两个相同类别的训练样本图像中的其中一个训练样本图像作为original,然后计算该随机选取得到的original与另一类别的图像diff之间的余弦角度,即为第二余弦角度。

[0098] 步骤a53,将所述第一余弦角度、所述第二余弦角度代入预设损失函数,计算得到损失值;

[0099] 将第一余弦角度、第二余弦角度代入预设损失函数,计算得到损失值。

[0100] 其中,该预设损失函数如下:

$$[0101] \quad J = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\cos(\text{angle\_same})-0.1}}{e^{\cos(\text{angle\_same})-0.1} + \sum_{\text{angle\_diff}} e^{\cos(\text{angle\_diff})}};$$

[0102] 其中,J为损失值,N表示批处理数据的个数, $\text{angle\_same} = 0.9\theta_{\text{same}} + 0.3$ , $\text{angle\_diff} = \theta_{\text{diff}} + 0.2$ 。通过上述预设损失函数,可以使得同类样本之间的特征距离更小、不同类样本之间的特征距离更大,从而可以更好扩大不同类样本之间的区别,使得训练得到的手势识别模型可实现对图像差异度极小的手势的精确识别。

[0103] 步骤a54,根据所述损失值进行特征提取层的后向传播和参数更新,直至训练完成,得到训练好的特征提取层;

[0104] 然后,根据该损失值进行特征提取层的后向传播和参数更新,直至训练完成,得到训练好的特征提取层。具体的,基于该损失值进行特征提取层的后向传播过程,以计算出特征提取层中各网络层参数的梯度,通过随机梯度下降法更新特征提取层中各层的参数;反复迭代更新,直至达到预设训练停止条件时,训练完成,得到最优参数,进而得到含该最优参数的特征提取层,即为训练好的特征提取层。其中,该预设训练停止条件可以为训练数达到预设次数,或损失值小于预设阈值等,可根据实际需要具体设定。

[0105] 在训练特征提取层的过程中,可以根据上述方式得到最终训练好的特征提取层,即根据该损失值进行特征提取层的后向传播和参数更新,直至达到预设训练停止条件时,训练完成,得到训练好的特征提取层。为提升训练效果,更精准获取手势图像的细粒度差异,从而进一步提高手势识别模型的鲁棒性和手势识别结果的准确性,还可以通过困难样本挖掘的方法对特征提取层的权值进行再次训练。

[0106] 进一步地,步骤a54包括:

[0107] 步骤a541,根据所述损失值进行所述特征提取层的后向传播和参数更新,直至达到预设训练停止条件时,得到初步训练好的特征提取层;

[0108] 步骤a542,对所述训练样本数据进行困难样本挖掘,基于挖掘得到的困难样本对所述初步训练好的特征提取层进行再次训练,得到训练好的特征提取层。

[0109] 本实施例中,可以先根据损失值进行特征提取层的后向传播和参数更新,直至达

到预设训练停止条件时,得到初步训练好的特征提取层。然后,采用困难样本挖掘的方法对该初步训练好的特征提取层的权值进行再次训练。具体的,先对训练样本数据进行困难样本挖掘,然后基于挖掘得到的困难样本对初步训练好的特征提取层进行再次训练,得到训练好的特征提取层。

[0110] 在进行困难样本挖掘时,所遵循的原则公式为:

$$[0111] \quad J_{change} = -\frac{1}{N} \sum_{i=1}^N \sum_t p_t (1 - q_t)^2 \log(q_t)$$

[0112] 其中, $P_t$ 为手势识别人工标注的one-hot类别标签(即采用独热编码的方式采用N位来对N个类别标签进行编码,属于哪一类即设定对应位为1,其他位为0,得到对应的向量标签), $q_t$ 为当前模型预测的结果,当训练样本图像被错分时, $q_t$ 会很小, $(1 - q_t)^2$ 接近于1,此时样本为困难样本,手势别网络的损失函数受到的影响不大;当图片被正确分时, $(1 - q_t)^2$ 接近0,此时样本为困难样本,将减少对手势网络模型参数的训练。

[0113] 本实施例中,通过困难样本挖掘的方法对特征提取层的权值进行再次训练,可以加重困难样本的学习,提升训练效果,从而使得训练得到的手势识别模型可以更精准地获取待识别手势图像的细粒度差异,进一步提高手势识别模型的鲁棒性和手势识别结果的准确性。

[0114] 步骤a55,利用所述训练好的特征提取层和所述训练样本数据对所述分类器进行训练,得到训练好的分类器;其中,所述手势识别模型包括所述训练好的特征提取层和所述训练好的分类器。

[0115] 在特征提取层训练完成之后,利用训练好的特征提取层和训练样本数据对分类器进行训练,得到训练好的分类器;其中,手势识别模型包括训练好的特征提取层和训练好的分类器。

[0116] 在训练分类器时,即将训练好的特征提取层冻结,然后将训练样本数据输入至由训练好的特征提取层和分类器构成的手势识别模型中,以利用训练样本数据对分类器进行训练,得到训练好的分类器。手势识别模型的具体训练过程可参照下述第二实施例。

[0117] 本实施例中,在训练手势识别模型的过程中,通过训练样本组合对预设手势模型进行训练,同时采用特定的预设损失函数来完成手势特征学习,其目的均在于最小化类内(即同类之间)的特征距离和最大化类间(即不同类之间)的特征距离,从而可使得训练得到的手势识别模型可实现对图像差异度极小的手势的精确识别,可大大提高手势识别模型的鲁棒性和手势识别结果的准确性。此外,在对手势识别模型中的特征提取层进行训练时,通过困难样本挖掘的方法对初步训练好的特征提取层的权值进行再次训练,可以加重困难样本的学习,提升训练效果,从而使得训练得到的手势识别模型可以更精准地获取待识别手势图像的细粒度差异,进一步提高手势识别模型的鲁棒性和手势识别结果的准确性。

[0118] 进一步地,基于上述第一实施例,提出本发明手势识别方法的第三实施例。

[0119] 在本实施例中,步骤S10包括:

[0120] 步骤a11,获取实时图像,通过预先训练好的手势唤醒模型检测所述实时图像中是否存在用户区域;

[0121] 在终端的使用过程中,可在启动手势控制模式时,每隔预设时间(如1s)获取实时图像,即通过终端的摄像头拍摄得到实时图像。然后,通过预先训练好的手势唤醒模型检测

实时图像中是否存在用户区域。

[0122] 其中,手势唤醒模型可以用于识别图像中的用户区域和手部区域,其训练过程如下:获取训练样本集,训练样本集中包括训练样本图像(为便于与手势识别模型的训练样本图像进行区分,后续称为训练图像)及其标注标签,其中,标注标签包括用户区域和手部区域的框选(即位置框选)和标注(即类型标注);然后,利用训练样本集对预设手势唤醒模型进行训练,得到训练好的手势唤醒模型。其中,预设手势唤醒模型也可以基于现有的MobileNetV1裁剪得到的,由于现有的MobileNetV1模型内存占用率较高,实时性难以保证,因此,可通过测试实验对MobileNetV1的部分卷积层进行裁剪,同时在其网络末端3层特征图上添加检测结构,其中,该检测结构用于对训练图像中的区域是否有目标(包括用户区域和手部区域)、目标所属类别进行检测判断,以得到预设手势唤醒模型。

[0123] 需要说明的是,在获取训练样本集时,为避免训练样本集中的训练图像过少,可对训练样本集中的训练图像进行数据扩增处理,例如,可采用左右翻转、背景融合、光照变换和色度与饱和度变换等数据扩增手段,具体的数据扩增方法可参照现有技术。

[0124] 若所述实时图像中存在用户区域,则执行步骤a12:对所述实时图像中的用户区域进行裁剪,得到用户区域图像;

[0125] 若检测到实时图像中存在用户区域,则可以先对实时图像中的用户区域进行裁剪,得到用户区域图像,可便于后续缩小手部区域识别的范围。同时,通过上述方式,先识别用户区域,再识别用户区域图像中的手部区域,进而对手部区域的手部区域图像进行识别,可避免由于用户距离终端太远时手部区域过小可能无法被检测到的问题,同时,可在一定程度上增加用户在室内活动的自由度且人机交互方式更加友好。

[0126] 若实时图像中不存在用户区域,则直接通过手势唤醒模型检测该实时图像中是否存在手部区域。

[0127] 步骤a13,通过所述手势唤醒模型检测所述用户区域图像中是否存在手部区域;

[0128] 然后,通过手势唤醒模型检测用户区域图像中是否存在手部区域。即,将用户区域图像输入至手势唤醒模型中,得到输出结果,输出结果包括是否存在手部区域的结果及检测到的手部区域的位置。

[0129] 若所述用户区域图像中存在手部区域,则执行步骤a44:对所述用户区域图像中的手部区域进行裁剪,得到待识别手势图像。

[0130] 若用户区域图像中存在手部区域,则对用户区域图像中的手部区域进行裁剪,得到待识别手势图像。

[0131] 进一步地,步骤a14包括:

[0132] 步骤a141,获取所述用户区域图像中手部区域的第一属性数据;

[0133] 步骤a142,根据所述第一属性数据对所述手部区域进行筛选,得到筛选后的手部区域;

[0134] 步骤a143,对所述筛选后的手部区域进行裁剪,得到待识别手势图像。

[0135] 本实施例中,由于待识别手势图像可能不合格,导致最终无法识别到手势,因此,在检测到用户区域图像中包括手部区域之后,可先获取用户区域图像中手部区域的第一属性数据,其中,第一属性数据可以包括但不限于:清晰度信息和完整性信息。然后,根据第一属性数据对手部区域进行筛选,得到筛选后的手部区域。例如,可以根据清晰度信息筛选出

清晰度较高的区域(即清晰度大于预设清晰度阈值所对应的手部区域),根据完整性信息(手部区域中的人手是否有部分位于摄像头视角外)筛选出完整的手部区域。最后,对筛选后的手部区域进行裁剪,得到待识别手势图像。

[0136] 通过上述方式,可对检测得到的手部区域中的不合理区域进行滤除,避免浪费不必要的计算机资源,同时也可以避免不合理的手部区域对手势识别结果的影响,从而可在一定程度上提高手势识别结果的准确性。

[0137] 进一步地,基于上述第三实施例,提出本发明手势识别方法的第四实施例。

[0138] 在本实施例中,步骤S20包括:

[0139] 步骤a21,判断所述待识别手势图像的数量为一个还是为至少两个;

[0140] 本实施例中,在实际应用场景中,待识别手势图像可能包括多个,而最终往往只需确定得到一个手势类别,以基于该手势来执行对应的操作。因此,在获取到待识别手势图像之后,需基于待识别手势图像的数量来确定最终的识别策略。

[0141] 具体的,先判断待识别手势图像的数量为一个还是为至少两个。

[0142] 若所述待识别手势图像的数量为一个,则执行步骤a22:将所述待识别手势图像输入至预先训练好的手势识别模型,得到手势识别结果;

[0143] 若待识别手势图像的数量为一个,则将待识别手势图像输入至手势识别模型,得到手势识别结果。即,直接将待识别手势图像输入至手势识别模型中,得到手势识别结果,该手势识别结果可以为该待识别手势图像中的手势属于各手势类别的分类概率,当然也可以为该待识别手势图像中的手势的所属的手势类别(即分类概率最大值所对应的手势类别)。

[0144] 若所述待识别手势图像的数量为至少两个,则执行步骤a23:基于预设规则和预先训练好的手势识别模型对所述待识别手势图像进行识别,得到手势识别结果。

[0145] 若待识别手势图像的数量为至少两个,则基于预设规则和预先训练好的手势识别模型对待识别手势图像进行识别,得到手势识别结果。

[0146] 具体的,步骤a23包括:

[0147] 步骤a231,将各待识别手势图像分别输入至预先训练好的手势识别模型,得到对应的手势类别及其分类概率;

[0148] 步骤a232,获取各待识别手势图像的第二属性数据,所述第二属性数据包括手部检测概率、像素信息和清晰度信息;

[0149] 步骤a233,基于所述手势类别的分类概率和所述第二属性数据对各待识别手势图像进行评分,得到各待识别手势图像的综合评分;

[0150] 步骤a234,根据所述综合评分和各待识别手势图像的手势类别,确定手势识别结果。

[0151] 本实施例中,先将各待识别手势图像分别输入至预先训练好的手势识别模型,得到对应的手势类别及其分类概率。然后,获取各待识别手势图像的第二属性数据,第二属性数据包括手部检测概率、像素信息和清晰度信息。其中,手部检测概率即为通过手势唤醒模型对手部区域图像进行识别时,确定为手部区域的检测概率;像素信息包括待识别手势图像的像素个数和目标图像的总像素个数,当待识别手势区域是基于用户区域图像裁剪得到的,则该目标图像为用户区域图像,当待识别手势区域是基于实时图像裁剪得到的,则该目

标图像为实时图像；清晰度信息即为待识别手势图像的清晰度，可通过清晰度算法获取得到，例如Brenner（勃伦纳）梯度函数、Tenengrad梯度函数、Laplacian梯度函数、SMD (STD Mean Difference, 灰度方差) 函数等。

[0152] 然后，基于手势类别的分类概率和第二属性数据对各待识别手势图像进行评分，得到各待识别手势图像的综合评分。其中，综合评分的计算公式如下：

$$[0153] \quad score = 0.1 * D_{prob} + 0.1 * C_{prob} + 0.5 * \frac{S_{hand}}{S_{img}} + 0.3 * \frac{sv}{\max_{sv} - \min_{sv}}$$

[0154] 其中，score表示综合评分， $D_{prob}$ 表示手部检测概率， $C_{prob}$ 表示手势类别的分类概率， $S_{hand}$ 表示待识别手势图像的像素个数， $S_{img}$ 表示目标图像的总像素个数，sv表示待识别手势图像的清晰度，利用 $\max_{sv}$ 和 $\min_{sv}$ 对待识别手势图像的清晰度进行了归一化，其中， $\max_{sv}$ 和 $\min_{sv}$ 为预设值，是基于终端分别设定的一预设最大清晰度和预设最小清晰度。当然，可以理解，在具体实施时，综合评分计算公式中各参数的系数（所占权重）可基于实际需要进行具体设定，并不限于上述计算公式。

[0155] 最后，根据综合评分和各待识别手势图像的手势类别，确定手势识别结果。具体的，可将综合评分中的最大值所对应的待识别手势图像的手势类别，作为最终识别得到的手势类别。

[0156] 本实施例中，针对手部区域包括一个和多个的情况，分别采用了对应的规则进行识别处理，最终确定得到手势识别结果，以便于在实际应用过程中，基于手势识别结果执行对应的操作。

[0157] 进一步地，基于上述各实施例，提出本发明手势识别方法的第五实施例。

[0158] 在本实施例中，在上述步骤S20之后，该手势识别方法还包括：

[0159] 步骤A，获取当前应用程序，并获取所述当前应用程序对应的手势类别与操作指令之间的映射关系；

[0160] 在本实施例中，在得到手势识别结果之后，可基于该手势识别结果对终端的应用程序进行控制。具体的，先获取当前应用程序，然后，获取该当前应用程序对应的手势类别与操作指令之间的映射关系。其中，该映射关系是预先设定的，不同的应用程序对应不同的映射关系。

[0161] 步骤B，根据所述手势识别结果和所述映射关系，确定目标操作指令；

[0162] 步骤C，基于所述目标操作指令控制所述当前应用程序执行对应的操作。

[0163] 然后，根据手势识别结果和映射关系，确定目标操作指令，进而基于目标操作指令控制当前应用程序执行对应的操作。例如，若当前应用程序为音乐播放程序，手势识别结果为握拳手势，其对应的目标操作指令为暂停，则对该当前应用程序执行暂停播放的操作。

[0164] 当然，需要说明的是，在具体实施例中，若终端中仅包括一种应用程序，例如智能音箱，若仅可实现音乐的播放，对应的只存在一种映射关系，此时，则无需执行步骤A，只需根据手势识别结果和映射关系，确定目标操作指令；然后，基于目标操作指令控制当前应用程序执行对应的操作。

[0165] 通过上述方式，本实施例基于用户手势对各应用程序进行对应的控制操作，实现了用户手势交互功能，可提升用户的使用体验。

[0166] 本发明还提供一种计算机可读存储介质，该计算机可读存储介质上存储有手势识

别程序,所述手势识别程序被处理器执行时实现如以上任一项实施例所述的手势识别方法的步骤。

[0167] 本发明计算机可读存储介质的具体实施例与上述手势识别方法各实施例基本相同,在此不作赘述。

[0168] 需要说明的是,在本文中,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、物品或者系统不仅包括那些要素,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、物品或者系统所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者系统中还存在另外的相同要素。

[0169] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0170] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明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如上所述的一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本发明各个实施例所述的方法。

[0171] 以上仅为本发明的优选实施例,并非因此限制本发明的专利范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

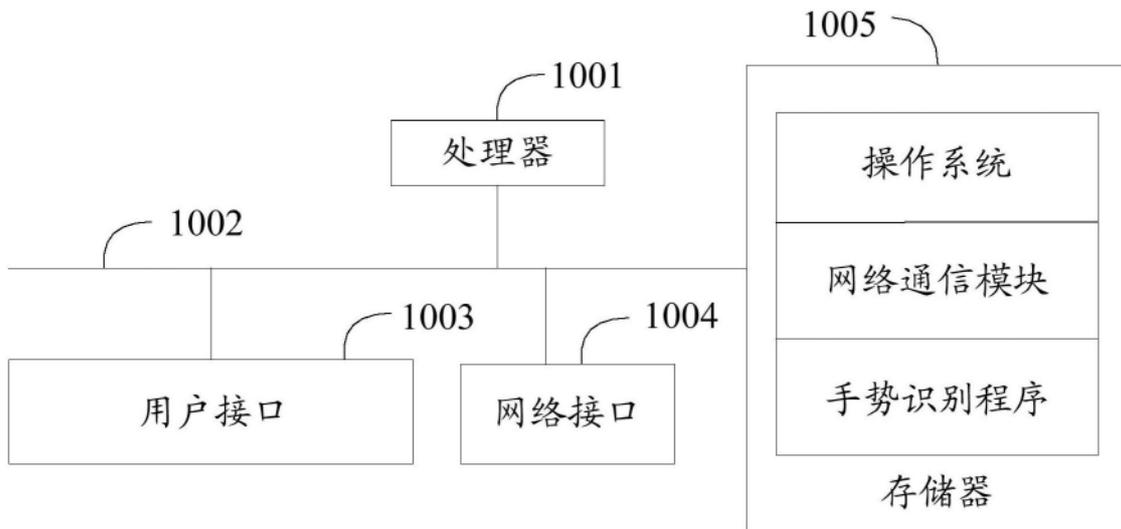


图1

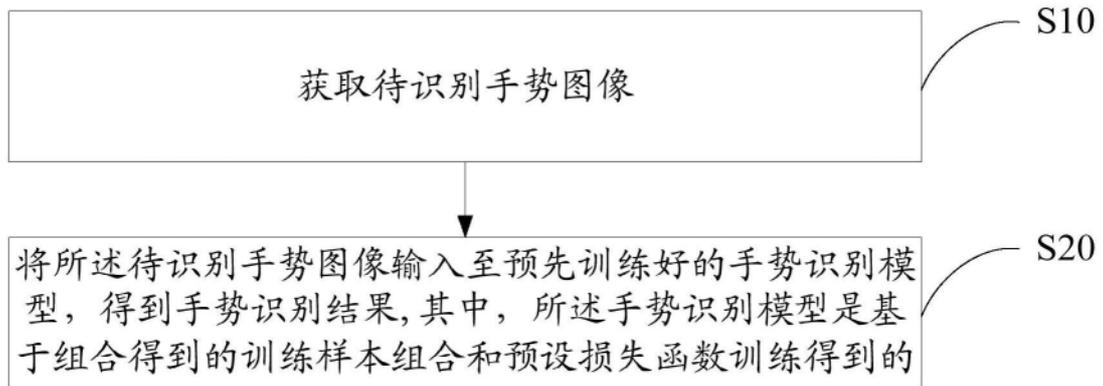


图2

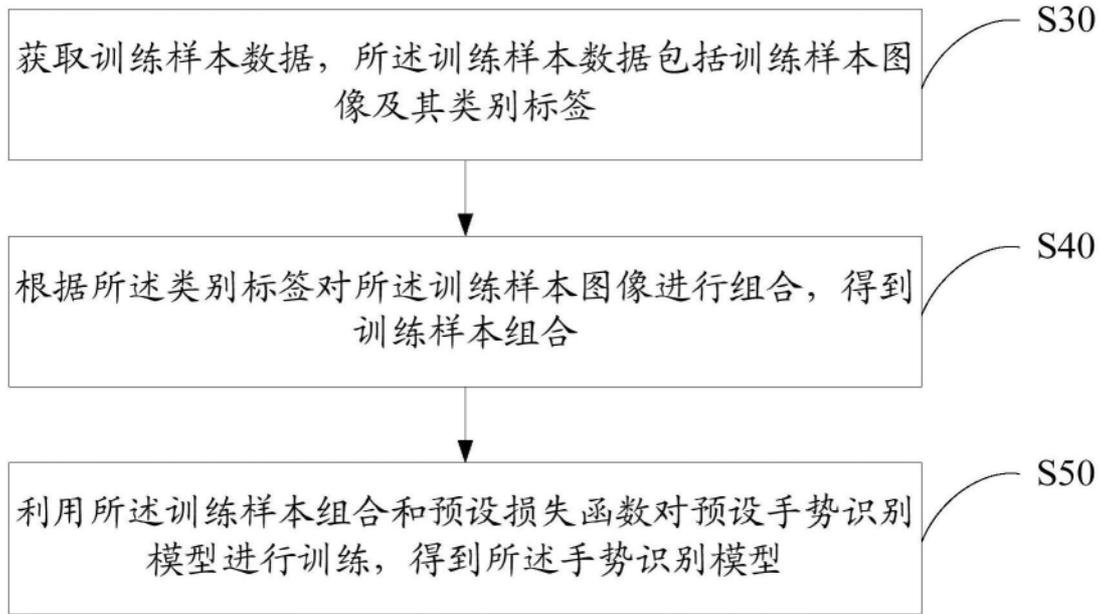


图3