



(12)发明专利

(10)授权公告号 CN 106203380 B

(45)授权公告日 2019.11.29

(21)申请号 201610576125.8

G06K 9/62(2006.01)

(22)申请日 2016.07.20

(56)对比文件

(65)同一申请的已公布的文献号
申请公布号 CN 106203380 A

CN 104755308 A,2015.07.01,
CN 104834907 A,2015.08.12,
CN 105446461 A,2016.03.30,
CN 102024151 A,2011.04.20,
CN 104571823 A,2015.04.29,
CN 102467905 A,2012.05.23,说明书第15-24段.

(43)申请公布日 2016.12.07

(73)专利权人 中国科学院计算技术研究所
地址 100190 北京市海淀区中关村科学院南路6号

CN 102662462 A,2012.09.12,
CN 104094287 A,2014.10.08,
CN 103092332 A,2013.05.08,

(72)发明人 陈益强 于汉超 钟习 杨晓东
胡子昂

审查员 赵会玲

(74)专利代理机构 北京泛华伟业知识产权代理有限公司 11280
代理人 王勇 苏晓丽

(51)Int.Cl.

G06K 9/00(2006.01)

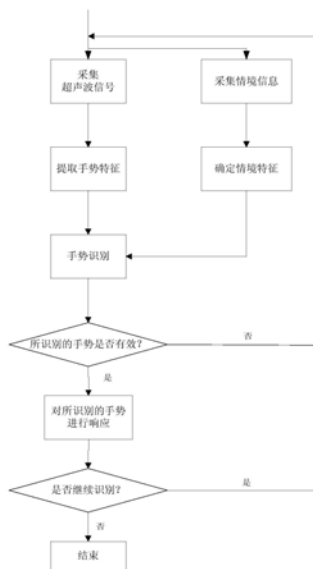
权利要求书2页 说明书9页 附图2页

(54)发明名称

超声波手势识别方法及系统

(57)摘要

本发明提供了一种融合情境感知信息的超声波手势识别方法,该方法同时采集超声波信号和与当前情境相关的情境信息,从采集的超声波信号获取手势特征,利用预先训练好的手势分类模型获取该手势特征属于预设的各种手势的概率;基于采集的情境信息确定各种手势在当前情境的下发生的概率;以及根据上述两种概率计算在当前情境中该手势特征属于预设的各种手势的概率并将其中最大概率对应的手势识别为与所采集的超声波信号对应的手势。该方法将手势信号与情境信息相融合,使用情境信息过滤用户的误操作手势、修正识别错误的手势,减少无效甚至错误的响应,从而提高了手势识别的准确率和鲁棒性,增强了人机交互体验。



1. 一种超声波手势识别方法,该方法包括
 - a) 发射超声波信号;
 - b) 采集超声波信号并同时采集与当前情境相关的情境信息,其中所述情境信息包括影响手势发生的一个或多个情境因素;
 - c) 基于发射和采集的超声波信号获取手势特征;
 - d) 利用预先训练好的手势分类模型获取该手势特征的手势概率向量,所述手势概率向量由所述手势特征属于预设的各种手势的概率构成;
 - e) 基于所采集的情境信息和预定的情境特征矩阵来确定与所述手势特征相关的情境概率向量,其中所述情境特征矩阵由所述预设的各种手势在各个情境因素下发生的概率构成,所述情境概率向量由预设的各种手势在当前情境下发生的概率构成;
 - f) 基于所述手势概率向量和所述情境概率向量确定在当前情境中所述手势特征属于预设的各种手势的概率并将其中最大概率对应的手势识别为与所采集的超声波信号对应的手势。
2. 根据权利要求1所述的方法,在步骤f) 还包括判断该最大概率是否在所识别的手势对应的置信区间中,如果在,则对该手势进行响应。
3. 根据权利要求1所述的方法,其中还包括训练手势分类模型的步骤,所述手势分类模型采用基于机器学习的方法来实现,所述手势分类模型的输入为手势特征,输出为手势概率向量。
4. 根据权利要求2所述的方法,其中所述置信区间是通过下列步骤得到的:
 - 确定用来进行训练的样本数据集,所述样本数据集中包括手势特征样本,与该手势特征样本相关的情境信息;
 - 对于样本数据集中的每个手势特征样本:
 - 根据预定的各种手势给该手势特征样本标注其所属的手势种类,利用预先训练好的手势分类模型获取该手势特征样本的手势概率向量,
 - 基于与该手势特征样本相关的情境信息和预定的情境特征矩阵来确定与该手势特征样本相关的情境概率向量,该情境概率向量由预设的各种手势在采集该手势特征样本时的情境下发生的概率构成;
 - 基于所述手势概率向量和所述情境概率向量得到所述手势特征在其相关情境中属于预设的各种手势的概率;
 - 以及基于各个手势特征样本所得到的概率及其所标注的手势种类来设置并调整各种手势对应的置信区间。
5. 一种超声波手势识别系统,该系统包括
 - 超声波收发装置,用于发射和采集超声波信号;
 - 情境信息采集装置,用于采集与当前情境相关的情境信息,其中所述情境信息包括影响手势发生的一个或多个情境因素;
 - 手势分类装置,用于基于发射的和采集的超声波信号获取手势特征,以及利用预先训练好的手势分类模型获取该手势特征的手势概率向量,所述手势概率向量由所述手势特征属于预设的各种手势的概率构成;
 - 情境映射装置,用于基于所采集的情境信息和预定的情境特征矩阵来确定与所述手势

特征相关的情境概率向量,其中所述情境特征矩阵由所述预设的各种手势在各个情境因素下发生的概率构成,所述情境概率向量由预设的各种手势在当前情境下发生的概率构成;

手势识别装置,用于基于所述手势概率向量和所述情境概率向量确定在当前情境中所述手势特征属于预设的各种手势的概率并将其中最大概率对应的手势识别为与所采集的超声波信号对应的手势。

6.根据权利要求5所述的系统,所述手势识别装置还用于判断该最大概率是否在所识别的手势对应的置信区间中,如果在,则对该手势进行响应。

7.根据权利要求5所述的系统,其中还包括训练装置,用于训练手势分类模型,所述手势分类模型采用基于机器学习的方法来实现,所述手势分类模型的输入为手势特征,输出为手势概率向量。

8.根据权利要求6所述的系统,其中所述置信区间是通过下列步骤得到的:

确定用来进行训练的样本数据集,所述样本数据集中包括手势特征样本,与该手势特征样本相关的情境信息;

对于样本数据集中的每个手势特征样本:

根据预定的各种手势给该手势特征样本标注其所属的手势种类,利用预先训练好的手势分类模型获取该手势特征样本的手势概率向量,

基于与该手势特征样本相关的情境信息和预定的情境特征矩阵来确定与该手势特征样本相关的情境概率向量,该情境概率向量由预设的各种手势在采集该手势特征样本时的情境下发生的概率构成;

基于所述手势概率向量和所述情境概率向量得到所述手势特征在其相关情境中属于预设的各种手势的概率;

以及基于各个手势特征样本所得到的概率及其所标注的手势种类来设置并调整各种手势对应的置信区间。

超声波手势识别方法及系统

技术领域

[0001] 本发明属于人机交互领域,尤其涉及非接触式手势识别方法。

背景技术

[0002] 随着诸如智能手机、手表、手环等可穿戴智能移动终端的普及,手势识别在人机交互领域扮演着越来越重要的角色。智能手表、手环等可穿戴设备由于设备屏幕小或者不配备显示屏,导致传统的按键或触摸屏式的手势识别方法难以直接应用,非接触式手势识别方法应运而生。传统的非接触式手势识别方法主要基于计算机视觉、数据手套、运动传感器等,存在对光照敏感、计算复杂度高、操作不方便、操作范围有限等问题。

[0003] 超声波手势识别方法是一种新型的非接触式手势识别方法,具有对光照不敏感、算法复杂度低、耗费资源少、感应范围广以及对设备硬件要求低等优点,更适合于可穿戴智能移动设备。近年来,超声波手势识别随着智能可穿戴设备的发展越来越受到国内外研究人员的关注。超声波手势识别方法通常是通过监测手部运动引发的超声波多普勒效应来识别手势。但日常生活中手部运动很灵活,而且误操作发生的可能性较大,现有的超声波手势识别方法的识别准确率并不理想。

发明内容

[0004] 因此,本发明的目的在于克服上述现有技术的缺陷,提供一种新的超声波手势识别方法,其兼顾手部运动发生时的情境因素,利用情境信息排除手部误操作的影响以提高手势识别的准确率和鲁棒性。

[0005] 本发明的目的是通过以下技术方案实现的:

[0006] 一方面,本发明提供了一种超声波手势识别方法,包括:

[0007] a) 发射超声波信号;

[0008] b) 采集超声波信号并同时采集与当前情境相关的情境信息,其中所述情境信息包括影响手势发生的一个或多个情境因素;

[0009] c) 基于发射的和采集的超声波信号获取手势特征;

[0010] d) 利用预先训练好的手势分类模型获取该手势特征的手势概率向量,所述手势概率向量由所述手势特征属于预设的各种手势的概率构成;

[0011] e) 基于所采集的情境信息和预定的情境特征矩阵来确定与所述手势特征相关的情境概率向量,其中所述情境特征矩阵由所述预设的各种手势在各个情境因素下发生的概率构成,所述情境概率向量由预设的各种手势在当前情境下发生的概率构成;

[0012] f) 基于所述手势概率向量和所述情境概率向量确定在当前情境中所述手势特征属于预设的各种手势的概率并将其中最大概率对应的手势识别为与所采集的超声波信号对应的手势。

[0013] 在一个实施例中,在步骤f) 还可包括判断该最大概率是否在所识别的手势对应的置信区间中,如果在,则对该手势进行响应。

[0014] 在一个实施例中,该方法还可包括训练手势分类模型的步骤,所述手势分类模型采用基于机器学习的方法来实现,所述手势分类模型的输入为手势特征,输出为手势概率向量。

[0015] 在一个实施例中,所述置信区间可以通过下列步骤得到的:

[0016] 确定用来进行训练的样本数据集,所述样本数据集中包括手势特征样本,与该手势特征样本相关的情境信息;

[0017] 对于样本数据集中的每个手势特征样本:

[0018] 根据预定的各种手势给该手势特征样本标注其所属的手势种类,利用预先训练好的手势分类模型获取该手势特征样本的手势概率向量,

[0019] 基于与该手势特征样本相关的情境信息和预定的情境特征矩阵来确定与该手势特征样本相关的情境概率向量,该情境概率向量由预设的各种手势在采集该手势特征样本时的情境下发生的概率构成;

[0020] 基于所述手势概率向量和所述情境概率向量得到所述手势特征在其相关情境中属于预设的各种手势的概率;

[0021] 以及基于各个手势特征样本所得到的概率及其所标注的手势种类来设置并调整各种手势对应的置信区间。

[0022] 另一方面,本发明提供了一种超声波手势识别系统,包括:

[0023] 超声波收发装置,用于采集和发射超声波信号;

[0024] 情境信息采集装置,用于采集与当前情境相关的情境信息,其中所述情境信息包括影响手势发生的一个或多个情境因素;

[0025] 手势分类装置,用于基于发射的和采集的超声波信号获取手势特征,以及利用预先训练好的手势分类模型获取该手势特征的手势概率向量,所述手势概率向量由所述手势特征属于预设的各种手势的概率构成;

[0026] 情境映射装置,用于基于所采集的情境信息和预定的情境特征矩阵来确定与所述手势特征相关的情境概率向量,其中所述情境特征矩阵由所述预设的各种手势在各个情境因素下发生的概率构成,所述情境概率向量由预设的各种手势在当前情境下发生的概率构成;

[0027] 手势识别装置,用于基于所述手势概率向量和所述情境概率向量确定在当前情境中所述手势特征属于预设的各种手势的概率并将其中最大概率对应的手势识别为与所采集的超声波信号对应的手势。

[0028] 与现有技术相比,本发明的优点在于:

[0029] 将手势信号与情境信息相融合,使用情境信息过滤用户的误操作手势、修正识别错误的手势,减少无效甚至错误的响应,从而提高了手势识别的准确率和鲁棒性,增强了人机交互体验。

附图说明

[0030] 以下参照附图对本发明实施例作进一步说明,其中:

[0031] 图1为根据本发明实施例的超声波手势识别方法的流程示意图;

[0032] 图2为根据本发明实施例的训练融合情境信息的手势识别模型的流程示意图;

[0033] 图3a为安静环境下根据本发明实施例的手势识别方法与传统无情境信息的手势识别方法的识别准确率对比示意图；

[0034] 图3b为标准环境下根据本发明实施例的手势识别方法与传统无情境信息的手势识别方法的识别准确率对比示意图；

[0035] 图3c为喧闹环境下根据本发明实施例的手势识别方法与传统无情境信息的手势识别方法的识别准确率对比示意图。

具体实施方式

[0036] 为了使本发明的目的,技术方案及优点更加清楚明白,以下结合附图通过具体实施例对本发明进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。

[0037] 人们的行为表现往往会受其所处的情境的影响,相同手势在不同情境下发生概率可能不同。而现有的超声波识别方法只关注于手部运动,采集到相关手势数据即进行识别,而并不考虑当前手势是否适合用户当前所处的环境,忽略了用户的真正意图,由此经常导致无效甚至错误的响应。在本发明中,提供了一种融合情境感知信息的超声波手势识别方法,同时采集手势数据和当前情境信息,融合两者作为手势识别的输入信息,利用情境信息过滤误操作手势、修正识别错误的手势等,以尽量减少无效或错误响应,识别出符合用户真正意图的手势。以播放音乐的应用为例,当前的环境声音和音乐播放音量都会影响用户调整音量的行为。例如,当环境声音较大、播放音量较低时,调高播放音量的可能性较大;反之,调低播放音量的可能性较大。而如果当前环境很安静并且当前播放音量较高,但检测到用户的手部运动为调高音量的手势,那么可以不对该手势进行响应,因为该手势可能是由于手部无意识的运动产生的误操作。对于超声波手势识别方法的不同应用场景,所设定的手势种类以及影响手势发生的相关情境因素有很大差异。本发明的方法对于具体的手势种类以及影响手势发生的情境因素不进行任何限制,本领域技术人员可以根据具体的应用场景来对此进行相应的选择和设置。

[0038] 图1给出了根据本发明一个实施例的超声波手势识别方法的流程示意。该方法实际上可以分为在线实时手势识别和离线训练手势识别模型两个阶段。如图1所示,在线实时手势识别主要包括(1)发射超声波以及采集超声波信号与当前情境相关的情境信息;(2)提取手势特征和情境特征;(3)基于所提取的手势特征和情境特征进行手势识别。

[0039] 更具体地,首先采集超声波信号并同时采集与当前情境相关的情境信息,这里,情境信息可以包括影响手势发生的一个或多个情境因素,其随具体应用场景而不同。在不同情境下,用户可能的手势集合不同,相同手势发生的概率也不相同。在特定情境下,如果一个手势符合情境信息的要求,那么这种手势发生的概率较大;反之,这种手势发生的概率较小。在特定情境下,影响手势发生的情境因素可能有一个,也可能有多个。为说明方便,下面假定有M种手势和L个情境因素,可以根据具体应用场景和具体的情境因素来预先确定各种手势在各情境因素下发生的概率。这里为描述方便,以矩阵的方式来表示这些概率,假定情境特征矩阵A为由各种手势在各情境因素下发生的概率构成的矩阵,可记为 $(a_{mk})_{M \times L}$,其中 $a_{mk} = p_{mk}(c_k)$ 表示第m种手势在第k个情境因素 c_k 下发生的概率($1 \leq m \leq M, 1 \leq k \leq L$)。具体概率计算方式可以依具体应用场景和具体的情境因素不同。

[0040] 这里为方便,仍以播放音乐的应用为例,假定有“前伸”和“后拉”两种手势,分别对应“调高播放音量”和“调低播放音量”两种用户调整音量的行为。在音乐播放活动中,当前的环境声音和音乐播放音量都会影响用户调整音量的行为。例如,当环境声音较大、播放音量较低时,调高播放音量的可能性较大;反之,调低播放音量的可能性较大。这里将用户所处的环境音量、当前播放音量设定为影响手势产生的两个情境因素。其中,对于环境音量,对应安静(声音低于20dB)、标准(声音处于20dB-40dB)、喧闹(声音高于40dB)共3种环境。在标准环境中调高和调低音量的可能性相差不大。在一个示例中,可以选择适当的概率函数来设置各种手势在各情境因素下发生的概率。例如假设“前伸”和“后拉”两种手势在典型标准环境音28dB情况下发生的概率均为0.5,那么如果检测到当前环境音量为15dB,则调低音量的可能性增加而调高音量的可能性减小,“前伸”手势在当前环境中发生的概率例如可以为 $0.5-0.5*15/28$,而“后拉”手势在当前环境中发生的概率例如可以为 $0.5+0.5*15/28$ 。在又一示例中,也可以通过分段函数的方式来指定各种手势在各情境因素下发生的概率。例如,将环境声音的取值范围分段,不同的范围设置不同的概率。例如,当环境声音处于20dB-25dB时,将“前伸”手势发生的概率设定为0.45,当环境声音处于25db-30db之间时,将“前伸”手势发生的概率设置为0.5,可以此类推。上述概率计算方法仅是举例说明而非进行任何限制。具体的概率设置可以实际需求和具体应用场景和情境因素等而具体确定。

[0041] 继续参考图1,在采集到超声波信号之后,从发射的超声波和接收的超声波信号的多普勒频移获取表征该采集的超声波信号的手势特征,典型的手势特征包括但不限于最大值、最小值、均值、标准方差、过零率、百分位数、关联系数、功率谱密度、频域熵、谱峰位置等。

[0042] 同时,基于采集的当前情境的情境信息和上述设定的情境特征矩阵来获取与该手势特征相关的情境特征,该情境特征可以采用由预设的各种手势在当前情境下发生的概率构成的向量来表示,也可以称为情境概率向量。假设当前采集的情境信息记为 $c = (c_1, c_2, \dots, c_k, \dots, c_L)$, c_k 表示影响当前采集的第k个情境因素,那么通过预先确定的情境特征矩阵A可以获得预定的M种手势在当前情境c下发生的概率,这些概率组成了一个M维的向量,即得到与当前采集的超声波的手势特征相关的情境概率向量 $C = [p_1(c), \dots, p_m(c), \dots, p_M(c)]$,其中 $p_m(c)$ 表示第m种手势在当前情境c下发生的概率,即 $p_m(c) = p(c_1, c_2, \dots, c_k, \dots, c_L)$,其概率值例如为 $p_1(c_1) \cdot p_2(c_2) \cdot \dots \cdot p_M(c_L)$ 。可以看出,与该手势特征相关的情境概率向量C由表示预定的各种手势在采集该手势数据时的当前情境c下发生的概率构成。

[0043] 然后,将所提取的手势特征作为预先训练好的手势分类模型的输入,通过该手势分类模型可得到该手势的手势概率向量,该手势概率向量由所述手势特征属于预设的各种手势的概率构成,其可记为 $G = [G_1, G_2, \dots, G_i, \dots, G_M]^T$,其中M表示预先设定的手势类别数, G_i 表示所采集的超声波信号被识别为第i种手势的概率($1 \leq i \leq M$)。而至于手势分类模型可以采用多种基于机器学习的分类模型来的训练,例如贝叶斯模型、决策树、极速学习机(ELM)等等,只要确定用于训练的手势特征和待识别的手势类别即可,具体的训练过程会在下文中进行简单介绍。

[0044] 这样,最终所识别的手势是基于该手势特征的手势概率向量及与其相关的情境概率向量来确定的。在如上文介绍的,得到该手势特征属于预设的各种手势的概率和各种手势在当前情境下发生的概率之后,便可以计算得出该手势特征在当前情境中属于预设的各

种手势的概率,其中最大概率对应的手势可被确定为最终识别的与所采集的超声波信号对应的手势。可以看出,该识别方法不仅考虑了手势数据本身还兼顾了采集该手势数据时的情境信息,如果一个手势符合情境,那么这个手势在手势概率向量和情境概率向量中对应的概率都很大,相乘之后得到的概率将较大,而如果一个手势不符合情境,那么该手势在手势概率向量和情境概率向量中对应的概率相乘会得到较小的值,这就体现了情境信息对识别错误的手势的修正作用。仍以播放音乐的应用为例,假定通过手势分类模型判断所采集的超声波信号为“前伸”手势(即调高音量)的概率最大,而检测的当前环境音非常低,即“前伸”手势在当前环境音的条件下发生的概率很小,那么综合这两个概率得到的所采集的超声波信号在当前环境音的条件下为“前伸”手势的概率会被修正为相对较小。

[0045] 优选地,为了进一步减少无效或错误的响应,该方法还包括判断上述最终识别出的手势是否有效的步骤,如果无效则不对该手势进行响应,如果有效,则对该手势进行响应。例如可以通过判定最终识别出的手势对应的概率是否在预先训练好的该手势对应的置信区间(其在下文离线训练阶段中会有详细介绍)中,如果在,则对该手势进行响应。这样可以进一步避免对误操作手势的错误响应。

[0046] 下面参考图2来介绍离线训练手势识别模型的阶段。参考图2,手势识别模型包括手势分类、情境分析和决策输出三个模块,离线训练阶段与上文介绍的在线识别过程基本上类似,主要包括下列步骤:

[0047] 1,获取样本数据

[0048] 可利用可穿戴智能移动设备(如智能手机、手表、手环等)内置的麦克风和扬声器来获取样本数据,其中扬声器用于发射超声波声音信号,麦克风用于采集声音信号。可使用各种设备内置的或附加的传感器采集特定情境下影响手势识别的相关情境因素。由不同的用户在各种情境下执行预定的各类手势,从而采集足够丰富的样本数据。

[0049] 2,手势分类

[0050] 从所采集的超声波数据中提取手势特征作为手势样本,并为各个手势样本标注其所属的手势类别号,利用这些标注后的手势样本作为训练数据集来训练手势分类模型,如上文所介绍的,这里可采用多种基于机器学习的分类模型来作为手势分类模型。对任意手势样本 x_j ,该手势分类模型输出的手势概率向量为 $G_j = [G_{j1}, G_{j2}, \dots, G_{jm}]^T \in R^m$, m 是手势类别数, R 表示实数。图2给出是利用极限学习机ELM实现的手势分类模型的示例。假设有 N 个任意手势样本 (x_i, t_i) ,其中 x_i 是一个 n 维的手势输入向量, $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$, t_i 是 m 维的目标向量, $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ 。一个有 L 个隐层节点的单隐层前馈神经网络表示为:

[0051]

$$f_L(x_j) = \sum_{i=1}^L \beta_i G(a_i, b_i, x_j), j = 1, \dots, N \quad (1)$$

[0052] 其中, a_i 和 b_i 是输入层到隐层节点之间的学习参数, β_i 是连接第 i 个隐层节点和输出节点之间的权重, $G(a_i, b_i, x_j)$ 为激活函数。该手势分类模型的学习目标是使输出误差最小,即存在 a_i, b_i, β_i 使得:

[0053]

$$f_L(x_j) = \sum_{i=1}^L \beta_i G(a_i, b_i, x_j) = t_j, j = 1, \dots, N \quad (2)$$

[0054] 通过上述过程求解出 β 。那么对于特定的手势样本 x_j ,和 x_j 相对应的实际手势输出

y_j 表示为:

[0055]

$$y_j = \sum_{i=1}^L \beta_i G(a_i, b_i, x_j), j = 1, \dots, N \quad (3)$$

[0056] 最后,通过公式(4)对 y_j 进行置信度向量求解:

[0057]

$$G_j = \text{norm} \left(\frac{1}{\text{abs}(y_j - 1)} \right) \quad (4)$$

[0058] 其中,norm函数表示归一化,abs函数表示取绝对值。

[0059] 3,情境分析

[0060] 在特定情境下,如果一个手势符合情境信息的要求,那么这种手势发生的概率较大;反之,这种手势发生的概率较小。在特定情境下,影响手势发生的情境因素可能有一个,也可能有多个。情境信息是指影响手势发生的相关情境因素。如上文介绍的,假定有M种用户手势和L个情境因素,那么可以确定情境特征矩阵A,记为 $(a_{mk})_{M \times L}$,其中 $a_{mk} = p_{mk}(c_k)$ 表示第m种手势在第k个情境因素 c_k 下发生的概率($1 \leq m \leq M, 1 \leq k \leq L$)。如上文讨论的,具体概率函数p依具体应用场景和具体的情境因素而定。

[0061] 假设采集手势样本 x_j 时同时采集的情境信息记为 $c_j = (c_{j1}, c_{j2}, \dots, c_{jk}, \dots, c_{jL})$, c_{jk} 表示影响手势样本 x_j 的第k个情境因素,那么通过预先确定的情境特征矩阵A可以获得预定的各种手势在当前情境 c_j 下发生的概率,这些概率组成了一个m维的向量,即得到与手势样本 x_j 相关的情境概率向量 $C_j = [p_1(c_j), \dots, p_m(c_j), \dots, p_M(c_j)]$,其中 $p_m(c_j)$ 表示第m种手势在当前情境 c_j 下发生的概率,即 $p_m(c_j) = p_m(c_{j1}, c_{j2}, \dots, c_{jk}, \dots, c_{jL})$,其概率值例如为 $p_1(c_1) \cdot p_2(c_2) \cdot \dots \cdot p_M(c_L)$ 。

[0062] 4,决策输出

[0063] 在得到手势样本 x_j 对应的手势概率向量 G_j 和情境概率向量 C_j 就可以得到手势样本 x_j 在当前情境 c_j 下作为预定的各种手势发生的概率向量 $P_j = G_j \cdot C_j$ 。这样得到的概率向量 P_j 中最大概率对应的手势就是针对该手势样本 x_j 所识别的手势,将该识别结果与该手势样本 x_j 所标注的手势类别相比较则可判断出本次识别结果是正确还是错误。其中对于误操作手势,通常被识别成任何类别的手势的概率都差不多,其手势分类模型输出中对应的最大概率接近均分概率,在和情境概率向量中相应概率相乘后的值通常居中,即位于一定区间中。因此对于训练样本集中各个手势样本及其对应情境信息样本通过上述识别过程得到的各识别结果进行统计分析,根据实验观察值得出误操作手势区间的上下限统计结果,记为 $\text{threshold}_{\text{low}}$ 和 $\text{threshold}_{\text{high}}$ 。这样,经训练后得到的手势识别模型如公式(5)表示,对于任一手势样本 x_j ,该手势识别模型的决策输出结果 O_j 如下:

$$O_j = \begin{cases} 0, & \max(G_j \cdot C_j) \leq \text{threshold}_{\text{high}} \\ k, & \text{else, 其中 } (G_j \cdot C_j)_k = \max(G_j \cdot C_j) \end{cases} \quad (5);$$

[0065] 其中,“ \cdot ”运算表示向量对应元素相乘, $O_j = 0$ 表示不响应当前的用户手势;当经上述过程得到的最大概率大于 $\text{threshold}_{\text{high}}$ 时,输出该最大概率值对应的第k种手势作为所识别的手势并对其进行响应(这里也可以理解为经训练该第k种手势的置信区间为 $(\text{threshold}_{\text{high}}, 1]$)。这是因为所识别的符合情境的手势,其对应概率相对较大,而不符合

情境的手势被修正后,其对应概率相对较小。因此,该方法在一定程度上修正手势识别错误的影响。

[0066] 或者,在优选的实施例中,经训练后得到的手势识别模型如公式(6)表示,对于任一手势样本 x_j ,该手势识别模型的决策输出结果 O_j 如下:

$$[0067] \quad O_j = \begin{cases} 0, & \text{threshold}_{low} \leq \max(G_j \cdot C_j) \leq \text{threshold}_{high} \\ k, & \text{else, 其中 } (G_j \cdot C_j)_k = \max(G_j \cdot C_j) \end{cases} \quad (6);$$

[0068] 其中,“ \cdot ”运算表示向量对应元素相乘, $O_j=0$ 表示不响应当前的用户手势;如果上述过程得到的最大概率大于 threshold_{high} 时,输出该最大概率值对应的第 k 种手势作为所识别的手势并对其进行响应。并且当经上述过程得到的最大概率小于 threshold_{low} 时,也会输出该最大概率值对应的第 k 种手势作为所识别的手势并对其进行响应,这主要是考虑了下面的因素:如果手势识别出的手势不符合情境,那么手势概率和情境概率点乘得到的概率向量的最大值通常小于 threshold_{low} 。例如,如果手势分类模型将本应属于第2种手势的超声波信号识别为属于第3种手势(即在手势概率向量中第3种手势对应的概率最大),而同时该第3种手势由于不符合情境,所以在情境概率向量中对应的概率会很小,那么其手势概率与情境概率的乘积可能不会是最大的。但是,如果一个手势不符合情境,但它的手势概率很大,乘以情境概率可能还是最大的,那么可以估计这是用户有意为之,是允许发生的和进行响应的。

[0069] 为了进一步说明本发明的效果,发明人进行了下列实验:具体应用设置为在安静、标准、喧闹等3种环境下的播放音乐,利用智能可穿戴设备内置的麦克风和扬声器作为超声波的收发装置,通过多普勒效应获得超声波手势识别信息;同时,对用户所处的环境音量、当前播放音量以及用户调整音量的行为进行感知。

[0070] 在该实验中定义了“前伸”和“后拉”两种手势,分别对应“调高播放音量”和“调低播放音量”两种用户调整音量的行为;将安静、标准、喧闹3种环境分别设置为表示声音低于20dB的环境、声音处于20dB-40dB的环境和声音高于40dB的环境。在实验中发现,在上述三种环境下最舒适的音乐播放音量分别为2dB、6dB和10dB。

[0071] 在音乐播放活动中,当前的环境声音和音乐播放音量都会影响用户调整音量的行为。例如,当环境声音较大、播放音量较低时,调高播放音量的可能性较大;反之,调低播放音量的可能性较大。不同环境下用户调整音量的手势与环境声音、音乐播放音量的具体对应关系如表1所示。

[0072] 表1

[0073]

手势类别	调整音量行为	环境声音相关性	播放音量相关性
前伸手势	调高播放音量	正相关	负相关
后拉手势	调低播放音量	负相关	正相关

[0074] 实验中邀请10位志愿者(7男3女)在上述3种声音环境下进行超声波手势识别实验。每次一个志愿者无干扰地进行实验,实验过程实时录像。实验中志愿者坐在椅子上,左手持智能设备,麦克风和扬声器朝向志愿者,右手进行手势操作。每位志愿者在每种声音环境下根据实际音乐播放音量需要进行相应的手势操作,共采集60个手势样本,其中“前伸”、“后拉”、“误操作”手势各占约1/3,“误操作”手势指手部无意识的运动产生的手势,而非用

户真正想表达的操作手势。志愿者和研究人员分别对采集到的手势数据进行标定,其中,志愿者按照自己的意图对所做的手势进行标定。

[0075] 实验中使用扬声器连续播放21kHz的超声波信号,采样频率为48kHz,每4096个采样点为一帧,平均手势长度为10帧。每一帧的时域信号经海明窗处理后进行4096个采样点的快速傅里叶变换(FFT)。根据多普勒公式,在上述采样条件下,超声波频率偏移范围不超过 $[-489, 500]$,频率分辨率为11.7Hz,故特征向量的维度为86。为减少误差,向发射频率左右两侧分别扩展,最终特征向量的维度为101。在发射频率两侧各选取50个频率点,记为 g_t , $g_t \in R^{101}$ 。对手势样本 x_j 得到10帧长的特征矩阵 $V_j = [g_1^T, \dots, g_{10}^T] \in R^{1010}$ 。

[0076] 作为情境特征的环境声音状态使用手势特征提取过程中FFT之后低于20kHz的1706个频率点进行评估,记作 w_t , $w_t \in R^{1706}$,对手势样本 x_j 得到环境声音矩阵 $W_j = [w_1^T, \dots, w_{10}^T]$,当前环境声音 $d_j = \max(W_j)$ 。本实验标准环境下环境声音 $e = 28\text{dB}$,将 e 设为基准环境声音。在标准状态下,每种手势随机发生,不受环境声音影响。当环境变化时,符合环境声音的手势概率增加,不符合环境声音的手势概率减少,对任意手势样本 x_j 表示为:

$$[0077] \quad p(c_{j1}) = \left[\frac{e}{e+d_j}, \frac{d_j}{e+d_j} \right]^T, \text{ 其中,标准环境下 } d_j = e.$$

[0078] 在任意手势样本 x_j 采集过程中同时记录当前的音乐播放音量 $v_j\text{dB}$ 。实验中首先利用手势信号的低频部分识别当前环境声音对应的环境,从而选择最舒适的音乐播放音量 $s\text{dB}$ 。当音乐播放音量低于 $s\text{dB}$ 时,调高播放音量的概率增加;而当音乐播放音量高于 $s\text{dB}$ 时,调低播放音量的概率增加,对任意手势样本 x_j 表示为:

$$[0079] \quad p(c_{j2}) = \left[\frac{s}{s+v_j}, \frac{v_j}{s+v_j} \right]^T.$$

[0080] 这样,与手势样本 x_j 对应的情境概率向量 $C_j = p(c_j) = p(c_{j1}, c_{j2}) = p(c_{j1}) \cdot p(c_{j2})$ 。

[0081] 将上述得到的手势样本 x_j 的特征矩阵 V_j 和情境概率向量提供给上文介绍的训练好的手势识别模型(即公式(6)),即可得到相应的手势识别结果。这里假设经上述离线训练阶段得到的手势识别模型中的 $\text{threshold}_{\text{low}}$ 和 $\text{threshold}_{\text{high}}$ 分别为0.2和0.31。

[0082] 本发明的融合情境感知信息的手势识别方法(ContextAware Information Fused Gesture Recognition Method)CAI-GRM的手势分类模型由ELM算法实现,本实验中ELM算法的激活函数为Sigmoid函数,隐层节点的个数为500。CAI-GRM在3种环境下手势识别的测试准确率结果如表2所示。

[0083] 表2

[0084]

类别	前伸手势	后拉手势	误操作手势	全部手势
安静环境	100.00%	100.00%	90.18%	96.73%
标准环境	100.00%	100.00%	96.98%	98.99%
喧闹环境	100.00%	100.00%	91.95%	97.32%

[0085] 由表2可以看出CAI-GRM手势识别方法在3种环境下整体识别准确率都在96%以上,其中“前伸”、“后拉”两种有效手势的识别完全正确,对于误操作手势,识别的准确率也在90%以上,这表明该方法能过滤大部分的误操作手势。

[0086] 为比较CAI-GRM与传统无情境信息的手势识别方法的识别准确率,传统无情境信

息的手势识别方法(附图中标记为ELM)使用ELM算法在相同的数据集上进行无情境信息的手势识别,结果如图3a(安静环境下测试结果)、3b(标准环境下测试结果)、3c(喧闹环境下测试结果)所示。由图3a-3c可以看出,CAI-GRM的识别准确率普遍高于无情境信息的ELM手势识别方法。对于“前伸”和“后拉”两种手势,CAI-GRM的识别准确率都高于无情境信息的手势识别方法,这表明CAI-GRM手势识别方法能使用情境信息对识别错误的手势进行修正,从而提高手势识别的准确率。对于“误操作”手势,CAI-GRM过滤掉绝大部分的误操作手势,对于判断为误操作的手势CAI-GRM不予响应,从而减少无效的操作,提高手势识别的准确率和鲁棒性;而无情境信息的手势识别方法对任意给定的手势输入都会产生相应的输出,对误操作也会给出相应的输出手势,完全不能过滤误操作手势,易导致无效甚至错误的响应。通过比较可以看出,融合情境感知信息的手势识别方法能够根据情境信息判断当前识别出的手势的合理性,能够较好地理解用户的操作意图、提高手势识别的准确率和鲁棒性。

[0087] 虽然本发明已经通过优选实施例进行了描述,然而本发明并非局限于这里所描述的实施例,在不脱离本发明范围的情况下还包括所做出的各种改变以及变化。

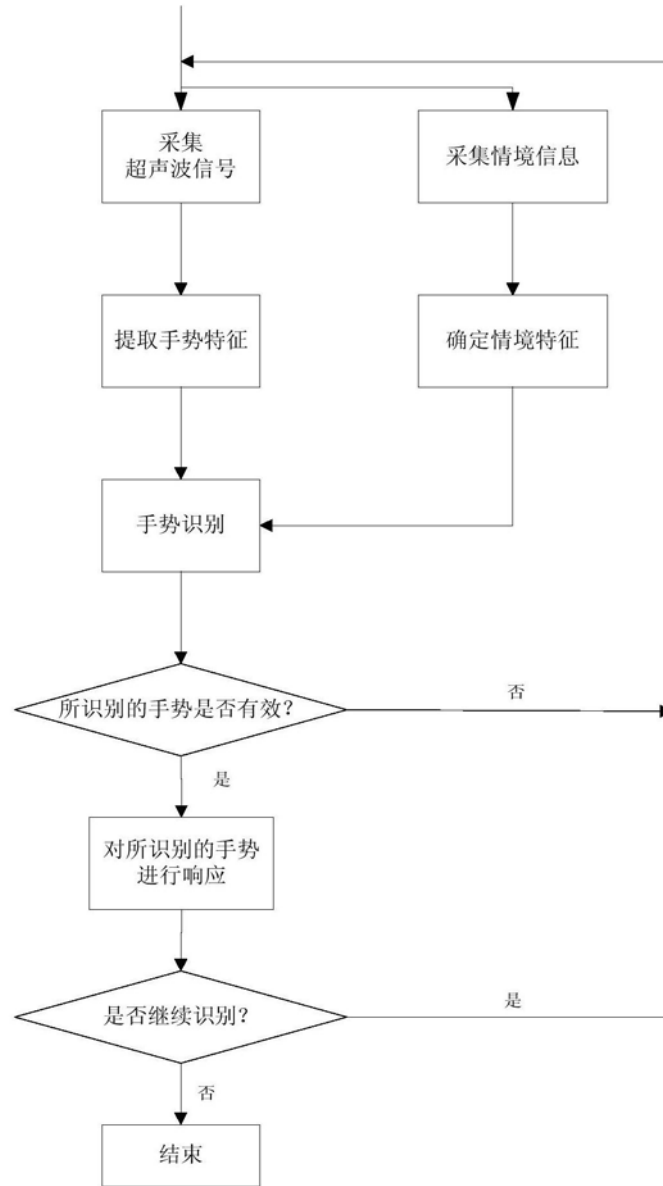


图1

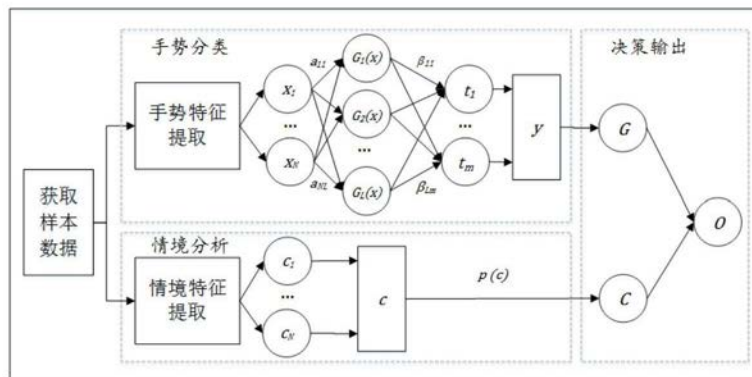


图2

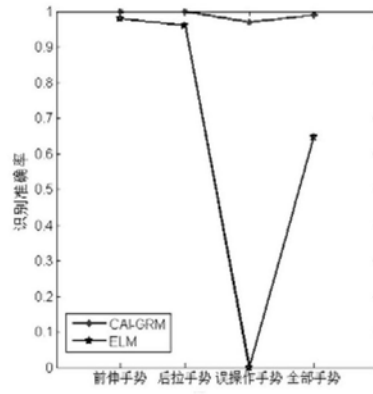


图3a

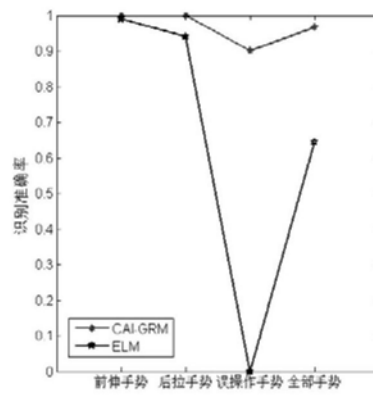


图3b

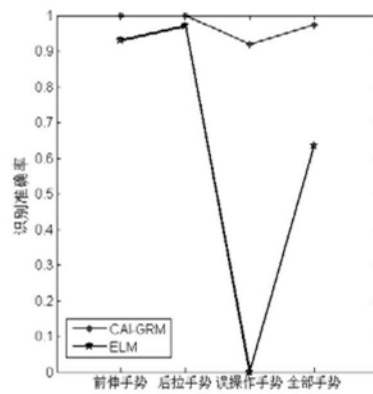


图3c