



(12)发明专利申请

(10)申请公布号 CN 108010526 A

(43)申请公布日 2018.05.08

(21)申请号 201711312402.5

(22)申请日 2017.12.08

(71)申请人 北京奇虎科技有限公司

地址 100088 北京市西城区新街口外大街
28号D座112室(德胜园区)

(72)发明人 毕宇鹏

(74)专利代理机构 北京市立方律师事务所

11330

代理人 刘延喜

(51) Int. Cl.

G10L 15/22(2006.01)

G10L 15/18(2013.01)

H04N 5/232(2006.01)

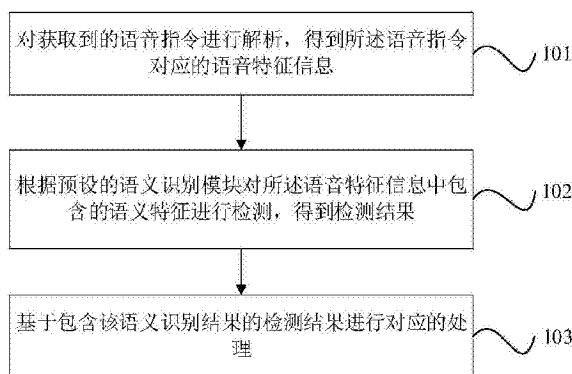
权利要求书2页 说明书10页 附图1页

(54)发明名称

语音处理方法及装置

(57)摘要

本发明涉及计算机技术领域,提供了一种语音处理方法及装置,该语音处理方法,包括:对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息;根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,所述检测结果中包含有语义匹配度最高的语义识别结果;基于包含该语义识别结果的检测结果进行对应的处理。实现了基于语音的处理,并且通过语音指令的控制,实现了无需人为操作即可实现对应操作的处理过程,降低了人为劳动力,同时实现了对于复杂语音指令有效处理,增加了处理范围,并通过这种免去人为操作过程的处理,进一步提升了用户的使用感受。



1. 一种语音处理方法,其特征在于,包括:
 - 对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息;
 - 根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,所述检测结果中包含有语义匹配度最高的语义识别结果;
 - 基于包含该语义识别结果的检测结果进行对应的处理。
2. 如权利要求1所述的方法,其特征在于,所述语音特征信息包括语义特征,所述根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,包括:
 - 根据预设的语义识别模块对所述语义特征进行识别,得到多个语义识别结果;
 - 并在得到的多个语义识别结果中确认语义匹配度最高的语义识别结果。
3. 如权利要求1或2所述的方法,其特征在于,所述基于包含该语义识别结果的检测结果进行对应的处理,包括:
 - 基于所述包含该语义识别结果的检测结果按照所述语音指令进行对应的处理;或,
 - 基于所述包含该语义识别结果的检测结果不做任何处理。
4. 如权利要求3所述的方法,其特征在于,所述基于所述包含该语义识别结果的检测结果按照所述语音指令进行对应的处理,包括:
 - 确定所述语音指令对应的指示信息;
 - 根据所述指示信息做相应的处理。
5. 如权利要求4所述的方法,其特征在于,所述指示信息包括如下任一项:
 - 基于网络直播平台和/或多媒体采集设备中的特定指令;
 - 基于多媒体设备中的播放和/或暂停指令。
6. 如权利要求1-5中任一项所述的方法,其特征在于,还包括:
 - 获取当前用户触发的动作和/或人脸;
 - 对当前用户触发的动作和/或人脸进行识别检测,得到识别结果;
 - 其中,所述基于包含该语义识别结果的检测结果进行对应的处理,包括:
 - 基于包含该语义识别结果的检测结果,并结合基于动作和/或人脸识别结果,进行对应的处理。
7. 如权利要求1-6中任一项所述的方法,其特征在于,还包括:
 - 根据预设的语音唤醒模块对所述语音特征信息进行检测,得到检测结果。
8. 一种语音处理装置,其特征在于,包括:
 - 解析单元,用于对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息;
 - 第一处理单元,用于根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,所述检测结果中包含有语义匹配度最高的语义识别结果;基于包含该语义识别结果的检测结果进行对应的处理。
9. 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计算机程序,该程序被处理器执行时实现权利要求1-7中任一项所述的方法。
10. 一种计算设备,包括:处理器、存储器、通信接口和通信总线,所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信;

所述存储器用于存放至少一可执行指令,所述可执行指令使所述处理器执行如权利要求1-7中任一项所述的语音处理方法对应的操作。

语音处理方法及装置

技术领域

[0001] 本发明涉及计算机技术领域,特别是涉及一种语音处理方法及装置。

背景技术

[0002] 随着消费类电子产品快速的发展,电子产品的功能性也越发强大。语音作为人类最基本的方式,将语音识别技术应用到消费类电子产品中,实现通过自然语音来控制此类产品的功能是未来发展的趋势。

[0003] 随着科技发展,尤其手机终端与多媒体终端设备的科技、智能化发展,人们在使用这些设备时,也不再仅仅是局限于其最初的基本功能,而是在追求越发智能化、人性化、便捷化、个性化的功能需求。

[0004] 如何能通过语音识别技术实现满足上述功能需求的技术方案,成为了当前亟待解决的技术问题。

发明内容

[0005] 本发明提供语音处理方法及装置,以实现基于语音指令的对应处理,同时通过多场景的应用,增加了处理范围,并且有效提升用户的使用感受。

[0006] 本发明提供了一种语音处理方法,包括:

[0007] 对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息;

[0008] 根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,所述检测结果中包含有语义匹配度最高的语义识别结果;

[0009] 基于包含该语义识别结果的检测结果进行对应的处理。

[0010] 优选地,所述语音特征信息包括语义特征,所述根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,包括:

[0011] 根据预设的语义识别模块对所述语义特征进行识别,得到多个语义识别结果;

[0012] 并在得到的多个语义识别结果中确认语义匹配度最高的语义识别结果。

[0013] 优选地,所述基于包含该语义识别结果的检测结果进行对应的处理,包括:

[0014] 基于所述包含该语义识别结果的检测结果按照所述语音指令进行对应的处理;
或,

[0015] 基于所述包含该语义识别结果的检测结果不做任何处理。

[0016] 优选地,所述基于所述包含该语义识别结果的检测结果按照所述语音指令进行对应的处理,包括:

[0017] 确定所述语音指令对应的指示信息;

[0018] 根据所述指示信息做相应的处理。

[0019] 优选地,所述指示信息包括如下任一项:

[0020] 基于网络直播平台和/或多媒体采集设备中的特定指令;

[0021] 基于多媒体设备中的播放和/或暂停指令。

- [0022] 优选地,所述特定指令包括如下任一项:
- [0023] 拍照;
- [0024] 摄像;
- [0025] 拍照中添加特效信息;
- [0026] 摄像中添加特效信息。
- [0027] 优选地,还包括:
- [0028] 获取当前用户触发的动作和/或人脸;
- [0029] 对当前用户触发的动作和/或人脸进行识别检测,得到识别结果;
- [0030] 其中,所述基于包含该语义识别结果的检测结果进行对应的处理,包括:
- [0031] 基于包含该语义识别结果的检测结果,并结合基于动作和/或人脸识别结果,进行对应的处理。
- [0032] 优选地,还包括:
- [0033] 根据预设的语音唤醒模块对所述语音特征信息进行检测,得到检测结果。
- [0034] 优选地,所述根据预设的语音唤醒模块对所述语音特征信息进行检测,包括:
- [0035] 根据所述语音唤醒模块对语音特征信息进行匹配,确定所述语音唤醒模块中是否存储有与语音特征信息匹配的目标语音特征信息;
- [0036] 并在匹配成功时,得到该匹配的目标语音特征信息。
- [0037] 优选地,在根据预设的语音唤醒模块对所述语音特征信息进行检测时,所述对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息,包括:
- [0038] 对所述语音指令进行声学特征提取,得到该语音指令对应的梅尔频率倒谱系数MFCC特征信息。
- [0039] 本发明还提供了一种语音处理装置,包括:
- [0040] 解析单元,用于对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息;
- [0041] 第一处理单元,用于根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,所述检测结果中包含有语义匹配度最高的语义识别结果;基于包含该语义识别结果的检测结果进行对应的处理。
- [0042] 优选地,所述语音特征信息包括语义特征,
- [0043] 所述第一处理单元,还用于根据预设的语义识别模块对所述语义特征进行识别,得到多个语义识别结果;并在得到的多个语义识别结果中确认语义匹配度最高的语义识别结果。
- [0044] 优选地,
- [0045] 所述第一处理单元,用于基于所述包含该语义识别结果的检测结果按照所述语音指令进行对应的处理;或,基于所述包含该语义识别结果的检测结果不做任何处理。
- [0046] 优选地,所述第一处理单元,具体用于确定所述语音指令对应的指示信息;根据所述指示信息做相应的处理。
- [0047] 优选地,所述指示信息包括如下任一项:
- [0048] 基于网络直播平台和/或多媒体采集设备中的特定指令;
- [0049] 基于多媒体设备中的播放和/或暂停指令。

- [0050] 优选地,所述特定指令包括如下任一项:
- [0051] 拍照;
- [0052] 摄像;
- [0053] 拍照中添加特效信息;
- [0054] 摄像中添加特效信息。
- [0055] 优选地,还包括:
- [0056] 获取单元,用于获取当前用户触发的动作和/或人脸;
- [0057] 第二处理单元,用于对当前用户触发的动作和/或人脸进行识别检测,得到识别结果;
- [0058] 所述第一处理单元,还用于基于包含该语义识别结果的检测结果,并结合基于动作和/或人脸识别结果,进行对应的处理。
- [0059] 优选地,
- [0060] 所述第一处理单元,还用于根据预设的语音唤醒模块对所述语音特征信息进行检测,得到检测结果。
- [0061] 优选地,
- [0062] 所述第一处理单元,用于根据所述语音唤醒模块对所述语音特征信息进行匹配,确定所述语音唤醒模块中是否存储有与语音特征信息匹配的目标语音特征信息;并在匹配成功时,得到该匹配的目标语音特征信息。
- [0063] 优选地,所述解析单元,具体用于对所述语音指令进行声学特征提取,得到该语音指令对应的梅尔频率倒谱系数MFCC特征信息。
- [0064] 本发明还提供了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计算机程序,该程序被处理器执行时实现上述的方法。
- [0065] 本发明还提供了一种计算设备,包括:处理器、存储器、通信接口和通信总线,所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信;
- [0066] 所述存储器用于存放至少一可执行指令,所述可执行指令使所述处理器执行上述的语音处理方法对应的操作。
- [0067] 与现有技术相比,本发明至少具有以下优点:
- [0068] 通过对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息的处理,实现了对所需求语音指令的特征提取,为后续对于该提取的特征的检测处理提供了保障;并且通过预设的语义识别模块对该提取出的语音特征信息中包含的语义特征的检测,再根据包含有语义匹配度最高的语义识别结果的检测结果来进行对应的处理,实现了基于语音指令的对应处理,实现了无需人为操作即可实现拍照的过程,降低了人为劳动力,同时实现了在复杂应用场景中对语音指令的有效处理,增加了处理范围;同时通过语音唤醒模块与语义识别模块的结合处理,提升了语音识别的准确度;也通过这种免去人为操作过程的处理,进一步提升了用户的使用感受。

附图说明

- [0069] 图1是本发明提供的语音处理方法的流程示意图;
- [0070] 图2是本发明提供的语音处理装置的结构图。

具体实施方式

[0071] 本发明提出一种语音处理方法及装置,下面结合附图,对本发明具体实施方式进行详细说明。

[0072] 下面详细描述本发明的实施例,所述实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,仅用于解释本发明,而不能解释为对本发明的限制。

[0073] 本技术领域技术人员可以理解,除非特意声明,这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是,本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件,但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。应该理解,当我们称元件被“连接”或“耦接”到另一元件时,它可以直接连接或耦接到其他元件,或者也可以存在中间元件。此外,这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的措辞“和/或”包括一个或多个相关联的列出项的全部或任一单元和全部组合。

[0074] 本技术领域技术人员可以理解,除非另外定义,这里使用的所有术语(包括技术术语和科学术语),具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是,诸如通用字典中定义的那些术语,应该被理解为具有与现有技术的上下文中的意义一致的意义,并且除非像这里一样被特定定义,否则不会用理想化或过于正式的含义来解释。

[0075] 本发明提供了一种语音处理方法,如图1所示,包括:

[0076] 步骤101,对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息。

[0077] 其中,本发明还包括:

[0078] 获取当前用户触发的动作和/或人脸;

[0079] 对当前用户触发的动作和/或人脸进行识别检测,得到识别结果。

[0080] 针对上述的动作验证过程,可以是手势验证过程,通过对当前用户触发的手势动作进行识别检测,得到对应的检测结果,进而实现根据检测结果进行对应的指令处理,如双手环抱构成心形,则在检测成功后在当前界面显示心形图案。

[0081] 当然,该手势验证过程还可以作为解锁验证,具体的,

[0082] 具体的,在显示界面向用户显示手势验证请求,请求当前用户输入预定的手势动作,并在当前显示界面的指定区域内随机生成多个不重合的采集点,然后采集用户触发各采集点生成的连线图,构成手势验证码,并对该构成的手势验证码与预先存储的解锁手势动作进行对比分析验证,得到验证结果;若验证结果为手势验证码与预先存储的解锁手势动作匹配,则确定验证成功,解锁当前界面,以待后续随时采集该用户的语音指令;若验证结果为手势验证码与预先存储的解锁手势动作不匹配,则确定验证失败,无法解锁当前界面,并在该界面显示“验证失败”的指示信息。

[0083] 其中,上述所提及的手势验证过程仅是为了说明本发明的动作验证过程所列举的一个实施例,对于其他能够达到本发明的动作验证过程相同技术效果的动作验证过程均在本发明的保护范围之内。

[0084] 对于基于人脸的验证过程,具体的,通过对当前用户触发的人脸进行识别检测,得到对应的检测结果,进而实现根据检测结果进行对应的指令处理,如当前用户露出笑脸,则在检测成功后在当前界面中笑脸的对应位置处分别显示小酒窝。

[0085] 当然,该人脸验证过程也同样可以作为解锁验证,具体的,在显示界面向用户显示人脸验证请求,请求当前用户触发人脸验证,并在当前显示界面提供指定的输入区域,然后采集用户提供的人脸信息,构成人脸验证码,并对该构成的人脸验证码与预先存储的解锁人脸信息进行对比分析验证,得到验证结果;若验证结果为人脸验证码与预先存储的解锁人脸信息匹配,则确定验证成功,解锁当前界面,以待后续随时采集该用户的语音指令;若验证结果为人脸验证码与预先存储的解锁人脸信息不匹配,则确定验证失败,无法解锁当前界面,并在该界面显示“验证失败”的指示信息。

[0086] 通过该验证作为指令输入的一种,增加了指令的处理方式,将指令并不仅限于语音指令,故而提升了用户的使用感受;通过该验证作为解锁过程的添加,保障了设备的安全,提升了其安全性。

[0087] 当然,在实际处理过程中,上述的验证过程可以根据当前用户的使用需要由用户自行设定,并不限定一定要在获取语音指令之前进行上述的验证过程。

[0088] 步骤102,根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果。

[0089] 其中,该检测结果中包含有语义匹配度最高的语义识别结果。

[0090] 优选地,所述语音特征信息包括语义特征,所述根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,包括:

[0091] 根据预设的语义识别模块对所述语义特征进行识别,得到多个语义识别结果;

[0092] 并在得到的多个语义识别结果中确认语义匹配度最高的语义识别结果。

[0093] 具体的,预先根据大量具有语义的语料训练该语义识别模块,从而根据该预先训练的语义识别模块对该语义特征进行分析,以根据分析结果找到与该语义特征的语义匹配度最高目标语义特征。

[0094] 对于其训练过程,可以包括:选取大量的样本,进行特征抽取,得到各样本数据的语义特征;对语义特征进行神经网络的深度学习过程处理,从而构建语义识别模块。其中,该神经网络可以为CNN(Convolutional Neural Network,卷积神经网络)、DNN(Deep Neural Network,深层神经网络)或RNN(Recurrent neural Network、循环神经网络)。

[0095] 对于上述语义识别模块的构建,根据处理的需要,可以构建所需的模块,即选取的样本数据决定了构建的模块所能检测的数据。

[0096] 具体的,通过语义识别模块对该语义特征进行语义识别,得到多个不同的语义识别结果,并通过该语义识别模块来对得到的多个语义识别结果进行确认,从该多个语义识别结果中筛选出语义匹配度最高的语义识别结果。通过该语义识别模块的处理,实现了对该语音指令的验证,提高了对语音指令处理的精准度。

[0097] 其中,上述动作和/或人脸的验证过程,可以发生在语义识别模块对语义特征的识别检测之前或之后,也可以和语义识别模块对语义特征的识别检测同时进行处理。由于对动作和/或人脸的验证速度高于语义识别模块对语义特征的识别检测速度,故优选地,可先对动作和/或人脸进行验证,再通过语义识别模块对语义特征进行验证。例如,在相机开启

之后,先接收到用户触发的动作“手摆出‘V’的姿势”,对其进行识别,得到识别结果;之后再接收到用户发送的语音指令“我想照相”,经语义识别模块的语义特征识别检测,确认需要进行拍照处理,从而实现了快速“拍照”操作。当然,上述实施例仅是为了说明本发明方案所列举的一个优选实施例,对于其他任意能够实现上述本发明的方案都在本发明的保护范围之内。

[0098] 步骤103,基于包含该语义识别结果的检测结果进行对应的处理。

[0099] 其中,所述基于包含该语义识别结果的检测结果进行对应的处理,包括:

[0100] 基于包含该语义识别结果的检测结果,并结合基于动作和/或人脸识别结果,进行对应的处理。

[0101] 进一步地,该基于包含该语义识别结果的检测结果进行对应的处理,包括两种方式,即处理与不处理:

[0102] (1) 基于所述包含该语义识别结果的检测结果按照所述语音指令进行对应的处理。

[0103] 具体的,该基于所述包含该语义识别结果的检测结果按照所述语音指令进行对应的处理,包括:

[0104] 确定所述语音指令对应的指示信息;

[0105] 根据所述指示信息做相应的处理。

[0106] 更进一步地,所述指示信息包括如下任一项:

[0107] 基于网络直播平台 and/或 多媒体采集设备中的特定指令;

[0108] 基于多媒体设备中的播放和/或暂停指令。

[0109] 其中,所述特定指令包括如下任一项:

[0110] 拍照;

[0111] 摄像;

[0112] 拍照中添加特效信息;

[0113] 摄像中添加特效信息。

[0114] 其中,上述特效信息可以是拍照过程中在人的面部添加动物的胡须,头部添加动物耳朵,还可以是在人物的背景中添加下雪、下玫瑰雨特效,当然,上述特效信息同样适用于摄像过程中。对于上述特效信息,并不仅局限于上述所列举的几个例子,对于其他任意能够实现上述所举例子中各特效效果具有相同效果的,均在本发明的保护范围之内。

[0115] (2) 基于所述包含该语义识别结果的检测结果不做任何处理。

[0116] 顾名思义,也即检测结果为未匹配上,所以不再做任何的处理,直接结束流程。当然,还可以不结束该流程,而选择发送提示消息,告知当前用户无法识别该语音指令或者未匹配到该语音指令,以使当前用户可以尝试调整该语音指令或者重新发送该语音指令。

[0117] 进一步地,在本方案中,在对该获取到的语音指令进行处理时,还包括:

[0118] 根据预设的语音唤醒模块对所述语音特征信息进行检测,得到检测结果。

[0119] 优选地,所述根据预设的语音唤醒模块对所述语音特征信息进行检测,包括:

[0120] 根据所述语音唤醒模块对语音特征信息进行匹配,确定所述语音唤醒模块中是否存储有与语音特征信息匹配的目标语音特征信息;

[0121] 并在匹配成功时,得到该匹配的目标语音特征信息。

[0122] 其中,在根据预设的语音唤醒模块对所述语音特征信息进行检测时,该对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息,包括:

[0123] 对所述语音指令进行声学特征提取,得到该语音指令对应的梅尔频率倒谱系数MFCC(Mel Frequency CepstrumCoefficient)特征信息。

[0124] 具体的,对所述语音指令经过预滤波、预加重、分帧、加窗后可以得到每帧语音的时域信号,对每一帧时频信号做离散傅里叶变换(DFT)得到频域信号,完成时域转换成频域,求频域信号的平方,即能量谱;通过使用M个Mel带通滤波器对其进行滤波,计算第m个滤波器输出的能量对数叠加,再经离散余弦变换(DCT)即可得到Mel倒谱系数MFCC。

[0125] 该语音唤醒模块可以是基于表征语音指令的各预设词汇的MFCC特征信息的语音特征信息的数据训练生成。

[0126] 对于其训练过程,可以包括:选取特定的唤醒词样本(如拍照、录像、下玫瑰雨等),进行特征抽取得到MFCC特征信息;对MFCC特征信息进行神经网络的深度学习过程处理,从而构建语音唤醒模块。其中,该神经网络可以为CNN(Convolutional Neural Network,卷积神经网络)、DNN(Deep Neural Network,深层神经网络)或RNN(Recurrent neural Network、循环神经网络)。

[0127] 若基于上述语音唤醒模块处理得到的处理结果为匹配成功时,得到该匹配的目标语音特征信息;从而实现了对该语音指令的有效识别。

[0128] 若基于上述语音唤醒模块处理得到的处理结果为匹配失败,则无法得到该匹配的目标语音特征信息,流程结束。

[0129] 在对所提取的语音特征信息进行检测时,通过该预设的语义识别模块和语音唤醒模块来分别进行检测,得到各自对应的检测结果,最终匹配得到所需的特征信息;通过上述两个模块的搭配检测,实现了对语音指令的精确检测匹配,提高了对语音指令处理的精准度。

[0130] 基于上述本发明所提供的语音处理方法,下面以三个具体的优选实施例对该方法做具体阐述,当然,该三个优选实施例仅是为了说明本发明方案所优选的实施方式,并不能代表本发明技术方案的全部。其中,上述本发明的语音处理方法可以应用于网络直播平台(可以是手机上的直播平台,也可以是电脑上的直播平台)中,也可以应用于多媒体采集设备(如手机终端的照相功能、摄像功能)中,还可以应用于多媒体设备(如电视)中。

[0131] 实施例一

[0132] 在用户打开手机的相机后,当在任意时刻采集到用户发送的语音指令“我要拍照”后,对该语音指令“我要拍照”进行解析,提取特征后得到该语音指令对应的语义特征,根据预先训练好的语义识别模块对该语义特征进行识别检测,得到对应的多个语义识别结果“拍照”、“我要”、“我要拍”、“我要拍照”、“要拍照”等,通过对得到的各语义识别结果进行识别,确定出与该语义特征语义匹配度最高的语义识别结果“拍照”,得到该目标语音的目标语音特征;并通过对其进行转换,得到可识别的目标语音“拍照”。之后,根据用户输入的手势动作(两个手指比划个‘耶’的姿势)进行对应的验证处理,并在对当前用户提供的手势动作验证通过后,基于该解析得到的目标语音“拍照”,实现在该手机屏幕上执行“拍照”处理。通过上述实施例,实现了基于语音的处理,并且通过语音指令的控制,实现了无需人为操作即可实现拍照的过程,降低了人为劳动力,同时实现了对于复杂语音指令有效处理,增加了

处理范围,也通过这种免去人为操作过程的处理,进一步提升了用户的使用感受。

[0133] 实施例二

[0134] 在用户使用手机的直播平台时,展示该直播平台对应的操作显示界面;当在任意时刻采集到用户发送的语音指令“下玫瑰雨”时,对该语音指令“下玫瑰雨”进行解析,提取特征后得到该语音指令对应的语义特征,根据预先训练好的语义识别模块对该解析得到的语义特征进行识别检测,得到对应的多个语义识别结果“下”、“玫瑰”、“玫瑰雨”、“下玫瑰雨”、“下玫瑰”等,通过对该得到的多个语义识别结果进行识别,确定出与该语义特征语义匹配度最高的语义识别结果“下玫瑰雨”;同时,在解析时还提取到该语音指令对应的MFCC特征信息,根据预先训练好的语音唤醒模块对该MFCC特征信息进行匹配检测,确定出与该MFCC特征信息匹配的目标语音的目标特征信息,得到该目标语音的目标语音特征“下玫瑰雨”;并结合上述语义识别模块的语义识别结果对该得到目标语音的目标语音特征进行进一步地验证,确认该目标语音的目标语音特征“拍照”与该语义匹配度最高的语义识别结果“下玫瑰雨”一致,为对应接收到的语音指令的语音特征,进而对该特征“拍照”进行转换得到可识别的目标语音“下玫瑰雨”,基于该确认的目标语音“下玫瑰雨”,在该直播平台中进行对应的下玫瑰雨处理。通过上述实施例,实现了基于语音的处理,并且通过语音指令的控制,实现了无需人为操作即可实现拍照的过程,降低了人为劳动力,同时通过语音唤醒模块与语义识别模块的结合处理,提升了识别的准确度,同时实现了对于复杂语音指令有效处理,增加了处理范围,也通过这种免去人为操作过程的处理,进一步提升了用户的使用感受。

[0135] 当然,在上述直播平台的实施例中,该语音指令还可以是“我要拍照”,通过对应的识别检测处理,实现在该直播平台中调用摄像头进行对应的拍照处理。

[0136] 实施例三

[0137] 当前用户打开电视,使电视处于打开状态,当用户准备去厨房做饭时,发送语音指令“暂停”,电视采集到用户所发送的该语音指令“暂停”,对该语音指令“暂停”进行解析,提取特征后得到对应的语义特征,根据预先训练好的语义识别模块对该语义特征进行匹配识别,确定出与该语义特征语义匹配度最高的语义识别结果,得到该目标语音的目标语音特征;并通过对其进行转换,得到可识别的目标语音“暂停”。进而基于该解析得到的目标语音“暂停”,实现在该电视上执行对应的暂停播放当前节目的处理。通过上述实施例,实现了基于语音的处理,以及根据语音指令的控制,实现了无需人为操作即可实现拍照的过程,降低了人为劳动力,同时实现了在复杂应用场景中对语音指令的有效处理,增加了处理范围,也通过这种免去人为操作过程的处理,进一步提升了用户的使用感受。

[0138] 基于上述本发明所提供的语音处理方法,本发明还提供了一种语音处理装置,如图2所示,包括:

[0139] 解析单元21,用于对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息;

[0140] 第一处理单元22,用于根据预设的语义识别模块对所述语音特征信息中包含的语义特征进行检测,得到检测结果,所述检测结果中包含有语义匹配度最高的语义识别结果;基于包含该语义识别结果的检测结果进行对应的处理。

[0141] 优选地,所述语音特征信息包括语义特征,

[0142] 所述第一处理单元22,还用于根据预设的语义识别模块对所述语义特征进行识别,得到多个语义识别结果;并在得到的多个语义识别结果中确认语义匹配度最高的语义识别结果。

[0143] 优选地,

[0144] 所述第一处理单元22,用于基于所述包含该语义识别结果的检测结果按照所述语音指令进行对应的处理;或,基于所述包含该语义识别结果的检测结果不做任何处理。

[0145] 优选地,所述第一处理单元22,具体用于确定所述语音指令对应的指示信息;根据所述指示信息做相应的处理。

[0146] 优选地,所述指示信息包括如下任一项:

[0147] 基于网络直播平台和/或多媒体采集设备中的特定指令;

[0148] 基于多媒体设备中的播放和/或暂停指令。

[0149] 优选地,所述特定指令包括如下任一项:

[0150] 拍照;

[0151] 摄像;

[0152] 拍照中添加特效信息;

[0153] 摄像中添加特效信息。

[0154] 优选地,还包括:

[0155] 获取单元23,用于获取当前用户触发的动作和/或人脸;

[0156] 第二处理单元24,用于对当前用户触发的动作和/或人脸进行识别检测,得到识别结果;

[0157] 所述第一处理单元22,还用于基于包含该语义识别结果的检测结果,并结合基于动作和/或人脸识别结果,进行对应的处理。

[0158] 优选地,

[0159] 所述第一处理单元22,还用于根据预设的语音唤醒模块对所述语音特征信息进行检测,得到检测结果。

[0160] 优选地,

[0161] 所述第一处理单元22,用于根据所述语音唤醒模块对所述语音特征信息进行匹配,确定所述语音唤醒模块中是否存储有与语音特征信息匹配的目标语音特征信息;并在匹配成功时,得到该匹配的目标语音特征信息。

[0162] 优选地,所述解析单元21,具体用于对所述语音指令进行声学特征提取,得到该语音指令对应的梅尔频率倒谱系数MFCC特征信息。

[0163] 本发明还提供了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计算机程序,该程序被处理器执行时实现上述的方法。

[0164] 本发明还提供了一种计算设备,包括:处理器、存储器、通信接口和通信总线,所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信;

[0165] 所述存储器用于存放至少一可执行指令,所述可执行指令使所述处理器执行上述的语音处理方法对应的操作。

[0166] 与现有技术相比,本发明至少具有以下优点:

[0167] 通过对获取到的语音指令进行解析,得到所述语音指令对应的语音特征信息的处

理,实现了对所需求语音指令的特征提取,为后续对于该提取的特征的检测处理提供了保障;并且通过预设的语义识别模块对该提取出的语音特征信息的检测,再根据检测结果来进行对应的处理,实现了基于语音指令的对应处理,实现了无需人为操作即可实现拍照的过程,降低了人为劳动力,同时实现了在复杂应用场景中对语音指令的有效处理,增加了处理范围;同时通过语音唤醒模块与语义识别模块的结合处理,提升了语音识别的准确度;也通过这种免去人为操作过程的处理,进一步提升了用户的使用感受。

[0168] 本技术领域技术人员可以理解,可以用计算机程序指令来实现这些结构图和/或框图和/或流图中的每个框以及这些结构图和/或框图和/或流图中的框的组合。本技术领域技术人员可以理解,可以将这些计算机程序指令提供给通用计算机、专业计算机或其他可编程数据处理方法的处理器来实现,从而通过计算机或其他可编程数据处理方法的处理器来执行本发明公开的结构图和/或框图和/或流图的框或多个框中指定的方案。

[0169] 其中,本发明装置的各个模块可以集成于一体,也可以分离部署。上述模块可以合并为一个模块,也可以进一步拆分成多个子模块。

[0170] 本领域技术人员可以理解附图只是一个优选实施例的示意图,附图中的模块或流程并不一定是实施本发明所必须的。

[0171] 本领域技术人员可以理解实施例中的装置中的模块可以按照实施例描述进行分布于实施例的装置中,也可以进行相应变化位于不同于本实施例的一个或多个装置中。上述实施例的模块可以合并为一个模块,也可以进一步拆分成多个子模块。

[0172] 上述本发明序号仅仅为了描述,不代表实施例的优劣。

[0173] 以上公开的仅为本发明的几个具体实施例,但是,本发明并非局限于此,任何本领域的技术人员能思之的变化都应落入本发明的保护范围。

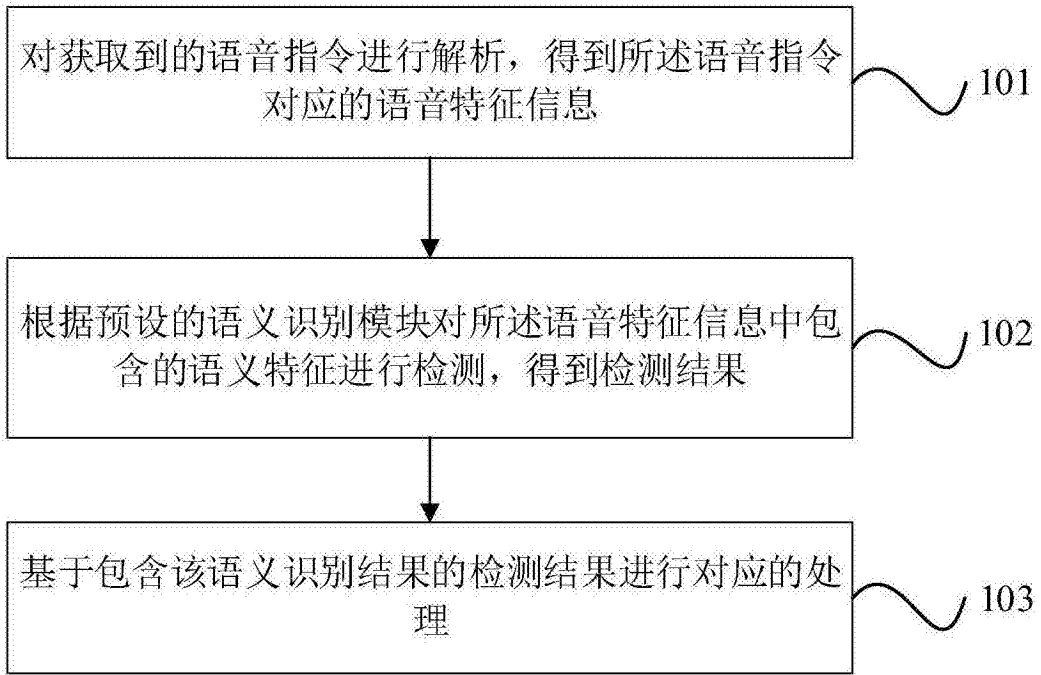


图1

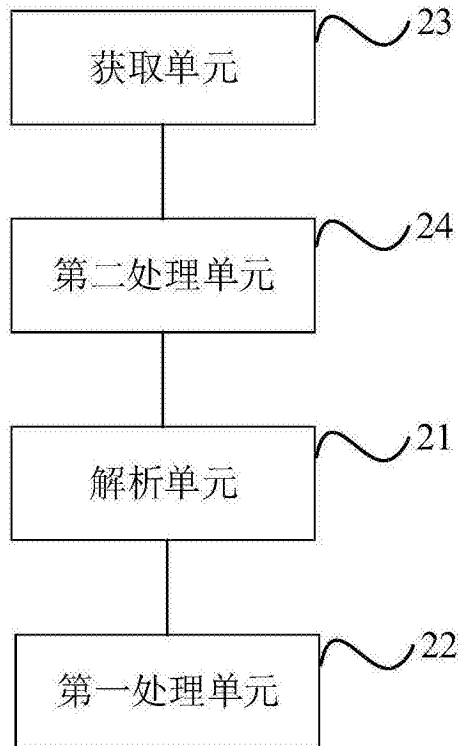


图2