



# (12)发明专利申请

(10)申请公布号 CN 109147825 A

(43)申请公布日 2019.01.04

(21)申请号 201810899641.3

(22)申请日 2018.08.09

(71)申请人 湖南永爱生物科技有限公司

地址 410205 湖南省长沙市长沙高新开发区桐梓坡西路408号麓谷林语小区I区综合体1栋501房

(72)发明人 徐国雄 刘分 刘任仲

(51)Int.Cl.

G10L 25/63(2013.01)

G10L 25/57(2013.01)

H04N 21/4788(2011.01)

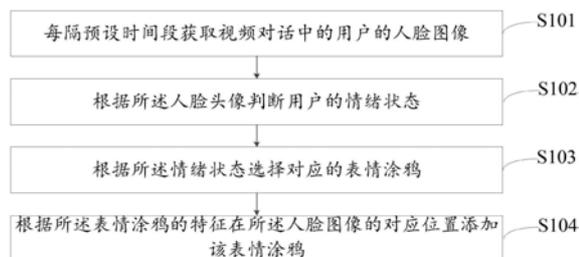
权利要求书2页 说明书6页 附图2页

## (54)发明名称

基于语音识别的人脸表情装饰方法、装置、存储介质及电子设备

## (57)摘要

本发明提供了一种基于语音识别的人脸表情装饰方法、装置、存储介质及电子设备,用于给视频对话过程中的用户添加表情涂鸦,该方法包括以下步骤:每隔预设时间段获取视频对话中的用户的对话音频数据;根据所述对话音频数据判断用户的情绪状态;根据所述情绪状态选择对应的表情涂鸦;根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。



1. 一种基于语音识别的人脸表情装饰方法,用于给视频对话过程中的用户添加表情涂鸦,其特征在于,该方法包括以下步骤:

每隔预设时间段获取视频对话中的用户的对话音频数据;

根据所述对话音频数据判断用户的情绪状态;

根据所述情绪状态选择对应的表情涂鸦;

根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。

2. 根据权利要求1所述的基于语音识别的人脸表情装饰方法,其特征在于,所述根据所述对话音频数据判断用户的情绪状态的步骤包括:

提取音频数据的当前音调信息以及当前音量信息;

对所述音频数据进行语音识别,提取出其中关于情绪的关键词汇;

结合该当前音调信息、当前音量信息以及该关键词汇判断出该用户的情绪状态。

3. 根据权利要求1所述的基于语音识别的人脸表情装饰方法,其特征在于,所述表情涂鸦包括多个表情元素,所述多个表情元素用于共同描述一种情绪状态,各个表情元素对应于人脸部的不同器官区域。

4. 根据权利要求3所述的基于语音识别的人脸表情装饰方法,其特征在于,所述根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦的步骤包括:

根据所述表情涂鸦的各个表情元素所对应的器官区域,将该表情涂鸦的每一个表情元素添加到与每一表情元素对应的器官区域的对应位置。

5. 根据权利要求1所述的基于语音识别的人脸表情装饰方法,其特征在于,所述情绪状态包括:严肃状态、愉悦状态、微怒状态、大怒状态、伤心状态、狂喜状态。

6. 一种基于语音识别的人脸表情装饰装置,其特征在于,包括:

获取模块,用于每隔预设时间段获取视频对话中的用户的对话音频数据;

判断模块,用于根据所述对话音频数据判断用户的情绪状态;

选择模块,用于根据所述情绪状态选择对应的表情涂鸦;

添加模块,用于根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。

7. 根据权利要求6所述的基于语音识别的人脸表情装饰装置,其特征在于,所述判断模块包括:

提取单元,用于提取音频数据的当前音调信息以及当前音量信息;

语音识别单元,用于对所述音频数据进行语音识别,提取出其中关于情绪的关键词汇;

判断单元,用于结合该当前音调信息、当前音量信息以及该关键词汇判断出该用户的情绪状态。

8. 根据权利要求6所述的基于语音识别的人脸表情装饰装置,其特征在于,所述表情涂鸦包括多个表情元素,所述多个表情元素用于共同描述一种情绪状态,各个表情元素对应于人脸部的不同器官区域;所述添加模块用于根据所述表情涂鸦的各个表情元素所对应的器官区域,将该表情涂鸦的每一个表情元素添加到与每一表情元素对应的器官区域的对应位置。

9. 一种存储介质,所述存储介质中存储有计算机程序,当所述计算机程序在计算机上运行时,使得所述计算机执行权利要求1-5任一项所述的方法。

10. 一种电子设备,包括处理器和存储器,所述存储器中存储有计算机程序,所述处理器通过调用所述存储器中存储的所述计算机程序,用于执行权利要求1-5任一项方法。

## 基于语音识别的人脸表情装饰方法、装置、存储介质及电子设备

### 技术领域

[0001] 本申请涉及通信领域,具体涉及一种基于语音识别的人脸表情装饰方法、装置、存储介质及电子设备。

### 背景技术

[0002] 现有技术,在视频对话例如QQ视频对话过程中,可以在对话的人脸图像中加入各种各样的表情涂鸦来增强对话的趣味性,但是,现有技术都是用户手动添加与自己情绪对应的表情涂鸦,操作麻烦,智能化程度不高,用户体验不够。

[0003] 因此,现有技术存在缺陷,急需改进。

### 发明内容

[0004] 本申请实施例提供一种基于语音识别的人脸表情装饰方法、装置、存储介质及电子设备,可以提高用户体验。

[0005] 本申请实施例提供了一种基于语音识别的人脸表情装饰方法,用于给视频对话过程中的用户添加表情涂鸦,该方法包括以下步骤:

[0006] 每隔预设时间段获取视频对话中的用户的对话音频数据;

[0007] 根据所述对话音频数据判断用户的情绪状态;

[0008] 根据所述情绪状态选择对应的表情涂鸦;

[0009] 根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。

[0010] 在本发明所述的基于语音识别的人脸表情装饰方法中,所述根据所述对话音频数据判断用户的情绪状态的步骤包括:

[0011] 提取音频数据的当前音调信息以及当前音量信息;

[0012] 对所述音频数据进行语音识别,提取出其中关于情绪的关键词汇;

[0013] 结合该当前音调信息、当前音量信息以及该关键词汇判断出该用户的情绪状态。

[0014] 在本发明所述的基于语音识别的人脸表情装饰方法中,所述表情涂鸦包括多个表情元素,所述多个表情元素用于共同描述一种情绪状态,各个表情元素对应于人脸部的不同器官区域。

[0015] 在本发明所述的基于语音识别的人脸表情装饰方法中,所述根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦的步骤包括:

[0016] 根据所述表情涂鸦的各个表情元素所对应的器官区域,将该表情涂鸦的每一个表情元素添加到与每一表情元素对应的器官区域的对应位置。

[0017] 在本发明所述的基于语音识别的人脸表情装饰方法中,所述情绪状态包括:严肃状态、愉悦状态、微怒状态、大怒状态、伤心状态、狂喜状态。

[0018] 一种基于语音识别的人脸表情装饰装置,包括:

[0019] 获取模块,用于每隔预设时间段获取视频对话中的用户的对话音频数据;

- [0020] 判断模块,用于根据所述对话音频数据判断用户的情绪状态;
- [0021] 选择模块,用于根据所述情绪状态选择对应的表情涂鸦;
- [0022] 添加模块,用于根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。
- [0023] 在本发明所述的基于语音识别的人脸表情装饰装置中,所述判断模块包括:
- [0024] 提取单元,用于提取音频数据的当前音调信息以及当前音量信息;
- [0025] 语音是被单元,用于对所述音频数据进行语音识别,提取出其中关于情绪的关键词汇;
- [0026] 判断单元,用于结合该当前音调信息、当前音量信息以及该关键词汇判断出该用户的情绪状态。
- [0027] 在本发明所述的基于语音识别的人脸表情装饰装置中,所述表情涂鸦包括多个表情元素,所述多个表情元素用于共同描述一种情绪状态,各个表情元素对应于人脸部的不同器官区域;所述添加模块用于根据所述表情涂鸦的各个表情元素所对应的器官区域,将该表情涂鸦的每一个表情元素添加到与每一表情元素对应的器官区域的对应位置。
- [0028] 一种存储介质,所述存储介质中存储有计算机程序,当所述计算机程序在计算机上运行时,使得所述计算机执行上述任一项所述的方法。
- [0029] 一种电子设备,包括处理器和存储器,所述存储器中存储有计算机程序,所述处理器通过调用所述存储器中存储的所述计算机程序,用于执行上述任一项方法。
- [0030] 由上可知,本发明通过每隔预设时间段获取视频对话中的用户的对话音频数据;根据所述对话音频数据判断用户的情绪状态;根据所述情绪状态选择对应的表情涂鸦;根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦,从而使得视频过程中可以自动添加符合用户情绪状态的表情涂鸦,具有提高用户体验的有益效果。

## 附图说明

- [0031] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍。显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。
- [0032] 图1是本申请实施例提供的基于语音识别的人脸表情装饰方法的流程示意图。
- [0033] 图2是本申请实施例提供的基于语音识别的人脸表情装饰装置的结构示意图。
- [0034] 图3是本申请实施例提供的电子设备的结构示意图。

## 具体实施方式

- [0035] 下面详细描述本申请的实施方式,所述实施方式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方式是示例性的,仅用于解释本申请,而不能理解为对本申请的限制。
- [0036] 在本申请的描述中,需要理解的是,术语“中心”、“纵向”、“横向”、“长度”、“宽度”、“厚度”、“上”、“下”、“前”、“后”、“左”、“右”、“竖直”、“水平”、“顶”、“底”、“内”、“外”、“顺时针”、“逆时针”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于

描述本申请和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此不能理解为对本申请的限制。此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个所述特征。在本申请的描述中,“多个”的含义是两个或两个以上,除非另有明确具体的限定。

[0037] 在本申请的描述中,需要说明的是,除非另有明确的规定和限定,术语“安装”、“相连”、“连接”应做广义理解,例如,可以是固定连接,也可以是可拆卸连接,或一体地连接;可以是机械连接,也可以是电连接或可以相互通讯;可以是直接相连,也可以通过中间媒介间接相连,可以是两个元件内部的连通或两个元件的相互作用关系。对于本领域的普通技术人员而言,可以根据具体情况理解上述术语在本申请中的具体含义。

[0038] 在本申请中,除非另有明确的规定和限定,第一特征在第二特征之“上”或之“下”可以包括第一和第二特征直接接触,也可以包括第一和第二特征不是直接接触而是通过它们之间的另外的特征接触。而且,第一特征在第二特征“之上”、“上方”和“上面”包括第一特征在第二特征正上方和斜上方,或仅仅表示第一特征水平高度高于第二特征。第一特征在第二特征“之下”、“下方”和“下面”包括第一特征在第二特征正下方和斜下方,或仅仅表示第一特征水平高度小于第二特征。

[0039] 下文的公开提供了许多不同的实施方式或例子用来实现本申请的不同结构。为了简化本申请的公开,下文中对特定例子的部件和设置进行描述。当然,它们仅仅为示例,并且目的不在于限制本申请。此外,本申请可以在不同例子中重复参考数字和/或参考字母,这种重复是为了简化和清楚的目的,其本身不指示所讨论各种实施方式和/或设置之间的关系。此外,本申请提供了的各种特定的工艺和材料的例子,但是本领域普通技术人员可以意识到其他工艺的应用和/或其他材料的使用。

[0040] 本申请的说明书和权利要求书以及上述附图中的术语“第一”、“第二”、“第三”等(如果存在)是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应当理解,这样描述的对象在适当情况下可以互换。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含。例如,包含了一系列步骤的过程、方法或包含了一系列模块或单元的装置、终端、系统不必限于清楚地列出的那些步骤或模块或单元,还可以包括没有清楚地列出的步骤或模块或单元,也可以包括对于这些过程、方法、装置、终端或系统固有的其它步骤或模块或单元。

[0041] 参考图1,图1为本发明本申请实施例提供了一种基于语音识别的人脸表情装饰方法,应用于手机、PAD等电子设备中。该基于语音识别的人脸表情装饰方法,用于给视频对话过程中的用户添加表情涂鸦,该方法包括以下步骤:

[0042] S101、每隔预设时间段获取视频对话中的用户的对话音频数据。

[0043] 由于人的情绪会随着视频对话的持续进行可能不断发生变化,因此,每隔预设时间段就需要采集一次用户的对话音频数据,以便于后续的情绪状态识别。在本实施例中,例如可以每隔10秒获取一次对话音频数据。当然,其并不限于此。该对话音频数据的时长为5秒到10秒。

[0044] S102、根据所述对话音频数据判断用户的情绪状态。

[0045] 在该步骤中,可以采用现有技术中的语音识别技术来进行人情绪识别。当然,在本

发明中,可以结合用户的语调、音量以及对话中的关于情绪的一些词语,例如“气死”、“你他X”、“滚蛋”、“呵呵”等。

[0046] 在一些实施例中,该步骤S102包括:提取音频数据的当前音调信息以及当前音量信息;对所述音频数据进行语音识别,提取出其中关于情绪的关键词汇;结合该当前音调信息、当前音量信息以及该关键词汇判断出该用户的情绪状态。

[0047] 其中,结合当前音调信息、当前音量信息进行判断时,还要结合用户的平均音调信息以及平均音量信息来判断,看当前音调信息、当前音量信息与平均音调信息以及平均音量信息的差值,来判断,并结合提取出的关键词,例如,用户以很高的音调以及音量说“呵呵”,说明其实用户已经生气了。不同的词语在不同的音调以及音量下说出来对应的情绪状态是不一样的。

[0048] S103、根据所述情绪状态选择对应的表情涂鸦。

[0049] 其中,表情涂鸦包括多个表情元素,所述多个表情元素用于共同描述一种情绪状态,各个表情元素对应于人脸部的不同器官区域。例如,对应于大怒状态的表情涂鸦,其包括围绕人脸的火焰表情元素以及眼睛区域火眼元素等。对应于伤心状态,则有在人眼睛以及眼睛以下区域的眼泪,嘴角下垂等。在此不一一例举。

[0050] S104、根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。

[0051] 在该步骤S104中,根据表情涂鸦的各个表情元素所对应的器官区域,将该表情涂鸦的每一个表情元素添加到与每一表情元素对应的器官区域的对应位置。例如,当用户大怒时,将火眼元素添加到用户的眼睛位置处,将火焰表情元素添加到头顶以及人脸上。

[0052] 由上可知,本发明通过每隔预设时间段获取视频对话中的用户的对话音频数据;根据所述对话音频数据判断用户的情绪状态;根据所述情绪状态选择对应的表情涂鸦;根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦,从而使得视频过程中可以自动添加符合用户情绪状态的表情涂鸦,具有提高用户体验的有益效果。

[0053] 本发明还提供了一种基于语音识别的人脸表情装饰装置,包括:获取模块201、判断模块202、选择模块203以及添加模块204。

[0054] 其中,该获取模块201用于每隔预设时间段获取视频对话中的用户的对话音频数据;由于人的情绪会随着视频对话的持续进行可能不断发生变化,因此,每隔预设时间段就需要采集一次用户的对话音频数据,以便于后续的情绪状态识别。在本实施例中,例如可以每隔10秒获取一次对话音频数据。当然,其并不限于此。该对话音频数据的时长为5秒到10秒。

[0055] 其中,该判断模块202用于根据所述对话音频数据判断用户的情绪状态;可以采用现有技术中的语音识别技术来进行人情绪识别。当然,在本发明中,可以结合用户的语调、音量以及对话中的关于情绪的一些词语,例如“气死”、“你他X”、“滚蛋”、“呵呵”等。在一些实施例中,该判断模块202包括:提取单元,用于提取音频数据的当前音调信息以及当前音量信息;语音识别单元,用于对所述音频数据进行语音识别,提取出其中关于情绪的关键词汇;判断单元,用于结合该当前音调信息、当前音量信息以及该关键词汇判断出该用户的情绪状态。其中,结合当前音调信息、当前音量信息进行判断时,还要结合用户的平均音调信息以及平均音量信息来判断,看当前音调信息、当前音量信息与平均音调信息以及平均音量信息的差值,来判断,并结合提取出的关键词,例如,用户以很高的音调以及音量说“呵

呵”，说明其实用户已经生气了。不同的词语在不同的音调以及音量下说出来对应的情绪状态是不一样的。

[0056] 其中，该选择模块203用于根据所述情绪状态选择对应的表情涂鸦；其中，表情涂鸦包括多个表情元素，所述多个表情元素用于共同描述一种情绪状态，各个表情元素对应于人脸部的不同器官区域。例如，对应于大怒状态的表情涂鸦，其包括围绕人脸的火焰表情元素以及眼睛区域火眼元素等。对应于伤心状态，则有在人眼睛以及眼睛以下区域的眼泪，嘴角下垂等。在此不一一例举。

[0057] 其中，该添加模块204用于根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。根据表情涂鸦的各个表情元素所对应的器官区域，将该表情涂鸦的每一个表情元素添加到与每一表情元素对应的器官区域的对应位置。例如，当用户大怒时，将火眼元素添加到用户的眼睛位置处，将火焰表情元素添加到头顶以及人脸上。

[0058] 由上可知，本发明通过每隔预设时间段获取视频对话中的用户的对话音频数据；根据所述对话音频数据判断用户的情绪状态；根据所述情绪状态选择对应的表情涂鸦；根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦，从而使得视频过程中可以自动添加符合用户情绪状态的表情涂鸦，具有提高用户体验的有益效果。

[0059] 本申请实施例还提供一种存储介质，该存储介质中存储有计算机程序，当该计算机程序在计算机上运行时，该计算机执行上述任一实施例所述的基于语音识别的人脸表情装饰方法，从而实现以下功能：每隔预设时间段获取视频对话中的用户的对话音频数据；根据所述对话音频数据判断用户的情绪状态；根据所述情绪状态选择对应的表情涂鸦；根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。

[0060] 请参照图3，本申请实施例还提供一种电子设备。该电子设备可以是智能手机、平板电脑等设备。如示，电子设备300包括处理器301和存储器302。其中，处理器301与存储器302电性连接。处理器301是终端300的控制中心，利用各种接口和线路连接整个终端的各个部分，通过运行或调用存储在存储器302内的计算机程序，以及调用存储在存储器302内的数据，执行终端的各种功能和处理数据，从而对终端进行整体监控。

[0061] 在本实施例中，电子设备300中的处理器301会按照如下的步骤，将一个或一个以上的计算机程序的进程对应的指令加载到存储器302中，并由处理器301来运行存储在存储器302中的计算机程序，从而实现各种功能：每隔预设时间段获取视频对话中的用户的对话音频数据；根据所述对话音频数据判断用户的情绪状态；根据所述情绪状态选择对应的表情涂鸦；根据所述表情涂鸦的特征在所述人脸图像的对应位置添加该表情涂鸦。

[0062] 存储器302可用于存储计算机程序和数据。存储器302存储的计算机程序中包含有可在处理器中执行的指令。计算机程序可以组成各种功能模块。处理器301通过调用存储在存储器302的计算机程序，从而执行各种功能应用以及数据处理。

[0063] 需要说明的是，本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成，该程序可以存储于计算机可读存储介质中，该存储介质可以包括但不限于：只读存储器 (ROM, Read Only Memory)、随机存取存储器 (RAM, Random Access Memory)、磁盘或光盘等。

[0064] 以上对本申请实施例所提供的数据分布存储方法、装置、存储介质及电子设备进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施

例的说明只是用于帮助理解本申请的方法及其核心思想;同时,对于本领域的技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本申请的限制。

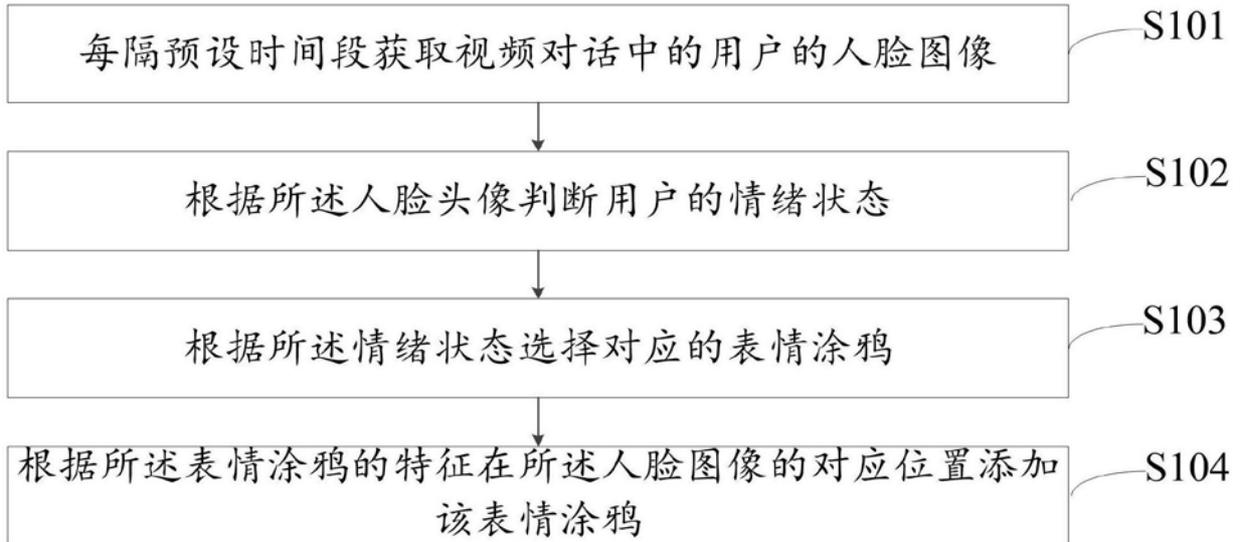


图1



图2

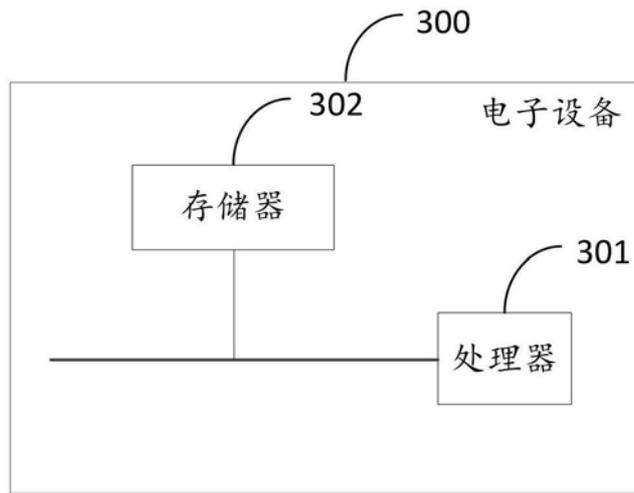


图3