



(12) 发明专利

(10) 授权公告号 CN 111107421 B

(45) 授权公告日 2022. 04. 01

(21) 申请号 201911417875.0

(22) 申请日 2019.12.31

(65) 同一申请的已公布的文献号
申请公布号 CN 111107421 A

(43) 申请公布日 2020.05.05

(73) 专利权人 北京达佳互联信息技术有限公司
地址 100085 北京市海淀区上地西路6号1
幢1层101D1-7

(72) 发明人 谢纨楠 范威 李杰

(74) 专利代理机构 北京柏杉松知识产权代理事
务所(普通合伙) 11413
代理人 丁芸 马敬

(51) Int. Cl.
H04N 21/433 (2011.01)
H04N 21/442 (2011.01)

(56) 对比文件

CN 108900891 A, 2018.11.27

CN 107959864 A, 2018.04.24

CN 108377396 A, 2018.08.07

US 10028004 B1, 2018.07.17

CN 106792122 A, 2017.05.31

审查员 周立秋

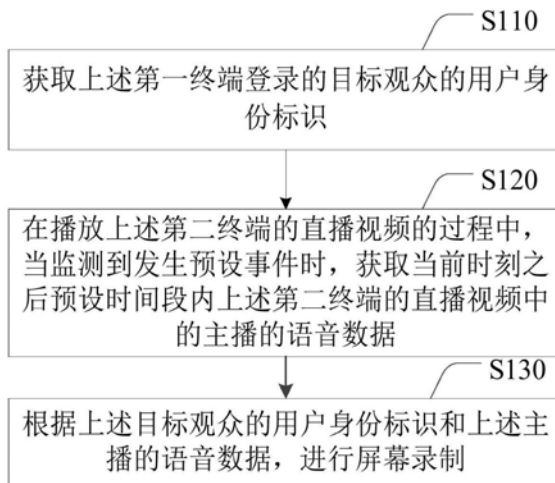
权利要求书3页 说明书13页 附图3页

(54) 发明名称

视频处理方法、装置、终端设备及存储介质

(57) 摘要

本公开的实施例提供的视频处理方法、装置、终端设备、存储介质及包含指令的计算机程序产品,涉及计算机技术领域,通过在播放第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内所述第二终端的直播视频中的主播的语音数据,根据所述目标观众的用户身份标识和所述主播的语音数据,进行屏幕录制,从而实现了在第一终端自动录制直播过程中的精彩内容,提高观众互动体验。



1. 一种视频处理方法,其特征在于,应用于第一终端,所述第一终端用于在第二终端直播时,播放所述第二终端的直播视频,所述方法包括:

获取所述第一终端登录的目标观众的用户身份标识;

在播放所述第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内所述第二终端的直播视频中的主播的语音数据;

根据所述目标观众的用户身份标识和所述主播的语音数据,判断所述主播的语音数据是否与所述目标观众有关;

当所述主播的语音数据与所述目标观众有关时,启动屏幕录制;

所述根据所述目标观众的用户身份标识和所述主播的语音数据,判断所述主播的语音数据是否与所述目标观众有关,包括:

分别对所述目标观众的用户身份标识和所述主播的语音数据进行特征提取,得到目标观众用户身份标识的语音特征和主播的语音特征;

根据所述目标观众用户身份标识的语音特征和所述主播的语音特征,判断所述主播的语音数据是否与所述目标观众有关。

2. 根据权利要求1所述的方法,其特征在于,所述预设事件包括:所述目标观众发送给所述主播的礼物金额超过预设金额;或者

所述目标观众发送给所述主播的礼物数量超过预设数量;或者

所述目标观众发送给所述主播的礼物等级超过预设等级。

3. 根据权利要求1所述的方法,其特征在于,所述分别对所述目标观众的用户身份标识和所述主播的语音数据进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征,包括:

基于梅尔倒谱系数MFCC方法对所述主播的语音数据进行特征提取,得到主播的语音数据特征;

分别对所述目标观众的用户身份标识和所述主播的语音数据特征进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征。

4. 根据权利要求3所述的方法,其特征在于,所述根据所述目标观众的用户身份标识和所述主播的语音数据,判断所述主播的语音数据是否与所述目标观众有关,包括:

根据预设算法将所述目标观众用户身份标识的语音特征和所述主播的语音特征进行语音相似度计算,得到目标语音相似度;

当所述目标语音相似度超过预设阈值时,确定所述主播的语音数据与所述目标观众有关。

5. 根据权利要求1所述的方法,其特征在于,所述启动屏幕录制,包括:

从当前时刻之前的第一预设时长的时刻开始,启动屏幕录制;或者

从当前时刻开始,启动屏幕录制;或者

从当前时刻之后的第二预设时长的时刻开始,启动屏幕录制。

6. 根据权利要求1所述的方法,其特征在于,在所述启动屏幕录制屏幕的步骤之后,所述方法还包括:

在屏幕录制预设时长后,结束屏幕录制;或者

当监测到所述第一终端发生第二预设事件时,结束屏幕录制。

7. 根据权利要求6所述的方法,其特征在于,所述当监测到所述第一终端发生第二预设事件时,结束屏幕录制,包括:

当监测到所述第一终端的屏幕展示的预设特效的次数达到预设次数阈值时,结束屏幕录制;或者

当监测到所述第一终端接收到预设音频信号时,结束屏幕录制。

8. 根据权利要求1所述的方法,其特征在于,所述当所述主播的语音数据与所述目标观众有关时,所述方法还包括:

在所述第一终端展示预设下载选项;

在获取所述目标观众针对所述预设下载选项的指定操作时,将录屏视频保存到本地。

9. 一种视频处理装置,其特征在于,应用于第一终端,所述第一终端用于在第二终端直播时,播放所述第二终端的直播视频,所述装置包括:

第一获取模块,被配置为获取所述第一终端登录的目标观众的用户身份标识;

第二获取模块,被配置为在播放所述第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内所述第二终端的直播视频中的主播的语音数据;

录制模块,被配置为根据所述目标观众的用户身份标识和所述主播的语音数据,判断所述主播的语音数据是否与所述目标观众有关;当所述主播的语音数据与所述目标观众有关时,启动屏幕录制;

所述录制模块具体用于:

分别对所述目标观众的用户身份标识和所述主播的语音数据进行特征提取,得到目标观众用户身份标识的语音特征和主播的语音特征;

根据所述目标观众用户身份标识的语音特征和所述主播的语音特征,判断所述主播的语音数据是否与所述目标观众有关。

10. 根据权利要求9所述的装置,其特征在于,所述预设事件包括:所述目标观众发送给所述主播的礼物金额超过预设金额;或者

所述目标观众发送给所述主播的礼物数量超过预设数量;或者

所述目标观众发送给所述主播的礼物等级超过预设等级。

11. 根据权利要求9所述的装置,其特征在于,所述录制模块具体用于:

基于梅尔倒谱系数MFCC装置对所述主播的语音数据进行特征提取,得到主播的语音数据特征;

分别对所述目标观众的用户身份标识和所述主播的语音数据特征进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征。

12. 根据权利要求11所述的装置,其特征在于,所述录制模块具体用于:

根据预设算法将所述目标观众用户身份标识的语音特征和所述主播的语音特征进行语音相似度计算,得到目标语音相似度;

当所述目标语音相似度超过预设阈值时,确定所述主播的语音数据与所述目标观众有关。

13. 根据权利要求9所述的装置,其特征在于,所述录制模块具体用于:

从当前时刻之前的第一预设时长的时刻开始,启动屏幕录制;或者

从当前时刻开始,启动屏幕录制;或者

从当前时刻之后的第二预设时长的时刻开始,启动屏幕录制。

14. 根据权利要求9所述的装置,其特征在于,所述装置还包括结束模块,被配置为:

在屏幕录制预设时长后,结束屏幕录制;或者

当监测到所述第一终端发生第二预设事件时,结束屏幕录制。

15. 根据权利要求14所述的装置,其特征在于,所述结束模块具体用于:

当监测到所述第一终端的屏幕展示的预设特效的次数达到预设次数阈值时,结束屏幕录制;或者

当监测到所述第一终端接收到预设音频信号时,结束屏幕录制。

16. 根据权利要求9所述的装置,其特征在于,所述装置还包括:

展示模块,被配置为在所述第一终端展示预设下载选项;

存储模块,被配置为在获取所述目标观众针对所述预设下载选项的指定操作时,将录屏视频保存到本地。

17. 一种终端设备,其特征在于,包括:处理器、通信接口、存储器和通信总线,其中:

所述处理器,通信接口,存储器通过通信总线完成相互间的通信;

存储器,被配置为存放计算机程序;

处理器,被配置为执行存储器上所存放的程序时,实现如权利要求1-8中任一项所述的视频处理方法。

18. 一种存储介质,其特征在于,所述存储介质内存储有计算机程序,所述计算机程序被处理器执行时实现权利要求1-8中任一所述的视频处理方法。

视频处理方法、装置、终端设备及存储介质

技术领域

[0001] 本公开涉及计算机技术领域，尤其涉及视频处理方法、装置、终端设备及存储介质。

背景技术

[0002] 随着网络技术的发展，同一时间透过网络系统在不同交流平台观看视频的网络直播技术越来越流行，主播可以采用网络直播方式通过网络直播平台发布网络直播数据，观众通过电视、手机、平板电脑等网络直播平台进行同步观看。为了提高用户体验，通过网络直播平台，观众可以通过发送弹幕消息、发送礼物等行为实现与主播的互动，在观众和主播互动过程中经常会遇到值得纪念或分享的精彩内容，目前，观众只能手动触发视频录制来记录这些内容，手动录制往往会导致错失直播视频录制时机，而且手动录制直播内容的方法为用户带来诸多不便，观众体验差。

发明内容

[0003] 本公开提供一种视频处理方法、装置、终端设备、存储介质及包含指令的计算机程序产品，以实现在第一终端自动录制直播过程中的精彩内容，提高观众互动体验。

[0004] 本公开的技术方案如下：

[0005] 根据本公开实施例的第一方面，本公开提供一种视频处理方法，应用于第一终端，所述第一终端用于在第二终端直播时，播放所述第二终端的直播视频，所述方法包括：

[0006] 获取所述第一终端登录的目标观众的用户身份标识；

[0007] 在播放所述第二终端的直播视频的过程中，当监测到发生预设事件时，获取当前时刻之后预设时间段内所述第二终端的直播视频中的主播的语音数据；

[0008] 根据所述目标观众的用户身份标识和所述主播的语音数据，进行屏幕录制。

[0009] 可选的，所述预设事件包括：所述目标观众发送给所述主播的礼物金额超过预设金额；或者

[0010] 所述目标观众发送给所述主播的礼物数量超过预设数量；或者

[0011] 所述目标观众发送给所述主播的礼物等级超过预设等级。

[0012] 可选的，所述根据所述目标观众的用户身份标识和所述主播的语音数据，进行屏幕录制，包括：

[0013] 根据所述目标观众的用户身份标识和所述主播的语音数据，判断所述主播的语音数据是否与所述目标观众有关；

[0014] 当所述主播的语音数据与所述目标观众有关时，启动屏幕录制。

[0015] 可选的，所述根据所述目标观众的用户身份标识和所述主播的语音数据，判断所述主播的语音数据是否与所述目标观众有关，包括：

[0016] 分别对所述目标观众的用户身份标识和所述主播的语音数据进行特征提取，得到目标观众用户身份标识的语音特征和主播的语音特征；

[0017] 根据所述目标观众用户身份标识的语音特征和所述主播的语音特征,判断所述主播的语音数据是否与所述目标观众有关。

[0018] 可选的,所述分别对所述目标观众的用户身份标识和所述主播的语音数据进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征,包括:

[0019] 基于梅尔倒谱系数MFCC方法对所述主播的语音数据进行特征提取,得到主播的语音数据特征;

[0020] 分别对所述目标观众的用户身份标识和所述主播的语音数据特征进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征。

[0021] 可选的,所述根据所述目标观众的用户身份标识和所述主播的语音数据,判断所述主播的语音数据是否与所述目标观众有关,包括:

[0022] 根据预设算法将所述目标观众用户身份标识的语音特征和所述主播的语音特征进行语音相似度计算,得到目标语音相似度;

[0023] 当所述目标语音相似度超过预设阈值时,确定所述主播的语音数据与所述目标观众有关。

[0024] 可选的,所述启动屏幕录制,包括:

[0025] 从当前时刻之前的第一预设时长的时刻开始,启动屏幕录制;或者

[0026] 从当前时刻开始,启动屏幕录制;或者

[0027] 从当前时刻之后的第二预设时长的时刻开始,启动屏幕录制。

[0028] 可选的,在所述启动屏幕录制屏幕的步骤之后,所述方法还包括:

[0029] 在屏幕录制预设时长后,结束屏幕录制;或者

[0030] 当监测到所述第一终端发生第二预设事件时,结束屏幕录制。

[0031] 可选的,所述当监测到所述第一终端发生第二预设事件时,结束屏幕录制,包括:

[0032] 当监测到所述第一终端的屏幕展示的预设特效的次数达到预设次数阈值时,结束屏幕录制;或者

[0033] 当监测到所述第一终端接收到预设音频信号时,结束屏幕录制。

[0034] 可选的,所述当所述主播的语音数据与所述目标观众有关时,所述方法还包括:

[0035] 在所述第一终端展示预设下载选项;

[0036] 在获取所述目标观众针对所述预设下载选项的指定操作时,将录屏视频保存到本地。

[0037] 根据本公开实施例的第二方面,本公开提供一种视频处理装置,应用于第一终端,所述第一终端用于在第二终端直播时,播放所述第二终端的直播视频,所述装置包括:

[0038] 第一获取模块,被配置为获取所述第一终端登录的目标观众的用户身份标识;

[0039] 第二获取模块,被配置为在播放所述第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内所述第二终端的直播视频中的主播的语音数据;

[0040] 录制模块,被配置为根据所述目标观众的用户身份标识和所述主播的语音数据,进行屏幕录制。

[0041] 可选的,所述预设事件包括:所述目标观众发送给所述主播的礼物金额超过预设金额;或者

- [0042] 所述目标观众发送给所述主播的礼物数量超过预设数量;或者
- [0043] 所述目标观众发送给所述主播的礼物等级超过预设等级。
- [0044] 可选的,所述录制模块具体用于:
- [0045] 根据所述目标观众的用户身份标识和所述主播的语音数据,判断所述主播的语音数据是否与所述目标观众有关;
- [0046] 当所述主播的语音数据与所述目标观众有关时,启动屏幕录制。
- [0047] 可选的,所述录制模块具体用于:
- [0048] 分别对所述目标观众的用户身份标识和所述主播的语音数据进行特征提取,得到目标观众用户身份标识的语音特征和主播的语音特征;
- [0049] 根据所述目标观众用户身份标识的语音特征和所述主播的语音特征,判断所述主播的语音数据是否与所述目标观众有关。
- [0050] 可选的,所述录制模块具体用于:
- [0051] 基于梅尔倒谱系数MFCC装置对所述主播的语音数据进行特征提取,得到主播的语音数据特征;
- [0052] 分别对所述目标观众的用户身份标识和所述主播的语音数据特征进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征。
- [0053] 可选的,所述录制模块具体用于:
- [0054] 根据预设算法将所述目标观众用户身份标识的语音特征和所述主播的语音特征进行语音相似度计算,得到目标语音相似度;
- [0055] 当所述目标语音相似度超过预设阈值时,确定所述主播的语音数据与所述目标观众有关。
- [0056] 可选的,所述录制模块具体用于:
- [0057] 从当前时刻之前的第一预设时长的时刻开始,启动屏幕录制;或者
- [0058] 从当前时刻开始,启动屏幕录制;或者
- [0059] 从当前时刻之后的第二预设时长的时刻开始,启动屏幕录制。
- [0060] 可选的,所述装置还包括结束模块,被配置为:
- [0061] 在屏幕录制预设时长后,结束屏幕录制;或者
- [0062] 当监测到所述第一终端发生第二预设事件时,结束屏幕录制。
- [0063] 可选的,所述结束模块具体用于:
- [0064] 当监测到所述第一终端的屏幕展示的预设特效的次数达到预设次数阈值时,结束屏幕录制;或者
- [0065] 当监测到所述第一终端接收到预设音频信号时,结束屏幕录制。
- [0066] 可选的,所述装置还包括:
- [0067] 展示模块,被配置为在所述第一终端展示预设下载选项;
- [0068] 存储模块,被配置为在获取所述目标观众针对所述预设下载选项的指定操作时,将录屏视频保存到本地。
- [0069] 根据本公开实施例的第三方面,本公开提供一种终端设备,包括:处理器、通信接口、存储器和通信总线,其中,所述处理器,通信接口,存储器通过通信总线完成相互间的通信;

[0070] 存储器,被配置为存放计算机程序;

[0071] 处理器,被配置为执行存储器上所存放的程序时,实现上述第一方面任一项所述的视频处理方法。

[0072] 根据本公开实施例的第四方面,提供一种存储介质,所述存储介质内存储有计算机程序,所述计算机程序被处理器执行时实现上述第一方面任一所述的视频处理方法。

[0073] 根据本公开实施例的第五方面,提供了一种包含指令的计算机程序产品,当其在计算机上运行时,使得计算机执行上述第一方面任一所述的视频处理方法。

[0074] 本公开的实施例提供的视频处理方法、装置、终端设备、存储介质及包含指令的计算机程序产品,至少带来以下有益效果:

[0075] 通过在播放第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内所述第二终端的直播视频中的主播的语音数据,根据所述目标观众的用户身份标识和所述主播的语音数据,进行屏幕录制,从而实现了在第一终端自动录制直播过程中的精彩内容,提高观众互动体验。

[0076] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0077] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理,并不构成对本公开的不当限定。

[0078] 图1是根据一示例性实施例示出的一种视频处理方法的示意图;

[0079] 图2是根据一示例性实施例示出的一种视频处理装置的框图;

[0080] 图3是根据一示例性实施例示出的第一种终端设备的框图;

[0081] 图4是根据一示例性实施例示出的第二种终端设备的框图。

具体实施方式

[0082] 为了使本领域普通人员更好地理解本公开的技术方案,下面将结合附图,对本公开实施例中的技术方案进行清楚、完整地描述。

[0083] 需要说明的是,本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等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。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0084] 本公开实施例公开了一种视频处理方法、装置、终端设备、存储介质及包含指令的计算机程序产品,以下分别进行说明。

[0085] 图1是根据一示例性实施例示出的一种视频处理方法的示意图,上述视频处理方法应用于第一终端,上述第一终端用于在第二终端直播时,播放上述第二终端的直播视频,如图1所示,包括以下步骤:

[0086] 在步骤S110中,获取上述第一终端登录的目标观众的用户身份标识。

[0087] 本公开实施例的视频处理方法可以通过终端设备实现,具体的,该终端设备可以是智能手机或电脑等。

[0088] 目前,主播在第二终端进行直播时,观众用户可以通过第一终端观看主播的直播视频,目标观众登录上述第一终端后,获取上述第一终端登录的目标观众的用户身份标识,上述目标观众的用户身份标识可以为目标观众的用户名、目标观众的账号,例如,终端设备将上述第一终端登录的目标观众的账号下载至本地。

[0089] 在步骤S120中,在播放上述第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内上述第二终端的直播视频中的主播的语音数据。

[0090] 在视频直播的过程中,上述目标观众可以给主播发送礼物,其中礼物可以为单个,也可以为多个礼物,单个礼物设置有礼物等级和礼物金额,多个礼物时,包括礼物数量。

[0091] 在一种可能的实施方式中,上述预设事件包括:上述目标观众发送给上述主播的礼物金额超过预设金额;或者

[0092] 上述目标观众发送给上述主播的礼物数量超过预设数量;或者

[0093] 上述目标观众发送给上述主播的礼物等级超过预设等级。

[0094] 预设事件可以为上述目标观众发送给上述主播的礼物金额超过预设金额;或者,上述目标观众发送给上述主播的礼物数量超过预设数量;或者,上述目标观众发送给上述主播的礼物等级超过预设等级中的一项或多项时,获取当前时刻之后预设时间段内上述第二终端的直播视频中的主播的语音数据,其中预设时间段是可以根据实际需要进行设定的。例如,观众给主播发送的礼物有3个等级,礼物1,礼物2,礼物3,上述3个等级的礼物对应不同的价格,礼物1的价格为8元,礼物2的价格为18元,礼物3的价格为88元,目标观众A给主播发送了1个礼物3,上述目标观众发送给上述主播的礼物3的等级超过预设等级2,则表示监测到发生预设事件;例如,目标观众A给主播发送了100个礼物1,上述目标观众A发送给上述主播的礼物1的数量超过预设礼物1的数量88时,则表示监测到发生预设事件;例如,目标观众A给主播发送了30个礼物1,20个礼物2,其发送礼物的金额超过预设礼物金额400元时,则表示监测到发生预设事件。当监测到发生预设事件时,获取当前时刻之后预设时间段内上述第二终端的直播视频中的主播的语音数据,具体的,可以基于MFCC (Mel-Frequency Cepstral Coefficients,梅尔倒谱系数)方法对直播视频中的主播的语音数据进行特征提取,得到主播的语音数据,例如,基于梅尔倒谱系数MFCC方法从当前时刻之后20s内上述第二终端的直播视频中获取主播的语音数据,根据主播的语音数据判断从当前时刻之后20s内,主播是否对目标观众A进行口播,其中口播表示主播对观众说的话,具体的,当主播收到目标观众送的礼物后,对目标观众进行感谢等。

[0095] 在步骤S130中,根据上述目标观众的用户身份标识和上述主播的语音数据,进行屏幕录制。

[0096] 根据上述目标观众的用户身份标识和上述主播的语音数据,可以判断上述主播的语音数据是否和上述目标观众的有关,从而进行屏幕录制。

[0097] 在一种可能的实施方式中,上述根据上述目标观众的用户身份标识和上述主播的语音数据,进行屏幕录制,包括:

[0098] 根据上述目标观众的用户身份标识和上述主播的语音数据,判断上述主播的语音数据是否与上述目标观众有关;

[0099] 当上述主播的语音数据与上述目标观众有关时,启动屏幕录制。

[0100] 例如,根据上述目标观众的用户身份标识和上述主播的语音数据,判断上述主播的语音数据是否与上述目标观众有关。例如,目标观众的用户身份标识为A,基于预设转换模型将目标观众的用户身份标识转换为语音特征,得到目标观众用户身份标识的语音特征,根据目标观众用户身份标识的语音特征的类型,对主播的语音数据进行特征提取,得到主播的语音特征。例如,目标观众用户身份标识的语音特征为音频特征,则将主播的语音数据进行特征提取,得到相应的主播的语音数据的音频特征,例如,目标观众用户身份标识的语音特征为文字拼音文本特征,则将主播的语音数据转换为相应的拼音文本。根据预设算法将上述目标观众用户身份标识的语音特征和上述主播的语音特征进行语音相似度计算,得到目标语音相似度,具体的,根据动态时间规整算法将上述目标观众用户身份标识的语音特征和上述主播的语音特征进行语音相似度计算,得到目标语音相似度,上述目标语音相似度超过预设阈值时,判断上述主播的语音数据与上述目标观众有关,当目标语音相似度未超过预设阈值时,判断上述主播的语音数据与上述目标观众无关。例如,其目标语音相似度为0.7,超过预设阈值0.5,则,判断上述主播的语音数据与上述目标观众有关。根据上述目标观众的用户身份标识和上述主播的语音数据,判断上述主播的语音数据是否与上述目标观众有关,当上述主播的语音数据与上述目标观众有关时,进行屏幕录制,从而实现了在第一终端自动录制直播过程中的精彩内容,提高观众互动体验。

[0101] 当上述主播的语音数据与上述目标观众有关时,进行屏幕录制,即,目标观众A给主播发送了1个礼物3,上述目标观众发送给上述主播的礼物3的等级超过预设等级2,获取当前时刻之后20s内上述第二终端的直播视频中的主播的语音数据,将目标观众的用户身份标识转换为语音特征,同时,对上述主播的语音数据进行特征提取,转换为相应的语音特征,经过计算,目标语音相似度为0.7,超过预设阈值0.5,判断上述主播的语音数据与上述目标观众有关,进行屏幕录制。进一步的,对录制预设时长的视频,例如,录制20s的视频,更进一步的,可以在第一终端屏幕生成下载至本地按钮,观众通过操作上述按钮,将上述20s录屏视频保存至本地,另外,也可以将上述20s录屏视频自动下载,自动上传至目标观众相应的主页面等。

[0102] 通过在播放第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内上述第二终端的直播视频中的主播的语音数据,根据上述目标观众的用户身份标识和上述主播的语音数据,进行屏幕录制,从而实现了在第一终端自动录制直播过程中的精彩内容,提高观众互动体验。

[0103] 在一种可能的实施方式中,上述根据上述目标观众的用户身份标识和上述主播的语音数据,判断上述主播的语音数据是否与上述目标观众有关,包括:

[0104] 步骤一,分别对上述目标观众的用户身份标识和上述主播的语音数据进行特征提取,得到目标观众用户身份标识的语音特征和主播的语音特征;

[0105] 步骤二,根据上述目标观众用户身份标识的语音特征和上述主播的语音特征,判断上述主播的语音数据是否与上述目标观众有关。

[0106] 基于预设转换模型将目标观众的用户身份标识转换为语音特征,得到目标观众用户身份标识的语音特征,根据目标观众用户身份标识的语音特征的类型,对主播的语音数据进行特征提取,得到主播的语音特征。例如,目标观众用户身份标识的语音特征为音频特

征,则将主播的语音数据进行特征提取,得到相应的主播的语音数据的音频特征,例如,目标观众用户身份标识的语音特征为文字拼音文本特征,则将主播的语音数据转换为相应的拼音文本。

[0107] 根据预设算法将上述目标观众用户身份标识的语音特征和上述主播的语音特征进行语音相似度计算,得到目标语音相似度,具体的,根据动态时间规整算法将上述目标观众用户身份标识的语音特征和上述主播的语音特征进行语音相似度计算,得到目标语音相似度,上述目标语音相似度超过预设阈值时,判断上述主播的语音数据与上述目标观众有关,当目标语音相似度未超过预设阈值时,判断上述主播的语音数据与上述目标观众无关。例如,其目标语音相似度为0.7,超过预设阈值0.5,则,判断上述主播的语音数据与上述目标观众有关。

[0108] 在一种可能的实施方式中,上述分别对上述目标观众的用户身份标识和上述主播的语音数据进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征,包括:

[0109] 基于MFCC方法对上述主播的语音数据进行特征提取,得到主播的语音数据特征;

[0110] 分别对上述目标观众的用户身份标识和上述主播的语音数据特征进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征。

[0111] 基于MFCC方法对上述主播的语音数据进行特征提取,得到主播的语音数据,从而根据主播的语音数据判断在目标观众A给主播发送了超过预设等级的礼物后,主播是否对目标观众A进行口播。

[0112] 基于预设转换模型将目标观众的用户身份标识转换为语音特征,得到目标观众用户身份标识的语音特征,根据目标观众用户身份标识的语音特征的类型,对主播的语音数据进行特征提取,得到主播的语音特征。例如,目标观众用户身份标识的语音特征为音频特征,则将主播的语音数据进行特征提取,得到相应的主播的语音数据的音频特征,例如,目标观众用户身份标识的语音特征为文字拼音文本特征,则将主播的语音数据转换为相应的拼音文本。

[0113] 在一种可能的实施方式中,上述根据上述目标观众的用户身份标识和上述主播的语音数据,判断上述主播的语音数据是否与上述目标观众有关,包括:

[0114] 根据预设算法将上述目标观众用户身份标识的语音特征和上述主播的语音特征进行语音相似度计算,得到目标语音相似度;

[0115] 当上述目标语音相似度超过预设阈值时,确定上述主播的语音数据与上述目标观众有关。

[0116] 根据预设算法将上述目标观众用户身份标识的语音特征和上述主播的语音特征进行语音相似度计算,得到目标语音相似度,具体的,根据动态时间规整算法将上述目标观众用户身份标识的语音特征和上述主播的语音特征进行语音相似度计算,得到目标语音相似度,上述目标语音相似度超过预设阈值时,判断上述主播的语音数据与上述目标观众有关,当目标语音相似度未超过预设阈值时,判断上述主播的语音数据与上述目标观众无关。例如,其目标语音相似度为0.7,超过预设阈值0.5,则,判断上述主播的语音数据与上述目标观众有关。

[0117] 在一种可能的实施方式中,上述预设算法为动态时间规整算法。

[0118] 动态时间规整算法可以克服发同一音而发音时间长短的不同,采用对输入语音信号进行伸长或缩短直到与标准模式的长度一致,根据动态时间规整算法可以计算语音相似度计算,从而得到目标语音相似度,提高计算准确率。

[0119] 在一种可能的实施方式中,上述启动屏幕录制,包括:

[0120] 从当前时刻之前的第一预设时长的时刻开始,启动屏幕录制;或者

[0121] 从当前时刻开始,启动屏幕录制;或者

[0122] 从当前时刻之后的第二预设时长的时刻开始,启动屏幕录制。

[0123] 启动屏幕录制可以从当前时刻之前的第一预设时长的时刻开始,也可以从当前时刻开始,也可以从当前时刻之后的第二预设时长的时刻开始。例如,当上述主播的语音数据与上述目标观众有关时,从当前时刻之前的第一预设时长的时刻开始,启动屏幕录制,第一预设时长可以按照实际情况进行设定,例如设置为10秒、15秒或20秒等。根据需要,在观看主播直播过程中,为捕捉精彩瞬间,可缓存当下直播时刻前一段时间的视频,例如,在观看主播直播过程中,始终缓存录制当前直播时刻前20s的缓存视频,例如,观看了19s,则缓存录制此19s的视频,观看了60s,则缓存录制40s至60s的直播视频。当上述主播的语音数据与上述目标观众有关时,根据上述主播的语音数据与上述目标观众有关时的时刻,从上述第一终端的缓存中获取上述时刻前第一预设时长的缓存视频,比如获取上述时刻前20s的缓存视频,即从上述时刻前20s开始,启动屏幕录制。

[0124] 或者,当上述主播的语音数据与上述目标观众有关时,从当前时刻开始,启动屏幕录制。或者,当上述主播的语音数据与上述目标观众有关时,从当前时刻之后的第二预设时长的时刻开始,启动屏幕录制。

[0125] 在一种可能的实施方式中,在上述启动屏幕录制屏幕的步骤之后,上述方法还包括:

[0126] 在屏幕录制预设时长后,结束屏幕录制;或者

[0127] 当监测到上述第一终端发生第二预设事件时,结束屏幕录制。

[0128] 通过在播放第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内上述第二终端的直播视频中的主播的语音数据,根据上述目标观众的用户身份标识和上述主播的语音数据,进行屏幕录制,从而实现了在第一终端自动录制直播过程中的精彩内容,提高观众互动体验。

[0129] 在一种可能的实施方式中,上述当监测到上述第一终端发生第二预设事件时,结束屏幕录制,包括:

[0130] 当监测到上述第一终端的屏幕展示的预设特效的次数达到预设次数阈值时,结束屏幕录制;或者

[0131] 当监测到上述第一终端接收到预设音频信号时,结束屏幕录制。

[0132] 当启动屏幕录制后,在屏幕录制预设时长后,结束屏幕录制,即对屏幕进行预设时长的录屏,例如,当上述主播的语音数据与上述目标观众有关时,从当前时刻开始,启动屏幕录制,然后在屏幕录制20s后,结束屏幕录制,得到20s的录屏视频,其中预设时长是根据需要设置的。

[0133] 另外,当监测到第一终端发生第二预设事件时也可以结束屏幕录制,例如,当监测到上述第一终端的屏幕展示的预设特效的次数达到预设次数阈值时,结束屏幕录制,例如,

一个礼物特效的时间为5s,则监测到5个礼物特效时,结束屏幕录制。或者,当监测到上述第一终端接收到预设音频信号时,结束屏幕录制,例如,从当前时刻开始,启动屏幕录制,则第一终端接收到预设音频信号中包括主播对当前观众的致谢词,对当前观众的口播,当监测到主播提及非当前观众的名字时,结束屏幕录制。上述第二预设事件可根据实际需要进行设定。

[0134] 在一种可能的实施方式中,上述当上述主播的语音数据与上述目标观众有关时,存储上述录屏视频,包括:

[0135] 在上述第一终端展示预设下载选项;

[0136] 在获取上述目标观众针对上述预设下载选项的指定操作时,将录屏视频保存到本地。

[0137] 具体的,当上述主播的语音数据与上述目标观众有关时,在第一终端屏幕生成下载至本地按钮,观众通过操作上述按钮,在获取上述目标观众针对上述预设下载选项的指定操作时,将上述录屏视频保存至本地。

[0138] 在一种可能的实施方式中,当监测到发生预设事件时,进行预设时长的录屏,得到录屏视频,例如,进行20s的录屏,得到20s录屏视频。然后获取上述录屏视频中主播的语音数据,具体的,可以基于MFCC梅尔倒谱系数方法对上述录屏视频中主播的语音数据进行特征提取,得到主播的语音数据,例如,基于梅尔倒谱系数MFCC方法从上述20s录屏视频中获取主播的语音数据,从而根据主播的语音数据判断在目标观众A给主播发送了超过预设等级的礼物后,主播是否对目标观众A进行口播。根据上述目标观众的用户身份标识和上述录屏视频中主播的语音数据,判断上述主播的语音数据是否与上述目标观众有关。例如,目标观众的用户身份标识为A,基于预设转换模型将目标观众的用户身份标识转换为语音特征,得到目标观众用户身份标识的语音特征,根据目标观众用户身份标识的语音特征的类型,对主播的语音数据进行特征提取,得到主播的语音特征。例如,目标观众用户身份标识的语音特征为音频特征,则将主播的语音数据进行特征提取,得到相应的主播的语音数据的音频特征,例如,目标观众用户身份标识的语音特征为文字拼音文本特征,则将主播的语音数据转换为相应的拼音文本。当上述主播的语音数据与上述目标观众有关时,存储上述录屏视频,当上述主播的语音数据与上述目标观众无关时,将上述录屏视频删除,从而节省观众端内存空间。

[0139] 图2根据一示例性实施例示出的一种视频处理装置的框图,应用于第一终端,上述第一终端用于在第二终端直播时,播放上述第二终端的直播视频,参见图2,该装置包括:第一获取模块310,第二获取模块320,录制模块330。

[0140] 第一获取模块310,被配置为获取上述第一终端登录的目标观众的用户身份标识;

[0141] 第二获取模块320,被配置为在播放上述第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内上述第二终端的直播视频中的主播的语音数据;

[0142] 录制模块330,被配置为根据上述目标观众的用户身份标识和上述主播的语音数据,进行屏幕录制。

[0143] 在一种可能的实施方式中,上述预设事件包括:上述目标观众发送给上述主播的礼物金额超过预设金额;或者

- [0144] 上述目标观众发送给上述主播的礼物数量超过预设数量;或者
- [0145] 上述目标观众发送给上述主播的礼物等级超过预设等级。
- [0146] 在一种可能的实施方式中,上述录制模块330具体用于:
- [0147] 根据上述目标观众的用户身份标识和上述主播的语音数据,判断上述主播的语音数据是否与上述目标观众有关;
- [0148] 当上述主播的语音数据与上述目标观众有关时,启动屏幕录制。
- [0149] 在一种可能的实施方式中,上述录制模块330具体用于:
- [0150] 分别对上述目标观众的用户身份标识和上述主播的语音数据进行特征提取,得到目标观众用户身份标识的语音特征和主播的语音特征;
- [0151] 根据上述目标观众用户身份标识的语音特征和上述主播的语音特征,判断上述主播的语音数据是否与上述目标观众有关。
- [0152] 在一种可能的实施方式中,上述录制模块330具体用于:
- [0153] 基于梅尔倒谱系数MFCC装置对上述主播的语音数据进行特征提取,得到主播的语音数据特征;
- [0154] 分别对上述目标观众的用户身份标识和上述主播的语音数据特征进行特征提取,分别得到目标观众用户身份标识的语音特征和主播的语音特征。
- [0155] 在一种可能的实施方式中,上述录制模块330具体用于:
- [0156] 根据预设算法将上述目标观众用户身份标识的语音特征和上述主播的语音特征进行语音相似度计算,得到目标语音相似度;
- [0157] 当上述目标语音相似度超过预设阈值时,确定上述主播的语音数据与上述目标观众有关。
- [0158] 在一种可能的实施方式中,上述录制模块330具体用于:
- [0159] 从当前时刻之前的第一预设时长的时刻开始,启动屏幕录制;或者
- [0160] 从当前时刻开始,启动屏幕录制;或者
- [0161] 从当前时刻之后的第二预设时长的时刻开始,启动屏幕录制。
- [0162] 在一种可能的实施方式中,上述装置还包括结束模块,被配置为:
- [0163] 在屏幕录制预设时长后,结束屏幕录制;或者
- [0164] 当监测到上述第一终端发生第二预设事件时,结束屏幕录制。
- [0165] 在一种可能的实施方式中,上述结束模块具体用于:
- [0166] 当监测到上述第一终端的屏幕展示的预设特效的次数达到预设次数阈值时,结束屏幕录制;或者
- [0167] 当监测到上述第一终端接收到预设音频信号时,结束屏幕录制。
- [0168] 在一种可能的实施方式中,上述装置还包括:
- [0169] 展示模块,被配置为在上述第一终端展示预设下载选项;
- [0170] 存储模块,被配置为在获取上述目标观众针对上述预设下载选项的指定操作时,将录屏视频保存到本地。
- [0171] 关于上述实施例中的装置,其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。
- [0172] 本公开实施例还提供了一种终端设备,包括:处理器、通信接口、存储器和通信总

线,其中,处理器,通信接口,存储器通过通信总线完成相互间的通信,上述存储器,用于存放计算机程序;

[0173] 上述处理器,用于执行上述存储器存放的计算机程序时,实现如下步骤:

[0174] 获取上述第一终端登录的目标观众的用户身份标识;

[0175] 在播放上述第二终端的直播视频的过程中,当监测到发生预设事件时,获取当前时刻之后预设时间段内上述第二终端的直播视频中的主播的语音数据;

[0176] 根据上述目标观众的用户身份标识和上述主播的语音数据,进行屏幕录制。可选的,处理器,用于执行存储器上所存放的程序时,还可以实现上述任一应用于第一终端的视频处理方法。

[0177] 上述终端设备提到的通信总线可以是外设部件互连标准 (Peripheral Component Interconnect, PCI) 总线或扩展工业标准结构 (Extended Industry Standard Architecture, EISA) 总线等。该通信总线可以分为地址总线、数据总线、控制总线等。为便于表示,图中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0178] 通信接口用于上述终端设备与其他设备之间的通信。

[0179] 存储器可以包括随机存取存储器 (Random Access Memory, RAM),也可以包括非易失性存储器 (Non-Volatile Memory, NVM),例如至少一个磁盘存储器。可选的,存储器还可以是至少一个位于远离前述处理器的存储装置。

[0180] 上述的处理器可以是通用处理器,包括中央处理器 (Central Processing Unit, CPU)、网络处理器 (Network Processor, NP) 等;还可以是数字信号处理器 (Digital Signal Processing, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现场可编程门阵列 (Field-Programmable Gate Array, 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。

[0181] 图3为根据一示例性实施例示出的第一种终端设备的框图,参见图3,例如,终端设备800可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,医疗设备,健身设备,个人数字助理等。

[0182] 参照图3,终端设备800可以包括以下一个或多个组件:处理组件802,存储器804,电源组件806,多媒体组件808,音频组件810,输入/输出 (I/O) 的接口812,传感器组件814,以及通信组件816。

[0183] 处理组件802通常控制终端设备800的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理组件802可以包括一个或多个处理器820来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件802可以包括一个或多个模块,便于处理组件802和其他组件之间的交互。例如,处理组件802可以包括多媒体模块,以方便多媒体组件808和处理组件802之间的交互。

[0184] 存储器804被配置为存储各种类型的数据以支持在终端设备800的操作。这些数据的示例包括用于在终端设备800上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器804可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器 (SRAM),电可擦除可编程只读存储器 (EEPROM),可擦除可编程只读存储器 (EPROM),可编程只读存储器 (PROM),只读存储器 (ROM),磁存储器,快闪存储器,磁盘或光盘。

[0185] 电源组件806为终端设备800的各种组件提供电力。电源组件806可以包括电源管理系统,一个或多个电源,及其他与为终端设备800生成、管理和分配电力相关联的组件。

[0186] 多媒体组件808包括在上述终端设备800和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。上述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与上述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件808包括一个前置摄像头和/或后置摄像头。当终端设备800处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0187] 音频组件810被配置为输出和/或输入音频信号。例如,音频组件810包括一个麦克风(MIC),当终端设备800处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器804或经由通信组件816发送。在一些实施例中,音频组件810还包括一个扬声器,用于输出音频信号。

[0188] I/O接口812为处理组件802和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0189] 传感器组件814包括一个或多个传感器,用于为终端设备800提供各个方面的状态评估。例如,传感器组件814可以检测到终端设备800的打开/关闭状态,组件的相对定位,例如上述组件为终端设备800的显示器和小键盘,传感器组件814还可以检测终端设备800或终端设备800一个组件的位置改变,用户与终端设备800接触的存在或不存在,终端设备800方位或加速/减速和终端设备800的温度变化。传感器组件814可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件814还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件814还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0190] 通信组件816被配置为便于终端设备800和其他设备之间有线或无线方式的通信。终端设备800可以接入基于通信标准的无线网络,如WiFi,运营商网络(如2G、3G、4G或5G),或它们的组合。在一个示例性实施例中,通信组件816经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,上述通信组件816还包括近场通信(NFC)模块,以促进短程通信。例如,在NFC模块可基于射频识别(RFID)技术,红外数据协会(IrDA)技术,超宽带(UWB)技术,蓝牙(BT)技术和其他技术来实现。

[0191] 在示例性实施例中,终端设备800可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述实施例中任一上述的视频处理方法。

[0192] 图4是根据一示例性实施例示出的第二种终端设备的框图。例如,终端设备900可以被提供为一服务器。参照图4,终端设备900包括处理组件922,其进一步包括一个或多个处理器,以及由存储器932所代表的存储器资源,用于存储可由处理组件922的执行的指令,例如应用程序。存储器932中存储的应用程序可以包括一个或一个以上的每一个对应于一

组指令的模块。此外,处理组件922被配置为执行指令,以执行上述实施例中任一上述的视频处理方法。

[0193] 终端设备900还可以包括一个电源组件926被配置为执行终端设备900的电源管理,一个有线或无线网络接口950被配置为将终端设备900连接到网络,和一个输入输出(I/O)接口958。终端设备900可以操作基于存储在存储器932的操作系统,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™或类似操作系统。

[0194] 在本公开实施例中,还提供了一种存储介质,该存储介质中存储有指令,当其在计算机上运行时,使得计算机执行上述实施例中任一上述的应用于第一终端的视频处理方法。在示例性实施例中,还提供了一种包括指令的存储介质,例如包括指令的存储器804,上述指令可由终端设备800的处理器820执行以完成上述方法。可选地,例如,存储介质可以是非临时性计算机可读存储介质,例如,上述存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0195] 在本公开实施例中,还提供了一种包含指令的计算机程序产品,当其在计算机上运行时,使得计算机执行上述实施例中任一上述的应用于第一终端的视频处理方法。

[0196] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其他实施方案。本公开旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由下面的权利要求指出。

[0197] 应当理解的是,本公开并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限制。

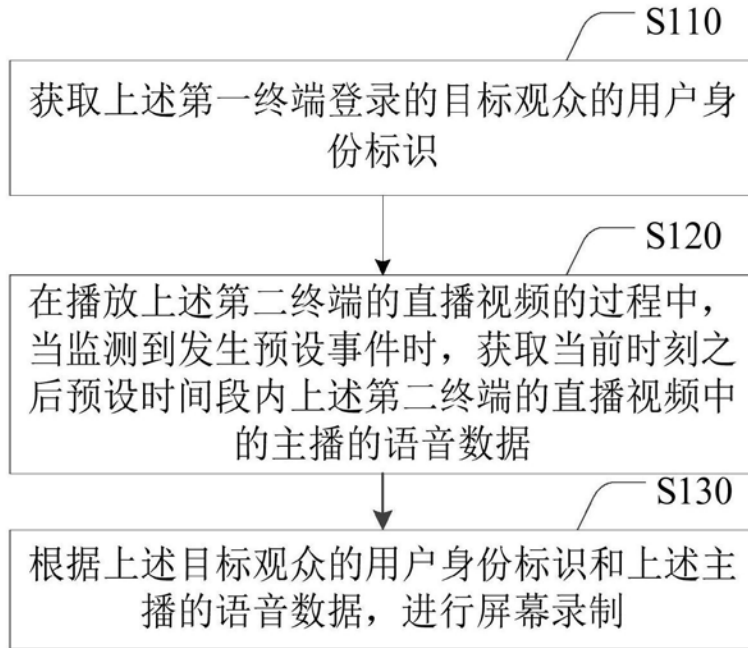


图1

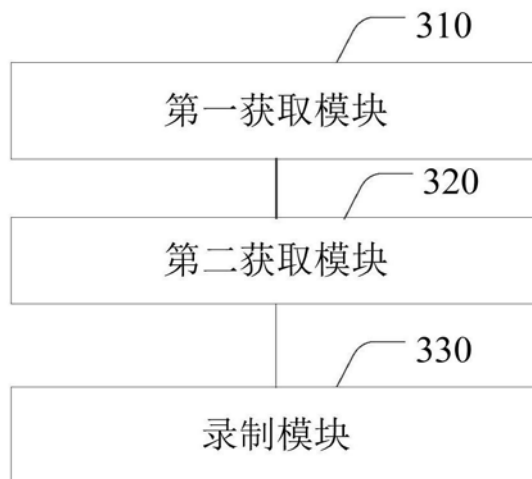


图2

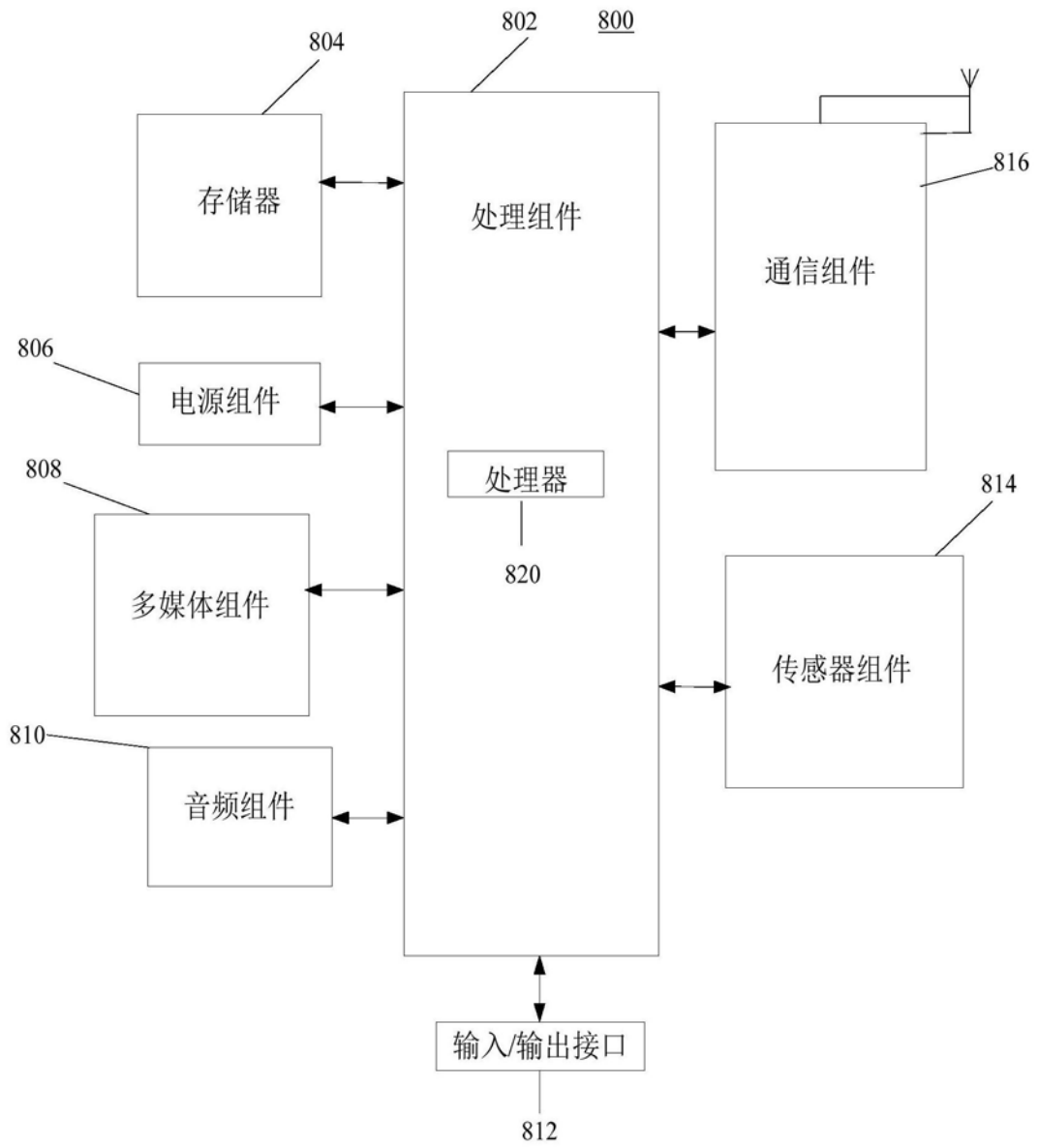


图3

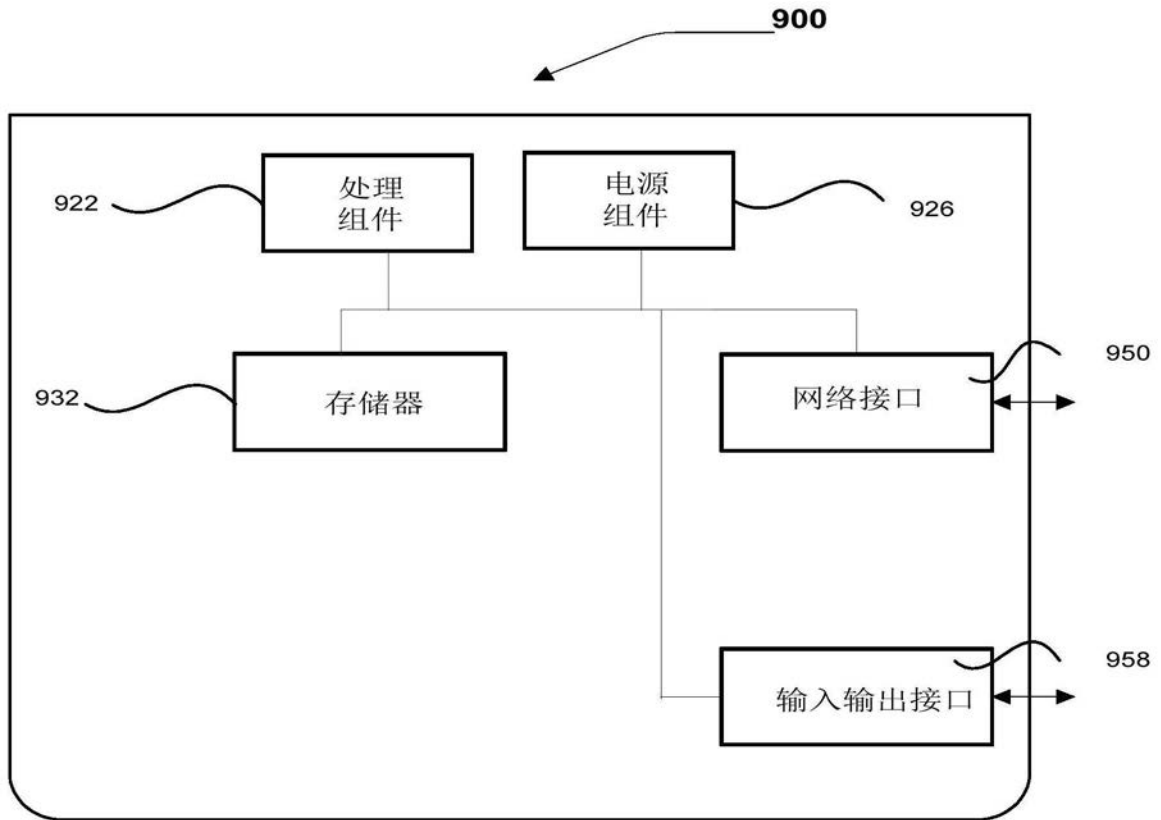


图4