(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107623830 A
(43)申请公布日 2018.01.23

(21)申请号 201610559604.9

(22)申请日 2016.07.15

(71)申请人 掌赢信息科技(上海)有限公司
地址 200063 上海市普陀区谈家渡路28号515室

(72)发明人 武俊敏

(51)Int.Cl.
H04N 7/14(2006.01)

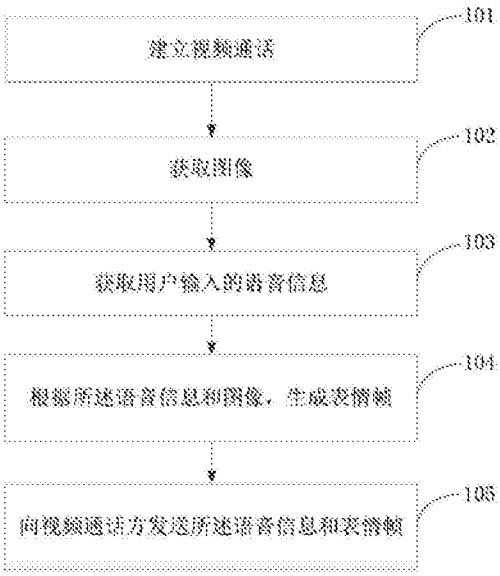
权利要求书1页 说明书9页 附图2页

(54)发明名称

一种视频通话方法及电子设备

(57)摘要

本发明实施例提供了一种视频通话方法及电子设备。该方法包括:建立视频通话;获取图像;获取用户输入的语音信息;根据所述语音信息和图像,生成表情帧;以及向视频通话方发送所述语音信息和表情帧。相比于纯粹的视频通话,本发明实施例节省了网络带宽,并且更具有趣味性和视觉效果,从而提高用户体验。



CN 107623830 A

1.一种视频通话方法,其特征在于,所述方法包括:

建立视频通话;

获取图像;

获取用户输入的语音信息;

根据所述语音信息和图像,生成表情帧;以及

向视频通话方发送所述语音信息和表情帧。

2.根据权利要求1所述的方法,其特征在于,所述获取图像包括:

获取预设图像或者用户输入的图像。

3.根据权利要求1所述的方法,其特征在于,所述语音信息的长度为预设帧率的倒数。

4.根据权利要求1所述的方法,其特征在于,所述根据语音信息和图像,生成表情帧包括:

获取所述语音信息的特征;

根据所述语音信息的特征选取对应的嘴部表情;以及

根据所述语音信息对应的嘴部表情和所述图像生成所述表情帧。

5.根据权利要求4所述的方法,其特征在于,所述根据所述语音信息的特征选取对应的嘴部表情包括:

根据所述语音信息的特征与预设的模型,在预设的表情库中选择与所述特征对应的嘴部表情。

6.一种电子设备,其特征在于,所述设备包括:

视频通话建立模块,用于建立视频通话;

图像获取模块,用于获取图像;

语音获取模块,用于获取用户输入的语音信息;

表情帧生成模块,用于根据所述语音信息和所述图像,生成表情帧;以及

发送模块,用于向视频通话方发送所述语音信息和表情帧。

7.根据权利要求6所述的电子设备,其特征在于,所述图像获取模块具体用于:

获取预设图像或者用户输入的图像。

8.根据权利要求6所述的电子设备,其特征在于,所述语音信息的长度为预设帧率的倒数。

9.根据权利要求6所述的电子设备,其特征在于,所述表情帧生成模块包括:

特征获取子模块,用于获取所述语音信息的特征;

嘴部表情选择子模块,用于根据所述语音信息的特征选取对应的嘴部表情;

表情帧生成子模块,用于根据所述语音信息对应的嘴部表情和所述图像生成所述表情帧。

10.根据权利要求9所述的电子设备,其特征在于,所述嘴部表情选择子模块具体用于:

根据所述语音信息的特征与预设的模型,在预设的表情库中选择与所述特征对应的嘴部表情。

一种视频通话方法及电子设备

技术领域

[0001] 本发明涉及通信领域，具体涉及一种视频通话方法及电子设备。

背景技术

[0002] 随着互联网的发展，人们越来越多地使用视频来进行通信。但是，目前的视频通信比较耗费网络带宽，并且比较单调，缺少趣味性。

发明内容

[0003] 本发明实施例提供了一种视频通话方法及电子设备，以便减少网络带宽消耗，增加趣味性，并且提高用户体验。

[0004] 根据本发明的第一方面，提供了一种视频通话方法，该方法包括：

[0005] 建立视频通话；

[0006] 获取图像；

[0007] 获取用户输入的语音信息；

[0008] 根据所述语音信息和图像，生成表情帧；以及

[0009] 向视频通话方发送所述语音信息和表情帧。

[0010] 结合本发明的第一方面，在第一种可能的实现方式中，所述获取图像包括：

[0011] 获取预设图像或者用户输入的图像。

[0012] 结合本发明的第一方面，在第二种可能的实现方式中，所述语音信息的长度为预设帧率的倒数。

[0013] 结合本发明的第一方面，在第三种可能的实现方式中，所述根据语音信息和图像，生成表情帧包括：

[0014] 获取所述语音信息的特征；

[0015] 根据所述语音信息的特征选取对应的嘴部表情；以及

[0016] 根据所述语音信息对应的嘴部表情和所述图像生成所述表情帧。

[0017] 结合本发明的第一方面的第三种可能的实现方式，在第四种可能的实现方式中，所述根据所述语音信息的特征选取对应的嘴部表情包括：

[0018] 根据所述语音信息的特征与预设的模型，在预设的表情库中选择与所述特征对应的嘴部表情。

[0019] 结合本发明的第一方面的第三种可能的实现方式，在第五种可能的实现方式中，所述根据所述语音信息的特征选取对应的嘴部表情包括：

[0020] 根据所述语音信息的特征，预设的模型以及该语音信息的上一语音信息对应的嘴部表情，在预设的表情库中选择与所述特征对应的嘴部表情。

[0021] 结合本发明的第一方面的第三种可能的实现方式，在第六种可能的实现方式中，所述根据所述语音信息对应的嘴部表情和所述图像生成对应的表情帧包括：

[0022] 将所述语音信息对应的嘴部表情和所述图像组合，生成对应的表情帧。

[0023] 根据本发明的第二方面，提供了一种电子设备，所述设备包括：

[0024] 视频通话建立模块，用于建立视频通话；

[0025] 图像获取模块，用于获取图像；

[0026] 语音获取模块，用于获取用户输入的语音信息；

[0027] 表情帧生成模块，用于根据所述语音信息和所述图像，生成表情帧；以及

[0028] 发送模块，用于向视频通话方发送所述语音信息和表情帧。

[0029] 结合本发明的第二方面，在第一种可能的实现方式中，所述图像获取模块具体用于：

[0030] 获取预设图像或者用户输入的图像。

[0031] 结合本发明的第二方面，在第二种可能的实现方式中，所述语音信息的长度为预设帧率的倒数。

[0032] 结合本发明的第二方面，在第三种可能的实现方式中，所述表情帧生成模块包括：

[0033] 特征获取子模块，用于获取所述语音信息的特征；

[0034] 嘴部表情选择子模块，用于根据所述语音信息的特征选取对应的嘴部表情；

[0035] 表情帧生成子模块，用于根据所述语音信息对应的嘴部表情和所述图像生成所述表情帧。

[0036] 结合本发明的第二方面的第三种可能的实现方式，在第四种可能的实现方式中，所述嘴部表情选择子模块具体用于：

[0037] 根据所述语音信息的特征与预设的模型，在预设的表情库中选择与所述特征对应的嘴部表情。

[0038] 结合本发明的第二方面的第三种可能的实现方式，在第五种可能的实现方式中，所述嘴部表情选择子模块具体用于：

[0039] 根据所述语音信息的特征，预设的模型以及该语音信息的上一语音信息对应的嘴部表情，在预设的表情库中选择与所述特征对应的嘴部表情。

[0040] 结合本发明的第二方面的第三种可能的实现方式，在第六种可能的实现方式中，所述表情帧生成子模块具体用于：

[0041] 将所述语音信息对应的嘴部表情和所述图像组合，生成对应的表情帧。

[0042] 根据本发明的第三方面，提供了一种电子设备，该电子设备包括：

[0043] 存储器、音频获取模块、网络接口模块以及与存储器、音频获取模块、网络接口模块连接的处理器，其中，存储器用于存储一组程序代码，处理器调用存储器所存储的程序代码用于执行以下操作：

[0044] 建立视频通话；

[0045] 获取图像；

[0046] 获取用户输入的语音信息；

[0047] 根据所述语音信息和图像，生成表情帧；以及

[0048] 向视频通话方发送所述语音信息和表情帧。

[0049] 结合本发明的第三方面，在第一种可能的实现方式中，处理器调用存储器所存储的程序代码用于执行以下操作：

[0050] 获取预设图像或者用户输入的图像。

[0051] 结合本发明的第三方面，在第二种可能的实现方式中，所述语音信息的长度为预设帧率的倒数。

[0052] 结合本发明的第三方面，在第三种可能的实现方式中，处理器调用存储器所存储的程序代码用于执行以下操作：

[0053] 获取所述语音信息的特征；

[0054] 根据所述语音信息的特征选取对应的嘴部表情；以及

[0055] 根据所述语音信息对应的嘴部表情和所述图像生成所述表情帧。

[0056] 结合本发明的第三方面的第三种可能的实现方式，在第四种可能的实现方式中，处理器调用存储器所存储的程序代码用于执行以下操作：

[0057] 根据所述语音信息的特征与预设的模型，在预设的表情库中选择与所述特征对应的嘴部表情。

[0058] 结合本发明的第三方面的第三种可能的实现方式，在第五种可能的实现方式中，处理器调用存储器所存储的程序代码用于执行以下操作：

[0059] 根据所述语音信息的特征，预设的模型以及该语音信息的上一语音信息对应的嘴部表情，在预设的表情库中选择与所述特征对应的嘴部表情。

[0060] 结合本发明的第三方面的第三种可能的实现方式，在第六种可能的实现方式中，处理器调用存储器所存储的程序代码用于执行以下操作：

[0061] 将所述语音信息对应的嘴部表情和所述图像组合，生成所述表情帧。

[0062] 本发明实施例提供了一种视频通话方法及电子设备。通过在视频通话过程中根据用户输入的语音信息，对图像进行动画化，以生成表情帧，并向视频通话方发送语音信息和表情帧，可以使得视频通话方能够观看表情帧并听到语音信息。相比于纯粹的视频通话，这种方式节省了网络带宽，并且更具有趣味性和视觉效果，从而提高用户体验。另外根据语音对图像进行动画化生成表情帧的方法能够实时的通过语音来生成对应的表情帧，无需获取面部信息，具有效率高、速度快、限制少、资源消耗少的优点。

附图说明

[0063] 为了更清楚地说明本发明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发明的一些实施例，对于本领域普通技术人员来讲，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。

[0064] 图1示出了根据本发明实施例的一种视频通话方法的流程图；

[0065] 图2示出了根据本发明实施例的根据所述语音信息和图像生成表情帧步骤的流程图；

[0066] 图3示出了根据本发明实施例的一种电子设备的框图；

[0067] 图4示出了根据本发明实施例的一种电子设备的框图。

具体实施方式

[0068] 为使本发明的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本

发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0069] 本发明实施例提供了一种视频通话方法及电子设备。通过在视频通话过程中根据用户输入的语音信息，对图像进行动画化，以生成表情帧，并向视频通话方发送语音信息和表情帧，可以使得视频通话方能够观看表情帧并听到语音信息。相比于纯粹的视频通话，这种方式节省了网络带宽，并且更具有趣味性和视觉效果，从而提高用户体验。另外根据语音对图像进行动画化生成表情帧的方法能够实时的通过语音来生成对应的表情帧，无需获取面部信息，具有效率高、速度快、限制少、资源消耗少的优点。

[0070] 图1示出了根据本发明实施例的一种视频通话方法的流程图。该方法可在终端中执行。终端可包括但不限于，移动电话、膝上型或笔记本计算机、台式计算机、个人数字助理(PDA)、游戏控制终端。如图1所示，该方法可包括以下步骤：

[0071] 步骤101：建立视频通话。

[0072] 视频通话可涉及两方或多方。可通过常规视频通话建立方法建立视频通话，例如SIP信令，本发明实施例对视频通话建立方式不加以限定。本发明实施例默认视频通话接收方接受视频通话建立请求。

[0073] 步骤102：获取图像。

[0074] 在一个实施例中，获取图像包括获取预设图像。在一个例子中，预设图像是默认的图像，每次均默认获取该图像进行操作。在另一个例子中，预设图像是用户上次输入的图像，这样每次用户输入新的图像则更新预设图像。

[0075] 在另一个实施例中，获取图像包括获取用户输入的图像。在一个例子中，用户输入的图像是用户从本地图像库中选择的图像，使得用户可以选择任意的本地图像来生成表情帧。在另一个例子中，用户输入的图像是用户通过摄像头获取的照片，使得用户可以选择即时拍摄的照片来生成表情帧。这里的照片可以是自拍照片，也可以是非自拍照片。在还一个例子中，用户输入的图像是用户从显示的一组预设图像中选择的图像。

[0076] 步骤103：获取用户输入的语音信息。

[0077] 具体的，通过终端包括或耦合的音频获取模块(例如麦克风)获取用户即时输入的语音信息。可包括通过麦克风实时获取音频信息，从该音频信息中分离出语音信息。通常来说，人的语音的频率范围在300Hz至4000Hz之间，因此可以通过对音频信息进行滤波，分离出频率范围在300Hz至4000Hz之间的信息作为人的语音信息。可选的，还可以进一步通过声音的强度来分离语音信息，因为人的语音一般在40dB至60dB之间，因此可以根据声音的dB来对音频信息进行过滤，分离出强度在在40dB至60dB之间的音频信息。可选的，还可以对分离出的语音信息进行降噪等处理，得到更加精确的语音信息。

[0078] 具体的，所述语音信息的长度为预设帧率的倒数。示例性的，当预设帧率为30帧/秒时，每个语音片段的长度为1/30秒；当预设帧率为60帧/秒时，每个语音片段的长度为1/60秒。本发明实施例对具体的预设帧率和语音信息的长度不加以限定。

[0079] 步骤104：根据所述语音信息和图像，生成表情帧。

[0080] 具体的，步骤104可包括以下步骤：

[0081] 获取所述语音信息的特征；

[0082] 根据所述语音信息的特征选取对应的嘴部表情；以及

[0083] 根据所述语音信息对应的嘴部表情和所述图像生成所述表情帧。

[0084] 在下文中将参照图2详细描述上述步骤。

[0085] 步骤105:向视频通话方发送所述语音信息和表情帧。

[0086] 视频通话方可以是一个或多个。例如,可以向一个或多个社交应用用户关联的一个或多个终端发送所述语音信息和表情帧。再如,可以向一个或多个社交应用群组关联的多个终端发送所述语音信息和表情帧。发送可包括经过服务器中转的方式或者不经过服务器中转的方式。

[0087] 在一个实施例中,该方法还包括显示表情帧的步骤,使得用户可以直观地感受视频通话中自己的表情帧。

[0088] 本发明实施例提供了一种视频通话方法及电子设备。通过在视频通话过程中根据用户输入的语音信息,对图像进行动画化,以生成表情帧,并向视频通话方发送语音信息和表情帧,可以使得视频通话方能够观看表情帧并听到语音信息。相比于纯粹的视频通话,这种方式节省了网络带宽,并且更具有趣味性和视觉效果,从而提高用户体验。

[0089] 图2示出了根据本发明实施例的根据所述语音信息和图像生成表情帧步骤的流程图。如图2中所示,该步骤包括以下步骤:

[0090] 步骤201:获取所述语音信息的特征。

[0091] 示例性的,该特征可以是MFCC(Mel Frequency Cepstral Coefficents,梅尔频率倒谱系数)特征。本发明实施例对具体的特征不加以限定。

[0092] 步骤202:根据所述语音信息的特征选取对应的嘴部表情。

[0093] 在一个实施例中,该过程可以包括:

[0094] 根据所述特征与预设模型,在预设的表情库中选择与所述特征对应的嘴部表情。

[0095] 预设模型可以是通过有监督学习训练得到的。训练可采用以下方案中的任一个方案进行:

[0096] 方案1:

[0097] a、收集训练数据。

[0098] 收集大量的包含语音和嘴部的形状对应关系的数据,例如电影、电视片段。

[0099] b、对收集到的数据进行预处理。

[0100] 将收集到的数据中带有人脸嘴部的视频帧挑选出来。

[0101] 将这些视频帧中嘴部的形状和对应的语音信息的MFCC特征提取出来。

[0102] c、根据这些嘴部的形状和对应的MFCC特征对随机森林(Random Forest)进行训练,得到训练后的随机森林作为预设模型。

[0103] 在根据所述特征选取对应的嘴部表情的过程中,将所述特征输入该训练后的随机森林,随机森林将判断该特征对应的嘴部的形状,并从预设的表情库中选取该嘴部的形状对应的嘴部表情作为所述特征对应的嘴部表情。

[0104] 方案2:

[0105] a、收集训练数据。

[0106] 收集大量的包含语音和嘴部开闭状态对应关系的数据,例如电影、电视片段。

[0107] b、对收集到的数据进行预处理。

[0108] 将收集到的数据中带有人脸嘴部的视频帧挑选出来。

[0109] 将这些视频帧中嘴部的开闭状态和对应的语音信息的MFCC特征提取出来。

[0110] c、根据这些嘴部的开闭状态和对应的MFCC特征对SVM(Support Vector Machine，支持向量机)进行训练，得到训练后的SVM作为预设模型。

[0111] 在根据所述特征选取对应的嘴部表情的过程中，将所述特征输入该训练后的SVM，SVM将判断该特征对应的嘴部状态是开还是闭，如果对应的状态是开，则从预设的表情库中选取嘴部状态为开的表情作为所述特征对应的嘴部表情，如果对应的状态是闭，则从预设的表情库中选取嘴部状态为闭的表情作为所述特征对应的嘴部表情。

[0112] 方案3:

[0113] a、收集训练数据。

[0114] 收集大量的包含语音和嘴部的形状对应关系的数据，例如电影、电视片段。

[0115] b、对收集到的数据进行预处理。

[0116] 将收集到的数据中带有人脸嘴部的视频帧挑选出来。

[0117] 将这些视频帧中嘴部的形状对应的人脸的特征点和该嘴部的形状对应的语音信息的MFCC特征提取出来。

[0118] c、根据这些人脸的特征点和对应的MFCC特征对GMM(Gaussian Mixture Model)模型进行训练，得到训练后的GMM模型作为预设模型。

[0119] 在根据所述特征选取对应的嘴部表情的过程中，将所述特征输入该训练后的GMM模型，GMM模型将判断该特征对应的人脸的特征点，并从预设的表情库中选取该人脸的特征点对应的嘴部表情作为所述特征对应的嘴部表情。

[0120] 方案4：

[0121] a、收集训练数据。

[0122] 收集大量的包含语音和嘴部的形状对应关系的数据，例如电影、电视片段。

[0123] b、对收集到的数据进行预处理。

[0124] 将收集到的数据中带有人脸嘴部的视频帧挑选出来。

[0125] 将这些视频帧中嘴部的形状对应的人脸的特征点和该嘴部的形状对应的语音信息的MFCC特征提取出来。

[0126] c、根据这些人脸的特征点和对应的MFCC特征对3层神经网络(Neural Networks)进行训练，得到训练后的3层神经网络作为预设模型。

[0127] 在根据所述特征选取对应的嘴部表情的过程中，将所述特征输入该训练后的3层神经网络，3层神经网络将判断该特征对应的人脸的特征点，并从预设的表情库中选取该人脸的特征点对应的嘴部表情作为所述特征对应的嘴部表情。

[0128] 在另一个实施例中，该过程可以包括：

[0129] 根据所述语音信息的特征，预设的模型以及该语音信息的上一语音信息对应的嘴部表情，在预设的表情库中选择与所述特征对应的嘴部表情。

[0130] 在该实施例中，预设模型的训练方式可以参照上面SVM所述，在此不再赘述。

[0131] 在根据所述特征选取对应的嘴部表情的过程中，将所述特征输入该训练后的SVM，SVM将判断该特征对应的嘴部状态是开的概率，记为p，则该嘴部状态是闭的概率为1-p。

[0132] 如果p超过预设的阈值，则判定对应的嘴部状态是开，否则判定对应的嘴部状态是闭。该阈值的初始值为0.5，并根据当所述特征对应的语音片段的上一语音片段对应的表情

的嘴部状态来对该阈值进行动态的调整。

[0133] 示例性的，当所述特征对应的语音片段的上一语音片段对应的表情的嘴部状态是开时，将该阈值调整为0.3，即所述特征对应的p大于0.3即判定其对应的嘴部状态是开。

[0134] 如果SVM判定该特征对应的状态是开，则从预设的表情库中选取嘴部状态为开的表情作为所述特征对应的表情，如果SVM判定该特征对应的状态是闭，则从预设的表情库中选取嘴部状态为闭的表情作为所述特征对应的表情。

[0135] 步骤204：根据所述语音信息对应的嘴部表情和所述图像生成对应的表情帧。

[0136] 具体的，该过程可以为：

[0137] 识别所述图像中的人脸嘴部区域。

[0138] 示例性的，可以根据主动外观模型(Active Appearance Model)、主动形状模型(Active Shape Model)或者其他方式，从所述图像中获取人脸嘴部区域的特征点。

[0139] 根据所述嘴部表情，驱动所述图像中的人脸嘴部区域。

[0140] 示例性的，计算所述嘴部表情中嘴部特征点与所述图像中的人脸嘴部区域中对应的特征点的位置偏差，根据所述偏差，生成所述图像中的人脸嘴部区域中每个特征点的移动参数，并根据所述移动参数驱动所述图像中的人脸嘴部区域。

[0141] 根据所述图像以及驱动后的所述图像中的人脸嘴部区域生成对应的表情帧。

[0142] 示例性的，以所述驱动后的所述图像中的人脸嘴部区域替换驱动前的所述图像中的人脸嘴部区域，并生成新的图像，根据所述新的图像生成对应的表情帧。

[0143] 上述识别图像中的人脸嘴部区域的过程中，如果未检测到人脸，则以默认的人脸嘴部区域作为生成表情帧的基础。

[0144] 根据所述嘴部表情，驱动所述默认的人脸嘴部区域。

[0145] 示例性的，计算所述嘴部表情中嘴部特征点与所述默认的人脸嘴部区域中对应的特征点的位置偏差，根据所述偏差，生成所述默认的人脸嘴部区域中每个特征点的移动参数，并根据所述移动参数驱动所述默认的人脸嘴部区域。

[0146] 根据所述默认的人脸嘴部区域以及驱动后的所述默认的人脸嘴部区域生成对应的表情帧。

[0147] 示例性的，以所述驱动后的所述默认的人脸嘴部区域替换驱动前的所述默认的人脸嘴部区域，并生成新的图像，根据所述新的图像生成对应的表情帧。

[0148] 本发明实施例能够实时的通过语音来生成对应的表情帧，无需获取面部信息，具有效率高、速度快、限制少、资源消耗少的优点。通过SVM能够快速的对嘴部的开闭状态进行判断，从而有效地提高识别的速度。通过随机森林能够快速的识别出嘴部的形状，从而有效地提高识别的速度。通过SVM能够快速的识别出嘴部的形状，从而有效地提高识别的速度，进一步地根据上一帧的嘴部状态对当前帧的嘴部状态进行判断，有效地提高了识别的准确率。通过GMM模型能够快速的识别出嘴部的形状，从而有效地提高识别的速度。通过神经网络能够快速的识别出嘴部的形状，从而有效地提高识别的速度。

[0149] 图3示出了根据本发明实施例的一种电子设备的框图。如图3所示，该电子设备包括：视频通话建立模块301，用于建立视频通话；图像获取模块302，用于获取图像；语音获取模块303，用于获取用户输入的语音信息；表情帧生成模块304，用于根据所述语音信息和图像，生成表情帧；以及发送模块305，用于向其他电子设备发送所述语音信息和表情帧。

[0150] 具体的，所述图像获取模块302用于获取预设图像或者用户输入的图像。

[0151] 具体的，所述语音信息的长度为预设帧率的倒数。

[0152] 具体的，所述表情帧生成模块304包括：

[0153] 特征获取子模块3041，用于获取所述语音信息的特征；

[0154] 嘴部表情选择子模块3042，用于根据所述语音信息的特征选取对应的嘴部表情；

[0155] 表情帧生成子模块3043，用于根据所述语音信息对应的嘴部表情和所述图像生成表情帧。

[0156] 可选的，所述嘴部表情选择子模块3043用于根据所述语音信息的特征与预设的模型，在预设的表情库中选择与所述特征对应的嘴部表情。

[0157] 可选的，所述嘴部表情选择子模块3043用于根据所述语音信息的特征，预设的模型以及该语音信息的上一语音信息对应的嘴部表情，在预设的表情库中选择与所述特征对应的嘴部表情。

[0158] 具体的，所述表情帧生成子模块3043用于将所述语音信息对应的嘴部表情和所述图像组合，生成表情帧。

[0159] 可选的，该电子设备还包括表情帧显示模块，用于显示表情帧，从而使得用户可以直观地感受视频通话过程中自己的表情帧。

[0160] 本发明实施例提供了一种电子设备。通过在视频通话过程中根据用户输入的语音信息，对图像进行动画化，以生成表情帧，并向视频通话方发送语音信息和表情帧，可以使得视频通话方能够观看表情帧并听到语音信息。相比于纯粹的视频通话，这种方式节省了网络带宽，并且更具有趣味性和视觉效果，从而提高用户体验。

[0161] 图4示出了根据本发明实施例的一种电子设备。如图4所示，该电子设备包括存储器401、音频获取模块402、网络接口模块403以及与存储器401、音频获取模块402、网络接口模块403连接的处理器404，其中，存储器401用于存储一组程序代码，处理器404调用存储器401所存储的程序代码用于执行以下操作：

[0162] 建立视频通话；

[0163] 获取图像；

[0164] 获取用户输入的语音信息；

[0165] 根据所述语音信息和图像，生成表情帧；以及

[0166] 向视频通话方发送所述语音信息和表情帧。

[0167] 具体的，处理器404调用存储器401所存储的程序代码用于执行以下操作：

[0168] 获取预设图像或者用户输入的图像。

[0169] 具体的，所述语音信息的长度为预设帧率的倒数。

[0170] 具体的，处理器404调用存储器401所存储的程序代码用于执行以下操作：

[0171] 获取所述语音信息的特征；

[0172] 根据所述语音信息的特征选取对应的嘴部表情；以及

[0173] 根据所述语音信息对应的嘴部表情和所述图像生成所述表情帧。

[0174] 可选的，处理器404调用存储器401所存储的程序代码用于执行以下操作：

[0175] 根据所述语音信息的特征与预设的模型，在预设的表情库中选择与所述特征对应的嘴部表情。

[0176] 可选的，处理器404调用存储器401所存储的程序代码用于执行以下操作：

[0177] 根据所述语音信息的特征，预设的模型以及该语音信息的上一语音信息对应的嘴部表情，在预设的表情库中选择与所述特征对应的嘴部表情。

[0178] 具体的，处理器404调用存储器401所存储的程序代码用于执行以下操作：

[0179] 将所述语音信息对应的嘴部表情和所述图像组合，生成所述表情帧。

[0180] 本发明实施例提供了一种电子设备。通过在视频通话过程中根据用户输入的语音信息，对图像进行动画化，以生成表情帧，并向视频通话方发送语音信息和表情帧，可以使得视频通话方能够观看表情帧并听到语音信息。相比于纯粹的视频通话，这种方式节省了网络带宽，并且更具有趣味性和视觉效果，从而提高用户体验。

[0181] 上述所有可选技术方案，可以采用任意结合形成本发明的可选实施例，在此不再一一赘述。

[0182] 需要说明的是：上述实施例提供的电子设备在执行视频通话方法时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将设备的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的电子设备与视频通话方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

[0183] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

[0184] 以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

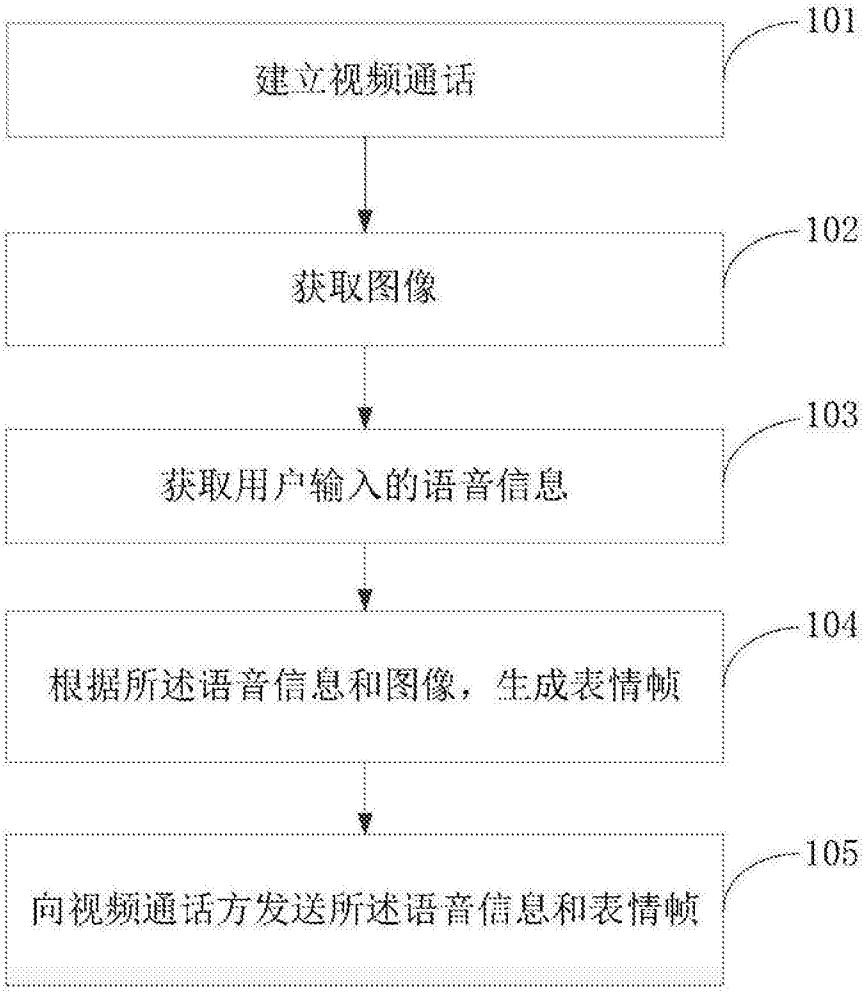

101
建立视频通话
102
获取图像
103
获取用户输入的语音信息
104
根据所述语音信息和图像，生成表情帧
105
向视频通话方发送所述语音信息和表情帧

图1

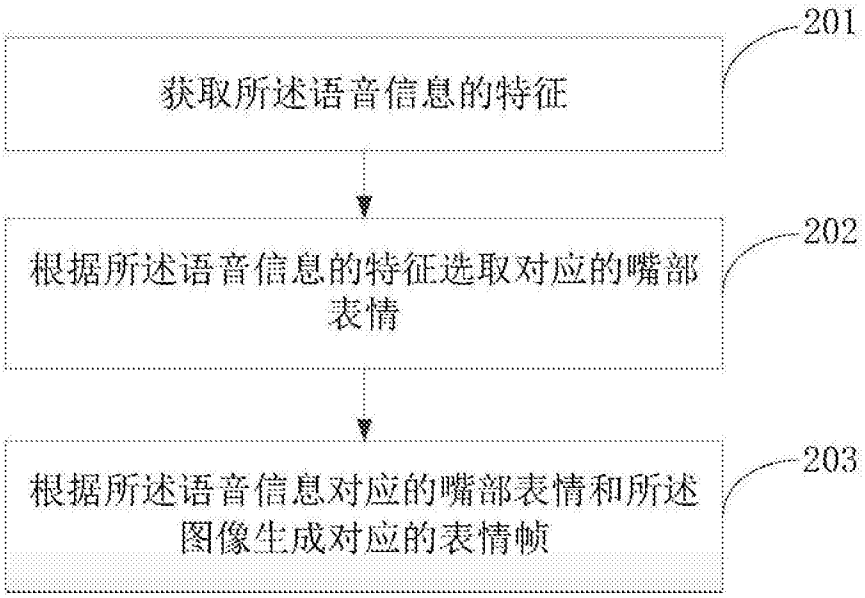

201
获取所述语音信息的特征
202
根据所述语音信息的特征选取对应的嘴部表情
203
根据所述语音信息对应的嘴部表情和所述图像生成对应的表情帧

图2

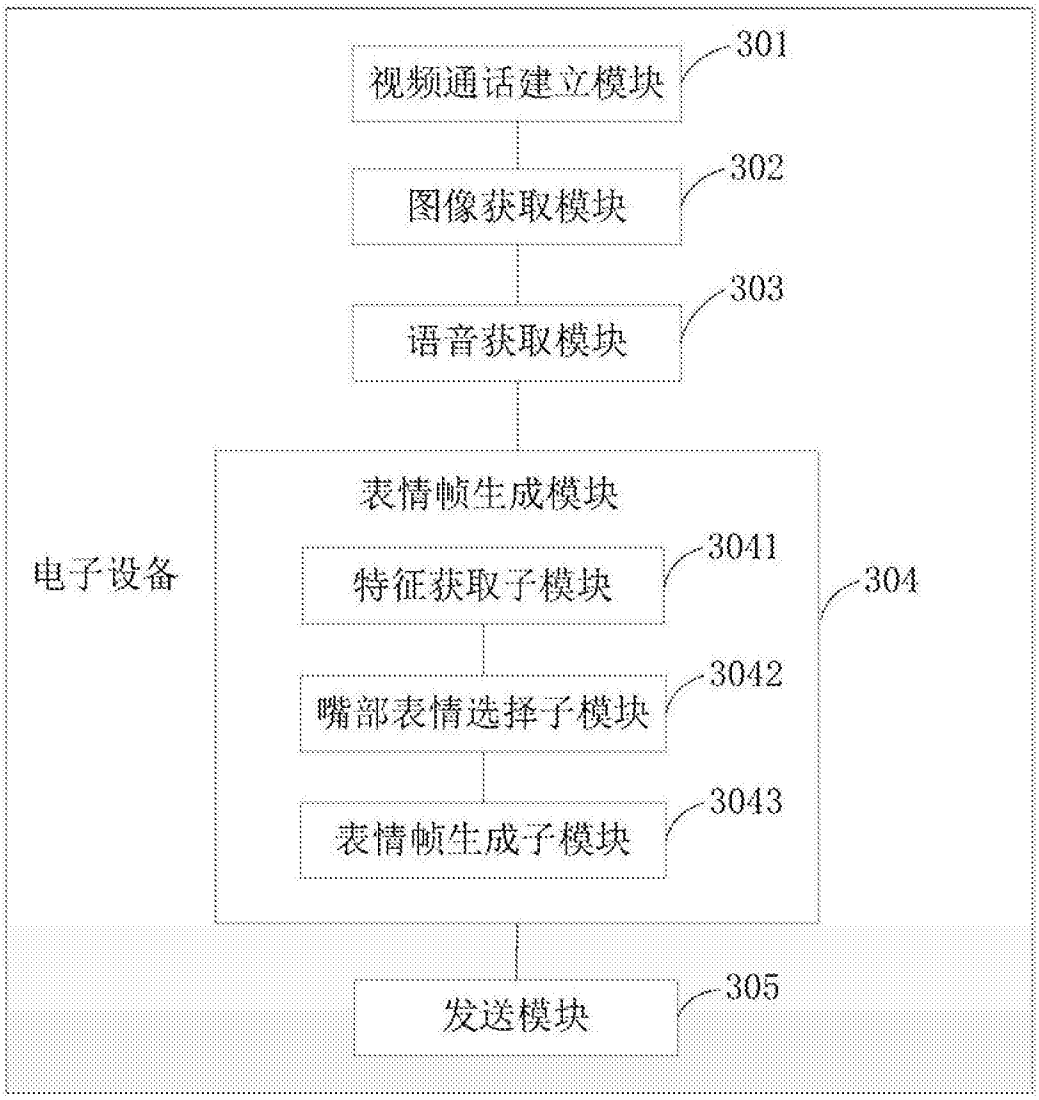

电子设备
视频通话建立模块
301
图像获取模块
302
语音获取模块
303
表情帧生成模块
特征获取子模块
3041
嘴部表情选择子模块
3042
表情帧生成子模块
3043
304
发送模块
305

图3

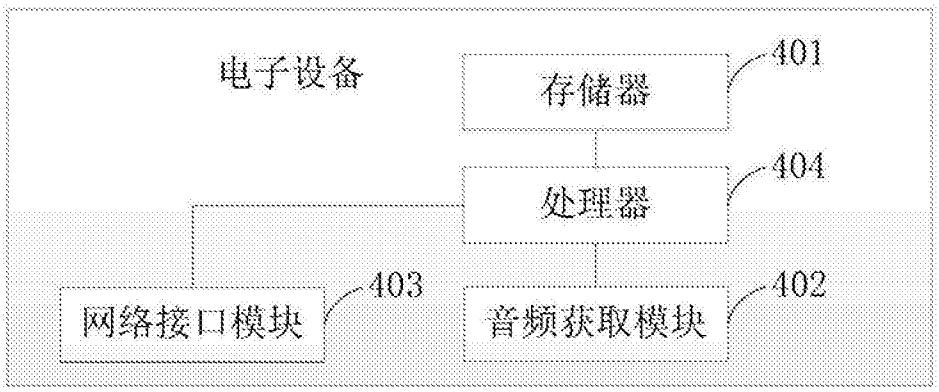

电子设备
存储器
401
处理器
404
网络接口模块
403
音频获取模块
402

图4