(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 111294665 A
(43)申请公布日 2020.06.16

(21)申请号 202010088388.0

(22)申请日 2020.02.12

(71)申请人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦

(72)发明人 刘玉强 鲍冠伯 彭哲

(74)专利代理机构 北京鸿德海业知识产权代理事务所(普通合伙) 11412
代理人 田宏宾

(51)Int.Cl.
H04N 21/81(2011.01)
G06T 17/00(2006.01)

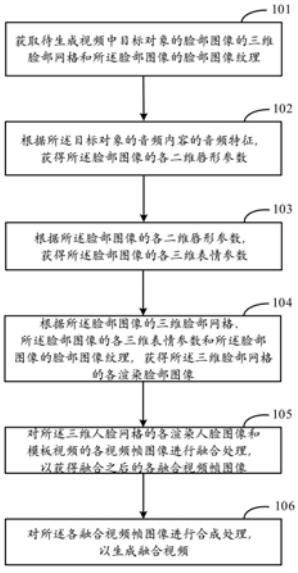
权利要求书4页 说明书17页 附图3页

(54)发明名称

视频的生成方法、装置、电子设备及可读存储介质

(57)摘要

本申请公开了一种视频的生成方法、装置、电子设备及可读存储介质，涉及计算机视觉技术。具体实现方案获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理；根据所述目标对象的音频内容的音频特征，获得所述脸部图像的各二维唇形参数；根据所述脸部图像的各二维唇形参数，获得所述脸部图像的各三维表情参数；根据所述脸部图像的三维脸部网格、所述脸部图像的各三维表情参数和所述脸部图像的脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像；对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理，以获得融合之后的各融合视频帧图像；以及对所述各融合视频帧图像进行合成处理，以生成融合视频。



CN 111294665 A

1.一种视频的生成方法,其特征在于,包括:

获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理;

根据所述目标对象的音频内容的音频特征,获得所述脸部图像的各二维唇形参数;

根据所述脸部图像的各二维唇形参数,获得所述脸部图像的各三维表情参数;

根据所述脸部图像的三维脸部网格、所述脸部图像的各三维表情参数和所述脸部图像的脸部图像纹理,获得所述三维脸部网格的各渲染脸部图像;

对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理,以获得融合之后的各融合视频帧图像;

对所述各融合视频帧图像进行合成处理,以生成融合视频。

2.根据权利要求1所述的方法,其特征在于,所述获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理,包括:

根据所述目标对象的图像内容,获得所述脸部图像的三维脸部网格;

根据所述脸部图像的三维脸部网格与所述目标对象的图像内容的投影关系,获得所述脸部图像的脸部图像纹理。

3.根据权利要求2所述的方法,其特征在于,所述根据所述目标对象的图像内容,获得所述脸部图像的三维脸部网格,包括:

根据所述目标对象的图像内容,获得所述目标对象的图像内容中人脸图像的关键点;

根据所述人脸图像的关键点,获得所述脸部图像的三维脸部网格。

4.根据权利要求1所述的方法,其特征在于,所述获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理,包括:

获取基础卡通形象的脸部形状和脸部纹理;

根据所述基础卡通形象的脸部形状和脸部纹理,获得所述脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理。

5.根据权利要求1所述的方法,其特征在于,所述根据所述目标对象的音频内容的音频特征,获得所述脸部图像的各二维唇形参数,包括:

获取所述目标对象的神经网络;

根据所述目标对象的音频内容的音频特征,利用所述目标对象的神经网络,获得所述脸部图像的各二维唇形参数。

6.根据权利要求5所述的方法,其特征在于,所述获取所述目标对象的神经网络之前,还包括:

利用训练对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数,进行模型训练处理,以获得通用的神经网络;

利用所述目标对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数,进行模型调整处理,以获得所述目标对象的神经网络。

7.根据权利要求1-6中任一项所述的方法,其特征在于,所述根据所述脸部图像的各二维唇形参数,获得所述脸部图像的各三维表情参数,包括:

获取所述脸部图像的各表情基;

根据所述脸部图像的各二维唇形参数、所述脸部图像的三维脸部网格和所述脸部图像

的各表情基，获得所述脸部图像的各二维唇形参数所对应的所述脸部图像的各三维唇形参数的表示，所述脸部图像的各三维唇形参数通过所述脸部图像的三维脸部网格和所述脸部图像的各表情基的线性加权表示；

根据所述脸部图像的各二维唇形参数和所述脸部图像的各三维唇形参数，获得所述脸部图像的各表情基的权重参数，以作为所述脸部图像的各三维表情参数。

8.根据权利要求7所述的方法，其特征在于，所述获取所述脸部图像的各表情基，包括：

获取所述人脸图像的三维脸部网格；

根据标准的各表情基、标准的三维脸部网格和所述人脸图像的三维脸部网格，获得所述人脸图像的各表情基。

9.根据权利要求7所述的方法，其特征在于，所述根据所述脸部图像的各二维唇形参数和所述脸部图像的各三维唇形参数，获得所述脸部图像的各表情基的权重参数，以作为所述脸部图像的各三维表情参数，包括：

确定优化问题，所述优化问题的目标函数为所述脸部图像的各二维唇形参数与所述脸部图像的各三维唇形参数的投影参数之间的差值的最小值函数，所述优化问题的约束条件为所述脸部图像的各表情基的权重参数的取值范围；其中，所述脸部图像的各三维唇形参数的投影参数为所述脸部图像的各表情基的各三维唇形参数的投影参数与所述脸部图像的各表情基的权重参数的乘积；

利用最小二乘法，对所述优化问题进行求解，以获得所述脸部图像的各表情基的权重参数。

10.一种视频的生成装置，其特征在于，包括：

网格纹理获取单元，用于获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理；

唇形参数获取单元，用于根据所述目标对象的音频内容的音频特征，获得所述脸部图像的各二维唇形参数；

表情参数获取单元，用于根据所述脸部图像的各二维唇形参数，获得所述脸部图像的各三维表情参数；

网格渲染单元，用于根据所述脸部图像的三维脸部网格、所述脸部图像的各三维表情参数和所述脸部图像的脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像；

图像融合单元，用于对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理，以获得融合之后的各融合视频帧图像；

视频合成单元，用于对所述各融合视频帧图像进行合成处理，以生成融合视频。

11.根据权利要求10所述的装置，其特征在于，所述网格纹理获取单元，具体用于

根据所述目标对象的图像内容，获得所述脸部图像的三维脸部网格；以及

根据所述脸部图像的三维脸部网格与所述目标对象的图像内容的投影关系，获得所述脸部图像的脸部图像纹理。

12.根据权利要求11所述的装置，其特征在于，所述网格纹理获取单元，具体用于

根据所述目标对象的图像内容，获得所述目标对象的图像内容中人脸图像的关键点；以及

根据所述人脸图像的关键点，获得所述脸部图像的三维脸部网格。

13.根据权利要求10所述的装置,其特征在于,所述网格纹理获取单元,具体用于

获取基础卡通形象的脸部形状和脸部纹理;以及

根据所述基础卡通形象的脸部形状和脸部纹理,获得所述脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理。

14.根据权利要求10所述的装置,其特征在于,所述唇形参数获取单元,具体用于

获取所述目标对象的神经网络;以及

根据所述目标对象的音频内容的音频特征,利用所述目标对象的神经网络,获得所述脸部图像的各二维唇形参数。

15.根据权利要求14所述的装置,其特征在于,所述唇形参数获取单元,还用于

利用训练对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数,进行模型训练处理,以获得通用的神经网络;以及

利用所述目标对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数,进行模型调整处理,以获得所述目标对象的神经网络。

16.根据权利要求10-15中任一项所述的装置,其特征在于,所述表情参数获取单元,具体用于

获取所述脸部图像的各表情基;

根据所述脸部图像的各二维唇形参数、所述脸部图像的三维脸部网格和所述脸部图像的各表情基,获得所述脸部图像的各二维唇形参数所对应的所述脸部图像的各三维唇形参数的表示,所述脸部图像的各三维唇形参数通过所述脸部图像的三维脸部网格和所述脸部图像的各表情基的线性加权表示;以及

根据所述脸部图像的各二维唇形参数和所述脸部图像的各三维唇形参数,获得所述脸部图像的各表情基的权重参数,以作为所述脸部图像的各三维表情参数。

17.根据权利要求16所述的装置,其特征在于,所述表情参数获取单元,具体用于

获取所述人脸图像的三维脸部网格;以及

根据标准的各表情基、标准的三维脸部网格和所述人脸图像的三维脸部网格,获得所述人脸图像的各表情基。

18.根据权利要求16所述的装置,其特征在于,所述表情参数获取单元,具体用于

确定优化问题,所述优化问题的目标函数为所述脸部图像的各二维唇形参数与所述脸部图像的各三维唇形参数的投影参数之间的差值的最小值函数,所述优化问题的约束条件为所述脸部图像的各表情基的权重参数的取值范围;其中,所述脸部图像的各三维唇形参数的投影参数为所述脸部图像的各表情基的各三维唇形参数的投影参数与所述脸部图像的各表情基的权重参数的乘积;以及

利用最小二乘法,对所述优化问题进行求解,以获得所述脸部图像的各表情基的权重参数。

19.一种电子设备,其特征在于,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-9中任一项所述的方法。

20.一种存储有计算机指令的非瞬时计算机可读存储介质，其特征在于，所述计算机指令用于使所述计算机执行权利要求1-9中任一项所述的方法。

视频的生成方法、装置、电子设备及可读存储介质

技术领域

[0001] 涉及计算机技术，具体涉及计算机视觉技术，尤其涉及一种视频的生成方法、装置、电子设备及可读存储介质。

背景技术

[0002] 随着互联网的深入发展，终端能够集成越来越多的功能，从而使得应用于终端上的应用(Application，APP)层出不穷。有些应用中会涉及视频的内容表达，通常可以采用人工方式，进行视频的录制，以生成具有各种内容表达的视频。

[0003] 然而，由于完全依赖人工录制，使得视频生成的效率较低。尤其是对于一些具有固定内容表达的视频，例如新闻播报、学科教学等内容表达的视频，这些视频所表达的内容是固定的，完全采用人工录制的方式，不但效率特别地，而且还会造成不必要的人力资源的浪费。

发明内容

[0004] 本申请的多个方面提供一种视频的生成方法、装置、电子设备及可读存储介质，用以提高视频生成的效率。

[0005] 本申请的一方面，提供一种视频的生成方法，包括：

[0006] 获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理；

[0007] 根据所述目标对象的音频内容的音频特征，获得所述脸部图像的各二维唇形参数；

[0008] 根据所述脸部图像的各二维唇形参数，获得所述脸部图像的各三维表情参数；

[0009] 根据所述脸部图像的三维脸部网格、所述脸部图像的各三维表情参数和所述脸部图像的脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像；

[0010] 对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理，以获得融合之后的各融合视频帧图像；

[0011] 对所述各融合视频帧图像进行合成处理，以生成融合视频。

[0012] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理，包括：

[0013] 根据所述目标对象的图像内容，获得所述脸部图像的三维脸部网格；

[0014] 根据所述脸部图像的三维脸部网格与所述目标对象的图像内容的投影关系，获得所述脸部图像的脸部图像纹理。

[0015] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述根据所述目标对象的图像内容，获得所述脸部图像的三维脸部网格，包括：

[0016] 根据所述目标对象的图像内容，获得所述目标对象的图像内容中人脸图像的关键点；

[0017] 根据所述人脸图像的关键点，获得所述脸部图像的三维脸部网格。

[0018] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理，包括：

[0019] 获取基础卡通形象的脸部形状和脸部纹理；

[0020] 根据所述基础卡通形象的脸部形状和脸部纹理，获得所述脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理。

[0021] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述根据所述目标对象的音频内容的音频特征，获得所述脸部图像的各二维唇形参数，包括：

[0022] 获取所述目标对象的神经网络；

[0023] 根据所述目标对象的音频内容的音频特征，利用所述目标对象的神经网络，获得所述脸部图像的各二维唇形参数。

[0024] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述获取所述目标对象的神经网络之前，还包括：

[0025] 利用训练对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数，进行模型训练处理，以获得通用的神经网络；

[0026] 利用所述目标对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数，进行模型调整处理，以获得所述目标对象的神经网络。

[0027] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述根据所述脸部图像的各二维唇形参数，获得所述脸部图像的各三维表情参数，包括：

[0028] 获取所述脸部图像的各表情基；

[0029] 根据所述脸部图像的各二维唇形参数、所述脸部图像的三维脸部网格和所述脸部图像的各表情基，获得所述脸部图像的各二维唇形参数所对应的所述脸部图像的各三维唇形参数的表示，所述脸部图像的各三维唇形参数通过所述脸部图像的三维脸部网格和所述脸部图像的各表情基的线性加权表示；

[0030] 根据所述脸部图像的各二维唇形参数和所述脸部图像的各三维唇形参数，获得所述脸部图像的各表情基的权重参数，以作为所述脸部图像的各三维表情参数。

[0031] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述获取所述脸部图像的各表情基，包括：

[0032] 获取所述人脸图像的三维脸部网格；

[0033] 根据标准的各表情基、标准的三维脸部网格和所述人脸图像的三维脸部网格，获得所述人脸图像的各表情基。

[0034] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述根据所述脸部图像的各二维唇形参数和所述脸部图像的各三维唇形参数，获得所述脸部图像的各表情基的权重参数，以作为所述脸部图像的各三维表情参数，包括：

[0035] 确定优化问题，所述优化问题的目标函数为所述脸部图像的各二维唇形参数与所述脸部图像的各三维唇形参数的投影参数之间的差值的最小值函数，所述优化问题的约束条件为所述脸部图像的各表情基的权重参数的取值范围；其中，所述脸部图像的各三维唇形参数的投影参数为所述脸部图像的各表情基的各三维唇形参数的投影参数与所述脸部图像的各表情基的权重参数的乘积；

[0036] 利用最小二乘法，对所述优化问题进行求解，以获得所述脸部图像的各表情基的权重参数。

[0037] 本申请的另一方面，提供一种视频的生成装置，包括：

[0038] 网格纹理获取单元，用于获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理；

[0039] 唇形参数获取单元，用于根据所述目标对象的音频内容的音频特征，获得所述脸部图像的各二维唇形参数；

[0040] 表情参数获取单元，用于根据所述脸部图像的各二维唇形参数，获得所述脸部图像的各三维表情参数；

[0041] 网格渲染单元，用于根据所述脸部图像的三维脸部网格、所述脸部图像的各三维表情参数和所述脸部图像的脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像；

[0042] 图像融合单元，用于对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理，以获得融合之后的各融合视频帧图像；

[0043] 视频合成单元，用于对所述各融合视频帧图像进行合成处理，以生成融合视频。

[0044] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述网格纹理获取单元，具体用于

[0045] 根据所述目标对象的图像内容，获得所述脸部图像的三维脸部网格；以及

[0046] 根据所述脸部图像的三维脸部网格与所述目标对象的图像内容的投影关系，获得所述脸部图像的脸部图像纹理。

[0047] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述网格纹理获取单元，具体用于

[0048] 根据所述目标对象的图像内容，获得所述目标对象的图像内容中人脸图像的关键点；以及

[0049] 根据所述人脸图像的关键点，获得所述脸部图像的三维脸部网格。

[0050] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述网格纹理获取单元，具体用于

[0051] 获取基础卡通形象的脸部形状和脸部纹理；以及

[0052] 根据所述基础卡通形象的脸部形状和脸部纹理，获得所述脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理。

[0053] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述唇形参数获取单元，具体用于

[0054] 获取所述目标对象的神经网络；以及

[0055] 根据所述目标对象的音频内容的音频特征，利用所述目标对象的神经网络，获得所述脸部图像的各二维唇形参数。

[0056] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述唇形参数获取单元，还用于

[0057] 利用训练对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数，进行模型训练处理，以获得通用的神经网络；以及

[0058] 利用所述目标对象的图像数据、该图像数据所对应的音频数据和该图像数据所对

应的各二维唇形参数，进行模型调整处理，以获得所述目标对象的神经网络。

[0059] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述表情参数获取单元，具体用于

[0060] 获取所述脸部图像的各表情基；

[0061] 根据所述脸部图像的各二维唇形参数、所述脸部图像的三维脸部网格和所述脸部图像的各表情基，获得所述脸部图像的各二维唇形参数所对应的所述脸部图像的各三维唇形参数的表示，所述脸部图像的各三维唇形参数通过所述脸部图像的三维脸部网格和所述脸部图像的各表情基的线性加权表示；以及

[0062] 根据所述脸部图像的各二维唇形参数和所述脸部图像的各三维唇形参数，获得所述脸部图像的各表情基的权重参数，以作为所述脸部图像的各三维表情参数。

[0063] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述表情参数获取单元，具体用于

[0064] 获取所述人脸图像的三维脸部网格；以及

[0065] 根据标准的各表情基、标准的三维脸部网格和所述人脸图像的三维脸部网格，获得所述人脸图像的各表情基。

[0066] 如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述表情参数获取单元，具体用于

[0067] 确定优化问题，所述优化问题的目标函数为所述脸部图像的各二维唇形参数与所述脸部图像的各三维唇形参数的投影参数之间的差值的最小值函数，所述优化问题的约束条件为所述脸部图像的各表情基的权重参数的取值范围；其中，所述脸部图像的各三维唇形参数的投影参数为所述脸部图像的各表情基的各三维唇形参数的投影参数与所述脸部图像的各表情基的权重参数的乘积；以及

[0068] 利用最小二乘法，对所述优化问题进行求解，以获得所述脸部图像的各表情基的权重参数。

[0069] 本发明的另一方面，提供一种电子设备，包括：

[0070] 至少一个处理器；以及

[0071] 与所述至少一个处理器通信连接的存储器；其中，

[0072] 所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行如上所述的方面和任一可能的实现方式的方法。

[0073] 本发明的另一方面，提供一种存储有计算机指令的非瞬时计算机可读存储介质，所述计算机指令用于使所述计算机执行如上所述的方面和任一可能的实现方式的方法。

[0074] 由上述技术方案可知，本申请实施例通过基于待生成视频中目标对象的音频内容，获得该音频内容所对应的所述目标对象的脸部图像的各二维唇形参数，进而将所述脸部图像的各二维唇形参数和所述脸部图像的各表情基进行规则化解算，建立所述脸部图像的各二维唇形参数到所述脸部图像的各表情基的权重参数的映射，进而，再进一步利用所述脸部图像的各表情基的权重参数，结合该目标对象的脸部图像的三维脸部网格和脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像，使得能够根据各渲染脸部图像和模板视频的各视频帧图像获得融合视频，无需人工参与，从而有效提高了视频生成的效率。

[0075] 另外，采用本申请所提供的技术方案，无需大量的三维标注数据来学习新形象的音频特征与表情基的权重参数之间的映射关系，能够进一步有效地提高视频生成的效率。

[0076] 另外，采用本申请所提供的技术方案，只需要目标对象的少量图像数据以及该图像数据所对应的音频数据，就能够完成目标对象的神经网络的训练，从而进一步有效地提高了视频生成的效率。

[0077] 另外，采用本申请所提供的技术方案，可以虚拟人物形象，通常可以根据单张或者多张目标对象的图像，获得该目标对象的脸部图像的三维脸部网格和脸部图像纹理，并通过音频生成的二维唇动所映射到的表情应用到脸部图像的三维脸部网格，从而实现虚拟人物形象的视频产生。

[0078] 另外，采用本申请所提供的技术方案，通过获取标准的三维脸部网格与所述人脸图像的三维脸部网格之间的差异，进而将该差异传递到标准的各表情基，最终生成人脸图像的各表情基，从而实现了自动化生成人脸图像的各表情基。

[0079] 另外，采用本申请所提供的技术方案，能够有效地提高用户的体验。

[0080] 上述方面或可能的实现方式所具有的其他效果将在下文中结合具体实施例加以说明。

附图说明

[0081] 为了更清楚地说明本申请实施例中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。附图仅仅用于更好地理解本方案，不构成对本申请的限定。其中：

[0082] 图1为本申请一实施例提供的视频的生成方法的流程示意图；

[0083] 图2为本申请另一实施例提供的视频的生成装置的结构示意图；

[0084] 图3是用来实现本申请实施例提供的视频的生成方法的电子设备的示意图。

具体实施方式

[0085] 以下结合附图对本申请的示范性实施例做出说明，其中包括本申请实施例的各种细节以助于理解，应当将它们认为仅仅是示范性的。因此，本领域普通技术人员应当认识到，可以对这里描述的实施例做出各种改变和修改，而不会背离本申请的范围和精神。同样，为了清楚和简明，以下的描述中省略了对公知功能和结构的描述。

[0086] 显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的全部其他实施例，都属于本申请保护的范围。

[0087] 需要说明的是，本申请实施例中所涉及的终端可以包括但不限于手机、个人数字助理(Personal Digital Assistant,PDA)、无线手持设备、平板电脑(Tablet Computer)、个人电脑(Personal Computer,PC)、MP3播放器、MP4播放器、可穿戴设备(例如，智能眼镜、智能手表、智能手环等)等。

[0088] 另外，本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。另

外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

[0089] 图1为本申请一实施例提供的视频的生成方法的流程示意图，如图1所示。

[0090] 101、获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理；

[0091] 102、根据所述目标对象的音频内容的音频特征，获得所述脸部图像的各二维唇形参数。

[0092] 103、根据所述脸部图像的各二维唇形参数，获得所述脸部图像的各三维表情参数。

[0093] 104、根据所述脸部图像的三维脸部网格、所述脸部图像的各三维表情参数和所述脸部图像的脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像。

[0094] 105、对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理，以获得融合之后的各融合视频帧图像。

[0095] 106、对所述各融合视频帧图像进行合成处理，以生成融合视频。

[0096] 需要说明的是，101～106的执行主体的部分或全部可以为位于本地终端的应用，或者还可以为设置在位于本地终端的应用中的插件或软件开发工具包(Software Development Kit，SDK)等功能单元，或者还可以为位于网络侧服务器中的处理引擎，或者还可以为位于网络侧的分布式系统，例如，网络侧的视频处理平台中的处理引擎或者分布式系统等，本实施例对此不进行特别限定。

[0097] 可以理解的是，所述应用可以是安装在终端上的本地程序(nativeApp)，或者还可以是终端上的浏览器的一个网页程序(webApp)，本实施例对此不进行限定。

[0098] 这样，通过基于待生成视频中目标对象的音频内容，获得该音频内容所对应的所述目标对象的脸部图像的各二维唇形参数，进而将所述脸部图像的各二维唇形参数和所述脸部图像的各表情基进行规则化解算，建立所述脸部图像的各二维唇形参数到所述脸部图像的各表情基的权重参数的映射，进而，再进一步利用所述脸部图像的各表情基的权重参数，结合该目标对象的脸部图像的三维脸部网格和脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像，使得能够根据各渲染脸部图像和模板视频的各视频帧图像获得融合视频，无需人工参与，从而有效提高了视频生成的效率。

[0099] 本申请中，待生成视频中目标对象可以是人物形象，或者还可以是卡通形象，本申请对此不进行特别限定。

[0100] 可选地，在本实施例的一个可能的实现方式中，对于目标对象为人物形象的场景来说，在101之前，还可以进一步预先获取该目标对象的图像内容。

[0101] 在一个具体的实现过程中，具体可以获取操作人员所提供的一张或者多张该目标对象的图像，进而，则可以根据所获取的一张或者多张该目标对象的图像，获得该目标对象的图像内容。

[0102] 在另一个具体的实现过程中，具体可以获取操作人员所提供的一个该目标对象的视频，进而，则可以从视频中，抽取视频的视频帧图像。具体来说，具体可以采用视频解码方式(如x264等视频解码方式)，对视频进行视频解码处理，以获得视频的数据流，根据视频的数据流获取各帧的图像数据，进而，再进一步采用图像编码方式(如png、jpg等图像编码方式)，对各帧的图像数据进行图像编码处理，以获得视频的视频帧图像。

[0103] 在获得视频的视频帧图像之后，则可以根据所述视频帧图像，获得该目标对象的图像内容。

[0104] 具体地，在101中，在获得目标对象的图像内容之后，具体可以根据所述目标对象的图像内容，获得所述脸部图像的三维脸部网格，进而，则可以根据所述脸部图像的三维脸部网格与所述目标对象的图像内容的投影关系，获得所述脸部图像的脸部图像纹理，此处可以称为脸部图像的UVmap。

[0105] 例如，具体可以根据所述目标对象的图像内容，获得所述目标对象的图像内容中人脸图像的关键点，进而，则可以根据所述人脸图像的关键点，获得所述脸部图像的三维脸部网格。

[0106] 在该例子中，具体可以对所述目标对象的图像内容中人脸图像进行关键点检测处理，从而获得所述目标对象的图像内容中人脸图像的关键点。

[0107] 具体来说，所述关键点检测处理是在人脸检测的基础之上，对人脸上的特征点例如人脸框、眼睛、鼻子、嘴巴等进行定位。目前，具体可以对人脸的150个特征点进行检测处理。具体可以包括三个步骤，即检测一次边框检测和检测两次人脸关键点检测。

[0108] 例如，可以对所述目标对象的图像内容中人脸图像定位出人脸面部的关键区域位置，作为关键点，包括眉毛、眼睛、鼻子、嘴巴、脸部轮廓等。

[0109] 具体来说，具体可以根据所获得的所述人脸图像的关键点，利用公开的三维脸部网格模型，例如三维可变形模型(3D Morphable Model，3DMM)等，获得所述脸部图像的三维脸部网格，记作M_{id}。

[0110] 可选地，在本实施例的一个可能的实现方式中，对于目标对象为卡通形象的场景来说，在101之前，还可以进一步预先获取该目标对象的形状融合变形器(blend shape)。

[0111] 具体地，在101中，在获得目标对象的形状融合变形器(blend shape)之后，具体可以基于该目标对象的形状融合变形器(blend shape)，则可以获取基础卡通形象的脸部形状和脸部纹理，进而，则可以根据所述基础卡通形象的脸部形状和脸部纹理，获得所述脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理。

[0112] 可选地，在本实施例的一个可能的实现方式中，在102之前，还可以进一步获取所述目标对象的音频内容。

[0113] 在一个具体的实现过程中，具体可以获取所述目标对象的文本，进而，则可以进一步利用文本转换技术，对所述文本进行语音转换处理，获得所述目标对象的音频内容。

[0114] 在另一个具体的实现过程中，具体可以直接获取所述目标对象的音频内容。

[0115] 可选地，在本实施例的一个可能的实现方式中，在102中，具体可以获取所述目标对象的神经网络，进而，则可以根据所述目标对象的音频内容的音频特征，利用所述目标对象的神经网络，获得所述脸部图像的各二维唇形参数。

[0116] 其中，所述目标对象的神经网络，可以采用循环神经网络(Recurrent Neural Network，RNN)，或者还可以采用其他神经网络，本实施例对此不进行特别限定。

[0117] 所谓的所述脸部图像的各二维唇形参数，是指脸部图像中唇部轮廓点的位置数据，用于描述唇部动作形状。

[0118] 在该实现方式之前，还可以进一步利用训练对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数，进行模型训练处理，以获得通用的神

经网络。

[0119] 在实际应用中，不同人物形象的表情虽然整体形变类似，但是都可能会存在轻微差异。因此，可以利用目标对象的图像数据和该图像数据所对应的音频数据，进行模型调整处理，以获得所述目标对象的神经网络。

[0120] 具体地，在获得通用的神经网络之后，针对特定的目标对象，则可以进一步利用所述目标对象的音视频数据即图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数，进行模型调整处理，以获得所述目标对象的神经网络。

[0121] 例如，可以获取闻联播不同的发言人的音视频数据即图像数据、以及该图像数据所对应的音频数据，训练音频特征与二维唇形参数的通用的神经网络。此时，可以采用现有的标注技术或者标注工具，对图像数据中人脸图像中唇部进行标注处理，以获得训练数据中音频特征所对应的各二维唇形参数，通常可以有64个唇形标注点。不同发言人使用相同的编码器(encoder)之后，接不同的解码器(decoder)进行解码处理，输出不同发言人所对应的各自的二维唇形参数。

[0122] 获取到新的发言人音视频数据即图像数据、以及该图像数据所对应的音频数据之后，可以进一步对通用的神经网络进行微调，训练语音特征到新发言人的二维唇形参数，此时，只需要使用不到1小时的音视频数据来进行迁移唇形动作。

[0123] 这样，只需要目标对象的少量图像数据以及该图像数据所对应的音频数据，就能够完成目标对象的神经网络的训练，从而进一步有效地提高了视频生成的效率。

[0124] 可选地，在本实施例的一个可能的实现方式中，在103中，具体可以获取所述脸部图像的各表情基，进而，则可以根据所述脸部图像的各二维唇形参数、所述脸部图像的三维脸部网格和所述脸部图像的各表情基，获得所述脸部图像的各二维唇形参数所对应的所述脸部图像的各三维唇形参数的表示，所述脸部图像的各三维唇形参数通过所述脸部图像的三维脸部网格和所述脸部图像的各表情基的线性加权表示。然后，则可以根据所述脸部图像的各二维唇形参数和所述脸部图像的各三维唇形参数，获得所述脸部图像的各表情基的权重参数，以作为所述脸部图像的各三维表情参数。

[0125] 在一个具体的实现过程中，具体可以获取所述人脸图像的三维脸部网格，进而，则可以根据标准的各表情基、标准的三维脸部网格和所述人脸图像的三维脸部网格，获得所述人脸图像的各表情基。

[0126] 具体地，具体可以获取所述人脸图像的三维脸部网格与所述标准的三维脸部网格之间在顶点法线方向上的位移，进而，将该位移传递到标准的各表情基，以生成所述人脸图像的各表情基。

[0127] 例如，可以采用公开人脸网格数据集所提供的萨里人脸模型(Surrey Face Model，SFM)，作为无表情的标准的三维脸部网格，记作M_{mean}；通过美术建模人员，建立表达唇部动作变化的27个标准的表情基，记作M_{bs_j}(j＝1,2,…,27)；进一步获取根据所获得的所述人脸图像的关键点，利用公开的三维脸部网格模型，例如三维可变形模型(3D Morphable Model，3DMM)等，所获得的所述脸部图像的三维脸部网格，记作M_{id}。在获得无表情的标准的三维脸部网格M_{mean}和27个标准的表情基M_{bs_i}(j＝1,2,…,27)，以及所述脸部图像的三维脸部网格M_{id}之后，可以获取所述人脸图像的三维脸部网格M_{id}与所述标准的三维脸部网格M_{mean}之间在顶点法线方向上的位移，进而，将该位移传递到标准的各表情基M_{bs_j}(j

＝1,2,…,27)，以生成所述人脸图像的各表情基$M_{id_bs_j}$(j＝1,2,…,27)。

[0128] 这样，通过获取标准的三维脸部网格与所述人脸图像的三维脸部网格之间的差异，进而将该差异传递到标准的各表情基，最终生成人脸图像的各表情基，从而实现了自动化生成人脸图像的各表情基。

[0129] 可以理解的是，所述脸部图像的三维脸部网格，可以理解为没有任何表情的人脸图像的无表情的三维脸部网格；所述人脸图像的各表情基，则可以理解为带各种表情的所述人脸图像的带表情的三维唇部网格。

[0130] 通常，人脸图像的脸部表情变化，可以用一个表情基矩阵描述，表情基矩阵是由一组表情基的向量，及各表情基的权重系数组成，其中，各表情基用来描述脸部表情变化。

[0131] 在另一个具体的实现过程中，在获得所述脸部图像的各二维唇形参数和所述脸部图像的三维脸部网格之后，具体可以将所述脸部图像的各二维唇形参数，对应到所述脸部图像的三维脸部网格上，类似地，还可以进一步对应到所述人脸图像的各表情基上，从而获得所述脸部图像的各三维唇形参数的表示。

[0132] 所谓的脸部图像的各二维唇形参数，是指所述脸部图像中唇部轮廓的位置参数，由指定数量(如64个等)的2D点的坐标组成。

[0133] 所谓的脸部图像的各三维唇形参数，是指所述脸部图像中唇部轮廓的位置参数在所述人脸图像的带表情的三维脸部网格上所对应的位置参数，由指定数量(如64个等)的3D点的坐标组成。其中，所述人脸图像的带表情的三维脸部网格可以为所述人脸图像的无表情的三维脸部网格和所述人脸图像的各表情基的线性加权。

[0134] 也就是说，所述脸部图像的三维唇形参数，可以理解为所述脸部图像中唇部轮廓的位置参数在所述人脸图像的无表情的三维脸部网格上所对应的位置参数，以及所述脸部图像中唇部轮廓的位置参数在所述人脸图像的各表情基上所对应的位置参数，的线性加权。

[0135] 具体地，所述脸部图像的各三维唇形参数可以通过所述脸部图像的三维脸部网格和所述脸部图像的各表情基的线性加权表示，即所述脸部图像的各三维唇形参数的表示为所述脸部图像的三维脸部网格和所述脸部图像的各表情基的线性加权。

[0136] 这样，所述脸部图像的各二维唇形参数，还可以通过所述脸部图像的各三维唇形参数的投影参数来表示。

[0137] 例如，所述脸部图像的各二维唇形参数可以根据唇部在水平方向的长度对齐到统一尺度。

[0138] 第i帧的二维唇形参数用$\mathrm{p}_{2\mathrm{d}}^{\mathrm{i}}$表示，是一个128维的列向量。$(x_k,y_k)$是第k个二维唇形点的坐标，其中k＝0,1,…63。因此，$p_{2d}^{i}=(\mathrm{x}_0,\mathrm{y}_0,\ldots,\mathrm{x}_{63},\mathrm{y}_{63})^{\mathrm{T}}$。

[0139] 第i帧的三维唇形参数的投影参数用$\mathrm{p}_{3\mathrm{d}}^{\mathrm{i}}{}'=\mathrm{M}*\alpha$表示，是一个128×28维的矩阵。其中，M为所述人脸图像的各二维唇形参数在所述人脸图像的无表情的三维脸部网格上所对应的各三维唇形参数的投影参数，以及在所述人脸图像的各表情基上所对应的各三维唇形参数的投影参数所组成的矩阵，是一个128×28维的矩阵；α为所述脸部图像的各表情基的权重参数，是一个28维的列向量，对每一个元素α_j限定在0.0到1.0之间，其中，j＝1，

2,…27。

[0140] $(x_{3d_k}^{j},\ y_{3d_k}^{j})$是第j个表情基的第k个三维唇形参数(即3D点的坐标)的在屏幕坐标系下正交投影的位置参数,其中,j=1,2,…27,k=0,1,…63。

[0141] $(x_{3d_k}^{0},\ y_{3d_k}^{0})$是所述脸部图像的三维脸部网格的第k个三维唇形参数(即3D点的坐标)的在屏幕坐标系下正交投影的位置参数,其中,k=0,1,…63。

[0142] 因此,$M=\begin{pmatrix} x_{3d_0}^{0} & x_{3d_0}^{1} & \cdots & x_{3d_0}^{27} \\ y_{3d_0}^{0} & y_{3d_0}^{1} & \cdots & y_{3d_0}^{27} \\ x_{3d_1}^{0} & x_{3d_1}^{1} & \cdots & y_{3d_1}^{27} \\ \cdots & & \cdots & \cdots \\ x_{3d_{63}}^{0} & x_{3d_{63}}^{1} & \cdots & x_{3d_{63}}^{27} \\ y_{3d_{63}}^{0} & y_{3d_{63}}^{1} & \cdots & y_{3d_{63}}^{27} \end{pmatrix}$。

[0143] 在另一个具体的实现过程中,在获得所述脸部图像的各二维唇形参数所对应的所述脸部图像的各三维唇形参数的表示之后,还可以进一步将映射问题转换成求解优化问题。

[0144] 具体地,具体可以确定优化问题,即

[0145] $\min ||p_{2d}^{i}-M*\alpha||_2$;

[0146] s.t.0≤α_j≤1.0。

[0147] 其中,所述优化问题的目标函数为所述脸部图像的各二维唇形参数与所述脸部图像的各三维唇形参数的投影参数之间的差值的最小值函数,所述优化问题的约束条件为所述脸部图像的各表情基的权重参数的取值范围;其中,所述脸部图像的各三维唇形参数的投影参数为所述脸部图像的各表情基的各三维唇形参数的投影参数与所述脸部图像的各表情基的权重参数的乘积。

[0148] 然后,在确定所述优化问题之后,则可以利用最小二乘法,对所述优化问题进行求解,以获得所述脸部图像的各表情基的权重参数。

[0149] 可选地,在本实施例的一个可能的实现方式中,在104中,具体可以将所述脸部图像的各三维表情参数,应用到所述脸部图像的三维脸部网格之上,以获得带表情的脸部图像的各三维脸部网格。然后,则可以对所述带表情的脸部图像的各三维脸部网格进行投影处理。最后,则可以利用所述脸部图像的脸部图像纹理,对所述投影处理的投影结果进行渲染处理,以获得所述三维脸部网格的各渲染脸部图像。

[0150] 其中,所述脸部图像的各三维表情参数在应用过程中,可以将其作为所述脸部图像的各表情基的权重系数,对所述脸部图像的各表情基,进行线性叠加,从而获得可用于所述脸部图像的三维脸部网格上的三维表情基网格。

[0151] 通常,所述脸部图像的各三维表情参数是以具体的人物形象作为训练对象,利用训练对象的连续的视频数据和该视频数据所对应的音频数据所学习到的音频特征与脸部图像的各三维表情参数之间映射关系,因此,所获得的脸部图像的各三维表情参数可以直接应用在人物形象作为目标对象时所获得的脸部图像的各三维脸部网格上。

[0152] 那么，对于卡通形象作为目标对象时所获得的脸部图像的各三维脸部网格，则不能直接应用于所获得的脸部图像的各三维表情参数，而是需要通过对卡通形象作为目标对象时所获得的脸部图像的各三维脸部网格与人物形象作为目标对象时所获得的脸部图像的各三维脸部网格进行对应顶点标注处理，以获得卡通形象的各三维脸部网格与人物形象的各三维脸部网格之间的对应关系。基于所获得的对应关系，则可以将所获得的脸部图像的各三维表情参数应用在卡通形象作为目标对象时所获得的脸部图像的各三维脸部网格上，从而实现了将所获得的脸部图像的各三维表情参数由人物形象迁移到卡通形象上，

[0153] 具体来说，在获得所述带表情的脸部图像的各三维脸部网格之后，可以进一步根据相机正视图的坐标关系，将其投影到一个二维平面上。

[0154] 例如，具体可以假设观测到带表情的脸部图像的各三维脸部网格的观测者的眼睛为一个质点，沿质点到带表情的脸部图像的各三维脸部网格的中心的连线为z轴，投影就是将z轴坐标消除，从而得到一个二维坐标平面和带表情的脸部图像的各三维脸部网格在二维坐标平面内的各个网格点坐标。因为每三个带表情的脸部图像的三维脸部网格的坐标点都能形成一个的三角形，每个坐标点都可以根据所述脸部图像的脸部图像纹理得到采样的纹理颜色，因此，在获得所述投影处理的投影结果之后，带表情的脸部图像的各三维脸部网格的各三角形内像素值都可以用三个点的像素值进行插值填充。其中，所述插值可以为像素值的线形插值，或者还可以为像素值的样条插值，或者还可以为像素值的其他插指，本实施例对此不进行特别限定。

[0155] 可选地，在本实施例的一个可能的实现方式中，在105中，具体可以利用预先设置的人脸图像蒙版，对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理，以获得融合之后的各融合视频帧图像。

[0156] 具体地，具体可以采用现有的各种融合方式，例如，阿尔法融合(alpha blending)方式、泊松融合(Possion Blending)方式等，进行融合处理。

[0157] 其中，本实施例中的所采用的人脸图像蒙版，可以为根据人脸纹理的人脸五官位置预先设置的固定形状的人脸图像蒙版，其中预设了保留用于融合的人脸图像范围。

[0158] 这样，通过采用根据人脸纹理的人脸五官位置预先设置的固定形状的人脸图像蒙版，配合三维人脸网格的各渲染人脸图像和所述模板视频的各视频帧图像进行融合处理，能够有效提升融合视频中用户人脸边缘与模板视频中模板背景之间的融合效果。

[0159] 在获得融合之后的各融合视频帧图像之后，还可以进一步对各融合视频帧图像进行图像解码处理，以获得原始各帧的图像数据，再将各帧的图像数据拼接成数据流，进而进行视频编码处理，以生成融合视频。

[0160] 采用本申请所提供的技术方案，能够有效地降低一些具有固定内容表达的视频的人力成本，例如新闻播报、学科教学等内容表达的视频，不需要视频录制和过多的人工参与，仅需要一个文本或者一段音频，便能实现实时播报。此外，还能够进一步满足在某些仅有语音播报的场景下对形象的需求。

[0161] 本实施例中，通过基于待生成视频中目标对象的音频内容，获得该音频内容所对应的所述目标对象的脸部图像的各二维唇形参数，进而将所述脸部图像的各二维唇形参数和所述脸部图像的各表情基进行规则化解算，建立所述脸部图像的各二维唇形参数到所述脸部图像的各表情基的权重参数的映射，进而，再进一步利用所述脸部图像的各表情基的

权重参数，结合该目标对象的脸部图像的三维脸部网格和脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像，使得能够根据各渲染脸部图像和模板视频的各视频帧图像获得融合视频，无需人工参与，从而有效提高了视频生成的效率。

[0162] 另外，采用本申请所提供的技术方案，无需大量的三维标注数据来学习新形象的音频特征与表情基的权重参数之间的映射关系，能够进一步有效地提高视频生成的效率。

[0163] 另外，采用本申请所提供的技术方案，只需要目标对象的少量图像数据以及该图像数据所对应的音频数据，就能够完成目标对象的神经网络的训练，从而进一步有效地提高了视频生成的效率。

[0164] 另外，采用本申请所提供的技术方案，可以虚拟人物形象，通常可以根据单张或者多张目标对象的图像，获得该目标对象的脸部图像的三维脸部网格和脸部图像纹理，并通过音频生成的二维唇动所映射到的表情应用到脸部图像的三维脸部网格，从而实现虚拟人物形象的视频产生。

[0165] 另外，采用本申请所提供的技术方案，通过获取标准的三维脸部网格与所述人脸图像的三维脸部网格之间的差异，进而将该差异传递到标准的各表情基，最终生成人脸图像的各表情基，从而实现了自动化生成人脸图像的各表情基。

[0166] 另外，采用本申请所提供的技术方案，能够有效地提高用户的体验。

[0167] 需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请并不受所描述的动作顺序的限制，因为依据本申请，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本申请所必须的。

[0168] 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其他实施例的相关描述。

[0169] 图2为本申请另一实施例提供的视频的生成装置的结构示意图，如图2所示。本实施例的视频的生成装置200可以包括网格纹理获取单元201、唇形参数获取单202、表情参数获取单元203、网格渲染单元204、图像融合单元205和视频合成单元205。其中，网格纹理获取单元201，用于获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理；唇形参数获取单元202，用于根据所述目标对象的音频内容的音频特征，获得所述脸部图像的各二维唇形参数；表情参数获取单元203，用于根据所述脸部图像的各二维唇形参数，获得所述脸部图像的各三维表情参数；网格渲染单元204，用于根据所述脸部图像的三维脸部网格、所述脸部图像的各三维表情参数和所述脸部图像的脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像；图像融合单元205，用于对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理，以获得融合之后的各融合视频帧图像；视频合成单元206，用于对所述各融合视频帧图像进行合成处理，以生成融合视频。

[0170] 需要说明的是，本实施例所提供的视频的生成装置的执行主体的部分或全部可以为位于本地终端的应用，或者还可以为设置在位于本地终端的应用中的插件或软件开发工具包(Software Development Kit，SDK)等功能单元，或者还可以为位于网络侧服务器中的处理引擎，或者还可以为位于网络侧的分布式系统，例如，网络侧的视频处理平台中的处理

引擎或者分布式系统等，本实施例对此不进行特别限定。

[0171] 可以理解的是，所述应用可以是安装在终端上的本地程序(nativeApp)，或者还可以是终端上的浏览器的一个网页程序(webApp)，本实施例对此不进行限定。

[0172] 可选地，在本实施例的一个可能的实现方式中，所述网格纹理获取单元201，具体可以用于根据所述目标对象的图像内容，获得所述脸部图像的三维脸部网格；以及根据所述脸部图像的三维脸部网格与所述目标对象的图像内容的投影关系，获得所述脸部图像的脸部图像纹理。

[0173] 在一个具体的实现过程中，所述网格纹理获取单元201，具体可以用于根据所述目标对象的图像内容，获得所述目标对象的图像内容中人脸图像的关键点；以及根据所述人脸图像的关键点，获得所述脸部图像的三维脸部网格。

[0174] 可选地，在本实施例的一个可能的实现方式中，所述网格纹理获取单元201，具体可以用于获取基础卡通形象的脸部形状和脸部纹理；以及根据所述基础卡通形象的脸部形状和脸部纹理，获得所述脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理。

[0175] 可选地，在本实施例的一个可能的实现方式中，所述唇形参数获取单元202，具体可以用于获取所述目标对象的神经网络；以及根据所述目标对象的音频内容的音频特征，利用所述目标对象的神经网络，获得所述脸部图像的各二维唇形参数。

[0176] 在一个具体的实现过程中，所述唇形参数获取单元202，还可以进一步用于利用训练对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数，进行模型训练处理，以获得通用的神经网络；以及利用所述目标对象的图像数据、该图像数据所对应的音频数据和该图像数据所对应的各二维唇形参数，进行模型调整处理，以获得所述目标对象的神经网络。

[0177] 可选地，在本实施例的一个可能的实现方式中，所述表情参数获取单元203，具体可以用于获取所述脸部图像的各表情基；根据所述脸部图像的各二维唇形参数、所述脸部图像的三维脸部网格和所述脸部图像的各表情基，获得所述脸部图像的各二维唇形参数所对应的所述脸部图像的各三维唇形参数的表示，所述脸部图像的各三维唇形参数通过所述脸部图像的三维脸部网格和所述脸部图像的各表情基的线性加权表示；以及根据所述脸部图像的各二维唇形参数和所述脸部图像的各三维唇形参数，获得所述脸部图像的各表情基的权重参数，以作为所述脸部图像的各三维表情参数。

[0178] 在一个具体的实现过程中，所述表情参数获取单元203，具体可以用于获取所述人脸图像的三维脸部网格；以及根据标准的各表情基、标准的三维脸部网格和所述人脸图像的三维脸部网格，获得所述人脸图像的各表情基。

[0179] 在另一个具体的实现过程中，所述表情参数获取单元203，具体可以用于确定优化问题，所述优化问题的目标函数为所述脸部图像的各二维唇形参数与所述脸部图像的各三维唇形参数的投影参数之间的差值的最小值函数，所述优化问题的约束条件为所述脸部图像的各表情基的权重参数的取值范围；其中，所述脸部图像的各三维唇形参数的投影参数为所述脸部图像的各表情基的各三维唇形参数的投影参数与所述脸部图像的各表情基的权重参数的乘积；以及利用最小二乘法，对所述优化问题进行求解，以获得所述脸部图像的各表情基的权重参数。

[0180] 需要说明的是，图1对应的实施例中的方法可以由本实施例提供的视频的生成装

置实现。详细描述可以参见图1对应的实施例中的相关内容，此处不再赘述。

[0181] 本实施例中，通过唇形参数获取单元基于待生成视频中目标对象的音频内容，获得该音频内容所对应的所述目标对象的脸部图像的各二维唇形参数，进而由表情参数获取单元将所述脸部图像的各二维唇形参数和所述脸部图像的各表情基进行规则化解算，建立所述脸部图像的各二维唇形参数到所述脸部图像的各表情基的权重参数的映射，进而，再进一步由网格渲染单元利用所述脸部图像的各表情基的权重参数，结合网格纹理获取单元所获取的该目标对象的脸部图像的三维脸部网格和脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像，使得图像融合单元和视频合成单元能够根据各渲染脸部图像和模板视频的各视频帧图像获得融合视频，无需人工参与，从而有效提高了视频生成的效率。

[0182] 另外，采用本申请所提供的技术方案，无需大量的三维标注数据来学习新形象的音频特征与表情基的权重参数之间的映射关系，能够进一步有效地提高视频生成的效率。

[0183] 另外，采用本申请所提供的技术方案，只需要目标对象的少量图像数据以及该图像数据所对应的音频数据，就能够完成目标对象的神经网络的训练，从而进一步有效地提高了视频生成的效率。

[0184] 另外，采用本申请所提供的技术方案，可以虚拟人物形象，通常可以根据单张或者多张目标对象的图像，获得该目标对象的脸部图像的三维脸部网格和脸部图像纹理，并通过音频生成的二维唇动所映射到的表情应用到脸部图像的三维脸部网格，从而实现虚拟人物形象的视频产生。

[0185] 另外，采用本申请所提供的技术方案，通过获取标准的三维脸部网格与所述人脸图像的三维脸部网格之间的差异，进而将该差异传递到标准的各表情基，最终生成人脸图像的各表情基，从而实现了自动化生成人脸图像的各表情基。

[0186] 另外，采用本申请所提供的技术方案，能够有效地提高用户的体验。

[0187] 根据本申请的实施例，本申请还提供了一种电子设备和一种存储有计算机指令的非瞬时计算机可读存储介质。

[0188] 如图3所示，是用来实现本申请实施例提供的视频的生成方法的电子设备的示意图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本申请的实现。

[0189] 如图3所示，该电子设备包括：一个或多个处理器301、存储器302，以及用于连接各部件的接口，包括高速接口和低速接口。各个部件利用不同的总线互相连接，并且可以被安装在公共主板上或者根据需要以其它方式安装。处理器可以对在电子设备内执行的指令进行处理，包括存储在存储器中或者存储器上以在外部输入/输出装置(诸如，耦合至接口的显示设备)上显示图形用户界面(GUI)的图形信息的指令。在其它实施方式中，若需要，可以将多个处理器和/或多条总线与多个存储器和多个存储器一起使用。同样，可以连接多个电子设备，各个设备提供部分必要的操作(例如，作为服务器阵列、一组刀片式服务器、或者多处理器系统)。图3中以一个处理器301为例。

[0190] 存储器302即为本申请所提供的非瞬时计算机可读存储介质。其中，所述存储器存

储有可由至少一个处理器执行的指令，以使所述至少一个处理器执行本申请所提供的视频的生成方法。本申请的非瞬时计算机可读存储介质存储计算机指令，该计算机指令用于使计算机执行本申请所提供的视频的生成方法。

[0191] 存储器302作为一种非瞬时计算机可读存储介质，可用于存储非瞬时软件程序、非瞬时计算机可执行程序以及单元，如本申请实施例中的视频的生成方法对应的程序指令/单元(例如，附图2所示的网格纹理获取单元201、唇形参数获取单元202、表情参数获取单元203、网格渲染单元204、图像融合单元205和视频合成单元206)。处理器301通过运行存储在存储器302中的非瞬时软件程序、指令以及单元，从而执行服务器的各种功能应用以及数据处理，即实现上述方法实施例中的视频的生成方法。

[0192] 存储器302可以包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需要的应用程序；存储数据区可存储根据实现本申请实施例提供的视频的生成方法的电子设备的使用所创建的数据等。此外，存储器302可以包括高速随机存取存储器，还可以包括非瞬时存储器，例如至少一个磁盘存储器件、闪存器件、或其他非瞬时固态存储器件。在一些实施例中，存储器302可选包括相对于处理器301远程设置的存储器，这些远程存储器可以通过网络连接至实现本申请实施例提供的视频的生成方法的电子设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0193] 视频的生成方法的电子设备还可以包括：输入装置303和输出装置304。处理器301、存储器302、输入装置303和输出装置304可以通过总线或者其他方式连接，图3中以通过总线连接为例。

[0194] 输入装置303可接收输入的数字或字符信息，以及产生与实现本申请实施例提供的视频的生成方法的电子设备的用户设置以及功能控制有关的键信号输入，例如触摸屏、小键盘、鼠标、轨迹板、触摸板、指示杆、一个或者多个鼠标按钮、轨迹球、操纵杆等输入装置。输出装置304可以包括显示设备、辅助照明装置(例如，LED)和触觉反馈装置(例如，振动电机)等。该显示设备可以包括但不限于，液晶显示器(LCD)、发光二极管(LED)显示器和等离子体显示器。在一些实施方式中，显示设备可以是触摸屏。

[0195] 此处描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、专用专用集成电路(ASIC)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0196] 这些计算程序(也称作程序、软件、软件应用、或者代码)包括可编程处理器的机器指令，并且可以利用高级过程和/或面向对象的编程语言、和/或汇编/机器语言来实施这些计算程序。如本文使用的，术语“机器可读介质”和“计算机可读介质”指的是用于将机器指令和/或数据提供给可编程处理器的任何计算机程序产品、设备、和/或装置(例如，磁盘、光盘、存储器、可编程逻辑装置(PLD))，包括，接收作为机器可读信号的机器指令的机器可读介质。术语“机器可读信号”指的是用于将机器指令和/或数据提供给可编程处理器的任何信号。

[0197] 为了提供与用户的交互，可以在计算机上实施此处描述的系统和技术，该计算机具有：用于向用户显示信息的显示装置(例如，阴极射线管(CRT)或者液晶显示器(LCD)监视器)；以及键盘和指向装置(例如，鼠标或者轨迹球)，用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈(例如，视觉反馈、听觉反馈、或者触觉反馈)；并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0198] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如，作为数据服务器)、或者包括中间件部件的计算系统(例如，应用服务器)、或者包括前端部件的计算系统(例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如，通信网络)来将系统的部件相互连接。通信网络的示例包括：局域网(LAN)、广域网(WAN)和互联网。

[0199] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。

[0200] 根据本申请实施例的技术方案，通过基于待生成视频中目标对象的音频内容，获得该音频内容所对应的所述目标对象的脸部图像的各二维唇形参数，进而将所述脸部图像的各二维唇形参数和所述脸部图像的各表情基进行规则化解算，建立所述脸部图像的各二维唇形参数到所述脸部图像的各表情基的权重参数的映射，进而，再进一步利用所述脸部图像的各表情基的权重参数，结合该目标对象的脸部图像的三维脸部网格和脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像，使得能够根据各渲染脸部图像和模板视频的各视频帧图像获得融合视频，无需人工参与，从而有效提高了视频生成的效率。

[0201] 另外，采用本申请所提供的技术方案，无需大量的三维标注数据来学习新形象的音频特征与表情基的权重参数之间的映射关系，能够进一步有效地提高视频生成的效率。

[0202] 另外，采用本申请所提供的技术方案，只需要目标对象的少量图像数据以及该图像数据所对应的音频数据，就能够完成目标对象的神经网络的训练，从而进一步有效地提高了视频生成的效率。

[0203] 另外，采用本申请所提供的技术方案，可以虚拟人物形象，通常可以根据单张或者多张目标对象的图像，获得该目标对象的脸部图像的三维脸部网格和脸部图像纹理，并通过音频生成的二维唇动所映射到的表情应用到脸部图像的三维脸部网格，从而实现虚拟人物形象的视频产生。

[0204] 另外，采用本申请所提供的技术方案，通过获取标准的三维脸部网格与所述人脸图像的三维脸部网格之间的差异，进而将该差异传递到标准的各表情基，最终生成人脸图像的各表情基，从而实现了自动化生成人脸图像的各表情基。

[0205] 另外，采用本申请所提供的技术方案，能够有效地提高用户的体验。

[0206] 应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本发申请中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本申请公开的技术方案所期望的结果，本文在此不进行限制。

[0207] 上述具体实施方式，并不构成对本申请保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本申请的精神和原则之内所作的修改、等同替换和改进等，均应包含在本申请保护范围之内。

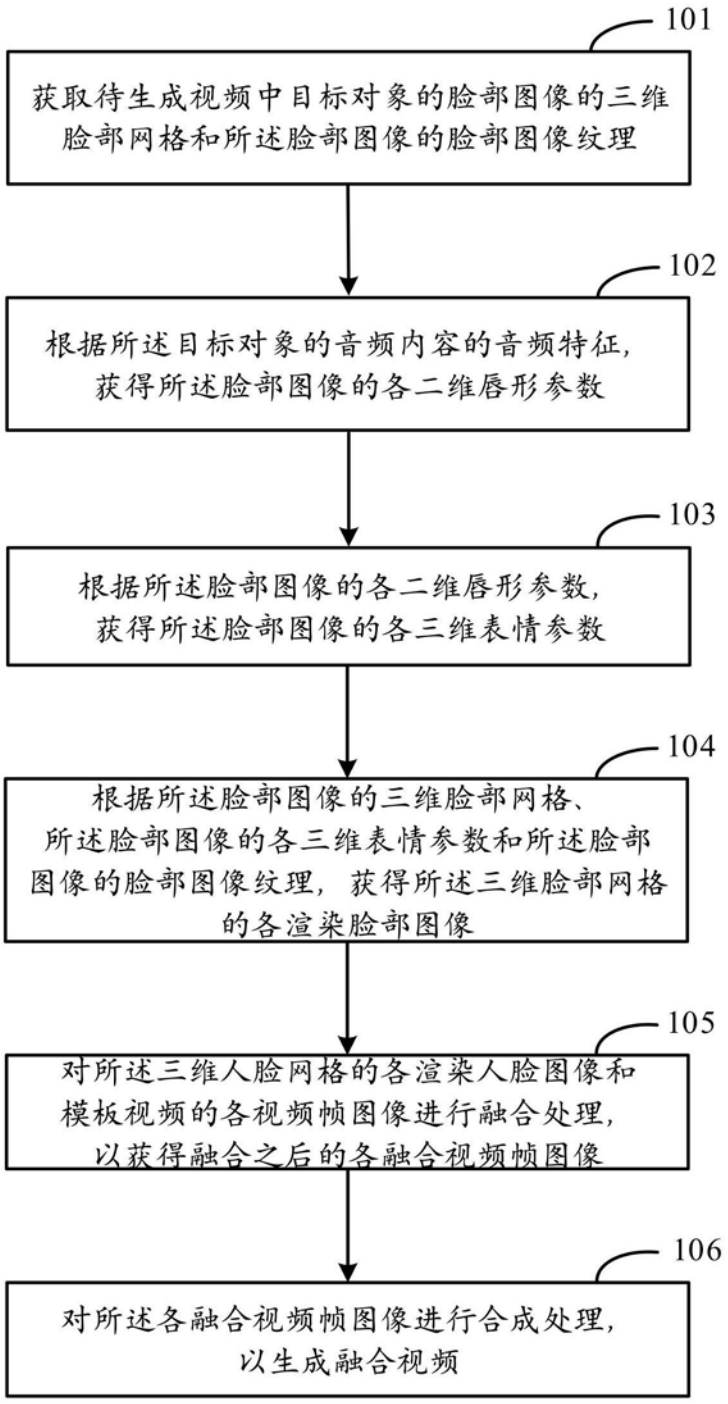

101
获取待生成视频中目标对象的脸部图像的三维脸部网格和所述脸部图像的脸部图像纹理
102
根据所述目标对象的音频内容的音频特征，获得所述脸部图像的各二维唇形参数
103
根据所述脸部图像的各二维唇形参数，获得所述脸部图像的各三维表情参数
104
根据所述脸部图像的三维脸部网格、所述脸部图像的各三维表情参数和所述脸部图像的脸部图像纹理，获得所述三维脸部网格的各渲染脸部图像
105
对所述三维人脸网格的各渲染人脸图像和模板视频的各视频帧图像进行融合处理，以获得融合之后的各融合视频帧图像
106
对所述各融合视频帧图像进行合成处理，以生成融合视频

图1

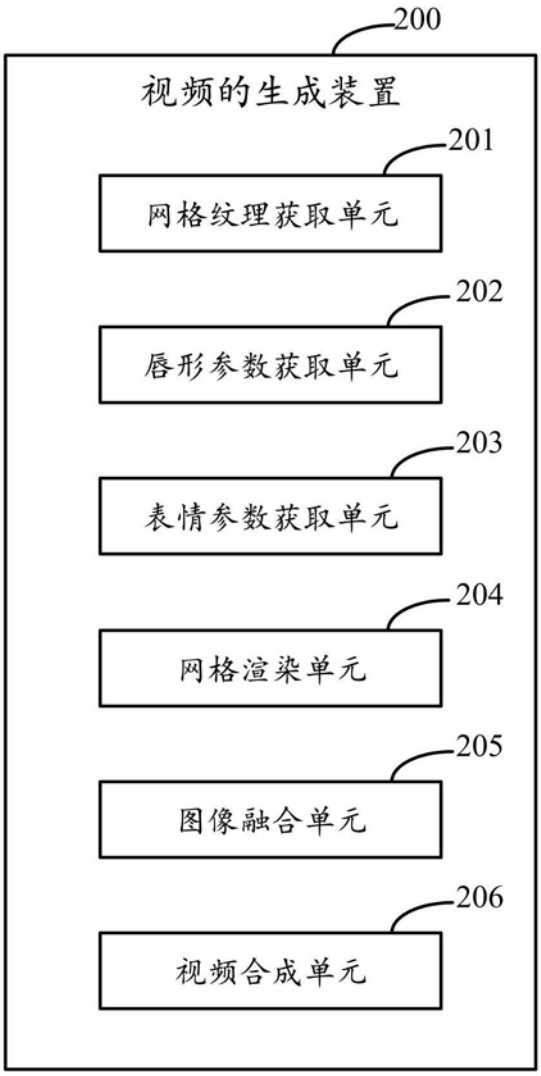

200
视频的生成装置
201
网格纹理获取单元
202
唇形参数获取单元
203
表情参数获取单元
204
网格渲染单元
205
图像融合单元
206
视频合成单元

图2

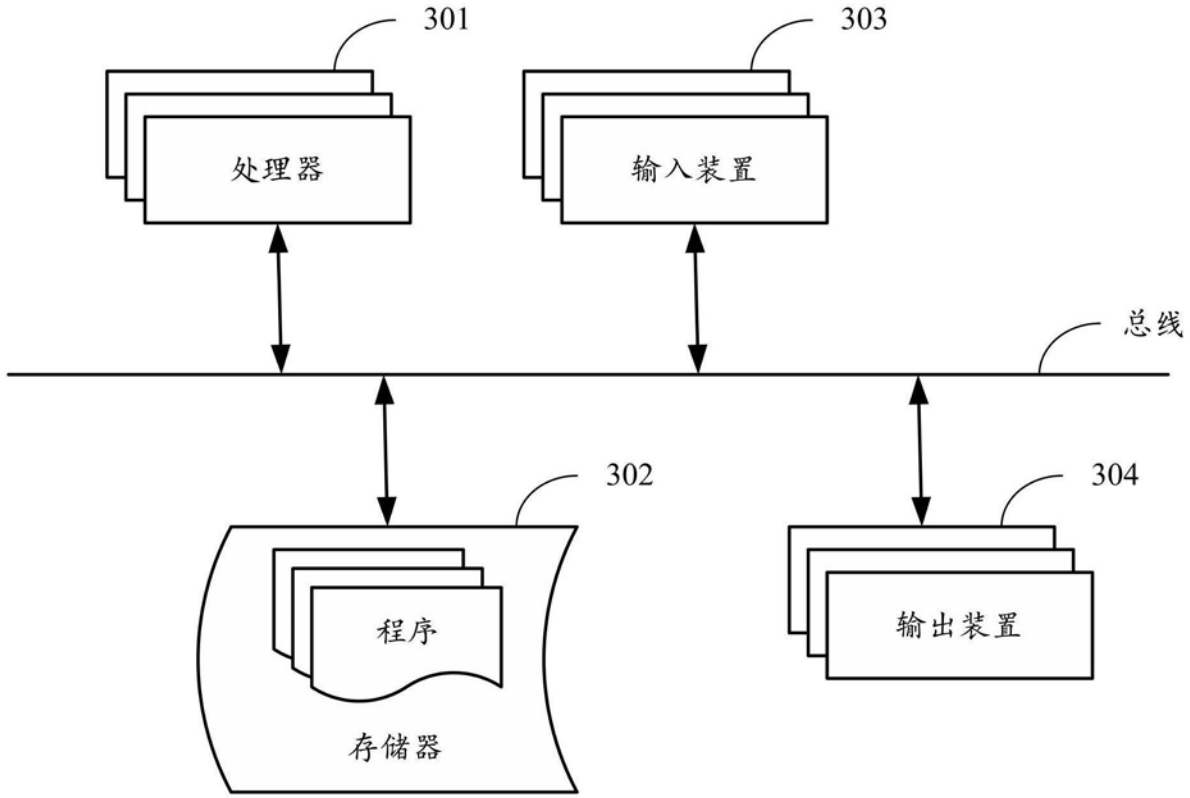


图3