

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 1 月 9 日 (09.01.2020)



(10) 国际公布号
WO 2020/007185 A1

- (51) 国际专利分类号:
H04N 5/232 (2006.01)
- (21) 国际申请号: PCT/CN2019/091359
- (22) 国际申请日: 2019 年 6 月 14 日 (14.06.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810723144.8 2018年7月4日 (04.07.2018) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。
- (72) 发明人: 周景锦 (ZHOU, Jingjin); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 程培 (CHENG, Pei); 中

国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 傅斌 (FU, Bin); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 高雨 (GAO, Yu); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。

(74) 代理人: 北京三高永信知识产权代理有限责任公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,

(54) Title: IMAGE PROCESSING METHOD AND APPARATUS, STORAGE MEDIUM AND COMPUTER DEVICE

(54) 发明名称: 图像处理方法、装置、存储介质和计算机设备

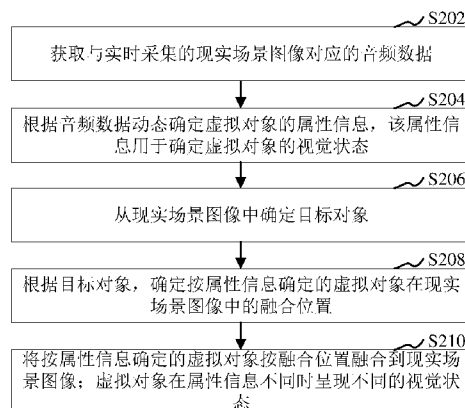


图 2

- S202 Acquire audio data corresponding to a real scene image acquired in real time
- S204 Dynamically determine attribute information concerning a virtual object according to the audio data, the attribute information being used for determining a visual state of the virtual object
- S206 Determine a target object from the real scene image
- S208 According to the target object, determine a fusion position, in the real scene image, of the virtual object determined according to the attribute information
- S210 Fuse the virtual object determined according to the attribute information into the real scene image according to the fusion position; the virtual object presents different visual states when the attribute information is different

(57) Abstract: The present application relates to an image processing method and apparatus, a storage medium and a computer device. The method comprises: acquiring audio data corresponding to a real scene image acquired in real time; dynamically determining attribute information concerning a virtual object according to the audio data, the attribute information being used for determining a visual state of the virtual object; determining a target object from the real scene image; according to the target object, determining a fusion position, in the real scene image, of the virtual object determined according to the attribute information; fusing the virtual object determined according to the attribute information into the real scene image according to the fusion position; the virtual object presenting different visual states when the attribute information is different. The solution provided in the present application can improve the interactivity between a user and a virtual object.

MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 本申请涉及一种图像处理方法、装置、存储介质和计算机设备, 方法包括: 获取与实时采集的
现实场景图像对应的音频数据; 根据音频数据动态确定虚拟对象的属性信息, 属性信息用于确定虚拟对象
的视觉状态; 从现实场景图像中确定目标对象; 根据目标对象, 确定按属性信息确定的虚拟对象在现实场
景图像中的融合位置; 将按属性信息确定的虚拟对象按融合位置融合到现实场景图像; 虚拟对象在属性信
息不同时呈现不同的视觉状态。本申请提供的方案可以提升用户与虚拟对象之间的交互性。

图像处理方法、装置、存储介质和计算机设备

本申请要求于 2018 年 7 月 4 日提交、申请号为 201810723144.8、发明名称为“图像处理方法、装置、存储介质和计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及图像处理技术领域，特别是涉及一种图像处理方法、装置、存储介质和计算机设备。

背景技术

随着图像处理技术和计算机技术的快速发展，出现了各式各样的用于录制视频的客户端，比如用户可通过客户端录制虚拟和现实相结合的视频。

目前在录制视频时，用户可在客户端的录制界面中自由选取虚拟对象，将这些虚拟对象添加到视频对应图像帧的相应位置，使得虚拟对象可跟随图像帧中运动目标的移动而移动。

然而，通过上述方式所录制的视频，虚拟对象仅能够跟随运动目标的移动而移动，交互性差。

发明内容

基于此，提供一种图像处理方法、装置、存储介质和计算机设备，能够解决虚拟对象仅能够跟随运动目标的移动而移动，使交互性差的技术问题。

一种图像处理方法，其特征在于，应用于图像处理系统，所述方法包括：

获取与实时采集的现实场景图像对应的音频数据；

根据所述音频数据动态确定虚拟对象的属性信息，所述属性信息用于确定所述虚拟对象的视觉状态；

从所述现实场景图像中确定目标对象；

根据所述目标对象，确定按所述属性信息确定的虚拟对象在所述现实场景图像中的融合

位置；

将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像；所述虚拟对象在属性信息不同时呈现不同的视觉状态。

一种图像处理装置，其特征在于，包括：

音频数据获取模块，用于获取与实时采集的现实场景图像对应的音频数据；

属性信息确定模块，用于根据所述音频数据动态确定虚拟对象的属性信息，所述属性信息用于确定所述虚拟对象的视觉状态；

目标对象确定模块，用于从所述现实场景图像中确定目标对象；

融合位置确定模块，用于根据所述目标对象，确定按所述属性信息确定的虚拟对象在所述现实场景图像中的融合位置；

融合模块，用于将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像；所述虚拟对象在属性信息不同时呈现不同的视觉状态。

一种存储介质，其特征在于，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行上述的图像处理方法的步骤。

一种计算机设备，其特征在于，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如下操作：

获取与实时采集的现实场景图像对应的音频数据；

根据所述音频数据动态确定虚拟对象的属性信息，所述属性信息用于确定所述虚拟对象的视觉状态；

从所述现实场景图像中确定目标对象；

根据所述目标对象，确定按所述属性信息确定的虚拟对象在所述现实场景图像中的融合位置；

将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像；所述虚拟对象在属性信息不同时呈现不同的视觉状态。

上述的图像处理方法、装置、存储介质和计算机设备，获取与实时采集的现实场景图像对应的音频数据，通过音频数据动态确定虚拟对象的属性，从而实现了根据音频数据对虚拟对象属性的控制。通过现实场景图像中的目标对象，确定虚拟对象在现实场景图像中的融合

位置，按照融合位置将根据属性信息确定的虚拟对象融合到现实场景图像中，由于虚拟对象的属性信息由音频数据控制，当音频数据发生变化时，融合到现实场景图像中的虚拟对象的属性信息也随之产生变化，提高了交互性。

附图说明

图 1 为一个实施例中应用图像处理方法的系统结构图；

图 2 为一个实施例中图像处理方法的流程示意图；

图 3 为一个实施例中将虚拟对象融合到现实场景图像的示意图；

图 4 为一个实施例中将虚拟对象融合到现实场景图像的示意图；

图 5 为一个实施例中根据音频数据确定属性信息，将具有该属性信息的虚拟对象融合到现实场景图像的示意图；

图 6 为一个实施例中根据音频数据的参数值确定虚拟对象的属性信息的步骤的流程示意图；

图 7 为一个实施例中对音频数据进行抽样、量化和编码的示意图；

图 8 为一个实施例中根据编码音频数据确定频率值的步骤的流程示意图；

图 9 为一个实施例中按照属性调整量调整虚拟对象，并将调整后的虚拟对象融合到现实场景图像的步骤的流程示意图；

图 10 为一个实施例中按照属性变化目标值调整虚拟对象，并将调整后的虚拟对象融合到现实场景图像的步骤的流程示意图；

图 11 为一个实施例中根据目标对象的特征确定虚拟对象在现实场景图像的融合位置的步骤的流程示意图；

图 12 为一个实施例中目标对象脸部特征点的示意图；

图 13 为另一个实施例中图像处理方法的流程示意图；

图 14 为一个实施例中图像处理装置的结构框图；

图 15 为一个实施例中计算机设备的结构框图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

图 1 为一个实施例中图像处理方法的应用环境图。参照图 1，该图像处理方法应用于图像处理系统。该图像处理系统可以是一个终端或多个终端的组合，该终端可以是智能手机、电脑或其它可支持 AR（Augmented Reality，增强现实）技术的设备。如图 1 所示，图像处理系统可以包括：摄像头、场景产生器、图像合成器和显示器；其中：

摄像头，用于获取目标对象对应环境的现实场景图像，将获取到的现实场景图像发送至图像合成器，以与增强现实模型的虚拟对象进行合成操作。

场景产生器，用于根据现实场景图像中目标对象的位置信息确定虚拟对象的融合位置，如通过分析目标对象的特征确定虚拟对象的融合位置，然后将虚拟对象发送至图像合成器。

图像合成器，用于将关于目标对象的现实场景图像和虚拟对象，按照融合位置进行融合，将融合的结果输出至显示屏。

显示器，用于将图像合成器输出的融合图像进行显示，形成目标对象和用于增强现实模型的虚拟对象共同显示的效果。

如图2所示，在一个实施例中，提供了一种图像处理方法。本实施例以该方法应用于上述图1中的终端来举例说明。参照图2，该图像处理方法具体包括如下步骤：

S202，获取与实时采集的现实场景图像对应的音频数据。

其中，现实场景可以指用户通过某种媒介所观看到的现实画面，在该现实画面中至少包括以下之一：人物、自然风景、人文风景和人类的智慧作品等。人类的智慧作品指的是人类通过劳动和智慧所创造的作品。

举例来说，现实场景为用户通过裸眼所观看到的人与自然的画面，或者用户通过 3D 眼镜所观看到的立体电影的画面。现实场景图像可以通过终端采集到的关于现实场景的图像。例如，现实场景图像为通过图 1 中的摄像头实时采集到的现实场景的图像。终端采集多个现实场景图像后，将多个现实场景图像按照采集时间进行组合，可得到视频。

音频数据为时域的音频信号，音频数据中携带有语音、音乐和音效的有规律的声波的频

率、幅度变化信息等。根据声波的特征，可把音频数据分类为规则音频和不规则音频。其中规则音频又可以分为语音、音乐和音效。音频数据的来源可以是终端从外界采集所得，也可以是从与现实场景图像的背景音频中读取所得。

在一个实施例中，S202 具体可以包括：在实时采集现实场景图像时，终端从当前环境中实时采集与现实场景图像对应的音频数据；或者，终端从实时采集的现实场景图像的背景音频中，读取与现实场景图像所对应时间戳对应的音频数据。

例如，在实时采集现实场景图像时，终端通过麦克风采集当前环境中的音乐或说话的声音。或者，由于在开发过程中，开发人员设置在实时采集的现实场景图像时自动播放背景音乐，终端实时采集现实场景图像时会生成现实场景图像对应的时间戳，在自动播放的背景音乐中读取与现实场景图像所对应时间戳对应的音频数据。或者，在实时采集的现实场景图像时，音乐客户端同时在播放音乐，终端则实时采集现实场景图像时会生成现实场景图像对应的时间戳，在播放的音乐中读取与现实场景图像所对应时间戳对应的音频数据。

S204，根据音频数据动态确定虚拟对象的属性信息，该属性信息用于确定虚拟对象的视觉状态。

其中，虚拟对象可以包括图像素材，该图像素材如静态贴纸或动态贴纸等。此外，虚拟对象还可以包括用于增强目标对象显示效果的虚拟道具。例如，虚拟道具可以是用于装扮目标对象的各种虚拟挂件和虚拟背景。

虚拟对象可以包括一个或多个属性，每个属性具有一个属性信息，该属性信息用于确定虚拟对象的视觉状态。

虚拟对象的属性信息可以包括以下至少一种：属性调整量和属性变化目标值。属性调整量包括虚拟对象的缩放比例、旋转角度和偏移量等。属性变化目标值包括虚拟对象的颜色 RGB 值等。属性调整量用于表示虚拟对象的属性的调整幅度，根据属性调整量能够对虚拟对象相应的属性进行调整，确定调整后的属性。属性变化目标值用于表示虚拟对象的属性的目标属性值，根据属性变化目标值能够对虚拟对象相应的属性进行调整，以使调整后的属性为该属性变化目标值。

需要说明的是，所确定虚拟对象的属性信息为实时采集现实场景图像当下的属性信息，则对于不同的现实场景图像，所确定虚拟对象的属性信息可能会不同。例如，假设用户拍摄

一个具有 AR 效果的视频,若视频中的第 i 帧现实场景图像为当前的现实场景图像,第 i 帧现实场景图像对应时刻的虚拟对象的属性信息为当前的属性信息。视频中的第 $i-1$ 帧现实场景图像为上一时刻的现实场景图像,第 $i-1$ 帧现实场景图像对应时刻的虚拟对象的属性信息为上一时刻的属性信息。其中, i 为大于等于 1 的正整数。

在一个实施例中,终端根据音频数据动态确定虚拟对象的属性信息,根据确定的属性信息调整虚拟对象,获得按属性信息确定的虚拟对象。

在一个实施例中,音频数据具有音量值、频率值和音色等参数值。终端根据音频数据动态确定虚拟对象的属性信息。当终端将虚拟对象的原始属性信息更新为所确定的属性信息时,便可实现对虚拟对象的调整。例如,终端根据音频数据的频率值来调整虚拟对象的缩放比例,或调整虚拟对象的颜色 RGB 值。

S206,从现实场景图像中确定目标对象。

其中,目标对象可以是自然界中具有生命的物体,如人类、动物、植物等。

在一个实施例中,S206 具体可以包括:从现实场景图像中识别生物特征;当生物特征满足预设条件时,则将现实场景图像中与生物特征对应的生物对象确定为目标对象。

其中,生物特征可以是生物的轮廓特征,也可以是生物的细节特征,如人的脸部特征等。

在一个实施例中,预设条件可以包括预设生物特征或生物特征的预设完整度阈值。终端从现实场景图像中识别生物特征;当生物特征满足预设生物特征,和/或生物特征的完整度达到预设完整度阈值时,则将现实场景图像中与生物特征对应的对象确定为目标对象。

即,当生物特征满足预设生物特征时,则将现实场景图像中与生物特征对应的对象确定为目标对象。或者,当生物特征的完整度达到预设完整度阈值时,则将现实场景图像中与生物特征对应的对象确定为目标对象。或者,当生物特征满足预设生物特征,且生物特征的完整度达到预设完整度阈值时,则将现实场景图像中与生物特征对应的对象确定为目标对象。

S208,根据目标对象,确定按属性信息确定的虚拟对象在现实场景图像中的融合位置。

其中,融合位置可以指:在融合到现实场景图像时,虚拟对象的中心点或关键位置点在现实场景图像中所处的位置。融合位置可以是一个位置点,也可以是多个位置点。针对不同的虚拟对象或同一虚拟对象的不同部位,可以确定不同的融合位置,例如不同的虚拟对象可以融合在目标对象的不同部位,同一个虚拟对象的不同部位可以融合在目标对象的不同部位。

在一个实施例中，终端检测目标对象的特征，从目标对象的特征中选取与虚拟对象匹配的特征，根据所选取的特征确定按属性信息确定的虚拟对象在现实场景图像中的融合位置。

其中，虚拟对象匹配的特征可以根据虚拟对象的类型确定，不同类型的虚拟对象可以匹配相同的特征或不同的特征。作为一个示例，可以设置多种类型的虚拟对象，并为每种类型的虚拟对象设置匹配的特征，从而建立虚拟对象与特征之间的匹配关系，后续即可根据该匹配关系确定虚拟对象匹配的特征。例如 AR 潜水眼镜匹配的特征为眼部特征，AR 兔子装扮的虚拟道具匹配的特征为嘴部特征。

根据所选取的特征确定按属性信息确定的虚拟对象在现实场景图像中的融合位置，可以包括：将所选取的特征所在的位置确定为虚拟对象在现实场景图像中的融合位置。

作为一个示例，如图 3 所示，当虚拟对象为 AR 潜水眼镜的虚拟道具时，终端检测目标对象的特征，从检测到的特征中确定眼部特征，根据眼部特征确定虚拟对象的融合位置为用户的眼部位置。

作为另一个示例，如图 4 所示，当虚拟对象为 AR 兔子装扮的虚拟道具时，对于 AR 兔子装扮中的兔牙虚拟道具，终端检测目标对象的特征，从检测到的特征中确定嘴部特征，根据嘴部特征确定虚拟对象的融合位置为用户的嘴部位置。对于 AR 兔子装扮中的耳朵虚拟道具，终端检测目标对象的特征，从检测到的特征中确定头部特征（如头发），根据头部特征确定虚拟对象的融合位置为用户的头部位置。

或者，对于同一虚拟对象来说，该虚拟对象中可以包括多个部位，则不同的部位可以匹配不同的特征。作为一个示例，当虚拟对象为雨衣虚拟道具时，雨衣虚拟道具中的头部匹配的特征为头部特征，雨衣虚拟道具中的胳膊匹配的特征为胳膊特征。

在一个实施例中，终端按属性信息调整虚拟对象，从而使虚拟对象进行尺寸的缩放、或更换颜色 RGB 值、或改变旋转角度等。需要说明的是，终端按属性信息调整的虚拟对象，是以上一帧融合后的现实场景图像中的虚拟对象为参考，从而实现对当前虚拟对象进行缩放、或更换颜色 RGB 值、或改变旋转角度等操作。

S210，将按属性信息确定的虚拟对象按融合位置融合到现实场景图像；虚拟对象在属性信息不同时呈现不同的视觉状态。

在一个实施例中，终端确定虚拟对象的中心点或关键位置点，将虚拟对象的中心点或关

键位置点融合至与融合位置匹配的区域，从而将虚拟对象融合到现实场景图像中。其中，中心点或关键位置点用于确定虚拟对象所要融合的区域，用于与对应的融合位置匹配，即在融合时，将虚拟对象的中心点置于对应的融合位置，或者将虚拟对象的关键位置点置于对应的融合位置。

如图 3 所示，p 表示虚拟对象的关键位置点，该关键位置点为 AR 潜水眼镜的虚拟道具的三个位置。在将 AR 潜水眼镜的虚拟道具融合到现实场景图像中时，根据图中 3 中的 p 点将 AR 潜水眼镜的虚拟道具对准目标对象的两眼和鼻子，从而实现 AR 潜水眼镜的虚拟道具的准确融合，具体效果如图 3 (b) 所示。

如图 4 所示，q 表示虚拟对象的关键位置点，分别位于 AR 兔子装扮的兔牙虚拟道具中的上部，以及，两个 AR 兔子装扮的耳朵虚拟道具的下部。在将 AR 兔子装扮的虚拟道具融合到现实场景图像时，根据关键位置点将 AR 兔子装扮的虚拟道具融合到目标对象的嘴唇部位和头部，从而实现 AR 兔子装扮的虚拟道具的准确融合，具体效果如图 4 (b) 所示。

虚拟对象在属性信息不同时呈现不同的视觉状态，举例来说，假设用户拍摄一个具有 AR 效果的视频，若视频中的第 i 帧现实场景图像为当前的现实场景图像，第 i 帧现实场景图像对应时刻的虚拟对象的属性信息为当前的属性信息。视频中的第 i-1 帧现实场景图像为上一时刻的现实场景图像，第 i-1 帧现实场景图像对应时刻的虚拟对象的属性信息为上一时刻的属性信息。其中，i 为大于等于 1 的正整数。当各帧现实场景图像和对应属性信息的虚拟对象按照时间组合，那么可获得一个具有 AR 效果的视频，在该视频中，虚拟对象的属性信息随着音频数据的变化而发生变化。如音频数据的音量值或频率值发生变化，虚拟对象的尺寸、或颜色 RGB 值、或方位随之发生变化。

作为一个示例，如图 5 所示，假设图 5 (a) 为上一时刻融合后的现实场景图像，m 为原始尺寸的虚拟对象。当所获取的音频数据的参数值发生变化，如音量大小发生变化，或频率值发生变化，那么，对应的虚拟对象的尺寸也发生变化，变化后的虚拟对象如图 5 (b) 中的 n 所示，n 为尺寸放大的虚拟对象。

上述实施例中，获取与实时采集的现实场景图像对应的音频数据，通过音频数据动态确定虚拟对象的属性，从而实现了根据音频数据对虚拟对象属性的控制。通过现实场景图像中的目标对象，确定虚拟对象在现实场景图像中的融合位置，按照融合位置将根据属性信息确

定的虚拟对象融合到现实场景图像中，由于虚拟对象的属性信息由音频数据控制，当音频数据发生变化时，融合到现实场景图像中的虚拟对象的属性信息也随之产生变化，提高了交互性。

在一个实施例中，如图 6 所示，S204 具体可以包括：

S602，获取音频数据的参数值。

其中，音频数据的参数值包括音频数据的音量值、频率值和音色等。音量值可以是以下任一种：平均频率值、最大音量值或最小音量值。频率值可以是以下任一种：平均频率值、最大频率值或最小频率值。

具体地，终端通过分析音频数据，获得音频数据中的音量值、频率值和音色等参数值。

在一个实施例中，S602 具体可以包括：对音频数据进行抽样；将抽样所得的结果量化并编码，获得编码音频数据；根据所获得的编码音频数据确定音频数据的参数值。

具体地，终端采用 PCM（Pulse Code Modulation，脉冲编码调制）的方式，对所采集的时间上连续的时域音频数据进行抽样、量化和编码，获得二进制的编码音频数据。终端根据所获得的编码音频数据确定音频数据的参数值，如确定音频数据的音量值。

作为一个示例，如图 7 所示，对音频数据 $U(t)$ 进行抽样，使时间上连续的时域音频数据离散化。将离散化的音频数据进行量化，获得 M 进制的量化音频数据，其中， M 为大于 2 的正整数。将量化后的音频数据进行编码，得到二进制的编码音频数据。

S604，确定参数值与虚拟对象的属性信息之间的预设映射关系。

在一个实施例中，不同类型的参数值对应不同的预设映射关系，终端根据参数值的类型，确定参数值与虚拟对象的属性信息之间的预设映射关系。其中，类型包括音量类型、频率类型和音色类型。与类型对应的，参数值包括音量值、频率值和音色。

例如，以参数值为音量值，以属性信息为缩放比例为例，终端确定音量与虚拟对象的缩放比例之间的预设映射关系，该预设映射关系可以是函数关系式，如下所示：

$$f(x) = \begin{cases} 1.0, & x < 50 \\ 1.0 + 3.0 \times \frac{x-50}{70}, & x \geq 50 \end{cases}$$

其中， x 为音频数据的音量值，范围可以是 0~120 分贝（db）， $f(x)$ 为虚拟对象的缩放比例。

S606, 根据预设映射关系, 将参数值映射为虚拟对象的属性信息。

在一个实施例中, 终端确定预设映射关系时, 将参数值作为变量输入该预设映射关系, 得到虚拟对象的对应属性信息。例如, 当音量值为 40db 时, 将音量值 40 输入函数 $f(x)$, 得到的属性信息为 1 的缩放比例, 即对原始的虚拟对象不做任何放大或缩小处理。又例如, 当音量值为 120db 时, 将音量值 120 输入函数 $f(x)$, 得到的属性信息为 4 的缩放比例, 即将原始的虚拟对象放大到原来的 4 倍。可以看出, 虚拟对象的属性信息随着音量数据的参数值变化而变化, 从而根据音量数据的参数值实现对虚拟对象的控制, 从而使虚拟对象呈现不同的视觉状态。

上述实施例中, 确定参数值与虚拟对象的属性信息之间的预设映射关系, 当获得对应的参数值的大小时, 便可以通过该预设映射关系获得对应的虚拟对象的属性信息, 从而实现对虚拟对象的调整, 使虚拟对象呈现不同的视觉状态, 提升了虚拟对象的多样化变化。

在一个实施例中, 虚拟对象的属性信息可以通过音频数据的参数值确定, 该参数值可以是频率值或音量值, 当参数值为频率值时; 如图 8 所示, 上述根据所获得的编码音频数据确定音频数据的参数值的步骤, 具体可以包括:

S802, 将时域的编码音频数据转换为频域音频数据。

在一个实施例中, 编码音频数据为离散化的音频信号。终端根据离散傅里叶变换, 将编码音频数据转换为频域音频数据。其中, 频域音频数据包含有音频数据的幅值 (即音量值)、频率值和相位。

S804, 将频域音频数据分段, 获得多个子频域音频数据。

在一个实施例中, 终端按照设定的步长, 对频域音频数据进行分段, 获得多个子频域音频数据。例如, 使用 512 点傅里叶变换时, 最多可以将 0 到截止频率 (若抽样率 48kHz 时, 截止频率为 24kHz) 的频段等分为 256 个频段, 然后执行 S806, 即确定各频段内的振幅。

在一个实施例中, 终端按照不等分的方式, 将频域音频数据划分为不等长的多个频段, 获得多个子频域音频数据。

S806, 确定各子频域音频数据的振幅。

在一个实施例中, 各子频域音频数据中包含了幅值、频率值和相位, 终端确定各子频域音频数据中的振幅, 从而获取到各子频域音频数据中的音量值。

S808, 从各子频域音频数据中选取振幅最大的子频域音频数据。

其中, 音频数据的振幅大表示音频数据的功率较大, 对于所获取的音频数据而言, 当音频数据的功率大, 对应的有用信号多, 当音频数据的功率小, 对应的有用信号少。例如, 终端通过麦克风采集音频数据, 当所采集的音频数据的功率小, 说明当前所采集的音频数据可能为噪声信号, 因此, 可以选取振幅最大的子频域音频数据。

具体地, 当确定各子频域音频数据的振幅时, 将各子频域音频数据之间的振幅进行比较, 获取振幅最大的子频域音频数据。

在一个实施例中, 终端对各子频域音频数据按照振幅的大小进行排列, 从排列的各子频域音频数据中选取振幅最大的子频域音频数据。

S810, 按照所选取的子频域音频数据确定音频数据对应的频率值。

上述实施例中, 对频域音频数据进行分段, 可以通过分段后所得子频域音频数据确定频率值, 一方面, 通过该频率值可以实现对虚拟对象的调整, 另一方面, 分段后的频域音频数据, 在计算过程中可降低计算量, 加快计算速率。

在一个实施例中, 虚拟对象的属性信息可以通过音频数据的参数值确定, 该参数值可以是频率值或音量值, 当参数值为音量值时, 根据所获得的编码音频数据确定音频数据的参数值的步骤, 具体可以包括: 根据所获得的编码音频数据确定音量值; 或者, 将时域的编码音频数据转换为频域音频数据; 根据频域音频数据确定音量值。

时域的编码音频数据的幅值可以表示音频数据的音量值。在一个实施例中, 终端将时域的编码音频数据的幅值, 确定为音频数据的音量值。

在一个实施例中, 编码音频数据为离散化的音频信号。终端根据离散傅里叶变换, 将编码音频数据转换为频域音频数据。其中, 频域音频数据包含有音频数据对应波形的幅值(即音量值)、频率值和相位。终端将频域音频数据中的幅值确定为音频数据的音量值。

在一个实施例中, 终端将编码音频数据转换为频域音频数据后, 按照设定的步长对频域音频数据进行分段, 获得多个子频域音频数据。终端根据各子频域音频数据确定对应的幅值, 将最大的幅值确定为音频数据的音量值, 或者, 将平均幅值确定为音频数据的音量值。

上述实施例中, 根据所获得的编码音频数据, 或者根据由时域的编码音频数据转换的频域音频数据这两种方式确定音量值, 得到用于确定虚拟对象视觉状态的属性信息, 通过该音

量值可以实现对虚拟对象的调整。

在一个实施例中，虚拟对象的属性信息可以包括以下至少一种：属性调整量和属性变化目标值。属性调整量可以包括虚拟对象的缩放比例、旋转角度和偏移量等，属性变化目标值可以包括虚拟对象的颜色 RGB 值。当属性信息为属性调整量时，上述的按属性信息确定的虚拟对象，是按属性调整量调整相应属性后的虚拟对象，如图 9 所示，S210 具体可以包括：

S902，确定虚拟对象所具有的、且与属性调整量对应的属性。

其中，属性指的是虚拟对象所拥有的属性，包括缩放、颜色、旋转和偏移等。与属性调整量对应的属性包括缩放、旋转和偏移等。对应的，属性调整量为属性对应的具体值。

在一个实施例中，终端根据音频数据的参数值，确定虚拟对象所具有的、且与属性调整量对应的属性。

S904，按照属性调整量调整虚拟对象的属性，得到调整属性后的虚拟对象。

在一个实施例中，终端根据映射关系，确定与音频数据的参数值对应的属性调整量后，按照属性调整量调整虚拟对象，使虚拟对象的属性发生变化，得到调整属性后的虚拟对象。

例如，若属性调整量为缩放比例，根据缩放比例调整虚拟对象的尺寸大小，获得调整尺寸后的虚拟对象。

S906，将调整属性后的虚拟对象按融合位置融合到现实场景图像。

上述实施例中，通过属性调整量调整虚拟对象的属性，将调整属性后的虚拟对象按融合位置融合到现实场景图像，获得随音频数据的参数变化的虚拟对象，实现根据音频数据调整虚拟对象，提升了虚拟对象的多样化变化，增强用户的体验。

在一个实施例中，虚拟对象的属性信息可以包括以下至少一种：属性调整量和属性变化目标值。属性调整量可以包括虚拟对象的缩放比例、旋转角度和偏移量等，属性变化目标值可以包括虚拟对象的颜色。当属性信息为属性变化目标值时，上述的按属性信息确定的虚拟对象，是将相应属性变化至属性变化目标值后的虚拟对象，如图 10 所示，S210 具体可以包括：

S1002，确定虚拟对象所具有的、且与属性变化目标值对应的属性。

其中，与属性变化目标值对应的属性包括虚拟对象的颜色。对应的，属性变化目标值为属性对应的具体值，如颜色 RGB 值。

在一个实施例中，终端根据音频数据的参数值，确定虚拟对象所具有的、且与属性变化目标值对应的属性。

S1004，将虚拟对象的属性变化至属性变化目标值，得到属性变化后的虚拟对象。

在一个实施例中，终端根据映射关系，确定与音频数据的参数值对应的属性变化目标值后，按照属性变化目标值调整虚拟对象，使虚拟对象的属性发生变化，得到调整属性后的虚拟对象。

例如，若属性变化目标值为目标颜色 RGB 值，根据目标颜色 RGB 值调整虚拟对象的显示颜色，使虚拟对象的原始显示颜色调整为目标颜色 RGB 值所对应的颜色，如原始为红色，通过目标颜色 RGB 值的调整后获得蓝色的虚拟对象。

S1006，将属性变化后的虚拟对象按融合位置融合到现实场景图像。

上述实施例中，通过属性变化目标值调整虚拟对象的属性，将调整属性后的虚拟对象按融合位置融合到现实场景图像，获得随音频数据的参数变化的虚拟对象，实现根据音频数据调整虚拟对象，提升了虚拟对象的多样化变化，增强用户的体验。

在一个实施例中，当属性信息为第一属性的属性调整量和第二属性的属性变化目标值时，上述的按属性信息确定的虚拟对象，是按属性调整量调整第一属性，且将相应的第二属性变化至属性变化目标值后的虚拟对象。S210 具体可以包括：确定虚拟对象所具有的、且与属性调整量对应的第一属性，确定虚拟对象所具有的、且与属性变化目标值对应的第二属性，按照属性调整量调整虚拟对象的第一属性，将虚拟对象的第二属性变化至属性变化目标值，得到属性变化后的虚拟对象，将属性变化后的虚拟对象按融合位置融合到现实场景图像。

在一个实施例中，如图 11 所示，S208 具体可以包括：

S1102，检测目标对象的特征。

具体地，终端通过特征点检测方式，如级联回归 CNN（Convolutional Neural Networks，卷积神经网络）、或 Dlib、Libfacedetect、或 Seetaface 等特征点检测方式检测目标对象的特征。

作为一个示例，如图 12 所示，为目标对象的脸部特征点的检测结果，为了描述方便，采用数字标记识别得到的各个脸部特征点，例如图 12 中所示的 1~17 表示目标对象的脸部边缘特征点，18~22 表示目标对象的左眉部特征点，23~27 表示目标对象的右眉部特征点，28~36 表示目标对象的鼻子特征点，37~42 表示目标对象的左眼特征点，43~48 表示目标对象的右眼

特征点，49~68 表示目标对象的嘴唇特征点。需要指出的是，以上仅为示例，在可选实施例中可以在以上脸部特征点中仅识别部分或更多的特征点，或采用其他方式标记各个特征点，均属于本申请实施例的范畴。

S1104，在所检测到的特征中查找与具有属性的虚拟对象匹配的特征。

不同的虚拟对象所对应现实场景图像中的融合位置也不同。如图 3 所示，AR 潜水眼镜的虚拟道具对应现实场景图像中的融合位置，应该是目标对象的眼部位置。如图 4 所示，AR 兔子装扮的虚拟道具的耳朵对应现实场景图像中的融合位置，应该是目标对象的头部位置；而 AR 兔子装扮的虚拟道具的兔牙对应现实场景图像中的融合位置，应该是目标对象的牙齿部位（或嘴唇部位）。如图 5 所示，AR 小猫装扮对应现实场景图像中的融合位置，应该是目标对象的两边脸部位置。

在一个实施例中，终端确定虚拟对象的功能或用途，根据功能或用途确定虚拟对象所要挂载在目标对象的部位，进而确定虚拟对象与目标对象的特征之间的关系。终端在所检测到的特征中，根据所确定的关系获得与具有属性的虚拟对象匹配的特征。

S1106，根据匹配的特征，确定具有属性的虚拟对象在现实场景图像中的融合位置。

上述实施例中，通过目标对象的特征点，确定虚拟对象在现实场景图像中的融合位置，以便于根据该融合位置将虚拟对象融合在现实场景图像中，获得视觉状态发生变化的虚拟对象，提升了虚拟对象的多样性变化。

在一个实施例中，该方法还可以包括：提取音频数据的音频特征；当音频特征符合第一触发条件时，则执行以下至少一种：新增虚拟对象；切换虚拟对象；切换视觉状态的类型。

其中，音频特征可以包括至少以下之一：音频数据的音量值、频率值、音色、对数功率谱和梅尔频率倒谱系数等。对数功率谱和梅尔频率倒谱系数可以反映出音频数据的功率值以及说话人的风格特征和语音表现力等特征。语音表现力可以是语音的声调、轻重和节奏等特征。对应的第一触发条件可以包括音量值达到预设音量阈值，或频率值达到预设频率阈值，或音色满足音色条件，或功率值达到功率阈值，或说话人的风格特征满足风格特征条件，或说话人的语音表现力满足表现力条件等。

视觉状态的类型可以是虚拟对象的显示尺寸、显示颜色和运动轨迹等。

在一个实施例中，终端对时域的音频数据进行分帧和加窗处理，获得各帧的音频数据。

终端对各帧的音频数据进行傅里叶变换, 获得对应的频谱。终端根据各帧的频谱计算出功率谱, 然后对该功率谱进行对数运算, 获得对数功率谱。终端可以将该对数功率谱确定为语音特征, 或者, 将对数功率谱经过离散余弦变换所得的结果确定为语音特征。

例如, 假设所采集的语音的信号表达式为 $x(n)$, 分帧和加窗后的语音为 $x'(n) = x(n) \times h(n)$, 对加窗后的语音 $x'(n) = x(n) \times h(n)$ 进行离散傅里叶变换, 得到对应的频谱信号为:

$$X(k) = \sum_{n=1}^{N-1} x(n) \times h(n) e^{-j2\pi kn/N}, 1 \leq k \leq N$$

其中, N 表示离散傅里叶变换的点数。

获得各帧语音的频谱时, 终端计算出对应的功率谱, 并求出功率谱的对数值得到对数功率谱, 从而得到对应的语音特征。

或者, 获得对数功率谱后, 终端将对数功率谱输入梅尔尺度的三角滤波器, 经离散余弦变换后得到梅尔频率倒谱系数, 所得的梅尔频率倒谱系数为:

$$C(n) = \sum_{m=1}^{N-1} x(n) \cos\left(\frac{\pi n(m-0.5)}{M}\right), n=1, 2, \dots, L$$

其中, L 阶指的是梅尔频率倒谱系数阶数, 可以取值取 12-16。 M 指的是三角滤波器个数。

上述实施例中, 通过提取音频数据的音频特征, 在音频特征满足对应的触发条件时, 可在原来虚拟对象的基础上新增虚拟对象, 或将原来的虚拟对象切换为其它虚拟对象, 或将原来所呈现的视觉状态进行切换, 使融合在现实场景图像中的虚拟对象多样化, 以及所呈现的视觉状态多样化, 进而提高了用户与虚拟对象的交互性。

在一个实施例中, 该方法还可以包括: 根据音频数据进行识别, 获得识别结果; 确定与识别结果匹配的动态效果类型; 按照动态效果类型和属性信息确定虚拟对象所呈现的视觉状态; 视觉状态与动态效果类型相匹配。

其中, 识别结果可以指音频类型和音频数据中关键字的文本特征。音频类型可以指音乐的类型, 如轻音乐、摇滚音乐和爵士音乐等音乐类型。文本特征可以指关键字的轻音和重音等。重音可通过增加音强或音高来表示。

动态效果可以是虚拟对象在动态变化过程中所显示出来的效果。具体地, 动态效果可以是以下中的任一种或多种的组合: 旋转、移动、透明与非透明之间变化和颜色变化等。例如,

虚拟对象随音频数据的变化而发生旋转，或边移动边旋转等。对应的，动态效果类型可以包括旋转类型、移动类型、透明与非透明之间变化类型和颜色变化类型。

在一个实施例中，终端识别出音频数据所对应的音乐类型时，获取与音乐类型对应的动态效果，按照获取的动态效果和属性信息确定虚拟对象所呈现的视觉状态。例如，当所获取的音频数据为摇滚音乐类型时，动态效果可以是比较动感的效果。

在一个实施例中，终端识别出音频数据中关键字的文本特征，根据识别的文本特征选择对应的动态效果。例如，当识别出音频中的关键字为重音时，将虚拟对象的动态效果切换为重音所对应的动态效果。

上述实施例中，通过音频数据的识别结果确定对应的动态效果，使虚拟对象随音频数据的变化而呈现不同的动态效果，提升了用户与虚拟对象之间的交互性。

对于传统的图像处理方案中，大部分相机/短视频类的应用程序都有动态展示虚拟对象的能力，即虚拟对象会跟随人脸的运动而运动；也都有播放背景音乐或者接收麦克风的能力，即录制视频的时候可以带有背景音乐或麦克风声音。但是目前还没有一款应用程序可以实时获取声音，进而分析声音的属性来实时调节虚拟对象的属性信息。

为了解决上述问题，本申请实施例提供了一种图像处理方法，通过该图像处理方法，可实现根据音乐变化动态的调整虚拟对象（如虚拟对象随音乐变化），使虚拟对象跟随音乐的音量值或频率值变化而发生颜色，或尺寸，或旋转角度的变化，如图 13 所示，该图像处理方法包括以下步骤：

S1302，获取音频数据。

终端获取音频数据的方式可以是：一是通过终端的麦克风采集音频数据的方式，另一是从终端播放的相应背景音乐读取音频数据的方式。通过终端的麦克风采集音频数据的方式是采集外界的音频数据，如利用手机常用的话筒功能采集用户发出的语音。从终端播放的相应背景音乐读取音频数据的方式，是终端解码所播放背景音乐的音频格式文件，从而获得音频数据。需要说明的是，可以将上述两种方式所获得的音频数据中的一种作为输入源，也可以将上述两种方式所获得的音频数据之间的混合作为输入源。终端通过 PCM 编码调制方式，将所获取的音频数据编码成二进制的编码音频数据。音频数据也可称为音频信号，在本申请实

施例中不做区分。

其中, PCM 是一种常见的编码方式, 是将模拟的音频数据按照预设的时间间隔进行抽样, 使模拟的音频数据离散化, 然后对抽样值进行量化, 将量化后的抽样值进行编码, 获得按二进制码表示抽样脉冲的幅值。

S1304, 解析音频数据, 获得对应的参数值, 参数值如频率值和音量值。

终端获得经过 PCM 编码后的编码音频数据, 从该编码音频数据中解析出与声音相关的属性信息, 该属性信息可以包括: 音量值、频率值和音色等。

音量值可以通过音频数据的响度或对应波形的幅值表示, 表征一段时间内音频数据的音量大小, 计算公式如下:

$$\bar{v} = 20 \log_{10} \left(\frac{1}{N} \sum_{i=1}^N \text{abs}(v_i) \right)$$

其中, v_i 表示经过 PCM 编码之后所得的编码音频数据中一个抽样点的振幅, N 表示抽样点的个数, 本实施例中 N 可以取值 1024, 也可以为其它数值。对于抽样率为 48k 的音频数据, 1 秒钟可计算 48 次音量值。

频率值可以是音频数据在单位时间内对应波形上下震动的次数, 单位为 Hz。音频数据可以被分解为不同频率值、不同幅值的正弦波的叠加, 利用 FFT (Fast Fourier Transformation, 快速傅里叶变换) 算法可以将 PCM 编码后所得的编码音频数据转化为频域音频数据, 通过该频域音频数据可获得频率值和音量值 (即幅值)。

将 PCM 编码后所得的编码音频数据转化为频域音频数据时, 可使用 512 点 FFT, 最多可将 0 到截止频率 (对于 48kHz 抽样率, 截止频率为 24kHz) 的频段等分为 256 个频段, 并计算出每个频段的幅值, 从而得到音频数据的音量值。此外, 获取振幅最大的频段, 根据振幅最大频段中的子频域音频数据确定频数据对应的频率值。

无论是音量值的计算还是 FFT 的计算, 在终端都可以实现实时计算。

S1306, 选取对应的映射式, 将获得的参数值输入映射式。

终端获得每个时段内的频域音频数据。以 48kHz 抽样率的音频数据作为输入源为例, 1 秒钟内会计算 48 次。终端根据不同的需求, 设计不同的映射式。其中, 该映射式即为本申请实施例所述的映射关系。该映射式的输入变量为音量值或频率值, 输出为虚拟对象的属性信息, 属性信息如颜色、缩放比例、旋转角度等。以音频数据的音量值为输入, 以虚拟对象

的缩放比例为输出为例，可以设计下述分段的映射式：

$$f(x) = \begin{cases} 1.0, & x < 50 \\ 1.0 + 3.0 \times \frac{x-50}{70}, & x \geq 50 \end{cases}$$

其中， x 为音频数据的音量值，范围可以是 0~120db， $f(x)$ 为虚拟对象的缩放比例。

根据实际需求，可以配置各种不同的映射式。其中，映射式在 3 个维度内可配：1) 根据映射式的输入类型配置映射式，如输入类型为音量值或频率值；2) 根据映射式的输出类型配置映射式，输出类型可以是虚拟对象的缩放比例、颜色、旋转角度、偏移等各个不同维度的属性信息；3) 根据函数的输入与输出之间的对应关系配置映射式。

S1308，输出虚拟对象的属性信息。

根据上述映射式，当分贝值小于 50db 时，虚拟对象保持默认大小 1.0。当分贝值大于 50db 时，虚拟对象的缩放比例随分贝值的增大而增大，当分贝值为 120db 时，缩放比例为 4.0。

如图 5 所示，图 5(a) 为默认大小，即缩放比例为 1.0；图 5(b) 为缩放大小约为 2.0 时的效果。

S1310，采集现实场景图像。

终端可通过摄像头实时采集现实场景图像。其中，现实场景图像可以是摄像头实时所采集视频中的一帧图像。

S1312，检测现实场景图像中对象的特征。

终端对现实场景图像中的目标对象做特征检测，如人脸特征检测。

其中，检测的方式可以是：采用开源的 opencv 或 dlib 的人脸配准点 SDK，或使用优图、商汤等提供的人脸特征点检测 SDK 进行特征检测。

S1314，将改变属性信息的虚拟对象与现实场景图像融合。

终端调整属性信息后的虚拟对象融合到现实场景图像中目标对象的固定区域（以目标对象的某个脸部特征点位锚点），即可实现虚拟对象实时跟随人脸，并随音频数据的属性信息变化而发生变化。

以音频数据的音量值控制虚拟对象的缩放比例为例，效果是虚拟对象可以实时跟随人脸，并且虚拟对象的尺寸会随着麦克风采集到的音量值，或者背景音乐的音量值实时发生变化。

S1316, 输出包括有虚拟对象的现实场景图像。

通过上述实施例, 可以很大程度上增加自拍/短视频类 APP 的可玩性, 虚拟对象会随着音频数据的参数值变化而发生大小、或颜色、或旋转角度等变化, 增加了虚拟对象的多样化变化, 提高了用户与虚拟对象之间的交互性。

图 2 为一个实施例中图像处理方法的流程示意图。应该理解的是, 虽然图 2 的流程图中的各个步骤按照箭头的指示依次显示, 但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明, 这些步骤的执行并没有严格的顺序限制, 这些步骤可以以其它的顺序执行。而且, 图 2 中的至少一部分步骤可以包括多个子步骤或者多个阶段, 这些子步骤或者阶段并不必然是在同一时刻执行完成, 而是可以在不同的时刻执行, 这些子步骤或者阶段的执行顺序也不必然是依次进行, 而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

如图 14 所示, 在一个实施例中, 提供了一种图像处理装置, 该图像处理装置具体包括: 音频数据获取模块 1402、属性信息确定模块 1404、目标对象确定模块 1406、融合位置确定模块 1408 和融合模块 1410; 其中:

音频数据获取模块 1402, 用于获取与实时采集的现实场景图像对应的音频数据;

属性信息确定模块 1404, 用于根据音频数据动态确定虚拟对象的属性信息, 该属性信息用于确定虚拟对象的视觉状态;

目标对象确定模块 1406, 用于从现实场景图像中确定目标对象;

融合位置确定模块 1408, 用于根据目标对象, 确定按属性信息确定的虚拟对象在现实场景图像中的融合位置;

融合模块 1410, 用于将按属性信息确定的虚拟对象按融合位置融合到现实场景图像; 虚拟对象在属性信息不同时呈现不同的视觉状态。

在一个实施例中, 音频数据获取模块 1402 还用于在实时采集现实场景图像时, 从当前环境中实时采集与现实场景图像对应的音频数据; 或者, 从实时采集的现实场景图像的背景音频中, 读取与现实场景图像所对应时间戳对应的音频数据。

在一个实施例中，目标对象确定模块 1406 还用于从现实场景图像中识别生物特征；当生物特征满足预设条件时，则将现实场景图像中与生物特征对应的生物对象确定为目标对象。

上述实施例中，获取与实时采集的现实场景图像对应的音频数据，通过音频数据动态确定虚拟对象的属性，从而实现了根据音频数据对虚拟对象属性的控制。通过现实场景图像中的目标对象，确定虚拟对象在现实场景图像中的融合位置，按照融合位置将根据属性信息确定的虚拟对象融合到现实场景图像中，由于虚拟对象的属性信息由音频数据控制，当音频数据发生变化时，融合到现实场景图像中的虚拟对象的属性信息也随之产生变化，提高了交互性。

在一个实施例中，属性信息确定模块 1404 还用于获取音频数据的参数值；确定参数值与虚拟对象的属性信息之间的预设映射关系；根据预设映射关系，将参数值映射为虚拟对象的属性信息。

在一个实施例中，属性信息确定模块 1404 对音频数据进行抽样；将抽样所得的结果量化并编码，获得编码音频数据；根据所获得的编码音频数据确定音频数据的参数值。

上述实施例中，确定参数值与虚拟对象的属性信息之间的预设映射关系，当获得对应的参数值的大小时，便可以通过该预设映射关系获得对应的虚拟对象的属性信息，从而实现对虚拟对象的调整，使虚拟对象呈现不同的视觉状态，提升了虚拟对象的多样化变化。

在一个实施例中，参数值包括频率值；属性信息确定模块 1404 还用于将时域的编码音频数据转换为频域音频数据；将频域音频数据分段，获得多个子频域音频数据；确定各子频域音频数据的振幅；从各子频域音频数据中选取振幅最大的子频域音频数据；按照所选取的子频域音频数据确定音频数据对应的频率值。

上述实施例中，对频域音频数据进行分段，可以通过分段后所得子频域音频数据确定频率值，一方面，通过该频率值可以实现对虚拟对象的调整，另一方面，分段后的频域音频数据，在计算过程中可降低计算量，加快计算速率。

在一个实施例中，参数值包括音量值；属性信息确定模块 1404 还用于根据所获得的编码音频数据确定音量值；或者，将时域的编码音频数据转换为频域音频数据；根据频域音频数据确定音量值。

上述实施例中，根据所获得的编码音频数据，或者根据由时域的编码音频数据转换的频

域音频数据这两种方式确定音量值，得到用于确定虚拟对象视觉状态的属性信息，通过该音量值可以实现对虚拟对象的调整。

在一个实施例中，属性信息包括属性调整量；按属性信息确定的虚拟对象，是按属性调整量调整相应属性后的虚拟对象；融合模块 1410 还用于确定虚拟对象所具有的、且与属性调整量对应的属性；按照属性调整量调整虚拟对象的属性，得到调整属性后的虚拟对象；将调整属性后的虚拟对象按融合位置融合到现实场景图像。

上述实施例中，通过属性调整量调整虚拟对象的属性，将调整属性后的虚拟对象按融合位置融合到现实场景图像，获得随音频数据的参数变化的虚拟对象，实现根据音频数据调整虚拟对象，提升了虚拟对象的多样化变化，增强用户的体验。

在一个实施例中，属性信息包括属性变化目标值；按属性信息确定的虚拟对象，是将相应属性变化至属性变化目标值后的虚拟对象；融合模块 1410 还用于确定虚拟对象所具有的、且与属性变化目标值对应的属性；将虚拟对象的属性变化至属性变化目标值，得到属性变化后的虚拟对象；将属性变化后的虚拟对象按融合位置融合到现实场景图像。

上述实施例中，通过属性变化目标值调整虚拟对象的属性，将调整属性后的虚拟对象按融合位置融合到现实场景图像，获得随音频数据的参数变化的虚拟对象，实现根据音频数据调整虚拟对象，提升了虚拟对象的多样化变化，增强用户的体验。

在一个实施例中，融合位置确定模块 1408 还用于检测目标对象的特征；在所检测到的特征中查找与具有属性的虚拟对象匹配的特征；根据匹配的特征，确定具有属性的虚拟对象在现实场景图像中的融合位置。

上述实施例中，通过目标对象的特征点，确定虚拟对象在现实场景图像中的融合位置，以便于根据该融合位置将虚拟对象融合在现实场景图像中，获得视觉状态发生变化的虚拟对象，提升了虚拟对象的多样性变化。

图 15 示出了一个实施例中计算机设备的内部结构图。该计算机设备具体可以是图 1 中的终端。如图 15 所示，该计算机设备包括该计算机设备包括通过系统总线连接的处理器、存储器、网络接口、输入装置和显示屏。其中，存储器包括非易失性存储介质和内存储器。该计算机设备的非易失性存储介质存储有操作系统，还可存储有计算机程序，该计算机程序被处理器执行时，可使得处理器实现图像处理方法。该内存储器中也可储存有计算机程序，该

计算机程序被处理器执行时，可使得处理器执行图像处理方法。计算机设备的显示屏可以是液晶显示屏或者电子墨水显示屏，计算机设备的输入装置可以是显示屏上覆盖的触摸层，也可以是计算机设备外壳上设置的按键、轨迹球或触控板，还可以是外接的键盘、触控板或鼠标等。

本领域技术人员可以理解，图 15 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在一个实施例中，本申请提供的图像处理装置可以实现为一种计算机程序的形式，计算机程序可在如图 15 所示的计算机设备上运行。计算机设备的存储器中可存储组成该 14 装置的各个程序模块，比如，图 14 所示的音频数据获取模块 1402、属性信息确定模块 1404、目标对象确定模块 1406、融合位置确定模块 1408 和融合模块 1410。各个程序模块构成的计算机程序使得处理器执行本说明书中描述的本申请各个实施例的图像处理方法中的步骤。

例如，图 15 所示的计算机设备可以通过如图 14 所示的图像处理装置中的音频数据获取模块 1402 执行 S202。计算机设备可通过属性信息确定模块 1404 执行 S204。计算机设备可通过目标对象确定模块 1406 执行 S206。计算机设备可通过融合位置确定模块 1408 执行 S208。计算机设备可通过融合模块 1410 执行 S210。

在一个实施例中，提供了一种计算机设备，包括存储器和处理器，存储器存储有计算机程序，计算机程序被处理器执行时，使得处理器执行以下步骤：取与实时采集的现实场景图像对应的音频数据；根据音频数据动态确定虚拟对象的属性信息，该属性信息用于确定虚拟对象的视觉状态；从现实场景图像中确定目标对象；根据目标对象，确定按属性信息确定的虚拟对象在现实场景图像中的融合位置；将按属性信息确定的虚拟对象按融合位置融合到现实场景图像；虚拟对象在属性信息不同时呈现不同的视觉状态。

在一个实施例中，计算机程序被处理器执行获取与实时采集的现实场景图像对应的音频数据的步骤时，使得处理器具体执行以下步骤：在实时采集现实场景图像时，从当前环境中实时采集与现实场景图像对应的音频数据；或者，从实时采集的现实场景图像的背景音频中，读取与现实场景图像所对应时间戳对应的音频数据。

在一个实施例中，计算机程序被处理器执行根据音频数据动态确定虚拟对象的属性信息

的步骤时，使得处理器具体执行以下步骤：获取音频数据的参数值；确定参数值与虚拟对象的属性信息之间的预设映射关系；根据预设映射关系，将参数值映射为虚拟对象的属性信息。

在一个实施例中，计算机程序被处理器执行获取音频数据的参数值的步骤时，使得处理器具体执行以下步骤：对音频数据进行抽样；将抽样所得的结果量化并编码，获得编码音频数据；根据所获得的编码音频数据确定音频数据的参数值。

在一个实施例中，参数值包括频率值，计算机程序被处理器执行根据所获得的编码音频数据确定音频数据的参数值的步骤时，使得处理器具体执行以下步骤：将时域的编码音频数据转换为频域音频数据；将频域音频数据分段，获得多个子频域音频数据；确定各子频域音频数据的振幅；从各子频域音频数据中选取振幅最大的子频域音频数据；按照所选取的子频域音频数据确定音频数据对应的频率值。

在一个实施例中，参数值包括音量值；计算机程序被处理器执行根据所获得的编码音频数据确定音频数据的参数值的步骤时，使得处理器具体执行以下步骤：根据所获得的编码音频数据确定音量值；或者，将时域的编码音频数据转换为频域音频数据；根据频域音频数据确定音量值。

在一个实施例中，属性信息包括属性调整量；按属性信息确定的虚拟对象，是按属性调整量调整相应属性后的虚拟对象；计算机程序被处理器执行将按属性信息确定的虚拟对象按融合位置融合到现实场景图像的步骤时，使得处理器具体执行以下步骤：确定虚拟对象所具有的、且与属性调整量对应的属性；按照属性调整量调整虚拟对象的属性，得到调整属性后的虚拟对象；将调整属性后的虚拟对象按融合位置融合到现实场景图像。

在一个实施例中，属性信息包括属性变化目标值；按属性信息确定的虚拟对象，是将相应属性变化至属性变化目标值后的虚拟对象；计算机程序被处理器执行将按属性信息确定的虚拟对象按融合位置融合到现实场景图像的步骤时，使得处理器具体执行以下步骤：确定虚拟对象所具有的、且与属性变化目标值对应的属性；将虚拟对象的属性变化至属性变化目标值，得到属性变化后的虚拟对象；将属性变化后的虚拟对象按融合位置融合到现实场景图像。

在一个实施例中，计算机程序被处理器执行从现实场景图像中确定目标对象的步骤时，使得处理器具体执行以下步骤：从现实场景图像中识别生物特征；当生物特征满足预设条件时，则将现实场景图像中与生物特征对应的生物对象确定为目标对象。

在一个实施例中，计算机程序被处理器执行根据目标对象，确定具有属性的虚拟对象在现实场景图像中的融合位置的步骤时，使得处理器具体执行以下步骤：检测目标对象的特征；在所检测到的特征中查找与具有属性的虚拟对象匹配的特征；根据匹配的特征，确定具有属性的虚拟对象在现实场景图像中的融合位置。

在一个实施例中，提供了一种计算机可读存储介质，存储有计算机程序，计算机程序被处理器执行时，使得处理器执行以下步骤：取与实时采集的现实场景图像对应的音频数据；根据音频数据动态确定虚拟对象的属性信息，该属性信息用于确定虚拟对象的视觉状态；从现实场景图像中确定目标对象；根据目标对象，确定按属性信息确定的虚拟对象在现实场景图像中的融合位置；将按属性信息确定的虚拟对象按融合位置融合到现实场景图像；虚拟对象在属性信息不同时呈现不同的视觉状态。

在一个实施例中，计算机程序被处理器执行获取与实时采集的现实场景图像对应的音频数据的步骤时，使得处理器具体执行以下步骤：在实时采集现实场景图像时，从当前环境中实时采集与现实场景图像对应的音频数据；或者，从实时采集的现实场景图像的背景音频中，读取与现实场景图像所对应时间戳对应的音频数据。

在一个实施例中，计算机程序被处理器执行根据音频数据动态确定虚拟对象的属性信息的步骤时，使得处理器具体执行以下步骤：获取音频数据的参数值；确定参数值与虚拟对象的属性信息之间的预设映射关系；根据预设映射关系，将参数值映射为虚拟对象的属性信息。

在一个实施例中，计算机程序被处理器执行获取音频数据的参数值的步骤时，使得处理器具体执行以下步骤：对音频数据进行抽样；将抽样所得的结果量化并编码，获得编码音频数据；根据所获得的编码音频数据确定音频数据的参数值。

在一个实施例中，参数值包括频率值，计算机程序被处理器执行根据所获得的编码音频数据确定音频数据的参数值的步骤时，使得处理器具体执行以下步骤：将时域的编码音频数据转换为频域音频数据；将频域音频数据分段，获得多个子频域音频数据；确定各子频域音频数据的振幅；从各子频域音频数据中选取振幅最大的子频域音频数据；按照所选取的子频域音频数据确定音频数据对应的频率值。

在一个实施例中，参数值包括音量值；计算机程序被处理器执行根据所获得的编码音频数据确定音频数据的参数值的步骤时，使得处理器具体执行以下步骤：根据所获得的编码音

频数据确定音量值；或者，将时域的编码音频数据转换为频域音频数据；根据频域音频数据确定音量值。

在一个实施例中，属性信息包括属性调整量；按属性信息确定的虚拟对象，是按属性调整量调整相应属性后的虚拟对象；计算机程序被处理器执行将按属性信息确定的虚拟对象按融合位置融合到现实场景图像的步骤时，使得处理器具体执行以下步骤：确定虚拟对象所具有的、且与属性调整量对应的属性；按照属性调整量调整虚拟对象的属性，得到调整属性后的虚拟对象；将调整属性后的虚拟对象按融合位置融合到现实场景图像。

在一个实施例中，属性信息包括属性变化目标值；按属性信息确定的虚拟对象，是将相应属性变化至属性变化目标值后的虚拟对象；计算机程序被处理器执行将按属性信息确定的虚拟对象按融合位置融合到现实场景图像的步骤时，使得处理器具体执行以下步骤：确定虚拟对象所具有的、且与属性变化目标值对应的属性；将虚拟对象的属性变化至属性变化目标值，得到属性变化后的虚拟对象；将属性变化后的虚拟对象按融合位置融合到现实场景图像。

在一个实施例中，计算机程序被处理器执行从现实场景图像中确定目标对象的步骤时，使得处理器具体执行以下步骤：从现实场景图像中识别生物特征；当生物特征满足预设条件时，则将现实场景图像中与生物特征对应的生物对象确定为目标对象。

在一个实施例中，计算机程序被处理器执行根据目标对象，确定具有属性的虚拟对象在现实场景图像中的融合位置的步骤时，使得处理器具体执行以下步骤：检测目标对象的特征；在所检测到的特征中查找与具有属性的虚拟对象匹配的特征；根据匹配的特征，确定具有属性的虚拟对象在现实场景图像中的融合位置。

在一个实施例中，计算机程序被处理器执行时，使得处理器具体执行以下步骤：提取音频数据的音频特征；当音频特征符合第一触发条件时，则执行以下至少一种：新增虚拟对象；切换虚拟对象；切换视觉状态的类型。

在一个实施例中，计算机程序被处理器执行时，使得处理器具体执行以下步骤：根据音频数据进行识别，获得识别结果；确定与识别结果匹配的动态效果类型；按照动态效果类型和属性信息确定虚拟对象所呈现的视觉状态；视觉状态与动态效果类型相匹配。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一非易失性计算机可读取存储介质

中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权 利 要 求 书

1、一种图像处理方法，其特征在于，应用于图像处理系统，所述方法包括：

获取与实时采集的现实场景图像对应的音频数据；

根据所述音频数据动态确定虚拟对象的属性信息，所述属性信息用于确定所述虚拟对象的视觉状态；

从所述现实场景图像中确定目标对象；

根据所述目标对象，确定按所述属性信息确定的虚拟对象在所述现实场景图像中的融合位置；

将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像；所述虚拟对象在属性信息不同时呈现不同的视觉状态。

2、根据权利要求 1 所述的方法，其特征在于，所述获取与实时采集的现实场景图像对应的音频数据包括：

在实时采集现实场景图像时，从当前环境中实时采集与所述现实场景图像对应的音频数据；或者，

从实时采集的现实场景图像的背景音频中，读取与所述现实场景图像所对应时间戳对应的音频数据。

3、根据权利要求 1 所述的方法，其特征在于，所述根据所述音频数据动态确定虚拟对象的属性信息包括：

获取所述音频数据的参数值；

确定参数值与虚拟对象的属性信息之间的预设映射关系；

根据所述预设映射关系，将所述参数值映射为所述虚拟对象的属性信息。

4、根据权利要求 3 所述的方法，其特征在于，所述获取所述音频数据的参数值包括：

对所述音频数据进行抽样；

将抽样所得的结果量化并编码，获得编码音频数据；

根据所获得的编码音频数据确定所述音频数据的参数值。

5、根据权利要求 3 所述的方法，其特征在于，所述参数值包括频率值；所述根据所获

得的编码音频数据确定所述音频数据的参数值包括：

将时域的所述编码音频数据转换为频域音频数据；

将所述频域音频数据分段，获得多个子频域音频数据；

确定各子频域音频数据的振幅；

从各子频域音频数据中选取振幅最大的子频域音频数据；

按照所选取的子频域音频数据确定所述音频数据对应的频率值。

6、根据权利要求 3 所述的方法，其特征在于，所述参数值包括音量值；所述根据所获得的编码音频数据确定所述音频数据的参数值包括：

根据所获得的编码音频数据确定音量值；或者，

将时域的所述编码音频数据转换为频域音频数据；根据所述频域音频数据确定音量值。

7、根据权利要求 1 所述的方法，其特征在于，所述属性信息包括属性调整量；所述按所述属性信息确定的虚拟对象，是按所述属性调整量调整相应属性后的虚拟对象；

所述将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像包括：

确定所述虚拟对象所具有的、且与所述属性调整量对应的属性；

按照所述属性调整量调整所述虚拟对象的所述属性，得到调整属性后的虚拟对象；

将调整属性后的虚拟对象按所述融合位置融合到所述现实场景图像。

8、根据权利要求 1 所述的方法，其特征在于，所述属性信息包括属性变化目标值；所述按所述属性信息确定的虚拟对象，是将相应属性变化至所述属性变化目标值后的虚拟对象；

所述将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像包括：

确定所述虚拟对象所具有的、且与所述属性变化目标值对应的属性；

将所述虚拟对象的所述属性变化至所述属性变化目标值，得到属性变化后的虚拟对象；

将属性变化后的虚拟对象按所述融合位置融合到所述现实场景图像。

9、根据权利要求 1 至 8 任一项所述的方法，其特征在于，所述从所述现实场景图像中确定目标对象包括：

从所述现实场景图像中识别生物特征；

当所述生物特征满足预设条件时，则

将所述现实场景图像中与所述生物特征对应的生物对象确定为目标对象。

10、根据权利要求 1 至 8 任一项所述的方法，其特征在于，所述根据所述目标对象，确定具有所述属性的虚拟对象在所述现实场景图像中的融合位置包括：

检测所述目标对象的特征；

在所检测到的特征中查找与具有所述属性的虚拟对象匹配的特征；

根据所述匹配的特征，确定具有所述属性的虚拟对象在所述现实场景图像中的融合位置。

11、根据权利要求 1 至 8 任一项所述的方法，其特征在于，还包括：

提取所述音频数据的音频特征；

当所述音频特征符合第一触发条件时，则执行以下至少一种：

新增虚拟对象；

切换虚拟对象；

切换所述视觉状态的类型。

12、根据权利要求 1 至 8 任一项所述的方法，其特征在于，还包括：

根据所述音频数据进行识别，获得识别结果；

确定与所述识别结果匹配的动态效果类型；

按照所述动态效果类型和所述属性信息确定所述虚拟对象所呈现的视觉状态；所述视觉状态与所述动态效果类型相匹配。

13、一种图像处理装置，其特征在于，包括：

音频数据获取模块，用于获取与实时采集的现实场景图像对应的音频数据；

属性信息确定模块，用于根据所述音频数据动态确定虚拟对象的属性信息，所述属性信息用于确定所述虚拟对象的视觉状态；

目标对象确定模块，用于从所述现实场景图像中确定目标对象；

融合位置确定模块，用于根据所述目标对象，确定按所述属性信息确定的虚拟对象在所述现实场景图像中的融合位置；

融合模块，用于将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像；所述虚拟对象在属性信息不同时呈现不同的视觉状态。

14、根据权利要求 13 所述的装置，其特征在于，所述音频数据获取模块还用于在实时

采集现实场景图像时，从当前环境中实时采集与所述现实场景图像对应的音频数据；或者，从实时采集的现实场景图像的背景音频中，读取与所述现实场景图像所对应时间戳对应的音频数据。

15、根据权利要求 14 所述的装置，其特征在于，属性信息确定模块还用于获取所述音频数据的参数值；确定参数值与虚拟对象的属性信息之间的预设映射关系；根据所述预设映射关系，将所述参数值映射为虚拟对象的属性信息。

16、一种存储介质，其特征在于，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如权利要求 1 至 12 中任一项所述方法的步骤。

17、一种计算机设备，其特征在于，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行以下步骤：

获取与实时采集的现实场景图像对应的音频数据；

根据所述音频数据动态确定虚拟对象的属性信息，所述属性信息用于确定所述虚拟对象的视觉状态；

从所述现实场景图像中确定目标对象；

根据所述目标对象，确定按所述属性信息确定的虚拟对象在所述现实场景图像中的融合位置；

将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像；所述虚拟对象在属性信息不同时呈现不同的视觉状态。

18、根据权利要求 17 所述的计算机设备，其特征在于，所述计算机程序被所述处理器执行获取与实时采集的现实场景图像对应的音频数据的步骤时，使得所述处理器执行以下步骤：

在实时采集现实场景图像时，从当前环境中实时采集与所述现实场景图像对应的音频数据；或者，

从实时采集的现实场景图像的背景音频中，读取与所述现实场景图像所对应时间戳对应的音频数据。

19、根据权利要求 17 所述的计算机设备，其特征在于，所述计算机程序被所述处理器执行根据音频数据动态确定虚拟对象的属性信息的步骤时，使得所述处理器执行以下步骤：

获取所述音频数据的参数值；

确定参数值与虚拟对象的属性信息之间的预设映射关系；

根据所述预设映射关系，将所述参数值映射为所述虚拟对象的属性信息。

20、根据权利要求 19 所述的计算机设备，其特征在于，所述计算机程序被所述处理器执行获取所述音频数据的参数值的步骤时，使得所述处理器执行以下步骤：

对所述音频数据进行抽样；

将抽样所得的结果量化并编码，获得编码音频数据；

根据所获得的编码音频数据确定所述音频数据的参数值。

21、根据权利要求 19 所述的计算机设备，其特征在于，所述参数值包括频率值；所述计算机程序被所述处理器执行根据所获得的编码音频数据确定所述音频数据的参数值的步骤时，使得所述处理器执行以下步骤：

将时域的所述编码音频数据转换为频域音频数据；

将所述频域音频数据分段，获得多个子频域音频数据；

确定各子频域音频数据的振幅；

从各子频域音频数据中选取振幅最大的子频域音频数据；

按照所选取的子频域音频数据确定所述音频数据对应的频率值。

22、根据权利要求 19 所述的计算机设备，其特征在于，所述参数值包括音量值；所述计算机程序被所述处理器执行根据所获得的编码音频数据确定所述音频数据的参数值的步骤时，使得所述处理器执行以下步骤：

根据所获得的编码音频数据确定音量值；或者，

将时域的所述编码音频数据转换为频域音频数据；根据所述频域音频数据确定音量值。

23、根据权利要求 17 所述的计算机设备，其特征在于，所述属性信息包括属性调整量；所述按所述属性信息确定的虚拟对象，是按所述属性调整量调整相应属性后的虚拟对象；

所述计算机程序被所述处理器执行将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像的步骤时，使得所述处理器执行以下步骤：

确定所述虚拟对象所具有的、且与所述属性调整量对应的属性；

按照所述属性调整量调整所述虚拟对象的所述属性，得到调整属性后的虚拟对象；

将调整属性后的虚拟对象按所述融合位置融合到所述现实场景图像。

24、根据权利要求 17 所述的计算机设备，其特征在于，所述属性信息包括属性变化目标值；所述按所述属性信息确定的虚拟对象，是将相应属性变化至所述属性变化目标值后的虚拟对象；

所述计算机程序被所述处理器执行将按所述属性信息确定的虚拟对象按所述融合位置融合到所述现实场景图像的步骤时，使得所述处理器执行以下步骤：

确定所述虚拟对象所具有的、且与所述属性变化目标值对应的属性；

将所述虚拟对象的所述属性变化至所述属性变化目标值，得到属性变化后的虚拟对象；

将属性变化后的虚拟对象按所述融合位置融合到所述现实场景图像。

25、根据权利要求 17 至 24 任一项所述的计算机设备，其特征在于，所述计算机程序被所述处理器执行从所述现实场景图像中确定目标对象的步骤时，使得所述处理器执行以下步骤：

从所述现实场景图像中识别生物特征；

当所述生物特征满足预设条件时，则

将所述现实场景图像中与所述生物特征对应的生物对象确定为目标对象。

26、根据权利要求 17 至 24 任一项所述的计算机设备，其特征在于，所述计算机程序被所述处理器执行根据所述目标对象，确定具有所述属性的虚拟对象在所述现实场景图像中的融合位置的步骤时，使得所述处理器执行以下步骤：

检测所述目标对象的特征；

在所检测到的特征中查找与具有所述属性的虚拟对象匹配的特征；

根据所述匹配的特征，确定具有所述属性的虚拟对象在所述现实场景图像中的融合位置。

27、根据权利要求 17 至 24 任一项所述的计算机设备，其特征在于，所述计算机程序被所述处理器执行时，使得所述处理器执行以下步骤：

提取所述音频数据的音频特征；

当所述音频特征符合第一触发条件时，则执行以下至少一种：

新增虚拟对象；

切换虚拟对象；

切换所述视觉状态的类型。

28、根据权利要求 17 至 24 任一项所述的计算机设备，其特征在于，所述计算机程序被所述处理器执行时，使得所述处理器执行以下步骤：

根据所述音频数据进行识别，获得识别结果；

确定与所述识别结果匹配的动态效果类型；

按照所述动态效果类型和所述属性信息确定所述虚拟对象所呈现的视觉状态；所述视觉状态与所述动态效果类型相匹配。

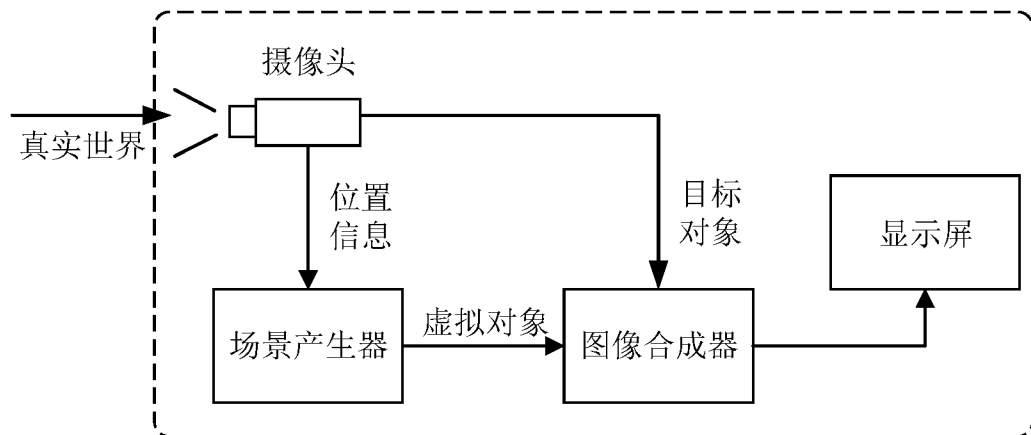


图 1

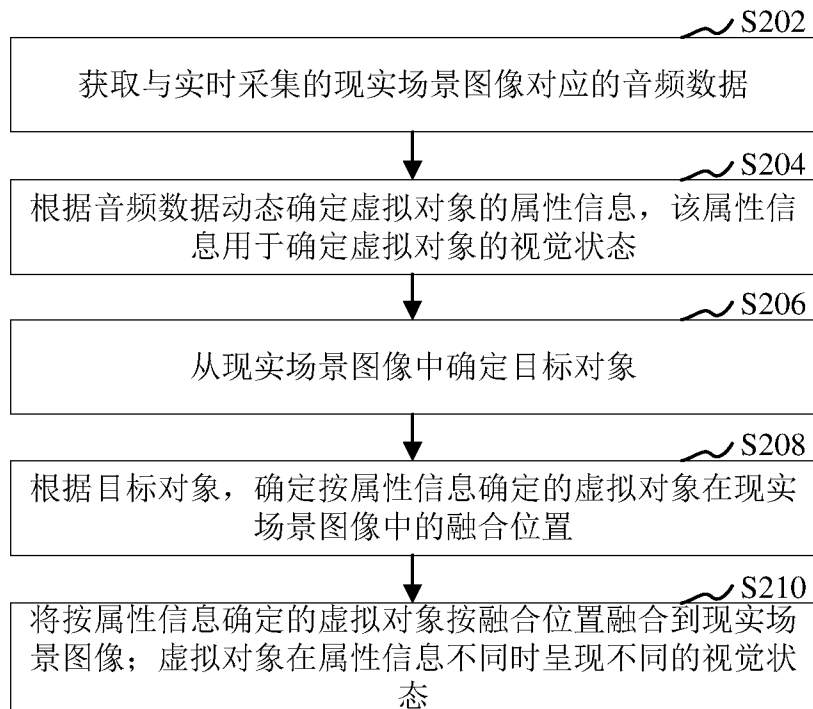


图 2

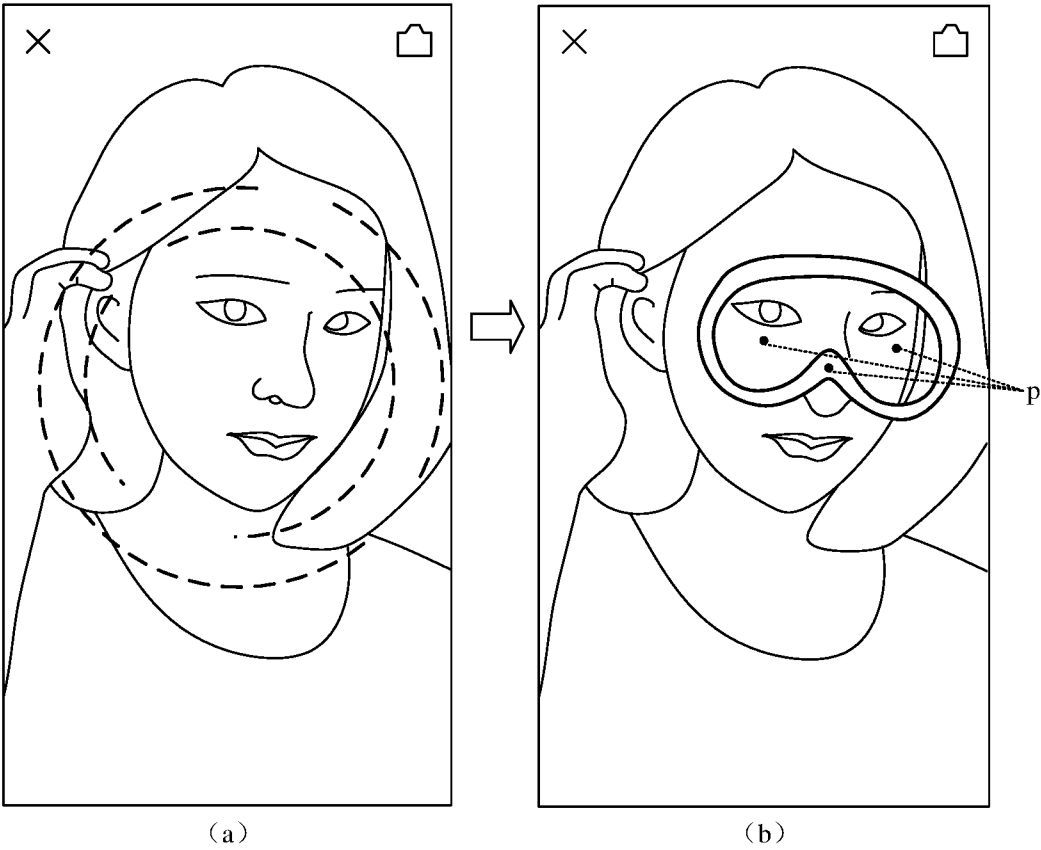


图 3

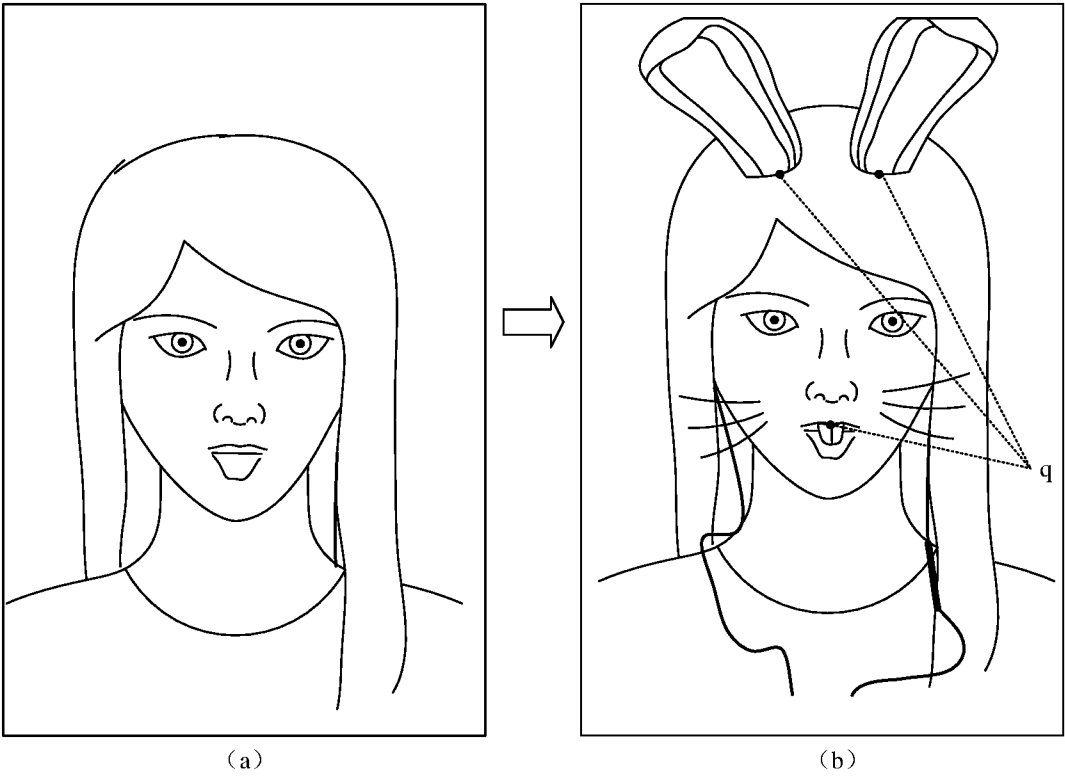


图 4

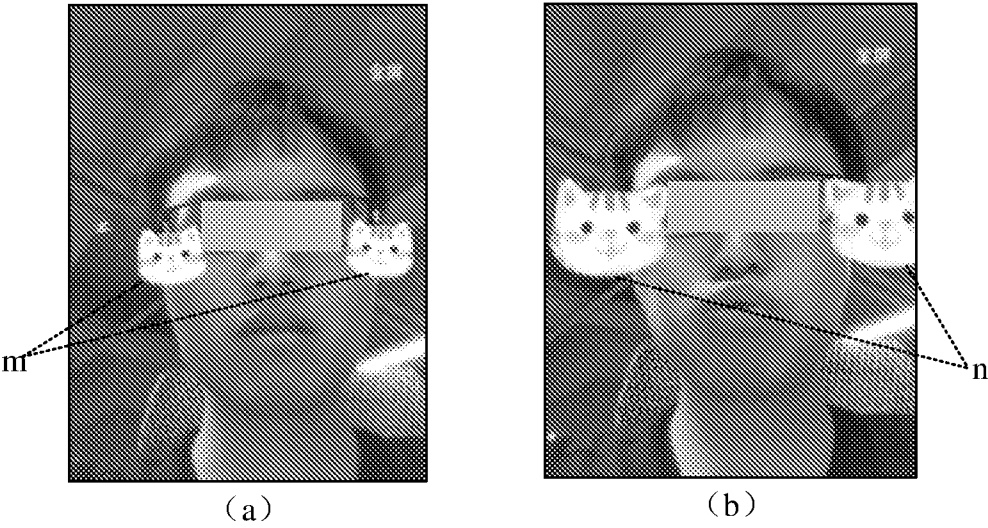


图 5

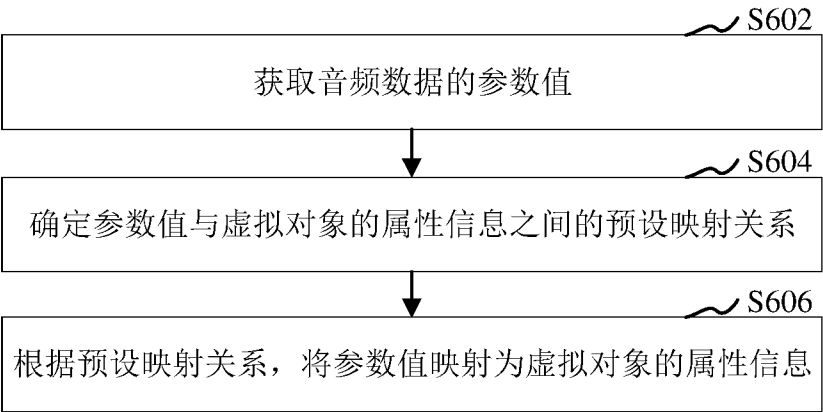


图 6

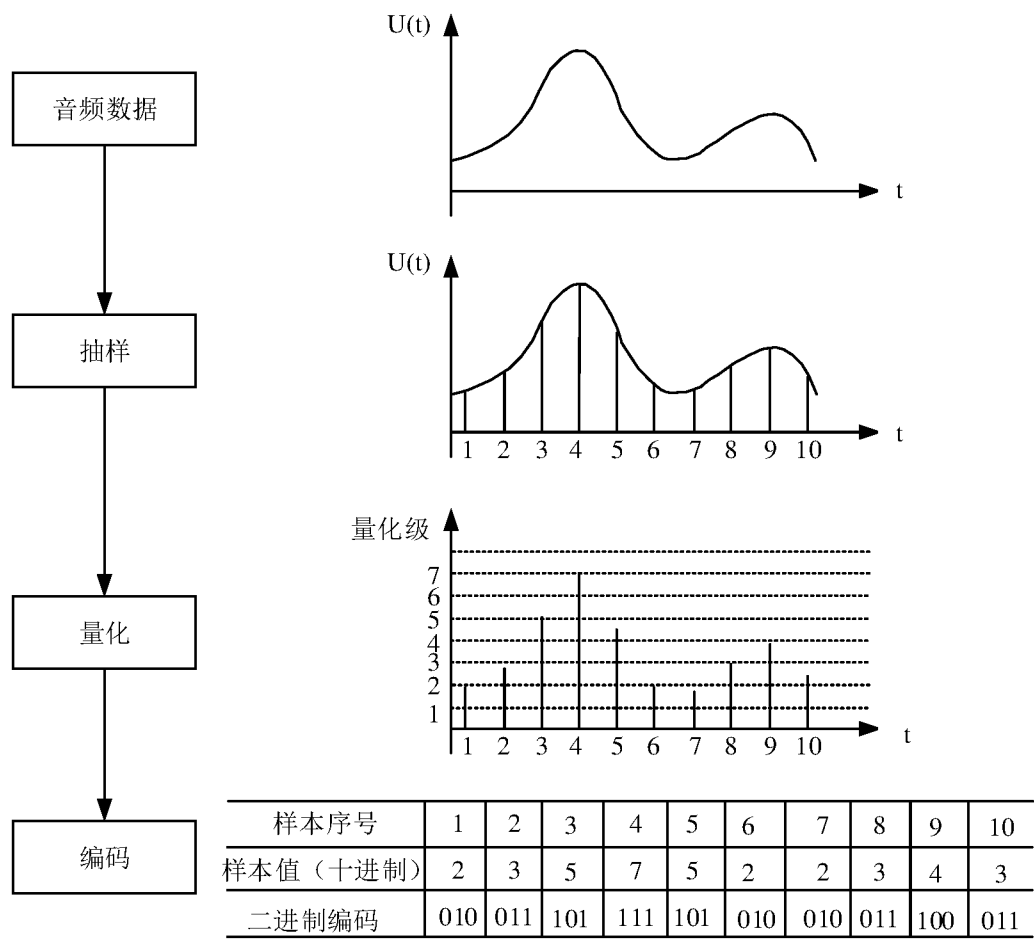


图 7

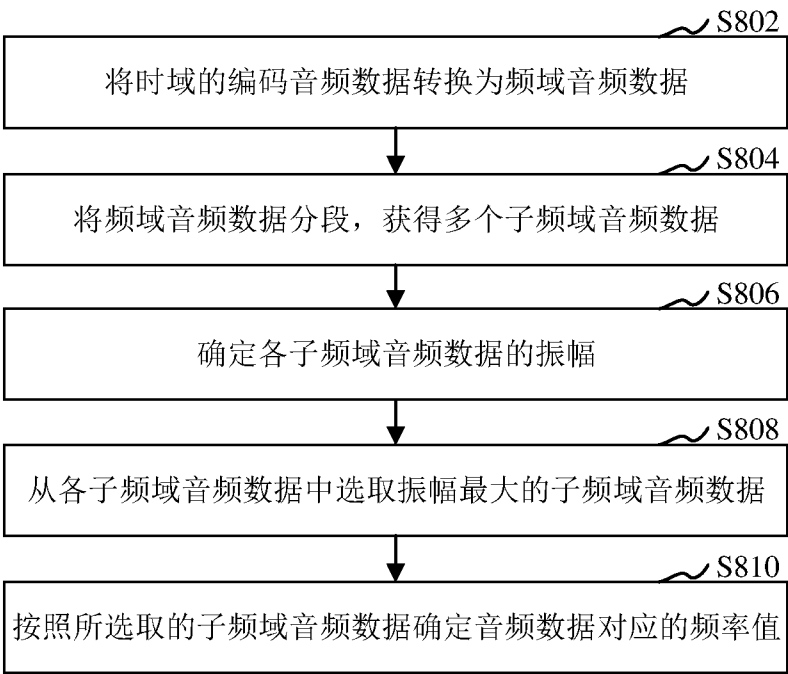


图 8

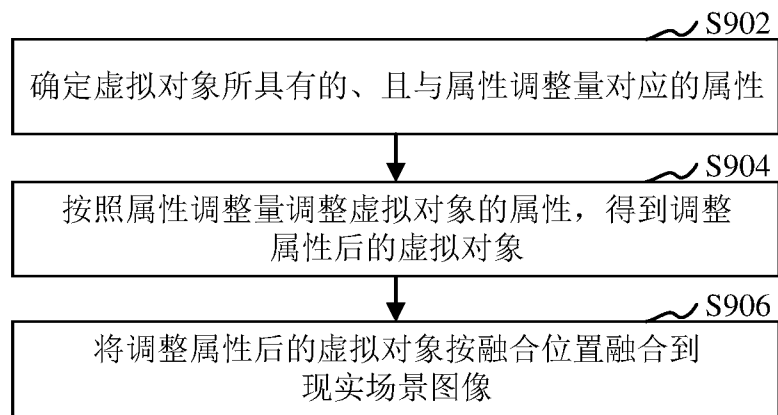


图 9

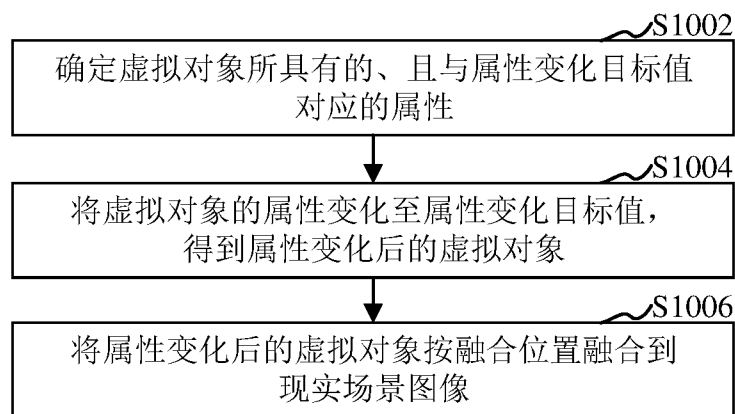


图 10

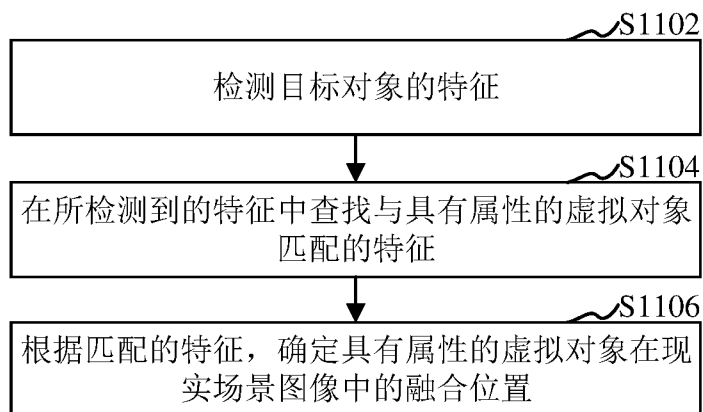


图 11

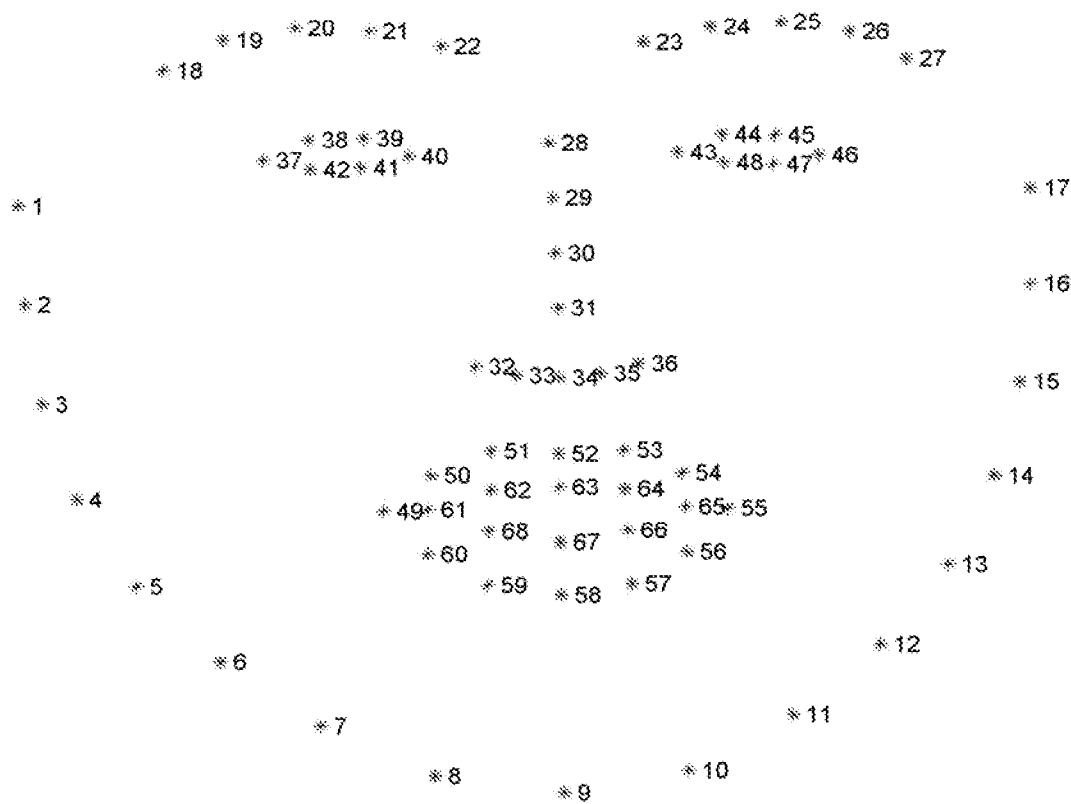


图 12

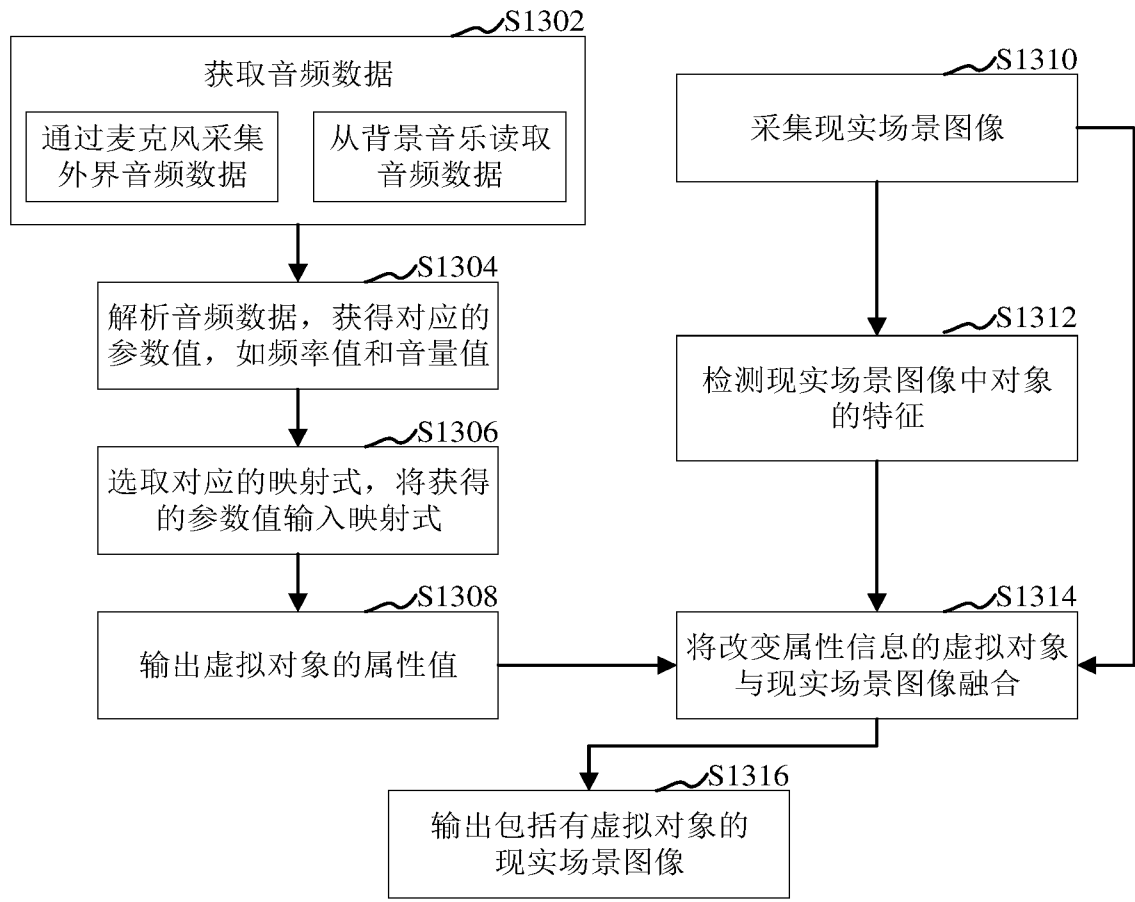


图 13

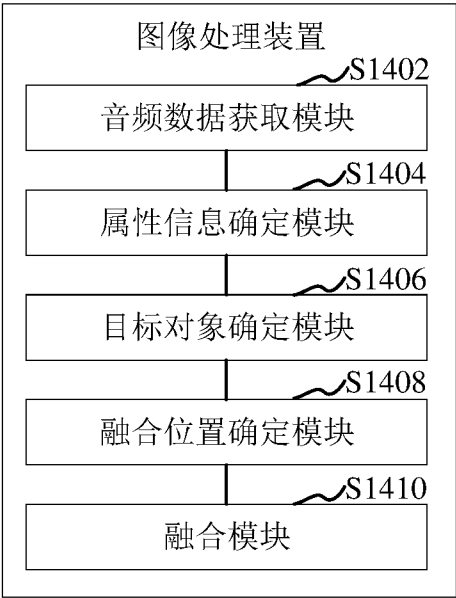


图 14

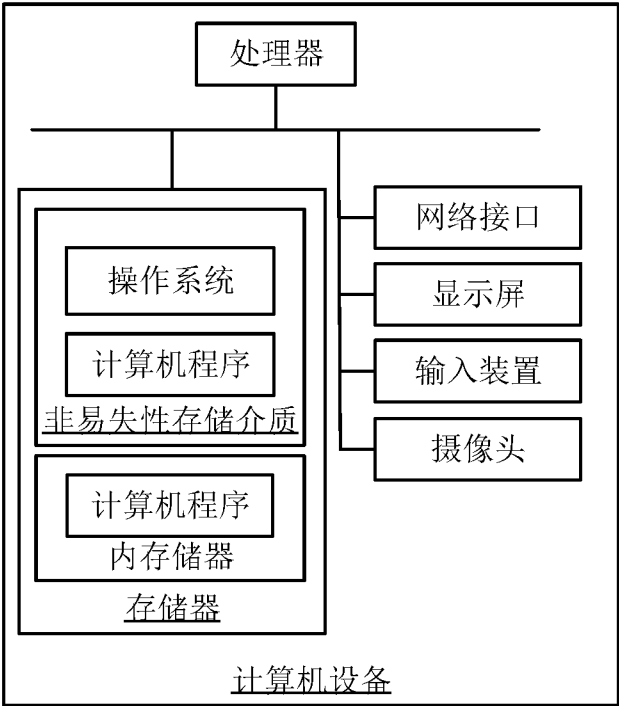


图 15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/091359

A. CLASSIFICATION OF SUBJECT MATTER

H04N 5/232(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N; G06F; G06T; G06K; G10L; H04R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNKI, CNPAT: virtual, real, stat+, time, life, attribute, audio, visual, fusion, position, 虚拟, 现实, 融合, 位置, 实时, 属性, 状态, 视觉, 目标, 音频, 视频, 图像

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 107343211 A (BEIJING SENSETIME SCIENCE AND TECHNOLOGY DEVELOPMENT CO., LTD.) 10 November 2017 (2017-11-10) description, paragraphs 0042-0175, and figures 1-7	1-28
Y	CN 107329980 A (FUJIAN STAR-NET E-VIDEO INFORMATION SYSTEM CO., LTD.) 07 November 2017 (2017-11-07) description, paragraphs 0071-0140, and figures 1-5	1-28
PX	CN 108769535 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 06 November 2018 (2018-11-06) claims 1-17, and description, paragraphs 0037-0228	1-28
A	CN 104036789 A (BEIJING ZHIGU RUITUO TECHNOLOGY SERVICE CO., LTD.) 10 September 2014 (2014-09-10) entire document	1-28
A	US 2014064517 A1 (ACER INCORPORATED) 06 March 2014 (2014-03-06) entire document	1-28

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

02 September 2019

Date of mailing of the international search report

19 September 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/091359

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107343211	A	10 November 2017	US	2018108165	A1	19 April 2018
				WO	2018033137	A1	22 February 2018
CN	107329980	A	07 November 2017	None			
CN	108769535	A	06 November 2018	None			
CN	104036789	A	10 September 2014	US	2016330512	A1	10 November 2016
				WO	2015101247	A1	09 July 2015
US	2014064517	A1	06 March 2014	EP	2706762	A2	12 March 2014
				TW	201412092	A	16 March 2014

国际检索报告

国际申请号

PCT/CN2019/091359

A. 主题的分类

H04N 5/232 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

H04N; G06F; G06T; G06K; G10L; H04R

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

WPI, EPODOC, CNKI, CNPAT: virtual, real, stat+, time, life, attribute, audio, visual, fusion, position, 虚拟, 现实, 融合, 位置, 实时, 属性, 状态, 视觉, 目标, 音频, 视频, 图像

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	CN 107343211 A (北京市商汤科技开发有限公司) 2017年 11月 10日 (2017 - 11 - 10) 说明书第0042-0175段, 附图1-7	1-28
Y	CN 107329980 A (福建星网视易信息系统有限公司) 2017年 11月 7日 (2017 - 11 - 07) 说明书第0071-0140段, 附图1-5	1-28
PX	CN 108769535 A (腾讯科技深圳有限公司) 2018年 11月 6日 (2018 - 11 - 06) 权利要求1-17, 说明书第0037-0228段	1-28
A	CN 104036789 A (北京智谷睿拓技术服务有限公司) 2014年 9月 10日 (2014 - 09 - 10) 全文	1-28
A	US 2014064517 A1 (ACER INCORPORATED) 2014年 3月 6日 (2014 - 03 - 06) 全文	1-28

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 9月 2日

国际检索报告邮寄日期

2019年 9月 19日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

郎亦虹

传真号 (86-10) 62019451

电话号码 53961711

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/091359

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	107343211	A	2017年 11月 10日	US	2018108165	A1	2018年 4月 19日
				WO	2018033137	A1	2018年 2月 22日
CN	107329980	A	2017年 11月 7日	无			
CN	108769535	A	2018年 11月 6日	无			
CN	104036789	A	2014年 9月 10日	US	2016330512	A1	2016年 11月 10日
				WO	2015101247	A1	2015年 7月 9日
US	2014064517	A1	2014年 3月 6日	EP	2706762	A2	2014年 3月 12日
				TW	201412092	A	2014年 3月 16日

表 PCT/ISA/210 (同族专利附件) (2015年1月)