(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
2017 年 9 月 21 日 (21.09.2017)

WIPO | PCT

(10) 国际公布号
WO 2017/157272 A1

(51) 国际专利分类号:
H04N 21/44 (2011.01) *G06T 7/60* (2017.01)

(21) 国际申请号: PCT/CN2017/076576

(22) 国际申请日: 2017 年 3 月 14 日 (14.03.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610143985.2 2016 年 3 月 14 日 (14.03.2016) 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: **汪倩怡 (WANG, Qianyi)**; 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)。 **戴阳刚 (DAI, Yanggang)**; 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)。 **应磊 (YING, Lei)**; 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)。 **吴发强 (WU, Faqiang)**; 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)。 **崔凌睿 (CUI, Lingrui)**; 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)。 **邬振海 (WU, Zhenhai)**; 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: **深圳市深佳知识产权代理事务所（普通合伙） (SHENPAT INTELLECTUAL PROPERTY AGENCY)**; 中国广东省深圳市国贸大厦 15 楼西座 1521 室, Guangdong 518014 (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,

[见续页]

(54) Title: INFORMATION PROCESSING METHOD AND TERMINAL

(54) 发明名称 : 一种信息处理方法及终端

(57) Abstract: Disclosed are an information processing method and a terminal. The method comprises: a terminal acquiring a first operation, so as to trigger the acquisition of first media information; during the process of acquiring the first media information, when the terminal detects an expression change, satisfying a pre-set condition, in a human face region or a user action change in an acquisition frame, reporting, to a server, an obtained variation as key information; the terminal receiving second media information pushed by the server and corresponding to the key information; and performing video synthesis on the first media information and the second media information.

(57) 摘要: 本申请公开了一种信息处理方法及终端，其中，所述方法包括：终端获取第一操作，以触发第一媒体信息的采集；终端在采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将得到的变化量作为关键信息上报给服务器；终端接收服务器推送的与所述关键信息对应的第二媒体信息；将第一媒体信息和第二媒体信息进行视频合成。

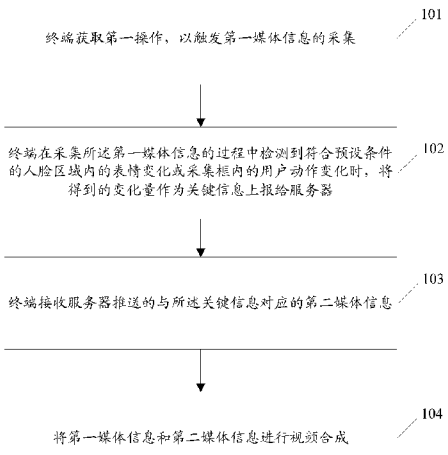

图 2

101 A terminal acquiring a first operation so as to trigger the acquisition of first media information
102 During the process of acquiring the first media information, when the terminal detects an expression change, satisfying a pre-set condition, in a human face region or a user action change in an acquisition frame, reporting, to a server, an obtained variation as key information
103 The terminal receiving second media information pushed by the server and corresponding to the key information
104 Performing video synthesis on the first media information and the second media information

WO 2017/157272 A1

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, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, 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))。

一种信息处理方法及终端

本申请要求于 2016 年 3 月 14 日提交中国专利局，申请号为 201610143985.2，发明名称为“一种信息处理方法及终端”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

本申请涉及通讯技术，尤其涉及一种信息处理方法及终端。

背景技术

用户利用终端如手机中的应用来录制视频，将录制的视频进行信息分享，是常见的信息处理方案。录制的视频还可以附加一些与视频内容相关或不相关的其他信息，来得到合成的视频。

要得到该合成的视频，需要执行附加其他信息的操作是非常复杂和繁琐的，需要用户去素材库选择这些信息，比如，用户需要去素材库选择与视频内容中某段视频信息相关的信息。即便选择的该信息是符合相关性需求的，还要进一步考虑：将该信息附加到视频内容中的哪个位置，哪个时间点等等因素，这种交互模式非常复杂，需要多次交互，势必导致处理效率低下，且来回交互往复，处理的时间成本也很高，最终达到的合成视频的效果也可能差强人意，并不符合真实的用户需求，用户可能重新再合成一次。那么，利用终端进行视频合成的信息处理成本会持续增加。然而，相关技术中，对于该问题，尚无有效解决方案。

发明内容

有鉴于此，本申请实施例提供了一种信息处理方法及终端，至少解决了现有技术存在的问题。

根据本申请的一个方面，提供了一种信息处理方法，所述方法包括：

终端获取第一操作，以触发第一媒体信息的采集；

终端在采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域

内的表情变化或采集框内的用户动作变化时，将检测到的表情变化或用户动作变化的变化量作为关键信息上报给服务器；

终端接收所述服务器推送的与所述关键信息对应的第二媒体信息；以及

将第一媒体信息和第二媒体信息进行视频合成。

根据本申请的另一方面，提供了一种终端，所述终端包括：

触发单元，用于获取第一操作，以触发第一媒体信息的采集；

检测单元，用于采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将检测到的表情变化或用户动作变化的变化量作为关键信息上报给服务器；

接收单元，用于接收所述服务器推送的与所述关键信息对应的第二媒体信息；以及

合成单元，用于将第一媒体信息和第二媒体信息进行视频合成。

根据本申请的另一方面，提供了一种非易失性存储介质，存储有程序，当所述非易失性存储介质存储的程序被包括一个或多个处理器的计算机设备执行时，可使所述计算机设备执行如上所述的信息处理方法。

本申请实施例的信息处理方法包括：终端获取第一操作，以触发第一媒体信息的采集；终端在采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将得到的变化量作为关键信息上报给服务器；终端接收服务器推送的与所述关键信息对应的第二媒体信息；将第一媒体信息和第二媒体信息按照预设配置进行视频合成。

采用本申请实施例，实时采集第一媒体信息的过程中，检测到表情变化或用户动作变化时，基于变化量从服务器得到对应的第二媒体信息，将第一媒体信息和第二媒体信息按照预设配置进行视频合成，从而在第一媒体信息采集结束后，重新回放合成的视频。在合成的视频中，在第一媒体信息的指定位置和指定时间显示有对应的第二媒体信息。由于，第二媒体信息无需用户手动选择和添加，因此，简化了操作流程，提高了处理效率；根据采集第一媒体信息的过程中得到的检测结果（如表情变化或用户动作变化）请求索取的对应第二媒体信息也更符合真实的用户需求。通过以上方法，首先第二

媒体信息的内容本身比较精准，其次第二媒体信息出现的位置和时间也能配合上如表情变化或用户动作变化等检测结果，因此，位置和时间点也精准。不仅减少了多次交互，也不需要后续再次调整和重新合成，降低了视频合成的信息处理成本和时间成本。

附图说明

图1为本申请实施例中进行信息交互的各方硬件实体的示意图；

图2为本申请实施例一的一个实现流程示意图；

图3为应用本申请实施例的应用场景示意图；

图4为应用本申请实施例的触发视频录制的示意图；

图5-6为采用现有技术的多个场景示意图；

图7-12为应用本申请实施例的多个场景示意图；

图13为本申请实施例二的一个实现流程示意图；

图14为本申请实施例三的一个组成结构示意图；

图15为本申请实施例四的一个硬件组成结构示意图；

图16为应用本申请实施例的RGB与透明度分开存储的场景示意图；以及

图17为应用本申请实施例的一个实例的系统架构图。

具体实施方式

下面结合附图对技术方案的实施作进一步的详细描述。

图 1 为本申请实施例中进行信息交互的各方硬件实体的示意图，图 1 中包括：服务器 11 和终端设备 21，22，23 和 24，其中终端设备 21，22，23 和 24 通过有线网络或者无线网络与服务器进行信息交互。终端设备可以包括手机、台式机、PC 机、一体机等。其中，终端设备中安装有满足用户日常和工作所需的多种应用。如果用户喜欢拍照和录视频，会在终端设备中安装诸如图片处理应用，视频处理应用等应用；出于社交分享的需求，也会安装社交应用。此外，还可以将运用图片处理应用和视频处理应用得到的处理结果通过社交应用进行信息分享。采用本申请实施例，基于上述图 1 所示的系统，终端设备定期从服务器获取各个应用的更新数据包在本地保存，当需要使用

终端设备上的应用，则开启应用（如视频处理应用），获取第一操作，如开启视频录制的操作，从而触发诸如视频的第一媒体信息的采集。终端设备在采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将得到的变化量作为关键信息上报给服务器。例如，该人脸区域内的表情变化可以为微笑，以及用户动作变化可以为眨眼睛或比划剪刀手。终端接收服务器推送的与所述关键信息对应的诸如贴纸的第二媒体信息；将第一媒体信息和第二媒体信息进行视频合成。采用本申请实施例，实时采集第一媒体信息的过程中，检测到表情变化或用户动作变化时，基于变化量从服务器得到对应的第二媒体信息，将第一媒体信息和第二媒体信息进行视频合成，从而在第一媒体信息采集结束后，重新回放合成的视频。在合成的视频中，在第一媒体信息的指定位置和指定时间显示有对应的第二媒体信息。由于，第二媒体信息无需用户手动选择和添加，因此，简化了操作流程，提高了处理效率；根据采集第一媒体信息的过程中得到的检测结果（如表情变化或用户动作变化）请求索取的对应第二媒体信息也更符合真实的用户需求。通过以上方法，首先第二媒体信息的内容本身比较精准，其次第二媒体信息出现的位置和时间也能配合上如表情变化或用户动作变化等检测结果，因此，位置和时间点也精准。不仅减少了多次交互，也不需要后续再次调整和重新合成，降低了视频合成的信息处理成本和时间成本。

上述图 1 的例子只是实现本申请实施例的一个系统架构实例，本申请实施例并不限于上述图 1 所述的系统结构，基于该系统架构，提出本申请各个实施例。

实施例一

本申请实施例的信息处理方法，如图 2 所示，所述方法包括:

步骤 101、终端获取第一操作，以触发第一媒体信息的采集。

一个应用场景中，如图 3 所示，用户躺在沙发上正在使用如手机 11 的终端设备。手机 11 的用户界面如图 4 所示，其中包含各种类型的应用图标，如音乐播放图标，功能设置图标，邮件收发图标等等。用户执行第一操作，如用手指点击 A1 标识的视频处理应用图标，进入视频录制的处理过程，从而触

发如视频的第一媒体信息的采集。比如，可以录制一段室内的场景，或者给自己进行自拍等等。

步骤 102、终端在采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将得到的变化量作为关键信息上报给服务器。

这里，仍然结合步骤 101 中的应用场景，在视频录制的处理过程中，通过人脸识别定位机制或表情识别机制，终端可以捕获人脸区域内的表情变化，例如，微笑，哭泣，皱眉等等。此外，终端设备还可以检测采集框（或称取景框）内的用户动作变化，例如比剪刀手。这种检测不限于人脸区域。还可以将人脸区域中的表情变化与用户动作变化进行组合识别，比如，将剪刀手和脸部表情中的微笑相结合进行组合识别。

在人脸识别的过程中，人脸识别技术是基于人的脸部特征，对视频录制中的人脸图像或者视频流进行采集，首先判断视频流中是否存在人脸，如果存在人脸，则进一步的给出脸的位置和大小，及定位出各个主要面部器官的位置信息，得到人脸中五官的各自位置和初始形态，当形态发生变化，如微笑时上下嘴唇的位置会相对初始形态产生位移和形变，则说明人脸五官的面部表情出现变化，也可以通过表情识别机制来识别出表情的变化。本申请实施例的人脸识别有别于常规的人脸识别，常规的人脸识别是为了通过构建的人脸识别系统来识别出用户的身份，是将识别出的人脸与已知人脸进行比对，以便于身份确认以及身份查找。

在表情识别过程中，可以分为四个阶段：如人脸图像的获取与预处理；人脸检测；表情特征提取；以及表情分类。如果仅仅通过人脸识别和定位机制，会存在不精确的问题，而表情识别机制是一种更加准确的处理策略。表情识别与人脸识别密切相关，如在人脸检测中的定位和人脸跟踪这些环节上是类似的，但特征提取上不同。举例来说，人脸识别提取的特征主要关注于不同人脸的个体差异和特性，而面部表情作为干扰信号存在，因此人脸识别不过多关注面部表情。而本申请实施例是需要关注表情的变化来触发对应的第二媒体信息，因此可以忽略个体差异，而关注于提取人脸在不同表情模式

下的差异特征的特征提取。其可以与个体差异相结合，也可以为了提高表情识别精度而将个体差异作为干扰信号处理，即不过多关注个体差异。特征提取是人脸表情识别中的核心步骤，决定着最终的识别结果，影响识别率的高低。其中，所述特征提取可以分为：静态图像特征提取和运动图像特征提取。就静态图像特征提取而言，提取的是表情的形变特征（或称为表情的暂态特征），就运动图像特征提取而言，对于运动图像，不仅要提取每一帧的表情形变特征，还要提取连续序列的运动特征。形变特征提取可以依赖中性表情或模型，从而把产生的表情与中性表情做比较来提取出形变特征，而运动特征的提取则直接依赖于表情产生的面部变化。表情有多种划分方式，1）如按照基本表情划分，如高兴、悲伤、惊讶、恐惧、愤怒和厌恶等，建立不同的人脸表情图像库以便后续的匹配和识别。2）按照情绪分类，如愉快，不愉快，激动，平静，紧张，轻松等。

步骤 103、终端接收服务器推送的与所述关键信息对应的第二媒体信息。

该步骤的一种具体实现可以为：步骤 102 将关键信息上报给服务器之后，服务器根据关键信息从素材库匹配对应的第二媒体信息，例如贴纸信息，并将第二媒体信息推送给终端，以便后续在步骤 104 中与第一媒体信息进行视频合成。无需用户手动选择贴纸信息，而是根据关键信息匹配后自动推送给终端，在终端采集第一媒体信息（如视频）的过程中自动合成（如将视频和贴纸信息相叠加）视频处理结果，在第一媒体信息（如视频）的指定位置和指定时间显示贴纸信息。

步骤 104、将第一媒体信息和第二媒体信息进行视频合成。

在本申请实施例一实施方式中，所述关键信息还包括：第一媒体信息中的文字信息。

该信息处理方法还包括：在采集所述第一媒体信息的过程中检测该文字信息，并将其作为关键信息上报给服务器。

现有技术中，如图 5 中的文字信息，具体的，在视频信息中包含 A2 所标识的文字信息“红红火火”。在视频信息录制完成后添加如 A2’所标识贴纸信息“红红火火”。该贴纸信息是通过终端多次与服务器的交互，从服务器素材

库手工选取的，之后，在将该贴纸信息附加到已经录制完成的视频信息中。

如图 6 所示为现有技术的另一个应用场景。具体的，在视频信息中包含 A3 所标识的文字信息“男朋友”。在视频信息录制完成后添加如 A3’所标识贴纸信息“男朋友”。该贴纸信息是通过终端多次与服务器的交互，从服务器素材库手工选取的，之后，在将该贴纸信息附具到已经录制完成的视频信息中。这种处理，非常繁琐，需要多次用户交互，后续找的贴纸也不一定就是用户真正需要的，即便是用户真正需要的，也需要用户手工在已经录制完成的视频信息上手动添加，比如将贴纸移动到视频信息的合适位置上等等。

而采用本申请实施例，如图 7 所示的视频中包括 A4 所标识的文字信息“吃不胖”，其作为关键信息被发送到服务器。基于该关键信息得到的匹配贴纸信息如 A4’所标识。如图 8 所示的视频中包括 A5 所标识的文字信息“男朋友”，其作为关键信息被发送个服务器。基于该关键信息得到的匹配贴纸信息如 A5’所标识。在图 7 和 8 中，B1 用于标识视频录制过程中的控制按钮，B2 用于标识视频录制结束后的回放按钮。如图 9 所示为一个录制视频过程中将贴纸信息与视频在合适的位置和时间点进行视频合成后，回放该视频的示意图。在图 9 中，对应录制的视频信息中播放语音为“祝姐妹们过年吃不胖”时，在视频界面可以显示对应语音的文字信息，在这个时间点，在视频界面上还显示有合成的贴纸信息以卷轴形式打开并显示的动态贴纸效果“吃不胖”。图 10 为另一个采用本申请实施例录制视频后，将贴纸信息与视频在合适的位置和时间点进行视频合成后，回放该视频的示意图，其中，当对应录制的视频信息中播放语音为“年终奖多多”时，在视频界面可以显示对应语音的文字信息，在这个时间点，在视频界面上还显示有合成的贴纸信息，以动态贴纸效果显示“年终奖多多”，并配合有货币单位的指示符，例如￥，将其与“年终奖多多”的文字相结合。

当录制的视频中有文字信息“年终奖多多”时，除了如图 10 所示的显示对应内容的贴纸信息之外，还可以通过识别人脸表情或用户动作而得到其他的贴纸形态。如图 11 所示，当录制的视频信息中播放语音为如 A6 标识的“年终奖多多”时，可以使用户动作和语音相结合。例如，该用户动作可以为用

户开心的眨眼。在这种情况下，在视频界面除了可以显示如图 10 所示的贴纸信息之外，还可以在这个“开心的眨眼”的时间段，在视频界面上还显示有其他的贴纸信息，如 A6’标识的“眼睛变成两个￥”。除了眨眼之外，该用户动作还可以为打响指。通过该用户动作触发显示如图 11 中的 A6’标识的“眼睛变成两个￥”或者显示如图 10 所示的贴纸信息“年终奖多多”。

图 12 所示为另一个采用本申请实施例的应用实例。在图 12 中，还可以通过识别人脸表情得到其他的贴纸形态。如图 12 所示，当对应录制的视频信息中播放语音为如 A7 标识的“我真有这么漂亮吗”时，识别出人脸脸颊的位置，在人脸脸颊的位置叠加如 A7’所示的贴纸信息。具体地，该贴纸信息为五官贴纸类型中的红脸蛋、腮红或红晕。那么，在视频界面在出现“我真有这么漂亮吗”的时间段，在视频界面上还显示有合成的贴纸信息，人脸上有红晕。

实施例二

本申请实施例的信息处理方法，如图 13 所示，所述方法包括:

步骤 201、终端开启应用，获取第一操作，触发第一媒体信息的采集。

一个应用场景中，如图 3 所示，用户躺在沙发上正在使用如手机 11 的终端设备。手机 11 的用户界面如图 4 所示，其中包含各种类型的应用图标，如音乐播放图标，功能设置图标，邮件收发图标等等。用户执行第一操作，如用手指点击 A1 标识的视频处理应用图标，进入视频录制的处理过程，从而触发如视频的第一媒体信息的采集。比如，可以录制一段室内的场景，或者给自己进行自拍等等。

步骤 202、终端在采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将得到的变化量作为关键信息上报给服务器。

这里，仍然结合步骤 201 中的应用场景，在视频录制的处理过程中，通过人脸识别定位机制，表情识别机制，终端设备可以捕获人脸区域内的表情变化，如微笑，哭泣，皱眉等等。此外，终端设备还可以检测采集框（或称

取景框）内的用户动作变化，例如比剪刀手。这种检测不限于人脸区域。还可以将人脸区域中的表情变化与用户动作变化进行组合识别，比如，将剪刀手和脸部表情中的微笑相结合进行组合识别。

在人脸识别的过程中，人脸识别技术是基于人的脸部特征，对视频录制中的人脸图像或者视频流进行采集，首先判断视频流中是否存在人脸，如果存在人脸，则进一步的给出脸的位置和大小，及定位出各个主要面部器官的位置信息，得到人脸中五官的各自位置和初始形态，当形态发生变化，如微笑时上下嘴唇的位置会相对初始形态产生位移和形变，则说明人脸五官的面部表情出现变化，也可以通过表情识别机制来识别出表情的变化。本申请实施例的人脸识别有别于常规的人脸识别，常规的人脸识别是为了通过构建的人脸识别系统来识别出用户的身份，是将识别出的人脸与已知人脸进行比对，以便于身份确认以及身份查找。

在表情识别过程中，可以分为四个阶段：如人脸图像的获取与预处理；人脸检测；表情特征提取；以及表情分类。如果仅仅通过人脸识别和定位机制，会存在不精确的问题，而表情识别机制是一种更加准确的处理策略，表情识别与人脸识别密切相关，如在人脸检测中的定位和人脸跟踪这些环节上是类似的，但特征提取上不同。举例来说，人脸识别提取的特征主要关注于不同人脸的个体差异和特性，而面部表情作为干扰信号存在。也就是说，不过多关注面部表情。而本申请实施例是需要关注表情的变化来触发对应的第二媒体信息，因此，可以忽略个体差异，而关注于提取人脸在不同表情模式下的差异特征的特征提取。其可以与个体差异相结合，也可以为了提高表情识别精度而将个体差异作为干扰信号处理，即不过多关注个体差异。特征提取是人脸表情识别中的核心步骤，决定着最终的识别结果，影响识别率的高低。其中，所述特征提取可以分为：静态图像特征提取和运动图像特征提取。就静态图像特征提取而言，提取的是表情的形变特征（或称为表情的暂态特征），就运动图像特征提取而言，对于运动图像，不仅要提取每一帧的表情形变特征，还要提取连续序列的运动特征。形变特征提取可以依赖中性表情或模型，从而把产生的表情与中性表情做比较来提取出形变特征，而运动特征

的提取则直接依赖于表情产生的面部变化。表情有多种划分方式，1）如按照基本表情划分，如高兴、悲伤、惊讶、恐惧、愤怒和厌恶等，建立不同的人脸表情图像库以便后续的匹配和识别。2）按照情绪分类，如愉快，不愉快，激动，平静，紧张，轻松等。

步骤 203、服务器从素材库中选取与关键信息对应的第二媒体信息和第二媒体信息的描述文件。

步骤 204、终端接收服务器推送的与所述关键信息对应的第二媒体信息和第二媒体信息的描述文件。

该步骤的一种具体实现可以为：在步骤 202 中将关键信息上报给服务器之后，服务器根据关键信息从素材库匹配对应的第二媒体信息，例如贴纸信息，并将该第二媒体信息推送给终端，以便后续在步骤 205 中与第一媒体信息进行视频合成。无需用户手动选择贴纸信息，而是根据关键信息匹配后自动推送给终端，在终端采集第一媒体信息（如视频）的过程中自动合成（如将视频和贴纸信息相叠加）视频处理结果，在第一媒体信息（如视频）的指定位置和指定时间显示贴纸信息。

这里，步骤 204 中，与所述关键信息对应的第二媒体信息的描述文件和第二媒体信息可以同时发送或者分别发送，取决于当时的网络状况，如果网络状况好，则同时发送，如果网络状况不好，为了避免网络不好，丢失数据，可以分别发放。

步骤 205、将第一媒体信息和第二媒体信息进行视频合成。

在本申请实施例一实施方式中，所述关键信息还包括：第一媒体信息中的文字信息。

因此该方法还包括：在采集所述第一媒体信息的过程中检测该文字信息，并将其作为关键信息上报给服务器。

现有技术中，如图 5 中的文字信息，具体的，在视频信息中包含 A2 所标识的文字信息“红红火火”。在视频信息录制完成后添加如 A2’所标识贴纸信息“红红火火”。该贴纸信息是通过终端多次与服务器的交互，从服务器素材库手工选取的，之后，在将该贴纸信息附加到已经录制完成的视频信息中。

如图 6 所示为现有技术的另一个应用场景，具体的，在视频信息中包含 A3 所标识的文字信息“男朋友”。在视频信息录制完成后添加如 A3’所标识贴纸信息“男朋友”。该贴纸信息是通过终端多次与服务器的交互，从服务器素材库手工选取的，之后，在将该贴纸信息附加到已经录制完成的视频信息中。这种处理，非常繁琐，需要多次用户交互，后续找的贴纸也不一定就是用户真正需要的，即便是用户真正需要的，也需要用户手工在已经录制完成的视频信息上手动添加，比如将贴纸移动到视频信息的合适位置上等等。

而采用本申请实施例，如图 7 所示的视频中包括 A4 所标识的文字信息“吃不胖”，其作为关键信息被发送到服务器。基于该关键信息得到的匹配贴纸信息如 A4’所标识。如图 8 所示的视频中包括 A5 所标识的文字信息“吃不胖”，其作为关键信息被发送到服务器。基于该关键信息得到的匹配贴纸信息如 A5’所标识。在图 7 和 8 中，B1 用于标识视频录制过程中的控制按钮，B2 用于标识视频录制结束后的回放按钮。如图 9 所示为一个录制视频过程中将贴纸信息与视频在合适的位置和时间点进行视频合成后，回放该视频的示意图。在图 9 中，对应录制的视频信息中播放语音为“祝姐妹们过年吃不胖”时，在视频界面可以显示对应语音的文字信息，在这个时间点，在视频界面上还显示有合成的贴纸信息以卷轴形式打开并显示的动态贴纸效果“吃不胖”。图 10 为另一个采用本申请实施例录制视频后，将贴纸信息与视频在合适的位置和时间点进行视频合成后，回放该视频的示意图，其中，当对应录制的视频信息中播放语音为“年终奖多多”时，在视频界面可以显示对应语音的文字信息，在这个时间点，在视频界面上还显示有合成的贴纸信息，以动态贴纸效果显示“年终奖多多”，并配合有货币单位的指示符，例如￥，将其与“年终奖多多”的文字相结合。

当录制的视频中有文字信息“年终奖多多”时，除了如图 10 所示的显示对应内容的贴纸信息之外，还可以通过识别人脸表情或用户动作而得到其他的贴纸形态。如图 11 所示，当录制的视频中播放语音为如 A6 标识的“年终奖多多”时，可以使用户动作和语音相结合。例如，该用户动作可以为用户开心的眨眼。在这种情况下，在视频界面除了可以如图 10 所示的显示对应的

文字信息之外，还可以在这个“开心的眨眼”的时间段，在视频界面上还显示有其他的贴纸信息，如 A6’标识的“眼睛变成两个￥”。除了眨眼之外，该用户动作该可以是打响指。通过该用户动作触发显示如图 11 中的 A6’标识的“眼睛变成两个￥”或者如图 10 所示的贴纸信息“年终奖多多”。

图 12 所示为另一个采用本申请实施例的应用实例。在图 12 中，还可以通过识别人脸表情得到其他的贴纸形态。如图 12 所示，当对应录制的视频信息中播放语音为如 A7 标识的“我真有这么漂亮吗”时，识别出人脸脸颊的位置，在人脸脸颊的位置叠加如 A7’所示的贴纸信息。具体地，该贴纸信息为五官贴纸类型中的红脸蛋、腮红或红晕。那么，在视频界面在出现“我真有这么漂亮吗”的时间段，在视频界面上还显示有合成的贴纸信息，人脸上有红晕。

在本申请实施例一实施方式中，所述将第一媒体信息和第二媒体信息进行视频合成，包括:

第一种实现方案：响应所述表情变化或所述用户动作变化，获取对应的特征检测结果，将所述第二媒体信息按照所述特征检测结果和所述第二媒体信息的描述文件的配置与第一媒体信息进行视频合成，并在指定时间点或时间段内将所述第二媒体信息显示在所述第一媒体信息指定的位置处。

第二种实现方案：响应所述文字信息，将所述第二媒体信息按照所述第二媒体信息的描述文件的配置与第一媒体信息进行视频合成，并在指定时间点或时间段内将所述第二媒体信息显示在所述第一媒体信息指定的位置处。

两种方案的区别在于：第一种方案，需要得到特征坐标（特征检测结果中的部分信息或全部信息），以便结合特征坐标，确定将贴纸信息放到视频信息中哪个合适的指定位置，第二媒体信息可以决定时间点，贴纸信息的摆放是有固定位置和固定时间要求的，根据这个指定位置和时间点，就可以实现在合适的位置，合适的时间点叠加贴纸信息到视频信息上，比如，如图 12 所示的“脸上出现腮红”这种场景；第二种方案中，可以不考虑特征坐标，第二媒体信息可以决定时间点，和贴纸的大小等属性，还可以包括贴纸信息的中心点位置，贴纸的摆放虽然也有固定位置和固定时间要求的，但是，相比

于第一种方案，更具有任意性，如图 7 所示，只要出现“吃不胖”就行，不限定“吃不胖”一定在人脸区域的哪个相对位置显示，而第一种方案中，是由表情变化或用户动作变化触发的贴纸请求，因此，务必要配合表情变化或用户动作变化来显示。

在本申请实施例一实施方式中，所述第二多媒体信息包括以下至少一类：1）由所述表情变化或所述用户动作变化触发显示的第一类贴纸信息，如五官贴纸和触发类贴纸；2）由排除所述表情变化或所述用户动作变化触发显示的第二类贴纸信息普通贴纸和背景贴纸。

在本申请实施例一实施方式中，所述响应所述表情变化或所述用户动作变化时，获取对应的特征检测结果，将所述第二媒体信息按照所述特征检测结果和所述第二媒体信息的描述文件的配置与第一媒体信息进行视频合成包括:

a1、响应所述表情变化或所述用户动作变化，将检测到的特征变化量上报服务器，以请求所述第一类贴纸信息和第一类贴纸信息的描述文件;

a2、检测所述表情变化或所述用户动作变化引起的特征坐标变化，由初始坐标定位到目标坐标，以根据目标坐标定位得到的位置点或者由初始坐标至目标坐标界定的位置区域来确定叠加所述第一类贴纸信息的位置;

a3、解析收到的所述第一类贴纸信息的描述文件，得到第一类贴纸信息的显示时间;

a4、按照所述确定的位置以及所述解析的第一类贴纸信息的显示时间，将第二媒体信息与第一媒体信息进行视频合成。

在本申请实施例一实施方式中，所述响应所述文字信息时，将所述第二媒体信息按照所述第二媒体信息的描述文件的配置与第一媒体信息进行视频合成，包括:

b1、响应所述文字信息，将检测到的文字信息上报服务器，以请求所述第二类贴纸信息和第二类贴纸信息的描述文件;

b2、解析收到的所述第二类贴纸信息的描述文件，得到第二类贴纸信息相对于第一媒体信息的位置，以及第二类贴纸信息的显示时间，其中，所述

位置包括第二类贴纸信息显示的中心点位置；

b3、按照所述得到的位置和所述显示时间，将第二媒体信息与第一媒体信息进行视频合成。

实施例三

根据本申请的实施例提供了一种终端。如图 14 所示，所述终端包括：触发单元 21，用于获取第一操作，以触发第一媒体信息的采集；检测单元 22，用于在采集所述第一媒体信息的过程中检测人脸区域内的表情变化或采集框内的用户动作变化时，将得到的变化量作为关键信息上报给服务器；及接收单元 23，用于接收服务器推送的与所述关键信息对应的第二媒体信息；及合成单元 24，用于将第一媒体信息和第二媒体信息进行视频合成。

一个应用场景中，如图 3 所示，用户躺在沙发上正在使用如手机 11 的终端设备。手机 11 的用户界面如图 4 所示，其中包含各种类型的应用图标，如音乐播放图标，功能设置图标，邮件收发图标等等。用户执行第一操作，如用手指点击 A1 标识的视频处理应用图标，进入视频录制的处理过程，从而触发如视频的第一媒体信息的采集。比如，可以录制一段室内的场景，或者给自己进行自拍等等。在视频录制的处理过程中，通过人脸识别定位机制或表情识别机制，终端可以捕获到人脸区域内的表情变化，例如微笑，哭泣，皱眉等等。此外，终端设备还可以检测采集框（或称取景框）内的用户动作变化，例如比剪刀手。还可以将人脸区域中的表情变化与用户动作变化进行组合识别，比如，将剪刀手和脸部表情中的微笑相结合进行组合识别。

在人脸识别的过程中，人脸识别技术是基于人的脸部特征，对视频录制中的人脸图像或者视频流进行采集，首先判断视频流中是否存在人脸，如果存在人脸，则进一步的给出脸的位置和大小，及定位出各个主要面部器官的位置信息，得到人脸中五官的各自位置和初始形态，当形态发生变化，如微笑时上下嘴唇的位置会相对初始形态产生位移和形变，则说明人脸五官的面部表情出现变化，也可以通过表情识别机制来识别出表情的变化。本申请实施例的人脸识别有别于常规的人脸识别，常规的人脸识别是为了通过构建的

人脸识别系统来识别出用户的身份，是将识别出的人脸与已知人脸进行比对，以便于身份确认以及身份查找。

在表情识别过程中，可以分为四个阶段：如人脸图像的获取与预处理；人脸检测；表情特征提取；和表情分类。如果仅仅通过人脸识别和定位机制，会存在不精确的问题，而表情识别机制是一种更加准确的处理策略。表情识别与人脸识别密切相关，如在人脸检测中的定位和人脸跟踪这些环节上是类似的，但特征提取上不同。举例来说，人脸识别提取的特征主要关注于不同人脸的个体差异和特性，而面部表情作为干扰信号存在，，因此人脸识别不过多关注面部表情。而本申请实施例是需要关注表情的变化来触发对应的第二媒体信息，因此可以忽略个体差异，而关注于提取人脸在不同表情模式下的差异特征的特征提取。其可以与个体差异相结合，也可以为了提高表情识别精度而将个体差异作为干扰信号处理，即不过多关注个体差异。特征提取是人脸表情识别中的核心步骤，决定着最终的识别结果，影响识别率的高低。其中，所述特征提取可以分为：静态图像特征提取和运动图像特征提取。就静态图像特征提取而言，提取的是表情的形变特征（或称为表情的暂态特征），就运动图像特征提取而言，对于运动图像，不仅要提取每一帧的表情形变特征，还要提取连续序列的运动特征。形变特征提取可以依赖中性表情或模型，从而把产生的表情与中性表情做比较来提取出形变特征，而运动特征的提取则直接依赖于表情产生的面部变化。表情有多种划分方式，1）如按照基本表情划分，如高兴、悲伤、惊讶、恐惧、愤怒和厌恶等，建立不同的人脸表情图像库以便后续的匹配和识别。2）按照情绪分类，如愉快，不愉快，激动，平静，紧张，轻松等。

在本申请实施例一实施方式中，所述关键信息还包括：第一媒体信息中的文字信息。

检测单元 22，还用于在采集所述第一媒体信息的过程中检测所述文字信息，并将所述文字信息作为关键信息上报给所述服务器。

现有技术中，如图 5 中的文字信息，具体的，在视频信息中包含 A2 所标识的文字信息“红红火火”。在视频信息录制完成后添加如 A2’所标识贴纸信

息“红红火火”。该贴纸信息是通过终端多次与服务器的交互，从服务器素材库手工选取的，之后，在将该贴纸信息附加到已经录制完成的视频信息中。

如图 6 所示为现有技术的另一个应用场景。具体的，在视频信息中包含 A3 所标识的文字信息“男朋友”。在视频信息录制完成后添加如 A3’所标识贴纸信息“男朋友”。该贴纸信息是通过终端多次与服务器的交互，从服务器素材库手工选取的，之后，在将该贴纸信息附具到已经录制完成的视频信息中。这种处理，非常繁琐，需要多次用户交互，后续找的贴纸也不一定就是用户真正需要的，即便是用户真正需要的，也需要用户手工在已经录制完成的视频信息上手动添加，比如将贴纸移动到视频信息的合适位置上等等。

而采用本申请实施例，如图 7 所示的视频中包括 A4 所标识的文字信息“吃不胖”，其作为关键信息被发送到服务器。基于该关键信息得到的匹配贴纸信息如 A4’所标识。如图 8 所示的视频中包括 A5 所标识的文字信息“男朋友”，其作为关键信息被发送到服务器。基于该关键信息得到的匹配贴纸信息如 A5’所标识。在图 7 和 8 中，B1 用于标识视频录制过程中的控制按钮，B2 用于标识视频录制结束后的回放按钮。如图 9 所示为一个录制视频过程中将贴纸信息与视频在合适的位置和时间点进行视频合成后，回放该视频的示意图，在图 9 中，对应录制的视频信息中播放语音为“祝姐妹们过年吃不胖”时，在视频界面可以显示对应语音的文字信息，在这个时间点，在视频界面上还显示有合成的贴纸信息以卷轴形式打开并显示的动态贴纸效果“吃不胖”。图 10 为另一个采用本申请实施例录制视频后，将贴纸信息与视频在合适的位置和时间点进行视频合成后，回放该视频的示意图，其中，当对应录制的视频信息中播放语音为“年终奖多多”时，在视频界面可以显示对应语音的文字信息，在这个时间点，在视频界面上还显示有合成的贴纸信息，以动态贴纸效果显示“年终奖多多”，并配合有货币单位的指示符，例如￥，将其与“年终奖多多”的文字相结合。

当录制的视频中有文字信息“年终奖多多”时，除了如图 10 所示的显示对应内容的贴纸信息之外，还可以通过识别人脸表情或用户动作而得到其他的贴纸形态。如图 11 所示，当录制的视频中播放语音为如 A6 标识的“年终

奖多多”时可以使用户动作和语音相结合。例如，该用户动作可以为用户开心的眨眼。在这种情况下，在视频界面除了可以如图 10 所示的显示贴纸信息之外，还可以在这个“开心的眨眼”的时间段，在视频界面上还显示有其他的贴纸信息，如 A6’标识的“眼睛变成两个￥”。除了眨眼之外，该还可以是用户动作还可以为打响指。通过该用户动作触发显示如图 11 中的 A6’标识的“眼睛变成两个￥”或者如图 10 所示的贴纸信息“年终奖多多”。

图 12 所示为另一个采用本申请实施例的应用实例。在图 12 中，还可以通过识别人脸表情得到其他的贴纸形态。如图 12 所示，当对应录制的视频信息中播放语音为如 A7 标识的“我真有这么漂亮吗”时，识别出人脸脸颊的位置，在人脸脸颊的位置叠加如 A7’所示的贴纸信息。具体地，该贴纸信息为五官贴纸类型中的红脸蛋、腮红或红晕。，那么，在视频界面在出现“我真有这么漂亮吗”的时间段，在视频界面上还显示有合成的贴纸信息，人脸上有红晕。

在本申请实施例一实施方式中，所述接收单元 24，进一步用于：接收服务器推送的与所述关键信息对应的第二媒体信息的描述文件。

所述描述文件包括：所述第二媒体信息相对于第一媒体信息的位置，以及第二媒体信息的显示时间。

在本申请实施例一实施方式中，所述合成单元 24 进一步用于根据所述描述文件将所述第一媒体信息与所述第二媒体信息进行视频合成，以在所述描述文件指定的显示时间内将所述第二媒体信息显示在所述描述文件指定的所述第一媒体信息的位置处。具体地，合成单元 24 包括两种具体实现：

第一种具体实现：响应所述表情变化或所述用户动作变化时，获取对应的特征检测结果，将所述第二媒体信息按照所述特征检测结果和所述第二媒体信息的描述文件的配置与第一媒体信息进行视频合成，并将所述第二媒体信息显示在所述第一媒体信息指定的位置和指定时间点或时间段内。

第二种具体实现：响应所述文字信息时，将所述第二媒体信息按照所述第二媒体信息的描述文件的配置与第一媒体信息进行视频合成，并将所述第二媒体信息显示在所述第一媒体信息指定的位置和指定时间点或时间段内。

在本申请实施例一实施方式中，所述第二多媒体信息包括以下至少一类：

由所述表情变化或所述用户动作变化触发显示的第一类贴纸信息；

由排除所述表情变化或所述用户动作变化触发显示的第二类贴纸信息。

在本申请实施例一实施方式中，所述合成单元 24，进一步用于：

响应所述表情变化或所述用户动作变化，将检测到的特征变化量上报服务器，以请求所述第一类贴纸信息和第一类贴纸信息的描述文件；

检测所述表情变化或所述用户动作变化引起的特征坐标变化，由初始坐标定位到目标坐标，以根据目标坐标定位得到的位置点或者由初始坐标至目标坐标界定的位置区域来叠加所述第一类贴纸信息；

解析收到的所述第一类贴纸信息的描述文件，得到第一类贴纸信息待显示的指定时间；

按照所述位置点或所述位置区域所指定的位置和所述指定时间，将第二媒体信息与第一媒体信息进行视频合成。

在本申请实施例一实施方式中，所述合成单元，进一步用于：

响应所述文字信息，将检测到的文字信息上报服务器后，以请求所述第二类贴纸信息和第二类贴纸信息的描述文件；

解析收到的所述第二类贴纸信息的描述文件，得到第一类贴纸信息待显示的指定位置和指定时间；所述指定位置包括第一类贴纸信息显示的中心点位置；

按照所述指定位置和所述指定时间，将第二媒体信息与第一媒体信息进行视频合成。

实施例四

这里需要指出的是，上述终端可以为 PC 这种电子设备，还可以为如 PAD，平板电脑，手提电脑这种便携电子设备、还可以为如手机这种智能移动终端，不限于这里的描述；所述服务器可以是通过集群系统构成的，为实现各单元功能而合并为一或各单元功能分体设置的电子设备，终端和服务器都至少包括用于存储数据的数据库和用于数据处理的处理器，或者包括设置于服务器

内的存储介质或独立设置的存储介质。

其中，对于用于数据处理的处理器而言，在执行处理时，可以采用微处理器、中央处理器（CPU，Central Processing Unit）、数字信号处理器（DSP，Digital Singnal Processor）或可编程逻辑阵列（FPGA，Field－Programmable Gate Array）实现；对于存储介质来说，包含操作指令，该操作指令可以为计算机可执行代码，通过所述操作指令来实现上述本申请实施例信息处理方法流程中的各个步骤。

该终端和该服务器作为硬件实体 S11 的一个示例如图 15 所示。所述装置包括处理器 41、存储介质 42 以及至少一个外部通信接口 43；所述处理器 41、存储介质 42 以及外部通信接口 43 均通过总线 44 连接。

这里需要指出的是：以上涉及终端和服务器项的描述，与上述方法描述是类似的，同方法的有益效果描述，不做赘述。对于本申请终端和服务器实施例中未披露的技术细节，请参照本申请方法实施例的描述。

以一个现实应用场景为例对本申请实施例阐述如下：

首先对应用场景进行介绍：一，在视频素材中，根据素材演绎的内容，经常会要加一些相关联的动态贴纸，让视频显得更加丰富。A）比如：拜年视频中，当表达恭喜发财时，会希望有一些从天而降的金币；B）又比如：当视频内容想发表主角娇羞状态时，会希望能在用户脸上加上红晕的特效。二，视频中会有和某个明星合演的需求，这个时候也可以将明星直接做成前景的贴纸，然后让用户和明星合影。现有技术中，经视频处理技术得到的示意图如图 5-6 所示。其中，如图 5 中的文字信息，具体的，在视频信息中包含 A2 所标识的文字信息“红红火火”。在视频信息录制完成后添加如 A2’所标识贴纸信息“红红火火”。该贴纸信息是通过终端多次与服务器的交互，从服务器素材库手工选取的，之后，在将该贴纸信息附具到已经录制完成的视频信息中。

如图 6 所示为现有技术的另一个应用场景。具体的，在视频信息中包含 A3 所标识的文字信息“男朋友”。在视频信息录制完成后添加如 A3’所标识贴纸信息“男朋友”。该贴纸信息是通过终端多次与服务器的交互，从服务器素

材库手工选取的，之后，在将该贴纸信息附具到已经录制完成的视频信息中。这种处理，非常繁琐，需要多次用户交互，后续找的贴纸也不一定就是用户真正需要的，即便是用户真正需要的，也需要用户手工在已经录制完成的视频信息上手动添加，比如将贴纸移动到视频信息的合适位置上等等。

上述图 5-6 的场景中，采用的现有视频处理技术为：应用（APP）会提供一些固定的贴纸，然后用户先录制视频，录制后，由用户自己去贴纸素材库里选择自己认为相关素材，然后通过复杂的交互，决定每张贴纸在什么时间点加上，加多久。且有些 APP 允许贴纸移动，然后要按住贴纸，拖动决定移动到哪个具体的位置，这样的后果是：需要终端与服务器间多次繁琐的交互，处理效率低下，在视频录制完成后手动选择贴纸再最终合成，视频处理成本高，浪费时间，还不一定符合用户需求。

针对上述应用场景，采用本申请实施例，是一种视频相关实时动效贴纸方案。采用本申请的人脸识别及定位机制，表情识别机制，视频合成处理机制，可以不需要用户通过复杂的操作，去一堆素材中选择某个和该视频素材相关的贴纸信息，而是选择了某段素材视频，在录制的过程中，就能看到对应的地方，在对应的时间，出现对应的贴纸信息，可以称为在视频录制的过程中就实时的在相应指定位置和指定时间点叠加上对应的贴纸信息，如图 7-12 所示。

采用本申请实施例，如图 7 所述的视频中包括 A4 所标识的文字信息“吃不胖”，其作为关键信息被发送到服务器。基于该关键信息得到的匹配贴纸信息如 A4’所标识。如图 8 所述的视频中包括 A5 所标识的文字信息“男朋友”，其作为关键信息被发送个服务器。基于该关键信息得到的匹配贴纸信息如 A5’所标识。在图 7-8 中，B1 用于标识视频录制过程中的控制按钮，B2 用于标识视频录制结束后的回放按钮。如图 9 所示为一个录制视频过程中将贴纸信息与视频在合适的位置和时间点进行视频合成后，回放该视频的示意图，在图 9 中，对应录制的视频信息中播放语音为“祝姐妹们过年吃不胖”时，在视频界面可以显示对应语音的文字信息，在这个时间点，在视频界面上还显示有合成的贴纸信息以卷轴形式打开并显示的动态贴纸效果“吃不胖”。图 10 为

另一个采用本申请实施例录制视频后，将贴纸信息与视频在合适的位置和时间点进行视频合成后，回放该视频的示意图，其中，当对应录制的视频信息中播放语音为“年终奖多多”时，在视频界面可以显示对应语音的文字信息，在这个时间点，在视频界面上还显示有合成的贴纸信息，以动态贴纸效果显示“年终奖多多”，并配合有货币单位的指示符，例如￥，将其与“年终奖多多”的文字相结合。

当录制的视频中有文字信息“年终奖多多”时，除了如图 10 所示的显示对应内容的贴纸信息之外，还可以通过识别人脸表情或用户动作而得到其他的贴纸形态。如图 11 所示，当录制的视频信息中播放语音为如 A6 标识的“年终奖多多”时可以使用户动作和语音相结合。例如，该用户动作可以为用户开心的眨眼。在这种情况下，在视频界面除了可以显示如图 10 所示的贴纸信息之外，还可以在这个“开心的眨眼”的时间段，在视频界面上还显示有其他的贴纸信息，如 A6’标识的“眼睛变成两个￥”。除了眨眼之外，该用户动作还可以为打响指。通过该用户动作触发显示如图 11 中的 A6’标识的“眼睛变成两个￥”或者如图 10 所示的“年终奖多多”。

图 12 所示为另一个采用本申请实施例的应用实例。在图 12 中，还可以通过识别人脸表情得到其他的贴纸形态。如图 12 所示，当对应录制的视频信息中播放语音为如 A7 标识的“我真有这么漂亮吗”时，识别出人脸脸颊的位置，在人脸脸颊的位置叠加如 A7’所示的贴纸信息。具体地，该贴纸信息为五官贴纸类型中的红脸蛋、腮红或红晕。那么，在视频界面在出现“我真有这么漂亮吗”的时间段，在视频界面上还显示有合成的贴纸信息，人脸上有红晕。

在录制完时，对应的贴纸信息也已经出现在视频里了。

这里需要指出的是：贴纸信息分为以下几种：

A）普通贴纸：比如天上掉下的金币，抖动的红包，盛开的一朵小花都属于这类贴纸；

B）五官贴纸：可以指定出现在五官某个具体的位置，并且会跟随五官移动的贴纸。如：脸蛋的红晕，眼镜等；

C）触发类贴纸：检测到某个具体的动作时，出现的变化的一组贴纸，出现的一组贴纸既可以是普通贴纸，也可以是五官贴纸；以及

D）背景贴纸：盖在视频最上方，并且重复播放的几帧视频，如图 8 中“吃不胖”的边框，如边框类贴纸。

对于上述贴纸信息的四种类型，触发类贴纸和五官类贴纸信息是如图 17 所示，需要贴纸信息与特征坐标相结合，之后再与录制视频合成，其需要与图 17 中的特征检测器和素材解析器发生关系，再与视频合成器发生关系，这是因为表情，动作，五官变化这种，坐标都是会变化的。而除去触发类贴纸和五官类贴纸信息之外，其他几种（普通贴纸和背景贴纸），都是贴纸信息直接与录制视频合成。也就是说，仅与图 17 中的视频合成器发生关系，因为坐标通常不会变。

要想实现上述图 7-12 所示的最终效果，在技术实现上，还包括以下内容：

一，将每个视频的贴纸信息作为素材的一部分，在素材包中，并且随素材下发。所述素材除了包括贴纸信息，还包括贴纸信息的描述文件等等。

二，动态的素材包括两部分：

A）贴纸信息的原始形态，主要有三种格式：i）静态贴图；ii）动态的图像互换格式（Graphics Interchange Format，Gif）图；iii）视频。其中，对于图片类的贴纸信息文件（如静态贴图和动态的 Gif 图），用可移植网络图形格式（Portable Network Graphic Format，PNG）图片等带透明度的图片进行叠加来实现视频合成即可；而对于多数视频类的贴纸信息文件（如视频），由于其是不带透明度的，因此当使用视频作为素材时，素材视频的分辨率是拍摄视频的两倍，其中一半的像素用来表示贴纸的 RGB 值，另一半像素用来表示贴纸的透明度。具体地，视频类贴纸信息的存储方式是：RGB 和透明通道分开，将拍摄的视频，分为一半是素材 RGB，一半是素材透明度进行存储，如图 16 所示。RGB 这种色彩模式是一种颜色标准，是通过对红(R)、绿(G)、蓝(B)三个颜色通道的变化以及它们相互之间的叠加来得到各式各样的颜色的，RGB 即是代表红、绿、蓝三个通道的颜色，这个标准几乎包括了人类视力所能感知的所有颜色。

在贴纸信息和拍摄的视频进行视频叠加的视频合成时，当贴纸信息上某个像素的透明度为 a 时，合成视频的 RGB 值 = a * 视频类贴纸信息的 RGB 值 +（1-a）* 拍摄视频的 RGB 值。

B）贴纸的描述文件，该描述文件包含的信息为：i）贴图出现的中心点位置；ii）贴图出现时间。从而，根据贴纸信息和贴纸描述文件，可以为终端主动推送贴纸，以在正在录制视频的合适的位置，合适的时间点，叠加上合适的动态贴纸，无需用户手动选择贴纸。其中，贴图出现时间包括：a）对于一次性播放的动态贴图，需要设定什么时候开始；b）对于重复播放的动态贴图，需要设定开始和结束时间。

C）五官类贴纸需要设定：五官类型的信息包括：i）头顶；ii）眼睛；iii）脸蛋；iv）嘴巴；v）鼻子。

D）触发类贴纸需要设定：触发条件，触发条件具体包括：i）张嘴；ii）眨眼；iii）微笑；iv）抖眉毛。

E）不同素材的层级关系。

三，素材预览时，解析动态贴纸的文件。

四，录制过程中，按照贴纸信息的描述文件，绘制动态贴纸，已达到实时可见的目的，如果是五官类和触发类贴纸，在系统构成中还包括了人脸检测算法。需要注意的是该人脸检测算法使用现有的人脸检测算法，算法本身不包括在专利内。绘制贴纸时根据人脸检测的结果将贴纸绘制在恰当的位置，如图 17 所示为整个系统的结构示意图。在一个实施例中，图 17 中的模块都位于终端侧。对于触发类贴纸和五官类贴纸信息，需要贴纸信息与特征坐标相结合，之后再与录制视频合成，即：需要与特征检测和素材解析器发生关系，再与视频合成器发生关系，这是因为表情，动作，五官变化这种，坐标都是会变化的。具体的，终端通过应用（如相机应用）拍摄原始视频，在拍摄过程中，终端通过特征检测器来检测原始视频中每一帧图像中的人脸区域或者取景框内用户动作的特征，分析出具体的特征参数及其对应的特征坐标。特征坐标包括初始坐标和形变后的目的坐标。终端收到基于服务器匹配特征参数后发送的贴纸信息和贴纸信息的描述文件后，通过素材解析器对贴纸信

息和贴纸信息的描述文件进行解析，以得到贴纸信息及其属性和叠加位置和叠加时间点等信息，将贴纸信息按照特征坐标，贴纸信息的描述文件指示的叠加位置和叠加时间点等信息，通过视频合成器将贴纸信息与正在拍摄的原始视频进行视频合成，生成含有贴纸信息的视频处理结果。而除去触发类贴纸和五官类贴纸信息之外，其他几种（普通贴纸和背景贴纸），都是贴纸信息直接与录制视频合成，即：仅与视频合成器发生关系，因为坐标通常不会变，具体的，终端通过应用（如相机应用）拍摄原始视频，在拍摄过程中，终端收到基于服务器匹配视频中文字信息后发送的贴纸信息和贴纸信息的描述文件后，通过素材解析器对贴纸信息和贴纸信息的描述文件进行解析，以得到贴纸信息及其属性和叠加位置和叠加时间点等信息，将贴纸信息按照贴纸信息的描述文件指示的叠加位置和叠加时间点等信息，通过视频合成器将贴纸信息与正在拍摄的原始视频进行视频合成，生成含有贴纸信息的视频处理结果。

五，最终在录制时将动态贴纸一起录制到视频中，已完成视频录制。

在本申请所提供的几个实施例中，应该理解到，所揭露的设备和方法，可以通过其它的方式实现。以上所描述的设备实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，如：多个单元或组件可以结合，或可以集成到另一个系统，或一些特征可以忽略，或不执行。另外，所显示或讨论的各组成部分相互之间的耦合、或直接耦合、或通信连接可以是通过一些接口，设备或单元的间接耦合或通信连接，可以是电性的、机械的或其它形式的。

上述作为分离部件说明的单元可以是、或也可以不是物理上分开的，作为单元显示的部件可以是、或也可以不是物理单元，即可以位于一个地方，也可以分布到多个网络单元上；可以根据实际的需要选择其中的部分或全部单元来实现本实施例方案的目的。

另外，在本申请各实施例中的各功能单元可以全部集成在一个处理单元中，也可以是各单元分别单独作为一个单元，也可以两个或两个以上单元集成在一个单元中；上述集成的单元既可以采用硬件的形式实现，也可以采用

硬件加软件功能单元的形式实现。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：移动存储设备、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

或者，本申请上述集成的单元如果以软件功能模块的形式实现并作为独立的产品销售或使用时，也可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请实施例的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机、服务器、或者网络设备等）执行本申请各个实施例所述方法的全部或部分。而前述的存储介质包括：移动存储设备、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以所述权利要求的保护范围为准。

权 利 要 求

1、一种信息处理方法，其特征在于，包括:

终端获取第一操作，以触发第一媒体信息的采集;

终端在采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将检测到的表情变化或用户动作变化的变化量作为关键信息上报给服务器;

终端接收所述服务器推送的与所述关键信息对应的第二媒体信息；以及

将第一媒体信息和第二媒体信息进行视频合成。

2、根据权利要求 1 所述的方法，其特征在于，所述关键信息还包括：所述第一媒体信息中的文字信息；以及

所述方法还包括:

在采集所述第一媒体信息的过程中检测所述文字信息，并将检测的文字信息作为关键信息上报给所述服务器。

3、根据权利要求 2 所述的方法，其特征在于，将第一媒体信息和第二媒体信息进行视频合成之前，所述方法还包括:

终端接收服务器推送的与所述关键信息对应的第二媒体信息的描述文件。

4、根据权利要求 3 所述的方法，其特征在于，所述描述文件包括：所述第二媒体信息相对于第一媒体信息的位置，以及第二媒体信息的显示时间。

5、根据权利要求 4 所述的方法，其特征在于，所述将第一媒体信息和第二媒体信息进行视频合成包括:

根据所述描述文件将所述第一媒体信息与所述第二媒体信息进行视频合成，以在所述描述文件指定的显示时间内将所述第二媒体信息显示在所述描述文件指定的所述第一媒体信息的位置处。

6、根据权利要求 2 至 5 中任一项所述的方法，其特征在于，所述第二多媒体信息包括以下至少一类:

由所述表情变化或所述用户动作变化触发显示的第一类贴纸信息；以及

由所述文字信息触发显示的第二类贴纸信息。

7、根据权利要求 6 所述的方法，其特征在于，当所述第二多媒体信息为第一类贴纸信息时，所述将第一媒体信息和第二媒体信息进行视频合成包括:

确定所述表情变化或所述用户动作变化的特征初始坐标和特征目标坐标，以根据所述特征目标坐标定位的位置点或者由所述特征初始坐标至所述特征目标坐标确定的位置区域来确定叠加所述第一类贴纸信息的位置;

解析收到的所述第一类贴纸信息的描述文件，得到第一类贴纸信息的显示时间;

按照所述确定的位置以及所述解析的第一类贴纸信息的显示时间，将第一类贴纸信息与第一媒体信息进行视频合成。

8、根据权利要求 6 所述的方法，其特征在于，当所述第二多媒体信息为第二类贴纸信息时，所述将第一媒体信息和第二媒体信息进行视频合成包括:

解析收到的所述第二类贴纸信息的描述文件，得到第二类贴纸信息相对于第一媒体信息的位置，以及第二类贴纸信息的显示时间；以及

按照所述得到的位置和所述显示时间，将第二类贴纸信息与第一媒体信息进行视频合成。

9、一种终端，包括:

触发单元，用于获取第一操作，以触发第一媒体信息的采集;

检测单元，用于采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将检测到的表情变化或用户动作变化的变化量作为关键信息上报给服务器;

接收单元，用于接收所述服务器推送的与所述关键信息对应的第二媒体信息；以及

合成单元，用于将第一媒体信息和第二媒体信息进行视频合成。

10、根据权利要求 9 所述的终端，其特征在于，所述关键信息还包括:所述第一媒体信息中的文字信息；以及

所述检测单元，还用于在采集所述第一媒体信息的过程中检测所述文字信息，并将所述文字信息作为关键信息上报给所述服务器。

11、根据权利要求 10 所述的终端，其特征在于，所述接收单元，进一步用于：接收服务器推送的与所述关键信息对应的第二媒体信息的描述文件。

12、根据权利要求 11 所述的终端，其特征在于，所述描述文件包括：所述第二媒体信息相对于第一媒体信息的位置，以及第二媒体信息的显示时间。

13、根据权利要求 12 所述的终端，其特征在于，所述合成单元，进一步用于：

根据所述描述文件将所述第一媒体信息与所述第二媒体信息进行视频合成，以在所述描述文件指定的显示时间内将所述第二媒体信息显示在所述描述文件指定的所述第一媒体信息的位置处。

14、根据权利要求 10 至 13 中任一项所述的终端，其特征在于，所述第二多媒体信息包括以下至少一类：

由所述表情变化或所述用户动作变化触发显示的第一类贴纸信息；以及

由所述文字信息触发显示的第二类贴纸信息。

15、根据权利要求 14 所述的终端，其特征在于，当所述第二多媒体信息为第一类贴纸信息时，所述合成单元，进一步用于：

确定所述表情变化或所述用户动作变化的特征初始坐标和特征目标坐标，以根据所述特征目标坐标定位的位置点或者由所述特征初始坐标至所述特征目标坐标确定的位置区域来确定叠加所述第一类贴纸信息的位置；

解析收到的所述第一类贴纸信息的描述文件，得到第一类贴纸信息的显示时间；以及

按照所述确定的位置以及所述解析的第一类贴纸信息的显示时间，将第一类贴纸信息与第一媒体信息进行视频合成。

16、根据权利要求 14 所述的终端，其特征在于，当所述第二多媒体信息为第二类贴纸信息时，所述合成单元，进一步用于：

解析收到的所述第二类贴纸信息的描述文件，得到第二类贴纸信息相对于第一媒体信息的位置，以及第二类贴纸信息的显示时间；以及

按照所述得到的位置和所述显示时间，将第二类贴纸信息与第一媒体信息进行视频合成。

17、一种非易失性存储介质，存储有程序，当所述非易失性存储介质存储的程序被包括一个或多个处理器的计算机设备执行时，可使所述计算机设备执行如权利要求 1 至 8 中任一项所述信息处理方法。

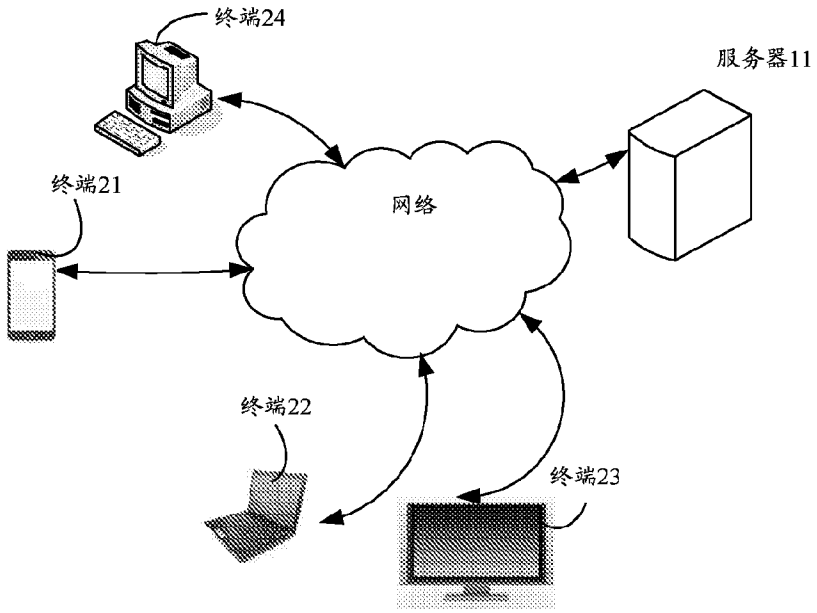

图 1

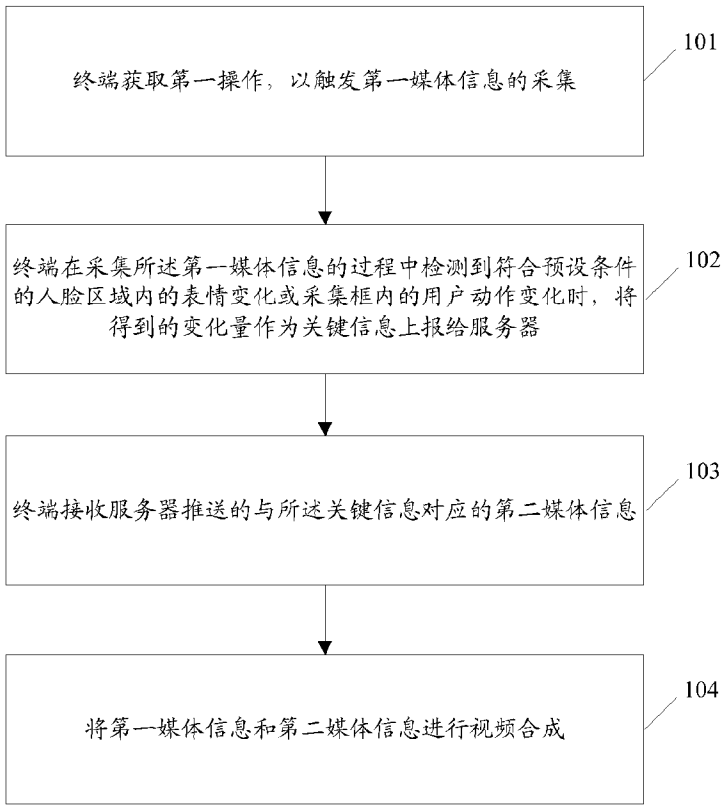
终端获取第一操作，以触发第一媒体信息的采集
101
终端在采集所述第一媒体信息的过程中检测到符合预设条件的人脸区域内的表情变化或采集框内的用户动作变化时，将得到的变化量作为关键信息上报给服务器
102
终端接收服务器推送的与所述关键信息对应的第二媒体信息
103
将第一媒体信息和第二媒体信息进行视频合成
104

图 2

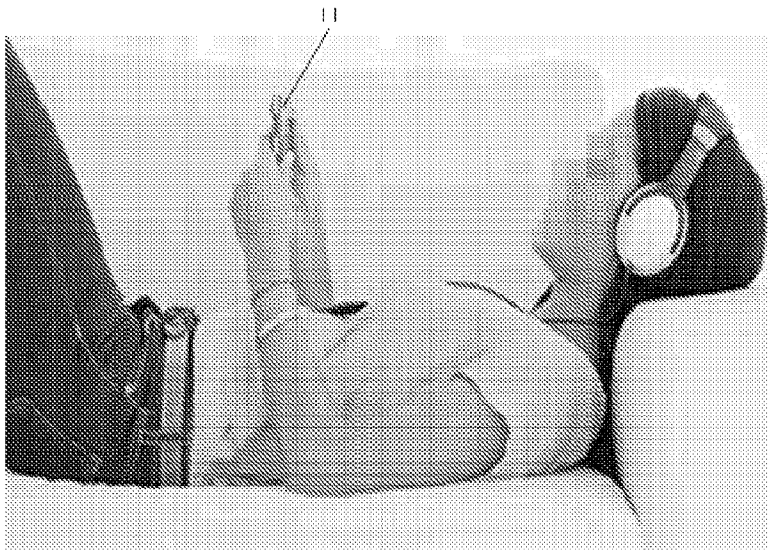
11

图 3

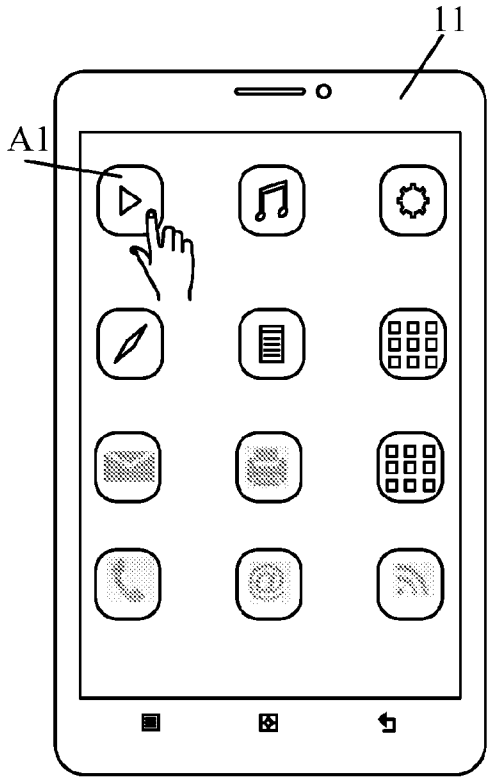

图 4

图 5

图 6

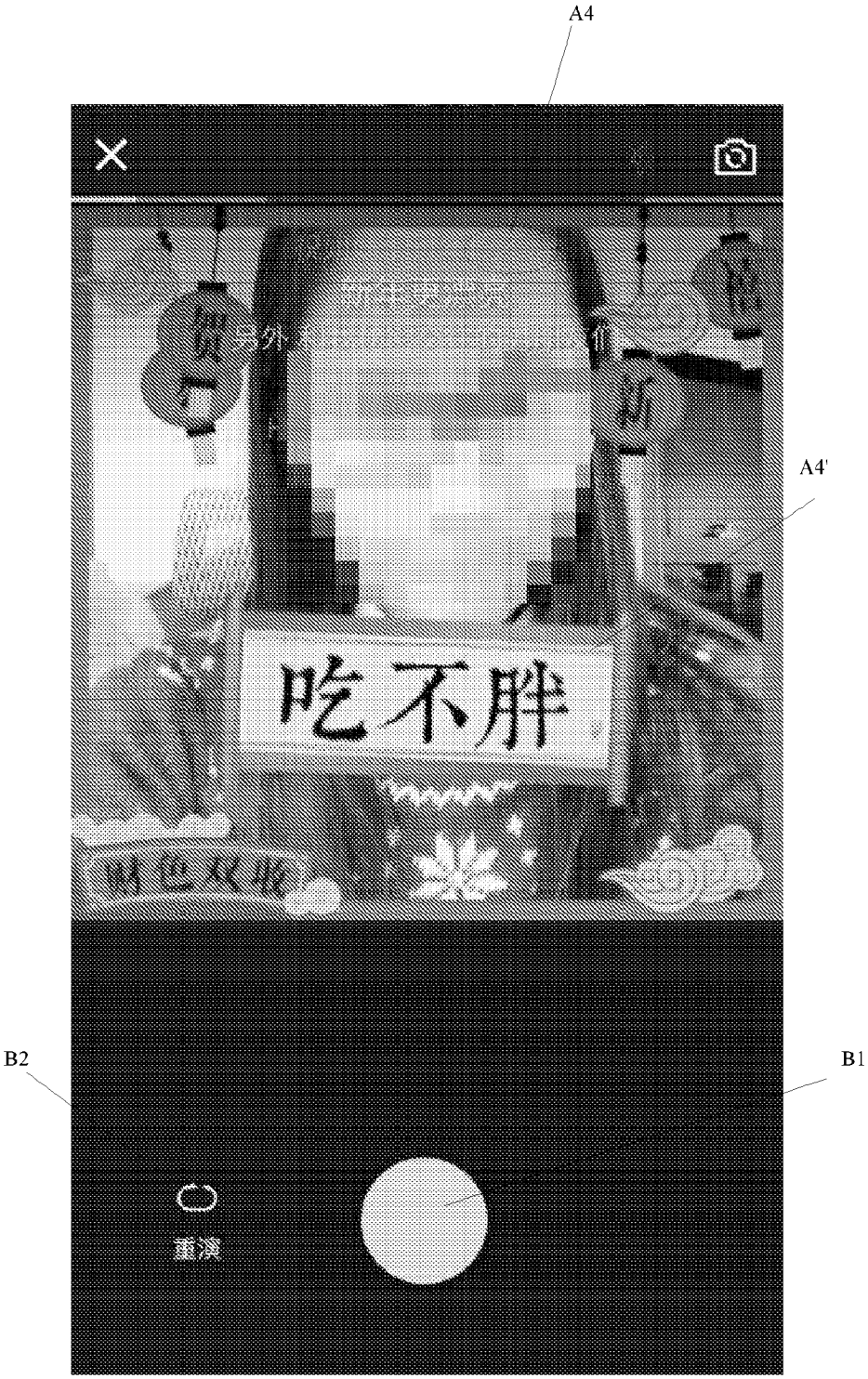

图 7

图 8

图 9

图 10

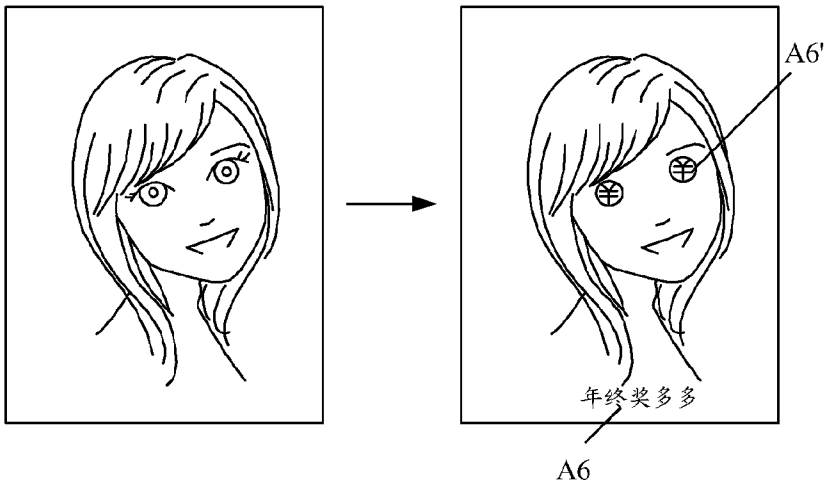

图 11

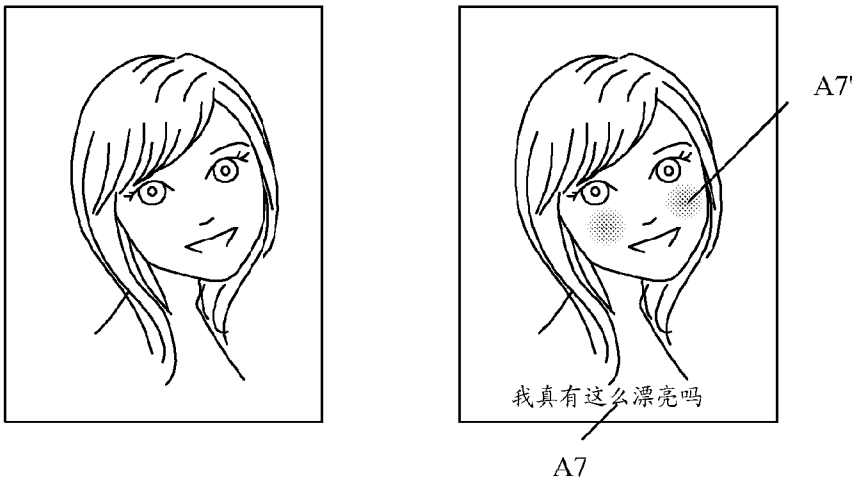
A7'
我真有这么漂亮吗
A7

图 12

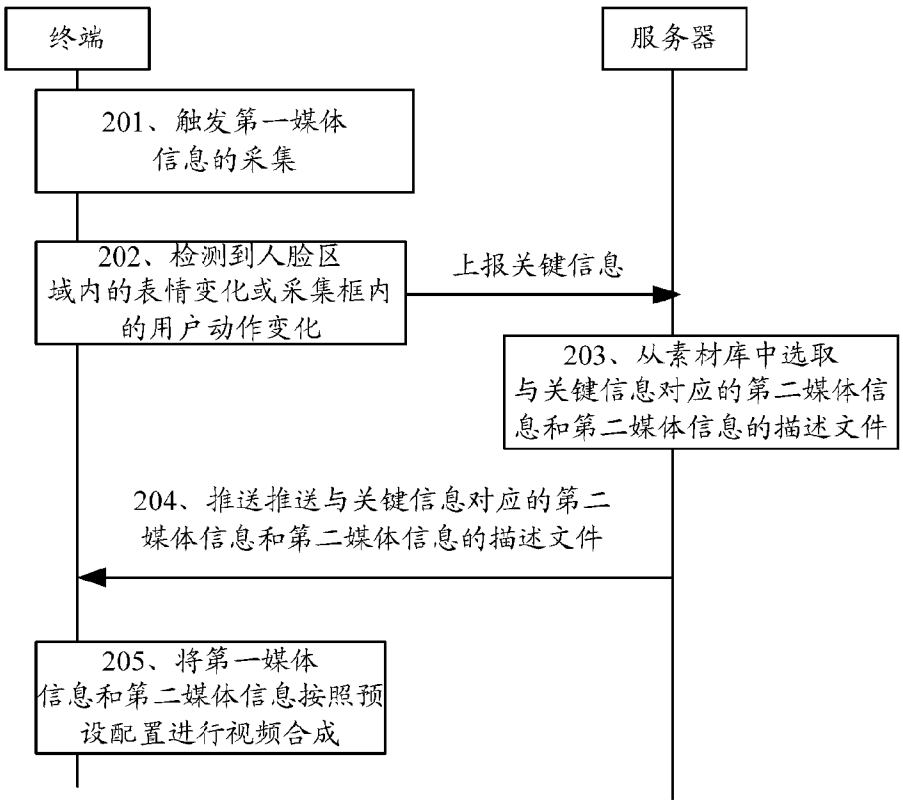
终端
服务器
201、触发第一媒体信息的采集
202、检测到人脸区域内的表情变化或采集框内的用户动作变化
上报关键信息
203、从素材库中选取与关键信息对应的第二媒体信息和第二媒体信息的描述文件
204、推送推送与关键信息对应的第二媒体信息和第二媒体信息的描述文件
205、将第一媒体信息和第二媒体信息按照预设配置进行视频合成

图 13

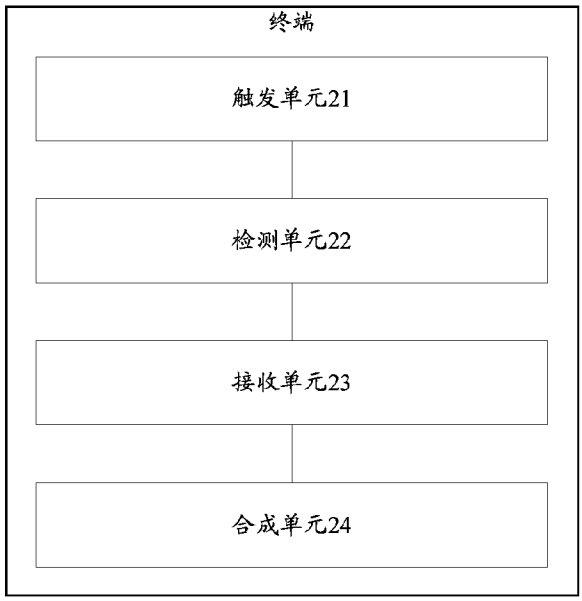
终端
触发单元21
检测单元22
接收单元23
合成单元24

图 14

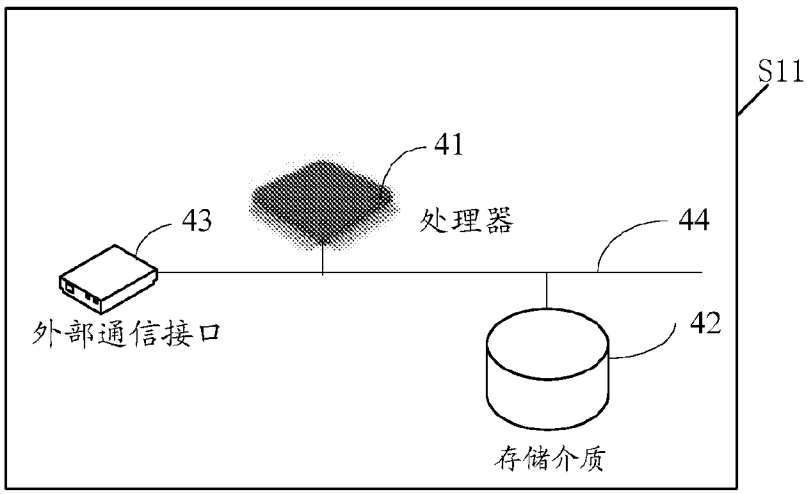
S11
41
处理器
43
44
外部通信接口
42
存储介质

图 15

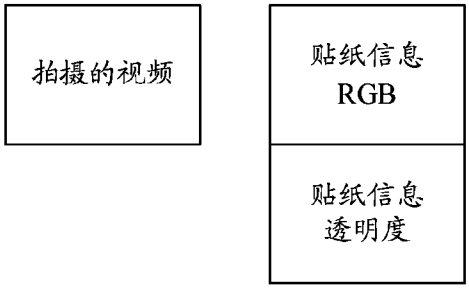
拍摄的视频
贴纸信息
RGB
贴纸信息
透明度

图 16

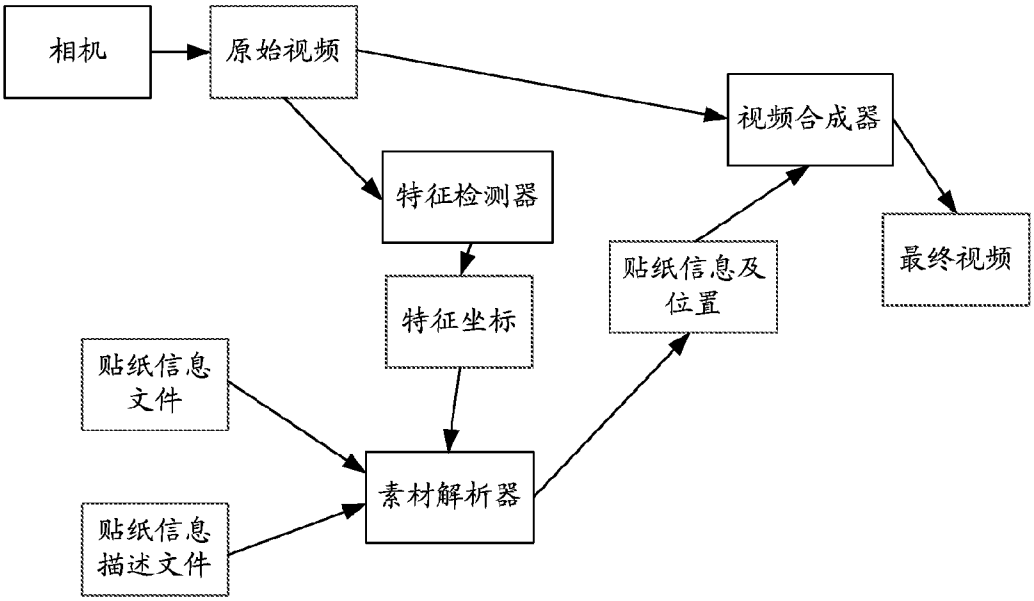
相机
原始视频
视频合成器
特征检测器
最终视频
贴纸信息及
位置
特征坐标
贴纸信息
文件
素材解析器
贴纸信息
描述文件

图 17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/076576

A. CLASSIFICATION OF SUBJECT MATTER

H04N 21/44 (2011.01) i; G06T 7/60 (2017.01) 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, G06T, G06F

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)

CNTXT; CNABS; DWPI; VEN, CNTXT, USTXT, EPTXT: flow, splicing, detection, sensing, determination, monitoring, posture, picture?, image?, media?, video, content, add+, combin+, compos+, overlay+, synthesiz+, emotion??, expression+, movement?, gesture

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2010066844 A (FUJIFILM C0RP.), 25 March 2010 (25.03.2010), figures 1-9, and description, paragraphs [0015]-[0036]	1-17
X	TW 201021550 A (ALTEK CORP.), 01 June 2010 (01.06.2010), figures 1 and 3, and description, page 8, line 8 to page 11, line 5	1-17
Y	CN 101453573 A (OLYMPUS IMAGING CORP.), 10 June 2009 (10.06.2009), claim 12, and description, page 10, line 14 to page 12, line 12	1-17
Y	CN 102427553 A (TCL CORPORATION), 25 April 2012 (25.04.2012), description, paragraphs [0080]-[0092]	1-17
PX	CN 105791692 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 20 July 2016 (20.07.2016), the whole document	1-17
A	CN 1870744 A (OKI ELECTRIC INDUSTRY CO., LTD.), 29 November 2006 (29.11.2006), the whole document	1-17

☐ 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	Date of mailing of the international search report
07 April 2017 (07.04.2017)	**01 June 2017 (01.06.2017)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **JIN, Yuan** Telephone No.: (86-10) **62411447**

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/076576

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2010066844 A	25 March 2010	None	
TW 201021550 A	01 June 2010	US 2010123804 A1	20 May 2010
CN 101453573 A	10 June 2009	CN 101453573 B	29 September 2010
		JP 2009141516 A	25 June 2009
CN 102427553 A	25 April 2012	None	
CN 105791692 A	20 July 2016	None	
CN 1870744 A	29 November 2006	KR 20060121679 A	29 November 2006
		US 2006281064 A1	14 December 2006
		JP 2006330958 A	07 December 2006
		TW 200703053 A	16 January 2007

国际检索报告

国际申请号
PCT/CN2017/076576

A. 主题的分类

H04N 21/44(2011.01)i; G06T 7/60(2017.01)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,G06T,G06F

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））

CNTXT;CNABS;DWPI;VEN,CNTXT,USTXT,EPTXT:图像,媒体,视频,流,内容,添加,合成,叠加,拼接,附加,检测,感测,确定,探测,监测,表情,动作,姿势,手势,picture?, image?, media?, video, content, add+, combin+,compos+, overlay+,synthesiz+,emotion??, expression+, movement?, gesture

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
X	JP 2010066844 A (FUJIFILM CORP) 2010年 3月 25日 (2010-03-25) 图1-9，说明书第0015-0036段	1-17
X	TW 201021550 A (ALTEK CORP) 2010年 6月 1日 (2010-06-01) 图1，3，说明书第8页第8行－第11页第5行	1-17
Y	CN 101453573 A (奥林巴斯映像株式会社) 2009年 6月 10日 (2009-06-10) 权利要求12，说明书第10页第14行－第12页第12行	1-17
Y	CN 102427553 A (TCL集团股份有限公司) 2012年 4月 25日 (2012-04-25) 说明书第0080－0092段	1-17
PX	CN 105791692 A (腾讯科技深圳有限公司) 2016年 7月 20日 (2016-07-20) 全文	1-17
A	CN 1870744 A (冲电气工业株式会社) 2006年 11月 29日 (2006-11-29) 全文	1-17

☐ 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	国际检索报告邮寄日期
2017年 4月 7日	2017年 6月 1日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 金源 电话号码 (86-10)62411447

表 PCT/ISA/210 (第2页) (2009年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/076576

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)
JP	2010066844	A	2010年 3月 25日	无			
TW	201021550	A	2010年 6月 1日	US	2010123804	A1	2010年 5月 20日
CN	101453573	A	2009年 6月 10日	CN	101453573	B	2010年 9月 29日
				JP	2009141516	A	2009年 6月 25日
CN	102427553	A	2012年 4月 25日	无			
CN	105791692	A	2016年 7月 20日	无			
CN	1870744	A	2006年 11月 29日	KR	20060121679	A	2006年 11月 29日
				US	2006281064	A1	2006年 12月 14日
				JP	2006330958	A	2006年 12月 7日
				TW	200703053	A	2007年 1月 16日