

(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
2024 年 12 月 19 日 (19.12.2024)



(10) 国际公布号
WO 2024/255839 A1

- (51) 国际专利分类号:
H04N 21/43 (2011.01)
- (21) 国际申请号: PCT/CN2024/099198
- (22) 国际申请日: 2024 年 6 月 14 日 (14.06.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202310715849.6 2023年6月15日 (15.06.2023) CN
- (71) 申请人: 北京字跳网络技术有限公司
(**BEIJING ZITIAO NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市海淀区紫金数码园4号楼2层0207, Beijing 100190 (CN)。
- (72) 发明人: 张瑞(**ZHANG, Rui**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局, Beijing 100028 (CN)。 龚立雪(**GONG, Lixue**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局, Beijing 100028 (CN)。 陈璐双(**CHEN, Lushuang**); 中国北

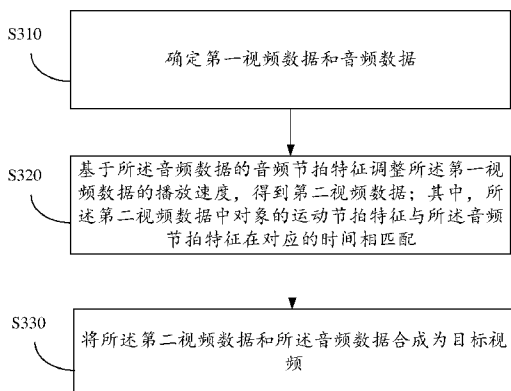
京市朝阳区七圣中街12号院融中心B1小邮局, Beijing 100028 (CN)。 关有国(**GUAN, Youguo**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局, Beijing 100028 (CN)。

(74) 代理人: 北京市金杜律师事务所 (**KING & WOOD MALLESONS**); 中国北京市朝阳区东三环中路1号环球金融中心办公楼东楼20层, Beijing 100020 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(54) **Title:** VIDEO PROCESSING METHOD AND APPARATUS, AND DEVICE, MEDIUM AND PROGRAM PRODUCT

(54) 发明名称: 视频处理方法、装置、设备、介质及程序产品



- S310 Determine first video data and audio data
- S320 Adjust the playing speed of the first video data on the basis of audio beat features of the audio data, so as to obtain second video data, wherein motion beat features of an object in the second video data match the audio beat features at corresponding times
- S330 Synthesize the second video data and the audio data into a target video

300

图 3

(57) **Abstract:** Provided in the present disclosure are a video processing method and apparatus, and a device, a storage medium and a program product. The method comprises: determining first video data and audio data; adjusting the playing speed of the first video data on the basis of audio beat features of the audio data, so as to obtain second video data, wherein visual beat features of an object in the second video data match the audio beat features at corresponding times; and synthesizing the second video data and the audio data into a target video.

(57) **摘要:** 本公开提供一种视频处理方法、装置、设备、存储介质及程序产品。该方法包括: 确定第一视频数据和音频数据; 基于所述音频数据的音频节拍特征调整所述第一视频数据的播放速度, 得到第二视频数据; 其中, 所述第二视频数据中对象的视觉节拍特征与所述音频节拍特征在对应的时间相匹配; 将所述第二视频数据和所述音频数据合成为目标视频。

[见续页]

WO 2024/255839 A1

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))。

视频处理方法、装置、设备、介质及程序产品

相关申请的交叉引用

本申请要求于 2023 年 6 月 15 日提交的，申请号为 202310715849.6、发明名称为“视频处理方法、装置、设备、介质及程序产品”的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本公开涉及计算机技术领域，尤其涉及一种视频处理方法、装置、设备、介质及程序产品。

10

背景技术

用户希望发布的视频数据的视觉效果和音乐的节奏相匹配。现有的应用程序虽然可以合成视频数据和音频数据，然而却只能从给定的音乐数据库中选择音乐进行合成，无法自由地选择其他音乐。而且，合成之后只有视频画面切换处的视觉效果于音乐的节奏匹配，无法使得整个视频数据的视觉效果与整个音频数据的节奏相匹配，导致最终的音视频数据的呈现效果不佳，不能满足用户的需求。

15

发明内容

本公开提出一种视频处理方法、装置、设备、存储介质及程序产品，以在一定程度上解决提供给视频数据的视觉效果与音频数据的节奏匹配不准确的技术问题。

20

本公开第一方面，提供了一种视频处理方法，包括：

确定第一视频数据和音频数据；

基于所述音频数据的音频节拍特征调整所述第一视频数据的播放速度，得到第二视频数据；其中，所述第二视频数据中对象的视觉节拍特征与所述音频节拍特征在对应的时间相匹配；

25

将所述第二视频数据和所述音频数据合成为目标视频。

本公开第二方面，提供了一种视频处理装置，包括：

数据确定模块，用于确定第一视频数据和音频数据；

视频处理模块，用于基于所述音频数据的音频节拍特征调整所述第一视频数据的播放速度，得到第二视频数据；其中，所述第二视频数据中对象的视觉
5 节拍特征与所述音频节拍特征在对应的时间相匹配；

合成模块，用于将所述第二视频数据和所述音频数据合成为目标视频。

本公开第三方面，提供了一种电子设备，其特征在于，包括一个或者多个
处理器、存储器；和一个或多个程序，其中所述一个或多个程序被存储在所述
10 存储器中，并且被所述一个或多个处理器执行，所述程序包括用于执行根据第
一方面或第二方面所述的方法的指令。

本公开第四方面，提供了一种包含计算机程序的非易失性计算机可读存储
介质，当所述计算机程序被一个或多个处理器执行时，使得所述处理器执行第
15 一方面或第二方面所述的方法。

本公开第五方面，提供了一种计算机程序产品，包括计算机程序指令，当
所述计算机程序指令在计算机上运行时，使得计算机执行第一方面所述的方法。

附图说明

为了更清楚地说明本公开或相关技术中的技术方案，下面将对实施例或相
关技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附
图仅仅是本公开的实施例，对于本领域普通技术人员来讲，在不付出创造性劳
25 动的前提下，还可以根据这些附图获得其他的附图。

图1为本公开实施例的视频处理架构的示意图。

图2为本公开实施例的示例性电子设备的硬件结构示意图。

图3为本公开实施例的视频处理方法的示意性流程图。

图4为本公开实施例的运动变化特征和节拍特征的时间对齐的示意性原理
30 图。

图5为本公开实施例的音频频谱图的示例。

图 6 为本公开实施例的音频节拍图的示例。

图 7 为本公开实施例的运动频谱图的示例。

图 8 为本公开实施例的视觉节拍图的示例。

图 9 为本公开实施例的节拍特征对齐的示例。

5 图 10 为本公开实施例的视频处理方法的示例。

图 11 为本公开实施例的视频处理装置的示意图。

具体实施方式

为使本公开的目的、技术方案和优点更加清楚明白，以下结合具体实施
10 例，并参照附图，对本公开进一步详细说明。

需要说明的是，除非另外定义，本公开实施例使用的技术术语或者科学术
语应当为本公开所属领域内具有一般技能的人士所理解的通常意义。本公开实
施例中使用的“第一”、“第二”以及类似的词语并不表示任何顺序、数量或者重
要性，而只是用来区分不同的组成部分。“包括”或者“包含”等类似的词语意指
15 出现该词前面的元件或者物件涵盖出现在该词后面列举的元件或者物件及其
等同，而不排除其他元件或者物件。“连接”或者“相连”等类似的词语并非限定
于物理的或者机械的连接，而是可以包括电性的连接，不管是直接的还是间接
的。“上”、“下”、“左”、“右”等仅用于表示相对位置关系，当被描述对象的绝
对位置改变后，则该相对位置关系也可能相应地改变。

20 可以理解的是，在使用本公开各实施例公开的技术方案之前，均应当依据
相关法律法规通过恰当的方式对本公开所涉及个人信息类型、使用范围、使
用场景等告知用户并获得用户的授权。

例如，在响应于接收到用户的主动请求时，向用户发送提示信息，以明确
地提示用户，其请求执行的操作将需要获取和使用到用户的个人信息。从而，
25 使得用户可以根据提示信息来自主地选择是否向执行本公开技术方案的操作
的电子设备、应用程序、服务器或存储介质等软件或硬件提供个人信息。

作为一种可选的但非限定性的实现方式，响应于接收到用户的主动请求，
向用户发送提示信息的方式例如可以是弹窗的方式，弹窗中可以以文字的方式
呈现提示信息。此外，弹窗中还可以承载供用户选择“同意”或者“不同意”向电
30 子设备提供个人信息的选择控件。

可以理解的是，上述通知和获取用户授权过程仅是示意性的，不对本公开

的实现方式构成限定，其它满足相关法律法规的方式也可应用于本公开的实现方式中。

图 1 示出了本公开实施例的视频处理架构的示意图。参考图 1，该视频处理架构 100 可以包括服务器 110、终端 120 以及提供通信链路的网络 130。服务器 110 和终端 120 之间可通过有线或无线的网络 130 连接。其中，服务器 110 可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、安全服务、CDN 等基础云计算服务的云服务器。

终端 120 可以是硬件或软件实现。例如，终端 120 为硬件实现时，可以是具有显示屏并且支持页面显示的各种电子设备，包括但不限于智能手机、平板电脑、电子书阅读器、膝上型便携计算机和台式计算机等等。终端 120 设备为软件实现时，可以安装在上述所列举的电子设备中；其可以实现成多个软件或软件模块(例如用来提供分布式服务的软件或软件模块)，也可以实现成单个软件或软件模块，在此不做具体限定。

需要说明的是，本申请实施例所提供的视频处理方法可以由终端 120 来执行，也可以由服务器 110 来执行。应了解，图 1 中的终端、网络和服务器的数目仅为示意，并不旨在对其进行限制。根据实现需要，可以具有任意数目的终端、网络和服务器的。

图 2 示出了本公开实施例所提供的示例性电子设备 200 的硬件结构示意图。如图 2 所示，电子设备 200 可以包括：处理器 202、存储器 204、网络模块 206、外围接口 208 和总线 210。其中，处理器 202、存储器 204、网络模块 206 和外围接口 208 通过总线 210 实现彼此之间在电子设备 200 的内部的通信连接。

处理器 202 可以是中央处理器 (Central Processing Unit, CPU)、视频处理器、神经网络处理器 (NPU)、微控制器 (MCU)、可编程逻辑器件、数字信号处理器 (DSP)、应用专用集成电路 (Application Specific Integrated Circuit, ASIC)、或者一个或多个集成电路。处理器 202 可以用于执行与本公开描述的技术相关的功能。在一些实施例中，处理器 202 还可以包括集成为单一逻辑组件的多个处理器。例如，如图 2 所示，处理器 202 可以包括多个处理器 202a、202b 和 202c。

存储器 204 可以配置为存储数据 (例如，指令、计算机代码等)。如图 2

所示，存储器 204 存储的数据可以包括程序指令（例如，用于实现本公开实施例的视频处理方法的程序指令）以及要处理的数据（例如，存储器可以存储其他模块的配置文件等）。处理器 202 也可以访问存储器 204 存储的程序指令和数据，并且执行程序指令以对要处理的数据进行操作。存储器 204 可以包括易失性存储装置或非易失性存储装置。在一些实施例中，存储器 204 可以包括随机访问存储器（RAM）、只读存储器（ROM）、光盘、磁盘、硬盘、固态硬盘（SSD）、闪存、存储棒等。

网络模块 206 可以配置为经由网络向电子设备 200 提供与其他外部设备的通信。该网络可以是能够传输和接收数据的任何有线或无线的网络。例如，该网络可以是有线网络、本地无线网络（例如，蓝牙、WiFi、近场通信（NFC）等）、蜂窝网络、因特网、或上述的组合。可以理解的是，网络的类型不限于上述具体示例。在一些实施例中，网络模块 306 可以包括任意数量的网络接口控制器（NIC）、射频模块、接收发器、调制解调器、路由器、网关、适配器、蜂窝网络芯片等的任意组合。

外围接口 208 可以配置为将电子设备 200 与一个或多个外围装置连接，以实现信息输入及输出。例如，外围装置可以包括键盘、鼠标、触摸板、触摸屏、麦克风、各类传感器等输入设备以及显示器、扬声器、振动器、指示灯等输出设备。

总线 210 可以被配置为在电子设备 200 的各个组件（例如处理器 202、存储器 204、网络模块 206 和外围接口 208）之间传输信息，诸如内部总线（例如，处理器-存储器总线）、外部总线（USB 端口、PCI-E 总线）等。

需要说明的是，尽管上述电子设备 200 的架构仅示出了处理器 202、存储器 204、网络模块 206、外围接口 208 和总线 210，但是在具体实施过程中，该电子设备 200 的架构还可以包括实现正常运行所必需的其他组件。此外，本领域的技术人员可以理解的是，上述电子设备 200 的架构中也可以仅包含实现本公开实施例方案所必需的组件，而不必包含图中所示的全部组件。

用户希望发布的音视频数据的视觉效果和音乐的节奏相匹配，可以称之为视频与音频的“卡点”。而现有的应用程序所提供的卡点功能只能从给定曲库选择，无法自由添加喜欢的音乐，且只将卡点用在视频转场处，对视频只做裁剪无节奏变换；或者，对视频进行固定间隔的变速，例如 500ms 区间内先加速再减速，变速后区间长度大致不变，无法使得整个视频数据的视觉效果与整个

音频数据的节奏相匹配，导致最终的音视频数据的呈现效果不佳，不能满足用户的需求。因此，如何使视频的视觉效果与音频的节奏相匹配成为了亟需解决的技术问题。

鉴于此，本公开实施例提供了一种视频处理方法、装置、设备、存储介质及程序产品。通过调整视频数据的视频帧之间的播放速度，使得视频数据中对象的视觉节拍特征与音频数据中的音频节拍特征相匹配。使得最终合成的视频数据的视觉节奏与音频数据的节奏同步，提高了音视频数据的呈现效果，提高用户的使用体验。

具体地，由于视频数据的视觉效果往往通过其所显示的对象呈现，则可以基于图像中对象的视觉节拍特征来表征视频数据的视频节奏，例如视频帧之间运动幅度、方向等。然后根据音频数据的音频节拍特征，调整视频数据中视频帧的播放速度，使得视觉节拍特征出现的时间发生变化，变为与音频节拍特征出现的时间同步，从而完成视频节奏与音频节奏的同步，实现任意视频和音频之间的节奏匹配。

参见图 3，图 3 示出了根据本公开实施例的视频处理方法的示意性流程图。本公开实施例的视频处理方法可以部署于客户端。图 3 中，视频处理方法 300 可以进一步包括如下步骤。

在步骤 S310，确定第一视频数据和音频数据。

其中，第一视频数据和音频数据可以来自同一音视频数据。在一些实施例中，用户可以选择第一音视频数据 A，然后基于第一音视频数据得到第一视频数据 A1 和音频数据 A2。例如，针对音视频数据 A，用户认为其中的视频视频节奏与音频节奏不匹配，则可以根据本公开实施例的方法，对第一音视频数据 A 进行处理，得到视频节奏与音频节奏向匹配的音视频数据。

第一视频数据和音频数据也可以来自不同的音视频数据。在一些实施例中，用户可以选择第一音视频数据 B，然后基于第一音视频数据得到第一视频数据 B1；以及用户可以选择音频数据 C。例如，针对音视频数据 B，用户希望采用音频数据 C 来替换音视频数据 B 中的音频数据，但替换后第一数据时间 B1 的视频节奏将与音频数据 C 的音频节奏不匹配，则可以根据本公开实施例的方法，对第一视频数据 B1 和音频数据 C 进行处理，使其视频节奏与音频节奏向匹配。

在步骤 S320，基于所述音频数据的音频节拍特征调整所述第一视频数据的

播放速度，得到第二视频数据；其中，所述第二视频数据中对象的视觉节拍特征与所述音频节拍特征在对应的时间相匹配。

其中，音频节拍特征可以表征音频数据的节奏。音频节拍特征具有强度，其中，强度大于或等于强度阈值的节拍特征可以是重节拍，强度小于强度阈值的节拍特征可以是轻节拍。视觉节拍特征可以表征视频数据的视频节奏；视觉节拍特征具有幅度，幅度越大表示视频节奏越强，幅度越小表示视频节奏越弱。

在一些实施例中，基于所述音频数据的音频节拍特征调整所述第一视频数据的播放速度，得到第二视频数据，包括：

基于所述音频数据进行音频节拍分析，得到所述音频节拍特征；

基于所述第一视频数据的光流信息，得到所述第一视频数据中对象的视觉节拍特征；

调整所述第一视频数据中视频帧之间的播放速度，以将所述视觉节拍特征与所述音频节拍特征对应的时间对齐，得到所述第二视频数据。

具体地，参见图 4，图 4 示出了根据本公开实施例的运动变化特征和节拍特征的时间对齐的示意性原理图。如图 4 所示，可以对音频数据进行节拍分析

(例如频谱分析等)，得到音频数据关于时间的音频节拍特征：时间 t_0 处的音频节拍特征 m_0 、时间 t_1 处的音频节拍特征 m_1 、时间 t_2 处的音频节拍特征 m_2 、……、时间 t_i 处的音频节拍特征 m_i (i 为自然数)、……。可以采用光流检测等技术得到第一视频数据中对象的光流信息，进而得到第一视频数据关于

时间的视觉节拍特征：时间 t_0' 处的视觉节拍特征 n_0 、时间 t_1' 处的视觉节拍特征 n_1 、时间 t_2' 处的视觉节拍特征 n_2 、……、时间 t_i' 处的视觉节拍特征 n_i 、……。

视觉节拍特征 n_0 、 n_1 、 n_2 、……、 n_i 、……可以分别对应视频帧 F_{n0} 、 F_{n1} 、 F_{n2} 、……、 F_{ni} 、……，视频帧 F_{n0} 、 F_{n1} 之间的播放速度为 v_0 ，视频帧 F_{n1} 、 F_{n2} 之间的播放速度为 v_1 ，……，视频帧 F_{ni} 、 F_{ni+1} 之间的播放速度为 v_i ，……。

可见，视觉节拍特征 m_0 - m_i 与视觉节拍特征 n_0 - n_i 之间并不同步，则可以改变视频帧之间的播放速度，以改变运动变化特征 n_0 - n_i 出现的时间点 t_0' - t_i' ，使得视觉节拍特征 n_0 - n_i 与音频节拍特征 m_0 - m_i 出现的时间点对齐，对齐后的视觉节拍特征 n_0 - n_i 分别对应于时间点 t_0 - t_i 。相应地，对齐后视觉节拍特征 n_0 - n_i 对应的视频帧 F_{n0} - F_{ni} 之间的播放速度也变为：视频帧 F_{n0} 、 F_{n1} 之间的播放速度为 v_0' ，视频帧 F_{n1} 、 F_{n2} 之间的播放速度为 v_1' ，……，视频帧 F_{ni} 、 F_{ni+1} 之间的播放速度为 v_i' ，……，即由播放速度曲线 v_0 - v_i 变为播放速度曲线

$v_0'-v_i'$ 。

在一些实施例中，基于所述音频数据进行音频节拍分析，得到所述音频数据的音频节拍特征，包括：

对所述音频数据进行频率分析得到所述音频数据关于时间的频谱特征，所述频谱特征包括音频频率和对应的声音强度；

基于所述频谱特征进行音频节拍识别，得到关于时间的音频节拍强度特征；

对所述音频节拍强度特征进行基于移动均线的峰值检测，确定峰值所对应的所述音频节拍强度特征为所述音频节拍特征。

具体地，频率分析可以包括傅里叶变换，例如短时傅里叶变换，将时域的音频数据转换为频域的音频频谱图。音频频谱图可以指音频的频率随时间变化的图像，能够描述某个时刻的各个频率处声音的分布情况。参见图 5，图 5 示出了根据本公开实施例的音频频谱图的示例。图 5 中，音频频谱图的横坐标可以是时间（单位为 s），纵坐标可以是频率（单位为 HZ），对于某个时刻 t 的某个频率（例如 0-8000HZ 中的某个频率）处，可以采用不同的颜色来表示不同的声音强度（单位为 dB）。

可以采用训练好的音频节拍识别网络对频谱特征进行音频节拍识别。其中，音频节拍识别网络可以基于训练数据对神经网络进行训练得到，训练数据可以包括音频频谱训练图和对应的音频节拍训练图，将音频频谱训练图作为输入层数据以及将对应的音频节拍训练图作为输出层数据对神经网络进行训练，可以得到训练好的音频节拍识别网络。具体地，可以基于音频节拍识别网络对音频频谱图中的频谱特征进行音频节拍识别，得到包括关于时间的音频节拍强度特征的音频节拍图，如图 6 所示，图 6 示出了根据本公开实施例的音频节拍图的示例。图 6 中，音频节拍图的横坐标可以是时间（单位为 s），纵坐标可以是音频节拍强度，某时刻的音频节拍强度越高表示该时刻为音频节拍点的可能性越高。

接着，可以对音频节拍强度特征进行基于移动均线的峰值检测，从而将得到的峰值所对应的音频节拍强度特征确定为音频节拍特征，如图 6 中的音频节拍特征 610。音频节拍点通常是频率和/或音量突变的点，据此可以基于移动均线的峰值检测确定最终的音频节拍特征点。具体地，针对每个时刻选取该时刻之前一定数量（例如当前时刻之前的若干个时间窗口）和/或该时刻之后一定数

量（例如当前时刻之后的若干个时间窗口）的音频节拍强度计算平均值，即可得到音频节拍移动均线。将该音频节拍移动均线对应的平均值作为阈值，判断每个时刻的音频节拍强度是否大于该时刻对应的阈值，如果音频节拍强度大于该时刻对应的阈值，则该时刻为音频节拍点，对应的音频节拍强度特征为音频节拍特征。

在一些实施例中，基于所述第一视频数据的光流信息，得到所述第一视频数据中对象的视觉节拍特征，包括：

针对所述第一视频数据进行光流检测，得到所述第一视频数据的中像素的光流信息；

基于所述光流信息得到所述第一视频数据关于时间的光流分布特征，所述光流分布特征包括多个光流方向和每个所述光流方向上对应的光流强度；

基于所述运动分布特征进行视觉节拍分析，得到所述第一视频数据中对象的视觉节拍特征。

具体地，可以将视频数据中移动对象的显著不连续变化作为视频局部显著峰值，并将其确定为视觉节拍特征。据此，可以采用能够反映视频帧之间的运动幅值方向等视觉特征的光流信息来进行视觉节拍分析。光流信息越大则视觉节拍特征的幅值越大，对应的视频节奏越强。据此，可以基于光流信息将视频数据进行视觉节拍识别处理得到与图5-图6中的音频数据的音频频谱图和音频节拍图类似的运动频谱图和视觉节拍图，如图7-图8所示，图7示出了根据本公开实施例的光流频谱图的示例，图8示出了根据本公开实施例的视觉节拍图的示例。

在一些实施例中，基于所述光流信息得到所述第一视频数据关于时间的光流分布特征，包括：

基于所述光流信息的角度得到每个所述像素的像素运动方向，以及基于所述光流信息的幅值得到每个所述像素的像素运动幅值；

针对每个所述光流方向，统计与所述光流方向的角度差在预设范围内的所述像素运动方向所对应的所述像素运动幅值之和，得到所述光流方向上对应的光流强度。

其中，如图7所示，光流频谱图的横坐标可以包括时间（单位为s），纵坐标可以包括光流方向，该光流方向可以是预设的。对于某一时刻的某个光流方向处，可以采用不同的颜色来表示不同的光流强度。具体地，对于视频数据中

的每个像素 (x, y) 在时刻 t 的光流信息 F_t ，其像素运动方向可以表示为光流信息 F_t 的角度 ϕ ，像素运动幅值可以表示为光流信息 F_t 的幅值 $|F_t(x, y)|$ 。统计所有像素的像素运动方向及像素运动幅值，将与同一光流方向的角度差在预设范围内的运动幅值求和，从而得到在各个光流方向上的光流幅值分布。

5 即，图 7 中的光流频谱图中的某一点 $D(t, \theta)$ 可以表示为：

$$D(t, \theta) = \sum_{x, y} |F_t(x, y)| \mathbb{1}_{\theta}(\angle F_t(x, y))$$

其中， $\mathbb{1}_{\theta}(\phi) := \begin{cases} 1 & \text{if } |\theta - \phi| \leq \frac{2\pi}{N_{bins}} \\ 0 & \text{otherwise} \end{cases}$ ， θ 为光流方向， ϕ 为像素运动方向，

N_{bins} 表示光流方向的个数，如图 7 中的纵坐标所示，图 7 中光流频谱图包括 6 个预设的光流方向。通过统计在光流方向 θ 附近（例如与光流方向 θ 相差预设角度 $2\pi/N_{bins}$ ）的所有光流信息的像素运动幅值之和，可以得到在光流方向 θ 上的光流强度。

在一些实施例中，基于所述光流分布特征进行视觉节拍分析，得到所述视觉节拍特征，包括：

15 基于所述光流分布特征进行视觉节拍识别，得到关于时间的视觉节拍强度特征；

对所述视觉节拍强度特征进行基于移动均线的峰值检测，确定峰值所对应的所述视觉节拍强度特征为所述视觉节拍特征。

具体地，与音频节拍识别过程类似，可以采用训练好的视觉节拍识别网络对光流分布特征进行视觉节拍识别。其中，视觉节拍识别网络可以基于视觉节拍训练数据对神经网络进行训练得到，视觉节拍训练数据可以包括用作输入层训练数据的光流频谱训练图和对应的用作输出层训练数据的视觉节拍训练图。基于该视觉节拍训练数据对神经网络进行训练，可以得到训练好的视觉节拍识别网络。具体地，可以基于视觉节拍识别网络对光流频谱图中的光流分布特征进行视觉节拍识别，得到包括关于时间的视觉节拍强度特征的视觉节拍图，如图 8 所示。图 8 中，视觉节拍图的横坐标可以是时间（单位为 s），纵坐标可以是视觉节拍强度，某时刻的视觉节拍强度越高表示该时刻为视觉节拍点的可能性越高。

接着，可以对视觉节拍强度特征进行基于移动均线的峰值检测，从而将得到的峰值所对应的视觉节拍强度特征确定为视觉节拍特征，如图 8 中的视觉节

拍特征 810。具体地，针对每个时刻选取该时刻之前一定数量（例如当前时刻之前的若干个时间窗口）和/或该时刻之后一定数量（例如当前时刻之后的若干个时间窗口）的视觉节拍强度计算平均值，即可得到视觉节拍移动均线。将该视觉节拍移动均线对应的平均值作为阈值，判断每个时刻的视觉节拍强度是否大于该时刻对应的阈值，如果视觉节拍强度大于该时刻对应的阈值，则该时刻为视觉节拍点，对应的视觉节拍强度特征为视觉节拍特征。

在一些实施例中，调整所述第一视频数据中视频帧之间的播放速度，以将所述视觉节拍特征与所述音频节拍特征对应的时间对齐，包括：

对所述视觉节拍特征和所述音频节拍特征进行过滤，得到过滤后的视觉节拍特征和音频节拍特征；

调整所述第一视频数据中视频帧的播放速度，使得过滤后的所述视觉节拍特征与所述音频节拍特征依次对齐。

其中，可以对视觉节拍特征、音频节拍特征进行过滤，例如剔除节拍间隔时间过短的点，使得两种节拍特征数量一致。例如，参见图 9，图 9 示出了根据本公开实施例的节拍特征对齐的示例。图 9 中，第 i 个视觉节拍特征，需要变速到第 i 个音频节拍特征对应的时刻。因此可以计算出一个视频变速曲线，纵坐标表示原始时刻，横坐标表示需要变速到的目标时刻。例如第 2 个视觉节拍特征 910 是在原始视频的 0.6s，第二个音频节拍特征是在音频的 2.4s，因此需要将该视觉节拍特征通过调整播放速度计慢速，使得视觉节拍特征 910 的时间到 2.4s。

在一些实施例中，针对所述第一视频数据进行光流检测之前，还包括：

重复所述第一视频数据的至少部分数据，以使得重复后的所述第一视频数据的对象具有周期性的光流变化。

具体地，第一视频数据可以具有固定的视频节奏，也可以不具有固定的视频节奏。当第一视频数据具有固定的视频节奏时，可以直接改变视频帧的播放速度来与音频节奏匹配。当第一视频数据不具有固定的视频节奏时，可以对第一视频数据进行裁剪使其变为具有固定的视频节奏。例如，将第一视频数据的部分数据移除，或者对部分数据进行重复多次。参见图 10，图 10 示出了根据本公开实施例的视频处理方法的示例。如图 10 所示，可以基于用户的操作确定音乐数据和第一视频数据，对音乐数据进行节奏分析得到音乐节奏，以及对第一视频数据进行视频节奏分析。其中，基于视频节奏分析确定第一视频数据

具有固定的节奏时，计算第一视频数据的第一播放速度曲线；基于视频节奏分析确定视频数据不具有固定的节奏时，可以对第一视频数据进行裁剪后计算裁剪后第一视频数据的第一播放速度曲线。可以改变第一视频数据中视频帧之间的播放速度，以使得运动变化特征与节拍特征出现的时间点对齐，则视频帧之间的播放速度由第一播放速度曲线变为第二播放速度曲线，得到具有第二播放速度曲线的第二视频数据。最后将音乐数据和第二视频数据合成，即可得到视频节奏与音乐节奏相匹配的音乐卡点视频。

在一些实施例中，调整所述第一视频数据中视频帧之间的播放速度之前，所述方法还包括：根据用户操作或所述视觉节拍特征设置所述音频节拍特征的时间和/或强度。

具体地，用户可以对节拍特征进行增加、移除、移动等操作来改变音频节拍特征在时间轴上的位置，即改变音频节拍特征出现的时间。也可以针对音频节拍特征的强度进行变更，例如，增大或减小音频节拍特征的强度。这样，用户可以根据自由地编辑最终音视频中视频节奏和音频节奏匹配的时间和强度，即实现对卡点效果的自由设置。

在步骤 S330，将所述第二视频数据和所述音频数据合成目标视频。

在一些实施例中，方法 300 还可以包括：采用特效播放所述目标视频中对应于所述节拍特征为重节拍的视频帧，所述重节拍包括所述节拍特征的强度大于或等于预设强度。

具体地，可以对重节拍对应的视频帧采用慢放、放大、抖动等视觉特效进行播放，配合与之匹配的音频节奏，以进一步增加视觉的冲击力，提高目标视频的呈现效果。

需要说明的是，本公开实施例的方法可以由单个设备执行，例如一台计算机或服务器等。本实施例的方法也可以应用于分布式场景下，由多台设备相互配合来完成。在这种分布式场景的情况下，这多台设备中的一台设备可以只执行本公开实施例的方法中的某一个或多个步骤，这多台设备相互之间会进行交互以完成所述的方法。

需要说明的是，上述对本公开的一些实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于上述实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要示出的特定顺序或者连续顺序才能实现期

望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

基于同一技术构思，与上述任意实施例方法相对应的，本公开还提供了一种视频处理装置，参见图 11，所述视频处理装置包括：

5 数据确定模块，用于确定第一视频数据和音频数据；

视频处理模块，用于基于所述音频数据的音频节拍特征调整所述第一视频数据的播放速度，得到第二视频数据；其中，所述第二视频数据中对象的视觉节拍特征与所述音频节拍特征在对应的时间相匹配；

合成模块，用于将所述第二视频数据和所述音频数据合成为目标视频。

10 为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本公开时可以把各模块的功能在同一个或多个软件和/或硬件中实现。

上述实施例的装置用于实现前述任一实施例中相应的视频处理方法，并且具有相应的方法实施例的有益效果，在此不再赘述。

基于同一技术构思，与上述任意实施例方法相对应的，本公开还提供了一种非暂态计算机可读存储介质，所述非暂态计算机可读存储介质存储计算机指令，所述计算机指令用于使所述计算机执行如上任一实施例所述的视频处理方法。

本实施例的计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据
20 结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带
25 磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。

上述实施例的存储介质存储的计算机指令用于使所述计算机执行如上任一实施例所述的视频处理方法，并且具有相应的方法实施例的有益效果，在此不再赘述。

30 所属领域的普通技术人员应当理解：以上任何实施例的讨论仅为示例性的，并非旨在暗示本公开的范围（包括权利要求）被限于这些例子；在本公开

的思路下，以上实施例或者不同实施例中的技术特征之间也可以进行组合，步骤可以以任意顺序实现，并存在如上所述的本公开实施例的不同方面的许多其它变化，为了简明它们没有在细节中提供。

5 另外，为简化说明和讨论，并且为了不会使本公开实施例难以理解，在所提供的附图中可以示出或可以不示出与集成电路（IC）芯片和其它部件的公知的电源/接地连接。此外，可以以框图的形式示出装置，以便避免使本公开实施例难以理解，并且这也考虑了以下事实，即关于这些框图装置的实施方式的细节是高度取决于将要实施本公开实施例的平台的（即，这些细节应当完全处于本领域技术人员的理解范围内）。在阐述了具体细节（例如，电路）以描述本
10 公开的示例性实施例的情况下，对本领域技术人员来说显而易见的是，可以在没有这些具体细节的情况下或者这些具体细节有变化的情况下实施本公开实施例。因此，这些描述应被认为是说明性的而不是限制性的。

尽管已经结合了本公开的具体实施例对本公开进行了描述，但是根据前面的描述，这些实施例的很多替换、修改和变型对本领域普通技术人员来说将是
15 显而易见的。例如，其它存储器架构（例如，动态 RAM（DRAM））可以使用所讨论的实施例。

本公开实施例旨在涵盖落入所附权利要求的宽泛范围之内的所有这样的替换、修改和变型。因此，凡在本公开实施例的精神和原则之内，所做的任何省略、修改、等同替换、改进等，均应包含在本公开的保护范围之内。

权 利 要 求 书

1、一种视频处理方法，包括：

确定第一视频数据和音频数据；

5 基于所述音频数据的音频节拍特征调整所述第一视频数据的播放速度，得到第二视频数据；其中，所述第二视频数据中对象的视觉节拍特征与所述音频节拍特征在对应的时间相匹配；

将所述第二视频数据和所述音频数据合成为目标视频。

2、根据权利要求 1 的方法，其中，基于所述音频数据的音频节拍特征调整所述
10 所述第一视频数据的播放速度，得到第二视频数据，包括：

基于所述音频数据进行音频节拍分析，得到所述音频节拍特征；

基于所述第一视频数据的光流信息，得到所述第一视频数据中对象的视觉节拍特征；

15 调整所述第一视频数据中视频帧之间的播放速度，以将所述视觉节拍特征与所述音频节拍特征对应的时间对齐，得到所述第二视频数据。

3、根据权利要求 2 的方法，其中，基于所述音频数据进行音频节拍分析，得到所述音频数据的音频节拍特征，包括：

20 对所述音频数据进行频率分析得到所述音频数据关于时间的频谱特征，所述频谱特征包括音频频率和对应的声音强度；

基于所述频谱特征进行音频节拍识别，得到关于时间的音频节拍强度特征；

对所述音频节拍强度特征进行基于移动均线的峰值检测，确定峰值所对应的所述音频节拍强度特征为所述音频节拍特征。

25 4、根据权利要求 2 的方法，其中，基于所述第一视频数据的光流信息，得到所述第一视频数据中对象的视觉节拍特征，包括：

针对所述第一视频数据进行光流检测，得到所述第一视频数据中像素的光流信息；

30 基于所述光流信息得到所述第一视频数据关于时间的光流分布特征，所述光流分布特征包括多个光流方向和每个所述光流方向上对应的光流强度；

基于所述光流分布特征进行视觉节拍分析，得到所述第一视频数据中对象

的视觉节拍特征。

5、根据权利要求 4 的方法，其中，基于所述光流信息得到所述第一视频数据关于时间的光流分布特征，包括：

5 基于所述光流信息的角度得到每个所述像素的像素运动方向，以及基于所述光流信息的幅值得到每个所述像素的像素运动幅值；

针对每个所述光流方向，统计与所述光流方向的角度差在预设范围内的所述像素运动方向所对应的所述像素运动幅值之和，得到所述光流方向上对应的光流强度。

10

6、根据权利要求 4 的方法，其中，基于所述光流分布特征进行视觉节拍分析，得到所述视觉节拍特征，包括：

基于所述光流分布特征进行视觉节拍识别，得到关于时间的视觉节拍强度特征；

15 对所述视觉节拍强度特征进行基于移动均线的峰值检测，确定峰值所对应的所述视觉节拍强度特征为所述视觉节拍特征。

7、根据权利要求 2 的方法，其中，调整所述第一视频数据中视频帧之间的播放速度，以将所述视觉节拍特征与所述音频节拍特征对应的时间对齐，包
20 括：

对所述视觉节拍特征和所述音频节拍特征进行过滤，得到过滤后的视觉节拍特征和音频节拍特征；

调整所述第一视频数据中视频帧的播放速度，使得过滤后的所述视觉节拍特征与所述音频节拍特征依次对齐。

25

8、根据权利要求 2 的方法，其中，针对所述第一视频数据进行光流检测之前，所述方法还包括：

重复所述第一视频数据的至少部分数据，以使得重复后的所述第一视频数据的对象具有周期性的光流变化；

30

或者，

调整所述第一视频数据中视频帧之间的播放速度之前，所述方法还包括：根据用户操作或所述视觉节拍特征设置所述音频节拍特征的时间和/或强度。

9、一种视频处理装置，包括：

数据确定模块，用于确定第一视频数据和音频数据；

5 视频处理模块，用于基于所述音频数据的音频节拍特征调整所述第一视频数据的播放速度，得到第二视频数据；其中，所述第二视频数据中对象的视觉节拍特征与所述音频节拍特征在对应的时间相匹配；

合成模块，用于将所述第二视频数据和所述音频数据合成为目标视频。

10 10、一种电子设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现如权利要求 1 至 8 任意一项所述的方法。

15 11、一种非暂态计算机可读存储介质，所述非暂态计算机可读存储介质存储计算机指令，所述计算机指令用于使计算机执行权利要求 1 至 8 任一所述方法。

12、一种计算机程序产品，包括计算机程序指令，当所述计算机程序指令在计算机上运行时，使得计算机执行权利要求 1 至 8 任一所述的方法。



100

图 1

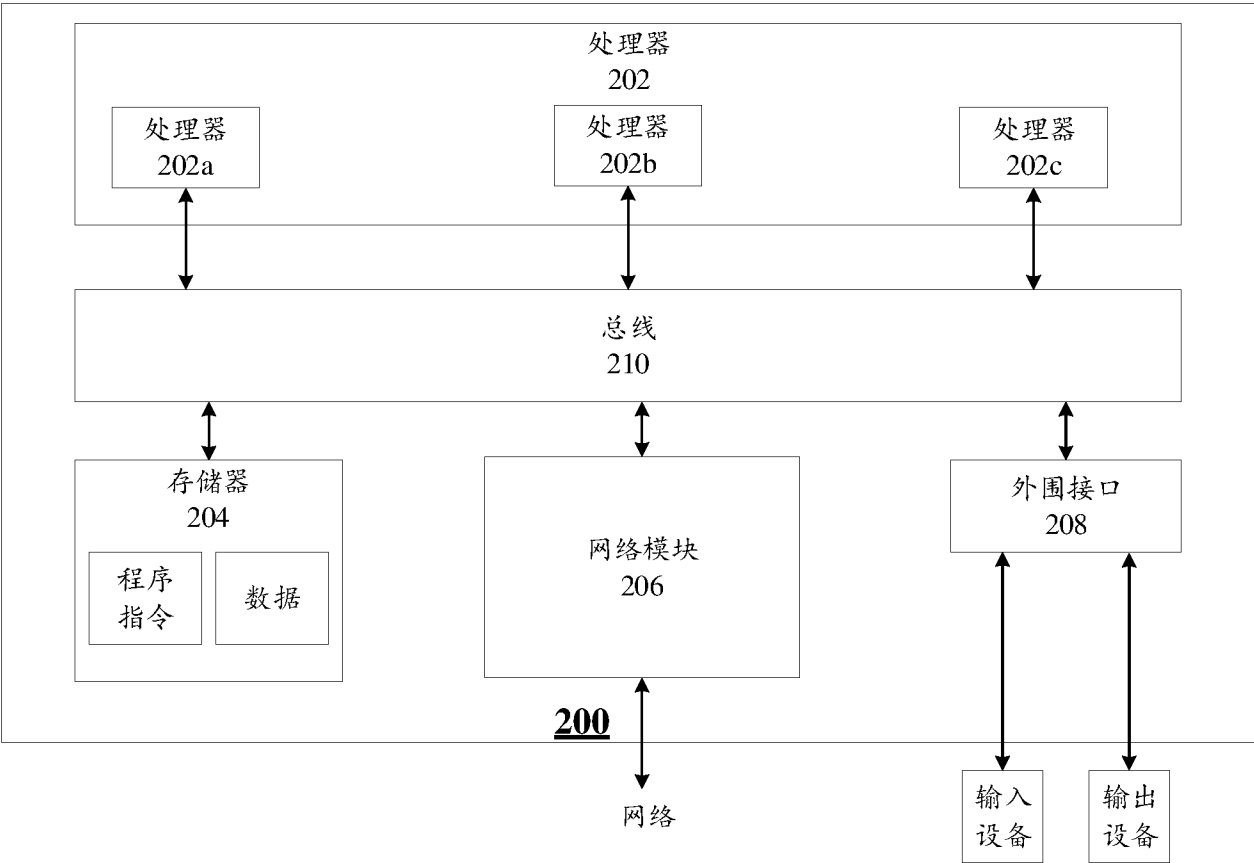


图 2

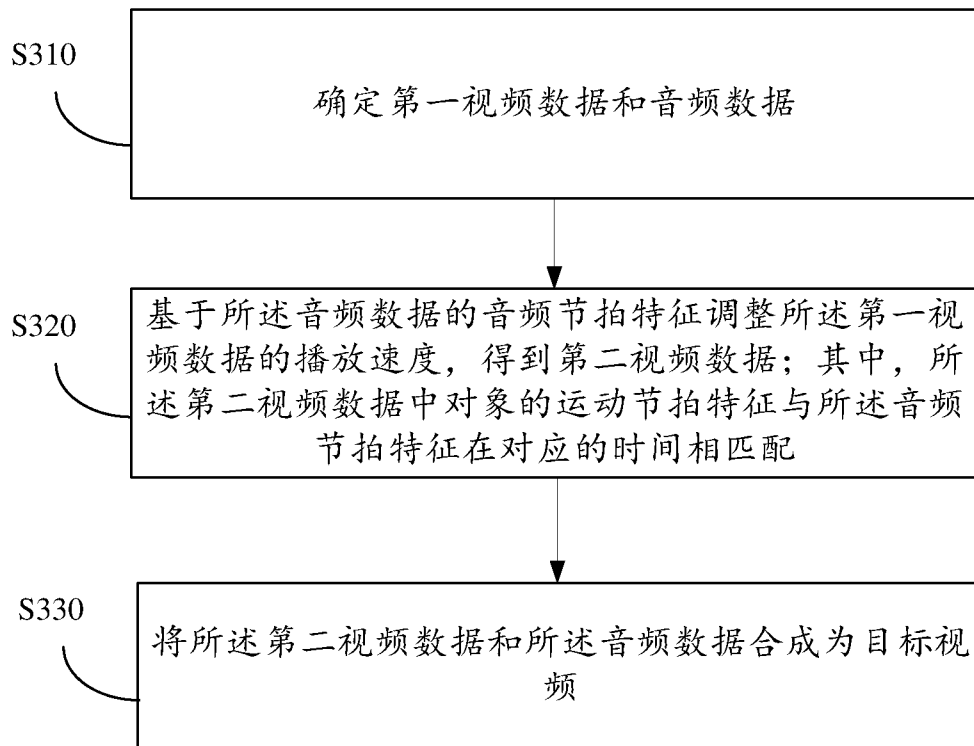
**300**

图 3

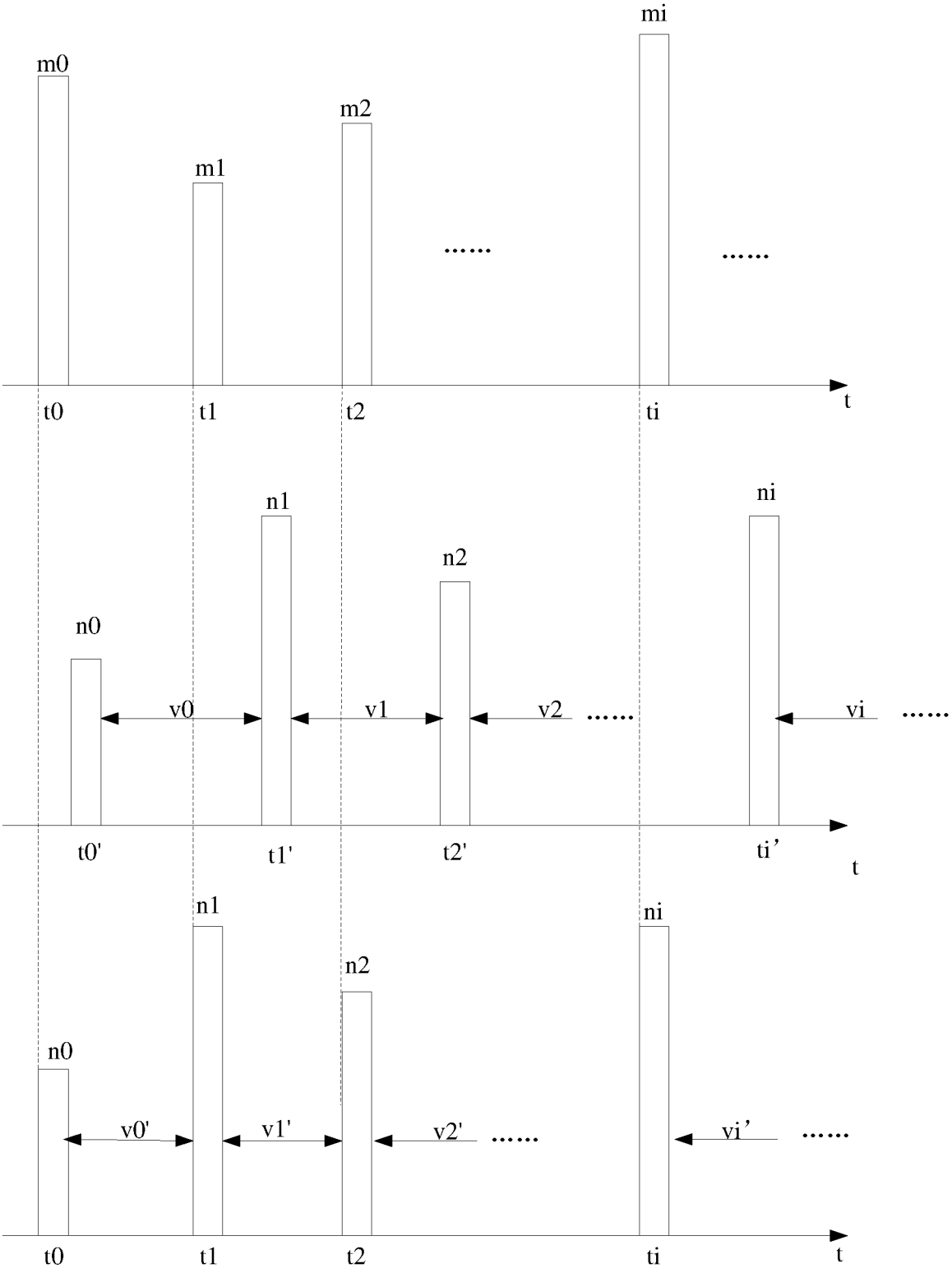


图 4

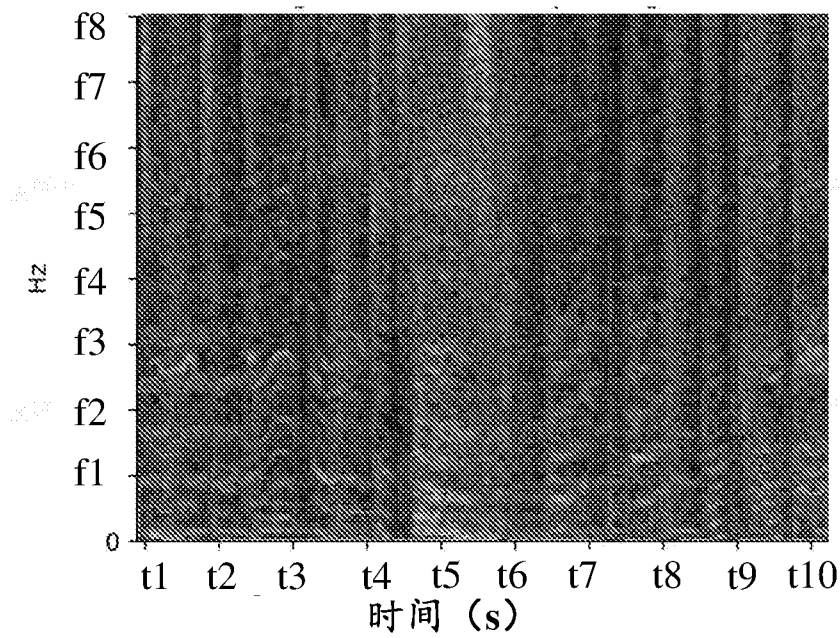


图 5

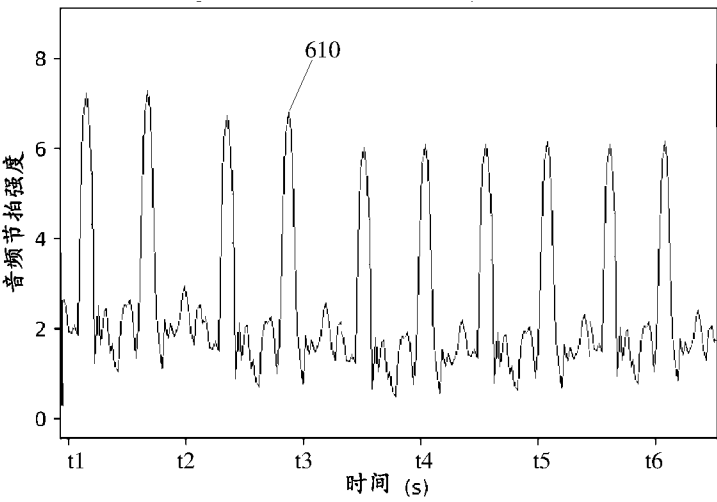


图 6

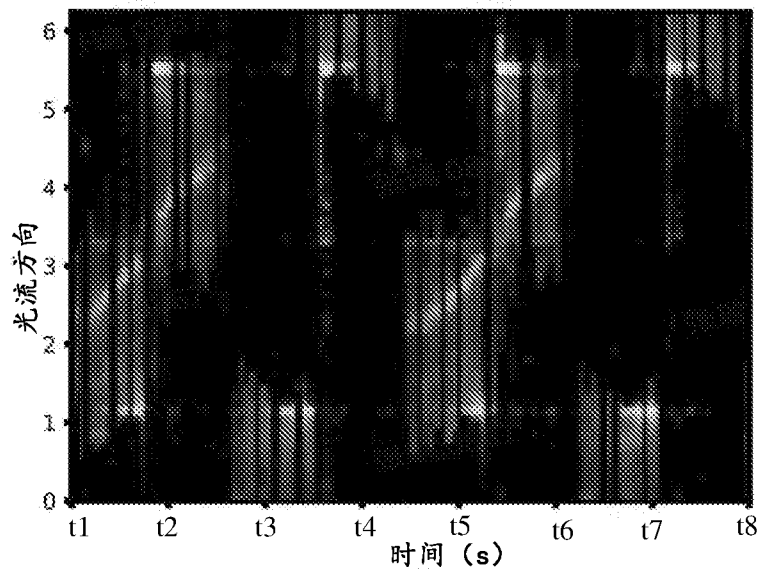


图 7

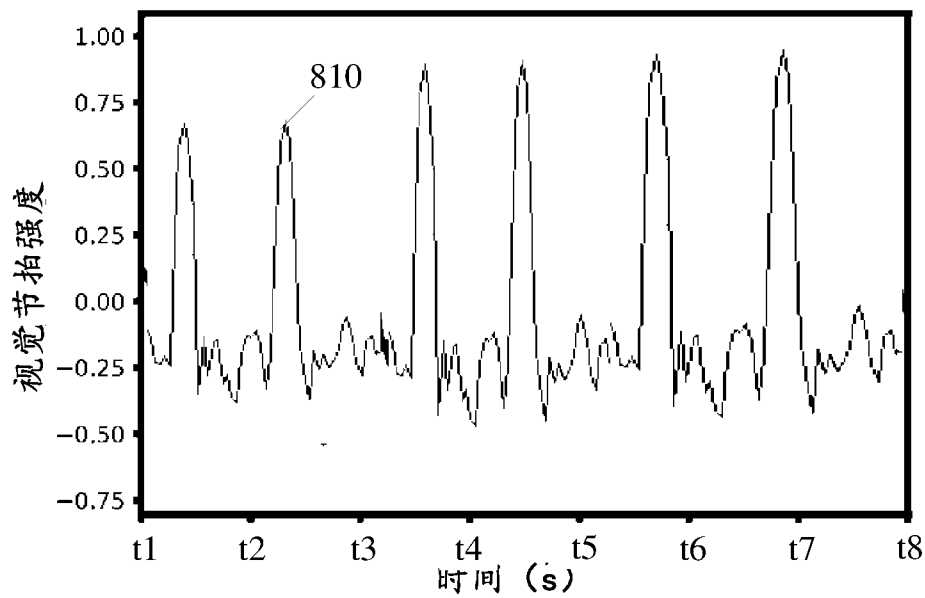


图 8

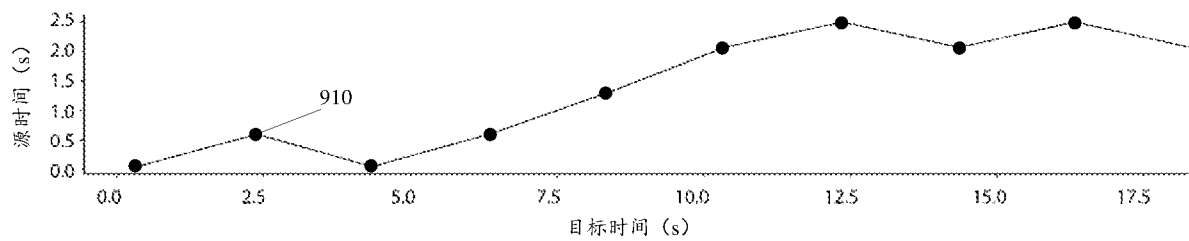


图 9

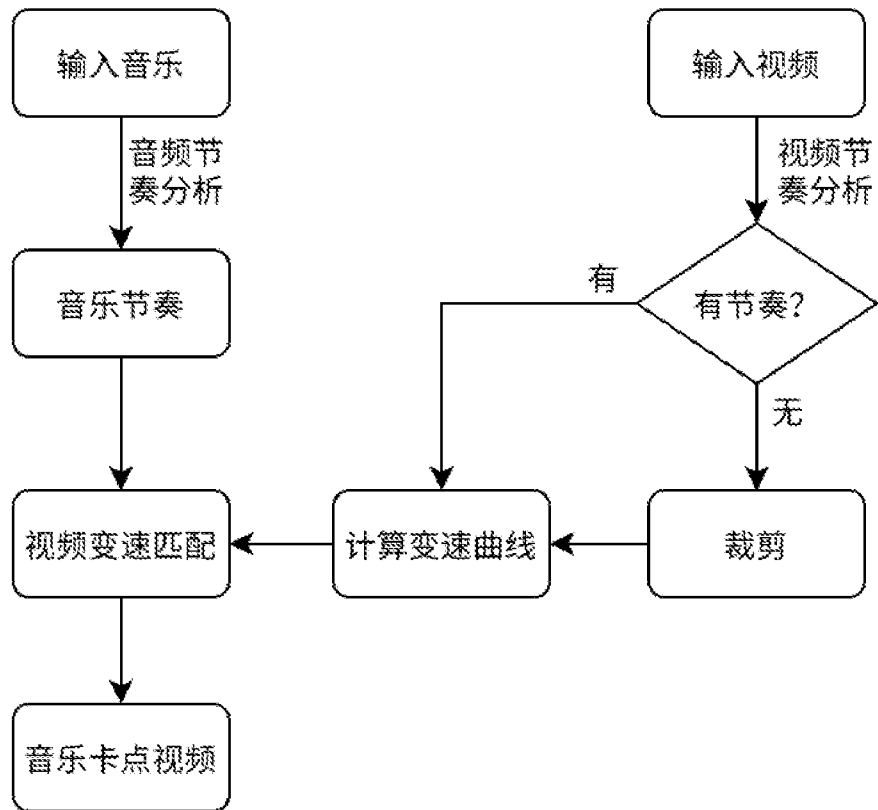
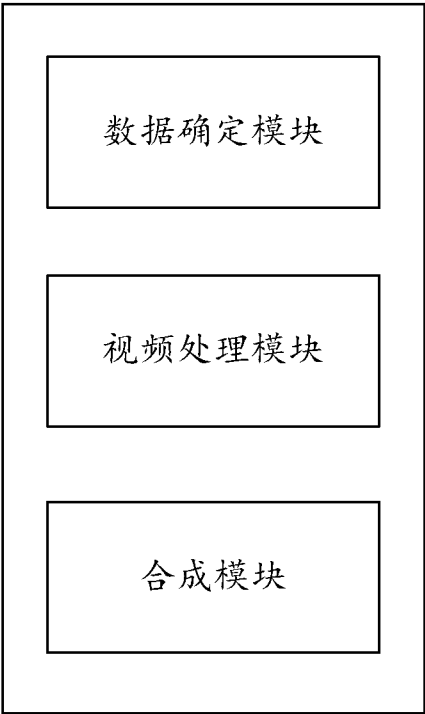


图 10



1100

图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/099198

A. CLASSIFICATION OF SUBJECT MATTER H04N21/43(2011.01)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) IPC: H04N21/- 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) CNABS, CNTXT, ENTXTC, VCN, VEN, ENTXT, JPTXT, WOTXT, CNKI, IEEE: 视频, 音频, 节奏, 节拍, 播放速度, 对齐, 匹配, 频率, 频谱, 光流, 合成, video, audio, beat?, rhythm, play+, speed, match+, frequency, spectrum, optical flow, merge, synthes+		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116761027 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 15 September 2023 (2023-09-15) description, paragraphs [0047]-[0103]	1-12
X	CN 115988262 A (BEIJING YOUZHUJU NETWORK TECHNOLOGY CO., LTD.) 18 April 2023 (2023-04-18) description, paragraphs [0035]-[0068]	1-12
X	CN 112380396 A (NETEASE (HANGZHOU) NETWORK CO., LTD.) 19 February 2021 (2021-02-19) description, paragraphs [0063]-[0108]	1-12
X	US 2022261573 A1 (ADOBE INC.) 18 August 2022 (2022-08-18) description, paragraphs [0004]-[0007]	1-12
A	CN 113473201 A (TENCENT MUSIC ENTERTAINMENT TECHNOLOGY (SHENZHEN) CO., LTD.) 01 October 2021 (2021-10-01) entire document	1-12
☐ 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 “D” document cited by the applicant in the international application “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 11 July 2024		Date of mailing of the international search report 02 August 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2024/099198

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116761027	A	15 September 2023	None			
CN	115988262	A	18 April 2023	None			
CN	112380396	A	19 February 2021	CN	112380396	B	26 April 2024
US	2022261573	A1	18 August 2022	US	11682238	B2	20 June 2023
CN	113473201	A	01 October 2021	None			

A. 主题的分类

H04N21/43(2011.01)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)

IPC: H04N21/-

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))

CNABS,CNXTXT,ENTXTC,VCN,VEN,ENTXT,JPTXT,WOTXT,CNKI,IEEE: 视频, 音频, 节奏, 节拍, 播放速度, 对齐, 匹配, 频率, 频谱, 光流, 合成, video, audio, beat?, rhythm, play+, speed, match+, frequency, spectrum, optical flow, merge, synthes+

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
PX	CN 116761027 A (北京字跳网络技术有限公司) 2023年9月15日 (2023 - 09 - 15) 说明书第[0047]-[0103]段	1-12
X	CN 115988262 A (北京有竹居网络技术有限公司) 2023年4月18日 (2023 - 04 - 18) 说明书第[0035]-[0068]段	1-12
X	CN 112380396 A (网易(杭州)网络有限公司) 2021年2月19日 (2021 - 02 - 19) 说明书第[0063]-[0108]段	1-12
X	US 2022261573 A1 (ADOBE INC.) 2022年8月18日 (2022 - 08 - 18) 说明书第[0004]-[0007]段	1-12
A	CN 113473201 A (腾讯音乐娱乐科技(深圳)有限公司) 2021年10月1日 (2021 - 10 - 01) 全文	1-12

☐ 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

“D” 申请人在国际申请中引证的文件

“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

2024年7月11日

国际检索报告邮寄日期

2024年8月2日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

张博

电话号码 (+86) 028-62969334

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/099198

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	116761027	A	2023年9月15日	无	
CN	115988262	A	2023年4月18日	无	
CN	112380396	A	2021年2月19日	CN 112380396	B 2024年4月26日
US	2022261573	A1	2022年8月18日	US 11682238	B2 2023年6月20日
CN	113473201	A	2021年10月1日	无	