

(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
2023 年 8 月 3 日 (03.08.2023)



(10) 国际公布号
WO 2023/142913 A1

(51) 国际专利分类号:
G06V 20/40 (2022.01)

(21) 国际申请号: PCT/CN2023/070262

(22) 国际申请日: 2023 年 1 月 4 日 (04.01.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202210113116.0 2022 年 1 月 29 日 (29.01.2022) CN

(71) 申请人: 北京有竹居网络技术有限公司
(**BEIJING YOUZHUJU NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市平谷区林荫北街 13 号信息大厦 802 室, Beijing 101299 (CN)。

(72) 发明人: 孙佳(**SUN, Jia**); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。袁泽寰(**YUAN, Zehuan**); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。

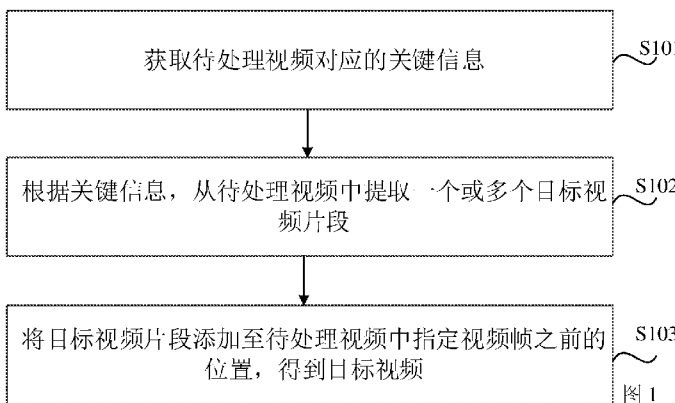
(74) 代理人: 北京知帆远景知识产权代理有限公司
(**ZHIFAN & PARTNERS**); 中国北京市海淀区阜成路 73 号裕惠大厦 B 座 805, Beijing 100142 (CN)。

(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, 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。

(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, 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,

(54) **Title:** VIDEO PROCESSING METHOD AND APPARATUS, READABLE MEDIUM AND ELECTRONIC DEVICE

(54) 发明名称: 视频处理方法、装置、可读介质及电子设备



S101 Obtain key information corresponding to video to be processed
S102 According to the key information, extract one or more target video clips from the video to be processed
S103 Add the target video clip to a position preceding a specified video frame in the video to be processed, and obtain a target video

(57) **Abstract:** A video processing method and apparatus, a readable medium and an electronic device. The method comprises: obtaining key information corresponding to a video to be processed; according to the key information, extracting one or more target video clips from the video to be processed; adding the target video clip to a position preceding a specified video frame in the video to be processed, and obtaining a target video, wherein the specified video frame is any one of a preset number of video frames in the video to be processed.

(57) **摘要:** 一种视频处理方法、装置、可读介质及电子设备。方法包括: 获取待处理视频对应的关键信息; 根据关键信息, 从待处理视频中提取一个或多个目标视频片段; 将目标视频片段添加至待处理视频中指定视频帧之前的位置, 得到目标视频, 指定视频帧为待处理视频中预设数量个视频帧中的任一视频帧。

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))。
- 在修改权利要求的期限届满之前进行，在收到该修改后将重新公布(细则48.2(h))。

视频处理方法、装置、可读介质及电子设备

本申请要求于 2022 年 01 月 29 日提交的，申请名称为“视频处理方法、装置、可读介质及电子设备”的、中国专利申请号为“202210113116.0”的优先权，该中国专利申请的全部内容通过引用结合在本申请中。

5 技术领域

本公开涉及计算机技术领域，具体地，涉及一种视频处理方法、装置、可读介质及电子设备。

背景技术

随着计算机技术的发展，多媒体的应用也越来越广泛，各种各样的视频在网络上不断涌现，在大量的网络视频中，用户难以根据自身需求快速确定所需要的视频。

发明内容

提供该发明内容部分以便以简要的形式介绍构思，这些构思将在后面的具体实施方式部分被详细描述。该发明内容部分并不旨在标识要求保护的技术方案的关键特征或必要特征，也不旨在用于限制所要求的保护的技术方案的范围。

15 第一方面，本公开提供了一种视频处理方法，所述方法包括：

获取待处理视频对应的关键信息；

根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段；

将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频，所述指定视频帧为所述待处理视频中前预设数量个视频帧中的任一视频帧。

20 第二方面，本公开提供了一种视频处理装置，所述装置包括：

信息获取模块，用于获取待处理视频对应的关键信息；

片段提取模块，用于根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段；

25 视频处理模块，用于将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频，所述指定视频帧为所述待处理视频中前预设数量个视频帧中的任一视频帧。

第三方面，本公开提供一种计算机可读介质，其上存储有计算机程序，该程序被处理装置执行时实现本公开第一方面所述方法的步骤。

第四方面，本公开提供一种电子设备，包括：

30 存储装置，其上存储有计算机程序；

处理装置，用于执行所述存储装置中的所述计算机程序，以实现本公开第一方面所述方

法的步骤。

采用上述技术方案，获取待处理视频对应的关键信息；根据关键信息，从待处理视频中提取一个或多个目标视频片段；将目标视频片段添加至待处理视频中指定视频帧之前的位置，得到目标视频，该指定视频帧为待处理视频中前预设数量个视频帧中的任一视频帧。这样，可以使得用户在浏览到目标视频后，能够快速观看到与目标视频的关键信息相似度最高的目标视频片段，使得用户能够快速甄别该目标视频是否符合自身需求，一方面能够提升用户选择视频的效率，另一方面也能够使得视频有效吸引有需求的用户观看。

本公开的其他特征和优点将在随后的具体实施方式部分予以详细说明。

附图说明

结合附图并参考以下具体实施方式，本公开各实施例的上述和其他特征、优点及方面将变得更加明显。贯穿附图中，相同或相似的附图标记表示相同或相似的元素。应当理解附图是示意性的，原件和元素不一定按照比例绘制。在附图中：

图 1 是根据一示例性实施例示出的一种视频处理方法的流程图。

图 2 是根据图 1 所示实施例示出的一种 S102 步骤的流程图。

图 3 是根据一示例性实施例示出的一种视频处理装置的框图。

图 4 是根据一示例性实施例示出的一种电子设备的框图。

具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

应当理解，本公开的方法实施方式中记载的各个步骤可以按照不同的顺序执行，和/或并行执行。此外，方法实施方式可以包括附加的步骤和/或省略执行示出的步骤。本公开的范围在此方面不受限制。

本文使用的术语“包括”及其变形是开放性包括，即“包括但不限于”。术语“基于”是“至少部分地基于”。术语“一个实施例”表示“至少一个实施例”；术语“另一实施例”表示“至少一个另外的实施例”；术语“一些实施例”表示“至少一些实施例”。其他术语的相关定义将在下文描述中给出。

需要注意，本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分，并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

需要注意，本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，本领域

技术人员应当理解，除非在上下文另有明确指出，否则应该理解为“一个或多个”。

本公开实施方式中的多个装置之间所交互的消息或者信息的名称仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

首先，对本公开的应用场景进行说明。本公开可以应用于视频处理场景，随着计算机技术的发展，各种各样的视频在网络上不断涌现。由于不同视频的内容和质量差异很大，当用户根据自身需求寻找某一类视频时，需要浏览大量视频寻找，由于视频内容是按照顺序播放的，关键信息都在视频播放过程中的视频帧中，用户需要浏览完整的视频或一大段视频后，才能够甄别某个视频是否符合自身需求，这样一方面会导致用户难以根据自身需求快速浏览得到所需要的视频；另一方面也会导致视频无法有效吸引有需求的用户观看。

为了解决上述问题，本公开提供了一种视频处理的方法、装置、可读介质及电子设备，根据待处理视频对应的关键信息，从待处理视频中提取一个或多个目标视频片段；并将目标视频片段添加至待处理视频中指定视频帧之前的位置，得到目标视频，该指定视频帧为待处理视频中前预设数量个视频帧中的任一视频帧。这样，可以使得用户在浏览到目标视频后，能够快速观看到与目标视频的关键信息相似度最高的目标视频片段，使得用户能够快速甄别该目标视频是否符合自身需求，一方面能够提升用户选择视频的效率；另一方面也能够使得视频有效吸引有需求的用户观看。

以下结合附图对本公开的具体实施方式进行详细说明。

图 1 是根据一示例性实施例示出的一种视频处理方法，如图 1 所示，该方法包括：

步骤 101、获取待处理视频对应的关键信息。

示例地，若该待处理视频为商品介绍视频，则该关键信息可以是该待处理视频所介绍的商品的卖点信息；若该待处理视频为科普知识视频，则该关键信息可以为该待处理视频所介绍的科普知识的知识点信息。

在本步骤中，可以将该待处理视频的视频标题作为关键信息，或者，将用户针对该待处理视频输入的信息作为关键信息，或者，将预先设置的热点词语作为关键信息。当然，也可以将上述视频标题、用户输入信息和热点词语组合作为关键信息。

步骤 102、根据关键信息，从待处理视频中提取一个或多个目标视频片段。

示例地，可以计算待处理视频的每个视频帧与上述关键信息的相似度，根据相似度最高的前 N 个视频帧组合得到一个或多个目标视频片段。

步骤 103、将目标视频片段添加至待处理视频中指定视频帧之前的位置，得到目标视频。

其中，该指定视频帧可以为待处理视频中前预设数量个视频帧中的任一视频帧。该预设数量可以是固定数值，例如 10 或 20；该预设数量也可以是根据待处理视频的视频帧总数确

定的数值，例如该预设数量可以是该待处理视频的视频帧总数的五分之一或十分之一。

例如，可以将待处理视频的第一帧作为该指定视频帧，这样，可以将目标视频片段添加至待处理视频的最前面。再例如，可以将标题帧之后的视频帧作为该指定视频帧，这样，可以将目标视频片段添加至待处理视频的标题帧之后，内容帧之前。

- 5 采用上述方法，获取待处理视频对应的关键信息；根据关键信息，从待处理视频中提取一个或多个目标视频片段；将目标视频片段添加至待处理视频中指定视频帧之前的位置，得到目标视频，该指定视频帧为待处理视频中前预设数量个视频帧中的任一视频帧。这样，可以使得用户在浏览到目标视频后，能够快速观看到与目标视频的关键信息相似度最高的目标视频片段，使得用户能够快速甄别该目标视频是否符合自身需求，一方面能够提升用户选择
- 10 视频的效率，另一方面也能够使得视频有效吸引有需求的用户观看。

图 2 是根据图 1 所示实施例示出的一种 S102 步骤的流程图，如图 2 所示，上述步骤 102 可以包括以下步骤：

步骤 1021、根据待处理视频的图像信息和音频信息，获取该待处理视频的一个或多个候选视频片段。

- 15 其中，该候选视频片段包括该待处理视频中的一段完整的语义信息。

可以通过相邻视频帧图像之间的图像相似度，以及相邻视频帧对应的音频帧之间的音频相似度，对待处理视频进行分割，得到一个或多个候选视频片段。

- 例如，可以将图像相似度小于等于第一预设图像相似阈值且音频相似度小于等于第一预设音频相似阈值的相邻视频帧作为分割帧，根据多个分割帧，对待处理视频进行分割，得到
- 20 一个或多个候选视频片段。

步骤 1022、获取每个候选视频片段与关键信息的相似度。

步骤 1023、将相似度最高的一个或多个候选视频片段作为目标视频片段。

- 示例地，可以通过预先训练的视频分析模型，确定每个候选视频片段与关键信息的相似度，根据相似度排序结果，将相似度最高的一个或预设数目个候选视频片段作为目标视频片
- 25 段。

这样，可以获取包括完整语义信息的视频片段，避免视频片段中的语义被截断导致无法向用户展示完整的语义信息，并通过相似度对比，将与关键信息的相似度最高的一个或多个候选视频片段作为目标视频片段，从而使得目标视频片段能够完整准确的展示待处理视频的关键信息。

- 30 在本公开的另一实施例中，上述步骤 1021 还可以通过以下方式获取该待处理视频的一个或多个候选视频片段：

S11、根据图像信息获取待处理视频的一个或多个待定视频片段。

其中，每个该待定视频片段包括一帧或多帧图像。

示例地，可以获取待处理视频的每帧图像中的图像文本；计算相邻帧图像之间的图像文本的文本相似度；根据该文本相似度，确定该待定视频片段。

5 例如，可以通过 OCR（Optical Character Recognition，光学字符识别），获取待处理视频的每帧图像中的待定文本；然后可以将该待定文本作为图像文本，也可以根据每个待定文本的文本面积过滤掉每帧图像中的无效小文本，例如，可以将文本面积小于或等于预设面积的待定文本作为无效小文本，将待定文本中除了无效小文本之外的其他文本作为上述图像文本。当然，也可以直接将文本面积大于预设面积的待定文本作为上述图像文本。

10 需要说明的是，OCR 是指通过检测暗、亮的模式确定字符形状，然后用字符识别方法将字符形状翻译成文本的过程，为了提高 OCR 识别文本的准确性，可以对每帧图像进行噪声去除、倾斜校正、字符切割等预处理过程，经过预处理后在进行字符识别得到上述待定文本。

进一步地，可以通过相邻视频帧对应的图像文本的文本相似度，对待处理视频进行分割，
15 得到一个或多个待定视频片段。例如，可以将图像相似度小于或等于第二预设图像相似阈值的相邻视频帧作为分割帧，根据多个分割帧，对待处理视频进行分割，得到一个或多个待定视频片段。

这样，可以将具有相同或相似图像文本的视频帧分割到一个待定视频片段中，从而使得每个待定视频片段具有相同或相似的语义。

20 S12、根据音频信息获取待处理视频的一个或多个待定音频片段。

示例地，可以获取音频信息对应的音频文本；对该音频文本进行断句推断，获取该音频文本中的断句信息；根据该断句信息，确定一个或多个该待定音频片段。

例如，可以通过 ASR（Automatic Speech Recognition，自动语音识别技术），获取待处理视频的音频信息对应的音频文本，以及每个音频文本中每个文本字符对应的时间戳。然后，
25 可以根据音频文本及时间戳进行断句推断，例如获取相邻文本字符之间的时间差，将时间差大于预设时间阈值的两个相邻文本字符作为断句字符，然后可以根据断句字符前后的语义确定断句标点，可以将该断句标点及该断句标点的位置信息作为上述断句信息；将两个断句标点之间的音频作为待定音频片段。

需要说明的是，上述断句字符和断句标点的获取，也可以通过预先训练的文本断句模型
30 对音频文本进行处理后得到，该文本断句模型的结构及训练方法可以参考现有技术中的实现方式，本公开对此不作限制。

S13、根据该待定视频片段和该待定音频片段，确定该候选视频片段。

在本步骤中，可以通过以下方式，根据待定音频片段对待定视频片段进行修正，得到该候选视频片段：

首先，确定待定音频片段与待定视频片段的对应关系。

5 在本步骤中，可以根据待定音频片段对应的音频文本和待定视频片段对应的图像文本的文本相似度，确定待定音频片段与待定视频片段的对应关系。

例如，针对每个待定视频片段，根据文本相似度对待定音频片段进行排序，将文本相似度最高的待定音频片段作为与该待定视频片段对应的待定音频片段。

10 进一步地，也可以根据待定音频片段和待定视频片段在待处理视频中的起止时间进行时间重合度计算，针对每个待处理视频，将时间重合度最高的待定音频片段作为与该待定视频片段对应的待定音频片段。

然后，针对每个待定视频片段，可以根据与该待定视频片段对应的待定音频片段，对该待定视频片段进行完整性修正，得到候选视频片段。

15 示例地，可以根据待定音频片段的时间信息，对待定视频片段进行完整性修正操作，该完整性修正操作可以包括视频片段起止时间修正或者相邻视频拍片段合并。

20 例如，第一待定视频片段的时间为视频开始播放后的第 15 秒至第 20 秒之间，若与该第一待定视频片段相对应的第一待定音频片段的时间为视频开始播放后的第 15 秒至第 21 秒之间的片段，则可以将该第一待定视频片段的时间修正为第 15 秒至第 20 秒。另外，为了避免视频片段重复，还可以将与该第一待定视频片段相邻的第二待定视频片段进行修正，例如将第二待定视频片段的时间从“第 21 秒至第 30 秒”修正为“第 22 秒至第 30 秒”

例如，第一待定视频片段的时间为视频开始播放后的第 15 秒至第 20 秒之间，与之相邻的第二待定视频片段的时间为视频开始播放后的第 21 秒至第 25 秒之间，若与该第一待定视频片段相对应的第一待定音频片段的时间为视频开始播放后的第 15 秒至第 25 秒之间的片段，则可以将该第二待定视频片段与该第一待定视频片段合并为新的第一待定视频片段。

25 这样，通过音频信息对视频片段进行完整性修正，可以使得得到的候选视频片段能够展示完整语义，避免候选视频片段中的语义被截断导致无法向用户展示完整的语义信息。

在本公开的另一实施例中，上述步骤 103 将目标视频片段添加至待处理视频中指定视频帧之前的位置，得到目标视频的方法可以包括：

首先，根据预设描述信息更新该目标视频片段。

30 该预设描述信息用于表征该目标视频片段对应该待处理视频的关键信息。例如，该预设描述信息可以是文本“视频简介”，将该文本增加至目标视频片段的每个视频帧中，以便提

示用户，该视频帧为视频简介部分，不是视频的主体内容。

然后，将更新后的目标视频片段添加至待处理视频中指定视频帧之前的位置，得到目标视频。

5 这样，可以使得用户能够区分展示关键信息的目标视频片段和视频的主体内容，避免由于视频内容跳跃导致用户对视频信息无法理解。

图 3 是根据一示例性实施例示出的一种视频处理装置的框图。如图 3 所示，该视频处理装置包括：

信息获取模块 301，用于获取待处理视频对应的关键信息；

10 片段提取模块 302，用于根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段；

视频处理模块 303，用于将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频，所述指定视频帧为所述待处理视频中前预设数量个视频帧中的任一视频帧。

15 在一些实施例中，所述片段提取模块 302，用于根据待处理视频的图像信息和音频信息，获取所述待处理视频的一个或多个候选视频片段；其中，所述候选视频片段包括所述待处理视频中的一段完整的语义信息；获取每个候选视频片段与所述关键信息的相似度；将所述相似度最高的一个或多个候选视频片段作为所述目标视频片段。

20 在一些实施例中，所述片段提取模块 302，用于根据所述图像信息获取所述待处理视频的一个或多个待定视频片段，每个所述待定视频片段包括一帧或多帧图像；根据所述音频信息获取所述待处理视频的一个或多个待定音频片段；根据所述待定视频片段和所述待定音频片段，确定所述候选视频片段。

在一些实施例中，所述片段提取模块 302，用于获取待处理视频的每帧图像中的图像文本；计算相邻帧图像之间的图像文本的文本相似度；根据所述文本相似度，确定所述待定视频片段。

25 在一些实施例中，所述片段提取模块 302，用于获取所述音频信息对应的音频文本；对所述音频文本进行断句推断，获取所述音频文本中的断句信息；根据所述断句信息，确定一个或多个所述待定音频片段。

30 在一些实施例中，所述片段提取模块 302，用于确定待定音频片段与待定视频片段的对应关系；针对每个待定视频片段，根据与该待定视频片段对应的待定音频片段，对该待定视频片段进行完整性修正，得到候选视频片段。

在一些实施例中，所述视频处理模块 303，用于根据预设描述信息更新所述目标视频片

段；所述预设描述信息用于表征所述目标视频片段对应所述待处理视频的关键信息；将更新后的目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频。

下面参考图 4，其示出了适于用来实现本公开实施例的电子设备（例如终端设备或服务器）900 的结构示意图。本公开实施例中的终端设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、PDA（个人数字助理）、PAD（平板电脑）、PMP（便携式多媒体播放器）、车载终端（例如车载导航终端）等等的移动终端以及诸如数字 TV、台式计算机等等的固定终端。图 4 示出的电子设备仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

如图 4 所示，电子设备 900 可以包括处理装置（例如中央处理器、图形处理器等）901，其可以根据存储在只读存储器（ROM）902 中的程序或者从存储装置 908 加载到随机访问存储器（RAM）903 中的程序而执行各种适当的动作和处理。在 RAM903 中，还存储有电子设备 900 操作所需的各种程序和数据。处理装置 901、ROM902 以及 RAM903 通过总线 904 彼此相连。输入/输出（I/O）接口 905 也连接至总线 904。

通常，以下装置可以连接至 I/O 接口 905：包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置 906；包括例如液晶显示器（LCD）、扬声器、振动器等的输出装置 907；包括例如磁带、硬盘等的存储装置 908；以及通信装置 909。通信装置 909 可以允许电子设备 900 与其他设备进行无线或有线通信以交换数据。虽然图 4 示出了具有各种装置的电子设备 900，但是应理解的是，并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在非暂态计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信装置 909 从网络上被下载和安装，或者从存储装置 908 被安装，或者从 ROM902 被安装。在该计算机程序被处理装置 901 执行时，执行本公开实施例的方法中限定的上述功能。

需要说明的是，本公开上述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、

磁存储器件、或者上述的任意合适的组合。在本公开中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中，计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：电线、光缆、RF（射频）等等，或者上述的任意合适的组合。

在一些实施方式中，客户端、服务器可以利用诸如 HTTP（HyperText Transfer Protocol，超文本传输协议）之类的任何当前已知或未来研发的网络协议进行通信，并且可以与任意形式或介质的数字数据通信（例如，通信网络）互连。通信网络的示例包括局域网（“LAN”），广域网（“WAN”），网际网（例如，互联网）以及端对端网络（例如，ad hoc 端对端网络），以及任何当前已知或未来研发的网络。

上述计算机可读介质可以是上述电子设备中所包含的；也可以是单独存在，而未装配入该电子设备中。

上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该电子设备执行时，使得该电子设备：获取待处理视频对应的关键信息；根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段；将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频，所述指定视频帧为所述待处理视频中前预设数量个视频帧中的任一视频帧。

可以以一种或多种程序设计语言或其组合来编写用于执行本公开的操作的计算机程序代码，上述程序设计语言包括但不限于面向对象的程序设计语言——诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言——诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网（LAN）或广域网（WAN）——连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个

模块、程序段、或代码的一部分，该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，

5 框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本公开实施例中所涉及到的模块可以通过软件的方式实现，也可以通过硬件的方式来实现。其中，模块的名称在某种情况下并不构成对该模块本身的限定，例如，信息获取

10 模块还可以被描述为“获取待处理视频对应的关键信息的模块”。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如，非限制性地，可以使用的示范类型的硬件逻辑部件包括：现场可编程门阵列（FPGA）、专用集成电路（ASIC）、专用标准产品（ASSP）、片上系统（SOC）、复杂可编程逻辑设备（CPLD）等等。

15 在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、

20 硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器（EPROM或快闪存储器）、光纤、便捷式紧凑盘只读存储器（CD-ROM）、光学储存设备、磁储存设备、或上述内容的任何合适组合。

根据本公开的一个或多个实施例，示例 1 提供了一种视频处理方法，所述方法包括：
获取待处理视频对应的关键信息；

25 根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段；

将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频，所述指定视频帧为所述待处理视频中前预设数量个视频帧中的任一视频帧。

根据本公开的一个或多个实施例，示例 2 提供了示例 1 所述的方法，所述根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段包括：

30 根据待处理视频的图像信息和音频信息，获取所述待处理视频的一个或多个候选视频片段；其中，所述候选视频片段包括所述待处理视频中的一段完整的语义信息；

获取每个候选视频片段与所述关键信息的相似度；

将所述相似度最高的一个或多个候选视频片段作为所述目标视频片段。

根据本公开的一个或多个实施例，示例 3 提供了示例 2 所述的方法，所述根据待处理视频的图像信息和音频信息，获取所述待处理视频的一个或多个候选视频片段包括：

5 根据所述图像信息获取所述待处理视频的一个或多个待定视频片段，每个所述待定视频片段包括一帧或多帧图像；

根据所述音频信息获取所述待处理视频的一个或多个待定音频片段；

根据所述待定视频片段和所述待定音频片段，确定所述候选视频片段。

根据本公开的一个或多个实施例，示例 4 提供了示例 3 所述的方法，所述根据所述图像
10 信息获取所述待处理视频的一个或多个待定视频片段包括：

获取待处理视频的每帧图像中的图像文本；

计算相邻帧图像之间的图像文本的文本相似度；

根据所述文本相似度，确定所述待定视频片段。

根据本公开的一个或多个实施例，示例 5 提供了示例 3 所述的方法，所述根据所述音频
15 信息获取所述待处理视频的一个或多个待定音频片段包括：

获取所述音频信息对应的音频文本；

对所述音频文本进行断句推断，获取所述音频文本中的断句信息；

根据所述断句信息，确定一个或多个所述待定音频片段。

根据本公开的一个或多个实施例，示例 6 提供了示例 3 所述的方法，所述根据所述待定
20 视频片段和所述待定音频片段，确定所述候选视频片段包括：

确定待定音频片段与待定视频片段的对应关系；

针对每个待定视频片段，根据与该待定视频片段对应的待定音频片段，对该待定视频片段进行完整性修正，得到候选视频片段。

根据本公开的一个或多个实施例，示例 7 提供了示例 1 至 6 中任一项所述的方法，所述
25 将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频包括：

根据预设描述信息更新所述目标视频片段；所述预设描述信息用于表征所述目标视频片段对应所述待处理视频的关键信息；

将更新后的目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频。

30 根据本公开的一个或多个实施例，示例 8 提供了一种视频处理装置，所述装置包括：
信息获取模块，用于获取待处理视频对应的关键信息；

片段提取模块，用于根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段；

视频处理模块，用于将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频，所述指定视频帧为所述待处理视频中前预设数量个视频帧中的任一视频帧。

根据本公开的一个或多个实施例，示例 9 提供了示例 8 所述的装置，所述片段提取模块，用于根据待处理视频的图像信息和音频信息，获取所述待处理视频的一个或多个候选视频片段；其中，所述候选视频片段包括所述待处理视频中的一段完整的语义信息；获取每个候选视频片段与所述关键信息的相似度；将所述相似度最高的一个或多个候选视频片段作为所述目标视频片段。

根据本公开的一个或多个实施例，示例 10 提供了示例 9 所述的装置，所述片段提取模块，用于根据所述图像信息获取所述待处理视频的一个或多个待定视频片段，每个所述待定视频片段包括一帧或多帧图像；根据所述音频信息获取所述待处理视频的一个或多个待定音频片段；根据所述待定视频片段和所述待定音频片段，确定所述候选视频片段。

根据本公开的一个或多个实施例，示例 11 提供了示例 10 所述的装置，所述片段提取模块，用于获取待处理视频的每帧图像中的图像文本；计算相邻帧图像之间的图像文本的文本相似度；根据所述文本相似度，确定所述待定视频片段。

根据本公开的一个或多个实施例，示例 12 提供了示例 10 所述的装置，所述片段提取模块，用于获取所述音频信息对应的音频文本；对所述音频文本进行断句推断，获取所述音频文本中的断句信息；根据所述断句信息，确定一个或多个所述待定音频片段。

根据本公开的一个或多个实施例，示例 13 提供了示例 10 所述的装置，所述片段提取模块，用于确定待定音频片段与待定视频片段的对应关系；针对每个待定视频片段，根据与该待定视频片段对应的待定音频片段，对该待定视频片段进行完整性修正，得到候选视频片段。

根据本公开的一个或多个实施例，示例 14 提供了示例 8 至 13 中任一项所述的装置，所述视频处理模块，用于根据预设描述信息更新所述目标视频片段；所述预设描述信息用于表征所述目标视频片段对应所述待处理视频的关键信息；将更新后的目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频。

以上描述仅为本公开的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本公开中所涉及的公开范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述公开构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开中公开的（但不限于）具有类似功能的技术特

征进行互相替换而形成的技术方案。

此外，虽然采用特定次序描绘了各操作，但是这不应理解为要求这些操作以所示出的特定次序或以顺序次序执行来执行。在一定环境下，多任务和并行处理可能是有利的。同样地，虽然在上面论述中包含了若干具体实现细节，但是这些不应被解释为对本公开的范围的限制。在单独的实施例的上下文中描述的某些特征还可以组合地实现在单个实施例中。相反地，在单个实施例的上下文中描述的各种特征也可以单独地或以任何合适的子组合的方式实现在多个实施例中。

尽管已经采用特定于结构特征和/或方法逻辑动作的语言描述了本主题，但是应当理解所附权利要求书中所限定的主题未必局限于上面描述的特定特征或动作。相反，上面所描述的特定特征和动作仅仅是实现权利要求书的示例形式。关于上述实施例中的装置，其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述，此处将不做详细阐述说明。

权利要求书

1.一种视频处理方法，其中，所述方法包括：

获取待处理视频对应的关键信息；

根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段；

5 将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频，
所述指定视频帧为所述待处理视频中前预设数量个视频帧中的任一视频帧。

2.根据权利要求1所述的方法，其中，所述根据所述关键信息，从所述待处理视频中提
取一个或多个目标视频片段包括：

10 根据待处理视频的图像信息和音频信息，获取所述待处理视频的一个或多个候选视频片
段；其中，所述候选视频片段包括所述待处理视频中的一段完整的语义信息；

获取每个候选视频片段与所述关键信息的相似度；

将所述相似度最高的一个或多个候选视频片段作为所述目标视频片段。

3.根据权利要求2所述的方法，其中，所述根据待处理视频的图像信息和音频信息，获
取所述待处理视频的一个或多个候选视频片段包括：

15 根据所述图像信息获取所述待处理视频的一个或多个待定视频片段，每个所述待定视频
片段包括一帧或多帧图像；

根据所述音频信息获取所述待处理视频的一个或多个待定音频片段；

根据所述待定视频片段和所述待定音频片段，确定所述候选视频片段。

20 4.根据权利要求3所述的方法，其中，所述根据所述图像信息获取所述待处理视频的一
个或多个待定视频片段包括：

获取待处理视频的每帧图像中的图像文本；

计算相邻帧图像之间的图像文本的文本相似度；

根据所述文本相似度，确定所述待定视频片段。

25 5.根据权利要求3所述的方法，其中，所述根据所述音频信息获取所述待处理视频的一
个或多个待定音频片段包括：

获取所述音频信息对应的音频文本；

对所述音频文本进行断句推断，获取所述音频文本中的断句信息；

根据所述断句信息，确定一个或多个所述待定音频片段。

30 6.根据权利要求3所述的方法，其中，所述根据所述待定视频片段和所述待定音频片段，
确定所述候选视频片段包括：

确定待定音频片段与待定视频片段的对应关系；

针对每个待定视频片段，根据与该待定视频片段对应的待定音频片段，对该待定视频片段进行完整性修正，得到候选视频片段。

- 5 7.根据权利要求 1 至 6 中任一项所述的方法，其中，所述将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频包括：

根据预设描述信息更新所述目标视频片段；所述预设描述信息用于表征所述目标视频片段对应所述待处理视频的关键信息；

将更新后的目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频。

- 10 8.一种视频处理装置，其中，所述装置包括：

信息获取模块，用于获取待处理视频对应的关键信息；

片段提取模块，用于根据所述关键信息，从所述待处理视频中提取一个或多个目标视频片段；

- 15 视频处理模块，用于将所述目标视频片段添加至所述待处理视频中指定视频帧之前的位置，得到目标视频，所述指定视频帧为所述待处理视频中前预设数量个视频帧中的任一视频帧。

9.一种计算机可读介质，其上存储有计算机程序，其中，该程序被处理装置执行时实现权利要求 1 至 7 中任一项所述方法的步骤。

10.一种电子设备，其包括：

- 20 存储装置，其上存储有计算机程序；

处理装置，用于执行所述存储装置中的所述计算机程序，以实现权利要求 1 至 7 中任一项所述方法的步骤。

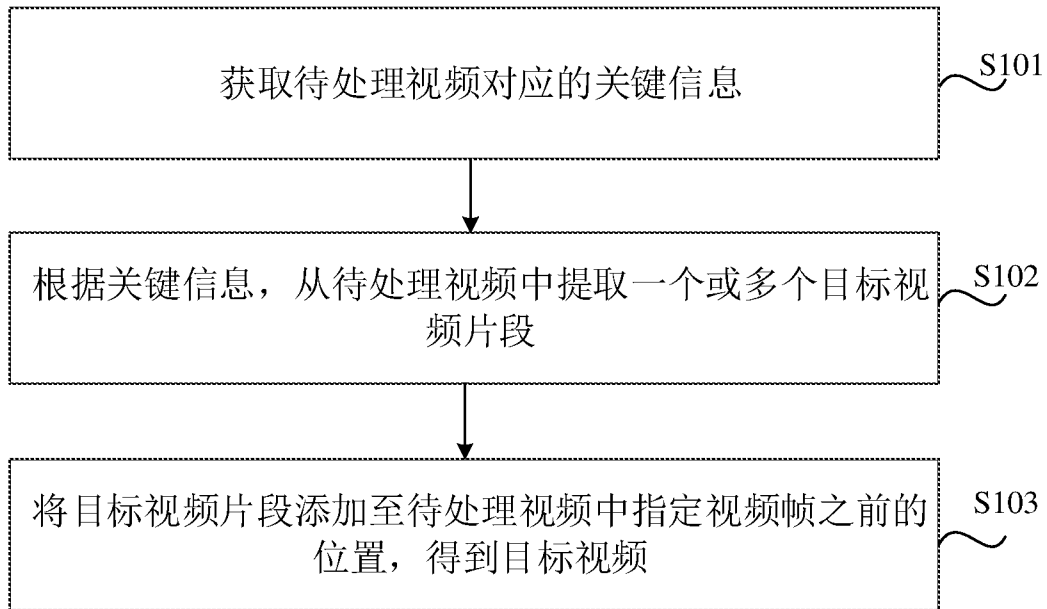


图 1

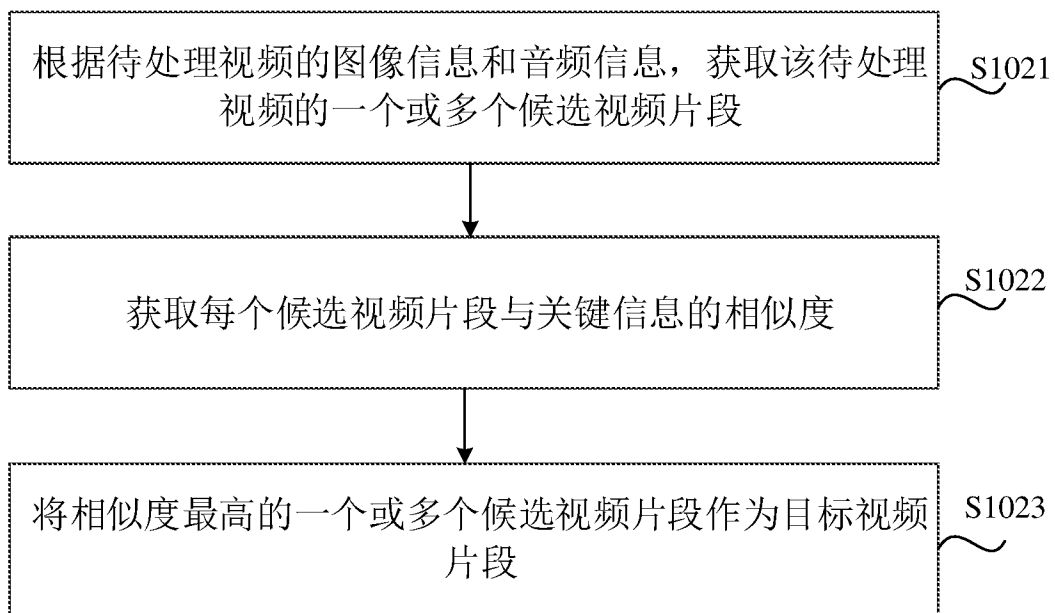


图 2

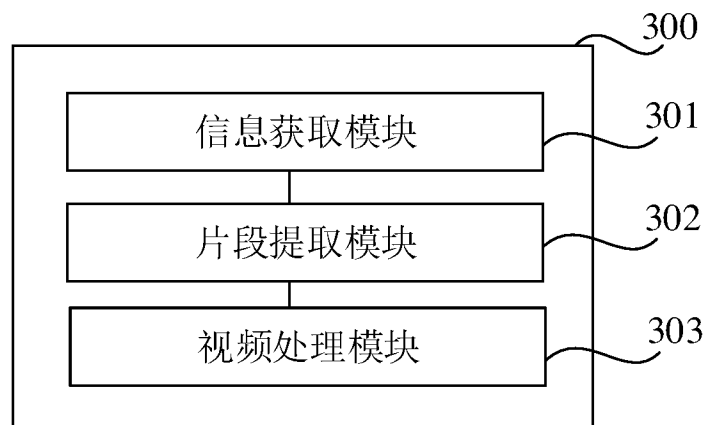


图 3

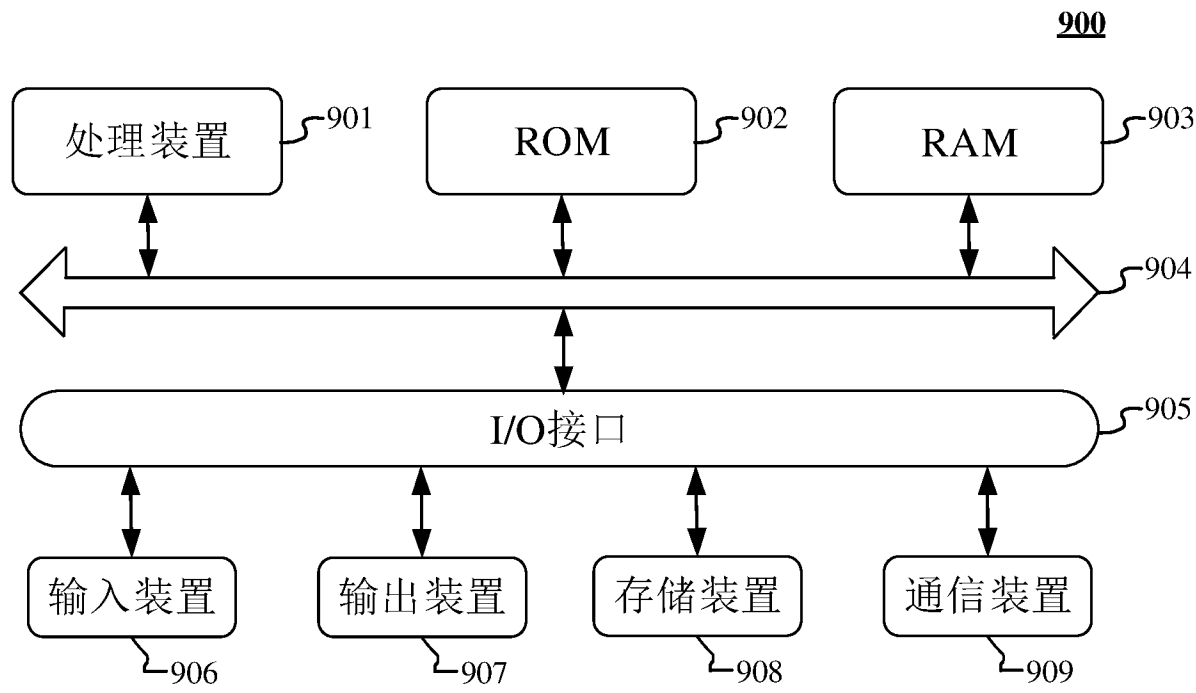


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/070262

A. CLASSIFICATION OF SUBJECT MATTER

G06V20/40(2022.01)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:G06V20/-

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, USTXT, VEN, 互联网检索, Internet search: 视频, 摘要, 概要, 精华, 浓缩, 热点, 预览, 提取, 选取, 获得, 识别, 筛选, 相似度, 相似性, 匹配度, 相关度, 相关性, 相似程度, 相关系数, 近似度, 相关程度, 关联性, 关键信息, 标题, 关键词, 标签, 关键字, 主题, 简介, Video??. Abstract, hotspot, preview, Extract+, identif+, recogn+, Similar+, key information, keyword, headline, title, theme, label

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 114445754 A (BEIJING YOUZHUJU NETWORK TECHNOLOGY CO., LTD.) 06 May 2022 (2022-05-06) claims 1-10	1-10
X	CN 113536036 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 22 October 2021 (2021-10-22) description, paragraphs [0098]-[0146]	1-10
X	US 2016070962 A1 (GOOGLE INC.) 10 March 2016 (2016-03-10) description, paragraphs [0052]-[0068]	1-10



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

16 May 2023

Date of mailing of the international search report

30 May 2023

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/CN2023/070262

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	114445754	A	06 May 2022	None			
CN	113536036	A	22 October 2021	HK	40053996	A0	18 February 2022
US	2016070962	A1	10 March 2016	WO	2016038522	A1	17 March 2016
				US	9953222	B2	24 April 2018
				US	2018239964	A1	23 August 2018
				US	10867183	B2	15 December 2020
				EP	3192273	A1	19 July 2017
				EP	3192273	A4	23 May 2018
				US	2021166035	A1	03 June 2021
				CN	107077595	A	18 August 2017

A. 主题的分类

G06V20/40 (2022.01) 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: G06V20/-

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, CNTXT, USTXT, VEN, 互联网检索: 视频, 摘要, 概要, 精华, 浓缩, 热点, 预览, 提取, 选取, 获得, 识别, 筛选, 相似度, 相似性, 匹配度, 相关度, 相关性, 相似程度, 相关系数, 近似度, 相关程度, 关联性, 关键信息, 标题, 关键词, 标签, 关键字, 主题, 简介, Video??, Abstract, hotspot, preview, Extract+, identif+, recogn+, Similar+, key information, keyword, headline, title, theme, label

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 114445754 A (北京有竹居网络技术有限公司) 2022年5月6日 (2022 - 05 - 06) 权利要求1-10	1-10
X	CN 113536036 A (腾讯科技 (深圳) 有限公司) 2021年10月22日 (2021 - 10 - 22) 说明书第[0098]-[0146]段	1-10
X	US 2016070962 A1 (GOOGLE INC.) 2016年3月10日 (2016 - 03 - 10) 说明书第[0052]-[0068]段	1-10

☐ 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

2023年5月16日

国际检索报告邮寄日期

2023年5月30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

刘洛

电话号码 (+86) 028-62967643

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/CN2023/070262

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	114445754	A	2022年5月6日	无			
CN	113536036	A	2021年10月22日	HK	40053996	A0	2022年2月18日
US	2016070962	A1	2016年3月10日	W0	2016038522	A1	2016年3月17日
				US	9953222	B2	2018年4月24日
				US	2018239964	A1	2018年8月23日
				US	10867183	B2	2020年12月15日
				EP	3192273	A1	2017年7月19日
				EP	3192273	A4	2018年5月23日
				US	2021166035	A1	2021年6月3日
				CN	107077595	A	2017年8月18日