

(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
2025 年 1 月 9 日 (09.01.2025)



(10) 国际公布号
WO 2025/007738 A1

(51) 国际专利分类号:
H04N 21/43 (2011.01)

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)。

(21) 国际申请号: PCT/CN2024/099889

(74) 代理人: 北京市柳沈律师事务所 (**LIU, SHEN & ASSOCIATES**); 中国北京市海淀区彩和坊路 10 号 1 号楼 10 层, Beijing 100080 (CN)。

(22) 国际申请日: 2024 年 6 月 18 日 (18.06.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310813430.4 2023 年 7 月 4 日 (04.07.2023) CN

(71) 申请人: 抖音视界有限公司 (**DOUYIN VISION CO., LTD.**) [CN/CN]; 中国北京市石景山区实兴大街 30 号院 3 号楼 2 层 B-0035 房间, Beijing 100041 (CN)。

(72) 发明人: 戴智勇 (**DAI, Zhiyong**); 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)。接宏恩 (**JIE, Hongen**); 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)。王继成 (**WANG, Jicheng**);

(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。

(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,

(54) **Title:** AUDIO-PICTURE SYNCHRONIZATION DETECTION METHOD AND APPARATUS, AND DEVICE AND STORAGE MEDIUM

(54) 发明名称: 音画同步检测方法、装置、设备和存储介质

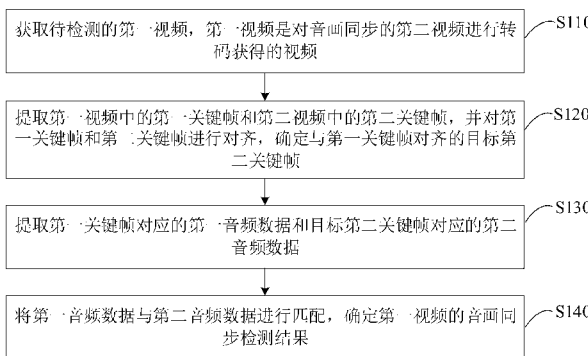


图 1

(57) **Abstract:** Provided in the present disclosure are an audio-picture synchronization detection method and apparatus, and a device and a storage medium. The method comprises: acquiring a first video to be subjected to detection, wherein the first video is a video obtained by means of transcoding a second video with audios and pictures synchronized (S110); extracting a first key frame in the first video and second key frames in the second video, and aligning the first key frame with the second key frames, so as to determine a target second key frame which is aligned with the first key frame (S120); extracting first audio data corresponding to the first key frame and second audio data corresponding to the target second key frame (S130); and matching the first audio data with the second audio data, so as to determine an audio-picture synchronization detection result for the first video (S140). The present disclosure makes it possible to perform automatic audio-picture synchronization detection on a transcoded video, thereby improving the efficiency for audio-picture synchronization detection and ensuring the accuracy of audio-picture synchronization detection.



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))。

(57) 摘要: 本公开提供了一种音画同步检测方法、装置、设备和存储介质。该方法包括: 获取待检测的第一视频, 第一视频是对音画同步的第二视频进行转码获得的视频 (S110); 提取第一视频中的第一关键帧和第二视频中的第二关键帧, 并对第一关键帧和第二关键帧进行对齐, 确定与第一关键帧对齐的目标第二关键帧 (S120); 提取第一关键帧对应的第一音频数据和目标第二关键帧对应的第二音频数据 (S130); 将第一音频数据与第二音频数据进行匹配, 确定第一视频的音画同步检测结果 (S140)。本公开可以对转码后的视频进行音画同步的自动检测, 从而提高了音画同步的检测效率, 并且保证了音画同步检测的准确性。

音画同步检测方法、装置、设备和存储介质

本申请要求于 2023 年 7 月 4 日提交的中国专利申请第 202310813430.4
5 号的优先权，该中国专利申请的全文通过引用的方式结合于此以作为本申请的一部分。

技术领域

本公开实施例涉及一种音画同步检测方法、装置、设备和存储介质。

10

背景技术

随着计算机技术的快速发展，往往需要对视频进行转码处理。然而，在视频转码过程中往往会出现音画不同步的情况，也就是音频和视频画面不能准确保持一致。例如，视频画面显示正在说话但没有相应的声音，极大降低了用户观看体验。目前，通常是通过人工观看视频的方式，自主判断是否存在因转码而导致音画不同步的情况。可见，这种人工检测方式费时费力，降低了音画同步的检测效率。

15

发明内容

20 本公开提供一种音画同步检测方法、装置、设备和存储介质，以对转码后的视频进行音画同步的自动检测，从而提高了音画同步的检测效率，并且保证了音画同步检测的准确性。

第一方面，本公开实施例提供了一种音画同步检测方法，包括：

25 获取待检测的第一视频，所述第一视频是对音画同步的第二视频进行转码获得的视频；

提取所述第一视频中的第一关键帧和所述第二视频中的第二关键帧，并对所述第一关键帧和所述第二关键帧进行对齐，确定与所述第一关键帧对齐的目标第二关键帧；

30 提取所述第一关键帧对应的第一音频数据和所述目标第二关键帧对应的第二音频数据；

将所述第一音频数据与所述第二音频数据进行匹配，确定所述第一视频的音画同步检测结果。

第二方面，本公开实施例还提供了一种音画同步检测装置，包括：

第一视频获取模块，用于获取待检测的第一视频，所述第一视频是对音画同步的第二视频进行转码获得的视频；

关键帧对齐模块，用于提取所述第一视频中的第一关键帧和所述第二视频中的第二关键帧，并对所述第一关键帧和所述第二关键帧进行对齐，确定与所述第一关键帧对齐的目标第二关键帧；

音频数据提取模块，用于提取所述第一关键帧对应的第一音频数据和所述目标第二关键帧对应的第二音频数据；

音画同步检测模块，用于将所述第一音频数据与所述第二音频数据进行匹配，确定所述第一视频的音画同步检测结果。

第三方面，本公开实施例还提供了一种电子设备，所述电子设备包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如本公开实施例任一所述的音画同步检测方法。

第四方面，本公开实施例还提供了一种包含计算机可执行指令的存储介质，所述计算机可执行指令在由计算机处理器执行时用于执行如本公开实施例任一所述的音画同步检测方法。

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。

图1是本公开实施例所提供的一种音画同步检测方法流程示意图；

图2是本公开实施例所涉及的一种第一音频数据和第二音频数据的波形匹配图；

图3是本公开实施例所提供的另一种音画同步检测方法流程示意图；

图 4 是本公开实施例所涉及的一种音频数据提取的示例；

图 5 是本公开实施例所提供的一种音画同步检测装置的结构示意图；

图 6 是本公开实施例所提供的一种电子设备的结构示意图。

5 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”。

15 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。

20 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”。

25 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。

图 1 为本公开实施例所提供的一种音画同步检测方法的流程示意图，本公开实施例适用于对转码后的视频进行音画同步检测的情况，该方法可以由音画同步检测装置来执行，该装置可以通过软件和/或硬件的形式实现，可选的，通过电子设备来实现，该电子设备可以是移动终端、PC 端或服务器等。

30 如图 1 所示，音画同步检测方法具体包括以下步骤：

S110、获取待检测的第一视频，第一视频是对音画同步的第二视频进行转码获得的视频。

其中，第一视频是指转码后的待检测视频。第二视频是指转码前的视频。第二视频中的音频和视频画面是同步的，从而可以将第二视频作为参考视频，
5 对第一视频进行音画同步的检测。

具体地，用户可以在客户端中制作视频并发布视频。服务器在接收到客户端发布上传的视频后，需要对上传视频进行转码处理，比如调整视频分辨率和码率等，从而可以提升画面质量，并将转码后的视频下发至其他客户端中进行播放。在上述应用场景中，可以将上传至服务器的视频作为音画同步的第二视
10 频，转码后的视频作为待检测的第一视频。通过以转码前的第二视频作为音画同步的参考视频，可以准确地检测出第一视频是否存在因视频转码而导致的音画不同步的情况。

S120、提取第一视频中的第一关键帧和第二视频中的第二关键帧，并对第一关键帧和第二关键帧进行对齐，确定与第一关键帧对齐的目标第二关键
15 帧。

其中，关键帧可以是指对视频内容起到决定性作用的视频帧，比如，存在场景切换的视频帧、存在运动变化的视频帧、关键动作所处于的视频帧等。第一关键帧可以是指第一视频中的关键帧。第一关键帧的数量可以为一个或多个。第二关键帧可以是指第二视频中的关键帧。第二关键帧的数量也可以为一个或多个。第一关键帧的数量与第二关键帧的数量可能相同，也可能不同。对
20 齐是指将具有相同画面的关键帧进行对齐的操作。目标第二关键帧是指与第一关键帧具有相同画面的第二关键帧。

具体地，对第一视频和第二视频进行关键帧的检测，提取出第一视频中的所有第一关键帧和第二视频中的所有第二关键帧。可以通过检测第一关键帧
25 与第二关键帧之间的图像相似度的方式，对第一关键帧和第二关键帧进行对齐处理。例如，针对每个第一关键帧而言，可以确定第一关键帧与每个第二关键帧之间的图像相似度，并可以将图像相似度最高且大于或等于预设相似度阈值的第二关键帧作为与该第一关键帧对齐的目标第二关键帧，从而获得与该第一关键帧具有相同画面的目标第二关键帧。

30 其中，图像相似度可以利用任意一种能够衡量两个图像相似度的指标进

行表征。例如，图像相似度利用结构相似性 SSIM (Structural Similarity) 指标进行表征。SSIM 指标的取值范围为 0 到 1，并且 SSIM 指标越大表明图像相似度越高。

需要说明的是，若所有第二关键帧中不存在与某个第一关键帧相同的目标第二关键帧，则表明无法对该第一关键帧进行对齐，此时可以忽略该第一关键帧，仅便利用能够对齐的第一关键帧进行后续的音画同步检测。

S130、提取第一关键帧对应的第一音频数据和目标第二关键帧对应的第二音频数据。

具体地，针对每个第一关键帧而言，可以在第一视频中的第一关键帧位置处，提取预设音频时长对应的第一音频数据。在第二视频中的目标第二关键帧位置处，也提取该预设音频时长对应的第二音频数据，从而提取出的第一音频数据和第二音频数据具有相同音频时长。例如，提取出的第一音频数据是以第一关键帧的播放时间戳为预设参考时刻，具有预设音频时长的音频数据。提取出的第二音频数据是以目标第二关键帧的播放时间戳为预设参考时刻，具有预设音频时长的音频数据。其中，预设参考时刻可以是指音频数据的开始时刻、中心时刻或者结束时刻。

需要说明的是，对齐后的第一关键帧和目标第二视频帧的画面内容是相同的，但其对应的播放时间戳可能是不同的。在提取出的第一音频数据和第二音频数据具有相同音频时长时，第一音频数据和第二音频数据需要在关键帧的相同位置处提取，以保证后续音画同步检测的准确性。例如，从第一视频中提取以第一关键帧的播放时间戳为开始时刻，具有预设音频时长的第一音频数据。从第二视频中提取以目标第二关键帧的播放时间戳为开始时刻，具有预设音频时长的第二音频数据。

S140、将第一音频数据与第二音频数据进行匹配，确定第一视频的音画同步检测结果。

具体地，图 2 给出了一种第一音频数据和第二音频数据的波形匹配图。如图 2 所示，第一音频数据和第二音频数据具有相同的音频时长，针对于此，可以基于图 2 中两个音频波形之间的相似程度确定第一音频数据与第二音频数据之间的匹配结果。例如，可以将第一音频数据和第二音频数据之间的协方差作为第一音频数据与第二音频数据之间的音频相似度，并在该音频相似度

大于或等于预设相似度阈值时，确定第一音频数据与第二音频数据匹配成功。在该音频相似度小于预设相似度阈值时，确定第一音频数据与第二音频数据匹配失败。若匹配成功，则表明在该第一关键帧位置处的音频和视频画面是同步的。若匹配失败，则表明在该第一关键帧位置处的音频和视频画面是不同步的，存在因转码而导致的音画不同步的情况。根据第一视频中的每个第一关键帧对应的匹配结果，可以确定第一视频的音画同步检测结果。例如，若第一视频中所有第一关键帧对应的匹配结果为匹配成功，则确定第一视频的音画同步检测结果为音画同步。若第一视频中存在至少一个第一关键帧对应的匹配结果为匹配失败，则确定第一视频的音画同步检测结果为音画不同步，第一视频中存在因转码而导致的音画不同步的情况。

本公开实施例的技术方案，通过将转码前音画同步的第二视频作为参考对转码后的第一视频进行检测，并将关键帧对齐后的第一关键帧对应的第一音频数据和目标第二关键帧对应的第二音频数据进行匹配，从而可以准确地确定第一视频是否存在因转码而导致的音画不同步的情况，实现了音画同步的自动检测，并且保证了音画同步检测的准确性。

作为一种可选的实施例，S120中的“对第一关键帧和第二关键帧进行对齐，确定与第一关键帧对齐的目标第二关键帧”，可以包括：获取第一关键帧可选的第二关键帧序列；按照第二关键帧序列的顺序，依次获取当前第二关键帧，并确定第一关键帧与当前第二关键帧之间的图像相似度；若图像相似度大于或等于预设相似度阈值，则将当前第二视频帧确定为与第一关键帧对齐的目标第二关键帧。

其中，图像相似度可以利用任意一种能够衡量两个图像相似度的指标进行表征。例如，图像相似度利用结构相似性 SSIM (Structural Similarity) 指标进行表征。SSIM 指标的取值范围为 0 到 1，并且 SSIM 指标越大表明图像相似度越高。

具体地，提取出的所有第一关键帧可以按照视频播放先后顺序进行排序，获得第一关键帧序列，比如{A1,A2,.....,An}。同理，提取出的所有第二关键帧可以按照视频播放先后顺序进行排序，获得第二关键帧序列，比如{B1,B2,.....,Bm}。可以按照第一关键帧序列的顺序，依次从第二关键帧序列中确定每个第一关键帧对齐的目标第二关键帧。

例如, 第一个第一关键帧 A1 可选的第二关键帧序列为 {B1,B2,……,Bm}, 此时按照该第二关键帧序列的顺序, 可以先将第二关键帧 B1 作为当前第二关键帧, 并检测第一关键帧 A1 与第二关键帧 B1 之间的图像相似度是否大于或等于预设相似度阈值; 若是, 则表明第一关键帧 A1 与第二关键帧 B1 画面相似, 此时可以直接将第二关键帧 B1 作为与第一关键帧 A1 对齐的目标第二关键帧; 若否, 则将下一个第二关键帧 B2 作为当前第二关键帧重新检测第一关键帧 A1 与第二关键帧 B2 之间的图像相似度是否大于或等于预设相似度阈值, 直到确定出与第一关键帧 A1 相匹配的第二关键帧为止, 从而完成第一关键帧 A1 的对齐。假设与第一关键帧 A1 对齐的目标第二关键帧为 B2, 则对第二个第一关键帧 A2 进行对齐时, 由于与第一关键帧 A2 对齐的目标第二关键帧是出现在第二关键帧 B2 后面的关键帧中, 从而可以将第二关键帧 B2 之后的第二关键帧作为第一关键帧 A2 可选的第二关键帧序列, 即 {B3,B4,……,Bm}, 并基于上述相同的检测过程, 从第二关键帧序列 {B3,B4,……,Bm} 中确定与第一关键帧 A2 对齐的目标第二关键帧。同理依次执行, 可以更加快速地获得与每个第一关键帧对齐的目标第二关键帧, 从而提高了关键帧对齐速度, 进一步提高了音画同步的检测效率。

图 3 为本公开实施例所提供的另一种音画同步检测方法的流程示意图, 本公开实施例在上述公开实施例的基础上, 提取的第一音频数据的音频时长大于第二音频数据的音频时长, 并在此基础上, 对步骤“将第一音频数据与第二音频数据进行匹配, 确定第一视频的音画同步检测结果”进行了优化。其中与上述各公开实施例相同或相应的术语的解释在此不再赘述。

如图 3 所示, 音画同步检测方法具体包括以下步骤:

S310、获取待检测的第一视频, 第一视频是对音画同步的第二视频进行转码获得的视频。

S320、提取第一视频中的第一关键帧和第二视频中的第二关键帧, 并对第一关键帧和第二关键帧进行对齐, 确定与第一关键帧对齐的目标第二关键帧。

S330、提取第一关键帧对应的第一音频数据和目标第二关键帧对应的第二音频数据。

其中, 第一音频数据对应的音频时长为第一预设时长, 第二音频数据对应

的音频时长为第二预设时长。第一预设时长大于第二预设时长。例如，第一预设时长为 500ms，第二预设时长为 20ms。

具体地，针对每个第一关键帧而言，可以在第一视频中的第一关键帧位置处，提取第一预设时长对应的第一音频数据。在第二视频中的目标第二关键帧位置处，提取第二预设时长对应的第二音频数据，从而提取出的第一音频数据的音频时长大于第二音频数据的音频时长，以便后续在第一音频数据中进行第二音频数据的匹配。

示例性地，第一音频数据的提取需要在第一关键帧位置处的左右范围内进行提取，以便后续进行准确的音频匹配。例如，第一音频数据可以是以第一关键帧的播放时间戳为中心时刻，具有第一预设时长的音频数据。图 4 给出了一种第一音频数据提取的示例。参见图 4，从第一视频的音频中提取以每个第一关键帧的播放时间戳为中心时刻，具有第一预设时长 500ms 的第一音频数据。

示例性地，用于作为参考的第二音频数据可以存在多种提取方式，其可以在目标第二关键帧位置处的左面、左面或者左右范围内进行提取。例如，第二音频数据可以是以目标第二关键帧的播放时间戳为预设参考时刻，具有第二预设时长的音频数据。其中，预设参考时刻可以为以下其中之一：开始时刻、中心时刻和结束时刻。图 4 也给出了一种第二音频数据提取的示例。参见图 4，从第二视频的音频中提取以每个目标第二关键帧的播放时间戳为开始时刻，具有第二预设时长 20ms 的第二音频数据。

S340、从第一音频数据中确定具有第二预设时长且与第二音频数据相匹配的目标音频数据。

其中，目标音频数据可以是指第一音频数据中与第二音频数据匹配成功的音频数据。也就是说，目标音频数据与第二音频数据可以近似于相同数据。

具体地，由于第一音频数据的时长大于第二音频数据的时长，从而可以对第一音频数据进行遍历，提取出多个具有第二预设时长的待选音频数据，比如每隔 1ms 提取一个具有第二预设时长的待选音频数据。可以确定每个待选音频数据与第二音频数据之间的音频相似度，并将音频相似度最大的待选音频数据作为与第二音频数据相匹配的目标音频数据。

示例性地，S340 可以包括：基于预设滑动时长和第二预设时长，对第一

音频数据进行滑动，获取当次滑动对应的具有第二预设时长的当前音频数据；确定当前音频数据与第二音频数据之间的音频相似度；基于当前音频数据对应的音频相似度，确定与第二音频数据相匹配的目标音频数据。

具体地，可以基于预设滑动时长，在第一音频数据进行逐次滑动，获得具有第二预设时长的当前音频数据。例如，在预设滑动时长为 1ms 时，参见图 4，第一次滑动获得的当前音频数据为第一音频数据中的第 0ms 到第 20ms 的音频数据，第二次滑动获得的当前音频数据为第一音频数据中的第 1ms 到第 21ms 的音频数据，依次类推。在每次滑动后，可以确定当次滑动获得的当前音频数据与第二音频数据之间的协方差，并将该协方差作为当前音频数据对应的音频相似度。若该音频相似度大于或等于预设相似度阈值，则可以确定当前音频数据与第二音频数据相匹配，此时将当前音频数据作为目标音频数据。若该音频相似度小于预设相似度阈值，则可以确定当前音频数据与第二音频数据不匹配，此时需要通过滑动更新当前音频数据，并检测下次滑动获得的当前音频数据是否与第二音频数据相匹配，直到确定出相匹配的目标音频数据时，停止滑动。通过上述滑动匹配的方式无需遍历第一音频数据中具有第二预设时长的所有音频数据，可以在保证同步检测准确性的基础上更加快速地获得与第二音频数据相匹配的目标音频数据，从而进一步提高了音画同步检测效率。

S350、以第二音频数据和目标第二关键帧为音画同步基准，确定目标音频数据和第一关键帧之间的音画偏移量。

其中，音画偏移量可以是指目标音频数据的声音与第一关键帧的画面之间的时间差。

具体地，由于第二视频是音画同步的，从而第二音频数据和目标第二关键帧是音画同步的，因此可以将第二音频数据与目标第二关键帧之间的第一时差作为音画同步的参考时差，对目标音频数据与第一关键帧之间的目标时差进行检测，若目标时差等于参考时差，则表明目标音频数据与第一关键帧是音画同步的，此时的音画偏移量为 0。若目标时差不等于参考时差，则可以基于目标时差与参考时差之间的差值，确定具体的音画偏移量。

示例性地，S350 可以包括：基于第二音频数据的开始时间戳和目标第二关键帧的播放时间戳，确定音画同步时的参考时差；基于目标音频数据的开始

时间戳和第一关键帧的播放时间戳，确定第一关键帧对应的目标时差；基于目标时差和参考时差，确定第一关键帧对应的音画偏移量。

具体地，可以将第二音频数据的开始时间戳减去目标第二关键帧的播放时间戳，获得的时间差值确定为音画同步时的参考时差。对于图 4 中的第二
5 音频数据的提取方式而言，第二音频数据的开始时间戳即为目标第二关键帧的播放时间戳，从而音画同步时的参考时差为 0。将目标音频数据的开始时间戳减去第一关键帧的播放时间戳，获得的时间差值确定为第一关键帧对应的目标时差。将目标时差减去参考时差，获得的时间差值确定为第一关键帧对应的音画偏移量。参见图 4，参考时差为 0，目标时差即为音画偏移量。若音画
10 偏移量大于 0，则表明音频超前视频画面。若音画偏移量小于 0，则表明音频滞后视频画面。

S360、基于第一关键帧对应的音画偏移量，确定第一视频的音画同步检测结果。

具体地，基于第一视频中的每个第一关键帧对应的音画偏移量，可以确定
15 第一视频整体上的音画同步检测结果。

示例性地，S360 可以包括：若第一视频中每个第一视频帧对应的音画偏移量均处于预设允许范围内，则确定第一视频的音画同步检测结果为音画同步；若第一视频中存在至少一个第一视频帧对应的音画偏移量未处于预设允许范围内，则确定第一视频的音画同步检测结果为音画不同步。

20 其中，预设允许范围可以是预先基于检测标准设置的，允许音画偏移的时差范围。例如，预设允许范围为： $[-185, 90]$ ，其表明相当于视频画面而言，音频滞后 185ms 或者超前 90ms 都是可接受的，认为也是音画同步的。

具体地，检测第一视频中每个第一视频帧对应的音画偏移量是否处于预设允许范围内，若所有的第一视频帧对应的音画偏移量均处于预设允许范围
25 内，则确定第一视频的音画同步检测结果为音画同步，即转码后的第一视频也是音画同步的，视频转码并未导致音画不同步的情况。若存在至少一个第一视频帧对应的音画偏移量未处于预设允许范围内，则确定第一视频的音画同步检测结果为音画不同步，也就是说，在未处于预设允许范围内的第一视频帧位置处出现了因转码而导致的音画不同步的情况，此时可以将未处于预设允许
30 范围内的第一视频帧对应的音画偏移量进行输出，以提醒用户音画不同步的

具体位置以及音画不同步的程度大小，使得用户进行快速处理。

本公开实施例的技术方案，通过提取音频时长更长的第一音频数据，从第一音频数据中再次提取出与第二音频数据相匹配的目标音频数据，并以第二音频数据和目标第二关键帧为音画同步基准，确定目标音频数据和第一关键帧之间的音画偏移量，并基于第一关键帧对应的音画偏移量，可以更加准确地确定出第一视频的音画同步检测结果，从而进一步提高了音画同步检测的准确性。

图 5 为本公开实施例所提供的一种音画同步检测装置的结构示意图，如图 5 所示，该装置具体包括：第一视频获取模块 510、关键帧对齐模块 520、音频数据提取模块 530 和音画同步检测模块 540。

其中，第一视频获取模块 510，用于获取待检测的第一视频，所述第一视频是对音画同步的第二视频进行转码获得的视频；关键帧对齐模块 520，用于提取所述第一视频中的第一关键帧和所述第二视频中的第二关键帧，并对所述第一关键帧和所述第二关键帧进行对齐，确定与所述第一关键帧对齐的目标第二关键帧；音频数据提取模块 530，用于提取所述第一关键帧对应的第一音频数据和所述目标第二关键帧对应的第二音频数据；音画同步检测模块 540，用于将所述第一音频数据与所述第二音频数据进行匹配，确定所述第一视频的音画同步检测结果。

本公开实施例所提供的技术方案，通过将转码前音画同步的第二视频作为参考对转码后的第一视频进行检测，并将关键帧对齐后的第一关键帧对应的第一音频数据和目标第二关键帧对应的第二音频数据进行匹配，从而可以准确地确定第一视频是否存在因转码而导致的音画不同步的情况，实现了音画同步的自动检测，并且保证了音画同步检测的准确性。

在上述技术方案的基础上，关键帧对齐模块 520，具体用于：

获取所述第一关键帧可选的第二关键帧序列；按照所述第二关键帧序列的顺序，依次获取当前第二关键帧，并确定所述第一关键帧与当前第二关键帧之间的图像相似度；若所述图像相似度大于或等于预设相似度阈值，则将当前第二视频帧确定为与所述第一关键帧对齐的目标第二关键帧。

在上述各技术方案的基础上，所述第一音频数据对应的音频时长为第一预设时长，所述第二音频数据对应的音频时长为第二预设时长，其中，所述第

一预设时长大于第二预设时长。

在上述各技术方案的基础上，所述第一音频数据是以所述第一关键帧的播放时间戳为中心时刻，具有第一预设时长的音频数据；

5 所述第二音频数据是以所述目标第二关键帧的播放时间戳为预设参考时刻，具有第二预设时长的音频数据；其中，所述预设参考时刻为以下其中之一：开始时刻、中心时刻和结束时刻。

在上述各技术方案的基础上，音画同步检测模块 540，包括：

目标音频数据确定单元，用于从所述第一音频数据中确定具有所述第二预设时长且与所述第二音频数据相匹配的目标音频数据；

10 音画偏移量确定单元，用于以所述第二音频数据和所述目标第二关键帧为音画同步基准，确定所述目标音频数据和所述第一关键帧之间的音画偏移量；

音画同步检测单元，用于基于所述第一关键帧对应的音画偏移量，确定所述第一视频的音画同步检测结果。

15 在上述各技术方案的基础上，目标音频数据确定单元，具体用于：

基于预设滑动时长和所述第二预设时长，对所述第一音频数据进行滑动，获取当次滑动对应的具有所述第二预设时长的当前音频数据；确定所述当前音频数据与所述第二音频数据之间的音频相似度；基于所述当前音频数据对应的音频相似度，确定与所述第二音频数据相匹配的目标音频数据。

20 在上述各技术方案的基础上，音画偏移量确定单元，具体用于：

基于所述第二音频数据的开始时间戳和所述目标第二关键帧的播放时间戳，确定音画同步时的参考时差；基于所述目标音频数据的开始时间戳和所述第一关键帧的播放时间戳，确定所述第一关键帧对应的目标时差；基于所述目标时差和所述参考时差，确定所述第一关键帧对应的音画偏移量。

25 在上述各技术方案的基础上，音画同步检测单元，具体用于：

若所述第一视频中每个第一视频帧对应的音画偏移量均处于预设允许范围内，则确定所述第一视频的音画同步检测结果为音画同步；若所述第一视频中存在至少一个第一视频帧对应的音画偏移量未处于预设允许范围内，则确定所述第一视频的音画同步检测结果为音画不同步。

30 本公开实施例所提供的音画同步检测装置可执行本公开任意实施例所提

供的音画同步检测方法，具备执行音画同步检测方法相应的功能模块和有益效果。

值得注意的是，上述装置所包括的各个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，各功能单元的具体名称也只是为了便于相互区分，并不用于限制本公开实施

例的保护范围。

图 6 为本公开实施例所提供的一种电子设备的结构示意图。下面参考图 6，其示出了适于用来实现本公开实施例的电子设备（例如图 6 中的终端设备或服务器）500 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 6 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 6 所示，电子设备 500 可以包括处理装置（例如中央处理器、图形处理器等）501，其可以根据存储在只读存储器（ROM）502 中的程序或者从存储装置 508 加载到随机访问存储器（RAM）503 中的程序而执行各种适当的动作和处理。在 RAM 503 中，还存储有电子设备 500 操作所需的各种程序和数据。处理装置 501、ROM 502 以及 RAM 503 通过总线 504 彼此相连。编辑/输出（I/O）接口 505 也连接至总线 504。

通常，以下装置可以连接至 I/O 接口 505：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 506；包括例如液晶显示器（LCD）、扬声器、振动器等等的输出装置 507；包括例如磁带、硬盘等的存储装置 508；以及通信装置 509。通信装置 509 可以允许电子设备 500 与其他设备进行无线或有线通信以交换数据。虽然图 6 示出了具有各种装置的电子设备 500，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 509 从网络上被下载和安装，或者从存储装置 508 被安装，或者从 ROM 502 被安装。在该计算机程序被处理装置 501 执行时，执行本公开实施例的方法中限定的上述功能。

5 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。

 本公开实施例提供的电子设备与上述实施例提供的音画同步检测方法属于同一发明构思，未在本实施例中详尽描述的技术细节可参见上述实施例，并且本实施例与上述实施例具有相同的有益效果。

10 本公开实施例提供了一种计算机存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述实施例所提供的音画同步检测方法。

 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——但不限于——电、磁、光、电磁、红外线、或半导体的系统、
15 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
20 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
25 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
30 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。

- 5 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:获取待检测的第一视频,所述第一视频是对音画同步的第二视频进行转码获得的视频;提取所述第一视频中的
10 第一关键帧和所述第二视频中的第二关键帧,并对所述第一关键帧和所述第二关键帧进行对齐,确定与所述第一关键帧对齐的目标第二关键帧;提取所述第一关键帧对应的第一音频数据和所述目标第二关键帧对应的第二音频数据;将所述第一音频数据与所述第二音频数据进行匹配,确定所述第一视频的音画同步检测结果。

- 15 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
25 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
30

的是，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单元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，单元的名称在某种情况下并不构成对该单元本身的限定，例如，第一获取单元还可以被描述为“获取至少两个网际协议地址的单元”。

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。

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、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，示例一提供了一种音画同步检测方法，包括：

获取待检测的第一视频，所述第一视频是对音画同步的第二视频进行转码获得的视频；

25 提取所述第一视频中的第一关键帧和所述第二视频中的第二关键帧，并对所述第一关键帧和所述第二关键帧进行对齐，确定与所述第一关键帧对齐的目标第二关键帧；

提取所述第一关键帧对应的第一音频数据和所述目标第二关键帧对应的第二音频数据；

30 将所述第一音频数据与所述第二音频数据进行匹配，确定所述第一视频

的音画同步检测结果。

根据本公开的一个或多个实施例，示例二提供了一种音画同步检测方法，还包括：

5 可选的，对所述第一关键帧和所述第二关键帧进行对齐，确定与所述第一关键帧对齐的目标第二关键帧，包括：

获取所述第一关键帧可选的第二关键帧序列；

按照所述第二关键帧序列的顺序，依次获取当前第二关键帧，并确定所述第一关键帧与当前第二关键帧之间的图像相似度；

10 若所述图像相似度大于或等于预设相似度阈值，则将当前第二视频帧确定为与所述第一关键帧对齐的目标第二关键帧。

根据本公开的一个或多个实施例，示例三提供了一种音画同步检测方法，还包括：

15 可选的，所述第一音频数据对应的音频时长为第一预设时长，所述第二音频数据对应的音频时长为第二预设时长，其中，所述第一预设时长大于第二预设时长。

根据本公开的一个或多个实施例，示例四提供了一种音画同步检测方法，还包括：

可选的，所述第一音频数据是以所述第一关键帧的播放时间戳为中心时刻，具有第一预设时长的音频数据；

20 所述第二音频数据是以所述目标第二关键帧的播放时间戳为预设参考时刻，具有第二预设时长的音频数据；其中，所述预设参考时刻为以下其中之一：

开始时刻、中心时刻和结束时刻。

根据本公开的一个或多个实施例，示例五提供了一种音画同步检测方法，还包括：

25 可选的，所述将所述第一音频数据与所述第二音频数据进行匹配，确定所述第一视频的音画同步检测结果，包括：

从所述第一音频数据中确定具有所述第二预设时长且与所述第二音频数据相匹配的目标音频数据；

30 以所述第二音频数据和所述目标第二关键帧为音画同步基准，确定所述目标音频数据和所述第一关键帧之间的音画偏移量；

基于所述第一关键帧对应的音画偏移量，确定所述第一视频的音画同步检测结果。

根据本公开的一个或多个实施例，示例六提供了一种音画同步检测方法，还包括：

- 5 可选的，所述从所述第一音频数据中确定具有所述第二预设时长且与所述第二音频数据相匹配的目标音频数据，包括：

 基于预设滑动时长和所述第二预设时长，对所述第一音频数据进行滑动，获取当次滑动对应的具有所述第二预设时长的当前音频数据；

 确定所述当前音频数据与所述第二音频数据之间的音频相似度；

- 10 基于所述当前音频数据对应的音频相似度，确定与所述第二音频数据相匹配的目标音频数据。

根据本公开的一个或多个实施例，示例七提供了一种音画同步检测方法，还包括：

- 15 可选的，所述以所述第二音频数据和所述目标第二关键帧为音画同步基准，确定所述目标音频数据和所述第一关键帧之间的音画偏移量，包括：

 基于所述第二音频数据的开始时间戳和所述目标第二关键帧的播放时间戳，确定音画同步时的参考时差；

 基于所述目标音频数据的开始时间戳和所述第一关键帧的播放时间戳，确定所述第一关键帧对应的目标时差；

- 20 基于所述目标时差和所述参考时差，确定所述第一关键帧对应的音画偏移量。

根据本公开的一个或多个实施例，示例八提供了一种音画同步检测方法，还包括：

- 25 可选的，所述基于所述第一关键帧对应的音画偏移量，确定所述第一视频的音画同步检测结果，包括：

 若所述第一视频中每个第一视频帧对应的音画偏移量均处于预设允许范围内，则确定所述第一视频的音画同步检测结果为音画同步；

 若所述第一视频中存在至少一个第一视频帧对应的音画偏移量未处于预设允许范围内，则确定所述第一视频的音画同步检测结果为音画不同步。

- 30 根据本公开的一个或多个实施例，示例九提供了一种音画同步检测装置，

包括:

第一视频获取模块,用于获取待检测的第一视频,所述第一视频是对音画同步的第二视频进行转码获得的视频;

5 关键帧对齐模块,用于提取所述第一视频中的第一关键帧和所述第二视频中的第二关键帧,并对所述第一关键帧和所述第二关键帧进行对齐,确定与所述第一关键帧对齐的目标第二关键帧;

音频数据提取模块,用于提取所述第一关键帧对应的第一音频数据和所述目标第二关键帧对应的第二音频数据;

10 音画同步检测模块,用于将所述第一音频数据与所述第二音频数据进行匹配,确定所述第一视频的音画同步检测结果。

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
15 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,
20 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,
25 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。

权利要求书

1、一种音画同步检测方法，包括：

5 获取待检测的第一视频，所述第一视频是对音画同步的第二视频进行转码获得的视频；

提取所述第一视频中的第一关键帧和所述第二视频中的第二关键帧，并对所述第一关键帧和所述第二关键帧进行对齐，确定与所述第一关键帧对齐的目标第二关键帧；

10 提取所述第一关键帧对应的第一音频数据和所述目标第二关键帧对应的第二音频数据；

将所述第一音频数据与所述第二音频数据进行匹配，确定所述第一视频的音画同步检测结果。

15 2、根据权利要求1所述的音画同步检测方法，其中，对所述第一关键帧和所述第二关键帧进行对齐，确定与所述第一关键帧对齐的目标第二关键帧，包括：

获取所述第一关键帧可选的第二关键帧序列；

按照所述第二关键帧序列的顺序，依次获取当前第二关键帧，并确定所述第一关键帧与当前第二关键帧之间的图像相似度；

20 若所述图像相似度大于或等于预设相似度阈值，则将当前第二视频帧确定为与所述第一关键帧对齐的目标第二关键帧。

3、根据权利要求1所述的音画同步检测方法，其中，所述第一音频数据对应的音频时长为第一预设时长，所述第二音频数据对应的音频时长为第二预设时长，其中，所述第一预设时长大于所述第二预设时长。

25 4、根据权利要求3所述的音画同步检测方法，其中，所述第一音频数据是以所述第一关键帧的播放时间戳为中心时刻，具有第一预设时长的音频数据；

所述第二音频数据是以所述目标第二关键帧的播放时间戳为预设参考时刻，具有第二预设时长的音频数据；其中，所述预设参考时刻为以下其中之一：开始时刻、中心时刻和结束时刻。

30 5、根据权利要求3所述的音画同步检测方法，其中，所述将所述第一音

频数据与所述第二音频数据进行匹配，确定所述第一视频的音画同步检测结果，包括：

从所述第一音频数据中确定具有所述第二预设时长且与所述第二音频数据相匹配的目标音频数据；

5 以所述第二音频数据和所述目标第二关键帧为音画同步基准，确定所述目标音频数据和所述第一关键帧之间的音画偏移量；

基于所述第一关键帧对应的音画偏移量，确定所述第一视频的音画同步检测结果。

6、根据权利要求5所述的音画同步检测方法，其中，所述从所述第一音频数据中确定具有所述第二预设时长且与所述第二音频数据相匹配的目标音频数据，包括：

基于预设滑动时长和所述第二预设时长，对所述第一音频数据进行滑动，获取当次滑动对应的具有所述第二预设时长的当前音频数据；

确定所述当前音频数据与所述第二音频数据之间的音频相似度；

15 基于所述当前音频数据对应的音频相似度，确定与所述第二音频数据相匹配的目标音频数据。

7、根据权利要求5或6所述的音画同步检测方法，其中，所述以所述第二音频数据和所述目标第二关键帧为音画同步基准，确定所述目标音频数据和所述第一关键帧之间的音画偏移量，包括：

20 基于所述第二音频数据的开始时间戳和所述目标第二关键帧的播放时间戳，确定音画同步时的参考时差；

基于所述目标音频数据的开始时间戳和所述第一关键帧的播放时间戳，确定所述第一关键帧对应的目标时差；

25 基于所述目标时差和所述参考时差，确定所述第一关键帧对应的音画偏移量。

8、根据权利要求5-7任一项所述的音画同步检测方法，其中，所述基于所述第一关键帧对应的音画偏移量，确定所述第一视频的音画同步检测结果，包括：

30 若所述第一视频中每个第一视频帧对应的音画偏移量均处于预设允许范围内，则确定所述第一视频的音画同步检测结果为音画同步；

若所述第一视频中存在至少一个第一视频帧对应的音画偏移量未处于预设允许范围内，则确定所述第一视频的音画同步检测结果为音画不同步。

9、一种音画同步检测装置，包括：

第一视频获取模块，用于获取待检测的第一视频，所述第一视频是对音画

5 同步的第二视频进行转码获得的视频；

关键帧对齐模块，用于提取所述第一视频中的第一关键帧和所述第二视频中的第二关键帧，并对所述第一关键帧和所述第二关键帧进行对齐，确定与所述第一关键帧对齐的目标第二关键帧；

10 音频数据提取模块，用于提取所述第一关键帧对应的第一音频数据和所述目标第二关键帧对应的第二音频数据；

音画同步检测模块，用于将所述第一音频数据与所述第二音频数据进行匹配，确定所述第一视频的音画同步检测结果。

10、一种电子设备，包括：

一个或多个处理器；

15 存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-8 中任一所述的音画同步检测方法。

11、一种包含计算机可执行指令的存储介质，其中，所述计算机可执行指令在由计算机处理器执行时用于执行如权利要求 1-8 中任一所述的音画同步
20 检测方法。

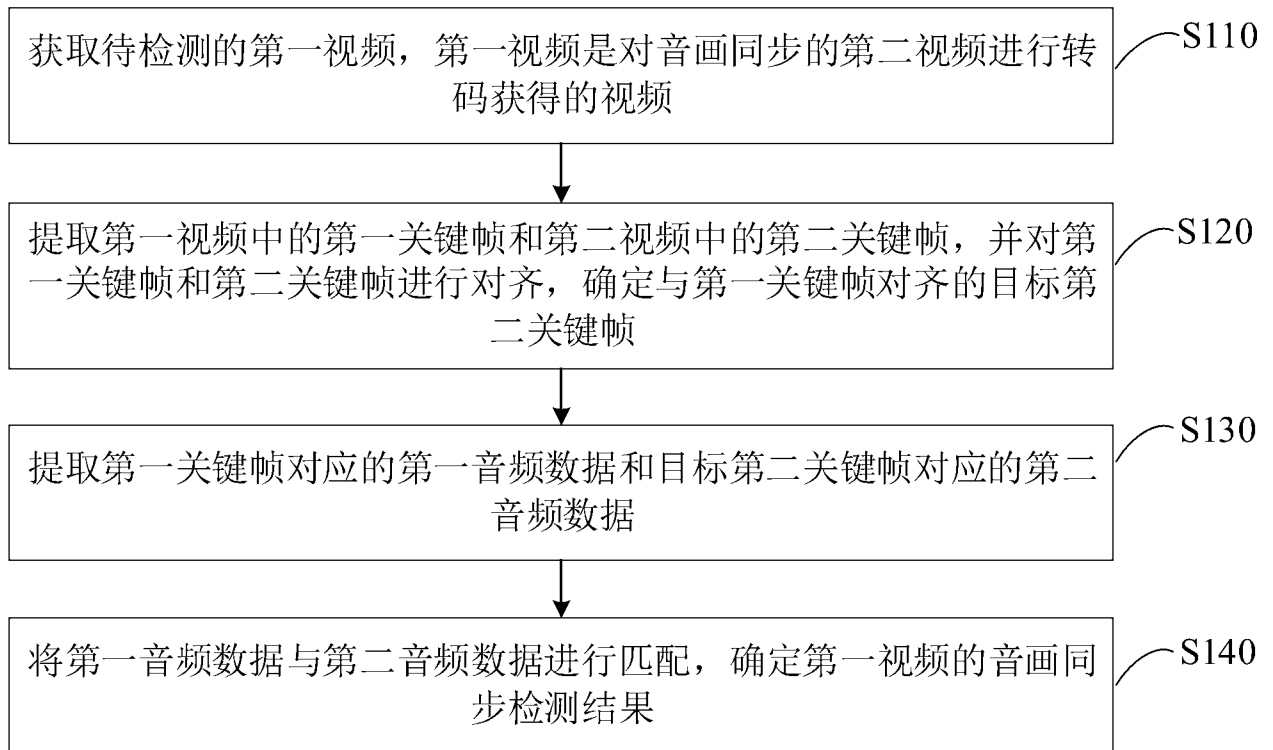


图 1

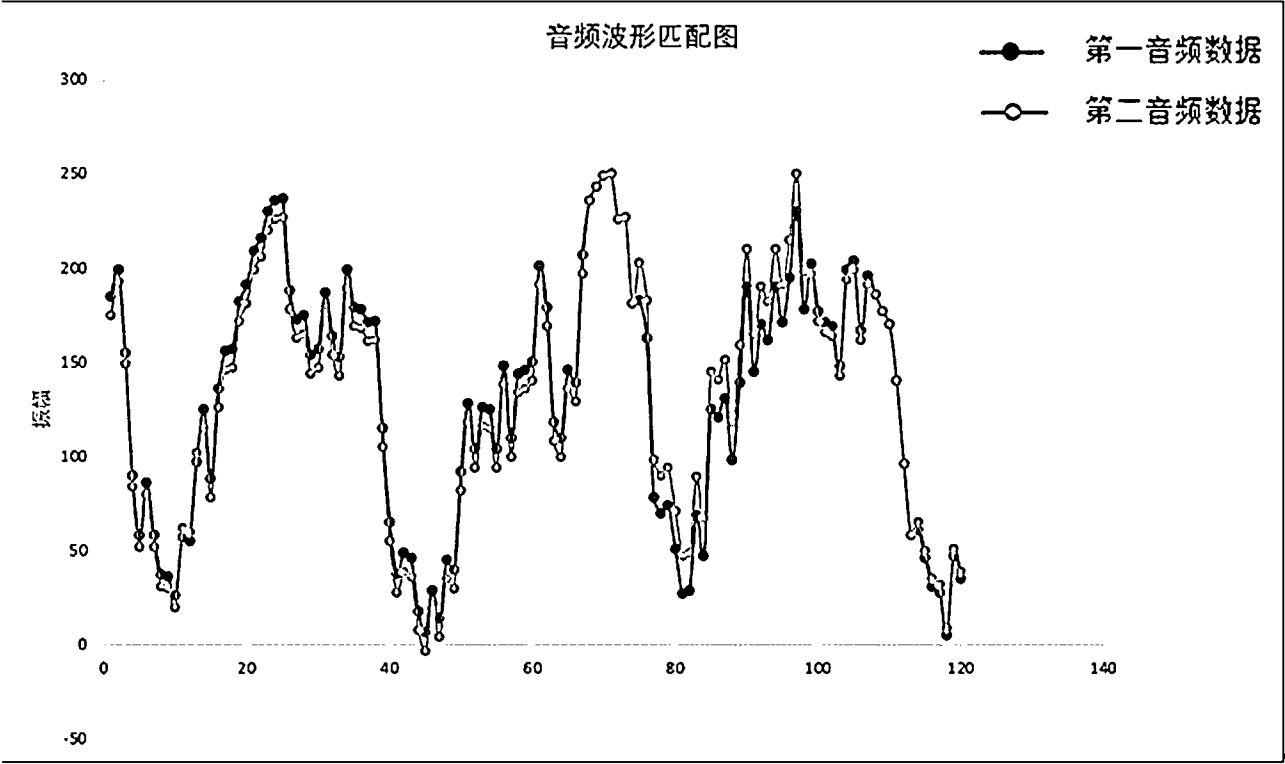


图 2

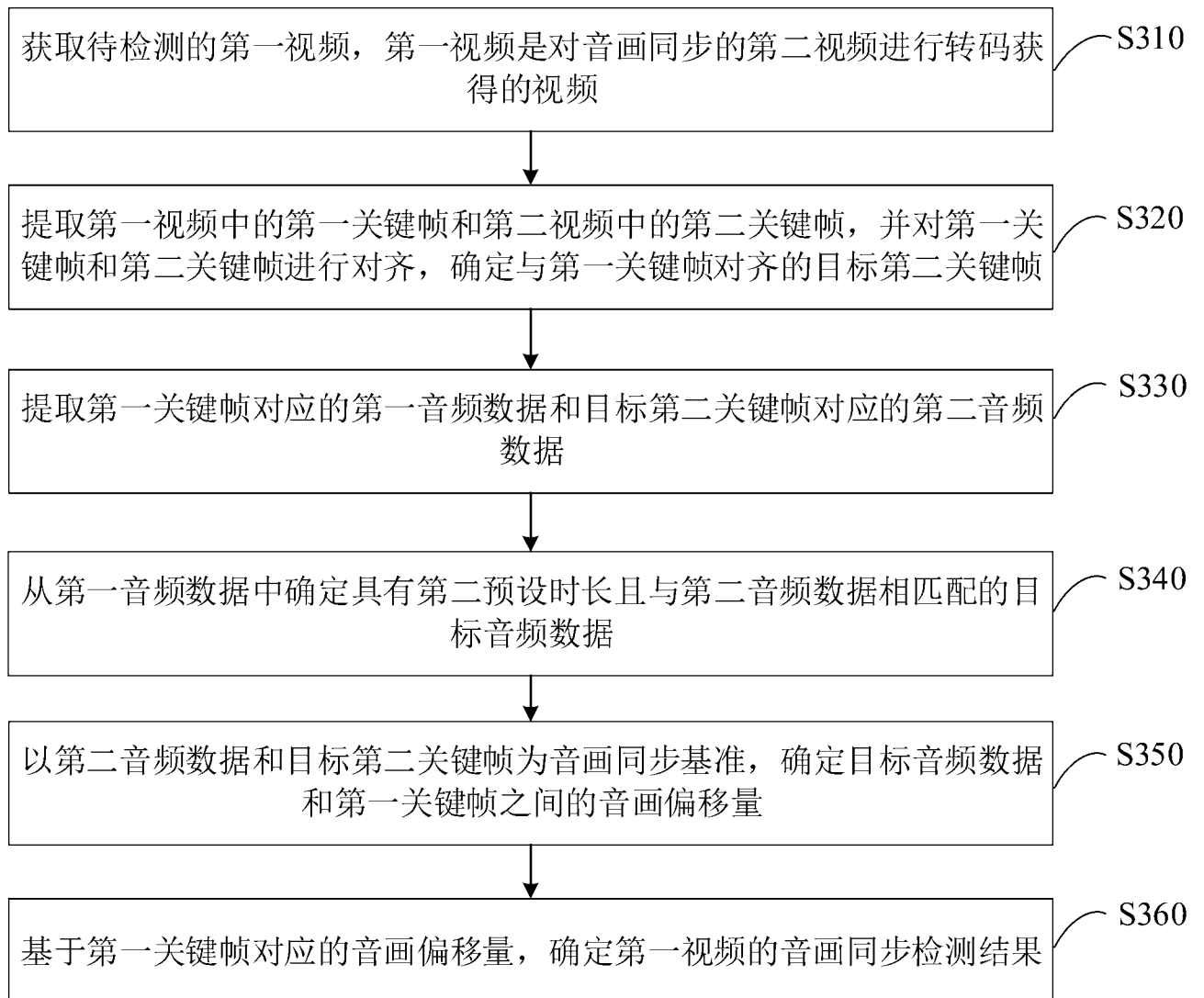


图 3

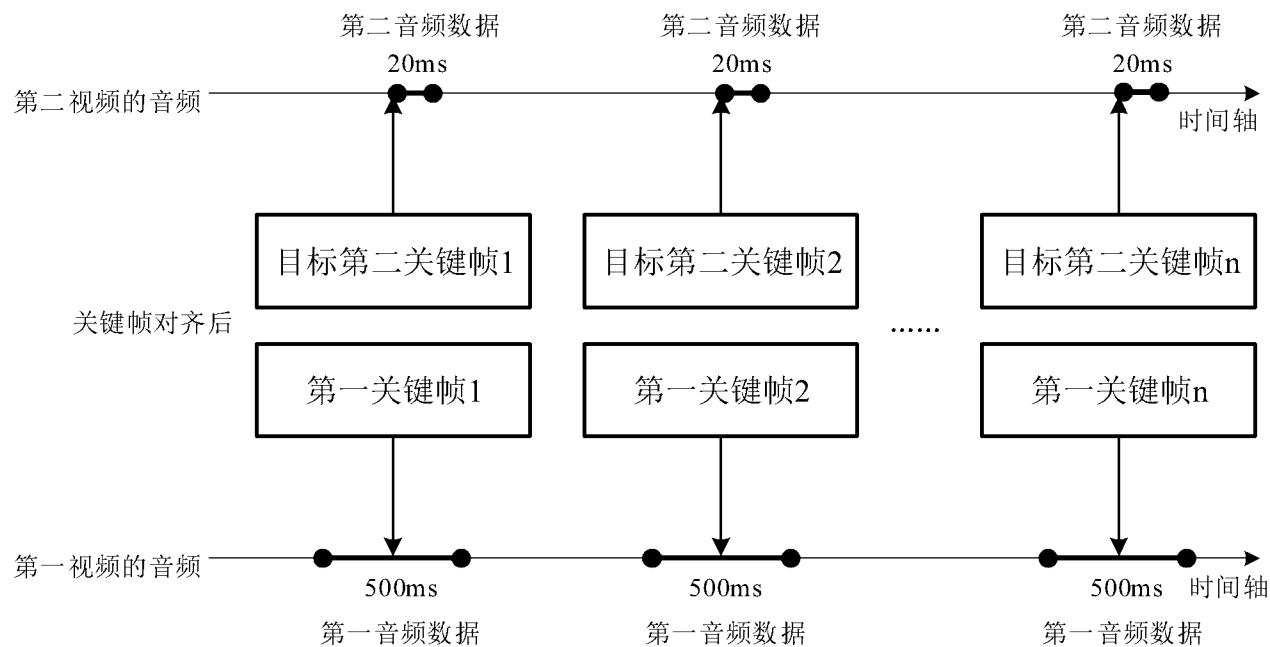


图 4

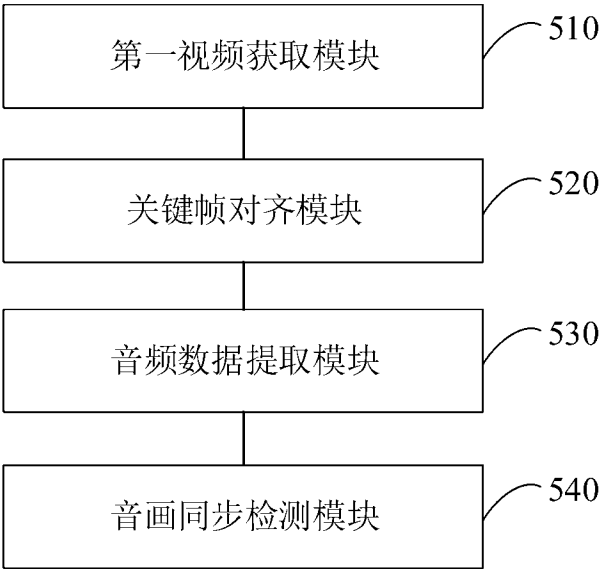


图 5

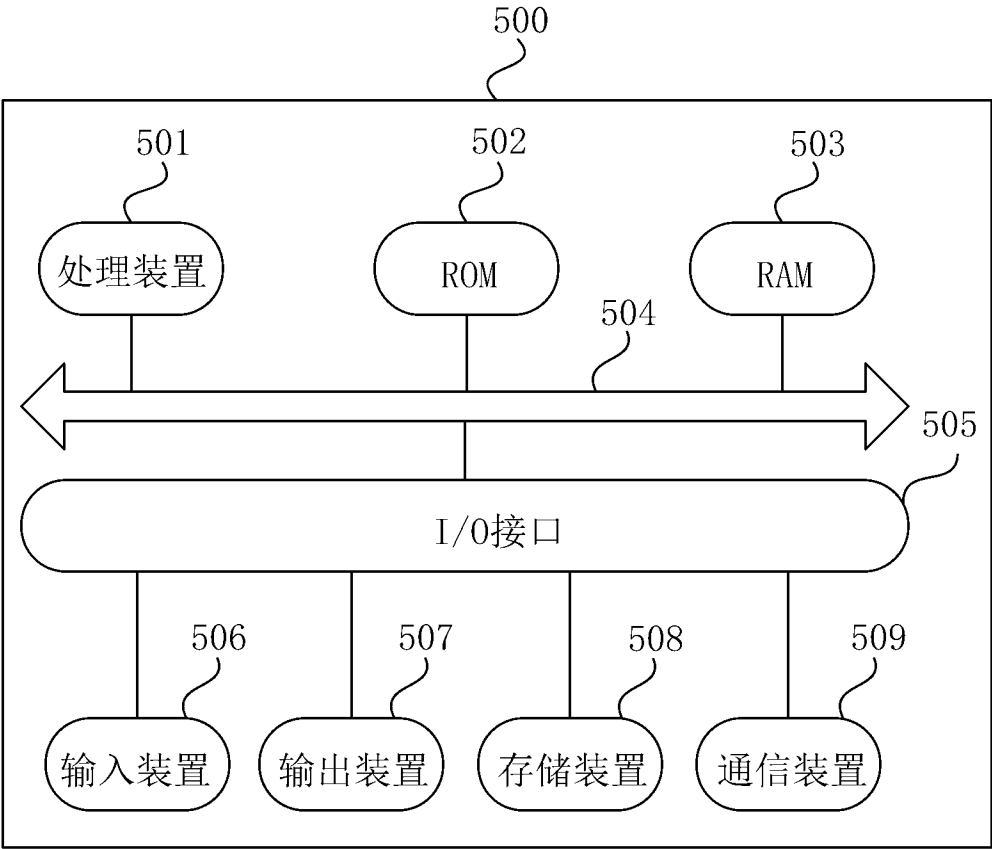


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/099889

A. CLASSIFICATION OF SUBJECT MATTER H04N 21/43(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: H04N 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; ENTXT; ENTXTC; CNKI: 同步, 音频, 视频, 原始, 参考, 转码, 时间戳, 匹配, 比对, 相似, 关键帧, 差, synchronize, audio, video, original, reference, transcode, timestamp, match, compare, similar, key frame, difference		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116708892 A (DOUYIN VISION CO., LTD.) 05 September 2023 (2023-09-05) claims 1-11	1-11
X	US 11395031 B1 (AT&T INTELLECTUAL PROPERTY I, L.P. et al.) 19 July 2022 (2022-07-19) description, paragraphs [0027]-[0089], and figures 1-5	1-11
X	CN 115776588 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 10 March 2023 (2023-03-10) description, paragraphs [0058]-[0081] and [0123], and figures 2-3B	1-11
A	CN 112597794 A (JI PENGFEI) 02 April 2021 (2021-04-02) entire document	1-11
A	CN 112434263 A (HANGZHOU ANCUN NETWORK TECHNOLOGY CO., LTD.) 02 March 2021 (2021-03-02) entire document	1-11
A	CN 102056026 A (CHINA MOBILE COMMUNICATIONS GROUP DESIGN INSTITUTE CO., LTD.) 11 May 2011 (2011-05-11) entire document	1-11
☒ 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 12 August 2024		Date of mailing of the international search report 28 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.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 115412733 A (SHANGHAI BILIBILI TECHNOLOGY CO., LTD.) 29 November 2022 (2022-11-29) entire document	1-11
A	US 2014153909 A1 (BROADCOM CORP.) 05 June 2014 (2014-06-05) entire document	1-11
A	US 2018077445 A1 (FACEBOOK, INC.) 15 March 2018 (2018-03-15) entire document	1-11
A	WO 2022144543 A1 (DARABASE LTD) 07 July 2022 (2022-07-07) entire document	1-11

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2024/099889

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116708892	A	05 September 2023	None			
US	11395031	B1	19 July 2022	US	2022360843	A1	10 November 2022
CN	115776588	A	10 March 2023	WO	2023036275	A1	16 March 2023
CN	112597794	A	02 April 2021	None			
CN	112434263	A	02 March 2021	None			
CN	102056026	A	11 May 2011	CN	102056026	B	03 April 2013
CN	115412733	A	29 November 2022	None			
US	2014153909	A1	05 June 2014	US	9635334	B2	25 April 2017
US	2018077445	A1	15 March 2018	US	10237608	B2	19 March 2019
WO	2022144543	A1	07 July 2022	US	2024129580	A1	18 April 2024
				CA	3203735	A1	07 July 2022
				EP	4272450	A1	08 November 2023

A. 主题的分类

H04N 21/43(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: H04N

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; ENTXT; ENTXTC; CNKI: 同步, 音频, 视频, 原始, 参考, 转码, 时间戳, 匹配, 比对, 相似, 关键帧, 差, synchronize, audio, video, original, reference, transcode, timestamp, match, compare, similar, key frame, difference

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 116708892 A (抖音视界有限公司) 2023年9月5日 (2023 - 09 - 05) 权利要求1-11	1-11
X	US 11395031 B1 (AT & T I P I L P E T A L.) 2022年7月19日 (2022 - 07 - 19) 说明书第[0027]-[0089]段, 附图1-5	1-11
X	CN 115776588 A (北京字跳网络技术有限公司) 2023年3月10日 (2023 - 03 - 10) 说明书第[0058]-[0081]、[0123]段, 附图2-3B	1-11
A	CN 112597794 A (季鹏飞) 2021年4月2日 (2021 - 04 - 02) 全文	1-11
A	CN 112434263 A (杭州安存网络科技有限公司) 2021年3月2日 (2021 - 03 - 02) 全文	1-11
A	CN 102056026 A (中国移动通信集团设计院有限公司) 2011年5月11日 (2011 - 05 - 11) 全文	1-11
A	CN 115412733 A (上海哔哩哔哩科技有限公司) 2022年11月29日 (2022 - 11 - 29) 全文	1-11

☒ 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8月12日

国际检索报告邮寄日期

2024年8月28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

胡西

电话号码 (+86) 027-59371621

PCT/ISA/210 表(第2页) (2022年7月)

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
A	US 2014153909 A1 (BROADCOM CORP) 2014年6月5日 (2014 - 06 - 05) 全文	1-11
A	US 2018077445 A1 (FACEBOOK INC) 2018年3月15日 (2018 - 03 - 15) 全文	1-11
A	WO 2022144543 A1 (DARABASE LTD) 2022年7月7日 (2022 - 07 - 07) 全文	1-11

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/099889

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	116708892	A	2023年9月5日	无	
US	11395031	B1	2022年7月19日	US	2022360843 A1 2022年11月10日
CN	115776588	A	2023年3月10日	WO	2023036275 A1 2023年3月16日
CN	112597794	A	2021年4月2日	无	
CN	112434263	A	2021年3月2日	无	
CN	102056026	A	2011年5月11日	CN	102056026 B 2013年4月3日
CN	115412733	A	2022年11月29日	无	
US	2014153909	A1	2014年6月5日	US	9635334 B2 2017年4月25日
US	2018077445	A1	2018年3月15日	US	10237608 B2 2019年3月19日
WO	2022144543	A1	2022年7月7日	US	2024129580 A1 2024年4月18日
				CA	3203735 A1 2022年7月7日
				EP	4272450 A1 2023年11月8日