

(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
2022 年 8 月 11 日 (11.08.2022)



(10) 国际公布号
WO 2022/166258 A1

- (51) 国际专利分类号:
H04N 21/845 (2011.01) **H04N 13/359** (2018.01)
H04N 21/44 (2011.01)
- (21) 国际申请号: PCT/CN2021/124610
- (22) 国际申请日: 2021 年 10 月 19 日 (19.10.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110160081.1 2021年2月5日 (05.02.2021) CN
- (71) 申请人: 深圳市优必选科技股份有限公司
(**UBTECH ROBOTICS CORP LTD**) [CN/CN]; 中国
广东省深圳市南山区学苑大道1001号南山智园
C1栋16、22楼, Guangdong 518000 (CN)。
- (72) 发明人: 林灿然(**LIN, Canran**); 中国广东省深圳
市南山区学苑大道1001号南山智园C1栋16、
22楼, Guangdong 518000 (CN)。 程骏(**CHENG,**

Jun); 中国广东省深圳市南山区学苑大道1001
号南山智园C1栋16、22楼, Guangdong 518000
(CN)。 郭渺辰(**GUO, Miao Chen**); 中国广东省深
圳市南山区学苑大道1001号南山智园C1栋16、22
楼, Guangdong 518000 (CN)。 邵池(**SHAO, Chi**);
中国广东省深圳市南山区学苑大道1001号南
山智园C1栋16、22楼, Guangdong 518000 (CN)。
张惊涛(**ZHANG, Jingtao**); 中国广东省深圳市
南山区学苑大道1001号南山智园C1栋16、
22楼, Guangdong 518000 (CN)。 钱程浩(**QIAN,**
Chenghao); 中国广东省深圳市南山区学苑大
道1001号南山智园C1栋16、22楼, Guangdong
518000 (CN)。 庞建新(**PANG, Jianxin**); 中国广
东省深圳市南山区学苑大道1001号南山智园C1
栋16、22楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一联合知识产权代
理有限公司 (**SHENZHEN ZHONGYI UNION**
INTELLECTUAL PROPERTY AGENCY CO.,LTD.);

(54) **Title:** BEHAVIOR RECOGNITION METHOD AND APPARATUS, TERMINAL DEVICE, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 行为识别方法、装置、终端设备及计算机可读存储介质

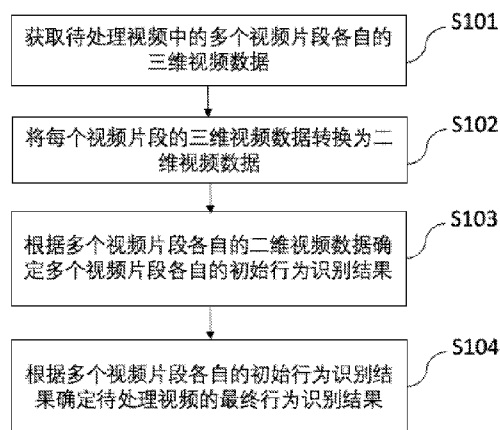


图 1

- S101 Acquire respective three-dimensional video data of a plurality of video clips in a video to be processed
- S102 Convert the three-dimensional video data of each video clip into two-dimensional video data
- S103 Determine respective initial behavior recognition results of the plurality of video clips according to the respective two-dimensional video data of the plurality of video clips
- S104 According to the respective initial behavior recognition results of the plurality of video clips, determine a final behavior recognition result of the video to be processed

(57) **Abstract:** The present application is applicable to the technical field of image processing, and provides a behavior recognition method and apparatus, a terminal device, and a computer-readable storage medium. The method comprises: acquiring respective three-dimensional video data of a plurality of video clips in a video to be processed; converting the three-dimensional video data of each video clip into two-dimensional video data; determining respective initial behavior recognition results of the plurality of video clips according to the respective two-dimensional video data of the plurality of video clips; and according to the respective initial behavior recognition results of the plurality of video clips, determining a final behavior recognition result of the video to be processed. By means of the method, the data processing amount in a video behavior recognition task can be effectively reduced, and the accuracy rate of a behavior recognition result is also improved.

(57) **摘要:** 本申请适用于图像处理技术领域, 提供了一种行为识别方法、装置、终端设备及计算机可读存储介质, 包括: 获取待处理视频中的多个视频片段各自的三维视频数据; 将每个所述视频片段的所述三维视频数据转换为二维视频数据; 根据所述多个视频片段各自的所述二维视频数据确定所述多个视频片段各自的初始行为识别结果; 根据所述多个视频片段各自的所述初始行为识别结果确定所述待处理视频的最终行为识别结果。通过上述方法, 能够有效减少视频行为识别任务中的数据处理量, 同时提高行为识别结果的准确率。

中国广东省深圳市福田区莲花街道紫荆社区深南大道6008号深圳特区报业大厦33层, Guangdong 518034 (CN)。

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

行为识别方法、装置、终端设备及计算机可读存储介质

本申请要求于 2021 年 02 月 05 日在中国专利局提交的、申请号为 202110160081.1 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

5 本申请属于图像处理技术领域，尤其涉及一种行为识别方法、装置、终端设备及计算机可读存储介质。

背景技术

10 基于视频的行为识别技术是指，通过对视频数据的分析，识别出视频中的行为类型的技术。由于视频是由多帧图像按照时序组合而成的，相比于图像的二维数据，视频增加了时序维度上的数据。因此，基于视频的行为识别技术需要对视频的三维数据进行分析处理。

随着深度学习技术的发展，该项技术逐渐渗透到基于视频的行为识别技术中。现有技术中，在图像识别任务中 2D 卷积处理的基础上，扩展到 3D 卷积处理，即利用 3D 卷积处理方法处理视频的三维数据。但是这种方法数据处理量较大，网络训练时难以收敛，无法保证识别结果的准确率。

技术问题

15 本申请实施例提供了一种行为识别方法、装置、终端设备及计算机可读存储介质，可以减少视频行为识别的数据处理量、提高识别准确率。

技术方案

20 第一方面，本申请实施例提供了一种行为识别方法，包括：
获取待处理视频中的多个视频片段各自的三维视频数据；
将每个所述视频片段的所述三维视频数据转换为二维视频数据；
根据所述多个视频片段各自的所述二维视频数据确定所述多个视频片段各自的初始行为识别结果；

25 根据所述多个视频片段各自的所述初始行为识别结果确定所述待处理视频的最终行为识别结果。

在本申请实施例中，将待处理视频的三维视频数据转换为二维视频数据，相当于将视频的三维数据处理任务转换为了二维数据处理任务，大大减少了数据处理量；另外，由于三维视频数据中包含了时序特征，因此，上述方法中不仅能够提取出视频的图像特征信息，还能够提取出视频中图像之间的时序特征信息，根据图像特征信息和时序特征信息综合识别视频中的行为类别，有效提高了识别结果的准确度。

30 在第一方面的一种可能的实现方式中，所述获取待处理视频中的多个视频片段各自的三维视频数据，包括：

对所述待处理视频进行视频抽帧处理，获得所述多个视频片段，其中，所述视频片段中包括多帧图像；

35 对于每个所述视频片段，根据所述视频片段中包含的各帧图像上的像素生成所述视频片段的所述三维视频数据，其中，所述三维视频数据的尺寸为 $H \times W \times T$ ，所述 H 为所述视频片段中各帧图像在宽度方向上包含的像素个数，所述 W 为所述视频片段中各帧图像在长度方向上包含的像素个数，所述 T 为所述视频片段中包含的图像的帧数。

40 在第一方面的一种可能的实现方式中，所述将每个所述视频片段的所述三维视频数据转换为二维视频数据，包括：

对于每个所述视频片段，将所述视频片段的所述三维视频数据中的每两维数据组合成一组所述二维视频数据，获得所述视频片段的三组所述二维视频数据。

在第一方面的一种可能的实现方式中，所述根据所述多个视频片段各自的所述二维视频数据确定所述多个视频片段各自的初始行为识别结果，包括：

对于每个所述视频片段，提取所述视频片段的三组所述二维视频数据各自的初始特征信息；

5 将所述视频片段的三组所述二维视频数据各自的所述初始特征信息融合为所述视频片段的融合特征信息；

根据所述视频片段的所述融合特征信息确定所述视频片段的初始行为识别结果。

在第一方面的一种可能的实现方式中，所述将所述视频片段的三组所述二维视频数据各自的所述初始特征信息融合为所述视频片段的融合特征信息，包括：

10 将所述视频片段的三组所述二维视频数据各自的所述初始特征信息拼接为特征拼接向量；

对所述特征拼接向量进行平均池化处理，得到池化特征信息；

根据预设的行为类别数量转换所述池化特征向量的尺寸，获得所述融合特征信息。

15 在第一方面的一种可能的实现方式中，所述视频片段的所述初始行为结果包括所述视频片段所属的行为类型；

所述根据所述多个视频片段各自的所述初始行为识别结果确定所述待处理视频的最终行为识别结果，包括：

根据所述多个视频片段各自的所述初始行为结果统计属于每种所述行为类型的所述视频片段的片段个数；

20 根据所述行为类型对应的所述片段个数，确定所述多个视频片段各自对应的所述行为类型中的目标类型；

将所述目标类型确定为所述待处理视频的所述最终行为识别结果。

第二方面，本申请实施例提供了一种行为识别装置，包括：

数据获取单元，用于获取待处理视频中的多个视频片段各自的三维视频数据；

25 数据转换单元，用于将每个所述视频片段的所述三维视频数据转换为二维视频数据；

片段识别单元，用于根据所述多个视频片段各自的所述二维视频数据确定所述多个视频片段各自的初始行为识别结果；

识别结果单元，用于根据所述多个视频片段各自的所述初始行为识别结果确定所述待处理视频的最终行为识别结果。

30 第三方面，本申请实施例提供了一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现如上述第一方面中任一项所述的行为识别方法。

第四方面，本申请实施例提供了一种计算机可读存储介质，本申请实施例提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如上述第一方面中任一项所述的行为识别方法。

35 第五方面，本申请实施例提供了一种计算机程序产品，当计算机程序产品在终端设备上运行时，使得终端设备执行上述第一方面中任一项所述的行为识别方法。

可以理解的是，上述第二方面至第五方面的有益效果可以参见上述第一方面中的相关描述，在此不再赘述。

40 附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例或现有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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

45 图1是本申请实施例提供的行为识别方法的流程示意图；

图 2 是本申请实施例提供的行为识别的数据处理流程示意图；

图 3 是本申请实施例提供的的数据转换过程的示意图；

图 4 是本申请实施例提供的行为识别装置的结构框图；

图 5 是本申请实施例提供的终端设备的结构示意图。

5 本发明的实施方式

以下描述中，为了说明而不是为了限定，提出了诸如特定系统结构、技术之类的具体细节，以便透彻理解本申请实施例。然而，本领域的技术人员应当清楚，在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中，省略对众所周知的系统、装置、电路以及方法的详细说明，以免不必要的细节妨碍本申请的描述。

应当理解，当在本申请说明书和所附权利要求书中使用时，术语“包括”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

在本申请说明书中描述的参考“一个实施例”或“一些实施例”等意味着在本申请的一个或多个实施例中包括结合该实施例描述的特定特征、结构或特点。由此，在本说明书中的不同之处出现的语句“在一个实施例中”、“在一些实施例中”、“在其他一些实施例中”、“在另外一些实施例中”等不是必然都参考相同的实施例，而是意味着“一个或多个但不是所有的实施例”，除非是以其他方式另外特别强调。

参见图 1，是本申请实施例提供的行为识别方法的流程示意图。作为示例而非限定，所述方法可以包括以下步骤：

S101，获取待处理视频中的多个视频片段各自的三维视频数据。

一幅图像中的像素构成了二维数据，二维数据的尺寸为 $H \times W$ ， H 为图像在宽度方向上包含的像素个数， W 为图像在长度方向上包含的像素个数。例如，一幅图像中长度方向上包含的像素个数为 10，宽度方向上包含的像素个数为 5，该图像的二维数据的尺寸为 5×10 ，即该图像中包含 50 个像素，这 50 个像素组成了 5×10 的二维数据，这一组二维数据中包含了该图像的特征信息。

视频是由多帧图像按照时序排列而成，相比于图像，视频多了时序维度的特征信息。即在图像的二维数据的基础上，增加了一维时序数据。因此，视频可以由三维数据来描述。

可选的，可以根据待处理视频中所有帧的图像生成待处理视频的三维视频数据。

但实际应用中，一段视频中相邻的几帧图像中通常包含相同或相似的内容。如果将待处理视频中所有帧图像的像素均包含在三维视频数据中，将会造成大量的数据冗余，后续的数据处理量也会较大。

为了降低待处理视频本身的数据冗余，可选的，可以对待处理视频进行抽样，将抽样得到的图像的像素构成待处理视频的三维视频数据。例如：待处理视频中有 100 帧图像，以 4 帧图像为间隔进行抽样，抽出 20 帧图像，将这 20 帧图像的像素构成待处理视频的三维视频数据。

如果抽样频率较高，得到的图像数量较多，数据处理量仍然较大；如果抽样频率较低，得到的图像数据较少，数据处理量较少，但是也会丢失较多的图像信息。

为了在减少数据处理量的同时，保留较多的图像信息，在本申请实施例中，获取每个视频片段的三维视频数据的方式包括：

对待处理视频进行视频抽帧处理，获得多个视频片段，每个视频片段中包括多帧图像；对于每个视频片段，根据视频片段中包含的各帧图像上的像素生成视频片段的三维视频数据。

其中，三维视频数据的尺寸为 $H \times W \times T$ ， H 为视频片段中各帧图像在宽度方向上包含的像素个数， W 为视频片段中各帧图像在长度方向上包含的像素个数， T 为视频片段中包含的图像的帧数。

视频抽帧处理，可以是每隔 n 帧图像抽取一帧图像，然后将抽取得到的图像按照时序划分为多个图像组，每个图像组为一个视频片段；还可以是每隔 n 帧图像抽取 m 帧图像，将这 m 帧图像确定为一个视频片段。

通过上述方法，不仅通过视频抽帧处理减少了待处理视频本身的数据冗余，还将抽取的图像划分为视频片段，以保留了相邻图像之间的时序特征信息和图像相关信息，为后续

参见图 2，是本申请实施例提供的行为识别的数据处理流程示意图。如图 2 所示，将待处理视频作为输入视频，待处理视频的尺寸为 $3 \times H \times W \times L$ 。其中， L 为待处理视频中包含的图像的总帧数；3 表示 RGB 三个颜色通道，在像素值中可以体现三个颜色通道的信息，因此，可以忽略该尺寸 3，即待处理视频的尺寸为 $H \times W \times L$ 。对输入视频进行抽帧组合（即视频抽帧处理，如图 2 中的以 T 帧为一组），获得多个视频片段，每个视频片段的尺寸为 $H \times W \times T$ 。

S102，将每个视频片段的三维视频数据转换为二维视频数据。

可选的，三维视频数据转换为二维视频数据的一种实现方式为：将三维视频数据中某一维度上的数据添加到另外两个维度上，以构成二维视频数据。

例如，参见图 3，是本申请实施例提供的数据转换过程的示意图。

如图 3 的 (a) 所示，一个视频片段中包含 4 帧图像 I、II、III、IV，4 帧图像按照时序组合。现将 4 张图像拼接为一幅大的拼接图像 V，这张大的拼接图像 v 上的像素构成了视频片段的二维视频数据。

从上述事例中可以看出，上述方式虽然能够保留图像信息，但是无法保留图像之间的时序信息。

为了同时保留图像信息和时序信息，在本申请实施例中，三维视频数据转换为二维视频数据的一种实现方式为：

对于每个视频片段，将视频片段的三维视频数据中的每二维数据组合成一组二维视频数据，获得视频片段的三组二维视频数据。

具体的，如图 2 所示，将三维视频数据 $H \times W \times T$ 拆分为 $H \times W$ 、 $H \times T$ 和 $W \times T$ 这三组二维视频数据。

示例性的，如图 3 中的 (b) 所示，一个视频片段中包含 4 帧图像 I、II、III、IV，假设每帧图像的尺寸为 2×3 （即 $H=2$ ， $W=3$ ， $T=4$ ，每帧图像中包含 6 个像素），该视频片段的三维视频数据的尺寸为 $2 \times 3 \times 4$ （如图 3 中的 (c) 所示，该三维视频数据可以看作是一个大的长方体，每帧图像中的各个像素可以看作是大的长方体的体素（即小的正方体），如图中标记为 1 的正方体表示视频片段的第 1 帧图像中的像素，标记为 2 的正方体表示视频片段的第 2 帧图像中的像素，以此类推，标记为 4 的正方体表示视频片段的第 4 帧图像中的像素）。

如图 3 中的 (c) 所示，将三维视频数据拆分为一组 2×3 的二维视频数据，该组中包括长方体中 abcd 横截面上的像素。将三维视频数据拆分为一组 2×4 的二维视频数据，该组中包括长方体中 abef 横截面上的像素。将三维视频数据拆分为一组 3×4 的二维视频数据，该组中包括长方体中 bcge 横截面上的像素。

S103，根据多个视频片段各自的二维视频数据确定多个视频片段各自的初始行为识别结果。

对于每个视频片段，可以将该视频片段的三组二维视频数据输入到训练后到识别模型中，输出该视频片段的初始行为识别结果。

但是上述方法中，识别模型的输入数据为三组二维数据，输入数据较多，在训练识别模型时，数据处理量较大。

为了解决上述问题，在一个实施例中，S103 的实现方式包括：

对于每个视频片段，提取视频片段的三组二维视频数据各自的初始特征信息；将视频

片段的三组二维视频数据各自的初始特征信息融合为视频片段的融合特征信息；根据视频片段的融合特征信息确定视频片段的初始行为识别结果。

示例性的，如图 2 所示，对三组二维视频数据分别进行二维卷积处理，获得每组二维视频数据的初始特征信息。例如：对二维数据 $H \times W$ 进行 $3 \times 3 \times 1$ 的卷积处理和 $3 \times 3 \times 1$ 的池化处理，获得 1×1 的初始特征信息。由于该二维数据的 T 维度为 1，因此，实际上相当于对二维数据 $H \times W$ 进行 3×3 的二维卷积处理和二维池化处理。同理，对于 $H \times T$ 的二维视频数据，使用 $3 \times 1 \times 3$ 的卷积处理和 $3 \times 1 \times 3$ 的池化处理；对于 $W \times T$ 的二维视频数据，使用 $1 \times 3 \times 3$ 的卷积处理和 $1 \times 3 \times 3$ 的池化处理。

需要说明的是，上述只是初始特征信息的获取方式的示例。实际应用中，对每组二维视频数据均可以进行多次卷积处理和池化处理，在此不做具体限定。

可选的，将三组二维视频数据各自的初始特征信息融合为融合特征信息的过程包括：

将视频片段的三组二维视频数据各自的初始特征信息拼接为特征拼接向量；对特征拼接向量进行平均池化处理，得到池化特征信息；根据预设的行为类别数量转换池化特征向量的尺寸，获得融合特征信息。

如图 2 所示，三组二维视频数据各自的初始特征信息为 1×1 的数值，通过 Concat 操作，将初始特征信息拼接在一起，得到尺寸为 $C_{out} \times 1 \times 1 \times 3$ 的特征拼接向量，其中， $C_{out} = 3$ ，即 RGB 三个颜色通道。然后在特征信息拼接的维度上，使用平均池化处理，得到尺寸为 $C_{out} \times 1 \times 1 \times 1$ 的池化特征信息。通过一个全连接层，将 C_{out} 维度改变为 k （其中 k 表示预设的行为类别数量，即所需识别的行为的类别数），得到尺寸为 $k \times 1 \times 1 \times 1$ 的融合特征信息。最后通过 softmax 层计算出该融合特征信息属于各种行为类别的概率值，并将最大概率值对应的行为类别确定为视频片段的初始识别结果。

S104，根据多个视频片段各自的初始行为识别结果确定待处理视频的最终行为识别结果。

其中，视频片段的初始行为结果包括视频片段所属的行为类型。

可选的，可以采用投票的方式，根据初始行为识别结果确定待处理视频的最终行为识别结果。具体的：

根据多个视频片段各自的初始行为结果统计属于每种行为类型的视频片段的片段个数；根据行为类型对应的片段个数，确定多个视频片段各自对应的行为类型中的目标类型；将目标类型确定为待处理视频的最终行为识别结果。

示例性的，待处理视频中包含 3 个视频片段，第一个视频片段所属的行为类别为 A，第二个视频片段所属的行为类别为 B，第三个视频片段所属的行为类别为 A。属于行为类别 A 的视频片段的片段个数为 2，属于行为类别 B 的视频片段段个数为 1，因此，目标类型为 A，即待处理视频的最终行为结果为 A。

在本申请实施例中，将待处理视频的三维视频数据转换为二维视频数据，相当于将视频的三维数据处理任务转换为了二维数据处理任务，大大减少了数据处理量；另外，由于三维视频数据中包含了时序特征，因此，上述方法中不仅能够提取出视频的图像特征信息，还能够提取出视频中图像之间的时序特征信息，根据图像特征信息和时序特征信息综合识别视频中的行为类别，有效提高了识别结果的准确度。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不对本申请实施例的实施过程构成任何限定。

对应于上文实施例所述的行为识别方法，图 4 是本申请实施例提供的行为识别装置的结构框图，为了便于说明，仅示出了与本申请实施例相关的部分。

参照图 4，该装置包括：

数据获取单元 41，用于获取待处理视频中的多个视频片段各自的三维视频数据。

数据转换单元 42，用于将每个所述视频片段的所述三维视频数据转换为二维视频数据。

片段识别单元 43，用于根据所述多个视频片段各自的所述二维视频数据确定所述多个

视频片段各自的初始行为识别结果。

识别结果单元 44，用于根据所述多个视频片段各自的所述初始行为识别结果确定所述待处理视频的最终行为识别结果。

可选的，数据获取单元 41 还用于：

- 5 对所述待处理视频进行视频抽帧处理，获得所述多个视频片段，其中，所述视频片段中包括多帧图像；对于每个所述视频片段，根据所述视频片段中包含的各帧图像上的像素生成所述视频片段的所述三维视频数据，其中，所述三维视频数据的尺寸为 $H \times W \times T$ ，所述 H 为所述视频片段中各帧图像在宽度方向上包含的像素个数，所述 W 为所述视频片段中各帧图像在长度方向上包含的像素个数，所述 T 为所述视频片段中包含的图像的帧数。

- 10 可选的，数据转换单元 42 还用于：

对于每个所述视频片段，将所述视频片段的所述三维视频数据中的每二维数据组合成一组所述二维视频数据，获得所述视频片段的三组所述二维视频数据。

可选的，片段识别单元 43 还用于：

- 15 对于每个所述视频片段，提取所述视频片段的三组所述二维视频数据各自的初始特征信息；将所述视频片段的三组所述二维视频数据各自的所述初始特征信息融合为所述视频片段的融合特征信息；根据所述视频片段的所述融合特征信息确定所述视频片段的初始行为识别结果。

可选的，片段识别单元 43 还用于：

- 20 将所述视频片段的三组所述二维视频数据各自的所述初始特征信息拼接为特征拼接向量；对所述特征拼接向量进行平均池化处理，得到池化特征信息；根据预设的行为类别数量转换所述池化特征向量的尺寸，获得所述融合特征信息。

可选的，所述视频片段的所述初始行为结果包括所述视频片段所属的行为类型。

可选的，识别结果单元 44 还用于：

- 25 根据所述多个视频片段各自的所述初始行为结果统计属于每种所述行为类型的所述视频片段的片段个数；根据所述行为类型对应的所述片段个数，确定所述多个视频片段各自对应的所述行为类型中的目标类型；将所述目标类型确定为所述待处理视频的所述最终行为识别结果。

- 30 需要说明的是，上述装置/单元之间的信息交互、执行过程等内容，由于与本申请方法实施例基于同一构思，其具体功能及带来的技术效果，具体可参见方法实施例部分，此处不再赘述。

另外，图 4 所示的行为识别装置可以是内置于现有的终端设备内的软件单元、硬件单元、或软硬结合的单元，也可以作为独立的挂件集成到所述终端设备中，还可以作为独立的终端设备存在。

- 35 所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中，也可以是各个单元单独物理存在，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，也可以采用软件功能单元的形式实现。另外，各功能单元、模块的具体名称也只是为了便于相互区分，并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

- 45 图 5 是本申请实施例提供的终端设备的结构示意图。如图 5 所示，该实施例的终端设备 5 包括：至少一个处理器 50（图 5 中仅示出一个）处理器、存储器 51 以及存储在所述存储器 51 中并可在所述至少一个处理器 50 上运行的计算机程序 52，所述处理器 50 执行所述计算机程序 52 时实现上述任意各个行为识别方法实施例中的步骤。

所述终端设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。该终端设备可包括，但不仅限于，处理器、存储器。本领域技术人员可以理解，图 5 仅仅是终端设备 5 的举例，并不构成对终端设备 5 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如还可以包括输入输出设备、网络接入设备等。

5 所称处理器 50 可以是中央处理单元(Central Processing Unit, CPU)，该处理器 50 还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

10 所述存储器 51 在一些实施例中可以是所述终端设备 5 的内部存储单元，例如终端设备 5 的硬盘或内存。所述存储器 51 在另一些实施例中也可以是所述终端设备 5 的外部存储设备，例如所述终端设备 5 上配备的插接式硬盘，智能存储卡 (Smart Media Card, SMC)，安全数字 (Secure Digital, SD) 卡，闪存卡 (Flash Card) 等。进一步地，所述存储器 15 51 还可以既包括所述终端设备 5 的内部存储单元也包括外部存储设备。所述存储器 51 用于存储操作系统、应用程序、引导装载程序(Boot Loader)、数据以及其他程序等，例如所述计算机程序的程序代码等。所述存储器 51 还可以用于暂时地存储已经输出或者将要输出的数据。

20 本申请实施例还提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现可实现上述各个方法实施例中的步骤。

本申请实施例提供了一种计算机程序产品，当计算机程序产品在终端设备上运行时，使得终端设备执行时实现可实现上述各个方法实施例中的步骤。

25 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读存储介质中。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一计算机可读存储介质中，该计算机程序在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机程序包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质至少可以包括：能够将计算机程序代码携带到装置/终端设备的任何实体或装置、记录介质、计算机 30 存储器、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质。例如 U 盘、移动硬盘、磁碟或者光盘等。在某些司法管辖区，根据立法和专利实践，计算机可读介质不可以是电载波信号和电信信号。

35 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述或记载的部分，可以参见其它实施例的相关描述。

40 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

45 在本申请所提供的实施例中，应该理解到，所揭露的装置/终端设备和方法，可以通过其它的方式实现。例如，以上所描述的装置/终端设备实施例仅仅是示意性的，例如，所述模块或单元的划分，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，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以通过一些接口，装置或单元的间接耦合或通讯连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者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者全部单元来实现本实施例方案的目的。

- 5 以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权利要求书

1. 一种行为识别方法，其特征在于，所述方法包括：

获取待处理视频中的多个视频片段各自的三维视频数据；

将每个所述视频片段的所述三维视频数据转换为二维视频数据；

5 根据所述多个视频片段各自的所述二维视频数据确定所述多个视频片段各自的初始行为识别结果；

根据所述多个视频片段各自的所述初始行为识别结果确定所述待处理视频的最终行为识别结果。

2. 如权利要求 1 所述的行为识别方法，其特征在于，所述获取待处理视频中的多个视频片段各自的三维视频数据，包括：

10 对所述待处理视频进行视频抽帧处理，获得所述多个视频片段，其中，所述视频片段中包括多帧图像；

对于每个所述视频片段，根据所述视频片段中包含的各帧图像上的像素生成所述视频片段的所述三维视频数据，其中，所述三维视频数据的尺寸为 $H \times W \times T$ ，所述 H 为所述视频片段中各帧图像在宽度方向上包含的像素个数，所述 W 为所述视频片段中各帧图像在长度方向上包含的像素个数，所述 T 为所述视频片段中包含的图像的帧数。

3. 如权利要求 1 所述的行为识别方法，其特征在于，所述将每个所述视频片段的所述三维视频数据转换为二维视频数据，包括：

20 对于每个所述视频片段，将所述视频片段的所述三维视频数据中的每两维数据组合成一组所述二维视频数据，获得所述视频片段的三组所述二维视频数据。

4. 如权利要求 3 所述的行为识别方法，其特征在于，所述根据所述多个视频片段各自的所述二维视频数据确定所述多个视频片段各自的初始行为识别结果，包括：

对于每个所述视频片段，提取所述视频片段的三组所述二维视频数据各自的初始特征信息；

25 将所述视频片段的三组所述二维视频数据各自的所述初始特征信息融合为所述视频片段的融合特征信息；

根据所述视频片段的所述融合特征信息确定所述视频片段的初始行为识别结果。

5. 如权利要求 4 所述的行为识别方法，其特征在于，所述将所述视频片段的三组所述二维视频数据各自的所述初始特征信息融合为所述视频片段的融合特征信息，包括：

30 将所述视频片段的三组所述二维视频数据各自的所述初始特征信息拼接为特征拼接向量；

对所述特征拼接向量进行平均池化处理，得到池化特征信息；

根据预设的行为类别数量转换所述池化特征向量的尺寸，获得所述融合特征信息。

6. 如权利要求 1 所述的行为识别方法，其特征在于，所述视频片段的所述初始行为结果包括所述视频片段所属的行为类型；

35 所述根据所述多个视频片段各自的所述初始行为识别结果确定所述待处理视频的最终行为识别结果，包括：

根据所述多个视频片段各自的所述初始行为结果统计属于每种所述行为类型的所述视频片段的片段个数；

40 根据所述行为类型对应的所述片段个数，确定所述多个视频片段各自对应的所述行为类型中的目标类型；

将所述目标类型确定为所述待处理视频的所述最终行为识别结果。

7. 一种行为识别装置，其特征在于，所述装置包括：

数据获取单元，用于获取待处理视频中的多个视频片段各自的三维视频数据；

45 数据转换单元，用于将每个所述视频片段的所述三维视频数据转换为二维视频数据；

片段识别单元，用于根据所述多个视频片段各自的所述二维视频数据确定所述多个视频片段各自的初始行为识别结果；

识别结果单元，用于根据所述多个视频片段各自的所述初始行为识别结果确定所述待处理视频的最终行为识别结果。

5 8. 如权利要求 7 所述的行为识别装置，其特征在于，所述数据获取单元还用于：

对所述待处理视频进行视频抽帧处理，获得所述多个视频片段，其中，所述视频片段中包括多帧图像；

10 对于每个所述视频片段，根据所述视频片段中包含的各帧图像上的像素生成所述视频片段的所述三维视频数据，其中，所述三维视频数据的尺寸为 $H \times W \times T$ ，所述 H 为所述视频片段中各帧图像在宽度方向上包含的像素个数，所述 W 为所述视频片段中各帧图像在长度方向上包含的像素个数，所述 T 为所述视频片段中包含的图像的帧数。

9. 一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现如权利要求 1 至 6 任一项所述的方法。

15 10. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求 1 至 6 任一项所述的方法。

附图

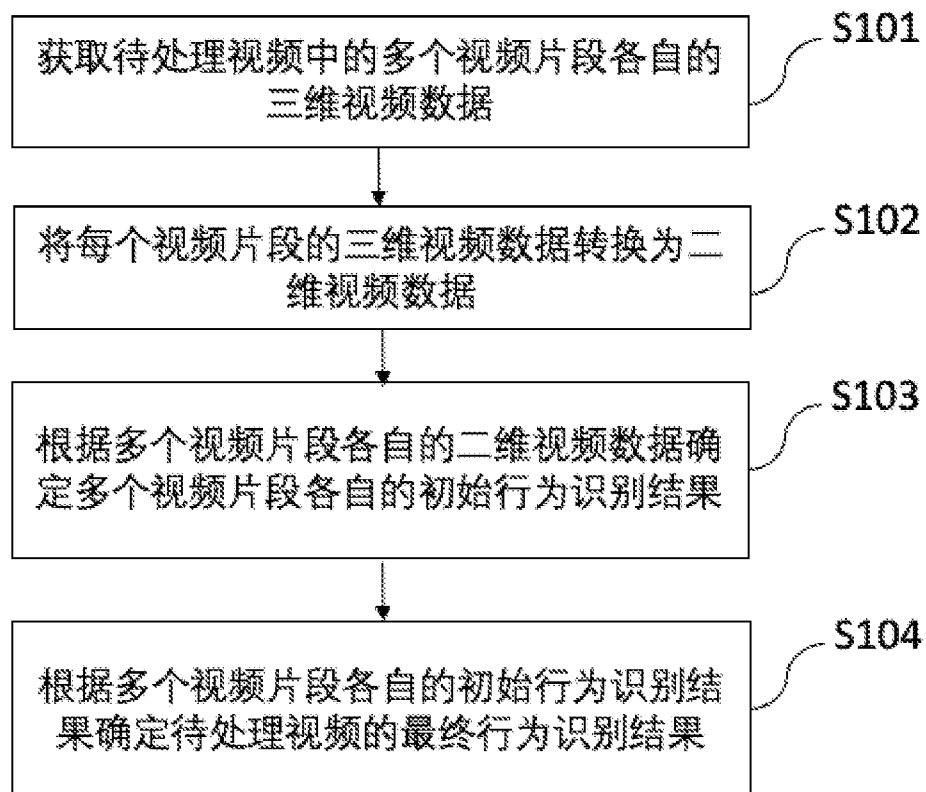


图 1

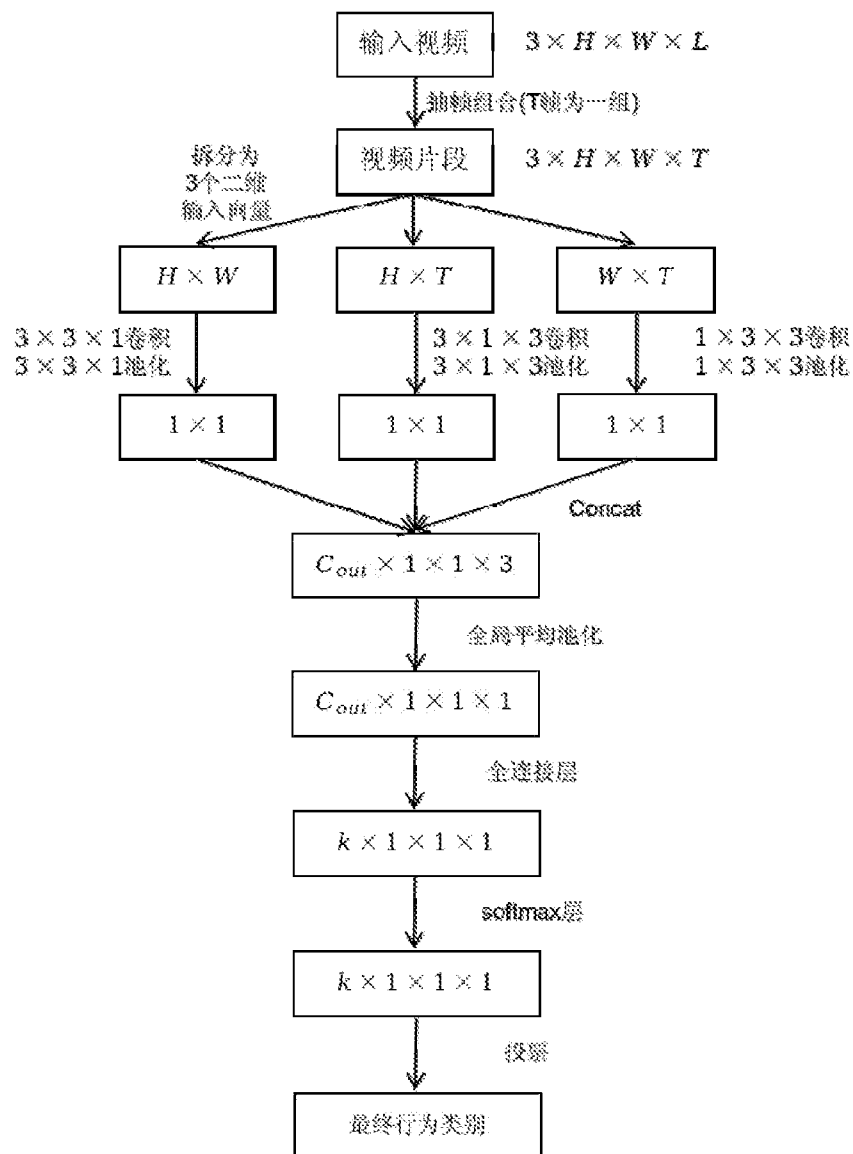


图 2

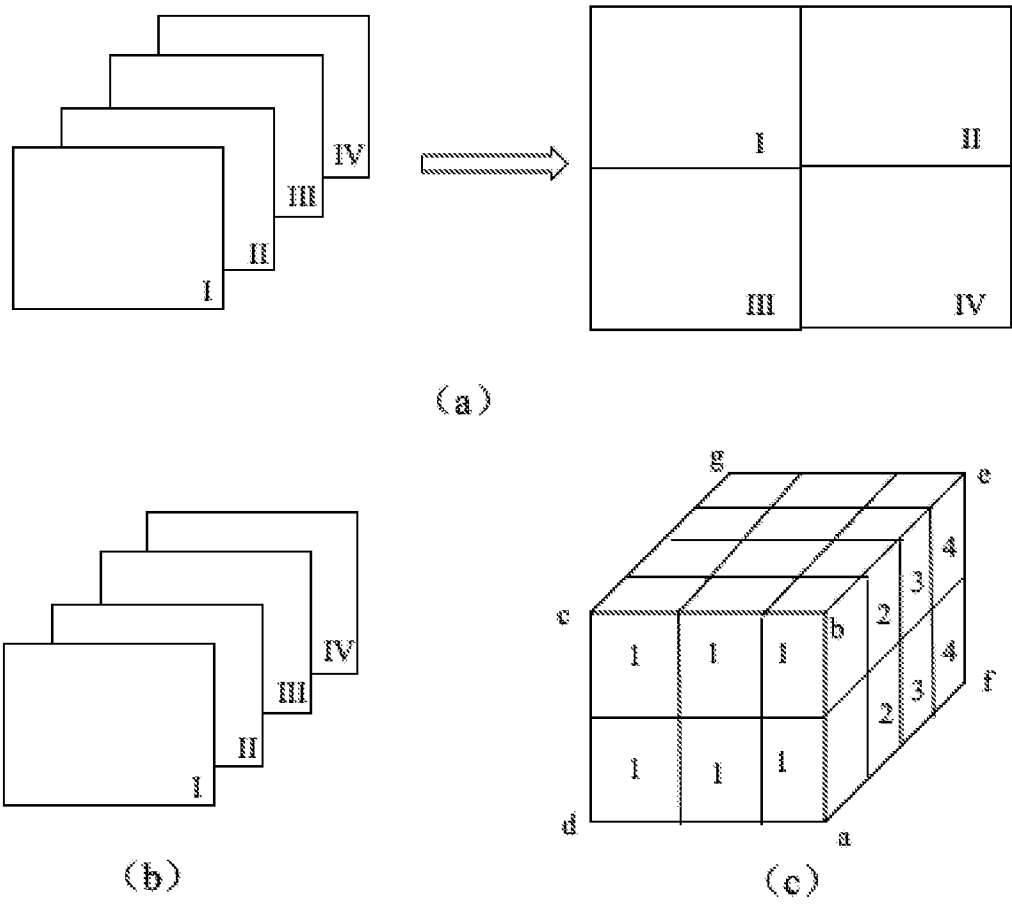


图 3

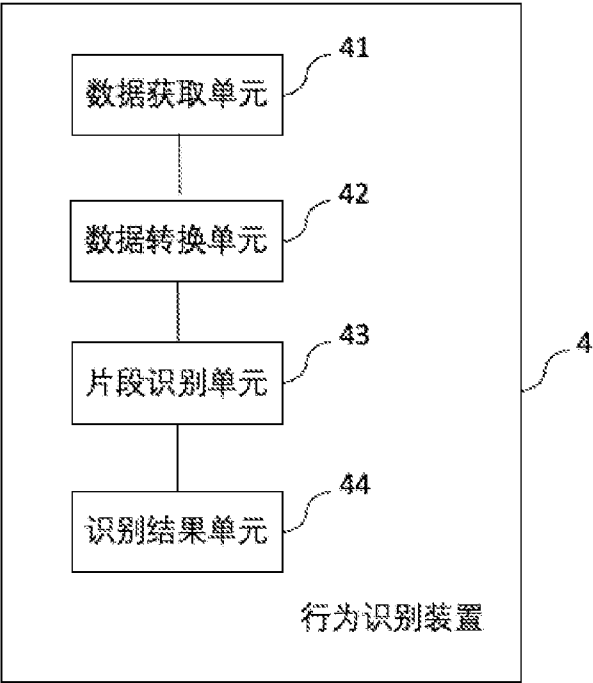


图 4

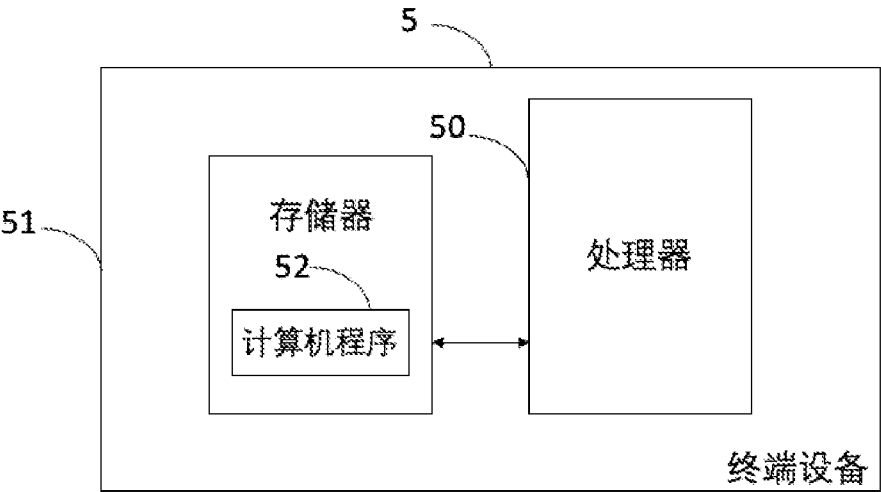


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/124610

A. CLASSIFICATION OF SUBJECT MATTER

H04N 21/845(2011.01)i; H04N 21/44(2011.01)i; H04N 13/359(2018.01)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; G06K

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)

CNKI, CNPAT, EPODOC, WPI, 3GPP: 三维, 二维, 视频, 片段, 行为, 识别, 辨别, 3D, three-dimension, 2D, two-dimension, video, behavior, recognition, identity, behavior recognition

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113015022 A (UBTECH ROBOTICS CORP.) 22 June 2021 (2021-06-22) claims 1-10, description paragraphs [0048]-[0108]	1-10
X	CN 109697434 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 30 April 2019 (2019-04-30) description paragraphs [0042]-[0183]	1-10
A	CN 109165573 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 08 January 2019 (2019-01-08) entire document	1-10
A	CN 111814922 A (SOBEY DIGITAL TECHNOLOGY CO., LTD.) 23 October 2020 (2020-10-23) entire document	1-10
A	CN 111291718 A (SHANGHAI SHANGTANG INTELLIGENT TECHNOLOGY CO., LTD.) 16 June 2020 (2020-06-16) entire document	1-10
A	US 2018084283 A1 (JAUNT INC.) 22 March 2018 (2018-03-22) entire document	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

“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

06 January 2022

Date of mailing of the international search report

18 January 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/124610

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113015022	A	22 June 2021	None			
CN	109697434	A	30 April 2019	None			
CN	109165573	A	08 January 2019	None			
CN	111814922	A	23 October 2020	None			
CN	111291718	A	16 June 2020	TW	202133036	A	01 September 2021
				WO	2021169924	A1	02 September 2021
US	2018084283	A1	22 March 2018	US	2019246096	A1	08 August 2019

国际检索报告

国际申请号

PCT/CN2021/124610

A. 主题的分类 H04N 21/845(2011.01)i; H04N 21/44(2011.01)i; H04N 13/359(2018.01)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; G06K 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)) CNKI, CNPAT, EP0DOC, WPI, 3GPP: 三维, 二维, 视频, 片段, 行为, 识别, 辨别, 3D, three-dimension, 2D, two-dimension, video, behavior, recognition, identity, behavior recognition		
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 113015022 A (深圳市优必选科技股份有限公司) 2021年6月22日 (2021 - 06 - 22) 权利要求1-10, 说明书第[0048]-[0108]段	1-10
X	CN 109697434 A (腾讯科技深圳有限公司) 2019年4月30日 (2019 - 04 - 30) 说明书第[0042]-[0183]段	1-10
A	CN 109165573 A (百度在线网络技术北京有限公司) 2019年1月8日 (2019 - 01 - 08) 全文	1-10
A	CN 111814922 A (成都索贝数码科技股份有限公司) 2020年10月23日 (2020 - 10 - 23) 全文	1-10
A	CN 111291718 A (上海商汤智能科技有限公司) 2020年6月16日 (2020 - 06 - 16) 全文	1-10
A	US 2018084283 A1 (JAUNT INC.) 2018年3月22日 (2018 - 03 - 22) 全文	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 “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 2022年1月6日		国际检索报告邮寄日期 2022年1月18日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 魏玲 电话号码 86-(10)-53961737

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/124610

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	113015022	A	2021年6月22日	无			
CN	109697434	A	2019年4月30日	无			
CN	109165573	A	2019年1月8日	无			
CN	111814922	A	2020年10月23日	无			
CN	111291718	A	2020年6月16日	TW	202133036	A	2021年9月1日
				WO	2021169924	A1	2021年9月2日
US	2018084283	A1	2018年3月22日	US	2019246096	A1	2019年8月8日