

(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 11 月 3 日 (03.11.2022)



(10) 国际公布号
WO 2022/227490 A1

(51) 国际专利分类号:

G06K 9/00 (2022.01)

(21) 国际申请号:

PCT/CN2021/130036

(22) 国际申请日:

2021 年 11 月 11 日 (11.11.2021)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202110449968.2 2021年4月25日 (25.04.2021) CN

(71) 申请人: 上海商汤智能科技有限公司 (SHANGHAI SENSETIME INTELLIGENT TECHNOLOGY CO., LTD.) [CN/CN]; 中国上海市徐汇区桂平路 391 号 3 号楼 1605A 室, Shanghai 200233 (CN)。

(72) 发明人: 苏海昇 (SU, Haisheng); 中国上海市徐汇区桂平路 391 号 3 号楼 1605A 室, Shanghai 200233 (CN)。苏婧 (SU, Jing); 中国上海市徐汇区桂平路 391 号 3 号楼 1605A 室, Shanghai 200233 (CN)。王栋梁 (WANG, Dongliang); 中国上海市徐汇区桂平路 391 号 3 号楼 1605A 室, Shanghai 200233 (CN)。

(74) 代理人: 北京派特恩知识产权代理有限公司 (CHINA PAT INTELLECTUAL PROPERTY OFFICE); 中国北京市海淀区海淀南路 21 号中关村知识产权大厦 B 座 2 层, Beijing 100080 (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))。

(54) Title: BEHAVIOR RECOGNITION METHOD AND APPARATUS, DEVICE, STORAGE MEDIUM, COMPUTER PROGRAM, AND PROGRAM PRODUCT

(54) 发明名称: 行为识别方法、装置、设备、存储介质、计算机程序及程序产品

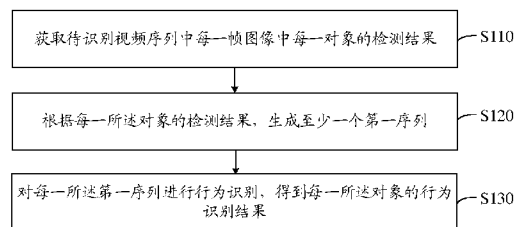


图 1B

S110 Obtain a detection result of each object in each frame of image in a video sequence to be recognized
S120 Generate at least one first sequence according to the detection result of each object
S130 Perform behavior recognition on each first sequence to obtain a behavior recognition result of each object

(57) Abstract: Disclosed in embodiments of the present disclosure are a behavior recognition method and apparatus, a device, a storage medium, a computer program, and a program product. The method comprises: obtaining a detection result of each object in each frame of image in a video sequence to be recognized; generating at least one first sequence according to the detection result of each object, each first sequence being a trajectory sequence comprising a same object; and performing behavior recognition on each first sequence to obtain a behavior recognition result of each object.

(57) 摘要: 本公开实施例公开了一种行为识别方法、装置、设备、存储介质、计算机程序及程序产品, 其中所述方法包括: 获取待识别视频序列中每一帧图像中每一对象的检测结果; 根据每一所述对象的检测结果, 生成至少一个第一序列; 其中, 每一所述第一序列为包括同一对象的轨迹序列; 对每一所述第一序列进行行为识别, 得到每一所述对象的行为识别结果。

WO 2022/227490 A1

行为识别方法、装置、设备、存储介质、计算机程序及程序产品

相关申请的交叉引用

本公开基于申请号为 202110449968.2、申请日为 2021 年 04 月 25 日、申请名称为“行为识别方法及装置、设备和存储介质”的中国专利申请提出，并要求该中国专利申请的
5 优先权，该中国专利申请的全部内容在此以全文引入的方式引入本公开。

技术领域

本公开涉及计算机视觉领域，涉及但不限于行为识别方法、装置、设备、存储介质、计算机程序及程序产品。

背景技术

10 相关技术中的行为识别方法首先对输入的视频序列进行全图的数据增强或其他预处理，然后直接送入到分类模型中进行预测。也就是说，输入整个视频序列到网络模型，让网络模型去感知行为发生的执行主体的发生位置。这种方式依赖于执行主体占比画面很大的前提假设，只适用于以执行主体为中心的视频行为识别，该类数据常见于公开的视频学术数据集中。

15 然而在大规模视频数据下，行为发生的执行主体的数量不可预知，且视频中的每一帧图像往往包含更多的信息，覆盖的视野也更大，造成识别效果较低、计算量大。

发明内容

本公开实施例提供一种行为识别方法、装置、设备、存储介质、计算机程序及程序产品。

20 本公开实施例的技术方案是这样实现的：

第一方面，本公开实施例提供一种行为识别方法，所述方法包括：获取待识别视频序列中每一帧图像中每一对象的检测结果；根据每一所述对象的检测结果，生成至少一个第一序列；其中，每一所述第一序列为包括同一对象的轨迹序列；对每一所述第一序列进行行为识别，得到每一所述对象的行为识别结果。

25 这样，首先获取每一帧图像中每一对象的检测结果，以确定单个对象在待识别视频序列中准确的轨迹序列即第一序列，然后对至少一个第一序列进行行为识别。通过利用视频分类算法与检测算法的结合，将单帧行为识别算法拓展为一个具有时序关系的行为识别算法。将视频分类问题转换成轨迹序列的识别问题，能够减小模型的识别难度，并提升有效信息的获取。同时拓展了视频理解模型的边界，发挥了视频理解模型的潜力，
30 也是对现有视频理解模型落地应用的体现。

第二方面，本公开实施例提供一种行为识别装置，包括获取模块、第一确定模块和识别模块，其中：所述获取模块，配置为获取待识别视频序列中每一帧图像中每一对象的检测结果；所述第一确定模块，配置为根据每一所述对象的检测结果，确定所述待识

别视频序列中每一第一序列；所述识别模块，配置为对每一所述第一序列进行行为识别，得到每一所述对象的行为识别结果。

第三方面，本公开实施例提供一种行为识别设备，包括存储器和处理器，所述存储器存储有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上述行为识别方法中的步骤。

第四方面，本公开实施例提供一种计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时实现上述行为识别方法中的步骤。

第五方面，本公开还提供一种计算机程序，包括计算机可读代码，当所述计算机可读代码在电子设备中运行时，所述电子设备中的处理器执行用于实现上述第一方面，或第一方面中任一种可能的实施方式中的步骤。

第六方面，本公开还提供一种计算机程序产品，所述计算机程序产品包括一条或多条指令，所述一条或多条指令适于由处理器加载并执行上述第一方面，或第一方面中任一种可能的实施方式中的步骤。

附图说明

为了更清楚地说明本公开实施例中的技术方案，下面将对实施例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的前提下，还可以根据这些附图获得其它的附图，其中：

图 1A 为本公开实施例提供的行为识别的一种网络架构示意图；

图 1B 为本公开实施例提供的一种行为识别方法的流程示意图；

图 2 为本公开实施例提供的确定待识别视频序列的流程示意图；

图 3 为本公开实施例提供的又一种行为识别方法的流程示意图；

图 4 为本公开实施例提供的生成至少一个第一序列的流程示意图；

图 5 为本公开实施例提供的再一种行为识别方法的流程示意图；

图 6A 为本公开实施例提供的单人行为识别算法的系统框图；

图 6B 为本公开实施例提供的单人行为识别算法的逻辑流程图；

图 6C 为本公开实施例提供的预处理后的输入轨迹序列样例示意图；

图 6D 为本公开实施例提供的空间运动掩膜裁剪的效果示意图；

图 7 为本公开实施例提供的一种行为识别装置的组成结构示意图；

图 8 为本公开实施例提供的一种行为识别设备的硬件实体示意图。

具体实施方式

为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合本公开实施例中的附图，对本公开实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本公开一部分实施例，而不是全部的实施例。以下实施例用于说明本公开，但不用来限制本公开的范围。基于本公开中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本公开保护的范围。

在以下的描述中，涉及到“一些实施例”，其描述了所有可能实施例的子集，但是可以理解，“一些实施例”可以是所有可能实施例的相同子集或不同子集，并且可以在不冲突的情况下相互结合。

需要指出，本公开实施例所涉及的术语“第一\第二\第三”仅仅是是区别类似的对象，不代表针对对象的特定排序，可以理解地，“第一\第二\第三”在允许的情况下可以互换特定的顺序或先后次序，以使这里描述的本公开实施例能够以除了在这里图示或描述的以外的顺序实施。

本技术领域技术人员可以理解，除非另外定义，这里使用的所有术语（包括技术术语和科学术语），具有与本公开实施例所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是，诸如通用字典中定义的那些术语，应该被理解为具有与现有技术的上下文中的意义一致的意义，并且除非像这里一样被特定定义，否则不会用理想化或过于正式的含义来解释。

视频中的行为识别是计算机视觉领域的一个重要应用，在智慧城市领域有着广泛的应用。随着现代通信能力的增强，在智慧城市的建设进程中，视频源越来越丰富。但是对这些视频源的利用仅仅停留在看或查的能力上，大部分决策能力还是需要大量的人力完成。随着云计算带来的算力提升，深度学习带来的智力的提升，可以充分利用深度学习的优势使得采集视频源的设备有一定的自动化决策能力。本公开实施例可应用于危险地带及时检测不安全事故如是否有人摔倒或者有小动物落水等行为，以便及时采取保护措施。

图 1A 为本公开实施例提供的行为识别的一种网络架构示意图，如图 1A 所示，该网络架构中包括：摄像头 101、对象检测模块 102、预处理模块 103 和视频识别模块 104：其中，对象检测模块 102、预处理模块 103 和视频识别模块 104 可以设置于服务器 100 中，为实现支撑一个示例性应用，摄像头 101 通过网络和服务器 100 建立通信连接。通过摄像头 101 采集特定场景下的视频，然后采样得到待识别视频序列 11 即包含目标对象的多帧图像，将待识别视频序列 11 输入到对象检测模块 102 中；对象检测模块 102 可以充分利用相关检测算法如帧间差分法、背景减除法、光流法等等实现对待识别视频序列 11 中目标对象的定位和分析，得到带检测结果（标注目标对象的检测框和对象标识）的多帧图像 12；然后通过预处理模块 103 对带检测结果的多帧图像 12 进行处理，基于每一帧图像中每一对象的检测框和对象标识，生成至少一个第一序列 13，该第一序列为的单个对象的轨迹序列；将至少一个第一序列 13 输入到视频识别模块 104 中；视频识别模块 104 可以充分利用相关视频理解模型对每一第一序列进行行为识别，最后输出视频层面的识别结果。基于该网络架构，可以设计一种包括对象定位和行为识别两个阶段的行为识别方法框架。其中对象定位阶段可以灵活的使用现有所有对象检测算法，行为识别阶段充分利用现有视频理解模型，拓展了视频理解模型的边界，发挥了视频理解模型的潜力，也是对现有视频理解模型落地应用的体现。

本公开实施例提供一种行为识别方法，应用于服务器、终端设备或其他设备。其中，

所述终端设备包括但不限于手机、笔记本电脑、平板电脑和掌上上网设备、多媒体设备、流媒体设备、移动互联网设备、可穿戴设备或其他类型的设备。图 1B 为本公开实施例提供的一种行为识别方法的流程示意图，如图 1B 所示，所述方法至少包括以下步骤：

步骤 S110，获取待识别视频序列中每一帧图像中每一对象的检测结果。

5 这里，待识别视频序列为对视频源进行采样得到的。视频源中的每一帧图像，往往覆盖的视野比较广阔，包含较多的信息，如行人、车辆、动物、建筑物以及其他复杂的背景信息。

可以理解的，待识别视频序列为由多帧图像构成的帧序列，其中每一帧图像中可以包含至少一个对象或者未包含对象，不同的帧图像包含的对象也不一定相同。这里的对象可以是行人，也可以是移动车辆、动物等，在实施中根据实际场景确定，本公开实
10 施例对此不作限定。

可以理解的，图像中的每一对象可以是符合一定条件的对象，在实施的过程中，对象满足一定的条件可以是：对象所在的图像区域在图像画面中所占的比例大于一定的比例阈值，或者，对象的类型属于某一预设类型等的，例如对象所在的图像区域在图像画
15 面中所占的比例大于 10%，对象的类型属于移动对象，例如动物，可自由移动的机器设备如扫地机等。

图像中的对象可以认为是潜在的行为发生的执行主体，由此，行为识别方法采用分类模型识别时，分类模型将对图像中的每一对象进行预测。

可以通过相关图像或视频处理技术实现对视频图像中对象的检测和定位分析，例如对象检测算法对该待识别视频序列进行预处理，得到多帧带检测框的图像，然后提取每
20 一帧图像中所述对象的检测框。其中，检测算法比如模板匹配，对于视频中运动目标检测，可以通过帧间差分法、背景减除法、光流法等等实现，本公开实施例对此不作限定。

这里，所述对象的检测结果至少包括对象的检测框和与所述检测框关联的对象标识。对于待识别视频序列中出现的不同对象，通过对象检测后会得到每一对象的检测结果和
25 与检测框关联的唯一对象标识。即不同对象的检测框通过相应的对象标识区分，以便后续自动识别发生行为的主体并及时处理。

步骤 S120，根据每一所述对象的检测结果，生成至少一个第一序列。

这里，所述第一序列为包括同一对象的轨迹序列，即单个行为主体在待识别视频序列所持续的时间段内的位置序列。

30 基于所述待识别视频序列中帧图像所包含对象的检测框，筛选出属于每一单个对象的检测框，然后将每一单个对象的检测框按时间顺序拼接即生成一个第一序列。

通过利用相关检测算法进行行为主体的定位，可以获得准确的行为发生区间即检测框所在的位置区间，从而进一步生成每一个第一序列。

步骤 S130，对每一所述第一序列进行行为识别，得到每一所述对象的行为识别结果。

35 这里，将上一步骤得出的一个或多个第一序列一起送入行为识别模型中，获取不同行为的判别分数。也就是说，将待识别视频序列中的多个单个对象的行为轨迹序列进行

独立行为识别，确认每一第一序列经过网络模型后的判别分数，如果存在至少一个判别分数高于某个给定的阈值，则输出该第一序列对应的单个对象的行为，例如一个老人跌倒、一个小孩躺卧、一只小狗落水等行为；如果不存在任一个判别分数高于某个给定的阈值，则确定该待识别视频序列中不存在单个对象的行为。

5 需要说明的是，步骤 S110 中可以灵活的使用相关检测算法进行行为主体的准确定位和分析，步骤 S130 中的行为识别过程中可以充分利用现有视频理解模型，结合这两个步骤构建了基于对象定位阶段和行为识别阶段两阶段的行为识别的算法框架，可以适应于更多覆盖视野大、包含更多信息的视频信息，如室外的城市街道场景以及室内的轨道交通等场景下获取的视频源。

10 在本公开实施例中，首先，获取待识别视频序列中每一帧图像中每一对象的检测结果；然后，根据每一所述对象的检测结果，生成所述待识别视频序列中至少一个第一序列；最后，对每一所述第一序列进行行为识别，得到每一所述对象的行为识别结果；如此，通过获取每一帧图像中每一对象的检测结果，以确定单个对象在待识别视频序列中准确的轨迹序列，然后对每一第一序列进行行为识别，将单帧行为识别算法拓展为一个具有时序关系的行为识别算法。从而将视频分类问题转换成轨迹序列的识别问题，能够减小模型的识别难度，并提升有效信息的获取。同时拓展了视频理解模型的边界，发挥了视频理解模型的潜力，也是对现有视频理解模型落地应用的体现。

图 2 为本公开实施例提供的确定待识别视频序列的流程示意图，如图 2 所述，所述流程至少包括以下步骤：

20 步骤 S210，按照特定时间间隔，从视频源中选取至少两帧图像。

这里，所述特定时间间隔为根据线上部署的识别模型所允许的检测周期/频率（例如 3 秒一次检测频率）和设定的采样帧数对应的采样间隔，例如采样 8 帧，则特定时间间隔为 3/8 秒。在实施中可以根据实际情况设定，本公开实施例对此不作限定。

步骤 S220，将所述至少两帧图像存入帧序列缓冲区。

25 这里，所述帧序列缓冲区的长度固定，也就是固定需要检测的视频序列的采样帧数。

步骤 S230，获取所述帧序列缓冲区中的帧图像作为所述待识别视频序列。

这里，通过特定时间间隔获取每一帧图像存入固定长度的帧序列缓冲区，并当存入的图像帧数超过帧序列缓冲区的最长长度时，将帧序列缓冲区中时间最早的第一帧图像弹出，从而形成长度固定的滑动窗。

30 本公开实施例通过间隔特定时间间隔采样视频序列中的多帧图像，构造滑窗式帧序列缓冲区，一方面是能高效处理离线长视频和在线视频流，并行处理每个窗口内的视频帧序列；另一方面，以固定大小的窗口作为基本处理单元可以保证采样的合理性，捕捉动作执行的全局过程。

35 在一些其他实施例中，所述对象的检测结果至少包括所述对象的检测框和与所述检测框关联的对象标识；图 3 为本公开实施例提供的又一种行为识别方法的流程示意图，如图 3 所述，上述步骤 S120 “根据每一所述对象的检测结果，生成至少一个第一序列”

可以通过以下步骤实现:

步骤 S310, 从每一所述对象的检测框中, 选择在空间位置上独立的检测框作为候选检测框。

5 这里, 所述在空间位置上独立的检测框可以为每一帧图像中稀疏区域的检测框, 或者与其他检测框之间无重叠部分的检测框。

通常认为独立的对象容易发生单个对象行为, 因此从每一帧图像包含的所有对象的检测框中, 筛选出每一单个对象的候选检测框。

可以通过以下方式实现“从每一所述对象的检测框中, 选择在空间位置上独立的检测框作为候选检测框”的过程:

10 方式一: 在所述每一帧图像中包括一个检测框的情况下, 将所述每一帧对象中的一个检测框作为所述候选检测框。

这里, 每一帧图像中包括一个检测框说明该检测框对应的对象为单个对象, 容易发生单个对象行为。

方式二: 在所述每一帧图像中包括至少两个检测框的情况下, 确定所述每一帧图像中的至少两个检测框之间的交并比; 根据所述交并比, 生成所述每一帧图像对应的邻接矩阵; 将所述邻接矩阵中匹配次数为零的检测框作为所述候选检测框。

20 这里, 每一帧图像中包括至少两个检测框说明该待识别视频序列中包含多个对象, 这多个对象在每一帧图像中的空间位置可以聚集在同一位置附近, 也可以相对分散的分布在不同位置, 也可以部分聚集部分分散分布。本公开实施例期望找出其中分散的单个对象的检测框进行后续处理。

每两个检测框之间的交并比 (Intersection over Union, IoU) 是两个区域重叠的部分除以两个区域的集合部分得出的结果。邻接矩阵中 (i, j) 的值代表帧图像中检测框 i 和检测框 j 的交并比。将第 i 行中数值大于 0 的个数 (除了自身) 作为检测框 i 的匹配次数。

25 需要说明的是, 如果邻接矩阵中不存在与检测框 i 之间的交并比大于 0 的其他检测框, 那么检测框 i 的匹配次数为 0, 说明该检测框 i 与所在帧图像中其他检测框之间没有重叠区域, 可以作为单个对象的候选检测框。

30 在一些其他的实施例中, 在所述每一帧图像中包括至少两个检测框的情况下, 首先对每一所述对象的检测框, 向外扩展特定比例, 得到扩展后的检测框; 从所述扩展后的检测框中筛选至少两个第一检测框; 其中, 所述第一检测框的面积大于所述扩展后的检测框中其他检测框; 然后, 确定所述至少两个第一检测框之间的交并比。例如, 在所述特定比例为 1.5 倍的情况下, 将每一所述对象的检测框的长和宽都扩展 1.5 倍。这样能够增大图像分辨率, 从而能更好地计算相互重叠的检测框之间的交并比。

步骤 S320, 根据所述候选检测框和与所述候选检测框关联的对象标识, 生成所述至少一个第一序列。

35 这里, 所述候选检测框可以在待识别视频序列中的每一帧图像中出现, 也可以只在一部分帧中出现。将所有待识别视频序列中出现的与每一对象标识关联的候选检测框按

照时间顺序连接起来，即可生成每一第一序列。

在一些其他的实施例中，所述候选检测框为从所述至少两个第一检测框中选择的，首先对所述候选检测框，按照所述特定比例进行内缩，得到原始尺寸的候选检测框；然后根据所述原始尺寸的候选检测框和与所述候选检测框关联的对象标识，生成所述至少一个第一序列。这样，对于从所述至少两个第一检测框中选择的候选检测框，在生成每一第一序列之前需要按照原比例内缩，确保最终进入行为识别阶段的轨迹序列为单个对象原始的运动轨迹，避免产生额外的计算量。

在实施中，图4为本公开实施例提供的生成至少一个第一序列的流程示意图，如图4所示，可以通过以下步骤确定每一所述第一序列：

步骤 S3201，对与每一所述对象标识关联的候选检测框在空间位置上取并集，得到与每一所述对象标识对应的最小包围框。

这里，对于所述候选检测框在不同帧图像上的空间位置可能是移动的，通过取并集合并成一个较大的区域作为最小包围框。例如，将滑窗内的视频帧序列中属于同一个 ID（Identity document，身份标识）的行人检测框在空间位置上取并集，得到这个行人在视频帧序列中的最小包围框。

步骤 S3202，根据所述最小包围框，截取所述每一帧图像中与所述最小包围框对应的区域图像。

这样，以最小包围框截取每一帧图像中的区域图像，能够避免行为主体的相对位置丢失，对于空间相似但运动节奏不同的行为识别有较好的性能提升。

值得注意的是，在截取区域图像的时候，可能存在某些帧图像中并未出现某个特定对象。这种情况下，可以用视频帧序列中其他帧图像中属于该特定对象的候选检测框进行取并集操作，得到属于该特定对象的最小包围框，并用最小包围框同时截取纯色背景的区域图像。

示例地，采样8帧图像，属于同一个ID的行人可能只出现在8帧中的前7帧，对于不出现该行人的第8帧用-1进行补全。即用该行人在前7帧中计算出的最小包围框对纯色背景进行裁剪，从而根据所有裁剪的区域图像生成该行人最终的轨迹序列，有效解决了基于单人持续检测容易出现目标丢失问题。

步骤 S3203，确定所述区域图像中除去所述候选检测框之外的噪声区域。

这里，截取后的区域图像中除去候选检测框之外的区域，对行为识别的识别结果意义不大，可以作为噪声区域。

步骤 S3204，利用纯色背景填充所述噪声区域，得到目标区域图像。

这里，由于在检测定位阶段只将单个对象的候选检测框提取出来，丢失了相对运动信息，例如对于跑步、走路这类单人行为动作，由于其动作形式和动作幅度相似，很难见其区分。本公开实施例对提取的检测框保持空间位置的同时，将周围的噪声区域用纯色背景填充，即利用空间运动掩膜（mask）保持行为主体的空间相对位置信息。

步骤 S3205，按照所述每一帧图像的时间戳，依次连接所述目标区域图像，得到与

每一所述对象标识对应的第一序列。

这里，将属于每一对象标识的目标区域图像按照所述每一帧图像的时间戳连接起来，即得到与每一所述对象标识对应的第一序列。

本公开实施例通过对筛选的单个对象的候选检测框在空间位置上取并集，得到每个
5 单个对象的最小包围框，从而根据最小包围框截取每一帧图像的区域图像，并对区域图
像的噪声背景进行填充，最终连接处理后的区域图像得到每一第一序列。借助待识别视
频序列的检测结果，将待识别视频序列拆分成若干个对象人的轨迹序列，将视频分类问
题转换成轨迹序列的识别问题，能够减小模型的识别难度，并提升有效信息的获取。同
时，利用空间运动掩膜解决以往算法丢失相对运动信息问题，特别是对于跑步、走路等
10 空间相似但空间运动节奏不同的动作有较好的性能提升。

图 5 为本公开实施例提供的再一种行为识别方法的流程示意图，如图 5 所述，所述方法包括以下步骤：

步骤 S510，获取待识别视频序列中每一帧图像中每一对象的检测结果。

这里，对于滑动窗口内的视频帧序列，通过相关检测算法获取每一帧图像中每一对
15 象的检测结果即每一对象的检测框和与检测框关联的对象标识。

步骤 S520，根据每一所述对象的检测结果，生成至少一个第一序列。

这里，所述第一序列为包括同一对象的轨迹序列。根据上一步骤获得的每一帧图像
中每一对象的检测结果，将属于同一单个对象的检测框（或者基于检测框确定的最小包
围区域）分别连接起来，形成每一单个对象的轨迹序列。

20 步骤 S530，将每一所述第一序列送入到行为识别模型中，得到每一所述第一序列的
判别分数。

这里，将每一第一序列经过行为识别模型，得到各自的分类结果，即如果判别分数
高于某个特定阈值，则输出异常；否则为正常。

步骤 S540，根据每一所述第一序列的判别分数，确定每一所述对象的行为识别结果。

25 这里，在存在任一判别分数大于行为阈值的情况下，确定所述待识别视频序列中存
在单个对象行为；或者在不存在任一判别分数大于特定阈值的情况下，确定所述待识别
视频序列为不存在单个对象行为。

步骤 S550，在所述第一序列的判别分数大于行为阈值的情况下，确定相应第一序列
对应的单个对象行为发生的空间位置和行为类别。

30 这里，通过上一步骤识别出待识别视频序列中的单个对象行为，然后根据该单个对
象对应检测框在帧图像中的位置坐标，标记单个对象行为发生的空间位置和行为类别。
行为类别为针对单个对象行为，经过训练好的行为识别模型能输出的类别，如跌倒、攀
爬、发传单等行为。

步骤 S560，根据所述空间位置和行为类别，确定告警内容。

35 这里，通过对单个对象行为的执行主体进行标记，可以进一步将标记的空间位置
和行为类别生成告警内容。

步骤 S570, 根据告警内容向所述空间位置对应的终端设备发送告警通知。

可以理解的, 不同的位置区域有对应的持有终端设备的管理人员, 通过终端设备接收告警系统的通知, 可以快速定位和处理该位置区域内发生的单个对象行为。

在本公开实施例中, 借助待识别视频序列的每一对象的检测结果, 将待识别视频序列拆分成若干个对象的轨迹序列, 将视频分类问题转换成轨迹序列的识别问题, 能够减小模型的识别难度, 同时提升有效信息的获取。在室外的城市街道场景以及室内的轨道交通等场景下发生危害生命安全的行为后, 系统就会自动识别出行为主体并告警, 给相关需求的人员提供了一个高效而便捷的检测能力。

下面结合一个具体实施例对上述行为识别方法进行说明, 然而值得注意的是, 该具体实施例仅是为了更好地说明本公开, 并不构成对本公开的不当限定。

本公开实施例以对象为行人为例进行说明。对于单人行为识别的主要任务是首先对出现在视频源中的单个行人进行检测, 锁定行为主体, 然后对该行为主体进行持续检测, 识别出具体行为。当行为主体执行危害生命安全的行为时需要对该行人预警。这样的流程使得采集视频源的设备具有自主化和智能化的判别能力, 可以给日常城市管理带来极大的效率提升。

传统的行为识别方法通常对输入的视频序列进行全图的数据增强或其他预处理后送入到分类模型中进行预测, 然而这种方式只适配置为以人为中心的视频行为识别, 该类数据常见于公开的视频学术数据集中。对于其他视频源中的图像来说, 每帧图像中往往包含更多的信息, 覆盖的视野也更大。为此, 本公开实施例利用行人检测算法来进行行为主体的定位, 并获得准确的行为发生区间, 以此构建单人行为识别算法。

基于单人行为识别问题, 本公开实施例通过调用上游检测组件获取视频图像中的行人检测框, 基于行人检测框来生成多个行人检测框轨迹序列, 对每个行人的轨迹序列进行行为识别。图 6A 为本公开实施例提供的单人行为识别算法的系统框图, 如图 6A 所示, 将标有行人检测框的视频序列 61 经过预处理, 得到三个单人各自的轨迹序列即轨迹序列 611、轨迹序列 612 和轨迹序列 613, 再经过已训练的行为识别网络 62 分别对每个轨迹序列进行行为识别, 得到轨迹序列 611 对应的判别分数 621、轨迹序列 612 对应的判别分数 622 和轨迹序列 613 对应的判别分数 623, 最后, 再将视频序列 61 中所有的轨迹序列的判别分数进行合并, 得到该视频序列 61 的识别结果 63。

图 6B 为本公开实施例提供的单人行为识别算法的逻辑流程图, 如图 6B 所示, 该流程包括以下步骤:

步骤 S601, 构造滑窗式的视频帧序列缓冲区。

对于获取的一段待识别视频, 根据预设的采样间隔获取一帧 (frame) 图像, 并放入视频帧序列缓冲区 (frame_buffer)。在实施中, 根据线上允许的检测周期/频率 (例如 3 秒一次检测频率), 明确固定采样帧数下的采样间隔, 例如采样 8 帧图像的情况下, 采样间隔为 3/8 秒。当该视频帧序列缓冲区中的图像帧数超过最大缓冲区范围 (max_buffer_size) 时, 将该视频帧序列缓冲区中的第一帧图像弹出。

构造滑窗式视频帧序列缓冲区一方面能高效处理离线长视频和在线视频流，并行处理每个窗口内的视频帧序列。另一方面以固定大小的窗口作为基本处理单元可以保证采样的合理性，捕捉动作执行的全局过程。

步骤 S602，定义行人轨迹序列的数量。

- 5 对于当前视频帧序列缓冲区内的每一帧图像，首先对所有的行人检测框的宽和高均扩展 1.5 倍。然后根据所有的行人检测框，得到邻接矩阵。例如按面积选取前 10 个行人检测框，构建 10x10 的邻接矩阵，其中邻接矩阵中元素 (i, j) 的值代表检测框 i 和检测框 j 之间的交并比。如果不存在行人检测框，则跳过生成邻接矩阵的步骤。

- 10 计算检测框 i 的匹配次数时，可以将第 i 行中邻接矩阵中 (i, j) 的值大于 0 的个数（忽略与自身的交并比）作为匹配次数。根据邻接矩阵，从每一帧图像中筛选出匹配次数为 0 的行人检测框作为稀疏检测框，从而每一帧图像中稀疏检测框的数量即为行人轨迹序列的数量。通常认为独立的行人容易发生单人行为，因此本公开实施例对与其他行人检测框之间无重叠的行人检测框进行分析处理，能够提高单人行为识别的效率。

步骤 S603，确定行人的轨迹序列。

- 15 对上一步确定的稀疏检测框进行帧间行人检测框的 ID 匹配，根据不同的 ID，可以得到不同行人的轨迹序列。

- 20 首先将确定的稀疏检测框内缩 1.5 倍来恢复分辨率。然后对于每个 ID 在当前视频帧序列缓冲区内所有帧图像的稀疏检测框，计算最小包围框，也就是将视频帧序列缓冲区内属于同一个行人的稀疏检测框在空间位置上取并集，得到该行人在当前视频帧序列中的最小包围框。最终将最小包围框裁剪（crop）每一帧图像出来对应的区域图像连接，生成每个行人的轨迹序列。图 6C 为本公开实施例提供的预处理后的输入轨迹序列样例示意图，如图 6C 所示，为以小孩为执行主体的轨迹序列，可以看出，该轨迹序列是基于采集的 8 帧图像，然后利用包含小孩识别结果的最小包围框，分别截取 8 帧图像得到的区域图像连接起来形成的。

- 25 对于该视频帧序列缓冲区中的所有帧图像，可能存在一些帧缺失同一 ID 的行人的情况，例如采样 8 帧图像，属于同一个 ID 的行人可能只出现在 8 帧中的前 7 帧，对于不出现该行人的第 8 帧用 -1 进行补全。即用该行人在前 7 帧中计算出的最小包围框对纯色背景进行裁剪，从而根据所有裁剪的区域图像生成该行人最终的轨迹序列。有效解决了基于单人持续检测容易出现目标丢失问题。

- 30 步骤 S604，将多个行人的轨迹序列送入行为识别模型，确定识别结果。

将上一步确定的多个行人的轨迹序列送入行为识别模型中获取不同行为的判别分数。观察每个行人的轨迹序列经过行为识别模型后的输出分数，如果高于某个给定的阈值，则输出对应的单人行为，否则表示当前视频帧序列不存在单人行为。

- 35 本公开实施例提供的单人行为识别算法可以广泛应用于智能视频分析中的行为识别中。例如，在室外的城市街道场景以及室内的轨道交通等场景下发生危害生命安全的行为后，系统就会自动识别出行为主体并报警，给相关需求的人员提供了一个高效而便捷

的检测能力。

对于跑步，走路这类单人行为动作。由于其动作形式和动作幅度相似，很难见其区分。在行人检测框提取阶段，由于将行人裁剪出来，丢失了相对运动信息。为了解决这种相对运动信息丢失问题，设计了一种 MaskAttention 裁剪法来保持行人裁剪同时又不丢失相对运动信息，将裁减的行人区域保持其空间位置同时周围用纯色背景填充。如图 6D 所示，对于每帧用最小包围框裁剪的区域图像，该区域图像中除行人检测框外的部分都用纯色背景填充。由于最小包围框是根据多帧图像里的行人检测框取并集得到的，因此，最小包围框和行人检测框的相对位置已知，从而保持了行人的相对运动信息。

本公开实施例充分利用现有行人识别算法并设计了一套两阶段的单人行为识别算法框架，行人定位阶段可以灵活的使用现有所有行人识别算法，行为识别阶段充分利用现有视频理解模型，拓展了视频理解模型的边界，发挥了视频理解模型的潜力，也是对现有视频理解模型落地应用的体现。不同于以往的单帧行为识别算法，通过利用视频分类算法与检测算法的结合，将单帧行为识别算法拓展为一个具有时序关系的行为识别算法。

本公开实施例同时使用了空间运动掩膜裁剪法来保持行为主体的空间相对位置信息。由于行人检测导致行为主体的相对位置丢失，利用空间运动掩膜解决以往算法丢失相对运动信息问题。特别是对于跑步、走路等空间相似但空间运动节奏不同的动作有较好的性能提升。

基于前述的实施例，本公开实施例再提供一种行为识别装置，所述识别装置包括所包括的各模块、以及各模块所包括的各单元，可以通过行为识别设备中的处理器来实现；当然也可通过具体的逻辑电路实现；在实施的过程中，处理器可以为中央处理器（Central Processing Unit, CPU）、微处理器（Micro Processing Unit, MPU）、数字信号处理器（Digital Signal Processor, DSP）或现场可编程门阵列（Field Programmable Gate Array, FPGA）等。

图 7 为本公开实施例提供的一种行为识别装置的组成结构示意图，如图 7 所示，所述识别装置 700 包括获取模块 710、第一确定模块 720 和识别模块 730，其中：

所述获取模块 710，配置为获取待识别视频序列中每一帧图像中每一对象的检测结果；

所述第一确定模块 720，配置为根据每一所述对象的检测结果，生成至少一个第一序列；其中，每一所述第一序列为包括同一对象的轨迹序列；

所述识别模块 730，配置为对每一所述第一序列进行行为识别，得到每一所述对象的行为识别结果。

在一些可能的实施例中，所述识别装置 700 还包括选取模块、存储模块和第二确定模块，其中：所述选取模块，配置为按照特定时间间隔，从视频源中选取至少两帧图像；所述存储模块，配置为将所述至少两帧图像存入帧序列缓冲区；其中，所述帧序列缓冲区的长度固定；所述第二确定模块，配置为获取所述帧序列缓冲区中的帧图像作为所述待识别视频序列。

在一些可能的实施例中，所述对象的检测结果至少包括所述对象的检测框和与所述检测框关联的对象标识；所述第一确定模块 720 包括第一确定子模块和第二确定子模块，其中：所述第一确定子模块，配置为从每一所述对象的检测框中，选择在空间位置上独立的检测框作为候选检测框；所述第二确定子模块，配置为根据所述候选检测框和与所述候选检测框关联的对象标识，生成所述至少一个第一序列。

在一些可能的实施例中，所述每一帧图像中包括一个检测框，所述第一确定子模块还配置为将所述每一帧图像中的一个检测框作为所述候选检测框。

在一些可能的实施例中，所述每一帧图像中包括至少两个检测框，所述第一确定子模块包括第一确定单元、生成单元和第二确定单元，其中：所述第一确定单元，配置为确定所述每一帧图像中的至少两个检测框之间的交并比；所述生成单元，配置为根据所述交并比，生成所述每一帧图像对应的邻接矩阵；所述第二确定单元，配置为将所述邻接矩阵中匹配次数为零的检测框作为所述候选检测框。

在一些可能的实施例中，所述第一确定子模块还包括扩展单元和选取单元，其中：所述扩展单元，配置为对每一所述对象的检测框，向外扩展特定比例，得到扩展后的检测框；所述选取单元，配置为从所述扩展后的检测框中筛选至少两个第一检测框；其中，所述第一检测框的面积大于所述扩展后的检测框中其他检测框；相应地，所述第一确定单元，还配置为确定所述至少两个第一检测框之间的交并比。

在一些可能的实施例中，所述第二确定子模块包括第三确定单元、截取单元和连接单元，其中：所述第三确定单元，配置为对与每一所述对象标识关联的候选检测框在空间位置上取并集，得到与每一所述对象标识对应的最小包围框；所述截取单元，配置为根据所述最小包围框，截取所述每一帧图像中与所述最小包围框对应的区域图像；所述连接单元，配置为按照所述每一帧图像的时间戳，依次连接所述区域图像，得到与每一所述对象标识对应的第一序列。

在一些可能的实施例中，所述候选检测框为从所述至少两个第一检测框中选择的，所述第二确定子模块还包括内缩单元和第四确定单元，其中：所述内缩单元，配置为对所述候选检测框，按照所述特定比例进行内缩，得到原始尺寸的候选检测框；所述第四确定单元配置为根据所述原始尺寸的候选检测框和与所述候选检测框关联的对象标识，生成所述至少一个第一序列。

在一些可能的实施例中，所述第二确定子模块还包括第五确定单元和填充单元，其中：所述第五确定单元，配置为确定所述区域图像中除去所述候选检测框之外的噪声区域；所述填充单元，配置为利用纯色背景填充所述噪声区域，得到目标区域图像；相应地，所述连接单元，还配置为按照所述每一帧图像的时间戳，依次连接所述目标区域图像，得到与每一所述对象标识对应的第一序列。

在一些可能的实施例中，所述识别模块 730 包括识别子模块和第三确定子模块，其中：所述识别子模块，配置为将每一所述第一序列送入到行为识别模型中，得到每一所述第一序列的判别分数；所述第三确定子模块，配置为根据每一所述第一序列的判别分

数，确定所述每一所述对象的行为识别结果。

在一些可能的实施例中，所述识别装置 700 还包括第三确定模块、第四确定模块和发送模块，其中：所述第三确定模块，配置为在所述第一序列的判别分数大于行为阈值的情况下，确定相应第一序列对应的单个对象行为发生的空间位置和行为类别；所述第四确定模块，配置为根据所述空间位置和行为类别，确定告警内容；所述发送模块，配置为根据所述告警内容向所述空间位置对应的终端设备发送告警通知，以使持有所述终端设备的管理人员处理所述单个对象行为。

这里需要指出的是：以上装置实施例的描述，与上述方法实施例的描述是类似的，具有同方法实施例相似的有益效果。对于本公开装置实施例中未披露的技术细节，请参照本公开方法实施例的描述而理解。

需要说明的是，本公开实施例中，如果以软件功能模块的形式实现上述行为识别方法，并作为独立的产品销售或使用，也可以存储在一个计算机可读取存储介质中。基于这样的理解，本公开实施例的技术方案本质上或者说对相关技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得行为识别设备（可以是具有摄像头的智能手机、平板电脑等）执行本公开各个实施例所述方法的全部或部分。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read Only Memory, ROM）、磁碟或者光盘等各种可以存储程序代码的介质。这样，本公开实施例不限制于任何特定的硬件和软件结合。

对应地，本公开实施例提供一种计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时实现上述实施例中任一所述行为识别方法中的步骤。对应地，本公开实施例中，还提供了一种芯片，所述芯片包括可编程逻辑电路和/或程序指令，当所述芯片运行时，用于实现上述实施例中任一所述行为识别方法中的步骤。对应地，本公开实施例中，还提供了一种计算机程序产品，当该计算机程序产品被行为识别设备的处理器执行时，其用于实现上述实施例中任一所述行为识别方法中的步骤。

本公开实施例还提供一种计算机程序产品，该计算机程序产品承载有程序代码，所述程序代码包括的指令可用于执行上述方法实施例中任一所述行为识别方法中的步骤。其中，上述计算机程序产品可以具体通过硬件、软件或其结合的方式实现。在一个可选实施例中，所述计算机程序产品具体体现为计算机存储介质，在另一个可选实施例中，计算机程序产品具体体现为软件产品，例如软件开发包（Software Development Kit, SDK）等等。

本公开实施例还提供一种计算机程序，包括计算机可读代码，当所述计算机可读代码在电子设备中运行时，所述电子设备中的处理器执行用于实现上述方法实施例中任一所述行为识别方法。

基于同一技术构思，本公开实施例提供一种行为识别设备，用于实施上述方法实施例记载的行为识别方法。图 8 为本公开实施例提供的一种行为识别设备的硬件实体示意图，如图 8 所示，所述识别设备 800 包括存储器 810 和处理器 820，所述存储器 810 存储

有可在处理器 820 上运行的计算机程序, 所述处理器 820 执行所述程序时实现本公开实施例任一所述行为识别方法中的步骤。

存储器 810 用于存储由处理器 820 可执行的指令和应用, 还可以缓存待处理器 820 以及识别设备 800 中各模块待处理或已经处理的数据 (例如, 图像数据、音频数据、语音通信数据和视频通信数据), 可以通过闪存 (FLASH) 或随机访问存储器 (Random Access Memory, RAM) 实现。

处理器 820 执行程序时实现上述任一项的行为识别方法的步骤。处理器 820 通常控制识别设备 800 的总体操作。

上述处理器可以为特定用途集成电路 (Application Specific Integrated Circuit, ASIC)、数字信号处理器 (Digital Signal Processor, DSP)、数字信号处理装置 (Digital Signal Processing Device, DSPD)、可编程逻辑装置 (Programmable Logic Device, PLD)、现场可编程门阵列 (Field Programmable Gate Array, FPGA)、中央处理器 (Central Processing Unit, CPU)、控制器、微控制器、微处理器中的至少一种。可以理解地, 实现上述处理器功能的电子器件还可以为其它, 本公开实施例不作具体限定。

上述计算机存储介质/存储器可以是只读存储器 (Read Only Memory, ROM)、可编程只读存储器 (Programmable Read-Only Memory, PROM)、可擦除可编程只读存储器 (Erasable Programmable Read-Only Memory, EPROM)、电可擦除可编程只读存储器 (Electrically Erasable Programmable Read-Only Memory, EEPROM)、磁性随机存取存储器 (Ferromagnetic Random Access Memory, FRAM)、快闪存储器 (Flash Memory)、磁表面存储器、光盘、或只读光盘 (Compact Disc Read-Only Memory, CD-ROM) 等存储器; 也可以是包括上述存储器之一或任意组合的各种行为识别设备, 如移动电话、计算机、平板设备、个人数字助理等。

这里需要指出的是: 以上存储介质和设备实施例的描述, 与上述方法实施例的描述是类似的, 具有同方法实施例相似的有益效果。对于本公开存储介质和设备实施例中未披露的技术细节, 请参照本公开方法实施例的描述而理解。

应理解, 说明书通篇中提到的“一个实施例”或“一实施例”意味着与实施例有关的特定特征、结构或特性包括在本公开的至少一个实施例中。因此, 在整个说明书各处出现的“在一个实施例中”或“在一实施例中”未必一定指相同的实施例。此外, 这些特定的特征、结构或特性可以任意适合的方式结合在一个或多个实施例中。应理解, 在本公开的各种实施例中, 上述各过程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, 而不对本公开实施例的实施过程构成任何限定。上述本公开实施例序号仅仅为了描述, 不代表实施例的优劣。

需要说明的是, 在本文中, 术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含, 从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素, 而且还包括没有明确列出的其他要素, 或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下, 由语句“包括一个……”限定的要素,

并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

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

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

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

或者，本公开上述集成的单元如果以软件功能模块的形式实现并作为独立的产品销售或使用，也可以存储在一个计算机可读取存储介质中。基于这样的理解，本公开实施例的技术方案本质上或者说对相关技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得设备自动测试线执行本公开各个实施例所述方法的全部或部分。而前述的存储介质包括：移动存储设备、ROM、磁碟或者光盘等各种可以存储程序代码的介质。

本公开所提供的几个方法实施例中所揭露的方法，在不冲突的情况下可以任意组合，得到新的方法实施例。

本公开所提供的几个方法或设备实施例中所揭露的特征，在不冲突的情况下可以任意组合，得到新的方法实施例或设备实施例。

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

工业实用性

在本公开实施例中，首先获取待识别视频序列中每一帧图像中每一对象的检测结果；然后根据每一所述对象的检测结果，生成至少一个第一序列；其中，每一所述第一序列为包括同一对象的轨迹序列；最后对每一所述第一序列进行行为识别，得到每一所述对象的行为识别结果。通过利用视频分类算法与检测算法的结合，将单帧行为识别算法拓展为一个具有时序关系的行为识别算法。将视频分类问题转换成轨迹序列的识别问题，能够减小模型的识别难度，并提升有效信息的获取。同时拓展了视频理解模型的边界，发挥了视频理解模型的潜力，也是对现有视频理解模型落地应用的体现。

权利要求书

1、一种行为识别方法，所述方法包括：

获取待识别视频序列中每一帧图像中每一对象的检测结果；

根据每一所述对象的检测结果，生成至少一个第一序列；其中，每一所述第一序列
5 为包括同一对象的轨迹序列；

对每一所述第一序列进行行为识别，得到每一所述对象的行为识别结果。

2、如权利要求 1 所述的方法，其中，所述待识别视频序列是通过以下过程得到的：

按照特定时间间隔，从视频源中选取至少两帧图像；

将所述至少两帧图像存入帧序列缓冲区；其中，所述帧序列缓冲区的长度固定；

10 获取所述帧序列缓冲区中的帧图像作为所述待识别视频序列。

3、如权利要求 1 或 2 所述的方法，其中，所述对象的检测结果至少包括所述对象的检测框和与所述检测框关联的对象标识；

所述根据每一所述对象的检测结果，生成至少一个第一序列，包括：

从每一所述对象的检测框中，选择在空间位置上独立的检测框作为候选检测框；

15 根据所述候选检测框和与所述候选检测框关联的对象标识，生成所述至少一个第一序列。

4、如权利要求 3 所述的方法，其中，所述每一帧图像中包括一个检测框，所述从每一所述对象的检测框中，选择在空间位置上独立的检测框作为候选检测框，包括：

将所述每一帧图像中的一个检测框作为所述候选检测框。

20 5、如权利要求 3 所述的方法，其中，所述每一帧图像中包括至少两个检测框，所述从每一所述对象的检测框中，选择在空间位置上独立的检测框作为候选检测框，包括：

确定所述每一帧图像中的至少两个检测框之间的交并比；

根据所述交并比，生成所述每一帧图像对应的邻接矩阵；

将所述邻接矩阵中匹配次数为零的检测框作为所述候选检测框。

25 6、如权利要求 3 至 5 任一项所述的方法，其中，所述根据所述候选检测框和与所述候选检测框关联的对象标识，生成所述至少一个第一序列，包括：

对与每一所述对象标识关联的候选检测框在空间位置上取并集，得到与每一所述对象标识对应的最小包围框；

根据所述最小包围框，截取所述每一帧图像中与所述最小包围框对应的区域图像；

30 按照所述每一帧图像的时间戳，依次连接所述区域图像，得到与每一所述对象标识对应的第一序列。

7、如权利要求 5 所述的方法，其中，所述方法还包括：

对每一所述对象的检测框，向外扩展特定比例，得到扩展后的检测框；

从所述扩展后的检测框中筛选至少两个第一检测框；其中，所述第一检测框的面积

大于所述扩展后的检测框中的其他检测框;

所述确定所述每一帧图像中的至少两个检测框之间的交并比, 包括:

确定所述至少两个第一检测框之间的交并比。

8、如权利要求 7 所述的方法, 其中, 所述候选检测框为从所述至少两个第一检测框
5 中选择的, 所述根据所述候选检测框和与所述候选检测框关联的对象标识, 生成所述至少一个第一序列, 包括:

对所述候选检测框, 按照所述特定比例进行内缩, 得到原始尺寸的候选检测框;

根据所述原始尺寸的候选检测框和与所述候选检测框关联的对象标识, 生成所述至少一个第一序列。

9、如权利要求 6 所述的方法, 其中, 所述方法还包括:

确定所述区域图像中除去所述候选检测框之外的噪声区域;

利用纯色背景填充所述噪声区域, 得到目标区域图像;

按照所述每一帧图像的时间戳, 依次连接所述目标区域图像, 得到与每一所述对象标识对应的第一序列。

10、如权利要求 3 至 9 任一项所述的方法, 其中, 所述对每一所述第一序列进行行为识别, 得到每一所述对象的行为识别结果, 包括:

将每一所述第一序列送入到行为识别模型中, 得到每一所述第一序列的判别分数;

根据每一所述第一序列的判别分数, 确定每一所述对象的行为识别结果。

11、如权利要求 10 所述的方法, 其中, 所述方法还包括:

在所述第一序列的判别分数大于行为阈值的情况下, 确定相应第一序列对应的单个对象行为发生的空间位置和行为类别;

根据所述空间位置和行为类别, 确定告警内容;

根据所述告警内容向所述空间位置对应的终端设备发送告警通知。

12、一种行为识别装置, 其中, 所述装置包括获取模块、第一确定模块和识别模块,
25 其中:

所述获取模块, 配置为获取待识别视频序列中每一帧图像中每一对象的检测结果;

所述第一确定模块, 配置为根据每一所述对象的检测结果, 生成至少一个第一序列;
其中, 每一所述第一序列为包括同一对象的轨迹序列;

所述识别模块, 配置为对每一所述第一序列进行行为识别, 得到每一所述对象的行为识别结果。
30

13、如权利要求 12 所述的装置, 其中, 所述识别装置还包括:

选取模块, 配置为按照特定时间间隔, 从视频源中选取至少两帧图像;

存储模块, 配置为将所述至少两帧图像存入帧序列缓冲区; 其中, 所述帧序列缓冲区的长度固定;

第二确定模块, 配置为获取所述帧序列缓冲区中的帧图像作为所述待识别视频序列。
35

14、如权利要求 12 或 13 所述的装置, 其中, 所述对象的检测结果至少包括所述对

象的检测框和与所述检测框关联的对象标识；所述第一确定模块包括：

第一确定子模块，配置为从每一所述对象的检测框中，选择在空间位置上独立的检测框作为候选检测框；

第二确定子模块，配置为根据所述候选检测框和与所述候选检测框关联的对象标识，
5 生成所述至少一个第一序列。

15、如权利要求 14 所述的装置，其中，所述每一帧图像中包括一个检测框，所述第一确定子模块还配置为将所述每一帧图像中的一个检测框作为所述候选检测框。

16、如权利要求 14 所述的装置，其中，所述每一帧图像中包括至少两个检测框，所述
10 第一确定子模块包括：

第一确定单元，配置为确定所述每一帧图像中的至少两个检测框之间的交并比；

生成单元，配置为根据所述交并比，生成所述每一帧图像对应的邻接矩阵；

第二确定单元，配置为将所述邻接矩阵中匹配次数为零的检测框作为所述候选检测
框。

17、如权利要求 14 至 16 任一项所述的装置，其中，所述第二确定子模块包括：

15 第三确定单元，配置为对与每一所述对象标识关联的候选检测框在空间位置上取并集，得到与每一所述对象标识对应的最小包围框；

截取单元，配置为根据所述最小包围框，截取所述每一帧图像中与所述最小包围框对应的区域图像；

连接单元，配置为按照所述每一帧图像的时间戳，依次连接所述区域图像，得到与
20 每一所述对象标识对应的第一序列。

18、如权利要求 16 所述的装置，所述第一确定子模块还包括：

扩展单元，配置为对每一所述对象的检测框，向外扩展特定比例，得到扩展后的检测框；

25 选取单元，配置为从所述扩展后的检测框中筛选至少两个第一检测框；其中，所述第一检测框的面积大于所述扩展后的检测框中其他检测框；

所述第一确定单元，还配置为确定所述至少两个第一检测框之间的交并比。

19、如权利要求 18 所述的装置，其中，所述候选检测框为从所述至少两个第一检测框中选择的，所述第二确定子模块还包括：

30 内缩单元，配置为对所述候选检测框，按照所述特定比例进行内缩，得到原始尺寸的候选检测框；

第四确定单元，配置为根据所述原始尺寸的候选检测框和与所述候选检测框关联的对象标识，生成所述至少一个第一序列。

20、如权利要求 17 所述的装置，其中，所述第二确定子模块还包括：

第五确定单元，配置为确定所述区域图像中除去所述候选检测框之外的噪声区域；

35 填充单元，配置为利用纯色背景填充所述噪声区域，得到目标区域图像；

所述连接单元，还配置为按照所述每一帧图像的时间戳，依次连接所述目标区域图

像，得到与每一所述对象标识对应的第一序列。

21、如权利要求 14 至 20 任一项所述的装置，其中，所述识别模块包括：

识别子模块，配置为将每一所述第一序列送入到行为识别模型中，得到每一所述第一序列的判别分数；

5 第三确定子模块，配置为根据每一所述第一序列的判别分数，确定所述每一所述对象的行为识别结果。

22、如权利要求 21 所述的装置，其中，所述识别装置还包括：

第三确定模块，配置为在所述第一序列的判别分数大于行为阈值的情况下，确定相应第一序列对应的单个对象行为发生的空间位置和行为类别；

10 第四确定模块，配置为根据所述空间位置和行为类别，确定告警内容；

发送模块，配置为根据所述告警内容向所述空间位置对应的终端设备发送告警通知，以使持有所述终端设备的管理人员处理所述单个对象行为。

23、一种行为识别设备，包括存储器和处理器，所述存储器存储有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权利要求 1 至 11 任一项所述方法中的
15 步骤。

24、一种计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时实现权利要求 1 至 11 任一项所述方法中的步骤。

25、一种计算机程序，包括计算机可读代码，当所述计算机可读代码在电子设备中运行时，所述电子设备中的处理器执行用于权利要求 1 至 11 中任一项所述的方法。

20 26、一种计算机程序产品，所述计算机程序产品包括一条或多条指令，所述一条或多条指令适于由处理器加载并执行如权利要求 1 至 11 任一项所述方法中的步骤。

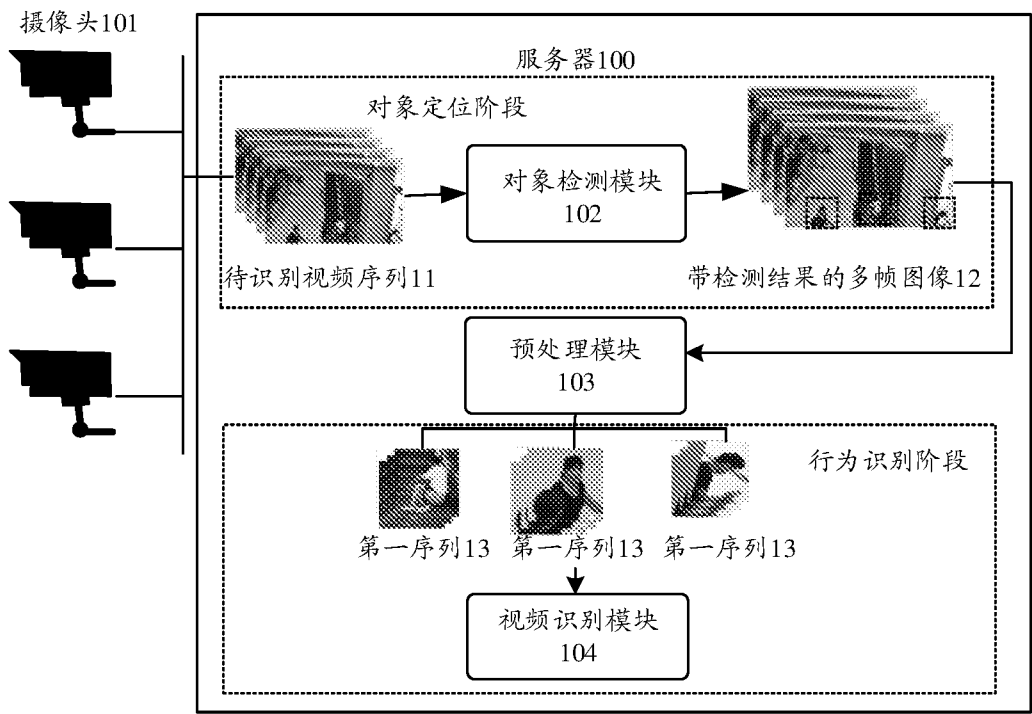


图 1A

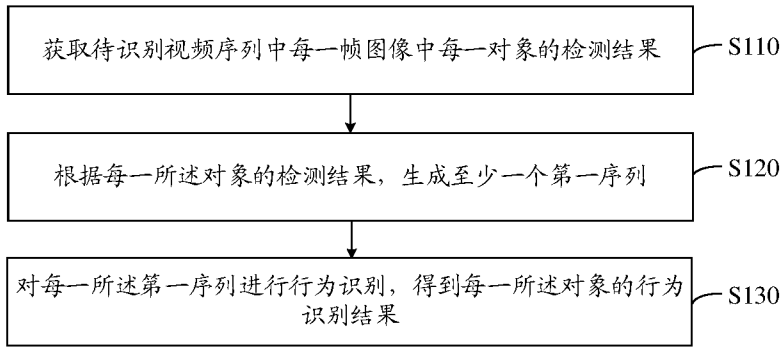


图 1B

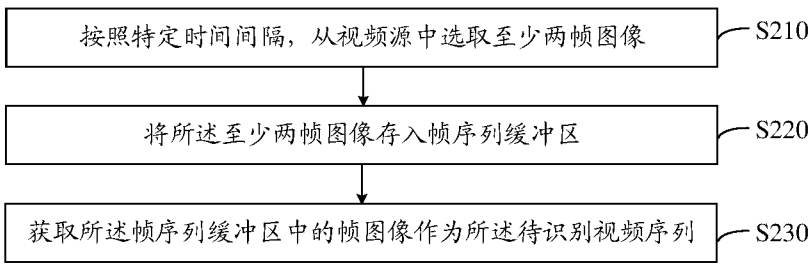


图 2

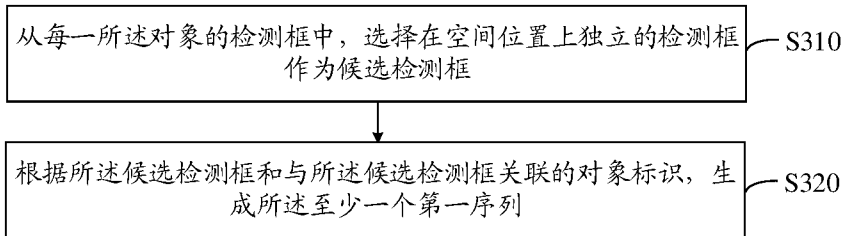


图 3

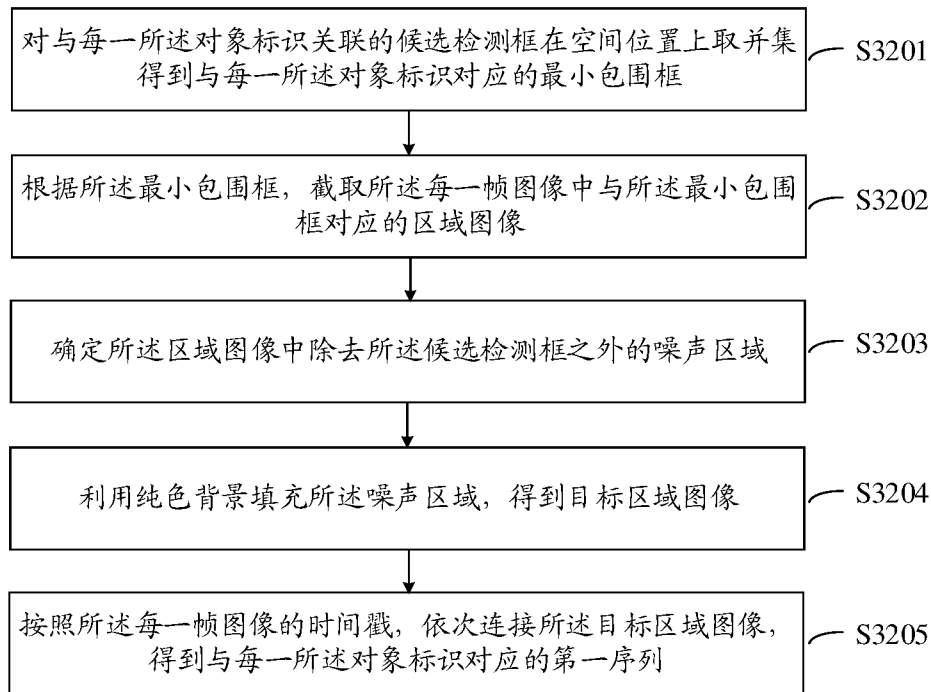


图 4

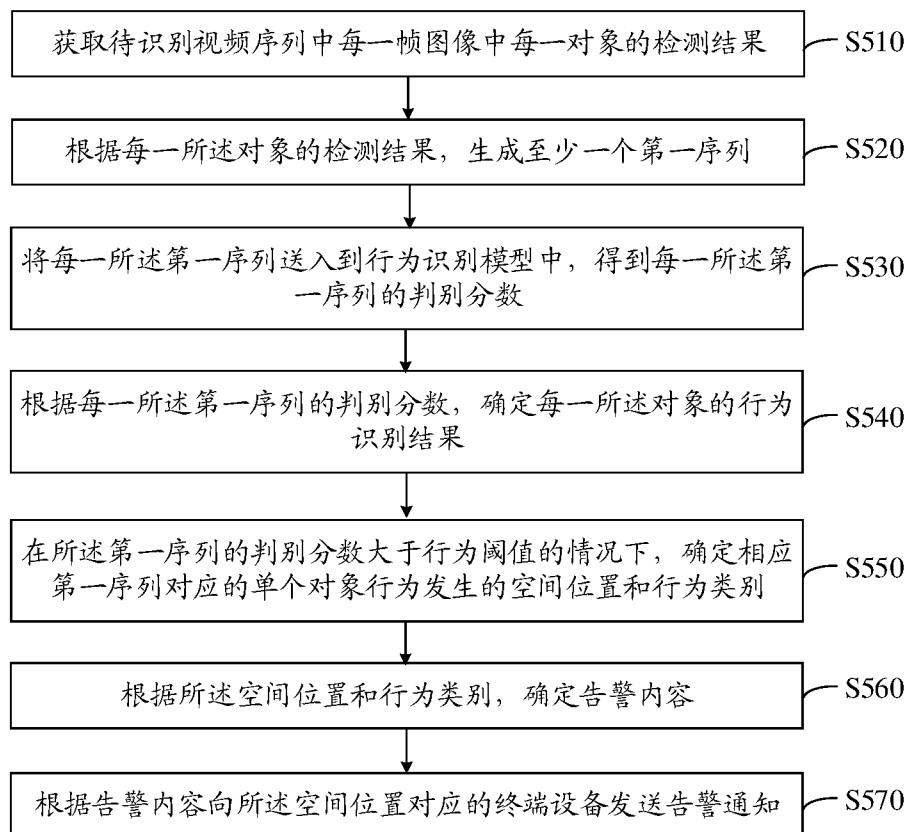


图 5

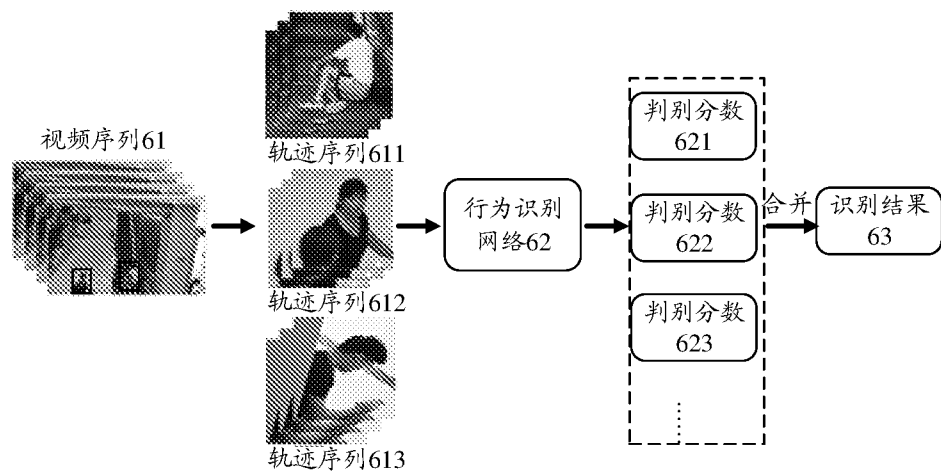


图 6A

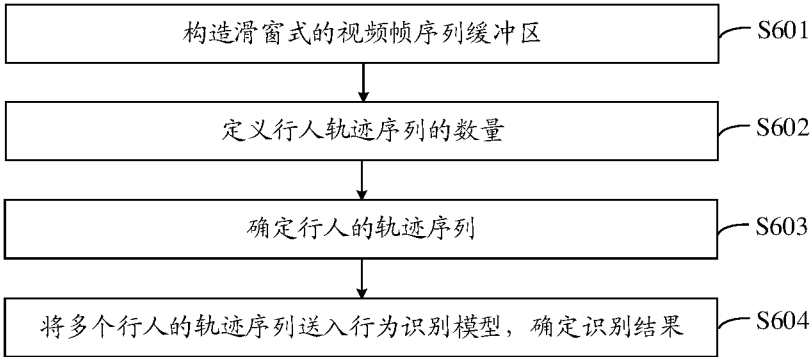


图 6B



图 6C

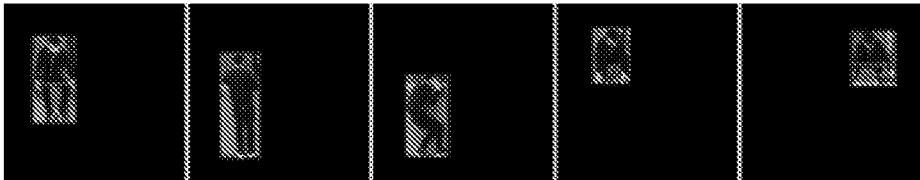


图 6D

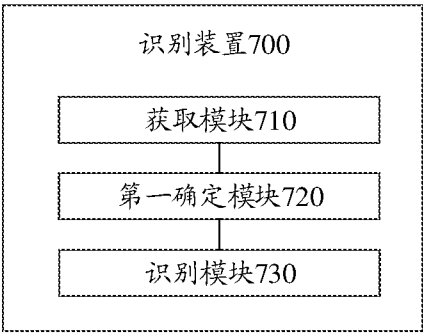


图 7

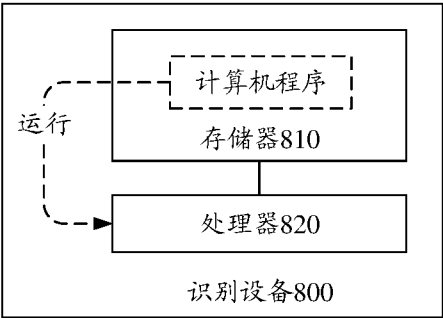


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/130036

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(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)

G06K, G06N, G06V

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)

DWPI, CNABS, WPABS, CNTXT, ENTXT, CNKI: 商汤, 行为, 动作, 轨迹, 追踪, 跟踪, 视频, 图像, 帧, 识别, 序列, 检测框, 区域, 矩阵, 裁剪, behavior, action, track, tracking, tracker, trace, MOT, multiple object tracking, video, image, frame, recognition, identification, sequence, detection box, area, block, matrix, crop

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113111839 A (SHANGHAI SHANGTANG INTELLIGENT TECHNOLOGY CO., LTD.) 13 July 2021 (2021-07-13) description, paragraphs 49-164	1-26
X	CN 111652043 A (BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD.) 11 September 2020 (2020-09-11) description, paragraphs 51-131	1-26
A	CN 108446669 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 24 August 2018 (2018-08-24) entire document	1-26
A	CN 112150514 A (SHANGHAI EYE CONTROL TECHNOLOGY CO., LTD.) 29 December 2020 (2020-12-29) entire document	1-26
A	CN 112287741 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD.; BEIJING JINGDONG CENTURY TRADING CO., LTD.) 29 January 2021 (2021-01-29) entire document	1-26



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

11 February 2022

Date of mailing of the international search report

17 February 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

International application No.

PCT/CN2021/130036**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 112016413 A (NANJING LEADING TECHNOLOGY CO., LTD.) 01 December 2020 (2020-12-01) entire document	1-26
A	US 2016132754 A1 (UNIV JOHNS HOPKINS) 12 May 2016 (2016-05-12) entire document	1-26

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/130036

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113111839	A	13 July 2021	None			
CN	111652043	A	11 September 2020	None			
CN	108446669	A	24 August 2018	None			
CN	112150514	A	29 December 2020	None			
CN	112287741	A	29 January 2021	None			
CN	112016413	A	01 December 2020	None			
US	2016132754	A1	12 May 2016	WO	2013177586	A1	28 November 2013

A. 主题的分类 G06K 9/00 (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) G06K, G06N, G06V 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)) DWPI, CNABS, WPABS, CNTXT, ENTXT, CNKI: 商汤, 行为, 动作, 轨迹, 追踪, 跟踪, 视频, 图像, 帧, 识别, 序列, 检测框, 区域, 矩阵, 裁剪, behavior, action, track, tracking, tracker, trace, MOT, multiple object tracking, video, image, frame, recognition, identification, sequence, detection box, area, block, matrix, crop		
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 113111839 A (上海商汤智能科技有限公司) 2021年7月13日 (2021 - 07 - 13) 说明书第49-164段	1-26
X	CN 111652043 A (北京三快在线科技有限公司) 2020年9月11日 (2020 - 09 - 11) 说明书第51-131段	1-26
A	CN 108446669 A (腾讯科技深圳有限公司) 2018年8月24日 (2018 - 08 - 24) 全文	1-26
A	CN 112150514 A (上海眼控科技股份有限公司) 2020年12月29日 (2020 - 12 - 29) 全文	1-26
A	CN 112287741 A (北京京东尚科信息技术有限公司 北京京东世纪贸易有限公司) 2021年1月29日 (2021 - 01 - 29) 全文	1-26
A	CN 112016413 A (南京领行科技股份有限公司) 2020年12月1日 (2020 - 12 - 01) 全文	1-26
A	US 2016132754 A1 (UNIV JOHNS HOPKINS) 2016年5月12日 (2016 - 05 - 12) 全文	1-26
☐ 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2月11日		国际检索报告邮寄日期 2022年2月17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		受权官员 田竞 电话号码 86-10-62411749

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/130036

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	113111839	A	2021年7月13日	无	
CN	111652043	A	2020年9月11日	无	
CN	108446669	A	2018年8月24日	无	
CN	112150514	A	2020年12月29日	无	
CN	112287741	A	2021年1月29日	无	
CN	112016413	A	2020年12月1日	无	
US	2016132754	A1	2016年5月12日	W0 2013177586 A1	2013年11月28日