

(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
2018 年 10 月 25 日 (25.10.2018)



(10) 国际公布号
WO 2018/192570 A1

(51) 国际专利分类号:
G06K 9/00 (2006.01)

(21) 国际申请号: PCT/CN2018/083924

(22) 国际申请日: 2018 年 4 月 20 日 (20.04.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710263004.2 2017年4月20日 (20.04.2017) CN

(71) 申请人: 北京市商汤科技开发有限公司(BEIJING SENSETIME TECHNOLOGY DEVELOPMENT CO., LTD) [CN/CN]; 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。

(72) 发明人: 汤晓鸥(TANG, Xiaouu); 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。 熊元骏(XIONG, Yuanjun); 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。 赵岳(ZHAO, Yue); 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。 王利民(WANG, Limin); 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。 武智融(WU, Zhirong); 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。 林达华(LIN, Dahua); 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。

(74) 代理人: 北京思源智汇知识产权代理有限公司 (BEIJING SIYUANZHUI INTELLECTUAL

(54) Title: TIME DOMAIN MOTION DETECTION METHOD AND SYSTEM, ELECTRONIC DEVICE AND COMPUTER STORAGE MEDIUM

(54) 发明名称: 时域动作检测方法和系统、电子设备、计算机存储介质

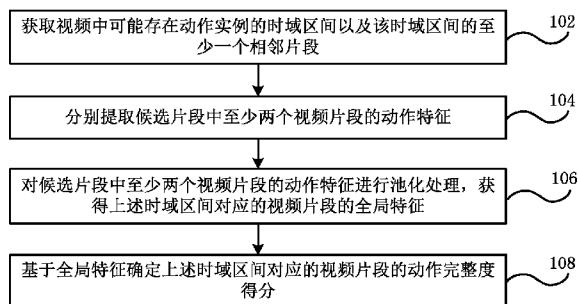


图 1

- 102 ACQUIRE A TIME DOMAIN INTERVAL WHERE A MOTION INSTANCE POSSIBLY EXISTS, IN A VIDEO, AND AT LEAST ONE ADJACENT CLIP OF THE TIME DOMAIN INTERVAL
- 104 RESPECTIVELY EXTRACT MOTION FEATURES OF AT LEAST TWO VIDEO CLIPS, IN CANDIDATE CLIPS
- 106 CARRY OUT POOLING PROCESSING ON THE MOTION FEATURES OF THE AT LEAST TWO VIDEO CLIPS, IN THE CANDIDATE CLIPS, AND ACQUIRE A GLOBAL FEATURE OF THE VIDEO CLIP CORRESPONDING TO THE ABOVE-MENTIONED TIME DOMAIN INTERVAL
- 108 DETERMINE, BASED ON THE GLOBAL FEATURE, A MOTION INTEGRITY SCORE OF THE VIDEO CLIP CORRESPONDING TO THE ABOVE-MENTIONED TIME DOMAIN INTERVAL

(57) Abstract: Disclosed are a time domain motion detection method and system, an electronic device and a computer storage medium, the method comprising: acquiring a time domain interval where a motion instance exists, in a video, and at least one adjacent clip of the time domain interval; respectively extracting motion features of at least two video clips, in candidate clips, wherein the candidate clips comprise a video clip corresponding to the time domain interval and an adjacent clip thereof; carrying out pooling processing on the motion features of the at least two video clips, in the candidate clips, and acquiring a global feature of the video clip corresponding

PROPERTY AGENCY); 中国北京市海淀区
中关村东路18号财智国际大厦B座502
室, Beijing 100083 (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,
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,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, 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))。

to the time domain interval; and determining, based on the global feature, a motion integrity score of the video clip corresponding to the time domain interval. The embodiments of the present application are beneficial to accurately determining whether an integrated motion instance exists in a time domain interval, thus improving the accuracy of motion integrity recognition.

(57) 摘要: 本申请实施例公开了一种时域动作检测方法和系统、电子设备、计算机存储介质, 其中, 方法包括: 获取视频中存在动作实例的时域区间以及所述时域区间的至少一个相邻片段; 分别提取候选片段中至少两个视频片段的动作特征, 其中, 所述候选片段包括所述时域区间对应的视频片段及其相邻片段; 对所述候选片段中至少两个视频片段的动作特征进行池化处理, 获得所述时域区间对应的视频片段的全局特征; 基于所述全局特征确定所述时域区间对应的视频片段的动作完整度得分。本申请实施例有益于准确的确定时域区间是否包含完整的动作实例, 提高了动作完整度识别的准确率。

时域动作检测方法和系统、电子设备、计算机存储介质

本申请要求在 2017 年 4 月 20 日提交中国专利局、申请号为 CN 201710263004.2、发明名称为“时域动作检测方法和系统、电子设备、计算机存储介质”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

本申请涉及计算机视觉技术，尤其涉及一种时域动作检测方法和系统、电子设备、计算机存储介质。

背景技术

理解人的行为和动作是计算机视觉技术的一项重要任务。时域动作检测技术，旨在视频中检测出与人物活动相关的动作片段，该视频可以是任何方式的视频，例如可能但不限于未经剪辑、较长的视频。时域动作检测技术在安全监控、网络视频分析、视频直播分析、乃至无人驾驶等领域都具有很大的应用价值。

发明内容

本申请实施例提供一种时域动作检测方案。

根据本申请实施例的一个方面，提供一种时域动作检测方法，包括：

获取视频中存在动作实例的时域区间以及所述时域区间的至少一个相邻片段；

分别提取候选片段中至少两个视频片段的动作特征，其中，所述候选片段包括所述时域区间对应的视频片段及其相邻片段；

对所述候选片段中至少两个视频片段的动作特征进行池化处理，获得所述时域区间对应的视频片段的全局特征；

基于所述全局特征确定所述时域区间对应的视频片段的动作完整度得分。

根据本申请实施例的另一个方面，提供一种时域动作检测系统，包括：

第一获取模块，用于获取视频中存在动作实例的时域区间以及所述时域区间的至少一个相邻片段；

特征提取模块，用于分别提取候选片段中至少两个视频片段的动作特征，其中，所述候选片段包括所述时域区间对应的视频片段及其相邻片段；

池化处理模块，用于对所述候选片段中至少两个视频片段的动作特征进行池化处理，获得所述时域区间对应的视频片段的全局特征；

第一确定模块，用于基于所述全局特征确定所述时域区间对应的视频片段的动作完整度得分。

根据本申请实施例的又一个方面，提供一种电子设备，包括上述任一实施例所述的时域动作检测系统。

根据本申请实施例的再一个方面，提供的另一种电子设备，包括：处理器和存储器；

所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行本申请任一实施例所述的时域动作检测方法对应的操作。

根据本申请实施例的再一个方面，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设备上运行时，所述设备中的处理器执行用于实现本申请任一实施例所述的时域动作检测方法中各步骤的指令。

根据本申请实施例的再一个方面，提供一种计算机存储介质，用于存储计算机可读的指令，所述指令被执行时实现本申请任一实施例所述的时域动作检测方法中各步骤的操作。

基于本申请上述实施例提供的时域动作检测方法和系统、电子设备、计算机程序、计算机存储介质，通过获取视频中存在动作实例的时域区间以及时域区间的至少一个相邻片段，提取该时域区间以及时域区间的至少一个相邻片段的动作特征并进行池化处理，获得该时域区间对应的视频片段的全局特征，基于全局特征确定时域区间对应的视频片段的动作完整度得分。本申请实施例基于包含时域区间的上下文信息的全局特征来确定时域区间对应的视频片段的动作完整度，从而有益于准确的确定时域区间是否包含完整的动作实例，提高了动作完整度识别的准确率。

下面通过附图和实施例，对本申请的技术方案做进一步的详细描述。

附图说明

构成说明书的一部分的附图描述了本申请的实施例，并且连同描述一起用于解释本申请的原理。

参照附图，根据下面的详细描述，可以更加清楚地理解本申请，其中：

图 1 为本申请时域动作检测方法一个实施例的流程图。

图 2 为本申请实施例中获得时序动作度序列一个实施例的流程图。

图 3 为本申请时域动作检测方法又一个实施例的流程图。

图 4 为本申请实施例中获得全局特征一个实施例的流程图。

图 5 为本申请实施例中获得全局特征另一个实施例的流程图。

图 6 为本申请实施例中确定完整度得分一个实施例的流程图。

图 7 为本申请时域动作检测方法又一个实施例的流程图。

图 8 为本申请时域动作检测系统一个实施例的结构示意图。

图 9 为本申请实施例中第一获取模块一个实施例的结构示意图。

图 10 为本申请时域动作检测系统另一个实施例的结构示意图。

图 11 为本申请实施例中第二获取模块一个实施例的结构示意图。

图 12 为本申请实施例中池化模块一个实施例的结构示意图。

图 13 为本申请实施例中第一确定模块一个实施例的结构示意图。

图 14 为本申请实施例中第三确定模块一个实施例的结构示意图。

图 15 为本申请时域动作检测系统又一个实施例的结构示意图。

图 16 为本申请电子设备一个应用实施例的结构示意图

具体实施方式

现在将参照附图来详细描述本申请的各种示例性实施例。应注意到：除非另外具体说明，否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本申请的范围。

同时，应当明白，为了便于描述，附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作为对本申请及其应用或使用的任何限制。

对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论，但在适当情况下，所述技术、方法和设备应当被视为说明书的一部分。

应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步讨论。

本申请实施例可以应用于终端设备、计算机系统、服务器等电子设备，其可与众多其它通用或专用计算机系统环境或配置一起操作。适于与终端设备、计算机系统、服务器等电子设备一起使用的众所周知的计算机系统、环境和/或配置的例子包括但不限于：个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境，等等。

终端设备、计算机系统、服务器等电子设备可以在由计算机系统执行的计算机系统可执行指令（诸如程序模块）的一般语境下描述。通常，程序模块可以包括例程、程序、目标程序、组件、逻辑、数据结构等等，它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施，分布式云计算环境中，任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中，程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

在本申请实施例中，卷积神经网络（Convolutional Neural Network, CNN）是已训练好的卷积神经网络，其能够实现本申请各实施例中记载的相应操作，例如，对输入的图像或视频片段进行动作度估计、分类、完整度识别，以获得输入的图像或视频片段的动作度值、相对于每个预设动作类别的得分、相对于每个动作类别的完整度得分、计算候选片段的置信度值等。

在本申请实施例中，视频片段指的是视频中的一部分，可以包括视频中的多帧连续图像。

在本申请实施例中，动作实例指的是某动作类别的具体动作。

在本申请实施例中，动作度（Actionness）表征的是某帧图像或某视频片段可能属于某动作实例的程度。

在本申请实施例中，完整度表征的是候选片段包含某动作类别完整的动作实例的程度。

在本申请实施例中，置信度表征的指某个候选片段是一个某动作类别动作的完整动作实例的概率。其可选的数学形式为该候选片段属于某动作类别的概率乘以该候选片段为该动作类别完整动作实例的概率。

图 1 为本申请时域动作检测方法一个实施例的流程图。如图 1 所示，该实施例的时域动作检测方法包括：

102，获取视频中可能存在动作实例的时域区间以及该时域区间的至少一个相邻片段。

其中，时域区间包括起始时间和结束时间，例如可以表示为起始时间和结束时间对（起始时间，结束

时间)。

在一个可选示例中,该操作 102 可以由处理器调用存储器存储的相应指令执行,也可以由被处理器运行的第一获取模块执行。

示例性地,上述至少一个相邻片段包括:视频中时序位于上述时域区间之前的第一相邻片段,和/或,视频中时序位于上述时域区间之后的第二相邻片段。其中的第一相邻片段和第二相邻片段分别包括至少一个视频片段。其中,第一相邻片段和第二相邻片段的长度可以根据实际需求设置,并且可以实时调整和更新,示例性地,第一相邻片段和第二相邻片段的长度可以是时域区间长度的一定比例,例如 20%~70%。

104,分别提取候选片段中至少两个视频片段的动作特征。

其中,候选片段包括上述时域区间对应的视频片段及其相邻片段。

可选地,操作 104 可以分别提取候选片段中各视频片段的动作特征。

在一个可选示例中,该操作 104 可以由处理器调用存储器存储的相应指令执行,也可以由被处理器运行的特征提取模块执行。

106,对候选片段中至少两个视频片段的动作特征进行池化处理,获得上述时域区间对应的视频片段的全局特征。

在一个可选示例中,该操作 106 可以由处理器调用存储器存储的相应指令执行,也可以由被处理器运行的池化处理模块执行。

108,基于全局特征确定上述时域区间对应的视频片段的动作完整度得分。

在一个可选示例中,该操作 108 可以由处理器调用存储器存储的相应指令执行,也可以由被处理器运行的第一确定模块执行。

基于本申请上述实施例提供的时域动作检测方法,通过获取视频中存在动作实例的时域区间以及时域区间的至少一个相邻片段,提取该时域区间以及时域区间的至少一个相邻片段的动作特征并进行池化处理,获得该时域区间对应的视频片段的全局特征,基于全局特征确定时域区间对应的视频片段的动作完整度得分。本申请实施例基于包含时域区间的上下文信息的全局特征来确定时域区间对应的视频片段的动作完整度,从而有益于准确的确定时域区间是否包含完整的动作实例,提高了动作完整度识别的准确率。

可选地,上述操作 102 中,获取视频中存在动作实例的时域区间以及时域区间的至少一个相邻片段,具体可以包括:

分别对视频中的至少两个视频片段进行动作度估计,获得时序动作度序列;

基于时序动作度序列进行动作位置预测,获得视频中存在动作实例的时域区间,时域区间包括起始时间和结束时间;

从视频中提取时域区间之前的第一相邻片段和/或时域区间之后的第二相邻片段。

每个时域区间之前和之后的相邻片段中包含该时域区间的上下文信息,提取每个时域区间之前和之后的相邻片段,便可以获得每个动作位置预测的时域区间的上下文信息,以便结合该上下文信息准确识别时域区间是否包括了某动作类别的一个完整的动作实例,从而更精确的定位一个完整的动作实例。

图 2 为本申请实施例中获得时序动作度序列一个实施例的流程图。如图 2 所示,该实施例可以通过如下操作实现:分别针对视频中的任一视频片段执行:

202,提取一帧图像作为原始图像,对原始图像进行动作度估计,得到第一动作度值;提取视频片段的光流,并对获得的光流场图片进行合并,得到拼合光流场图像,对拼合光流场图像进行动作度估计,得到第二动作度值。

例如,从视频中每隔若干帧(如 6 帧)提取一帧原始图像,以该帧原始图像及其之后的五帧图像作为一个视频片段,提取该帧原始图像及其之后的五帧图像的光流,得到五帧光流场图片。或者以视频中的每六帧图像构成一个视频片段,按照预设规则或者随机从每个视频片段中提取一帧图像作为原始图像。

其中的光流场图片,例如可以是基于 8 位位图、共 256 个离散的色阶的灰度图像,灰度图像的中值为 128。

由于光流场是一个向量场,当使用灰度图像表示光流场图片时,需要用两幅标量场图片表示光流场图片,即分别对应于光流场图片坐标轴的 X 方向和 Y 方向幅度的两幅标量场图片。

示例性地,假设每个视频片段包括 N 帧图像,则分别基于 N 帧图像中的每相邻的两帧图像进行计算,获得 N-1 组光流场图片,其中 N-1 组光流场图片中的每一组光流场图片分别包括一帧横向光流图像及一帧纵向光流图像。将光流场图片的像素值按照均值为 128 的标准,线性映射到 0 到 255 的区间,然后将这 N-1 组光流场图片作为单独的通道拼合为一张 N-1 通道的拼合光流场图像。

在一个可选示例中,该操作 202 可以由处理器调用存储器存储的相应指令执行,也可以由被处理器运行的动作估计子模块中的图像获取单元、第一子卷积神经网络和第二子卷积神经网络执行。

在一个可选的示例中，可以分别通过两个预先训练好的卷积神经网络（本申请实施例中的第一子卷积神经网络和第二子卷积神经网络），对原始图像进行动作度估计得到第一动作度值、对拼合光流场图像进行动作度估计得到第二动作度值的操作。

可选地，可以通过标注有参考第一动作度值的样本图像对第一卷积神经网络模型进行训练，来获得第一子卷积神经网络；通过标注有参考第二动作度值的样本拼合光流场图像对第二卷积神经网络模型进行训练，来获得第二子卷积神经网络。

可选地，通过第一卷积神经网络模型对样本视频中的样本图像进行动作度估计、输出第一预测动作度值，基于第一预测动作度值与参考第一动作度值之间的差值对第一卷积神经网络模型进行训练，即：对第一卷积神经网络模型的网络参数值进行调整，直至第一预测动作度值与参考第一动作度值之间的差值满足预设条件，例如小于第一预设阈值，从而获得训练好的第一子卷积神经网络。

通过第二卷积神经网络模型对样本拼合光流场图像进行动作度估计、输出第二预测动作度值，基于第二预测动作度值与参考第二动作度值之间的差值对第二卷积神经网络模型进行训练，即：对第二卷积神经网络模型的网络参数值进行调整，直至第二预测动作度值与参考第二动作度值之间的差值满足预设条件，例如小于第二预设阈值，从而获得训练好的第二子卷积神经网络。

204，由第一动作度值与第二动作度值，获取视频片段的动作度值。

例如，将第一动作度值与第二动作度值相加，由于同时结合了光流场图像的动作度值，以获得更加准确的视频片段的动作度值。

通过操作 204 可以得到视频中至少一个视频片段的动作度值，视频中所有视频片段的动作度值在时序上构成动作度序列，即：根据时间变化的动作度值，称为时序动作度序列。例如，假设视频中包含 M 个视频片段，其经过卷积神经网络计算得出的动作度估值分别为 $A1 \sim AM$ ，那么该视频的时序动作度序列可表示为 $A1, A2, A3, \dots, AM$ 。

在一个可选示例中，该操作 204 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的动作估计子模块中的计算单元执行。

可选地，在获得待检测视频的时序动作度序列的另一个实施例中，通过操作 204 获取视频片段的动作度值之后，还可以选择性地执行：

206，对视频片段的动作度值进行归一化处理，将视频片段的动作度值映射到 $[0,1]$ 的范围内，得到归一化的动作度值。则，基于实施例中获得的动作度序列可以为：由归一化的动作度值形成的时序动作度序列，其形式为动作度值随时间变化的一维函数，值域范围为 $0 \sim 1$ 。

例如，可利用 Sigmoid 函数或 Tanh 函数将至少一个视频片段的动作度值映射到 $[0,1]$ 的范围内。将时序动作度序列中的至少一个动作度值归一化后，可更规范的表示各动作度值及不同动作度值之间的差异，以便于后续准确地进行动作位置预测。

在一个可选示例中，该操作 206 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的动作估计子模块中的归一化单元执行。

在本申请的一个示例中，通过上述实施例获得时序动作度序列后，可以利用预先设置的动作位置预测算法，例如滑窗法、时序动作度聚合算法等，基于时序动作度序列，聚合得到视频中的动作预测时域区间。

仍以以上包含 M 个片段的视频为例，对于时序动作度序列 $A1, A2, A3, \dots, AM$ ，可通过时序动作度聚合算法，聚合得到一个或多个动作实例，可表示为例如 $A1A2$ 、 $A5A6A7$ 等等，其中 $A1A2$ 表示视频中的第一个片段和第二个片段为一个动作实例， $A5A6A7$ 表示视频中的第五、六、七个片段为一个动作实例。由于 $A1, A2, A3, \dots, AM$ 是在时序上的序列，因此，序列中每个元素均具有对应的时间坐标，从而可以得到每个动作实例的发生时间区间，即，在时域中的预测时域区间，其表示可能存在动作实例的一系列视频片段在时域中的位置。

由此，根据视频中多个片段分别对应的动作度值生成视频的时序动作度序列，再对时序动作度序列进行聚合，从而能够得到视频中一个或多个动作实例的发生时间（如，开始时间和结束时间）。根据本申请实施例，能够在各种视频中，包括在经过剪辑的视频和未经剪辑的原始视频中，可检测到每个动作实例的开始时间和结束时间，实现时域动作检测。

另外，在本申请时域动作检测方法的另一实施例中，还可以包括如下操作：

基于上述时域区间对应的视频片段的动作特征，获取该时域区间对应的视频片段的至少一个动作类别的分类得分；以及

根据该时域区间对应的视频片段的至少一个动作类别的分类得分，确定时域区间对应的视频片段的检测动作类别。

本申请实施例中，可以根据实际应用需要设置动作类别的数量，例如应用于多分类（ N 个可能类别）

时, 上述至少一个动作类别的分类得分可包括 $N+1$ 个: 时域区间对应的视频片段分别对应于 N 个分类的得分和 1 个背景类的得分。则, 可以选取 $N+1$ 个分类得分中数值最高的分类得分对应的都工作类别作为该时域区间对应的视频片段的检测动作类别。

另外, 本申请实施例也可应用于识别视频中是否存在特定类别的动作, 则上述至少一个动作类别的分类得分可包括分别对应于视频中是、否存在特定类别的动作的两个得分。

可选地, 在本申请时域动作检测方法的一实施例中, 还可以输出时域区间以及时域区间对应的视频片段的检测动作类别, 输出信息可以表示为三元组 (动作类别, 起始时间, 结束时间)。

由此, 基于本申请实施例的时域动作检测方法, 可以获得检测视频中可能包含动作的时域区间和该动作所属的动作类别, 以便用于准确的定位至视频中相应动作类别的位置。

图 3 为本申请时域动作检测方法再一个实施例的流程图。如图 3 所示, 该实施例的时域动作检测方法包括:

302, 获取视频中可能存在动作实例的时域区间以及该时域区间的至少一相邻片段。

其中, 时域区间包括起始时间和结束时间, 例如可以表示为起始时间和结束时间对 (起始时间, 结束时间)。

在一个可选示例中, 该操作 302 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的第一获取模块执行。

示例性地, 上述至少一相邻片段包括: 视频中时序位于上述时域区间之前的第一个相邻片段, 和/或, 视频中时序位于上述时域区间之后的第二相邻片段。其中的第一相邻片段和第二相邻片段分别包括至少一个视频片段。

304, 分别提取候选片段中至少两个视频片段的动作特征。

其中, 候选片段包括上述时域区间对应的视频片段及其相邻片段。

之后, 同时执行操作 306 和 310。

在一个可选示例中, 该操作 304 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的特征提取模块执行。

306, 对候选片段中至少两个视频片段的动作特征进行池化处理, 获得上述时域区间对应的视频片段的全局特征。

在一个可选的示例中, 可以对提取的候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理, 获得上述时域区间对应的视频片段的全局特征。

在一个可选示例中, 该操作 306 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的池化处理模块执行。

308, 基于全局特征确定上述时域区间对应的视频片段的动作完整度得分。

之后, 执行操作 314。

在一个可选示例中, 该操作 308 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的第一确定模块执行。

310, 基于上述时域区间对应的视频片段的动作特征, 获取该时域区间对应的视频片段的至少一个动作类别的分类得分。

在一个可选示例中, 该操作 310 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的第二获取模块执行。

示例性地, 该操作 310 可以通过如下方式实现:

分别基于上述时域区间对应的至少一个视频片段的动作特征, 获取该时域区间对应的至少一个视频片段分别属于上述至少一个动作类别的得分;

分别对该时域区间对应的至少一个视频片段属于同一动作类别的得分求和, 获得该时域区间对应的视频片段的至少一个动作类别的分类得分。

312, 根据上述时域区间对应的视频片段的至少一个动作类别的分类得分, 确定该时域区间对应的视频片段的检测动作类别。

在一个可选示例中, 该操作 312 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的第二确定模块执行。

314, 基于上述时域区间对应的视频片段的分类得分与动作完整度得分, 确定该时域区间对应的视频片段是否包含对应的至少一个动作类别完整的动作实例。

在一个可选示例中, 该操作 314 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的第三确定模块执行。

若该时域区间对应的视频片段包含对应的至少一个动作类别完整的动作实例，可以输出该时域区间以及时域区间对应的视频片段的检测动作类别，输出信息可以表示为三元组（检测动作类别，起始时间，结束时间）。否则，若该时域区间对应的视频片段未包含对应的至少一个动作类别完整的动作实例，便可以不输出该时域区间以及时域区间对应的视频片段的检测动作类别。

基于本申请上述实施例提供的时域动作检测方法，对视频进行动作位置预测获得可能包括动作的位置（即：时域区间）后，提取每个时域区间之前和之后的相邻片段，便可以获得每个动作位置预测的时域区间的上下文信息，由该时域区间对应的视频片段属于至少一个动作类别的分类得分和动作完整度得分，确定该时域区间对应的视频片段是否包括所属动作类别完整的动作实例，从而获得视频中包括完整动作实例的动作位置检测结果。基于本申请实施例，可以识别出视频中包括至少一个动作类别或特定动作类别完整的动作实例的视频片段，实现对至少一个动作类别或特定动作类别完整的动作实例的准确识别和精确定位；并且，可以同时获取时域区间对应的视频片段属于至少一个动作类别的得分和动作完整度得分，提高了整个时域动作检测过程的执行效率，节省检测时间。

在一个可选的示例中，操作 306 中，可以基于预先设置的配置规则，对候选片段中至少两个视频片段的动作特征进行结构化的时域金字塔式池化处理，获得上述时域区间对应的视频片段的全局特征。

本申请实施例中，金字塔式是指：候选片段中不同长度的视频片段的动作特征位于不同的层，位于不同层的、不同长度的视频片段的动作特征形成一个金字塔形状。例如，金字塔的第 1 层（最底层）包括整个候选片段的动作特征，金字塔的第 2 层（金字塔从下往上层数依次增加）包括候选片段中时域区间对应的视频片段的动作特征。作为本申请实施例金字塔式结构的一种特殊示例，金字塔式结构也可以仅包括一个层，该层包括整个候选片段的动作特征。

本申请实施例，提出了结构化的时域金字塔式池化方法，分别抽取时域区间前后和内部的时间-空间特征进行多层次时域金字塔式池化，可以构造任意长度的候选片段的结构化全局特征，表达时序上的动作特征，从而可以同时基于至少一个粒度的动作特征及动作特征在时间上的变化，更精确的定位发生某类动作类别动作实例的视频片段。

在其中一个可选的示例中，一种可能的配置规则可以包括：池化层数 K ，以及时域区间对应的视频片段、第一相邻片段和/或第二相邻片段在各层的分割部分数 B_K ，其中， K 、 B_K 的取值分别为大于或等于 1 的整数。时域区间对应的视频片段、第一相邻片段和第二相邻片段这三阶段的池化层数 K 可以相同或不同，在同一层的分割部分数 B_K 也可以相同或不同。

则在一种可选的示例中，预设池化层数 K 的取值为 1。相应地，对提取的候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理，包括：

对于预设分割部分数 B_K 的取值为 1 的任一第一待池化片段，由任一第一待池化片段中至少一个视频片段的动作特征获取任一第一待池化片段的池化特征；

对于预设分割部分数 B_K 的取值大于 1 的任一第二待池化片段，将任一第二待池化片段中的所有视频片段分割为 B_K 个部分，分别由 B_K 个部分中各部分视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，得到任一第二待池化片段的池化特征；

其中，第一待池化片段包括时域区间对应的视频片段、第一相邻片段和第二相邻片段中的任意一个或多个；第二待池化片段包括候选片段中第一待池化片段外的其他待池化片段。例如，若第一待池化片段包括时域区间对应的视频片段，则第二待池化片段包括第一相邻片段和第二相邻片段；若第一待池化片段包括时域区间对应的视频片段和第一相邻片段，则第二待池化片段包括第二相邻片段。

在第二种可能的示例中，预设池化层数 K 的取值大于 1。相应地，对提取的候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理，包括：

分别针对第 k 个池化层：对于预设分割部分数 B_K 的取值为 1 的任一第一待池化片段，由任一第一待池化片段中至少一个视频片段的动作特征获取任一第一待池化片段在第 k 层的池化特征；对于预设分割部分数 B_K 的取值大于 1 的任一第二待池化片段，将任一第二待池化片段中的所有视频片段分割为 B_K 个部分，分别由 B_K 个部分中各部分视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，得到任一第二待池化片段在第 k 层的池化特征；

分别针对至少一个待池化片段：按照预设方式对 K 个池化层的池化特征进行合并，得到至少一个待池化片段的池化特征；

其中，第一待池化片段包括时域区间对应的视频片段、第一相邻片段和第二相邻片段中的任意一个或多个；第二待池化片段包括候选片段中第一待池化片段外的其他待池化片段。

在基于上述第一种和第二种示例的另一个可选的示例中，对提取的候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理之后，还可以包括：合并时域金字塔式池化后得到的池化特征，来获得

时域区间对应的视频片段的全局特征。

可选地，可以分别对第一相邻片段的池化特征、时域区间对应的视频片段的池化特征和第二相邻片段的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征。

本申请任一实施例中，时序结构是指作的不同阶段构成组合所形成的时间关系上的结构。全局特征即整个候选片段的动作特征，其包含了比相应时域区间更全面的动作特征。

在一个可选的示例中，另一种可能的配置规则可以包括：池化层数 K 和候选片段的分割部分数 B_K ，其中， K 、 B_K 的取值分别为大于或等于 1 的整数。

则在第三种可能的示例中，预设池化层数 K 的取值大于 1。相应地，对候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理，包括：

分别针对第 k 个池化层：识别预设分割部分数 B_K 的取值是否为 1；

若预设分割部分数 B_K 的取值为 1，由候选片段中至少一个视频片段的动作特征获取候选片段在第 k 层的池化特征；

若预设分割部分数 B_K 的取值大于 1，将候选片段中的所有视频片段分割为 B_K 个部分，分别由 B_K 个部分中各部分视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，得到候选片段在第 k 层的池化特征。

在基于上述第三种示例的另一个可选的示例中，对提取的候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理之后，也可以包括：合并时域金字塔式池化后得到的池化特征，来获得时域区间对应的视频片段的全局特征。

可选地，可以按照预设方式对 K 个池化层的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征。

基于另一种可能的配置规则，在第四种可能的示例中，预设池化层数 K 的取值为 1。相应地，对提取的候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理，包括：

识别预设分割部分数 B_K 的取值是否为 1；

若预设分割部分数 B_K 的取值为 1，分别针对至少两个待池化片段中的任一待池化片段，由任一待池化片段中至少两个视频片段的动作特征获取任一待池化片段的池化特征，待池化片段包括时域区间内视频片段、第一相邻片段和第二相邻片段；对所有待池化片段的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征；

若预设分割部分数 B_K 的取值大于 1，将候选片段中的至少一个视频片段分割为 B_K 个部分，分别由 B_K 个部分中各部分中视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征。

本申请上述多个示例，示例性地提出了多种基于时域区间及其相邻区间的时间-空间特征进行时域金字塔式池化的方法，可以构造出针对任意长度片段、不同层次、不同粒度的全局特征，以便于根据实际需求和计算资源选择相应的时域金字塔式池化的方法来获得相应的全局特征，检测时域区间包含某类动作的完整性。

图 4 为本申请实施例中获得全局特征一个实施例的流程图。该实施例中的配置规则可以包括：池化层数 K ，时域区间内视频片段、第一相邻片段和/或第二相邻片段在各层的分割部分数 B_K ，其中， K 、 B_K 的取值分别为大于或等于 1 的整数。时域区间内视频片段、第一相邻片段和第二相邻片段这三阶段的池化层数 K 可以相同或不同，在同一层的分割部分数 B_K 也可以相同或不同。

在一个可选示例中，图 4 所示获得全局特征的操作可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的池化处理子模块执行。

如图 4 所示，该实施例中，对候选片段中至少两个视频片段的动作特征进行池化处理，以及合并时域金字塔式池化后得到的池化特征，获得上述时域区间对应的视频片段的全局特征，包括：

402，识别配置规则中的池化层数 K 的取值是否为 1。

若池化层数 K 的取值为 1，执行操作 404。否则，若池化层数 K 的取值大于 1，执行操作 410。

404，分别针对候选片段中的任一待池化片段，识别该任一待池化片段的分割部分数 B_K 的取值是否为 1。

其中，待池化片段包括时域区间对应的视频片段、第一相邻片段和第二相邻片段。

对于分割部分数 B_K 的取值为 1 的待池化片段，称为第一待池化片段，对于其中的任一第一待池化片段，执行操作 406。对于分割部分数 B_K 的取值大于 1 的待池化片段，称为第二待池化片段，对于其中的任一第二待池化片段，执行操作 408。

其中，第一待池化片段包括时域区间对应的视频片段、第一相邻片段和第二相邻片段中的任意一个或

多个；第二待池化片段包括候选片段中第一待池化片段外的其他待池化片段。

406，由任一第一待池化片段中至少一个视频片段的动作特征获取该任一第一待池化片段的池化特征。之后，执行操作 418。

408，将任一第二待池化片段中的所有视频片段分割为 B_K 个部分，分别由 B_K 个部分中各部分中视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，得到该任一第二待池化片段的池化特征。

之后，执行操作 418。

410，分别针对第 k 个池化层：识别至少一个待池化片段的分割部分数 B_K 的取值是否为 1。

其中的待池化片段包括时域区间对应的视频片段、第一相邻片段和第二相邻片段， k 的取值大于 0 且不大于 K 。

对于分割部分数 B_K 的取值为 1 的任一第一待池化片段，称为第一待池化片段，对于其中的任一第一待池化片段，执行操作 412。对于分割部分数 B_K 的取值大于 1 的待池化片段，称为第二待池化片段，对于其中的任一第二待池化片段，执行操作 414。

412，由该任一第一待池化片段中至少一个视频片段的动作特征获取该任一第一待池化片段在第 k 层的池化特征。

之后，执行操作 416。

414，将该任一第二待池化片段中的至少一个视频片段分割为 B_K 个部分，分别由该 B_K 个部分中各部分中视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，得到该任一第二待池化片段在第 k 层的池化特征。

416，分别针对至少一个待池化片段：按照预设方式对 K 个池化层的池化特征进行合并，得到至少一个待池化片段的池化特征。

418，分别对第一相邻片段的池化特征、时域区间对应的视频片段的池化特征和第二相邻片段的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征。

以下以一个应用为例，对本申请结构化的时域金字塔式池化方法进行进一步的详细说明，可以如下：

对于动作位置预测得到的任意一个时域区间（起始时间，结束时间），它的内部、之前、之后一共有三个阶段（即：本申请实施例中的时域区间内视频片段、第一相邻片段和第二相邻片段），每个阶段内包括若干小的视频片段，对每一个视频片段，经过一个用于动作分类的卷积神经网络进行特征提取，可以得到一个动作特征，可以体现为一组特征向量 $\mathbf{v}_t$ 。对时域区间内所有的视频片段 $s, \dots, e$ 的特征向量 $\mathbf{v}_s, \dots, \mathbf{v}_e$ ，第 k 层时域金字塔将这些视频片段 $s, \dots, e$ 作为一个整体等分成 B_K 个部分，每个部分的起止片段分别表示为 $s_{ki}, \dots, e_{ki}$ 。那么对于第 k 层第 i 个部分，求取 $\mathbf{v}_{s_{ki}} \dots \mathbf{v}_{e_{ki}}$ 的平均值（另外也可以按照预设规则去其最大值或最小值）作为用于表示该部分信息的池化向量，将这 B_K 个部分的池化向量按照视频片段的时序进行合并作为第 k 层池化向量；再对时域区间内视频片段在所有 K 层的池化向量按照视频片段的时序进行合并，例如以平均的方式合并，得到一个时域区间的池化向量。同样，也可以对第一相邻片段和第二相邻片段进行类似操作，最终获得的时序结构化、金字塔结构的池化向量（即：全局特征）是时域区间内视频片段、第一相邻片段和第二相邻片段这三个阶段的池化向量合并而成，其中不同阶段的池化层数 K 和分割部分数 B_K 的配置规则可以不同，例如，配置规则为由金字塔的第 1 层至更高层，每层的分割部分数依次增多。

图 5 为本申请实施例中获得全局特征另一个实施例的流程图。

在一个可选示例中，图 5 所示获得全局特征的操作可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的池化处理子模块执行。

该实施例中，配置规则可以包括：池化层数 K 和候选片段的分割部分数 B_K ，其中， K 、 B_K 的取值分别为大于或等于 1 的整数。如图 5 所示，该实施例包括：

502，识别配置规则中的池化层数 K 的取值是否为 1。

若池化层数 K 的取值为 1，执行操作 504。否则，若池化层数 K 的取值大于 1，执行操作 510。

504，识别分割部分数 B_K 的取值是否为 1。

若分割部分数 B_K 的取值 1，执行操作 506。若分割部分数 B_K 的取值大于 1，执行操作 508。

506，分别针对至少一个待池化片段中的任一待池化片段，由该任一待池化片段中至少一个视频片段的动作特征获取该任一待池化片段的池化特征，并对至少一个待池化片段的池化特征进行合并，获得上述时域区间对应的视频片段时序结构化的全局特征。

其中，待池化片段包括时域区间对应的视频片段、第一相邻片段和第二相邻片段。

之后，不执行本实施例的后续流程。

508，将候选片段中的至少一个视频片段分割为 B_K 个部分，例如，可以将候选片段中的所有视频片段平均分割为 B_K 个部分，分别由 B_K 个部分中各部分中视频片段的动作特征获取对应部分的池化特征，并对该 B_K 个部分的池化特征进行合并，得到上述时域区间对应片段时序结构化的全局特征。

之后，不执行本实施例的后续流程。

510，分别针对第 k 个池化层：识别分割部分数 B_K 的取值是否为1，其中， k 的取值大于0且不大于 K 。

若分割部分数 B_K 的取值为1，执行操作512。若分割部分数 B_K 大于1，执行操作514。

512，由候选片段中至少两个视频片段的动作特征获取该候选片段在第 k 层的池化特征。

之后，执行操作516。

514，将候选片段中的所有视频片段分割为 B_K 个部分，分别由 B_K 个部分中各部分中视频片段的动作特征获取对应部分的池化特征，并对该 B_K 个部分的池化特征进行合并，得到候选片段在第 k 层的池化特征。

516，按照预设方式对 K 个池化层的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征。

以下以一个应用为例，对本申请结构化的时域金字塔式池化方法进行进一步的详细说明，可以如下。该应用示例中， K 的取值为2， B_K 的取值为2：

对一个候选片段，金字塔的第1层（即：最底层）为整个候选片段，先处理第1层，即：对整个候选片段内至少两个视频片段的特征向量进行平均合并，得到一个池化向量；再处理第2层，即：将该候选片段分为前后50%两部分，在每个部分内部分别进行特征向量的平均合并，得到两组池化向量；

然后将金字塔两层的所有池化向量，共3组，按照时序关系进行拼接，得到一个更高维度的特征向量，即该时域区间对应的视频片段的全局特征。这里拼接向量指将两个池化向量首尾相接，得到一个维度更高的池化向量。

图6为本申请实施例中确定时域区间对应的视频片段的完整度得分一个实施例的流程图。如图6所示，该实施例中，基于全局特征确定时域区间对应的视频片段的动作完整度得分，可以包括：

602，对全局特征进行分类，得到上述候选片段对应的至少一个动作类别。

对 N 种可能的动作类别，分类结果将得到 $N+1$ 个动作类别得分，其中的第 $N+1$ 类为背景类。其中， N 为大于零的整数。

在一个可选示例中，该操作602可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的分类子模块执行。

604，分别通过上述候选片段对应的至少一个动作类别相关的分类器，基于全局特征识别上述时域区间对应的视频片段包括当前动作类别完整的动作实例的程度，得到上述时域区间对应的视频片段分别对应至少一个动作类别的动作完整度得分。

即：对 N 种可能的动作类别，就对应 N 个分类器，每个分类器将接收到的全局特征与预先学习的对应动作类别完整动作的特征进行比对，输出一个得分，即动作完整度。该动作完整度得分表示一个时域区间包括某类动作类别完整的动作实例的程度（也即：可能性）。

在一个可选示例中，该操作604可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的至少一个分类器执行。

图7为本申请时域动作检测方法又一个实施例的流程图。如图7所示，该实施例的时域动作检测方法包括：

702，分别对视频中的至少两个视频片段进行动作度估计，获得时序动作度序列。

在一个可选示例中，该操作702可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的动作估计子模块执行。

704，基于上述时序动作度序列进行动作位置预测，获得视频中可能存在动作实例的时域区间，该时域区间包括起始时间和结束时间，例如可以表示为起始时间和结束时间对（起始时间，结束时间）。

在一个可选示例中，该操作704可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的位置预测子模块执行。

根据动作位置预测结果，分别以至少一个动作实例的时域区间作为当前时域区间，执行操作706~728。

706，提取当前时域区间内所有视频片段，以及当前时域区间之前预设长度的第一相邻片段和之后预设长度的第二相邻片段，同一动作实例的时域区间内所有视频片段、第一相邻片段和第二相邻片段构成同一动作实例的候选片段。

其中，第一相邻片段和第二相邻片段分别包括至少一个视频片段。

在一个可选示例中，该操作706可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运

行的提取子模块执行。

708, 以当前时域区间对应的候选片段作为当前候选片段, 分别提取当前候选片段中至少两个视频片段的动作特征。

之后, 同时执行操作 710 和 714。

710, 分别获取当前时域区间内至少一个视频片段分别对应至少一个动作类别的得分。

712, 分别对当前时域区间内所有视频片段对应同一动作类别的得分求和, 获得当前时域区间对应的视频片段对应至少一个动作类别的分类得分。

在一个可选示例中, 该操作 708~712 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的第二获取模块执行。

714, 基于当前时域区间对应的视频片段的动作特征、第一相邻片段的动作特征和第二相邻片段的动作特征, 获取当前时域区间对应的视频片段时序结构化的全局特征。

在一个可选示例中, 该操作 714 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的合并子模块执行。

716, 对全局特征进行分类, 得到当前候选片段对应的至少一个动作类别。

在一个可选示例中, 该操作 716 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的分类子模块执行。

718, 通过当前候选片段对应的至少一个动作类别相关的分类器, 基于全局特征识别当前时域区间对应的视频片段包括当前动作类别的完整动作实例的可能性, 得到当前时域区间对应的视频片段分别对应至少一个动作类别的动作完整度得分。

其中, 每一个动作类别对应一个分类器, 即: 该动作类别相关的分类器。

在一个可选示例中, 该操作 718 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的至少一个分类器执行。

分别针对当前时域区间对应的视频片段和当前候选片段对应的所有动作类别中的至少一个动作类别, 执行操作 720~724。

720, 基于时域区间对应的视频片段对应至少一个动作类别的分类得分与动作完整度得分, 获取当前时域区间对应的视频片段是否包括对应至少一个动作类别完整的动作实例的置信度值。

示例性地, 该操作 720 中, 可以以当前时域区间对应的视频片段对应的至少一个动作类别作为当前动作类别, 对当前时域区间对应视频片段对应当前动作类别的分类得分与动作完整度得分求取乘积或者求和, 获取当前时域区间对应的视频片段是否包括当前动作类别完整的动作实例的置信度值。

在一个可选示例中, 该操作 720 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的第二计算子模块执行。

722, 分别针对当前时域区间对应的视频片段和当前候选片段对应的所有动作类别中的至少一个动作类别, 基于获取到的置信度值是否大于预设阈值, 确定当前时域区间对应的视频片段是否包含对应动作类别完整的动作实例。

本申请实施例中的预设阈值在实际应用中可以根据对检测结果的质量要求设置, 并且可以根据需要调整。

在一个可选示例中, 该操作 722 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的确定子模块执行。

若当前时域区间对应的视频片段包含对应动作类别完整的动作实例, 执行操作 726。否则, 不执行本实施例的后续流程。

724, 对于包含对应动作类别完整的动作实例的视频片段作为目标视频片段, 获取目标视频片段的动作位置检测结果。

在一个可选示例中, 该操作 724 可以由处理器调用存储器存储的相应指令执行, 也可以由被处理器运行的确定子模块执行。

其中, 目标视频片段的动作位置检测结果包括完整的动作实例对应的检测动作类别、该动作实例的起始时间和结束时间, 可以表示为三元组(动作类别, 起始时间, 结束时间), 视频的动作位置检测结果包括至少一个目标视频片段的动作位置检测结果。

可选地, 若通过操作 722, 目标视频片段包括完整的动作实例对应的动作类别大于一个, 则可以选择对应于最大置信度值的一个动作实例对应的检测动作类别、该动作实例的起始时间和结束时间作为目标视频片段的动作位置检测结果。

726, 分别针对至少一个目标视频片段包括的至少一个完整的动作实例: 基于该目标视频片段的全局

特征，对目标视频片段的动作位置检测结果中的时域区间进行位置回归，获得起止位置偏移量，该起止位置偏移量包括起始时间的位置偏移量和结束时间的位置偏移量。

在一个可选示例中，该操作 726 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的位置回归模块执行。

728，通过起止位置偏移量对动作位置检测结果中的时域区间进行调整，获得调整后的动作位置检测结果，可以表示为三元组（动作类别，起始时间，结束时间）。

在一个可选示例中，该操作 728 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的输出模块执行。

可选地，对一个视频进行时域动作检测时，可以在获得一个时域区间的动作位置检测结果时，即实时输出该时域区间的动作位置检测结果；也可以在获得整个视频的动作位置检测结果后，输出整个视频的动作位置检测结果，其可能包括一个或多个时域区间的动作位置检测结果。

示例性地，可以将目标视频片段的全局特征输入一预先训练好的线性回归模型，由该线性回归模型预测得到位置偏移量，并通过起止位置偏移量对动作位置检测结果中的时域区间进行调整，从而获得更加准确的动作位置检测结果，以便从整个视频中得到与实际动作实例的起止时间更接近的动作实例视频片段。

本申请实施例中，可以预先通过样本视频对初始线性回归模型进行训练，使得训练好的线性回归模型接收到视频、候选片段的全局特征及其包括的时域区间后，可以基于全局特征精确的定位出某动作类别完整的动作实例的准确时域区间，从而输出输入的时域区间中起始时间的位置偏移量和结束时间的位置偏移量。输入的时域区间与准确时域区间一致时，起始时间的位置偏移量和结束时间的位置偏移量均为 0。其中的样本视频包含某动作类别完整的动作实例、以及该类动作类别完整的动作实例精确的时域区间。

本申请实施例通过位置回归对视频中的动作起止时间（即：时域区间）进行更精细的调整，进一步提升了在高交并比（Intersection-over-Union, IOU）下的检测准确率。

基于本申请上述实施例提供的时域动作检测方法，可以识别出视频中包括至少一个动作类别或特定动作类别完整的动作实例的视频片段，实现对至少一个动作类别或特定动作类别完整的动作实例的准确识别和精确定位；并且，可以同时获取时域区间对应的视频片段属于至少一个动作类别的得分和动作完整度得分，提高了整个时域动作检测过程的执行效率，节省检测时间。本申请实施例可以应用于对各种长度的视频（包括剪辑出的视频和未经剪辑的视频）的检测，本申请实施例应用于检测未经剪辑、较长的视频时，对视频中动作类别的识别和位置的精确定位效果尤其显著。

本申请实施例的时域动作检测方法，可用于对视频进行各种预设特定动作类别（可称为：目标动作类别）动作实例的检测，例如，应用于车辆驾驶场景中通过监控视频监控司机是否存在疲劳驾驶行为（如：点头、闭眼等动作）时，可以设定点头、闭眼等疲劳驾驶行为动作分别为一个目标动作类别，将监控视频作为视频，基于本申请实施例的时域动作检测方法检测监控视频中是否存在完整的点头、闭眼等目标动作类别的完整动作，每个动作分别称为一个动作实例。若监控视频中存在一个完整的点头、闭眼等目标动作类别的完整动作，则确认司机发生了疲劳驾驶行为。

本申请上述实施例中，操作 706~722 可以通过一个预先训练好的神经网络实现。

示例性地，可以通过样本视频片段对初始神经网络模型进行训练，使得神经网络模型输出的动作位置检测结果与样本视频片段预先标注的参考动作位置信息（动作类别，起始时间，结束时间）之间的差别小于预设条件，从而获得训练好的神经网络，其中的样本视频片段预先标注有参考动作位置信息（动作类别，起始时间，结束时间）。

本申请实施例的时域动作检测方法应用于各种应用场景时，基于时域区间对应的视频片段的分类得分与动作完整度得分，确定该时域区间对应的视频片段是否包含目标动作类别完整的动作实例，可以通过如下方式实现：识别时域区间对应的视频片段对应至少一个动作类别的分类得分中，是否存在分类得分大于预设第一数值的预设动作类别（此时认为时域区间对应的视频片段属于该预设动作类别）；以及识别时域区间对应的视频片段对应至少一个动作类别的动作完整度得分中，是否存在动作完整度得分大于预设第二数值的预设动作类别（此时认为时域区间对应的视频片段包含了该预设动作类别的完整动作实例）；

响应于存在分类得分大于预设第一数值的预设动作类别和/或动作完整度得分大于预设第二数值的预设动作类别，以该分类得分大于预设第一数值的预设动作类别和/或动作完整度得分大于预设第二数值的预设动作类别作为目标动作类别，基于该时域区间对应的视频片段对应的目标动作类别的分类得分与动作完整度得分，确定该时域区间对应的视频片段是否包含目标动作类别完整的动作实例。

相应地，该应用实施例中，输出的动作位置检测结果可以包括：时域区间对应的视频片段包含完整的动作实例所属的目标动作类别、起始时间和结束时间。

本申请实施例提供的任一种时域动作检测方法可以由任意适当的具有数据处理能力的设备执行，包括

但不限于：终端设备和服务器等。或者，本申请实施例提供的任一种时域动作检测方法可以由处理器执行，如处理器通过调用存储器存储的相应指令来执行本申请实施例提及的任一种时域动作检测方法。下文不再赘述。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

图8为本申请时域动作检测系统一个实施例的结构示意图。本申请各实施例的时域动作检测系统可用于实现上述各实施例的时域动作检测方法。如图8所示，该实施例的时域动作检测系统包括：第一获取模块，特征提取模块，池化处理模块和第一确定模块。其中：

第一获取模块，用于获取视频中存在动作实例的时域区间以及所述时域区间的至少一个相邻片段。

其中，上述时域区间的至少一个相邻片段包括：视频中时序位于该时域区间之前的第一相邻片段，和/或，视频中时序位于该时域区间之后的第二相邻片段。第一相邻片段和第二相邻片段分别包括至少一个视频片段。

特征提取模块，用于分别提取候选片段中至少两个视频片段的动作特征，其中，候选片段包括时域区间对应的视频片段及其相邻片段。

池化处理模块，用于对候选片段中至少两个视频片段的动作特征进行池化处理，获得上述时域区间对应的视频片段的全局特征。

第一确定模块，用于基于上述全局特征确定上述时域区间对应的视频片段的动作完整度得分。

基于本申请上述实施例提供的时域动作检测系统，通过获取视频中存在动作实例的时域区间以及时域区间的至少一个相邻片段，提取该时域区间以及时域区间的至少一个相邻片段的动作特征并进行池化处理，获得该时域区间对应的视频片段的全局特征，基于全局特征确定时域区间对应的视频片段的动作完整度得分。本申请实施例基于包含时域区间的上下文信息的全局特征来确定时域区间对应的视频片段的动作完整度，从而有益于准确的确定时域区间是否包含完整的动作实例，提高了动作完整度识别的准确率。

图9为本申请实施例中第一获取模块一个具体实施例的结构示意图。如图9所示，该实施例的第一获取模块可以包括：

动作估计子模块，用于分别对视频中的至少一个视频片段进行动作度估计，获得时序动作度序列。

位置预测子模块，用于基于上述时序动作度序列进行动作位置预测，获得视频中存在动作实例的时域区间，其中的时域区间包括起始时间和结束时间。

提取子模块，用于从视频中提取上述时域区间之前的第一相邻片段和/或所述时域区间之后的第二相邻片段。

示例性地，动作估计子模块可以包括：

图像获取单元，用于分别针对视频中的任一视频片段：提取一帧图像作为原始图像，以及提取任一视频片段的光流，并对获得的光流场图片进行合并，得到拼合光流场图像；

第一子卷积神经网络，用于对原始图像进行动作度估计，得到第一动作度值；

第二子卷积神经网络，用于对拼合光流场图像进行动作度估计，得到第二动作度值；

计算单元，用于基于同一视频片段的第一动作度值与第二动作度值，确定任一视频片段的动作度值。视频中所有视频片段基于时序关系的动作度值形成时序动作度序列。

进一步地，在另一个可选的示例中，动作估计子模块还可以包括：归一化单元，用于对计算单元获取的视频片段的动作度值进行归一化处理，得到归一化的动作度值。相应地，该实施例中，时序动作度序列可以包括：由视频中所有视频片段归一化的动作度值形成的时序动作度序列。

图10为本申请时域动作检测系统另一个实施例的结构示意图。如图10所示，与上述实施例的时域动作检测系统相比，本实施例的时域动作检测系统还包括：第二获取模块和第二确定模块。其中：

第二获取模块，用于基于时域区间对应的视频片段的动作特征，获取该时域区间对应的视频片段的至少一个动作类别的分类得分。

第二确定模块，用于根据时域区间对应的视频片段的至少一个动作类别的分类得分，确定该时域区间对应的视频片段的检测动作类别。

进一步地，再参见图10，在又一个实施例中，时域动作检测系统还可以包括：输出模块，用于输出时域区间以及时域区间对应的视频片段的检测动作类别。

图11为本申请实施例中第二获取模块一个实施例的结构示意图。如图11所示，该实施例中，第二获取模块可以包括：

第一获取子模块，用于分别基于时域区间对应的至少一个视频片段的动作特征，获取该时域区间对应

的至少一个视频片段分别属于至少一个动作类别的得分；

第一计算子模块，用于分别对上述时域区间对应的至少一个视频片段属于同一动作类别的得分求和，获得该时域区间对应的视频片段的至少一个动作类别的分类得分。

图 12 为本申请实施例中池化模块一个实施例的结构示意图。如图 12 所示，该实施例中，池化模块可以包括：池化处理子模块，用于基于预先设置的配置规则，对提取的候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理。

在其中一种配置规则中，包括：池化层数 K 和候选片段的分割部分数 B_K ，其中， K 、 B_K 的取值分别为大于或等于 1 的整数。

基于该配置规则，在其中一个可选的示例中，池化层数 K 的取值为 1。池化处理子模块可以用于：

识别预设分割部分数 B_K 的取值是否为 1；

若预设分割部分数 B_K 的取值为 1，分别针对至少一个待池化片段中的任一待池化片段，由任一待池化片段中各视频片段的动作特征获取任一待池化片段的池化特征，待池化片段包括时域区间内视频片段、第一相邻片段和第二相邻片段；对所有待池化片段的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征；以及

若预设分割部分数 B_K 的取值大于 1，将候选片段中的所有视频片段分割为 B_K 个部分，分别由 B_K 个部分中各部分视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征

可选地，再参见图 12，在池化模块的另一个实施例中，还可以选择性地包括合并子模块，用于合并时域金字塔式池化后得到的池化特征。

在一个可选的示例中，配置规则可以包括：池化层数 K 和候选片段的分割部分数 B_K ，其中， K 池化层数 K 的取值大于 1，分割部分数 B_K 的取值分别为大于或等于 1 的整数。

该可选的示例中，池化处理子模块可以用于：分别针对第 k 个池化层：识别预设分割部分数 B_K 的取值是否为 1；若预设分割部分数 B_K 的取值为 1，由候选片段中至少一个视频片段的动作特征获取候选片段在第 k 层的池化特征；若预设分割部分数 B_K 的取值大于 1，将候选片段中的所有视频片段分割为 B_K 个部分，分别由 B_K 个部分的各部分视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，得到候选片段在第 k 层的池化特征；

合并子模块，用于按照预设方式对 K 个池化层的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征。

在一个可选的示例中，配置规则可以包括：池化层数 K ，以及时域区间对应的视频片段、第一相邻片段和/或第二相邻片段在至少一层的分割部分数 B_K ，其中，设池化层数 K 的取值为 1，分割部分数 B_K 的取值分别为大于或等于 1 的整数。

则在该可选示例中，池化处理子模块可以用于：对于预设分割部分数 B_K 的取值为 1 的任一第一待池化片段，由任一第一待池化片段中至少一个视频片段的动作特征获取任一第一待池化片段的池化特征；对于预设分割部分数 B_K 的取值大于 1 的任一第二待池化片段，将任一第二待池化片段中的所有视频片段分割为 B_K 个部分，分别由至少一个部分中视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，得到任一第二待池化片段的池化特征；第一待池化片段包括时域区间对应的视频片段、第一相邻片段和第二相邻片段中的任意一个或多个；第二待池化片段包括候选片段中第一待池化片段外的其他待池化片段；

合并子模块，可以用于分别对第一相邻片段的池化特征、时域区间对应的视频片段的池化特征和第二相邻片段的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征。

在一个可选的示例中，配置规则可以包括：池化层数 K ，以及时域区间对应的视频片段、第一相邻片段和/或第二相邻片段在至少一层的分割部分数 B_K ，其中，设池化层数 K 的取值大于 1，分割部分数 B_K 的取值分别为大于或等于 1 的整数。

则在该可选示例中，池化处理子模块可以用于：分别针对第 k 个池化层：对于预设分割部分数 B_K 的取值为 1 的任一第一待池化片段，由任一第一待池化片段中至少一个视频片段的动作特征获取任一第一待池化片段在第 k 层的池化特征；对于预设分割部分数 B_K 的取值大于 1 的任一第二待池化片段，将任一第二待池化片段中的所有视频片段分割为 B_K 个部分，分别由至少一个部分中视频片段的动作特征获取对应部分的池化特征，并对 B_K 个部分的池化特征进行合并，得到任一第二待池化片段在第 k 层的池化特征；以及分别针对至少两个待池化片段：按照预设方式对 K 个池化层的池化特征进行合并，得到至少两个待池化片段的池化特征；第一待池化片段包括时域区间对应的视频片段、第一相邻片段和第二相邻片段中的任意一个或多个；第二待池化片段包括候选片段中第一待池化片段外的其他待池化片段；

合并子模块，可以用于分别对第一相邻片段的池化特征、时域区间对应的视频片段的池化特征和第二相邻片段的池化特征进行合并，获得时域区间对应的视频片段时序结构化的全局特征。

图 13 为本申请实施例中第一确定模块一个实施例的结构示意图。如图 13 所示，该实施例中，第一确定模块可以包括：

分类子模块，用于对全局特征进行分类，得到候选片段对应的至少一个动作类别；

至少一个分类器，分别与候选片段对应的至少一个动作类别相关，用于基于全局特征识别时域区间对应的视频片段包括当前动作类别完整的动作实例的程度，得到时域区间对应的视频片段分别对应至少一个动作类别的动作完整度得分。

可选地，再参见图 10，在时域动作检测系统的又一实施例中，还可以包括：第三确定模块，用于基于时域区间对应的视频片段的分类得分与动作完整度得分，确定时域区间对应的视频片段是否包含对应的至少一个动作类别完整的动作实例。

图 14 为本申请实施例中第三确定模块一个实施例的结构示意图。如图 14 所示，该实施例中，第三确定模块可以包括：

第二计算子模块，用于分别针对时域区间对应的视频片段对应的任一动作类别，对时域区间对应的视频片段对应任一动作类别的分类得分与动作完整度得分求取乘积或者求和，获得时域区间对应的视频片段是否包含任一动作类别完整的动作实例的置信度值；

确定子模块，用于基于置信度值是否大于预设阈值，确定时域区间对应的视频片段是否包含任一动作类别完整的动作实例。

可选地，确定子模块，还可用于在时域区间对应的视频片段包含完整的动作实例对应的动作类别大于一个时，选择对应于最大置信度值的一个动作类别作为时域区间对应的视频片段的检测动作类别。

进一步地，再参见图 10，在时域动作检测系统的再一实施例中，还可以包括：位置回归模块，用于在时域区间对应的视频片段包含至少一个动作类别完整的动作实例时，基于全局特征对时域区间进行位置回归，获得起止位置偏移量，并通过起止位置偏移量对时域区间进行调整，其中的起止位置偏移量包括时域区间中起始时间和结束时间的位置偏移量。

相应地，该实施例中，输出模块，可以用于输出调整后的时域区间以及时域区间对应的视频片段的检测动作类别。

图 15 为本申请时域动作检测系统又一实施例的结构示意图。如图 15 所示，与上述各实施例相比，该实施例的时域动作检测系统还包括：识别模块，用于识别时域区间对应的视频片段的至少一个动作类别的分类得分中，是否存在分类得分大于预设第一数值的预设动作类别；以及识别动作完整度得分中，是否存在动作完整度得分大于预设第二数值的预设动作类别。

相应地，该实施例中，第三确定模块，可以用于在存在分类得分大于预设第一数值的预设动作类别和/或动作完整度得分大于预设第二数值的预设动作类别时，以分类得分大于预设第一数值的预设动作类别和/或动作完整度得分大于预设第二数值的预设动作类别作为目标动作类别，基于时域区间对应的视频片段对应的目标动作类别的分类得分与动作完整度得分，确定时域区间对应的视频片段是否包含目标动作类别完整的动作实例；

输出模块，可以用于输出目标动作类别及其时域区间。

另外，本申请实施例中还提供了一种电子设备，包括本申请上述任一实施例的时域动作检测系统。

此外，本申请实施例中还提供了另一种电子设备，包括：

处理器和存储器；

所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行上述任一实施例的时域动作检测方法对应的操作。

本申请各实施例中的电子设备，例如可以是移动终端、个人计算机（PC）、平板电脑、服务器等。

图 16 为本申请电子设备一个应用实施例的结构示意图。图 16 示出了适于用来实现本申请实施例的终端设备或服务器的电子设备的结构示意图。如图 16 所示，该电子设备包括一个或多个处理器、通信部等，所述一个或多个处理器例如：一个或多个中央处理单元（CPU），和/或 一个或多个图像处理器（GPU）等，处理器可以根据存储在只读存储器（ROM）中的可执行指令或者从存储部分加载到随机访问存储器（RAM）中的可执行指令而执行各种适当的动作和处理。通信部可包括但不限于网卡，所述网卡可包括但不限于 IB（Infiniband）网卡，处理器可与只读存储器和/或随机访问存储器中通信以执行可执行指令，通过总线与通信部相连、并经通信部与其他目标设备通信，从而完成本申请实施例提供的任一项方法对应的操作，例如，获取视频中存在动作实例的时域区间以及时域区间的至少一个相邻片段；分别提取候选片段中至少两个视频片段的动作特征，其中，候选片段包括时域区间对应的视频片段及其相邻片段；对候选片段中至少

两个视频片段的动作特征进行池化处理，获得时域区间对应的视频片段的全局特征；基于全局特征确定时域区间对应的视频片段的动作完整度得分。

此外，在 RAM 中，还可存储有装置操作所需的各种程序和数据。CPU、ROM 以及 RAM 通过总线彼此相连。在有 RAM 的情况下，ROM 为可选模块。RAM 存储可执行指令，或在运行时向 ROM 中写入可执行指令，可执行指令使处理器执行上述通信方法对应的操作。输入/输出（I/O）接口也连接至总线。通信部可以集成设置，也可以设置为具有多个子模块（例如多个 IB 网卡），并在总线链接上。

以下部件连接至 I/O 接口：包括键盘、鼠标等的输入部分；包括诸如阴极射线管（CRT）、液晶显示器（LCD）等以及扬声器等的输出部分；包括硬盘等的存储部分；以及包括诸如 LAN 卡、调制解调器等等的网络接口卡的通信部分。通信部分经由诸如因特网的网络执行通信处理。驱动器也根据需要连接至 I/O 接口。可拆卸介质，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器上，以便于从其上读出的计算机程序根据需要被安装入存储部分。

需要说明的，如图 16 所示的架构仅为一种可选实现方式，在实践过程中，可根据实际需要对上述图 16 的部件数量和类型进行选择、删减、增加或替换；在不同功能部件设置上，也可采用分离设置或集成设置等实现方式，例如 GPU 和 CPU 可分离设置或者可将 GPU 集成在 CPU 上，通信部可分离设置，也可集成设置在 CPU 或 GPU 上，等等。这些可替换的实施方式均落入本申请公开的保护范围。

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，其包括有形地包含在机器可读介质上的计算机程序，计算机程序包含用于执行流程图所示的方法的程序代码，程序代码可包括对应执行本申请实施例提供的方法步骤对应的指令，例如，获取视频中存在动作实例的时域区间以及时域区间的至少一个相邻片段的指令；分别提取候选片段中至少两个视频片段的动作特征的指令，其中，候选片段包括时域区间对应的视频片段及其相邻片段；对候选片段中至少两个视频片段的动作特征进行池化处理的指令，获得时域区间对应的视频片段的全局特征的指令；基于全局特征确定时域区间对应的视频片段的动作完整度得分的指令。在这样的实施例中，该计算机程序可以通过通信部分从网络上被下载和安装，和/或从可拆卸介质被安装。在该计算机程序被中央处理单元（CPU）执行时，执行本申请的方法中限定的上述功能。

另外，本申请实施例还提供了一种计算机程序，包括计算机可读代码，当所述计算机可读代码在设备上运行时，所述设备中的处理器执行用于实现本申请任一实施例所述的时域动作检测方法中各步骤的指令。

另外，本申请实施例还提供了一种计算机存储介质，用于存储计算机可读的指令，该指令被执行时实现本申请任一实施例所述的时域动作检测方法中各步骤的操作。

基于本申请上述各实施例的电子设备和计算机存储介质，基于包含时域区间的上下文信息的全局特征来确定时域区间对应的视频片段的动作完整度，从而有益于准确的确定时域区间是否包含完整的动作实例，提高了动作完整度识别的准确率。

本申请实施例，例如可以应用于处理如下场景：

场景 1：视频分享网站可以分析用户上传的视频，基于视频中包含的完整动作实例所属的动作类别，自动生成视频内容的相关标签（例如，跑步、跳高、竞走、撑杆跳、跳远及三级跳等体育运动类的动作类别标签，以及日常生活活动等动作类别标签），将视频中最相关的片段（即：包含该动作类别完整的动作实例的视频片段）而不是冗长的整个视频推送给其他感兴趣的潜在用户，促进视频的高效搜索和快速传播。

场景 2：用户视频分享网站或视频直播网站，需要分析用户上传的视频或直播视频中是否有违规的动作 / 行为，可基于本申请实施例提供的时域动作检测方法对用户上传的视频或直播视频进行检测，如存在违规的动作 / 行为，有益于准确定位这些动作/行为发生的时间及其动作类别。

场景 3：监控场景，在医院、护理所乃至家中，基于本申请实施例提供的时域动作检测方法，通过监控视频对照顾病人、长者或儿童的服务人员进行监管，以便出现虐待行为时快速做出反应。

场景 4：在机器人 / 自动驾驶汽车与使用者的交互过程中，需要检测某些特定动作的发生或结束来引导机器人的特定行为，这就基于本申请实施例提供的时域动作检测方法，准确检测监控视频中的某些特定动作（即：特定动作类别的动作）、精确定位这些特定动作的开始时间和结束时间。

本申请实施例具有以下总的有益技术效果：

可以识别出待检测视频中包括各动作类别或特定动作类别完整的动作实例的视频片段，实现对各动作类别或特定动作类别完整的动作实例的准确识别和精确定位；

可以同时获取时域区间对应的视频片段属于各动作类别的分类得分和动作完整度得分，提高了整个时域动作检测过程的执行效率，节省检测时间；

提出了结构化的时域金字塔池化方法，分别抽取时域区间前后和内部的时间-空间特征进行多层次时域

金字塔池化，可以构造任意长度的候选片段的结构化全局特征，表达时序上的动作特征，从而可以同时基于各粒度的动作特征及动作特征在时间上的变化，更精确的定位发生某类动作类别动作实例的视频片段；

通过位置回归对待检测视频的动作起止时间进行更精细的调整，进一步提升了在高交叠率（IOU）下的检测准确率。

本说明书中各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似的部分相互参见即可。对于系统实施例而言，由于其与方法实施例基本对应，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

可能以许多方式来实现本申请的方法和装置。例如，可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本申请的方法和装置。用于所述方法的步骤的上述顺序仅是为了进行说明，本申请的方法的步骤不限于以上具体描述的顺序，除非以其它方式特别说明。此外，在一些实施例中，还可将本申请实施为记录在记录介质中的程序，这些程序包括用于实现根据本申请的方法的机器可读指令。因而，本申请还覆盖存储用于执行根据本申请的方法的程序的记录介质。

本申请的描述是为了示例和描述起见而给出的，而并不是无遗漏的或者将本申请限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施例是为了更好说明本申请的原理和实际应用，并且使本领域的普通技术人员能够理解本申请从而设计适于特定用途的带有各种修改的各种实施例。

权利要求

1、一种时域动作检测方法，其特征在于，包括：

获取视频中存在动作实例的时域区间以及所述时域区间的至少一个相邻片段；

分别提取候选片段中至少两个视频片段的动作特征，其中，所述候选片段包括所述时域区间对应的视频片段及其相邻片段；

对所述候选片段中至少两个视频片段的动作特征进行池化处理，获得所述时域区间对应的视频片段的全局特征；

基于所述全局特征确定所述时域区间对应的视频片段的动作完整度得分。

2、根据权利要求1所述的方法，其特征在于，所述至少一个相邻片段包括：所述视频中时序位于所述时域区间之前的第一相邻片段，和/或，所述视频中时序位于所述时域区间之后的第二相邻片段；所述第一相邻片段和所述第二相邻片段分别包括至少一个视频片段。

3、根据权利要求1或2所述的方法，其特征在于，所述获取视频中存在动作实例的时域区间以及所述时域区间的至少一个相邻片段，包括：

分别对所述视频中的至少一个视频片段进行动作度估计，获得时序动作度序列；

基于所述时序动作度序列进行动作位置预测，获得所述视频中存在动作实例的时域区间，所述时域区间包括起始时间和结束时间；

从所述视频中提取所述时域区间之前的第一相邻片段和/或所述时域区间之后的第二相邻片段。

4、根据权利要求3所述的方法，其特征在于，所述分别对所述视频中的至少一个视频片段进行动作度估计，获得时序动作度序列，包括：

分别针对所述视频中的任一视频片段：提取一帧图像作为原始图像，对所述原始图像进行动作度估计，得到第一动作度值；提取所述任一视频片段的光流，并对获得的光流场图片进行合并，得到拼合光流场图像，对所述拼合光流场图像进行动作度估计，得到第二动作度值；

由所述第一动作度值与所述第二动作度值，获取所述任一视频片段的动作度值；

所述视频中所有视频片段基于时序关系的动作度值形成所述时序动作度序列。

5、根据权利要求4所述的方法，其特征在于，所述获取所述任一视频片段的动作度值之后，还包括：

对所述任一视频片段的动作度值进行归一化处理，得到归一化的动作度值；

所述时序动作度序列包括：由所述归一化的动作度值形成的时序动作度序列。

6、根据权利要求1至5任意一项所述的方法，其特征在于，还包括：

基于所述时域区间对应的视频片段的动作特征，获取所述时域区间对应的视频片段的至少一个动作类别的分类得分；

根据所述时域区间对应的视频片段的至少一个动作类别的分类得分，确定所述时域区间对应的视频片段的检测动作类别。

7、根据权利要求6所述的方法，其特征在于，还包括：

输出所述时域区间以及所述时域区间对应的视频片段的检测动作类别。

8、根据权利要求6或7所述的方法，其特征在于，所述基于所述时域区间对应的视频片段的动作特征，获取所述时域区间对应的视频片段的至少一个动作类别的分类得分，包括：

分别基于所述时域区间对应的至少一个视频片段的动作特征，获取所述时域区间对应的至少一个视频片段分别属于所述至少一个动作类别的得分；

分别对所述时域区间对应的所述至少一个视频片段属于同一动作类别的得分求和，获得所述时域区间对应的视频片段的至少一个动作类别的分类得分。

9、根据权利要求1至8任意一项所述的方法，其特征在于，所述对所述候选视频片段中至少两个视频片段的动作特征进行池化处理，包括：

对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理。

10、根据权利要求9所述的方法，其特征在于，所述对提取的候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理之后，还包括：

合并时域金字塔式池化后得到的池化特征。

11、根据权利要求10所述的方法，其特征在于，所述对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理之前，还包括：

预设池化层数K的取值为1；

所述对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理，包括：

对于预设分割部分数 B_K 的取值为1的任一第一待池化片段，由所述任一第一待池化片段中至少一个视

频片段的动作特征获取所述任一第一待池化片段的池化特征;

对于预设分割部分数 B_K 的取值大于1的任一第二待池化片段,将所述任一第二待池化片段中的所有视频片段分割为 B_K 个部分,分别由所述 B_K 个部分中各部分视频片段的动作特征获取对应部分的池化特征,并对所述 B_K 个部分的池化特征进行合并,得到所述任一第二待池化片段的池化特征;

所述第一待池化片段包括所述时域区间对应的视频片段、所述第一相邻片段和所述第二相邻片段中的任意一个或多个;所述第二待池化片段包括所述候选片段中所述第一待池化片段外的其他待池化片段。

12、根据权利要求10所述的方法,其特征在于,所述对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理之前,还包括:

预设池化层数 K 的取值大于1;

所述对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理,包括:

分别针对第 k 个池化层:对于预设分割部分数 B_K 的取值为1的任一第一待池化片段,由所述任一第一待池化片段中至少一个视频片段的动作特征获取所述任一第一待池化片段在第 k 层的池化特征;对于预设分割部分数 B_K 的取值大于1的任一第二待池化片段,将所述任一第二待池化片段中的所有视频片段分割为 B_K 个部分,分别由所述 B_K 个部分中各部分视频片段的动作特征获取对应部分的池化特征,并对所述 B_K 个部分的池化特征进行合并,得到所述任一第二待池化片段在第 k 层的池化特征;

分别针对至少一个待池化片段:按照预设方式对 K 个池化层的池化特征进行合并,得到至少一个待池化片段的池化特征;

所述第一待池化片段包括所述时域区间对应的视频片段、所述第一相邻片段和所述第二相邻片段中的任意一个或多个;所述第二待池化片段包括所述候选片段中所述第一待池化片段外的其他待池化片段。

13、根据权利要求10-12任一所述的方法,其特征在于,所述合并时域金字塔式池化后得到的池化特征,获得所述时域区间对应的视频片段的全局特征,包括:

分别对所述第一相邻片段的池化特征、所述时域区间对应的视频片段的池化特征和所述第二相邻片段的池化特征进行合并,获得所述时域区间对应的视频片段时序结构化的所述全局特征。

14、根据权利要求10所述的方法,其特征在于,所述对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理之前,还包括:

预设池化层数 K 的取值大于1;

所述对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理,包括:

分别针对第 k 个池化层:识别预设分割部分数 B_K 的取值是否为1;

若预设分割部分数 B_K 的取值为1,由所述候选片段中至少一个视频片段的动作特征获取所述候选片段在第 k 层的池化特征;

若预设分割部分数 B_K 的取值大于1,将所述候选片段中的所有视频片段分割为 B_K 个部分,分别由所述 B_K 个部分中各部分中视频片段的动作特征获取对应部分的池化特征,并对所述 B_K 个部分的池化特征进行合并,得到所述候选片段在第 k 层的池化特征。

15、根据权利要求14所述的方法,其特征在于,所述合并时域金字塔式池化后得到的池化特征,获得所述时域区间对应的视频片段的全局特征,包括:

按照预设方式对 K 个池化层的池化特征进行合并,获得所述时域区间对应的视频片段时序结构化的所述全局特征。

16、根据权利要求10所述的方法,其特征在于,所述对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理之前,还包括:

预设池化层数 K 的取值为1;

所述对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理,包括:

识别预设分割部分数 B_K 的取值是否为1;

若预设分割部分数 B_K 的取值为1,分别针对至少两个待池化片段中的任一待池化片段,由所述任一待池化片段中至少两个视频片段的动作特征获取所述任一待池化片段的池化特征,所述待池化片段包括所述时域区间内视频片段、所述第一相邻片段和所述第二相邻片段;对所有待池化片段的池化特征进行合并,获得所述时域区间对应的视频片段时序结构化的所述全局特征;

若预设分割部分数 B_K 的取值大于1,将所述候选片段中的所有视频片段分割为 B_K 个部分,分别由所述 B_K 个部分中各部分中视频片段的动作特征获取对应部分的池化特征,并对所述 B_K 个部分的池化特征进行合并,获得所述时域区间对应的视频片段时序结构化的所述全局特征。

17、根据权利要求1至16任一所述的方法,其特征在于,所述基于所述全局特征确定所述时域区间对应的视频片段的动作完整度得分,包括:

对所述全局特征进行分类，得到所述候选片段对应的至少一个动作类别；

分别通过所述候选片段对应的至少一个动作类别相关的分类器，基于所述全局特征识别所述时域区间对应的视频片段包括当前动作类别完整的动作实例的程度，得到所述时域区间对应的视频片段分别对应至少一个动作类别的动作完整度得分。

18、根据权利要求 1 至 17 任意一项所述的方法，其特征在于，还包括：

基于所述时域区间对应的视频片段的分类得分与动作完整度得分，确定所述时域区间对应的视频片段是否包含对应的至少一个动作类别完整的动作实例。

19、根据权利要求 18 所述的方法，其特征在于，所述基于所述时域区间对应的视频片段的分类得分与动作完整度得分，确定所述时域区间对应的视频片段是否包含对应的至少一个动作类别完整的动作实例，包括：

分别针对所述时域区间对应的视频片段对应的任一动作类别，对所述时域区间对应的视频片段对应所述任一动作类别的分类得分与动作完整度得分求取乘积或者求和，获得所述时域区间对应的视频片段是否包含所述任一动作类别完整的动作实例的置信度值；

基于所述置信度值是否大于预设阈值，确定所述时域区间对应的视频片段是否包含所述任一动作类别完整的动作实例。

20、根据权利要求 19 所述的方法，其特征在于，还包括：

若所述时域区间对应的视频片段包含完整的动作实例对应的动作类别大于一个，选择对应于最大置信度值的一个动作类别作为所述时域区间对应的视频片段的检测动作类别。

21、根据权利要求 18 所述的方法，其特征在于，所述基于所述时域区间对应的视频片段的分类得分与动作完整度得分，确定所述时域区间对应的视频片段是否包含对应至少一个动作类别完整的动作实例，包括：

识别所述时域区间对应的视频片段的至少一个动作类别的分类得分中，是否存在分类得分大于预设第一数值的预设动作类别；以及识别所述动作完整度得分中，是否存在动作完整度得分大于预设第二数值的预设动作类别；

响应于存在分类得分大于预设第一数值的预设动作类别和/或动作完整度得分大于预设第二数值的预设动作类别，以分类得分大于预设第一数值的预设动作类别和/或动作完整度得分大于预设第二数值的预设动作类别作为目标动作类别，基于所述时域区间对应的视频片段对应的目标动作类别的分类得分与动作完整度得分，确定所述时域区间对应的视频片段是否包含所述目标动作类别完整的动作实例；

所述输出所述时域区间以及所述时域区间对应的视频片段的检测动作类别时，输出所述目标动作类别及其时域区间。

22、根据权利要求 18 至 21 任意一项所述的方法，其特征在于，还包括：

若所述时域区间对应的视频片段包含至少一个动作类别完整的动作实例，基于所述全局特征对所述时域区间进行位置回归，获得起止位置偏移量，所述起止位置偏移量包括时域区间中起始时间和结束时间的位置偏移量；

通过所述起止位置偏移量对所述时域区间进行调整；

所述输出所述时域区间以及所述时域区间对应的视频片段的检测动作类别时，输出调整后的时域区间以及所述时域区间对应的视频片段的检测动作类别。

23、一种时域动作检测系统，其特征在于，包括：

第一获取模块，用于获取视频中存在动作实例的时域区间以及所述时域区间的至少一个相邻片段；

特征提取模块，用于分别提取候选片段中至少两个视频片段的动作特征，其中，所述候选片段包括所述时域区间对应的视频片段及其相邻片段；

池化处理模块，用于对所述候选片段中至少两个视频片段的动作特征进行池化处理，获得所述时域区间对应的视频片段的全局特征；

第一确定模块，用于基于所述全局特征确定所述时域区间对应的视频片段的动作完整度得分。

24、根据权利要求 23 所述的系统，其特征在于，所述至少一个相邻片段包括：所述视频中时序位于所述时域区间之前的第一相邻片段，和/或，所述视频中时序位于所述时域区间之后的第二相邻片段；所述第一相邻片段和所述第二相邻片段分别包括至少一个视频片段。

25、根据权利要求 23 或 24 所述的系统，其特征在于，所述第一获取模块包括：

动作估计子模块，用于分别对视频中的至少一个视频片段进行动作度估计，获得时序动作度序列；

位置预测子模块，用于基于所述时序动作度序列进行动作位置预测，获得所述视频中存在动作实例的时域区间，所述时域区间包括起始时间和结束时间；

提取子模块,用于从所述视频中提取所述时域区间之前的第一相邻片段和/或所述时域区间之后的第二相邻片段。

26、根据权利要求 25 所述的系统,其特征在于,所述动作估计子模块包括:

图像获取单元,用于分别针对所述视频中的任一视频片段:提取一帧图像作为原始图像,以及提取所述任一视频片段的光流,并对获得的光流场图片进行合并,得到拼合光流场图像;

第一子卷积神经网络,用于对所述原始图像进行动作度估计,得到第一动作度值;

第二子卷积神经网络,用于对所述拼合光流场图像进行动作度估计,得到第二动作度值;

计算单元,用于由同一视频片段的第一动作度值与第二动作度值,获取所述任一视频片段的动作度值;

所述视频中所有视频片段基于时序关系的动作度值形成所述时序动作度序列。

27、根据权利要求 26 所述的系统,其特征在于,所述动作估计子模块还包括:

归一化单元,用于对所述计算单元获取的视频片段的动作度值进行归一化处理,得到归一化的动作度值;

所述时序动作度序列包括:由归一化的动作度值形成的时序动作度序列。

28、根据权利要求 13 至 27 任意一项所述的系统,其特征在于,还包括:

第二获取模块,用于基于所述时域区间对应的视频片段的动作特征,获取所述时域区间对应的视频片段的至少一个动作类别的分类得分;

第二确定模块,用于根据所述时域区间对应的视频片段的至少一个动作类别的分类得分,确定所述时域区间对应的视频片段的检测动作类别。

29、根据权利要求 28 所述的系统,其特征在于,还包括:

输出模块,用于输出所述时域区间以及所述时域区间对应的视频片段的检测动作类别。

30、根据权利要求 28 或 29 所述的系统,其特征在于,所述第二获取模块包括:

第一获取子模块,用于分别基于所述时域区间对应的至少一个视频片段的动作特征,获取所述时域区间对应的至少一个视频片段分别属于所述至少一个动作类别的得分;

第一计算子模块,用于分别对所述时域区间对应的至少一个视频片段属于同一动作类别的得分求和,获得所述时域区间对应的视频片段的至少一个动作类别的分类得分。

31、根据权利要求 23 至 30 任意一项所述的系统,其特征在于,所述池化处理模块包括:

池化处理子模块,用于对所述候选片段中至少两个视频片段的动作特征进行时域金字塔式池化处理。

32、根据权利要求 31 所述的系统,其特征在于,所述池化处理模块还包括:

合并子模块,用于合并时域金字塔式池化后得到的池化特征。

33、根据权利要求 32 所述的系统,其特征在于,所述池化处理模块,还用于预设池化层数 K 的取值为 1;

所述池化处理子模块用于:对于预设分割部分数 B_K 的取值为 1 的任一第一待池化片段,由所述任一第一待池化片段中至少一个视频片段的动作特征获取所述任一第一待池化片段的池化特征;对于预设分割部分数 B_K 的取值大于 1 的任一第二待池化片段,将所述任一第二待池化片段中的所有视频片段分割为 B_K 个部分,分别由各部分中视频片段的动作特征获取对应部分的池化特征,并对所述 B_K 个部分的池化特征进行合并,得到任一第二待池化片段的池化特征;所述第一待池化片段包括所述时域区间对应的视频片段、所述第一相邻片段和所述第二相邻片段中的任意一个或多个;所述第二待池化片段包括所述候选片段中所述第一待池化片段外的其他待池化片段;

所述合并子模块,用于分别对所述第一相邻片段的池化特征、所述时域区间对应的视频片段的池化特征和所述第二相邻片段的池化特征进行合并,获得所述时域区间对应的视频片段时序结构化的所述全局特征。

34、根据权利要求 32 所述的系统,其特征在于,所述池化处理模块,还用于预设池化层数 K 的取值大于 1;

所述池化处理子模块用于:

分别针对第 k 个池化层:对于预设分割部分数 B_K 的取值为 1 的任一第一待池化片段,由所述任一第一待池化片段中至少一个视频片段的动作特征获取所述任一第一待池化片段在第 k 层的池化特征;对于预设分割部分数 B_K 的取值大于 1 的任一第二待池化片段,将所述任一第一待池化片段中的所有视频片段分割为 B_K 个部分,分别由各部分中视频片段的动作特征获取对应部分的池化特征,并对所述 B_K 个部分的池化特征进行合并,得到所述任一第一待池化片段在第 k 层的池化特征;以及分别针对至少一个待池化片段:按照预设方式对 K 个池化层的池化特征进行合并,得到至少一个待池化片段的池化特征;所述第一待池化片段包括所述时域区间对应的视频片段、所述第一相邻片段和所述第二相邻片段中的任意一个或多个;所述第

二待池化片段包括所述候选片段中所述第一待池化片段外的其他待池化片段；

所述合并子模块，用于分别对所述第一相邻片段的池化特征、所述时域区间对应的视频片段的池化特征和所述第二相邻片段的池化特征进行合并，获得所述时域区间对应的视频片段时序结构化的所述全局特征。

35、根据权利要求 32 所述的系统，其特征在于，所述池化处理模块，还用于预设池化层数 K 的取值大于 1；

所述池化处理子模块用于：分别针对第 k 个池化层：识别预设分割部分数 B_K 的取值是否为 1；若预设分割部分数 B_K 的取值为 1，由所述候选片段中至少一个视频片段的动作特征获取所述候选片段在第 k 层的池化特征；若预设分割部分数 B_K 的取值大于 1，将所述候选片段中的所有视频片段分割为 B_K 个部分，分别由所述 B_K 个部分各部分视频片段的动作特征获取对应部分的池化特征，并对所述 B_K 个部分的池化特征进行合并，得到所述候选片段在第 k 层的池化特征；

所述合并子模块，用于按照预设方式对 K 个池化层的池化特征进行合并，获得所述时域区间对应的视频片段时序结构化的所述全局特征。

36、根据权利要求 32 所述的系统，其特征在于，所述池化处理模块，还用于预设池化层数 K 的取值为 1；

所述池化处理子模块用于：

识别预设分割部分数 B_K 的取值是否为 1；

若预设分割部分数 B_K 的取值为 1，分别针对至少两个待池化片段中的任一待池化片段，由所述任一待池化片段中至少两个视频片段的动作特征获取所述任一待池化片段的池化特征，所述待池化片段包括所述时域区间内视频片段、所述第一相邻片段和所述第二相邻片段；对所有待池化片段的池化特征进行合并，获得所述时域区间对应的视频片段时序结构化的所述全局特征；

若预设分割部分数 B_K 的取值大于 1，将所述候选片段中的所有视频片段分割为 B_K 个部分，分别由所述 B_K 个部分中各部分视频片段的动作特征获取对应部分的池化特征，并对所述 B_K 个部分的池化特征进行合并，获得所述时域区间对应的视频片段时序结构化的所述全局特征。

37、根据权利要求 23 至 36 任意一项所述的系统，其特征在于，所述第一确定模块包括：

分类子模块，用于对所述全局特征进行分类，得到所述候选片段对应的至少一个动作类别；

至少一个分类器，分别与所述候选片段对应的至少一个动作类别相关，用于基于所述全局特征识别所述时域区间对应的视频片段包括当前动作类别完整的动作实例的程度，得到所述时域区间对应的视频片段分别对应至少一个动作类别的动作完整度得分。

38、根据权利要求 23 至 37 任意一项所述的系统，其特征在于，还包括：

第三确定模块，用于基于所述时域区间对应的视频片段的分类得分与动作完整度得分，确定所述时域区间对应的视频片段是否包含对应的至少一个动作类别完整的动作实例。

39、根据权利要求 38 所述的系统，其特征在于，所述第三确定模块包括：

第二计算子模块，用于分别针对所述时域区间对应的视频片段对应的任一动作类别，对所述时域区间对应的视频片段对应所述任一动作类别的分类得分与动作完整度得分求取乘积或者求和，获得所述时域区间对应的视频片段是否包含所述任一动作类别完整的动作实例的置信度值；

确定子模块，用于基于所述置信度值是否大于预设阈值，确定所述时域区间对应的视频片段是否包含所述任一动作类别完整的动作实例。

40、根据权利要求 38 或 39 所述的系统，其特征在于，所述确定子模块，还用于在所述时域区间对应的视频片段包含完整的动作实例对应的动作类别大于一个时，选择对应于最大置信度值的一个动作类别作为所述时域区间对应的视频片段的检测动作类别。

41、根据权利要求 38 所述的系统，其特征在于，还包括：

识别模块，用于识别所述时域区间对应的视频片段的至少一个动作类别的分类得分中，是否存在分类得分大于预设第一数值的预设动作类别；以及识别所述动作完整度得分中，是否存在动作完整度得分大于预设第二数值的预设动作类别；

所述第三确定模块，用于在存在分类得分大于预设第一数值的预设动作类别和/或动作完整度得分大于预设第二数值的预设动作类别时，以分类得分大于预设第一数值的预设动作类别和/或动作完整度得分大于预设第二数值的预设动作类别作为目标动作类别，基于所述时域区间对应的视频片段对应的目标动作类别的分类得分与动作完整度得分，确定所述时域区间对应的视频片段是否包含所述目标动作类别完整的动作实例的操作；

所述输出模块用于输出所述目标动作类别及其时域区间。

42、根据权利要求 38 至 41 任意一项所述的系统，其特征在于，还包括：

位置回归模块，用于在所述时域区间对应的视频片段包含至少一个动作类别完整的动作实例时，基于所述全局特征对所述时域区间进行位置回归，获得起止位置偏移量，并通过所述起止位置偏移量对所述时域区间进行调整；所述起止位置偏移量包括时域区间中起始时间和结束时间的位置偏移量；

所述输出模块，用于输出调整后的时域区间以及所述时域区间对应的视频片段的检测动作类别。

43、一种电子设备，其特征在于，包括权利要求 23 至 42 任意一项所述的时域动作检测系统。

44、一种电子设备，其特征在于，包括：处理器和存储器；

所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行如权利要求 1-22 任一项所述的时域动作检测方法对应的操作。

45、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设备上运行时，所述设备中的处理器执行用于实现如权利要求 1-22 任一项所述的时域动作检测方法中各步骤的指令。

46、一种计算机存储介质，用于存储计算机可读的指令，其特征在于，所述指令被执行时实现如权利要求 1-22 任一项所述的时域动作检测方法中各步骤的操作。

1/6

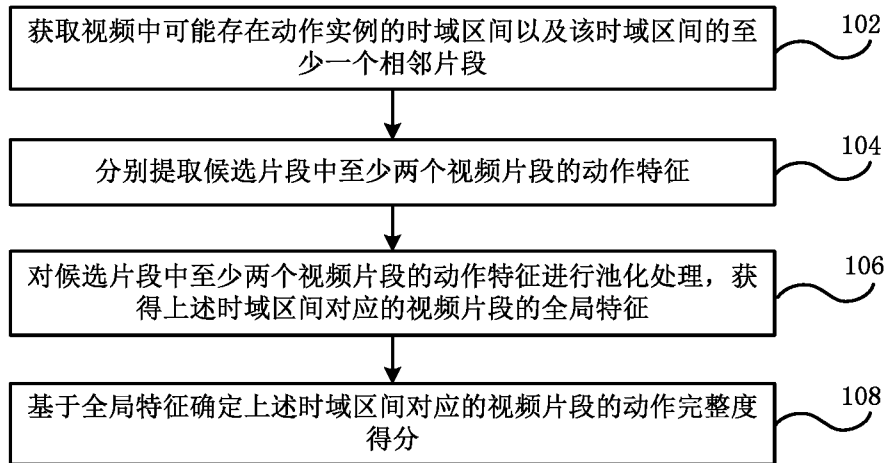


图 1

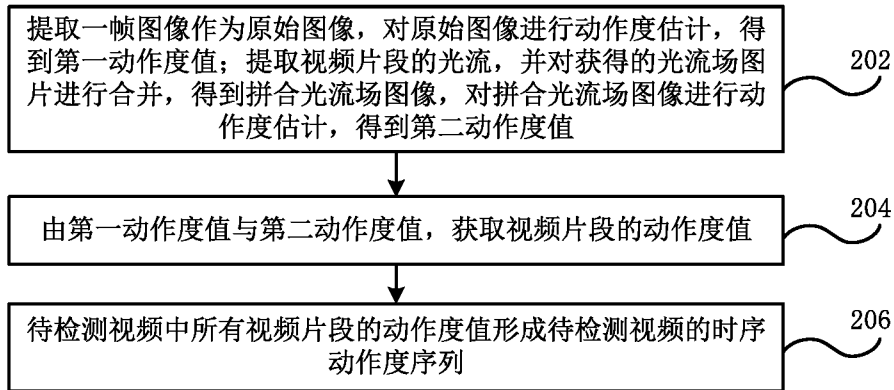


图 2

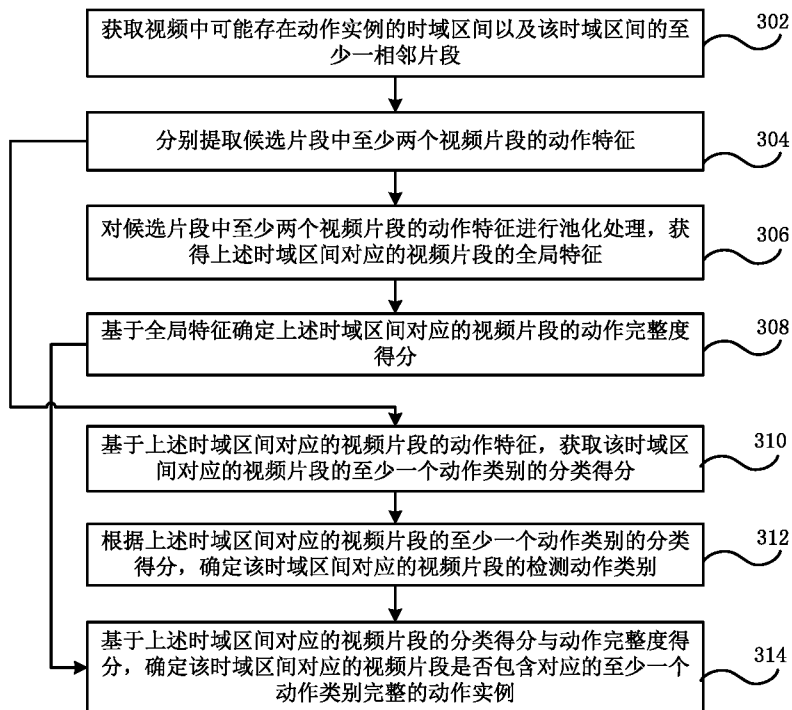


图 3

2/6

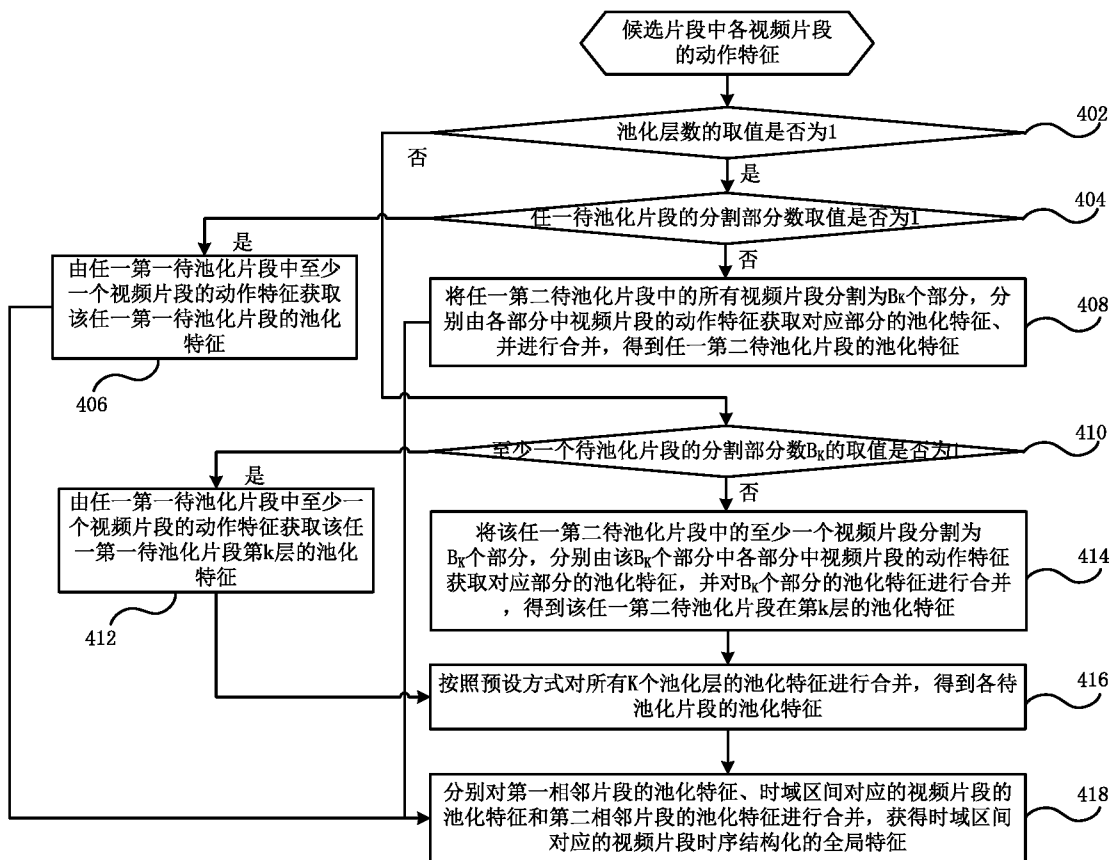


图 4

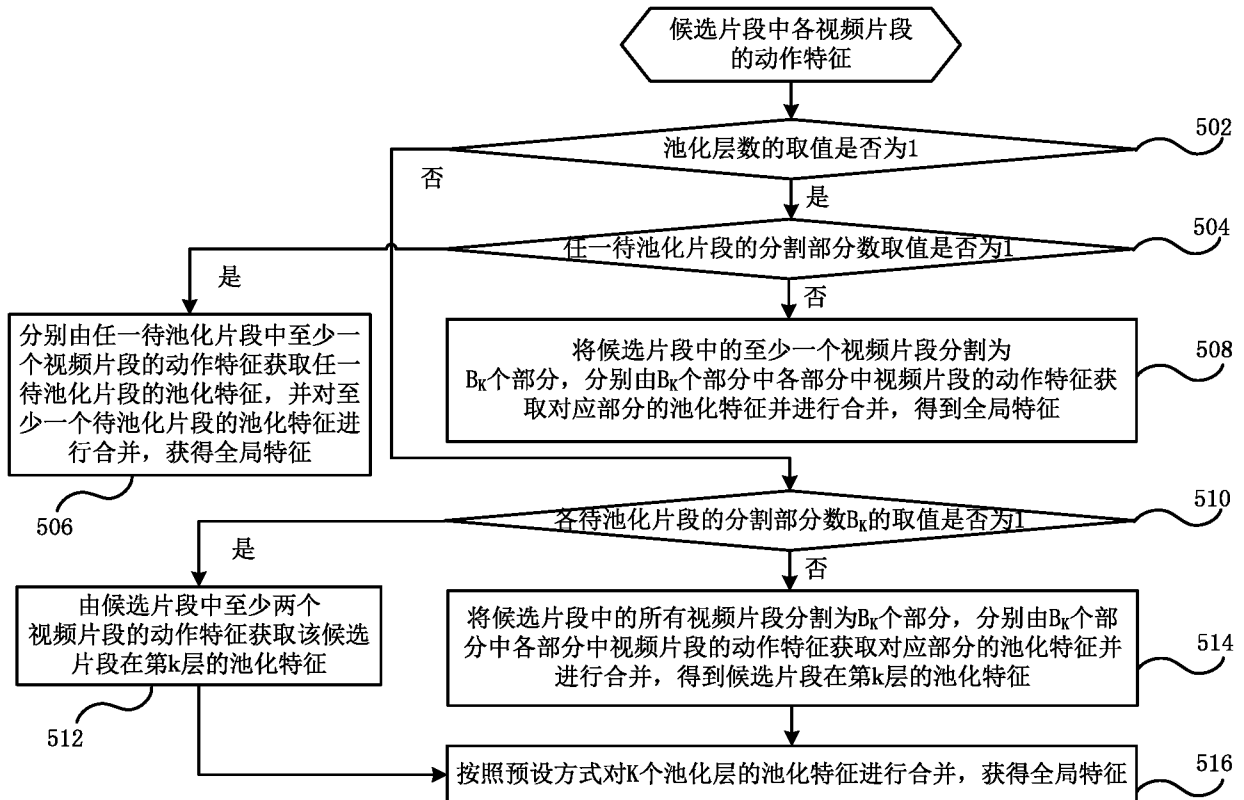


图 5

3/6

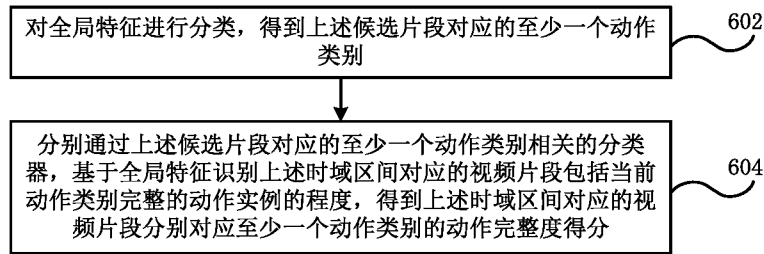


图 6

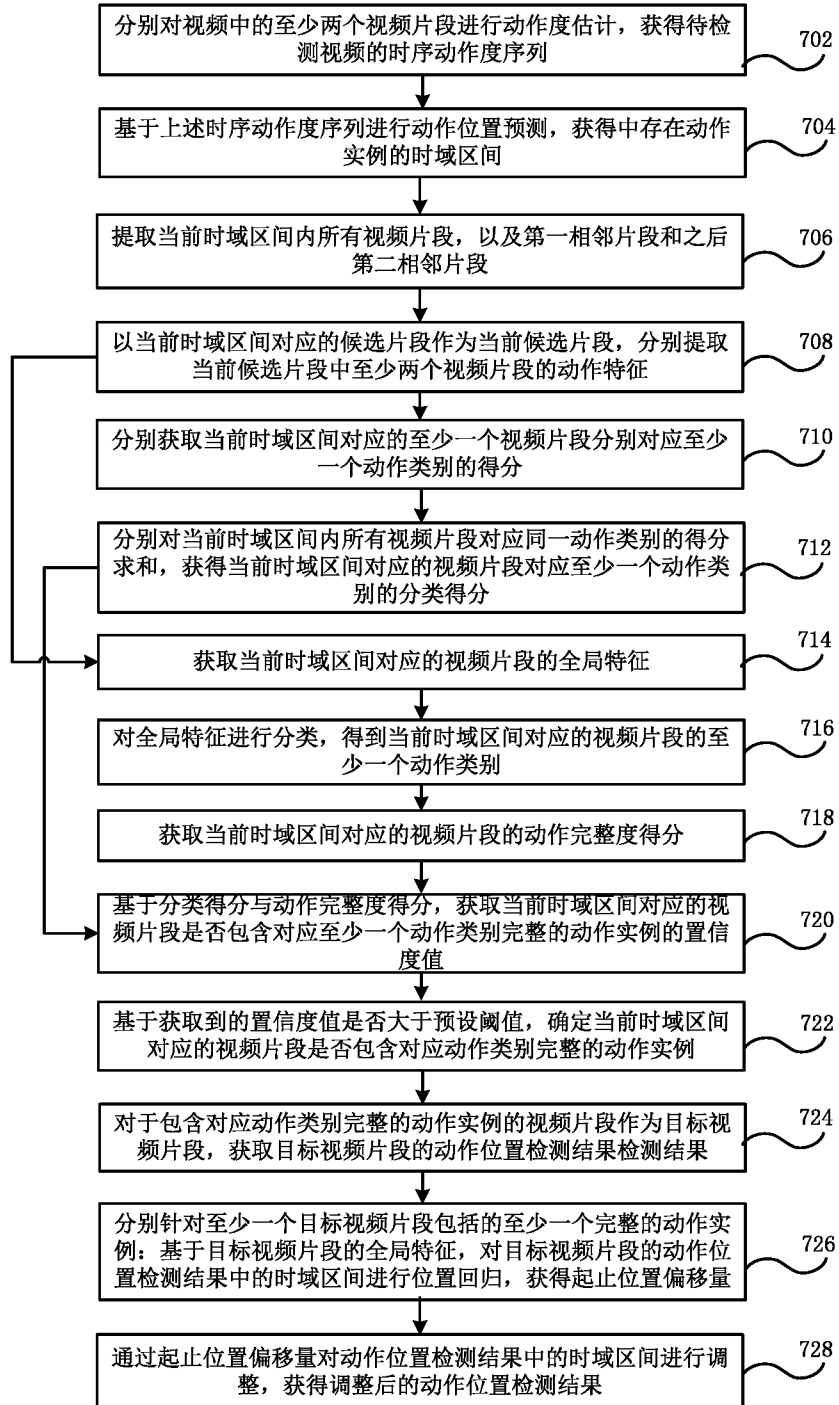


图 7

4/6

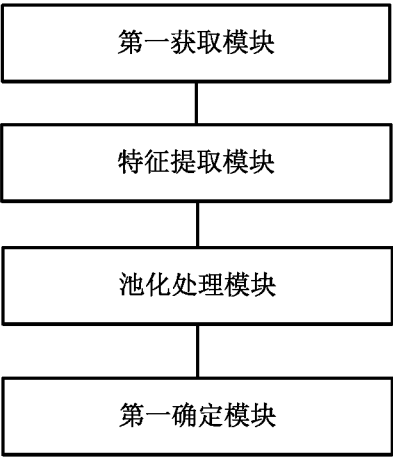


图 8

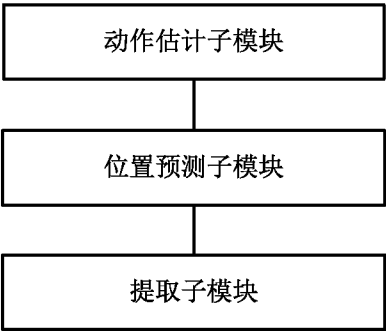


图 9

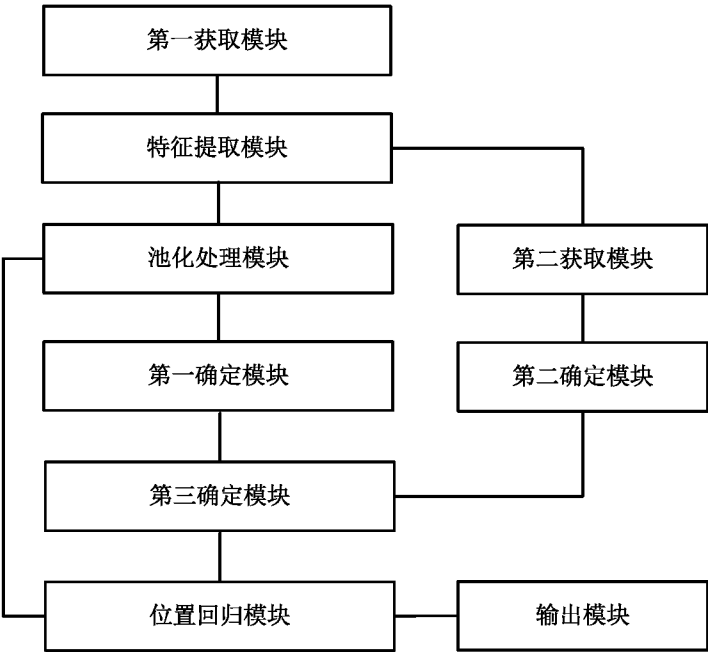


图 10

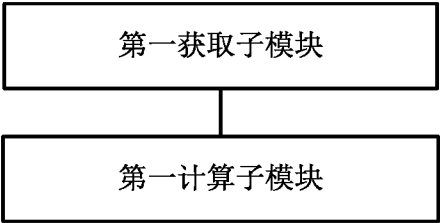


图 11



图 12

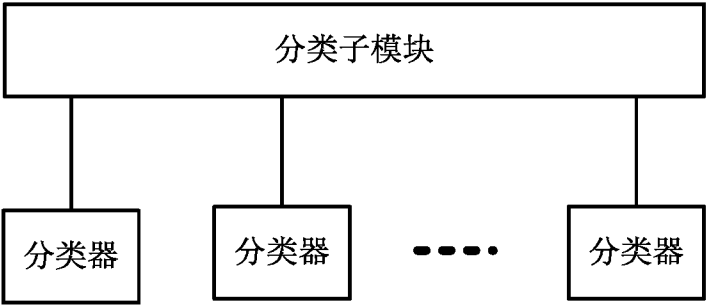


图 13



图 14

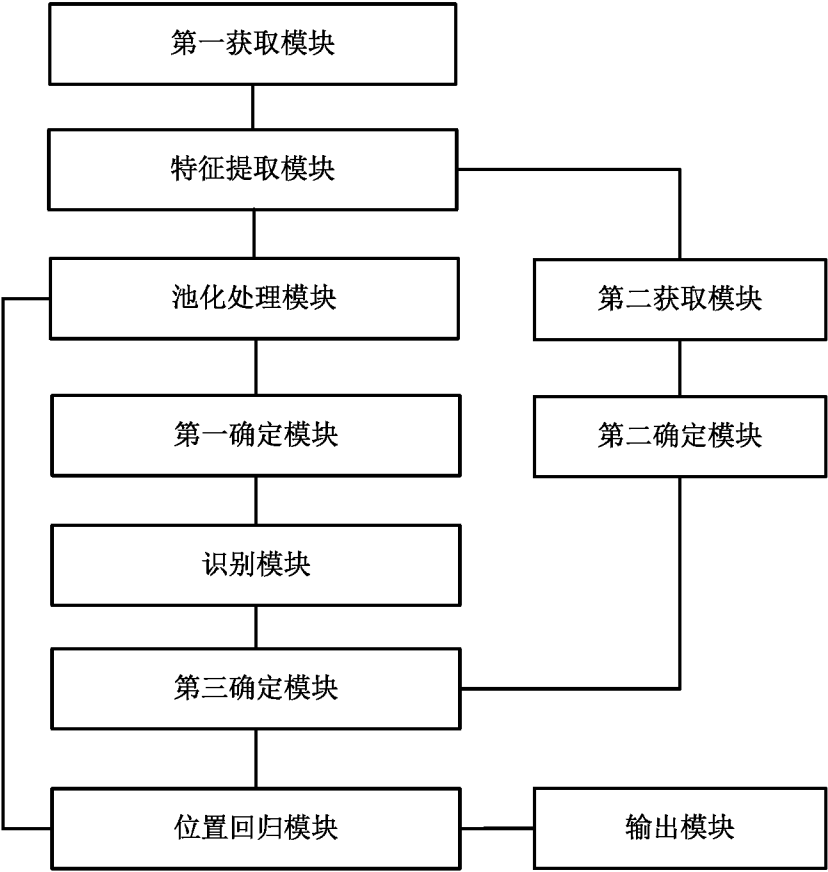


图 15

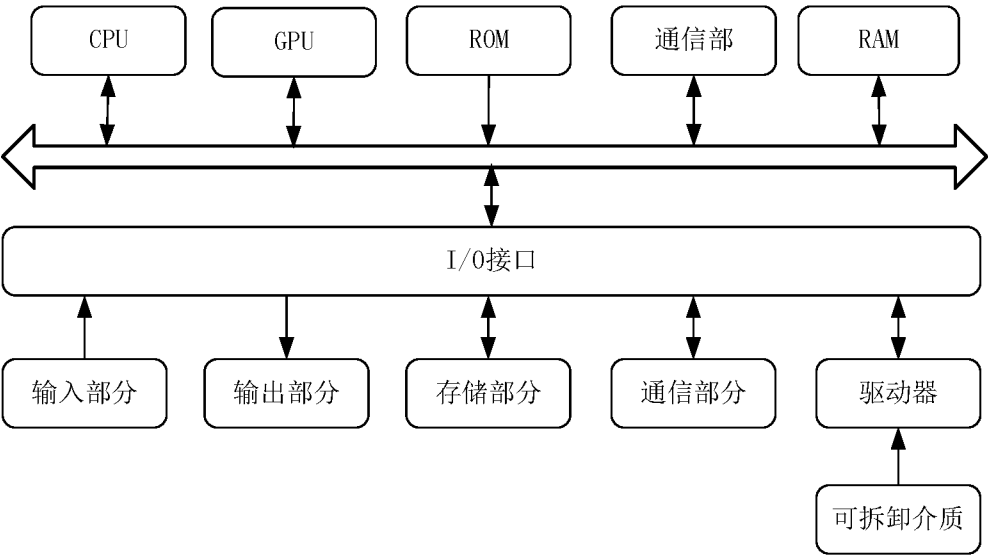


图 16

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/083924

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00 (2006.01) 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

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)

CNXTX; CNABS; SIPOABS; DWPI; CNKI: 时域, 池化, 视频, 动作, 行为, time domain, pooling, video, action, behavior

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106203255 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY), 07 December 2016 (07.12.2016), entire document	1-46
A	CN 105354528 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES), 24 February 2016 (24.02.2016), entire document	1-46
A	US 2017083798 A1 (CANON KABUSHIKI KAISHA), 23 March 2017 (23.03.2017), entire document	1-46
A	CN 104933417 A (SOOCHOW UNIVERSITY), 23 September 2015 (23.09.2015), entire document	1-46

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 11 June 2018	Date of mailing of the international search report 05 July 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer WANG, Yankun Telephone No. 86-(10)-62412330

International application No.
PCT/CN2018/083924

Form PCT/ISA/210 (patent family annex) (January 2015)

国际检索报告

国际申请号

PCT/CN2018/083924

A. 主题的分类

G06K 9/00(2006.01)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

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))

CNTXT;CNABS;SIPOABS;DWPI;CNKI: 时域, 池化, 视频, 动作, 行为, time domain, pooling, video, action, behavior

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	CN 106203255 A (华中科技大学) 2016年 12月 7日 (2016 - 12 - 07) 全文	1-46
A	CN 105354528 A (中国科学院深圳先进技术研究院) 2016年 2月 24日 (2016 - 02 - 24) 全文	1-46
A	US 2017083798 A1 (CANON KABUSHIKI KAISHA) 2017年 3月 23日 (2017 - 03 - 23) 全文	1-46
A	CN 104933417 A (苏州大学) 2015年 9月 23日 (2015 - 09 - 23) 全文	1-46

☐ 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

2018年 6月 11日

国际检索报告邮寄日期

2018年 7月 5日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(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

王艳坤

电话号码 86-(10)-62412330

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/083924

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	106203255	A	2016年 12月 7日	无	
CN	105354528	A	2016年 2月 24日	无	
US	2017083798	A1	2017年 3月 23日	无	
CN	104933417	A	2015年 9月 23日	无	