

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2023 年 2 月 9 日 (09.02.2023)



(10) 国际公布号
WO 2023/010758 A1

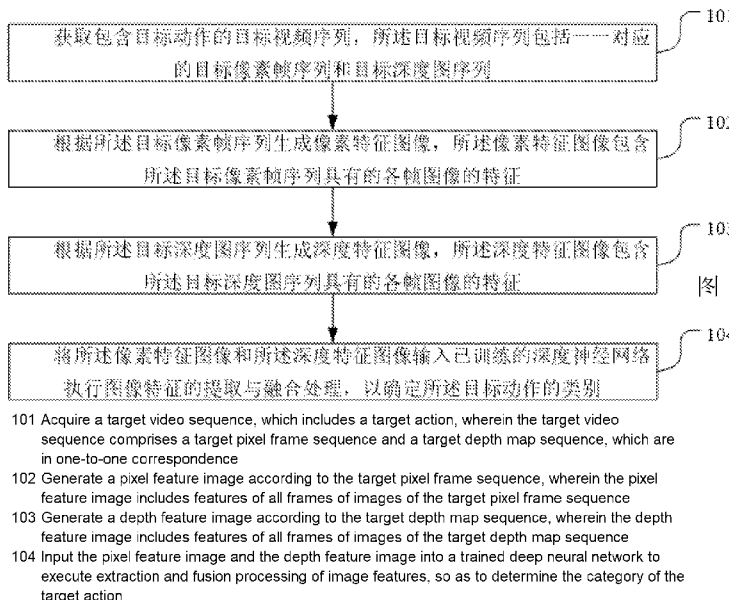
- (51) 国际专利分类号:
G06K 9/00 (2022.01) *G06N 3/04* (2006.01)
G06K 9/62 (2022.01)
- (21) 国际申请号: PCT/CN2021/138566
- (22) 国际申请日: 2021 年 12 月 15 日 (15.12.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110889116.5 2021年8月4日 (04.08.2021) CN
- (71) 申请人: 中国科学院深圳先进技术研究院 (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY CHINESE ACADEMY OF SCIENCES) [CN/CN]; 中国广东省深圳市

南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。

- (72) 发明人: 任子良 (REN, Ziliang); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。程俊 (CHENG, Jun); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。张锬石 (ZHANG, Qieshi); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。高向阳 (GAO, Xiangyang); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。康宇航 (KANG, Yuhang); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。

(54) Title: ACTION DETECTION METHOD AND APPARATUS, AND TERMINAL DEVICE AND STORAGE MEDIUM

(54) 发明名称: 一种动作检测方法、装置、终端设备和存储介质



(57) Abstract: An action detection method and apparatus, and a terminal device and a storage medium. The method comprises: acquiring a target video sequence, which includes a target action, wherein the target video sequence comprises a target pixel frame sequence and a target depth map sequence, which are in one-to-one correspondence (101); generating a pixel feature image according to the target pixel frame sequence, wherein the pixel feature image includes features of all frames of images of the target pixel frame sequence (102); generating a depth feature image according to the target depth map sequence, wherein the depth feature image includes features of all frames of images of the target depth map sequence (103); and inputting the pixel feature image and the depth feature image into a trained deep neural network to execute extraction and fusion processing of image features, so as to determine the category of the target

(74) 代理人: 北京中巡通大知识产权代理有限公司
(BEIJING ZHONG XUN TONG DA INTELLECTUAL
PROPERTY AGENCY CO., LTD.); 中国北京市
西城区榆树馆胡同东口1层106, Beijing
100195 (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))。

action (104). By means of the method, the interference of an environmental factor on a detection sample can be weakened to a certain extent, thereby improving the accuracy of action detection.

(57) 摘要: 一种动作检测方法、装置、终端设备和存储介质。该方法包括: 获取包含目标动作的目标视频序列, 所述目标视频序列包括一一对应的目标像素帧序列和目标深度图序列 (101); 根据所述目标像素帧序列生成像素特征图像, 所述像素特征图像包含所述目标像素帧序列具有的各帧图像的特征 (102); 根据所述目标深度图序列生成深度特征图像, 所述深度特征图像包含所述目标深度图序列具有的各帧图像的特征 (103); 将所述像素特征图像和所述深度特征图像输入已训练的深度神经网络执行图像特征的提取与融合处理, 以确定所述目标动作的类别 (104)。采用该方法能够在一定程度上减弱环境因素对检测样本的干扰, 从而提高动作检测的准确率。

一种动作检测方法、装置、终端设备和存储介质

技术领域

5 本申请涉及图像处理技术领域，尤其涉及一种动作检测方法、装置、终端设备和存储介质。

背景技术

 动作检测作为模式识别的研究分支，广泛应用于视频安全监控、视频检索
10 和健康医疗等领域。动作检测是指对视频片段中的目标（例如人体）进行识别与跟踪，以确定该目标的动作类别。

 目前，通常采用基于RGB图像的动作检测方法，该方法通过对RGB图像序列的像素特征进行分析以实现动作检测。然而，作为检测样本的RGB图像容易受到光照变化等环境因素的干扰，导致动作检测的准确率较低。

15

发明内容

 有鉴于此，本申请实施例提供了一种动作检测方法、装置、终端设备和存储介质，能够提高动作检测的准确率。

 本申请实施例的第一方面提供了一种动作检测方法，包括：

20 获取包含目标动作的目标视频序列，所述目标视频序列包括一一对应的目标像素帧序列和目标深度图序列；

 根据所述目标像素帧序列生成像素特征图像，所述像素特征图像包含所述目标像素帧序列具有的各帧图像的特征；

 根据所述目标深度图序列生成深度特征图像，所述深度特征图像包含所述
25 目标深度图序列具有的各帧图像的特征；

 将所述像素特征图像和所述深度特征图像输入已训练的神经网络执行

图像特征的提取与融合处理，以确定所述目标动作的类别。

在本申请实施例中，首先，获取包含目标动作的目标视频序列，该目标视频序列包括一一对应的像素帧序列和深度图序列；然后，根据像素帧序列生成包含其中各帧图像的图像特征的像素特征图像，以及根据深度图序列生成包含其中各帧图像的图像特征的深度特征图像；最后，将像素特征图像和深度特征图像输入一个已训练的深度神经网络进行图像特征的提取与融合处理，从而确定目标动作的类别。上述过程将视频图像的像素特征和深度特征融合，利用像素信息和深度信息的互补性，能够在一定程度上减弱环境因素对检测样本的干扰，从而提高动作检测的准确率。

在本申请的一种实现方式中，所述深度神经网络包括特征提取模块和特征融合模块，所述将所述像素特征图像和所述深度特征图像输入已训练的深度神经网络执行图像特征的提取与融合处理，以确定所述目标动作的类别，可以包括：

将所述像素特征图像和所述深度特征图像输入所述特征提取模块进行处理，得到图像语义特征；

将所述图像语义特征输入所述特征融合模块进行处理，得到融合后的图像特征；

基于所述融合后的图像特征确定所述目标动作的类别。

在本申请的一种实现方式中，所述根据所述目标像素帧序列生成像素特征图像，可以包括：

在时间维度上对所述目标像素帧序列执行稀疏采样处理，得到第一图像序列；

将所述第一图像序列包含的各帧图像融合，得到所述像素特征图像；

所述根据所述目标深度图序列生成深度特征图像，可以包括：

在时间维度上对所述目标深度图序列执行稀疏采样处理，得到第二图像序列；

将所述第二图像序列包含的各帧图像融合，得到所述深度特征图像。

进一步的，所述将所述第一图像序列包含的各帧图像融合，得到所述像素特征图像，可以包括：

对所述第一图像序列包含的各帧图像按照对应位置像素点执行图像特征的
5 叠加、求平均和取整操作，得到所述像素特征图像；

所述将所述第二图像序列包含的各帧图像融合，得到所述深度特征图像，
可以包括：

将所述第二图像序列包含的各帧图像分别转换成各帧灰度图像；

对所述各帧灰度图像按照对应位置像素点执行图像特征的叠加、求平均和
10 取整操作，得到所述深度特征图像。

在本申请的一种实现方式中，在确定所述目标动作的类别之后，还可以包
括：

从预设的规范动作视频库中查找与所述目标动作的类别对应的基准视频序
列，所述基准视频序列包含规范化的所述目标动作；

15 在时间维度上对所述基准视频序列执行稀疏采样处理，得到第三图像序列，
所述第三图像序列包含的图像帧数和所述第一图像序列包含的图像帧数相同；

分别标注出所述第一图像序列包含的各帧图像中目标物体具有的指定部位
的位置，以及所述第三图像序列包含的各帧图像中所述目标物体具有的所述指
定部位的位置，所述目标物体为执行所述目标动作的物体；

20 根据所述第一图像序列包含的各帧图像中所述目标物体具有的所述指定部
位的位置，构建得到所述指定部位对应的第一运动轨迹曲线；

根据所述第三图像序列包含的各帧图像中所述目标物体具有的所述指定部
位的位置，构建得到所述指定部位对应的第二运动轨迹曲线；

根据所述第一运动轨迹曲线和所述第二运动轨迹曲线，计算得到曲线误差；

25 根据所述曲线误差确定所述目标视频序列包含的所述目标动作的规范化程
度。

进一步的，所述根据所述第一运动轨迹曲线和所述第二运动轨迹曲线，计算得到曲线误差，可以包括：

分别计算所述第一运动轨迹曲线中的每个目标位置点和其在所述第二运动轨迹曲线中的对应位置点之间的距离，得到每个所述目标位置点的误差；

5 将每个所述目标位置点的误差叠加，得到所述曲线误差。

在本申请的一种实现方式中，所述获取包含目标动作的目标视频序列，可以包括：

获取包含多个动作的原始视频序列，所述原始视频序列包括一一对应的原始像素帧序列和原始深度图序列；

10 对所述原始像素帧序列执行视频的动作分割处理，得到多个包含单一动作的像素帧序列分段，并从所述多个像素帧序列分段中选取一个像素帧序列分段，作为所述目标像素帧序列；

对所述原始深度图序列执行视频的动作分割处理，得到多个包含单一动作的深度图序列分段，并从所述多个深度图序列分段中选取与所述目标像素帧序列对应的深度图序列分段，作为所述目标深度图序列。

本申请实施例的第二方面提供了一种动作检测装置，包括：

视频序列获取模块，用于获取包含目标动作的目标视频序列，所述目标视频序列包括一一对应的目标像素帧序列和目标深度图序列；

像素特征生成模块，用于根据所述目标像素帧序列生成像素特征图像，所述像素特征图像包含所述目标像素帧序列具有的各帧图像的特征；

深度特征生成模块，用于根据所述目标深度图序列生成深度特征图像，所述深度特征图像包含所述目标深度图序列具有的各帧图像的特征；

动作检测模块，用于将所述像素特征图像和所述深度特征图像输入已训练的神经网络执行图像特征的提取与融合处理，以确定所述目标动作的类别。

25 本申请实施例的第三方面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，所述处理器执行

所述计算机程序时实现如本申请实施例的第一方面提供的动作检测方法。

本申请实施例的第四方面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如本申请实施例的第一方面提供的动作检测方法。

5 本申请实施例的第五方面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，使得终端设备执行本申请实施例的第一方面所述的动作检测方法。

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。

10

附图说明

15 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。

图 1 是本申请实施例提供的一种动作检测方法的流程图；

图 2 是对像素帧序列和深度图序列分别进行时空信息表示，获得对应的像素特征图像和深度特征图像的示意图；

图 3 是本申请实施例提供的一种深度神经网络的结构示意图；

20 图 4 是本申请实施例提供的一种特征交互模块的结构示意图；

图 5 是采用视觉显著性算法将人体分割标注成多个指定部位的示意图；

图 6 是本申请实施例提供的一种人体动作检测与动作规范化评价方法的流程示意图；

图 7 是本申请实施例提供的一种动作检测装置的结构图；

25 图 8 是本申请实施例提供的一种终端设备的示意图。

具体实施方式

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，

5 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。另外，在本申请说明书和所附权利要求书的描述中，术语“第一”、“第二”、“第三”等仅用于区分描述，而不能理解为指示或暗示相对重要性。

本申请提出一种基于计算机视觉的动作检测方法，通过将视频图像的像素
10 特征和深度特征融合，能够有效利用像素信息和深度信息的互补性，极大地提高整体动作识别效率和模型的抗干扰能力。另外，在识别出动作之后，还能进一步对该动作的规范性进行评价。本申请更具体的技术实现细节，请参照下文所述的方法实施例。

应当理解，本申请各个方法实施例的执行主体为各种类型的终端设备或服务
15 器，比如手机、平板电脑、笔记本电脑、台式电脑和各类可穿戴设备等。

请参阅图 1，示出了本申请实施例提供的一种动作检测方法，包括：

101、获取包含目标动作的目标视频序列，所述目标视频序列包括一一对应的目标像素帧序列和目标深度图序列；

首先，获取一个包含目标动作的目标视频序列。其中，目标动作是待识别
20 的动作，其可以是任何目标物体（例如人体、动物或者机器人等）执行的一个任意类型的动作，例如人体的坐下、打保龄球和俯卧撑等动作。另外，该目标视频序列包含两部分的图像序列，第一部分是包含像素特征的目标像素帧序列（具体可以是 RGB 图像序列或者灰度图像序列等），第二部分是包含深度特征的目标深度图序列，两部分图像序列是一一对应的，即目标像素帧序列的第一
25 帧图像与目标深度图序列的第一帧图像对应，目标像素帧序列的第二帧图像与目标深度图序列的第二帧图像对应，以此类推。在实际操作中，可以采用某些

指定类型的摄像头（例如微软的 kinect 摄像头等），拍摄得到 RGB 图像序列及其对应的深度图像序列，两个图像序列中的每帧图像都是一一对应的。

在本申请的一种实现方式中，所述获取包含目标动作的目标视频序列，可以包括：

5 （1）获取包含多个动作的原始视频序列，所述原始视频序列包括一一对应的原始像素帧序列和原始深度图序列；

 （2）对所述原始像素帧序列执行视频的动作分割处理，得到多个包含单一动作的像素帧序列分段，并从所述多个像素帧序列分段中选取一个像素帧序列分段，作为所述目标像素帧序列；

10 （3）对所述原始深度图序列执行视频的动作分割处理，得到多个包含单一动作的深度图序列分段，并从所述多个深度图序列分段中选取与所述目标像素帧序列对应的深度图序列分段，作为所述目标深度图序列。

 在某些应用场合中，获取到的通常是包含多个不同动作的原始视频序列，例如针对某个人体在一定时间内的活动视频，其可能包含走路、坐下、站立和
15 跑步等多个不同动作。针对这些应用场合，可以采用动作分割的方法，将原始视频序列分割成一个个包含单一动作的视频序列分段，然后针对每个视频序列分段可以分别采用本申请提出的动作检测方法识别出各自包含的动作，从而实现整个原始视频序列包含的所有动作的识别。

 具体的，获取到的原始像素帧序列同样包括一一对应的原始像素帧序列和
20 原始深度图序列，可以分别对原始像素帧序列和原始深度图序列执行视频序列的动作分割处理，分别得到多个包含单一动作的像素帧序列分段和多个包含单一动作的深度图序列分段。然后，从该多个像素帧序列分段中选取一个像素帧序列分段（包含当前需要识别的目标动作的像素帧序列分段），作为目标像素帧序列；从该多个深度图序列分段中选取与目标像素帧序列对应的深度图序列
25 分段（即包含当前需要识别的目标动作的深度图序列分段），作为目标深度图序列。

在实际操作中,可以采用基于运动量(Quantity of movement, QOM)的动作分割方法。对于包含多个不同动作的原始视频序列,其包含的每一帧图像具有相对于其相邻帧图像和第一帧图像的相关移动信息,可以根据其中对应的运动量来检测每个动作的开始帧和结束帧,从而实现动作分割。例如,假设某个原始视频序列为 I , 可以定义其包含的第 t 帧图像的 QOM 为:

$$QOM_t = \sum_{m, n} \Psi(I_t(m, n), I_1(m, n))$$

其中, (m, n) 表示图像中的像素坐标, $\Psi(x, y)$ 定义为:

$$\Psi(x, y) = \begin{cases} 1, & \text{if } |x - y| \geq Threshold_{QOM} \\ 0, & \text{otherwise} \end{cases}$$

$Threshold_{QOM}$ 是一个预设的参数,根据经验值一般可以设为 60。另外设置一个参数 $Threshold_{inter}$ 作为动作内分割的门限值,可以通过滑窗的方法来对其进行迭代更新。假设某个动作的平均长度为 L 帧图像,则可以以其前 12.5% 和后 12.5% 帧图像的平均 QOM 值作为 $Threshold_{inter}$ 的候选数值。然后,通过比对应选取 L 帧图像长度内的最小 QOM 值所对应的帧作为动作分界帧(开始帧和结束帧),以此完成动作分割。总的来说,基于 QOM 的动作分割方法,主要通过运动量变化和时间尺度的结合,来实现视频序列的动作检测与分割,最终获得的是仅包含单一动作的视频序列分段。

102、根据所述目标像素帧序列生成像素特征图像,所述像素特征图像包含所述目标像素帧序列具有的各帧图像的特征;

在获得目标像素帧序列之后,可以基于该目标像素帧序列具有的各帧图像生成一幅像素特征图像。该像素特征图像包含目标像素帧序列具有的每帧图像的特征,可以用于表征目标像素帧序列的整体像素特征。例如,假设目标像素帧序列为 N 帧 RGB 图像,则可以基于该 N 帧 RGB 图像生成一幅类似 RGB 的像素特征图像,用于表征该 N 帧 RGB 图像的整体像素特征。这个过程获得的

像素特征图像，可以称作目标像素帧序列的时空信息表示。

在本申请的一种实现方式中，所述根据所述目标像素帧序列生成像素特征图像，可以包括：

5 (1) 在时间维度上对所述目标像素帧序列执行稀疏采样处理，得到第一图像序列；

(2) 将所述第一图像序列包含的各帧图像融合，得到所述像素特征图像。

可以采用将目标像素帧序列包含的各帧图像融合的方式获得像素特征图像，然而，由于目标像素帧序列包含很多帧图像，若将这些图像都进行融合会产生大量的计算量，影响算法运行速度。因此，可以先在时间维度上对目标像素帧序列执行稀疏采样处理，从而去除帧间的冗余信息，降低计算量。另外，该稀疏采样可以采用平均采样的方式，以避免产生动作表示不均匀，丢失空间维度信息的问题。例如，假设目标像素帧序列包含 100 帧图像，则可以使用稀疏平均采样的方式，从中提取出第 5 帧、第 15 帧、第 25 帧...第 95 帧总共 10 帧图像，作为得到的第一图像序列，即可以用该 10 帧图像来表示整个目标动作。然后，可以采用图像叠加等方式将该第一图像序列包含的各帧图像融合，从而得到对应的像素特征图像。

具体的，所述将所述第一图像序列包含的各帧图像融合，得到所述像素特征图像，可以包括：

20 对所述第一图像序列包含的各帧图像按照对应位置像素点执行图像特征的叠加、求平均和取整操作，得到所述像素特征图像。

假设第一图像序列是 RGB 图像序列，则这个过程可以将 RGB 三通道视作一个矢量矩阵，即该 RGB 图像序列中的每帧 RGB 图像都具有一个对应的矢量矩阵，将这些矢量矩阵包含的像素特征叠加、求平均再取整后，就可以得到一个最终的矢量矩阵，其即为该像素特征图像对应的矢量矩阵，也可以称作该 RGB 图像序列的时空信息表示样本。例如，假设第一图像序列为 $\langle I_1, I_2, I_3, \dots, I_T \rangle$ ，即共包含 T 帧图像，则其对应的像素特征图像 M 可以表示为：

$$M = \frac{1}{T} \sum_{i=1}^T I_i$$

103、根据所述目标深度图序列生成深度特征图像，所述深度特征图像包含所述目标深度图序列具有的各帧图像的特征；

与步骤 102 类似的，在获得目标深度图序列之后，同样可以基于该目标深度图序列具有的各帧图像生成一幅深度特征图像。该深度特征图像包含目标深度图序列具有的每帧图像的特征，可以用于表征目标深度图序列的整体深度特征。例如，假设目标深度图序列为 N 帧深度图像，则可以基于该 N 帧深度图像生成一幅近似的深度图像，用于表征该 N 帧深度图像的整体深度特征。这个过程获得的深度特征图像，可以称作目标深度图序列的时空信息表示。

在本申请的一种实现方式中，所述根据所述目标深度图序列生成深度特征图像，可以包括：

(1) 在时间维度上对所述目标深度图序列执行稀疏采样处理，得到第二图像序列；

(2) 将所述第二图像序列包含的各帧图像融合，得到所述深度特征图像。

与根据目标像素帧序列生成像素特征图像的方法类似，同样可以在时间维度上对目标深度图序列执行稀疏采样处理，从而去除帧间的冗余信息，降低计算量。例如，假设目标深度图序列包含 100 帧深度图像，则可以使用稀疏平均采样的方式，从中提取出第 5 帧、第 15 帧、第 25 帧...第 95 帧（帧标可以与步骤 102 中第一图像序列的帧标一一对应）总共 10 帧深度图像，作为得到的第二图像序列，即可以用该 10 帧深度图像来表示整个目标动作。然后，可以采用图像叠加等方式将该第二图像序列包含的各帧深度图像融合，从而得到对应的深度特征图像。

具体的，所述将所述第二图像序列包含的各帧图像融合，得到所述深度特征图像，可以包括：

(1) 将所述第二图像序列包含的各帧图像分别转换成各帧灰度图像；

(2) 对所述各帧灰度图像按照对应位置像素点执行图像特征的叠加、求平均和取整操作，得到所述深度特征图像。

对于由深度图像构成的第二图像序列，由于其表示的是距离信息，故需要先通过缩放等方式将其包含的各帧深度图像转换成灰度值为 0-255 的灰度图像，此时可视作单通道的矢量矩阵，从而方便地实现图像融合。也即，第二图像序列中的每帧深度图像都具有一个对应的单通道的矢量矩阵，将这些矢量矩阵包含的像素特征叠加、求平均再取整后，就可以得到一个最终的矢量矩阵，其即为该深度特征图像对应的矢量矩阵，也可以称作该第二图像序列的时空信息表示样本。

如图 2 所示，是对像素帧序列和深度图序列分别进行时空信息表示，获得对应的像素特征图像和深度特征图像的示意图。可以看出，获得的像素特征图像包含该像素帧序列的整体像素特征；获得的深度特征图像包含该深度图序列的整体深度特征。

104、将所述像素特征图像和所述深度特征图像输入已训练的深度神经网络执行图像特征的提取与融合处理，以确定所述目标动作的类别。

在通过步骤 102 获得像素特征图像以及通过步骤 103 获得深度特征图像之后，可以将该两帧特征图像输入一个预先训练完成的深度神经网络进行处理。该深度神经网络通过执行图像特征的提取与融合等方式，能够联合学习特征图像中的像素特征和深度特征，最终输出一个类别标签，从而确定目标动作的类别。在实际操作中，该深度神经网络可以选用 Resnet、inception 和 VGG 等成熟的网络模型架构，本申请不对该深度神经网络的类型和结构进行限定。

在本申请的一种实现方式中，所述深度神经网络包括特征提取模块和特征融合模块，所述将所述像素特征图像和所述深度特征图像输入已训练的深度神经网络执行图像特征的提取与融合处理，以确定所述目标动作的类别，可以包括：

(1) 将所述像素特征图像和所述深度特征图像输入所述特征提取模块进行

处理，得到图像语义特征；

(2) 将所述图像语义特征输入所述特征融合模块进行处理，得到融合后的图像特征；

(3) 基于所述融合后的图像特征确定所述目标动作的类别。

5 可以采用多级卷积层和池化层的结构构建特征提取模块，用于提取像素特征图像和深度特征图像的图像语义特征；然后，通过包含卷积神经单元的特征融合模块来实现图像语义特征的融合（主要采用特征的点乘、加权求和或者求最大值等方式实现融合），得到融合后的图像特征。该融合后的图像特征是目标动作与其它动作之间的区别性特征，故可以基于该特征实现动作识别，即确定目标动作的类别。

10 该深度神经网络的一种结构示意图如图 3 所示。在图 3 中，该深度神经网络包含特征提取模块和特征融合模块两部分，其中特征提取模块主要由多级级联的卷积层、特征交互模块和全连接层构成，特征融合模块主要由多级级联的卷积神经单元（用图中 3 的圆圈表示）构成。该特征交互模块的一种结构示意图如图 4 所示，该模块主要包含两个卷积核 1×1 的卷积层，将两个模态的图像特征（像素特征和深度特征）输入该模块之后，能够对图像特征进行中层语义特征和高层语义特征（中层语义特征一般指在网络模型参数学习过程中所学习到的特征，高层语义特征一般指网络模型学习完成后输出的能够对样本进行分类的特征）的互补学习。

20 假设像素特征图像是三通道的 RGB 图像，深度特征图像是单通道的深度图像，为了适应该深度神经网络，可以将两个模态的特征图像进行联结，获得四通道的样本。例如，RGB 图像可以表示为三通道的矢量矩阵，深度图像可以表示为单通道的矢量矩阵，那么将两个矢量矩阵拼接，即可得到一个四通道的矢量矩阵。相应的，该深度神经网络的输入端通道数也是四个，将四通道的样本输入该深度神经网络后，经过多级卷积层、池化层以及前文所述的交互学习模块的处理，能够学习到样本的中层语义特征和高层语义特征；之后，通过特

征融合模块将高层语义特征融合，得到目标动作的区别性特征，最终基于该区别性特征完成动作分类。

在某些应用场景中，除识别出用户的动作类型之外，还需要进一步检测该动作是否规范，并给出相应的规范性评价结果，以便纠正用户的错误动作。有鉴于此，在本申请的一种实现方式中，在确定所述目标动作的类别之后，还可以包括：

(1)从预设的规范动作视频库中查找与所述目标动作的类别对应的基准视频序列，所述基准视频序列包含规范化的所述目标动作；

(2)在时间维度上对所述基准视频序列执行稀疏采样处理，得到第三图像序列，所述第三图像序列包含的图像帧数和所述第一图像序列包含的图像帧数相同；

(3)分别标注出所述第一图像序列包含的各帧图像中目标物体具有的指定部位的位置，以及所述第三图像序列包含的各帧图像中所述目标物体具有的所述指定部位的位置，所述目标物体为执行所述目标动作的物体；

(4)根据所述第一图像序列包含的各帧图像中所述目标物体具有的所述指定部位的位置，构建得到所述指定部位对应的第一运动轨迹曲线；

(5)根据所述第三图像序列包含的各帧图像中所述目标物体具有的所述指定部位的位置，构建得到所述指定部位对应的第二运动轨迹曲线；

(6)根据所述第一运动轨迹曲线和所述第二运动轨迹曲线，计算得到曲线误差；

(7)根据所述曲线误差确定所述目标视频序列包含的所述目标动作的规范化程度。

可以预先在指定的存储区域(例如某个数据库)构建一个规范动作视频库，将规范化的各种类型动作的视频序列(例如人体标准跑步动作的视频序列、人体标准俯卧撑动作的视频序列等)存入该规范动作视频库。在确定目标动作的类别之后，可以从该规范动作视频库中查找与该目标动作的类型对应的基准视

频序列，该基准视频序列包含规范化的所述目标动作。例如，若目标动作为跑步，则基准视频序列为人体标准跑步动作的视频序列。

然后，在时间维度上对基准视频序列执行稀疏采样处理，得到第三图像序列，需要注意第三图像序列包含的图像帧数和前文所述的第一图像序列包含的图像帧数是相同的。例如，假设第一图像序列为包含 T 帧图像的 $\langle I_1, I_2, I_3, \dots, I_T \rangle$ ，
5 则第三图像序列可以表示为包含 T 帧图像的 $\langle N_1, N_2, N_3, \dots, N_T \rangle$ 。

接着，分别标注出第一图像序列包含的各帧图像中目标物体具有的指定部位的位置，以及第三图像序列包含的各帧图像中目标物体具有的该指定部位的位置；其中，该目标物体是执行目标动作的物体，该指定部位的数量可以是一个，也可以是多个（为了提高后续曲线误差计算的准确率，一般需要设置多个）。
10 例如，若目标物体是人体，则可以采用视觉显著性算法等方式将人体分割标注成头部、躯干、左手、右手、左脚和右脚等 6 个指定部位，如图 5 所示。

之后，根据第一图像序列包含的各帧图像中该目标物体具有的各个指定部位的位置，基于时间维度构建出各个指定部位分别对应的第一运动轨迹曲线。
15 例如，针对人体的 6 个指定部位，可以将第一图像序列包含的各帧图像中人体头部中心点（可以采用归一化的操作，将人体头部表示为一个中心点，具体可以计算人体头部包含的所有坐标点的平均坐标值作为中心点，针对其它指定部位均可以采用相同的操作方式）的坐标连接起来，得到人体头部对应的第一运动轨迹曲线；将第一图像序列包含的各帧图像中人体躯干中心点的坐标连接起来，
20 得到人体躯干对应的第一运动轨迹曲线，以此类推，分别构建出 6 个指定部位对应的第一运动轨迹曲线。

同样的，可以根据第三图像序列包含的各帧图像中该目标物体具有的各个指定部位的位置，基于时间维度构建得到各个指定部位分别对应的第二运动轨迹曲线。也即，针对人体的 6 个指定部位，可以将第三图像序列包含的各帧图
25 像中人体头部中心点的坐标连接起来，得到人体头部对应的第二运动轨迹曲线；将第三图像序列包含的各帧图像中人体躯干中心点的坐标连接起来，得到人体

躯干对应的第二运动轨迹曲线，以此类推，分别构建出 6 个指定部位对应的第二运动轨迹曲线。

接下来，根据构建的第一运动轨迹曲线和第二运动轨迹曲线，可以计算得到曲线误差。假设指定部位只有一个，即第一运动轨迹曲线和第二运动轨迹曲线都只有一条，那么曲线误差可以是该第一运动轨迹曲线和该第二运动轨迹曲线之差。具体的，可以分别计算第一运动轨迹曲线中的每个目标位置点和其在第二运动轨迹曲线中的对应位置点之间的距离，得到每个目标位置点的误差，然后将每个目标位置点的误差叠加，得到曲线误差。这里的目标位置点可以是运动轨迹曲线连接的节点，例如，假设指定部位为人体头部，则第一运动轨迹曲线中的各个目标位置点可以为第一图像序列包含的各帧图像中的人体头部中心点，各个目标位置点在第二运动轨迹曲线中的对应位置点可以为第三图像序列包含的各帧图像中的人体头部中心点。该曲线误差可以采用以下公式计算：

$$err = \sum_{t=1}^T \|R_t^N - R_t^I\|^2$$

其中，err 表示曲线误差， $t=1,2,3...T$ ，T 表示第一图像序列和第三图像序列的图像帧数， $R_t^N - R_t^I$ 表示第一运动轨迹曲线中的目标位置点和其在第二运动轨迹曲线中的对应位置点之间的距离。

假设指定部位有多个，即第一运动轨迹曲线和第二运动轨迹曲线都有多条，则可以采用以上方法分别计算得到各个指定部位的曲线误差分量，最后再将各个指定部位的曲线误差分量相加，得到总的曲线误差。例如，针对人体的 6 个指定部位，可以分别计算得到 6 个曲线误差分量： $err_{\text{头部}}$ 、 $err_{\text{躯干}}$ 、 $err_{\text{左手}}$ 、 $err_{\text{右手}}$ 、 $err_{\text{左脚}}$ 和 $err_{\text{右脚}}$ ，则总的曲线误差 $err_{\text{总}}$ 可以表示为：

$$err_{\text{总}} = err_{\text{头部}} + err_{\text{躯干}} + err_{\text{左手}} + err_{\text{右手}} + err_{\text{左脚}} + err_{\text{右脚}}$$

在计算得到曲线误差之后，可以根据该曲线误差确定目标视频序列包含的

目标动作的规范化程度。具体的，该曲线误差可以用于表征目标视频序列包含的目标动作和基准视频序列包含的规范化目标动作（标准动作）之间的偏差程度，也即若该曲线误差越小，则表示目标视频序列包含的目标动作和标准动作之间的偏差越小，即该目标动作的规范化程度越高。在实际操作中，可以先对该曲线误差执行归一化处理，这里可以采用计算误差倒数的方法来处理，即：

$$1/\text{err}_{\text{总}} = 1/\text{err}_{\text{头部}} + 1/\text{err}_{\text{躯干}} + 1/\text{err}_{\text{左手}} + 1/\text{err}_{\text{右手}} + 1/\text{err}_{\text{左脚}} + 1/\text{err}_{\text{右脚}}$$

然后，可以通过 softmax 函数将归一化后的曲线误差转换为概率值，该概率值越大则表示待测样本（即目标视频序列）和规范化动作样本（即基准视频序列）越接近，即待测样本的规范化程度越高。反之，则表示待测样本和规范化动作样本的偏差越大，即待测样本的规范化程度越低。

在本申请实施例中，首先，获取包含目标动作的目标视频序列，该目标视频序列包括一一对应的像素帧序列和深度图序列；然后，根据像素帧序列生成包含其中各帧图像的图像特征的像素特征图像，以及根据深度图序列生成包含其中各帧图像的图像特征的深度特征图像；最后，将像素特征图像和深度特征图像输入一个已训练的深度神经网络进行图像特征的提取与融合处理，从而确定目标动作的类别。上述过程将视频图像的像素特征和深度特征融合，利用像素信息和深度信息的互补性，能够在一定程度上减弱环境因素对检测样本的干扰，从而提高动作检测的准确率。

为便于理解本申请实施例提出的动作检测与动作规范化评价方法，以下列举一个实际的应用场景。如图 6 所示，是本申请实施例提出的一种人体动作检测与动作规范化评价方法的流程示意图。

在图 6 中，首先输入原始视频序列，然后对该原始视频序列执行 QOM 动作分割处理，得到包含单一人体动作的目标视频序列；接着，对目标视频序列执行稀疏平均采样处理，得到第一图像序列；然后，获取第一图像序列的时空信息表示得到像素特征图像，将其和对应的深度特征图像输入深度神经网络进行识别，得到动作类别；之后，从规范动作视频库中查找与该动作类别对应的

基准视频序列，对基准视频序列执行稀疏平均采样处理，得到第三图像序列；
然后，采用显著性检测的方式，分别标注出第一图像序列包含的各帧图像中人体头部、躯干、左手、右手、左脚和右脚 6 个指定部位的位置，并据此构建得到对应的第一运动轨迹曲线，以及标注出第三图像序列包含的各帧图像中该 6
5 个指定部位的位置，并据此构建得到对应的第二运动轨迹曲线；接下来，根据两部分运动轨迹曲线计算得到对应的曲线误差，最后对曲线误差执行归一化处理，通过 softmax 函数将归一化后的曲线误差转换为概率值，并根据该概率值的大小评估目标视频序列包含的人体动作的规范化程度。

综上所述，本申请实施例通过将视频图像的像素特征和深度特征融合，能够
10 够有效利用像素信息和深度信息的互补性，极大地提高整体动作识别效率和模型的抗干扰能力。而且，通过将待测样本的动作和规范化动作样本的动作进行向量化的误差比对，能够实现动作的规范化程度评价，在体育锻炼等领域中具有重大的应用价值。

应理解，上述各个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，而不应对本申请实施例的实施过程构成任何限定。

上面主要描述了一种动作检测方法，下面将对一种动作检测装置进行描述。

请参阅图 7，本申请实施例中一种动作检测装置的一个实施例包括：

视频序列获取模块 701，用于获取包含目标动作的目标视频序列，所述目标视频序列包括一一对应的目标像素帧序列和目标深度图序列；
20

像素特征生成模块 702，用于根据所述目标像素帧序列生成像素特征图像，所述像素特征图像包含所述目标像素帧序列具有的各帧图像的特征；

深度特征生成模块 703，用于根据所述目标深度图序列生成深度特征图像，所述深度特征图像包含所述目标深度图序列具有的各帧图像的特征；

25 动作检测模块 704，用于将所述像素特征图像和所述深度特征图像输入已训练的神经网络执行图像特征的提取与融合处理，以确定所述目标动作的

类别。

在本申请的一种实现方式中，所述深度神经网络包括特征提取模块和特征融合模块，所述动作检测模块可以包括：

特征提取单元，用于将所述像素特征图像和所述深度特征图像输入所述特征提取模块进行处理，得到图像语义特征；

特征融合单元，用于将所述图像语义特征输入所述特征融合模块进行处理，得到融合后的图像特征；

动作检测单元，用于基于所述融合后的图像特征确定所述目标动作的类别。

在本申请的一种实现方式中，所述像素特征生成模块可以包括：

第一稀疏采样处理单元，用于在时间维度上对所述目标像素帧序列执行稀疏采样处理，得到第一图像序列；

第一图像融合单元，用于将所述第一图像序列包含的各帧图像融合，得到所述像素特征图像；

所述深度特征生成模块可以包括：

第二稀疏采样处理单元，用于在时间维度上对所述目标深度图序列执行稀疏采样处理，得到第二图像序列；

第二图像融合单元，用于将所述第二图像序列包含的各帧图像融合，得到所述深度特征图像。

进一步的，所述第一图像融合单元可以包括：

像素特征处理子单元，用于对所述第一图像序列包含的各帧图像按照对应位置像素点执行图像特征的叠加、求平均和取整操作，得到所述像素特征图像；

所述第二图像融合单元可以包括：

灰度转换子单元，用于将所述第二图像序列包含的各帧图像分别转换成各帧灰度图像；

深度特征处理子单元，用于对所述各帧灰度图像按照对应位置像素点执行图像特征的叠加、求平均和取整操作，得到所述深度特征图像。

在本申请的一种实现方式中，所述动作检测装置还可以包括：

基准视频查找模块，用于从预设的规范动作视频库中查找与所述目标动作的类别对应的基准视频序列，所述基准视频序列包含规范化的所述目标动作；

稀疏采样处理模块，用于在时间维度上对所述基准视频序列执行稀疏采样
5 处理，得到第三图像序列，所述第三图像序列包含的图像帧数和所述第一图像序列包含的图像帧数相同；

显著性标注模块，用于分别标注出所述第一图像序列包含的各帧图像中目标物体具有的指定部位的位置，以及所述第三图像序列包含的各帧图像中所述
10 目标物体具有的所述指定部位的位置，所述目标物体为执行所述目标动作的物体；

第一曲线构建模块，用于根据所述第一图像序列包含的各帧图像中所述目标物体具有的所述指定部位的位置，构建得到所述指定部位对应的第一运动轨迹曲线；

第二曲线构建模块，用于根据所述第三图像序列包含的各帧图像中所述目标物体具有的所述指定部位的位置，构建得到所述指定部位对应的第二运动轨迹
15 曲线；

曲线误差计算模块，用于根据所述第一运动轨迹曲线和所述第二运动轨迹曲线，计算得到曲线误差；

规范化评价模块，用于根据所述曲线误差确定所述目标视频序列包含的所述
20 目标动作的规范化程度。

进一步的，所述曲线误差计算模块可以包括：

位置点误差计算单元，用于分别计算所述第一运动轨迹曲线中的每个目标位置点和其在所述第二运动轨迹曲线中的对应位置点之间的距离，得到每个所述目标位置点的误差；

误差叠加单元，用于将每个所述目标位置点的误差叠加，得到所述曲线误差。
25

在本申请的一种实现方式中，所述视频序列获取模块可以包括：

原始视频序列获取单元，用于获取包含多个动作的原始视频序列，所述原始视频序列包括一一对应的原始像素帧序列和原始深度图序列；

5 第一动作分割处理单元，用于对所述原始像素帧序列执行视频的动作分割处理，得到多个包含单一动作的像素帧序列分段，并从所述多个像素帧序列分段中选取一个像素帧序列分段，作为所述目标像素帧序列；

第二动作分割处理单元，用于对所述原始深度图序列执行视频的动作分割处理，得到多个包含单一动作的深度图序列分段，并从所述多个深度图序列分段中选取与所述目标像素帧序列对应的深度图序列分段，作为所述目标深度图
10 序列。

本申请实施例还提供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如图 1 表示的任意一种动作检测方法。

本申请实施例还提供一种计算机程序产品，当该计算机程序产品在终端设备上运行时，使得终端设备执行实现如图 1 表示的任意一种动作检测方法。
15

图 8 是本申请一实施例提供的终端设备的示意图。如图 8 所示，该实施例的终端设备 8 包括：处理器 80、存储器 81 以及存储在所述存储器 81 中并可在所述处理器 80 上运行的计算机程序 82。所述处理器 80 执行所述计算机程序 82 时实现上述各个动作检测方法的实施例中的步骤，例如图 1 所示的步骤 101 至
20 104。或者，所述处理器 80 执行所述计算机程序 82 时实现上述各装置实施例中各模块/单元的功能，例如图 7 所示模块 701 至 704 的功能。

所述计算机程序 82 可以被分割成一个或多个模块/单元，所述一个或者多个模块/单元被存储在所述存储器 81 中，并由所述处理器 80 执行，以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机程序指令段，该指令段用于描述所述计算机程序 82 在所述终端设备 8 中的执行过程。
25

所称处理器 80 可以是中央处理单元(Central Processing Unit, CPU)，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
5 器也可以是任何常规的处理器等。

所述存储器 81 可以是所述终端设备 8 的内部存储单元, 例如终端设备 8 的硬盘或内存。所述存储器 81 也可以是所述终端设备 8 的外部存储设备, 例如所述终端设备 8 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
10 存储器 81 还可以既包括所述终端设备 8 的内部存储单元也包括外部存储设备。所述存储器 81 用于存储所述计算机程序以及所述终端设备所需的其他程序和数据。所述存储器 81 还可以用于暂时地存储已经输出或者将要输出的数据。

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
15 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。

所属领域的技术人员可以清楚地了解到, 为描述的方便和简洁, 上述描述的系统, 装置和单元的具体工作过程, 可以参考前述方法实施例中的对应过程,
20 在此不再赘述。

在上述实施例中, 对各个实施例的描述都各有侧重, 某个实施例中沒有详

述或记载的部分，可以参见其它实施例的相关描述。

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
5 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。

在本申请所提供的实施例中，应该理解到，所揭露的装置和方法，可以通过其它的方式实现。例如，以上所描述的系统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
10 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是通过一些接口，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
15 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。

另外，在本申请各个实施例中的各功能单元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
20 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。

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取存储介质中。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，也可以通过计算机程序来指令相关的
25 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

5 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可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、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等。需要说明的是，所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减，例如在某些司法管辖区，根据立法和专利实践，计算机可读介质不包括是电载波信号和电信信号。

10 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. 一种动作检测方法，其特征在于，包括：

5 获取包含目标动作的目标视频序列，所述目标视频序列包括一一对应的目标像素帧序列和目标深度图序列；

根据所述目标像素帧序列生成像素特征图像，所述像素特征图像包含所述目标像素帧序列具有的各帧图像的特征；

根据所述目标深度图序列生成深度特征图像，所述深度特征图像包含所述目标深度图序列具有的各帧图像的特征；

10 将所述像素特征图像和所述深度特征图像输入已训练的神经网络执行图像特征的提取与融合处理，以确定所述目标动作的类别。

2. 如权利要求 1 所述的方法，其特征在于，所述神经网络包括特征提取模块和特征融合模块，所述将所述像素特征图像和所述深度特征图像输入已训练的神经网络执行图像特征的提取与融合处理，以确定所述目标动作的类别，包括：

将所述像素特征图像和所述深度特征图像输入所述特征提取模块进行处理，得到图像语义特征；

将所述图像语义特征输入所述特征融合模块进行处理，得到融合后的图像特征；

20 基于所述融合后的图像特征确定所述目标动作的类别。

3. 如权利要求 1 所述的方法，其特征在于，所述根据所述目标像素帧序列生成像素特征图像，包括：

在时间维度上对所述目标像素帧序列执行稀疏采样处理，得到第一图像序列；

25 将所述第一图像序列包含的各帧图像融合，得到所述像素特征图像；

所述根据所述目标深度图序列生成深度特征图像，包括：

在时间维度上对所述目标深度图序列执行稀疏采样处理，得到第二图像序列；

将所述第二图像序列包含的各帧图像融合，得到所述深度特征图像。

4. 如权利要求3所述的方法，其特征在于，所述将所述第一图像序列包含的各帧图像融合，得到所述像素特征图像，包括：

对所述第一图像序列包含的各帧图像按照对应位置像素点执行图像特征的叠加、求平均和取整操作，得到所述像素特征图像；

所述将所述第二图像序列包含的各帧图像融合，得到所述深度特征图像，包括：

- 10 将所述第二图像序列包含的各帧图像分别转换成各帧灰度图像；

对所述各帧灰度图像按照对应位置像素点执行图像特征的叠加、求平均和取整操作，得到所述深度特征图像。

5. 如权利要求3所述的方法，其特征在于，在确定所述目标动作的类别之后，还包括：

- 15 从预设的规范动作视频库中查找与所述目标动作的类别对应的基准视频序列，所述基准视频序列包含规范化的所述目标动作；

在时间维度上对所述基准视频序列执行稀疏采样处理，得到第三图像序列，所述第三图像序列包含的图像帧数和所述第一图像序列包含的图像帧数相同；

- 20 分别标注出所述第一图像序列包含的各帧图像中目标物体具有的指定部位的位置，以及所述第三图像序列包含的各帧图像中所述目标物体具有的所述指定部位的位置，所述目标物体为执行所述目标动作的物体；

根据所述第一图像序列包含的各帧图像中所述目标物体具有的所述指定部位的位置，构建得到所述指定部位对应的第一运动轨迹曲线；

- 25 根据所述第三图像序列包含的各帧图像中所述目标物体具有的所述指定部位的位置，构建得到所述指定部位对应的第二运动轨迹曲线；

根据所述第一运动轨迹曲线和所述第二运动轨迹曲线，计算得到曲线误差；

根据所述曲线误差确定所述目标视频序列包含的所述目标动作的规范化程度。

6. 如权利要求 5 所述的方法，其特征在于，所述根据所述第一运动轨迹曲线和所述第二运动轨迹曲线，计算得到曲线误差，包括：

5 分别计算所述第一运动轨迹曲线中的每个目标位置点和其在所述第二运动轨迹曲线中的对应位置点之间的距离，得到每个所述目标位置点的误差；

 将每个所述目标位置点的误差叠加，得到所述曲线误差。

7. 如权利要求 1 至 6 中任一项所述的方法，其特征在于，所述获取包含目标动作的目标视频序列，包括：

10 获取包含多个动作的原始视频序列，所述原始视频序列包括一一对应的原始像素帧序列和原始深度图序列；

 对所述原始像素帧序列执行视频的动作分割处理，得到多个包含单一动作的像素帧序列分段，并从所述多个包含单一动作的像素帧序列分段中选取一个像素帧序列分段，作为所述目标像素帧序列；

15 对所述原始深度图序列执行视频的动作分割处理，得到多个包含单一动作的深度图序列分段，并从所述多个包含单一动作的深度图序列分段中选取与所述目标像素帧序列对应的深度图序列分段，作为所述目标深度图序列。

8. 一种动作检测装置，其特征在于，包括：

20 视频序列获取模块，用于获取包含目标动作的目标视频序列，所述目标视频序列包括一一对应的目标像素帧序列和目标深度图序列；

 像素特征生成模块，用于根据所述目标像素帧序列生成像素特征图像，所述像素特征图像包含所述目标像素帧序列具有的各帧图像的特征；

 深度特征生成模块，用于根据所述目标深度图序列生成深度特征图像，所述深度特征图像包含所述目标深度图序列具有的各帧图像的特征；

25 动作检测模块，用于将所述像素特征图像和所述深度特征图像输入已训练的神经网络执行图像特征的提取与融合处理，以确定所述目标动作的类别。

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 7 中任一项所述的动作检测方法。

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 7 中任一项所述的动作检测方法。

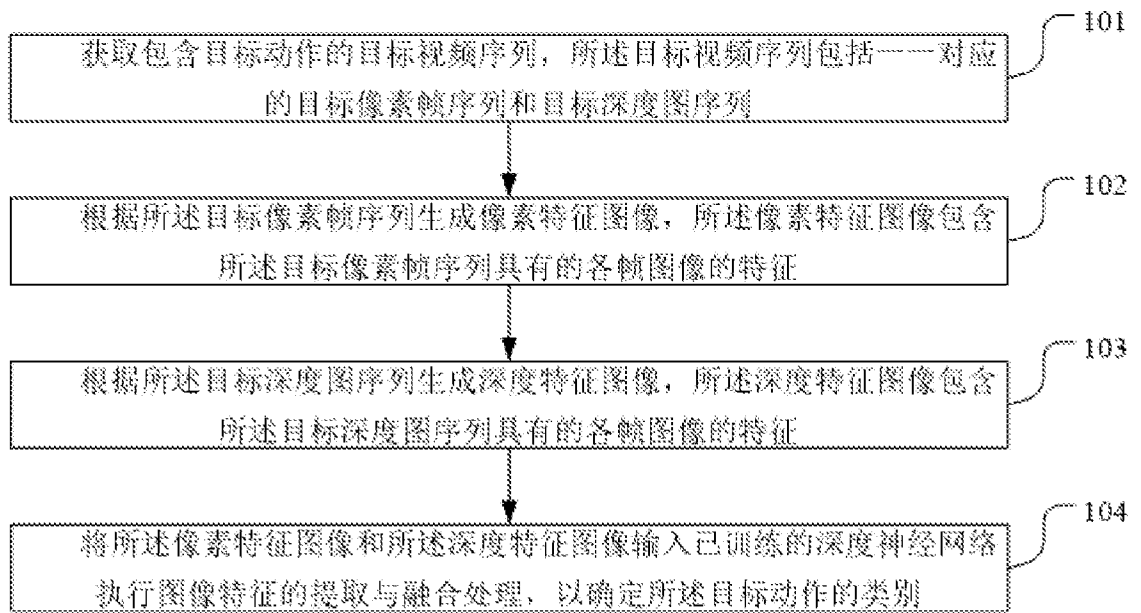


图 1

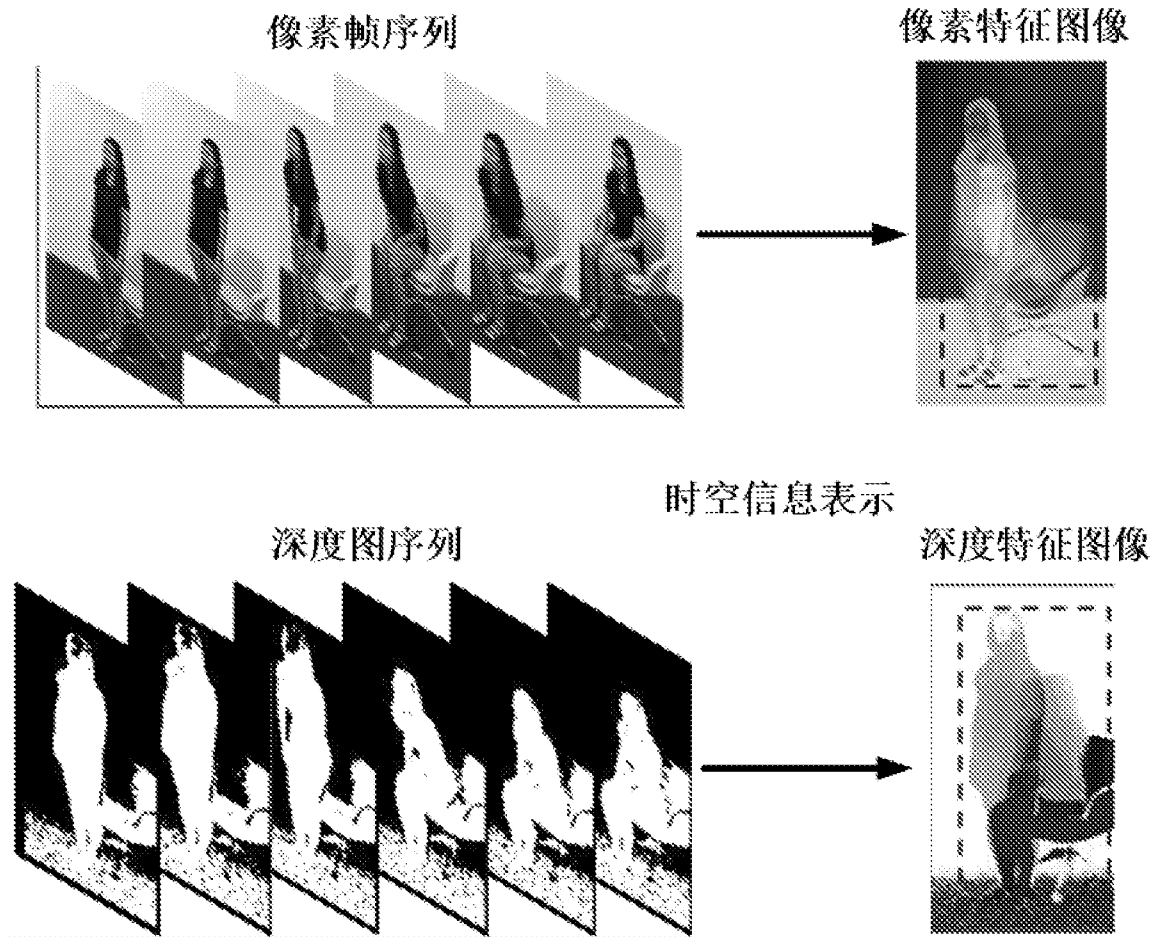


图 2

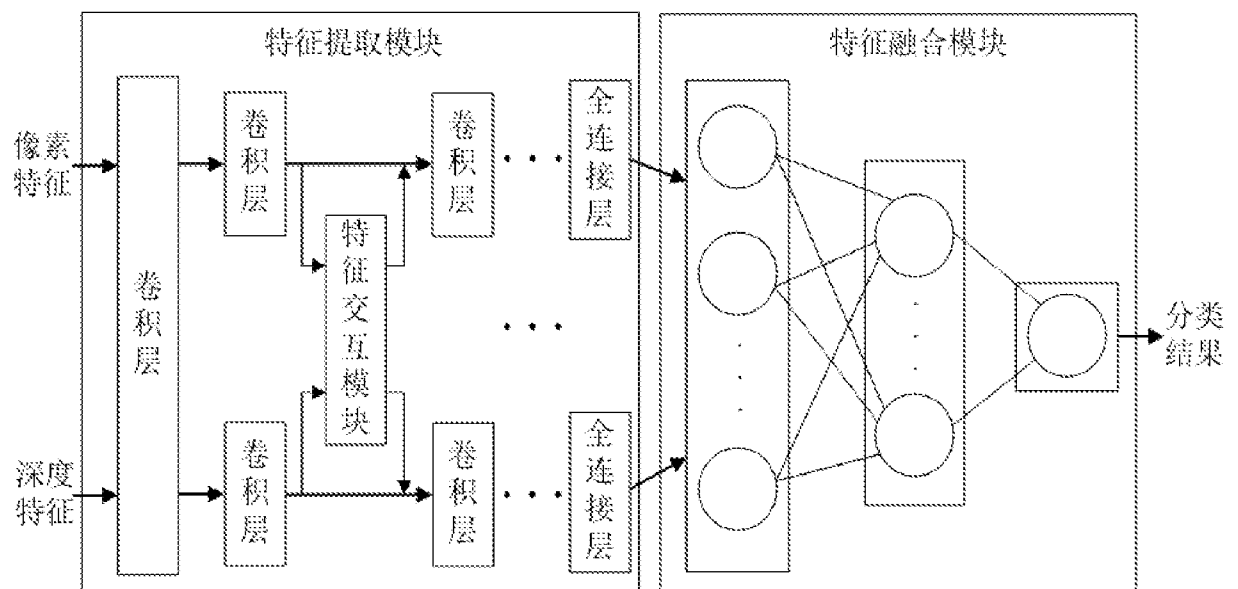


图 3

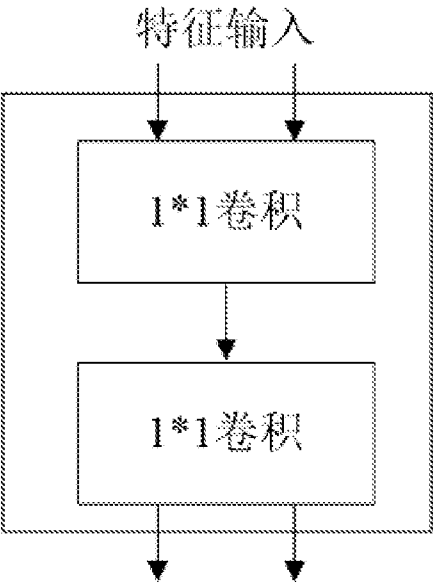


图 4

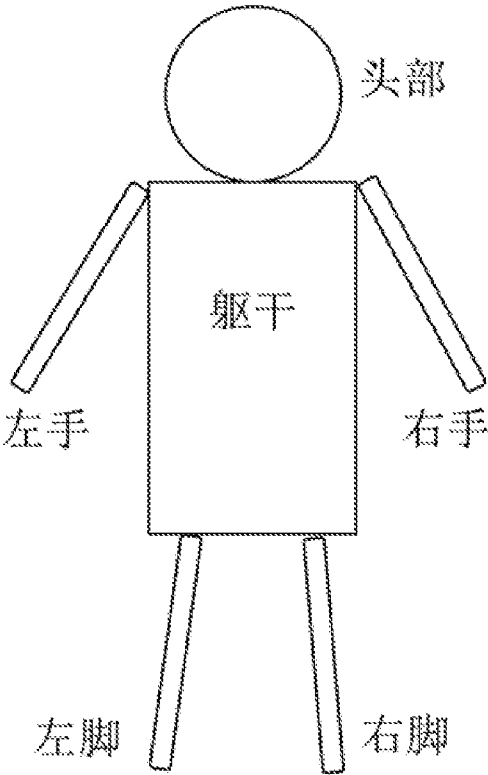


图 5

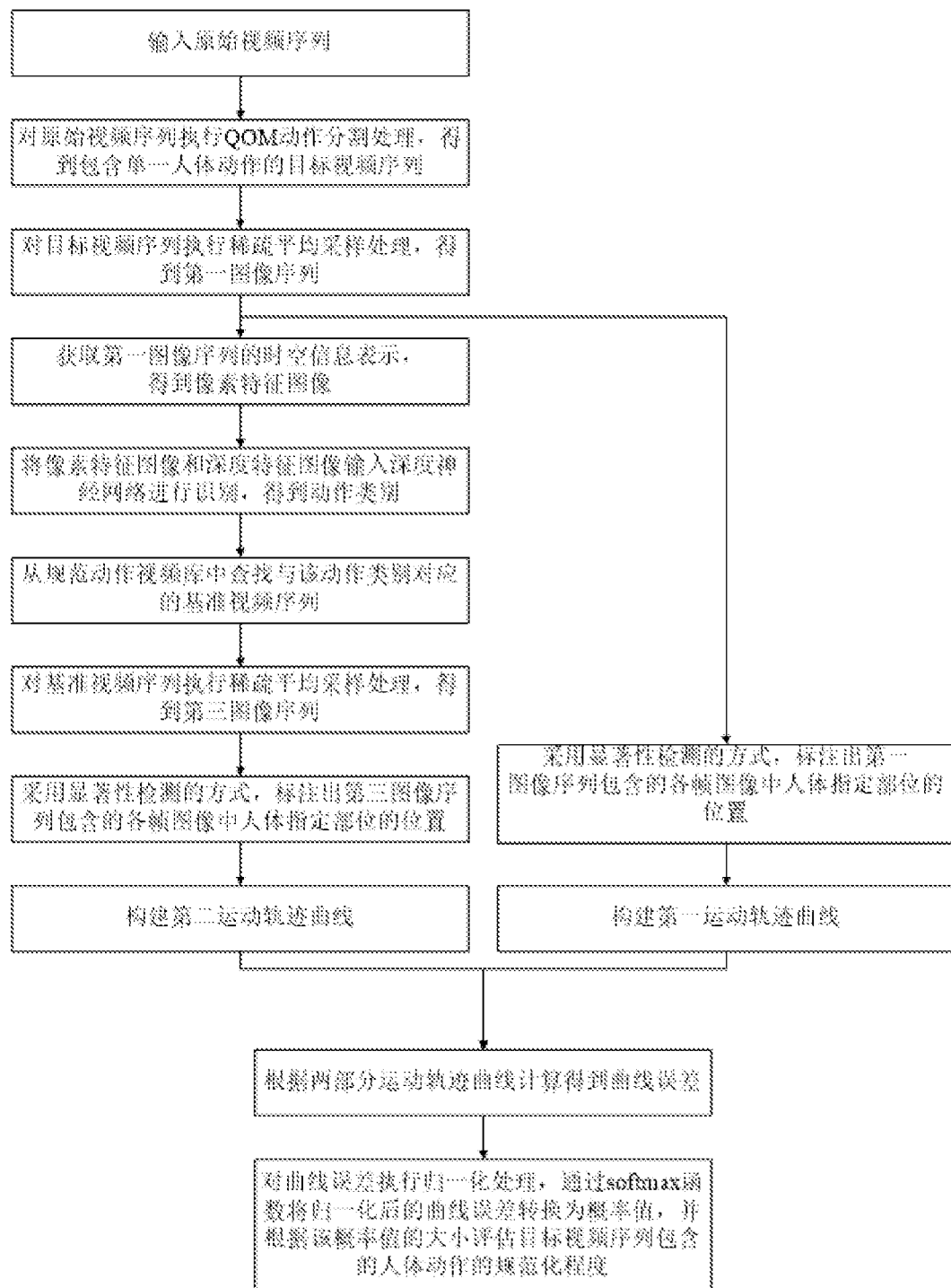


图 6

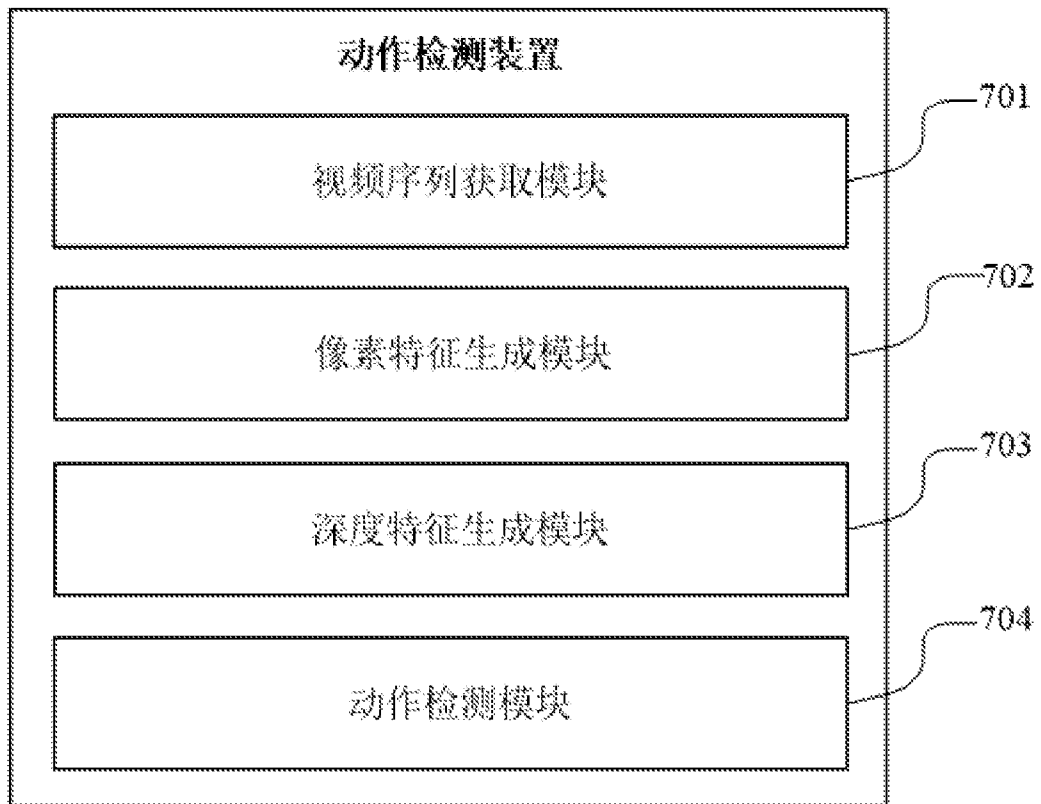


图 7

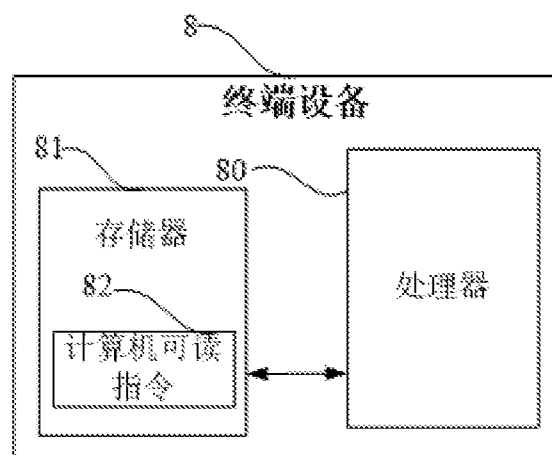


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/138566

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2022.01)i; G06K 9/62(2022.01)i; G06N 3/04(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 G06N

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)

CNPAT, CNKI, WPI, EPODOC, IEEE: 动作, 行为, 识别, 检测, 类别, 像素, 帧, 融合, 序列, 深度, 图, 灰度, action, behavior, recognis+, detect, sort, pixel, frame, fusion, serial, depth, iamge, gray w value

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113326835 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 31 August 2021 (2021-08-31) claims 1-10	1-10
X	CN 112257526 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 22 January 2021 (2021-01-22) description, paragraphs [0047]-[0116] and [0125]-[0134]	1-4, 7-10
A	CN 108108699 A (CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 01 June 2018 (2018-06-01) entire document	1-10
A	CN 112131928 A (ZHEJIANG UNIVERSITY OF TECHNOLOGY) 25 December 2020 (2020-12-25) entire document	1-10
A	US 2020265567 A1 (SAMSUNG ELECTRONICS CO., LTD.) 20 August 2020 (2020-08-20) 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

10 March 2022

Date of mailing of the international search report

29 March 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/CN2021/138566

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113326835	A	31 August 2021	None			
CN	112257526	A	22 January 2021	None			
CN	108108699	A	01 June 2018	None			
CN	112131928	A	25 December 2020	None			
US	2020265567	A1	20 August 2020	CN	113454981	A	28 September 2021
				WO	2020171373	A1	27 August 2020
				EP	3871405	A1	01 September 2021

国际检索报告

国际申请号

PCT/CN2021/138566

A. 主题的分类

G06K 9/00(2022.01)i; G06K 9/62(2022.01)i; G06N 3/04(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 G06N

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))

CNPAT, CNKI, WPI, EPDOC, IEEE: 动作, 行为, 识别, 检测, 类别, 像素, 帧, 融合, 序列, 深度, 图, 灰度, action, behavior, recognis+, detect, sort, pixel, frame, fusion, serial, depth, iamage, gray w value

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 113326835 A (中国科学院深圳先进技术研究院) 2021年8月31日 (2021 - 08 - 31) 权利要求1-10	1-10
X	CN 112257526 A (中国科学院深圳先进技术研究院) 2021年1月22日 (2021 - 01 - 22) 说明书第[0047]-[0116]、[0125]-[0134]段	1-4, 7-10
A	CN 108108699 A (重庆邮电大学) 2018年6月1日 (2018 - 06 - 01) 全文	1-10
A	CN 112131928 A (浙江工业大学) 2020年12月25日 (2020 - 12 - 25) 全文	1-10
A	US 2020265567 A1 (SAMSUNG ELECTRONICS CO., LTD.) 2020年8月20日 (2020 - 08 - 20) 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3月10日

国际检索报告邮寄日期

2022年3月29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

田民丽

电话号码 86-(10)-53961355

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/138566

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	113326835	A	2021年8月31日	无	
CN	112257526	A	2021年1月22日	无	
CN	108108699	A	2018年6月1日	无	
CN	112131928	A	2020年12月25日	无	
US	2020265567	A1	2020年8月20日	CN 113454981 A	2021年9月28日
				WO 2020171373 A1	2020年8月27日
				EP 3871405 A1	2021年9月1日