

动态手势识别方法和装置、手势交互控制方法和装置

本申请要求在 2017 年 12 月 22 日提交中国专利局、申请号为 CN201711417801.8、发明名称为“动态手势识别方法及装置”和在 2018 年 8 月 24 日提交中国专利局、申请号为 CN201810974244.8、发明名称为“动态动作检测方法、动态动作控制方法及装置”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

本申请实施例涉及图像处理技术，尤其是一种动态手势识别方法和装置、手势交互控制方法和装置。

背景技术

手势是图像与视频信息中一个重要的人机交互特征。手势识别算法核心任务是给定一张包含手的图片，判断出其中手势的类型的方法。

发明内容

本申请实施例提供一种动态手势识别技术方案以及手势交互控制技术方案。

根据本申请实施例一个方面，提供一种动态手势识别方法，包括：对待检测的视频流中的动态手势进行定位，得到动态手势框；从所述视频流的多帧图像帧中截取与所述动态手势框对应的图像块；基于所述截取出的图像块生成检测序列；根据所述检测序列进行动态手势识别。

根据本申请实施例另一个方面，提供一种动态手势识别建模方法，包括：采集不同动态手势类型的样本视频流；对所述不同动态手势类型的动态手势框进行标记；从样本视频流的多帧图像帧中截取与动态手势框的标注信息对应的图像块，构成图像序列；以动态手势类型作为监督数据，以所述图像序列作为训练数据，训练第一动态手势识别模型。

根据本申请实施例又一个方面，提供一种动态手势识别装置，包括：手势定位单元，用于对待检测的视频流中的动态手势进行定位，得到动态手势框；处理单元，用于从所述视频流的多帧图像帧中截取与所述动态手势框对应的图像块；检测序列生成单元，用于基于所述截取出的图像块生成检测序列；手势识别单元，用于根据所述检测序列进行动态手势识别。

根据本申请实施例还一个方面，提供一种动态手势识别模型建立装置，包括：第一动态手势识别模型建立单元；所述第一动态手势识别模型建立单元包括：样本采集子单元，用于采集不同动态手势类型的样本视频流；手势框标记子单元，用于对所述不同动态手势类型的动态手势框进行标记；图像序列构成子单元，用于从样本视频流的多帧图像帧中截取与动态手势框的标注信息对应的图像块，构成图像序列；训练子单元，用于以动态手势类型作为监督数据，以所述图像序列作为训练数据，训练第一动态手势识别模型。

根据本申请实施例的再一个方面，提供一种手势交互控制方法，包括：

获取视频流；

采用如上任意一项所述的动态手势识别方法确定所述视频流中的动态手势识别结果；

控制设备执行与所述动态手势识别结果相应的操作。

根据本申请实施例的又一个方面，提供一种手势交互控制装置，所述装置包括：

视频流获取模块，用于获取视频流；

结果获取模块，用于采用如上任意一项所述的动态手势识别装置确定所述视频流中的动态手势识别结果；

操作执行模块，用于控制设备执行与所述动态手势识别结果相应的操作。

根据本申请实施例的又一个方面，提供一种电子设备，包括处理器，所述处理器包括如上任意一项所述的动态手势识别装置，或如上任意一项所述的动态手势识别建模装置，或如上任意一项所述的手势交互控制装置。

根据本申请实施例的另一个方面，提供一种电子设备，包括：存储器，用于存储可执行指令；以及处理器，用于与所述存储器通信以执行所述可执行指令从而完成如上任意一项所述动态手势识别方法或如上任意一项所述的动态手势识别建模方法或如上任意一项所述的手势交互控制方法的操作。

根据本申请实施例的再一个方面，提供一种计算机可读存储介质，用于存储计算机可读取的指令，其特征在于，所述指令被执行时执行如上任意一项所述动态手势识别方法或如上任意一项所述的动态手势识别建模方法或如上任意一项所述的手势交互控制方法的操作。

根据本申请实施例的还一个方面，提供一种计算机程序产品，包括计算机可读代码，其特征在于，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上任意一项所述动态手势识别方法或如上任意一项所述的动态手势识别建模方法或如上任意一项所述的手势交互控制方法的指令。

基于本申请上述实施例提供的动态手势识别方法和装置、手势交互控制方法和装置，从视频流的多帧图像帧中截取与动态手势框对应的图像块，基于图像块生成的检测序列进行动态手势识别。由于是基于与动态手势框对应的图像块进行的动态手势识别，因此可以识别出一系列变化的动态手势。

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为本申请实施例提供的动态手势识别方法的一个流程图。

图2为本申请实施例提供的动态手势识别方法的另一个流程图。

图3为本申请实施例提供的建立第一动态手势识别模型的一个流程图。

图4为本申请实施例提供的建立第二动态手势识别模型的一个流程图。

图5为本申请实施例提供的动态手势识别装置的一个结构示意图。

图6为本申请实施例提供的动态手势识别装置的另一个结构示意图。

图7为本申请实施例提供的手势交互控制方法的一个流程图。

图8为本申请实施例提供的手势交互控制方法的一个应用示例流程图。

图9为本申请实施例提供的手势交互控制装置的一个结构示意图。

图10为本申请实施例提供的电子设备的一个结构示意图。

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计算机系统/服务器，其可与众多其它通用或专用计算系统环境或配置一

起操作。适于与计算机系统/服务器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。

计算机系统/服务器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。

在实现本申请的过程中，发明人发现，目前的手势识别问题，只是识别静态单张图像，识别单张图片就只能识别一些简单的静态手势，例如剪刀手、拳头、OK 等等，在人机交互的过程中，用静态手势操控机器不如动态手势自然，而且承载的信息也更少。因此，亟需一种对动态手势进行识别的方案。

图 1 为本申请实施例提供的动态手势识别方法的一个流程图。该方法可以由任意电子设备执行，例如终端设备、服务器、移动设备、车载设备等等。如图 1 所示，该实施例方法包括 S101-S104。

S101：对待检测的视频流中的动态手势进行定位，得到动态手势框。

在一个可选示例中，该步骤 S101 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手势定位单元 501 执行。

本申请实施例中的动态手势，是指由一系列动作（可以是连续或不连续的动作）构成的手势，是相对静态手势而言的，例如包括但不限于：挥手、点击、打手枪手势、抓取手势，等等。在可选的应用中，例如，挥手可以实现文件翻页的操作；点击可以实现类似鼠标点击的操作；打手枪可以制作特效或者衔接游戏；抓取可以拖拽东西，类似鼠标拖拽文件。

对动态手势进行定位确定动态手势框，目的是为了后续在该动态手势框中截取出待检测的图像。其中的动态手势框，是指涵盖动态手势中一系列动作的框，例如是一个矩形框，手势图像都包含在该矩形框中。

可以理解，由于动态手势包含一系列静态手势，因此，可以通过先确定静态手势框再进行放大的方式确定动态手势框，由此可保证动态手势框将其余相关联的静态手势都包含在其中。

在一种可选方式中，可以通过如下方式确定动态手势框：从待检测的视频流的图像帧中选取任意一帧图像帧中的静态手势，对该静态手势进行定位，确定出静态手势框；根据预置的放大比例，对静态手势框进行放大处理，确定出动态手势框。例如，从视频流某帧图像帧中选取一个静态手势框，按照预置放大比例（例如 120%）对该静态手势框进行放大，放大后的框即为动态手势框。其中，多帧图像帧的静态手势框可以满足：静态手势框位于动态手势框内，或者，静态手势框与动态手势框相同。

S102：从视频流的多帧图像帧中截取与动态手势框对应的图像块。

视频流中被截取出来的图像块可以是视频流中连续的帧，也可以是连续的关键帧或采样帧，只要是与动态手势框对应即可。

在一个可选示例中，该步骤 S102 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处理单元 502 执行。

S103：基于截取出的图像块生成检测序列。

截取出来的图像块的大小通常小于图像帧，并且包括有图像中的动态手势框。这样处理的好处在于，考虑了多帧图像帧的手部定位信息，另外，对图像帧中大于动态手势框的部分去除而不作考虑，从而起到降噪的效果。

在一个可选示例中，该步骤 S103 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检测序列生成单元 503 执行。

S104：根据检测序列进行动态手势识别。

在一个可选示例中，该步骤 S104 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手势识别单元 504 执行。

基于截取出的图像块生成检测序列，后续利用动态手势框截取并生成一个图像块序列（而不是原来视频流的多帧图像帧）进行手势识别。

可见，本申请实施例中，从视频流的多帧图像帧中截取与动态手势框对应的图像块，基于图像块生成的检测序列进行动态手势识别。由于是基于与动态手势框对应的图像块进行的动态手势识别，因此可以识别出一系列变化的动态手势。

在一种可选方式中，基于检测序列中的帧间图像差进行动态手势识别。可选地：首先确定检测序列中多个帧图像帧之间图像差；然后基于多个帧图像帧之间图像差生成图像差序列；最后根据检测序列和图像差序列进行动态手势识别。

进一步，为了保证较好的体现图像的时序相关性，本申请实施例一种可选方式中提出，不但依据图像进行手势识别，而且，依据图像差进行手势识别。图像差，也可以理解是像素差，是相邻两帧图像相同位置的像素做差得到的，由于是相邻两帧的相同位置进行像素差处理，因此可以体现动态手势的变化过程及趋势，较好的识别出手势动态变化过程。

当然，上述相邻帧间图像差仅是一个例子，帧间图像差不限于此种限制，还可以是不相邻帧间的图像差，例如间隔固定数目帧或随机帧之间的图像差。可见，帧间图像差为检测序列中两个相邻参考帧之间的图像差。参考帧为实际帧，也可以是关键帧。

如前所述，动态手势的类型可以包括但不限于：挥手、点击、打手枪手势、抓取手势等，在一种可选方式中，可以预先分别建立第一动态手势识别模型和第二动态手势识别模型，分别将截取的图像及计算的图像差输入到这两个模型中，输出得到至少一个动态手势类型的概率，概率较高（例如，概率最高）的动态手势类型即为本次识别的结果。

在一种可选方式中，可多次（多段）进行识别，并根据多次识别结果确定动态手势类型。例如，在依据截取的一段（称为第一段）图像进行识别后，再依据第二段图像进行识别，以及，第三段图像进行识别，最后，根据三次识别结果确定动态手势类型。因此，这种实现方式中，上述方法还包括如下步骤：多次截取获得预置帧数的图像，并多次执行图像差计算，以及多次根据截取的图像以及计算的图像差进行动态手势识别；根据多次动态手势识别得到的动态手势类型的概率，确定出最终的动态手势识别结果。例如，在一种可选方式中：对所有次数的动态手势识别的至少一个动态手势类型概率进行求和处理，确定求和概率较高（包括：概率最高或者概率从高到低排序中前 n 位中的某一概率， n 为大于 1 的整数）的动态手势类型作为最终的动态手势识别结果。通过对图像分段多次进行识别并对每次识别的概率求和后确定最终的识别结果，可以减少每一次的计算量，提高识别的实时速度，尤其适合动作时间跨度较大的动态手势。

图 2 为本申请实施例提供的动态手势识别方法的另一个流程图。在图 1 实施例基础上，图 2 实施例介绍了以卷积神经网络作为动态手势识别模型，针对检测视频流中的动态手势进行检测的过程。

如图 2 所示，该实施例方法包括 S201-S204。

S201：建立第一动态手势识别模型和第二动态手势识别模型。

在一个可选的方式中，参见图 3 和图 4，图 3 为本申请实施例提供的建立第一动态手势识别模型的一个流程图。图 4 为本申请实施例提供的建立第二动态手势识别模型的一个流程图。

参考图 3，建立第一动态手势识别模型的过程包括 S301-S304。

S301：采集不同动态手势类型的样本视频流。

在一个可选示例中，该步骤 S301 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样本采集子单元 6071 执行。

例如，采集已知动态手势类型（例如：挥手、点击、打手枪、抓取）的视频流，标记样本视频流的开始帧和结束帧。

S302: 对不同动态手势类型的动态手势框进行标记。

在一个可选示例中, 该步骤 S302 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手势框标记子单元 6072 执行。

动态手势框, 是指涵盖样本视频流中动态手势中一系列动作的框, 例如是一个矩形框, 该动态手势的各个静态手势图像都包含在该矩形框中。

可以理解, 由于动态手势包含一系列静态手势, 因此, 可以通过先确定静态手势框再进行放大的方式确定动态手势框, 由此可保证动态手势框将其余相关联的静态手势都包含在其中。

在一种可选方式中, 可以通过如下方式确定动态手势框: 从样本视频流的图像中选取任意一幅图像中的静态手势, 对该静态手势进行定位, 确定出静态手势框; 根据预置的放大比例, 对静态手势框进行放大处理, 确定出动态手势框。例如, 从视频流某帧图像中选取一个静态手势框, 按照预置放大比例 (例如 120%) 对该静态手势框进行放大, 放大后的框即为动态手势框。

S303: 从样本视频流的多帧图像帧中截取与动态手势框的标注信息对应的图像块, 构成图像序列。

在一个可选示例中, 该步骤 S303 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图像序列构成子单元 6073 执行。

S304: 以动态手势类型作为监督数据, 以图像序列作为训练数据, 训练第一动态手势识别模型。

在一个可选示例中, 该步骤 S304 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训练子单元 6074 执行。

在一种可选方式中, 通过如下步骤建立第一动态手势识别模型:

(1) 将图像序列分为至少一段; 例如, 将图像序列平均分为三段。

(2) 在至少一段中抽取出预置帧数的图像, 堆叠组成图像训练数据。

例如, 在每一段图像数据中抽取出 (随机或连续) 五帧图像进行堆叠, 构成图像训练数据。或者, 从至少一段图像数据中共抽取出 (随机或连续) 十帧图像进行堆叠, 构成图像训练数据。

可选地, 以卷积神经网络的三维的矩阵数据为例, 三个维度分别是通道、图像的高和宽, 比如, 一张灰度图的通道数就是 1, 一张 RGB 图像通道数就是 3, 这里的堆叠是通道堆叠, 比如有五张通道数是 1 的图像, 堆叠后就是一个通道数是 5 的三维矩阵。

(3) 以动态手势类型作为监督数据, 以图像序列作为训练数据, 训练第一动态手势识别模型。

参考图 4, 建立第二动态手势识别模型的过程包括 S401-S406。

S401: 采集不同动态手势类型的样本视频流。

在一个可选示例中, 该步骤 S401 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样本采集子单元 6081 执行。

例如, 采集已知动态手势类型 (例如: 挥手、点击、打手枪、抓取) 的视频流, 标记样本视频流的开始帧和结束帧。

S402: 对不同动态手势类型的动态手势框进行标记。

在一个可选示例中, 该步骤 S402 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手势框标记子单元 6082 执行。

动态手势框, 是指涵盖样本视频流中动态手势中一系列动作的框, 例如是一个矩形框, 该动态手势的各个静态手势图像都包含在该矩形框中。

可以理解, 由于动态手势包含一系列静态手势, 因此, 可以通过先确定静态手势框再进行放大的方式确定动态手势框, 由此可保证动态手势框将其余相关联的静态手势都包含在其中。

在一种可选方式中, 可以通过如下方式确定动态手势框: 从样本视频流的图像中选取任意一幅图像中的静态手势, 对该静态手势进行定位, 确定出静态手势框; 根据预置的放大比例, 对静态手势框进行放大处理, 确定出动态手势框。例如, 从视频流某帧图像中选取一个静态手势框, 按照预置放大比例 (例如 120%) 对该静态手势框进行放大, 放大后的框即为动态手势框。

S403: 从样本视频流的多帧图像帧中截取与动态手势框的标注信息对应的图像块, 构成图像序列。

在一个可选示例中, 该步骤 S403 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图像序列构成子单元 6083 执行。

S404: 确定图像序列中多个帧间图像差。

在一个可选示例中, 该步骤 S404 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图像差确定子单元 6084 执行。

为了保证较好的体现图像的时序相关性, 需要确定多个帧间图像差。图像差, 也可以理解是像素差, 是相邻两帧图像相同位置的像素做差得到的, 由于是相邻两帧的相同位置进行像素差处理, 因此可以体现动态手势的变化过程及趋势, 较好的识别出手势动态变化过程。

当然, 上述相邻帧间图像差仅是一个例子, 帧间图像差不限于此种限制, 还可以是不相邻帧间的图像差, 例如间隔固定数目帧或随机帧之间的图像差。可见, 帧间图像差为检测序列中两个相邻参考帧之间的图像差。参考帧为实际帧, 也可以是关键帧。

S405: 基于确定的多个帧间图像差生成图像差序列。

在一个可选示例中, 该步骤 S405 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图像差序列确定子单元 6085 执行。

S406: 以动态手势类型作为监督数据, 以图像差序列作为训练数据, 训练第二动态手势识别模型。

在一个可选示例中, 该步骤 S406 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训练子单元 6086 执行。

在一种可选方式中, 通过如下步骤建立第二动态手势识别模型:

- (1) 将图像差序列分为至少一段;
- (2) 从至少一段中抽取出预置帧数的图像, 堆叠组成图像差训练数据;
- (3) 以动态手势类型作为监督数据, 以图像差训练数据训练第二动态手势识别模型。

需要说明的是, 可以基于不同网络实现上述第一动态手势识别模型和第二动态手势识别模型。例如可以基于卷积神经网络实现第一动态手势识别模型和第二动态手势识别模型。可选的, 第一动态手势识别模型可以但不限于为第一神经网络模型, 第一神经网络模型基于样本视频流预先训练而得。第一神经网络模型可以包括但不限于卷积层、非线性层 (Relu)、池化层和/或分类层等。同理, 第二动态手势识别模型可以但不限于为第二神经网络模型, 第二神经网络模型基于样本视频流预先训练而得。第二神经网络模型可以包括但不限于卷积层、非线性层 (Relu)、池化层和/或分类层等。当然, 除了卷积神经网络之外, 还可以基于循环神经网络、增强学习神经网络或生成对抗神经网络等实现上述第一动态手势识别模型和第二动态手势识别模型, 本申请实施例对此不作限定。

S202: 将截取出的图像输入到第一动态手势识别模型, 以及, 将相邻两帧图像的图像差输入到第二动态手势识别模型, 识别出动态手势类型的预测概率。

本领域技术人员了解, 卷积神经网络的识别工作过程可以大致包括: 图像特征提取阶段以及对特征进行分类阶段。例如, 以将图像输入到第一动态识别模型为例子, 是将预置数量帧数 (例如 5 帧) 图像输入到第一动态识别模型, 经过卷积层、激活层、池化层的使用, 将图像中的特征提取出来, 然后, 经过分类器对特征进行分类, 最终得出动态手势类型的预测概率。

S203: 根据第一动态手势识别模型和第二动态手势识别模型的动态手势类型的预测概率, 确定动态手势识别结果。

在一种可选方式中, 可将第一动态手势识别模型和第二动态手势识别模型的至少一个动态手势类型的预测概率进行加权平均处理, 确定加权平均概率较高 (如, 加权平均概率最高) 的动态手势类型为本次动态手势识别结果。例如, 可以预先设置两个模型的加权系数, 在确定动态手势识别结果时, 按照各个模型的加权系数, 对两个模型的预测概括进行加权平均处理, 确定加权平均概率最高的动态手势类型为本次动态手势识别结果。当然, 除了加权平均法之外, 还可以采取其他方法 (例如调和平均法和平方

平均法等)对预测概率进行处理,最终确定识别结果。

通过本申请实施例的动态手势识别方法,分别对图像及图像差进行概率识别,从而得到至少一个动态手势类型的概率,并确定概率较大(如,概率最大)的动态手势类型为识别结果,其中,图像差可以较好的体现图像前后的时序相关性,可以实现动态手势的识别。

在一个可选方式中,通过对图像分段多次进行识别并对每次识别的概率求和后确定最终的识别结果,可以减少每一次的计算量,提高识别的实时速度,尤其适合动作时间跨度较大的动态手势。

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5为本申请实施例提供的动态手势识别装置的一个结构示意图。该实施例的装置可用于实现本申请上述各方法实施例。如图5所示,该实施例的装置包括:

手势定位单元501,用于对待检测的视频流中的动态手势进行定位,得到动态手势框。

处理单元502,用于从视频流的多帧图像帧中截取与动态手势框对应的图像块。

检测序列生成单元503,用于基于截取出的图像块生成检测序列。

手势识别单元504,用于根据检测序列进行动态手势识别。

图6为本申请实施例提供的动态手势识别装置的另一个结构示意图。该实施例的装置可用于实现本申请上述各方法实施例。如图6所示,该实施例的装置包括:

手势定位单元601,用于对待检测的视频流中的动态手势进行定位,得到动态手势框。

处理单元602,用于从视频流的多帧图像帧中截取与动态手势框对应的图像块。

检测序列生成单元603,用于基于截取出的图像块生成检测序列。

手势识别单元604,用于根据检测序列进行动态手势识别。

在一种可选方式中,手势定位单元601包括:

静态手势框定位单元6011,用于对视频流的多帧图像中的至少一帧图像帧进行静态手势定位,得到至少一帧图像帧的静态手势框;

动态手势框确定子单元6012,用于根据得到的至少一帧图像帧的静态手势框确定动态手势框。

在一种可选方式中,动态手势框确定子单元6012用于:对至少一帧图像帧的静态手势框进行放大处理,得到动态手势框。

在一种可选方式中,视频流的多帧图像帧中至少一帧图像帧的静态手势框满足:静态手势框位于动态手势框内,或者,静态手势框与动态手势框相同。

在一种可选方式中,手势识别单元604包括:

图像差确定子单元6041,用于确定检测序列中多个帧间图像差;

图像差序列确定子单元6042,用于基于确定的多个帧间图像差生成图像差序列;

动态手势识别子单元6043,用于根据检测序列和图像差序列进行动态手势识别。

在一种可选方式中,帧间图像差为检测序列中两个相邻参考帧之间的图像差。

在一种可选方式中,动态手势识别子单元6043用于:将检测序列输入第一动态手势识别模型,以获得第一动态手势识别模型输出的第一动态手势类别预测概率;将图像差序列输入第二动态手势识别模型,以获得第二动态手势识别模型输出的第二动态手势类别预测概率;根据第一动态手势类别预测概率和第二动态手势类别预测概率,确定动态手势识别结果。

在一种可选方式中,第一动态手势识别模型为第一神经网络,第二动态手势识别模型为第二神经网络,第一神经网络和第二神经网络的结构相同或不同。

在一种可选方式中,手势识别单元604还包括:

多次识别控制单元605,用于多次截取获得检测序列,并多次生成图像差序列,以及多次根据检测序列以及图像差序列进行动态手势识别;

识别结果确定单元 606，用于根据每次动态手势识别得到的动态手势类型的概率，确定动态手势识别结果。

在一种可选方式中，手势识别单元 604 还包括：第一动态手势识别模型建立单元 607；第一动态手势识别模型建立单元 607 包括：

样本采集子单元 6071，用于采集不同动态手势类型的样本视频流；

手势框标记子单元 6072，用于对不同动态手势类型的动态手势框进行标记；

图像序列构成子单元 6073，用于从样本视频流的多帧图像帧中截取与动态手势框的标注信息对应的图像块，构成图像序列；

训练子单元 6074，用于以动态手势类型作为监督数据，以图像序列作为训练数据，训练第一动态手势识别模型。

在一种可选方式中，训练子单元 6074 用于：将图像序列分为至少一段；从至少一段中抽取出预置帧数的图像，堆叠组成图像训练数据；以及，以动态手势类型作为监督数据，以图像训练数据训练第一动态手势识别模型。

在一种可选方式中，手势识别单元 604 还包括：第二动态手势识别模型建立单元 608；第二动态手势识别模型建立单元 608 包括：

样本采集子单元 6081，用于采集不同动态手势类型的样本视频流；

手势框标记子单元 6082，用于对不同动态手势类型的动态手势框进行标记；

图像序列构成子单元 6083，用于从样本视频流的多帧图像帧中截取与动态手势框的标注信息对应的图像块，构成图像序列；

图像差确定子单元 6084，用于确定图像序列中多个帧间图像差；

图像差序列确定子单元 6085，用于基于确定的多个帧间图像差生成图像差序列；

训练子单元 6086，用于以动态手势类型作为监督数据，以图像差序列作为训练数据，训练第二动态手势识别模型。

在一种可选方式中，训练子单元 6086 用于：将图像差序列分为至少一段；从至少一段中抽取出预置帧数的图像，堆叠组成图像差训练数据；以及，以动态手势类型作为监督数据，以图像差训练数据训练第二动态手势识别模型。

本实施例的动态手势识别装置可用于实现前述多个方法实施例中相应的动态手势识别方法，并具有相应的方法实施例的有益效果，在此不再赘述。

图7为本申请实施例提供的手势交互控制方法的一个流程图。该方法可以由任意电子设备执行，例如终端设备、服务器、移动设备、车载设备、无人机、机器人、无人车、电视机、车辆、家居设备或其他类型的智能设备等等。如图7所示，该手势交互控制方法，包括：

步骤S700，获取视频流。

在一个可选示例中，该步骤 S700 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视频流获取模块 100 执行。

步骤S710，采用上述任一项动态手势识别方法确定该视频流中的动态手势识别结果。

在一个可选示例中，该步骤 S710 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结果获取模块 200 执行。

步骤S720，控制设备执行与动态手势识别结果相应的操作。

在一个可选示例中，该步骤 S720 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操作执行模块 300 执行。

在一种可选的实现方式中，可以在需要进行控制操作的设备上设置拍摄装置，可以利用拍摄装置实时获取视频流。也可以获取拍摄装置拍摄完成的视频流。可以利用不同的拍摄装置拍摄视频流。拍摄装置可以包括双目摄像头、深度摄像头或普通摄像头。可以根据动态手势识别的需求，选用不同类型的摄

像头拍摄视频流。

在一种可选的实现方式中，步骤S720，包括：

根据预先确定的动态手势识别结果与操作指令之间的对应关系，获取与动态手势识别结果对应的操作指令；

根据操作指令控制设备执行相应操作。

在一种可选的实现方式中，可以预设动态手势识别结果和操作指令之间的对应关系。在对应关系中，可以是一个动态手势识别结果对应一个操作指令，也可以是多个动态手势识别结果对应一个操作指令。可以根据被操作的设备的类型和操作需求，确定操作指令的类型和内容。本申请实施例不限定操作指令的形式和具体内容。

在一种可选的实现方式中，可以输出操作指令，对被操作的设备进行控制。由于动态手势识别结果可以实时追踪视频流中的动作，输出的操作指令也可以实时追踪视频流中执行对象的动作，使得操作者可以相对精准的控制被操作的设备。

在本申请实施例中，可以根据对应关系和动态手势识别结果，确定与动态手势识别结果对应的操作指令并输出。基于动态手势识别结果的实时性和准确性，操作指令也能够实时追踪视频流中执行对象的动作，使得操作者能够更加精确的操作设备。

在一种可选的实现方式中，根据操作指令控制设备执行相应操作，包括：

根据操作指令对车辆的车窗、车门或车载系统进行控制。

在一种可选的实现方式中，可以在车辆中设置监控设备，拍摄车辆中的驾驶者或乘坐者的监控视频作为视频流。可以将拍摄到的视频流进行实时地动态手势识别。

在一种可选的实现方式中，控制设备执行与动态手势识别结果相应的操作，包括：

响应于动态手势识别结果为预定义动态动作，控制车辆执行与预定义动态动作对应的操作。

在一种可选的实现方式中，预定义动态动作包括动态手势，动态手势可以包括但不限于以下至少之一：单指顺/逆时针旋转、手掌左/右挥动、两指前戳、拇指和小指伸出、手掌朝下向下压、手掌朝上向上抬起、手掌向左/右扇风、拇指伸出左/右移动、手掌左/右长滑、掌心向上拳变掌、掌心向上掌变拳、掌心向下掌变拳、掌心向下掌变拳、单指滑动、多指向内捏住、单指双击、单指单击、多指双击、多指单击；

与预定义动态动作对应的操作可以包括但不限于以下至少之一：调节音量大/小，歌曲切换，歌曲暂停/继续，接听或启动电话、挂断或拒接电话、空调温度升高或降低、多屏互动、打开天窗、关闭天窗、锁紧门锁、解锁门锁、拖动导航、缩小地图、放大地图。

在一种可选的实现方式中，可以利用单指顺/逆时针旋转的动态手势，对车辆的音响设备进行调节音量大/小的操作。可以利用手掌左/右挥动的动态手势，对车辆的音响设备进行歌曲切换的操作。可以利用两指前戳的动态手势，对车辆的音响设备进行歌曲暂停/继续的操作。可以利用拇指和小指伸出的动态手势，对车辆的通信设备进行接听或启动电话的操作。可以利用手掌朝下向下压的动态手势，对车辆的通信设备进行挂断或拒接电话的操作。可以利用拇指伸出左/右移动的动态手势，对车辆的空调设备进行空调温度升高或降低的操作。可以利用手掌左/右长滑的动态手势，对车辆的显示屏幕进行多屏互动的操作。可以利用掌心向上拳变掌的动态手势，对车辆进行打开天窗的操作（例如，可以每次打开设定的长度，例如每次打开10厘米）。可以利用掌心向上掌变拳的动态手势，对车辆进行关闭天窗的操作。可以利用掌心向下掌变拳的动态手势，对车辆进行锁紧门锁的操作。可以利用掌心向下掌变拳的动态手势，对车辆进行解锁门锁的操作。可以利用单指滑动的动态手势，对车辆的导航设备进行拖动导航的操作。可以利用多指向内捏住的动态手势，对车辆的导航设备进行缩小地图的操作。可以利用单指双击的动态手势，对车辆的导航设备进行放大地图的操作。

在本申请实施例中，可以利用操作指令对车辆的车窗、车门或车载系统进行控制。可以利用在视频流中识别出的动态动作，对车辆自身或车辆上的车载系统进行不同的操作。基于本申请实施例中的动态

动作检测方法，可以使得操作者对车辆自身或车载系统进行相对精准地控制。

图8为本申请实施例提供的手势交互控制方法的一个应用示例流程图。如图8所示：

可以在车辆中配置深度摄像头，利用深度摄像头获取驾驶员的监控图像作为视频流。可以将拍摄到的监控图像进行实时的动作识别。在本申请实施例中，对驾驶员的手部动作进行动态手势识别。

一、获取在先检测的动态手势识别结果：

步骤S810，将第一队列和第二队列设置为空，动态手势识别结果也设置为空。将视频流中的帧图像按照时间段前后顺序依次加入第一队列。

步骤S820，检测第一队列的帧图像中是否有动态动作。如无动态动作，接步骤S830，如有动态动作，接步骤S840。

在一种可选的实现方式中，第一队列的长度可以为十帧帧图像。在视频流中的第一帧至第十帧帧图像加入第一队列后，可以判断第一队列中的十帧帧图像中是否有动态手势。例如，动态动作为手部动态动作。可选的，可以依次识别帧图像中是否有手部，当第一队列的帧图像中有手部的情况下，可以根据手部上的手指和/或手掌识别待识别图像中的手势。当至少一帧图像中的手势匹配时，可以确定至少一帧图像中的手势的动作轨迹和/或切换信息。在至少一帧图像中的手势的动作轨迹和/或切换信息也匹配的情况下，可以检测第一队列中的动态动作。

步骤S830，继续将视频流中的至少一帧图像按时间前后顺序依次加入第一队列中，跳转至步骤S820。

在一种可选的实现方式中，可以将第十一帧帧图像加入第一队列的后端，并将第一队列中最前端的帧帧图像移出第一队列。此时第一队列中包括第二帧至第十一帧帧图像，跳转至步骤S820后可以判断此时的第一队列中是否有动态动作。如无动态动作，可继续将第十二帧帧图像加入第一队列，并移出第二帧帧图像，直至根据步骤S820判断第一队列的帧图像中有动态动作。

步骤S840，将第一队列中的帧图像移至第二队列中。第一队列被清空，根据第二队列中的帧图像确定动态手势识别结果，得到在先检测的动态手势识别结果。

在一种可选的实现方式中，得到在先检测的动态手势识别结果后，第一队列为空，第二队列中有十帧帧图像，为视频流中的第二十三帧至第三十二帧帧图像。可以根据第二队列的帧图像中的动作确定动态手势识别结果。视频流中未进行动态手势识别的帧图像，可以作为待识别的帧图像进行后续分析。即自第三十三帧帧图像开始的帧图像，可以作为待识别的帧图像，进入后续动态动作检测的步骤S850。

二、动态手势识别步骤：

步骤S850，依次判断视频流中至少一个待识别的帧图像与动态手势识别结果是否匹配。如果匹配，接步骤S860，如果不匹配，接步骤S870。

在一种可选的实现方式中，可以根据第三十三帧帧图像与第二队列中的最后端的帧图像（第三十二帧帧图像），确定第三十三帧帧图像与动态手势识别结果是否匹配。可以先确定第三十三帧帧图像中的手势，与第二队列中的第三十二帧帧图像的手势是否一致。在手势一致的情况下，确定第三十三帧帧图像中的手势的动作轨迹和/或切换信息，与动态手势识别结果中手势的动作轨迹和/或切换信息是否匹配（其中，帧图像中的手势的动态轨迹与动态手势识别结果中手势的动态轨迹进行匹配，帧图像中的手势的切换信息与动态手势识别结果中手势的切换信息进行匹配）。在手势的动作轨迹和/或切换信息也匹配的情况下，可以确定第三十三帧帧图像与动态手势识别结果匹配。

步骤S860，如果匹配，将待识别的帧图像加入第二队列中。

在一种可选的实现方式中，将第三十三帧帧图像加入第二队列中后，第二队列中的帧图像更新为第二十四帧至第三十三帧。

步骤S870，如果不匹配，将待识别的帧图像加入第一队列中。

步骤S880，确定第一队列的帧图像中的动态手势是否与动态手势识别结果匹配。如果不匹配，接步骤S890。

步骤 S890，在第一队列的帧图像中的动态手势与动态手势识别结果不匹配的情况下，清空第二队

列,将第一队列中的帧图像移至第二队列中。并根据更新后的第二队列的帧图像中的动作更新动态手势识别结果。

可以理解,本申请实施例提及的上述各个方法实施例,在不违背原理逻辑的情况下,均可以彼此相互结合形成结合后的实施例,限于篇幅,本申请不再赘述。

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9为本申请实施例提供的手势交互控制装置的一个结构示意图。如图9所示,该手势交互控制装置包括:

视频流获取模块100,用于获取视频流;

结果获取模块200,用于采用上述的动态手势识别装置确定该视频流中的动态手势识别结果;

操作执行模块300,用于控制设备执行与动态手势识别结果相应的操作。

在一种可选的实现方式中,操作执行模块300,包括:

操作指令获取子模块,用于根据预先确定的动态手势识别结果与操作指令之间的对应关系,获取与动态手势识别结果对应的操作指令;

操作执行子模块,用于根据操作指令控制设备执行相应操作。

在一种可选的实现方式中,操作执行子模块,用于:

根据操作指令对车辆的车窗、车门或车载系统进行控制。

在一种可选的实现方式中,操作执行模块300,还用于:

响应于检测结果为预定义动态动作,控制车辆执行与预定义动态动作对应的操作。

在一种可选的实现方式中,预定义动态动作包括动态手势,动态手势可以包括但不限于以下至少之一:单指顺/逆时针旋转、手掌左/右挥动、两指前戳、拇指和小指伸出、手掌朝下向下压、手掌朝上向上抬起、手掌向左/右扇风、拇指伸出左/右移动、手掌左/右长滑、掌心向上拳变掌、掌心向上掌变拳、掌心向下掌变拳、掌心向下拳变掌、单指滑动、多指向内捏住、单指双击、单指单击、多指双击、多指单击;

与预定义动态动作对应的操作可以包括但不限于以下至少之一:调节音量大/小,歌曲切换,歌曲暂停/继续,接听或启动电话、挂断或拒接电话、空调温度升高或降低、多屏互动、打开天窗、关闭天窗、锁紧门锁、解锁门锁、拖动导航、缩小地图、放大地图。

本领域技术人员可以理解,在具体实施方式的上述方法中,各步骤的撰写顺序并不意味着严格的执行顺序而对实施过程构成任何限定,各步骤的具体执行顺序应当以其功能和可能的内在逻辑确定。

在一些实施例中,本申请实施例提供的装置具有的功能或包含的模块可以用于执行上文方法实施例描述的方法,其具体实现可以参照上文方法实施例的描述,为了简洁,这里不再赘述。

根据本申请实施例的又一个方面,提供一种电子设备,包括处理器,处理器包括本申请上述任意一个实施例提供的动态手势识别装置,或本申请上述任意一个实施例提供的动态手势识别建模装置,或本申请上述任意一个实施例提供的手势交互控制装置。

根据本申请实施例的另一个方面,提供一种电子设备,包括:存储器,用于存储可执行指令;以及处理器,用于与存储器通信以执行可执行指令从而完成本申请上述任意一个实施例提供的动态手势识别方法或本申请上述任意一个实施例提供的动态手势识别建模方法或本申请上述任意一个实施例提供的手势交互控制方法的操作。

根据本申请实施例的再一个方面,提供一种计算机可读存储介质,用于存储计算机可读的指令,其特征在于,指令被执行时执行本申请上述任意一个实施例提供的动态手势识别方法或本申请上述任意一个实施例提供的动态手势识别建模方法或本申请上述任意一个实施例提供的手势交互控制方法的操作。

根据本申请实施例的还一个方面，提供一种计算机程序产品，包括计算机可读代码，其特征在于，当计算机可读代码在设备上运行时，设备中的处理器执行用于实现本申请上述任意一个实施例提供的动态手势识别方法或本申请上述任意一个实施例提供的动态手势识别建模方法或本申请上述任意一个实施例提供的手势交互控制方法的指令。

本申请实施例还提供了一种电子设备，例如可以是移动终端、个人计算机（PC）、平板电脑、服务器等。下面参考图 10，其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 1000 的结构示意图：如图 10 所示，计算机系统 1000 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）1001，和/或 一个或多个图像处理器（GPU）1013 等，处理器可以根据存储在只读存储器（ROM）1002 中的可执行指令或者从存储部分 1008 加载到随机访问存储器（RAM）1003 中的可执行指令而执行各种适当的动作和处理。通信部 1012 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 602 和/或随机访问存储器 1030 中通信以执行可执行指令，通过总线 1004 与通信部 1012 相连、并经通信部 1012 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：对待检测的视频流中的动态手势进行定位，得到动态手势框；从视频流的多帧图像帧中截取与动态手势框对应的图像块；基于截取出的图像块生成检测序列；根据检测序列进行动态手势识别。

此外，在 RAM 1003 中，还可存储有装置操作所需的各种程序和数据。CPU1001、ROM1002 以及 RAM1003 通过总线 1004 彼此相连。在有 RAM1003 的情况下，ROM1002 为可选模块。RAM1003 存储可执行指令，或在运行时向 ROM1002 中写入可执行指令，可执行指令使处理器 1001 执行上述通信方法对应的操作。输入/输出（I/O）接口 1005 也连接至总线 1004。通信部 1012 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 1005：包括键盘、鼠标等的输入部分 1006；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 1007；包括硬盘等的存储部分 1008；以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 1009。通信部分 1009 经由诸如因特网的网络执行通信处理。驱动器 1010 也需要连接至 I/O 接口 1005。可拆卸介质 1011，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 1010 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 1008。

需要说明的，如图 10 所示的架构仅为一种可选实现方式，在具体实践过程中，可根据实际需要对上述图 10 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，根据本申请的实施例，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，本申请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，对待检测的视频流中的动态手势进行定位，得到动态手势框；从动态手势框中截取预置帧数的图像，并从截取出的图像中计算出相邻两帧图像的图像差；根据截取出的图像以及相邻两帧图像的图像差进行动态手势识别。在这样的实施例中，该计算机程序可以通过通信部分 1009 从网络上被下载和安装，和/或从可拆卸介质 1011 被安装。在该计算机程序被中央处理单元（CPU）1001 执行时，执行本申请的方法中限定的上述功能。

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。