

一种视频特征提取方法及装置

相关申请的交叉引用

本申请要求申请号为 201810271774.6、申请日为 2018 年 3 月 29 日的中国专利申请的优先权，该文献的全部内容以引用方式并入本文。

技术领域

本公开涉及视频处理技术领域，特别是涉及一种视频特征的提取方法及装置。

背景技术

在如今的多媒体信息社会中，用户每天向视频平台上传海量的视频，
这些视频中大部分是正常的有价值视频，然而也有一些问题视频，问题视频主要包括：和平台视频数据库中的已有视频重复的视频、与版权数据库中的视频重复的视频（例如，需要支付版权费的视频）以及某些不适宜或禁止展示的视频。因此需要对用户上传的海量视频进行快速的比对和过滤。而提高视频比对速度和准确性的核心技术是对视频帧的特征进行合理的提取和相似度判断。

为了提高比对速度和准确性，需要为一个视频生成一个能表征此视频的视频特征，目的是通过比对两个视频的视频特征来判断视频的相似程度。视频特征的提取方法和视频特征的优良程度决定了视频比对的效率和准确率。

发明内容

本公开的目的在于提供一种新的视频特征提取方法及装置。

本公开的目的是采用以下的技术方案来实现的。依据本公开提出的视频特征提取方法，包括以下步骤：对视频对象进行抽帧，得到一个或多个帧图像；对每个所述帧图像逐级地进行多种类型的池化，以得到所述帧图像的图像特征；其中，所述多种类型的池化包括最大池化、最小池化和平均池化；根据所述一个或多个帧图像的所述图像特征确定视频特征。

本公开的目的还可以采用以下的技术措施来进一步实现。

前述的视频特征提取方法，其中所述的对每个所述帧图像逐级地进行多种类型的池化包括：基于所述帧图像的多种颜色通道逐级地进行所述多种类型的池化。

前述的视频特征提取方法，其中所述的对每个所述帧图像逐级地进行多种类型的池化，以得到所述帧图像的图像特征包括：根据所述帧图像确定一个矩阵，利用所述多种类型的池化，逐级地生成更小的矩阵，直到缩

小为仅包含一个点的矩阵，根据所述仅包含一个点的矩阵确定所述图像特征。

前述的视频特征提取方法，其中所述对每个所述帧图像逐级地进行多种类型的池化，以得到所述帧图像的图像特征包括以下步骤：(a) 根据一个所述帧图像，确定一个具有第一矩阵维度和第二矩阵维度的第一矩阵；所述第一矩阵中的点与所述帧图像中的像素相对应；所述第一矩阵中的点的取值为第一向量，所述第一向量为 3 维的向量，用于表示对应像素的三个颜色通道的亮度；(b) 在所述第一矩阵上设置多个第一区块，每个所述第一区块包含多个所述第一向量；所述多个第一区块在第一矩阵维度上的数量少于所述第一矩阵在第一矩阵维度上所包含的点的数量，且所述多个第一区块在第二矩阵维度上的数量少于所述第一矩阵在第二矩阵维度上所包含的点的数量；对于每个所述第一区块，分别计算所述第一区块所包含的多个所述第一向量的各个维的最大值、最小值和平均值，得到一个 9 维的第二向量；(c) 根据所述多个第一区块所对应的所述第二向量，确定第二矩阵；所述第二矩阵中的点与所述第一区块相对应，所述第二矩阵中的点的取值为所述第二向量；(d) 重复步骤 (b) 和步骤 (c)，直到将所述第一矩阵缩小成一个取值为 3^N 维向量的点，其中的 N 为正整数；将所述 3^N 维向量确定为所述帧图像的图像特征。

前述的视频特征提取方法，其中所述的根据所述一个或多个帧图像的所述图像特征确定视频特征包括：对所述图像特征进行二值化处理，得到二值化图像特征；根据所述一个或多个帧图像的所述二值化图像特征，确定视频特征。

前述的视频特征提取方法，其中所述的对所述图像特征进行二值化处理，得到二值化图像特征包括以下步骤：根据所述图像特征生成多个小组，每个所述小组包含所述图像特征中的多个元素；分别对每个所述小组中的所述多个元素进行求和，得到每个所述小组的加和值；将所述多个小组两两配对，得到多个小组对；对于每个所述小组，比较所述小组对中的两个所述小组的所述加和值的大小，根据比较结果生成一个二值化的图像特征比特；根据所述多个小组对的所述图像特征比特，确定所述帧图像的二值化图像特征。

本公开的目的还采用以下技术方案来实现。依据本公开提出的视频特征库构建方法，包括以下步骤：按照前述的任意一项的视频特征提取方法提取视频对象的视频特征；将所述视频特征存储到视频特征库中。

本公开的目的还采用以下技术方案来实现。依据本公开提出的视频特征提取装置，包括：抽帧模块，用于对视频对象进行抽帧，得到一个或多个帧图像；图像特征确定模块，用于对每个所述帧图像逐级地进行多种类

型的池化，以得到所述帧图像的图像特征；其中，所述多种类型的池化包括最大池化、最小池化和平均池化；视频特征确定模块，用于根据所述一个或多个帧图像的所述图像特征向量，确定视频特征。

本公开的目的还可以采用以下的技术措施来进一步实现。

5 前述的视频特征提取装置，其还包括执行前述任一视频特征提取方法步骤的模块。

本公开的目的还采用以下技术方案来实现。依据本公开提出的音频指纹库构建装置，包括：视频特征提取模块，用于按照前述任意一项的视频特征提取方法提取视频对象的视频特征；视频特征存储模块，用于将所述
10 视频特征存储到视频特征库中；视频特征库，用于存储所述视频特征。

本公开的目的还采用以下技术方案来实现。依据本公开提出的一种视频特征提取硬件装置，包括：存储器，用于存储非暂时性计算机可读指令；以及处理器，用于运行所述计算机可读指令，使得所述处理器执行时实现前述任意一种视频特征提取方法。

15 本公开的目的还采用以下技术方案来实现。依据本公开提出的一种计算机可读存储介质，用于存储非暂时性计算机可读指令，当所述非暂时性计算机可读指令由计算机执行时，使得所述计算机执行前述任意一种视频特征提取方法。

本公开的目的还采用以下技术方案来实现。依据本公开提出的一种终端设备，包括前述任意一种视频特征提取装置。

上述说明仅是本公开技术方案的概述，为了能更清楚了解本公开的技术手段，而可依照说明书的内容予以实施，并且为了让本公开的上述和其他目的、特征和优点能够更明显易懂，以下特举较佳实施例，并配合附图，详细说明如下。

25

附图说明

图 1 是本公开一个实施例的视频特征提取方法的流程框图。

图 2 是本公开一个实施例提供的逐级进行多类型池化处理的流程框图。

图 3 是本公开一个实施例提供的利用随机投影法对图像特征进行二值
30 化处理的流程框图。

图 4 是利用本公开的方法来提取帧图像的图像特征的一个具体示例的流程示意图。

图 5 是本公开一个实施例的视频特征库构建方法的流程框图。

图 6 是本公开一个实施例的视频特征提取装置的结构框图。

35 图 7 是本公开一个实施例的视频特征库构建装置的结构框图。

图 8 是本公开一个实施例的视频特征提取硬件装置的硬件框图。

图 9 是本公开一个实施例的计算机可读存储介质的示意图。

图 10 是本公开一个实施例的终端设备的结构框图。

具体实施方式

5 为进一步阐述本公开为达成预定发明目的所采取的技术手段及功效，以下结合附图及较佳实施例，对依据本公开提出的视频特征提取方法及装置的具体实施方式、结构、特征及其功效，详细说明如后。

图 1 为本公开的视频特征提取方法一个实施例的示意性流程框图。请参阅图 1，本公开示例的视频特征提取方法，主要包括以下步骤：

10 步骤 S11，对视频对象进行抽帧，得到一个或多个帧图像。需注意，对视频对象的类型不做限制，可以是一段视频信号，也可以是一个视频文件。此后，处理进到步骤 S12。

步骤 S12，对每个帧图像，逐级地进行多种类型的池化（Pooling）处理，以得到该帧图像的图像特征。其中，池化（Pooling）是一种在卷积神经网络领域的降维方法，而所谓的多种类型的池化包括最大池化、最小池化
15 和平均池化。此后，处理进到步骤 S13。

具体地，可以基于帧图像的多种颜色通道逐级地进行多种类型的池化，以根据帧图像的多种颜色通道得到图像特征。

步骤 S13，根据该一个或多个帧图像对应的多个图像特征，确定该视频
20 对象的视频特征。具体地，可以将该多个图像特征按照帧图像的时间顺序组合在一起，得到该视频特征。

本公开提出的视频特征提取方法，通过逐级地对由视频抽帧得到的帧图像进行多种类型的池化以生成视频特征，能够大大提高视频特征提取的
25 准确性和提取的效率，并能提高得到的视频特征的优良程度和鲁棒性。

在本公开的一种实施例中，对帧图像逐级地进行多种类型的池化的包括：根据帧图像确定一个矩阵，利用多种类型的池化，逐级地生成更小的
30 矩阵，直到缩小为一个仅包含一个点的矩阵（或者，也可以将矩阵中的“点”称为矩阵中的“元素”），根据该仅包含一个点的矩阵确定该帧图像的图像特征。

图 2 为本公开的视频特征提取方法一个实施例提供的逐级进行多类型池化处理的示意性流程框图。具体地，请参阅图 2，本公开的视频特征提取方法一个实施例提供的步骤 S12 中的逐级进行多类型池化处理具体包括以下
35 步骤：

步骤（a），根据一个帧图像，确定一个具有第一矩阵维度和第二矩阵
35 维度（或者说，具有长度方向和宽度方向）的第一矩阵。不妨假设该帧图像的长度为 x 像素、宽度为 y 像素，其中的 x 和 y 为正整数。该第一矩阵

中的一个点（也可以将矩阵中的点称为矩阵中的元素，但为了与向量中的元素相区分，以下均将矩阵中的元素称为“点”）对应该帧图像中的一个像素，从而该第一矩阵为一个第一矩阵维度的长度为 x 、第二矩阵维度的长度为 y 的矩阵（即 $x*y$ 矩阵）；这里所说的矩阵的第一矩阵维度/第二矩阵维度的长度用于表示该矩阵在第一矩阵维度/第二矩阵维度上所包含的点的数量。该第一矩阵中的每个点的取值为一个 3 维的向量，将该 3 维的向量定义为第一向量，该第一向量用于表示该帧图像中的对应像素的三个颜色通道的亮度。需要注意的是，当视频对象的颜色模式为红绿蓝模式（RGB 模式）时，可以取红、绿、蓝三个颜色通道；但并非一定取红、绿、蓝三个颜色通道，例如，可以根据视频对象所使用的颜色模式进行选取；甚至所选取的颜色通道的数量也并非必须是三个，例如，可以选取红绿蓝三个颜色通道中的两个。此后，处理进到步骤（b）。

步骤（b），在第一矩阵上设置多个第一区块（事实上每个区块相当于一个池化窗，因此也可将第一区块称为第一池化窗），不妨设置 x_1*y_1 个第一区块，其中的 x_1 和 y_1 为正整数，每个第一区块包含多个该第一矩阵的点（或者说，包含多个第一向量）；该多个第一区块在第一矩阵维度上的数量少于该第一矩阵的第一矩阵维度的长度（或者说，少于该第一矩阵在第一矩阵维度上所包含的点的数量），且该多个第一区块在第二矩阵维度上的数量少于该第一矩阵的第二矩阵维度的长度（或者说，少于该第一矩阵在第二矩阵维度上所包含的点的数量），即有 x_1 的值小于 x ，且 y_1 的值小于 y 。对于每个第一区块，分别计算第一区块所包含的多个第一向量的各个维的最大值、最小值和平均值，得到该第一区块对应的一个 9 维的向量，将该 9 维的向量定义为第二向量。需要说明的是，各个第一区块之间可以部分相互重叠，即可以包含相同的点，也可以不相互重叠。此后，处理进到步骤（c）。

具体地，在设置第一区块时，可以均匀地将第一矩阵的第一矩阵维度分成 x_1 段，每段具有相同的长度，且相邻两段之间包含相同的点（部分重叠），按照同样的方式，将第一矩阵的第二矩阵维度分成 y_1 段，再将该 x_1 段与该 y_1 段进行组合，得到第一矩阵的 x_1*y_1 个第一区块。

需要说明的是，当设置的每个第一区块具有相同的大小和相同的间距时（相邻的两个第一区块可以重叠），前述的在第一矩阵上设置多个第一区块并计算各个第一区块的第二向量的过程，事实上等同于用一个池化窗按照一定间距扫描（或者说划过）整个第一矩阵，并在每次扫描中，计算该池化窗覆盖的区域的第二向量。

步骤（c），根据该多个 x_1*y_1 个第一区块以及每个第一区块对应的第二向量，确定第二矩阵；该第二矩阵中的一个点对应一个第一区块，当设置

了 $x_1 * y_1$ 个第一区块时,该第二矩阵就是一个第一矩阵维度的长度为 x_1 、第二矩阵维度的长度为 y_1 的矩阵(即 $x_1 * y_1$ 矩阵);该第二矩阵中的各个点的取值为对应的第一区块的该第二向量。此后,处理进到步骤(d)。

需要说明的是,在确定第二矩阵时,需要按照一定顺序进行第一区块与第二矩阵中的点的对应。作为一种具体示例,可以按照各个第一区块在第一矩阵中的位置顺序,对第二矩阵中的各个点进行排列。

步骤(d),重复步骤(b)和步骤(c):根据包含 $x_1 * y_1$ 个点且每个点的取值为9维向量的第二矩阵,得到包含 $x_2 * y_2$ 个点且每个点的取值为27维向量的第三矩阵(其中的 x_2 为小于 x_1 的正整数, y_2 为小于 y_1 的正整数);再根据包含 $x_2 * y_2$ 个点且每个点的取值为27维向量的第三矩阵,得到包含 $x_3 * y_3$ 个点且每个点的取值为81维向量的第三矩阵(其中的 x_3 为小于 x_2 的正整数, y_3 为小于 y_2 的正整数);...;直到将该第一矩阵(或者说,该帧图像)缩小成一个 $1*1$ 的第N矩阵(事实上,就是将矩阵降维成了一个点),其中的N为正整数,该第N矩阵仅包括一个点,该点的取值为一个 3^N 维的向量;将该 3^N 维向量确定为该帧图像的图像特征。

需要注意的是,在步骤(d)中,在各次的设置区块的过程中,应根据矩阵的大小采用相应的方式来设置区块,以适应矩阵的第一矩阵维度、第二矩阵维度的逐级减小。

在本公开的实施例中,还可以包括以下步骤:对所确定的图像特征进行二值化处理,得到二值化图像特征,该二值化图像特征为由0/1构成的比特串;再根据所得到的二值化图像特征,确定视频特征。

将图像特征进行二值化处理,能够压缩视频特征的存储,并加速视频比对的相似度计算过程,另外,进行二值化处理还有利于进行视频比对的索引库召回过程。

具体地,可以利用随机投影(random projection)法将图像特征转化为二值化形式的图像特征,该方法特别适用于将向量形式的图像特征二值化。图3为本公开的视频特征提取方法一个实施例提供的利用随机投影法对图像特征进行二值化的示意性框图。请参阅图3,本公开示例的利用随机投影法对图像特征进行二值化处理的过程,主要包括以下步骤:

步骤S21,为了生成长度为n的二值化图像特征,根据图像特征,生成2n个小组(group),每个小组包含该图像特征中的多个元素(也就是,每个小组包含图像特征的多个维度的数值)。其中的n为正整数。此后,处理进到步骤S22。

需要说明的是,一个小组具体包含哪些元素是任意的,并且两个不同小组可以包括一些相同的元素。不过,为了便于视频比对,每个小组包含具体包含哪些元素可以是预设的,或者对多个视频对象,可以采用同样的

方式来生成该小组。

在本示例中，每个小组所包含的元素的数量是相同的。但需要说明的是，事实上各个小组所包含的元素的数量也可以是不同的。

步骤 S22，分别对每个小组所包括的多个元素进行求和，以得到每个小组的加和值。此后，处理进到步骤 S23。

步骤 S23，将该 $2n$ 个小组两两配对，得到 n 个小组对。此后，处理进到步骤 S24。

具体地，可以预先将 $2n$ 个小组排序（或者将小组编号），并将相邻的两个小组配成一对。

步骤 S24，分别对 n 个小组对进行比较，比较每一个小组对中的两个小组的加和值的大小，根据比较的结果生成一个二值化的图像特征比特。此后，处理进到步骤 S25。

具体地，在预先已将小组排序（或者已将编号）的示例中，在一对小组中，如果排序靠前的小组的加和值大于排序靠后的小组的加和值，则生成一个取值为 1 的二值化图像特征比特，反之则生成一个取值为 0 的二值化图像特征比特。需要说明的是，不限制生成二值化图像特征比特的方式，比如，可以当排序靠前的小组的加和值小于排序靠后的小组的加和值时生成取值为 1 的二值化图像特征比特。

步骤 S25，根据该 n 个小组对的该 n 个二值化图像特征比特，组成该帧图像的长度为 n 的二值化图像特征。

图 4 为利用本公开的视频特征提取方法进行一次具体的提取帧图像的图像特征过程的示意性流程图。请参阅图 4，利用本公开的实施例提供的提取帧图像的图像特征的一个具体示例的步骤如下：

步骤 S31，对于从视频对象抽样得到的一个 243×243 的帧图像（长为 243 像素，宽为 243 像素），每个像素有红绿蓝 3 个通道，在图 4 中用 I、II、III 分别标识出了红、绿、蓝三个通道。根据帧图像定义第一矩阵：将第一矩阵中每一个点与在帧图像中的相同位置的像素相对应，根据每个像素的红绿蓝 3 个通道的亮度取值定义对应的点的取值，从而得到了一个 243×243 的第一矩阵，且第一矩阵中的点的取值为一个 3 维向量。

步骤 S32，用一个 13×13 的矩阵方块（或者，可将该矩阵方块称为池化窗）划过第一矩阵；

获取矩阵方块所覆盖的 13×13 个点的每一维（事实上就是红绿蓝三个颜色通道的亮度）的最大值、最小值和中间值，从而得到一个 9 维向量；

矩阵方块每次向第一矩阵的长度方向或向宽度方向移动 3 个点，矩阵方块沿长宽方向依次划过所有点，并计算矩阵方块所覆盖的多个点的每一维的最大值、最小值和中间值；

处理完整个第一矩阵后，得到 81×81 的第二矩阵，该第二矩阵中的点的取值为 9 维向量。

步骤 S33，重复步骤 S32，利用一个 10×10 的矩阵方块划过第二矩阵且每次划过 3 个点，得到 27×27 的第三矩阵，该第三矩阵中的点的取值为 27 维向量；利用一个 6×6 的矩阵方块划过第三矩阵且每次划过 2 个点，得到 9×9 的第四矩阵，该第四矩阵中的点的取值为 81 维向量；...，直到得到一个 1×1 的单点矩阵，该单点矩阵所包含的点的取值为 729 维向量，将该 729 维向量定义为池化向量。

步骤 S34，通过随机投影法，对该池化向量进行二值化处理，得到该帧图像的二值化的图像特征。

图 5 为本公开的视频特征库构建方法一个实施例的示意性流程图。请参阅图 5，本公开示例的视频特征库构建方法，主要包括以下步骤：

步骤 S41，按照前述的本公开示例的视频特征提取方法的步骤提取视频对象的视频特征。此后，处理进到步骤 S42。

步骤 S42，将视频对象的视频特征存储到视频特征库中。

需要说明的是，一个视频特征库中的视频特征应该是按照同样的特征提取方法得到的，即在步骤 S41 的视频特征提取的过程中，在步骤 S11 中基于同样的方式进行抽帧，在步骤 S12 中基于同样的方式逐级地对帧图像进行多种类型的池化，且在步骤 S13 中基于同样的方式将图像特征组成视频特征。另外，随着时间的推移，可以随时对视频特征库进行更新。

图 6 为本公开的视频特征提取装置一个实施例的示意性结构框图。请参阅图 6，本公开示例的视频特征提取装置 100 主要包括：

抽帧模块 110，用于对视频对象进行抽帧，得到一个或多个帧图像；

图像特征确定模块 120，用于对每个帧图像，逐级地进行多种类型的池化，得到该帧图像的图像特征；其中，所谓的多种类型的池化包括最大池化、最小池化和平均池化；

视频特征确定模块 130，用于根据该一个或多个帧图像对应的多个图像特征确定视频特征。

具体地，其中的图像特征确定模块 120 还用于按照前述的本公开的视频特征提取方法的实施例所示出的具体步骤，逐级地对帧图像进行多类型池化。

进一步地，本公开示例的视频特征提取装置 100 还包括二值化模块（图中未示出），用于按照前述的本公开的视频特征提取方法的实施例所示出的具体步骤，对图像特征进行二值化处理。此时，该视频特征确定模块 130，用于根据二值化后的图像特征确定视频特征。

图 7 为本公开的视频特征库构建装置一个实施例的示意性结构图。请

参阅图 7, 本公开示例的视频特征库构建装置 200 主要包括:

视频特征提取模块 201, 包括前述的本公开示例的视频特征提取装置的抽帧模块 110、图像特征确定模块 120 及视频特征确定模块 130, 并可包括二值化模块, 用于按照前述的本公开示例的视频特征提取方法的步骤提取
5 视频对象的视频特征。

视频特征存储模块 202, 用于将视频特征存储到视频特征库中。

视频特征库 203, 用于存储各个视频对象的视频特征。

图 8 是图示根据本公开的实施例的视频特征提取硬件装置的硬件框图。如图 8 所示, 根据本公开实施例的视频特征提取硬件装置 300 包括存储器
10 301 和处理器 302。视频特征提取硬件装置 300 中的各组件通过总线系统和/或其它形式的连接机构(未示出)互连。

该存储器 301 用于存储非暂时性计算机可读指令。具体地, 存储器 301 可以包括一个或多个计算机程序产品, 该计算机程序产品可以包括各种形式的计算机可读存储介质, 例如易失性存储器和/或非易失性存储器。该易
15 失性存储器例如可以包括随机存取存储器(RAM)和/或高速缓冲存储器(cache)等。该非易失性存储器例如可以包括只读存储器(ROM)、硬盘、闪存等。

该处理器 302 可以是中央处理单元(CPU)或者具有数据处理能力和/或指令执行能力的其它形式的处理单元, 并且可以控制视频特征提取硬件
20 装置 300 中的其它组件以执行期望的功能。在本公开的一个实施例中, 该处理器 302 用于运行该存储器 301 中存储的该计算机可读指令, 使得该视频特征提取硬件装置 300 执行前述的本公开各实施例的视频特征提取方法的全部或部分步骤。

图 9 是图示根据本公开的实施例的计算机可读存储介质的示意图。如
25 图 9 所示, 根据本公开实施例的计算机可读存储介质 400, 其上存储有非暂时性计算机可读指令 401。当该非暂时性计算机可读指令 401 由处理器运行时, 执行前述的本公开各实施例的视频特征提取方法的全部或部分步骤。

图 10 是图示根据本公开实施例的终端设备的硬件结构示意图。终端设备可以以各种形式来实施, 本公开中的终端设备可以包括但不限于诸如移动电话、智能电话、笔记本电脑、数字广播接收器、PDA(个人数字助理)、
30 PAD(平板电脑)、PMP(便携式多媒体播放器)、导航装置、车载终端设备、车载显示终端、车载电子后视镜等等的移动终端设备以及诸如数字 TV、台式计算机等等的固定终端设备。

如图 10 所示, 终端设备 1100 可以包括无线通信单元 1110、A/V(音频/视频)输入单元 1120、用户输入单元 1130、感测单元 1140、输出单元 1150、
35 存储器 1160、接口单元 1170、控制器 1180 和电源单元 1190 等等。图 10

示出了具有各种组件的终端设备，但是应理解的是，并不要求实施所有示出的组件。可以替代地实施更多或更少的组件。

其中，无线通信单元 1110 允许终端设备 1100 与无线通信系统或网络之间的无线电通信。A/V 输入单元 1120 用于接收音频或视频信号。用户输入单元 1130 可以根据用户输入的命令生成键输入数据以控制终端设备的各种操作。感测单元 1140 检测终端设备 1100 的当前状态、终端设备 1100 的位置、用户对于终端设备 1100 的触摸输入的有无、终端设备 1100 的取向、终端设备 1100 的加速或减速移动和方向等等，并且生成用于控制终端设备 1100 的操作的命令或信号。接口单元 1170 用作至少一个外部装置与终端设备 1100 连接可以通过的接口。输出单元 1150 被构造为以视觉、音频和/或触觉方式提供输出信号。存储器 1160 可以存储由控制器 1180 执行的处理和控制的软件程序等等，或者可以暂时地存储已经输出或将要输出的数据。存储器 1160 可以包括至少一种类型的存储介质。而且，终端设备 1100 可以与通过网络连接执行存储器 1160 的存储功能的网络存储装置协作。控制器 1180 通常控制终端设备的总体操作。另外，控制器 1180 可以包括用于再现或回放多媒体数据的多媒体模块。控制器 1180 可以执行模式识别处理，以将在触摸屏上执行的手写输入或者图片绘制输入识别为字符或图像。电源单元 1190 在控制器 1180 的控制下接收外部电力或内部电力并且提供操作各元件和组件所需的适当的电力。

本公开提出的视频特征提取方法的各种实施方式可以以使用例如计算机软件、硬件或其任何组合的计算机可读介质来实施。对于硬件实施，本公开提出的视频特征提取方法的各种实施方式可以通过使用特定用途集成电路 (ASIC)、数字信号处理器 (DSP)、数字信号处理装置 (DSPD)、可编程逻辑装置 (PLD)、现场可编程门阵列 (FPGA)、处理器、控制器、微控制器、微处理器、被设计为执行这里描述的功能的电子单元中的至少一种来实施，在一些情况下，本公开提出的视频特征提取方法的各种实施方式可以在控制器 1180 中实施。对于软件实施，本公开提出的视频特征提取方法的各种实施方式可以与允许执行至少一种功能或操作的单独的软件模块来实施。软件代码可以由以任何适当的编程语言编写的软件应用程序（或程序）来实施，软件代码可以存储在存储器 1160 中并且由控制器 1180 执行。

以上，根据本公开实施例的视频特征提取方法、装置、硬件装置、计算机可读存储介质以及终端设备，通过逐级地对由视频抽帧得到的帧图像进行多种类型的池化以生成视频特征，能够大大提高视频特征提取的准确性和提取的效率，并能提高得到的视频特征的优良程度和鲁棒性，进而使得基于本公开的视频特征提取方法得到的视频特征进行的视频比对、视频检索、视频消重以及视频内容监测具有更高的准确率、更高的效率和更好

的鲁棒性。

以上结合具体实施例描述了本公开的基本原理，但是，需要指出的是，在本公开中提及的优点、优势、效果等仅是示例而非限制，不能认为这些优点、优势、效果等是本公开的各个实施例必须具备的。另外，上述公开
5 的具体细节仅是为了示例的作用和便于理解的作用，而非限制，上述细节并不限制本公开为必须采用上述具体的细节来实现。

本公开中涉及的器件、装置、设备、系统的方框图仅作为例示性的例子并且不意图要求或暗示必须按照方框图示出的方式进行连接、布置、配置。如本领域技术人员将认识到的，可以按任意方式连接、布置、配置这
10 些器件、装置、设备、系统。诸如“包括”、“包含”、“具有”等等的词语是开放性词汇，指“包括但不限于”，且可与其互换使用。这里所使用的词汇“或”和“和”指词汇“和/或”，且可与其互换使用，除非上下文明确指示不是如此。这里所使用的词汇“诸如”指词组“诸如但不限于”，且可与其互换使用。

另外，如在此使用的，在以“至少一个”开始的项的列举中使用的“或”
15 指示分离的列举，以便例如“A、B 或 C 的至少一个”的列举意味着 A 或 B 或 C，或 AB 或 AC 或 BC，或 ABC（即 A 和 B 和 C）。此外，措辞“示例的”不意味着描述的例子是优选的或者比其他例子更好。

还需要指出的是，在本公开的系统和方法中，各部件或各步骤是可以
20 分解和/或重新组合的。这些分解和/或重新组合应视为本公开的等效方案。

可以不脱离由所附权利要求定义的教导的技术而进行对在此所述的技术的各种改变、替换和更改。此外，本公开的权利要求的范围不限于以上所述的处理、机器、制造、事件的组成、手段、方法和动作的具体方面。可以利用与在此所述的相应方面进行基本相同的功能或者实现基本相同的结果的当前存在的或者稍后要开发的处理、机器、制造、事件的组成、手
25 段、方法或动作。因而，所附权利要求包括在其范围内的这样的处理、机器、制造、事件的组成、手段、方法或动作。

提供所公开的方面的以上描述以使本领域的任何技术人员能够做出或者使用本公开。对这些方面的各种修改对于本领域技术人员而言是非常显
30 而易见的，并且在此定义的一般原理可以应用于其他方面而不脱离本公开的范围。因此，本公开不意图被限制到在此示出的方面，而是按照与在此公开的原理和新颖的特征一致的最宽范围。

为了例示和描述的目的已经给出了以上描述。此外，此描述不意图将本公开的实施例限制到在此公开的形式。尽管以上已经讨论了多个示例方面和实施例，但是本领域技术人员将认识到其某些变型、修改、改变、添
35 加和子组合。