

视频片段的识别方法、装置、设备以及存储介质

本申请要求于 2022 年 3 月 8 日提交中国专利局，申请号为 202210219364.3、发明名称为“视频片段的识别方法、装置、设备以及存储介质”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

本申请涉及计算机技术领域，特别涉及一种视频片段的识别方法、装置、设备以及存储介质。

背景技术

随着计算机技术的发展，视频呈海量增长之势，上网观看视频的用户越来越多。视频包括电视剧，而电视剧通常有片头和片尾。为了方便用户观看电视剧，视频平台会提供跳过片头和片尾的功能，而跳过片头和片尾的基础是确定电视剧中片头和片尾的位置。

相关技术中，电视剧片头和片尾位置的确定都是采用人工标注的方法实现的，即由人工观看电视剧，然后标记电视剧的片头和片尾的位置。

但是，采用人工标注的方法需要消耗大量的时间和人力资源，导致确定电视剧片头和片尾位置的效率较低。

发明内容

本申请各实施例提供了一种视频片段的识别方法、装置、设备、存储介质和计算机程序产品，技术方案如下：

一种视频片段的识别方法，所述方法包括：

获取第一视频的视频帧特征和至少一个第二视频的视频帧特征，并基于第一视频的视频帧特征以及至少一个第二视频的视频帧特征，确定多个视频帧对，所述视频帧对包括相似度符合相似度条件的第一视频帧和第二视频帧，所述第一视频帧属于所述第一视频，所述第二视频帧属于所述至少一个第二视频；

基于所述多个视频帧对的出现时间差值，将所述多个视频帧对中的第一视频帧进行融合，得到所述第一视频中的至少一个候选视频片段，所述出现时间差值是指所述视频帧对中的两个视频帧在视频中的出现时间之间的差值；及

获取目标时间范围，并基于所述至少一个候选视频片段以及所述目标时间范围，确定所述第一视频中的至少一个目标视频片段，所述目标视频片段处于所述第一视频的所述目标时间范围内。

一种视频片段的识别装置，所述装置包括：

视频帧对确定模块，用于获取第一视频的视频帧特征和至少一个第二视频的视频帧特

征，并基于第一视频的视频帧特征以及至少一个第二视频的视频帧特征，确定多个视频帧对，所述视频帧对包括相似度符合相似度条件的第一视频帧和第二视频帧，所述第一视频帧属于所述第一视频，所述第二视频帧属于所述至少一个第二视频；

融合模块，用于基于所述多个视频帧对的出现时间差值，将所述多个视频帧对中的第一视频帧进行融合，得到所述第一视频中的至少一个候选视频片段，所述出现时间差值是指所述视频帧对中的两个视频帧在视频中的出现时间之间的差值；及

目标视频片段确定模块，用于获取目标时间范围，并基于所述至少一个候选视频片段以及目标时间范围，确定所述第一视频中的至少一个目标视频片段，所述目标视频片段处于所述第一视频的所述目标时间范围内。

一种计算机设备，所述计算机设备包括一个或多个处理器和一个或多个存储器，所述一个或多个存储器中存储有至少一条计算机程序，所述计算机程序由所述一个或多个处理器加载并执行以实现所述视频片段的识别方法。

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中存储有至少一条计算机程序，所述计算机程序由处理器加载并执行以实现所述视频片段的识别方法。

提供了一种计算机程序产品，包括计算机程序，该计算机程序被处理器执行时实现上述视频片段的识别方法。

附图说明

图1是本申请实施例提供的一种视频片段的识别方法的实施环境的示意图；

图2是本申请实施例提供的一种视频片段的识别方法的流程图；

图3是本申请实施例提供的一种视频片段的识别方法的流程图；

图4是本申请实施例提供的一种提取视频帧特征的方法的流程图；

图5是本申请实施例提供的一种第一子片段和第二子片段的示意图；

图6是本申请实施例提供的一种不同重合方式的第一子片段的示意图；

图7是本申请实施例提供的一种候选视频片段融合的示意图；

图8是本申请实施例提供的一种视频片段的识别方法的流程图；

图9是本申请实施例提供的一种片段挖掘系统的流程图；

图10是本申请实施例提供的一种获取电视剧片头和片尾的方法的流程图；

图11是本申请实施例提供的一种片段数据库的存储方式的示意图；

图12是本申请实施例提供的一种获取电视剧片头和片尾的方法的流程图；

图13是本申请实施例提供的一种识别侵权视频的方法的流程图；

图14是本申请实施例提供的一种视频片段的识别方法的流程图；

图15是本申请实施例提供的一种视频片段的识别装置结构示意图；

图16是本申请实施例提供的一种终端的结构示意图；

图17是本申请实施例提供的一种服务器的结构示意图。

具体实施方式

本申请中术语“第一”“第二”等字样用于对作用和功能基本相同的相同项或相似项进行区分，应理解，“第一”、“第二”、“第n”之间不具有逻辑或时序上的依赖关系，也不对数量和执行顺序进行限定。

人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

机器学习(Machine Learning, ML)是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识子模型使之不断改善自身的性能。

汉明距离(Hamming Distance):用于衡量二进制特征间的距离,通过统计数值不同的特征位数量作为距离实现,如(1000)与(0011)的汉明距离为3。

需要说明的是,本申请所涉及的信息(包括但不限于用户设备信息、用户个人信息等)、数据(包括但不限于用于分析的数据、存储的数据、展示的数据等)以及信号,均为经用户授权或者经过各方充分授权的,且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。

图1是本申请实施例提供的一种视频片段的识别方法的实施环境示意图,参见图1,该实施环境中可以包括终端110和服务器140。

终端110通过无线网络或有线网络与服务器140相连。可选地,终端110是车载终端、智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表以及智能电视等,但并不局限于此。终端110安装和运行有支持视频片段识别的应用程序。

服务器140是独立的物理服务器,或者是多个物理服务器构成的服务器集群或者分布式系统,或者是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、分发网络(Content Delivery Network, CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。服务器140为该终端110上运行的应用程序提供后台服务。

本申请实施例对终端110和服务器140的数量不做限制。

在介绍完本申请实施例的实施环境之后,下面将结合上述实施环境,对本申请实施例的应用场景进行介绍,在下述说明过程中,终端也即是上述实施环境中的终端110,服务器也即是上述实施环境中的服务器140。

本申请实施例提供的视频片段的识别方法能够应用在识别视频的片头和片尾的场景下,比如,应用在识别电视剧的片头和片尾的场景下,或者应用在识别纪录片的片头和片尾的场景下,或者应用在识别短视频集合的片头和片尾的场景下等。

以本申请实施例提供的视频片段的识别方法应用在识别电视剧片头和片尾的场景下

为例，技术人员通过终端选择需要进行片头片尾识别的电视剧，该电视剧包括多个视频，每个视频为电视剧中的一集。在通过终端选中该电视剧的情况下，服务器能够采用本申请实施例各提供的技术方案，基于该电视剧中的多个视频来进行处理，得到该多个视频中的片头和片尾。在对该多个视频进行处理的过程中，服务器基于第一视频的视频帧特征以及至少一个第二视频的视频帧特征，确定多个视频帧对，每个视频帧对包括相似度符合相似度条件的第一视频帧和第二视频帧，第一视频帧属于第一视频，第二视频帧属于该至少一个第二视频，也就是说，每个视频帧对包括第一视频中的一个视频帧以及第二视频中的一个视频帧，第一视频和该至少一个第二视频均属于该多个视频。服务器基于该多个视频帧对的出现时间差值，将该多个视频帧对中的第一视频帧进行融合，得到该第一视频中的至少一个候选视频片段，出现时间差值是指视频帧对中的两个视频帧在视频中的出现时间之间的差值，也即是视频帧对中第一视频帧在第一视频中的出现时间与第二视频帧在第二视频中的出现时间之间的差值。服务器基于至少一个候选视频片段以及目标时间范围，确定第一视频中的至少一个目标视频片段，由于是应用在识别电视剧片头和片尾的场景下，那么该目标时间段也即是片头或者片尾所在的时间段，确定出的目标视频片段也即是第一视频的片头或者片尾。

需要说明的是，上述是以本申请实施例提供的视频片段的识别方法应用在识别电视剧片头和片尾的场景下为例进行说明的，上述其他应用场景的实施过程与上述说明属于同一发明构思，实施过程不再赘述。

另外，本申请实施例提供的视频片段的识别方法除了能够应用在上述识别电视剧的片头和片尾的场景、识别纪录片的片头和片尾的场景以及识别短视频集合的片头和片尾的场景之外，也能够应用在识别其他类型视频的片头和片尾的场景中，本申请实施例对此不做限定。

介绍完本申请实施例的实施环境和应用场景之后，下面对本申请实施例提供的视频片段的识别方法进行说明，参见图 2，本申请实施例提供的技术方案可以由计算机设备执行，其中，计算机设备为终端或服务器执行，本申请实施例提供的技术方案也可以由终端和服务器共同执行。在本申请下述实施例中，以执行主体为服务器为例进行说明，容易理解地，虽然下述实施例是以服务器为例进行说明，但是，本申请的各实施例也可由终端执行。也即，本申请各实施例提供的技术方案实则是由计算机设备执行的。

视频片段的识别方法包括：

201、服务器获取第一视频的视频帧特征和至少一个第二视频的视频帧特征，并基于第一视频的视频帧特征以及至少一个第二视频的视频帧特征，确定多个视频帧对，该视频帧对包括相似度符合相似度条件的第一视频帧和第二视频帧，该第一视频帧属于该第一视频，该第二视频帧属于该至少一个第二视频。

其中，第一视频和至少一个第二视频属于同一个视频集合，比如，第一视频和第二视频为同一部电视剧的不同集。视频帧特征为视频帧的嵌入特征，比如为深度哈希特征。第

一视频帧和第二视频帧之间的相似度通过第一视频帧的视频帧特征以及第二视频帧的视频帧特征来确定。每个视频帧对包括一个第一视频帧和一个第二视频帧，且视频帧对中的第一视频帧和第二视频帧之间的相似度符合相似度条件，也即是视频帧对中的第一视频帧和第二视频帧为相似度较高的两个视频帧。

202、服务器基于该多个视频帧对的出现时间差值，将该多个视频帧对中的第一视频帧进行融合，得到该第一视频中的至少一个候选视频片段，该出现时间差值是指该视频帧对中的两个视频帧在视频中的出现时间之间的差值。

其中，视频帧对中的第一视频帧是与第二视频帧之间相似度较高的视频帧，而候选视频片段是由多个视频帧对中的第一视频帧融合得到的，那么候选视频片段也即是第一视频中与至少一个第二视频具有重合内容的视频片段。出现时间差值能够反映第一视频帧和第二视频帧在第一视频和第二视频中出现时间的偏差。

203、服务器获取目标时间范围，并基于该至少一个候选视频片段以及目标时间范围，确定该第一视频中的至少一个目标视频片段，该目标视频片段处于该第一视频的目标时间范围内。

其中，目标时间范围是指视频中时间范围，目标时间范围由技术人员根据实际情况进行设置，本申请实施例对此不做限定。

通过本申请实施例提供的技术方案，基于视频帧特征之间的相似度，确定包含相似视频帧的视频帧对。基于出现时间差值来对视频帧对中的第一视频帧进行融合，得到至少一个候选视频片段。最终从至少一个候选视频片段中确定出处于目标时间范围的目标视频片段。确定目标片段的过程无需人工参与，由计算机设备直接基于第一视频和至少一个第二视频就能够自动进行，效率较高。

上述步骤 201-203 是对本申请实施例提供的视频片段的识别方法的简单介绍，下面将结合一些例子，对本申请实施例提供的视频片段的识别方法进行更加详细的说明，参见图 3，本申请实施例提供的技术方案可以由终端或服务器执行，也可以由终端和服务器共同执行，在本申请实施例中，以执行主体为服务器为例进行说明，方法包括：

301、服务器对第一视频和至少一个第二视频进行特征提取，得到第一视频的视频帧特征以及至少一个第二视频的视频帧特征。

在其中一种可能的实施方式中，服务器将第一视频和至少一个第二视频输入特征提取模型，通过该特征提取模型对该第一视频和该至少一个第二视频进行特征提取，得到该第一视频的视频帧特征以及该至少一个第二视频的视频帧特征。

其中，服务器通过特征提取模型对第一视频和至少一个第二视频进行特征提取的过程，也即是对第一视频的第一视频帧以及第二视频的第二视频帧进行特征提取的过程，在这种情况下，该特征提取模型为一个图像特征提取模型。

在这种实施方式下，通过特征提取模型对该第一视频和该至少一个第二视频进行特征提取，得到第一视频的视频帧特征以及至少一个第二视频的视频帧特征，从而实现对第一

视频和至少一个第二视频进行抽象表达，提高后续的运算效率。

为了对上述实施方式进行说明，下面通过三个例子对上述实施方式进行说明。

例 1、服务器将该第一视频和该至少一个第二视频输入特征提取模型，通过特征提取模型对多个第一视频帧和多个第二视频帧进行卷积和池化，得到该多个第一视频帧的视频帧特征以及多个第二视频帧的视频帧特征，其中，多个第一视频帧为第一视频的视频帧，多个第二视频帧为至少一个第二视频的视频帧。

下面以服务器对第一视频进行特征提取的方法进行说明，服务器将该第一视频的多个第一视频帧输入特征提取模型，通过该特征提取模型的卷积层，对该多个第一视频帧进行卷积，得到该多个第一视频帧的特征图。服务器通过该特征提取模型的池化层，对该多个第一视频帧的特征图进行最大池化或者平均池化中的任一项，得到该多个第一视频帧的视频帧特征。在一些实施例中，服务器以矩阵的形式来表示第一视频帧，以向量的形式来表示视频帧特征，在对第一视频帧进行卷积的过程中，采用卷积核在第一视频帧上进行滑动的方式来实现。

在一些实施例中，该特征提取模型为基于卷积神经网络（Convolutional Neural Networks, CNN）的特征提取器，比如为采用大规模开源数据集 imagenet（图网）上预训练的神经网络 Resnet-101（残差网络 101），该神经网络 Resnet101 的结构参见表 1。该神经网络 Resnet-101 的池化（Pooling）层的输出结果为视频帧特征，其中，101 是指模型的层数，该视频帧特征为一个 1×2048 的向量。

表 1

Layer name	Output size	ResNet-101	
Conv1	300×500	7×7, 64, stride 2	
Conv2_x	150×250	3×3 max pool, stride 2	
		1×1, 64	×3 blocks
		3×3, 64	
		1×1, 256	
Conv3_x	76×126	1×1, 128	×4 blocks
		3×3, 128	
		1×1, 512	
Conv4_x	38×63	1×1, 256	×23 blocks
		3×3, 256	
		1×1, 1024	
Conv5_x	19×32	1×1, 512	×3 blocks
		3×3, 512	
		1×1, 2048	
Pool	1×2048	Max pool	

Class	1×1000	full connection	
-------	--------	-----------------	--

其中，Layer name 为特征提取模型 ResNet-101 中各个层面的名称，Output size 为输出的特征图的尺寸，max pool 指最大值池化，stride 是指步长，blocks 是指层，一层可能包括多个卷积核，Conv 是指卷积层，Pool 是指池化层，Class 是指分类层，full connection 是指全连接，在上述提取视频帧特征的过程中，不使用 Class 层。

需要说明的是，上述是以特征提取模型为 ResNet-101 为例进行说明的，在其他可能的实施方式中，该特征提取模型还可以为其他结构，本申请实施例对此不做限定。

另外，上述特征提取过程是基于卷积来实现的，得到的视频帧特征用于表达视频帧的图像纹理的特征，这样的视频帧特征也被称为视频帧的底层特征。在其他可能的实施方式中，该特征提取模型还能够提取视频帧的语义特征，得到的视频帧特征能够反映视频帧的语义，下面对服务器通过该特征提取模型来提取视频帧的语义特征的方法进行说明。

例 2、服务器将该第一视频和该至少一个第二视频输入特征提取模型，通过特征提取模型，基于注意力机制对多个第一视频帧和多个第二视频帧进行编码，得到该多个第一视频帧的视频帧特征以及多个第二视频帧的视频帧特征，其中，多个第一视频帧为第一视频的视频帧，多个第二视频帧为至少一个第二视频的视频帧，通过该特征提取模型获取的视频帧特征也即是对应视频帧的语义特征。在这种实施方式下，该特征提取模型为语义特征编码器，比如为 Transformer 编码器。

下面以服务器对多个第一视频进行特征提取的方法进行说明，服务器将该第一视频的多个第一视频帧输入特征提取模型，通过该特征提取模型，对该多个第一视频帧进行嵌入编码，得到多个嵌入向量，一个嵌入向量对应于一个第一视频帧，嵌入向量用于表示第一视频帧在第一视频中的位置以及第一视频帧的内容。服务器将多个嵌入向量输入特征提取模型，通过特征提取模型的三个线性变换矩阵，对多个嵌入向量进行线性变换，得到每个第一视频帧对应的查询（Query）向量、键（Key）向量以及值（Value）向量。服务器通过特征提取模型，基于多个第一视频帧对应的查询向量以及键向量，获取多个第一视频帧的注意力权重。服务器通过特征提取模型，基于每个第一视频帧的注意力权重和每个第一视频帧的值向量，获取每个第一视频帧的注意力编码向量，注意力编码向量也即是第一视频帧的视频帧特征。

比如，服务器通过特征提取模型，将每个嵌入向量分别与三个线性变换矩阵相乘，得到每个第一视频帧分别对应的查询向量、键向量以及值向量。对于多个第一视频帧中的第一个第一视频帧，服务器通过特征提取模型，基于第一个第一视频帧的查询向量，与多个第一视频帧的键向量，确定多个第一视频帧对第一个第一视频帧之间的多个注意力权重。对于多个第一视频帧中的第一个第一视频帧，服务器通过特征提取模型，将多个第一视频帧对第一个第一视频帧的注意力权重，与多个第一视频帧的值向量进行加权求和，得到第一个第一视频帧的注意力编码向量，也即是第一个第一视频帧的视频帧特征。

上述例 1 和例 2 分别以该特征提取模型提取视频帧的底层特征以及语义特征为例进

行说明的，在其他可能的实施方式中，服务器还能够通过该特征提取模型同时获取视频帧的底层特征以及语义特征，下面通过例 3 进行说明。

例 3、服务器将该第一视频和该至少一个第二视频输入特征提取模型，通过特征提取模型对多个第一视频帧和多个第二视频帧进行卷积和池化，得到该多个第一视频帧的底层特征以及多个第二视频帧的底层特征，其中，多个第一视频帧为第一视频的视频帧，多个第二视频帧为至少一个第二视频的视频帧。服务器通过该特征提取模型，基于注意力机制对多个第一视频帧和多个第二视频帧进行编码，得到该多个第一视频帧的语义特征以及多个第二视频帧的语义特征。服务器将各个第一视频帧的底层特征和语义特征进行融合，得到各个第一视频帧的视频帧特征。服务器将各个第二视频帧的底层特征和语义特征进行融合，得到各个第二视频帧的视频帧特征。

举例来说，该特征提取模型包括第一子模型和第二子模型，该第一子模型用于提取视频帧的底层特征，该第二子模型用于提取视频帧的语义特征。服务器将该第一视频和该至少一个第二视频输入特征提取模型之后，通过该第一子模型来获取该多个第一视频帧的底层特征以及多个第二视频帧的底层特征，通过第二子模型来获取该多个第一视频帧的语义特征以及多个第二视频帧的语义特征。服务器将各个视频帧的底层特征和语义特征进行融合时，可以采用加权求和的方式，加权求和的权重由技术人员根据实际情况进行设置，比如设置为 0.5，本申请实施例对此不做限定。服务器通过该第一子模型和该第二子模型获取视频帧的底层特征和语义特征的方法分别与上述例 1 和例 2 同理，因此在此不再赘述。

需要说明的是，上述是以特征提取模型提取视频帧的底层特征和语义特征为例进行说明的，随着科学技术的发展，服务器还能够采用其他结构的特征提取模型来获取视频帧特征，本申请实施例对此不做限定。

在一些实施例中，第一视频和至少一个第二视频是属于同一个视频集合中的视频，其中，第一视频是待确定目标视频片段的视频，该至少一个第二视频为该视频集合中除该第一视频以外的全部视频，或者，该至少一个第二视频为从该视频集合中抽取的视频，抽取时屏蔽该第一视频。在该至少一个第二视频为从视频集合中抽取的视频的情况下，服务器从该视频集合中随机抽取目标视频数量个第二视频，在抽取过程中，屏蔽该第一视频，也即是抽取出的目标视频数量个第二视频中不包括该第一视频，该目标视频数量由技术人员根据实际情况进行设置，本申请实施例对此不做限定。服务器将该第一视频和该至少一个第二视频分别组成至少一个视频对，每个视频对包括该第一视频和该至少一个第二视频中的一个第二视频。

比如，在该视频集合包括 46 个视频的情况下，对于每个第一视频 i ，服务器从该视频集合的剩余视频中随机抽取 10 个第二视频 r ，将该第一视频 i 和该 10 个第二视频 r 分别组成 10 个视频对，在后续处理过程中，以视频对为单位来进行，其中，10 也即是目标视频数量。

另外，在一些实施例中，服务器对该第一视频和该至少一个第二视频进行特征提取之

前,对该第一视频和该至少一个第二视频进行抽帧,得到该第一视频的多个第一视频帧和各个第二视频的多个第二视频帧。通过对视频进行抽帧,能够减少后续特征提取过程的运算量,能够提升特征提取的效率。

以第一视频为例,服务器以目标间隔从第一视频中进行抽帧,得到该第一视频的多个第一视频帧,其中,目标间隔是指第一视频的目标播放时长,比如 1s,或者,该目标间隔是指目标数量的帧间隔,比如 25 帧。在该目标间隔是指第一视频的目标播放时长的情况下,服务器每隔目标播放时长从该第一视频中抽取一帧作为第一视频帧。在第一视频为 6s,目标播放时长为 1s 的情况下,服务器从该第一视频中抽取 6 个第一视频帧。在该目标时间间隔是指目标数量的帧间隔的情况下,服务器每隔目标数量的视频帧从该第一视频中进行抽取,得到多个第一视频帧。在第一视频包括 100 个视频帧,目标数量为 10 的情况下,服务器从该第一视频中抽取 10 个第一视频帧。比如,参见图 4,服务器以目标间隔从第一视频 400 中进行抽帧,得到该第一视频的多个第一视频帧 401。服务器将该第一视频的多个第一视频帧 401 输入特征提取模型 402,通过该特征提取模型 402 输出该多个第一视频帧 401 的视频帧特征 403。

需要说明的是,上述步骤 301 为可选步骤,既可以是服务器提前执行的,也可以是服务器在执行本申请实施例提供的技术方案时执行的,本申请实施例对此不做限定。

302、服务器基于第一视频的视频帧特征以及至少一个第二视频的视频帧特征,确定多个视频帧对,该视频帧对包括相似度符合相似度条件的第一视频帧和第二视频帧,该第一视频帧属于该第一视频,该第二视频帧属于该至少一个第二视频。

在其中一种可能的实施方式中,服务器确定多个第一视频帧的视频帧特征与多个第二视频帧的视频帧特征之间的相似度。服务器将相似度符合目标条件的第一视频帧和第二视频帧确定为一个视频帧对,每个视频帧对包括一个第一视频帧和一个第二视频帧。

其中,视频帧特征之间的相似度通过欧氏距离或者余弦相似度,本申请实施例对此不做限定。

在这种实施方式下,服务器能够基于第一视频帧和第二视频帧之间的相似度来确定多个视频帧对,由于视频帧对中的视频帧为不同视频中相似度较高的视频帧,后续基于视频帧对就能够快捷地确定出相似的视频片段,从而最终确定出目标视频片段,效率较高。

在相似度为欧式距离的情况下,服务器确定多个第一视频帧的视频帧特征与多个第二视频帧的视频帧特征之间的欧式距离。服务器将欧式距离小于或等于距离阈值的第一视频帧和第二视频帧确定为一个视频帧对。其中,距离阈值由技术人员根据实际情况进行设置,本申请实施例对此不做限定。在距离阈值为 0.5 的情况下,在任一第一视频帧的视频帧特征与任一第二视频帧的视频帧特征之间的欧式距离小于或等于 0.5 的情况下,服务器将该第一视频帧和该第二视频帧确定为一个视频帧对。

在相似度为余弦相似度的情况下,服务器确定多个第一视频帧的视频帧特征与多个第二视频帧的视频帧特征之间的余弦相似度。服务器将余弦相似度大于或等于相似度阈值的

第一视频帧和第二视频帧确定为一个视频帧对。在相似度阈值为 0.8 的情况下，在任一第一视频帧的视频帧特征与任一第二视频帧的视频帧特征之间的余弦相似度大于或等于 0.8 的情况下，服务器将该第一视频帧和该第二视频帧确定为一个视频帧对。

在其中一些实施例中，在服务器将第一视频和至少一个第二视频组成至少一个视频对的情况下，服务器以视频对为单位来确定视频对中第一视频的视频帧特征以及第二视频的视频帧特征之间的相似度，从而确定视频对下的多个视频帧对。比如，对于视频对 (i, r) 来说，服务器确定第一视频 i 的视频帧特征与第二视频 r 的视频帧特征之间的相似度。服务器将相似度符合目标条件的第一视频帧和第二视频帧确定为一个视频帧对。也即是，对于第一视频 i 中每个第一视频帧 j，确定第一视频帧 j 与第二视频 r 中每个第二视频帧的视频帧特征之间的欧式距离。服务器将欧式距离小于 t_0 的第二视频帧作为第一视频帧 j 的相似帧，该第一视频帧 j 与该相似帧组成一个视频帧对。服务器将获取到的第一视频帧 j 的相似帧存储在第一列表中，该第一列表也被称为相似帧列表 (sim-id-list)。在一些实施例中，服务器将帧的标识存储在该第一列表中，帧的标识用于指示帧所属的视频以及帧在视频中的位置。比如，对于 $j=1$ 的第一视频帧，相似帧列表 sim-id-list 为 [1, 2, 3]，表示与第二视频 r 的第 1、2、3 秒对应的视频帧为相似帧，其中， $j=1$ 表示第一视频中第 1 秒对应的视频帧。

在其中一个实施例中，基于第一视频的视频帧特征以及至少一个第二视频的视频帧特征，确定多个视频帧对包括：获取视频集合，并将所述视频集合中的待确定目标视频片段的视频作为第一视频；将所述视频集合中的、不同于所述第一视频的至少一个视频，作为至少一个第二视频；将所述第一视频和所述至少一个第二视频分别组成至少一个视频对；所述视频对包括所述第一视频和所述至少一个第二视频中的一个第二视频；将处于相同视频对中的第一视频和第二视频进行视频帧特征的相似度计算，得到相似度计算结果；根据所述相似度计算结果，将处于相同视频对中的、相似度符合相似度条件的第一视频帧和第二视频帧作为一对视频帧对；所述第一视频帧属于所述第一视频，所述第二视频帧属于所述至少一个第二视频。

可选地，在步骤 302 之后，在确定出的视频帧对的数量为 0 的情况下，服务器确定该第一视频中不存在目标视频片段。

303、服务器确定多个视频帧对的出现时间差值。

在其中一种可能的实施方式中，服务器将该多个视频帧对中第一视频帧在第一视频中的出现时间与该视频帧对中第二视频帧在第二视频中的出现时间相减，得到该多个视频帧对的出现时间差值。在一些实施例中，服务器将该多个视频帧对出现时间差值存储在第二列表中，该第二列表也被称为出现时间差值列表 (diff-time-list)，在后续处理过程中，能够直接从该第二列表中调用对应的出现时间差值。比如，对于 $j=1$ 的第一视频帧，相似帧列表 sim-id-list 为 [1, 2, 3]，那么对应的出现时间差值列表 diff-time-list 为 [0, 1, 2]。

304、服务器基于该多个视频帧对的出现时间差值，将该多个视频帧对划分为多个视频帧组，同一个该视频帧组中的视频帧对对应于同一个出现时间差值，该出现时间差值是指该视频帧对中的两个视频帧在视频中的出现时间之间的差值。

在其中一种可能的实施方式中，针对多个视频帧对中的任一视频帧对，服务器确定视频帧对中的第一视频帧的第一出现时间和第二视频帧的第二出现时间；第一出现时间是指第一视频帧在第一视频中出现的时间，第二出现时间是指第二视频帧在第二视频中出现的时间；服务器将视频帧对中的第一视频帧的第一出现时间减去第二视频帧的第二出现时间，得到视频帧对的出现时间差值；服务器将出现时间差值相同的视频帧对划分为一个初始视频帧组，并将初始视频帧组中的视频帧对的出现时间差值，作为初始视频帧组所对应的出现时间差值。服务器基于多个初始视频帧组对应的出现时间差值，将该多个初始视频帧组进行融合，得到该多个视频帧组。

其中，初始视频帧组包括多个出现时间差值相同的视频帧对，不同初始视频帧组对应于不同的出现时间差值，其中，初始视频帧组对应的出现时间差值是指该初始视频帧组中视频帧对的出现时间差值。

在其中一种可能的实施方式中，在基于多个视频帧对的出现时间差值，将多个视频帧对划分为多个视频帧组之前，上述方法还包括：针对多个视频帧对中的任一视频帧对，将视频帧对中的第一视频帧的第一出现时间减去视频帧对中的第二视频帧的第二出现时间，得到视频帧对的出现时间差值；第一出现时间是指第一视频帧在第一视频中出现的时间，第二出现时间是指第二视频帧在第二视频中出现的时间。

在这种实施方式下，出现时间差值相同的视频帧对中的视频帧可能会构成完整的视频片段，通过将视频帧对聚合为视频帧组，便于后续确定相似的视频片段。

举例来说，服务器获取预设的配置信息，并按照配置信息中的目标顺序对该多个初始视频帧组进行排序，对于多个候选视频帧组中任两个相邻的候选视频帧组，在两个相邻的候选视频帧组之间的匹配时间差值符合匹配时间差值条件的情况下，将两个相邻的候选视频帧组融合为一个视频帧组，匹配时间差值是指两个相邻的候选视频帧组对应的出现时间差值之间的差值。

服务器根据预配置信息中的目标顺序对多个初始视频帧组进行排序，得到多个候选视频帧组。在该多个候选视频帧组中任两个相邻的候选视频帧组之间的匹配时间差值符合匹配时间差值条件的情况下，服务器将该两个相邻的候选视频帧组融合为一个视频帧组，该匹配时间差值是指该两个相邻的候选视频帧组对应的出现时间差值之间的差值。

为了对上述举例中描述的技术过程进行更加清楚的说明，下面将分为两个部分对上述举例进行进一步说明。

第一部分、服务器按照目标顺序对该多个初始视频帧组进行排序，得到多个候选视频帧组。

在其中一种可能的实施方式中，服务器按照对应出现时间差值从小到大的顺序对该多

个初始视频帧组进行排序，得到多个候选视频帧组。在这种情况下，目标顺序是指出现时间差值从大至小的顺序。在一些实施例中，在任一初始视频帧组中，服务器按照视频帧对中第一视频帧在第一视频帧中出现时间的先后进行排序。

在这种实施方式下，服务器按照从大至小的顺序对该多个初始视频帧组进行排序，在得到的多个候选视频帧组中，任两个候选视频帧组对应的出现时间差值均较为接近，便于后续的融合过程。

举例来说，以多个初始视频帧组为[3, 5], [11, 12], [2, 4], [4, 6], [6, 9], [7, 10], [10, 11]，其中，每个括号代表一个视频帧对[i, r]，括号中的前一个数字为第一视频帧 i 的标识，第二个数字为第二视频帧 r 的标识，该标识为视频帧在视频中的出现时间。对于视频帧对[3, 5]来说，出现时间差值为 $5-3=2$ ，对于视频帧对[6, 9]来说，出现时间差值为 $9-6=3$ 。服务器按照对应出现时间差值从小到大的顺序对该多个初始视频帧组进行排序，得到多个候选视频帧组[10, 11], [11, 12], [2, 4], [3, 5], [4, 6], [6, 9], [7, 10]。

在其中一种可能的实施方式中，服务器按照对应出现时间差值从小至大的顺序对该多个初始视频帧组进行排序，得到多个候选视频帧组。在这种情况下，目标顺序是指出现时间差值从小至大的顺序。在一些实施例中，在任一初始视频帧组中，服务器按照视频帧对中第一视频帧在第一视频帧中出现时间的先后进行排序。

在这种实施方式下，服务器按照从小至大的顺序对该多个初始视频帧组进行排序，在得到的多个候选视频帧组中，任两个候选视频帧组对应的出现时间差值均较为接近，便于后续的融合过程。

在一些实施例中，在采用第一列表来存储视频帧对，采用第二列表存储出现差值的情况下，服务器基于第一列表和第二列表生成第三列表，该第三列表用于存储视频帧对以及出现差值，该第三列表能够存储多个初始视频帧组，比如，该第三列表的形式为第三列表 (match-dt-list): {d: {count, start-id, match-id-list}, ...}，其中，d 为出现时间差值，d: {count, start-id, match-id-list} 表示出现时间差值为 d 的初始视频帧组，count 为该初始视频帧组中视频帧对的数量，start-id 为第一视频帧的最小标识，match-id-list 为视频帧对。

第二部分、在该多个候选视频帧组中任两个相邻的候选视频帧组之间的匹配时间差值符合匹配时间差值条件的情况下，服务器将该两个相邻的候选视频帧组融合为一个视频帧组。

在其中一种可能的实施方式中，该两个相邻的候选视频帧组包括第一候选视频帧组和第二候选视频帧组，在该第一候选视频帧组对应的出现时间差值与该第二候选视频帧组对应的出现时间差值之间的匹配时间差值小于或等于匹配差值阈值的情况下，服务器将该第一候选视频帧组中的视频帧对添加至该第二候选视频帧组，得到该视频帧组。

其中，将该多个候选视频帧组融合为多个视频帧组包括多个迭代过程，在将该第一候

选视频帧组与该第二候选视频帧组融合为一个视频帧组之后，服务器还能够确定新融合的该视频帧组与后一个候选视频帧组之间的匹配时间差值，在该匹配时间差值符合匹配时间差值条件的情况下，将该新融合的该视频帧组与后一个候选视频帧组进行再一次融合，融合过程与融合该第一候选视频帧组与该第二候选视频帧组的过程属于同一发明构思，实现过程不再赘述。当然，在该匹配时间差值不符合匹配时间差值条件的情况下，服务器再确定该后一个候选视频帧组与后两个候选视频帧组之间的匹配时间差值，从而基于匹配时间差值进行进一步地处理。匹配差值阈值由技术人员根据实际情况进行设置，本申请实施例对此不做限定。

在这种实施方式下，通过基于出现时间差值来融合候选视频帧组，能够减少候选视频帧组的数量，从而减少后续处理的运算量，提高运算效率。

在其中一种可能的实施方式中，两个相邻的候选视频帧组包括第一候选视频帧组和第二候选视频帧组，将两个相邻的候选视频帧组融合为一个视频帧组包括：在第一候选视频帧组与第二候选视频帧组之间的匹配时间差值小于或等于匹配差值阈值的情况下，将第一候选视频帧组中的视频帧对添加至第二候选视频帧组；及基于第二候选视频帧组对应的出现时间差值，采用参考第二视频帧替换目标第二视频帧，得到视频帧组；目标第二视频帧为新添加至第二候选视频帧组中的第二视频帧；参考第二视频帧为第二视频中与目标第一视频帧之间的出现时间差值为目标差值的第二视频帧；目标差值为第二候选视频帧组对应的出现时间差值；目标第一视频帧为目标第二视频帧所属视频帧对中的第一视频帧。

举例来说，服务器确定该第一候选视频帧组和第二候选视频帧组的匹配时间差值。在该匹配时间差值小于或等于匹配差值阈值的情况下，服务器基于该第二候选视频帧组对应的出现时间差值，采用参考第二视频帧替换目标第二视频帧，得到该视频帧组，该目标第二视频帧为新添加至该第二候选视频帧组中的第二视频帧，该参考第二视频帧为该第二视频中与目标第一视频帧之间的出现时间差值为该第二候选视频帧组对应的出现时间差值的第二视频帧，该目标第一视频帧为该目标第二视频帧所属视频帧对中的第一视频帧。

在这种实施方式下，在将第一候选视频帧组中的视频帧对添加至第二候选视频帧组之后，服务器还能够根据第二候选视频帧组的出现时间差值对新添加至第二候选视频帧组中的视频帧对进行调整，以使得调整之后的视频帧对的出现时间差值与该第二候选视频帧组相同，保持视频帧对的出现差值与视频帧组的出现差值之间的一致性。

为了更加清楚的进行说明，下面以第一候选视频帧组对应的出现时间差值为3，包括[6, 9]，[7, 10]两个视频帧对，第二候选视频帧组对应的出现时间差值为2，包括[2, 4]，[3, 5]，[4, 6]三个视频帧对，匹配差值阈值为3为例进行说明。由于第一候选视频帧组与第二候选视频帧组之间的匹配时间差值为1，那么服务器确定该匹配时间差值小于该匹配差值阈值，需要对该第一候选视频帧组和第二候选视频帧组进行合并。服务器将第一候选视频帧组中的两个视频帧对[6, 9]和[7, 10]添加至第二候选视频帧组，该第二候选视频帧组变为[2, 4]，[3, 5]，[4, 6]，[6, 9]，[7, 10]，由于第二候选视频帧组对应的

出现时间差值为 2，那么服务器基于该出现时间差值 2，将添加至第二候选视频帧组中两个视频帧对[6, 9]和[7, 10]中的第二视频帧进行调整，得到两个新的视频帧对[6, 8]和[7, 9]。对新加入第二候选视频帧组的第二视频帧进行调整之后，该第二候选视频帧组变为[2, 4], [3, 5], [4, 6], [6, 8], [7, 9]，每个视频帧对的出现时间差值均为 2。

需要说明的是，上述是以服务器将第一候选视频帧组中的视频帧对添加至第二候选视频帧组中为例进行说明的，在其他可能的实施方式中，服务器也能够将第二候选视频帧中的视频帧对添加至第一候选视频帧组。

在一些实施例中，服务器基于第一候选视频帧组和第二候选视频帧组中视频帧对的数量来确定将第一候选视频帧组中的视频帧对添加至第二候选视频帧组，还是将第二候选视频帧组中的视频帧对添加至第一候选视频帧组。比如，在第一候选视频帧组中视频帧对的数量大于第二候选视频帧组中视频帧对的数量，服务器将该第二候选视频帧组中的视频帧对添加至该第一候选视频帧组。在第二候选视频帧组中视频帧对的数量大于第一候选视频帧组中视频帧对的数量，服务器将该第一候选视频帧组中的视频帧对添加至该第二候选视频帧组。在第二候选视频帧组中视频帧对的数量等于第一候选视频帧组中视频帧对的数量，服务器将该第一候选视频帧组中的视频帧对添加至该第二候选视频帧组。或者，在第二候选视频帧组中视频帧对的数量等于第一候选视频帧组中视频帧对的数量，服务器将该第二候选视频帧组中的视频帧对添加至该第一候选视频帧组。

在这种情况下，服务器能够根据候选视频帧组中视频帧对的数量来确定合并候选视频帧组的方式，将包括视频帧数量较少的候选视频帧组添加至包括视频帧数量较多的视频帧组，以减少运算量，提高效率。

305、对于该多个视频帧组中的任一视频帧组，服务器按照该视频帧组中视频帧对的第一视频帧在该第一视频中的出现时间，将该视频帧组中视频帧对的第一视频帧融合为一个候选视频片段。

在其中一种可能的实施方式中，服务器比较该视频帧组中任两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间，得到两个相邻的视频帧对的出现时间差。在该两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间之间的差值符合出现时间条件的情况下，服务器将该两个相邻的视频帧对添加至临时帧列表。在该两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间之间的差值不符合出现时间条件的情况下，服务器将该临时帧列表中的视频帧对融合为一个参考视频片段。服务器基于多个参考视频片段，确定该至少一个候选视频片段。

其中，临时帧列表用于存储出现时间之间的差值符合出现时间条件的视频帧对。在一些实施例中，出现时间之间的差值符合出现时间条件是指，出现时间之间的差值小于或等于出现时间差值阈值，出现时间差值阈值由技术人员根据实际情况进行设置，比如设置为 8s，本申请实施例对此不做限定。

为了对上述实施方式进行更加清楚的说明，下面将分为四个部分对上述实施方式进行说明。

第一部分、服务器比较该视频帧组中任两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间。

在一些实施例中，服务器以第一视频帧在第一视频中的出现时间作为第一视频帧的标识，以第二视频帧在第二视频中的出现时间作为第二视频帧的标识在这种情况下，服务器比较任两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间时，比较这两个第一视频帧的标识即可。比如，该视频帧组包括视频帧对[2, 4], [3, 5], [4, 6], [6, 8], [7, 9]，服务器依次比较视频帧对的第一视频帧在第一视频中的出现时间。在第一次比较过程中，服务器比较第一个视频帧对[2, 4]的第一视频帧2与第二个视频帧对[3, 5]的第一视频帧3在第一视频中的出现时间。

第二部分、在该两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间之间的差值符合出现时间条件的情况下，服务器将该两个相邻的视频帧对添加至临时帧列表。

在其中一种可能的实施方式中，在该两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间之间的差值小于或等于出现时间差值阈值的情况下，服务器将该两个相邻的视频帧对添加至临时帧列表。比如，还是以该视频帧组视频帧对[2, 4], [3, 5], [4, 6], [6, 8], [7, 9]为例，对于视频帧对[2, 4]和[3, 5]来说，在出现时间差值阈值为3的情况下，由于[2, 4]和[3, 5]中的第一视频帧在第一视频中的出现时间差值为 $3-2=1$ ，因此服务器将该两个视频帧对添加至临时帧列表(Tmplist)， $Tmplist=[[2, 4], [3, 5]]$ 。

服务器确定将视频帧对添加至临时帧列表包括多个迭代过程，在任一迭代过程中，服务器比较当前视频帧对的第一视频帧与上一个视频帧对的第一视频帧在第一视频中的出现时间差值，这里当前视频帧对是指当前正在处理的视频帧对，上一个视频帧对是指上一次迭代过程中处理的视频帧对。比如，服务器在将视频帧对[2, 4]和[3, 5]添加至临时帧列表之后，进一步确定视频帧对[3, 5]和[4, 6]的第一视频帧在第一视频中的出现时间差值与出现时间差值阈值之间的关系，由于[3, 5]和[4, 6]中的第一视频帧在第一视频中的出现时间差值为 $4-3=1$ ，因此服务器将视频帧对[4, 6]添加至临时帧列表(Tmplist)， $Tmplist=[[2, 4], [3, 5], [4, 6]]$ 。通过多个迭代过程，得到临时帧列表 $Tmplist=[[2, 4], [3, 5], [4, 6], [6, 8], [7, 9]]$ 。

第三部分、在该两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间之间的差值不符合出现时间条件的情况下，服务器将该临时帧列表中的视频帧对融合为参考视频片段。

其中，参考视频片段包括第一子片段和第二子片段，第一子片段是由视频帧对中的第一视频帧构成的，第二子片段是由视频帧对中的第二视频帧构成的。

在其中一种可能的实施方式中，在该两个相邻的视频帧对的第一视频帧在该第一视频中的出现时间之间的差值大于出现时间差值阈值的情况下，服务器将该临时帧列表中的第

一视频帧融合为第一子片段，将该临时帧列表中的第二视频帧融合为第二子片段，该第一子片段和该第二子片段构成该参考视频片段。由于视频帧对中的第一视频帧和第二视频帧是相似度较高的视频帧，那么第一子片段和第二子片段也即是相似度较高的片段。比如，参见图 5，示出了第一子片段 501 和第二子片段 502 的形式，第一子片段 501 开头的第一个视频帧和第二子片段 502 开头的第一个视频帧构成一个视频帧对，第一子片段 501 结尾的第一个视频帧和第二子片段 502 结尾的第一个视频帧构成另一个视频帧对。在一些实施例中，一个参考视频片段中的第一子片段和第二子片段也被称为匹配段。

比如，该两个相邻的视频帧对为[9, 11]和[2, 4]，该两个视频帧对的第一视频帧在该第一视频中的出现时间之间的差值为 $9-2=7$ ，那么服务器将临时帧列表中的第一视频帧融合为一个参考视频片段。比如，临时帧列表 $Tmplist = [[2, 4], [3, 5], [4, 6], [6, 8], [7, 9]]$ ，那么服务器将该临时帧列表中的第一视频帧[2,], [3,], [4,], [6,], [7,]融合为第一子片段(2, 7)，将该临时帧列表中的第二视频帧[, 4], [, 5], [, 6], [, 8], [, 9]融合为第二子片段(4, 9)，该第一子片段(2, 7)和该第二子片段(4, 9)构成该参考视频片段(2, 7, 4, 9)，其中，该参考视频片段的格式为(src-startTime, src-endTime, ref-startTime, ref-endTime)，其中，src-startTime 是指第一子片段的开头，也即是临时帧列表中序号最小的第一视频帧，src-endTime 是指第一子片段的结尾，也即是临时帧列表中序号最大的第一视频帧，ref-startTime 是指第二子片段的开头，也即是临时帧列表中序号最小的第二视频帧，ref-endTime 是指第二子片段的结尾，也即是临时帧列表中序号最大的第二视频帧，序号是指视频帧的标识，表示视频帧在视频中的位置，序号越小，表示视频帧在视频中的位置越靠前，序号越大，表示视频帧在视频中的位置越靠后。在一些实施例中，服务器将参考视频片段存储在匹配段列表 match-duration-list 中。由于确定视频帧对时遍历了第一视频和第二视频的所有视频帧，可能会出现某一视频帧与多个视频帧相似的情况，从而会出现 match-duration-list 中存在两个参考视频片段的时间有交叠。

在一些实施例中，对视频帧组中的视频帧对进行遍历，确定当前遍历至的当前视频帧对和上一个遍历至的上一个视频帧对；当前视频帧对和上一个视频帧对为视频帧组中的两个相邻的视频帧对；比较当前视频帧对和上一个视频帧对的第一视频帧在第一视频中的出现时间，得到第一视频帧的出现时间的差值；在第一视频帧的出现时间的差值符合出现时间条件的情况下，将当前视频帧对和上一个视频帧对添加至临时帧列表；在第一视频帧的出现时间的差值不符合出现时间条件的情况下，将临时帧列表中的视频帧对融合为参考视频片段，并在融合后清空临时帧列表；确定下一个遍历至的视频帧对，将下一个遍历至的视频帧对作为新的当前视频帧对，返回比较当前视频帧对和上一个视频帧对的第一视频帧在第一视频中的出现时间的步骤继续执行，直至最后一个遍历至的视频帧对，并在临时帧列表中具有视频帧对时，将临时列表中的视频帧对融合为参考视频片段，基于多个参考视频片段，确定至少一个候选视频片段。其中，第一视频帧的出现时间的差值指的是，视

视频帧组中的两个相邻视频帧对的第一视频帧在第一视频中的出现时间的差值。在一些实施例中，该参考视频片段还能够携带第一子片段对应的出现时间差值、第一子片段的时长以及第一子片段包括的视频帧的数量等信息，以便于服务器调用。

另外，除了上述第三部分提供的方式之外，本申请实施例提供了另一种触发将临时帧列表中的视频帧对融合为参考视频片段的方法。

在其中一种可能的实施方式中，在当前处理的视频帧对为该视频帧组中最后一个视频帧对的情况下，服务器将视频帧对添加至临时帧列表，将该临时帧列表中的视频帧对融合为参考视频片段。比如，该视频帧组包括[2, 4], [3, 5], [4, 6], [6, 8], [7, 9]五个视频帧对，在服务器处理视频帧对[7, 9]时，由于该视频帧对[7, 9]是该视频帧组中的最后一个视频帧对，服务器将该视频帧对[7, 9]加入临时帧列表，将该临时帧列表中的视频帧对融合为参考视频片段，融合过程参见上一种实施方式的描述，在此不再赘述。

由于出现时间差值较小的视频帧可能会构成一个相对完整的视频片段，因此，将出现时间差值较小的视频帧进行融合，可得到相对完整的参考视频，相比于通过零碎的视频帧来确定目标视频片段，本申请实施例可便于后续基于相对完整的参考视频来确定一个更为完整的目标视频片段。

第四部分、服务器基于多个参考视频片段，确定该至少一个候选视频片段。

其中，该多个参考视频片段包括第一重合视频片段和/或第二重合视频片段，该第一重合视频片段是指属于该多个参考视频片段中第一参考视频片段的参考视频片段，该第二重合视频片段是指与该多个参考视频片段中第二参考视频片段部分重合的参考视频片段。

其中，第一重合视频片段属于该第一参考视频片段是指，第一重合视频片段的内容被该第一参考视频片段完全包含，或者或第一参考视频片段完全包含了该第一重合视频片段。

为了对上述第四部分的内容进行更加清楚的说明，下面对服务器从多个参考视频片段中确定第一重合视频片段的方法进行说明。

在其中一种可能的实施方式中，服务器基于该多个参考视频片段中第一子片段在第一视频中的出现时间，从该多个参考视频片段中确定第一重合视频片段。

其中，第一子片段也即是第一视频帧构成的视频片段，出现时间包括第一子片段在第一视频中的开始时间和结束时间。

举例来说，对于该多个参考视频片段中的参考视频片段 A_i 和参考视频片段 B_i ，服务器比较该参考视频片段 A_i 的第一子片段在第一视频中的出现时间以及该参考视频片段 B_i 的第一子片段在第一视频中的出现时间，在该参考视频片段 B_i 的第一子片段在第一视频中的出现时间是该参考视频片段 A_i 的第一子片段在第一视频中的出现时间的子集的情况下，确定该参考视频片段 B_i 为第一重合视频片段。比如，参见图6，该多个参考视频片段包括参考视频片段 A_i 和参考视频片段 B_i ，服务器比较该参考视频片段 A_i 的第一子片段 m_i 在该第一视频中的出现时间以及该参考视频片段 B_i 的第一子片段 n_i 在该第一视频中的出现时间。在该第一子片段 n_i 的开始时间在第一子片段 m_i 之后，且该第一子片段 n_i 的结束时间在第

一子片段 m_1 之前的情况下, 服务器将该参考视频片段 B_1 确定为第一重合视频片段, 该参考视频片段 A_1 也即是上述第一参考视频片段。

对服务器从多个参考视频片段中确定第一重合视频片段的方法进行说明之后, 下面对服务器从多个参考视频片段中确定第二重合视频片段的方法进行说明。

在其中一种可能的实施方式中, 服务器基于该多个参考视频片段中第一子片段在第一视频中的出现时间, 从该多个参考视频片段中确定第二重合视频片段。

举例来说, 对于该多个参考视频片段中的参考视频片段 A_2 和参考视频片段 B_2 , 服务器比较该参考视频片段 A_2 的第一子片段在第一视频中的出现时间以及该参考视频片段 B_2 的第一子片段在第一视频中的出现时间, 在该参考视频片段 B 的第一子片段在第一视频中的出现时间与该参考视频片段 A 的第一子片段在第一视频中的出现时间的存在交集的情况下, 将参考视频片段 A 和参考视频片段 B 中时长较短的参考视频片段确定为第二重合视频片段。比如, 参见图 6, 该多个参考视频片段包括参考视频片段 A_2 和参考视频片段 B_2 , 服务器比较该参考视频片段 A_2 的第一子片段 m_2 在该第一视频中的出现时间以及该参考视频片段 B_2 的第一子片段 n_2 在该第一视频中的出现时间。在该第一子片段 n_2 的开始时间在第一子片段 m_2 的开始时间之后, 结束时间之前, 且该第一子片段 n_2 的结束时间在第一子片段 m_2 之后, 或者该第一子片段 n_2 的开始时间在第一子片段 m_2 之前, 且该第一子片段 n_2 的结束时间在第一子片段 m_2 的结束时间之前, 开始时间之后, 参考视频片段 B_2 的时长小于参考视频片段 A_2 的情况下, 服务器将该参考视频片段 B_2 确定为第二重合视频片段, 该参考视频片段 A_2 也即是上述第二参考视频片段。

在介绍完服务器确定第一重合视频片段和第二重合视频片段的方法之后, 下面对上述第四部分提供的步骤进行说明。

在其中一种可能的实施方式中, 在该多个参考视频片段包括该第一重合视频片段的情况下, 服务器将该第一重合视频片段删除, 得到该至少一个候选视频片段。

在这种实施方式下, 服务器能够将重复的第一重合视频片段从多个参考视频片段中删除, 以减少得到的候选视频片段的数量, 减少运算量, 提高运算效率。

在其中一种可能的实施方式中, 在该多个参考视频片段包括该第二重合视频片段的情况下, 服务器将该第二重合视频片段与该第二参考片段之间的重合部分删除, 得到该至少一个候选视频片段。

在这种实施方式下, 服务器能够将第二重合视频片段与第二参考片段之间的重合部分删除, 以减少得到的候选视频片段的长度, 减少运算量, 提高运算效率。

在上述实施方式的基础上, 可选地, 服务器还能够执行下述步骤:

在一些实施例, 将该第二重合视频片段与该第二参考片段之间的重合部分删除之后, 服务器比较第三类参考视频片段的时长与目标时长, 该第三类参考视频片段是指删除重合部分的该第二重合视频片段。在该第三类参考视频片段的时长大于或等于该目标时长的情况下, 服务器保留该第三类参考视频片段。在该第三类参考视频片段的时长小于该目标时

长的情况下，服务器删除该第三类参考视频片段。

其中，目标时长由技术人员根据实际情况进行设置，本申请实施例对此不做限定。在服务器保留该第三类参考视频片段的情况下，也即是采用该第三类参考视频片段替换了原本的第二重合视频片段。

下面通过两个例子对上述实施方式进行说明：

例 1、对于该多个参考视频片段中的参考视频片段 A_2 和参考视频片段 B_2 ，在该参考视频片段 A_2 的第一子片段 m_2 和该参考视频片段 B_2 的第一子片段 n_2 具有部分重合，且第一子片段 m_2 的开始时间早于第一子片段 n_2 的情况下，服务器将第一子片段 n_2 的开始时间移动到第一子片段 m_2 的结束时间，得到子片段 l_1 ，该子片段 l_1 为第三类参考视频片段的第一子片段。在该子片段 l_1 的时长小于或等于该目标时长的情况下，服务器删除该子片段 l_1 ，同时删除该子片段 l_1 所属的第三类参考视频片段。在该子片段 l_1 的时长大于该目标时长的情况下，服务器保留该子片段 l_1 ，同时保留该子片段 l_1 所属的第三类参考视频片段。

例 2、对于该多个参考视频片段中的参考视频片段 A_2 和参考视频片段 B_2 ，在该参考视频片段 A_2 的第一子片段 m_2 和该参考视频片段 B_2 的第一子片段 n_2 具有部分重合，且第一子片段 n_2 的开始时间早于第一子片段 m_2 的情况下，服务器将第一子片段 n_2 的结束时间移动到第一子片段 m_2 的开始时间，得到子片段 l_2 ，该子片段 l_2 为第三类参考视频片段的第一子片段。在该子片段 l_2 的时长小于或等于该目标时长的情况下，服务器删除该子片段 l_2 ，同时删除该子片段 l_2 所属的第三类参考视频片段。在该子片段 l_2 的时长大于该目标时长的情况下，服务器保留该子片段 l_2 ，同时保留该子片段 l_2 所属的第三类参考视频片段。

由于在第三类参考视频片段的时长小于目标时长的情况下，可认为该第三类参考视频片段包括的视频帧数量较少，可能为误生成的参考视频片段，通过将该参考视频片段删除，可提升后续基于剩余的参考视频片段所生成的目标视频片段的准确性。

306、服务器基于该至少一个候选视频片段，确定该至少一个目标候选视频片段，该目标候选视频片段在该至少一个候选视频片段中的出现次数符合次数条件。

在其中一种可能的实施方式中，服务器基于该至少一个候选视频片段，确定至少一个参考候选视频片段。服务器确定每个参考候选视频片段在该至少一个参考候选视频片段中的出现次数。服务器将出现次数符合该出现次数条件的参考候选视频片段确定为目标候选视频片段。

其中，参考候选视频片段在该至少一个参考候选视频片段的出现次数是指，该至少一个参考候选视频片段中该参考候选视频片段的数量。比如，该至少一个参考候选视频片段为 1, 2, 3, 1, 4, 5，那么对于参考候选视频片段 1 来说，出现次数就为 2。

为了对上述实施方式进行说明，下面将分为三个部分对上述实施方式进行说明。

第一部分、服务器基于该至少一个候选视频片段，确定至少一个参考候选视频片段。

其中，该至少一个候选视频片段包括第三重合视频片段和/或第四重合视频片段，该第三重合视频片段是指属于该至少一个候选视频片段中第一候选视频片段的候选视频片

段,该第四重合视频片段是指与该至少一个候选视频片段中第二候选视频片段部分重合的候选视频片段。

为了对上述第一部分的内容进行更加清楚的说明,下面对服务器从至少一个候选视频片段中确定第三重合视频片段的方法进行说明。

在其中一种可能的实施方式中,服务器基于该至少一个候选视频片段中第一子片段在第一视频中的出现时间,从该至少一个候选视频片段中确定第三重合视频片段。

其中,候选视频片段包括第一子片段和第二子片段,第一子片段是由视频帧对中的第一视频帧构成的,第二子片段是由视频帧对中的第二视频帧构成的。

举例来说,该至少一个候选视频片段为两个候选视频片段,对于该至少一个候选视频片段中的候选视频片段 C_1 和候选视频片段 D_1 ,服务器比较该候选视频片段 C_1 的第一子片段在第一视频中的出现时间以及该候选视频片段 D_1 的第一子片段在第一视频中的出现时间,在该候选视频片段 D_1 的第一子片段在第一视频中的出现时间是该候选视频片段 C_1 的第一子片段在第一视频中的出现时间的子集的情况下,确定该候选视频片段 D_1 为第三重合视频片段。

比如,该至少一个候选视频片段为两个候选视频片段,包括候选视频片段 C_1 和候选视频片段 D_1 ,服务器比较该候选视频片段 C_1 的第一子片段 o_1 在该第一视频中的出现时间以及该候选视频片段 D_1 的第一子片段 p_1 在该第一视频中的出现时间。在该第一子片段 p_1 的开始时间在第一子片段 o_1 之后,且该第一子片段 p_1 的结束时间在第一子片段 o_1 之前的情况下,服务器将该候选视频片段 D_1 确定为第三重合视频片段,该候选视频片段 C_1 也即是上述第一候选视频片段。

对服务器从至少一个候选视频片段中确定第三重合视频片段的方法进行说明之后,下面对服务器从至少一个候选视频片段中确定第四重合视频片段的方法进行说明。

在其中一种可能的实施方式中,服务器基于该至少一个候选视频片段中第一子片段在第一视频中的出现时间,从该至少一个候选视频片段中确定第四重合视频片段。

举例来说,该至少一个候选视频片段为两个候选视频片段,对于该至少一个候选视频片段中的候选视频片段 C_2 和候选视频片段 D_2 ,服务器比较该候选视频片段 C_2 的第一子片段在第一视频中的出现时间以及该候选视频片段 D_2 的第一子片段在第一视频中的出现时间,在该候选视频片段 D_2 的第一子片段在第一视频中的出现时间与该候选视频片段 C_2 的第一子片段在第一视频中的出现时间的存在交集的情况下,将候选视频片段 C_2 和候选视频片段 D_2 中时长较短的候选视频片段确定为第四重合视频片段。

比如,该至少一个候选视频片段为两个候选视频片段,包括候选视频片段 C_2 和候选视频片段 D_2 ,服务器比较该候选视频片段 C_2 的第一子片段 o_2 在该第一视频中的出现时间以及该候选视频片段 D_2 的第一子片段 p_2 在该第一视频中的出现时间。在该第一子片段 p_2 的开始时间在第一子片段 o_2 的开始时间之后,结束时间之前,且该第一子片段 p_2 的结束时间在第一子片段 o_2 之后,或者该第一子片段 p_2 的开始时间在第一子片段 o_2 之前,且该第一

子片段 p_2 的结束时间在第一子片段 o_2 的结束时间之前，开始时间之后，候选视频片段 D_2 的时长小于候选视频片段 C_2 的情况下，服务器将该候选视频片段 D_2 确定为第四重合视频片段，该候选视频片段 C_2 也即是上述第二候选视频片段。

在介绍完服务器确定第三重合视频片段和第四重合视频片段的方法之后，下面对上述第一部分提供的步骤进行说明。

在其中一种可能的实施方式中，在该至少一个候选视频片段包括该第三重合视频片段的情况下，服务器将该第三重合视频片段删除，得到该至少一个参考候选视频片段。在一些实施例中，在删除该第三重合视频片段之前，服务器将该第三重合视频片段的出现次数叠加到该第一候选视频片段上。由于第三重合视频片段被第一候选视频片段完全包含，那么将该第三重合视频片段的出现次数叠加到该第一候选视频片段上能够提高该第一候选视频片段在后续处理中的权重。

在这种实施方式下，服务器能够将重复的第三重合视频片段从至少一个候选视频片段中删除，以减少得到的参考候选视频片段的数量，减少运算量，提高运算效率。

下面通过一个具体的例子来进行说明。

在该候选视频片段 D_1 的第一子片段 o_1 是该候选视频片段 C_1 的第一子片段 p_1 的子集，且第一子片段 o_1 的时长 $> 0.5 * \text{第一子片段 } p_1$ 的情况下，则服务器删除第一子片段 o_1 ，同时删除该候选视频片段 D_1 ，将该候选视频片段 D_1 的出现次数叠加到该候选视频片段 C_1 上。

在上述实施方式的基础上，可选地，服务器将第三重合视频片段的出现次数叠加到该第一候选视频片段上之前，还能够确定该第三重合视频片段的时长和第一候选视频片段的时长，基于该第三重合视频片段的时长和第一候选视频片段的时长来确定是否将该第三重合视频片段的出现次数叠加到该第一候选视频片段上。

比如，服务器确定该第三重合视频片段的时长和第一候选视频片段的时长。服务器确定该第三重合视频片段的时长与第一候选视频片段的时长之间的第一比值，在该第一比值大于或等于比值阈值的情况下，服务器将该第三重合视频片段的出现次数叠加到该第一候选视频片段上；在该第一比值小于比值阈值的情况下，服务器不将该第三重合视频片段的出现次数叠加到该第一候选视频片段上，其中，比值阈值由技术人员根据实际情况进行设置，比如设置为 0.5，本申请实施例对此不做限定。

在其中一种可能的实施方式中，在该至少一个候选视频片段包括该第四重合视频片段，且该第四重合视频片段与该第二候选视频片段之间的重合度符合重合度条件的情况下，服务器确定该第四重合视频片段的出现次数。服务器基于重合度符合重合度条件的各第四重合视频片段的出现次数，确定该至少一个参考候选视频片段。

其中，重合度是指，重合的视频片段的时长与被比较的视频片段的时长之间的比值。比如，对于第四重合视频片段和第二候选视频片段，第二候选视频片段为被比较的视频片段，确定第四重合视频片段和第二候选视频片段之间的重合度时，将第四重合视频片段与第二候选视频片段之间重合的视频片段的时长与第二候选视频片段的时长相除即可得到。

重合度符合重合度条件是指，重合度大于或等于重合度阈值。

下面通过两种实施方式对上述实施方式中服务器基于该第四重合视频片段的出现次数，确定该至少一个参考候选视频片段的方法进行说明。

实施方式 1、在该第四重合视频片段的出现次数大于或等于第一出现次数阈值的情况下，服务器将该第四重合视频片段与第二候选视频片段进行融合，得到该至少一个参考候选视频片段。在一些实施例中，分别将每个重合度符合重合度条件的第四重合视频片段与相应的第二候选视频片段进行融合，得到至少一个参考候选视频片段。在一些实施例中，在该第四重合视频片段与第二候选视频片段进行融合之前，服务器将该第四重合视频片段的出现次数叠加到该第二候选视频片段上。

其中，第一出现次数阈值由技术人员根据实际情况进行设置，比如设置为 3，本申请实施例对此不做限定。出现次数大于或等于第一出现次数阈值表示该第四重合视频片段不可忽略，需要进行进一步处理以提高获取目标视频片段的准确性。

下面对上述实施方式中服务器将该第四重合视频片段与第二候选视频片段进行融合的方法进行说明。

在一些实施例中，以第四重合视频片段的时长小于该第二候选视频片段为例，服务器将从该第四重合视频片段中将与该第二候选视频片段之间的重复部分删除，将剩余部分添加到该第二候选视频片段上，得到一个候选视频片段。比如，参见图 7，第四重合视频片段 701 的时长小于该第二候选视频片段 702，第四重合视频片段 704 的时长小于该第二候选视频片段 705。在该第四重合视频片段 701 的结束时间晚于该第二候选视频片段 702 的情况下，服务器将该第四重合视频片段 701 与该第二候选视频片段 702 融合，得到一个候选视频片段 703。在该第四重合视频片段 704 的开始时间早于该第二候选视频片段 705 的情况下，服务器将该第四重合视频片段 704 与该第二候选视频片段 705 融合，得到一个候选视频片段 706。

通过将第四重合视频片段与第二候选视频片段进行融合，可减少视频片段的数量，从而减少运算量，提高运算效率。

实施方式 2、在该第四重合视频片段的出现次数小于该第一出现次数阈值的情况下，服务器将该第四重合视频片段删除，得到该至少一个参考候选视频片段。服务器将该第四重合视频片段的出现次数叠加到该第二候选视频片段上。

其中，出现次数小于第一出现次数阈值表示该第四重合视频片段可以忽略，服务器将该第四重合视频片段删除即可。

通过将部分第四重合视频片段删除，可减少视频片段的数量，从而减少运算量，提高运算效率。

在其中一种可能的实施方式中，在该至少一个候选视频片段包括该第四重合视频片段，且该第四重合视频片段与该第二候选视频片段之间的重合度不符合该重合度条件的情况下，服务器将该第四重合视频片段删除，得到该至少一个参考候选视频片段。在一些实施

例中,在将该第四重合视频片段删除之前,服务器将该第四重合视频片段的出现次数叠加到该第二候选视频片段上。

在其中一种可能的实施方式中,在该至少一个候选视频片段包括该第四重合视频片段,且该第四重合视频片段的时长小于该第二候选视频片段的情况下,服务器将该第四重合视频片段删除,得到该至少一个参考候选视频片段。在一些实施例中,在将该第四重合视频片段删除之前,服务器将该第四重合视频片段的出现次数叠加到该第二候选视频片段上。

在一些实施例中,至少一个参考候选视频片段被服务器存储在匹配列表(match-list)中以便调用。

通过将重合度不符合重合度条件,或者时长小于第二候选视频片段时长的第四重合视频片段进行删除,可减少视频片段的数量,从而减少后续的运算量,提高运算效率。

第二部分、服务器确定参考候选视频片段在该至少一个参考候选视频片段的出现次数。

通过上述第一部分的处理过程,服务器基于至少一个候选视频片段,确定至少一个参考候选视频片段,确定过程中涉及出现次数的合并和删除,服务器重新确定该至少一个参考候选视频片段的出现次数。在一些实施例中,服务器能够将该至少一个参考候选视频片段的出现次数存储在出现次数列表(count-list)中以便调用。

比如,在确定第一视频中的目标视频片段时,服务器采用三个第二视频来进行挖掘,为了方便进行说明,将该第一视频命名为*i*,将该三个第二视频分别命名为*vid1*、*vid2*以及*vid3*。采用上述各个步骤之后,服务器基于该第一视频*i*和第二视频*vid1*确定了两个候选视频片段[(2, 7, 4, 9), (10, 11, 11, 12)],基于该第一视频*i*和第二视频*vid2*确定了一个候选视频片段[(2, 7, 4, 9)],基于该第一视频*i*和第二视频*vid3*确定了一个候选视频片段[(2, 7, 4, 10)]。服务器对这四个候选视频片段进行统计,确定候选视频片段(2, 7, 4, 9)的出现次数为2次,(2, 7, 4, 10)的出现次数为1次,(10, 11, 11, 12)的出现次数为1次。通过上述第一部分的方式融合这四个候选视频片段之后,得到两个参考候选视频片段[(2, 7, 4, 9), (10, 11, 11, 12)],且参考候选视频片段(2, 7, 4, 9)的出现次数为3,参考候选视频片段(10, 11, 11, 12)的出现次数为1,以次数列表(count-list)来进行存储时,即count-list=[3, 1]。

第三部分、服务器将出现次数符合该出现次数条件的参考候选视频片段确定为目标候选视频片段。

在一种可能的实施方式中,服务器将出现次数大于或等于第二出现次数阈值的参考候选视频片段确定为目标候选视频片段。

其中,第二出现次数阈值与该至少一个参考候选视频片段的数量正相关,也即是该至少一个参考候选视频片段的数量越多,该第二出现次数阈值也就越大;该至少一个参考候选视频片段的数量越少,该第二出现次数阈值也就越小。在一些实施例中,该第二出现次数阈值为目标比值与该至少一个参考候选视频片段的数量的乘积,该目标比值为小于1的正数。

比如，在得到的两个参考候选视频片段 $[(2, 7, 4, 9), (10, 11, 11, 12)]$ ，且参考候选视频片段 $(2, 7, 4, 9)$ 的出现次数为3，参考候选视频片段 $(10, 11, 11, 12)$ 的出现次数为1，第二出现次数阈值为3的情况下，服务器将参考候选视频片段 $(10, 11, 11, 12)$ 删除，最终保留参考候选视频片段 $(2, 7, 4, 9)$ ，以及出现次数3。以匹配列表(match-list)和次数列表(count-list)来进行存储时，即 $\text{match-list} = (2, 7, 4, 9)$ ； $\text{count-list} = [3]$ 。

307、对于任一目标候选视频片段，在目标候选视频片段在该第一视频中的出现时间处于目标时间范围的情况下，服务器将该目标候选视频片段确定为该第一视频中的目标视频片段。

其中，该目标时间范围由技术人员根据实际情况进行设置，比如，在本申请实施例提供的技术方案应用在识别视频的片头和片尾的场景下时，该目标时间范围为视频的片头和片尾可能存在的时间范围，在这种情况下，目标时间范围包括第一时间范围和第二时间范围，第一时间范围为片头可能存在的范围，第二时间范围为片尾可能存在的范围。比如，将视频前 $1/5$ 时间设置为片头时间，也即是第一时间范围，后 $1/5$ 时间为片尾时间，也即是第二时间范围，对于10分钟的视频，设定片头仅可能出现在前2分钟，片尾在后2分钟。其中， $1/5$ 是技术人员根据实际情况设置的，针对不同类型的视频可以进行相应的调整，比如，针对15分钟左右的少儿动漫可取 $1/5$ ，针对电视剧45分钟长视频可取 $1/8$ 。

需要说明的是，上述步骤301-307是以服务器确定第一视频中的目标视频片段为例进行说明的，在该第一视频和该至少一个第二视频属于同一个视频集合的情况下，服务器能够采用与上述步骤301-307同理的方法来确定该视频集合中其他视频的目标视频片段，其他视频是指除第一视频以外的视频。

下面结合图8对本申请实施例提供的技术方案进行说明。

参见图8，在本申请实施例中，服务器基于视频帧特征之间的相似度进行匹配，得到多个视频帧对。服务器基于出现时间差值将该多个视频帧对划分为多个初始视频帧组。服务器基于出现时间差值将该多个初始视频帧组融合为多个候选视频帧组。服务器将该多个候选视频帧组融合为多个视频帧组。服务器基于该多个视频帧组，输出第一视频的目标视频片段。

在一些实施例中，上述步骤301-307可以由一个片段挖掘系统来实现，在本申请实施例提供的技术方案应用在识别视频片头片尾的场景下时，该片段挖掘系统也即是片头片尾挖掘系统。参见图9该视频片段挖掘系统提供了如下功能。提取多个视频的视频帧特征。对于每个视频，将该视频与该多个视频中的其他视频组成视频对。基于多个视频对来进行匹配，得到多个视频帧对。将多个视频帧对进行融合，得到多个视频帧组。基于该多个视频帧组，确定目标视频片段在该视频中的位置。基于该目标视频片段在该视频中的位置，获取该目标视频片段。在本申请实施例提供的技术方案应用在识别视频片头片尾的场景下时，该目标视频片段也即是该视频的片头或者片尾。

参见图 10 本申请实施例提供的技术方案应用在识别电视剧片头片尾的场景下时，获取电视剧，该电视剧包括多个视频。将该多个视频输入片段挖掘系统，通过该片段挖掘系统输出该多个视频的片头和片尾。在一些实施例中，该片段挖掘系统能够输出该多个视频的片头和片尾的时间戳。

308、服务器将第一视频中的目标视频片段存储在片段数据库中。

在一种可能的实施方式中，服务器对该第一视频的目标视频片段进行特征提取，得到该目标视频片段的视频帧特征。服务器将该目标视频片段的视频帧特征存储在该片段数据库中。在一些实施例中，服务器将该目标视频片段的视频帧特征关联到该第一视频，比如，服务器将该目标视频片段的视频帧特征的标识设置为第一视频的标识。在该第一视频属于某个视频集合的情况下，服务器将该第一视频的标识关联到该视频集合的标识，以便于后续的查询过程。

其中，对目标视频片段进行特征提取得到该目标视频片段的视频帧特征与上述步骤 301 属于同一发明构思，实现过程参见上述步骤 301 的描述，在此不再赘述。

比如，目标视频片段为 (2, 7)，服务器从该第一视频中获取 2-7 秒对应的目标视频片段，从该目标视频片段中抽取多个参考视频帧。服务器对该多个参考视频帧进行特征提取，得到该多个参考视频帧的视频帧特征。服务器将该多个参考视频帧的视频帧特征存储在片段数据库中。服务器将该多个参考视频帧的视频帧特征与第一视频的标识 Vid1 相关联，将第一视频的标识 Vid1 与该第一视频所属的视频集合的标识 Cid1 相关联。图 11 示出了一种片段数据库的存储形式，参见图 11，在数据库 1100 中，em1-emN 是视频帧特征，vid1-vidK 是不同视频的标识，N 和 K 均为正整数。

服务器将第一视频中的目标视频片段存储在片段数据库中之后，还能够利用该片段数据库进行视频片段检索，方法如下：

在其中一种可能的实施方式中，服务器对待识别的目标视频的多个目标视频帧进行特征提取，得到该多个目标视频帧的视频帧特征。服务器基于该多个目标视频帧的视频帧特征、该第一视频帧的视频帧特征以及该至少一个第二视频的视频帧特征，确定该目标视频的至少一个目标视频片段。

其中，服务器对目标视频的多个目标视频帧进行特征提取，得到该多个目标视频帧的视频帧特征的过程与上述步骤 301 属于同一发明构思，实现过程参见上述步骤 301 的描述，在此不再赘述。服务器基于该多个目标视频帧的视频帧特征、该第一视频帧的视频帧特征以及该至少一个第二视频的视频帧特征，确定该目标视频的至少一个目标视频片段的过程，与上述步骤 302-307 属于同一发明构思，实现过程参见上述步骤 302-307 的描述，在此不再赘述。在一些实施例中，在该片段数据库进行视频片段检索是由视频检索系统实现的。在一些实施例中，该第一视频帧的视频帧特征以及该至少一个第二视频的视频帧特征存储在片段数据库中。

通过时间段匹配算法设计，实现基于图像 embedding 特征的相似视频段匹配方法，可

支持变长（体现在匹配逻辑中，相同出现时间差值下合并匹配帧为时间段时并不要求合并的帧必须前后连续）、位置变化（体现在匹配逻辑中，当出现时间差值为 0，则位置无变化，出现时间差值大于 0 则位置可以有变化）的相似视频段匹配。该方法耗时小，性能优。

基于视频时间段匹配的方法产生的片头片尾挖掘方案可以实现复杂变长、位置变化等复杂视频片头片尾识别定位，解决现有方案中无法解决的困难情况。

借助基于时间段匹配的片头片尾检索方案联合，可以实现实时（10 分钟内）的片头片尾挖掘方案，对应用友好。

上述视频片段识别的方法能够应用在识别视频片段片头片尾的场景下，还能够应用在识别侵权视频的场景下，下面将分别对这两种应用场景进行介绍。

在该视频片段的检索方法应用在检索视频片段片头片尾的场景下，将待识别的目标视频输入该视频检索系统，由该视频检索系统对该目标视频进行特征提取，得到该多个目标视频帧的视频帧特征。通过该视频检索系统，基于该多个目标视频帧的视频帧特征在片段数据库中进行匹配，得到该目标视频中的目标视频片段，该目标视频片段也即是该目标视频的片头或者片尾。

以识别电视剧中新更新视频的片头和片尾为例，比如，该电视剧已经更新了 10 集，通过上述步骤 301-307 获取了这 10 集的片头和片尾，通过上述步骤 308 将这 10 集的片头和片尾存储在片段数据库中。在该电视剧更新第 11 集时，将该第 11 集作为该目标视频，将该目标视频输入该视频检索系统，由该视频检索系统对该目标视频进行特征提取，得到该多个目标视频帧的视频帧特征。通过该视频检索系统，基于该多个目标视频帧的视频帧特征在片段数据库中进行匹配，得到该目标视频中的目标视频片段，该目标视频片段也即是该目标视频的片头或者片尾。在该片段数据库中将视频帧特征与视频的标识和视频集合的标识关联的情况下，能够基于视频集合的标识在有限的范围内进行匹配，从而提高确定目标视频片段的效率，其中，该视频集合也即是该电视剧。

下面将结合图 12 进一步说明。

确定待进行片头片尾识别的电视剧。获取该电视剧中的多个视频。将该多个视频输入片段挖掘系统 1201，由该片段挖掘系统 1201 输出该多个视频的片头和片尾。将该多个视频的片头和片尾存储在片段数据库 1202 中。在该电视剧更新了目标视频的情况下，将该目标视频输入视频检索系统 1203，由视频检索系统 1203 采用该目标视频在该片段数据库 1202 中进行检索，得到该目标视频的片头和片尾。本申请实施例提供的技术方案对同一视频集合中的视频挖掘片头片尾采用视频相同时间段检索的方法，即对同一视频集合，通过检索以及时序定位找到相同的视频片段，作为挖掘到的片头片尾。交叉排重，是指视频集合内部的视频经过相互检索找到重复的视频片段。视频排重检索的目的是对第一视频，检索出其与库存视频相同的视频片段。

需要说明的是，一个视频可能有多个片头或片尾满足上列要求，这属于正常情况，对

于片头曲+本集花絮+同一广告植入+正片类型的电视剧，片头曲、广告植入是多个视频中可匹配到的，但花絮由于每集都不同，故不会被匹配到，所以会出现 2 个片头。

在该视频片段的检索方法应用在识别侵权视频的场景下，将待识别的目标视频输入该视频检索系统，由该视频检索系统对该目标视频进行特征提取，得到该多个目标视频帧的视频帧特征，其中，该目标视频也即是待进行侵权识别的视频。通过该视频检索系统，基于该多个目标视频帧的视频帧特征在片段数据库中进行匹配，得到该目标视频中的目标视频片段，该目标视频片段也即是该目标视频的片头或者片尾。将该目标视频片段从该目标视频中删除，基于删除目标视频片段后的目标视频来进行侵权识别，侵权识别的目的是确定删除目标视频片段后的目标视频与指定视频的内容是否相同。其中，侵权识别由侵权识别系统来实现，侵权识别系统能够对查询视频在侵权保护视频数据库中进行排重，如果发现重复，则表示侵权。然而由于仅需要保护正片内容，常规影视剧的片头片尾不在侵权排重范围内，采用本申请实施例提供的技术方案能够实现对影视剧进行片头片尾识别。

下面将结合图 13 进一步说明。

确定待进行侵权识别的电视剧。获取该电视剧中的多个视频，将该多个视频存储在侵权保护视频数据库 1301 中。将该多个视频输入片段挖掘系统 1302，由该片段挖掘系统 1302 输出该多个视频的片头和片尾。将该多个视频的片头和片尾存储在片段数据库 1303 中。在需要对目标视频进行侵权识别的情况下，将该目标视频输入视频检索系统 1304，由视频检索系统 1304 采用该目标视频在该片段数据库 1303 中进行检索，得到该目标视频的片头和片尾。将该目标视频的片头和片尾删除，通过侵权识别系统 1305 来输出该目标视频的侵权结果，侵权结果包括侵权和不侵权。

在一些实施例中，基于上述方式基于目标视频在片段数据库中进行查询之后，在得到该目标视频的多个目标视频片段的情况下，服务器将该多个目标视频片段中最长的目标视频片段确定为最终的目标视频片段，在本申请实施例提供的技术方案应用在识别视频的片头和片尾的情况下，该目标视频片段也即是该目标视频的片头和片尾，该过程参见图 14。

另外，视频检索系统与片段挖掘系统可以同时提供对外接口，即检索入库、挖掘入库以同时开放由用户指定要使用的具体功能。也可以仅提供一个识别接口，后台根据库存中是否已经有该视频标识对应电视剧的片头片尾进行检索还是挖掘的判断，由后台触发要使用的具体功能，该具体功能包括检索和挖掘。

上述所有可选技术方案，可以采用任意结合形成本申请的可选实施例，在此不再一一赘述。

通过本申请实施例提供的技术方案，基于视频帧特征之间的相似度，确定包含相似视频帧的视频帧对。基于出现时间差值来对视频帧对中的第一视频帧进行融合，得到至少一个候选视频片段。最终从至少一个候选视频片段中确定出处于目标时间范围的目标视频片段。确定目标片段的过程无需人工参与，由计算机设备直接基于第一视频和至少一个第二视频就能够自动进行，效率较高。

通过上述视频段匹配算法设计,实现基于视频帧特征的相似视频片段匹配方法,可支持变长(体现在匹配逻辑中,相同出现时间差值下合并视频帧对时并不要求合并的帧必须前后连续)、位置变化(体现在匹配逻辑中,当出现时间差值为0,则位置无变化,出现时间差值大于0则位置可以有变化)的相似视频段匹配。该方法耗时小,性能优。

图15是本申请实施例提供的一种视频片段的识别装置的结构示意图,参见图15,装置包括:视频帧对确定模块1501、融合模块1502以及目标视频片段确定模块1503。

视频帧对确定模块1501,用于基于第一视频的视频帧特征以及至少一个第二视频的视频帧特征,确定多个视频帧对,该视频帧对包括相似度符合相似度条件的第一视频帧和第二视频帧,该第一视频帧属于该第一视频,该第二视频帧属于该至少一个第二视频。

融合模块1502,用于基于该多个视频帧对的出现时间差值,将该多个视频帧对中的第一视频帧进行融合,得到该第一视频中的至少一个候选视频片段,该出现时间差值是指该视频帧对中的两个视频帧在视频中的出现时间之间的差值。

目标视频片段确定模块1503,用于基于该至少一个候选视频片段以及目标时间范围,确定该第一视频中的至少一个目标视频片段,该目标视频片段处于该第一视频的该目标时间范围内。

在其中一种可能的实施方式中,该融合模块1502,用于基于该多个视频帧对的出现时间差值,将该多个视频帧对划分为多个视频帧组,同一个该视频帧组中的视频帧对对应于同一个出现时间差值。对于该多个视频帧组中的任一视频帧组,按照该视频帧组中视频帧对的第一视频帧在该第一视频中的出现时间,将该视频帧组中视频帧对的第一视频帧融合为一个该候选视频片段。

在其中一种可能的实施方式中,该融合模块1502,用于针对多个视频帧对中的任一视频帧对,将视频帧对中的第一视频帧的第一出现时间减去视频帧对中的第二视频帧的第二出现时间,得到视频帧对的出现时间差值;第一出现时间是指第一视频帧在第一视频中出现的时间,第二出现时间是指第二视频帧在第二视频中出现的时间;将出现时间差值相同的视频帧对划分为一个初始视频帧组,并将初始视频帧组中的视频帧对的出现时间差值,作为初始视频帧组所对应的出现时间差值。基于多个初始视频帧组对应的出现时间差值,将该多个初始视频帧组进行融合,得到该多个视频帧组。

在其中一种可能的实施方式中,该融合模块1502,用于按照目标顺序对该多个初始视频帧组进行排序,得到多个候选视频帧组。对于该多个候选视频帧组中任两个相邻的候选视频帧组,在两个相邻的候选视频帧组之间的匹配时间差值符合匹配时间差值条件的情况下,将该两个相邻的候选视频帧组融合为一个视频帧组,该匹配时间差值是指该两个相邻的候选视频帧组对应的出现时间差值之间的差值。

在其中一种可能的实施方式中,该两个相邻的候选视频帧组包括第一候选视频帧组和第二候选视频帧组,该融合模块1502,用于在该第一候选视频帧组对应的出现时间差值与该第二候选视频帧组对应的出现时间差值之间的匹配时间差值小于或等于匹配差值阈值

的情况下，将该第一候选视频帧组中的视频帧对添加至该第二候选视频帧组，得到该视频帧组。

在其中一种可能的实施方式中，两个相邻的候选视频帧组包括第一候选视频帧组和第二候选视频帧组，该融合模块 1502，用于在第一候选视频帧组与第二候选视频帧组之间的匹配时间差值小于或等于匹配差值阈值的情况下，将该第一候选视频帧组中的视频帧对添加至该第二候选视频帧组。基于该第二候选视频帧组对应的出现时间差值，采用参考第二视频帧替换目标第二视频帧，得到该视频帧组，该目标第二视频帧为新添加至该第二候选视频帧组中的第二视频帧，该参考第二视频帧为该第二视频中与目标第一视频帧之间的出现时间差值为该第二候选视频帧组对应的出现时间差值的第二视频帧，该目标第一视频帧为该目标第二视频帧所属视频帧对中的第一视频帧。

在其中一种可能的实施方式中，该融合模块 1502，用于对视频帧组中的视频帧对进行遍历，确定当前遍历至的当前视频帧对和上一个遍历至的上一个视频帧对。当前视频帧对和上一个视频帧对为视频帧组中的两个相邻的视频帧对。比较当前视频帧对和上一个视频帧对的第一视频帧在第一视频中的出现时间，得到第一视频帧的出现时间的差值。在第一视频帧的出现时间的差值符合出现时间条件的情况下，将当前视频帧对和上一个视频帧对添加至临时帧列表；在第一视频帧的出现时间的差值不符合出现时间条件的情况下，将临时帧列表中的视频帧对融合为参考视频片段，并在融合后清空临时帧列表。确定下一个遍历至的视频帧对，将下一个遍历至的视频帧对作为新的当前视频帧对，返回比较当前视频帧对和上一个视频帧对的第一视频帧在第一视频中的出现时间的步骤继续执行，直至最后一个遍历至的视频帧对。基于多个参考视频片段，确定该至少一个候选视频片段。

在其中一种可能的实施方式中，该多个参考视频片段包括第一重合视频片段，该第一重合视频片段是指属于该多个参考视频片段中第一参考视频片段的参考视频片段，该融合模块 1502，用于在该多个参考视频片段包括该第一重合视频片段的情况下，将该第一重合视频片段删除，得到该至少一个候选视频片段。

在其中一种可能的实施方式中，该多个参考视频片段包括第二重合视频片段，该第二重合视频片段是指与该多个参考视频片段中第二参考视频片段部分重合的参考视频片段，该融合模块 1502，用于在该多个参考视频片段包括该第二重合视频片段的情况下，将该第二重合视频片段与该第二参考片段之间的重合部分删除，得到该至少一个候选视频片段。

在其中一种可能的实施方式中，该融合模块 1502 还用于：比较第三类参考视频片段的时长与目标时长，该第三类参考视频片段是指删除重合部分的该第二重合视频片段。在该第三类参考视频片段的时长大于或等于该目标时长的情况下，保留该第三类参考视频片段。在该第三类参考视频片段的时长小于该目标时长的情况下，删除该第三类参考视频片段。

在其中一种可能的实施方式中，该目标视频片段确定模块 1503，用于基于该至少一个候选视频片段，确定该至少一个目标候选视频片段，该目标候选视频片段在该至少一个候

选视频片段中的出现次数符合次数条件。

对于任一该目标候选视频片段，在该目标候选视频片段在第一视频中的出现时间处于该目标时间范围的情况下，将该目标候选视频片段确定为该第一视频中的目标视频片段。

在其中一种可能的实施方式中，该目标视频片段确定模块 1503，用于基于所述至少一个候选视频片段，确定至少一个参考候选视频片段。确定每个参考候选视频片段在该至少一个参考候选视频片段的出现次数。将出现次数符合该出现次数条件的参考候选视频片段确定为目标候选视频片段。

在其中一种可能的实施方式中，该至少一个候选视频片段包括第三重合视频片段该第三重合视频片段是指属于该至少一个候选视频片段中第一候选视频片段的候选视频片段，该目标视频片段确定模块 1503，用于在该至少一个候选视频片段包括该第三重合视频片段的情况下，将该第三重合视频片段删除，得到该至少一个参考候选视频片段。

在其中一种可能的实施方式中，该至少一个候选视频片段包括第四重合视频片段，该第四重合视频片段是指与该至少一个候选视频片段中第二候选视频片段部分重合的候选视频片段，目标视频片段确定模块 1503，用于在该至少一个候选视频片段包括该第四重合视频片段，且该第四重合视频片段与该第二候选视频片段之间的重合度符合重合度条件的情况下，确定该第四重合视频片段的出现次数。基于重合度符合重合度条件的各第四重合视频片段各自对应的出现次数，确定该至少一个参考候选视频片段。

在其中一种可能的实施方式中，该至少一个候选视频片段包括第四重合视频片段，该第四重合视频片段是指与该至少一个候选视频片段中第二候选视频片段部分重合的候选视频片段，目标视频片段确定模块 1503，用于在该至少一个候选视频片段包括该第四重合视频片段，且该第四重合视频片段与该第二候选视频片段之间的重合度不符合该重合度条件的情况下，将该第四重合视频片段删除，得到该至少一个参考候选视频片段。

在一种可能的实施方式中，该至少一个候选视频片段包括第四重合视频片段，该第四重合视频片段是指与该至少一个候选视频片段中第二候选视频片段部分重合的候选视频片段，目标视频片段确定模块 1503，用于在该至少一个候选视频片段包括该第四重合视频片段，且该第四重合视频片段的时长小于该第二候选视频片段的情况下，将该第四重合视频片段删除，得到该至少一个参考候选视频片段。

在其中一种可能的实施方式中，该目标视频片段确定模块 1503，用于对于至少一个候选视频片段中任一符合重合度条件的第四重合视频片段，在该第四重合视频片段的出现次数大于或等于第一出现次数阈值的情况下，将该第四重合视频片段与第二候选视频片段进行融合，得到该至少一个参考候选视频片段。

在其中一种可能的实施方式中，该目标视频片段确定模块 1503，用于对于至少一个候选视频片段中任一符合重合度条件的第四重合视频片段，在该第四重合视频片段的出现次数小于该第一出现次数阈值的情况下，将该第四重合视频片段删除，得到该至少一个参考候选视频片段。

在其中一种可能的实施方式中，该装置还包括：

特征提取模块，用于对待识别的目标视频的多个目标视频帧进行特征提取，得到该多个目标视频帧的视频帧特征。

该目标视频片段确定模块 1503，还用于基于该多个目标视频帧的视频帧特征、该第一视频帧的视频帧特征以及该至少一个第二视频的视频帧特征，确定该目标视频的至少一个目标视频片段。

需要说明的是：上述实施例提供的视频片段的识别装置在识别视频片段时，仅以上述各功能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模块完成，即将计算机设备的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的视频片段的识别装置与视频片段的识别方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

通过本申请实施例提供的技术方案，基于视频帧特征之间的相似度，确定包含相似视频帧的视频帧对。基于出现时间差值来对视频帧对中的第一视频帧进行融合，得到至少一个候选视频片段。最终从至少一个候选视频片段中确定出处于目标时间范围的目标视频片段。确定目标片段的过程无需人工参与，由计算机设备直接基于第一视频和至少一个第二视频就能够自动进行，效率较高。

本申请实施例提供了一种计算机设备，用于执行上述方法，该计算机设备可以实现为终端或者服务器，下面先对终端的结构进行介绍：

图 16 是本申请实施例提供的一种终端的结构示意图。

通常，终端 1600 包括有：一个或多个处理器 1601 和一个或多个存储器 1602。

处理器 1601 可以包括一个或多个处理核心，比如 4 核心处理器、8 核心处理器等。处理器 1601 可以采用 DSP (Digital Signal Processing, 数字信号处理)、FPGA (Field-Programmable Gate Array, 现场可编程门阵列)、PLA (Programmable Logic Array, 可编程逻辑阵列) 中的至少一种硬件形式来实现。处理器 1601 也可以包括主处理器和协处理器，主处理器是用于对在唤醒状态下的数据进行处理的处理单元，也称 CPU (Central Processing Unit, 中央处理器)；协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中，处理器 1601 可以在集成有 GPU (Graphics Processing Unit, 图像处理单元)，GPU 用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中，处理器 1601 还可以包括 AI (Artificial Intelligence, 人工智能) 处理器，该 AI 处理器用于处理有关机器学习的计算操作。

存储器 1602 可以包括一个或多个计算机可读存储介质，该计算机可读存储介质可以是非暂态的。存储器 1602 还可包括高速随机存取存储器，以及非易失性存储器，比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中，存储器 1602 中的非暂态的计算机可读存储介质用于存储至少一个计算机程序，该至少一个计算机程序用于被处理器 1601 所执行以实现本申请中方法实施例提供的视频片段的识别方法。

在一些实施例中,终端 1600 还可选包括有:外围设备接口 1603 和至少一个外围设备。处理器 1601、存储器 1602 和外围设备接口 1603 之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口 1603 相连。具体地,外围设备包括:射频电路 1604、显示屏 1605、摄像头组件 1606、音频电路 1607 和电源 1608 中的至少一种。

外围设备接口 1603 可被用于将 I/O (Input /Output, 输入/输出) 相关的至少一个外围设备连接到处理器 1601 和存储器 1602。在一些实施例中,处理器 1601、存储器 1602 和外围设备接口 1603 被集成在同一芯片或电路板上;在一些其他实施例中,处理器 1601、存储器 1602 和外围设备接口 1603 中的任意一个或两个可以在单独的芯片或电路板上实现,本实施例对此不加以限定。

射频电路 1604 用于接收和发射 RF (Radio Frequency, 射频) 信号,也称电磁信号。射频电路 1604 通过电磁信号与通信网络以及其他通信设备进行通信。射频电路 1604 将电信号转换为电磁信号进行发送,或者,将接收到的电磁信号转换为电信号。可选地,射频电路 1604 包括:天线系统、RF 收发器、一个或多个放大器、调谐器、振荡器、数字信号处理器、编解码芯片组、用户身份模块卡等等。

显示屏 1605 用于显示 UI (User Interface, 用户界面)。该 UI 可以包括图形、文本、图标、视频及其它们的任意组合。当显示屏 1605 是触摸显示屏时,显示屏 1605 还具有采集在显示屏 1605 的表面或表面上方的触摸信号的能力。该触摸信号可以作为控制信号输入至处理器 1601 进行处理。此时,显示屏 1605 还可以用于提供虚拟按钮和/或虚拟键盘,也称软按钮和/或软键盘。

摄像头组件 1606 用于采集图像或视频。可选地,摄像头组件 1606 包括前置摄像头和后置摄像头。通常,前置摄像头设置在终端的前面板,后置摄像头设置在终端的背面。

音频电路 1607 可以包括麦克风和扬声器。麦克风用于采集用户及环境的声波,并将声波转换为电信号输入至处理器 1601 进行处理,或者输入至射频电路 1604 以实现语音通信。

电源 1608 用于为终端 1600 中的各个组件进行供电。电源 1608 可以是交流电、直流电、一次性电池或可充电电池。

在一些实施例中,终端 1600 还包括有一个或多个传感器 1609。该一个或多个传感器 1609 包括但不限于:加速度传感器 1610、陀螺仪传感器 1611、压力传感器 1612、光学传感器 1613 以及接近传感器 1614。

加速度传感器 1610 可以检测以终端 1600 建立的坐标系的三个坐标轴上的加速度大小。

陀螺仪传感器 1611 可以终端 1600 的机体方向及转动角度,陀螺仪传感器 1611 可以与加速度传感器 1610 协同采集用户对终端 1600 的 3D 动作。

压力传感器 1612 可以设置在终端 1600 的侧边框和/或显示屏 1605 的下层。当压力传

感器 1612 设置在终端 1600 的侧边框时,可以检测用户对终端 1600 的握持信号,由处理器 1601 根据压力传感器 1612 采集的握持信号进行左右手识别或快捷操作。当压力传感器 1612 设置在显示屏 1605 的下层时,由处理器 1601 根据用户对显示屏 1605 的压力操作,实现对 UI 界面上的可操作性控件进行控制。

光学传感器 1613 用于采集环境光强度。在一个实施例中,处理器 1601 可以根据光学传感器 1613 采集的环境光强度,控制显示屏 1605 的显示亮度。

接近传感器 1614 用于采集用户与终端 1600 的正面之间的距离。

本领域技术人员可以理解,图 16 中示出的结构并不构成对终端 1600 的限定,可以包括比图示更多或更少的组件,或者组合某些组件,或者采用不同的组件布置。

上述计算机设备还可以实现为服务器,下面对服务器的结构进行介绍:

图 17 是本申请实施例提供的一种服务器的结构示意图,该服务器 1700 可因配置或性能不同而产生比较大的差异,可以包括一个或多个处理器(Central Processing Units, CPU) 1701 和一个或多个存储器 1702,其中,所述一个或多个存储器 1702 中存储有至少一条计算机程序,所述至少一条计算机程序由所述一个或多个处理器 1701 加载并执行以实现上述各个方法实施例提供的方法。当然,该服务器 1700 还可以具有有线或无线网络接口、键盘以及输入输出接口等部件,以便进行输入输出,该服务器 1700 还可以包括其他用于实现设备功能的部件,在此不做赘述。

在示例性实施例中,还提供了一种计算机可读存储介质,该计算机可读存储介质中存储有至少一条计算机程序,该计算机程序由处理器加载并执行以实现上述实施例中的视频片段的识别方法。例如,该计算机可读存储介质可以是只读存储器(Read-Only Memory, ROM)、随机存取存储器(Random Access Memory, RAM)、只读光盘(Compact Disc Read-Only Memory, CD-ROM)、磁带、软盘和光数据存储设备等。

在示例性实施例中,还提供了一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现上述视频片段的识别方法。

在一些实施例中,本申请实施例所涉及的计算机程序可被部署在一个计算机设备上执行,或者在位于一个地点的多个计算机设备上执行,又或者,在分布在多个地点且通过通信网络互连的多个计算机设备上执行,分布在多个地点且通过通信网络互连的多个计算机设备可以组成区块链系统。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来指令相关的硬件完成,该程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

上述仅为本申请的可选实施例,并不用以限制本申请,凡在本申请的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

权利要求书

1. 一种视频片段的识别方法，由计算机设备执行，所述方法包括：

获取第一视频的视频帧特征和至少一个第二视频的视频帧特征，并基于所述第一视频的视频帧特征以及所述至少一个第二视频的视频帧特征，确定多个视频帧对，所述视频帧对包括相似度符合相似度条件的第一视频帧和第二视频帧，所述第一视频帧属于所述第一视频，所述第二视频帧属于所述至少一个第二视频；

基于所述多个视频帧对的出现时间差值，将所述多个视频帧对中的第一视频帧进行融合，得到所述第一视频中的至少一个候选视频片段，所述出现时间差值是指所述视频帧对中的两个视频帧在视频中的出现时间之间的差值；及

获取目标时间范围，并基于所述至少一个候选视频片段以及所述目标时间范围，确定所述第一视频中的至少一个目标视频片段，所述目标视频片段处于所述第一视频的所述目标时间范围内。

2. 根据权利要求1所述的方法，其特征在于，所述基于所述多个视频帧对的出现时间差值，将所述多个视频帧对中的第一视频帧进行融合，得到所述第一视频中的至少一个候选视频片段包括：

基于所述多个视频帧对的出现时间差值，将所述多个视频帧对划分为多个视频帧组，同一个所述视频帧组中的视频帧对对应于同一个出现时间差值；及

对于所述多个视频帧组中的任一视频帧组，按照所述视频帧组中视频帧对的第一视频帧在所述第一视频中的出现时间，将所述视频帧组中视频帧对的第一视频帧融合为一个所述候选视频片段。

3. 根据权利要求2所述的方法，其特征在于，在所述基于所述多个视频帧对的出现时间差值，将所述多个视频帧对划分为多个视频帧组之前，所述方法还包括：

针对所述多个视频帧对中的任一视频帧对，确定所述视频帧对中的第一视频帧的第一出现时间和第二视频帧的第二出现时间；所述第一出现时间是指第一视频帧在第一视频中出现的时间，所述第二出现时间是指第二视频帧在第二视频中出现的时间；

将所述视频帧对中的第一视频帧的第一出现时间减去第二视频帧的第二出现时间，得到所述视频帧对的出现时间差值；

所述基于所述多个视频帧对的出现时间差值，将所述多个视频帧对划分为多个视频帧组包括：将出现时间差值相同的视频帧对划分为一个初始视频帧组，并将所述初始视频帧组中的视频帧对的出现时间差值，作为所述初始视频帧组所对应的出现时间差值；及

基于多个初始视频帧组对应的出现时间差值，将所述多个初始视频帧组进行融合，得到所述多个视频帧组。

4. 根据权利要求3所述的方法，其特征在于，所述基于多个初始视频帧组对应的出现时间差值，将所述多个初始视频帧组进行融合，得到所述多个视频帧组包括：

获取预设的配置信息，所述配置信息包括目标顺序；

按照所述目标顺序对所述多个初始视频帧组进行排序，得到多个候选视频帧组；及

对于所述多个候选视频帧组中任两个相邻的候选视频帧组，在所述两个相邻的候选视频帧组之间的匹配时间差值符合匹配时间差值条件的情况下，将所述两个相邻的候选视频帧组融合为一个视频帧组，所述匹配时间差值是指所述两个相邻的候选视频帧组对应的出现时间差值之间的差值。

5. 根据权利要求 4 所述的方法，其特征在于，所述两个相邻的候选视频帧组包括第一候选视频帧组和第二候选视频帧组，所述将所述两个相邻的候选视频帧组融合为一个视频帧组包括：

在所述第一候选视频帧组与所述第二候选视频帧组之间的匹配时间差值小于或等于匹配差值阈值的情况下，将所述第一候选视频帧组中的视频帧对添加至所述第二候选视频帧组；及

基于所述第二候选视频帧组对应的出现时间差值，采用参考第二视频帧替换目标第二视频帧，得到所述视频帧组；所述目标第二视频帧为新添加至所述第二候选视频帧组中的第二视频帧；所述参考第二视频帧为所述第二视频中与目标第一视频帧之间的出现时间差值为目标差值的第二视频帧；所述目标差值为所述第二候选视频帧组对应的出现时间差值；所述目标第一视频帧为所述目标第二视频帧所属视频帧对中的第一视频帧。

6. 根据权利要求 2 所述的方法，其特征在于，所述按照所述视频帧组中视频帧对的第一视频帧在所述第一视频中的出现时间，将所述视频帧组中视频帧对的第一视频帧融合为一个所述候选视频片段包括：

对所述视频帧组中的视频帧对进行遍历，确定当前遍历至的当前视频帧对和上一个遍历至的上一个视频帧对；所述当前视频帧对和所述上一个视频帧对为所述视频帧组中的两个相邻的视频帧对；

比较所述当前视频帧对和所述上一个视频帧对的第一视频帧在所述第一视频中的出现时间，得到第一视频帧的出现时间的差值；

在所述第一视频帧的出现时间的差值符合出现时间条件的情况下，将所述当前视频帧对和所述上一个视频帧对添加至临时帧列表；

在所述第一视频帧的出现时间的差值不符合出现时间条件的情况下，将所述临时帧列表中的视频帧对融合为参考视频片段，并在融合后清空所述临时帧列表；

确定下一个遍历至的视频帧对，将所述下一个遍历至的视频帧对作为新的当前视频帧对，返回比较所述当前视频帧对和所述上一个视频帧对的第一视频帧在所述第一视频中的出现时间的步骤继续执行，直至最后一个遍历至的视频帧对；及

基于多个参考视频片段，确定所述至少一个候选视频片段。

7. 根据权利要求 6 所述的方法，其特征在于，所述多个参考视频片段包括第一重合视频片段，所述第一重合视频片段是指属于所述多个参考视频片段中第一参考视频片段的参考视频片段；所述基于多个参考视频片段，确定所述至少一个候选视频片段包括：

在所述多个参考视频片段包括所述第一重合视频片段的情况下，将所述第一重合视频片段删除，得到所述至少一个候选视频片段。

8. 根据权利要求 6 所述的方法，其特征在于，所述多个参考视频片段包括第二重合视频片段，所述第二重合视频片段是指，与所述多个参考视频片段中的第二参考视频片段部分重合的参考视频片段；所述基于多个参考视频片段，确定所述至少一个候选视频片段包括：

在所述多个参考视频片段包括所述第二重合视频片段的情况下，将所述第二重合视频片段与所述第二参考片段之间的重合部分删除，得到所述至少一个候选视频片段。

9. 根据权利要求 8 所述的方法，其特征在于，所述在所述多个参考视频片段包括所述第二重合视频片段的情况下，将所述第二重合视频片段与所述第二参考片段之间的重合部分删除之后，所述方法还包括：

比较第三类参考视频片段的时长与目标时长，所述第三类参考视频片段是指删除重合部分的所述第二重合视频片段；

在所述第三类参考视频片段的时长大于或等于所述目标时长的情况下，保留所述第三类参考视频片段；及

在所述第三类参考视频片段的时长小于所述目标时长的情况下，删除所述第三类参考视频片段。

10. 根据权利要求 1 所述的方法，其特征在于，所述基于所述至少一个候选视频片段以及所述目标时间范围，确定所述第一视频中的至少一个目标视频片段包括：

基于所述至少一个候选视频片段，确定至少一个目标候选视频片段，所述目标候选视频片段在所述至少一个候选视频片段中的出现次数符合次数条件；及

对于任一所述目标候选视频片段，在所述目标候选视频片段在所述第一视频中的出现时间处于所述目标时间范围的情况下，将所述目标候选视频片段确定为所述第一视频中的目标视频片段。

11. 根据权利要求 10 所述的方法，其特征在于，所述基于所述至少一个候选视频片段，确定至少一个目标候选视频片段包括：

基于所述至少一个候选视频片段，确定至少一个参考候选视频片段；

确定每个所述参考候选视频片段在所述至少一个参考候选视频片段的出现次数；及

将出现次数符合所述出现次数条件的参考候选视频片段确定为目标候选视频片段。

12. 根据权利要求 11 所述的方法，其特征在于，所述至少一个候选视频片段包括第三重合视频片段，所述第三重合视频片段是指属于所述至少一个候选视频片段中第一候选视频片段的候选视频片段，所述基于所述至少一个候选视频片段，确定至少一个参考候选视频片段包括：

在所述至少一个候选视频片段包括所述第三重合视频片段的情况下，将所述第三重合视频片段删除，得到所述至少一个参考候选视频片段。

13. 根据权利要求 11 所述的方法, 其特征在于, 所述至少一个候选视频片段包括第四重合视频片段, 所述第四重合视频片段是指与所述至少一个候选视频片段中第二候选视频片段部分重合的候选视频片段, 所述基于所述至少一个候选视频片段, 确定至少一个参考候选视频片段包括:

在所述至少一个候选视频片段包括所述第四重合视频片段, 且所述第四重合视频片段与所述第二候选视频片段之间的重合度符合重合度条件的情况下, 确定所述第四重合视频片段的出现次数; 基于重合度符合重合度条件的各所述第四重合视频片段各自对应的出现次数, 确定所述至少一个参考候选视频片段。

14. 根据权利要求 11 所述的方法, 其特征在于, 所述至少一个候选视频片段包括第四重合视频片段, 所述第四重合视频片段是指与所述至少一个候选视频片段中第二候选视频片段部分重合的候选视频片段, 所述基于所述至少一个候选视频片段, 确定至少一个参考候选视频片段包括:

在所述至少一个候选视频片段包括所述第四重合视频片段, 且所述第四重合视频片段与所述第二候选视频片段之间的重合度不符合所述重合度条件的情况下, 将所述第四重合视频片段删除, 得到所述至少一个参考候选视频片段。

15. 根据权利要求 11 所述的方法, 其特征在于, 所述至少一个候选视频片段包括第四重合视频片段, 所述第四重合视频片段是指与所述至少一个候选视频片段中第二候选视频片段部分重合的候选视频片段, 所述基于所述至少一个候选视频片段, 确定至少一个参考候选视频片段包括:

在所述至少一个候选视频片段包括所述第四重合视频片段, 且所述第四重合视频片段的时长小于所述第二候选视频片段的情况下, 将所述第四重合视频片段删除, 得到所述至少一个参考候选视频片段。

16. 根据权利要求 13 所述的方法, 其特征在于, 所述基于重合度符合重合度条件的各所述第四重合视频片段各自对应的出现次数, 确定所述至少一个参考候选视频片段包括:

对于所述至少一个候选视频片段中任一个符合重合度条件的第四重合视频片段, 在所述第四重合视频片段的出现次数大于或等于第一出现次数阈值的情况下, 将所述第四重合视频片段与第二候选视频片段进行融合, 得到所述至少一个参考候选视频片段。

17. 根据权利要求 13 所述的方法, 其特征在于, 所述基于重合度符合重合度条件的各所述第四重合视频片段各自对应的出现次数, 确定所述至少一个参考候选视频片段包括:

对于所述至少一个候选视频片段中任一个符合重合度条件的第四重合视频片段, 在所述第四重合视频片段的出现次数小于所述第一出现次数阈值的情况下, 将所述第四重合视频片段删除, 得到所述至少一个参考候选视频片段。

18. 根据权利要求 1 所述的方法, 其特征在于, 所述方法还包括:

获取待识别的目标视频, 并对所述待识别的目标视频的多个目标视频帧进行特征提取, 得到所述多个目标视频帧的视频帧特征; 及

基于所述多个目标视频帧的视频帧特征、所述第一视频帧的视频帧特征以及所述至少一个第二视频的视频帧特征，确定所述目标视频的至少一个目标视频片段。

19. 一种视频片段的识别装置，其特征在于，所述装置包括：

视频帧对确定模块，用于获取第一视频的视频帧特征和至少一个第二视频的视频帧特征，并基于第一视频的视频帧特征以及至少一个第二视频的视频帧特征，确定多个视频帧对，所述视频帧对包括相似度符合相似度条件的第一视频帧和第二视频帧，所述第一视频帧属于所述第一视频，所述第二视频帧属于所述至少一个第二视频；

融合模块，用于基于所述多个视频帧对的出现时间差值，将所述多个视频帧对中的第一视频帧进行融合，得到所述第一视频中的至少一个候选视频片段，所述出现时间差值是指所述视频帧对中的两个视频帧在视频中的出现时间之间的差值；及

目标视频片段确定模块，用于获取目标时间范围，并基于所述至少一个候选视频片段以及目标时间范围，确定所述第一视频中的至少一个目标视频片段，所述目标视频片段处于所述第一视频的所述目标时间范围内。

20. 一种计算机设备，其特征在于，所述计算机设备包括一个或多个处理器和一个或多个存储器，所述一个或多个存储器中存储有至少一条计算机程序，所述计算机程序由所述一个或多个处理器加载并执行以实现如权利要求 1 至权利要求 18 任一项所述的视频片段的识别方法。

21. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质中存储有至少一条计算机程序，所述计算机程序由处理器加载并执行以实现如权利要求 1 至权利要求 18 任一项所述的视频片段的识别方法。

22. 一种计算机程序产品，包括计算机程序，其特征在于，该计算机程序被处理器执行时实现权利要求 1 至权利要求 18 任一项所述的视频片段的识别方法。