(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 116156246 B
(45)授权公告日 2024.10.25

(21)申请号 202310025329.2
(22)申请日 2023.01.09
(65)同一申请的已公布的文献号
申请公布号 CN 116156246 A
(43)申请公布日 2023.05.23
(73)专利权人 北京水滴科技集团有限公司
地址 100102 北京市朝阳区利泽中二路2号C座二层201
(72)发明人 王月宝 黄明星 李银锋 刘海伦 许垒 沈鹏 胡尧 周晓波
(74)专利代理机构 北京中强智尚知识产权代理有限公司 11448
专利代理师 刘丽颖
(51)Int.Cl.
H04N 21/44(2011.01)
H04N 21/435(2011.01)
H04N 21/488(2011.01)
(56)对比文件
CN 111565338 A,2020.08.21
CN 111798543 A,2020.10.20
审查员 周思

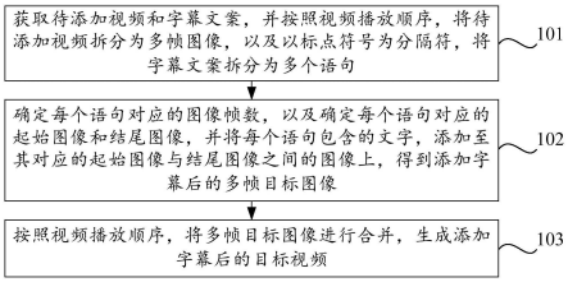
权利要求书2页 说明书8页 附图2页

(54)发明名称

一种视频字幕添加方法、装置、电子设备和可读存储介质

(57)摘要

本申请提供了一种视频字幕添加方法、装置、电子设备和可读存储介质，涉及视频处理技术领域。该方法包括：获取待添加视频和字幕文案，并按照视频播放顺序，将待添加视频拆分为多帧图像，以及以标点符号为分隔符，将字幕文案拆分为多个语句；确定每个语句对应的图像帧数，以及确定每个语句对应的起始图像和结尾图像，并将每个语句包含的文字，添加至其对应的起始图像与结尾图像之间的图像上，得到添加字幕后的多帧目标图像；按照视频播放顺序，将多帧目标图像进行合并，生成添加字幕后的目标视频。本申请实现了更加简便的视频字幕添加，提高了添加效率，降低了添加错误率，并且两句话的字幕之间具有一定的时间间隔，实现语句之间的停顿。



CN 116156246 B

1.一种视频字幕添加方法,其特征在于,包括:

获取待添加视频和字幕文案,并按照视频播放顺序,将所述待添加视频拆分为多帧图像,以及以标点符号为分隔符,将所述字幕文案拆分为多个语句;

确定每个所述语句对应的图像帧数,以及确定每个所述语句对应的起始图像和结尾图像,并将每个所述语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上,得到添加字幕后的多帧目标图像;

按照所述视频播放顺序,将多帧所述目标图像进行合并,生成添加字幕后的目标视频;

其中,对于第i个语句,将所述第i个语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上,具体包括:

将所述第i个语句包含的文字,添加至第一图像与第二图像之间的每帧图片上;

其中,所述第一图像为所述第i个语句的起始图像的后a个图像,所述第二图像为所述第i个语句的结尾图像的前b个图像,a、b均为大于或等于1,且小于或等于4的整数,i为大于或等于1,且小于或等于多个所述语句的总数量。

2.根据权利要求1所述的方法,其特征在于,确定每个所述语句对应的图像帧数,具体包括:

根据多帧所述图像的总数量和所述字幕文案的文字总字数,计算所述字幕文案的每个文字对应的图像帧数;

根据每个文字对应的图像帧数和所述语句包含的文字字数,计算每个所述语句对应的图像帧数。

3.根据权利要求1所述的方法,其特征在于,

所述字幕文案的文字总字数不包括所述标点符号的数量。

4.根据权利要求1所述的方法,其特征在于,在将每个所述语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上之前,还包括:

若所述语句对应的图像帧数不为整数,且图像帧数的小数部分大于或等于0.5,则所述图像帧数向上取整;

若所述语句对应的图像帧数不为整数,且图像帧数的小数部分小于0.5,则所述图像帧数向下取整。

5.根据权利要求1至4中任一项所述的方法,其特征在于,所述字幕文案为文案编辑人员所编辑的,且在所述以标点符号为分隔符,将所述字幕文案拆分为多个语句之前,还包括:

对所述字幕文案进行校正,保留有效文案。

6.根据权利要求1至4中任一项所述的方法,其特征在于,

所述目标视频的视频时长小于或等于3分钟。

7.一种视频字幕添加装置,其特征在于,包括:

获取模块,用于获取待添加视频和字幕文案;

拆分模块,用于按照视频播放顺序,将所述待添加视频拆分为多帧图像,以及以标点符号为分隔符,将所述字幕文案拆分为多个语句;

字幕添加模块,用于确定每个所述语句对应的图像帧数,以及确定每个所述语句对应的起始图像和结尾图像,并将每个所述语句包含的文字,添加至其对应的起始图像与结尾

图像之间的图像上,得到添加字幕后的多帧目标图像;

图像合并模块,用于按照所述视频播放顺序,将多帧所述目标图像进行合并,生成添加字幕后的目标视频;

其中,所述字幕添加模块,具体用于:

将第i个语句包含的文字,添加至第一图像与第二图像之间的每帧图片上;

其中,所述第一图像为所述第i个语句的起始图像的后a个图像,所述第二图像为所述第i个语句的结尾图像的前b个图像,a、b均为大于或等于1,且小于或等于4的整数,i为大于或等于1,且小于或等于多个所述语句的总数量。

8.一种电子设备,其特征在于,包括处理器和存储器,所述存储器存储有在所述处理器上运行的程序或指令,所述程序或指令被所述处理器执行时实现如权利要求1至6中任一项所述的视频字幕添加方法的步骤。

9.一种可读存储介质,其上存储有程序或指令,其特征在于,所述程序或指令被处理器执行时实现如权利要求1至6中任一项所述的视频字幕添加方法的步骤。

一种视频字幕添加方法、装置、电子设备和可读存储介质

技术领域

[0001] 本申请涉及视频处理技术领域，尤其是涉及到一种视频字幕添加方法、视频字幕添加装置、电子设备和可读存储介质。

背景技术

[0002] 近年来，自媒体的发展如火如荼，各大短视频平台应运而生，用户利用短视频生成技术实现短视频的生成，进而进行发布。具体地，将文案文本转译为对应的音频，利用嘴形驱动技术，用音频让某一形象“动起来”，最后再利用语音识别技术将音频识别为文字，将文字添加至视频。

[0003] 但是，语音识别算法的训练需要大量训练数据，且为更好的效果，针对不同的话术场景需要不同的训练数据，开发过程耗时费力；并且，由于语音识别算法存在错误率，这就导致生成的字幕存在一定量的错别字，则需要人工进行检查修改。

发明内容

[0004] 有鉴于此，本申请提供了一种视频字幕添加方法、视频字幕添加装置、电子设备和可读存储介质，实现了更加简便的视频字幕添加，提高了视频字幕添加效率，降低了视频字幕添加错误率。

[0005] 第一方面，本申请实施例提供了一种视频字幕添加方法，包括：获取待添加视频和字幕文案，并按照视频播放顺序，将待添加视频拆分为多帧图像，以及以标点符号为分隔符，将字幕文案拆分为多个语句；确定每个语句对应的图像帧数，以及确定每个语句对应的起始图像和结尾图像，并将每个语句包含的文字，添加至其对应的起始图像与结尾图像之间的图像上，得到添加字幕后的多帧目标图像；按照视频播放顺序，将多帧目标图像进行合并，生成添加字幕后的目标视频。

[0006] 根据本申请实施例的上述方法，还可以具有以下附加技术特征：

[0007] 在上述技术方案中，可选地，根据每个语句向每帧图像添加字幕，得到添加字幕后的多帧目标图像，具体包括：确定每个语句对应的图像帧数，以及确定每个语句对应的起始图像和结尾图像；将每个语句包含的文字，添加至其对应的起始图像与结尾图像之间的图像上，得到添加字幕后的多帧目标图像。

[0008] 在上述任一技术方案中，可选地，确定每个语句对应的图像帧数，具体包括：根据多帧图像的总数量和字幕文案的文字总字数，计算字幕文案的每个文字对应的图像帧数；根据每个文字对应的图像帧数和语句包含的文字字数，计算每个语句对应的图像帧数。

[0009] 在上述任一技术方案中，可选地，字幕文案的文字总字数不包括标点符号的数量。

[0010] 在上述任一技术方案中，可选地，在将每个语句包含的文字，添加至其对应的起始图像与结尾图像之间的图像上之前，还包括：若语句对应的图像帧数不为整数，且图像帧数的小数部分大于或等于0.5，则图像帧数向上取整；若语句对应的图像帧数不为整数，且图像帧数的小数部分小于0.5，则图像帧数向下取整。

[0011] 在上述任一技术方案中,可选地,对于第i个语句,将第i个语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上,具体包括:将第i个语句包含的文字,添加至第一图像与第二图像之间的每帧图片上;其中,第一图像为第i个语句的起始图像的后a个图像,第二图像为第i个语句的结尾图像的前b个图像,a、b均为大于或等于1,且小于或等于4的整数,i为大于或等于1,且小于或等于多个语句的总数量。

[0012] 在上述任一技术方案中,可选地,目标视频的视频时长小于或等于3分钟。

[0013] 在上述任一技术方案中,可选地,字幕文案为文案编辑人员所编辑的,且在以标点符号为分隔符,将字幕文案拆分为多个语句之前,还包括:对字幕文案进行校正,保留有效文案。

[0014] 第二方面,本申请实施例提供了一种视频字幕添加装置,包括:获取模块,用于获取待添加视频和字幕文案;拆分模块,用于按照视频播放顺序,将待添加视频拆分为多帧图像,以及以标点符号为分隔符,将字幕文案拆分为多个语句;字幕添加模块,用于确定每个语句对应的图像帧数,以及确定每个语句对应的起始图像和结尾图像,并将每个语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上,得到添加字幕后的多帧目标图像;图像合并模块,用于按照视频播放顺序,将多帧目标图像进行合并,生成添加字幕后的目标视频。

[0015] 第三方面,本申请实施例提供了一种电子设备,该电子设备包括处理器和存储器,存储器存储可在处理器上运行的程序或指令,程序或指令被处理器执行时实现如第一方面的方法的步骤。

[0016] 第四方面,本申请实施例提供了一种可读存储介质,该可读存储介质上存储程序或指令,程序或指令被处理器执行时实现如第一方面的方法的步骤。

[0017] 第五方面,本申请实施例提供了一种芯片,该芯片包括处理器和通信接口,通信接口和处理器耦合,处理器用于运行程序或指令,实现如第一方面的方法。

[0018] 第六方面,本申请实施例提供一种计算机程序产品,该程序产品被存储在存储介质中,该程序产品被至少一个处理器执行以实现如第一方面的方法。

[0019] 在本申请实施例中,在获取到待添加视频后,按照视频播放顺序,也即图像帧的时间顺序,将待添加视频拆分为M帧图像,以及获取需要添加为视频字幕的字幕文案,该字幕文案由文字、字母、标点符号等构成,并将字幕文案按照其包含的标点符号拆分为Z个语句,M为大于1的整数,Z为大于1的整数。进一步地,将图像帧数与字幕字数映射起来,从而实现向图像添加字幕,得到添加字幕后的多帧目标图像。最后,再按照视频播放顺序,也即图像帧的时间顺序,将多帧目标图像进行合并,得到添加字幕后的目标视频。并且,在根据每个语句向每帧图像添加字幕时,对于任一语句,确定该语句对应的图像帧数,再确定该语句对应的起始图像和结尾图像,最后将该语句包含的文字,添加至起始图像与结尾图像之间的图像上。也就是说,该语句对应的起始图像和结尾图像上不添加文字,使得每个语句之间至少有两帧图像上没有文字,由此使得两句话的字幕之间具有一定的时间间隔,实现语句之间的停顿。

[0020] 本申请实施例中,一方面,添加至视频的字幕直接来自于原字幕文案,方法操作简单,字幕添加速度快。相比于相关技术中利用语音识别模型的方法,无需收集语音数据训练语音识别模型,避免视频开发过程耗时费力,且降低了视频字幕添加错误率。另一方面,在

语句对应的起始图像与结尾图像之间的图像上添加文字,使得两句话的字幕之间具有一定的时间间隔,实现语句之间的停顿,保证更好的视觉感觉。

[0021] 上述说明仅是本申请技术方案的概述,为了能够更清楚了解本申请的技术手段,而可依照说明书的内容予以实施,并且为了让本申请的上述和其它目的、特征和优点能够更明显易懂,以下特举本申请的具体实施方式。

附图说明

[0022] 此处所说明的附图用来提供对本申请的进一步理解,构成本申请的一部分,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。在附图中:

[0023] 图1示出了本申请实施例的视频字幕添加方法的流程示意图;

[0024] 图2示出了本申请实施例的视频字幕添加装置的结构框图;

[0025] 图3示出了本申请实施例的电子设备的结构框图。

具体实施方式

[0026] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚地描述,显然,所描述的实施例是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员获得的所有其他实施例,都属于本申请保护的范围。

[0027] 本申请的说明书和权利要求书中的术语“第一”、“第二”等是用于区别类似的对象,而不用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施,且“第一”、“第二”等所区分的对象通常为一类,并不限定对象的个数,例如第一对象可以是一个,也可以是多个。此外,说明书以及权利要求中“和/或”表示所连接对象的至少其中之一,字符“/”,一般表示前后关联对象是一种“或”的关系。

[0028] 下面结合附图,通过具体的实施例及其应用场景对本申请实施例提供的视频字幕添加方法、视频字幕添加装置、电子设备和可读存储介质进行详细地说明。

[0029] 本申请实施例提供了一种视频字幕添加方法,如图1所示,该方法包括:

[0030] 步骤101,获取待添加视频和字幕文案,并按照视频播放顺序,将待添加视频拆分为多帧图像,以及以标点符号为分隔符,将字幕文案拆分为多个语句;

[0031] 步骤102,确定每个语句对应的图像帧数,以及确定每个语句对应的起始图像和结尾图像,并将每个语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上,得到添加字幕后的多帧目标图像;

[0032] 步骤103,按照视频播放顺序,将多帧目标图像进行合并,生成添加字幕后的目标视频。

[0033] 在该实施例中,视频是由一帧一帧的图像按照时间顺序构成,例如,时长10秒帧率为24的视频由240张图像构成,在获取到待添加视频后,按照视频播放顺序,也即图像帧的时间顺序,将待添加视频拆分为M帧图像,M为大于1的整数。以及,获取需要添加为视频字幕的字幕文案,该字幕文案由文字、字母、标点符号等构成,并将字幕文案按照其包含的标点符号拆分为Z个语句,Z为大于1的整数。

[0034] 进一步地,将图像帧数与字幕字数映射起来,从而实现向图像添加字幕,得到添加

字幕后的多帧目标图像。最后,再按照视频播放顺序,也即图像帧的时间顺序,将多帧目标图像进行合并,得到添加字幕后的目标视频。

[0035] 示例性地,获取文案编辑人员编辑的字幕文案,以及获取视频编辑人员提供的动画视频,将字幕文案转译为对应的音频,向动画视频添加音频,并且,利用原字幕文案通过上述方式对动画视频添加字幕,最终得到添加字幕后的动画视频。

[0036] 在本申请的一个实施例中,目标视频可以为短视频,其视频时长小于或等于3分钟。

[0037] 值得注意的是,在根据每个语句向每帧图像添加字幕时,对于任一语句,确定该语句对应的图像帧数,再确定该语句对应的起始图像和结尾图像,最后将该语句包含的文字,添加至起始图像与结尾图像之间的图像上。也就是说,该语句对应的起始图像和结尾图像上不添加文字,使得每个语句之间至少有两帧图像上没有文字,由此使得两句话的字幕之间具有一定的时间间隔,实现语句之间的停顿。

[0038] 本申请实施例中,一方面,添加至视频的字幕直接来自于原字幕文案,方法操作简单,字幕添加速度快。相比于相关技术中利用语音识别模型的方法,无需收集语音数据训练语音识别模型,避免视频开发过程耗时费力,且降低了视频字幕添加错误率。另一方面,在语句对应的起始图像与结尾图像之间的图像上添加文字,使得两句话的字幕之间具有一定的时间间隔,实现语句之间的停顿,保证更好的视觉感觉。

[0039] 在本申请的一个实施例中,字幕文案为文案编辑人员所编辑的,也即,添加至视频的字幕直接来自于原字幕文案,无需进行音频识别而得到字幕文案,从而提高了字幕添加准确性。

[0040] 并且,在以标点符号为分隔符,将字幕文案拆分为多个语句之前,还包括:对字幕文案进行校正,保留有效文案。其中,对字幕文案进行校正包括过滤掉重复文字、修改错别字等操作,例如。对于文案编辑人员输入的“我家的的门前”,将其中一个多余的“的”字去掉,则得到的有效的字幕文案为“我家的门前”。

[0041] 本申请实施例,通过对文案编辑人员输入的字幕文案进行校正,保证字幕文案的准确性。

[0042] 在本申请的一个实施例中,确定每个语句对应的图像帧数,具体包括:根据多帧图像的总数量和字幕文案的文字总字数,计算字幕文案的每个文字对应的图像帧数;根据每个文字对应的图像帧数和语句包含的文字字数,计算每个语句对应的图像帧数。

[0043] 在该实施例中,待添加视频语速均匀,视频帧数为M,也即多帧图像的总数量为M,字幕文案的文字总字数为N,显然M>N。将视频帧数除以字幕文案的文字总字数,得到每个文字对应的图像帧数,也即,M/N表示每个文字对应几帧图像。

[0044] 再通过每个文字对应的图像帧数与语句包含的文字字数的乘积,得到每个语句对应的图像帧数,例如,包含的文字字数为p的语句,其对应的图像帧数为(M/N)×p。依此计算字幕文案中每一个语句对应的图像帧数,并映射到待添加视频的视频帧上,由此实现添加字幕。

[0045] 本申请实施例,将字幕文案中的语句,根据语句长度(也即,语句包含的文字字数)分配视频帧并映射到视频帧上,实现对视频的字幕添加,能够简化操作流程以及缩短开发时间,且降低错误率。

[0046] 在本申请的一个实施例中,字幕文案的文字总字数不包括标点符号的数量。标点符号作为语句之间的分隔符,用于间隔两个语句,其数量不计入字幕文案的文字总字数,使得在显示字幕时,两句话的字幕之间具有一定的时间间隔,实现语句之间的停顿。

[0047] 在本申请的一个实施例中,在将每个语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上之前,还包括:若语句对应的图像帧数不为整数,且图像帧数的小数部分大于或等于0.5,则图像帧数向上取整;若语句对应的图像帧数不为整数,且图像帧数的小数部分小于0.5,则图像帧数向下取整。

[0048] 在该实施例中,如果语句对应的图像帧数不为整数,则对其小数部分进行四舍五入处理。经统计在一篇文案中,每个语句小数部分的四舍、五入是交错、有规律出现的,且四舍与五入出现的频数随着语句数量的增多趋于相等。

[0049] 因此,小数部分的这种取整方法,也作为一种修正机制,增加了字幕与语音节奏一致的稳定性。

[0050] 在本申请的一个实施例中,对于第i个语句,将第i个语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上,具体包括:将第i个语句包含的文字,添加至第一图像与第二图像之间的每帧图片上;其中,第一图像为第i个语句的起始图像的后a个图像,第二图像为第i个语句的结尾图像的前b个图像,a、b均为大于或等于1,且小于或等于4的整数,i为大于或等于1,且小于或等于多个语句的总数量。

[0051] 在该实施例中,在将字幕添加至某一语句图像帧时,从该语句的起始图像的后a个图像开始,到该语句的结尾图像的前b个图像结束,也就是说,在两个语句之间具有a+b帧图像是不添加字幕的,由此使得两句话的字幕之间具有一定的时间间隔。

[0052] 其中,a、b均为大于或等于1,且小于或等于4的整数,也即两个语句之间间隔的图像帧数大于或等于2,且小于或等于8。a、b的值与语句包含的文字字数相关,语句包含的文字字数越多,a、b的值越大,语句包含的文字字数越少,a、b的值越小。

[0053] 这里,需要说明的是,任一语句的对应的图像帧数均大于8。

[0054] 通过上述方式,能够使得两句话的字幕之间具有一定的时间间隔,实现语句之间的停顿,保证更好的视觉感觉。

[0055] 在一个具体实施例中,对于图像帧数为M的短视频、文字总字数为N的字幕文案,将字幕文案添加至短视频的方法包括以下步骤:

[0056] (1)以选定标点符号(例如,句号、问号、叹号、逗号等)为分隔符,将字幕文案拆分为多个语句,并按原文顺序,用自然数对语句编号,共得到Z条语句。

[0057] (2)将短视频拆分为多个图像,按视频播放顺序,用自然数对每一帧图像编号,共计M张图像。

[0058] (3)计算$p_i \times (M/N)$,其中,i=1,2,.....,Z,Z为语句的数量,p_i表示第i条语句包含的字数。当$p_i \times (M/N)$的小数部分≥0.5时,向上取整,反之向下取整,取整后的$p_i \times (M/N)$计作P_i。

[0059] (4)确定第i条语句在短视频中对应的视频帧的起、末位置:

[0060] $i_{begin}=\sum_0^{i-1}P_j$,其中,$P_0=0$

[0061] $i_{end}=\sum_0^{i}P_j$

[0062] 第i条语句在视频帧索引中对应的起、末索引分别为i_{begin}、i_{end}。

[0063] (5)将第i条语句的文本内容,用opencv写入到帧编号处于[$i_{begin+2}$,i_{end-2}]中的每一帧图像上,由此能够使得两句话之间具有一定的时间间隔。

[0064] 执行完如上的所述操作,便完成了短视频每一帧的字幕添加,然后再将所有帧合成为视频,便完成了短视频的字幕添加。

[0065] 进一步地,作为上述视频字幕添加方法的具体实现,本申请实施例提供了一种视频字幕添加装置。如图2所示,该视频字幕添加装置200包括:获取模块201、拆分模块202、字幕添加模块203以及图像合并模块204。

[0066] 其中,获取模块201,用于获取待添加视频和字幕文案;拆分模块202,用于按照视频播放顺序,将待添加视频拆分为多帧图像,以及以标点符号为分隔符,将字幕文案拆分为多个语句;字幕添加模块203,用于确定每个语句对应的图像帧数,以及确定每个语句对应的起始图像和结尾图像,并将每个语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上,得到添加字幕后的多帧目标图像;图像合并模块204,用于按照视频播放顺序,将多帧目标图像进行合并,生成添加字幕后的目标视频。

[0067] 在该实施例中,视频是由一帧一帧的图像按照时间顺序构成,例如,时长10秒帧率为24的视频由240张图像构成,在获取到待添加视频后,按照视频播放顺序,也即图像帧的时间顺序,将待添加视频拆分为M帧图像,M为大于1的整数。以及,获取需要添加为视频字幕的字幕文案,该字幕文案由文字、字母、标点符号等构成,并将字幕文案按照其包含的标点符号拆分为Z个语句,Z为大于1的整数。

[0068] 进一步地,将图像帧数与字幕字数映射起来,从而实现向图像添加字幕,得到添加字幕后的多帧目标图像。最后,再按照视频播放顺序,也即图像帧的时间顺序,将多帧目标图像进行合并,得到添加字幕后的目标视频。

[0069] 在本申请的一个实施例中,目标视频可以为短视频,其视频时长小于或等于3分钟。

[0070] 值得注意的是,在根据每个语句向每帧图像添加字幕时,对于任一语句,确定该语句对应的图像帧数,再确定该语句对应的起始图像和结尾图像,最后将该语句包含的文字,添加至起始图像与结尾图像之间的图像上。也就是说,该语句对应的起始图像和结尾图像上不添加文字,使得每个语句之间至少有两帧图像上没有文字,由此使得两句话的字幕之间具有一定的时间间隔,实现语句之间的停顿。

[0071] 本申请实施例中,一方面,添加至视频的字幕直接来自于原字幕文案,方法操作简单,字幕添加速度快。相比于相关技术中利用语音识别模型的方法,无需收集语音数据训练语音识别模型,避免视频开发过程耗时费力,且降低了视频字幕添加错误率。另一方面,在语句对应的起始图像与结尾图像之间的图像上添加文字,使得两句话的字幕之间具有一定的时间间隔,实现语句之间的停顿,保证更好的视觉感觉。

[0072] 进一步地,字幕添加模块203,具体用于:确定每个语句对应的图像帧数,以及确定每个语句对应的起始图像和结尾图像;将每个语句包含的文字,添加至其对应的起始图像与结尾图像之间的图像上,得到添加字幕后的多帧目标图像。

[0073] 进一步地,字幕添加模块203,具体用于:根据多帧图像的总数量和字幕文案的文字总字数,计算字幕文案的每个文字对应的图像帧数;根据每个文字对应的图像帧数和语句包含的文字字数,计算每个语句对应的图像帧数。

[0074] 进一步地,字幕文案的文字总字数不包括标点符号的数量。

[0075] 进一步地,字幕添加模块203,具体用于:若语句对应的图像帧数不为整数,且图像帧数的小数部分大于或等于0.5,则图像帧数向上取整;若语句对应的图像帧数不为整数,且图像帧数的小数部分小于0.5,则图像帧数向下取整。

[0076] 进一步地,字幕添加模块203,具体用于:将第i个语句包含的文字,添加至第一图像与第二图像之间的每帧图片上;其中,第一图像为第i个语句的起始图像的后a个图像,第二图像为第i个语句的结尾图像的前b个图像,a、b均为大于或等于1,且小于或等于4的整数,i为大于或等于1,且小于或等于多个语句的总数量。

[0077] 进一步地,字幕文案为文案编辑人员所编辑的,且该装置还包括:校正模块,用于对字幕文案进行校正,保留有效文案。

[0078] 进一步地,目标视频的视频时长小于或等于3分钟。

[0079] 本申请实施例中的视频字幕添加装置可以是电子设备,也可以是电子设备中的部件,例如集成电路或芯片。该电子设备可以是终端,也可以为除终端之外的其他设备。示例性的,电子设备可以为手机、平板电脑、笔记本电脑、掌上电脑、车载电子设备、移动上网装置(Mobile Internet Device,MID)、超级移动个人计算机(Ultra-Mobile Personal Computer,UMPC)、上网本或者个人数字助理(Personal Digital Assistant,PDA)等,还可以为服务器、网络附属存储器(Network Attached Storage,NAS)、个人计算机(Personal Computer,PC)、电视机(Television,TV)等,本申请实施例不作具体限定。

[0080] 本申请实施例中的视频字幕添加装置200可以为具有操作系统的装置。该操作系统可以为安卓(Android)操作系统,可以为ios操作系统,还可以为其他可能的操作系统,本申请实施例不作具体限定。

[0081] 本申请实施例提供的视频字幕添加装置200能够实现图1的视频字幕添加方法实施例实现的各个过程,为避免重复,这里不再赘述。

[0082] 本申请实施例还提供一种电子设备,如图3所示,该电子设备300包括处理器301和存储器302,存储器302上存储有可在处理器301上运行的程序或指令,该程序或指令被处理器301执行时实现上述视频字幕添加方法实施例的各个步骤,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0083] 需要说明的是,本申请实施例中的电子设备包括上述的移动电子设备和非移动电子设备。

[0084] 存储器302可用于存储软件程序以及各种数据。存储器302可主要包括存储程序或指令的第一存储区和存储数据的第二存储区,其中,第一存储区可存储操作系统、至少一个功能所需的应用程序或指令(比如声音播放功能、图像播放功能等)等。此外,存储器302可以包括易失性存储器或非易失性存储器,或者,存储器302可以包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器(Read-Only Memory,ROM)、可编程只读存储器(Programmable ROM,PROM)、可擦除可编程只读存储器(Erasable PROM,EPROM)、电可擦除可编程只读存储器(Electrically EPROM,EEPROM)或闪存。易失性存储器可以是随机存取存储器(Random Access Memory,RAM),静态随机存取存储器(Static RAM,SRAM)、动态随机存取存储器(Dynamic RAM,DRAM)、同步动态随机存取存储器(Synchronous DRAM,SDRAM)、双倍数据速率同步动态随机存取存储器(Double Data Rate SDRAM,DDRSDRAM)、增

强型同步动态随机存取存储器(Enhanced SDRAM,ESDRAM)、同步连接动态随机存取存储器(Synch link DRAM,SLDRAM)和直接内存总线随机存取存储器(Direct Rambus RAM,DRRAM)。本申请实施例中的存储器302包括但不限于这些和任意其它适合类型的存储器。

[0085] 处理器301可包括一个或多个处理单元;可选的,处理器301集成应用处理器和调制解调处理器,其中,应用处理器主要处理涉及操作系统、用户界面和应用程序等的操作,调制解调处理器主要处理无线通信信号,如基带处理器。可以理解的是,上述调制解调处理器也可以不集成到处理器301中。

[0086] 本申请实施例还提供一种可读存储介质,可读存储介质上存储有程序或指令,该程序或指令被处理器执行时实现上述视频字幕添加方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0087] 本申请实施例还提供了一种芯片,芯片包括处理器和通信接口,通信接口和处理器耦合,处理器用于运行程序或指令,实现上述视频字幕添加方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0088] 应理解,本申请实施例提到的芯片还可以称为系统级芯片、系统芯片、芯片系统或片上系统芯片等。

[0089] 本申请实施例还提供一种计算机程序产品,该程序产品被存储在存储介质中,该程序产品被至少一个处理器执行以实现如上述视频字幕添加方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0090] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。此外,需要指出的是,本申请实施方式中的方法和装置的范围不限按示出或讨论的顺序来执行功能,还可包括根据所涉及的功能按基本同时的方式或按相反的顺序来执行功能,例如,可以按不同于所描述的次序来执行所描述的方法,并且还可以添加、省去、或组合各种步骤。另外,参照某些示例所描述的特征可在其他示例中被组合。

[0091] 上面结合附图对本申请的实施例进行了描述,但是本申请并不局限于上述的具体实施方式,上述的具体实施方式仅仅是示意性的,而不是限制性的,本领域的普通技术人员在本申请的启示下,在不脱离本申请宗旨和权利要求所保护的范围情况下,还可做出很多形式,均属于本申请的保护之内。

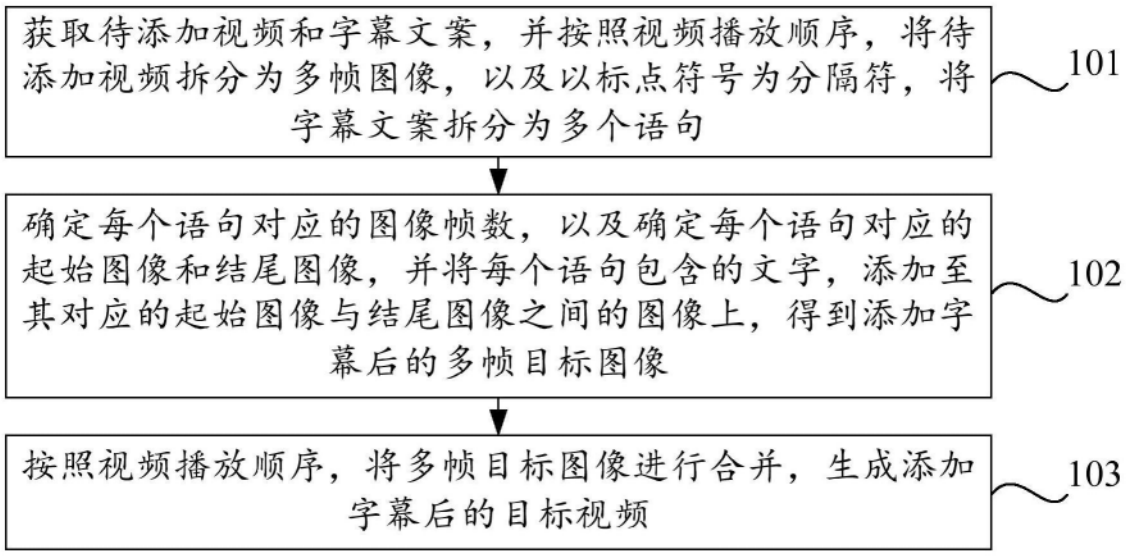

获取待添加视频和字幕文案，并按照视频播放顺序，将待添加视频拆分为多帧图像，以及以标点符号为分隔符，将字幕文案拆分为多个语句
101
确定每个语句对应的图像帧数，以及确定每个语句对应的起始图像和结尾图像，并将每个语句包含的文字，添加至其对应的起始图像与结尾图像之间的图像上，得到添加字幕后的多帧目标图像
102
按照视频播放顺序，将多帧目标图像进行合并，生成添加字幕后的目标视频
103

图1

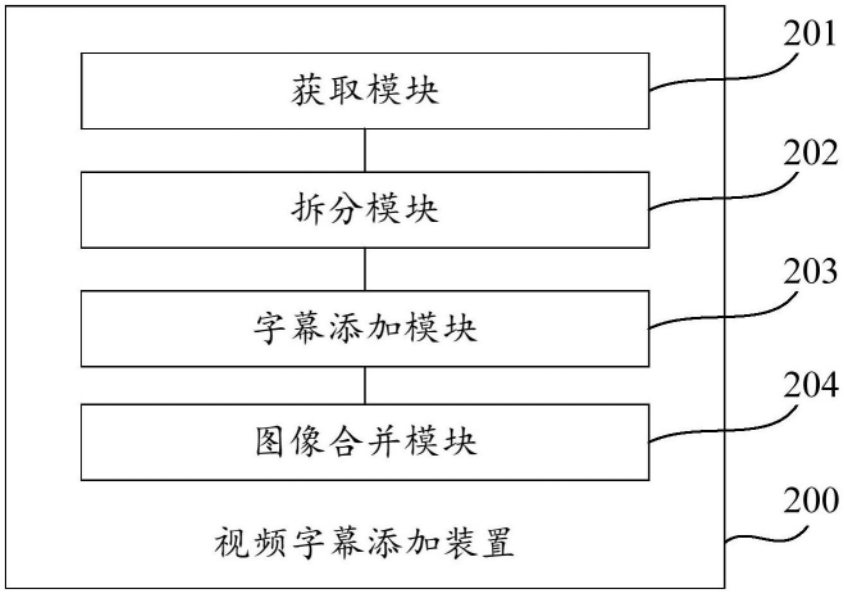

获取模块
201
拆分模块
202
字幕添加模块
203
图像合并模块
204
视频字幕添加装置
200

图2

图3