(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113033552 B
(45)授权公告日 2024.02.02

(21)申请号 202110299717.0

(22)申请日 2021.03.19

(65)同一申请的已公布的文献号
申请公布号 CN 113033552 A

(43)申请公布日 2021.06.25

(73)专利权人 北京字跳网络技术有限公司
地址 100190 北京市海淀区紫金数码园4号楼2层0207

(72)发明人 肖学锋

(74)专利代理机构 泰和泰律师事务所 51219
专利代理师 祝海燕

(51)Int.Cl.
G06V 20/62(2022.01)
G06V 20/40(2022.01)
G06V 10/82(2022.01)
G06N 3/0464(2023.01)
G06N 3/08(2023.01)

(56)对比文件
CN 101729784 A,2010.06.09
CN 107392086 A,2017.11.24
CN 107465911 A,2017.12.12
EP 2860696 A1,2015.04.15
US 2003152271 A1,2003.08.14
WO 2018127539 A1,2018.07.12
Khare, Vijeta,等.A new Histogram Oriented Moments descriptor for multi-oriented moving text detection in video.《Expert Systems with Applications》.2015,(42),7627-7640.
夏利民,等.基于密度轨迹与句法规则的复杂行为识别.《小型微型计算机系统》.2016,(07),239-243.
金红,等.基于内容检索的视频处理技术.《中国图象图形学报》.2000,(04),10-17.

审查员 徐欢欢

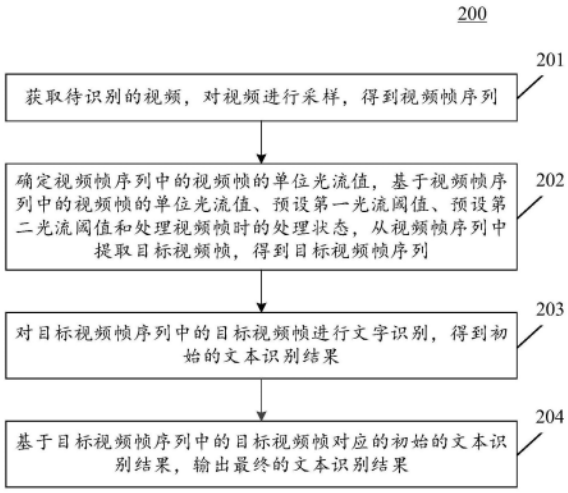
权利要求书5页 说明书15页 附图4页

(54)发明名称

文本识别方法、装置和电子设备

(57)摘要

本公开实施例公开了文本识别方法、装置和电子设备。该方法的一具体实施方式包括:获取待识别的视频,对视频进行采样,得到视频帧序列;确定视频帧序列中的视频帧的单位光流值,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,得到目标视频帧序列;对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。该实施方式提高了视频文本识别的准确性。



CN 113033552 B

1.一种文本识别方法,其特征在于,包括:

获取待识别的视频,对所述视频进行采样,得到视频帧序列,其中,所述视频帧序列中的视频帧按照在所述视频中由前到后的顺序进行排列,所述视频中呈现有文字;

确定所述视频帧序列中的视频帧的单位光流值,基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧,得到目标视频帧序列,其中,所述处理状态包括陷入状态和非陷入状态,所述第一光流阈值用来判断该视频帧是否处于初步静止状态,初步静止状态用于表征视频内容由较快的变化速度变为较慢的变化速度,所述第二光流阈值用来判断该视频帧是否处于绝对静止状态,绝对静止状态用于表征视频内容的变化速度很慢,若该视频帧处于初步静止状态,则处理该视频帧时的处理状态为陷入状态,若该视频帧未处于初步静止状态,则处理该视频帧时的处理状态为非陷入状态;

对所述目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;

基于所述目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

2.根据权利要求1所述的方法,其特征在于,所述目标视频帧序列中的目标视频帧的数目为至少两个;以及

所述基于所述目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果,包括:

针对所述目标视频帧序列中的每组相邻帧,确定从该组相邻帧中识别出的初始的文本识别结果之间的编辑距离,响应于确定出所述编辑距离小于预设编辑距离阈值,从该组相邻帧中选取置信度最高的视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出。

3.根据权利要求1所述的方法,其特征在于,所述基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧,包括:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于所述第一光流阈值,确定该视频帧的单位光流值是否小于预设第二光流阈值;

若是,则从所述视频帧序列中提取出该视频帧。

4.根据权利要求1所述的方法,其特征在于,所述基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧,包括:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于所述第一光流阈值,确定该视频帧的单位光流值是否小于预设第二光流阈值;

若否,则将所述处理状态更改为陷入状态。

5.根据权利要求1所述的方法,其特征在于,所述基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧,包括:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值;

若是,则从所述视频帧序列中提取出该视频帧以及将所述处理状态更改为非陷入状态。

6.根据权利要求1所述的方法,其特征在于,所述基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧,包括:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值;

若否,则基于该视频帧的单位光流值,确定在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

7.根据权利要求6所述的方法,其特征在于,所述基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧,包括:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态、该视频帧的单位光流值大于预设第一光流阈值且该视频帧的前一帧的单位光流值小于所述第一光流阈值,从所述视频帧序列中提取目标视频帧,以及将所述处理状态更改为非陷入状态,其中,所述目标视频帧为在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

8.根据权利要求1所述的方法,其特征在于,所述基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧,包括:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第二光流阈值且该视频帧的前一帧的单位光流值大于所述第二光流阈值,从所述视频帧序列中提取出该视频帧。

9.根据权利要求1所述的方法,其特征在于,所述对所述目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果,包括:

针对所述目标视频帧序列中的目标视频帧,确定该目标视频帧中文本框的位置,利用所述文本框的位置,从该目标视频帧中裁剪出文本区域,从所述文本区域中识别文本,得到初始的文本识别结果。

10.根据权利要求9所述的方法,其特征在于,所述确定该目标视频帧中文本框的位置,包括:

将该目标视频帧的尺寸调整到预设尺寸;

将尺寸调整后的目标视频帧输入预先训练的文本框检测模型中,得到所述尺寸调整后的目标视频帧中文本框的位置信息;

利用文本框在所述尺寸调整后的目标视频帧中的位置信息,确定文本框在该目标视频帧中的位置。

11.根据权利要求9所述的方法,其特征在于,所述从所述文本区域中识别文本,得到初始的文本识别结果,包括:

将所述文本区域输入预先训练的文本识别网络中,得到初始的文本识别结果,其中,所述文本识别网络为卷积神经网络与连续时间序列分类算法相结合的网络框架。

12.一种文本识别装置,其特征在于,包括:

获取单元,用于获取待识别的视频,对所述视频进行采样,得到视频帧序列,其中,所述视频帧序列中的视频帧按照在所述视频中由前到后的顺序进行排列,所述视频中呈现有文字;

提取单元,用于确定所述视频帧序列中的视频帧的单位光流值,基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧,得到目标视频帧序列,其中,所述处理状态包括陷入状态和非陷入状态,所述第一光流阈值用来判断该视频帧是否处于初步静止状态,初步静止状态用于表征视频内容由较快的变化速度变为较慢的变化速度,所述第二光流阈值用来判断该视频帧是否处于绝对静止状态,绝对静止状态用于表征视频内容的变化速度很慢,若该视频帧处于初步静止状态,则处理该视频帧时的处理状态为陷入状态,若该视频帧未处于初步静止状态,则处理该视频帧时的处理状态为非陷入状态;

识别单元,用于对所述目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;

输出单元,用于基于所述目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

13.根据权利要求12所述的装置,其特征在于,所述目标视频帧序列中的目标视频帧的数目为至少两个;以及

所述输出单元进一步用于通过如下方式基于所述目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果:

针对所述目标视频帧序列中的每组相邻帧,确定从该组相邻帧中识别出的初始的文本识别结果之间的编辑距离,响应于确定出所述编辑距离小于预设编辑距离阈值,从该组相邻帧中选取置信度最高的视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出。

14.根据权利要求12所述的装置,其特征在于,所述提取单元进一步用于通过如下方式基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于所述第一光流阈值,确定该视频帧的单位光流值是否小于预设第二光流阈值;

若是,则从所述视频帧序列中提取出该视频帧。

15.根据权利要求12所述的装置,其特征在于,所述提取单元进一步用于通过如下方式基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于所述第一光流阈值,确定该视频帧的单位光流值是否小于预设第二光流阈值;

若否,则将所述处理状态更改为陷入状态。

16.根据权利要求12所述的装置,其特征在于,所述提取单元进一步用于通过如下方式基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值;

若是,则从所述视频帧序列中提取出该视频帧以及将所述处理状态更改为非陷入状态。

17.根据权利要求12所述的装置,其特征在于,所述提取单元进一步用于通过如下方式基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值;

若否,则基于该视频帧的单位光流值,确定在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

18.根据权利要求17所述的装置,其特征在于,所述提取单元进一步用于通过如下方式基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态、该视频帧的单位光流值大于预设第一光流阈值且该视频帧的前一帧的单位光流值小于所述第一光流阈值,从所述视频帧序列中提取目标视频帧,以及将所述处理状态更改为非陷入状态,其中,所述目标视频帧为在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

19.根据权利要求12所述的装置,其特征在于,所述提取单元进一步用于通过如下方式基于所述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从所述视频帧序列中提取目标视频帧:

针对所述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第二光流阈值且该视频帧的前一帧的单位光流值大于所述第二光流阈值,从所述视频帧序列中提取出该视频帧。

20.根据权利要求12所述的装置,其特征在于,所述识别单元进一步用于通过如下方式对所述目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果:

针对所述目标视频帧序列中的目标视频帧,确定该目标视频帧中文本框的位置,利用所述文本框的位置,从该目标视频帧中裁剪出文本区域,从所述文本区域中识别文本,得到初始的文本识别结果。

21.根据权利要求20所述的装置,其特征在于,所述识别单元进一步用于通过如下方式确定该目标视频帧中文本框的位置:

将该目标视频帧的尺寸调整到预设尺寸;

将尺寸调整后的目标视频帧输入预先训练的文本框检测模型中,得到所述尺寸调整后的目标视频帧中文本框的位置信息;

利用文本框在所述尺寸调整后的目标视频帧中的位置信息,确定文本框在该目标视频帧中的位置。

22.根据权利要求20所述的装置,其特征在于,所述识别单元进一步用于通过如下方式从所述文本区域中识别文本,得到初始的文本识别结果:

将所述文本区域输入预先训练的文本识别网络中,得到初始的文本识别结果,其中,所述文本识别网络为卷积神经网络与连续时间序列分类算法相结合的网络框架。

23.一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,其上存储有一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-11中任一所述的方法。

24.一种计算机可读介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现如权利要求1-11中任一所述的方法。

文本识别方法、装置和电子设备

技术领域

[0001] 本公开实施例涉及计算机技术领域,具体涉及文本识别方法、装置和电子设备。

背景技术

[0002] 目前,随着信息化建设的全面开展,文字识别技术已经进入行业应用开发的成熟阶段。在对视频中的文字进行识别的过程中,通常会首先从视频中提取待识别的视频帧,再对待识别的视频帧中的文字进行识别。因此,如何从视频中提取待识别的视频帧,是视频文字识别的关键步骤。

发明内容

[0003] 提供该公开内容部分以便以简要的形式介绍构思,这些构思将在后面的具体实施方式部分被详细描述。该公开内容部分并不旨在标识要求保护的技术方案的关键特征或必要特征,也不旨在用于限制所要求的保护的技术方案的范围。

[0004] 本公开实施例提供了一种文本识别方法、装置和电子设备,提高了视频文本识别的准确性。

[0005] 第一方面,本公开实施例提供了一种文本识别方法,该方法包括:获取待识别的视频,对视频进行采样,得到视频帧序列,其中,视频帧序列中的视频帧按照在视频中由前到后的顺序进行排列,视频中呈现有文字;确定视频帧序列中的视频帧的单位光流值,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,得到目标视频帧序列,其中,处理状态包括陷入状态和非陷入状态;对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

[0006] 第二方面,本公开实施例提供了一种文本识别装置,该装置包括:获取单元,用于获取待识别的视频,对视频进行采样,得到视频帧序列,其中,视频帧序列中的视频帧按照在视频中由前到后的顺序进行排列,视频中呈现有文字;提取单元,用于确定视频帧序列中的视频帧的单位光流值,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,得到目标视频帧序列,其中,处理状态包括陷入状态和非陷入状态;识别单元,用于对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;输出单元,用于基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

[0007] 第三方面,本公开实施例提供了一种电子设备,包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如第一方面所述的文本识别方法。

[0008] 第四方面,本公开实施例提供了一种计算机可读介质,其上存储有计算机程序,该程序被处理器执行时实现如第一方面所述的文本识别方法的步骤。

[0009] 本公开实施例提供的文本识别方法、装置和电子设备，通过首先获取待识别的视频，对上述视频进行采样，得到视频帧序列；之后，确定上述视频帧序列中的视频帧的单位光流值，基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从上述视频帧序列中提取目标视频帧，得到目标视频帧序列；而后，对上述目标视频帧序列中的目标视频帧进行文字识别，得到初始的文本识别结果；最后，基于上述目标视频帧序列中的目标视频帧对应的初始的文本识别结果，输出最终的文本识别结果。通过稠密光流计算的这种可以从视频中抽取出较为清晰的视频帧，对这些较为清晰的视频帧进行文字识别，提高了视频文本识别的准确性。

附图说明

[0010] 结合附图并参考以下具体实施方式，本公开各实施例的上述和其他特征、优点及方面将变得更加明显。贯穿附图中，相同或相似的附图标记表示相同或相似的元素。应当理解附图是示意性的，原件和元素不一定按照比例绘制。

[0011] 图1是本公开的各个实施例可以应用于其中的示例性系统架构图；

[0012] 图2是根据本公开的文本识别方法的一个实施例的流程图；

[0013] 图3是根据本公开的文本识别方法的又一个实施例的流程图；

[0014] 图4是根据本公开的文本识别装置的一个实施例的结构示意图；

[0015] 图5是适于用来实现本公开实施例的电子设备的计算机系统的结构示意图。

具体实施方式

[0016] 下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

[0017] 应当理解，本公开的方法实施方式中记载的各个步骤可以按照不同的顺序执行，和/或并行执行。此外，方法实施方式可以包括附加的步骤和/或省略执行示出的步骤。本公开的范围在此方面不受限制。

[0018] 本文使用的术语“包括”及其变形是开放性包括，即“包括但不限于”。术语“基于”是“至少部分地基于”。术语“一个实施例”表示“至少一个实施例”；术语“另一实施例”表示“至少一个另外的实施例”；术语“一些实施例”表示“至少一些实施例”。其他术语的相关定义将在下文描述中给出。

[0019] 需要注意，本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分，并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

[0020] 需要注意，本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，本领域技术人员应当理解，除非在上下文另有明确指出，否则应该理解为“一个或多个”。

[0021] 本公开实施方式中的多个装置之间所交互的消息或者信息的名称仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

[0022] 图1示出了可以应用本公开的文本识别方法的实施例的示例性系统架构100。

[0023] 如图1所示，系统架构100可以包括摄像头101，网络1021、1022、1023，终端设备103

和服务器104。网络1021用以在摄像头101和终端设备103之间提供通信链路的介质。网络1022用以在摄像头101和服务器104之间提供通信链路的介质。网络1023用以在终端设备103和服务器104之间提供通信链路的介质。网络1021、1022、1023可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

[0024] 摄像头101又称为电脑相机、电脑眼、电子眼等，是一种视频输入设备，被广泛地应用于视频会议、实时监控等各个方面。在这里，摄像头101也可以为无人机的摄像头。

[0025] 终端设备103可以通过网络1021与摄像头101交互，以发送或接收消息等，例如，终端设备103可以从摄像头101中获取待识别的视频。终端设备103可以通过网络1023与服务器104交互，以发送或接收消息等，例如，服务器104可以从终端设备103中获取待识别的视频。终端设备103上可以安装有各种通讯客户端应用，例如视频拍摄类应用、视频处理类应用、即时通讯软件等。

[0026] 终端设备103可以从摄像头101中获取待识别的视频，对上述视频进行采样，得到视频帧序列；之后，可以确定上述视频帧序列中的视频帧的单位光流值，基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从上述视频帧序列中提取目标视频帧，得到目标视频帧序列；而后，可以对上述目标视频帧序列中的目标视频帧进行文字识别，得到初始的文本识别结果；最后，可以基于上述目标视频帧序列中的目标视频帧对应的初始的文本识别结果，输出最终的文本识别结果。

[0027] 终端设备103可以是硬件，也可以是软件。当终端设备103为硬件时，可以是具有摄像头并且支持信息交互的各种电子设备，包括但不限于智能手机、平板电脑、膝上型便携计算机等。当终端设备103为软件时，可以安装在上述所列举的电子设备中。其可以实现成多个软件或软件模块(例如用来提供分布式服务的多个软件或软件模块)，也可以实现成单个软件或软件模块。在此不做具体限定。

[0028] 服务器104可以是提供各种服务的服务器。例如，可以是对视频中的文本进行识别的服务器。服务器104可以首先从摄像头101获取待识别的视频，或者从终端设备103获取待识别的视频，对上述视频进行采样，得到视频帧序列；之后，可以确定上述视频帧序列中的视频帧的单位光流值，基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从上述视频帧序列中提取目标视频帧，得到目标视频帧序列；而后，可以对上述目标视频帧序列中的目标视频帧进行文字识别，得到初始的文本识别结果；最后，可以基于上述目标视频帧序列中的目标视频帧对应的初始的文本识别结果，输出最终的文本识别结果。

[0029] 需要说明的是，服务器104可以是硬件，也可以是软件。当服务器104为硬件时，可以实现成多个服务器组成的分布式服务器集群，也可以实现成单个服务器。当服务器104为软件时，可以实现成多个软件或软件模块(例如用来提供分布式服务)，也可以实现成单个软件或软件模块。在此不做具体限定。

[0030] 需要说明的是，本公开实施例所提供的文本识别方法可以由服务器104执行，则文本识别装置可以设置于服务器104中。本公开实施例所提供的文本识别方法也可以由终端设备103执行，则文本识别装置可以设置于终端设备103中。

[0031] 还需要说明的是，在本公开实施例所提供的文本识别方法由服务器104执行的情

况下,若服务器104可以从摄像头101中获取待识别的视频,此时示例性系统架构100可以不存在网络1021、1023和终端设备103。若服务器104可以从终端设备103中获取待识别的视频,此时示例性系统架构100可以不存在网络1021、1022和摄像头101。若服务器104的本地可以存储有待识别的视频,此时示例性系统架构100可以不存在网络1021、1022、1023,摄像头101和终端设备103。

[0032] 还需要说明的是,在本公开实施例所提供的文本识别方法由终端设备103执行的情况下,若终端设备103可以从摄像头101中获取待识别的视频,此时示例性系统架构100可以不存在网络1022、1023和服务器104。若终端设备103可以从服务器104中获取待识别的视频,此时示例性系统架构100可以不存在网络1021、1022和摄像头101。若终端设备103的本地可以存储有待识别的视频,此时示例性系统架构100可以不存在网络1021、1022、1023,摄像头101和服务器104。

[0033] 应该理解,图1中的摄像头、网络、终端设备和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的摄像头、网络、终端设备和服务器。

[0034] 继续参考图2,示出了根据本公开的文本识别方法的一个实施例的流程200。该文本识别方法,包括以下步骤:

[0035] 步骤201,获取待识别的视频,对视频进行采样,得到视频帧序列。

[0036] 在本实施例中,文本识别方法的执行主体(例如,图1中的终端设备103或服务器104)可以获取待识别的视频,对视频进行采样,得到视频帧序列。上述执行主体可以按照预设的采样率对上述视频进行采样。采样率也可以称为采样频率或者采样速度,通常指的是每秒从连续信号(这里为视频)中提取并组成离散信号的采样个数。

[0037] 在这里,上述视频帧序列中的视频帧可以按照在上述视频中由前到后的顺序进行排列。上述视频中通常呈现有文字。

[0038] 步骤202,确定视频帧序列中的视频帧的单位光流值,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,得到目标视频帧序列。

[0039] 在本实施例中,上述执行主体可以确定上述视频帧序列中的视频帧的单位光流值。具体地,针对上述视频帧序列中的视频帧,上述执行主体可以对该视频帧进行稠密光流(Dense Optical Flow)计算,得到该视频帧中各个像素点的光流值。稠密光流是一种针对图像进行逐点匹配的图像配准方法,稠密光流计算图像上所有的点的偏移量,从而形成一个稠密的光流场。在这里,可以将该视频帧与该视频帧的前一帧进行逐点匹配,从而计算该视频帧上所有像素点相对于前一帧的相应像素点的偏移量,得到该视频帧中各个像素点的光流值。而后,上述执行主体可以确定该视频帧中各个像素点的光流值的平方和,可以将上述平方和与该视频帧的面积的比值确定为该视频帧的单位光流值。

[0040] 之后,上述执行主体可以基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧,得到目标视频帧序列。上述目标视频帧通常是上述视频帧序列中较为清晰的视频帧。

[0041] 上述第一光流阈值可以用来判断该视频帧是否处于初步静止状态,初步静止状态可以用于表征视频内容由一个较快的变化速度变为一个较慢的变化速度。若该视频帧的单

位光流值小于上述第一光流阈值且该视频帧的前一帧的单位光流值大于上述第一光流阈值,则可以说明该视频帧处于初步静止状态。

[0042] 上述第二光流阈值可以用来判断该视频帧是否处于绝对静止状态,绝对静止状态可以用于表征视频内容的变化速度很慢。若该视频帧的单位光流值小于上述第二光流阈值,则可以说明该视频帧处于绝对静止状态。

[0043] 处理视频帧时的处理状态可以包括陷入(trap)状态和非陷入状态。若该视频帧处于初步静止状态,则处理该视频帧时的处理状态可以为陷入状态。若该视频帧未处于初步静止状态,则处理该视频帧时的处理状态可以为非陷入状态。

[0044] 步骤203,对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果。

[0045] 在本实施例中,上述执行主体可以对上述目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果。作为示例,针对上述目标视频帧序列中的每个目标视频帧,上述执行主体可以将该目标视频帧输入预先训练的文本识别模型中,得到该目标视频帧中的文本识别结果作为初始的文本识别结果。上述文本识别模型可以用于表征帧与帧中的文本识别结果之间的对应关系。

[0046] 步骤204,基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

[0047] 在本实施例中,上述执行主体可以基于上述目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。作为示例,上述执行主体可以将上述目标视频帧序列中的目标视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出。

[0048] 本公开的上述实施例提供的方法通过对视频帧序列中的相邻两帧进行稠密光流计算,从而可以从视频中抽取出较为清晰的视频帧,对这些较为清晰的视频帧进行文字识别,提高了视频文本识别的准确性。

[0049] 在一些可选的实现方式中,上述执行主体可以通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧:针对上述视频帧序列中的视频帧,上述执行主体可以确定处理该视频帧时的处理状态是否为非陷入状态,确定该视频帧的单位光流值是否小于预设第一光流阈值,以及确定该视频帧的前一帧的单位光流值是否大于上述第一光流阈值。若确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于上述第一光流阈值且该视频帧的前一帧的单位光流值大于第一光流阈值,则上述执行主体可以确定该视频帧的单位光流值是否小于预设第二光流阈值。即在确定出该视频帧处于初步静止状态的情况下,确定该视频帧是否处于绝对静止状态。若确定出该视频帧的单位光流值小于上述第二光流阈值,则上述执行主体可以从上述视频帧序列中提取出该视频帧。

[0050] 在一些可选的实现方式中,上述执行主体可以通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧:针对上述视频帧序列中的视频帧,上述执行主体可以确定处理该视频帧时的处理状态是否为非陷入状态,确定该视频帧的单位光流值是否小于预设第一光流阈值,以及确定该视频帧的前一帧的单位光流值是否大于上述第一光

流阈值。若确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于上述第一光流阈值且该视频帧的前一帧的单位光流值大于第一光流阈值，则上述执行主体可以确定该视频帧的单位光流值是否小于预设第二光流阈值。即在确定出该视频帧处于初步静止状态的情况下，确定该视频帧是否处于绝对静止状态。若确定出该视频帧的单位光流值大于等于上述第二光流阈值，则上述执行主体可以将上述处理状态更改为陷入状态。若该视频帧的单位光流值小于上述第一光流阈值且该视频帧的前一帧的单位光流值大于上述第一光流阈值，则可以说明该视频帧处于初步静止状态，若该视频帧处于初步静止状态，则处理该视频帧时的处理状态为陷入状态，因此，将上述处理状态更改为陷入状态。

[0051] 在一些可选的实现方式中，上述执行主体可以通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从上述视频帧序列中提取目标视频帧：针对上述视频帧序列中的视频帧，上述执行主体可以确定处理该视频帧时的处理状态是否为陷入状态。若处理该视频帧时的处理状态为陷入状态，则上述执行主体可以确定该视频帧的单位光流值是否小于预设第二光流阈值。若该视频帧的单位光流值小于上述第二光流阈值，则上述执行主体可以从上述视频帧序列中提取出该视频帧，以及可以将上述处理状态更改为非陷入状态。

[0052] 在一些可选的实现方式中，上述执行主体可以通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从视频帧序列中提取目标视频帧：针对上述视频帧序列中的视频帧，上述执行主体可以确定处理该视频帧时的处理状态是否为陷入状态。若处理该视频帧时的处理状态为陷入状态，则上述执行主体可以确定该视频帧的单位光流值是否小于预设第二光流阈值。若该视频帧的单位光流值大于等于上述第二光流阈值，则上述执行主体可以基于该视频帧的单位光流值，确定在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。随着在视频帧序列中按顺序选取视频帧，如果处理状态一直为陷入状态，上述执行主体可以记录在陷入状态阶段所处理的视频帧中的最小单位光流值和最小单位光流值对应的视频帧。之后，上述执行主体可以将该视频帧的单位光流值与所记录的最小单位光流值进行比较。若该视频帧的单位光流值小于所记录的最小单位光流值，则可以利用该视频帧的单位光流值替换所记录的最小单位光流值。

[0053] 在一些可选的实现方式中，上述执行主体可以通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从视频帧序列中提取目标视频帧：针对上述视频帧序列中的视频帧，上述执行主体可以确定处理该视频帧时的处理状态是否为陷入状态，确定该视频帧的单位光流值是否大于预设第一光流阈值，以及确定该视频帧的前一帧的单位光流值是否小于上述第一光流阈值。若确定出处理该视频帧时的处理状态为陷入状态、该视频帧的单位光流值大于上述第一光流阈值且该视频帧的前一帧的单位光流值小于上述第一光流阈值，则上述执行主体可以从上述视频帧序列中提取目标视频帧，以及将上述处理状态更改为非陷入状态。在这里，上述目标视频帧可以为在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

[0054] 在一些可选的实现方式中，上述执行主体可以通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从视频帧序列中提取目标视频帧：针对上述视频帧序列中的视频帧，上述执行主体可以确

定处理该视频帧时的处理状态是否为非陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值,以及确定该视频帧的前一帧的单位光流值是否大于上述第二光流阈值。若确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第二光流阈值且该视频帧的前一帧的单位光流值大于第二光流阈值,则上述执行主体可以从上述视频帧序列中提取出该视频帧。

[0055] 在一些可选的实现方式中,上述执行主体可以通过如下方式对上述目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果:针对上述目标视频帧序列中的目标视频帧,上述执行主体可以确定该目标视频帧中文本框的位置。在这里,上述执行主体可以将该目标视频帧输入预先训练的文本框检测模型中,得到该目标视频帧中文本框的位置信息。上述文本框检测模型可以用于表征帧与帧中文本框的位置信息之间的对应关系。之后,上述执行主体可以利用上述文本框的位置,从该目标视频帧中裁剪出文本区域。而后,上述执行主体可以从上述文本区域中识别文本,得到初始的文本识别结果。在这里,上述执行主体可以利用OCR(Optical Character Recognition,光学字符识别)方式从上述文本区域中识别文本。

[0056] 在一些可选的实现方式中,上述执行主体可以通过如下方式确定该目标视频帧中文本框的位置:上述执行主体可以将该目标视频帧的尺寸调整到预设尺寸。上述预设尺寸通常为适合被文本框检测模型进行处理的图像的尺寸。之后,上述执行主体可以将尺寸调整后的目标视频帧输入预先训练的文本框检测模型中,得到上述尺寸调整后的目标视频帧中文本框的位置信息。上述文本框检测模型可以用于表征帧与帧中文本框的位置信息之间的对应关系。而后,上述执行主体可以利用文本框在上述尺寸调整后的目标视频帧中的位置信息,确定文本框在该目标视频帧中的位置。即上述执行主体可以将上述尺寸调整后的目标视频帧中的文本框映射到该目标视频帧中。

[0057] 在一些可选的实现方式中,上述执行主体可以通过如下方式从上述文本区域中识别文本,得到初始的文本识别结果:上述执行主体可以将上述文本区域输入预先训练的文本识别网络中,得到初始的文本识别结果。在这里,上述文本识别网络可以为卷积神经网络(Convolutional Neural Network,CNN)与连续时间序列分类算法(Connectionist Temporal Classification,CTC)相结合的网络框架。上述文本识别网络可以用于表征文本区域与文本区域中的文本识别结果之间的对应关系。

[0058] 进一步参考图3,其示出了文本识别方法的又一个实施例的流程300。该文本识别方法的流程300,包括以下步骤:

[0059] 步骤301,获取待识别的视频,对视频进行采样,得到视频帧序列。

[0060] 步骤302,确定视频帧序列中的视频帧的单位光流值,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,得到目标视频帧序列。

[0061] 步骤303,对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果。

[0062] 在本实施例中,步骤301-303可以按照与步骤201-203类似的方式执行,在此不再赘述。

[0063] 步骤304,针对目标视频帧序列中的每组相邻帧,确定从该组相邻帧中识别出的初

始的文本识别结果之间的编辑距离,响应于确定出编辑距离小于预设编辑距离阈值,从该组相邻帧中选取置信度最高的视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出。

[0064] 在本实施例中,针对上述目标视频帧序列中的每组相邻帧,文本识别方法的执行主体(例如,图1中的终端设备103或服务器104)可以确定从该组相邻帧中识别出的初始的文本识别结果之间的编辑距离。相邻帧可以由上述目标视频帧序列中的两个位置相邻的视频帧所组成。

[0065] 设A和B是两个字符串,将字符串A变换为字符串B所需要的最少字符操作数可以称为字符串A到字符串B的编辑距离。字符操作包括:删除一个字符、插入一个字符以及将一个字符改写为另一个字符。作为示例,若字符串A=abc,字符串B=abf,则在将字符串A变换为字符串B仅需将字符c修改为字符f,所以字符串A到字符串B的编辑距离为1。

[0066] 在这里,若相邻帧为视频帧M和视频帧N,从视频帧M中识别出的初始的文本识别结果为字符串m,从视频帧N中识别出的初始的文本识别结果为字符串n,则上述执行主体可以确定字符串m到字符串n的编辑距离。

[0067] 之后,上述执行主体可以确定上述编辑距离是否小于预设编辑距离阈值。若上述编辑距离小于上述编辑距离阈值,则上述执行主体可以从该组相邻帧中选取置信度最高的视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出。

[0068] 作为示例,若视频帧中的文本识别结果是利用预先训练的文本识别网络识别出的,则上述文本识别网络在输出视频帧中的文本识别结果的同时通常也会输出该文本识别结果对应的概率,这个概率通常可以表征从视频帧中识别出该文本识别结果的置信度。

[0069] 从图3中可以看出,与图2对应的实施例相比,本实施例中的文本识别方法的流程300体现了确定相邻帧的文本识别结果之间的编辑距离,若编辑距离小于预设编辑距离阈值,从相邻帧中选取置信度最高的视频帧对应的文本识别结果进行输出的步骤。由此,本实施例描述的方案可以进一步提高视频文本识别的准确性。

[0070] 进一步参考图4,作为对上述各图所示方法的实现,本公开提供了一种文本识别装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0071] 如图4所示,本实施例的文本识别装置400包括:获取单元401、提取单元402、识别单元403和输出单元404。其中,获取单元401用于获取待识别的视频,对视频进行采样,得到视频帧序列,其中,视频帧序列中的视频帧按照在视频中由前到后的顺序进行排列,视频中呈现有文字;提取单元402用于确定视频帧序列中的视频帧的单位光流值,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,得到目标视频帧序列,其中,处理状态包括陷入状态和非陷入状态;识别单元403用于对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;输出单元404用于基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

[0072] 在本实施例中,文本识别装置400的获取单元401、提取单元402、识别单元403和输出单元404的具体处理可以参考图2对应实施例中的步骤201、步骤202、步骤203和步骤204。

[0073] 在一些可选的实现方式中,上述目标视频帧序列中的目标视频帧的数目可以为至

少两个;以及上述输出单元404可以进一步用于通过如下方式基于上述目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果:针对上述目标视频帧序列中的每组相邻帧,上述输出单元404可以确定从该组相邻帧中识别出的初始的文本识别结果之间的编辑距离,响应于确定出上述编辑距离小于预设编辑距离阈值,上述输出单元404可以从该组相邻帧中选取置信度最高的视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出。

[0074] 在一些可选的实现方式中,上述提取单元402可以进一步用于通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧:针对上述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于上述第一光流阈值,上述提取单元402可以确定该视频帧的单位光流值是否小于预设第二光流阈值;若是,则上述提取单元402可以从上述视频帧序列中提取出该视频帧。

[0075] 在一些可选的实现方式中,上述提取单元402可以进一步用于通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧:针对上述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于上述第一光流阈值,上述提取单元402可以确定该视频帧的单位光流值是否小于预设第二光流阈值;若否,则上述提取单元402可以将上述处理状态更改为陷入状态。

[0076] 在一些可选的实现方式中,上述提取单元402可以进一步用于通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧:针对上述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,上述提取单元402可以确定该视频帧的单位光流值是否小于预设第二光流阈值;若是,则上述提取单元402可以从上述视频帧序列中提取出该视频帧以及将上述处理状态更改为非陷入状态。

[0077] 在一些可选的实现方式中,上述提取单元402可以进一步用于通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧:针对上述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,上述提取单元402可以确定该视频帧的单位光流值是否小于预设第二光流阈值;若否,则上述提取单元402可以基于该视频帧的单位光流值,确定在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

[0078] 在一些可选的实现方式中,上述提取单元402可以进一步用于通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧:针对上述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态、该视频帧的单位光流值大于预设第一光流阈值且该视频帧的前一帧的单位光流值小于上述第一光流阈值,上述提取单元402可以从上述视频帧序列中提取目标视频帧,以及将上述处理状态更改为非陷入状态,

其中,上述目标视频帧为在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

[0079] 在一些可选的实现方式中,上述提取单元402可以进一步用于通过如下方式基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧:针对上述视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第二光流阈值且该视频帧的前一帧的单位光流值大于上述第二光流阈值,上述提取单元402可以从上述视频帧序列中提取出该视频帧。

[0080] 在一些可选的实现方式中,上述识别单元403可以进一步用于通过如下方式对上述目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果:针对上述目标视频帧序列中的目标视频帧,上述识别单元403可以确定该目标视频帧中文本框的位置,利用上述文本框的位置,从该目标视频帧中裁剪出文本区域,从上述文本区域中识别文本,得到初始的文本识别结果。

[0081] 在一些可选的实现方式中,上述识别单元403可以进一步用于通过如下方式确定该目标视频帧中文本框的位置:上述识别单元403可以将该目标视频帧的尺寸调整到预设尺寸;之后,可以将尺寸调整后的目标视频帧输入预先训练的文本框检测模型中,得到上述尺寸调整后的目标视频帧中文本框的位置信息;而后,可以利用文本框在上述尺寸调整后的目标视频帧中的位置信息,确定文本框在该目标视频帧中的位置。

[0082] 在一些可选的实现方式中,上述识别单元403可以进一步用于通过如下方式从上述文本区域中识别文本,得到初始的文本识别结果:上述识别单元403可以将上述文本区域输入预先训练的文本识别网络中,得到初始的文本识别结果,其中,上述文本识别网络为卷积神经网络与连续时间序列分类算法相结合的网络框架。

[0083] 下面参考图5,其示出了适于用来实现本公开的实施例的电子设备(例如图1中的服务器或终端设备)500的结构示意图。本公开的实施例中的终端设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、PDA(个人数字助理)、PAD(平板电脑)、PMP(便携式多媒体播放器)、车载终端(例如车载导航终端)等等的移动终端以及诸如数字TV、台式计算机等等的固定终端。图5示出的电子设备仅仅是一个示例,不应对本公开的实施例的功能和使用范围带来任何限制。

[0084] 如图5所示,电子设备500可以包括处理装置(例如中央处理器、图形处理器等)501,其可以根据存储在只读存储器(ROM)502中的程序或者从存储装置508加载到随机访问存储器(RAM)503中的程序而执行各种适当的动作和处理。在RAM 503中,还存储有电子设备500操作所需的各种程序和数据。处理装置501、ROM 502以及RAM 503通过总线504彼此相连。输入/输出(I/O)接口505也连接至总线504。

[0085] 通常,以下装置可以连接至I/O接口505:包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置506;包括例如液晶显示器(LCD)、扬声器、振动器等的输出装置507;包括例如磁带、硬盘等的存储装置508;以及通信装置509。通信装置509可以允许电子设备500与其他设备进行无线或有线通信以交换数据。虽然图5示出了具有各种装置的电子设备500,但是应理解的是,并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。图5中示出的每个方框可以代表一个装置,也可以根

据需要代表多个装置。

[0086] 特别地,根据本公开的实施例,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,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信装置509从网络上被下载和安装,或者从存储装置508被安装,或者从ROM 502被安装。在该计算机程序被处理装置501执行时,执行本公开的实施例的方法中限定的上述功能。需要说明的是,本公开的实施例所述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开的实施例中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开的实施例中,计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:电线、光缆、RF(射频)等等,或者上述的任意合适的组合。

[0087] 上述计算机可读介质可以是上述电子设备中所包含的;也可以是单独存在,而未装配入该电子设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被该电子设备执行时,使得该电子设备:获取待识别的视频,对上述视频进行采样,得到视频帧序列,其中,上述视频帧序列中的视频帧按照在上述视频中由前到后的顺序进行排列,上述视频中呈现有文字;确定上述视频帧序列中的视频帧的单位光流值,基于上述视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从上述视频帧序列中提取目标视频帧,得到目标视频帧序列,其中,上述处理状态包括陷入状态和非陷入状态;对上述目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;基于上述目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

[0088] 可以以一种或多种程序设计语言或其组合来编写用于执行本公开的实施例的操作的计算机程序代码,所述程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0089] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0090] 根据本公开的一个或多个实施例,提供了一种文本识别方法,该方法包括:获取待识别的视频,对视频进行采样,得到视频帧序列,其中,视频帧序列中的视频帧按照在视频中由前到后的顺序进行排列,视频中呈现有文字;确定视频帧序列中的视频帧的单位光流值,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,得到目标视频帧序列,其中,处理状态包括陷入状态和非陷入状态;对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

[0091] 根据本公开的一个或多个实施例,目标视频帧序列中的目标视频帧的数目为至少两个;以及基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果,包括:针对目标视频帧序列中的每组相邻帧,确定从该组相邻帧中识别出的初始的文本识别结果之间的编辑距离,响应于确定出编辑距离小于预设编辑距离阈值,从该组相邻帧中选取置信度最高的视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出。

[0092] 根据本公开的一个或多个实施例,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,包括:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于第一光流阈值,确定该视频帧的单位光流值是否小于预设第二光流阈值;若是,则从视频帧序列中提取出该视频帧。

[0093] 根据本公开的一个或多个实施例,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,包括:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于第一光流阈值,确定该视频帧的单位光流值是否小于预设第二光流阈值;若否,则将处理状态更改为陷入状态。

[0094] 根据本公开的一个或多个实施例,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,包括:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值;若是,则从视频帧序列

中提取出该视频帧以及将处理状态更改为非陷入状态。

[0095] 根据本公开的一个或多个实施例,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,包括:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值;若否,则基于该视频帧的单位光流值,确定在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

[0096] 根据本公开的一个或多个实施例,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,包括:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态、该视频帧的单位光流值大于预设第一光流阈值且该视频帧的前一帧的单位光流值小于第一光流阈值,从视频帧序列中提取目标视频帧,以及将处理状态更改为非陷入状态,其中,目标视频帧为在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

[0097] 根据本公开的一个或多个实施例,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,包括:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第二光流阈值且该视频帧的前一帧的单位光流值大于第二光流阈值,从视频帧序列中提取出该视频帧。

[0098] 根据本公开的一个或多个实施例,对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果,包括:针对目标视频帧序列中的目标视频帧,确定该目标视频帧中文本框的位置,利用文本框的位置,从该目标视频帧中裁剪出文本区域,从文本区域中识别文本,得到初始的文本识别结果。

[0099] 根据本公开的一个或多个实施例,确定该目标视频帧中文本框的位置,包括:将该目标视频帧的尺寸调整到预设尺寸;将尺寸调整后的目标视频帧输入预先训练的文本框检测模型中,得到尺寸调整后的目标视频帧中文本框的位置信息;利用文本框在尺寸调整后的目标视频帧中的位置信息,确定文本框在该目标视频帧中的位置。

[0100] 根据本公开的一个或多个实施例,从文本区域中识别文本,得到初始的文本识别结果,包括:将文本区域输入预先训练的文本识别网络中,得到初始的文本识别结果,其中,文本识别网络为卷积神经网络与连续时间序列分类算法相结合的网络框架。

[0101] 根据本公开的一个或多个实施例,提供了一种文本识别装置,该装置包括:获取单元,用于获取待识别的视频,对视频进行采样,得到视频帧序列,其中,视频帧序列中的视频帧按照在视频中由前到后的顺序进行排列,视频中呈现有文字;提取单元,用于确定视频帧序列中的视频帧的单位光流值,基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧,得到目标视频帧序列,其中,处理状态包括陷入状态和非陷入状态;识别单元,用于对目标视频帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果;输出单元,用于基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果。

[0102] 根据本公开的一个或多个实施例,目标视频帧序列中的目标视频帧的数目为至少两个;以及输出单元进一步用于通过如下方式基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果,输出最终的文本识别结果:针对目标视频帧序列中的每组相邻帧,确

定从该组相邻帧中识别出的初始的文本识别结果之间的编辑距离,响应于确定出编辑距离小于预设编辑距离阈值,从该组相邻帧中选取置信度最高的视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出。

[0103] 根据本公开的一个或多个实施例,提取单元进一步用于通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于第一光流阈值,确定该视频帧的单位光流值是否小于预设第二光流阈值;若是,则从视频帧序列中提取出该视频帧。

[0104] 根据本公开的一个或多个实施例,提取单元进一步用于通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第一光流阈值且该视频帧的前一帧的单位光流值大于第一光流阈值,确定该视频帧的单位光流值是否小于预设第二光流阈值;若否,则将处理状态更改为陷入状态。

[0105] 根据本公开的一个或多个实施例,提取单元进一步用于通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值;若是,则从视频帧序列中提取出该视频帧以及将处理状态更改为非陷入状态。

[0106] 根据本公开的一个或多个实施例,提取单元进一步用于通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态,确定该视频帧的单位光流值是否小于预设第二光流阈值;若否,则基于该视频帧的单位光流值,确定在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

[0107] 根据本公开的一个或多个实施例,提取单元进一步用于通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为陷入状态、该视频帧的单位光流值大于预设第一光流阈值且该视频帧的前一帧的单位光流值小于第一光流阈值,从视频帧序列中提取目标视频帧,以及将处理状态更改为非陷入状态,其中,目标视频帧为在陷入状态阶段所处理的视频帧中最小单位光流值对应的视频帧。

[0108] 根据本公开的一个或多个实施例,提取单元进一步用于通过如下方式基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态,从视频帧序列中提取目标视频帧:针对视频帧序列中的视频帧,响应于确定出处理该视频帧时的处理状态为非陷入状态、该视频帧的单位光流值小于预设第二光流阈值且该视频帧的前一帧的单位光流值大于第二光流阈值,从视频帧序列中提取出该视频帧。

[0109] 根据本公开的一个或多个实施例,识别单元进一步用于通过如下方式对目标视频

帧序列中的目标视频帧进行文字识别,得到初始的文本识别结果:针对目标视频帧序列中的目标视频帧,确定该目标视频帧中文本框的位置,利用文本框的位置,从该目标视频帧中裁剪出文本区域,从文本区域中识别文本,得到初始的文本识别结果。

[0110] 根据本公开的一个或多个实施例,识别单元进一步用于通过如下方式确定该目标视频帧中文本框的位置:将该目标视频帧的尺寸调整到预设尺寸;将尺寸调整后的目标视频帧输入预先训练的文本框检测模型中,得到尺寸调整后的目标视频帧中文本框的位置信息;利用文本框在尺寸调整后的目标视频帧中的位置信息,确定文本框在该目标视频帧中的位置。

[0111] 根据本公开的一个或多个实施例,识别单元进一步用于通过如下方式从文本区域中识别文本,得到初始的文本识别结果:将文本区域输入预先训练的文本识别网络中,得到初始的文本识别结果,其中,文本识别网络为卷积神经网络与连续时间序列分类算法相结合的网络框架。

[0112] 根据本公开的一个或多个实施例,提供了一种电子设备,包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如上述文本识别方法。

[0113] 根据本公开的一个或多个实施例,提供了一种计算机可读介质,其上存储有计算机程序,该程序被处理器执行时实现如上述文本识别方法的步骤。

[0114] 描述于本公开的实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中,例如,可以描述为:一种处理器包括获取单元、提取单元、识别单元和输出单元。其中,这些单元的名称在某种情况下并不构成对该单元本身的限定,例如,获取单元还可以被描述为“获取待识别的视频,对视频进行采样,得到视频帧序列的单元”。

[0115] 以上描述仅为本公开的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本公开的实施例中所涉及的发明范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开的实施例中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

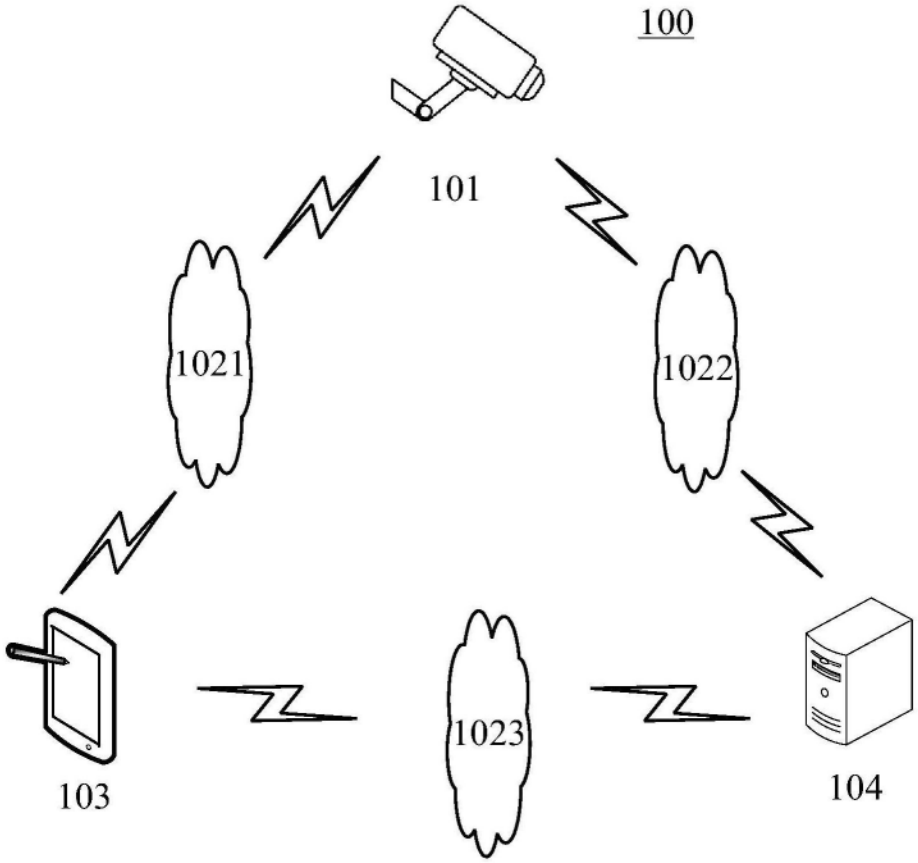


图1

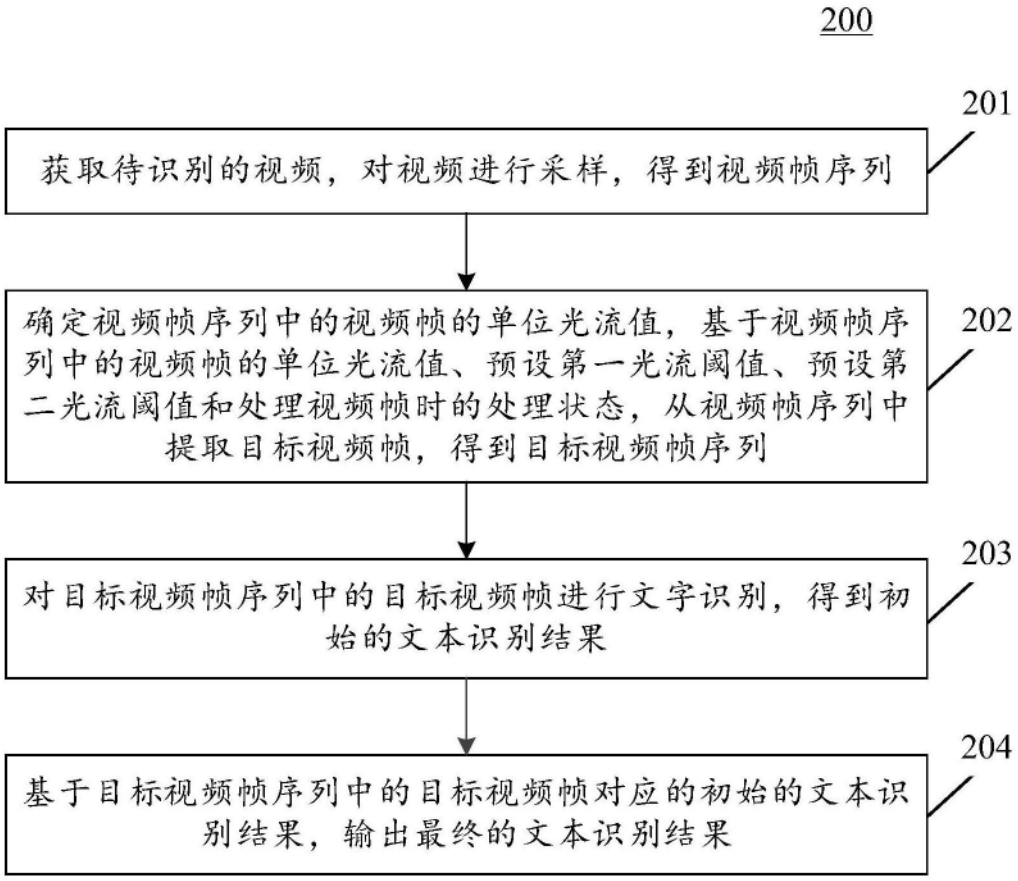

200
201
获取待识别的视频，对视频进行采样，得到视频帧序列
202
确定视频帧序列中的视频帧的单位光流值，基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从视频帧序列中提取目标视频帧，得到目标视频帧序列
203
对目标视频帧序列中的目标视频帧进行文字识别，得到初始的文本识别结果
204
基于目标视频帧序列中的目标视频帧对应的初始的文本识别结果，输出最终的文本识别结果

图2

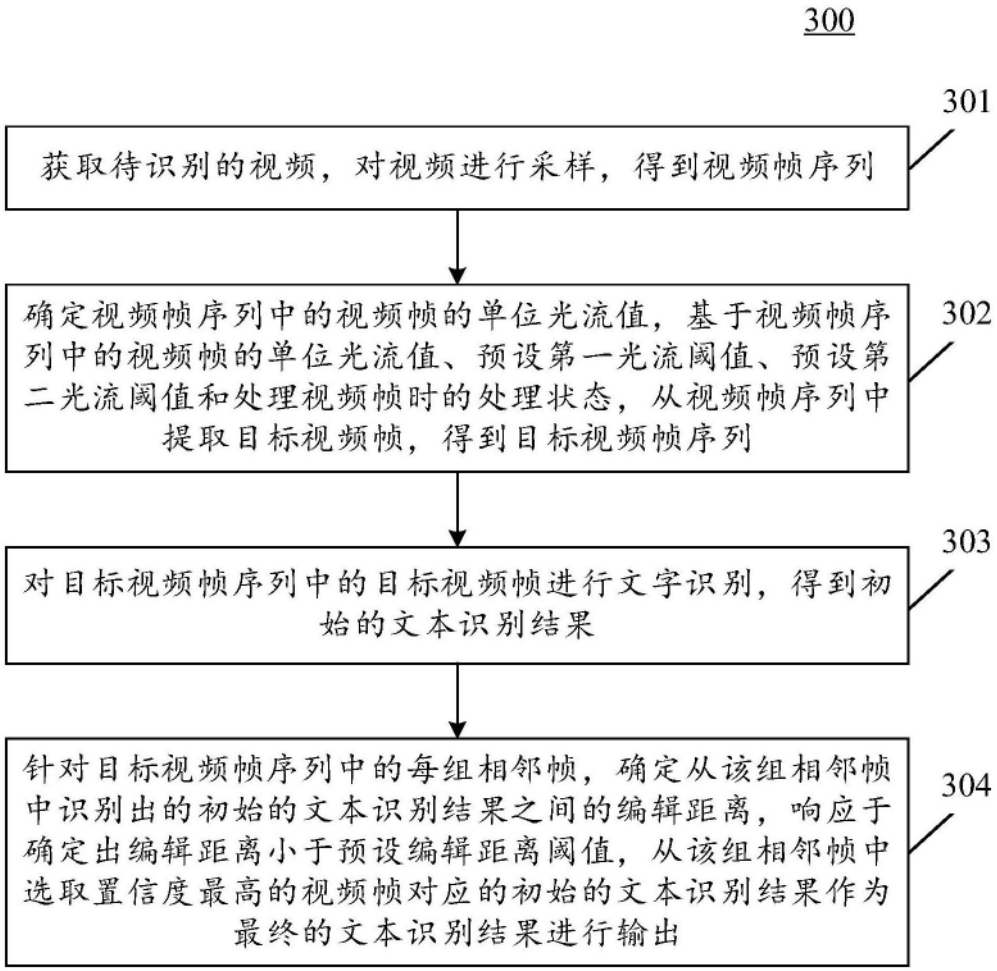

300
301
获取待识别的视频，对视频进行采样，得到视频帧序列
302
确定视频帧序列中的视频帧的单位光流值，基于视频帧序列中的视频帧的单位光流值、预设第一光流阈值、预设第二光流阈值和处理视频帧时的处理状态，从视频帧序列中提取目标视频帧，得到目标视频帧序列
303
对目标视频帧序列中的目标视频帧进行文字识别，得到初始的文本识别结果
304
针对目标视频帧序列中的每组相邻帧，确定从该组相邻帧中识别出的初始的文本识别结果之间的编辑距离，响应于确定出编辑距离小于预设编辑距离阈值，从该组相邻帧中选取置信度最高的视频帧对应的初始的文本识别结果作为最终的文本识别结果进行输出

图3

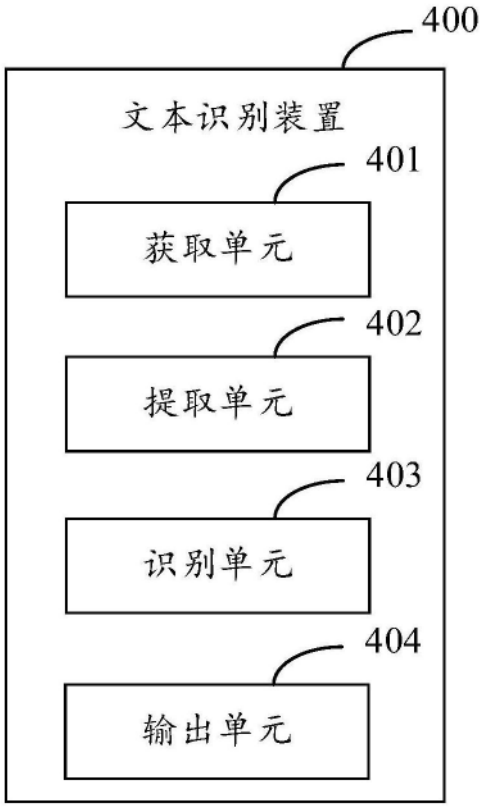

400
文本识别装置
401
获取单元
402
提取单元
403
识别单元
404
输出单元

图4

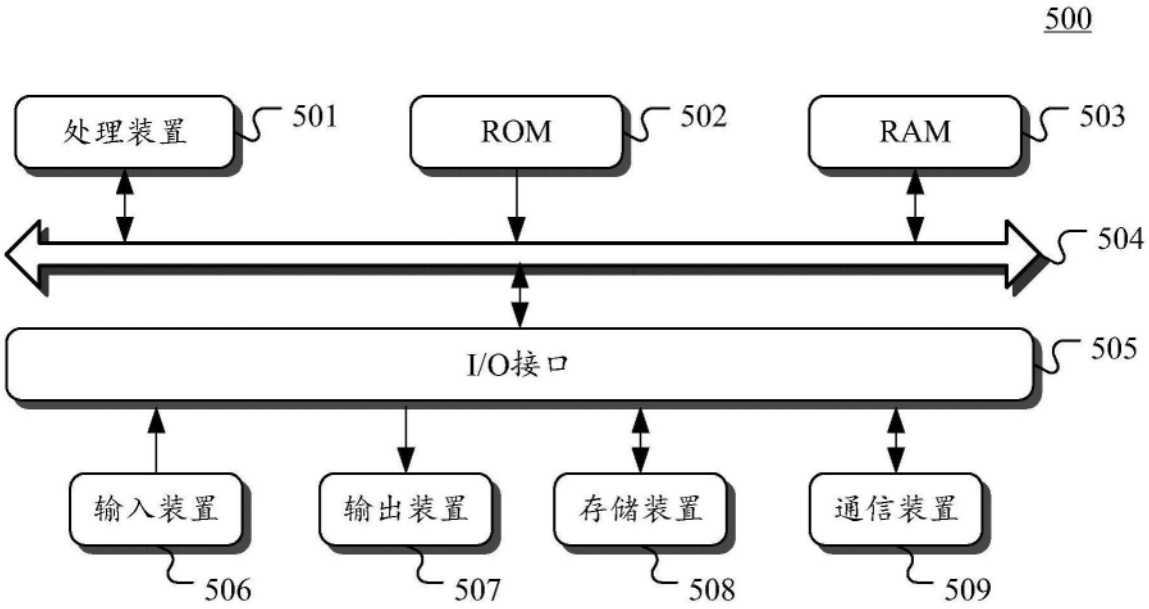

500
处理装置
501
ROM
502
RAM
503
504
I/O接口
505
输入装置
506
输出装置
507
存储装置
508
通信装置
509

图5