(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 110163181 B
(45)授权公告日 2021.07.06

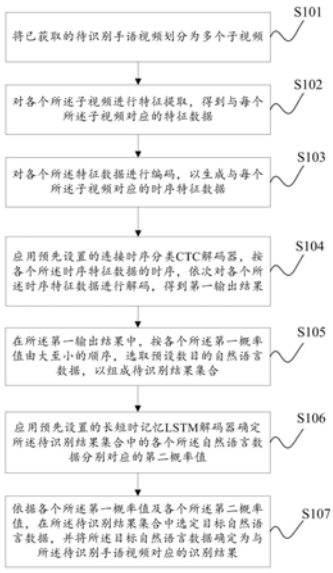
(21)申请号 201910456373.2
(22)申请日 2019.05.29
(65)同一申请的已公布的文献号
申请公布号 CN 110163181 A
(43)申请公布日 2019.08.23
(73)专利权人 中国科学技术大学
地址 230026 安徽省合肥市包河区金寨路96号
(72)发明人 李厚强 周文罡 蒲俊福
(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 李伟 王宝筠
(51)Int.Cl.
G06K 9/00(2006.01)
(56)对比文件
CN 109190578 A,2019.01.11
CN 108647603 A,2018.10.12
CN 108171198 A,2018.06.15
CN 105893942 A,2016.08.24
CN 109063615 A,2018.12.21
US 2019138607 A1,2019.05.09
Necati Cihan Camgoz等.SubUNets: End-to-end Hand Shape and Continuous Sign Language Recognition.《2017 IEEE International Conference on Computer Vision (ICCV)》.2017,
Runpeng Cui等.Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization.《2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)》.2017,
审查员 罗倩
权利要求书3页 说明书12页 附图3页

(54)发明名称

手语识别方法及装置

(57)摘要

本发明提供了一种手语识别方法,包括:对各个子视频进行特征提取得到与每个子视频对应的特征数据;对各个特征数据进行编码得到与每个子视频对应的时序特征数据;应用预先设置的CTC解码器,按各个时序特征数据的时序,依次对各个时序特征数据进行解码,得到第一输出结果;在第一输出结果中,按各个第一概率值由大至小的顺序选取预设数目的自然语言数据以组成待识别结果集合;应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据对应的第二概率值;依据第一概率值及第二概率值,在待识别结果集合中选定目标自然语言数据。通过CTC解码器及LSTM解码器共同解码,能有效的提升手语识别的精度。



CN 110163181 B

1.一种手语识别方法,其特征在于,包括:

将已获取的待识别手语视频划分为多个子视频;

对各个所述子视频进行特征提取,得到与每个所述子视频对应的特征数据;

对各个所述特征数据进行编码,以生成与每个所述子视频对应的时序特征数据;

应用预先设置的连接时序分类CTC解码器,按各个所述时序特征数据的时序,依次对各个所述时序特征数据进行解码,得到第一输出结果;所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值;

在所述第一输出结果中,按各个所述第一概率值由大至小的顺序,选取预设数目的自然语言数据,以组成待识别结果集合;

应用预先设置的长短时记忆LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值;

依据各个所述第一概率值及各个所述第二概率值,在所述待识别结果集合中选定目标自然语言数据,并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

2.根据权利要求1所述的方法,其特征在于,所述将已获取的待识别手语视频划分为多个子视频,包括:

调用预先设置的滑动窗,按预设的步长,从所述待识别手语视频的起始端依次提取与所述滑动窗的窗长匹配的子视频;

其中,所述窗长大于所述步长。

3.根据权利要求1所述的方法,其特征在于,所述应用预先设置的CTC解码器,按各个所述时序特征数据的时序,依次对各个所述时序特征数据进行解码,得到第一输出结果,包括:

将每个所述时序特征数据分别与预设的各个自然语言单词进行匹配,得到每个所述时序特征数据分别与各个所述自然语言单词对应的自然语言概率分布;

基于所述概率分布确定与待识别视频对应的各个自然语言数据的第一概率值;

将各个所述第一概率值组成第一输出结果。

4.根据权利要求1所述的方法,其特征在于,所述在所述第一输出结果中,按各个所述第一概率值由大至小的顺序,选取预设数目的自然语言数据,包括:

依据第一概率值的大小,对所述第一输出结果中的各个第一概率值进行排序,并由大至小的选取预设数目的所述第一概率值;确定已选取的各个所述第一概率值分别对应的自然语言数据;

或,

将所述第一输出结果中的各个第一概率值与预先设置的概率阈值进行比较,得到多个大于所述概率阈值的所述第一概率值;在所述多个大于所述概率阈值的所述第一概率值中,由大至小的选取预设数目的所述第一概率值,并确定已选取的各个所述第一概率值分别对应的自然语言数据。

5.根据权利要求1所述的方法,其特征在于,所述依据各个所述第一概率值及各个所述第二概率值,在所述待识别结果集合中选定目标自然语言数据,并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果,包括:

基于所述第一概率值及所述第二概率值对待识别结果集合中的各个所述自然语言数据进行评分,得到评分结果;

依据所述评分结果在所述待识别结果集合中确定与所述待识别手语视频对应的目标自然语言数据;

将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

6.一种手语识别装置,其特征在于,包括:

获取单元,用于将已获取的待识别手语视频划分为多个子视频;

提取单元,用于对各个所述子视频进行特征提取,得到与每个所述子视频对应的特征数据;

编码单元,用于对各个所述特征数据进行编码,得到与每个所述子视频对应的时序特征数据;

解码单元,用于应用预先设置的CTC解码器,按各个所述时序特征数据的时序,依次对各个所述时序特征数据进行解码,得到第一输出结果;所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值;

选取单元,用于在所述第一输出结果中,按各个所述第一概率值由大至小的顺序,选取预设数目的自然语言数据,以组成待识别结果集合;

执行单元,用于应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值;

识别单元,用于依据各个所述第一概率值及各个所述第二概率值,在所述待识别结果集合中选定目标自然语言数据,并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

7.根据权利要求6所述的装置,其特征在于,所述获取单元,包括:

提取子单元,用于调用预先设置的滑动窗,按预设的步长,从所述待识别手语视频的起始端依次提取与所述滑动窗的窗长匹配的子视频;

其中,所述窗长大于所述步长。

8.根据权利要求6所述的装置,其特征在于,所述解码单元,包括:

匹配子单元,用于将每个所述时序特征数据分别与预设的各个自然语言单词进行匹配,得到每个所述时序特征数据分别与各个所述自然语言单词对应的自然语言概率分布;

第一确定子单元,用于依据所述概率分布确定与待识别视频对应的各个自然语言数据的第一概率值;

第一执行子单元,将各个所述第一概率值组成第一输出结果。

9.根据权利要求6所述的装置,其特征在于,所述选取单元,包括:

第一排序子单元或第二排序子单元;

所述第一排序子单元,用于依据第一概率值的大小,对所述第一输出结果中的各个第一概率值进行排序,并由大至小的选取预设数目的所述第一概率值;确定已选取的各个所述第一概率值分别对应的自然语言数据;

所述第二排序子单元,用于将所述第一输出结果中的各个第一概率值与预先设置的概率阈值进行比较,得到多个大于所述概率阈值的所述第一概率值;在所述多个大于所述概率阈值的所述第一概率值中,由大至小的选取预设数目的所述第一概率值,并确定已选取

的各个所述第一概率值分别对应的自然语言数据。

10.根据权利要求6所述的装置,其特征在于,所述识别单元,包括:

评分子单元,用于依据所述第一概率值及所述第二概率值对待识别结果集合中的各个所述自然语言数据进行评分,得到评分结果;

第二确定子单元,用于依据所述评分结果在所述待识别结果集合中确定与所述待识别手语视频对应的目标自然语言数据;

第三确定子单元,用于将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

手语识别方法及装置

技术领域

[0001] 本发明涉及数据处理领域,特别涉及一种手语识别方法及装置。

背景技术

[0002] 随着信息技术的发展,基于机器学习的连续手语识别技术也得到了极大的发展。手语是聋哑人士获取信息及表达信息的一种重要方式,聋哑人士通常通过手语来与他人进行沟通,这需要交流的双方都要懂得手语,而正常人学习手语需耗费大量的时间和精力,故而大部分正常人的手语水平较低,难以与聋哑人士进行交流。因此,连续手语识别技术对聋哑人的沟通具有重大意义。

[0003] 然而,现有的基于机器学习的连续手语识别技术中,往往对包含手语的视频的识别准确率低,因此,如何提高手语视频的识别准确率成为本领域技术人员迫切解决的问题。

发明内容

[0004] 本发明所要解决的技术问题是提供一种手语识别方法,能够能基于连接时序分类(Connectionist Temporal Classification,CTC)解码器及长短时记忆(Long Short Term Memory,LSTM)解码器共同对待识别手语视频进行识别,有效的提升手语识别的准确率。

[0005] 本发明还提供了一种手语识别装置,用以保证上述方法在实际中的实现及应用。

[0006] 一种手语识别方法,包括:

[0007] 将已获取的待识别手语视频划分为多个子视频;

[0008] 对各个所述子视频进行特征提取,得到与每个所述子视频对应的特征数据;

[0009] 对各个所述特征数据进行编码,以生成与每个所述子视频对应的时序特征数据;

[0010] 应用预先设置的连接时序分类CTC解码器,按各个所述时序特征数据的时序,依次对各个所述时序特征数据进行解码,得到第一输出结果;所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值;

[0011] 在所述第一输出结果中,按各个所述第一概率值由大至小的顺序,选取预设数目的自然语言数据,以组成待识别结果集合;

[0012] 应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值;

[0013] 依据各个所述第一概率值及各个所述第二概率值,在所述待识别结果集合中选定目标自然语言数据,并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0014] 上述的方法,可选的,所述将已获取的待识别手语视频划分为多个子视频,包括:

[0015] 调用预先设置的滑动窗,按预设的步长,从所述待识别手语视频的起始端依次提取与所述滑动窗的窗长匹配的子视频;

[0016] 其中,所述窗长大于所述步长。

[0017] 上述的方法,可选的,应用预先设置的CTC解码器,按各个所述时序特征数据的时

序，依次对各个所述时序特征数据进行解码，得到第一输出结果，包括：

[0018] 将每个所述时序特征数据分别与预设的各个自然语言单词进行匹配，得到每个所述时序特征数据分别与各个所述自然语言单词对应的自然语言概率分布；

[0019] 基于所述概率分布确定与待识别视频对应的各个自然语言数据的第一概率值；

[0020] 将各个所述第一概率值组成第一输出结果。

[0021] 上述的方法，可选的，在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，包括：

[0022] 依据第一概率值的大小，对所述第一输出结果中的各个第一概率值进行排序，并由大至小的选取预设数目的所述第一概率值；确定已选取的各个所述第一概率值分别对应的自然语言数据；

[0023] 或，

[0024] 将所述第一输出结果中的各个第一概率值与预先设置的概率阈值进行比较，得到多个大于所述概率阈值的所述第一概率值；在所述多个大于所述概率阈值的所述第一概率值中，由大至小的选取预设数目的所述第一概率值，并确定已选取的各个所述第一概率值分别对应的自然语言数据。

[0025] 上述的方法，可选的，依据各个所述第一概率值及各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果，包括：

[0026] 基于所述第一概率值及所述第二概率值对待识别结果集合中的各个所述自然语言数据进行评分，得到评分结果；

[0027] 依据所述评分结果在所述待识别结果集合中确定与所述待识别手语视频对应的目标自然语言数据；

[0028] 将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0029] 一种手语识别装置，包括：

[0030] 获取单元，用于将已获取的待识别手语视频划分为多个子视频；

[0031] 提取单元，用于对各个所述子视频进行特征提取，得到与每个所述子视频对应的特征数据；

[0032] 编码单元，用于对各个所述特征数据进行编码，得到与每个所述子视频对应的时序特征数据；

[0033] 解码单元，用于应用预先设置的CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果；所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值；

[0034] 选取单元，用于在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，以组成待识别结果集合；

[0035] 执行单元，用于应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值；

[0036] 识别单元，用于依据各个所述第一概率值及各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0037] 上述的装置，可选的，所述获取单元，包括：

[0038] 提取子单元，用于调用预先设置的滑动窗，按预设的步长，从所述待识别手语视频的起始端依次提取与所述滑动窗的窗长匹配的子视频；

[0039] 其中，所述窗长大于所述步长。

[0040] 上述的装置，可选的，所述解码单元，包括：

[0041] 匹配子单元，用于将每个所述时序特征数据分别与预设的各个自然语言单词进行匹配，得到每个所述时序特征数据分别与各个所述自然语言单词对应的自然语言概率分布；

[0042] 第一确定子单元，用于依据所述概率分布确定与待识别视频对应的各个自然语言数据的第一概率值；

[0043] 第一执行子单元，将各个所述第一概率值组成第一输出结果。

[0044] 上述的装置，可选的，所述选取单元，包括：

[0045] 第一排序子单元或第二排序子单元；

[0046] 所述第一排序子单元，用于依据第一概率值的大小，对所述第一输出结果中的各个第一概率值进行排序，并由大至小的选取预设数目的所述第一概率值；确定已选取的各个所述第一概率值分别对应的自然语言数据；

[0047] 所述第二排序子单元，用于将所述第一输出结果中的各个第一概率值与预先设置的概率阈值进行比较，得到多个大于所述概率阈值的所述第一概率值；在所述多个大于所述概率阈值的所述第一概率值中，由大至小的选取预设数目的所述第一概率值，并确定已选取的各个所述第一概率值分别对应的自然语言数据。

[0048] 上述的装置，可选的，所述识别单元，包括：

[0049] 评分子单元，用于依据所述第一概率值及所述第二概率值对待识别结果集合中的各个所述自然语言数据进行评分，得到评分结果；

[0050] 第二确定子单元，用于依据所述评分结果在所述待识别结果集合中确定与所述待识别手语视频对应的目标自然语言数据；

[0051] 第三确定子单元，用于将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0052] 一种存储介质，所述存储介质包括存储的指令，其中，在所述指令运行时控制所述存储介质所在的设备执行上述的手语识别方法。

[0053] 一种电子设备，包括存储器，以及一个或者一个以上的指令，其中一个或者一个以上指令存储于存储器中，且经配置以由一个或者一个以上处理器执行上述的手语识别方法。

[0054] 经由上述方案可知，本发明提供了一种手语识别方法，包括：将已获取的待识别手语视频划分为多个子视频；对各个所述子视频进行特征提取，得到与每个所述子视频对应的特征数据；对各个所述特征数据进行编码，以生成与每个所述子视频对应的时序特征数据；应用预先设置的CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果；所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值；在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，以组成待识别结果集合；应用预先设置的LSTM解码

器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值;依据各个所述第一概率值及各个所述第二概率值,在所述待识别结果集合中选定目标自然语言数据,并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。应用本发明实施例提供的方法,能基于CTC解码器及LSTM解码器共同对待识别手语视频进行识别,有效的提升手语识别的精度。

附图说明

[0055] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0056] 图1为本发明提供的一种手语识别方法的方法流程图;

[0057] 图2为本发明提供的一种手语识别方法的又一方法流程图;

[0058] 图3为本发明提供的一种手语识别方法的又一方法流程图;

[0059] 图4为本发明提供的一种手语识别装置的结构示意图;

[0060] 图5为本发明提供的一种电子设备的结构示意图。

具体实施方式

[0061] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0062] 本发明可用于众多通用或专用的计算装置环境或配置中。例如:个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器装置、包括以上任何装置或设备的分布式计算环境等等。

[0063] 本发明实施例提供了一种手语识别方法,该方法可以应用在多种系统平台,其执行主体可以为计算机终端或各种移动设备的处理器,所述方法的方法流程图如图1所示,具体包括:

[0064] S101:将已获取的待识别手语视频划分为多个子视频。

[0065] 本发明实施例提供的方法中,所述待识别手语视频中包含手语动作信息。

[0066] 需要说明的是,所述待识别手语视频可由表征连续手语动作的图像序列组成。

[0067] S102:对各个所述子视频进行特征提取,得到与每个所述子视频对应的特征数据。

[0068] 本发明实施例提供的方法中,应用预先设置的卷积神经网络模型对每个子视频进行特征提取,得到与每个子视频对应的特征数据。

[0069] 其中,该卷积神经网络模型可以为三维残差卷积神经网络。

[0070] 本发明实施例提供的方法中,将各个子视频的视频尺寸调整为224×224,使用18层的三维残差卷积神经网络,提取各个子视频在该三维残差神经网络中的池化层的512维响应作为子视频的特征数据。

[0071] S103:对各个所述特征数据进行编码,以生成与每个所述子视频对应的时序特征

数据。

[0072] 本发明实施例提供的方法中，每个子视频的时序特征数据包含与该子视频对应的手语词汇的概率分布。

[0073] 本发明实施例提供的方法中，应用预先设置的编码器对各个特征数据进行编码，得到编码结果，并将该编码结果映射到词汇对数概率空间，得到与每个所述子视频对应的时序特征数据，可选的，应用编码器对特征数据进行映射得到的隐变量，该隐变量为编码结果。

[0074] 具体的，该编码器可以是双向长短时记忆网络，该双向长短时记忆网络的层数可以为两层。

[0075] S104：应用预先设置的连接时序分类CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果；所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值。

[0076] 本发明实施例提供的方法中，所述第一输出结果为应用预先设置的基于连接时序分类CTC解码器，对各个所述时序特征数据进行解码得到的与所述待识别手语视频对应的各个自然语言数据的概率分布。

[0077] S105：在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，以组成待识别结果集合。

[0078] 本发明实施例提供的方法中，通过按第一概率值由大至小的顺序选取预设数目的自然语言数据，可得到与该待识别手语视频关联度高的待识别自然语言数据集合。

[0079] 本发明实施例提供的方法中，所述自然语言数据可以为各种语言类型的语音数据或文字数据，语言类型可以为中文、英文、日文或法文等等。

[0080] S106：应用预先设置的长短时记忆LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值。

[0081] 本发明实施例提供的方法中，可选的，将各个自然语言数据及编码器的编码结果输入至LSTM解码器中，可得到各个自然语言数据的第二概率值。

[0082] S107：依据各个所述第一概率值及各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0083] 本发明实施例提供的方法中，依据待识别结果集合中的每个自然语言数据的第一概率值及第二概率值，分别对待识别结果集合中的每个自然语言数据进行评分，将评分最高的自然语言数据确定为目标自然语言数据，即待识别视频的识别结果。

[0084] 本发明实施例提供的手语识别方法，包括：将已获取的待识别手语视频划分为多个子视频；对各个所述子视频进行特征提取，得到与每个所述子视频对应的特征数据；对各个所述特征数据进行编码，以生成与每个所述子视频对应的时序特征数据；应用预先设置的CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果；所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值；在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，以组成待识别结果集合；应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值；依据各个所述第一概率值及

各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。应用本发明实施例提供的方法，能基于CTC解码器及LSTM解码器共同对待识别手语视频进行识别，有效的提升手语识别的精度。

[0085] 本发明实施例提供的方法中，在上述实施过程的基础上，具体的，将已获取的待识别手语视频划分为多个子视频，包括：

[0086] 调用预先设置的滑动窗，按预设的步长，从所述待识别手语视频的起始端依次提取与所述滑动窗的窗长匹配的子视频；

[0087] 其中，所述窗长大于所述步长。

[0088] 本发明实施例提供的方法中，该滑动窗的窗长可设置为8，该滑动窗的步长可以设置为4；每个子视频均和与其相邻的子视频存在重叠部分，能有效的避免应用滑窗对待识别手语视频的分割错误对手语的识别结果造成负面影响。

[0089] 本发明实施例提供的手语识别方法中，在上述实施过程的基础上，具体的，应用预先设置的CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果，如图2所示，具体包括：

[0090] S201：将每个所述时序特征数据分别与预设的各个自然语言单词进行匹配，得到每个所述时序特征数据分别与各个所述自然语言单词对应的自然语言概率分布。

[0091] 本发明实施例提供的方法中，将各个时序特征数据组成时序特征数据序列，并将该时序特征序列输入至预先设置的CTC解码器中，可得到每个时序特征数据对应的子视频属于预设的自然语言单词集合中的各个自然语言单词的概率值。

[0092] S202：基于所述概率分布确定与待识别视频对应的各个自然语言数据的第一概率值。

[0093] 本发明实施例提供的方法中，通过选取概率值较高的单自然语言单词确定待识别视频对应的各个自然语言数据，并依据组成每个自然语言数据的自然语言单词的概率值，确定该自然语言数据的第一概率值。

[0094] 本发明实施例提供的方法中，通过确定每个时序特征数据属于各个自然语言单词的概率，并依据每个自然语言单词的概率确定与待识别手语视频对应的各个自然语言数据的第一概率值。

[0095] S203：将各个所述第一概率值组成第一输出结果。

[0096] 本发明实施例提供的方法中，基于每个时序特征数据属于预设的各个自然语言单词的概率，确定预设数目的第一概率值，每个第一概率值对应表示待识别手语视频属于该对应概率值对应的自然语言数据的概率，其中，自然语言数据由多个自然语言单词组成。即依据每个时序特征数据属于自然语言单词的概率值选定预设数目的解码路径，基于每条解码路径确定与该解码路径对应的自然语言数据；每个解码路径对应一个自然语言数据。

[0097] 本发明实施例提供的方法中，CTC解码器中引入空白标签，用于表示当前输入至该CTC解码器的时序特征数据对应的子视频不属于预设的各个自然语言单词。

[0098] 本发明实施例提供的方法中，可以将各个时序特征数据组成的时序特征序列输入至CTC解码器中，用$\pi=(\pi_1,\ldots,\pi_T)$表示一条解码路径，对于待识别手语视频X，路径π的条件概率为：

[0099] $$p(\pi|V)=\prod_{t=1}^{N}p(\pi_t|v_t)=\prod_{t=1}^{N}Y_{t,\pi_t}$$

[0100] 通过定义一个多对一的映射$\mathcal{M}$,确定初始自然语言数据:再依次删除重复标签和空白标签,得到自然语言数据;对于一个长度为L的自然语言数据$s=(s_1,\cdots,s_L)$,s的条件概率是所有对应解码路径的概率总和,其计算公式如下:

[0101] $$p(s|V)=\sum_{\pi\in\mathcal{M}^{-1}(s)}p(\pi|V)$$

[0102] 其中,$\mathcal{M}^{-1}(s)=\{\pi|M(\pi)=s\}$是$\mathcal{M}$的逆映射。

[0103] 本发明实施例提供的手语识别方法中,在上述实施过程的基础上,具体的,在所述第一输出结果中,按各个所述第一概率值由大至小的顺序,选取预设数目的自然语言数据,包括:

[0104] 依据第一概率值的大小,对所述第一输出结果中的各个第一概率值进行排序,并由大至小的选取预设数目的所述第一概率值;确定已选取的各个所述第一概率值分别对应的自然语言数据;

[0105] 或,

[0106] 将所述第一输出结果中的各个第一概率值与预先设置的概率阈值进行比较,得到多个大于所述概率阈值的所述第一概率值;在所述多个大于所述概率阈值的所述第一概率值中,由大至小的选取预设数目的所述第一概率值,并确定已选取的各个所述第一概率值分别对应的自然语言数据。

[0107] 本发明实施例提供的方法中,每个第一概率值对应均对应一个自然语言数据。

[0108] 本发明实施例提供的方法中,在上述实施过程的基础上,具体的,应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值,包括:

[0109] 确定对各个特征数据进行编码的编码结果;

[0110] 获取当前待解码自然语言数据;

[0111] 将所述编码结果及所述自然语言数据输入至LSTM解码器中,得到与当前待解码的自然语言数据对应的第二概率值。

[0112] 本发明实施例提供的方法中,通过确定所述当前待解码自然语言数据对应的各个词向量;依次对每个词向量进行解码,并在对每个词向量进行解码时,确定与当前词向量对应的隐状态及预先设置的注意力向量;基于当前解码的词向量、所述词向量对应的隐状态及注意力向量,得到与当前待解码自然语言数据对应的解码结果;将每个词向量对应的解码结果映射到词汇对数概率空间,得到与当前待解码的自然语言数据对应的第二概率值。

[0113] 本发明实施例提供的方法中,与当前词向量对应的隐状态为当前时刻的前一时刻所述LSTM输出的隐状态。

[0114] 本发明实施例提供的方法中,对每个自然语言数据开始解码时,输入开始标志符,使LSTM根据当前时序特征数据、前一时刻隐状态及注意力向量开始解码,直到LSTM解码器输出解码结束标志符时,得到当前待解码的自然语言数据与该LSTM解码器对应的解码结

果，该解码结果的表达式如下：

[0115] $$d_k = \mathrm{Decoder}_{lstm}(c_k, s_k, h^d_{k-1})$$

[0116] 其中，c_k是注意力向量，s_k是当前输入时序特征数据，h^d_{k-1}是解码器隐状态。

[0117] 基于LSTM解码器的全连接层将该解码结果映射到词汇对数概率空间$z_k = W_{fc2} \cdot d_k + b_{fc2}$，得到待识别手语视频与LSTM解码器对应的概率分布，具体如下：

[0118] $Z = (Z_{k,l}) = [z_1, z_2, \cdots, z_L]^T$

[0119] 其中，L是解码出的句子长度，$Z_{k,l}$是当前词向量s_k属于手语词汇l的概率。

[0120] 本发明实施例提供的手语识别方法中，在上述实施过程的基础上，具体的，依据各个所述第一概率值及各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果，如图3所示，包括：

[0121] S301：基于所述第一概率值及所述第二概率值对待识别结果集合中的各个所述自然语言数据进行评分，得到评分结果。

[0122] 本发明实施例提供的方法中，将各个自然语言数据组成待识别结果集合$C = \{s^i | i = 1, \cdots, K\}$，基于待识别结果集合中的每个自然语言数据的第一概率值，及第二概率值，对待识别结果集合中的各个所述自然语言数据进行评分，对各个所述目标自然语言数据进行评分的评分公式如下：

[0123] $$r(s^i) = \alpha \ln p_{ctc}(s^i | V) + (1-\alpha) \ln P_{lstm}(s^i | V) + \beta \ln L_i$$

[0124] 其中，$r(s^i)$为自然语言数据s^i的评分结果，α是可调参数，L_i是s^i的长度，$\beta \ln L_i$是长度项，用于平衡长序列生成概率偏低的问题。

[0125] S302：依据所述评分结果在所述待识别结果集合中确定与所述待识别手语视频对应的目标自然语言数据。

[0126] 本发明实施例提供的方法中，根据得分r在待识别结果集合中挑出得分最高的自然语言数据作为目标自然语言数据，即，将得分最高的自然语言数据确定为与待识别手语视频对应的目标自然语言数据，其表达式为：

[0127] $$\hat{s} = \arg\max_s r(s)$$

[0128] S303：将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0129] 本发明实施例提供的方法中，在训练目标解码器时，基于最大似然准则分别得到CTC解码器及LSTM解码器的损失函数。具体过程如下：

[0130] 将已获取的训练视频划分为多个子视频；对各个所述子视频进行特征提取，得到与每个所述子视频对应的特征数据；对各个所述特征数据进行编码，得到各个所述特征数据的编码结果；将该编码结果及与该训练视频对应的词汇标签输入至LSTM解码器，得到所述训练视频对应的各个自然语言单词的第一概率分布；并将该编码结果映射到词汇对数概率空间，得到各个子视频对应的手语词汇概率分布；将所述手语词汇概率分布输入至CTC结果器，得到所述训练视频对应的各个自然语言单词的第二概率分布，使用软动态时间规整对第一概率分布及第二概率分布进行对齐，使LSTM解码器得到的第一概率分布及CTC解码器解码得到的第二概率分布趋于一致。

[0131] 本发明实施例提供的方法中，具体对齐过程如下：

[0132] 给定手语视频V和其对应的标注$s=(s_1,\cdots,s_L)$，CTC解码器对应的损失函数为：

[0133] $$\mathcal{L}_{ctc}=-\ln p_{ctc}(s|V)$$

[0134] 其中，$P_{ctc}(s|V)$是给定V的s的后验概率。

[0135] 对于LSTM解码器，s给定V的概率为：

[0136] $$p_{lstm}(s|V)=\prod_{i=1}^{L}p(s_i|s_{i-1})=\prod_{i=1}^{L}Z_{i,s_i}$$

[0137] 该LSTM解码器对应的损失函数为：

[0138] $$\mathcal{L}_{lstm}=-\ln p_{lstm}(s|V)$$

[0139] 此外，软动态时间规整的约束项为：

[0140] $$\mathcal{L}_{align}=\mathcal{D}_p(Y,Z)$$

[0141] 联合优化以下目标函数：

[0142] $$\mathcal{L}=\lambda\mathcal{L}_{ctc}+(1-\lambda)\mathcal{L}_{lstm}+\mathcal{L}_{align}+\mu\|\omega\|^2$$

[0143] 其中，λ是超参数，用于调节CTC解码器及LSTM解码器之间的平衡，$\mu||\omega||^2$是正则项，用于减缓网络过拟合现象。

[0144] 本发明实施例提供的方法中，通过软动态时间规整对CTC解码器的输出及LSTM解码器的输出进行对齐，给定第一输出结果$u=(u_1,\cdots,u_m)$和第二输出结果$v=(v_1,\cdots,v_n)$，记原始动态时间规整算法计算子串$u^i=(u_1,\cdots,u_i)$和$v_j=(v1,\cdots,v_j)$的距离为$D_{i,j}$，其计算式为：

[0145] $D_{i,j}=d_{i,j}+\min(D_{i-1,j},D_{i,j-1},D_{i-1,j-1})$

[0146] 其中

[0147] $d_{i,j}=||u_i-v_j||_2$

[0148] 本发明实施例提供的方法中，为了使原始动态时间规整算法可导用于网络优化求解，软动态时间规整算法引入近似的最小函数操作算子如下：

[0149] $$\min_i^{\gamma}\{a_i\}=\begin{cases}\min_i\{a_i\}, & \gamma=0\\ -\gamma\log\sum_i e^{-\frac{a_i}{\gamma}}, & \gamma>0\end{cases}$$

[0150] 由此，两种解码方式对应的概率分布Y和Z的软动态时间规整距离为：

[0151] $$\mathcal{D}_p=D_{N,L}(Y,Z)$$

[0152] 其中，N和L分别是两种解码解码方式解码长度。

[0153] 本发明实施例提供的方法中，利用回溯算法得到规整路径，规整路径是手语词汇和手语视频片段之间的对齐关系，该规整路径为：

[0154] $\Pi=\{(p,q)|p\leqslant N,q\leqslant L\}$

[0155] 第p个视频片段的规整标签ℓ_p为：

[0156] $\ell_p = s_q$

[0157] 上述各个具体的实现方式，及各个实现方式的衍生过程，均在本发明保护范围内。

[0158] 与图1所述的方法相对应，本发明实施例还提供了一种手语识别装置，用于对图1中方法的具体实现，本发明实施例提供的手语识别装置可以应用计算机终端或各种移动设备中，其结构示意图如图4所示，具体包括：

[0159] 获取单元401，用于将已获取的待识别手语视频划分为多个子视频；

[0160] 提取单元402，用于对各个所述子视频进行特征提取，以生成与每个所述子视频对应的特征数据；

[0161] 编码单元403，用于对各个所述特征数据进行编码，以生成与每个所述子视频对应的时序特征数据；

[0162] 解码单元404，用于应用预先设置的CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果；所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值；

[0163] 选取单元405，用于在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，以组成待识别结果集合；

[0164] 执行单元406，用于应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值；

[0165] 识别单元407，用于依据各个所述第一概率值及各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0166] 本发明实施例提供的手语识别装置，包括：将已获取的待识别手语视频划分为多个子视频；对各个所述子视频进行特征提取，得到与每个所述子视频对应的特征数据；对各个所述特征数据进行编码，以生成与每个所述子视频对应的时序特征数据；应用预先设置的CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果；所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值；在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，以组成待识别结果集合；应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值；依据各个所述第一概率值及各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。应用本发明实施例提供的方法，能基于CTC解码器及LSTM解码器共同对待识别手语视频进行识别，有效的提升手语识别的精度。

[0167] 本发明实施例提供的手语装置，可选的，所述获取单元401，包括：

[0168] 提取子单元，用于调用预先设置的滑动窗，按预设的步长，从所述待识别手语视频的起始端依次提取与所述滑动窗的窗长匹配的子视频；

[0169] 其中，所述窗长大于所述步长。

[0170] 本发明实施例提供的手语装置，可选的，所述解码单元404，包括：

[0171] 匹配子单元，用于将每个所述时序特征数据分别与预设的各个自然语言单词进行匹配，得到每个所述时序特征数据分别与各个所述自然语言单词对应的自然语言概率分

布；

[0172] 第一确定子单元，用于依据所述概率分布确定与待识别视频对应的各个自然语言数据的第一概率值；

[0173] 第一执行子单元，将各个所述第一概率值组成第一输出结果。

[0174] 本发明实施例提供的手语装置，可选的，所述选取单元405，包括：

[0175] 第一排序子单元或第二排序子单元；

[0176] 所述第一排序子单元，用于依据第一概率值的大小，对所述第一输出结果中的各个第一概率值进行排序，并由大至小的选取预设数目的所述第一概率值；确定已选取的各个所述第一概率值分别对应的自然语言数据；

[0177] 所述第二排序子单元，用于将所述第一输出结果中的各个第一概率值与预先设置的概率阈值进行比较，得到多个大于所述概率阈值的所述第一概率值；在所述多个大于所述概率阈值的所述第一概率值中，由大至小的选取预设数目的所述第一概率值，并确定已选取的各个所述第一概率值分别对应的自然语言数据。

[0178] 本发明实施例提供的手语装置，可选的，所述识别单元407，包括：

[0179] 评分子单元，用于依据所述第一概率值及所述第二概率值对待识别结果集合中的各个所述自然语言数据进行评分，得到评分结果；

[0180] 第二确定子单元，用于依据所述评分结果在所述待识别结果集合中确定与所述待识别手语视频对应的目标自然语言数据；

[0181] 第三确定子单元，用于将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0182] 本发明实施例还提供了一种存储介质，所述存储介质包括存储的指令，其中，在所述指令运行时控制所述存储介质所在的设备执行上述手语识别方法。

[0183] 本发明实施例还提供了一种电子设备，其结构示意图如图5所示，具体包括存储器501，以及一个或者一个以上的指令502，其中一个或者一个以上指令502存储于存储器501中，且经配置以由一个或者一个以上处理器503执行所述一个或者一个以上指令502进行以下操作：

[0184] 将已获取的待识别手语视频划分为多个子视频；

[0185] 对各个所述子视频进行特征提取，得到与每个所述子视频对应的特征数据；

[0186] 对各个所述特征数据进行编码，以生成与每个所述子视频对应的时序特征数据；

[0187] 应用预先设置的CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果；所述第一输出结果包含与所述待识别视频对应的各个自然语言数据的第一概率值；

[0188] 在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，以组成待识别结果集合；

[0189] 应用预先设置的LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值；

[0190] 依据各个所述第一概率值及各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果。

[0191] 需要说明的是，本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。对于装置类实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0192] 最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，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、物品或者设备不仅包括那些要素，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、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个.....”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0193] 为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，在实施本发明时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0194] 通过以上的实施方式的描述可知，本领域的技术人员可以清楚地了解到本发明可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在存储介质中，如ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备(可以是个人计算机，服务器，或者网络设备等)执行本发明各个实施例或者实施例的某些部分所述的方法。

[0195] 以上对本发明所提供的一种手语识别方法及装置进行了详细介绍，本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

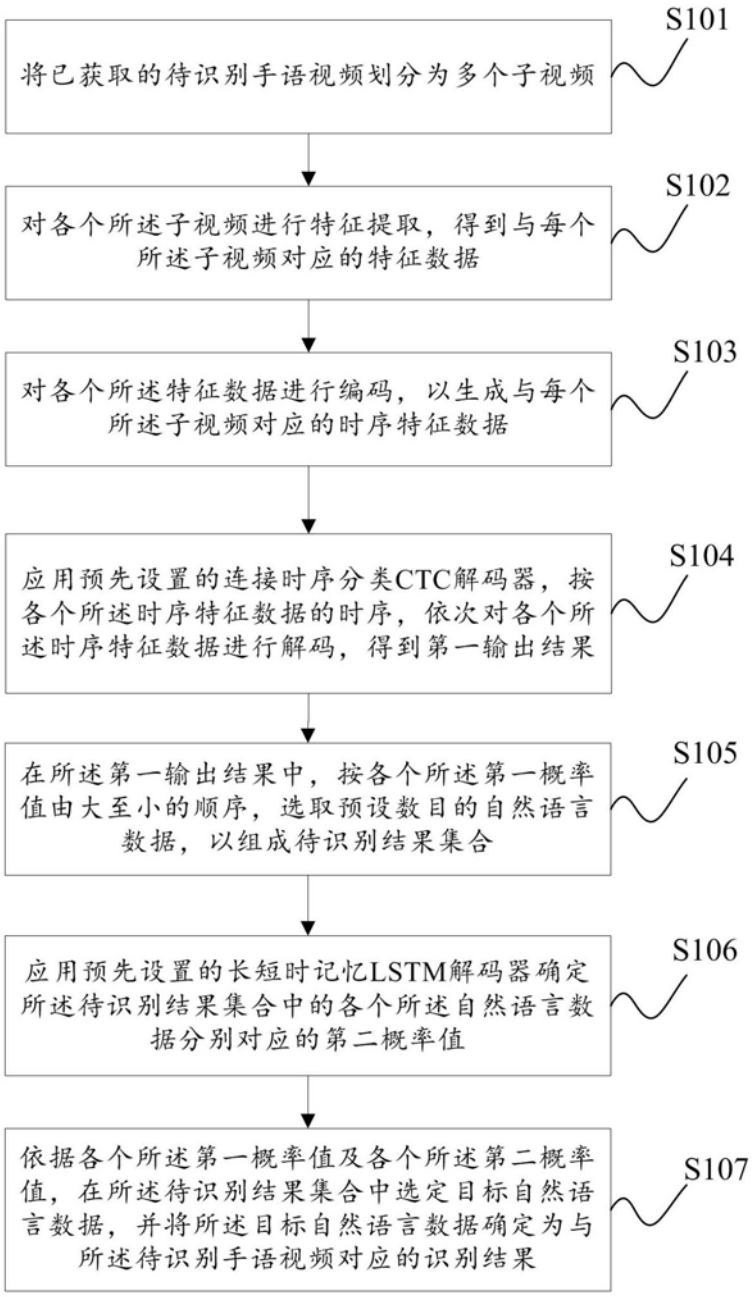

S101
将已获取的待识别手语视频划分为多个子视频
S102
对各个所述子视频进行特征提取，得到与每个所述子视频对应的特征数据
S103
对各个所述特征数据进行编码，以生成与每个所述子视频对应的时序特征数据
S104
应用预先设置的连接时序分类CTC解码器，按各个所述时序特征数据的时序，依次对各个所述时序特征数据进行解码，得到第一输出结果
S105
在所述第一输出结果中，按各个所述第一概率值由大至小的顺序，选取预设数目的自然语言数据，以组成待识别结果集合
S106
应用预先设置的长短时记忆LSTM解码器确定所述待识别结果集合中的各个所述自然语言数据分别对应的第二概率值
S107
依据各个所述第一概率值及各个所述第二概率值，在所述待识别结果集合中选定目标自然语言数据，并将所述目标自然语言数据确定为与所述待识别手语视频对应的识别结果

图1

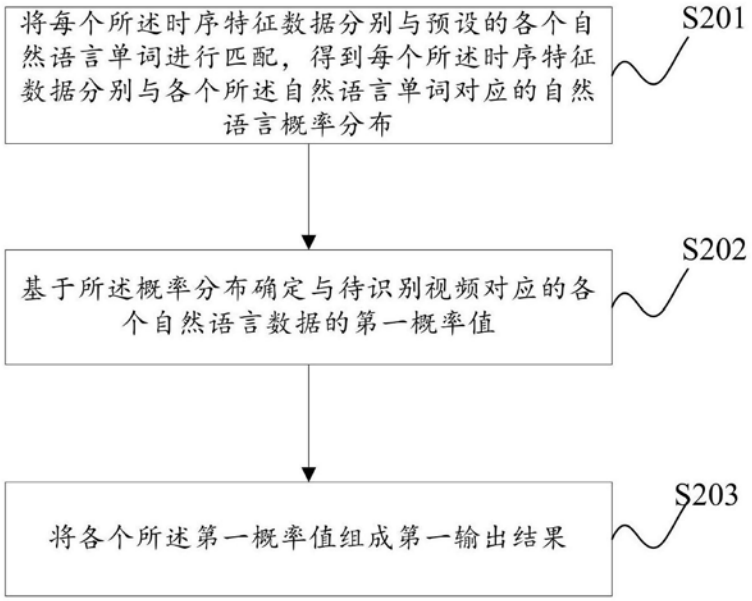

将每个所述时序特征数据分别与预设的各个自然语言单词进行匹配，得到每个所述时序特征数据分别与各个所述自然语言单词对应的自然语言概率分布
S201
基于所述概率分布确定与待识别视频对应的各个自然语言数据的第一概率值
S202
将各个所述第一概率值组成第一输出结果
S203

图2

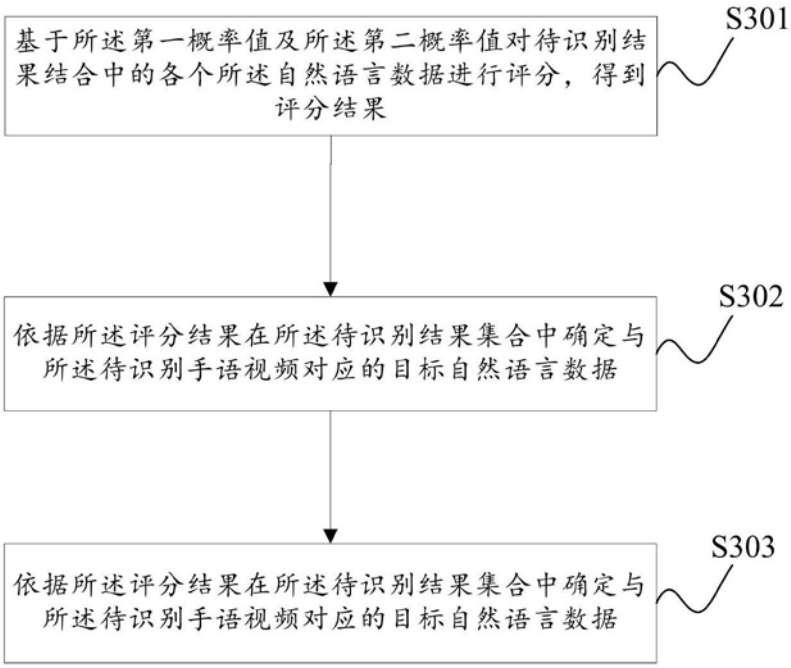

基于所述第一概率值及所述第二概率值对待识别结果结合中的各个所述自然语言数据进行评分，得到评分结果
S301
依据所述评分结果在所述待识别结果集合中确定与所述待识别手语视频对应的目标自然语言数据
S302
依据所述评分结果在所述待识别结果集合中确定与所述待识别手语视频对应的目标自然语言数据
S303

图3

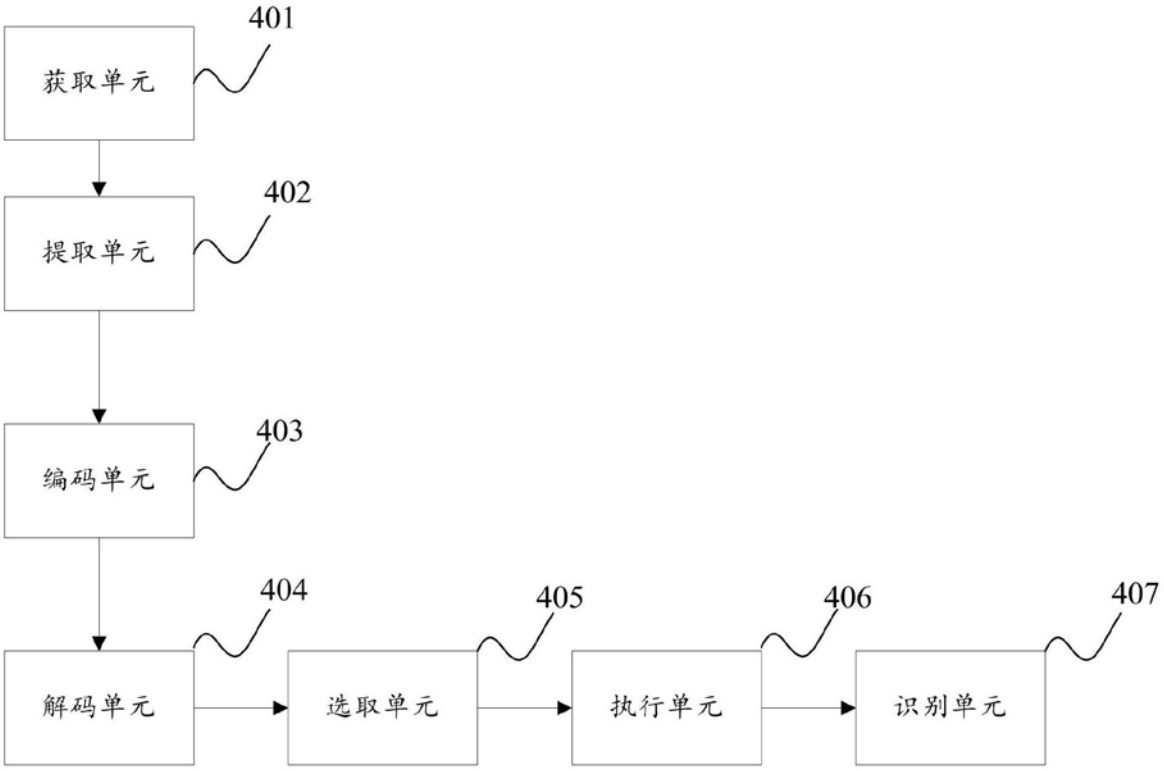

401
获取单元
402
提取单元
403
编码单元
404
解码单元
405
选取单元
406
执行单元
407
识别单元

图4

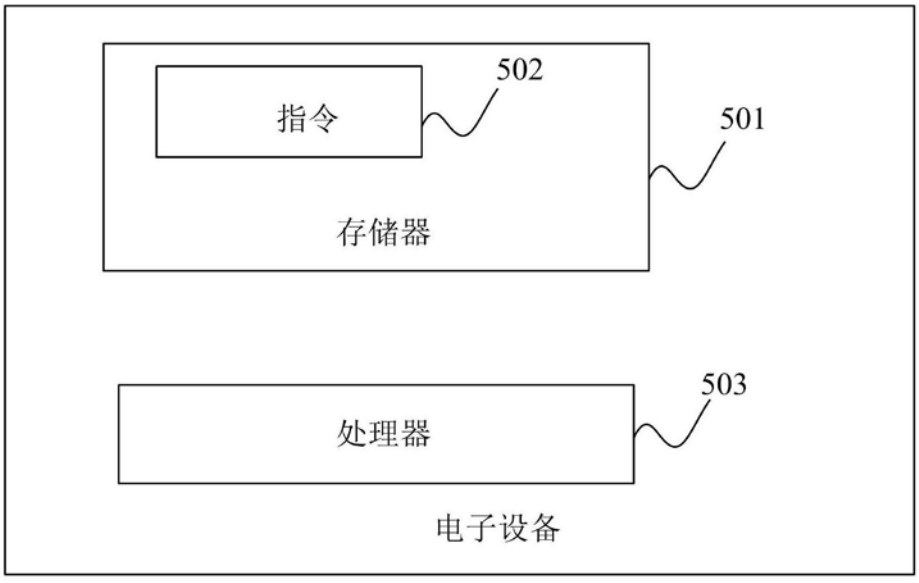

502
指令
501
存储器
503
处理器
电子设备

图5