

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2016 年 6 月 2 日 (02.06.2016)



(10) 国际公布号
WO 2016/082267 A1

- (51) 国际专利分类号:
G10L 15/25 (2013.01) G10L 15/26 (2006.01)
- (21) 国际申请号: PCT/CN2014/094624
- (22) 国际申请日: 2014 年 12 月 23 日 (23.12.2014)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201410714386.2 2014 年 11 月 28 日 (28.11.2014) CN
- (71) 申请人: 深圳创维-RGB 电子有限公司 (SHENZHEN SKYWORTH-RGB ELECTRONIC CO., LTD.) [CN/CN]; 中国广东省深圳市南山区深南大道创维大厦 A 座 13-16 楼, Guangdong 518057 (CN)。
- (72) 发明人: 付春元 (FU, Chunyuan); 中国广东省深圳市南山区深南大道创维大厦 A 座 13-16 楼, Guangdong 518057 (CN)。
- (74) 代理人: 深圳市世纪恒程知识产权代理事务所 (CENFO INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市南山区南山大道 3838 号设计产业园金栋二层 210-212 (原南头城工业村 11 栋), Guangdong 518052 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

(54) Title: VOICE RECOGNITION METHOD AND SYSTEM

(54) 发明名称: 语音识别方法和系统

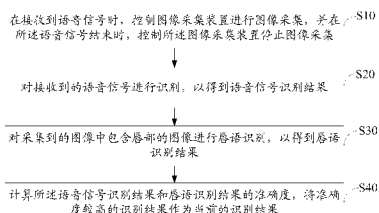


图 1 /FIG. 1

S10 Upon receipt of a voice signal, control an image acquisition device to acquire an image; control the image acquisition device to stop image acquiring when the voice signal is over
S20 Recognize the received voice signal to acquire a voice signal recognition result
S30 Conduct lip-reading recognition of the acquired image comprising a lip image to acquire a lip recognition result
S40 Calculate accuracy of the voice signal recognition result and a lip-reading recognition result, take the recognition result with a higher accuracy as the current voice recognition result

(57) Abstract: A voice recognition method and system, the method comprising: upon receipt of a voice signal, controlling an image acquisition device to acquire an image; controlling the image acquisition device to stop image acquiring when the voice signal is over (S10); recognizing the received voice signal to acquire a voice signal recognition result (S20); conducting lip-reading recognition of the acquired image comprising a lip image to acquire a lip recognition result (S30); calculating accuracy of the voice signal recognition result and a lip-reading recognition result, taking the recognition result with a higher accuracy as the current voice recognition result (S40). The technical solution improves the voice recognition accuracy.

(57) 摘要: 一种语音识别方法和系统, 在接收到语音信号时, 控制图像采集装置进行图像采集, 并在所述语音信号结束时, 控制所述图像采集装置停止图像采集 (S10); 对接收到的语音信号进行识别, 以得到语音信号识别结果 (S20); 对采集到的图像中包含唇部的图像进行唇语识别, 以得到唇语识别结果 (S30); 计算所述语音信号识别结果和唇语识别结果的准确度, 将准确度较高的识别结果作为当前的语音识别结果 (S40)。该技术方案提高了语音识别的准确性。

WO 2016/082267 A1

说明书

发明名称：语音识别方法和系统

[1] 技术领域

[2] 本发明涉及语音控制领域，尤其涉及语音识别方法和系统。

[3] 背景技术

[4] 随着语音交互的飞速发展，通过语音的方式控制终端（如电视以及空调器等），或者通过语音的方式进行数据输入已成为应用非常广泛的方式。目前，语音交互仍存在诸多问题，如语音识别不准确，易受环境影响较大，例如周围有人声嘈杂或者有背景音乐的话，语音采集装置采集到的语音信号包括人发出的语音信号以及周围的噪音信号，使得终端无法准确识别接受到的语音信号，导致语音识别不够准确。

[5] 发明内容

[6] 本发明的主要目的在于提出一种语音识别方法和系统，旨在解决语音识别不够准确的技术问题。

[7] 为实现上述目的，本发明提供的一种语音识别方法，所述语音识别方法包括以下步骤：

[8] 在接收到语音信号时，控制图像采集装置进行图像采集，并在所述语音信号结束时，控制所述图像采集装置停止图像采集；

[9] 对接收到的语音信号进行识别，以得到语音信号识别结果；

[10] 对采集到的图像中包含唇部的图像进行唇语识别，以得到唇语识别结果；

[11] 计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果。

[12] 优选地，所述对采集到的图像中包含唇部的图像进行唇语识别，以得到唇语识别结果的步骤包括：

[13] 确定采集到的图像中包含唇部的图像，将所述包含唇部的图像作为有效图像，并确定所述有效图像中唇部的位置；

[14] 根据每一帧所述有效图像的唇形以及上一帧所述有效图像的唇形确定用户输出

的字符；

- [15] 基于每一帧所述有效图像对应的字符组成唇语识别结果。
- [16] 优选地，所述确定采集到的图像中包含唇部的图像，将所述包含唇部的图像作为有效图像，并确定所述有效图像中唇部位置的步骤包括：
 - [17] 确定采集到的每帧图像中脸部轮廓；
 - [18] 将脸部轮廓内的各个像素点色度值与预存的人脸中各个像素点的色度值进行比对，以确定采集到的每帧图像中脸部位置；
 - [19] 确定脸部位置中眼部位置，并基于眼部位置以及唇部位置之间的相对位置确定唇部区域；
 - [20] 将唇部区域中各个像素点的RGB色度值进行比对；
 - [21] 在唇部区域存在RGB色度值满足预设条件的像素点时，确定该帧图像为包含唇部的图像，将所述包含唇部的图像作为有效图像；
 - [22] 基于唇部区域中各个像素点的RGB色度值确定唇部的位置。
- [23] 优选地，所述对接收到的语音信号进行识别，以得到语音信号识别结果的步骤包括：
 - [24] 将接收到的语音信号转换成字符串，并按照预设的关键词库，将所述字符串拆分为多个关键词；
 - [25] 标注各个所述关键词的词性，确定各个相邻的关键词之间的词性是否匹配；
 - [26] 在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词；
 - [27] 在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；
 - [28] 将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。
- [29] 优选地，所述对接收到的语音信号进行识别，以得到语音信号识别结果的步骤还包括：
 - [30] 在替换后的第二关键词与相邻关键词之间词性不匹配，且所述第二关键词存在

多个时，将所述第一关键词替换为其它第二关键词，并确定替换后的第二关键词与相邻关键词之间词性是否匹配，直至替换完所有的第二关键词，将转换得到的字符串作为当前的语音信号识别结果。

[31] 优选地，所述计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果的步骤包括：

[32] 将所述语音信号识别以及唇语识别结果拆分为多个关键词；

[33] 确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；

[34] 对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述唇语识别结果的准确度；

[35] 将准确度较高的识别结果作为当前的语音识别结果。

[36] 此外，为实现上述目的，本发明还提出一种语音识别系统，其特征在于，所述语音识别系统包括：

[37] 控制模块，用于在接收到语音信号时，控制图像采集装置进行图像采集，并在所述语音信号结束时，控制所述图像采集装置停止图像采集；

[38] 语音信号识别模块，用于对接收到的语音信号进行识别，以得到语音信号识别结果；

[39] 唇语识别模块，用于对采集到的图像中包含唇部的图像进行唇语识别，以得到唇语识别结果；

[40] 处理模块，用于计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果。

[41] 优选地，所述唇语识别模块包括：

[42] 唇部定位子模块，用于确定采集到的图像中包含唇部的图像，将所述包含唇部的图像作为有效图像，并确定所述有效图像中唇部位置；

[43] 确定子模块，用于根据每一帧所述有效图像的唇形以及上一帧所述有效图像的唇形确定用户输出的字符；

[44] 重组子模块，用于基于每一帧所述有效图像对应的字符组成唇语识别结果。

- [45] 优选地，所述唇部定位子模块包括：
- [46] 脸部轮廓确定单元，用于确定采集到的每帧图像中脸部轮廓；
- [47] 脸部位置定位单元，用于将确定的脸部轮廓内的各个像素点色度值与预存的人脸中各个像素点的色度值进行比对，以确定采集到的每帧图像中脸部位置；
- [48] 唇部区域定位单元，用于确定脸部位置中眼部位置，并基于眼部位置以及唇部位置之间的相对位置确定唇部区域；
- [49] 比对单元，用于将唇部区域中各个像素点的RGB色度值进行比对；
- [50] 处理单元，用于在唇部区域存在RGB色度值满足预设条件的像素点时，确定该帧图像为包含唇部的图像，将所述包含唇部的图像作为有效图像；
- [51] 唇部位置定位单元，用于基于唇部区域中各个像素点的RGB色度值确定唇部的位置。
- [52] 优选地，所述语音信号识别模块包括：
- [53] 转换子模块，用于将接收到的语音信号转换成字符串；
- [54] 拆分子模块，按照预设的关键词库，将所述字符串拆分为多个关键词；
- [55] 词性匹配子模块，用于标注各个所述关键词的词性，，确定各个相邻的关键词之间的词性是否匹配；
- [56] 确定子模块，用于在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词，以及在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；
- [57] 处理子模块，用于将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。
- [58] 优选地，所述处理子模块还用于在替换后的第二关键词与相邻关键词之间词性不匹配，且所述第二关键词存在多个时，将所述第一关键词替换为其它第二关键词，并确定替换后的第二关键词与相邻关键词之间词性是否匹配，直至替换完所有的第二关键词，并将转换得到的字符串作为当前的语音信号识别结果。

- [59] 优选地，所述处理模块包括：
- [60] 拆分子模块，用于将所述语音信号识别以及唇语识别结果拆分为多个关键词；
- [61] 关联度计算子模块，用于确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；
- [62] 准确度计算子模块，用于对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述唇语识别结果的准确度；
- [63] 处理子模块，用于将准确度较高的识别结果作为当前的语音识别结果。
- [64] 本发明提出的语音识别方法和系统，同时进行语音信号以及唇语的识别，并计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的识别结果，而不是仅仅单一识别语音信号，提高了语音识别的准确性。
- [65] 附图说明
- [66] 图1为本发明语音识别方法较佳实施例的流程示意图；
- [67] 图2为图1中步骤S20的细化流程示意图；
- [68] 图3为图1中步骤S30的细化流程示意图；
- [69] 图4为图3中步骤S31的细化流程示意图；
- [70] 图5为图1中步骤S40的细化流程示意图；
- [71] 图6为本发明语音识别系统较佳实施例的功能模块示意图；
- [72] 图7为图6中语音信号识别模块的细化功能模块示意图；
- [73] 图8为图6中唇语识别模块的细化功能模块示意图；
- [74] 图9为图8中唇部定位子模块的细化功能模块示意图；
- [75] 图10为图6中处理模块的细化功能模块示意图。
- [76] 本发明目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。
- [77] 具体实施方式
- [78] 应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

- [79] 本发明提供一种语音识别方法。
- [80] 参照图1，图1为本发明语音识别方法较佳实施例的流程示意图。
- [81] 本实施例提出的语音识别方法优选运行于被控终端（如电视机以及空调器等）中，被控终端基于语音识别接收进行相应的操作；或者语音识别方法可运行于控制终端，控制终端将语音信号识别结果对应的代码传输至相应的被控终端。
- [82] 本实施例提出一种语音识别方法，所述语音识别方法包括：
- [83] 步骤S10，在接收到语音信号时，控制图像采集装置进行图像采集，并在所述语音信号结束时，控制所述图像采集装置停止图像采集；
- [84] 在本实施例中，仅在接收到语音信号时才控制图像采集装置进行图像采集，而在未接受语音信号时处于休眠状态，以减少能耗，例如，在预设时间间隔内未接受到语音信号时，控制所述图像采集装置进入休眠状态。
- [85] 本领域技术人员可以理解的是，可控制图像采集装置实时或定时进行图像采集，在接收到语音信号时，确定接收到的语音信号的第一时间点以及语音信号结束的第二时间点，获取图像采集装置在该第一时间点以及第二时间点之间采集到的图像。
- [86] 步骤S20，对接收到的语音信号进行识别，以得到语音信号识别结果；
- [87] 在本实施例中，可通过将语音信号转换为字符信号得到语音信号识别结果。进一步地，为提高语音信号识别结果的准确性，可对语音信号转换的字符串进行纠错，具体纠错过程参照图2，所述步骤S20包括：
- [88] 步骤S21，将接收到的语音信号转换成字符串，并按照预设的关键词库，将所述字符串拆分为多个关键词；
- [89] 可预设包括多个关键词的关键词库，将语音信号转换得到的字符串与词库中存储的关键词进行比对，并确定预设的关键词库中与字符串匹配的关键词，并将该字符串拆分为各个匹配的关键词。本领域技术人员可以理解的是，关键词库中可不用设置数字类的关键词，在确定与字符串匹配的关键词后，可先提取字符串中匹配的关键词，并将字符串中剩余的不匹配的部分作为一个关键词。例如，语音信号转换得到的字符串为“电视机，切换至23频道”，则该字符串与预设的关键词库中的关键词匹配的为“电视机、切换、至以及频道”，则由字符串中直

接提取出“电视机、切换、至以及频道”，然后将剩余的“23”作为一个关键词。

- [90] 步骤S22，标注各个所述关键词的词性，确定各个相邻的关键词之间的词性是否匹配；
- [91] 该关键词的词性可为名词、动词、形容词、副词以及介词等，可预约各类词性的搭配，例如在相邻的关键词为动词+形容词时，则认为相邻的关键词之间词性不匹配，可能存在识别错误。
- [92] 步骤S23，在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词；
- [93] 步骤S24，在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；
- [94] 在本实施例中，可预设混淆音词库，该混淆音词库中可设置在语音信号转换为字符串时容易混淆的关键词，各个易混淆的关键词关联保存。在相邻的关键词不匹配时，可将该不匹配的关键词作为第一关键词与混淆音词库中的关键词进行比对，以对错误的关键词进行纠错。
- [95] 本领域技术人员可以理解的是，在所述混淆音词库中不存在所述不匹配的关键词时，可将转换得到的字符串作为当前的语音信号识别结果。
- [96] 步骤S25，将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。
- [97] 本领域技术人员可以理解的是，在替换后的第二关键词与相邻关键词之间词性不匹配，且所述第二关键词存在多个时，将所述第一关键词替换为其它第二关键词，并确定替换后的第二关键词与相邻关键词之间词性是否匹配，直至替换完所有的第二关键词，将转换得到的字符串作为当前的语音信号识别结果。
- [98] 步骤S30，对采集到的图像中包含唇部的图像进行唇语识别，以得到唇语识别结果；
- [99] 在本实施例中，唇语识别结果可根据每一帧图像中的唇形以及与上一帧的图像中的唇形确定得到，具体过程如图3所示，所述步骤S30包括：

- [100] 步骤S31，确定采集到的图像中包含唇部的图像，将所述包含唇部的图像作为有效图像，并确定所述有效图像中唇部的位置；
- [101] 在本实施例中，确定采集到的每帧图像中唇部的位置具体过程如图4所示，具体过程如下：
- [102] 步骤S311，确定采集到的每帧图像中脸部轮廓；
- [103] 由于采集到的图像中的各个像素点对应的色度值不同，可直接根据每帧图像中的像素点的色度值分布以及预设的脸部轮廓得到每帧图像中脸部位置。
- [104] 本领域技术人员可以理解的是，在图像采集装置的图像采集区域中有多个人存在时，可基于接收到语音信号对声源方向进行定位，基于确定的声源方向确定用户在采集到的图像中的位置，基于声源方向确定用户在采集到的图像中的位置属于现有技术，在此不再赘述。
- [105] 在采集到的图像中没有脸部轮廓时，直接将语音信号对应的语音信号识别结果作为当前的语音识别结果，或者可提示用户重新输入语音信号。
- [106] 步骤S312，将确定的脸部轮廓内的各个像素点色度值与预存的人脸中各个像素点的色度值进行比对，以确定采集到的每帧图像中脸部位置；
- [107] 确定脸部轮廓内各个像素点的YUV色度值至与预存的人脸中各个像素点的YUV色度值之间的相似度，在相似度大于预设值时，认为该像素点为脸部像素点，还相似度的计算公式为现有技术，在此不再赘述。
- [108] 步骤S313，确定脸部位置中眼部位置，并基于眼部位置以及唇部位置之间的相对位置确定唇部区域；
- [109] 在本实施例中，由于眼部的像素点的灰度值小于脸部其它位置的灰度值，可根据各个像素点之间的灰度值确定眼部位置，在眼部位置的下方以及脸部的下三分之一出，即可容易确定出唇部所在的区域。
- [110] 步骤S314，在唇部区域存在RGB色度值满足预设条件的像素点时，确定该帧图像为包含唇部的图像，将所述包含唇部的图像作为有效图像；
- [111] 步骤S315，基于唇部区域中各个像素点的RGB色度值确定唇部的位置
- [112] 但由于当前确定的唇部区域仅仅为初步确定，该区域内有唇部的像素点以及脸部的像素点，则需要在该区域内确定唇部位置。由于唇部像素点的RGB色度值

中，B（蓝色）分量的远远大于G（绿色）分量，则预设的条件可设置为B（蓝色）分量与G（绿色）分量之间的差值大于预设值，而脸部的像素点中B（蓝色）分量的小于G（绿色）分量，则可通过对各个像素点的B分量以及G分量进行比对，以确定唇部位置。

[113] 步骤S32，根据每一帧所述有效图像的唇形以及上一帧所述有效图像的唇形确定用户输出的字符；

[114] 步骤S33，基于每一帧所述有效图像对应的字符组成唇语识别结果。

[115] 本领域技术人员可以理解的是，采集到的图像中第一帧图像的上一帧图像的唇形默认为闭嘴唇形，用户可基于上一帧图像以及该帧图像对应的唇形可得出用户的唇部走势，将得到的唇部走势与预存的唇部走势进行比对，以得到当前输出的字符。按照每一帧图像的采集顺序，将每一帧图像的文件组合成唇语识别结果

[116] 步骤S40，计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的识别结果。

[117] 在本实施例中，计算所述语音信号识别结果和唇语识别结果的准确度的具体过程如图5所示，具体过程如下：

[118] 步骤S41，将所述语音信号识别以及唇语识别结果拆分为多个关键词；

[119] 关键词拆分的过程见上述语音信号关键词拆分过程，在此不再赘述。

[120] 步骤S42，确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词之间的第一关联度，并确定所述唇语识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；

[121] 在本实施例中，第一关联度的计算公式为： $I(x, y) = \log_2 p(x, y) / p(x) * p(y)$ ， $p(x)$ 为相邻的两个关键词x、y中，关键词x在字符串中出现的次数， $p(y)$ 为相邻的两个关键词x、y中，关键词y在字符串中出现的次数， $p(x, y)$ 为相邻的两个关键词x、y以相邻的方式在字符串中同时出现的次数。第二关联度与第一关联度计算的方式相同，在此不再赘述。

[122] 步骤S43，对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述唇语识别结果的准确度；

- [123] 在本实施例中，对字符串中各个相邻的关键词进行第一关联度计算，可得到多个第一关联度，对计算得到的关联度求和可得到字符串总体的准确度。
- [124] 步骤S44，将准确度较高的识别结果作为当前的语音识别结果。
- [125] 本实施例提出的语音识别方法，同时进行语音信号以及唇语的识别，并计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的识别结果，而不是仅仅单一识别语音信号，提高了语音识别的准确性。
- [126] 本发明进一步提供一种语音识别系统。
- [127] 参照图6，图6为本发明语音识别系统较佳实施例的功能模块示意图。
- [128] 需要强调的是，对本领域的技术人员来说，图6所示功能模块图仅仅是一个较佳实施例的示例图，本领域的技术人员围绕图6所示的语音识别系统的功能模块，可轻易进行新的功能模块的补充；各功能模块的名称是自定义名称，仅用于辅助理解该语音识别系统的各个程序功能块，不用于限定本发明的技术方案，本发明技术方案的核心是，各自定义名称的功能模块所要达成的功能。
- [129] 本实施例提出的语音识别系统优选运行于被控终端（如电视机以及空调器等）中，被控终端基于语音识别接收进行相应的操作；或者语音识别系统可运行于控制终端，控制终端将语音信号识别结果对应的代码传输至相应的被控终端。
- [130] 本实施例提出一种语音识别系统，所述语音识别系统包括：
- [131] 控制模块10，用于在接收到语音信号时，控制图像采集装置进行图像采集，并在所述语音信号结束时，控制所述图像采集装置停止图像采集；
- [132] 在本实施例中，控制模块10仅在接收到语音信号时才控制图像采集装置进行图像采集，而在未接受语音信号时处于休眠状态，以减少能耗，例如，在预设时间间隔内未接受到语音信号时，控制模块10控制所述图像采集装置进入休眠状态。
- [133] 本领域技术人员可以理解的是，控制模块10可控制图像采集装置实时或定时进行图像采集，在接收到语音信号时，确定接收到的语音信号的第一时间点以及语音信号结束的第二时间点，获取图像采集装置在该第一时间点以及第二时间点之间采集到的图像。
- [134] 语音信号识别模块20，用于对接收到的语音信号进行识别，以得到语音信号识

别结果；

- [135] 在本实施例中，语音信号识别模块20可通过将语音信号转换为字符信号得到语音信号识别结果。进一步地，为提高语音信号识别结果的准确性，可对语音信号转换的字符串进行纠错，参照图7，所述语音信号识别模块20包括：
- [136] 转换子模块21，用于将接收到的语音信号转换成字符串；
- [137] 拆分子模块22，按照预设的关键词库，将所述字符串拆分为多个关键词；
- [138] 可预设包括多个关键词的关键词库，拆分子模块22将语音信号转换得到的字符串与词库中存储的关键词进行比对，并确定预设的关键词库中与字符串匹配的关键词，并将该字符串拆分为各个匹配的关键词。本领域技术人员可以理解的是，关键词库中可不用设置数字类的关键词，在确定与字符串匹配的关键词后，拆分子模块22可先提取字符串中匹配的关键词，并将字符串中剩余的不匹配的部分作为一个关键词。例如，语音信号转换得到的字符串为“电视机，切换至23频道”，则该字符串与预设的关键词库中的关键词匹配的为“电视机、切换、至以及频道”，则由字符串中直接提取出“电视机、切换、至以及频道”，然后将剩余的“23”作为一个关键词。
- [139] 词性匹配子模块23，用于标注各个所述关键词的词性，，确定各个相邻的关键词之间的词性是否匹配；
- [140] 该关键词的词性可为名词、动词、形容词、副词以及介词等，可预约各类词性的搭配，例如在相邻的关键词为动词+形容词时，则词性匹配子模块23认为相邻的关键词之间词性不匹配，可能存在识别错误。
- [141] 确定子模块24，用于在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词，以及在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；
- [142] 在本实施例中，可预设混淆音词库，该混淆音词库中可设置在语音信号转换为字符串时容易混淆的关键词，各个易混淆的关键词关联保存。在相邻的关键词不匹配时，可将该不配的关键词作为第一关键词与混淆音词库中的关键词进行比对，以对错误的关键词进行纠错。

- [143] 本领域技术人员可以理解的是，在所述混淆音词库中不存在所述不匹配的关键词时，可将转换得到的字符串作为当前的语音信号识别结果。
- [144] 处理子模块25，用于将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。
- [145] 本领域技术人员可以理解的是，在替换后的第二关键词与相邻关键词之间词性不匹配，且所述第二关键词存在多个时，处理子模块25将所述第一关键词替换为其它第二关键词，并确定替换后的第二关键词与相邻关键词之间词性是否匹配，直至替换完所有的第二关键词，处理子模块25将转换得到的字符串作为当前的语音信号识别结果。
- [146] 唇语识别模块30，用于对采集到的图像中包含唇部的图像进行唇语识别，以得到唇语识别结果；
- [147] 在本实施例中，唇语识别结果可根据每一帧图像中的唇形以及与上一帧的图像中的唇形确定得到，参照图8所示，所述唇语识别模块30包括：
- [148] 唇部定位子模块31，用于确定采集到的图像中包含唇部的图像，将所述包含唇部的图像作为有效图像，并确定所述有效图像中唇部位置；
- [149] 在本实施例中，确定采集到的每帧图像中唇部的位置具体单元如图9所示，所述唇部定位子模块31包括：
- [150] 脸部轮廓确定单元311，用于确定采集到的每帧图像中脸部轮廓；
- [151] 由于采集到的图像中的各个像素点对应的色度值不同，脸部轮廓确定单元311可直接根据每帧图像中的像素点的色度值分布以及预设的脸部轮廓得到每帧图像中脸部位置。
- [152] 本领域技术人员可以理解的是，在图像采集装置的图像采集区域中有多个人存在时，脸部轮廓确定单元311可基于接收到语音信号对声源方向进行定位，基于确定的声源方向确定用户在采集到的图像中的位置，基于声源方向确定用户在采集到的图像中的位置属于现有技术，在此不再赘述。
- [153] 在采集到的图像中没有脸部轮廓时，处理模块40直接将语音信号对应的语音信

号识别结果作为当前的语音识别结果，或者可提示用户重新输入语音信号。

[154] 脸部位置定位单元312，用于将确定的脸部轮廓内的各个像素点色度值与预存的人脸中各个像素点的色度值进行比对，以确定采集到的每帧图像中脸部位置；

[155] 确定脸部轮廓内各个像素点的YUV色度值至与预存的人脸中各个像素点的YUV色度值之间的相似度，在相似度大于预设值时，认为该像素点为脸部像素点，还相似度的计算公式为现有技术，在此不再赘述。

[156] 唇部区域定位单元313，用于确定脸部位置中眼部位置，并基于眼部位置以及唇部位置之间的相对位置确定唇部区域；

[157] 在本实施例中，由于眼部的像素点的灰度值小于脸部其它位置的灰度值，可根据各个像素点之间的灰度值确定眼部位置，在眼部位置的下方以及脸部的下三分之一出，即可容易确定出唇部所在的区域。

[158] 比对单元314，用于将唇部区域中各个像素点的RGB色度值进行比对；

[159] 处理单元315，用于在唇部区域存在RGB色度值满足预设条件的像素点时，确定该帧图像为包含唇部的图像，将所述包含唇部的图像作为有效图像；

[160] 唇部位置定位单元316，用于基于唇部区域中各个像素点的RGB色度值确定唇部的位置。

[161] 但由于当前确定的唇部区域仅仅为初步确定，该区域内有唇部的像素点以及脸部的像素点，则需要在该区域内确定唇部位置。由于唇部像素点的RGB色度值中，B（蓝色）分量的远远大于G（绿色）分量，则预设的条件可设置为B（蓝色）分量与G（绿色）分量之间的差值大于预设值，而脸部的像素点中B（蓝色）分量的小于G（绿色）分量，则可通过对各个像素点的B分量以及G分量进行比对，以确定唇部位置。

[162] 确定子模块32，用于根据每一帧所述有效图像的唇形以及上一帧所述有效图像的唇形确定用户输出的字符；

[163] 重组子模块33，用于基于每一帧所述有效图像对应的字符组成唇语识别结果。

[164] 本领域技术人员可以理解的是，采集到的图像中第一帧图像的上一帧图像的唇形默认为闭嘴唇形，用户可基于上一帧图像以及该帧图像对应的唇形可得出用

户的唇部走势，将得到的唇部走势与预存的唇部走势进行比对，以得到当前输出的字符。按照每一帧图像的采集顺序，将每一帧图像的文件组合成唇语识别结果

[165] 处理模块40，用于计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果。

[166] 在本实施例中，参照图10，所述处理模块40包括：

[167] 拆分子模块41，用于将所述语音信号识别以及唇语识别结果拆分为多个关键词；

[168] 关键词拆分的过程见上述语音信号关键词拆分过程，在此不再赘述。

[169] 关联度计算子模块42，用于确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；

[170] 在本实施例中，第一关联度的计算公式为：， $p(x)$ 为相邻的两个关键词 x 、 y 中，关键词 x 在字符串中出现的次数， $p(y)$ 为相邻的两个关键词 x 、 y 中，关键词 y 在字符串中出现的次数， $p(x, y)$ 为相邻的两个关键词 x 、 y 以相邻的方式在字符串中同时出现的次数。第二关联度与第一关联度计算的方式相同，在此不再赘述。

[171] 准确度计算子模块43，用于对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述唇语识别结果的准确度；

[172] 在本实施例中，对字符串中各个相邻的关键词进行第一关联度计算，可得到多个第一关联度，对计算得到的关联度求和可得到字符串总体的准确度。

[173] 处理子模块44，用于将准确度较高的识别结果作为当前的语音识别结果。

[174] 本实施例提出的语音识别系统，同时进行语音信号以及唇语的识别，并计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的识别结果，而不是仅仅单一识别语音信号，提高了语音识别的准确性。

[175] 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、物品或者系统不仅

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、物品或者系统所固有的要素。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、物品或者系统中还存在另外的相同要素。

[176] 上述本发明实施例序号仅仅为了描述，不代表实施例的优劣。

[177] 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。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，该计算机软件产品存储在一个存储介质（如ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端设备（可以是手机，计算机，服务器，空调器，或者网络设备等等）执行本发明各个实施例所述的方法。

[178] 以上仅为本发明的优选实施例，并非因此限制本发明的专利范围，凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本发明的专利保护范围内。

权利要求书

- [权利要求 1] 一种语音识别方法，其特征在于，所述语音识别方法包括以下步骤：
- 在接收到语音信号时，控制图像采集装置进行图像采集，并在所述语音信号结束时，控制所述图像采集装置停止图像采集；
- 对接收到的语音信号进行识别，以得到语音信号识别结果；
- 对采集到的图像中包含唇部的图像进行唇语识别，以得到唇语识别结果；
- 计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果。
- [权利要求 2] 如权利要求1所述的语音识别方法，其特征在于，所述对采集到的图像中包含唇部的图像进行唇语识别，以得到唇语识别结果的步骤包括：
- 确定采集到的图像中包含唇部的图像，将所述包含唇部的图像作为有效图像，并确定所述有效图像中唇部的位置；
- 根据每一帧所述有效图像的唇形以及上一帧所述有效图像的唇形确定用户输出的字符；
- 基于每一帧所述有效图像对应的字符组成唇语识别结果。
- [权利要求 3] 如权利要求2所述的语音识别方法，其特征在于，所述确定采集到的图像中包含唇部的图像，将所述包含唇部的图像作为有效图像，并确定所述有效图像中唇部位置的步骤包括：
- 确定采集到的每帧图像中脸部轮廓；
- 将脸部轮廓内的各个像素点色度值与预存的人脸中各个像素点的色度值进行比对，以确定采集到的每帧图像中脸部位置；
- 确定脸部位置中眼部位置，并基于眼部位置以及唇部位置之间的相对位置确定唇部区域；
- 将唇部区域中各个像素点的RGB色度值进行比对；
- 在唇部区域存在RGB色度值满足预设条件的像素点时，确定该帧

图像为包含唇部的图像，将所述包含唇部的图像作为有效图像；
基于唇部区域中各个像素点的RGB色度值确定唇部的位置。

[权利要求 4]

如权利要求1所述的语音识别方法，其特征在于，所述对接收到的语音信号进行识别，以得到语音信号识别结果的步骤包括：
将接收到的语音信号转换成字符串，并按照预设的关键词库，将所述字符串拆分为多个关键词；
标注各个所述关键词的词性，确定各个相邻的关键词之间的词性是否匹配；
在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词；
在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；
将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。

[权利要求 5]

如权利要求4所述的语音识别方法，其特征在于，所述对接收到的语音信号进行识别，以得到语音信号识别结果的步骤还包括：
在替换后的第二关键词与相邻关键词之间词性不匹配，且所述第二关键词存在多个时，将所述第一关键词替换为其它第二关键词，并确定替换后的第二关键词与相邻关键词之间词性是否匹配，直至替换完所有的第二关键词，将转换得到的字符串作为当前的语音信号识别结果。

[权利要求 6]

如权利要求如权利要求2所述的语音识别方法，其特征在于，所述对接收到的语音信号进行识别，以得到语音信号识别结果的步骤包括：
将接收到的语音信号转换成字符串，并按照预设的关键词库，将

所述字符串拆分为多个关键词；

标注各个所述关键词的词性，确定各个相邻的关键词之间的词性是否匹配；

在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词；

在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；

将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。

[权利要求 7]

如权利要求如权利要求3所述的语音识别方法，其特征在于，所述对接收到的语音信号进行识别，以得到语音信号识别结果的步骤包括：

将接收到的语音信号转换成字符串，并按照预设的关键词库，将所述字符串拆分为多个关键词；

标注各个所述关键词的词性，确定各个相邻的关键词之间的词性是否匹配；

在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词；

在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；

将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。

- [权利要求 8] 如权利要求1所述的语音识别方法，其特征在于，所述计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果的步骤包括：
- 将所述语音信号识别以及唇语识别结果拆分为多个关键词；
- 确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；
- 对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述唇语识别结果的准确度；
- 将准确度较高的识别结果作为当前的语音识别结果。
- [权利要求 9] 如权利要求2所述的语音识别方法，其特征在于，所述计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果的步骤包括：
- 将所述语音信号识别以及唇语识别结果拆分为多个关键词；
- 确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；
- 对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述唇语识别结果的准确度；
- 将准确度较高的识别结果作为当前的语音识别结果。
- [权利要求 10] 如权利要求3所述的语音识别方法，其特征在于，所述计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果的步骤包括：
- 将所述语音信号识别以及唇语识别结果拆分为多个关键词；
- 确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语识别结果拆分成的关键词

中，各个相邻关键词的之间的第二关联度；
对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述语音信号识别结果的准确度；
将准确度较高的识别结果作为当前的语音识别结果。

[权利要求 11]

一种语音识别系统，其特征在于，所述语音识别系统包括：
控制模块，用于在接收到语音信号时，控制图像采集装置进行图像采集，并在所述语音信号结束时，控制所述图像采集装置停止图像采集；
语音信号识别模块，用于对接收到的语音信号进行识别，以得到语音信号识别结果；
唇语识别模块，用于对采集到的图像中包含唇部的图像进行唇语识别，以得到唇语识别结果；
处理模块，用于计算所述语音信号识别结果和唇语识别结果的准确度，将准确度较高的识别结果作为当前的语音识别结果。

[权利要求 12]

如权利要求11所述的语音识别系统，其特征在于，所述唇语识别模块包括：
唇部定位子模块，用于确定采集到的图像中包含唇部的图像，将所述包含唇部的图像作为有效图像，并确定所述有效图像中唇部位置；
确定子模块，用于根据每一帧所述有效图像的唇形以及上一帧所述有效图像的唇形确定用户输出的字符；
重组子模块，用于基于每一帧所述有效图像对应的字符组成唇语识别结果。

[权利要求 13]

如权利要求12所述的语音识别系统，其特征在于，所述唇部定位子模块包括：
脸部轮廓确定单元，用于确定采集到的每帧图像中脸部轮廓；
脸部位置定位单元，用于将确定的脸部轮廓内的各个像素点色度

值与预存的人脸中各个像素点的色度值进行比对，以确定采集到的每帧图像中脸部位置；

唇部区域定位单元，用于确定脸部位置中眼部位置，并基于眼部位置以及唇部位置之间的相对位置确定唇部区域；

比对单元，用于将唇部区域中各个像素点的RGB色度值进行比对；

处理单元，用于在唇部区域存在RGB色度值满足预设条件的像素点时，确定该帧图像为包含唇部的图像，将所述包含唇部的图像作为有效图像；

唇部位置定位单元，用于基于唇部区域中各个像素点的RGB色度值确定唇部的位置。

[权利要求 14]

如权利要求11所述的语音识别系统，其特征在于，所述语音信号识别模块包括：

转换子模块，用于将接收到的语音信号转换成字符串；

拆分子模块，按照预设的关键词库，将所述字符串拆分为多个关键词；

词性匹配子模块，用于标注各个所述关键词的词性，，确定各个相邻的关键词之间的词性是否匹配；

确定子模块，用于在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词，以及在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；

处理子模块，用于将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。

[权利要求 15]

如权利要求14所述的语音识别系统，其特征在于，所述处理子模

块还用于在替换后的第二关键词与相邻关键词之间词性不匹配，且所述第二关键词存在多个时，将所述第一关键词替换为其它第二关键词，并确定替换后的第二关键词与相邻关键词之间词性是否匹配，直至替换完所有的第二关键词，并将转换得到的字符串作为当前的语音信号识别结果。

[权利要求 16]

如权利要求12所述的语音识别系统，其特征在于，所述语音信号识别模块包括：

转换子模块，用于将接收到的语音信号转换成字符串；

拆分子模块，按照预设的关键词库，将所述字符串拆分为多个关键词；

词性匹配子模块，用于标注各个所述关键词的词性，，确定各个相邻的关键词之间的词性是否匹配；

确定子模块，用于在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词，以及在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；

处理子模块，用于将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。

[权利要求 17]

如权利要求13所述的语音识别系统，其特征在于，所述语音信号识别模块包括：

转换子模块，用于将接收到的语音信号转换成字符串；

拆分子模块，按照预设的关键词库，将所述字符串拆分为多个关键词；

词性匹配子模块，用于标注各个所述关键词的词性，，确定各个相邻的关键词之间的词性是否匹配；

确定子模块，用于在有相邻的关键词之间的词性不匹配时，将所述不匹配关键词作为第一关键词，并确定预设的混淆音词库是否存在所述第一关键词，以及在所述混淆音词库存在所述不匹配的关键词时，确定所述混淆音词库中所述第一关键词对应的第二关键词；

处理子模块，用于将所述第一关键词替换为第二关键词，并在替换后的第二关键词与相邻关键词之间词性匹配时，将替换后的第二关键词以及其它关键词重新组合成为语音信号识别结果，并将重新组合的唇语识别结果作为当前的语音信号识别结果。

[权利要求 18]

如权利要求11所述的语音识别系统，其特征在于，所述处理模块包括：

拆分子模块，用于将所述语音信号识别以及唇语识别结果拆分为多个关键词；

关联度计算子模块，用于确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；

准确度计算子模块，用于对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述语音信号识别结果的准确度；

处理子模块，用于将准确度较高的识别结果作为当前的语音识别结果。

[权利要求 19]

如权利要求12所述的语音识别系统，其特征在于，所述处理模块包括：

拆分子模块，用于将所述语音信号识别以及唇语识别结果拆分为多个关键词；

关联度计算子模块，用于确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语

识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；

准确度计算子模块，用于对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述语音信号识别结果的准确度；

处理子模块，用于将准确度较高的识别结果作为当前的语音识别结果。

[权利要求 20]

如权利要求13所述的语音识别系统，其特征在于，所述处理模块包括：

拆分子模块，用于将所述语音信号识别以及唇语识别结果拆分为多个关键词；

关联度计算子模块，用于确定所述语音信号识别结果拆分成的关键词中，各个相邻关键词的之间的第一关联度，并确定所述唇语识别结果拆分成的关键词中，各个相邻关键词的之间的第二关联度；

准确度计算子模块，用于对确定的第一关联度求和，得到所述语音信号识别结果的准确度，并对确定的第二关联度求和，得到所述语音信号识别结果的准确度；

处理子模块，用于将准确度较高的识别结果作为当前的语音识别结果。

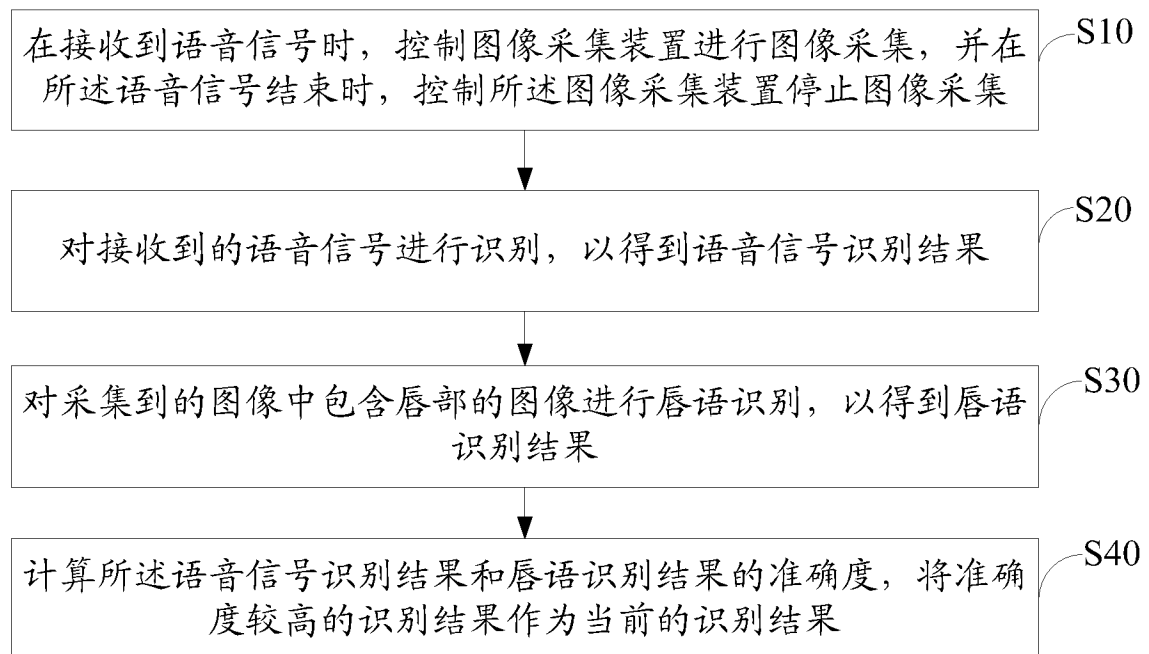


图 1

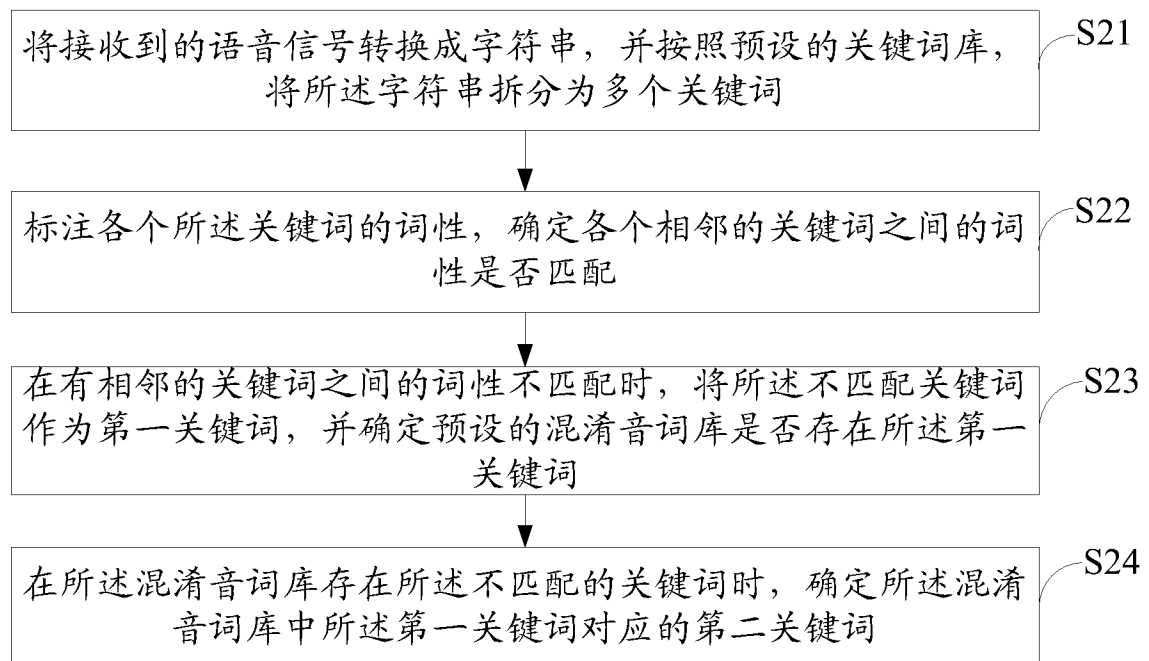
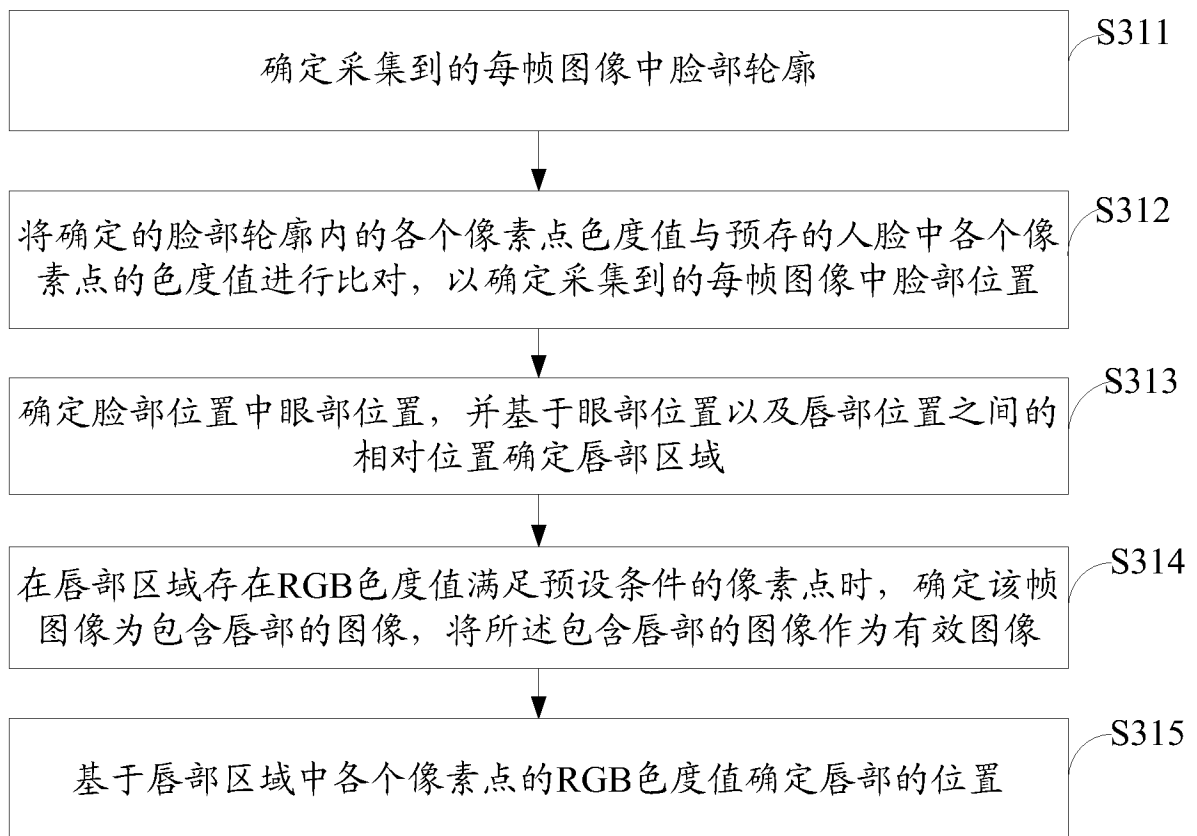
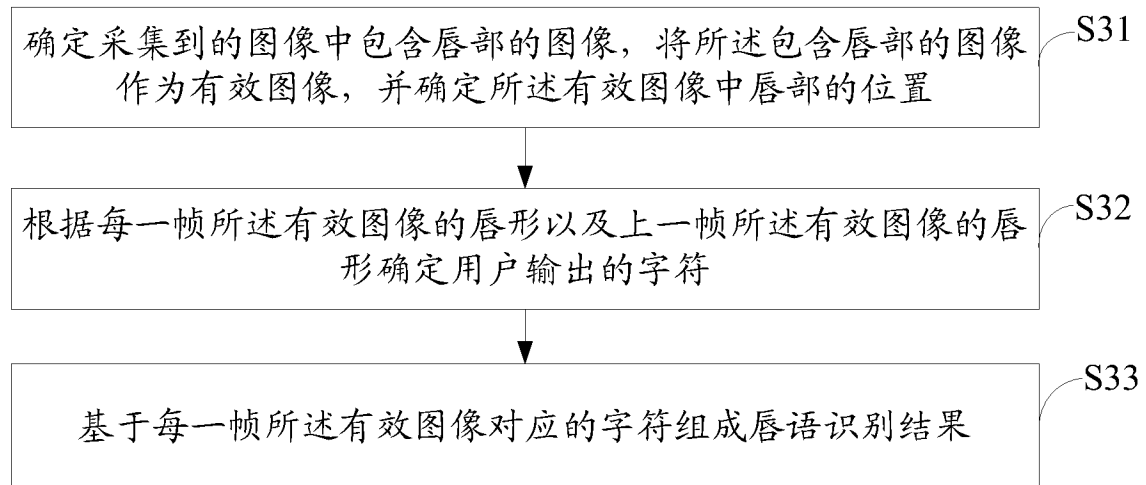


图 2



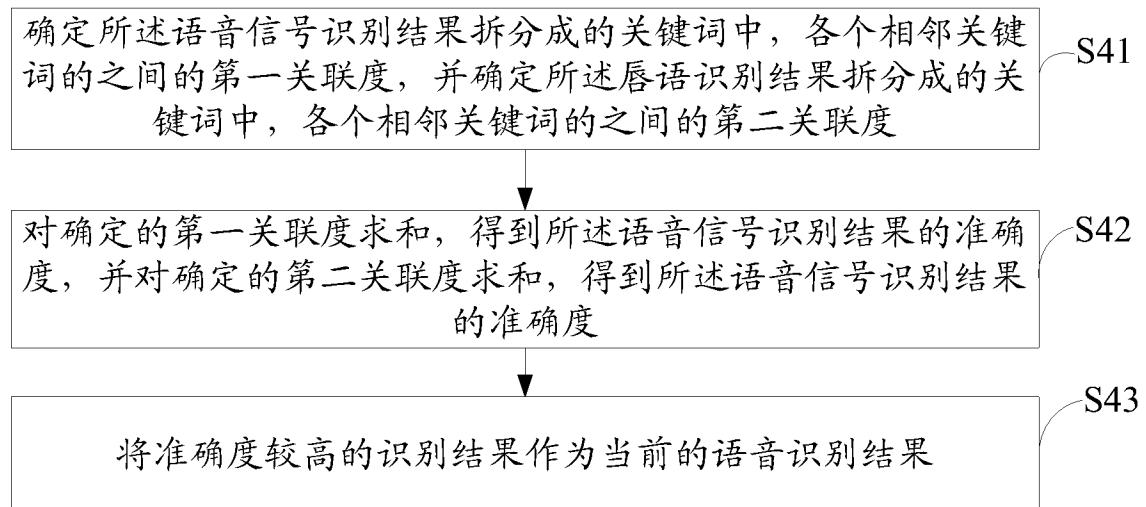


图 5

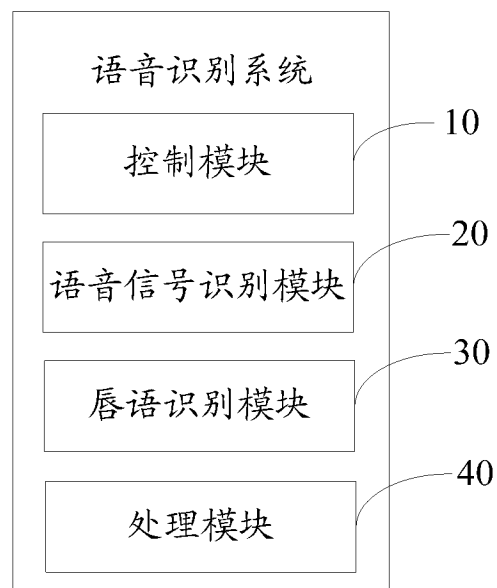


图 6



图 7

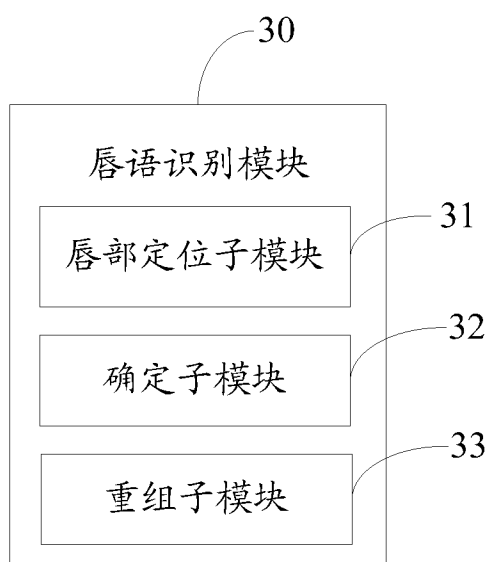


图 8

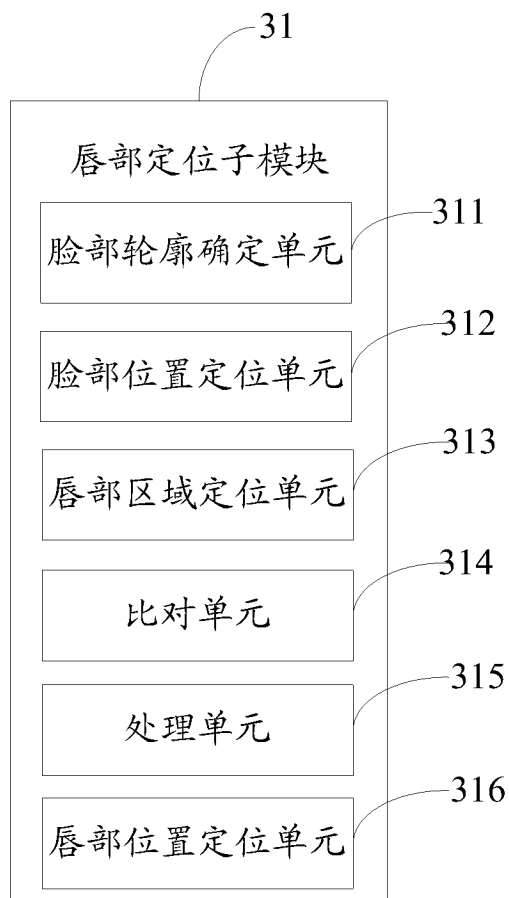


图 9

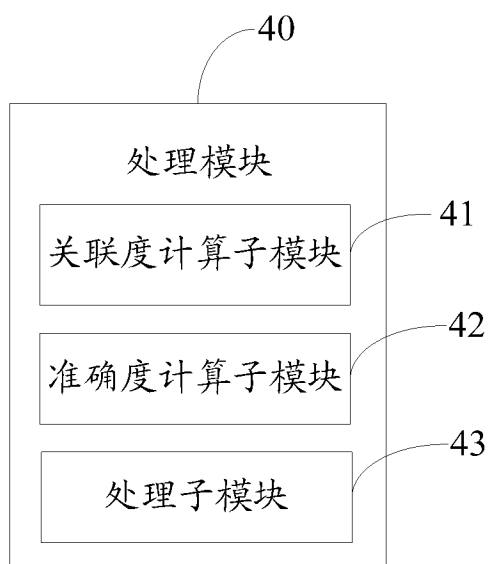


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/094624

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/25 (2013.01) i; G10L 15/26 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

SIPOABS, CNTXT, CJFD, VEN: recognition, LIP, MOUTH, FACE, IMAGE, KEYWORD, SPEECH, AUDIO, VIDEO

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2013054240 A1 (SAMSUNG ELECTRONICS CO., LTD.), 28 February 2013 (28.02.2013), figures 1-4, and description, paragraphs 36-56	1-20
Y	US 6633844 B1 (IBM CORP.), 14 October 2003 (14.10.2003), figures 1-2, and description, column 4, line 10 to column 6, line 21	1-20
Y	US 2003171932 A1 (JUANG, B. et al.), 11 September 2003 (11.09.2003), figures 1-3, and description, paragraphs 16-18	1-20
Y	US 2010063820 A1 (BROADCOM CORP.), 11 March 2010 (11.03.2010), figures 1-5, and description, paragraphs 19-44	1-20
Y	JP 2002304194 A (KUJIRADA MASANOBU), 18 October 2002 (18.10.2002), figure 1, and description, paragraphs 7-9	1-20
Y	CN 102023703 A (HYUNDAI MOTOR COMPANY et al.), 20 April 2011 (20.04.2011), figures 1-2, and description, paragraphs 47-69	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 21 August 2015 (21.08.2015)	Date of mailing of the international search report 09 September 2015 (09.09.2015)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer WANG, Xiaoping Telephone No.: (86-10) 62085739

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2014/094624

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
US 2013054240 A1	28 February 2013	KR 20130022607 A	07 March 2013
		EP 2562746 A1	27 February 2013
US 6633844 B1	14 October 2003	None	
US 2003171932 A1	11 September 2003	None	
US 2010063820 A1	11 March 2010	US 2004117191 A1	17 June 2004
		US 7587318 B2	08 September 2009
JP 2002304194 A	18 October 2002	None	
CN 102023703 A	20 April 2011	US 2011071830 A1	24 March 2011
		KR 20110032244 A	30 March 2011
		CN 102023703 B	11 March 2015
		US 8442820 B2	14 May 2013
		KR 1092820 B	12 December 2011

A. 主题的分类		
G10L 15/25 (2013.01)i; G10L 15/26 (2006.01)i		
按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类		
B. 检索领域		
检索的最低限度文献 (标明分类系统和分类号)		
G10L		
包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))		
SIPOABS, CNTXT, CJFD, VEN: 语音, 识别, 唇, 嘴, 脸, 图像, 关键词, LIP, MOUTH, FACE, IMAGE, KEYWORD, SPEECH, AUDIO, VIDEO		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	US 2013054240 A1 (SAMSUNG ELECTRONICS CO., LTD.) 2013年 2月 28日 (2013 - 02 - 28) 图1-4, 说明书第36-56段	1-20
Y	US 6633844 B1 (IBM CORP.) 2003年 10月 14日 (2003 - 10 - 14) 图1-2, 说明书第4栏第10行至第6栏第21行	1-20
Y	US 2003171932 A1 (JUANG B等) 2003年 9月 11日 (2003 - 09 - 11) 图1-3, 说明书第16-18段	1-20
Y	US 2010063820 A1 (BROADCOM CORP.) 2010年 3月 11日 (2010 - 03 - 11) 图1-5, 说明书第19-44段	1-20
Y	JP 2002304194 A (KUJIRADA MASANOBU) 2002年 10月 18日 (2002 - 10 - 18) 图1, 说明书第7-9段	1-20
Y	CN 102023703 A (现代自动车株式会社等) 2011年 4月 20日 (2011 - 04 - 20) 图1-2, 说明书第47-69	1-20
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期		国际检索报告邮寄日期
2015年 8月 21日		2015年 9月 9日
ISA/CN的名称和邮寄地址		授权官员
中华人民共和国国家知识产权局 (ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国		王晓萍
传真号 (86-10)62019451		电话号码 (86-10)62085739

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2014/094624

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
US	2013054240	A1	2013年 2月 28日	KR	20130022607	A	2013年 3月 7日
				EP	2562746	A1	2013年 2月 27日
US	6633844	B1	2003年 10月 14日	无			
US	2003171932	A1	2003年 9月 11日	无			
US	2010063820	A1	2010年 3月 11日	US	2004117191	A1	2004年 6月 17日
				US	7587318	B2	2009年 9月 8日
JP	2002304194	A	2002年 10月 18日	无			
CN	102023703	A	2011年 4月 20日	US	2011071830	A1	2011年 3月 24日
				KR	20110032244	A	2011年 3月 30日
				CN	102023703	B	2015年 3月 11日
				US	8442820	B2	2013年 5月 14日
				KR	1092820	B	2011年 12月 12日