(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112445450 A
(43)申请公布日 2021.03.05

(21)申请号 201910818670.7

(22)申请日 2019.08.30

(71)申请人 比亚迪股份有限公司
地址 518118 广东省深圳市坪山新区比亚迪路3009号

(72)发明人 周云霄

(74)专利代理机构 北京英创嘉友知识产权代理事务所(普通合伙) 11447
代理人 魏嘉熹

(51)Int.Cl.
G06F 3/16(2006.01)
G06F 3/0484(2013.01)
G06F 3/0488(2013.01)

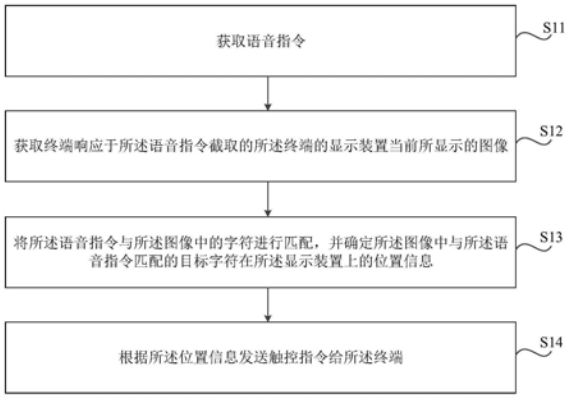
权利要求书2页 说明书8页 附图5页

(54)发明名称

基于语音控制终端的方法、装置、存储介质和电子设备

(57)摘要

本公开涉及一种基于语音控制终端的方法、装置、存储介质和电子设备，该方法包括：获取语音指令；获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像；将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息；根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。用于解决相关技术中，部分肢体残障人士难以通过触控屏幕控制终端的技术问题。



CN 112445450 A

1.一种基于语音控制终端的方法,其特征在于,所述方法包括:

获取语音指令;

获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像;

将所述语音指令与所述图像中的字符进行匹配,并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息;

根据所述位置信息发送触控指令给所述终端,以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。

2.根据权利要求1所述的方法,其特征在于,所述将所述语音指令与所述图像中的字符进行匹配,包括:

识别所述语音指令以得到所述语音指令中的目标关键词;

识别所述图像以得到所述图像中的目标字符;

基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配。

3.根据权利要求2所述的方法,其特征在于,所述识别所述语音指令以得到所述语音指令中的目标关键词,包括:

识别所述语音指令得到对应于该语音指令的文字信息;

基于预设规则从所述文字信息中提取出所述语音指令中的所述目标关键词。

4.根据权利要求2所述的方法,其特征在于,所述识别所述图像以得到所述图像中的目标字符,包括:

确定所述图像中包含所述目标字符的感兴趣区域;

对所述感兴趣区域进行二值化处理;

于所述二值化处理后的感兴趣区域内确定包含所述目标字符的最小矩形区域;

基于机器学习算法或光学字符识别OCR对所述最小矩形区域进行识别得到所述目标字符。

5.根据权利要求2所述的方法,其特征在于,所述基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配,包括:

确定与所述目标关键词满足预设关联规则的目标字符;

确定该目标字符为与该目标关键词所属的语音指令相匹配的字符。

6.根据权利要求2所述的方法,其特征在于,所述基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配之前,所述方法还包括:

若本次获取的所述图像与上一次获取的所述终端响应于上一条语音指令截取的所述终端的显示装置所显示的图像相同,不对所述图像进行识别,调用上一次识别图像得到的目标字符用于所述匹配。

7.一种基于语音控制终端的装置,其特征在于,所述装置包括:

语音指令获取模块,用于获取语音指令;

图像获取模块,用于获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像;

匹配模块,用于将所述语音指令与所述图像中的字符进行匹配,并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息;

发送模块,用于根据所述位置信息发送触控指令给所述终端,以控制所述终端模拟对

所述位置信息表征的触控区域进行触控操作。

8.根据权利要求7所述的装置,其特征在于,所述匹配模块包括:

语音指令识别子模块,用于识别所述语音指令以得到所述语音指令中的目标关键词;

图像识别子模块,用于识别所述图像以得到所述图像中的目标字符;

匹配子模块,用于基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配。

9.根据权利要求8所述的装置,其特征在于,所述语音指令识别子模块用于:

识别所述语音指令得到对应于该语音指令的文字信息;

基于预设规则从所述文字信息中提取出所述语音指令中的所述目标关键词。

10.根据权利要求8所述的装置,其特征在于,所述图像识别子模块用于:

确定所述图像中包含所述目标字符的感兴趣区域;

对所述感兴趣区域进行二值化处理;

于所述二值化处理后的感兴趣区域内确定包含所述目标字符的最小矩形区域;

基于机器学习算法或光学字符识别OCR对所述最小矩形区域进行识别得到所述目标字符。

11.根据权利要求8所述的装置,其特征在于,所述匹配子模块用于:

确定与所述目标关键词满足预设关联规则的目标字符;

确定该目标字符为与该目标关键词所属的语音指令相匹配的字符。

12.根据权利要求8所述的装置,其特征在于,所述装置还包括:

调用模块,用于在基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配之前,当本次获取的所述图像与上一次获取的所述终端响应于上一条语音指令截取的所述终端的显示装置所显示的图像相同时,不对所述图像进行识别,调用上一次识别图像得到的目标字符用于所述匹配。

13.一种计算机可读存储介质,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权利要求1至6中任一项所述方法的步骤。

14.一种电子设备,其特征在于,包括:

存储器,其上存储有计算机程序;

处理器,用于执行所述存储器中的所述计算机程序,以实现权利要求1至6中任一项所述方法的步骤。

基于语音控制终端的方法、装置、存储介质和电子设备

技术领域

[0001] 本公开涉及控制技术领域，具体地，涉及一种基于语音控制终端的方法、装置、存储介质和电子设备。

背景技术

[0002] 随着技术的发展，越来越多的终端例如手机，平板电脑等采用触控屏幕实现对终端的控制，通过手指在触控屏幕上执行点击、滑动等操作手势实现不同的触控操作。

[0003] 相关技术中，由于需要手指与终端的触控屏幕接触才能完成对终端的控制，因此对于部分肢体残障人士，难以通过触控屏幕实现对终端的控制。

发明内容

[0004] 本公开的目的是提供一种基于语音控制终端的方法、装置、存储介质和电子设备，用于解决相关技术中，部分肢体残障人士难以通过触控屏幕控制终端的技术问题。

[0005] 为了实现上述目的，本公开实施例的第一方面，提供一种基于语音控制终端的方法，所述方法包括：

[0006] 获取语音指令；

[0007] 获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像；

[0008] 将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息；

[0009] 根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。

[0010] 可选地，所述将所述语音指令与所述图像中的字符进行匹配，包括：

[0011] 识别所述语音指令以得到所述语音指令中的目标关键词；

[0012] 识别所述图像以得到所述图像中的目标字符；

[0013] 基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配。

[0014] 可选地，所述识别所述语音指令以得到所述语音指令中的目标关键词，包括：

[0015] 识别所述语音指令得到对应于该语音指令的文字信息；

[0016] 基于预设规则从所述文字信息中提取出所述语音指令中的所述目标关键词。

[0017] 可选地，所述识别所述图像以得到所述图像中的目标字符，包括：

[0018] 确定所述图像中包含所述目标字符的感兴趣区域；

[0019] 对所述感兴趣区域进行二值化处理；

[0020] 于所述二值化处理后的感兴趣区域内确定包含所述目标字符的最小矩形区域；

[0021] 基于机器学习算法或光学字符识别OCR对所述最小矩形区域进行识别得到所述目标字符。

[0022] 可选地，所述基于所述目标关键词与所述目标字符将所述语音指令与所述图像中

的字符进行匹配，包括：

[0023] 确定与所述目标关键词满足预设关联规则的目标字符；

[0024] 确定该目标字符为与该目标关键词所属的语音指令相匹配的字符。

[0025] 可选地，所述基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配之前，所述方法还包括：

[0026] 若本次获取的所述图像与上一次获取的所述终端响应于上一条语音指令截取的所述终端的显示装置所显示的图像相同，不对所述图像进行识别，调用上一次识别图像得到的目标字符用于所述匹配。

[0027] 本公开实施例的第二方面，提供一种基于语音控制终端的装置，所述装置包括：

[0028] 语音指令获取模块，用于获取语音指令；

[0029] 图像获取模块，用于获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像；

[0030] 匹配模块，用于将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息；

[0031] 发送模块，用于根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。

[0032] 可选地，所述匹配模块包括：

[0033] 语音指令识别子模块，用于识别所述语音指令以得到所述语音指令中的目标关键词；

[0034] 图像识别子模块，用于识别所述图像以得到所述图像中的目标字符；

[0035] 匹配子模块，用于基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配。

[0036] 可选地，所述语音指令识别子模块用于：

[0037] 识别所述语音指令得到对应于该语音指令的文字信息；

[0038] 基于预设规则从所述文字信息中提取出所述语音指令中的所述目标关键词。

[0039] 可选地，所述图像识别子模块用于：

[0040] 确定所述图像中包含所述目标字符的感兴趣区域；

[0041] 对所述感兴趣区域进行二值化处理；

[0042] 于所述二值化处理后的感兴趣区域内确定包含所述目标字符的最小矩形区域；

[0043] 基于机器学习算法或光学字符识别OCR对所述最小矩形区域进行识别得到所述目标字符。

[0044] 可选地，所述匹配子模块用于：

[0045] 确定与所述目标关键词满足预设关联规则的目标字符；

[0046] 确定该目标字符为与该目标关键词所属的语音指令相匹配的字符。

[0047] 可选地，所述装置还包括：

[0048] 调用模块，用于在基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配之前，当本次获取的所述图像与上一次获取的所述终端响应于上一条语音指令截取的所述终端的显示装置所显示的图像相同时，不对所述图像进行识别，调用上一次识别图像得到的目标字符用于所述匹配。

[0049] 本公开实施例的第三方面，提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述第一方面中任一项所述方法的步骤。

[0050] 本公开实施例的第四方面，提供一种电子设备，包括：

[0051] 存储器，其上存储有计算机程序；

[0052] 处理器，用于执行所述存储器中的所述计算机程序，以实现上述第一方面中任一项所述方法的步骤。

[0053] 通过上述技术方案，获取语音指令，以及终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像；将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息；根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。故即使对于肢体残障人士，也可以通过语音实现对触控区域的模拟触控操作，进而达到控制终端的目的。

[0054] 本公开的其他特征和优点将在随后的具体实施方式部分予以详细说明。

附图说明

[0055] 附图是用来提供对本公开的进一步理解，并且构成说明书的一部分，与下面的具体实施方式一起用于解释本公开，但并不构成对本公开的限制。在附图中：

[0056] 图1是根据一示例性实施例示出的一种终端与服务器通信的应用场景图。

[0057] 图2是根据一示例性实施例示出的一种基于语音控制终端的方法的流程图。

[0058] 图3是根据一示例性实施例示出的一种基于语音控制终端的方法包括的步骤中将所述语音指令与所述图像中的字符进行匹配的流程图。

[0059] 图4是根据一示例性实施例示出的一种基于语音控制终端的方法包括的步骤中识别所述语音指令以得到所述语音指令中的目标关键词的流程图。

[0060] 图5是根据一示例性实施例示出的一种基于语音控制终端的方法包括的步骤中识别所述图像以得到所述图像中的目标字符的流程图。

[0061] 图6是根据一示例性实施例示出的一种基于语音控制终端的方法包括的步骤中基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配的流程图。

[0062] 图7是根据一示例性实施例示出的一种基于语音控制终端的方法的另一流程图。

[0063] 图8是根据一示例性实施例示出的一种基于语音控制终端的装置的框图。

[0064] 图9是根据一示例性实施例示出的一种电子设备的框图。

具体实施方式

[0065] 以下结合附图对本公开的具体实施方式进行详细说明。应当理解的是，此处所描述的具体实施方式仅用于说明和解释本公开，并不用于限制本公开。

[0066] 图2是根据一示例性实施例示出的一种基于语音控制终端的方法的流程图，可以应用于如图1所示的服务器10，如图2所示，该方法包括：

[0067] S11，获取语音指令。

[0068] 举例来讲，如图1所示，用户发出语音指令后，可以由终端20接收，然后由终端20发

送给服务器10，使得语音指令被服务器10获取。当然，语音指令也可以通过其他设备由服务器10获取。如图1所示，终端20可以为手机。

[0069] S12，获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像。

[0070] 沿用上述例子，终端检测到语音指令后，截取终端的显示装置例如触摸显示屏当前显示的图像，然后将该图像发送至服务器，由服务器接收。为了使得终端可以响应于语音指令进行截图操作，可以采用类似于语音助手的方式，通过在语音指令中加入预设的关键词进行激活，例如预设的关键词可以为“Hi，ABC”，则打开相机的语音指令可以为“Hi，ABC，打开相机”。

[0071] S13，将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息。

[0072] 沿用上述例子，若用户当前正在终端进行扑克牌游戏，语音指令为“Hi，ABC，出一张K”，则服务器可以通过对上述步骤S12中截取的图像进行识别检测，确定图像中目标字符“K”的位置，并确定该目标字符“K”在终端的显示装置如显示触摸屏上的位置信息例如该目标字符“K”的坐标信息，以完成该语音指令与图像中的字符的匹配。需要说明的是，图像中的字符可以包括文字，符号，以及图标等具有象征意义的标识，本公开对其不作具体限制。

[0073] S14，根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。

[0074] 沿用上述例子，在完成步骤S13后，服务器生成包含上述位置信息的触控指令，并将该触控指令发送给终端，终端接收到该触控指令后，控制终端对上述目标字符“K”所在的触控区域进行模拟触控，完成语音指令所述的“出一张K”的操作，当然，如果出牌操作为向上滑动，则模拟触控动作为向上滑动，本公开对具体的触控操作不作具体限制，只要其能完成语音指令对应的操作即可。另一方面，位置信息表征的触控区域可以是目标字符所在的显示区域，也可以是与该位置信息具有一定关联的其它显示区域，例如目标字符在显示触摸屏的左上角，与该目标字符的位置信息相关联的触控区域可以位于显示触摸屏的右下角，则在对触摸显示屏进行模拟触控操作时，选择对右下角的触控区域进行模拟触控。为了进行模拟触控，可以在接收到触控指令时，获取终端的模拟触控权限，当然，也可以由用户自行配置相关应用自动开启模拟触控权限，以简化步骤。

[0075] 通过上述技术方案，获取语音指令，以及终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像；将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息；根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。故即使对于肢体残障人士，也可以通过语音实现对触控区域的模拟触控操作，进而达到控制终端的目的。

[0076] 可选地，如图3所示，本公开中，将所述语音指令与所述图像中的字符进行匹配，包括：

[0077] S131，识别所述语音指令以得到所述语音指令中的目标关键词。

[0078] 具体来讲，目标关键词可以是预设的关键词，例如可以建立包含目标关键词的词库，通过识别语音指令得到存在于词库中的目标关键词。

[0079] 可选地，本公开中，如图4所示，识别所述语音指令以得到所述语音指令中的目标关键词，包括：

[0080] S1311，识别所述语音指令得到对应于该语音指令的文字信息。

[0081] 具体来讲，终端将语音指令发送至服务器，服务器对该语音指令进行识别以将该语音指令转化为对应的文字信息。在一种具体的实施方式中，可以通过现有的语音识别技术例如百度线上识别端对语音指令进行识别，关于语音识别的具体方式，本公开不做具体限制。

[0082] S1312，基于预设规则从所述文字信息中提取出所述语音指令中的所述目标关键词。

[0083] 由于识别出的文字信息可能包含冗余错乱的信息，因此通过预设规则如正则表达式来对文字信息进行筛选，提取出所需的目标关键词。在一种可能的实施方式中，可以先对文字信息进行分词处理，然后依次检查分词处理后得到的词语是否存在于上述预设的包含目标关键词的词库中，若存在，则确定该词语为目标关键词。

[0084] 沿用上述例子，若“K”存在于上述预设的包含目标关键词的词库中，则对于语音指令“Hi，ABC，出一张K”，提取出的目标关键词为“K”。

[0085] S132，识别所述图像以得到所述图像中的目标字符。

[0086] 具体来讲，目标字符可以是预设的字符，例如可以建立包含目标字符的字符库，通过识别语音指令得到存在于字符库中的目标字符，该字符库可以与上述词库合并为一个。

[0087] 可选地，如图5所示，本公开中，识别所述图像以得到所述图像中的目标字符，包括：

[0088] S1321，确定所述图像中包含所述目标字符的感兴趣区域。

[0089] 感兴趣区域(ROI，region ofinterest)是从图像中选择的一个目标区域，作为后续的处理对象，如此可以减小图像处理的区域大小，减小计算量，加快计算速度。感兴趣区域的选取可以采用现有技术中的方法，本公开对其不作赘述。沿用上述例子，对于语音指令“Hi，ABC，出一张K”，确定的图像中的感兴趣区域可以为包含目标字符“K”的扑克牌所在的区域。

[0090] S1322，对所述感兴趣区域进行二值化处理。

[0091] 由于获取的图像多为彩色，为简化计算量，可以对感兴趣区域进行二值化处理，例如可以利用计算机视觉库OpenCV中的inRange()函数和threshold()函数进行二值化处理。其中，inRange()函数可以对多通道图像进行二值化，将像素值位于预设范围内的像素点赋值为像素值255或像素值0中的一者，将像素值位于上述预设范围外的像素点赋值为像素值255或像素值0中的另一者，预设范围可以根据实际情况设置。threshold()函数用于对单通道图像进行二值化，对于像素值大于或等于预设阈值的像素点，将其像素值赋值为像素值255或像素值0中的一者，对于像素值小于预设阈值的像素点，将其像素值赋值为像素值255或像素值0中的另一者。对于二值化的具体方式，本公开不作具体限制。

[0092] 沿用上述例子，由于扑克牌分红色和黑色两种，因此可以先利用上述inRange()函数分别对红色和黑色扑克牌各进行一次二值化处理，然后再利用上述threshold()函数进行一次二值化处理，如此相比于将图像直接转化为灰度图像再进行二值化的效果更佳，有助于提高识别的准确率。

[0093] 为便于进行二值化处理，在进行二值化处理之前，可以先将图像从一个颜色空间转换到另一个颜色空间，转换的过程中数据类型不变，即转换后的图像的数据类型和位深度与源图像一致。在一种可能的实施方式中，可以将图像从RGB(Red,Green,Blue)颜色空间转换到HSV(Hue,Saturation,Value)颜色空间。

[0094] S1323，于所述二值化处理后的感兴趣区域内确定包含所述目标字符的最小矩形区域。

[0095] 举例来讲，对于上述感兴趣区域进行二值化处理后的图像，可以先调用OpenCV中的findContours()函数获取目标字符“K”的轮廓，然后调用OpenCV中的boundingRect()函数确定该目标字符“K”所在的最小矩形区域。

[0096] S1324，基于机器学习算法或光学字符识别OCR对所述最小矩形区域进行识别得到所述目标字符。

[0097] 沿用上述例子，通过机器学习算法例如R-CNN算法或OCR对上述得到的最小矩形区域进行识别得到目标字符“K”，由于仅对最小矩形区域进行识别，减小了相关计算量，加快了计算速度。其中，OCR可以基于Python的pytesseract库实现。

[0098] S133，基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配。

[0099] 可选地，如图6所示，本公开中，基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配，包括：

[0100] S1331，确定与所述目标关键词满足预设关联规则的目标字符。

[0101] 具体来讲，预设关联规则可以为目标关键词与目标字符相同，或者目标关键词与目标字符的相似度大于相似度阈值。进而对于每一目标关键词，确定与该目标关键词对应的满足上述预设关联规则的目标字符。

[0102] S1332，确定该目标字符为与该目标关键词所属的语音指令相匹配的字符。

[0103] 执行上述步骤S1331后，将确定的目标字符作为与对应的目标关键词所述的语音指令相匹配的字符，以完成相应匹配工作。便于进行后续的模拟触控。需要说明的是，上述步骤S131至步骤S133以及相应的子步骤均可以在服务器中运行。

[0104] 图7是根据一示例性实施例示出的一种基于语音控制终端的方法的另一流程图，可以应用于如图1所示的服务器10，如图7所示，该方法包括：

[0105] S21，获取语音指令。

[0106] S22，获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像。

[0107] S23，识别所述语音指令以得到所述语音指令中的目标关键词。

[0108] S24，识别所述图像以得到所述图像中的目标字符。

[0109] S25，若本次获取的所述图像与上一次获取的所述终端响应于上一条语音指令截取的所述终端的显示装置所显示的图像相同，不对所述图像进行识别，调用上一次识别图像得到的目标字符用于所述匹配。

[0110] 在执行上述步骤S22获取终端的显示装置当前显示的图像之后，判断本次获取的图像与上一次获取的图像是否相同，上一次获取图像的具体方式可以参考上述步骤S12，若相同，则表明该图像在上一次进行模拟触控的过程中，可能已经按照上述步骤S132描述的

方式进行了识别并得到了相应的目标字符，存储在服务器中，因此为了简化步骤，减小计算量，对于本次获取的图像不进行上述识别操作，而直接调用上一次得到的存储在服务器中的目标字符用于后续步骤S26所述的相关匹配操作。

[0111] S26，基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配。

[0112] S27，根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。

[0113] 图8是根据一示例性实施例示出的一种基于语音控制终端的装置的框图，如图8所示，该装置100包括：

[0114] 语音指令获取模块110，用于获取语音指令；

[0115] 图像获取模块120，用于获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像；

[0116] 匹配模块130，用于将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息；

[0117] 发送模块140，用于根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。

[0118] 通过上述技术方案，获取语音指令，以及终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像；将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息；根据所述位置信息发送触控指令给所述终端，以控制所述终端模拟对所述位置信息表征的触控区域进行触控操作。故即使对于肢体残障人士，也可以通过语音实现对触控区域的模拟触控操作，进而达到控制终端的目的。

[0119] 可选地，所述匹配模块130包括：

[0120] 语音指令识别子模块，用于识别所述语音指令以得到所述语音指令中的目标关键词；

[0121] 图像识别子模块，用于识别所述图像以得到所述图像中的目标字符；

[0122] 匹配子模块，用于基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配。

[0123] 可选地，所述语音指令识别子模块用于：

[0124] 识别所述语音指令得到对应于该语音指令的文字信息；

[0125] 基于预设规则从所述文字信息中提取出所述语音指令中的所述目标关键词。

[0126] 可选地，所述图像识别子模块用于：

[0127] 确定所述图像中包含所述目标字符的感兴趣区域；

[0128] 对所述感兴趣区域进行二值化处理；

[0129] 于所述二值化处理后的感兴趣区域内确定包含所述目标字符的最小矩形区域；

[0130] 基于机器学习算法或光学字符识别OCR对所述最小矩形区域进行识别得到所述目标字符。

[0131] 可选地，所述匹配子模块用于：

[0132] 确定与所述目标关键词满足预设关联规则的目标字符；

[0133] 确定该目标字符为与该目标关键词所属的语音指令相匹配的字符。

[0134] 可选地，所述装置100还包括：

[0135] 调用模块，用于在基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配之前，当本次获取的所述图像与上一次获取的所述终端响应于上一条语音指令截取的所述终端的显示装置所显示的图像相同时，不对所述图像进行识别，调用上一次识别图像得到的目标字符用于所述匹配。

[0136] 关于上述实施例中的装置，其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述，此处将不做详细阐述说明。

[0137] 图9是根据一示例性实施例示出的一种电子设备的框图。例如，电子设备1900可以被提供为一服务器。参照图9，电子设备1900包括处理器1922，其数量可以为一个或多个，以及存储器1932，用于存储可由处理器1922执行的计算机程序。存储器1932中存储的计算机程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外，处理器1922可以被配置为执行该计算机程序，以执行上述的基于语音控制终端的方法。

[0138] 另外，电子设备1900还可以包括电源组件1926和通信组件1950，该电源组件1926可以被配置为执行电子设备1900的电源管理，该通信组件1950可以被配置为实现电子设备1900的通信，例如，有线或无线通信。此外，该电子设备1900还可以包括输入/输出(I/O)接口1958。电子设备1900可以操作基于存储在存储器1932的操作系统，例如Windows ServerTM，Mac OS XTM，UnixTM，LinuxTM等等。

[0139] 在另一示例性实施例中，还提供了一种包括程序指令的计算机可读存储介质，该程序指令被处理器执行时实现上述的基于语音控制终端的方法的步骤。例如，该计算机可读存储介质可以为上述包括程序指令的存储器1932，上述程序指令可由电子设备1900的处理器1922执行以完成上述的基于语音控制终端的方法。

[0140] 在另一示例性实施例中，还提供一种计算机程序产品，该计算机程序产品包含能够由可编程的装置执行的计算机程序，该计算机程序具有当由该可编程的装置执行时用于执行上述的基于语音控制终端的方法的代码部分。

[0141] 以上结合附图详细描述了本公开的优选实施方式，但是，本公开并不限于上述实施方式中的具体细节，在本公开的技术构思范围内，可以对本公开的技术方案进行多种简单变型，这些简单变型均属于本公开的保护范围。

[0142] 另外需要说明的是，在上述具体实施方式中所描述的各个具体技术特征，在不矛盾的情况下，可以通过任何合适的方式进行组合，为了避免不必要的重复，本公开对各种可能的组合方式不再另行说明。

[0143] 此外，本公开的各种不同的实施方式之间也可以进行任意组合，只要其不违背本公开的思想，其同样应当视为本公开所公开的内容。

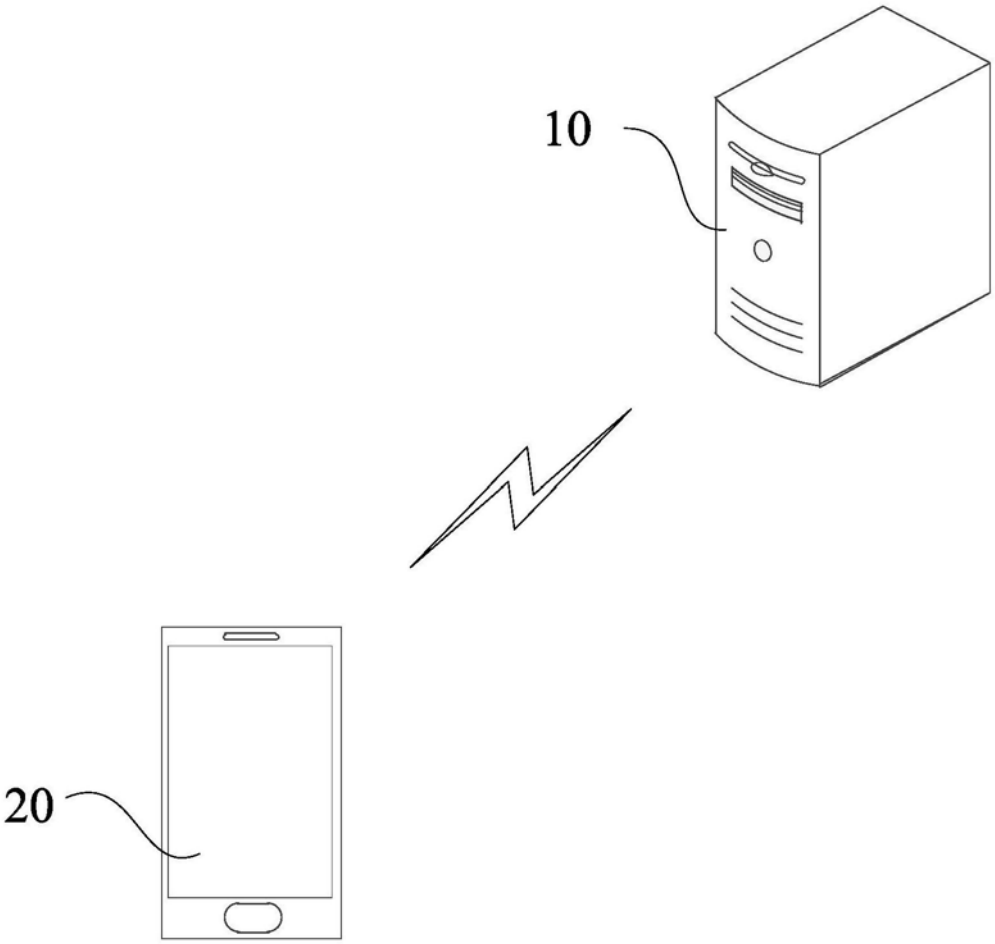


图1

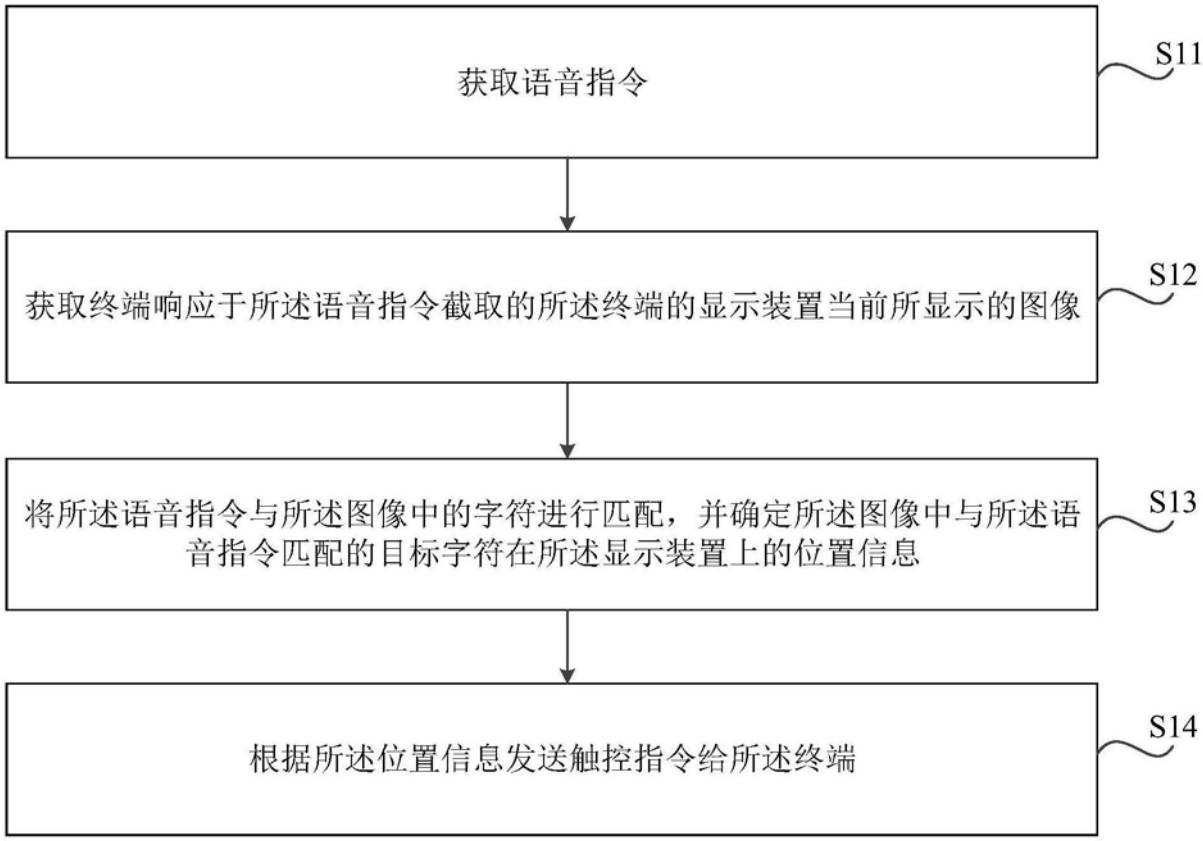

获取语音指令
S11
获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像
S12
将所述语音指令与所述图像中的字符进行匹配，并确定所述图像中与所述语音指令匹配的目标字符在所述显示装置上的位置信息
S13
根据所述位置信息发送触控指令给所述终端
S14

图2

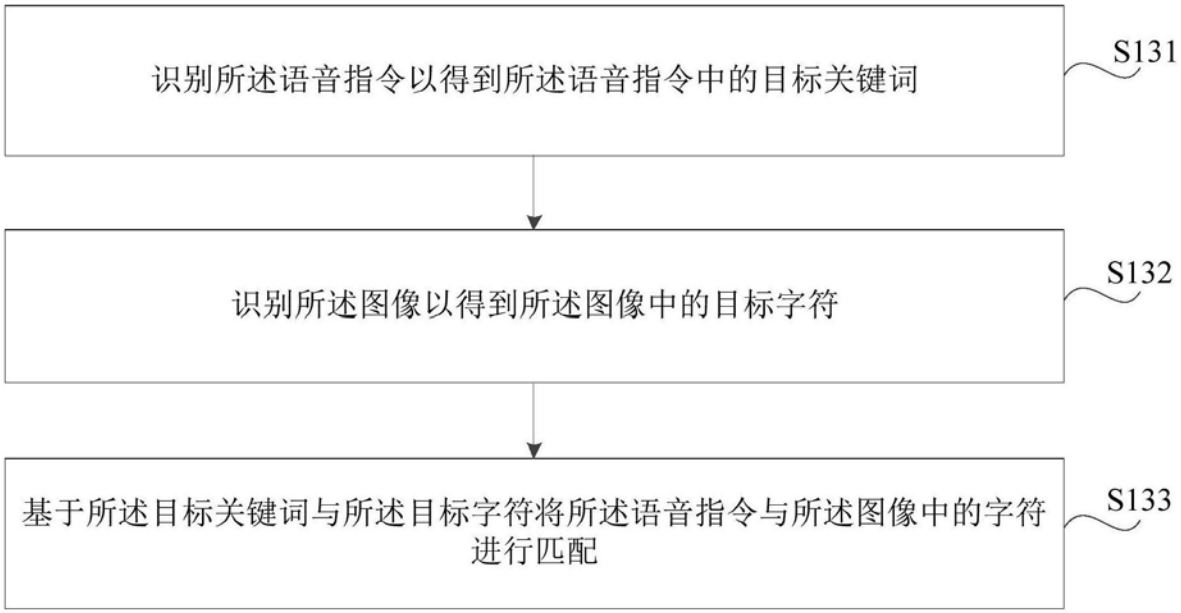

识别所述语音指令以得到所述语音指令中的目标关键词
S131
识别所述图像以得到所述图像中的目标字符
S132
基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配
S133

图3

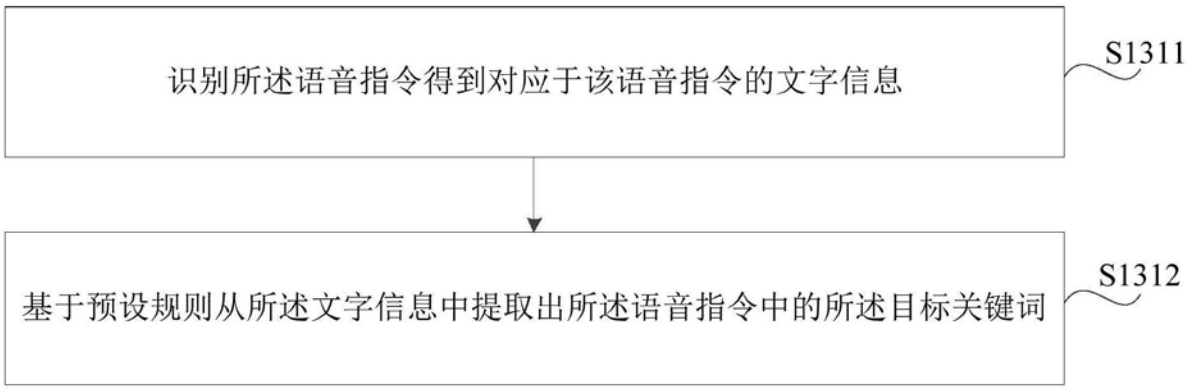

识别所述语音指令得到对应于该语音指令的文字信息
S1311
基于预设规则从所述文字信息中提取出所述语音指令中的所述目标关键词
S1312

图4

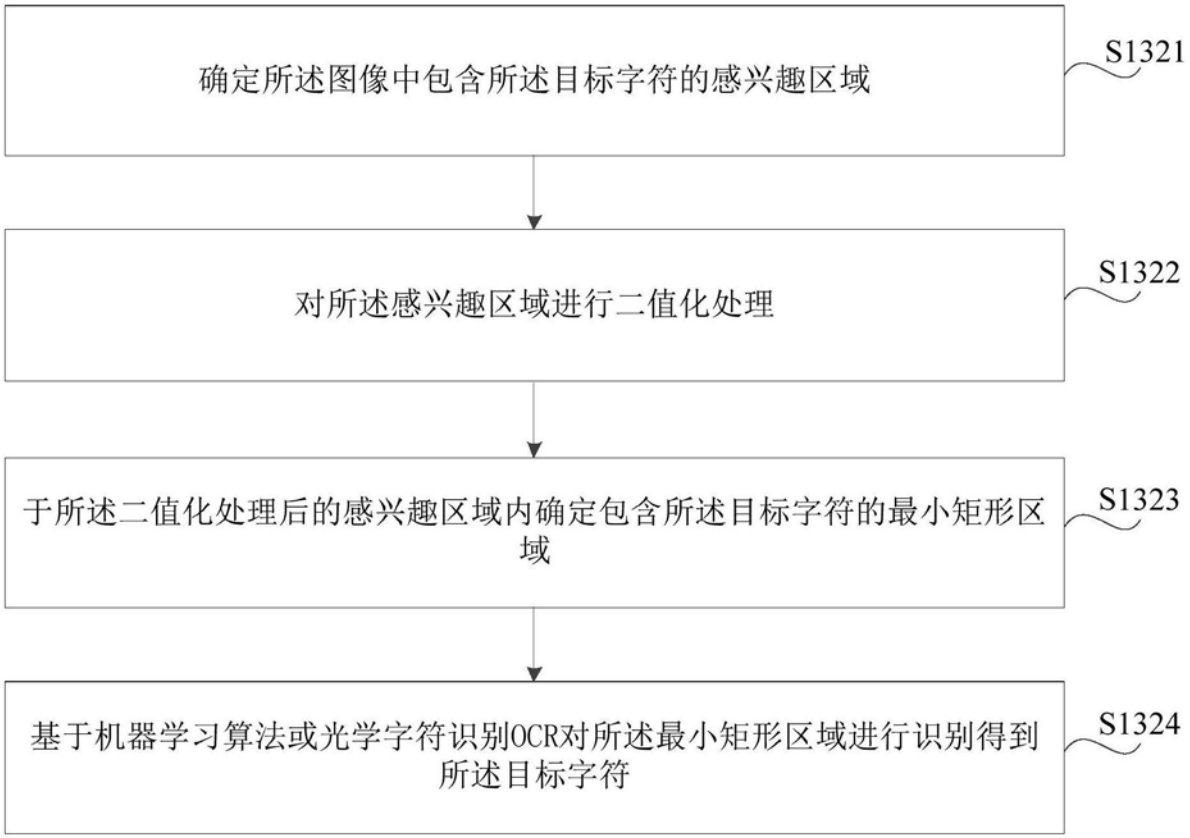

确定所述图像中包含所述目标字符的感兴趣区域
S1321
对所述感兴趣区域进行二值化处理
S1322
于所述二值化处理后的感兴趣区域内确定包含所述目标字符的最小矩形区域
S1323
基于机器学习算法或光学字符识别OCR对所述最小矩形区域进行识别得到所述目标字符
S1324

图5

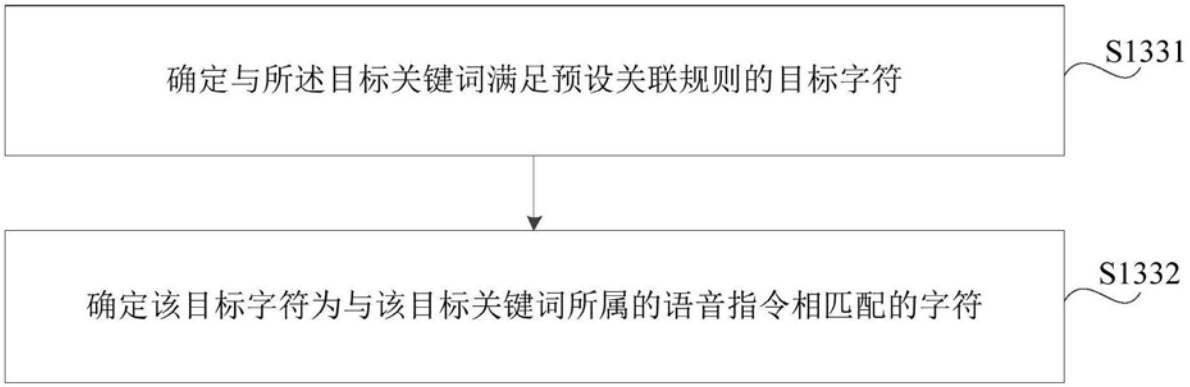

确定与所述目标关键词满足预设关联规则的目标字符
S1331
确定该目标字符为与该目标关键词所属的语音指令相匹配的字符
S1332

图6

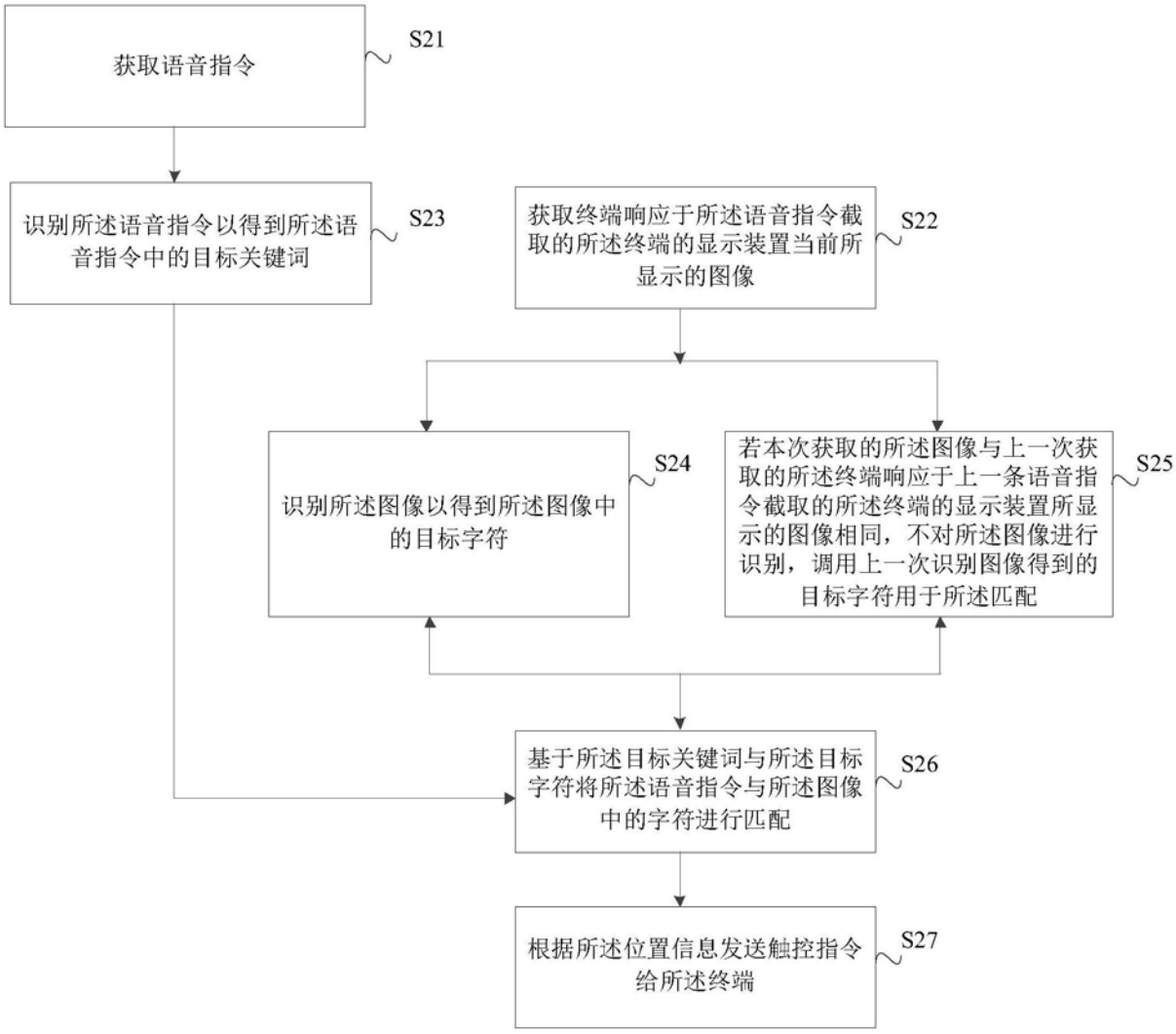

获取语音指令
S21
识别所述语音指令以得到所述语音指令中的目标关键词
S23
获取终端响应于所述语音指令截取的所述终端的显示装置当前所显示的图像
S22
识别所述图像以得到所述图像中的目标字符
S24
若本次获取的所述图像与上一次获取的所述终端响应于上一条语音指令截取的所述终端的显示装置所显示的图像相同，不对所述图像进行识别，调用上一次识别图像得到的目标字符用于所述匹配
S25
基于所述目标关键词与所述目标字符将所述语音指令与所述图像中的字符进行匹配
S26
根据所述位置信息发送触控指令给所述终端
S27

图7

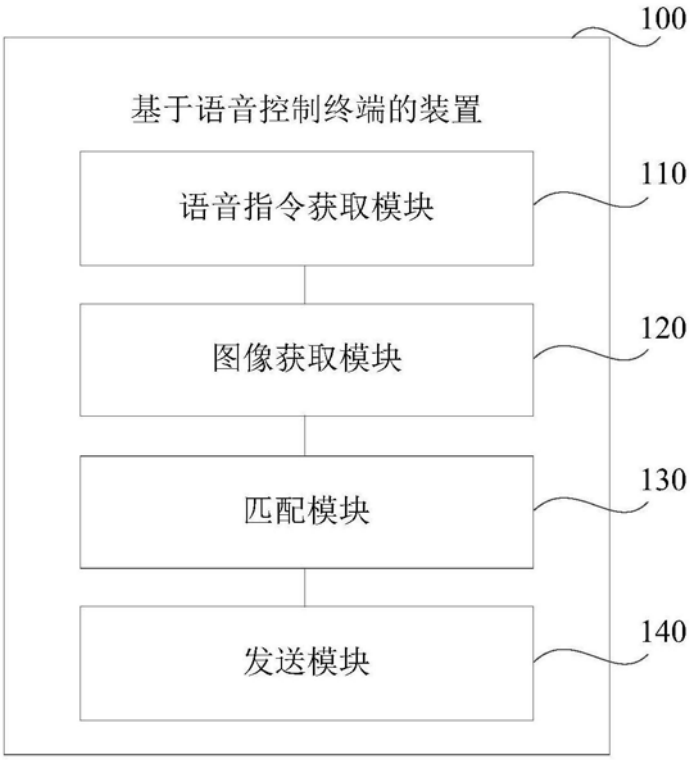

100
基于语音控制终端的装置
语音指令获取模块
110
图像获取模块
120
匹配模块
130
发送模块
140

图8

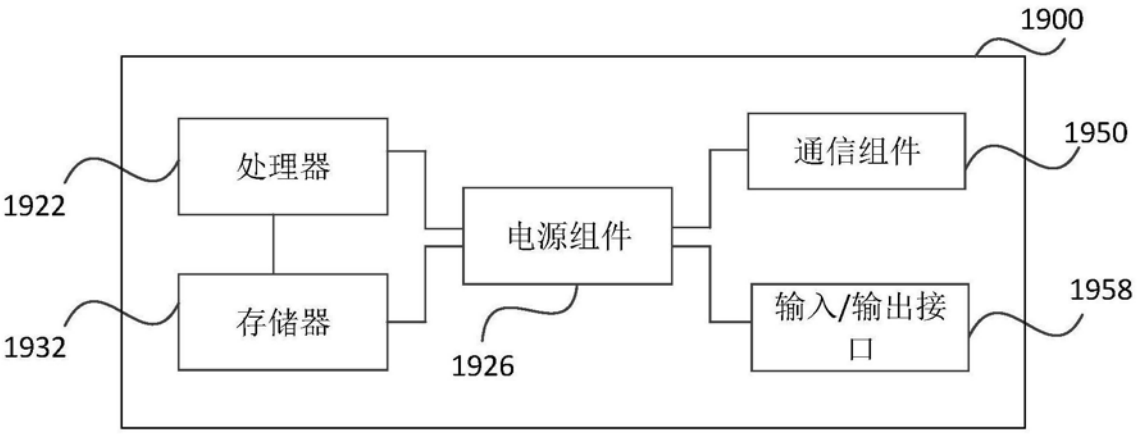

1900
处理器
1922
存储器
1932
电源组件
1926
通信组件
1950
输入/输出接口
1958

图9