



(12)发明专利

(10)授权公告号 CN 102737101 B

(45)授权公告日 2018.09.04

(21)申请号 201210091176.3

(22)申请日 2012.03.30

(65)同一申请的已公布的文献号
申请公布号 CN 102737101 A

(43)申请公布日 2012.10.17

(30)优先权数据
13/076,862 2011.03.31 US
13/077,233 2011.03.31 US
13/077,303 2011.03.31 US
13/077,396 2011.03.31 US
13/077,431 2011.03.31 US
13/077,455 2011.03.31 US
13/077,368 2011.03.31 US

(73)专利权人 微软技术许可有限责任公司
地址 美国华盛顿州

(72)发明人 L·P·赫克 M·金达昆塔
D·米特比 L·施蒂费尔曼

(74)专利代理机构 永新专利商标代理有限公司
72002

代理人 王英

(51)Int.Cl.
G06F 17/30(2006.01)

(56)对比文件
US 2010/0281435 A1,2010.11.04,
US 2010/0217604 A1,2010.08.26,
CN 1845052 A,2006.10.11,
CN 1963752 A,2007.05.16,
US 2010/0281435 A1,2010.11.04,

审查员 白利敏

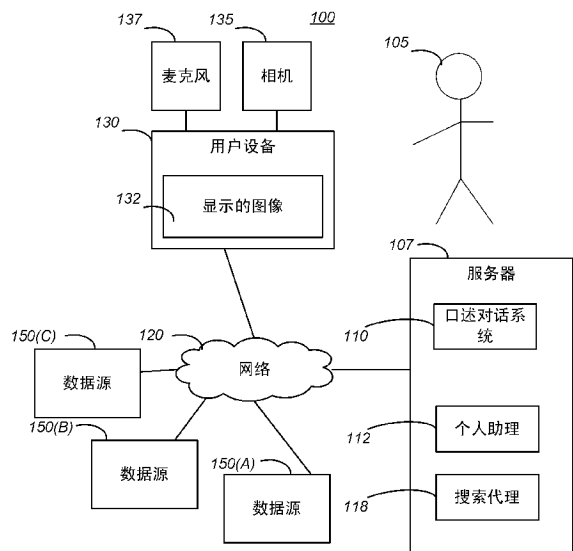
权利要求书2页 说明书5页 附图3页

(54)发明名称

用于自然用户界面系统的组合式激活

(57)摘要

本发明涉及用于自然用户界面系统的组合式激活。可提供用户交互激活。可对从用户接收的多个信号进行评估以确定该多个信号是否与视觉显示相关联。如果是的话,该多个信号可被翻译成代理动作,并且可检索与视觉显示相关联的上下文。可根据所检索的上下文来执行代理动作,并且可向用户显示与所执行的代理动作相关联的结果。



1. 一种用于提供用户(105)交互激活的方法(200),所述方法(200)包括:
 - 接收(210)来自用户(105)的多个信号;
 - 确定(220)所述多个信号是否与视觉显示相关联,其中,所述视觉显示包括静态图像、视频和/或游戏图像;
 - 响应于确定(220)所述多个信号与视觉显示相关联:
 - 将所述多个信号翻译(240)成代理动作,其中所述多个信号包括通过照相机标识的激活手势,其中所述翻译(240)包括利用由所述激活手势标识的所述视觉显示的子集的指示来创建所述代理动作;
 - 响应于所标识的指向所述视觉显示的激活手势,检索与所述视觉显示相关联的上下文,
 - 根据所检索的上下文和接收的信号执行(250)所述代理动作,以及
 - 向用户(105)显示(260)与所执行的代理动作相关联的结果。
2. 如权利要求1所述的方法(200),其特征在于,所述多个信号还包括:用户通过话筒口述的单词或通过键盘或鼠标的输入。
3. 如权利要求1所述的方法(200),其特征在于,根据所检索的上下文执行(250)代理动作包括根据所检索的上下文将多个结果缩小为所述多个结果的子集。
4. 如权利要求3所述的方法(200),其特征在于,还包括向用户(105)显示(260)所述多个结果的子集。
5. 如权利要求1所述的方法(200),其特征在于,还包括:
 - 接收(210)来自用户(105)的手势,其中所述手势包括所述多个信号中的至少之一;
 - 根据所述手势更新(240)所检索的上下文;以及
 - 根据所更新的上下文执行(250)所述代理动作。
6. 如权利要求1所述的方法(200),其中:
 - 所接收的多个信号包括语音信号。
7. 如权利要求6所述的方法,其特征在于,所述手势和所述语音信号是同时从用户(105)接收的。
8. 如权利要求6所述的方法,其特征在于,与所述视觉显示相关联的上下文是从与所述视觉显示相关联的多个元数据中检索的。
9. 一种用于提供用户交互激活的系统,所述系统包括:
 - 用于接收来自用户的多个信号的装置;
 - 用于确定所述多个信号是否与视觉显示相关联的装置,其中,所述视觉显示包括静态图像、视频和/或游戏图像;
 - 用于响应于确定所述多个信号与视觉显示相关联:
 - 将所述多个信号翻译成代理动作的装置,其中所述多个信号包括通过照相机标识的激活手势,其中该翻译(240)包括利用由所述激活手势标识的所述视觉显示的子集的指示来创建所述代理动作;
 - 响应于所标识的指向所述视觉显示的激活手势,检索与所述视觉显示相关联的上下文的装置;
 - 根据所检索的上下文和接收的信号执行所述代理动作的装置;以及

向用户显示与所执行的代理动作相关联的结果的装置。

用于自然用户界面系统的组合式激活

技术领域

[0001] 本发明涉及用户交互系统,更具体地,涉及用于自然用户界面系统的组合式激活。

背景技术

[0002] 自然用户界面系统的组合式激活可提供多模式自然用户界面激活系统,该系统可使用多种模式来激活或操作应用。在一些情形中,自然用户界面系统注重于单一模式的激活或操作。例如,用户通过语音命令或通过敲击屏幕来激活应用。然而,常规系统中的单一模式的命令激活可以是高度敏感的或容易出现各种类型的不准确,诸如无意的激活。

发明内容

[0003] 提供本概述以便以简化形式介绍将在以下详细描述中进一步描述的一些概念。此发明内容既不旨在标识所要求保护的主题的关键特征或必要特征。本发明内容也不旨在用于限制所要求保护的主题的范围。

[0004] 可提供用户交互激活。可对从用户接收的多个信号进行评估以确定该多个信号是否与视觉显示相关联。如果是的话,该多个信号可被翻译成代理动作,并且可检索与视觉显示相关联的上下文。可根据所检索的上下文来执行代理动作,并且可向用户显示与所执行的代理动作相关联的结果。

[0005] 以上概括描述和以下详细描述两者都提供了示例,并且只是说明性的。因此,以上概括描述和以下详细描述不应当被认为是限制性的。此外,除了本文中所阐述的那些特征或变体以外,还可以提供其他特征或变体。例如,实施例可涉及具体实施方式中所描述的各种特征组合和子组合。

附图说明

[0006] 合并在本公开中并构成其一部分的附图示出本发明的实施例。在附图中:

[0007] 图1是操作环境的框图;

[0008] 图2是一种用于提供用户交互激活的方法的流程图;以及

[0009] 图3是包括计算设备的系统的框图。

具体实施方式

[0010] 以下详细描述参考各个附图。只要可能,就在附图和以下描述中使用相同的附图标记来指示相同或相似的元件。尽管可能描述了本发明的实施例,但修改、改编、以及其他实现是可能的。例如,可对附图所示的元件进行置换、添加、或修改,并且可通过对所公开的方法置换、重新排序、或添加阶段来修改本文中所描述的方法。因此,以下详细描述并不限制本发明。相反,本发明的正确范围由所附权利要求书定义。

[0011] 口述对话系统(SDS)使得人们能够用他们的声音与计算机进行交互。驱动该SDS的主要组件可以包括对话管理器:该组件管理与用户的基于对话的会话。对话管理器可通过

多个输入源的组合来确定用户的意图,这多个输入源诸如语音识别和自然语言理解组件输出、来自先前对话轮次的上下文、用户上下文、和/或从知识库(例如搜索引擎)返回的结果。在确定意图后,对话管理器可采取动作,诸如向用户显示最终结果和/或继续与用户的对话以满足他们的意图。

[0012] 图1是操作环境100的框图,操作环境100包括用户105、服务器107、网络120、以及用户设备130。服务器107可包括口述对话系统(SDS) 110、个人助理程序112、和/或搜索代理118。SDS 110可用于经由网络120接收用户词组、查询、动作、和/或动作请求。网络120可包括专有网络(例如,企业内联网)、蜂窝网络和/或诸如因特网等公共网络。操作环境100还可包括多个数据源150(A)-(C)。用户设备130可用于提供显示的图像132,诸如与照片、视频、和/或游戏相关联的图像。用户设备130可被耦合到相机135,相机135可用于记录用户105以及捕捉用户105所作的动作和/或手势。用户设备130还可进一步用来捕捉用户105诸如通过话筒137口述的单词,和/或捕捉来自用户105的诸如通过键盘和/或鼠标(未绘出)的其它输入。根据本发明的其它实施例,相机135可以包括能够检测用户105的移动的任何运动检测设备。例如,相机135可以包括微软KINECT®运动捕捉设备,它包括多个相机和多个话筒。

[0013] 图2是阐述根据本发明的一实施例的用于提供用户查询的个性化的方法200中所涉及各概略阶段的流程图。方法200可使用计算设备300来实现,这将在下面参考图3予以更详细描述。在下文中将更详细地描述实现方法200的各阶段的方式。方法200可开始于起始框205并继续至阶段210,在那里,计算设备300可接收来自用户的多个信号。例如,SDS 110可接收口述查询并由相机135标识用户105所执行的第一手势。例如,用户可挥手并说一个命令,如“你好,xbox”。

[0014] 随后,方法200可前进到阶段220,在那里,计算设备300可确定该信号是否针对该系统。例如,用户105指向屏幕可包括一个激活手势,而用户105从相机135前走过可不包括激活手势。根据本发明的各实施例,用户105可将任何手势定义为相关联的手势。如果所标识的手势和/或语音信号被标识为不针对SDS 110,则方法200可在阶段270结束。

[0015] 如果信号针对该系统,则方法200可前进到阶段230,在那里,计算设备300可检索与视觉显示相关联的上下文。例如,元数据可与视频流相关联,视频流提供诸如标题、演员、描述、评级等信息。对于另一示例,可从数据源150(A)-(C)中的一个检索上下文。例如,数据源150(A)可包括电影信息网站。

[0016] 方法200可随后前进到步骤240,在那里,计算设备300可将接收到的信号翻译成代理动作。例如,相机135可捕捉用户105的指向手势,指向手势可用于指示视觉显示的子集。例如,在电影视频的当前帧中有三名演员,则相机可标识用户105正指向三名演员中的哪一名。指示可被用于创建与诸如“那个演员是谁?”之类的语音查询相关联的代理动作。因此,代理动作可能够选择性地标识三名演员中用户所指示的那一个。

[0017] 方法200随后可前进至阶段250,在那里,计算设备300可根据所检索到的上下文和所接收到的信号执行代理动作。例如,SDS 110可从数据源150(A)检索所显示的电影中的所有演员的列表,将结果缩小到信号被接收时所显示的三名演员,并且根据用户105指向哪个演员来标识具体的演员。

[0018] 方法200接着可前进到阶段260,在那里,计算设备300可向用户显示与所执行的查询相关联的结果。例如,可在用户设备130上显示一个字幕来提供查询的结果。随后,方法

200可在阶段270结束。

[0019] 根据本发明的一实施例可包括用于提供用户交互激活的系统。该系统可包括存储器存储和耦合到该存储器存储的处理单元。处理单元可用于接收来自用户的查询、检索与视觉显示相关联的上下文、根据所检索到的上下文执行查询、以及向用户显示与所执行的查询相关联的结果。视觉显示可包括例如静态图像、视频、和/或游戏图像。可用于根据所检索的上下文来检索查询可包括处理单元可用于根据所检索的上下文将多个结果缩小到所述多个结果的子集。处理单元还可进一步用于接收来自用户的手势、根据该手势(例如指向手势)更新所检索的上下文、以及根据所更新的上下文执行查询。可用于根据指向手势来更新所检索的上下文可包括处理单元可用于标识指向手势所指示的视觉显示的元素。

[0020] 根据本发明的另一实施例可包括用于提供用户交互激活的系统。该系统可包括存储器存储和耦合到该存储器存储的处理单元。处理单元可用于接收包括自然语音词组(例如,口述词组)的请求、检索与视觉显示相关联的上下文、标识用户所作出的手势、根据所检索的上下文和所标识的手势来执行与请求所关联的动作、以及向用户提供与所执行的动作相关联的结果。根据本方面的各实施例,自然语言词组可包括口述和/或会话语法而不是特别格式化的查询。例如,“那个建筑物是什么”可包括自然语言词组以及可与电影“盗梦空间”的视觉显示相关联。诸如可被提供给搜索引擎之类的可用来对比的格式化查询可包括“domain:imdb.com title:Inception time:1:32 ‘identify building’ coordinates:132, 425”。视觉显示可包括与用户相关联的记录设备所捕捉的图像。例如,用户可用相机拍摄一张数码照片并查看图像。用户的手势可包括激活手势。例如,用户105可直接指向相机135以指示用户105将要作出查询和/或动作。

[0021] 根据本发明的又一实施例可包括用于提供用户交互激活的系统。该系统可包括存储器存储和耦合到该存储器存储的处理单元。处理单元可用于接收来自用户的多个同时发生的信号,其中至少一个第一信号包括经由至少一个麦克风接收的语音信号并且至少一个第二信号包括经由至少一个相机接收的手势,并且处理单元可用于确定所述多个信号是否针对该系统。响应于确定所述多个信号针对该系统,处理单元可用于:接收来自用户的查询;检索与视觉显示相关联的上下文;标识经由相机从用户接收的第二手势;将所述多个信号翻译成与视觉显示相关联的至少一个代理动作,其中手势包括可用于选择视觉显示的子集的指向手势;根据所检索的上下文和所标识的第二手势来执行查询代理动作;以及向用户显示与所执行的查询代理动作相关联的结果。

[0022] 图3是包括计算设备300的系统的框图。根据本发明的一个实施例,上述存储器存储和处理单元可在诸如图3的计算设备300之类的计算设备中实现。可使用硬件、软件或固件的任何合适的组合来实现存储器存储和处理单元。例如,存储器存储和处理单元可用计算设备300或结合计算设备300的其他计算设备318中的任一个来实现。根据本发明的实施例,上述系统、设备和处理器是示例,而其他系统、设备和处理器可包括上述存储器存储和处理单元。此外,计算设备300可包括如上所述的操作环境100。系统100可在其他环境中操作,并且不限于计算设备300。

[0023] 参考图3,根据本发明的一实施例的系统可包括计算设备,诸如计算设备300。在基本配置中,计算设备300可包括至少一个处理单元302和系统存储器304。取决于计算设备的配置和类型,系统存储器304可包括,但不限于,易失性存储器(例如,随机存取存储器

(RAM)、非易失性存储器(例如,只读存储器(ROM)、闪存、或任何组合。系统存储器304可以包括操作系统305、一个或多个编程模块306,且可以包括个人助理程序112。例如,操作系统305可适用于控制计算设备300的操作。此外,本发明的实施例可结合图形库、其他操作系统、或任何其他应用程序来实践,并且不限于任何特定应用或系统。该基本配置在图3中由虚线308内的那些组件示出。

[0024] 计算设备300可具有附加特征或功能。例如,计算设备300还可包括附加数据存储设备(可移动和/或不可移动),诸如例如,磁盘、光盘、或磁带。这些附加存储在图3中由可移动存储309和不可移动存储310示出。计算机存储介质可包括以用于存储诸如计算机可读指令、数据结构、程序模块、或其他数据等信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。系统存储器304、可移动存储309和不可移动存储310都是计算机存储介质(即,存储器存储)的示例。计算机存储介质可包括,但不限于,RAM、ROM、电可擦除只读存储器(EEPROM)、闪存或其他存储器技术、CD-ROM、数字多功能盘(DVD)或其他光存储、磁带盒、磁带、磁盘存储或其他磁性存储设备、或者可用于存储信息且可由计算设备300访问的任何其他介质。任何此类计算机存储介质可以是设备300的一部分。计算设备300还可以具有输入设备312,如键盘、鼠标、笔、声音输入设备、触摸输入设备等。还可包括诸如显示器、扬声器、打印机等输出设备314。上述设备是示例,并且可使用其他设备。

[0025] 计算设备300还可包含可允许设备300诸如通过分布式计算环境中的网络(例如,内联网或因特网)来与其他计算设备318进行通信的通信连接316。通信连接316是通信介质的一个示例。通信介质通常由诸如载波或其他传输机制之类的已调制数据信号中的计算机可读指令、数据结构、程序模块、或其他数据来体现,并且包括任何信息传送介质。术语“已调制数据信号”可以描述以对该信号中的信息进行编码的方式设定或者改变其一个或多个特征的信号。作为示例而非限制,通信介质包括诸如有线网络或直接线连接等有线介质,以及诸如声学、射频(RF)、红外线和和其他无线介质等无线介质。如此处所使用的术语“计算机可读介质”可包括存储介质和通信介质两者。

[0026] 如上所述,可在系统存储器304中存储包括操作系统305在内的多个程序模块和数据文件。当在处理单元302上执行时,编程模块306(例如,个人助理程序112)可执行各过程,包括例如,如上所述的方法200的各阶段中的一个或多个。上述过程是一个示例,且处理单元302可执行其他过程。根据本发明的实施例可使用的其他编程模块可包括电子邮件和联系人应用程序、文字处理应用程序、电子表格应用程序、数据库应用程序、幻灯片演示应用程序、绘图或计算机辅助应用程序等。

[0027] 一般而言,根据本发明的实施例,程序模块可包括可执行特定任务或可实现特定抽象数据类型的例程、程序、组件、数据结构和其他类型的结构。此外,本发明的实施例可用其他计算机系统配置来实践,包括手持式设备、多处理器系统、基于微处理器的系统或可编程消费电子产品、小型机、大型计算机等。本发明的实施例还可在其中任务由通过通信网络链接的远程处理设备执行的分布式计算环境中实践。在分布式计算环境中,程序模块可位于本地和远程存储器存储设备两者中。

[0028] 此外,本发明的实施例可在包括分立电子元件的电路、包含逻辑门的封装或集成电路芯片、利用微处理器的电路、或在包含电子元件或微处理器的单个芯片上实践。本发明的实施例还可使用能够执行诸如例如,AND(与)、OR(或)和NOT(非)的逻辑运算的其他技术

来实践,包括但不限于,机械、光学、流体和量子技术。另外,本发明的实施例可在通用计算机或任何其他电路或系统中实践。

[0029] 例如,本发明的实施例可被实现为计算机过程(方法)、计算系统、或诸如计算机程序产品或计算机可读介质之类的制品。计算机程序产品可以是计算机系统可读并用于执行计算机过程的指令的计算机程序编码的计算机存储介质。计算机程序产品还可以是计算机系统可读并用于执行计算机过程的指令的计算机程序编码的载体上的传播信号。因此,本发明可以硬件和/或软件(包括固件、常驻软件、微码等)来体现。换言之,本发明的实施例可采用其上包含有供指令执行系统使用或结合其使用的计算机可使用或计算机可读程序代码的计算机可使用或计算机可读存储介质上的计算机程序产品的形式。计算机可使用或计算机可读介质可以是可包含、存储、通信、传播、或传输程序以供指令执行系统、装置或设备使用或结合其使用的任何介质。

[0030] 计算机可使用或计算机可读介质例如可以是、但不限于电、磁、光、电磁、红外、或半导体系统、装置、设备或传播介质。更具体的计算机可读介质示例(非穷尽列表),计算机可读介质可包括以下:具有一条或多条导线的电连接、便携式计算机盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或闪存)、光纤、以及便携式压缩盘只读存储器(CD-ROM)。注意,计算机可使用或计算机可读介质甚至可以是其上打印有程序的纸张或另一合适的介质,因为程序可经由例如对纸张或其他介质的光学扫描而电子地捕获,随后如有必要被编译、解释、或以其他合适的方式处理,并且随后存储在计算机存储器中。

[0031] 以上参考例如根据本发明的实施例的方法、系统和计算机程序产品的框图和/或操作示图描述了本发明的实施例。框中所注明的各功能/动作可按不同于任何流程图所示的次序出现。例如,取决于所涉及的功能/动作,连续示出的两个框实际上可基本同时执行,或者这些框有时可按相反的次序执行。

[0032] 尽管已描述了本发明的特定实施例,但也可能存在其他实施例。此外,虽然本发明的实施例被描述为与存储在存储器和其他存储介质中的数据相关联,但是数据还可被存储在其他类型的计算机可读介质上或从其读取,诸如辅助存储设备(像硬盘、软盘、或CD-ROM)、来自因特网的载波、或其他形式的RAM或ROM。此外,所公开的方法的各步骤可以任何方式修改,包括通过对各步骤重新排序和/或插入或删除步骤,而不背离本发明。

[0033] 包括此处所包括的代码中的版权在内的所有权利都归属于申请人并且是本申请人的财产。本申请人保持并保留此处所包括的代码中的所有权利,并且授予仅关于所授权专利的再现且未出于其他目的再现该材料的许可。

[0034] 尽管本说明书包括示例,但本发明的范围由所附权利要求书来指示。此外,尽管用对结构特征和/或方法动作专用的语言描述了本说明书,但权利要求书并不限于以上所描述的特征或动作。相反,以上所描述的特定特征和动作是作为本发明的实施例的示例来公开的。

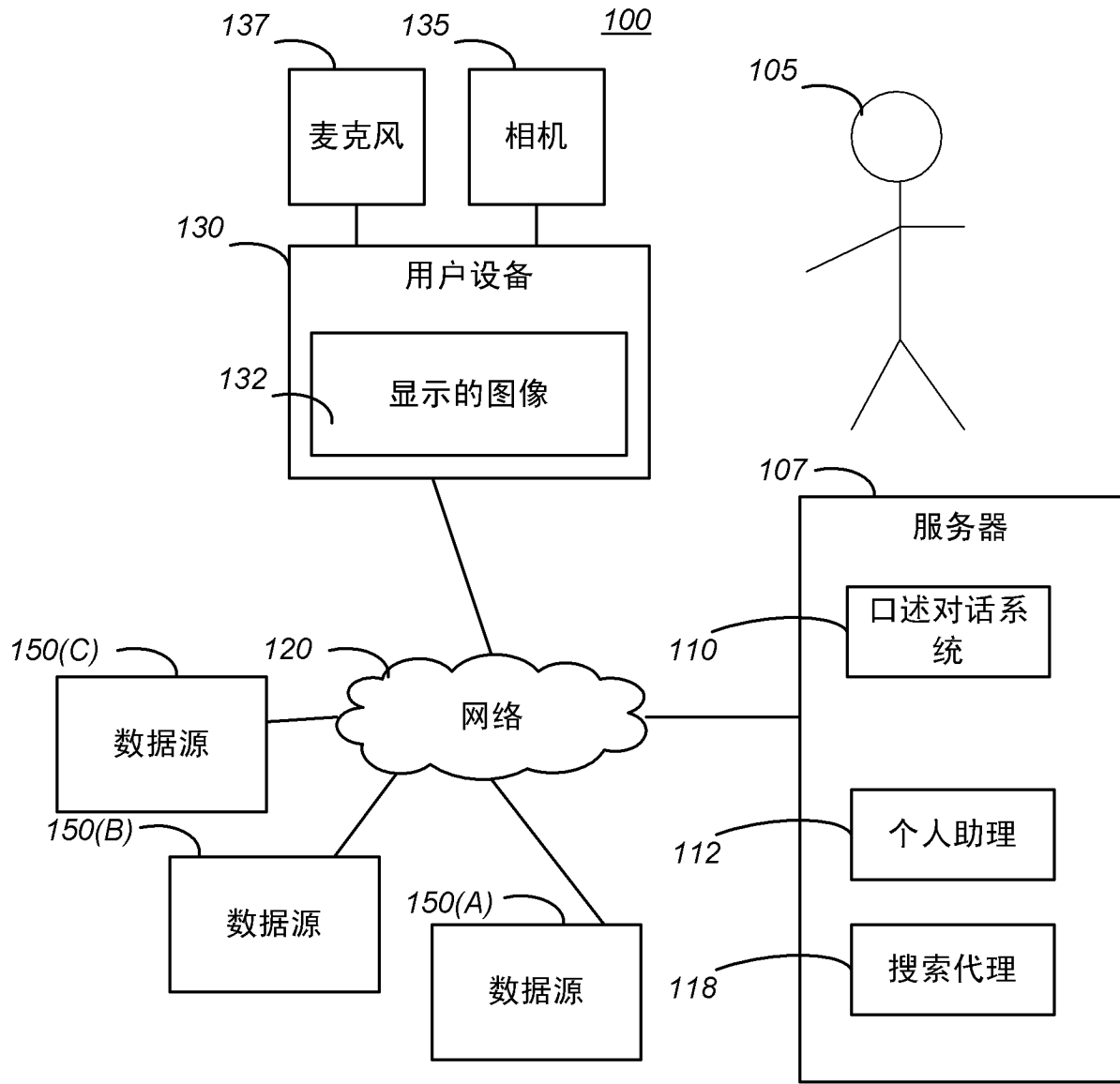


图1

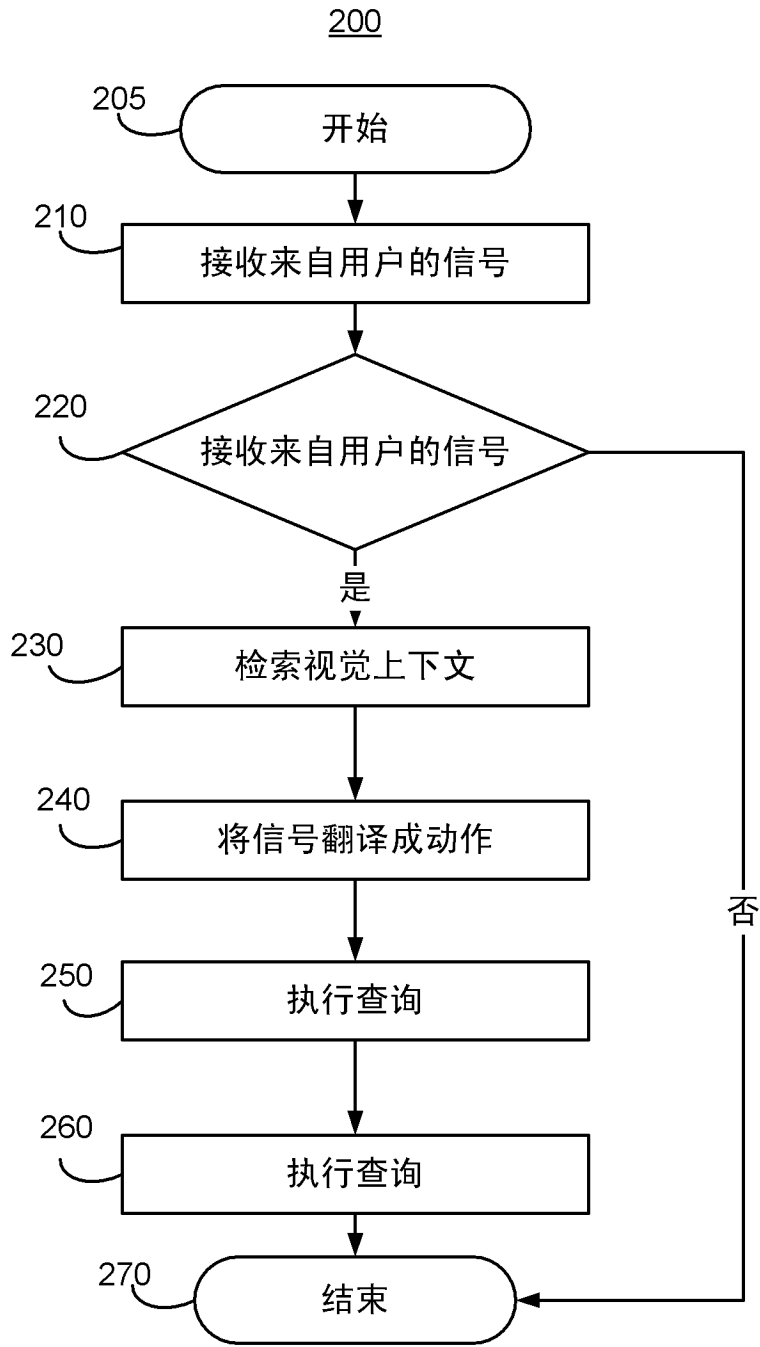


图2

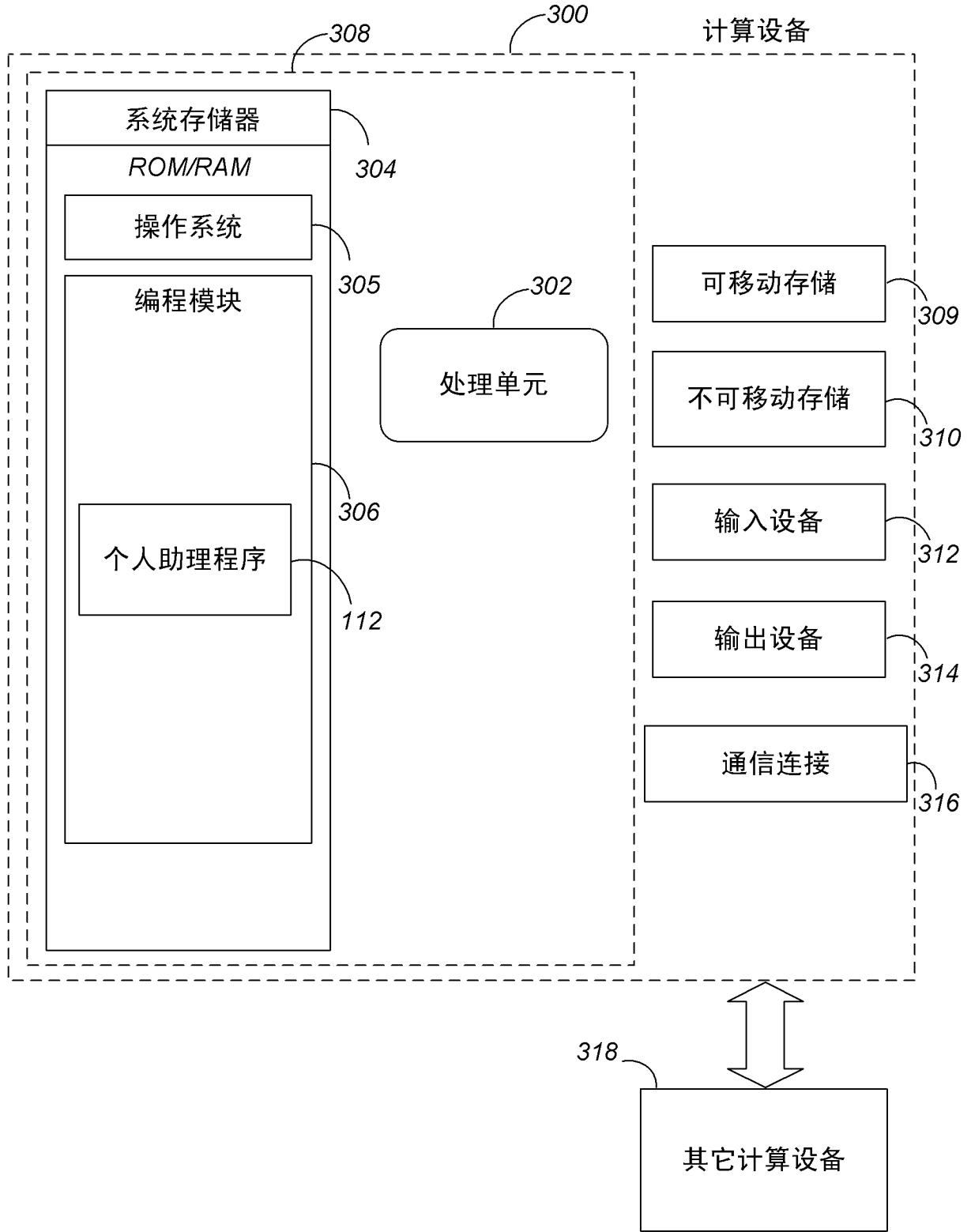


图3