



(12)发明专利

(10)授权公告号 CN 105706077 B

(45)授权公告日 2018.12.11

(21)申请号 201480061355.X

(22)申请日 2014.11.05

(65)同一申请的已公布的文献号
申请公布号 CN 105706077 A

(43)申请公布日 2016.06.22

(30)优先权数据

61/901,996 2013.11.08 US

14/075,956 2013.11.08 US

14/076,029 2013.11.08 US

(85)PCT国际申请进入国家阶段日
2016.05.09

(86)PCT国际申请的申请数据
PCT/US2014/064086 2014.11.05

(87)PCT国际申请的公布数据
W02015/069737 EN 2015.05.14

(73)专利权人 谷歌有限责任公司

地址 美国加利福尼亚州

(72)发明人 亚历山大·J·卡思贝尔特
约舒阿·J·埃斯泰勒

(74)专利代理机构 中原信达知识产权代理有限
责任公司 11219

代理人 周亚荣 安翔

(51)Int.Cl.

G06F 17/28(2006.01)

(56)对比文件

US 6539116 B2,2003.03.25,

US 6539116 B2,2003.03.25,

US 2006/0152479 A1,2006.07.13,

US 2009/0313536 A1,2009.12.17,

审查员 王国海

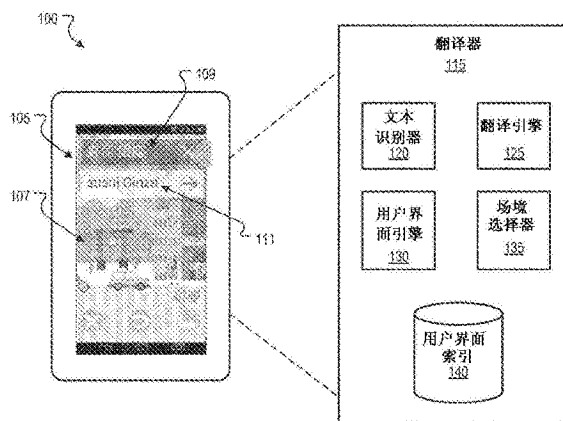
权利要求书4页 说明书25页 附图16页

(54)发明名称

呈现图像中描绘的文本的翻译

(57)摘要

包括编码在计算机存储介质上的计算机程序的方法、系统和装置,用于呈现图像描绘的文本的附加信息。一方面,方法包括接收图像。识别所述图像中描绘的文本。基于所述图像中描绘的所述文本的布置,为所述图像选择呈现场境。每个呈现场境对应于图像内的文本的特定布置。每个呈现场境具有用于呈现关于所述文本的附加信息的对应的用户界面。用于每个呈现场境的所述用户界面不同于用于其他呈现场境的所述用户界面。识别对应于所选择的呈现场境的用户界面。使用所识别的用户界面,为所述图像中描绘的所述文本的至少部分呈现附加信息。所述用户界面在所述图像之上的覆盖中呈现所述附加信息。



1. 一种由数据处理装置执行的方法,所述方法包括:

接收图像;

识别各自包括所述图像中描绘的文本的多个有区别的文本块;

基于相关的文本块中描绘的文本的视觉特征以及所述相关的文本块中描绘的所述文本的布置,识别多个所述相关的文本块集合,每个相关文本块集合包括具有相匹配视觉特征的文本;

基于所识别的集合,从多个呈现场境中为所述图像选择呈现场境,其中,每个呈现场境具有对应的用户界面,所述对应的用户界面用于呈现所识别的集合中的特定集合中所包括的所述文本的至少部分的翻译,其中,用于每个呈现场境的所述用户界面不同于用于其他呈现场境的用户界面,并且其中用于所述多个呈现场境中的第一呈现场境的所述用户界面包括与用于所述多个呈现场境中的第二呈现场境的所述用户界面不同的文本块集合的翻译;

识别对应于所选择的呈现场境的用户界面;以及

使用所选择的用户界面,呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译,而不呈现所识别的集合中的另一集合中所包括的文本的翻译。

2. 根据权利要求1所述的方法,其中,为所述图像选择所述呈现场境包括:

识别所述图像中描绘的多个单独的文本块;以及

基于所述图像中描绘的所述多个单独的文本块,为所述图像选择所述呈现场境。

3. 根据权利要求1所述的方法,其中,为所述图像选择所述呈现场境包括:

基于所述图像中描绘的所述文本的所述布置,确定所述文本的第一部分相较于所述文本的至少一个其他部分被更突出地呈现;以及

响应于所述确定,从所述多个呈现场境中选择突出场境。

4. 根据权利要求3所述的方法,其中:

识别对应于所选择的呈现场境的用户界面包括识别对应于所述突出场境的突出用户界面;以及

呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译包括在所述文本的所述第一部分之上的覆盖中呈现所述文本的所述第一部分的语言翻译。

5. 根据权利要求1所述的方法,其中,为所述图像选择所述呈现场境包括:

确定所述图像中描绘的所述文本包括地址;以及

响应于所述确定,从所述多个呈现场境中选择地图场境。

6. 根据权利要求5所述的方法,其中:

识别对应于所选择的呈现场境的所述用户界面包括识别对应于所述地图场境的地图用户界面;以及

呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译包括在地图的图像内呈现所述地址的语言翻译。

7. 根据权利要求5所述的方法,其中:

识别对应于所选择的呈现场境的用户界面包括识别对应于所述地图场境的地图用户界面;以及

呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译

包括呈现与所述地图的图像内的所述地址相对应的标题,所述标题以与呈现所述地图的所述图像的设备相关联的用户的语言被呈现。

8. 根据权利要求1所述的方法,其中,为所述图像选择所述呈现场境包括:

识别所述图像中描绘的多个单独的文本块;

基于所述单独的文本块的布置以及所述单独的文本块的呈现,确定所述多个单独的文本块属于文本的集合;以及

响应于所述确定,从所述多个呈现场境中选择集合场境。

9. 根据权利要求8所述的方法,其中:

识别对应于所选择的呈现场境的用户界面包括识别对应于所述集合场境的集合用户界面;以及

呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译包括在所述图像之上的覆盖中呈现每个单独的文本块中描绘的文本的语言翻译,每个语言翻译被呈现在语言翻译的对应文本块之上。

10. 一种系统,包括:

数据处理装置;以及

与所述数据处理装置数据通信的存储器存储装置,所述存储器存储装置存储可由所述数据处理装置执行的指令,并且在这样的执行时,使得所述数据处理装置执行操作,所述操作包括:

接收图像;

识别各自包括所述图像中描绘的文本的多个有区别的文本块;

基于相关的文本块中描绘的文本的视觉特征以及所述相关的文本块中描绘的所述文本的布置,识别多个所述相关的文本块集合,每个相关文本块集合包括具有相匹配视觉特征的文本;

基于所识别的集合,从多个呈现场境中为所述图像选择呈现场境,其中,每个呈现场境具有对应的用户界面,所述对应的用户界面用于呈现所识别的集合中的特定集合中所包括的所述文本的至少部分的翻译,其中,用于每个呈现场境的所述用户界面不同于用于其他呈现场境的用户界面,并且其中用于所述多个呈现场境中的第一呈现场境的所述用户界面包括与用于所述多个呈现场境中的第二呈现场境的所述用户界面不同的文本块集合的翻译;

识别对应于所选择的呈现场境的用户界面;以及

使用所选择的用户界面,呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译,而不呈现所识别的集合中的另一集合中所包括的文本的翻译。

11. 根据权利要求10所述的系统,其中,为所述图像选择所述呈现场境包括:

识别所述图像中描绘的多个单独的文本块;以及

基于所述图像中描绘的所述多个单独的文本块,为所述图像选择所述呈现场境。

12. 根据权利要求10所述的系统,其中,为所述图像选择所述呈现场境包括:

基于所述图像中描绘的所述文本的所述布置,确定所述文本的第一部分相较于所述文本的至少一个其他部分被更突出地呈现;以及

响应于所述确定,从所述多个呈现场境中选择突出场境。

13. 根据权利要求12所述的系统, 其中:

识别对应于所选择的呈现场境的用户界面包括识别对应于所述突出场境的突出用户界面; 以及

呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译包括在所述文本的所述第一部分之上的覆盖中呈现所述文本的所述第一部分的语言翻译。

14. 根据权利要求10所述的系统, 其中, 为所述图像选择所述呈现场境包括:

确定所述图像中描绘的所述文本包括地址; 以及

响应于所述确定, 从所述多个呈现场境中选择地图场境。

15. 根据权利要求14所述的系统, 其中:

识别对应于所选择的呈现场境的用户界面包括识别对应于所述地图场境的地图用户界面; 以及

呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译包括在地图的图像内呈现所述地址的语言翻译。

16. 根据权利要求14所述的系统, 其中:

识别对应于所选择的呈现场境的用户界面包括识别对应于所述地图场境的地图用户界面; 以及

呈现所选择的呈现场境的所述特定集合中所包括的所述文本的所述至少部分的翻译包括呈现与所述地图的图像内的所述地址相对应的标题, 所述标题以与呈现所述地图的所述图像的设备相关联的用户的语言被呈现。

17. 根据权利要求10所述的系统, 其中, 为所述图像选择所述呈现场境包括:

识别所述图像中描绘的多个单独的文本块;

基于所述单独的文本块的布置以及所述单独的文本块的呈现, 确定所述多个单独的文本块属于文本的集合; 以及

响应于所述确定, 从所述多个呈现场境中选择集合场境。

18. 一种编码有计算机程序的计算机存储介质, 所述程序包括指令, 所述指令在由数据处理装置执行时使得所述数据处理装置执行操作, 所述操作包括:

接收图像;

识别各自包括所述图像中描绘的文本的多个有区别的文本块;

基于相关的文本块中描绘的文本的视觉特征以及所述相关的文本块中描绘的所述文本的布置, 识别多个所述相关的文本块集合, 每个相关文本块集合包括具有相匹配视觉特征的文本;

基于所识别的集合, 从多个呈现场境中为所述图像选择呈现场境, 其中, 每个呈现场境具有对应的用户界面, 所述对应的用户界面用于呈现所识别的集合中的特定集合中所包括的所述文本的至少部分的翻译, 其中, 用于每个呈现场境的所述用户界面不同于用于其他呈现场境的用户界面, 并且其中用于所述多个呈现场境中的第一呈现场境的所述用户界面包括与用于所述多个呈现场境中的第二呈现场境的所述用户界面不同的文本块集合的翻译;

识别对应于所选择的呈现场境的用户界面; 以及

使用所选择的用户界面, 呈现所选择的呈现场境的所述特定集合中所包括的所述文本

的所述至少部分的翻译,而不呈现所识别的集合中的另一集合中所包括的文本的翻译。

呈现图像中描绘的文本的翻译

背景技术

[0001] 诸如智能电话的许多用户设备都装配有用于捕捉图片和数字图像的相机。用户经常在旅行中使用相机捕捉图片并经由多媒体消息或社交网络站点分享图片。用户还可以使用图像来获取关于与目标或位置相关的附加信息。例如，用户可以捕捉地标的图像并使用所述图像作为查询发起互联网搜索。

[0002] 当在讲与用户不同的语言的另一个国家旅行时，用户可能希望获取用户遇到的文本的翻译。例如，在另一个国家内旅行时用户可能遇到是以外语的街道标志，或者外语的项的列表。

发明内容

[0003] 本说明书描述了涉及基于与图像相关联的呈现场(context)呈现与图像中描绘的文本有关的附加信息(例如，文本的语言翻译)的技术。

[0004] 通常，本说明书中所描述的主题的一个创新的方面可被具体化在包括以下动作的方法中：接收图像；识别所述图像中描绘的文本；基于所述图像中描绘的所述文本的布置，从呈现场境中为所述图像选择一呈现场境，其中，每个呈现场境对应于图像内文本的特定布置，并且每个呈现场境具有对应的用户界面，所述对应的用户界面用于呈现关于所述图像中描绘的所述文本的附加信息，其中，用于每个呈现场境的用户界面不同于用于其他呈现场境的用户界面；识别对应于所选择的呈现场境的所述用户界面；以及使用所识别的用户界面来为所述图像中描绘的所述文本的至少部分呈现附加信息，所述用户界面在所述图像之上的覆盖中呈现所述附加信息。本方面的其他实施例包括被配置为执行所述方法的所述动作、被编码在计算机存储设备上的对应的系统、装置和计算机程序。

[0005] 这些和其他实施例可以分别可选地包括下列特征中的一个或多个。所述附加信息可以包括所识别的文本的至少部分的语言翻译。为所述图像选择所述呈现场境包括识别所述图像中描绘的多个单独的文本块以及基于所述图像中描绘的所述多个单独的文本块，为所述图像选择所述呈现场境。

[0006] 为所述图像选择所述呈现场境包括基于所述图像中描绘的所述文本的所述布置，确定所述文本的第一部分相较于所述文本的至少一个其他部分被更突出地呈现；以及响应于所述确定，从所述呈现场境中选择突出场境。

[0007] 识别对应于所选择的呈现场境的所述用户界面可以包括识别对应于所述突出场境的突出用户界面。为所述图像中描绘的所述文本的至少部分呈现附加信息可以包括在所述文本的所述第一部分之上的覆盖中呈现所述文本的所述第一部分的语言翻译。

[0008] 为所述图像选择所述呈现场境可以包括：确定所述图像中描绘的所述文本包括地址；以及响应于所述确定，从所述呈现场境中选择地图场境。识别对应于所选择的呈现场境的所述用户界面可以包括识别对应于所述地图场境的地图用户界面。为所述图像中描绘的所述文本的至少部分呈现附加信息可以包括在地图的图像内呈现所述地址的语言翻译。

[0009] 识别对应于所选择的呈现场境的所述用户界面可以包括识别对应于所述地图场

境的地图用户界面。为所述图像中描绘的所述文本的至少部分呈现附加信息可以包括呈现对应于所述地图的图像内的所述地址的标题。所述标题以与呈现所述地图的所述图像的设备相关联的用户语言被呈现。

[0010] 为所述图像选择所述呈现场境可以包括：识别所述图像中描绘的单独的文本块；基于所述单独的文本块的布置以及所述单独的文本块的呈现，确定所述多个单独的文本块属于文本的集合；以及响应于所述确定，从所述呈现场境中选择集合场境。

[0011] 识别对应于所选择的呈现场境的所述用户界面可以包括识别对应于所述集合场境的集合用户界面。为所述图像中描绘的所述文本的至少部分呈现附加信息可以包括在所述图像之上的覆盖中呈现每个单独的文本块中描绘的文本的语言翻译。每个语言翻译可以被呈现在语言翻译的对应文本块之上。

[0012] 通常，本说明书中所描述的主题的另一个方面可被具体化在包括以下动作的方法中：接收图像；识别所述图像中描绘的文本，所识别的文本在图像中所识别的一个或多个文本块中，每个文本块与图像中的其他文本块不同；从突出场境中为所述图像选择突出呈现场境，其中，每个突出呈现场境对应于在图像内呈现文本的每个文本块的相关突出度，并且每个突出呈现场境具有用于呈现与图像中描绘的所识别的文本有关的附加信息的对应的用户界面，对突出呈现场境的选择是基于在图像中描绘所识别的文本的一个或多个文本块的相关突出度；识别对应于所选择的突出呈现场境的用户界面；以及使用所识别的用户界面，为所述图像中描绘的所述文本的至少部分呈现附加信息，所述用户界面在所述图像之上的覆盖中呈现所述附加信息。本方面的其他实施例包括被配置为执行所述方法的所述动作、被编码在计算机存储设备上的对应的系统、装置和计算机程序。

[0013] 这些和其他实施例可以分别可选地包括下列特征中的一个或多个。为所述图像选择所述突出呈现场境可以包括：基于图像内的单个文本块的位置和单个文本块中所包括的文本的大小，确定所述图像中描绘的所述文本包括突出显示的文本的单个文本块；并且响应于所述确定从突出呈现场境中选择单个块场境。单个块场境可以对应于在突出显示的文本的单个文本块之上的覆盖中呈现突出显示的文本的单个文本块中包括的文本的语言翻译的用户界面。

[0014] 识别图像中描绘的文本可以包括：识别第一文本块和不同于第一文本块的第二文本块，第一文本块和第二文本块被描绘在图像中。为所述图像选择所述突出呈现场境可以包括：确定第一文本块相较于第二文本块在图像内被显示得更加突出；以及响应于所述确定，从突出呈现场境中选择主-次块场境。主-次块场境可以对应于呈现文本的主要块的语言翻译的用户界面。

[0015] 使用所识别的用户界面为所述图像中描绘的所述文本的至少部分呈现附加信息可以包括：在图像之上的覆盖中呈现包括在第一文本块中的文本的语言翻译。覆盖可以位于图像中的第一文本块之上。

[0016] 方面可以进一步包括在图像中的第二文本块的描绘处的主-次用户界面中呈现可选择的用户界面元素；以及响应于接收到对可选择的用户界面元素的选择，呈现包括在第二文本块中的文本的语言翻译。

[0017] 确定第一文本块相较于第二文本块在图像内被显示得更加突出可以包括：确定包括在第一文本块中的文本大于包括在第二文本块中的文本；以及确定第一文本块相较于第

二文本块位于离图像的中心更接近。

[0018] 识别图像中描绘的文本可以包括：识别第一文本块和不同于第一文本块的第二文本块。第一文本块和第二文本块可以被描绘在图像中。为所述图像选择所述突出呈现场境可以包括：确定第一文本块与第二文本块在图像内以基本上相等的突出度被显示；以及响应于所述确定，从突出呈现场境中选择同等突出场境。同等突出场境可以对应于呈现了具有基本上相等的突出度的文本的多个块的语言翻译的用户界面。

[0019] 使用所述识别的用户界面为所述图像中描绘的所述文本的至少部分呈现附加信息可以包括：在图像之上的第一覆盖中呈现包括在第一文本块中的文本的语言翻译，第一覆盖位于图像中的第一文本块之上；以及在图像之上的第二覆盖中呈现包括在第二文本块中的文本的语言翻译，第二覆盖位于图像中的第二文本块之上。

[0020] 确定第一文本块与第二文本块在图像内以基本上相等的突出度被显示可以包括：确定包括在第一文本块中的文本的大小基本上等于包括在第二文本块中的文本的大小。

[0021] 通常，本说明书中所描述的主题的另一个方面可被具体化在包括以下动作的方法中：接收图像；识别所述图像中描绘的文本，所识别的文本在图像内描绘的文本块中，每个文本块与图像中其他的文本块不同；从集合呈现场境中为图像选择集合呈现场境，其中每个集合呈现场境对应于图像内的文本块的布置并且每个集合呈现场境具有用于呈现与至少一个文本块有关的附加信息的对应用户界面，对集合呈现场境的选择是基于图像中描绘的文本块的布置；识别对应于所选择的集合呈现场境的用户界面；以及使用所识别的用户界面，呈现与文本块中的至少一个有关的附加信息。本方面的其他实施例可以包括被配置为执行所述方法的所述动作、被编码在计算机存储设备上的对应的系统、装置和计算机程序。

[0022] 这些和其他实施例可以分别可选地包括下列特征中的一个或多个。从集合呈现场境中选择所述集合呈现场境可以包括：基于将每个文本块的文本与每个其他文本块的文本的比较，确定文本块中的每个是彼此相关的；以及响应于所述确定而选择单个集合呈现场境。单个集合呈现场境可以具有在图像之上的一个或多个覆盖中呈现每个文本块的语言翻译的对应用户界面。

[0023] 从集合呈现场境中选择集合呈现场境可以包括：基于将每个文本块的文本与每个其他文本块的文本的比较，确定文本块中的两个或多个彼此不相关；以及响应于所述确定而选择多个项呈现场境，所述多个项呈现场境具有响应于所述确定而在单独的显示页面中呈现每个文本框的语言翻译的对应用户界面。

[0024] 对应用户界面可以包括用于每个文本块的用户界面元素。当特定文本块的用户界面元素被选择时，对应用户界面可以呈现特定文本块的语言翻译。

[0025] 从集合呈现场境中选择集合呈现场境可以包括：确定每个文本块的背景颜色与每个其他文本块的背景颜色基本上匹配；基于对背景颜色基本上匹配的确定，确定文本块彼此相关；以及响应于所述确定而选择单个集合呈现场境。单个集合呈现场境可以具有在图像之上的一个或多个覆盖中呈现每个文本块的语言翻译的对应用户界面。

[0026] 从集合呈现场境中选择集合呈现场境可以包括：将每个文本块的大小和位置与每个其他文本块的大小和位置比较；基于所述比较来确定文本块相关；以及响应于所述确定而选择单个集合呈现场境。单个集合呈现场境可以具有在图像之上的一个或多个覆盖中呈

现每个文本块的语言翻译的对应用户界面。

[0027] 确定文本块相关可以包括：确定文本块的大小基本上类似并且每对相邻的文本块与其他每对相邻的文本块在图像内以基本上类似的距离被隔开。

[0028] 从集合呈现场境中选择集合呈现场境可以包括：基于第一组文本块的布置，识别第一组类似的文本块；基于第二组文本块的布置，识别第二组类似的文本块，第二组文本块不同于第一组文本块；以及响应于识别第一组和第二组文本块，选择多个集合呈现场境，多个集合呈现场境对应于为第一组类似文本块呈现附加信息的用户界面，其与用于第二组文本块的附加信息分离。

[0029] 方面可以进一步包括翻译文本块中的至少一个的文本并至少基于翻译的文本的多个特征确定翻译的文本的可读性度量。可以基于可读性度量选择用户界面。

[0030] 方面可以进一步包括确定图像是菜单的图像并确定每个文本块与菜单中的特定项相关联。选择用户界面可以包括选择在特定文本块之上的覆盖中呈现每个特定文本块的翻译的用户界面。

[0031] 方面可以进一步包括，对于每个文本块：识别对应于文本块的图像中的价格，所述价格以特定货币被呈现；将所述价格转换为不同于特定货币的货币；以及在识别的价格之上或附近的覆盖中呈现转换后的价格。

[0032] 可以实现本说明书中所描述的主题的特定实施例，以便实现以下好处中的一个或多个。可以在图像的场境中以更易读的方式呈现图像中描绘的文本的翻译(或与图像中所识别的文本有关的其他信息)。可以基于用于图像的呈现场境动态地选择用于呈现文本的翻译的用户界面，使得翻译以对用户有用的方式被呈现。通过使用可读性度量来选择使得翻译能够以更加可读的形式被呈现的用户界面，改善了翻译的可读性和清楚度。当多个文本块在图像中被识别时，用户界面可以使用户能够快速且简单地在不同文本块的翻译之间导航。

[0033] 本说明书中所描述的主题的一个或多个实施例的细节在以下的附图和描述中被阐述。通过说明书、附图和权利要求，本主题的其他特征、方面和优势将变得显而易见。

附图说明

[0034] 图1是呈现与图像中描绘的文本相关的附加信息的示例用户设备的框图。

[0035] 图2描绘了用于捕捉图像并呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0036] 图3描绘了用于捕捉呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0037] 图4描绘了用于呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0038] 图5描绘了用于捕捉图像并呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0039] 图6描绘了用于呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0040] 图7描绘了用于呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0041] 图8描绘了用于呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0042] 图9是用于基于为图像选择的突出呈现场境来呈现与图像中描绘的文本相关的附

加信息的示例过程的流程图。

[0043] 图10描绘了用于呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0044] 图11描绘了用于捕捉图像并且呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0045] 图12描绘了用于呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0046] 图13是用于基于为图像选择的集合呈现场境来呈现与图像中描绘的文本相关的附加信息的示例过程的流程图。

[0047] 图14描绘了用于呈现被识别为具有地图场境的图像中描绘的文本的语言翻译的示例用户界面的屏幕截图。

[0048] 图15是用于基于为图像选择的地图呈现场境来呈现与图像中描绘的文本相关的附加信息的示例过程的流程图。

[0049] 图16是用于基于为图像选择的呈现场境来呈现与图像中描绘的文本相关的附加信息的示例过程的流程图。

[0050] 各种附图中相同的参考数字和标号指示相同的元素。

具体实施方式

[0051] 概述

[0052] 系统可以识别图像中描绘的文本，翻译所述文本，并将所述文本的翻译以对用户有用的方式呈现。然而，当图像描绘了多个不同的文本块时，简单地呈现图像描绘的全部文本的翻译对用户来说可能会令人混淆。例如，图像可以描绘购物中心中几个不同商店的标志。在单个文本块中呈现每个标志的翻译的用户界面可能使用户混淆，因为用户可能不能识别一个商店的名称在哪里结束以及其他商店的名称从哪里开始。而且，自动地呈现所有翻译会导致混乱的显示，这降低了用户体验。

[0053] 系统可以基于文本在图像内的布置和/或其他视觉特征确定如何呈现翻译。例如，系统可以在确定什么文本要翻译以及如何呈现翻译的文本时考虑图像描绘的文本的特征，诸如多个文本块之间的接近度、文本块之间的对齐、文本之间的相似性（例如，字体大小、语族、设计样式、颜色、背景色等等）、一个或多个集合中多个文本块的分组和/或文本识别中的置信水平。本文中所描述的这些以及其他特征可以提供对图像的场境以及它的文本的深入理解。例如，文本块关于图像的中心或焦点的相对位置可以指示用户最感兴趣的文本。

[0054] 在一些实现中，例如，基于文本在图像内的布置和/或其他视觉特征，系统可以从多个呈现场境中选择呈现场境。多个呈现场境可以包括一个或多个突出呈现场境、一个或多个集合呈现场境和/或一个或多个地图呈现场境。突出呈现场境是在图像内突出描绘一个或多个文本块的场境。例如，单个文本块可以位于图像的中心附近，或者两个类似的突出的文本块可以被放置在离图像的中心类似的距离处。集合呈现场境是在图像中描绘多个有区别的文本块的场境。例如，餐厅菜单的图像可以被归类为集合呈现场境，因为每个菜单项可被视为与图像中描绘的每个其他菜单项相关的单个的文本块。地图呈现场境是在图像中描绘一个或多个地址或标识位置的其他信息的场境。

[0055] 基于所选择的呈现场境，系统可以选择用于呈现与图像中描绘的文本相关的附加信息的用户界面。例如，基于为图像选择的呈现场境，系统可以在用于突出呈现场境的多个

用户界面、用于集合呈现场境的多个用户界面和/或用于地图呈现场境的多个用户界面之间进行选择。每种类型的呈现场境对于被识别为具有该类型的呈现场境的图像可以具有不同集合的规则或者用于选择用户界面的不同过程。

[0056] 示例用户设备

[0057] 图1是呈现与图像中描绘的文本相关的附加信息的示例用户设备100的框图。用户设备100是能够捕捉和/或呈现数码照片和图像的电子设备。示例用户设备100包括个人计算机、移动通信设备(例如,智能电话)、平板计算机以及可以捕捉和/或呈现图像的其他设备。为了呈现图像,用户设备100包括显示器105,诸如液晶显示器(“LCD”)、触摸屏或其他类型的显示器。在本示例中,显示器105正在呈现用户界面,所述用户界面用于呈现在图像107中识别的文本109的翻译111。

[0058] 用户设备100包括翻译器115,所述翻译器115使用户能捕捉对象的图像,诸如标志、菜单和广告牌,并接收对象描绘的文本的翻译。翻译器115包括文本识别器120,所述文本识别器120可以识别图像以及其他类型的文档中的文本。在一些实现中,文本识别器120使用光学字符识别(“OCR”)分析图像以识别图像描绘的文本。文本识别器120可以检测多种不同语言的文本。例如,文本识别器120可以包括能够识别多种语言的文本的OCR引擎,或用于多种不同语言中的每一个的OCR引擎。

[0059] 文本识别器120还可以检测文本的特征,诸如文本颜色、背景颜色、文本大小、字体类型和/或图像内文本的位置。例如,这些特征可以被用于识别有区别的文本块。例如,基于文本的两个部分具有不同的字体颜色、不同的背景颜色或被相互隔开(例如,至少隔开了阈值距离),文本识别器120可以确定图像中描绘的文本的两个部分被包括在两个有区别的文本块中。文本识别器120可以将关于所识别的文本的数据发送至翻译器115的翻译引擎125。该数据可以指定文本自身以及文本的所识别的特征。

[0060] 翻译引擎125可以将文本(或至少所述文本的部分)翻译为不同的语言。例如,翻译引擎125可以将文本109(中文)翻译为与用户设备100相关联的用户的口语或者用户所选择的另一种语言。用户可以使用用户界面指定语言,如下面更详细描述。在本示例中,文本109已经被翻译为英文并且文本109的英文翻译111被呈现给用户。在一些实现中,翻译引擎125可以如由用户设备100或由用户设备100的用户界面引擎130所指定的,只翻译文本的部分。例如,用户设备100可以选择用于翻译的一个或多个文本块。此外,在一些实现中,翻译引擎125可以将货币值从一种货币翻译为另一种。例如,翻译引擎125可以将以外币的形式描绘的货币量翻译为由用户指定的货币。

[0061] 用户界面引擎130可以选择用于呈现与在图像中识别的文本有关的附加信息的用户界面。附加信息可以包括在图像中识别的文本的语言翻译、在图像中识别的货币量(例如,在餐厅菜单的图像中识别的价格)的货币翻译和/或与文本有关的其他信息。用户界面可以定义如何呈现附加信息。例如,如图1中所示出的,可以在图像107之上的覆盖中呈现翻译111。在其他用户界面中,可以在只示出文本的单独的屏幕中呈现翻译,例如,由于翻译包含太多文本要在覆盖中呈现。基于为图像识别或选择的呈现场境,用户界面引擎130可以从用户界面的集合中选择用户界面。如下面所描述的,基于图像描绘的文本的布置或其他特征,场境选择器135可以为图像选择呈现场境。

[0062] 用户界面索引140可以存储呈现场境与用户界面的映射。例如,用户界面索引140

可以包括信息,所述信息识别用于突出呈现场境的用户界面、用于集合呈现场境的用户界面和/或用于地图呈现场境的用户界面。用户界面引擎130可以访问用户界面索引140以识别用于已经为图像选择的特定呈现场境的合适的用户界面,生成用于在用户设备100上呈现的用户界面,以及使用用户界面呈现图像描绘的文本的附加信息(例如,文本的翻译)。用于各个呈现场境的示例用户界面在图2-8、10-12和14中被图示,并且在下面被详细描述。

[0063] 用户界面生成器120还可以使用与附加信息相关联的可读性度量,以选择用于呈现附加信息的用户界面,所述附加信息与图像中描绘的文本有关。可读性度量可以指示如果使用特定用户界面呈现附加信息,所述附加信息可以有多可读。用于用户界面以及用于特定附加信息的可读性度量可以是基于可以使用用户界面呈现的附加信息的字符数量(或者百分比)。例如,只能显示附加信息的10%的用户界面可以具有低可读性度量,并且因此不可以被用于呈现附加信息。

[0064] 通过将用于附加信息的可读性度量与多个合适的用户界面中的每个进行比较,用户界面引擎130可以为图像选择用户界面。例如,假定为图像选择的用户界面场境是特定的突出呈现场境。用户界面引擎130可以确定用于与由图像描绘的文本有关的附加信息相对于可用于特定突出场景的每个用户界面的可读性度量。用户界面引擎130可以基于可读性度量从可用的用户界面中选择。例如,用户界面引擎130可以选择具有最高可读性度量的用户界面。

[0065] 通过另一个示例,用户界面引擎130可以识别用于所选择的呈现场境的优选用户界面并且确定附加信息相对于优选用户界面的可读性度量。例如,优选用户界面可以在图像之上的覆盖中呈现文本的翻译。用户界面引擎130可以将可读性度量与用于优选用户界面的可读性阈值相比较。如果可读性度量满足可读性阈值(例如,通过满足或超过可读性阈值),则用户界面引擎130可以选择优选用户界面并在图像之上的覆盖中呈现翻译。如果可读性度量不满足可读性阈值,则用户界面引擎130可以选择替选的用户界面,诸如在单独的屏幕中呈现翻译的用户界面,所述屏幕使用户能浏览或滚动通过多个屏幕或页面以查看翻译。

[0066] 基于图像描绘的文本的布置和/或其他特征,场境选择器135可以为图像选择呈现场境。可以从多个呈现场境中为图像选择呈现场境。例如,可以从一个或多个突出呈现场境、一个或多个集合呈现场境、一个或多个地图呈现场境和/或其他类型的呈现场境中选择呈现场境。突出呈现场境可以包括用于在其中描绘文本的单个块的图像的“单个块场境”、用于在其中相较于其他文本块更主要地描绘特定文本块的图像的“主-次场境”和/或用于在其中基本上均等突出地(例如,文本块具有在彼此阈值量之内的突出分值)描绘两个或更多个文本块的“均等突出场境”。集合呈现场境可以包括“单个集合”场境(用于在其中描绘多个相关文本块的图像)、“多个集合”场境(用于在其中描绘相关文本块的多个群组的图像)和/或“整合”场境(用于在其中应该将文本块的翻译一起呈现的图像)。

[0067] 下面详细描述呈现场境中的每一个和用于确定如何为特定图像选择呈现场境的技术。具体地,参照图3-8描述突出场境并且参照图9描述用于处理突出场境的过程;参照图10-12描述集合场境并且参照图13描述用于处理突出场境的过程;并且参照图14描述地图场境并且参照图15描述用于处理地图场境的过程。最后,参照图16描述用于处理图像并从各种场境中选择的过程以及相应地处理所选择的场境的过程。

[0068] 尽管示例用户设备100被图示为包括文本识别器120、翻译引擎125、用户界面引擎130、场境选择器135以及用户界面索引140,但这些元素中的一个或多个可以被存储在其他设备上和/或由其他设备执行,诸如与用户设备100通信的服务器。例如,用户设备100可以通过通信网络(例如,LAN、WAN或互联网)的方式将图像传输至服务器。服务器可以处理图像以为图像选择呈现场境以及用户界面并且将数据发送至用户设备100,这使得用户设备100使用所选择的用户界面呈现与图像中描绘的文本有关的附加信息。

[0069] 示例翻译用户界面

[0070] 图2描绘了示例用户界面的屏幕截图200A-200C,其用于捕捉图像并呈现图像中描绘的文本的语言翻译。在示例屏幕截图200A中,用户界面210使用户能选择用于翻译的语言。具体地,用户界面210使用户能选择初始语言212,所述初始语言的文本将在图像中被找到,并且使用户能选择文本将被翻译为的翻译语言214。在本示例中,用户已经选择了初始语言212为中文以及翻译语言214为英文。因此,在本示例中,用户想要将在图像中发现的中文文本翻译为英文文本。用户界面还包括相机按钮206,当所述相机按钮206被用户选择时,使得用户设备使用安装在用户设备上的相机或者可通信地被耦合至用户设备的相机捕捉图像。

[0071] 在示例屏幕截图200B中,用户界面230描绘了用户设备例如响应于对相机按钮206的选择所捕捉到的图像232。示例图像232包括使用中文字符的文本的多个部分,其已经被用户设备识别并且利用白色背景被突出显示。具体地,图像232包括位于图像232左上角附近的文本的第一部分234、位于图像232中心附近的文本的第二部分236以及位于图像232右下角附近的文本的第三部分238。文本234-238的每个部分都被描绘在图像232中单独的标志上。

[0072] 翻译屏幕用户界面250被描绘在屏幕截图200C中。翻译屏幕用户界面250描绘了图像232中所识别出的文本252以及图像232中所识别出的文本252的英文翻译254。文本的第一部分234已经被翻译为词“Free”;文本的第二部分236已经被翻译为“BigRiverInn2ndFloor”;并且文本的第三部分238已经被翻译为词“Last”。在这个示例用户界面250中,文本252已经被翻译的就好像文本252被包括在单个文本块中一样,即使文本252是被描绘在单独的有区别的文本块中。如同通过翻译文本254可以看到的,将多个有区别的文本块翻译为一个可以导致无意义的翻译,这会使用户迷惑或沮丧。通过评估图像描绘的文本的布置和/或其他视觉特征并为图像描绘的文本选择场境,可以提供更有意义的翻译。例如,图3-5图示了可以基于对文本的每个部分的突出度的分析来更好地呈现文本252的翻译的用户界面。

[0073] 突出呈现场境和用户界面

[0074] 图像内的文本的突出度可以被用于确定要对哪个文本提供翻译(或其他信息),因为用户常常将他们的相机聚焦在他们感兴趣的目标上。通过评估文本的突出度,翻译器115可以确定哪个文本是用户可能想要获取翻译的并提供该文本的翻译。例如,用户可以对特定的标志上的文本拍照以获取所述标志的翻译。在照片中,虽然特定的标志在照片中心附近,背景中仍可能存在其他标志。通过评估文本的突出度,翻译器可以识别特定标志的文本用于翻译,因为该文本相较于背景标志的文本被更加突出地描绘。翻译器115还可以确定不用翻译背景标志的文本,因为用户可能对那些标志的翻译不感兴趣。

[0075] 图3描绘了示例用户界面的屏幕截图300A和300B,其用于呈现图像中描绘的文本的语言翻译。屏幕截图300A类似于屏幕截图200B,并包括呈现包括使用中文字符的多个文本部分334-338的图像332的用户界面330,所述文本334-338已经被用户设备识别并且利用白色背景被突出显示。

[0076] 翻译屏幕用户界面350被描绘在屏幕截图300B中。示例翻译屏幕用户界面350只呈现图像332中所识别出的文本352的部分以及文本352的该部分的英文翻译354,而不是呈现图像332中描绘的文本的全部的翻译就好像它们被包括在一个文本块中一样。具体地,翻译屏幕用户界面350呈现文本的第二部分336以及文本的第二部分336的英文翻译354 (“BigRiverInn2ndFloor”)。

[0077] 翻译屏幕用户界面350还包括使用户能浏览附加信息的控件359。控件359可以被配置为在包括每个文本部分334-336的翻译的屏幕之间导航或者获取每个翻译的其他信息。例如,响应于对控件359的选择,用户设备可以显示呈现文本的第三部分338的翻译的另一个屏幕。对控件359的另一个选择可以使得用户设备显示呈现文本的第一部分334的翻译的屏幕。然而对控件359的另一个选择可以使得用户设备返回以呈现呈现翻译354的屏幕。

[0078] 在本示例中,基于相对于文本的第一部分334的突出度和文本的第三部分338的突出度的文本的第二部分336的突出度,诸如图1的翻译器115的翻译器已经选择了文本的第二部分用于翻译。例如,基于文本的部分的相对位置以及文本的部分的方向,翻译器115可以识别文本的每个部分属于有区别的文本块。基于该确定,翻译器115可以确定呈现场境是突出呈现场境并评估文本的每个部分的突出度。因为文本的第二部分在图像中心附近被显示并使用了比文本的其他两个部分大的字体,翻译器115可以确定文本的第二部分336相较于文本的其他两个部分被更加突出地呈现并且响应于所述确定而自动翻译文本的第二部分336。

[0079] 通过只呈现最突出的文本的翻译,用户设备更可能将用户寻找的翻译提供给用户,因为用户通常会使他们的相机聚焦使得期望的内容处于或靠近捕捉的图像的中心。这也避免了用可能是在背景中并被用户无意地捕捉到的文本的翻译使用户困惑。例如,当用户正尝试捕捉文本的第二部分336的图像用于翻译的时候,文本的第一部分334和文本的第三部分338可能已经被用户无意地捕捉到。

[0080] 除了自动地翻译更突出的文本以外,翻译器115还可以选择用于图像的呈现场境以及基于所选择的呈现场境选择用于呈现翻译的用户界面。在该示例中,翻译器115可以为图像332选择主-次场境,因为翻译器115已经确定图像332包括:包括文本的第二部分336的更主要的文本块(或主要的文本块)以及包括文本的第一部分334和文本的第三部分338的两个次要的文本块,所述文本的第一部分334和文本的第三部分338相较于文本的第二部分336比较在图像332中不突出地被描绘。

[0081] 例如,用于主-次场境的用户界面可以包括呈现图像中描绘的最突出的文本的翻译的翻译屏幕用户界面(例如,用户界面350);在图像上呈现覆盖的用户界面,所述覆盖包括最突出的文本的翻译而不呈现次要的文本块的翻译;在图像上呈现覆盖的用户界面,所述覆盖包括最突出的文本块的翻译以及用户界面元素,当选择所述用户界面元素时呈现次要的文本块的翻译;和/或在图像上呈现用于每个文本块的覆盖的用户界面,所述每个文本块呈现覆盖的对应文本块的翻译。

[0082] 例如,图4描绘了示例用户界面的屏幕截图400A和400B,其用于在图像上的覆盖中呈现文本块的语言翻译。屏幕截图400A类似于屏幕截图200B,并且包括呈现图像432的用户界面430,所述图像432包括位于图像432左上角附近的文本的第一部分434、位于图像432中心附近的文本的第二部分436以及位于图像432右下角附近的文本的第三部分438。屏幕截图400B中描绘的用户界面450呈现图像432以及图像432之上的覆盖454-458,其各自包括图像432描绘的文本的翻译,而不是在单独的屏幕中不带有图像432的情况下呈现文本的第二部分的翻译。具体地,覆盖454呈现文本的第一部分434的翻译;覆盖456呈现文本的第二部分456的翻译;并且覆盖458呈现文本的第三部分438的翻译。在本示例中,覆盖454-458分别位于覆盖所对应的文本之上(或代替文本)。这使得对于用户来说使用翻译来确定图像432中的每个标志的主题是简单的。例如,在图像432中描绘的街道上散步的用户可以很容易地看用户界面450并确定河大酒店位于街道上的哪里。

[0083] 每个覆盖454-458可以包括可选择的图标或者与可选择的图标相关联,所述可选择的图标使用户能够获取关于被翻译的文本的附加信息。例如,屏幕截图400C描绘了用户界面470,所述用户界面470使用户能够以初始语言或翻译的语言听文本的发音。关于文本的其他信息也可以被呈现,诸如被包括在关于文本所识别出的实体的文本或信息中的词的定义。响应于用户选择与覆盖454相关联的用户界面元素,用户界面470可以被呈现。

[0084] 在一些实现中,用于主-次场境的用户界面可能最初只包括用于最主要的文本的覆盖,而不是用于没那么主要的文本块的覆盖。这样的用户界面可以提供用户界面元素,所述用户界面元素使用户能请求图像中描绘的次要文本块的翻译。例如,图5描绘了使用户能选择文本用于翻译的示例用户界面的屏幕截图500A-500D。

[0085] 在屏幕截图500A中,用户界面510包括相机按钮,当所述相机按钮被选择时,使得用户设备捕捉图像512。示例图像512包括两个有区别的文本块514和516,其包括日语文本。在屏幕截图500B中,用户界面520在图像512之上的单个文本块523中呈现文本块514和516两者的文本。用户界面520还在图像512之上的文本块525中呈现文本块514和516两者的翻译。诸如图1中的翻译器115的翻译器可以响应于识别文本块514和516,为用户界面520自动地翻译两个文本块的文本。替选地,翻译器115可以只翻译文本块514的文本,因为它被用较大字体更主要地显示。或者,响应于确定两个文本块都不被主要地显示,因为文本块514和516两者都位于图像的角落,翻译器115可以不翻译文本块514或者516。下面参照图9详细描述用于确定图像中描绘的文本块的突出度的过程。

[0086] 在屏幕截图500C中,用户界面530为用户提供指令533以选择用于翻译的文本。在本示例中,用户已经在文本块516中突出了文本用于翻译。作为响应,翻译器115可以识别所选择的文本的翻译并在覆盖中提供翻译。例如,屏幕截图500D描绘了响应于对文本512的用户选择在图像512之上的覆盖549中呈现翻译的文本的用户界面540。用户还可以选择覆盖549以请求关于翻译的文本的附加信息。

[0087] 在一些实现中,用户界面可以提供用户界面元素,所述用户界面元素能被拖动至用户想翻译的文本。例如,用户界面元素可在触屏设备处被显示并且用户可以将用户界面元素拖动至期望的文本。作为响应,用户设备可以从翻译器115请求文本的翻译并且使用翻译器115所选择的用户界面呈现所述翻译。

[0088] 图6描绘了示例用户界面的屏幕截图600A和600B,其为主要的文本提供了翻译并

使用户能够选择次要的文本用于翻译。在屏幕截图600A中,用户界面610描绘了两个街道标志617和619的图像615。每个街道标志617和619都描绘了中文字符的文本。为了将图像615中的文本从中文字符翻译为英文,翻译器115可以识别包括中文字符的图像中的文本块。例如,翻译器115可以将街道标志617中的中文字符识别为文本块620并将街道标志619中的中文字符识别为文本块625。

[0089] 为了确定要翻译哪个文本以及如何呈现所翻译的文本,翻译器115可以评估每个所识别的文本块620和625的突出度。在本示例中,翻译器115可以确定文本块620相较于文本块625被更主要地描绘,因为文本块620相较于文本块625更接近于图像615的中心。此外,文本块620的文本大小大于文本块625的文本大小。

[0090] 响应于确定文本块620相较于文本块625被更主要地描绘,翻译器115可以确定自动地翻译文本块620的文本,而不会自动地翻译文本块625的文本。如屏幕截图600B中所示出的,翻译器115可以选择用户界面640,所述用户界面640在图像615中的文本块620的中文字符之上(或替代)的覆盖643中呈现文本块620的翻译。

[0091] 在本示例中,完整的翻译无法容纳进覆盖643中。用户界面640可以提供用户界面元素647,当所述用户界面元素647被选择时,使得用户界面640呈现全部翻译。在一些实现中,翻译器115可以确定翻译的可读性度量没有满足阈值,并且作为响应,在另一个屏幕中呈现翻译而不是呈现部分翻译。

[0092] 用户界面640还可以在自动地被翻译的次要的文本处提供诸如突出显示的块的用户界面元素。例如,用户界面640包括用户界面元素645以将没有被翻译的文本块625突出显示。响应于对用户界面元素645的用户选择,用户界面640可以呈现被包括在文本块625中的文本的翻译。该翻译可以被呈现在街道标志619或文本块625之上的覆盖中,或者被呈现在单独的屏幕或用户界面中。

[0093] 一些图像可以以突出的方式描绘文本的单个块。例如,用户可以使用用户设备以捕捉指令或单个标志的图像以获取所述指令或标志的翻译。在这样的情况下,翻译器115可以识别文本并自动地在特定用户界面中呈现文本的翻译。例如,响应于确定图像描绘了单个主要被显示的文本块,翻译器115可以为图像选择单个块场境。单个块场境的用户界面可以包括:在图像之上的覆盖中呈现被包括在单个文本块中的文本的翻译的用户界面;在没有图像的单独的屏幕中呈现被包括在单个文本块中的文本的翻译用户界面,诸如使得用户能浏览或滚动通过被翻译的文本的屏幕;和/或其他类型的用户界面。

[0094] 图7描绘了示例用户界面的屏幕截图700A和700B,所述示例用户界面呈现了图像中描绘的文本的语言翻译。屏幕截图700A描绘了呈现图像715的用户界面710。图像715描绘了具有中文字符的文本块720以及具有英文字符的文本块725。在本示例中,用户已经请求了从中文到英文的翻译。因此,翻译器115可以分析图像以识别任意中文字符并识别包括在文本块720中的文本,而忽略文本块725中的英文字符。翻译器115可以进一步分析包括在文本块720中的文本以确定文本块720是包括彼此相互关联的文本的单个文本块。例如,翻译器115可以比较文本大小、文本颜色、文本的背景颜色、文本的位置和接近度、和/或其他特征以确定包括在文本块720中的文本是相关的。作为响应,翻译器115可以为图像选择单个块场境并基于所述选择识别用户界面。

[0095] 如屏幕截图700B中所示出的,翻译器115已经选择了在图像715之上的覆盖760中

呈现文本块720的翻译的用户界面750。用户界面750还包括用户界面元素755,当用户选择了所述用户界面元素755时,用户界面750一次呈现一行的翻译文本。例如,用户能够浏览翻译的文本的每行,其中,每行都被呈现在单独的显示屏中而不是每个其他行中。

[0096] 一些图像可以描绘同等或类似突出的两个或更多个文本的块。例如,用户可以捕捉两个标志的图像,以确定要经过哪个方向或者哪扇门。在这样的情况下,呈现两个标志的翻译以帮助引导用户可能是有益的。

[0097] 图8描绘了用于呈现具有类似突出度的文本块的语言翻译的示例用户界面的屏幕截图800A-800C。在屏幕截图800A中,用户界面810呈现包括两个标志820和830的图像815,所述标志820和830包括中文字符。在本示例中,用户正在请求从中文字符到英文的翻译。因此,翻译器115可以识别包括在标志820中的中文字符以及包括在标志830中的中文字符。翻译器115还可以分析包括在每个标志820和830中的中文字符以确定每个标志的中文字符都被包括在有区别的文本块中。例如,翻译器115可以将包括在标志820中的中文字符识别为文本块825并将包括在标志825中的中文字符识别为不同于文本块825的文本块835。

[0098] 响应于识别图像815中的多个文本块825和835,翻译器115可以评估文本块的突出度并确定文本块825和835具有类似的突出度,因为每个文本块825和835的文本大小和颜色是类似的并且文本块825和835距离图像815中心的距离相似。因为文本块825和835具有类似的突出度,翻译器115可以为图像815选择同等突出场境并选择对应于所述同等突出场境的用户界面。

[0099] 如屏幕截图800B中所示出的,选择了用户界面840以呈现文本块的翻译。用户界面840在位于标志820上方的覆盖845中呈现文本块825中的中文字符的翻译,其中,文本块825中的中文字符被识别位于所述标志820处。类似地,用户界面840在位于标志830上方的覆盖847中呈现文本块835中的中文字符的翻译,其中,文本块835中的中文字符被识别位于所述标志830处。每个覆盖845和847也是可选择的以请求覆盖的对应文本的完整翻译。例如,如在屏幕截图800C中所示出的,在用户界面860中呈现“toilet”的中文的完整翻译是用户选择覆盖845的结果。

[0100] 诸如翻译器115的翻译器可以评估图像描绘的文本的布置和视觉特征以为图像选择突出呈现场境。翻译器115还可以为所识别的文本的至少一部分选择用于呈现附加信息的用户界面。可以基于所选择的突出呈现场境来选择用户界面。例如,如果单个块场境被选择,则对应于单个块场境的用户界面可以被选择并被用于呈现附加信息。

[0101] 图9是用于基于为图像识别的突出呈现场境来呈现关于图像中描绘的文本的附加信息的示例过程900的流程图。例如,通过诸如图1中的用户设备100和/或翻译器115的数据处理装置,可以实现过程900的操作。还可以通过存储在计算机存储介质上的指令实现过程900,其中通过数据处理装置对指令的执行使得所述数据处理装置执行该过程900的操作。

[0102] 接收图像(902)。例如,用户可以使用诸如智能电话、平板电脑或其他用户设备的用户设备对目标拍照片。标志可以包括对用户来说是外语的文本,其也可以在所接收的图像中被描绘。

[0103] 识别所接收的图像中的文本(904)。可以使用OCR或其他文本识别技术来识别所述文本。在一些实现中,只识别特定语言的文本。例如,用户可以指定特定语言的文本应被翻译为另一种语言,诸如用户理解的语言。在这个示例中,只有特定语言的文本可以在图像中

被识别。

[0104] 除了实际的文本,可以识别文本的特征(例如,视觉特征)。这些特征可以包括文本的颜色、文本之下的背景的颜色、文本的大小、文本的字体类型、和/或图像内的文本位置。

[0105] 在一些实现中,可以忽略被剪切或者不完整的文本。例如,如果只有标志的一部分被图像描绘并且标志的文本包括诸如代替“stop,”的“sto”的不完整的词,则标志的文本可能不会被识别用于提供附加信息或用于为图像选择突出呈现场境。翻译不完整的文本会使用户感到困惑或导致无意义的翻译。

[0106] 基于所识别的文本,识别一个或多个有区别的文本块(906)。可以评估文本的特征以识别有区别的文本块。例如,与其他文本临近(例如,在阈值距离之内)并具有类似的外观(例如,类似的大小、文本大小和/或背景颜色)的文本可被视为文本块A的部分。相反地,远离(例如,大于阈值距离)文本块A并且与文本块A的文本具有不同的文本大小、不同的文本颜色和/或不同的背景颜色的文本的另一部分可以被识别为另一个有区别的文本块B。文本的两个部分的视觉特征(例如,文本颜色、文本大小、字体类型和/或背景颜色)的相似性和文本的两个部分之间的距离可以被用于确定文本的两个部分是否属于共同的文本块。

[0107] 基于所识别的文本的突出度,选择用于图像的突出呈现场境(908)。可以从多个突出呈现场境中选择突出呈现场境。例如,多个突出呈现场境可以包括单个块场境、主-次场境、和/或同等突出场境。

[0108] 为了评估一个或多个文本块的突出度,可以为每个文本块计算突出度分值。每个文本块的突出度分值可以基于文本块在图像内的位置、包括在文本块中的文本的大小、包括在文本块中的文本的多个特征、包括在文本块中的文本的颜色、文本块的背景的颜色、和/或指示文本块的突出度的其他因素。位置可以是图像中的绝对位置(例如,使用坐标系)或者是从图像的中心到文本块的一个点(例如,文本块的中心或最接近图像的中心文本块的点)的距离。例如,更接近图像的中心文本块相较于更远离图像的中心文本块具有更高的突出度分值。类似地,具有大文本大小(或多特征)的文本块相较于具有较小的文本大小(或较少特征)的文本块可能具有更高的突出度分值。此外,较亮的文本或背景颜色相较于较暗或较单调的文本或背景颜色可能导致更高的突出度分值。

[0109] 如果对图像识别了单个有区别的文本块,可以为图像选择单个块场境。在一些实现中,在为图像选择单个块场境之前,可以对单个有区别的文本块的突出度进行评估。例如,如果单个有区别的文本块是图像的角落中的文本的小块,单个有区别的文本块可能具有低的突出度分值(例如,低于突出度分值阈值的突出度分值)。在这样的情况下,可以不选择单个块场境并且可以不自动翻译文本。

[0110] 如果对图像识别了多个有区别的文本块,则可以评估文本块的相关突出度以确定是选择主-次场境还是同等突出场境。例如,可以为每个有区别的文本块计算突出度分值并将所述突出度分值进行比较。

[0111] 如果文本块相较于为图像所识别的每个其他的文本块具有显著高的突出度分值,则可以为图像选择主-次场境。例如,如果文本块A具有大于每个其他文本块至少阈值量的突出度分值,则文本块A可被视为主要的文本块并且可以为图像选择主-次场境。如果为图像所识别的两个或更多个文本块具有类似的突出度,则可以为图像选择同等突出场境。例如,如果文本块B具有在文本块A的突出度分值内的阈值量的突出度分组并且文本块A和B的

突出度分值大于为图像所识别的任意其他文本块的突出度分值,则可以为图像选择同等突出场境。

[0112] 基于所选择的突出呈现场境,选择将要为其呈现附加信息的文本(910)。例如,如果为图像选择了单个块场境,可以选择包括在图像中所识别的单个有区别的文本块中的文本用于提供附加信息。该附加信息可以是包括在单个有区别的文本块中的文本的语言翻译。如果为图像选择了主-次场境,则可以选择包括在主要文本块(例如,具有最大突出度的文本块)中的文本用于提供附加信息。如果为图像选择了同等突出场境,则可以选择包括在具有类似突出度的文本块中的文本用于提供附加信息。

[0113] 为所选择的文本识别附加信息(912)。附加信息可以包括所选择的文本的语言翻译。例如,用户可以选择用户希望图像描绘的文本将要被翻译成的语言。对于包括货币形式的数量的文本,附加信息可以包括到用户所指定的货币的量的兑换。例如,如果所选择的文本包括日元价格,则附加信息可以包括对应于所述日元价格的美元数量。

[0114] 基于所选择的突出呈现场境和/或附加信息的可读性度量,选择用户界面(914)。每个突出呈现场境可以具有可以被用于(或有资格被用于)呈现用于图像中描绘的文本的附加信息的一个或多个对应的用户界面。此外,有资格被用于一个突出呈现场境的用户界面可以不同于用于每个其它的突出呈现场境的用户界面。

[0115] 单个块场境可以具有在图像之上的覆盖中呈现附加信息的用户界面(例如,图7中的用户界面)和/或使用户能够在一个或多个单独的屏幕中浏览附加信息的用户界面(例如,图3中的用户界面)。每个单独的屏幕可包括使得用户能够浏览附加屏幕以查看更多附加信息的控件。

[0116] 主-次场境可包括为图像中描绘的最突出的文本块呈现附加信息的用户界面(例如,图3中的用户界面);在图像上呈现包括与最突出的文本有关的附加信息的覆盖而不呈现次要文本块的附加信息的用户界面;在图像上呈现包括与最突出的文本块有关的附加信息和用户界面元素的覆盖的用户界面,当所述用户界面元素被选择时,呈现关于次要文本块的附加信息(例如,图6中的用户界面);和/或在图像上呈现用于呈现与覆盖的对应文本块有关的附加信息的每个文本块的覆盖的用户界面(例如,图4中的用户界面)。

[0117] 同等突出场境可以包括对已经被识别为具有同等突出度的文本块中的每个都在覆盖中呈现附加信息的用户界面(例如,图7中的用户界面)和/或使得用户能在一个或多个单独的屏幕中浏览同等突出的文本块(以及图像中所识别的任意其他文本块)的附加信息的用户界面(例如,图3中的用户界面)。

[0118] 每个突出呈现场境可以具有用于该场境的优选用户界面。例如,优选用户界面可以是在图像之上的覆盖中呈现附加信息的用户界面。通过另一个示例,用户可以能够为每个突出呈现场境选择优选用户界面。响应于对特定突出呈现场境的选择,可以选择用于特定突出呈现场境的优选用户界面。

[0119] 在一些实现中,附加信息的可读性度量被用于选择用于呈现附加信息的用户界面。如上面所述,如果附加信息被使用特定用户界面呈现至用户,可读性度量可以指示附加信息的可读性如何。用于用户界面并用于特定附加信息的可读性度量可以是基于可以使用用户界面被呈现的附加信息的多个特征。可读性度量可以等于在(a)能够使用用户界面被呈现的附加信息的特征数目;以及(b)附加信息中的特征的总数目之间的比率或与所述比

率成比例。例如,覆盖可能具有有限的空间量来呈现附加信息。空间量可以是基于覆盖将要为其呈现附加信息的文本块的大小。如果只有翻译的部分被呈现在覆盖中,这会使得用户困惑。在这样的情况下,在单独的屏幕中呈现附加信息可能是更好的,可以在所述单独的屏幕中呈现翻译的全部或至少是部分。

[0120] 可读性度量可以被用于确定是否为所选择的突出呈现场境使用优选用户界面(例如,在图像之上带有覆盖的用户界面)或者是否使用替选的用户界面(例如,使得用户能够浏览附加信息的用户界面)。例如,如果附加信息相对于优选突出呈现场境的可读性度量满足可读性阈值(例如,通过满足或超过可读性阈值),则可以选择优选用户界面。如果附加信息相对于优选突出呈现场境的可读性度量没有满足可读性阈值(例如,可读性度量小于可读性阈值),则可以选择允许附加信息的更多特征的呈现的替选的用户界面。

[0121] 可以在独立于优选用户界面的多个用户界面之间选择可读性度量。例如,可以选择有资格用于所选择的突出呈现场境并具有最高的与附加信息关联的可读性度量的用户界面以呈现附加信息。

[0122] 使用所选择的用户界面呈现附加信息(916)。例如,用户设备可以生成用户界面并使用所选择的用户界面呈现附加信息。

[0123] 集合呈现场境和用户界面

[0124] 一些图像可能在图像内具有文本的许多有区别的部分,所述文本的许多有区别的部分具有各种突出度水平。例如,餐厅菜单的图像可能包括用于图像中描绘的每个菜单项的文本列表。在该场境中,如果识别相关的文本块的集合并将每个集合的翻译(或其他附加信息)呈现在一起,可能对用户更加有益。例如,用户可能希望查看每个菜单项的翻译以确定点哪项,而不是只查看图像中心附近的菜单项。同时,将菜单项看做文本的一个块可能会使用户困惑,类似于图2中的用户界面250。诸如图1中的翻译器115的翻译器可以评估图像中的多个文本块的布置和/或视觉特征以确定如何将文本块的翻译呈现至用户,如参照图10-13更加详细地被描述的。

[0125] 图10描绘了用于捕捉图像并且呈现图像中描绘的文本的语言翻译的示例用户界面1000A-1000C的屏幕截图。在屏幕截图1000A中,用户界面1010描绘了例如响应于对图像捕捉按钮1012的用户选择所捕捉到的菜单的图像1015。图像1010描绘了中文文本的餐厅1020的名称以及菜单项1022的列表。图像1010还描绘了每个菜单项1022的价格1024。

[0126] 如用户界面1010的顶部附近所示出的,用户已经选择了将中文文本翻译为英文文本。翻译器115可以识别图像1015中的中文文本并基于文本的布置和呈现(例如,基于文本的大小、文本的颜色、背景颜色以及文本各部分之间的距离)确定文本的哪些部分是有区别的文本块。在本示例中,翻译器115已经将餐厅1020的名称识别为有区别的文本块并且将每个菜单项识别为有区别的文本块。

[0127] 翻译器115还可以评估文本块的布置和呈现以确定两个或更多个文本块是否相关并应该属于文本块的集合。具体地,翻译器115可以评估文本块的大小和位置、文本块的背景颜色、文本的风格以及语言模型以确定是否将文本块分组在集合内。用于确定将哪个文本块分组在集合内的过程在图13中被图示并在下面被描述。

[0128] 在本示例中,例如,由于文本和背景之间的相似性并且还由于邻近的菜单项之间类似的间隔,翻译器115已经将菜单项识别为单个集合。例如,由于餐厅1020的名称相对于

菜单项(例如,左对齐)方向上的差异(例如,在菜单上居中)以及餐厅1020的名称和菜单项1022之间的文本和背景颜色中的差异,餐厅1020的名称可以被排除在集合之外。

[0129] 响应于识别图像1010中的文本块的单个集合,翻译器115可以为图像1010选择单个集合场境。翻译器115还可以选择用于呈现图像1015描绘的文本的翻译的用户界面。可以从有资格或对应于单个集合场境的用户界面中选择该用户界面。例如,如屏幕截图1000B和1000C中所示出的,用于单个集合场境的用户界面可以使得用户能够在包括在集合中的每个文本块的翻译之间进行浏览。屏幕截图1000B中描绘的用户界面1030利用框1032突出显示餐厅的名称并利用框1034突出显示每个菜单项。响应于对诸如餐厅的名称或菜单项的突出显示的文本框的选择,用户界面1050可以被呈现。该用户界面1050以原始语言(中文)呈现餐厅的名称并以用户指定的语言(英文)呈现餐厅的名称的翻译1054。用户界面1050还包括使得用户能够浏览每个菜单项的翻译的导航控件1056。例如,对箭头1058的选择可以使得用户界面1050呈现下一个项的翻译并且对箭头1057的选择可以使得用户界面1050呈现前一个项的翻译。另一个用户界面元素1059允许用户请求关于所显示的翻译的更多信息。

[0130] 尽管在本示例中,除了菜单项的翻译以外,用户界面1050还允许用户浏览餐厅的名称的翻译,但是在一些实现中,用户界面1050可以只包括被包括在集合中的文本块。可以通过选择用户界面1030中突出显示的文本块,单独请求没有被包括在集合中的其他文本块的翻译。

[0131] 用于单个集合场境的用户界面还可以包括对集合的每个文本块都在图像之上呈现包括文本块的翻译的覆盖的用户界面。例如,图11描绘了用于捕捉图像并且呈现图像中描绘的文本的语言翻译的示例用户界面的屏幕截图1100A-1100C。屏幕截图1100A类似于图10中的屏幕截图1000A并且包括呈现图像1115的用户界面1110,所述图像1115具有餐厅1120的名称、菜单项1122和每个菜单项1122的价格1124。

[0132] 在屏幕截图1100B中,用户界面1130呈现图像1115以及包括餐厅的名称的翻译的覆盖1132、以及用于每个菜单项的覆盖1134。每个覆盖1132包括其相应的菜单项的翻译。每个覆盖1132和1134还被置于其相应的文本块之上,使得将翻译与实际的菜单比较并作出适当的选择对用户来说是容易的。

[0133] 每个覆盖1132和1134还可以是可选择以请求关于文本的附加信息。例如,如屏幕截图1100C中所示出的,响应于对覆盖的选择,用户界面1150可以被呈现。在本示例中,响应于对用于图像1115的底部附近所显示的“nutmilk”的覆盖的选择,用户界面1150被呈现。用户界面呈现“nutmilk”的中文原始文本1152、翻译1154以及使得用户能够浏览每个菜单项和餐厅名称的翻译的导航控件1156。

[0134] 为了确定是否在如图11中所示出的覆盖中或者在如图10中所示出的浏览模式中呈现翻译,翻译器115可以如上述地评估可读性度量。如果覆盖中的翻译的可读性度量中的每一个都满足可读性度量(例如,通过满足或超过可读性度量),翻译器115可以在覆盖中呈现翻译。否则,翻译器115可以在浏览模式中呈现翻译。

[0135] 一些图像可被识别为包括文本块的多个集合。例如,图12描绘了用于呈现被识别为具有文本块的多个集合的图像中描绘的文本的语言翻译的示例用户界面的屏幕截图1200A和1200B。在屏幕截图1200A中,用户界面1210在单个支撑1221上描绘了多个街道标志的图像1215。例如,用户可以在交叉口处并尝试基于标志确定走哪条路。

[0136] 翻译器115可以识别每个街道标志上描绘的文本并评估所述文本以确定每个街道标志包括与每个其他文本块中的文本块有区别的文本块。例如,翻译器115可以基于每个街道标志和每个其他街道标志的文本之间的距离以及邻近的街道标志之间的背景颜色中的不连续性,确定每个街道标志的文本是有区别的文本块。

[0137] 翻译器115还可以评估所识别的文本块的布置和呈现以识别可能相关的文本块的组。每个组可能接着被归类为集合。例如,翻译器115可以将被贴在支撑1221左侧的街道标志的文本块识别为第一集合1220,因为文本块具有相同的文本颜色、相同的背景颜色、相同的方向、类似的文本大小,并且竖直地以相似的距离隔开。类似地,翻译器115可以识别包括被贴在支撑1221右侧的街道标志的文本块的第二集合1225。由于由支撑1221提供的文本块之间的间隔,所述第二集合1225可被识别为与第一集合1220分开的集合。

[0138] 响应于为图像1210识别多个集合,翻译器115可为图像1210选择多个集合场境。翻译器115还可以选择用于呈现翻译或与文本块有关的其他附加信息的用户界面。用于多个集合场境的用户界面可以包括在覆盖中呈现翻译的用户界面和/或使得用户能够使用导航控件浏览翻译的用户界面。在用户界面的这两种类型之间的选择还可以基于可读性度量。

[0139] 在屏幕截图1200B中,用户界面1240呈现图像1215的变暗版本、街道标志中的一个的原始文本1242和该街道标志的翻译1244。用户界面1240还包括使得用户能够浏览街道标志的翻译的导航控件1246。在这个示例中,导航控件使得用户能够浏览街道标志中的八个的翻译。这八个街道标志包括第一集合1220的街道标志中的每一个,以及较低的第二集合1225的四个街道标志。第二集合1225中的最高的街道标志可以被排除在外,因为它的文本的部分被其他街道标志截掉了。因此,准确的翻译可能不可用并且翻译部分文本对用户没有帮助。

[0140] 在一些实现中,用户界面1240可以使得用户一次能够浏览两个集合中的一个。例如,响应于对第一集合中的文本块的选择,用户界面可以为第一集合1220的街道标志呈现文本块的翻译。导航控件1246可以接着使得用户能够只浏览第一集合的文本块的翻译,而不包括第二集合1225的文本块。用户界面1240还可以包括用户界面元素1249,当所述用户界面元素1249被选择时,将用户返回至描绘图像1215的用户界面1210。用户可以接着选择第二集合的街道标志的文本块并使用类似于用户界面1240的用户界面浏览第二集合的街道标志描绘的文本的翻译。

[0141] 诸如翻译器115的翻译器可以评估图像描绘的文本的布置和视觉特征以为图像选择集合呈现场境。翻译器115还可以为所识别的文本的至少一部分选择用于呈现附加信息的用户界面。可以基于所选择的集合呈现场境选择用户界面。例如,如果多个集合场境被选择,则诸如图12的用户界面1240的对应于多个集合场境的用户界面可以被选择。

[0142] 图13是基于为图像所选择的集合呈现场境用于呈现关于图像中描绘的文本的附加信息的示例过程1300的流程图。例如,通过诸如图1中的用户设备100的数据处理装置,可以实现过程1300的操作。还可以通过存储在计算机存储介质上的指令实现过程900,其中通过数据处理装置对指令的执行使得所述数据处理装置执行该过程1300的操作。

[0143] 接收图像(1302)。例如,用户可以使用诸如智能电话、平板电脑或其他用户设备的用户设备给目标拍照片。标志可以包括对用户来说是外语的文本,其也可以在所接收的图像中被描绘。

[0144] 识别所接收的图像中的文本(1304)。可以使用OCR或其他文本识别技术来识别所述文本。在一些实现中,只识别特定语言的文本。例如,用户可以指定特定语言的文本应被翻译为另一种语言,诸如用户理解的语言。在这个示例中,只有特定语言的文本可以在图像中被识别。

[0145] 除了实际的文本,可以识别文本的特征。这些特征可以包括文本的颜色、文本之下的背景的颜色、文本的大小、文本的字体类型、和/或图像内的文本位置。此外,可以忽略被剪切或者另外不完整的文本,如上面所描述的。

[0146] 基于所识别的文本,识别一个或多个有区别的文本块(1306)。可以评估文本的特征以识别有区别的文本块。例如,与其他文本接近(例如,在阈值距离内)并具有类似的外观(例如,类似的大小、文本大小和/或背景颜色)的文本可被视为文本块A的部分。相反地,远离(例如,大于阈值距离)文本块A并且与文本块A的文本具有不同的文本大小、不同的文本颜色和/或不同的背景颜色的文本的另一部分可以被识别为另一个有区别的文本块B。

[0147] 基于图像中的文本块的布置和/或视觉特征,选择用于图像的集合呈现场境(1308)。可以从多个集合呈现场境中选择集合呈现场境,诸如从单个集合场境、多个集合场境和/或集成集合场境。为了确定选择哪个集合场境,可以为图像识别一个或多个集合。

[0148] 可以评估为图像所识别的文本块的特征以识别集合。这些特征可以包括每个文本块的大小和位置、文本块的背景颜色、文本的风格(例如,字体、颜色、全部大写、全部小写、文本大小等等),并且文本块的方向可以被用于将文本块分组在集合内。具有类似特征的文本块可被一起分组在集合中。例如,具有相同字体、相同文本颜色、相同背景颜色并且彼此接近(例如,在彼此的阈值距离之内)的文本块可以被分组在集合内。相反地,具有不同字体和不同颜色的两个文本块可能不会一起被包括在集合中。

[0149] 多个文本块之间的相对距离还可以被用于将文本块分组在集合内。例如,菜单和列表经常相等或至少非常类似地隔开文本行。多个文本块之间类似的距离可以指示文本块是类似的,并且因此可被用于确定多个文本块属于一个集合。

[0150] 文本块的方向也可以被用于确定文本块是否应该一起被包括在集合中。例如,如果几个文本块朝向相对于图像的类似角度,则文本块可能是相关的。相反地,如果一个文本块的文本具有水平方向且另一个文本块的文本以相对于水平的一个角度偏向上,则文本块可能不是相关的,并且因此不一起被包括在集合中。

[0151] 还可以比较每个文本块的背景颜色以确定是否要将文本块包括在集合中。具有类似背景颜色的文本块可能是相关的并可以一起被包括在集合中。相反地,具有基本上不同颜色(例如,一个是蓝色且一个是红色)的文本块可能不是相关的并且可以不一起被包括在集合中。例如,菜单中的列表中呈现的菜单项通常具有相同字体颜色和背景颜色。在本示例中,菜单项可以被一起分组在集合内。

[0152] 语言模型也可以被用于将文本块分组在集合内。例如,语言模型可以被用于确定两个或更多个文本块的文本是否是相关的。如果文本块基于语言模型被确定是相关的,则文本块可以一起被包括在集合中。

[0153] 可以基于所识别出的集合来选择集合呈现场境。例如,如果识别了多个集合,则可以为图像选择多个集合场境。如果为图像识别了单个集合,则可以在单个集合场境和集成集合场境之间做出选择。当可以独立于每个其他文本块呈现被包括在集合中的每个文本块

的附加信息时,可以选择单个集合场境。例如,可以为菜单项的集合选择单个集合场境。当附加信息是相关的使得附加信息应该被一起呈现时,可以选择集成的文本块。例如,将街道标志的翻译一起示出可能更好,使得用户可以将翻译与它的对应的街道标志相匹配。

[0154] 为了在单个集合场境和集成集合场境之间选择,可以使用语言模型。例如,可以使用语言模型来确定文本是指菜单上的多个食物项或者文本是指多个街道标志。诸如数字或价格的标识符的存在也可以被用于在单个集合场境和集成集合场境之间选择。如果集合中的每个文本块都包括标识符,则每个文本块可被独立地呈现。因此,可以选择单个集合场境而不是集成集合场境。

[0155] 为文本识别附加信息(1310)。附加信息可以包括所选择的文本的语言翻译。例如,用户可以选择用户希望图像描绘的文本将要被翻译成的语言。对于包括货币形式的数量的文本,附加信息可以包括到用户所指定的货币的量的兑换。例如,如果所选择的文本包括日元价格,则附加信息可以包括对应于所述日元价格的美元数量。

[0156] 基于所选择的集合呈现场境和/或附加信息的可读性度量,选择用户界面(1312)。每个集合呈现场境可以具有可以被用于为图像中描绘的文本呈现附加信息的一个或多个对应的用户界面。此外,有资格被用于一个集合呈现场境的用户界面可以不同于用于每个其它集合呈现场境的用户界面。

[0157] 多个集合场境可以具有在图像之上的覆盖中呈现附加信息的用户界面(例如,类似于图11中的用户界面1130)和/或使用户能够在一个或多个单独的屏幕中浏览用于集合中的每个文本块的附加信息的用户界面(例如,图12中的用户界面1240)。如上面所描述的,每个集合可被单独浏览则每个其他集合在类似用户界面1240的用户界面中。

[0158] 单个集合场境还可以包括在图像之上的覆盖中呈现附加信息的用户界面(例如,类似于图11中的用户界面1130)和/或使用户能够在一个或多个单独的屏幕中浏览用于集合中的每个文本块的附加信息的用户界面(例如,图12中的用户界面1240)。在一些实现中,单个集合场境可以对应于自动默认为单独为集合中的每个文本块呈现附加信息的用户界面。例如,如果每个文本块都包括标识符,默认的或优选用户界面可以是单独为每个文本块呈现附加信息并允许用户浏览集合中的每个其他文本块的附加信息的用户界面。

[0159] 可以将集成集合场境与同时或在相同屏幕上为包括在集合中的每个文本块呈现附加信息的用户界面(例如,与允许用户浏览无法容纳在屏幕上的信息的用户界面元素)相关联。用于集成集合场境的用户界面可以包括在图像之上的一个或多个覆盖中呈现附加信息的用户界面以及在单独的屏幕上呈现附加信息的用户界面。在覆盖用户界面的示例中,图12中的每个街道标志的翻译可以被呈现在其对应的街道标志之上的覆盖中。以这种方式,用户可以容易地确定哪个翻译对应于哪个街道标志。

[0160] 每个集合呈现场境都可以具有用于该场境的优选用户界面。例如,用于多个集合场境和集成集合场境的优选用户界面可以是在图像之上的覆盖中呈现附加信息的用户界面。用于单个集合场境的优选用户界面可以是单独为每个文本块呈现附加信息的用户界面。

[0161] 在一些实现中,附加信息的可读性度量被用于选择用于呈现附加信息的用户界面。可读性度量被用于确定是否为所选择的集合呈现场境使用优选用户界面(例如,在图像之上带有覆盖的用户界面)或者使用替选的用户界面(例如,使得用户能够浏览附加信息的

用户界面)。例如,如果附加信息相对于优选集合呈现场境的可读性度量满足可读性阈值(例如,通过达到或超过可读性阈值),则可以选择优选用户界面。如果附加信息相对于优选集合呈现场境的可读性度量没有满足可读性阈值(例如,可读性度量小于可读性阈值),则可以选择允许附加信息的更多特征的呈现的替选的用户界面。

[0162] 可读性度量可以被用于在独立于优选用户界面的多个用户界面之间选择。例如,可以选择有资格用于所选择的集合呈现场境并导致最高的可读性度量的用户界面。

[0163] 使用所选择的用户界面来呈现附加信息(1314)。例如,用户设备可以生成用户界面并在所选择的界面中呈现附加信息。

[0164] 地图呈现场境和用户界面

[0165] 在一些场境中,基于图像中描绘的文本接收地图或到一个位置的驾驶路线对用户可以是有益的。例如,用户为事件的海报拍照,所述海报包括事件和所述事件的地址。呈现包括带有位置的翻译或地址的翻译的地图的用户界面可以帮助用户识别事件并确定如何到达所述事件。

[0166] 当地图用户界面的呈现对用户可能有益时,诸如图1中的翻译器115的翻译器可以选择地图呈现场境。例如,响应于检测到捕捉的图像描绘的一个或多个地址,翻译器115可以选择地图呈现场境。如果用户已经向翻译器115提供了位置信息,用于地图呈现场境的用户界面可以包括对应于所检测的地址和/或用户的当前位置的地图。

[0167] 地图呈现场境可以包括单个地址地图场境和多个地址地图场境。响应于在图像中检测到单个地址,可以选择单个地址地图场境。用于单个地址地图场境的用户界面可以包括地图,所述地图示出对应于所述单个地址的位置和/或到所述单个地址的驾驶路线。用户界面还可以呈现地址的语言翻译和/或所述地址所引用的位置或建筑的名称的语言翻译。用于多个地址地图场境的用户界面可以包括控件,所述控件使得用户能够在每个所检测的地址的信息之间进行浏览。例如,用户界面可以使得用户能够在分别示出地图和/或多个地址中特定地址的驾驶路线的卡或屏幕之间进行浏览。每个卡或屏幕还可以呈现地址的语言翻译和/或所述卡或屏幕对应的地址所引用的位置或建筑的名称的语言翻译。地图呈现场境用户界面中所提供的路线指引可以是用户指定的翻译语言。

[0168] 图14描绘了用于呈现与被识别为具有地图场境的图像中描绘的文本有关的附加信息的示例用户界面1400A-1400C的屏幕截图。在屏幕截图1400A中,例如响应于对图像捕捉按钮1414的用户选择,图像1412被捕捉并在用户界面1410中被呈现。示例图像1412包括关于事件的信息和事件的地址1416。

[0169] 翻译器115可以分析图像1412中的文本以识别地址。例如,翻译器115可以识别图像中的文本并将所识别的文本与已知的地址比较。如果存在匹配,翻译器115可以将文本识别为地址。响应于识别地址1414,可以提供允许用户请求并获取关于所述地址的附加信息的用户界面,诸如到所述地址的地图和/或路线指引。例如,用户界面1430可以如屏幕截图1400B中所示出地被呈现。

[0170] 在屏幕截图1400B中,用户界面1430使用地址1414周围的框1432将地址1414突出显示。用户界面1430还提供允许用户查看对应于地址1414的地图的“View on map”控件1434。响应于对控件1434的选择,用户界面1450可以被呈现为屏幕截图1400C中所示出的。用户界面1450呈现包括地址的位置的地图1454、包括地址的翻译的覆盖1452、识别地址1414的地

图1454上的标记1456、和识别地址的文本1458。此外,或备选地,用户界面1450可以呈现事件将要发生的建筑或位置的名称的翻译。

[0171] 在一些实现中,响应于识别图像中的地址,可以自动呈现与地址或位置的翻译一起呈现地图或驾驶路线的用户界面。例如,响应于检测到地址1414,在没有用户选择控件1434的情况下,用户界面1450可以被呈现。

[0172] 示例用户界面1450还包括浏览控件1458,如果存在图像1412中所识别出的另一个地址,所述浏览控件1458将使得用户能够浏览另一个卡或屏幕。例如,对浏览控件1458的选择可以使得用户设备能够呈现地图以及图像1412中所识别的另一个地址的翻译。

[0173] 图15是用于基于为图像所选择的地图呈现场境来呈现关于图像中描绘的文本的附加信息的示例过程1500的流程图。例如,通过诸如图1中的用户设备100和/或翻译器115的数据处理装置,可以实现过程1500的操作。还可以通过存储在计算机存储介质上的指令实现过程1500,其中通过数据处理装置对指令的执行使得所述数据处理装置执行该过程1500的操作。

[0174] 接收图像(1502)。例如,用户可以使用诸如智能电话、平板电脑或其他用户设备的用户设备给目标拍照片。标志可以包括对用户来说是外语的文本,其也可以在所接收的图像中被描绘。

[0175] 识别所接收的图像中的文本(1504)。可以使用OCR或其他文本识别技术来识别所述文本。在一些实现中,只识别特定语言的文本。例如,用户可以指定特定语言的文本应被翻译为另一种语言,诸如用户理解的语言。在这个示例中,只有特定语言的文本可以在图像中被识别。

[0176] 在所识别的文本中识别地址(1506)。例如,可以将所识别的文本与已知的地址比较以确定所识别的文本是否指代地址。如果图像描绘了多个地址,每个地址都可以被识别。

[0177] 基于所识别的地址,识别地图呈现场境(1508)。在一些实现中,基于图像中所识别的多个地址来选择地图场境。例如,响应于图像中识别了单个地址可以选择单个地址地图场境,而响应于图像中识别了多于一个地址可以选择多个地址地图场境。

[0178] 选择用于呈现关于所识别的地址的附加信息的用户界面。如上面所描述的,如果用户已经提供了到当前位置的访问,用于地图呈现场境的用户界面可以描绘示出对应于地址的位置和/或从用户的当前位置到所述地址的驾驶路线的地图。用于地图呈现场境的用户界面还可以在地图或驾驶路线之上的覆盖中包括所述地址或对应于所述地址的位置的名称的语言翻译。对于单个地址地图场境,用户界面可以包括用于单个地址的地图和/或驾驶路线。对于多个地址场境,用户界面可以包括用于为每个所识别的位置浏览地图和驾驶路线的控件,如上面参照图14所描述的。

[0179] 使用所选择的用户界面呈现地址的附加信息(1512)。例如,生成呈现到地址的地图或驾驶路线以及呈现地址或对应于所述地址的位置的名称的翻译的覆盖的用户界面。响应于识别图像中的地址,可以自动呈现该用户界面。在一些实现中,响应于识别图像中的地址,可以呈现用户界面从而允许用户查看地图用户界面。例如,可以生成并呈现类似于图14中的用户界面1430的用户界面。在对“View on map”控件1434的选择之后,可以呈现具有地图或驾驶路线的用户界面。

[0180] 用于为图像选择呈现场境的示例过程

[0181] 如上面所述的,除了在呈现场境的每个类别内(例如,在突出呈现类别或集合呈现类别内)选择呈现场境,翻译器115可以为图像选择呈现场境的类别。例如,响应于接收图像,翻译器115可以确定是否为图像选择突出呈现场境、为图像选择集合呈现场境或为图像选择地图呈现场境。基于图像描绘的文本的布置和/或视觉特征,可以做出该选择。

[0182] 基于所选择的呈现场境,翻译器115还可以选择用于呈现关于图像描绘的文本的附加信息的用户界面。如上面所述的,每个呈现场境都可以与一个或多个用户界面相关联并且用于一个呈现场境的用户界面可以不同于用于其他呈现场境的用户界面。例如,用于在覆盖中呈现主要文本块的翻译的用户界面不同于用于单独呈现单个集合场境中每个文本块的翻译的用户界面。

[0183] 图16是用于基于为图像所选择的呈现场境来呈现关于图像中描绘的文本的附加信息的示例过程1600的流程图。例如,通过诸如图1中的用户设备100和/或翻译器115的数据处理装置,可以实现过程1600的操作。还可以通过存储在计算机存储介质上的指令实现过程1600,其中通过数据处理装置对指令的执行使得所述数据处理装置执行该过程1600的操作。

[0184] 接收图像(1602)。例如,用户可以使用诸如智能电话、平板电脑或其他用户设备的用户设备给目标拍照片。标志可以包括对用户来说是外语的文本,其也可以在所接收的图像中被描绘。

[0185] 识别所接收的图像中的文本(1604)。可以使用OCR或其他文本识别技术来识别所述文本。在一些实现中,只识别特定语言的文本。例如,用户可以指定特定语言的文本应被翻译为另一种语言,诸如用户理解的语言。在这个示例中,只有特定语言的文本可以在图像中被识别。

[0186] 除了实际的文本,可以识别文本的特征。这些特征可以包括文本的颜色、文本之下的背景的颜色、文本的大小、文本的字体类型、和/或图像内的文本的位置。此外,可以忽略被剪切或者不完整的文本,如上面所描述的。

[0187] 基于所识别的文本,识别一个或多个有区别的文本块(1606)。可以评估文本的特征以识别有区别的文本块。例如,与其他文本临近(例如,在阈值距离之内)并具有类似的外观(例如,类似的大小、文本大小和/或背景颜色)的文本块可被视为文本块A的部分。相反地,远离(例如,大于阈值距离)文本块A并且与文本块A的文本具有不同的文本大小、不同的文本颜色和/或不同的背景颜色的文本的另一部分可以被识别为另一个有区别的文本块B。

[0188] 基于文本块的布置和/或文本块的视觉特征,为图像选择呈现场境(1608)。基于文本块,可以确定图像是否与呈现场境中的一个很好地匹配。例如,如果图像包括可以被分组在一个或多个集合内的多个文本块,可以选择集合呈现场境。在本示例中,可以分析图像以识别图像中的任意集合,并且如果识别了任意集合,则可识别多个集合。如果没有识别一个集合,则可以评估每个识别的文本块的突出度以确定是否应该为图像选择突出呈现场境。如果文本块中的一个或多个在图像中被主要地描绘,则可以为图像选择突出呈现场境。

[0189] 还可以分析所识别的文本块以确定文本块中是否包括地址。如果检测到地址,则可以响应于检测到所述地址而自动识别地图呈现场境。在一些实现中,当文本块中检测到小于阈值数目的地址,可以选择地图呈现场境。例如,图像可以包含许多地址并且集合场境或突出场境会更合适。如果具有地址的文本可以被分组在集合内,则可以选择集合呈现场

境。如果地址中的一个相较于其他被描绘为更加突出(例如,地址具有大于阈值量的突出度分值,其大于其他地址的突出度分值),则可以选择突出呈现场境。

[0190] 在一些实现中,可以为图像执行上面所描述的过程900、1300和1500中的一个或多个,并且可以基于呈现场境与图像匹配的程度来对每个呈现场境评分。可以接着为图像选择具有最高分值的呈现场境。

[0191] 可以基于图像中所识别的多个有区别的文本块来选择呈现场境。例如,如果图像中所识别的文本块的数目超过了文本块的阈值数目,可以选择集合呈现场境并且所述文本块可以被包括在集合中。如果图像中识别出小于阈值数目的文本块,可以选择突出呈现场境。基于所识别的每个文本块的突出度,可以选择突出呈现场境,如上面所描述的。

[0192] 基于所选择的突出呈现场境,选择将要对其呈现附加信息的文本(1610)。在一些呈现场境中,只为所识别的文本的部分呈现附加信息。例如,如果选择了主-次文本块,可以识别被识别为主要文本块的文本块的附加信息。因此,可以选择主要的文本块的文本用于呈现附加信息。如果选择了集合呈现场境,则可以识别所识别的集合中的每个文本块。如果选择了地图呈现场境,则可以选择图像中每个识别出的地址用于呈现附加信息。

[0193] 为所选择的文本识别附加信息(1612)。附加信息可以包括所选择的文本的语言翻译。例如,用户可以选择用户希望图像描绘的文本将要被翻译成的语言。对于包括货币形式的数量的文本,附加信息可以包括到用户所指定的货币的量的兑换。例如,如果所选择的文本包括日元价格,则附加信息可以包括对应于所述日元价格的美元数量。

[0194] 基于所选择的呈现场境和/或附加信息的可读性度量,选择用户界面(1614)。如上面所描述的,每个呈现场境可与有资格被用于呈现关于为其选择呈现场境的图像中所描述的文本的附加信息的一个或多个用户界面相关联。例如,基于用于所选择的呈现场境的优选用户界面和/或有资格的用户界面和附加信息的可读性度量,可以选择用于所选择的呈现场境的用户界面中的一个,如上面所描述的。

[0195] 使用所选择的用户界面来呈现附加信息(1616)。例如,用户设备可以生成用户界面并使用所选择的用户界面来呈现附加信息。

[0196] 附加实现细节

[0197] 本说明书中所描述的主题和操作的实施例可以以数字电子电路、或者以计算机软件、固件或硬件的形式被实现,包括本说明书中公开的结构和他们的结构等效物,或以他们中一个或多个的组合的形式被实现。本说明书中所描述的主题的实施例可被实现为一个或多个计算机程序,即,计算机程序指令的一个或多个模块,所述计算机程序指令被编码在计算机存储介质上用于由数据处理装置执行或控制数据处理装置的操作。替选地或附加地,程序指令可以被编码在人工生成的传播的信号上,例如机器生成的电子、光或电磁信号,其被生成以编码信息用于传输至合适的接收端装置,用于被数据处理装置执行。计算机存储介质可以是或被包括在计算机可读存储设备、计算机可读存储基板、随机或串行存取存储器阵列或设备、或他们中一个或多个的组合中。此外,计算机存储介质不是传播的信号,计算机存储介质可以是被编码在人工生成的传播的信号上的计算机程序指令的源或目的地。计算机存储介质还可以是一个或多个物理组件或介质(例如,多个CD、盘或其他存储设备),或被包括在所述一个或多个物理组件或介质之中。

[0198] 本说明书中所描述的操作可被实现为由数据处理装置在数据上执行,所述数据被

存储在一个或多个计算机可读存储设备上或从其他的源被接收。

[0199] 术语“数据处理装置”包括用于处理数据的所有种类的装置、设备和机器,包括通过示例形式的可编程处理器、计算机、片上系统或以上的多个或组合。装置可以包括专用逻辑电路,例如,FPGA(现场可编程门阵列)或ASIC(专用集成电路)。除了硬件以外,装置还可以包括为正在探讨的计算机程序创建执行环境的代码,例如,构成处理器固件、协议栈、数据库管理系统、操作系统、跨平台运行时环境、虚拟机或他们中一个或多个的组的代码。装置和执行环境可以实现多种不同的计算模型基础架构,诸如web服务、分布式计算和网络格式计算基础架构。

[0200] 计算机程序(也被称为程序、软件、软件应用、脚本或代码)可以任意形式的编程语言被编写,包括编译或解释型语言、声明性的或过程的语言,并且其可以以任意形式被部署,包括作为独立程序或模块、组件、子例程、对象或适合用在计算环境中的其他单元。计算机程序可以但不是必须对应于文件系统中的文件。程序可以被存储在保持其他程序或数据(例如,存储在标记语言文档中的一个或多个脚本)的文件的一部分中,在专用于正在讨论的程序的单个文件中,或在多个协调文件(例如,存储一个或多个模块、子程序、或部分代码的文件)中。计算机程序可被部署为在一个计算机上或位于一个地点或分布于多个地点且通过通信网络相连的多个计算机上被执行。

[0201] 本说明书中描述的过程和逻辑流程可通过执行一个或多个计算机程序的一个或多个可编程处理器被执行,以通过对输入数据进行操作并生成输出来执行动作。还可以通过专用逻辑电路执行所述过程和逻辑流程,装置也可以被实现为专用逻辑电路,例如FPGA(现场可编程门阵列)或ASIC(专用集成电路)。

[0202] 适合用于计算机程序的执行的处理器包括例如通用和专用微处理器,以及任意类型的数字计算机的任意一个或多个处理器。一般地,处理器将从只读存储器或随机存取存储器或两者接收指令和数据。计算机的基本元件是用于根据指令执行动作的处理器和用于存储指令和数据的一个或多个存储器设备。一般地,计算机也将包括或被操作耦接,以从用于存储数据的一个或多个大容量储存设备(例如,磁、磁光盘或光盘)接收数据,或向其传送数据,或两者。然而,计算机不需要具有这样的设备。此外,计算机可以被嵌入进另一设备中,例如,移动电话、个人数字助理(PDA)、移动音频或视频播放器、游戏机、全球定位系统(GPS)接收器或便携式存储设备(例如,通用串行总线(USB)闪存驱动),仅举几个例子。适合于存储计算机程序指令和数据的设备包括所有形式的非易失性存储器、介质和存储器设备,通过示例的形式包括半导体存储器设备,例如EPROM、EEPROM和闪速存储器设备;磁盘,例如内部硬盘或可拆卸盘;磁光盘;和CD-ROM以及DVD-ROM盘。处理器和存储器可以通过专用逻辑电路被补充,或被合并并在专用逻辑电路中。

[0203] 为了提供与用户的交互,本说明书中描述的主题的实施例可以在具有以下的计算机上被实现:用于向用户显示信息的显示设备,例如,CRT(阴极射线管)或LCD(液晶显示器)监视器;和键盘以及指示设备,例如鼠标或轨迹球,用户可以通过其向计算机提供输入。其他类型的设备还可被用于提供与用户的交互;例如,提供给用户的反馈可以是任意形式的感知反馈,例如视觉反馈、听觉反馈或触觉反馈;并且来自用户的输入可以以任意形式被接收,包括声学的、语音的、或触觉输入。此外,计算机可以通过发送文档至用户使用的设备并从用户使用的设备接收数据与用户交互;例如,通过响应于从用户的用户设备上的web浏览

器接收的请求而将网页发送至该web浏览器。

[0204] 本说明书中描述的主题的实施例可以被实现在计算系统中,所述计算系统包括后端组件,例如,作为数据服务器,或其包括中间件组件,例如应用服务器,或其包括前端组件,例如,具有用户可通过其与本说明书中描述的主题的实现交互的图形用户界面或web浏览器的用户计算机,或者一个或多个这样的后端、中间件或前端组件的任意组合。系统的组件可以通过任意形式或介质的数字数据通信例如通信网络被相互连接,例如,通信网络。通信网络的示例包括局域网(“LAN”)和广域网(“WAN”)、互联网络(例如,互联网)以及对等网络(例如,自组对等网络)。

[0205] 计算系统可以包括客户端和服务端。客户端和服务端一般远离彼此并典型地通过通信网络交互。客户端和服务端的关系通过在相应的计算机上运行并相互具有客户端-服务器关系的计算机程序发生。在一些实施例中,服务器将数据(例如,HTML页面)传输给客户端设备(例如,为了显示数据至与用户设备交互的用户或从所述用户接收用户输入的目的)。在客户端设备处生成的数据(例如,用户交互的结果)可以在服务端处从用户设备接收。

[0206] 虽然本说明书包含很多特定的实现细节,这些不应该被解释为对任意发明或可以要求保护的内容的范围的限制,但可作为特定于特定发明的特定实施例的特征的描述。在本说明书中在单独的实施例的上下文中描述的某些特征也可在单个实施例中结合地被实现。相反地,在单个实施例的上下文中描述的各种特征还可以单独在多个实施例中或在任意合适的子组合中被实现。此外,尽管特征可在上面被描述为在某些组合中起作用并且初始地如此要求,但是来自所要求保护的组合的一个或多个特征可以在某些情况下从组合中被去除,并且所要求保护的组合可以指向子组合或子组合的变化。

[0207] 类似地,虽然操作在附图中以特定的顺序被描绘,这不应被理解为要求这样的操作以示出的特定顺序或连续顺序被执行,或者所有图示说明的操作被执行,以获得期望的结果。在某些情况下,多任务和并行处理可以是有益的。此外,以上描述的实施例中的多个系统组件的分离不应该被理解为在所有实施例中都要求这样的分离,并且应理解所描述的程序组件和系统通常可以在单个软件产品中被整合在一起或打包进多个软件产品。

[0208] 因此,主题的特定实施例已经被描述。其他实施例在所附权利要求的范围内。在一些情况下,权利要求中记载的动作可以以不同顺序被执行且仍可以获得期望的结果。此外,附图中描绘的过程不必要求所示的特定顺序或连续顺序,以获得期望的结果。在某些实现中,多任务和并行处理可以是有益的。

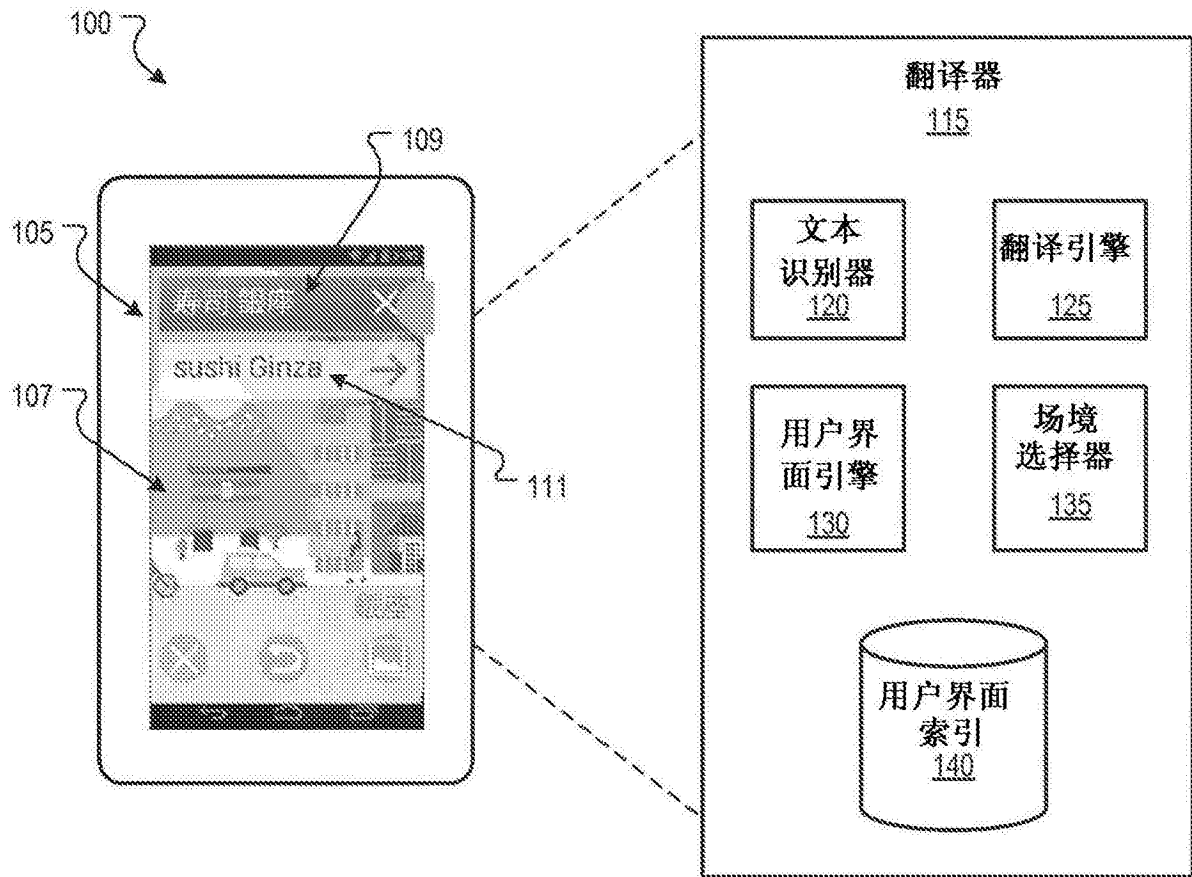


图1

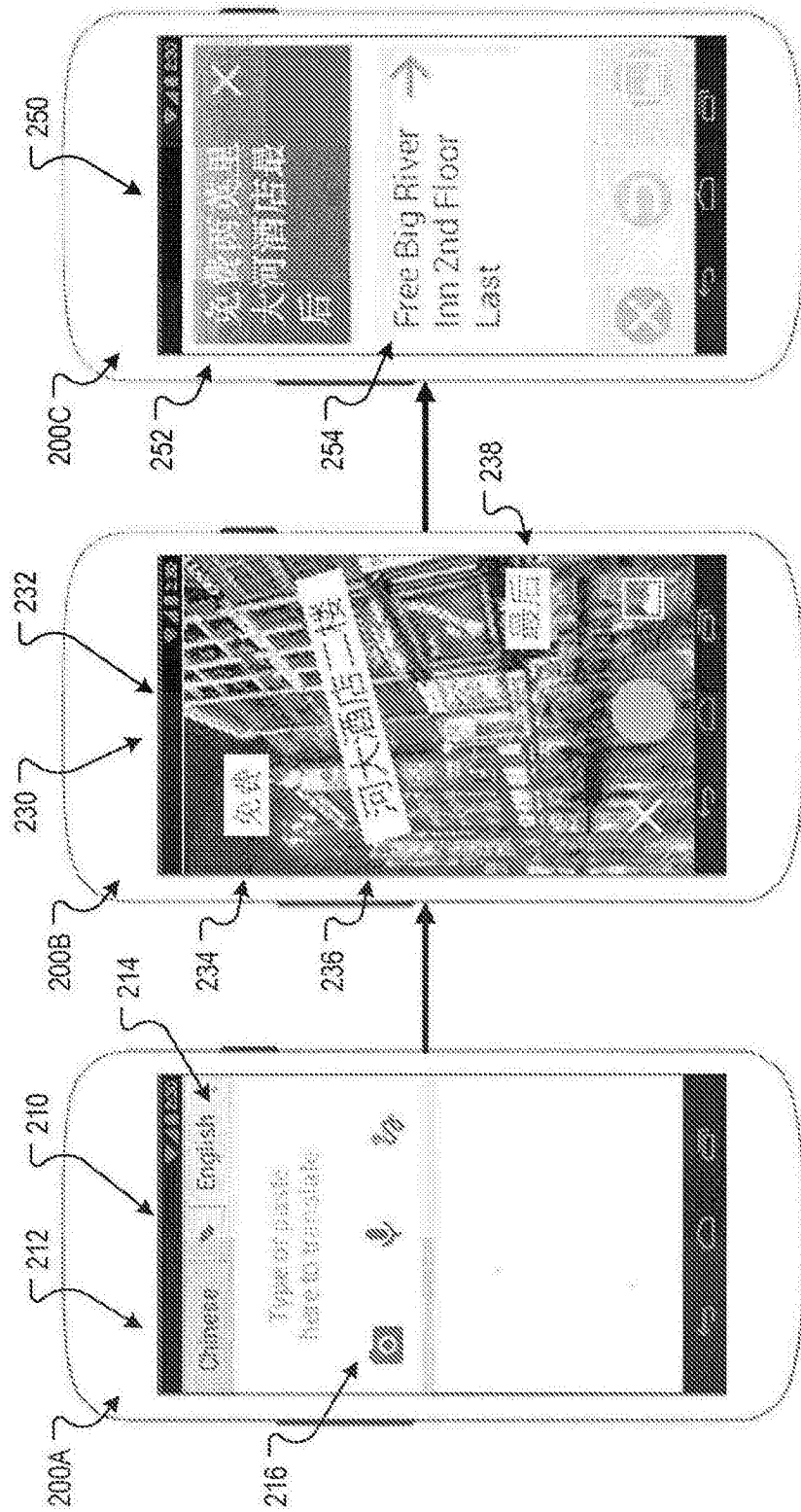


图2

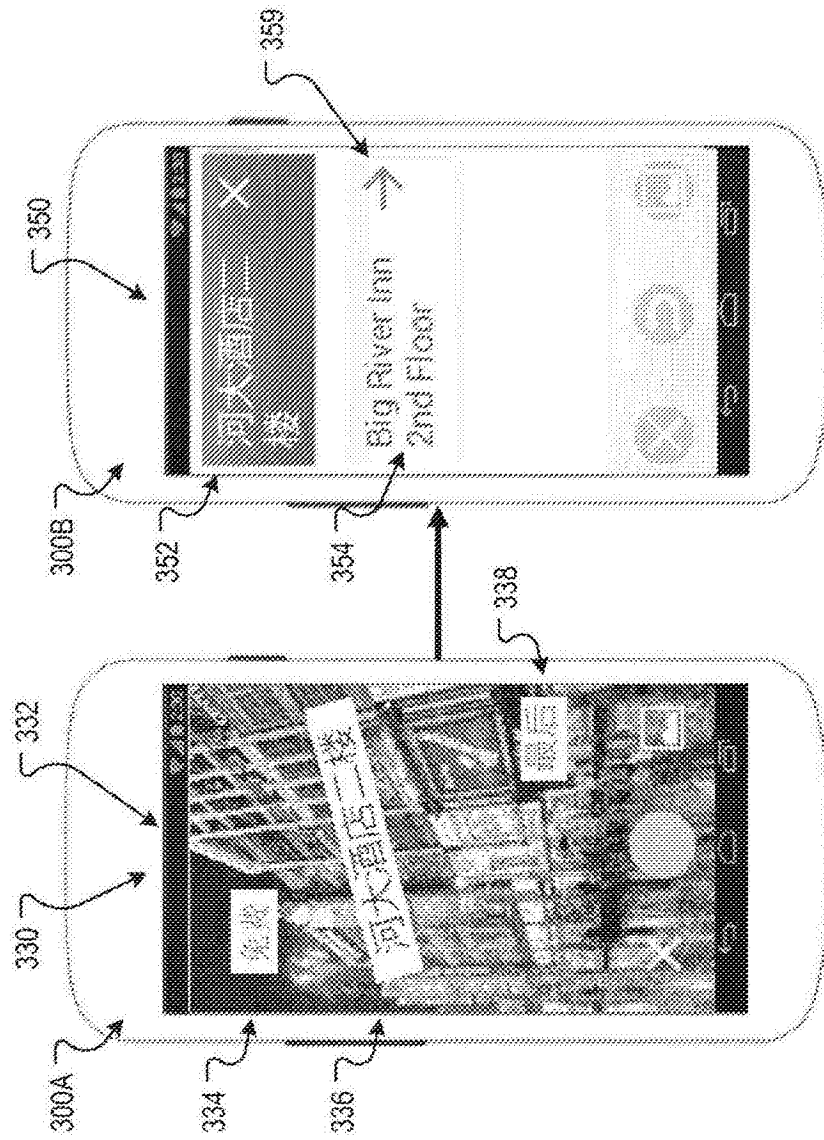


图3

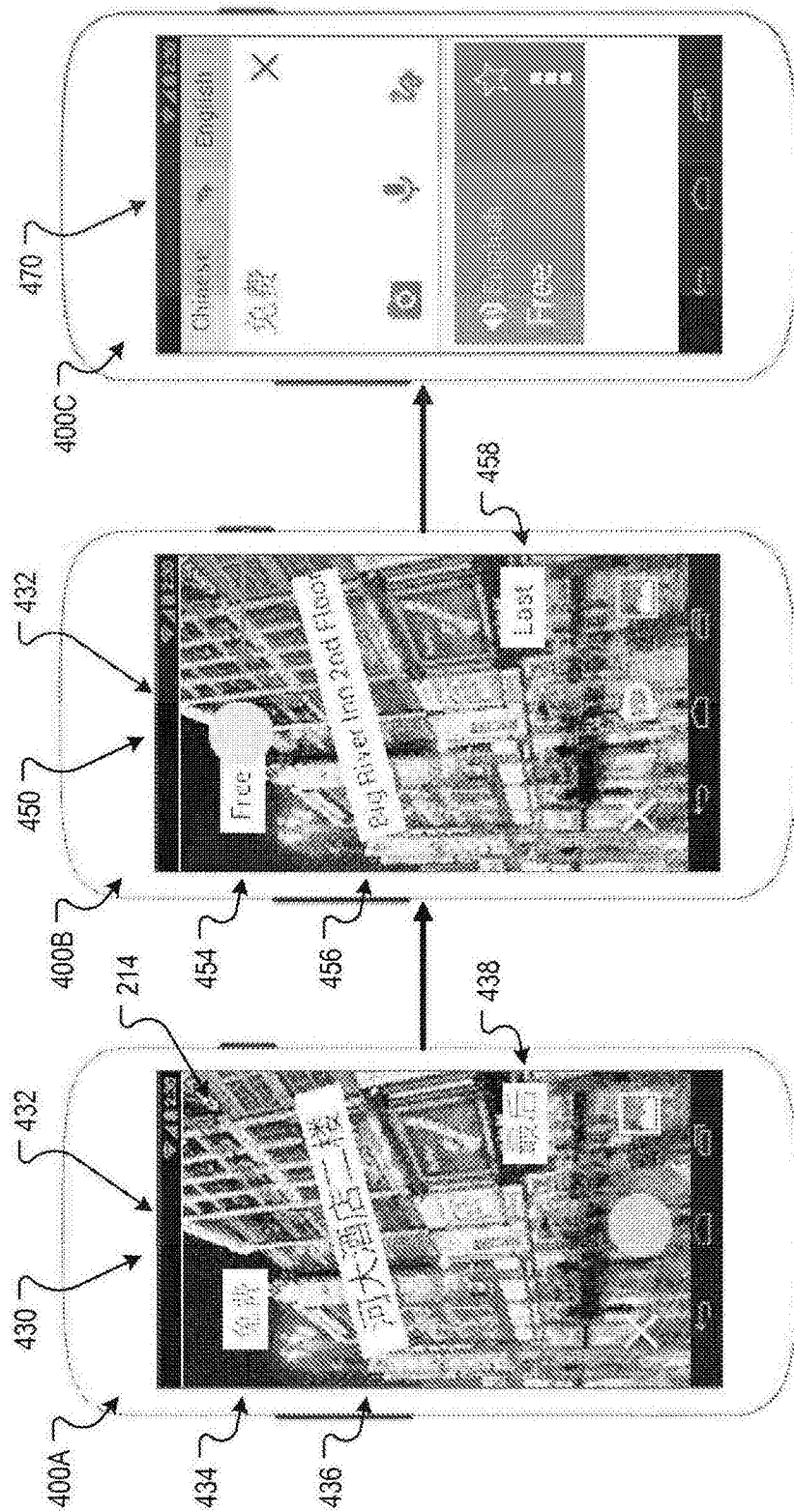


图4

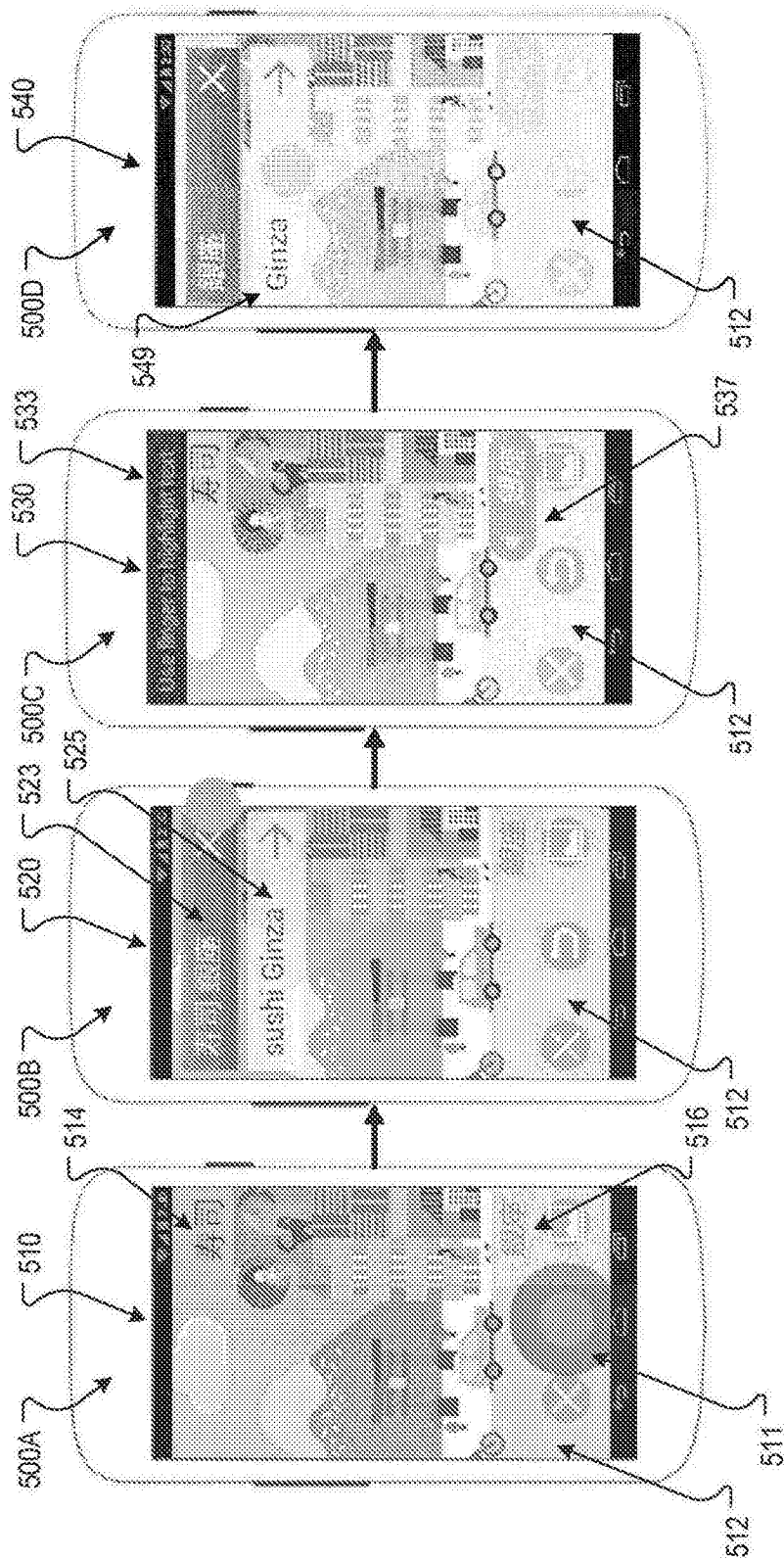


图5

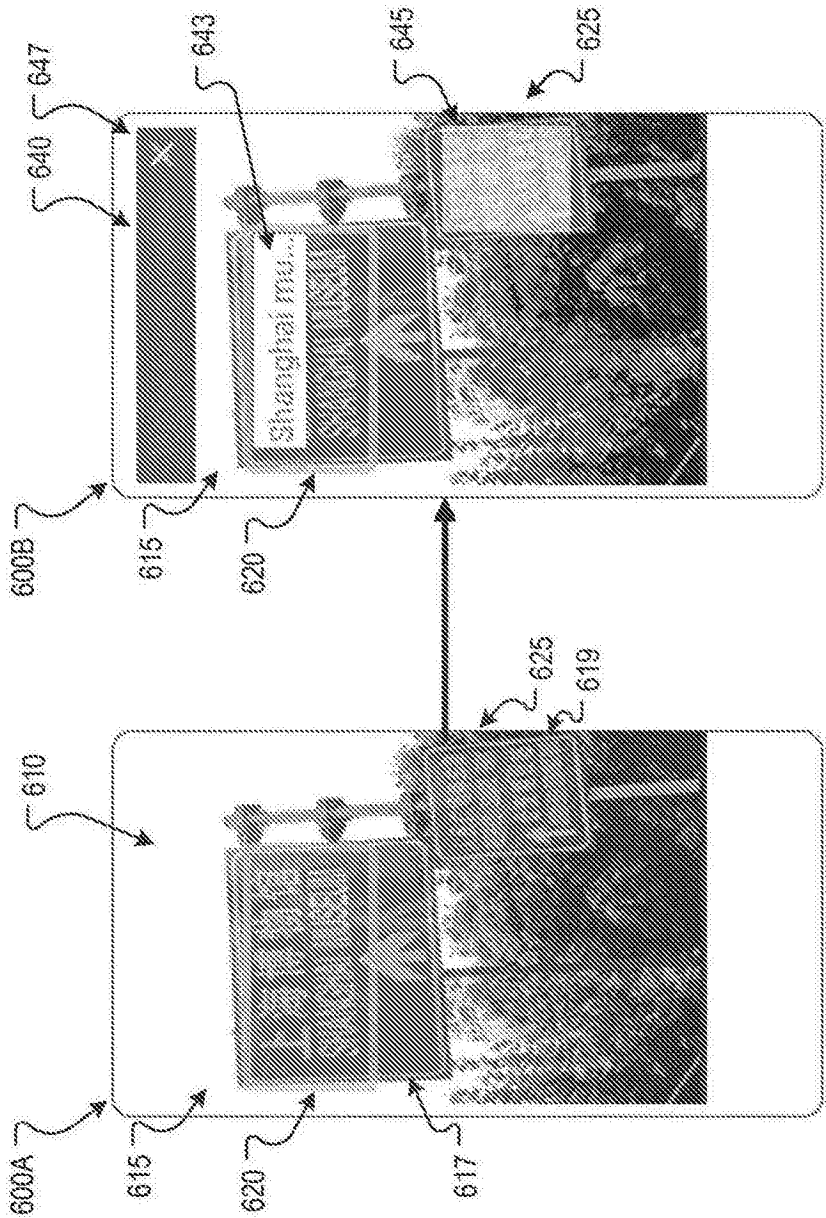


图6

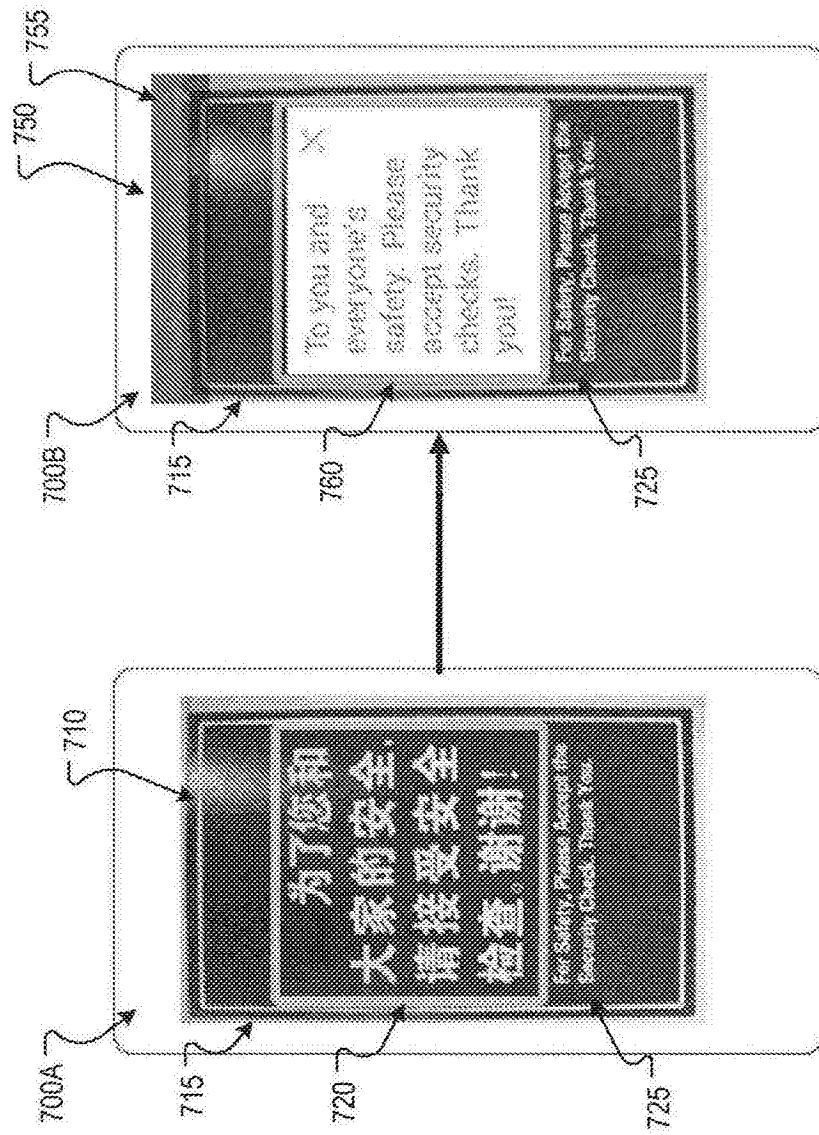


图7

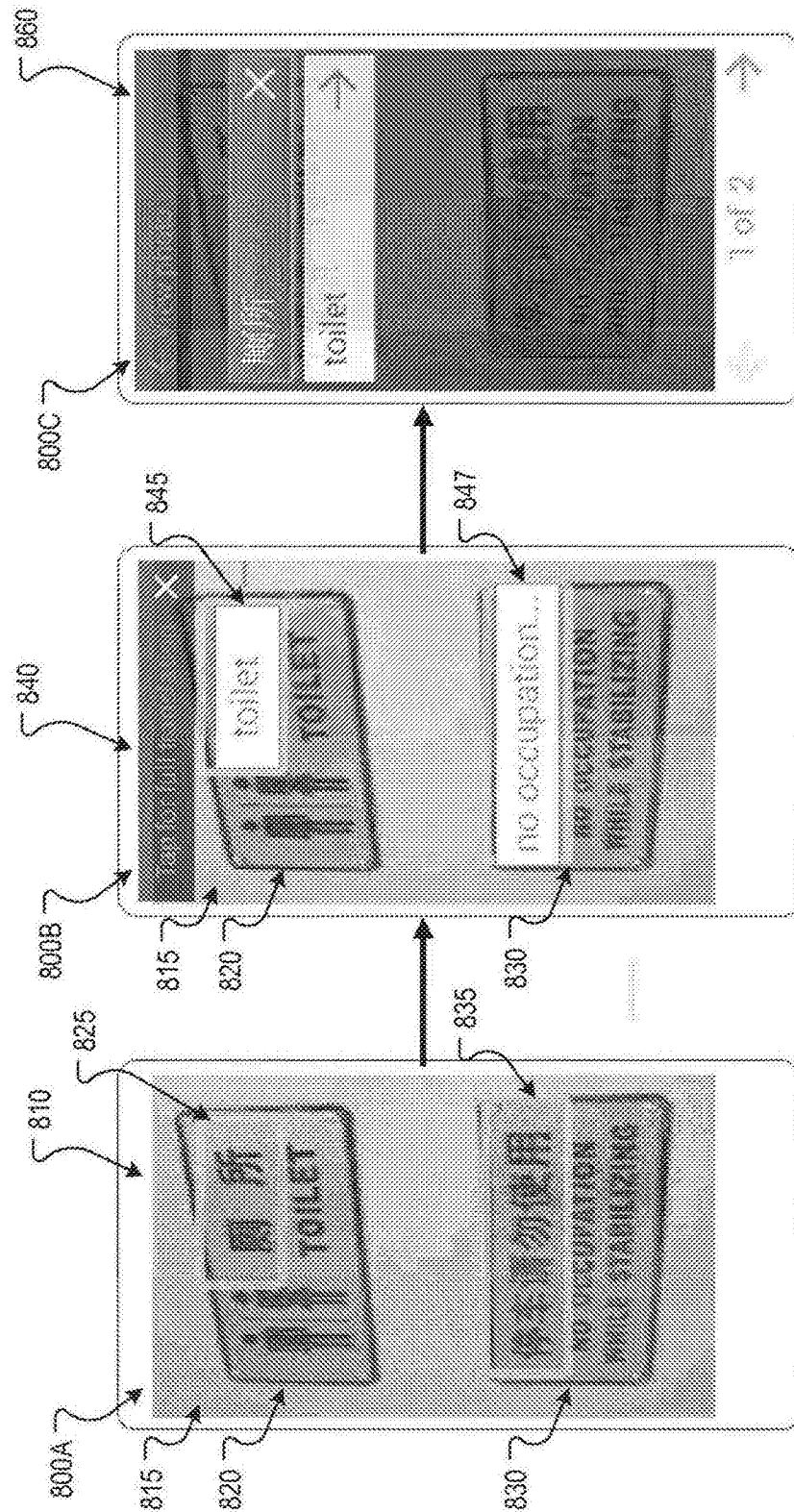


图8

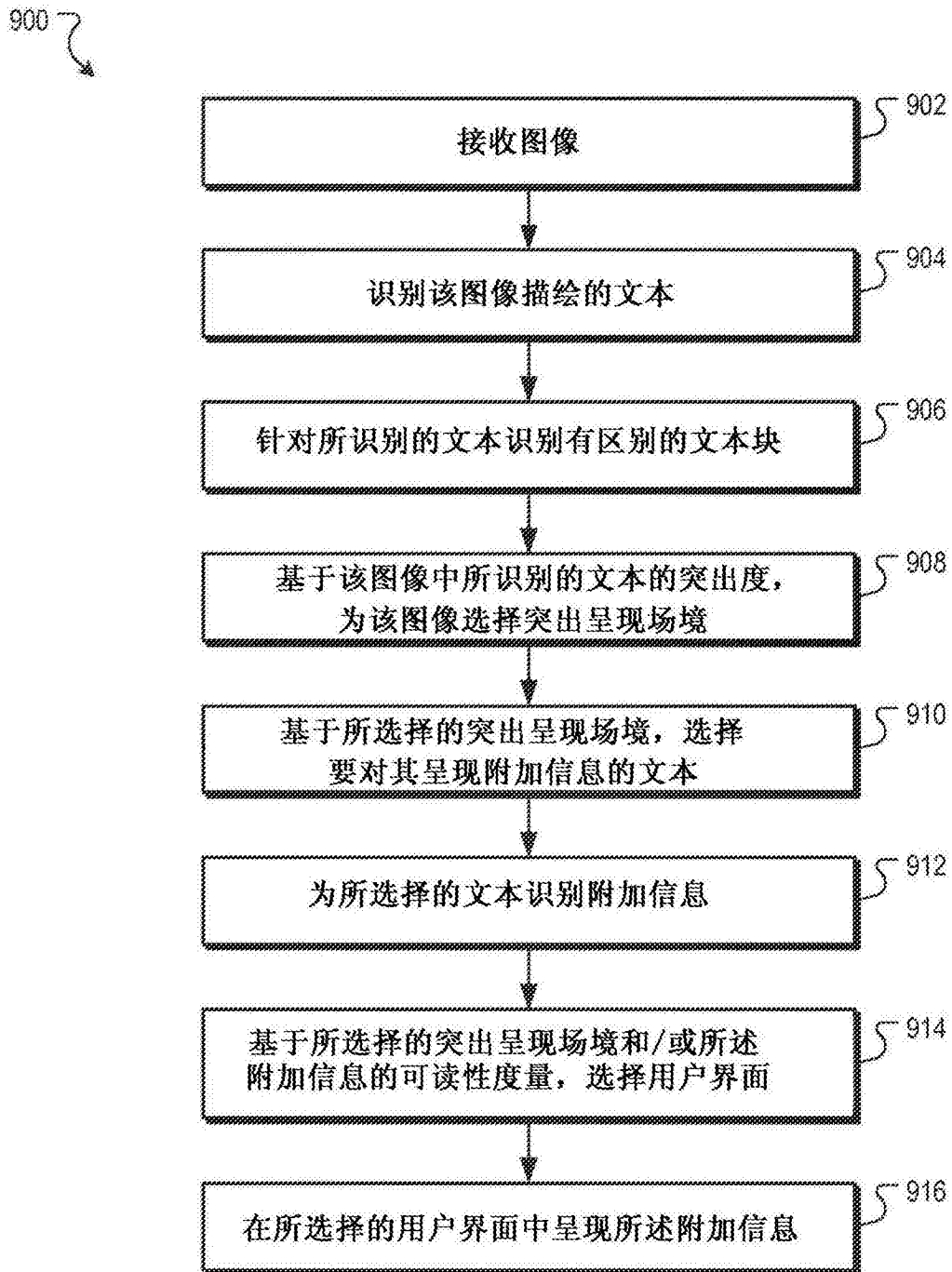


图9

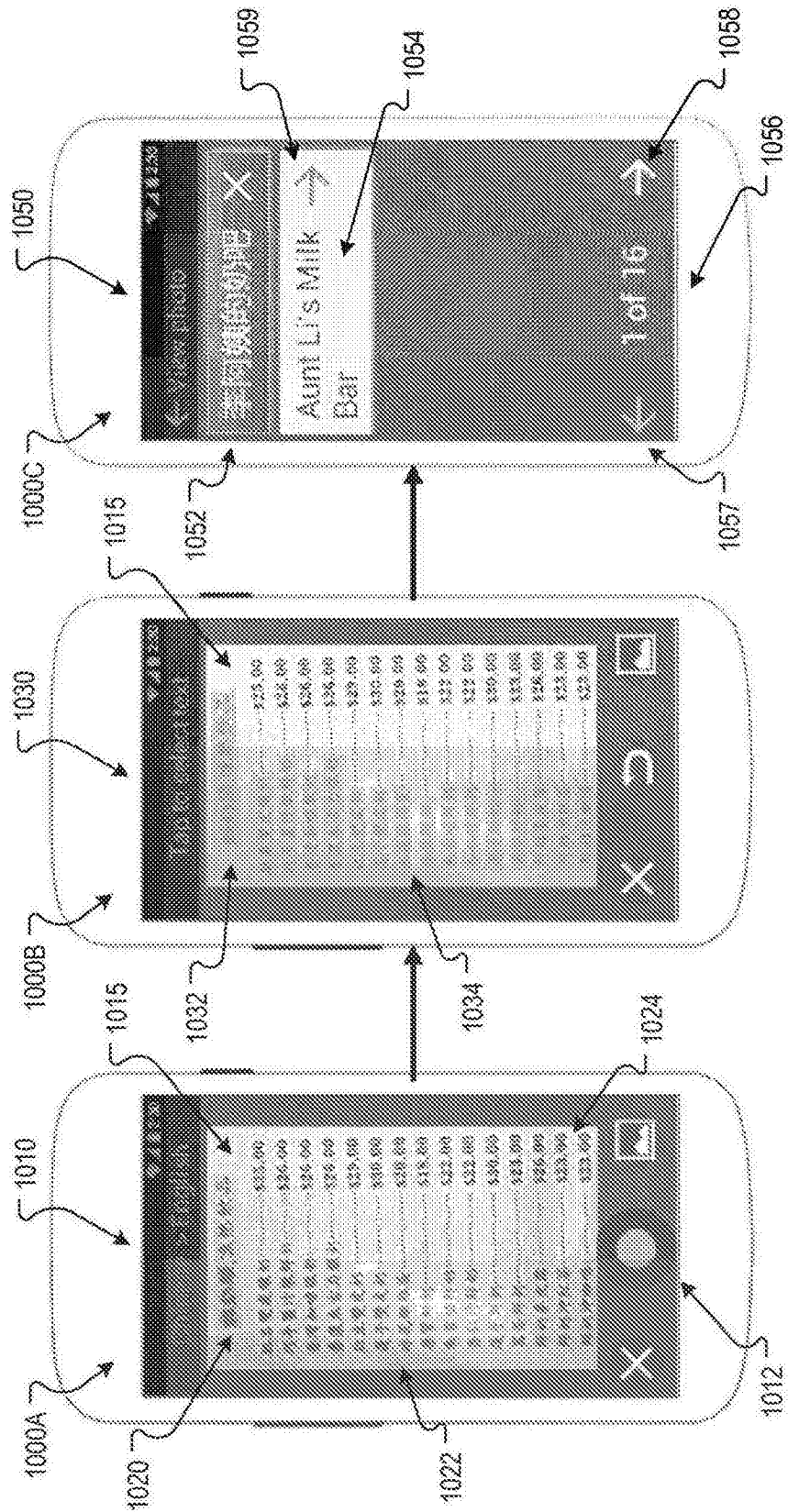


图10

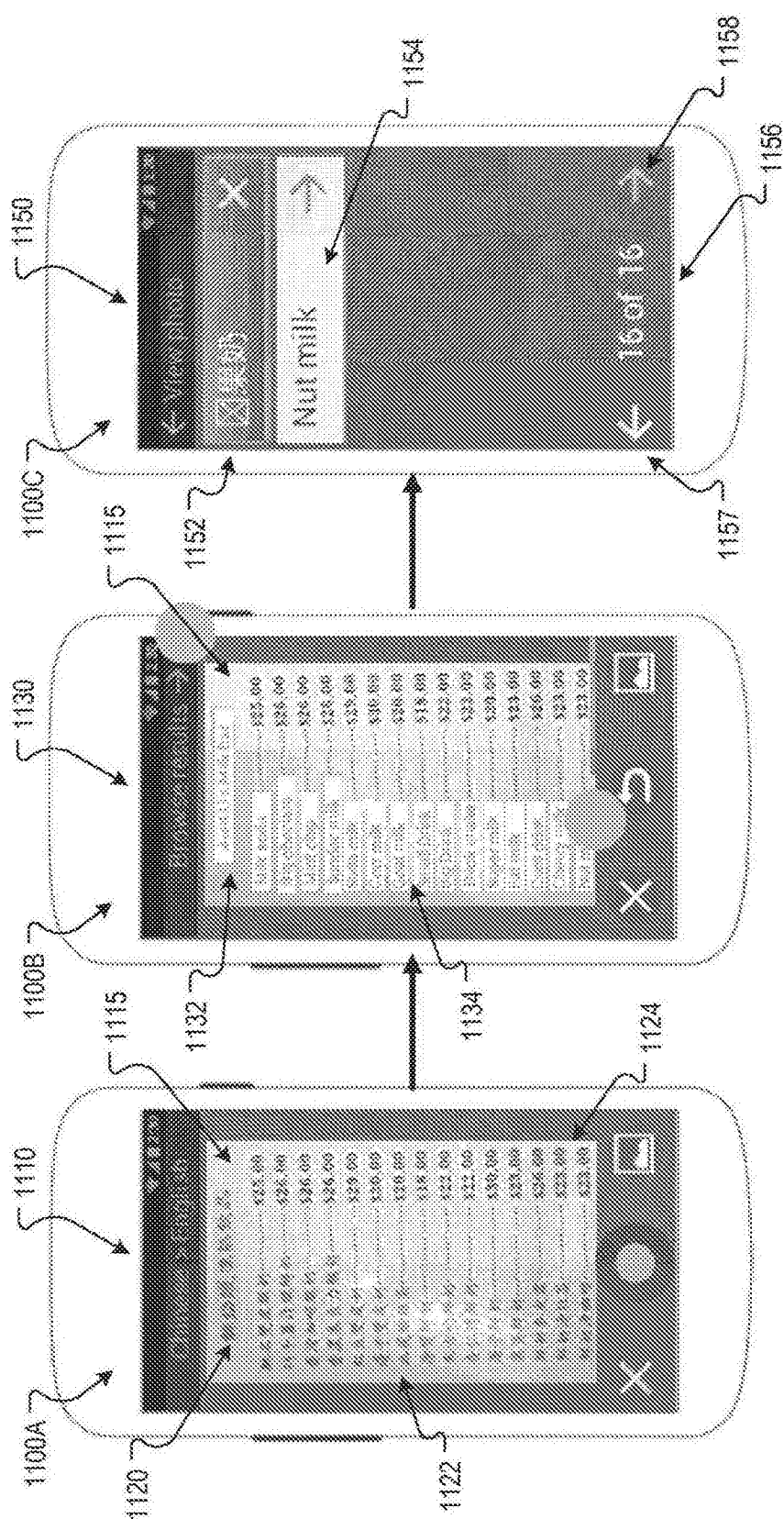


图 11

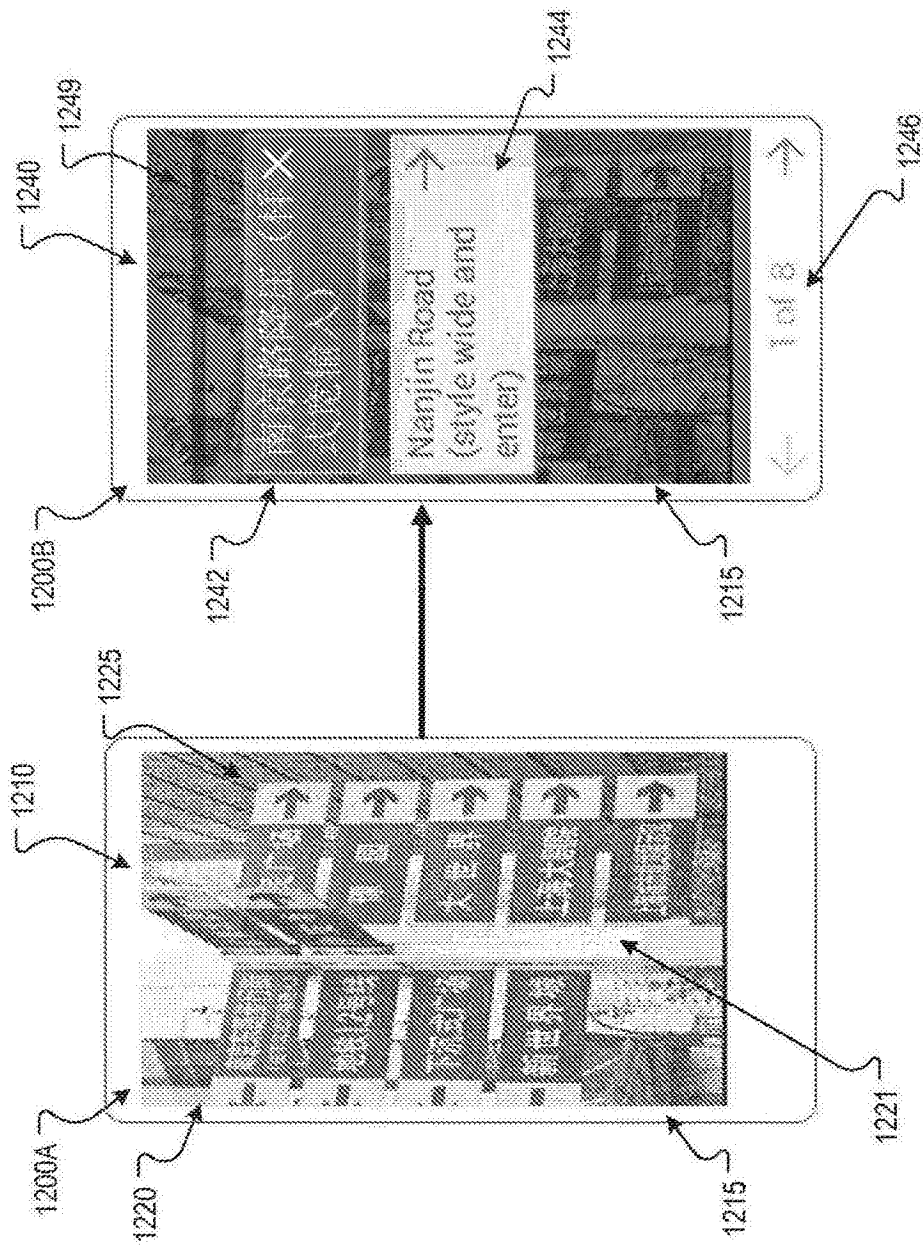


图12

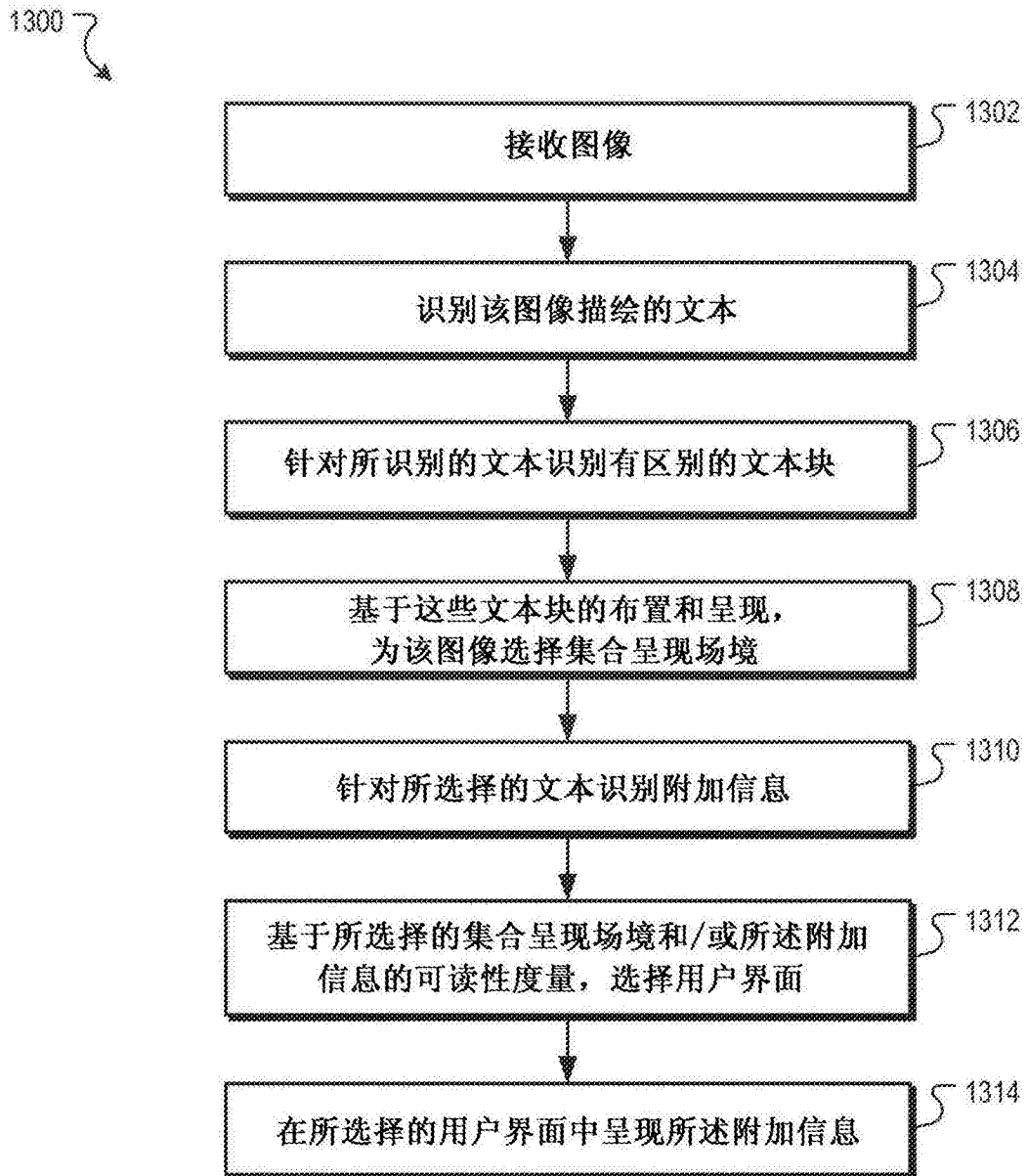


图13

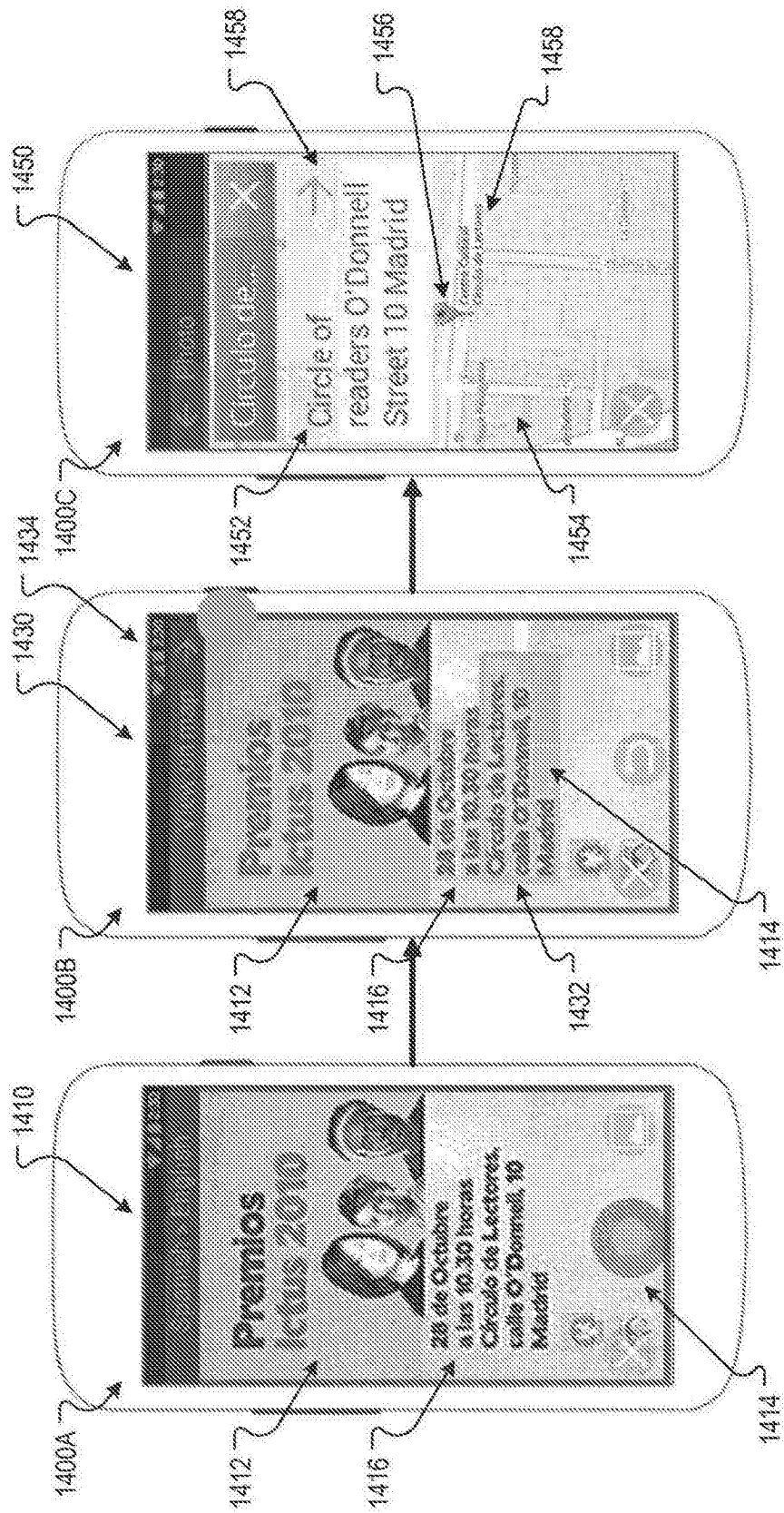


图14

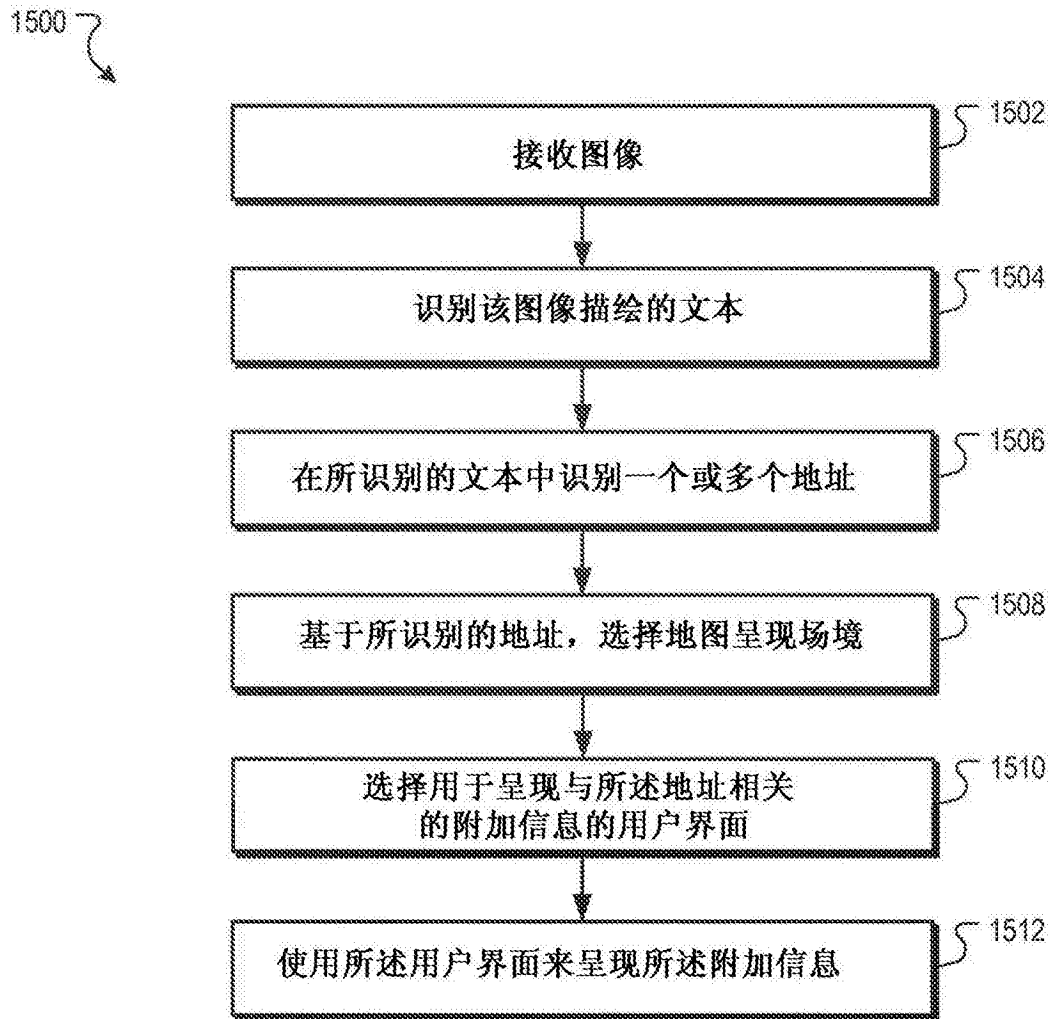


图15

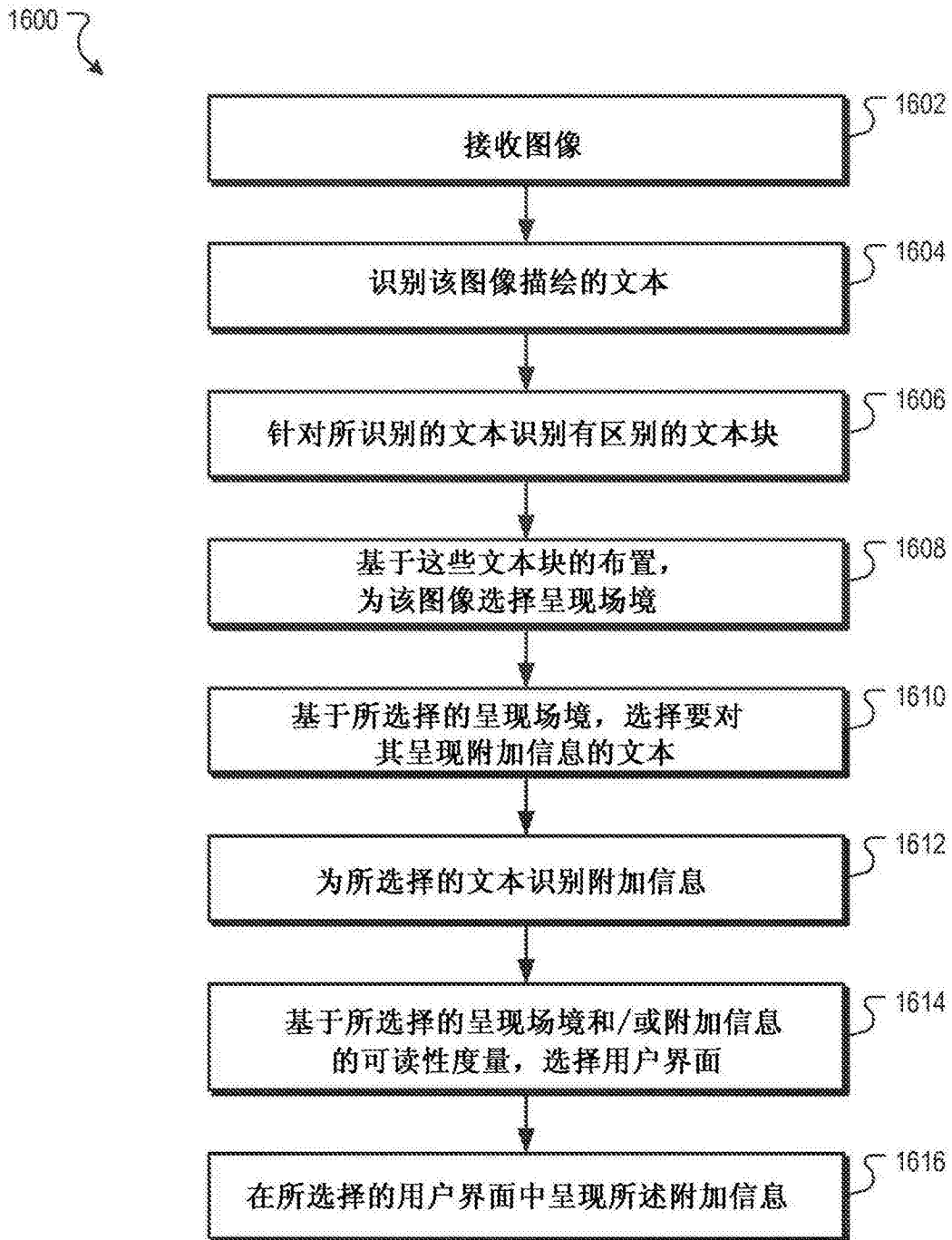


图16