



(12) 发明专利申请

(10) 申请公布号 CN 102667764 A

(43) 申请公布日 2012. 09. 12

(21) 申请号 201080045197. 0

(22) 申请日 2010. 08. 05

(30) 优先权数据

61/232, 397 2009. 08. 07 US

61/266, 122 2009. 12. 02 US

12/850, 513 2010. 08. 04 US

(85) PCT申请进入国家阶段日

2012. 04. 06

(86) PCT申请的申请数据

PCT/US2010/044604 2010. 08. 05

(87) PCT申请的公布数据

W02011/017558 EN 2011. 02. 10

(71) 申请人 谷歌公司

地址 美国加利福尼亚州

(72) 发明人 戴维·彼得鲁 西奥多·鲍尔

(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 17/30 (2006. 01)

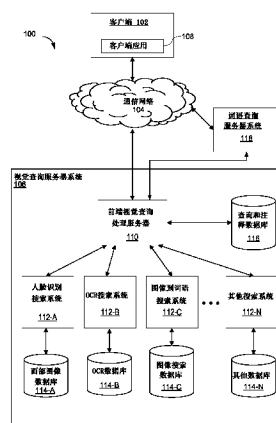
权利要求书 3 页 说明书 21 页 附图 15 页

(54) 发明名称

用于为视觉查询的多个区域展示搜索结果的
用户接口

(57) 摘要

从客户端系统向视觉查询搜索系统提交视觉查询,诸如照片、截屏、扫描的图像、视频帧。搜索系统通过将视觉查询发送给多个并行搜索系统来处理,多个并行搜索系统每一个实现不同的视觉查询搜索过程。从并行搜索系统接收多个结果。利用搜索结果,创建交互式结果文档,并将其发送给客户端系统。交互式结果文档具有用于视觉查询的子部分的至少一个视觉标识符,该至少一个视觉标识符带有到关于该子部分的至少一个搜索结果的可选择的链接。视觉标识符可以是围绕相应子部分的边界框,或在相应子部分之上的半透明标签。可选地,边界框或标签按照结果类型被颜色编码。



1. 一种处理视觉查询的计算机实现的方法,包括:

在具有一个或多个处理器和存储一个或多个程序以供所述一个或多个处理器执行的存储器的服务器系统处:

从客户端系统接收视觉查询;

通过将所述视觉查询发送给多个并行搜索系统以供同时处理来对所述视觉查询进行处理,其中所述多个搜索系统中的搜索系统实现多个视觉查询搜索过程中的相应视觉查询搜索过程;

从所述多个并行搜索系统中的一个或多个接收多个搜索结果;

创建交互式结果文档,所述交互式结果文档包括所述视觉查询的相应子部分的一个或多个视觉标识符,以及关于每一个视觉标识符的到所述搜索结果中的至少一个的至少一个用户可选择的链接;以及

将所述交互式结果文档发送给所述客户端系统。

2. 根据权利要求1所述的计算机实现的方法,其中至少一个搜索结果包括与所述视觉查询的对应子部分有关的数据。

3. 根据权利要求1-2中的任何一个所述的计算机实现的方法,进一步包括:将相应子部分的文本发送给基于文本的查询处理系统。

4. 根据权利要求1-3中的任何一个所述的计算机实现的方法,其中,当与相应视觉标识符相对应的所述视觉查询的所述子部分包含包括一个或多个词语的文本时,与所述相应视觉标识符相对应的搜索结果包括来自对所述文本中的所述词语的至少一个的词语查询搜索的结果。

5. 根据权利要求1-3中的任何一个所述的计算机实现的方法,其中,当与相应视觉标识符相对应的所述视觉查询的所述子部分包含人的脸时,与所述相应视觉标识符相对应的搜索结果包括以下的一个或多个:名字、称呼、联系信息、帐户信息、地址信息、与其脸包含在可选择的子部分中的所述人相关联的相关移动设备的当前位置、其脸包含在可选择的子部分中的所述人的其他图像以及所述人的脸的潜在图像匹配。

6. 根据权利要求1-3中的任何一个所述的计算机实现的方法,其中,当与相应视觉标识符相对应的所述视觉查询的所述子部分包含产品时,与所述相应视觉标识符相对应的搜索结果包括以下的一个或多个:产品信息、产品评论、发起对所述产品的购买的选项、发起对所述产品的出价的选项、类似产品的列表以及相关产品的列表。

7. 根据权利要求1-6中的任何一个所述的计算机实现的方法,其中对所述一个或多个视觉标识符中的相应视觉标识符进行格式化,以按照在所述视觉查询的所述相应子部分中的所识别的实体的类型、以在视觉上不同的方式展示。

8. 根据权利要求7所述的计算机实现的方法,其中对所述相应视觉标识符进行格式化,来以选自由以下组成的组中的在视觉上不同的方式展示:重叠的颜色、重叠的图案、标签背景颜色、标签背景图案、标签字体颜色以及边界颜色。

9. 根据权利要求1-8中的任何一个所述的计算机实现的方法,其中所述一个或多个视觉标识符中的相应视觉标识符包括由与所述视觉查询的所述相应子部分相关联的至少一个词语组成的标签,其中对所述标签进行格式化以在所述交互式结果文档中在所述相应子部分上或靠近所述相应子部分展示。

10. 根据权利要求 1-9 中的任何一个所述的计算机实现的方法,其中所述发送进一步包括:以搜索结果列表形式发送所述多个搜索结果的子集以与所述交互式结果文档一起展示。

11. 根据权利要求 10 所述的计算机实现的方法,进一步包括:

接收对所述至少一个用户可选择的链接的用户选择;以及

识别所述搜索结果列表中与所述选择的链接相对应的搜索结果。

12. 根据权利要求 1-11 中的任何一个所述的计算机实现的方法,其中所述一个或多个视觉标识符包括围绕所述视觉查询的相应子部分的一个或多个边界框。

13. 根据权利要求 12 所述的计算机实现的方法,其中所述边界框中的每一个勾勒出所述视觉查询的所述相应子部分。

14. 根据权利要求 12-13 中的任何一个所述的计算机实现的方法,其中至少一个边界框包括一个或多个更小的边界框。

15. 根据权利要求 12-14 中的任何一个所述的计算机实现的方法,其中所述边界框中的每一个包括到一个或多个搜索结果的用户可选择的链接,其中所述用户可选择的链接具有与所述边界框所环绕的所述视觉查询的所述子部分相对应的激活区域。

16. 根据权利要求 1-11 中的任何一个所述的计算机实现的方法,其中到所述搜索结果中的一个或多个的相应用户可选择的链接具有激活区域,所述激活区域对应于与对应的视觉标识符相关联的所述视觉查询的所述子部分。

17. 一种用于处理视觉查询的服务器系统,包括:

用于执行程序的一个或多个中央处理单元;

存储待由所述一个或多个中央处理单元执行的一个或多个程序的存储器;

所述一个或多个程序包括用于下述的指令:

从客户端系统接收视觉查询;

通过将所述视觉查询发送给多个并行搜索系统以供同时处理来对所述视觉查询进行处理,其中所述多个搜索系统中的搜索系统实现多个视觉查询搜索过程中的相应视觉查询搜索过程;

从所述多个并行搜索系统中的一个或多个接收多个搜索结果;

创建交互式结果文档,所述交互式结果文档包括所述视觉查询的相应子部分的一个或多个视觉标识符,以及关于每一个视觉标识符的到所述搜索结果中的至少一个的至少一个用户可选择的链接;以及

将所述交互式结果文档发送给所述客户端系统。

18. 根据权利要求 17 所述的系统,其中所述一个或多个视觉标识符包括围绕所述视觉查询的相应子部分的一个或多个边界框。

19. 根据权利要求 18 所述的系统,其中所述边界框中的每一个勾勒出所述视觉查询的所述相应子部分。

20. 根据权利要求 18-19 中的任何一个所述的系统,其中所述边界框中的每一个包括到一个或多个搜索结果的用户可选择的链接,其中所述用户可选择的链接具有与所述边界框所环绕的所述视觉查询的所述子部分相对应的激活区域。

21. 根据权利要求 17-20 中的任何一个所述的系统,其中所述一个或多个视觉标识符

中的相应视觉标识符包括由与所述视觉查询的所述相应子部分相关联的至少一个词语组成的标签,其中对所述标签进行格式化以在所述交互式结果文档中在所述相应子部分上或靠近所述相应子部分展示。

22. 一种非暂时性计算机可读存储介质,其存储被配置成由计算机执行的一个或多个程序,所述一个或多个程序包括用于下述的指令:

从客户端系统接收视觉查询;

通过将所述视觉查询发送给多个并行搜索系统以供同时处理来对所述视觉查询进行处理,其中所述多个搜索系统中的搜索系统实现多个视觉查询搜索过程中的相应视觉查询搜索过程;

从所述多个并行搜索系统中的一个或多个接收多个搜索结果;

创建交互式结果文档,所述交互式结果文档包括所述视觉查询的相应子部分的一个或多个视觉标识符,以及关于每一个视觉标识符的到所述搜索结果中的至少一个的至少一个用户可选择的链接;以及

将所述交互式结果文档发送给所述客户端系统。

23. 根据权利要求 22 所述的计算机可读存储介质,其中所述一个或多个视觉标识符包括围绕所述视觉查询的相应子部分的一个或多个边界框。

24. 根据权利要求 23 所述的计算机可读存储介质,其中所述边界框中的每一个勾勒出所述视觉查询的所述相应子部分。

25. 根据权利要求 23-24 中的任何一个所述的计算机可读存储介质,其中所述边界框中的每一个包括到一个或多个搜索结果的用户可选择的链接,其中所述用户可选择的链接具有与所述边界框所环绕的所述视觉查询的所述子部分相对应的激活区域。

26. 根据权利要求 22-25 中的任何一个所述的计算机可读存储介质,其中所述一个或多个视觉标识符中的相应视觉标识符包括由与所述视觉查询的所述相应子部分中的图像相关联的至少一个词语组成的标签,其中对所述标签进行格式化以在所述交互式结果文档中在所述相应子部分上或靠近所述相应子部分展示。

27. 一种非暂时性计算机可读存储介质,其存储被配置成由计算机系统的一个或多个处理器执行的一个或多个程序,所述一个或多个程序包括待由所述一个或多个处理器执行以执行权利要求 1-16 中的任何一个所述的方法的指令。

28. 一种服务器系统,包括:

一个或多个处理器;以及

存储一个或多个程序以供所述一个或多个处理器执行的存储器,所述一个或多个程序包括待由所述一个或多个处理器执行以执行权利要求 1-16 中的任何一个所述的方法的指令。

用于为视觉查询的多个区域展示搜索结果的用户接口

技术领域

[0001] 所公开的实施例一般地涉及通过用于处理视觉查询的多个并行搜索系统展示搜索结果。

背景技术

[0002] 基于文本或基于词语的搜索,其中用户将词或短语输入到搜索引擎中并且接收各种结果,是用于进行搜索的有用工具。然而,基于词语的查询要求用户能够输入相关词语。有时,用户可能希望知道关于图像的信息。例如,用户可能想要知道照片中的人的名字,或用户可能想要知道图片中的花或鸟的名称。因此,能够接收视觉查询并且提供搜索结果的系统将是期望的。

发明内容

[0003] 根据一些实施例,一种处理视觉查询的计算机实现的方法包括:在服务器系统上执行以下步骤,所述服务器系统具有一个或多个处理器和存储一个或多个程序以供该一个或多个处理器执行的存储器。接收来自客户端系统的视觉查询。通过将该视觉查询发送给多个并行搜索系统以供同时处理来对该视觉查询进行处理。多个搜索系统中的每一个实现多个视觉查询搜索过程中的不同视觉查询搜索过程。然后,服务器系统从多个并行搜索系统中的一个或多个接收多个搜索结果。其创建包括视觉查询的相应子部分的一个或多个视觉标识符的交互式结果文档。每一个视觉标识符具有到搜索结果中的至少一个的至少一个用户可选择的链接。最后,服务器系统将该交互式结果文档发送给客户端系统。在一些实施例中,搜索结果包括与视觉查询的对应子部分有关的数据。

[0004] 在一些实施例中,发送进一步包括:以搜索结果列表形式发送多个搜索结果的子集以与交互式结果文档一起展示。可选地,该方法进一步包括:接收对至少一个用户可选择的链接的用户选择;以及识别搜索结果列表中与所述选择的链接相对应的搜索结果。

[0005] 在一些实施例中,视觉标识符包括围绕视觉查询的相应子部分的一个或多个边界框。边界框可以是正方形或可以勾勒出视觉查询的相应子部分。可选地,一些边界框在其之内包括更小的边界框。

[0006] 在一些实施例中,边界框中的每一个包括到一个或多个搜索结果的用户可选择的链接,以及用户可选择的链接具有与边界框所环绕的视觉查询的子部分相对应的激活区域。即使在视觉标识符不是边界框的实施例中,相应用户可选择的链接也具有对应于与对应的视觉标识符相关联的视觉查询的子部分的激活区域。

[0007] 在一些实施例中,当可选择的子部分包含文本时,该方法进一步包括:将可选择的子部分的文本发送给基于文本的查询处理系统。在其他实施例中,当视觉查询的子部分对应于包含文本的相应视觉标识符时,与该相应视觉标识符相对应的搜索结果包括来自对文本中的词语的至少一个的词语查询搜索的结果。

[0008] 在一些实施例中,当与相应视觉标识符相对应的视觉查询的子部分包含人的脸

时,与该相应视觉标识符相对应的搜索结果包括名字、称呼、联系信息、帐户信息、地址信息、与其脸包含在可选择的子部分中的人相关联的相关移动设备的当前位置、其脸包含在可选择的子部分中的人的其他图像和 / 或该人的脸的潜在图像匹配。

[0009] 在一些实施例中,当与相应视觉标识符相对应的视觉查询的子部分包含产品时,与该相应视觉标识符相对应的搜索结果包括产品信息、产品评论、发起对产品的购买的选项、发起对产品的出价的选项、类似产品的列表和 / 或相关产品的列表。

[0010] 在一些实施例中,对一个或多个视觉标识符中的相应视觉标识符进行格式化,以按照在视觉查询的相应子部分中的所识别的实体的类型、以在视觉上不同的方式展示。可以对相应视觉标识符进行格式化,来以在视觉上不同的方式展示,诸如重叠的颜色、重叠的图案、标签背景颜色、标签背景图案、标签字体颜色以及边界颜色。

[0011] 在一些实施例中,一个或多个视觉标识符中的相应视觉标识符包括由与视觉查询的相应子部分中的图像相关联的至少一个词语组成的标签。对该标签进行格式化以在交互式结果文档中在该相应子部分上或靠近该相应子部分展示。

附图说明

[0012] 图 1 是图示包括视觉查询服务器系统的计算机网络的框图。

[0013] 图 2 是图示与一些实施例一致的用于对视觉查询作出响应的过程的流程图。

[0014] 图 3 是图示与一些实施例一致的用于用交互式结果文档对视觉查询作出响应的过程的流程图。

[0015] 图 4 是图示与一些实施例一致的在客户端和视觉查询服务器系统之间的通信的流程图。

[0016] 图 5 是图示与一些实施例一致的客户端系统的框图。

[0017] 图 6 是图示与一些实施例一致的前端视觉查询处理服务器系统的框图。

[0018] 图 7 是图示与一些实施例一致的利用来处理视觉查询的并行搜索系统中的通用搜索系统的框图。

[0019] 图 8 是图示与一些实施例一致的利用来处理视觉查询的 OCR 搜索系统的框图。

[0020] 图 9 是图示与一些实施例一致的利用来处理视觉查询的面部识别搜索系统的框图。

[0021] 图 10 是图示与一些实施例一致的利用来处理视觉查询的图像到词语搜索系统的框图。

[0022] 图 11 图示与一些实施例一致的具有示例性视觉查询的截屏的客户端系统。

[0023] 图 12A 和 12B 每一个图示与一些实施例一致的具有带有边界框的交互式结果文档的截屏的客户端系统。

[0024] 图 13 图示与一些实施例一致的具有按类型编码的交互式结果文档的截屏的客户端系统。

[0025] 图 14 图示与一些实施例一致的具有带有标签的交互式结果文档的截屏的客户端系统。

[0026] 图 15 图示与一些实施例一致的交互式结果文档和视觉查询与结果列表同时显示的截屏。

[0027] 整个附图,相同的参考数字指代对应的部分。

具体实施方式

[0028] 现将详细参考实施例,在附图中图示了所述实施例的示例。在下面的详细描述中,阐述了许多具体细节以提供对本发明的全面了解。然而,对本领域普通技术人员将是显而易见的是,在没有这些具体细节的情况下,可以实践本发明。在其他情况下,没有详细描述众所周知的方法、程序、组件、电路和网络,以免不必要地使实施例的方面模糊不清。

[0029] 还将理解的是,尽管词语第一、第二等在此可以用于描述各种元素,这些元素不应由这些词语限制。这些词语仅仅用于区分元素。例如,在不背离本发明的范围的情况下,第一联系人可以被称为第二联系人,以及类似地,第二联系人可以被称为第一联系人。第一联系人和第二联系人都是联系人,但是其不是同一联系人。

[0030] 在此对本发明的描述中所使用的术语仅用于描述特定实施例的目的,并且并不意在对本发明的限制。除非语境另外明确指示,如在本发明的描述和所附权利要求中所使用的,单数形式“一”、“一个”和“所述”还意在包括复数形式。还将理解的是,如在此所使用的词语“和 / 或”是指并且涵盖相关联的列出的项中的一个或多个的任何和所有可能组合。将进一步理解的是,词语“包括”当在本说明书中使用指示陈述的特征、完整物、步骤、操作、元素和 / 或组件的存在,但是并不排除一个或多个其他特征、完整物、步骤、操作、元素、组件和 / 或其组群的存在或添加。

[0031] 取决于语境,如在此所使用的,词语“如果”可以被解释成意为“在……时”或“当……时”或“响应于确定”或“响应于检测”。类似地,取决于语境,短语“如果确定”或“如果检测”可以被解释成意为“当确定时”或“响应于确定”或“当检测(陈述的条件或事件)时”或“响应于检测(陈述的条件或事件)”。

[0032] 图 1 是图示根据一些实施例的包括视觉查询服务器系统的计算机网络的框图。计算机网络 100 包括一个或多个客户端系统 102 和视觉查询服务器系统 106。一个或多个通信网络 104 使这些组件互连。通信网络 104 可以是多种网络中的任何网络,包括局域网(LAN)、广域网(WAN)、无线网络、有线网络、因特网或这样的网络的组合。

[0033] 客户端系统 102 包括用于接收视觉查询(例如,图 11 的视觉查询 1102)的客户端应用 108,其由客户端系统执行。视觉查询是作为查询提交给搜索引擎或搜索系统的图像。视觉查询的示例不受限制包括照片、扫描的文档和图像以及图画。在一些实施例中,客户端应用 108 选自由搜索应用、用于浏览器应用的搜索引擎插件和用于浏览器应用的搜索引擎扩展组成的集合。在一些实施例中,客户端应用 108 是“广泛的”搜索框,其允许用户将任何格式的图像拖放入该搜索框以被用作视觉查询。

[0034] 客户端系统 102 将查询发送给视觉查询服务器系统 106 并且从视觉查询服务器系统 106 接收数据。客户端系统 102 可以是能够与视觉查询服务器系统 106 进行通信的任何计算机或其他设备。示例不受限制包括台式和笔记本式计算机、大型计算机、服务器计算机、诸如移动电话和个人数字助理的移动设备、网络终端以及机顶盒。

[0035] 视觉查询服务器系统 106 包括前端视觉查询处理服务器 110。前端服务器 110 从客户端 102 接收视觉查询,并且将该视觉查询发送给多个并行搜索系统 112 以供同时处理。搜索系统 112 每一个实现不同的视觉查询搜索过程,并且必要时访问其对应的数据库 114

以通过其不同的搜索过程对视觉查询进行处理。例如,人脸识别搜索系统 112-A 将访问面部图像数据库 114-A 以查找与图像查询的面部匹配。如将参考图 9 更详细说明的,如果视觉查询包含人脸,则面部识别搜索系统 112-A 将返回来自面部图像数据库 114-A 的一个或多个搜索结果(例如,名字、匹配的人脸等)。在另一个示例中,光学字符识别(OCR)搜索系统 112-B 将视觉查询中的任何可识别的文本转换成文本以作为一个或多个搜索结果返回。在光学字符识别(OCR)搜索系统 112-B 中,如将参考图 8 更详细说明的,可以访问 OCR 数据库 114-B 以识别特定字体或文本图案。

[0036] 可以使用任何数量的并行搜索系统 112。一些示例包括面部识别搜索系统 112-A、OCR 搜索系统 112-B、图像到词语搜索系统 112-C(其可以识别对象或对象类别)、产品识别搜索系统(其可以被配置成识别诸如书籍封面和 CD 的二维图像,以及还可以被配置成识别诸如家具的三维图像)、条形码识别搜索系统(其识别一维和二维式样条形码)、命名实体识别搜索系统、地标识别(其可以被配置成识别如同埃菲尔铁塔的特定著名地标,以及还可以被配置成识别诸如广告牌的特定图像的语料库)、由客户端系统 102 中的 GPS 接收器或移动电话网络提供的地理位置信息辅助的地点识别、颜色识别搜索系统以及类似的图像搜索系统(其搜索并标识与视觉查询类似的图像)。更多的搜索系统可以被添加为在图 1 中由系统 112-N 所表示的、另外的并行搜索系统。除 OCR 搜索系统外,所有搜索系统在此被共同定义为执行图像匹配过程的搜索系统。包括 OCR 搜索系统的所有搜索系统被统称为按图像查询搜索系统。在一些实施例中,视觉查询服务器系统 106 包括面部识别搜索系统 112-A、OCR 搜索系统 112-B 和至少一个其他的按图像查询搜索系统 112。

[0037] 并行搜索系统 112 每一个分别对视觉搜索查询进行处理,并且将其结果返回给前端服务器系统 110。在一些实施例中,前端服务器 100 可以对搜索结果执行一个或多个分析,诸如以下的一个或多个:将结果聚合成复合文档、选择结果的子集来显示以及对结果进行排名,如将参考图 6 更详细说明的。前端服务器 110 将搜索结果通信给客户端系统 102。

[0038] 客户端系统 102 向用户展示一个或多个搜索结果。结果可以在显示器上、通过音频扬声器或用于向用户通信信息的任何其他装置来展示。用户可以以多种方式进行交互。在一些实施例中,用户的选择、注释和与搜索结果的其他交互被传送给视觉查询服务器系统 106,并且与视觉查询一起被记录在查询和注释数据库 116 中。查询和注释数据库中的信息可以用于改进视觉查询结果。在一些实施例中,将来自查询和注释数据库 116 的信息周期性推送到并行搜索系统 112,其将信息的任何相关部分合并入其各自的单独数据库 114。

[0039] 计算机网络 100 可选地包括用于响应于词语查询而执行搜索的词语查询服务器系统 118。相对于包含图像的视觉查询,词语查询是包含一个或多个词语的查询。词语查询服务器系统 118 可以用于生成对视觉查询服务器系统 106 中的各自搜索引擎所产生的信息进行补充的搜索结果。从词语查询服务器系统 118 返回的结果可以包括任何格式。词语查询服务器系统 118 可以包括文本文档、图像、视频等。虽然词语查询服务器系统 118 在图 1 中被示出为独立系统,可选地,视觉查询服务器系统 106 可以包括词语查询服务器系统 118。

[0040] 在下面参考图 2-4 中的流程图提供了关于视觉查询服务器系统 106 的操作的另外信息。

[0041] 图 2 是图示根据本发明的某些实施例的用于对视觉查询作出响应的视觉查询服务器系统方法的流程图。在图 2 中所示的操作中的每一个可以对应于存储在计算机存储器或计算机可读存储介质中的指令。

[0042] 视觉查询服务器系统从客户端系统接收视觉查询 (202)。客户端系统例如可以是台式计算设备、移动设备或另一个类似设备 (204)，如参考图 1 所说明的。在图 11 中示出了在示例客户端系统上的示例视觉查询。

[0043] 视觉查询是任何适当格式的图像文档。例如，视觉查询可以是照片、截屏、扫描的图像或帧或多个视频帧的序列 (206)。在一些实施例中，视觉查询是内容创作程序 (图 5 的 736) 所产生的图画。如此，在一些实施例中，用户“绘制”视觉查询，而在其他实施例中，用户扫描或拍摄视觉查询。一些视觉查询使用诸如 Acrobat 的图像生成应用、照片编辑程序、绘图程序或图像编辑程序来创建。例如，视觉查询可以来自：用户在其移动电话上拍摄其朋友的照片，然后将该照片作为视觉查询提交给服务器系统。视觉查询还可以来自：用户对杂志页面进行扫描，或获得在台式计算机上的网页的截屏，然后将扫描或截屏作为视觉查询提交给服务器系统。在一些实施例中，视觉查询通过浏览器应用的搜索引擎扩展、通过用于浏览器应用的插件或通过客户端系统 102 执行的搜索应用被提交给服务器系统 106。视觉查询还可以由 (客户端系统执行的) 支持或生成可以由客户端系统传送给位于远程的服务器的图像的其他应用程序提交。

[0044] 视觉查询可以是文本和非文本元素的组合 (208)。例如，查询可以是包含图像和文本，诸如一个人站在路标旁边，的杂志页面的扫描。视觉查询可以包括无论是由嵌入在客户端系统中的照相机拍摄的还是通过客户端系统扫描或接收的文档获得的、人的脸的图像。视觉查询还可以是仅包含文本的文档的扫描。视觉查询还可以是多个不同主题的图像，诸如森林中的数只鸟、人和物体 (例如，汽车、公园长凳等)、人和动物 (例如，宠物、农场动物、蝴蝶等)。视觉查询可以具有两个或多个不同的元素。例如，视觉查询可以包括在产品包装上的条形码和产品图像或产品名称。例如，视觉查询可以是包括书籍标题、封面艺术和条形码的书籍封面的图片。如在下面更详细论述的，在一些情况下，一个视觉查询将产生与该视觉查询的不同部分相对应的两个或多个不同的搜索结果。

[0045] 服务器系统如下对视觉查询进行处理。前端服务器系统将视觉查询发送给多个并行搜索系统以供同时处理 (210)。每一个搜索系统实现不同的视觉查询搜索过程，即独立搜索系统通过其自身处理方案对视觉查询进行处理。

[0046] 在一些实施例中，视觉查询被发送给其以供处理的搜索系统中的一个光学字符识别 (OCR) 搜索系统。在一些实施例中，视觉查询被发送给其以供处理的搜索系统中的一个面部识别搜索系统。在一些实施例中，运行不同的视觉查询搜索过程的多个搜索系统至少包括：光学字符识别 (OCR)、面部识别以及不同于 OCR 和面部识别的另一个按图像查询过程 (212)。另一个按图像查询过程选自包括但不限于以下的过程集：产品识别、条形码识别、对象或对象类别识别、命名实体识别以及颜色识别 (212)。

[0047] 在一些实施例中，命名实体识别作为 OCR 搜索系统的后期过程发生，其中对 OCR 的文本结果分析著名的人、场所、对象等，然后在词语查询服务器系统 (图 1 的 118) 中搜索被标识为是命名实体的词语。在其他实施例中，著名的地标、标志、人、专辑封面、商标等的图像由图像到词语搜索系统识别。在其他实施例中，利用与图像到词语搜索系统分离的不同

命名实体按图像查询过程。对象或对象类别识别系统识别如同“汽车”的通用结果类型。在一些实施例中，该系统还识别产品品牌、特定产品型号等，以及提供更具体的描述，如同“保时捷”。搜索系统中的部分可以是特定于特殊用户的搜索系统。例如，颜色识别和面部识别的特定版本可以由盲人使用的特殊搜索系统。

[0048] 前端服务器系统从并行搜索系统接收结果 (214)。在一些实施例中，结果附有搜索分值。对于一些视觉查询，搜索系统中的部分不会找到相关结果。例如，如果视觉查询是花的图片，则面部识别搜索系统和条形码搜索系统不会找到任何相关结果。在一些实施例中，如果没有相关结果被找到，则从该搜索系统接收空或零搜索分值 (216)。在一些实施例中，如果前端服务器在预定义时段（例如，0.2、0.5、1、2 或 5 秒）之后没有从搜索系统接收到结果，则其将仿佛该超时服务器产生了空搜索分值般对所接收到的结果进行处理，以及将对从其他搜索系统接收到的结果进行处理。

[0049] 可选地，当所接收到的搜索结果中的至少两个满足预定义标准时，对其进行排名 (218)。在一些实施例中，预定义标准中的一个排除无效结果。预定义标准是结果不是无效的。在一些实施例中，预定义标准中的一个排除具有落在预定义最小分值之下的（例如，关于相关性因素的）数字分值的的结果。可选地，对多个搜索结果进行过滤 (220)。在一些实施例中，仅在结果的总数超过预定义阈值的情况下，对结果进行过滤。在一些实施例中，除落在预定义最小分值之下的结果被排除之外，对所有结果进行排名。对于一些视觉查询，对结果的内容进行过滤。例如，如果结果中的部分包含私人信息或个人保护的信息，则滤除这些结果。

[0050] 可选地，视觉查询服务器系统创建复合搜索结果 (222)。其一个实施例是：如参考图 3 说明的，当多于一个搜索系统结果被嵌入在交互式结果文档中时。词语查询服务器系统（图 1 的 118）可以用来自词语搜索的结果来扩充来自并行搜索系统中的一个的结果，其中另外结果被链接到文档或信息源、或包含可能与视觉查询相关的另外信息的文本和/或图像。因此，例如，复合搜索结果可以包含 OCR 结果和到 OCR 文档中的命名实体的链接 (224)。

[0051] 在一些实施例中，OCR 搜索系统（图 1 的 112-B）或前端视觉查询处理服务器（图 1 的 110）识别文本中的可能相关的词。例如，其可以识别诸如名人或地点的命名实体。将命名实体作为查询词语提交给词语查询服务器系统（图 1 的 118）。在一些实施例中，将词语查询服务器系统所产生的词语查询结果作为“链接”嵌入在视觉查询结果中。在一些实施例中，将词语查询结果作为独立链接返回。例如，如果书籍封面的图片是视觉查询，则可能的是，对象识别搜索系统将为该书籍产生较高评分的命中。如此，将在词语查询服务器系统 118 上运行关于该书籍的标题的词语查询，以及将词语查询结果与视觉查询结果一起返回。在一些实施例中，在加标签组中展示词语查询结果以将其与视觉查询结果进行区分。可以分别搜索结果，或可以使用搜索查询中的所有识别的命名实体来执行搜索以产生特别相关的另外搜索结果。例如，如果视觉查询是扫描的关于巴黎的旅游小册子，则返回的结果可以包括到词语查询服务器系统 118 的用于发起对词语查询“巴黎圣母院”的搜索的链接。类似地，复合搜索结果包括来自关于所识别的著名图像的文本搜索的结果。例如，在同一旅游小册子情况下，还可以示出到关于在小册子中被示出为图片的著名目的地，如同“埃菲尔铁塔”和“卢浮宫”，的词语查询结果的活链接（即使词语“埃菲尔铁塔”和“卢浮宫”在小册子

本身中没有出现)。

[0052] 视觉查询服务器系统然后将至少一个结果发送给客户端系统 (226)。典型地,如果视觉查询处理服务器从多个搜索系统中的至少部分接收多个搜索结果,则其然后将多个搜索结果中的至少一个发送给客户端系统。对于一些视觉查询,仅仅一个搜索系统会返回相关结果。例如,在仅包含文本的图像的视觉查询中,仅仅 OCR 服务器的结果可能是相关的。对于一些视觉查询,来自一个搜索系统的仅仅一个结果可能是相关的。例如,仅仅与扫描的条形码有关的产品可能是相关的。在这些情况下,前端视觉处理服务器将仅仅返回相关搜索结果。对于一些视觉查询,将多个搜索结果发送给客户端系统,以及该多个搜索结果包括来自并行搜索系统中的多于一个的搜索结果 (228)。这可以在多于一个不同图像在视觉查询中时发生。例如,如果视觉查询是人骑马的图片,则可以与关于该马的对象识别结果一起显示关于该人的面部识别的结果。在一些实施例中,使经由图像搜索系统的关于特定查询的所有结果聚组并且一起展示。例如,在标题“面部识别结果”下显示最高 N 个面部识别结果,以及在标题“对象识别结果”下一起显示最高 N 个对象识别结果。替选地,如下所述,可以使来自特定图像搜索系统的搜索结果按图像区域聚组。例如,如果视觉查询包括两个人脸,则其两者均产生面部识别结果,将关于每一个人脸的结果展示为不同组。对于一些视觉查询(例如,包括文本和一个或多个对象两者的图像的视觉查询),搜索结果可以包括 OCR 结果和一个或多个图像匹配结果两者 (230)。

[0053] 在一些实施例中,用户可能希望了解更多关于特定搜索结果的信息。例如,如果视觉查询是海豚的图片,并且“图像到词语”搜索系统返回以下词语“水”、“海豚”、“蓝色”和“鳍状肢”,则用户可能希望运行对“鳍状肢”的基于文本的查询词语搜索。当用户希望运行对词语查询的搜索(例如,如通过用户点击或选择搜索结果中的对应链接所指示的)时,对查询词语服务器系统(图 1 的 118)进行访问,以及运行对所选择的词语的搜索。在客户端系统上独立或结合视觉查询结果显示对应的搜索词语结果 (232)。在一些实施例中,前端视觉查询处理服务器(图 1 的 110)自动(即除初始视觉查询外,没有接收任何用户命令)为视觉查询选择一个或多个最高潜在的文本结果,在词语查询服务器系统 118 上运行那些文本结果,然后将那些词语查询结果与视觉查询结果一起返回给客户端系统,作为将至少一个搜索结果返回给客户端系统的一部分 (232)。在上面的示例中,如果“鳍状肢”是海豚的视觉查询图片的第一词语结果,则前端服务器针对“鳍状肢”运行词语查询,并且将那些词语查询结果与视觉查询结果一起返回给客户端系统。该实施例,其中认为可能被用户选择的词语结果在将来自视觉查询的搜索结果发送给用户之前被自动执行,节省了用户时间。在一些实施例中,如上所述,将这些结果显示为复合搜索结果 (222)。在其他实施例中,替代复合搜索结果或除复合搜索结果外,所述结果是搜索结果列表的一部分。

[0054] 图 3 是图示用于用交互式结果文档对视觉查询作出响应的过程的流程图。在上面参考图 2 描述了前三个操作 (202、210、214)。从自并行搜索系统接收 (214) 的搜索结果,创建交互式结果文档 (302)。

[0055] 现将详细描述创建交互式结果文档 (302)。对于一些视觉查询,交互式结果文档包括视觉查询的各个子部分的一个或多个视觉标识符。每一个视觉标识符具有到搜索结果中的至少一个的至少一个用户可选择的链接。视觉标识符标识视觉查询的相应子部分。对于一些视觉查询,交互式结果文档仅仅具有带有到一个或多个结果的一个用户可选择的链接

的一个视觉标识符。在一些实施例中，到搜索结果中的一个或多个的相应用户可选择的链接具有激活区域，以及激活区域对应于与对应的视觉标识符相关联的视觉查询的子部分。

[0056] 在一些实施例中，视觉标识符是边界框 (304)。在一些实施例中，如图 12A 中所示，边界框围绕视觉查询的子部分。边界框不必是正方形或长方形框形状，而是可以是任何式样的形状，包括圆形的、椭圆形的、（例如，与视觉查询中的对象、实体或视觉查询的区域）等角的、不规则的或任何其他形状，如图 12B 中所示。对于一些视觉查询，边界框勾勒出视觉查询的子部分中的可标识的实体的边界 (306)。在一些实施例中，每一个边界框包括到一个或多个搜索结果的用户可选择的链接，其中用户可选择的链接具有与边界框所环绕的视觉查询的子部分相对应的激活区域。当在边界框内的空间（用户可选择的链接的激活区域）被用户选择时，返回与勾勒出的子部分中的图像相对应的搜索结果。

[0057] 在一些实施例中，如图 14 中所示，视觉标识符是标签 (307)。在一些实施例中，标签包括与视觉查询的相应子部分中的图像相关联的至少一个词语。对每一个标签进行格式化以在交互式结果文档中在相应子部分上或靠近相应子部分展示。在一些实施例中，标签是彩色编码的。

[0058] 在一些实施例中，对每一个相应的视觉标识符进行格式化，以按照在视觉查询的相应子部分中的所识别的实体的类型、以在视觉上不同的方式展示。例如，如图 13 中所示，围绕产品、人、商标和两个文本区域的边界框每一个用不同的交叉影线图案来展示，表示不同颜色透明边界框。在一些实施例中，对视觉标识符进行格式化，来以在视觉上不同的方式展示，诸如重叠的颜色、重叠的图案、标签背景颜色、标签背景图案、标签字体颜色以及边界颜色。

[0059] 在一些实施例中，交互式结果文档中的用户可选择的链接是到包含与视觉查询的对应子部分有关的一个或多个结果的文档或对象的链接 (308)。在一些实施例中，至少一个搜索结果包括与视觉查询的对应子部分有关的数据。如此，当用户选择与相应子部分相关联的可选择的链接时，该用户被指引到与视觉查询的相应子部分中的所识别的实体相对应的搜索结果。

[0060] 例如，如果视觉查询是条形码的照片，则可能有这样的照片部分，其是条形码附于之上的包装的不相关部分。交互式结果文档可以包括仅围绕条形码的边界框。当用户在勾勒出的条形码边界框内进行选择时，显示条形码搜索结果。条形码搜索结果可以包括一个结果，与该条形码相对应的产品名称，或条形码结果可以包括数个结果，诸如其中能够购买、评论等该产品的多种地点。

[0061] 在一些实施例中，当与相应视觉标识符相对应的视觉查询的子部分包含包括一个或多个词语的文本时，与该相应视觉标识符相对应的搜索结果包括来自对文本中的词语的至少一个的词语查询搜索的结果。在一些实施例中，当与相应视觉标识符相对应的视觉查询的子部分包含人的脸，其中为该脸找到满足预定义可靠性（或其他）标准的至少一个匹配（即搜索结果）时，与该相应视觉标识符相对应的搜索结果包括以下的一个或多个：名字、称呼、联系信息、帐户信息、地址信息、与其脸包含在可选择的子部分中的人相关联的相关移动设备的当前位置、其脸包含在可选择的子部分中的人的其他图像以及该人的脸的潜在图像匹配。在一些实施例中，当与相应视觉标识符相对应的视觉查询的子部分包含产品，其中为该产品找到满足预定义可靠性（或其他）标准的至少一个匹配（即搜索结果）时，

与该相应视觉标识符相对应的搜索结果包括以下的一个或多个：产品信息、产品评论、发起对产品的购买的选项、发起对产品的出价的选项、类似产品列表以及相关产品列表。

[0062] 可选地，交互式结果文档中的相应用户可选择的链接包括锚文本，其在文档中显示，而不必激活链接。锚文本提供与当链接被激活时所获取的信息有关的信息，诸如关键词或词语。可以将锚文本显示为标签 (307) 的一部分、或在边界框 (304) 的一部分中显示、或显示为当用户将光标悬停在用户可选择的链接上达诸如 1 秒的预确定时段时所显示的另外信息。

[0063] 可选地，交互式结果文档中的相应用户可选择的链接是到搜索引擎的链接，其用于搜索与基于文本的查询（在此有时称为词语查询）相对应的信息或文档。该链接的激活促使搜索引擎执行搜索，其中查询和搜索引擎由该链接指定（例如，搜索引擎由该链接中的 URL 指定，以及基于文本的搜索查询由该链接的 URL 参数指定），同时结果被返回给客户端系统。可选地，在该示例中的链接可以包括指定搜索查询中的文本或词语的锚文本。

[0064] 在一些实施例中，响应于视觉查询而产生的交互式结果文档可以包括与来自同一搜索系统的结果相对应的多个链接。例如，视觉查询可以是一群人的图像或图片。交互式结果文档可以包括围绕每一个人的边界框，其当被激活时为人群中的每一个脸从面部识别搜索系统返回结果。对于一些视觉查询，交互式结果文档中的多个链接对应于来自多于一个搜索系统的搜索结果 (310)。例如，如果人和狗的图片作为视觉查询被提交，则交互式结果文档中的边界框可以分别勾勒出该人和狗。当（在交互式结果文档中）该人被选择时，返回来自面部识别搜索系统的搜索结果，以及当（在交互式结果文档中）该狗被选择时，返回来自图像到词语搜索系统的结果。对于一些视觉查询，交互式结果文档包含 OCR 结果和图像匹配结果 (312)。例如，如果一个人站在标志旁边的图片作为视觉查询被提交，则交互式结果文档可以包括用于该人和用于该标志中的文本的视觉标识符。类似地，如果杂志的扫描被用作视觉查询，则交互式结果文档可以包括用于页面上的广告中的照片或商标的视觉标识符以及用于同样在该页面上的文章的文本的视觉标识符。

[0065] 在创建了交互式结果文档之后，将其发送给客户端系统 (314)。在一些实施例中，如在上面参考图 2 所论述的，结合来自一个或多个并行搜索系统的搜索结果列表发送交互式结果文档（例如，图 15 的文档 1200）。在一些实施例中，如图 15 中所示，在客户端系统处在来自一个或多个并行搜索系统的搜索结果列表之上或邻近所述搜索结果列表显示交互式结果文档 (315)。

[0066] 可选地，用户将通过选择结果文档中的视觉标识符来与结果文档进行交互。服务器系统从客户端系统接收关于交互式结果文档中的视觉标识符的用户选择的信息 (316)。如上所述，在一些实施例中，通过选择边界框内的激活区域来激活链接。在其他实施例中，通过视觉查询的子部分的视觉标识符的用户选择来激活链接，所述视觉标识符不是边界框。在一些实施例中，链接的视觉标识符是热按钮、位于子部分附近的标签、文本中加下划线的词、或视觉查询中的对象或主题的其他表示。

[0067] 在将搜索结果列表与交互式结果文档一起展示 (315) 的实施例中，当用户选择用户可选择的链接 (316) 时，标识搜索结果列表中的与所选择的链接相对应的搜索结果。在一些实施例中，光标将跳转或自动移动到与所选择的链接相对应的第一结果。在客户端 102 的显示器太小而不能显示交互式结果文档和整个搜索结果列表两者的一些实施例中，选择

交互式结果文档中的链接促使搜索结果列表滚动或跳转,以显示与所选择的链接相对应的至少第一结果。在一些其他实施例中,响应于对交互式结果文档中的链接的用户选择,对结果列表重新排序,使得在结果列表的顶部显示与该链接相对应的第一结果。

[0068] 在一些实施例中,当用户选择用户可选择的链接(316)时,视觉查询服务器系统将与视觉查询的对应子部分有关的结果的至少子集发送给客户端以向用户显示(318)。在一些实施例中,用户可以同时选择多个视觉标识符,并且将同时接收关于所有所选择的视觉标识符的结果子集。在其他实施例中,在对用户可选择的链接中的任何链接的用户选择之前,将对应于用户可选择的链接的搜索结果预载到客户端上,以响应于对交互式结果文档中的一个或多个链接的用户选择而几乎即刻向用户提供搜索结果。

[0069] 图4是图示在客户端和视觉查询服务器系统之间的通信的流程图。客户端102从用户/查询者接收视觉查询(402)。在一些实施例中,可以仅从已注册或“选择加入(opt-in)”到视觉查询系统的用户接受视觉查询。在一些实施例中,仅为已注册面部识别视觉查询系统的用户执行对面部识别匹配的搜索,而为任何人执行其他类型的视觉查询,不管其是否已“选择加入”到面部识别部分。

[0070] 如上所述,视觉查询的格式可以采取许多形式。视觉查询将可能包含位于视觉查询文档的子部分中的一个或多个主题。对于一些视觉查询,客户端系统102对视觉查询执行类型识别预处理(404)。在一些实施例中,客户端系统102在该预处理系统中搜索特定可识别的模式。例如,对于一些视觉查询,客户端可以识别颜色。对于一些视觉查询,客户端可以识别特定子部分可能包含文本(因为该区域由以淡色空间等环绕的较小暗色字符组成)。客户端可以包含任何数量的预处理类型识别器或类型识别模块。在一些实施例中,客户端将具有用于识别条形码的类型识别模块(条形码识别406)。可以通过识别长方形区域中的独特条纹图案来那么做。在一些实施例中,客户端将具有用于识别视觉查询的特定主题或子部分可能包含人脸的类型识别模块(人脸检测408)。

[0071] 在一些实施例中,将所识别的“类型”返回给用户以供验证。例如,客户端系统102可以返回陈述“在你的视觉查询中找到了条形码,你对接收条形码查询结果感兴趣吗?”的消息。在一些实施例中,消息甚至可以指示类型在其中被找到的视觉查询的子部分。在一些实施例中,该展示类似于参考图3所论述的交互式结果文档。例如,其可以勾勒出视觉查询的子部分,并且指示该子部分可能包含人脸,以及询问用户其是否对接收面部识别结果感兴趣。

[0072] 在客户端102执行视觉查询的可选预处理之后,客户端将该视觉查询发送给视觉查询服务器系统106,具体发送给前端视觉查询处理服务器110。在一些实施例中,如果预处理产生了相关结果,即如果类型识别模块中的一个产生了高于某一阈值的结果,指示查询或查询的子部分可能是特定类型(人脸、文本、条形码等)的,则客户端将向前传递关于预处理的结果的信息。例如,客户端可以指示人脸识别模块对视觉查询的特定子部分包含人脸有75%的把握。更一般地,预处理结果,如果有的话,包括一个或多个主题类型值(例如,条形码、人脸、文本等)。可选地,发送给视觉查询服务器系统的预处理结果包括以下的一个或多个:对于预处理结果中的每一个主题类型值,识别与该主题类型值相对应的视觉查询的子部分的信息,以及对于预处理结果中的每一个主题类型值,指示对该主题类型值和/或视觉查询的对应子部分的标识的置信水平的置信值。

[0073] 前端服务器 110 从客户端系统接收视觉查询 (202)。所接收到的视觉查询可以包含上述预处理信息。如上所述,前端服务器将视觉查询发送给多个并行搜索系统 (210)。如果前端服务器 110 接收了关于子部分包含了某一类型的主题的可能性的预处理信息,则前端服务器可以将该信息向前传递给并行搜索系统中的一个或多个。例如,其可以传递特定子部分可能是人脸的信息,使得面部识别搜索系统 112-A 能够首先对视觉查询的该分部进行处理。类似地,发送 (特定子部分可能是人脸的) 相同信息可以由其他并行搜索系统使用来忽略该子部分或首先对其他子部分进行分析。在一些实施例中,前端服务器不会将预处理信息传递给并行搜索系统,而是替代地使用该信息来扩充其对从并行搜索系统接收的结果进行处理的方式。

[0074] 如参考图 2 所说明的,对于一些视觉查询,前端服务器 110 从并行搜索系统接收多个搜索结果 (214)。前端服务器然后可以执行多种排名和过滤,并且可以创建交互式搜索结果文档,如参考图 2 和 3 说明的。如果前端服务器 110 接收了关于子部分包含了某一类型的主题的可能性的预处理信息,则其可以通过给予匹配经预处理所识别的主题类型的那些结果偏好来进行过滤和排序。如果用户指示了请求特定类型的结果,则前端服务器将在处理结果时考虑用户的请求。例如,如果用户仅请求了条形码信息,则前端服务器可以滤除所有其他结果,或前端服务器将在列出其他结果之前列出与所请求的类型有关的所有结果。如果交互式视觉查询文档被返回,则服务器可以对与用户指示了感兴趣的结果类型相关联的链接进行预搜索,而仅提供用于执行对在交互式结果文档中所指示的其他主题的相关搜索的链接。然后,前端服务器 110 将搜索结果发送给客户端系统 (226)。

[0075] 客户端 102 从服务器系统接收结果 (412)。当适当时,这些结果将包括匹配在预处理阶段中找到的结果类型的结果。例如,在一些实施例中,其将包括一个或多个条形码结果 (414) 或一个或多个面部识别结果 (416)。如果客户端的预处理模块已指示了特定类型的结果是可能的,并且该结果被找到,则将突出列出该类型的所找到的结果。

[0076] 可选地,用户将对结果中的一个或多个进行选择或注释 (418)。用户可以选择一个搜索结果、可以选择特定类型的搜索结果、和 / 或可以选择交互式结果文档的一部分 (420)。对结果的选择是所返回的结果与查询相关的隐式反馈。这样的反馈信息可以在未来的查询处理操作中被利用。注释提供也能够未来的查询处理操作中被利用的、关于所返回的结果的显示反馈。注释采取以下形式:对所返回的结果的部分的校正 (如同对错误 OCR 化的词的校正) 或独立注释 (自由形式或结构化的)。

[0077] 用户的对一个搜索结果的选择,一般从同一类型的数个结果选择“正确的”结果 (例如,选择来自面部识别服务器的正确的结果),是被称为解释中的选择的过程。用户的对特定类型的搜索结果的选择,一般从数个不同类型的所返回的结果选择感兴趣的“类型”的结果 (例如,选择杂志中的文章的经 OCR 化的文本,而不是关于同样在同一页面上的广告的视觉结果),是被称为对意图的消歧的过程。如参考图 8 详细说明的,用户可以类似地选择经 OCR 化的文档中的特定链接的词 (诸如所识别的命名实体)。

[0078] 替代地或另外地,用户可能希望对特定搜索结果进行注释。可以以自由形式风格或结构化格式完成该注释 (422)。注释可以是对结果的描述或可以是对结果的评论。例如,其可以指示结果中的主题的名称,或其可以指示“这是本好书”或“这个产品在购买一年内损坏”。注释的另一个示例是围绕视觉查询的子部分的用户绘制的边界框和标识该边界框

内的对象或主题的用户提供的文本。参考图 5 更详细说明了用户注释。

[0079] 将搜索结果的用户选择和其他注释发送给服务器系统 (424)。前端服务器 110 接收该选择和注释,并且对其进行进一步处理 (426)。如果信息是对交互式结果文档中的对象、子区域或词语的选择,如适当,可以请求关于该选择的进一步信息。例如,如果选择是对一个视觉结果的,则将请求关于该视觉结果的更多信息。如果选择是(来自 OCR 服务器或来自图像到词语服务器的)词,则将该词的文本搜索发送给词语查询服务器系统 118。如果选择是来自面部图像识别搜索系统的人的,则将请求该人的简档。如果选择是关于交互式搜索结果文档的特定部分的,则将请求潜在的视觉查询结果。

[0080] 参考图 5 说明的,如果服务器系统接收注释,则将该注释存储在查询和注释数据库 116 中。然后,将来自注释数据库 116 的信息周期性复制到并行服务器系统中的一个或多个的单独注释数据库,如在下面参考图 7-10 所论述的。

[0081] 图 5 是图示与本发明的一个实施例一致的客户端系统 102 的框图。客户端系统 102 典型地包括一个或多个处理单元 (CPU) 702、一个或多个网络或其他通信接口 704、存储器 712 以及用于使这些组件互连的一个或多个通信总线 714。客户端系统 102 包括用户接口 705。用户接口 705 包括显示设备 706,以及可选地包括诸如键盘、鼠标或其他输入按钮的输入装置 708。替选地或另外地,显示设备 706 包括触敏表面 709,在这种情况下,显示器 706/709 是触敏显示器。在具有触敏显示器 706/709 的客户端系统中,物理键盘是可选的(例如,当需要键盘输入时,可以显示软键盘)。此外,一些客户端系统使用麦克风和语音识别来补充或替代键盘。可选地,客户端 102 包括 GPS(全球定位卫星)接收器或用于确定客户端系统 102 的位置的其他位置检测装置 707。在一些实施例中,提供了视觉查询搜索服务,其要求客户端系统 102 支持视觉查询服务器系统接收指示客户端系统 102 的位置的位置信息。

[0082] 客户端系统 102 还包括图像捕捉设备 710,诸如照相机或扫描仪。存储器 712 包括高速随机存取存储器,诸如 DRAM、SRAM、DDR RAM 或其他随机存取固态存储器设备;以及可以包括非易失性存储器,诸如一个或多个磁盘存储设备、光盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 712 可以可选地包括位于远离 CPU 702 的地方的一个或多个存储设备。存储器 712 或替选地存储器 712 内的非易失性存储器设备包括非暂时性计算机可读存储介质。在一些实施例中,存储器 712 或存储器 712 的计算机可读存储介质存储以下的程序、模块和数据结构、或其子集:

[0083] • 操作系统 716,其包括用于处理各种基本系统服务和用于执行依赖硬件的任务的程序;

[0084] • 网络通信模块 718,其被用于经由一个或多个通信网络接口 704(有线的或无线的)和诸如因特网、其他广域网、局域网、城域网等等的的一个或多个通信网络将客户端计算机 102 连接到其他计算机;

[0085] • 图像捕捉模块 720,其用于处理图像捕捉设备/照相机 710 所捕捉到的相应图像,其中该相应图像可以作为视觉查询(例如,由客户端应用模块)发送给视觉查询服务器系统;

[0086] • 一个或多个客户端应用模块 722,其用于处理按图像进行查询的各个方面,包括但不限于:按图像查询提交模块 724,其用于将视觉查询提交给视觉查询服务器系统;可选

地,感兴趣区域选择模块 725,其检测对图像中的感兴趣区域的选择(诸如,在触敏显示器 706/709 上的手势),并且将该感兴趣区域准备为视觉查询;结果浏览器 726,其用于显示视觉查询的结果;以及可选地,注释模块 728,其带有:用于结构化注释文本输入的可选模块 730,诸如以一种形式进行填充;或用于自由形式注释文本输入的可选模块 732,其能够接受来自多种格式的注释;以及图像区域选择模块 734(在此有时被称为结果选择模块),其允许用户选择图像的特定子部分以进行注释;

[0087] • 可选的内容创作应用 736,其允许用户通过创建或编辑图像,而不是仅仅经由图像捕捉设备 710 捕捉一个来创作视觉查询;可选地,一个或这样的应用 736 可以包括使用户能够选择图像的子部分以用作视觉查询的指令;

[0088] • 可选的本地图像分析模块 738,其在将视觉查询发送给视觉查询服务器系统之前对该视觉查询进行预处理。本地图像分析可以识别图像的特定类型或在图像内的子区域。可以由这样的模块 738 识别的图像类型的示例包括以下的一个或多个:面部类型(在视觉查询内识别的面部图像)、条形码类型(在视觉查询内识别的条形码)以及文本类型(在视觉查询内识别的文本);以及

[0089] • 另外的可选客户端应用 740,诸如电子邮件应用、电话应用、浏览器应用、地图应用、即时消息应用、社交网络应用等。在一些实施例中,当适当的可动作搜索结果被选择时,可以启动或访问与该可动作搜索结果相对应的应用。

[0090] 可选地,允许用户选择图像的特定子部分以进行注释的图像区域选择模块 734 还允许用户选择搜索结果作为“正确的”命中,而不必对其进行进一步注释。例如,用户可以被展示有最高 N 个面部识别匹配,并且可以从该结果列表选择正确的人。对于一些搜索查询,将展示多于一种类型的结果,以及用户将选择一种类型的结果。例如,图像查询可以包括一个人站在树旁边,但是只有关于该人的结果对用户来说才是感兴趣的。因此,图像选择模块 734 允许用户指示哪种图像类型是“正确的”类型——即,其在接收中感兴趣的类型。用户还可能希望通过使用(用于以一种形式进行填充的)注释文本输入模块 730 或自由形式注释文本输入模块 732 来添加个人评注或描述性词,来对搜索结果进行注释。

[0091] 在一些实施例中,可选的本地图像分析模块 738 是客户端应用(图 1 的 108)的一部分。此外,在一些实施例中,可选的本地图像分析模块 738 包括用来执行本地图像分析以对视觉查询或其部分进行预处理或分类的一个或多个程序。例如,客户端应用 722 可以在将视觉查询提交给搜索引擎之前识别图像包含条形码、人脸或文本。在一些实施例中,当本地图像分析模块 738 检测到视觉查询包含特定类型的图像时,该模块询问用户其是否对对应类型的搜索结果感兴趣。例如,本地图像分析模块 738 可以基于人脸的一般特征(即,不用确定哪个人的脸)来检测人脸,并且在将查询发送给视觉查询服务器系统上之前向用户提供即时反馈。其可以返回如同“检测到人脸,你对获得这张人脸的面部识别匹配感兴趣吗?”的结果。这可以为视觉查询服务器系统(图 1 的 106)节省时间。对于一些视觉查询,前端视觉查询处理服务器(图 1 的 110)仅将视觉查询发送给与本地图像分析模块 738 所识别的图像类型相对应的搜索系统 112。在其他实施例中,对搜索系统 112 的视觉查询可以将视觉查询发送给所有搜索系统 112A-N,但是将对来自与本地图像分析模块 738 所识别的图像类型相对应的搜索系统 112 的结果进行排名。在一些实施例中,本地图像分析对视觉查询服务器系统的操作产生影响的方式取决于客户端系统的配置、或与用户或客户端系统

相关联的配置或处理参数。此外,任何特定视觉查询的实际内容和由本地图像分析产生的结果可以促使不同视觉查询在客户端系统和视觉查询服务器系统任一或两者处被不同地处理。

[0092] 在一些实施例中,以两个步骤执行条形码识别,其中对视觉查询是否包括条形码的分析在客户端系统上本地图像分析模块 738 处执行。然后,只有在客户端确定视觉查询可能包括条形码时,才将该视觉查询传递给条形码搜索系统。在其他实施例中,条形码搜索系统对每个视觉查询进行处理。

[0093] 可选地,客户端系统 102 包括另外的客户端应用 740。

[0094] 图 6 是图示与本发明的一个实施例一致的前端视觉查询处理服务器系统 110 的框图。前端服务器 110 典型地包括一个或多个处理单元 (CPU) 802、一个或多个网络或其他通信接口 804、存储器 812 以及用于使这些组件互连的一个或多个通信总线 814。存储器 812 包括高速随机存取存储器,诸如 DRAM、SRAM、DDR RAM 或其他随机存取固态存储器设备;以及可以包括非易失性存储器,诸如一个或多个磁盘存储设备、光盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 812 可以可选地包括位于远离 CPU 802 的地方的一个或多个存储设备。存储器 812 或替代地存储器 812 内的非易失性存储器设备包括非暂时性计算机可读存储介质。在一些实施例中,存储器 812 或存储器 812 的计算机可读存储介质存储以下的程序、模块和数据结构、或其子集:

[0095] • 操作系统 816,其包括用于处理各种基本系统服务和用于执行依赖硬件的任务的程序;

[0096] • 网络通信模块 818,其被用于经由一个或多个通信网络接口 804(有线的或无线的)和诸如因特网、其他广域网、局域网、城域网等等的的一个或多个通信网络将前端服务器系统 110 连接到其他计算机;

[0097] • 查询管理器 820,其用于处理来自客户端系统 102 的进入的视觉查询,并且将其发送给两个或多个并行搜索系统;如在本文档中其他地方所描述的,在一些特殊情况下,视觉查询可以针对搜索系统中的仅仅一个,诸如当视觉查询包括客户端生成的指令(例如,“仅面部识别搜索”)时;

[0098] • 结果过滤模块 822,其用于可选地对来自一个或多个并行搜索系统的结果进行过滤,并且将最高或“相关”结果发送给客户端系统 102 以供展示;

[0099] • 结果排名和格式化模块 824,其用于可选地对来自一个或多个并行搜索系统的结果进行排名,以及用于对结果进行格式化以供展示;

[0100] • 结果文档创建模块 826,其在适当时被用来创建交互式搜索结果文档;模块 826 可以包括子模块,包括但不限于:边界框创建模块 828 和链接创建模块 830;

[0101] • 标签创建模块 831,其用于创建为视觉查询的相应子部分的视觉标识符的标签;

[0102] • 注释模块 832,其用于从用户接收注释,并且将其发送给注释数据库 116;

[0103] • 可动作搜索结果模块 838,其用于响应于视觉查询,生成一个或多个可动作搜索结果元素,每一个被配置成启动客户端侧动作;可动作搜索结果元素的示例是用来发起电话呼叫、发起电子邮件消息、在地图上标出地址、进行餐馆预定以及提供购买产品的选项的按钮;以及

[0104] • 查询和注释数据库 116,其包括数据库本身 834 和对数据库的索引 836。

[0105] 结果排名和格式化模块 824 对从一个或多个并行搜索系统（图 1 的 112-A-112-N）返回的结果进行排名。如在上面已指出的，对于一些视觉查询，仅仅来自一个搜索系统的结果可能是相关的。在这样的情况下，仅仅对来自该一个搜索系统的相关搜索结果进行排名。对于一些视觉查询，数种类型的搜索结果可能是相关的。在这些情况下，在一些实施例中，结果排名和格式化模块 824 使来自具有最相关结果（例如，具有最高相关性分值的结果）的搜索系统的所有结果排名高于关于较不相关的搜索系统的结果。在其他实施例中，结果排名和格式化模块 824 使来自每一个相关搜索系统的最高结果排名高于剩余结果。在一些实施例中，结果排名和格式化模块 824 根据为搜索结果中的每一个计算的相关性分值对结果进行排名。对于一些视觉查询，除在并行视觉搜索系统上进行搜索外，执行扩充的文本查询。在一些实施例中，当文本查询同样被执行时，以在视觉上不同于视觉搜索系统结果的方式展示其结果。

[0106] 结果排名和格式化模块 824 还对结果进行格式化。在一些实施例中，以列表格式展示结果。在一些实施例中，用交互式结果文档展示结果。在一些实施例中，展示交互式结果文档和结果列表两者。在一些实施例中，查询类型指示结果如何被展示。例如，如果在视觉查询中检测到多于一个可搜索的主题，则产生交互式结果文档，而如果仅仅检测到一个可搜索的主题，则将仅仅以列表格式显示结果。

[0107] 结果文档创建模块 826 用于创建交互式搜索结果文档。交互式搜索结果文档可以具有一个或多个检测和搜索到的主题。边界框创建模块 828 创建围绕搜索到的主题中的一个或多个的边界框。边界框可以是长方形框，或可以勾勒出主题的形状。链接创建模块 830 创建到搜索结果的链接，所述搜索结果与其中的在交互式搜索结果文档中的相应主题相关联。在一些实施例中，在边界框区域内点击激活链接创建模块所插入的对应链接。

[0108] 查询和注释数据库 116 包含能够用来改进视觉查询结果的信息。在一些实施例中，用户可以在视觉查询结果已被展示之后对图像进行注释。此外，在一些实施例中，用户可以在将图像发送给视觉查询搜索系统之前对图像进行注释。预注释可以通过使结果集中、或与视觉查询搜索并行运行对注释的词基于文本的搜索来帮助视觉查询处理。在一些实施例中，可以使图片的经注释的版本公开（例如，当用户例如通过将图像和注释指明为不是私人的来准许了公开时），以被返回为潜在图像匹配命中。例如，如果用户拍摄了花的图片，并且通过给出关于该花的详细属和种信息来对该图像进行注释，则该用户可能想要该图像向执行查找该花的视觉查询研究的任何人展示。在一些实施例中，将来自查询和注释数据库 116 的信息周期性推送到并行搜索系统 112，其将信息的相关部分（如果有的话）合并到其各自的单独数据库 114 中。

[0109] 图 7 是图示用来处理视觉查询的并行搜索系统中的一个的框图，图 7 图示了与本发明的一个实施例一致的“通用”搜索系统 112-N。该服务器系统是通用的，仅仅因为其表示视觉查询搜索服务器 112-N 中的任何一个。通用服务器系统 112-N 典型地包括一个或多个处理单元（CPU）502、一个或多个网络或其他通信接口 504、存储器 512 以及用于使这些组件互连的一个或多个通信总线 514。存储器 512 包括高速随机存取存储器，诸如 DRAM、SRAM、DDR RAM 或其他随机存取固态存储器设备；以及可以包括非易失性存储器，诸如一个或多个磁盘存储设备、光盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 512 可以可选地包括位于远离 CPU 502 的地方的一个或多个存储设备。存储器 512 或备选地存

存储器 512 内的非易失性存储器设备包括非暂时性计算机可读存储介质。在一些实施例中，存储器 512 或存储器 512 的计算机可读存储介质存储以下的程序、模块和数据结构、或其子集：

[0110] • 操作系统 516, 其包括用于处理各种基本系统服务和用于执行依赖硬件的任务的程序；

[0111] • 网络通信模块 518, 其被用于经由一个或多个通信网络接口 504 (有线的或无线的) 和诸如因特网、其他广域网、局域网、城域网等等的的一个或多个通信网络将通用服务器系统 112-N 连接到其他计算机；

[0112] • 特定于特定服务器系统的搜索应用 520, 其例如可以是条形码搜索应用、颜色识别搜索应用、产品识别搜索应用和对象或对象类别搜索应用等；

[0113] • 如果特定搜索应用利用索引, 则可选的索引 522；

[0114] • 可选的图像数据库 524, 其用于存储与特定搜索应用相关的图像, 其中所存储的图像数据, 如果有的话, 取决于搜索过程类型；

[0115] • 可选的结果排名模块 526 (有时被称为相关性评分模块), 其用于对来自搜索应用的结果进行排名, 排名模块可以为来自搜索应用的每一个结果分配相关性分值, 以及如果没有结果达到预定义的最小分值, 则可以向前端视觉查询处理服务器返回指示来自该服务器系统的结果不相关的空或零值分值；以及

[0116] • 注释模块 528, 其用于从注释数据库 (图 1 的 116) 接收注释信息、确定注释信息中的任何信息是否与特定搜索应用相关, 以及将注释信息的任何确定的相关部分合并入相应的注释数据库 530。

[0117] 图 8 是图示与本发明的一个实施例一致的利用来处理视觉查询的 OCR 搜索系统 112-B 的框图。OCR 搜索系统 112-B 典型地包括一个或多个处理单元 (CPU) 602、一个或多个网络或其他通信接口 604、存储器 612 以及用于使这些组件互连的一个或多个通信总线 614。存储器 612 包括高速随机存取存储器, 诸如 DRAM、SRAM、DDR RAM 或其他随机存取固态存储器设备；以及可以包括非易失性存储器, 诸如一个或多个磁盘存储设备、光盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 612 可以可选地包括位于远离 CPU 602 的地方的一个或多个存储设备。存储器 612 或替代地存储器 612 内的非易失性存储器设备包括非暂时性计算机可读存储介质。在一些实施例中, 存储器 612 或存储器 612 的计算机可读存储介质存储以下的程序、模块和数据结构、或其子集：

[0118] • 操作系统 616, 其包括用于处理各种基本系统服务和用于执行依赖硬件的任务的程序；

[0119] • 网络通信模块 618, 其被用于经由一个或多个通信网络接口 604 (有线的或无线的) 和诸如因特网、其他广域网、局域网、城域网等等的的一个或多个通信网络将 OCR 搜索系统 112-B 连接到其他计算机；

[0120] • 光学字符识别 (OCR) 模块 620, 其试图识别视觉查询中的文本, 并且将字母图像转换成字符；

[0121] • 可选的 OCR 数据库 114-B, 其由 OCR 模块 620 利用来识别特定字体、文本图案以及对字母识别特有的其他特征；

[0122] • 可选的拼写检查模块 622, 其通过针对词典检查经转换的词, 并且对另外匹配词

典词的词中的潜在错误转换的字母进行替换,来改进字母图像到字符的转换;

[0123] • 可选的命名实体识别模块 624,其搜索经转换的文本内的命名实体、将所识别的命名实体作为词语查询中的词语发送给词语查询服务器系统(图 1 的 118)、以及将来自词语查询服务器系统的结果作为嵌入在经 OCR 化的文本中的链接与所识别的命名实体相关联地提供;

[0124] • 可选的文本匹配应用 632,其通过针对文本片段数据库检查经转换的片段(诸如经转换的句子和段落),并且对另外匹配文本匹配应用文本片段的经 OCR 化的文本片段中的潜在错误转换的字母进行替换,来改进字母图像到字符的转换,在一些实施例中,将文本匹配应用所找到的文本片段作为链接提供给用户(例如,如果用户扫描了纽约时报的一个页面,则文本匹配应用可以提供到纽约时报网站上的整个发表的文章的链接);

[0125] • 结果排名和格式化模块 626,其用于对经 OCR 化的结果进行格式化以供展示,以及对到命名实体的可选链接进行格式化,以及还可选地对来自文本匹配应用的任何相关结果进行排名;以及

[0126] • 可选的注释模块 628,其用于从注释数据库(图 1 的 116)接收注释信息、确定注释信息中的任何信息是否与 OCR 搜索系统相关,以及将注释信息的任何确定的相关部分合并入相应的注释数据库 630。

[0127] 图 9 是图示与本发明的一个实施例一致的利用来处理视觉查询的面部识别搜索系统 112-A 的框图。面部识别搜索系统 112-A 典型地包括一个或多个处理单元(CPU)902、一个或多个网络或其他通信接口 904、存储器 912 以及用于使这些组件互连的一个或多个通信总线 914。存储器 912 包括高速随机存取存储器,诸如 DRAM、SRAM、DDR RAM 或其他随机存取固态存储器设备;以及可以包括非易失性存储器,诸如一个或多个磁盘存储设备、光盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 912 可以可选地包括位于远离 CPU 902 的地方的一个或多个存储设备。存储器 912 或替代地存储器 912 内的非易失性存储器设备包括非暂时性计算机可读存储介质。在一些实施例中,存储器 912 或存储器 912 的计算机可读存储介质存储以下的程序、模块和数据结构、或其子集:

[0128] • 操作系统 916,其包括用于处理各种基本系统服务和用于执行依赖硬件的任务的程序;

[0129] • 网络通信模块 918,其被用于经由一个或多个通信网络接口 904(有线的或无线的)和诸如因特网、其他广域网、局域网、城域网等等的一个或多个通信网络将面部识别搜索系统 112-A 连接到其他计算机;

[0130] • 面部识别搜索应用 920,其用于在面部图像数据库 114-A 中搜索匹配在视觉查询中出现的人脸的面部图像,以及对社交网络数据库 922 搜索与在面部图像数据库 114-A 中找到的每一个匹配有关的信息;

[0131] • 面部图像数据库 114-A,其用于为多个用户存储一个或多个面部图像;可选地,面部图像数据库包括除用户外的人的面部图像,诸如家庭成员和用户认识的并已被标识为在包括在面部图像数据库 114-A 中的图像中出现的其他人;可选地,面部图像数据库包括从外部源获取的面部图像,所述外部源诸如在公共域合法的面部图像供应商;

[0132] • 可选地,社交网络数据库 922,其包含与社交网络的用户有关的信息,诸如名字、地址、职业、组成员、社交网络关系、移动设备的当前 GPS 位置、共享偏好、兴趣、年龄、家乡、

个人统计、工作信息等,如参考图 12A 更详细论述的;

[0133] • 结果排名和格式化模块 924,其用于对来自面部图像数据库 114-A 的潜在面部匹配进行排名(例如,将相关性和/或匹配质量分值分配给所述潜在面部匹配),以及对结果进行格式化以供展示;在一些实施例中,对结果的排名或评分利用从前述社交网络数据库检索到的相关信息;在一些实施例中,搜索格式化的结果包括潜在图像匹配以及来自社交网络数据库的信息子集;以及

[0134] • 注释模块 926,其用于从注释数据库(图 1 的 116)接收注释信息、确定注释信息中的任何信息是否与面部识别搜索系统相关,以及将注释信息的任何确定的相关部分存储入相应的注释数据库 928 中。

[0135] 图 10 是图示与本发明的一个实施例一致的利用来处理视觉查询的图像到词语搜索系统 112-C 的框图。在一些实施例中,图像到词语搜索系统识别视觉查询中的对象(实例识别)。在其他实施例中,图像到词语搜索系统识别视觉查询中的对象类别(类型识别)。在一些实施例中,图像到词语系统识别对象和对象类别两者。图像到词语搜索系统为视觉查询中的图像返回潜在词语匹配。图像到词语搜索系统 112-C 典型地包括一个或多个处理单元(CPU) 1002、一个或多个网络或其他通信接口 1004、存储器 1012 以及用于使这些组件互连的一个或多个通信总线 1014。存储器 1012 包括高速随机存取存储器,诸如 DRAM、SRAM、DDR RAM 或其他随机存取固态存储器设备;以及可以包括非易失性存储器,诸如一个或多个磁盘存储设备、光盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 1012 可以可选地包括位于远离 CPU 1002 的地方的一个或多个存储设备。存储器 1012 或替代地存储器 1012 内的非易失性存储器设备包括非暂时性计算机可读存储介质。在一些实施例中,存储器 1012 或存储器 1012 的计算机可读存储介质存储以下的程序、模块和数据结构、或其子集:

[0136] • 操作系统 1016,其包括用于处理各种基本系统服务和用于执行依赖硬件的任务的程序;

[0137] • 网络通信模块 1018,其被用于经由一个或多个通信网络接口 1004(有线的或无线的)和诸如因特网、其他广域网、局域网、城域网等等的的一个或多个通信网络将图像到词语搜索系统 112-C 连接到其他计算机;

[0138] • 图像到词语搜索应用 1020,其在图像搜索数据库 114-C 中搜索匹配视觉查询中的主题的图像;

[0139] • 图像搜索数据库 114-C,其能够被搜索应用 1020 搜索以找到类似于视觉查询的主题的图像;

[0140] • 词语到图像反向索引 1022,其存储用户在使用基于文本的查询搜索引擎 1006 来搜索图像时所使用的文本词语;

[0141] • 结果排名和格式化模块 1024,其用于对潜在图像匹配进行排名和/或对在词语到图像反向索引 1022 中标识的与潜在图像匹配相关联的词语进行排名;以及

[0142] • 注释模块 1026,其用于从注释数据库(图 1 的 116)接收注释信息、确定注释信息中的任何信息是否与图像到词语搜索系统 112-C 相关,以及将注释信息的任何确定的相关部分存储入相应的注释数据库 1028 中。

[0143] 更多地意在将图 5-10 作为可以在一组计算机系统中存在的各种特征的功能性描

述,而不是作为在此描述的实施例的结构性示意。实际上,并且如本领域普通技术人员所认识到的,可以使分离示出的项相组合并且可以使一些项分离。例如,可以在单个服务器上实现在这些图中分离示出的一些项,并且可以通过一个或多个服务器来实现单个项。用来实现视觉查询处理的系统的实际数量以及在它们之间如何分配特征因实施方式的不同而不同。

[0144] 在此描述的方法中的每一个可以由存储在非暂时性计算机可读存储介质中并且由一个或多个服务器或客户端的一个或多个处理器执行的指令控制。在上面标识的模块或程序(即,指令集)不必被实现为独立软件程序、例程或模块,因此,在各种实施例中,可以组合或另外重新安排这些模块的各种子集。在图 5-10 中所示的操作中的每一个可以对应于存储在计算机存储器或非暂时性计算机可读存储介质中的指令。

[0145] 图 11 图示了带有示例性视觉查询 1102 的截屏的客户端系统 102。在图 11 中所示的客户端系统 102 是移动设备,诸如蜂窝电话、便携式音乐播放器或便携式电子邮件设备。客户端系统 102 包括显示器 706 和一个或多个输入装置 708,诸如在该附图中所示的按钮。在一些实施例中,显示器 706 是触敏显示器 709。在具有触敏显示器 709 的实施例中,在显示器 709 上显示的软按钮可以可选地替代机电按钮 708 中的部分或全部。如在下面更详细说明的,触敏显示器在与视觉查询结果进行交互时也是有帮助的。客户端系统 102 还包括图像捕捉机制,诸如照相机 710。

[0146] 图 11 图示了视觉查询 1102,其是在商店货架上的包装的照片或视频帧。在次描述的实施例中,视觉查询是在两维中的每一个中在像素上具有与视觉查询的大小相对应的分辨率的二维图像。在该示例中的视觉查询 1102 是三维对象的二维图像。视觉查询 1102 包括背景元素、产品包装 1104 以及包装上的多种类型的实体,包括人图像 1106、商标图像 1108、产品图像 1110 以及多种文本元素 1112。

[0147] 如参考图 3 说明的,视觉查询 1102 被发送给前端服务器 110,其将视觉查询 1102 发送给多个并行搜索系统(112A-N)、接收结果以及创建交互式结果文档。

[0148] 图 12A 和 12B 每一个图示了带有交互式结果文档 1200 的实施例的截屏的客户端系统 102。交互式结果文档 1200 包括视觉查询 1102 的相应子部分的一个或多个视觉标识符 1202,其每一个包括到搜索结果子集的用户可选择的链接。图 12A 和 12B 图示了带有为边界框 1202(例如,边界框 1202-1、1202-2、1202-3)的视觉标识符的交互式结果文档 1200。在图 12A 和 12B 中所示的实施例中,用户通过轻击在由特定子部分的边界框 1202 勾勒出的空间内的激活区域来激活对该特定子部分相对应的搜索结果的显示。例如,用户将通过轻击环绕人图像的边界框 1306(图 13)来激活与该人图像相对应的搜索结果。在其他实施例中,使用鼠标或键盘而不是触敏显示器来选择可选择的链接。在一些实施例中,当用户预览边界框 1202 时(即,当用户单击、轻击一次或将指针悬停在边界框上时),显示第一对应的搜索结果。当用户选择边界框时(即,当用户双击、轻击两次或使用另一个机制来指示选择时),用户激活对多个对应的搜索结果的显示。

[0149] 在图 12A 和 12B 中,视觉标识符是环绕视觉查询的子部分的边界框 1202。图 12A 图示了为正方形或长方形的边界框 1202。图 12B 图示了勾勒出视觉查询的子部分中的可标识的实体的边界的边界框 1202,诸如用于饮料瓶的边界框 1202-3。在一些实施例中,各个边界框 1202 在其内包括更小的边界框 1202。例如,在图 12A 和 12B 中,标识包装的边界

框 1202-1 环绕标识商标的边界框 1202-2 和所有其他边界框 1202。在包括文本的一些实施例中,还包括用于文本词语中的部分的活动热链接 1204。图 12B 示出了示例,其中“Active Drink”和“United States”被显示为热链接 1204。与这些词语相对应的搜索结果是从词语查询服务器系统 118 接收到的结果,而与边界框相对应的结果是来自按图像查询搜索系统的结果。

[0150] 图 13 图示了带有按视觉查询中的所识别的实体的类型编码的交互式结果文档 1200 的截屏的客户端系统 102。图 11 的视觉查询包含人图像 1106、商标图像 1108、产品图像 1110 和多种文本元素 1112。如此,在图 13 中显示的交互式结果文档 1200 包括围绕人 1306、标签 1308、产品 1310 和两个文本区域 1312 的边界框 1202。图 13 的边界框每一个用不同的交叉影线来展示,其表示不同颜色透明边界框 1202。在一些实施例中,将边界框的视觉标识符(和/或交互式结果文档 1200 中的标签或其他视觉标识符)进行格式化来以在视觉上不同的方式展示,诸如重叠的颜色、重叠的图案、标签背景颜色、标签背景图案、标签字体颜色以及边界框边界颜色。参考图 13 中的边界框示出了用于特定识别的实体的类型编码,但是也可以将按类型编码应用于为标签的视觉标识符。

[0151] 图 14 图示了具有带有标签 1402 的交互式结果文档 1200 的截屏的客户端设备 102,标签 1402 是图 11 的视觉查询 1102 的相应子部分的视觉标识符。标签视觉标识符 1402 每一个包括到对应的搜索结果的子集的用户可选择的链接。在一些实施例中,通过在标签 1402 的区域内所显示的描述性文本来识别可选择的链接。一些实施例包括在一个标签 1402 内的多个链接。例如,在图 14 中,悬停在喝水的女人的图像上的标签包括到关于该女人的面部识别结果的链接和到关于该特定图片的图像识别结果(例如,使用相同图片的其他产品或广告的图像)的链接。

[0152] 在图 14 中,标签 1402 被显示为带有文本的部分透明的区域,其位于交互式结果文档的其相应子部分上。在其他实施例中,将相应标签放置在靠近但是并不位于交互式结果文档的其相应子部分上。在一些实施例中,以与参考图 13 所论述的相同的方式按类型对标签进行编码。在一些实施例中,用户通过轻击在由标签 1302 的边缘或外围所勾勒出的空间内的激活区域来激活对与对应于标签 1302 的特定子部分相对应的搜索结果的显示。在上面参考图 12A 和 12B 的边界框所论述的相同预览和选择功能也适用于为标签 1402 的视觉标识符。

[0153] 图 15 图示了交互式结果文档 1200 和原始视觉查询 1102 与结果列表 1500 同时显示的截屏。在一些实施例中,如图 12-14 中所示,交互式结果文档 1200 独自显示。在其他实施例中,如图 15 中所示,交互式结果文档 1200 与原始视觉查询同时显示。在一些实施例中,视觉查询结果列表 1500 与原始视觉查询 1102 和/或交互式结果文档 1200 一起同时显示。客户端系统的类型和在显示器 706 上的空间量可以确定结果列表 1500 是否与交互式结果文档 1200 同时显示。在一些实施例中,客户端系统 102(响应于提交给视觉查询服务器系统的视觉查询)接收结果列表 1500 和交互式结果文档 1200 两者,但是当用户滚动到交互式结果文档 1200 下面时,仅仅显示结果列表 1500。在这些实施例中的一些中,客户端系统 102 在不必再次查询服务器的情况下显示与用户选择的视觉标识符 1202/1402 相对应的结果,因为结果列表 1500 由客户端系统 102 响应于视觉查询而接收,然后被本地存储在客户端系统 102 处。

[0154] 在一些实施例中,将结果列表 1500 组织成类别 1502。每一个类别包含至少一个结果 1503。在一些实施例中,使类别标题高亮以将其与结果 1503 进行区别。类别 1502 根据其计算的类别权重来排序。在一些实施例中,类别权重是在该类别中的最高 N 个结果的权重的组合。如此,首先显示可能产生了更相关结果的类别。在对于同一所识别的实体,返回多于一个类别 1502(诸如图 15 中所示的面部图像识别匹配和图像匹配) 的实施例中,首先显示的类别具有更高的类别权重。

[0155] 如参考图 3 说明的,在一些实施例中,当交互式结果文档 1200 中的可选择的链接被客户端系统 102 的用户选择时,光标将自动移动到适当的类别 1502 或该类别中的第一结果 1503。替选地,当交互式结果文档中的可选择的链接被客户端系统 102 的用户选择时,将结果列表 1500 进行重新排序,使得首先显示与所选择的链接相关的类别。这例如通过使可选择的链接编码有识别对应的搜索结果的信息,或通过对搜索结果进行编码以指示对应可选择的链接或指示对应的结果类别来完成。

[0156] 在一些实施例中,搜索结果的类别对应于产生那些搜索结果的按图像查询搜索系统。例如,在图 15 中,类别中的部分是产品匹配 1506、标志匹配 1508、面部识别匹配 1510、图像匹配 1512。原始视觉查询 1102 和 / 或交互式结果文档 1200 可以类似地用诸如查询 1504 的类别标题显示。类似地,还可以将来自词语查询服务器所执行的任何词语搜索的结果显示为独立类别,诸如 web 结果 1514。在其他实施例中,视觉查询中的多于一个实体将从同一按图像查询搜索系统产生结果。例如,视觉查询可以包括两个不同的人脸,其将从面部识别搜索系统返回不同结果。如此,在一些实施例中,类别 1502 按所识别的实体而不是搜索系统来划分。在一些实施例中,在所识别的实体类别头部 1502 中显示所识别的实体的图像,使得关于该所识别的实体的结果与关于另一个所识别的实体的结果是可区分的,即使两者结果由同一按图像查询搜索系统产生。例如,在图 15 中,产品匹配类别 1506 包括两个实体产品实体以及同样两个实体类别 1502——盒装产品 1516 和瓶装产品 1518,其中每一个具有多个对应的搜索结果 1503。在一些实施例中,类别可以按所识别的实体和按图像查询系统的类型来划分。例如,在图 15 中,在产品匹配类别产品下有返回了相关结果的两个不同实体。

[0157] 在一些实施例中,结果 1503 包括缩略图像。例如,如在图 15 中关于面部识别匹配结果所示的,与诸如图像中的人的名字的一些文本描述一起显示了关于“Actress X(女演员 X)”和“Social Network Friend Y(社交网络朋友 Y)”的面部匹配的图片的较小版本(也称为缩略图像)。

[0158] 为了说明目的,参考具体实施例描述了前面的描述。然而,在上面的说明性论述并不意在是穷尽的或将本发明限制在所公开的精确形式。根据上面的教导,许多修改和变形是可能的。为了最佳说明本发明的原理和其实际应用,选择并描述了实施例,从而使本领域技术人员能够以适于预期的特定用途的各种修改最佳利用本发明和各种实施例。

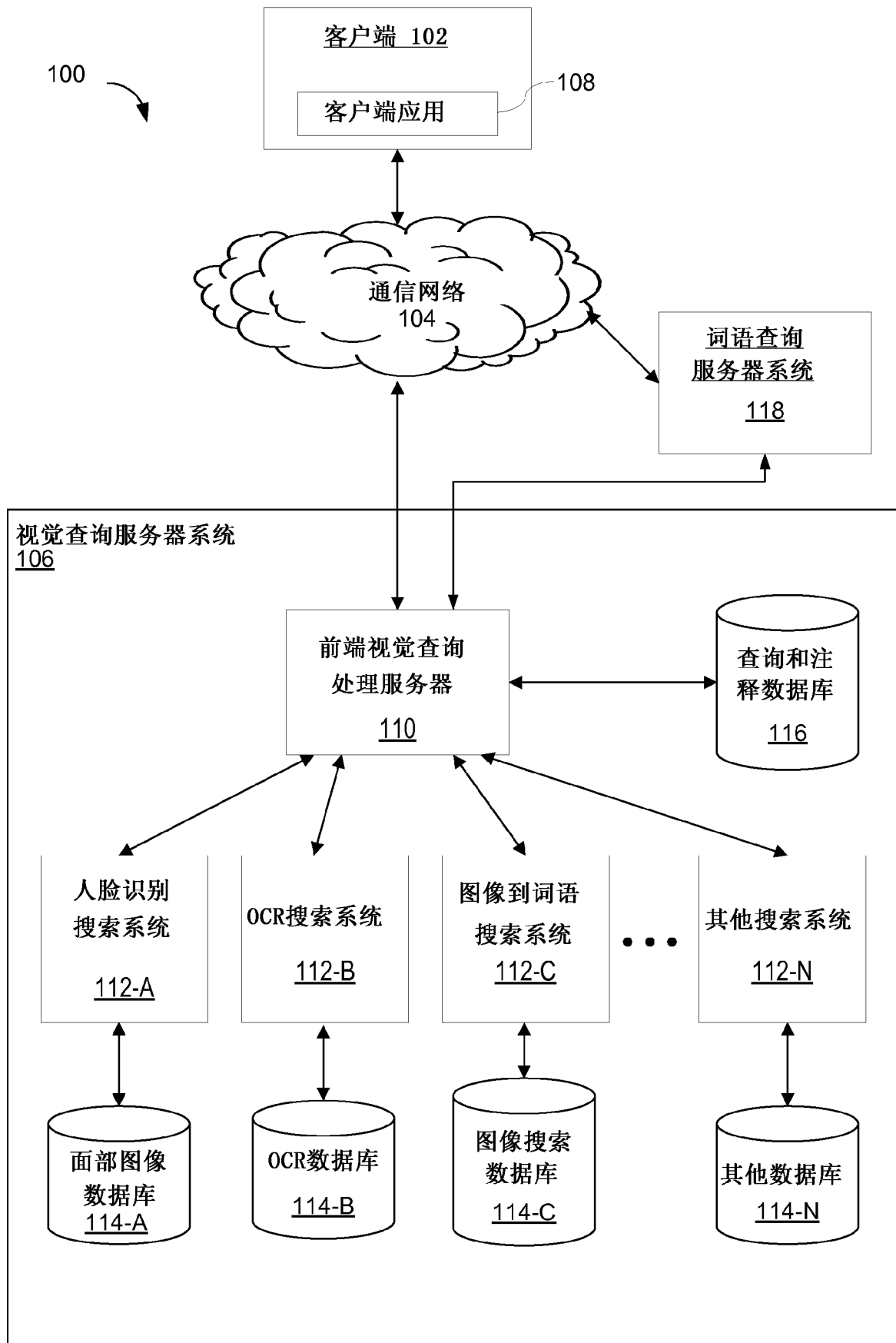


图 1

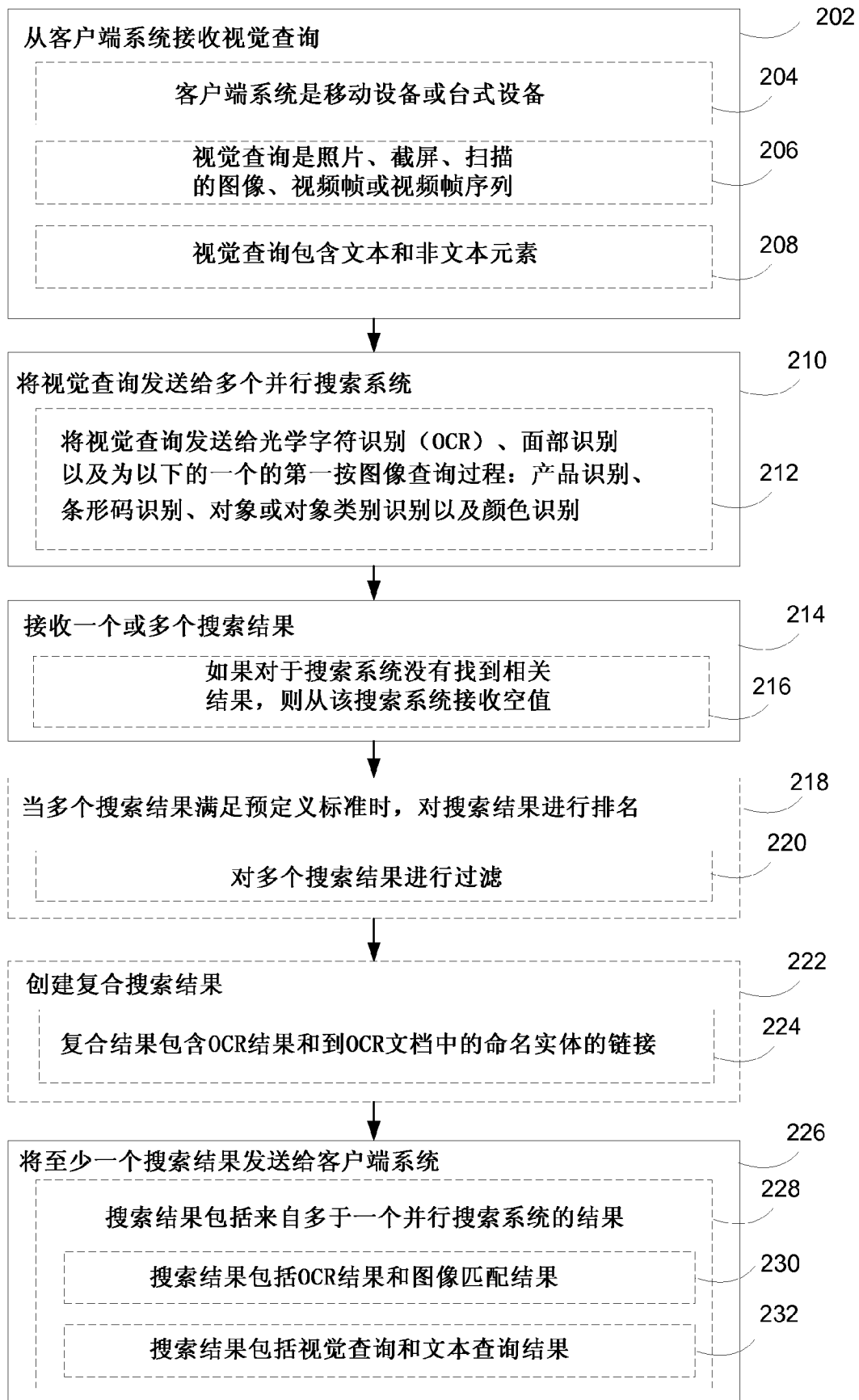


图 2

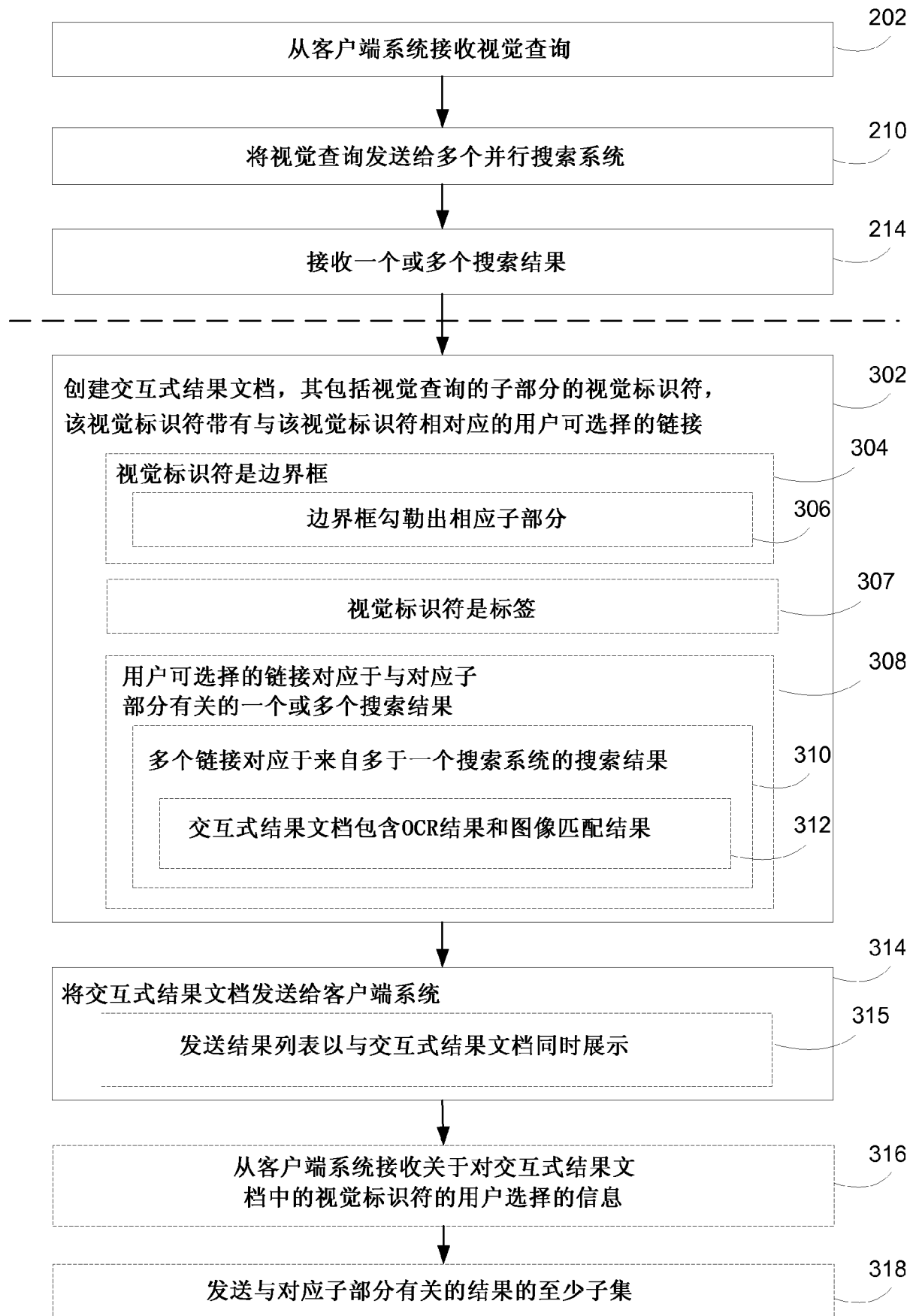


图 3

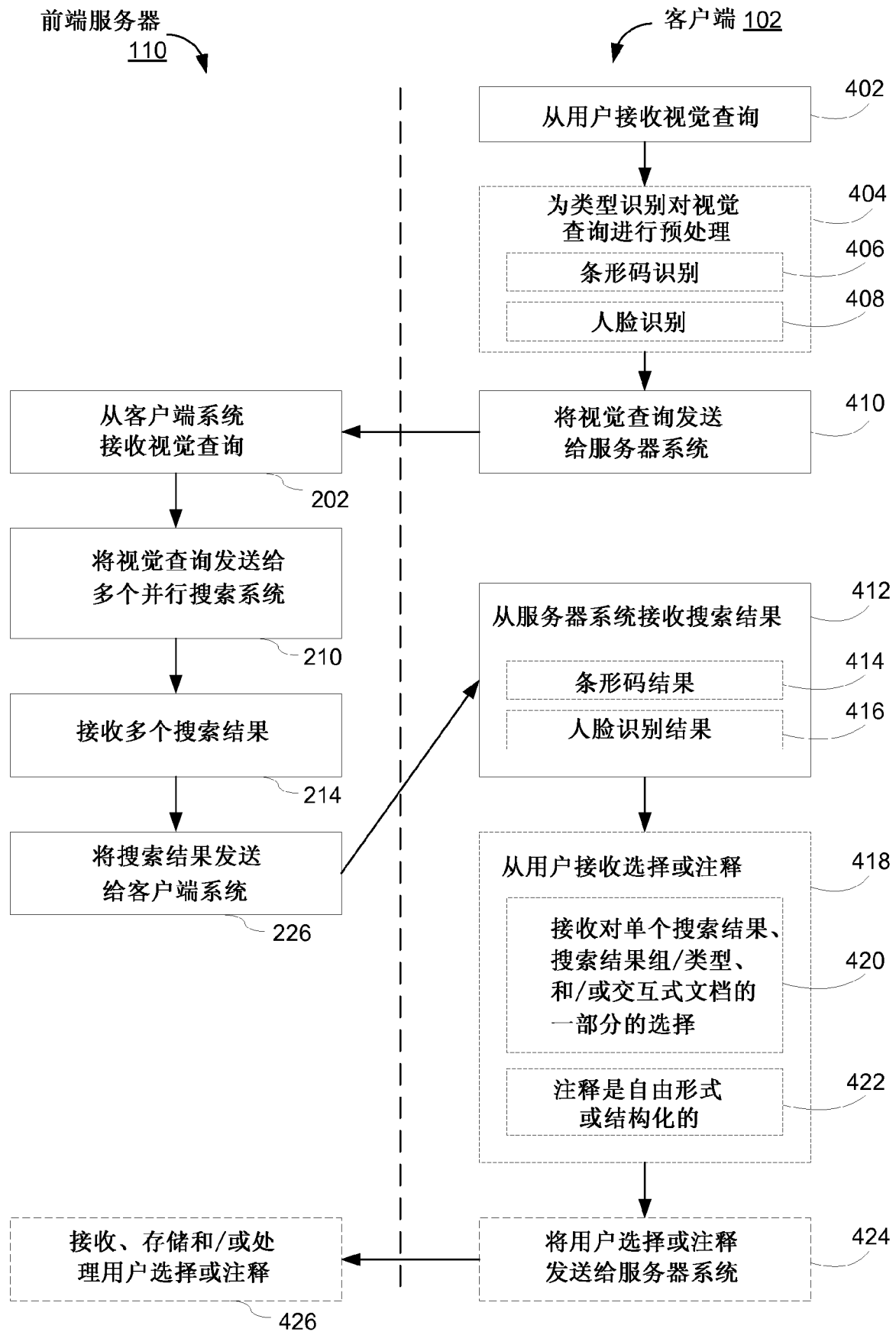


图 4

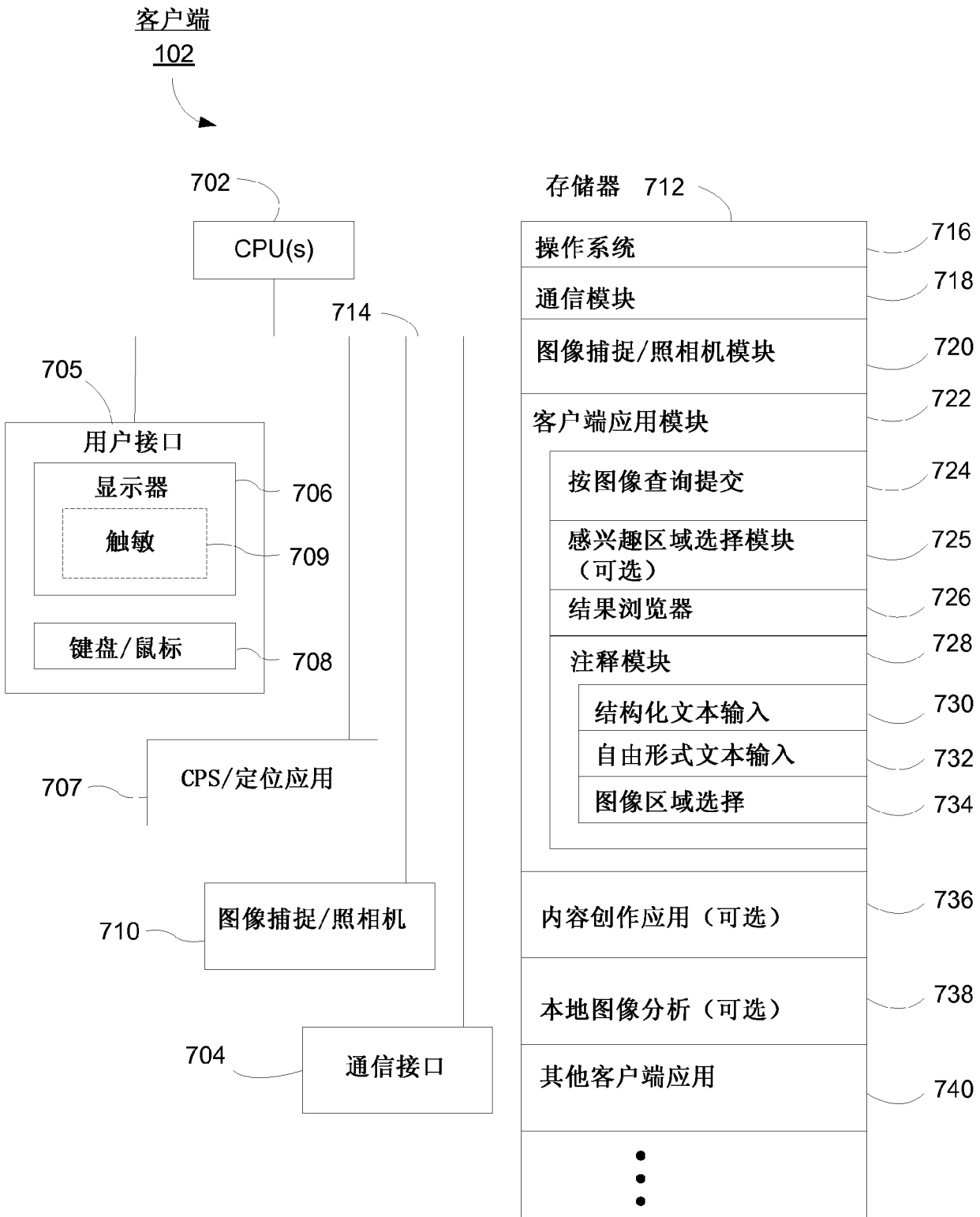


图 5

前端服务器搜索系统

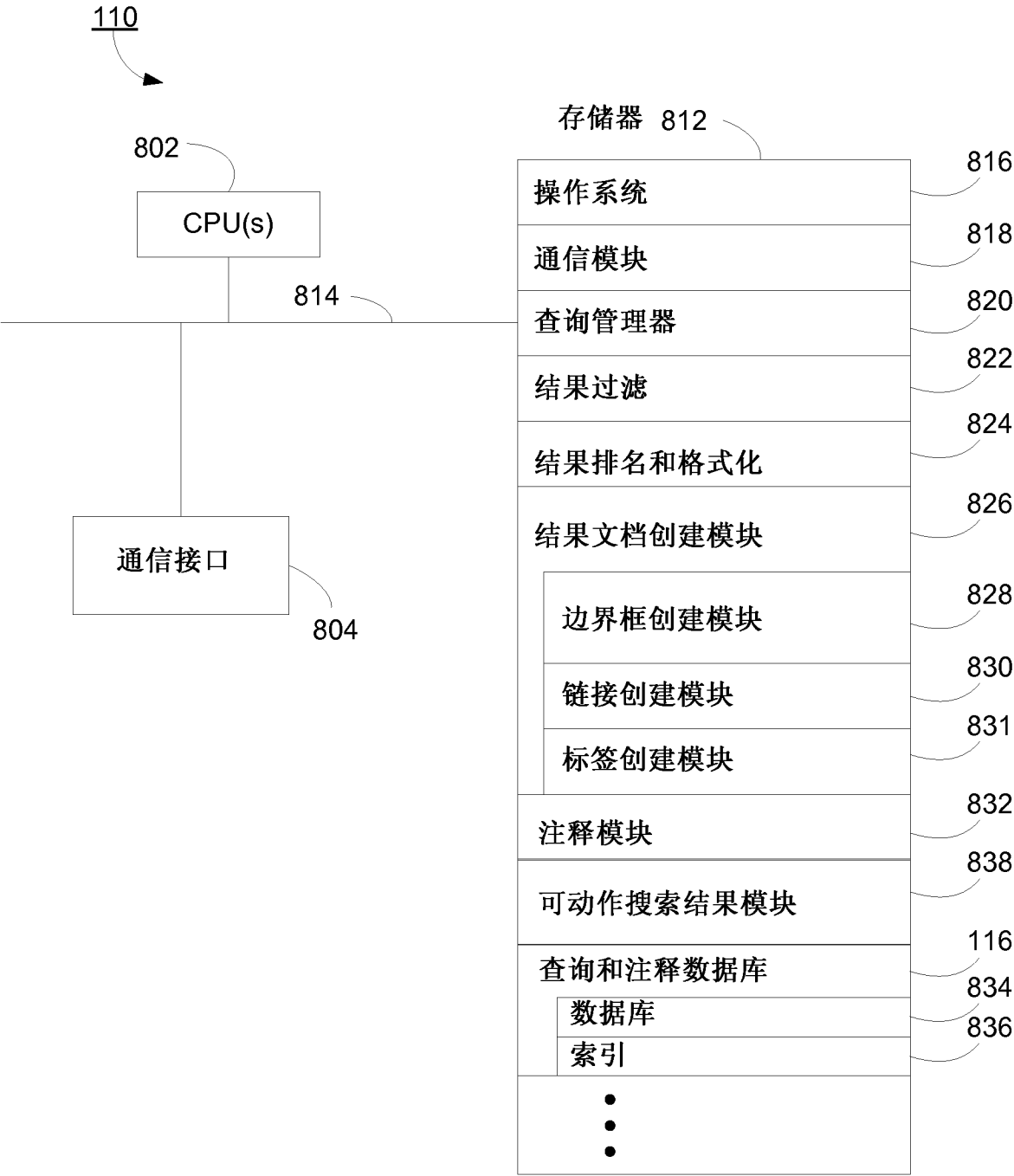


图 6

服务器
102

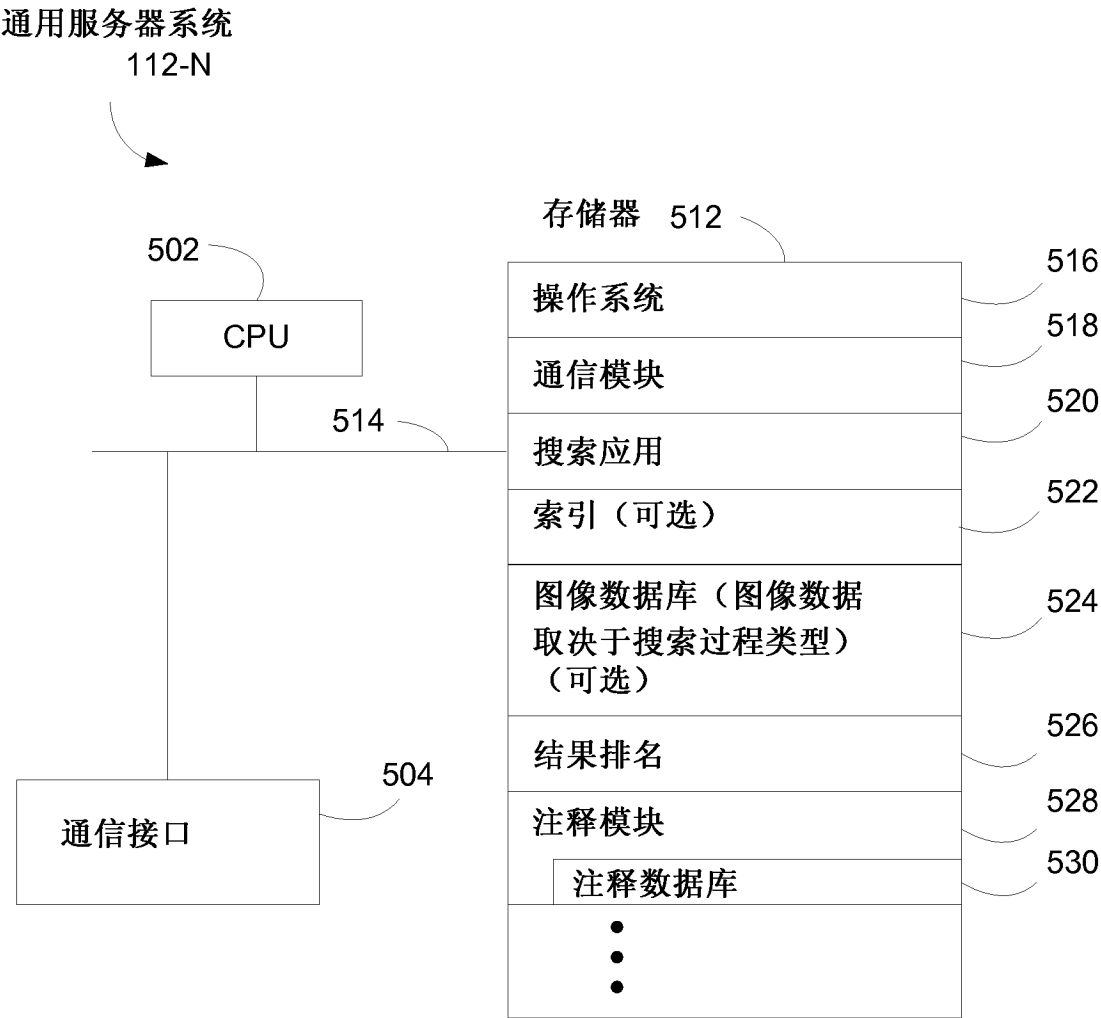


图 7

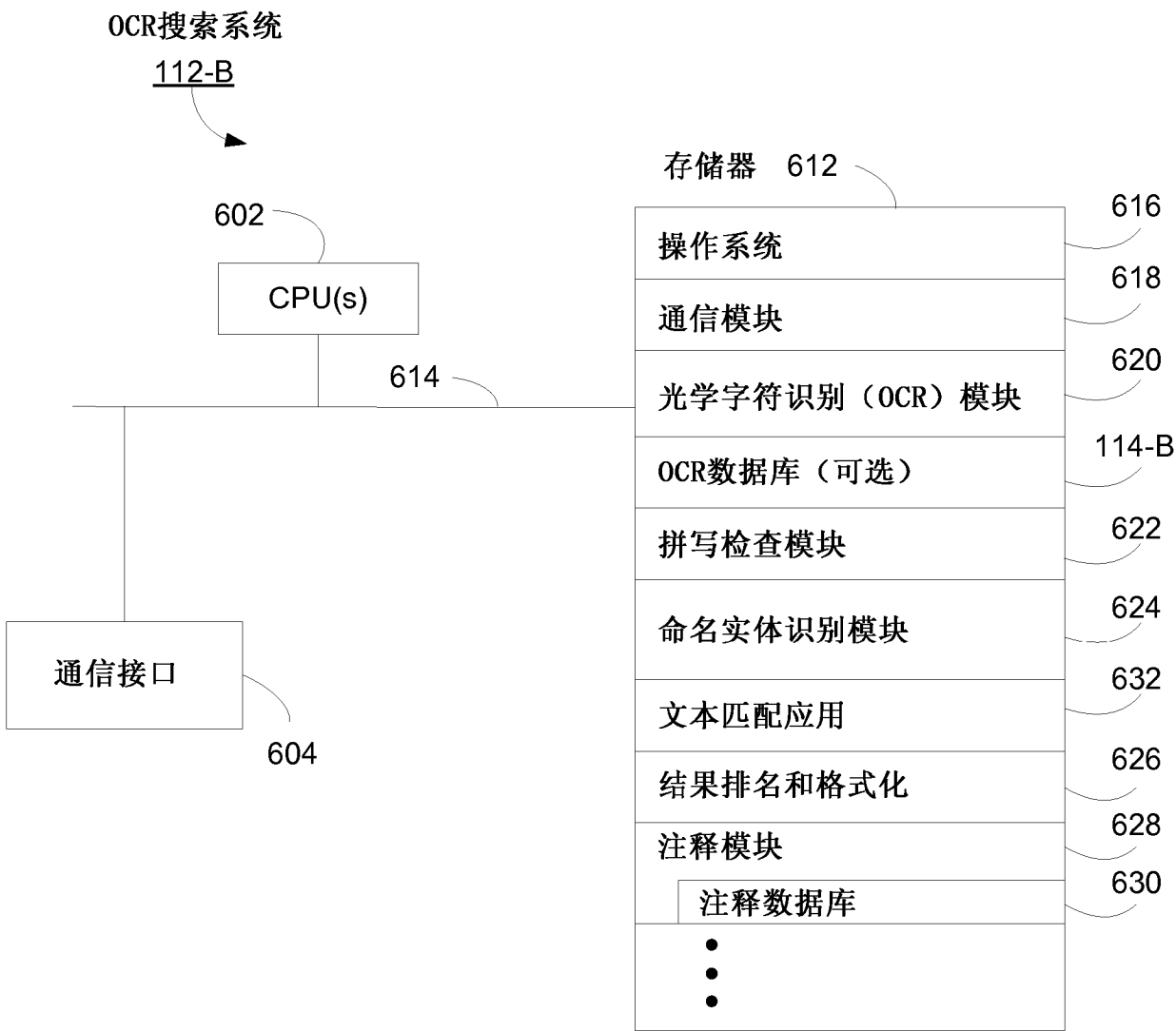


图 8

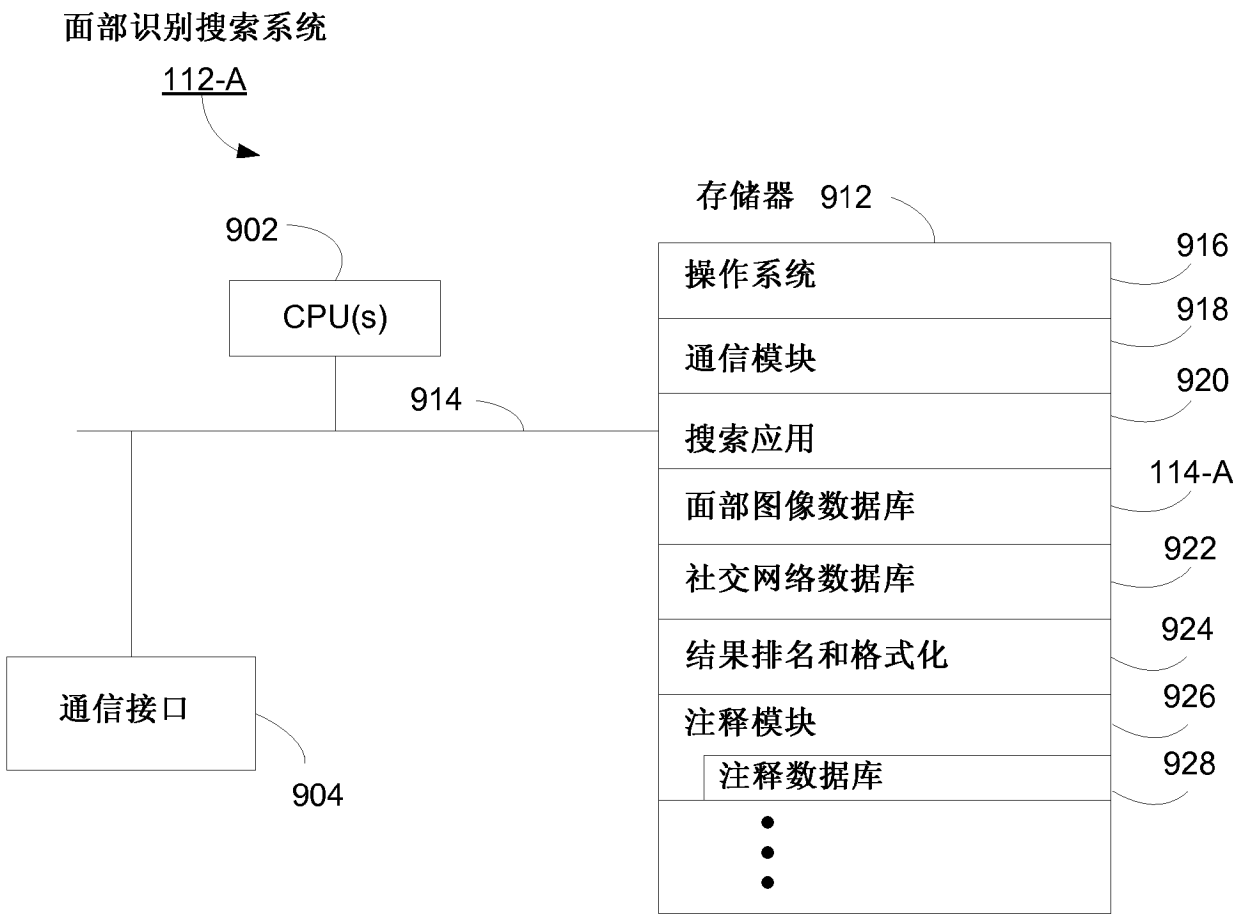


图 9

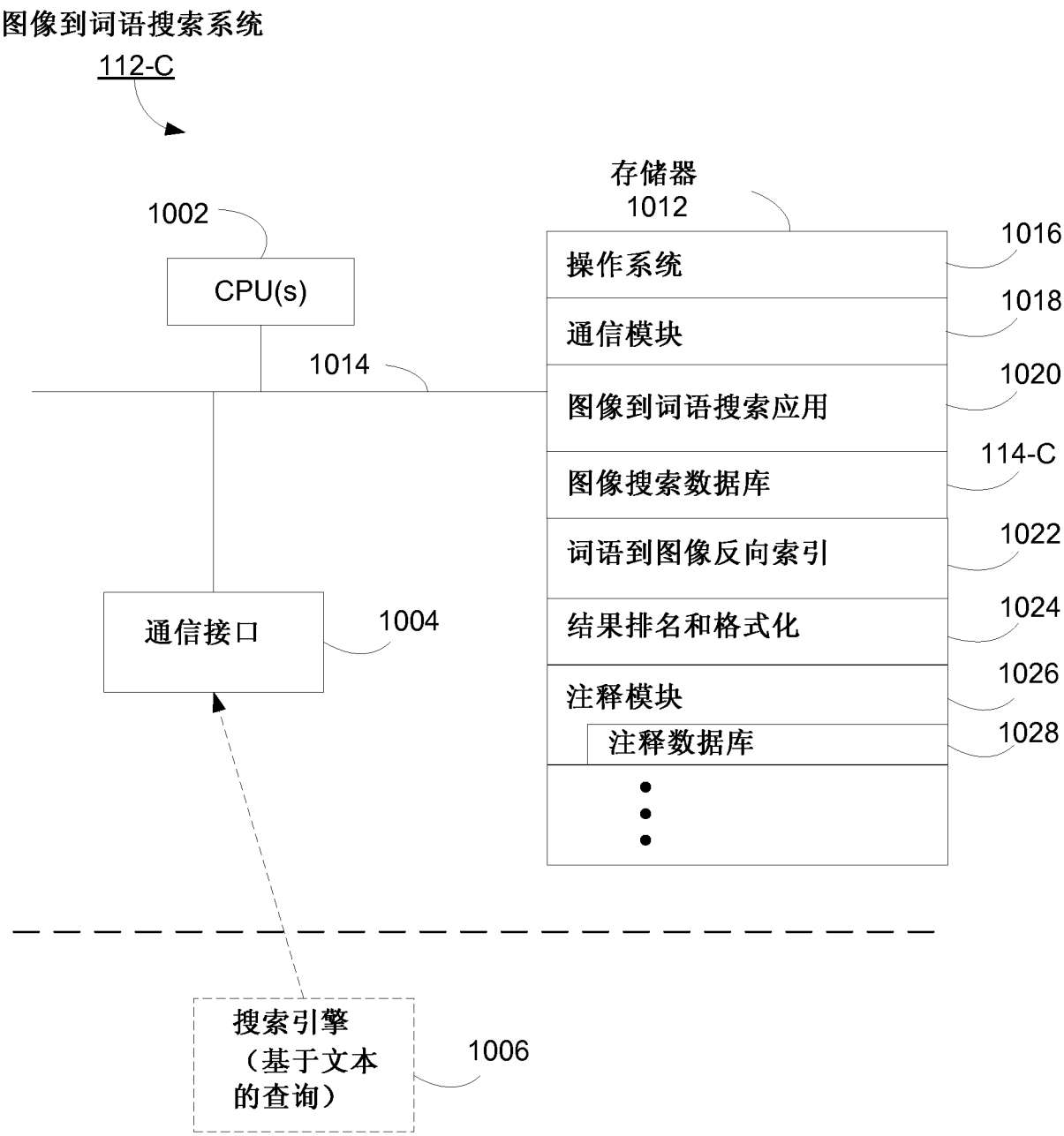


图 10

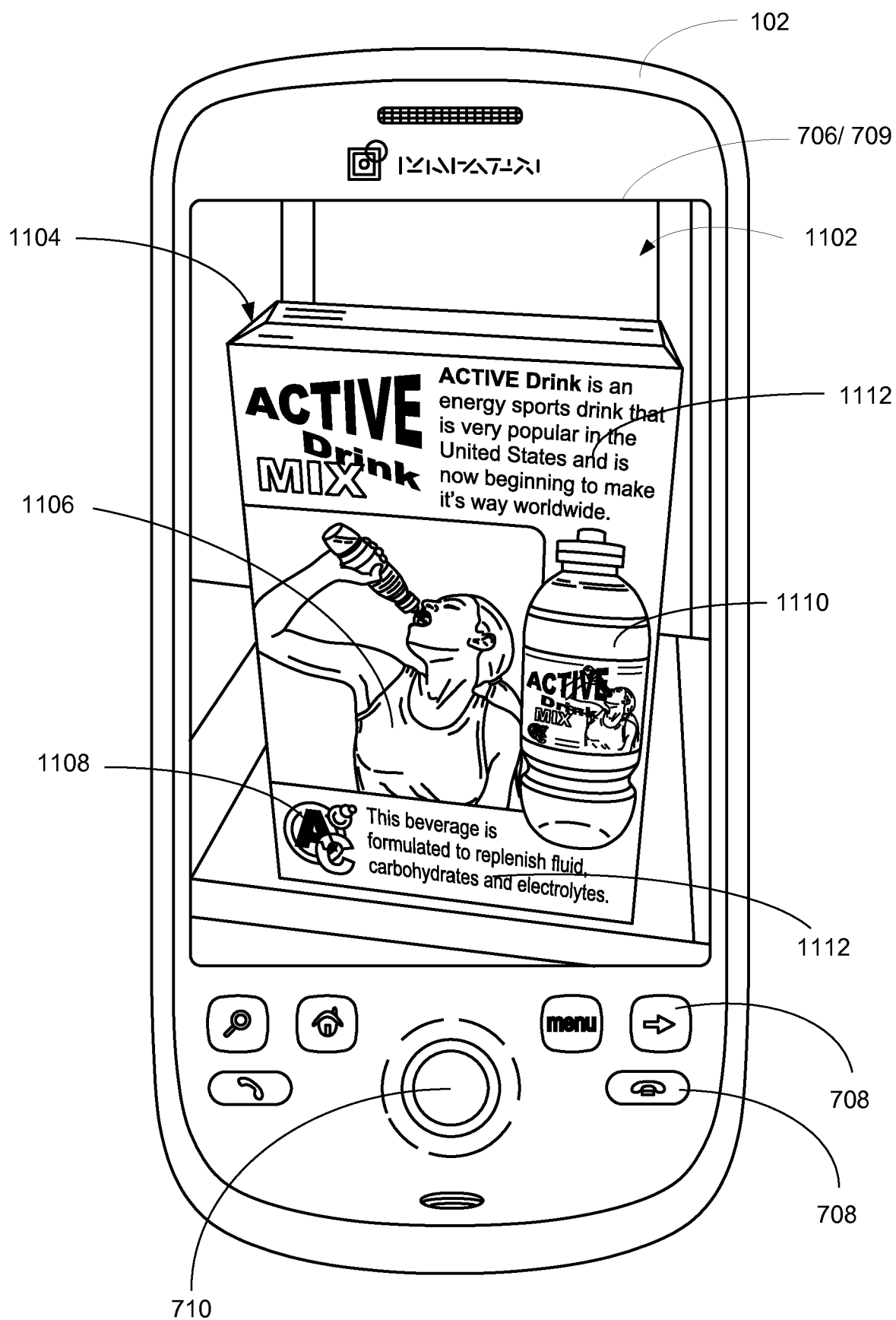


图 11

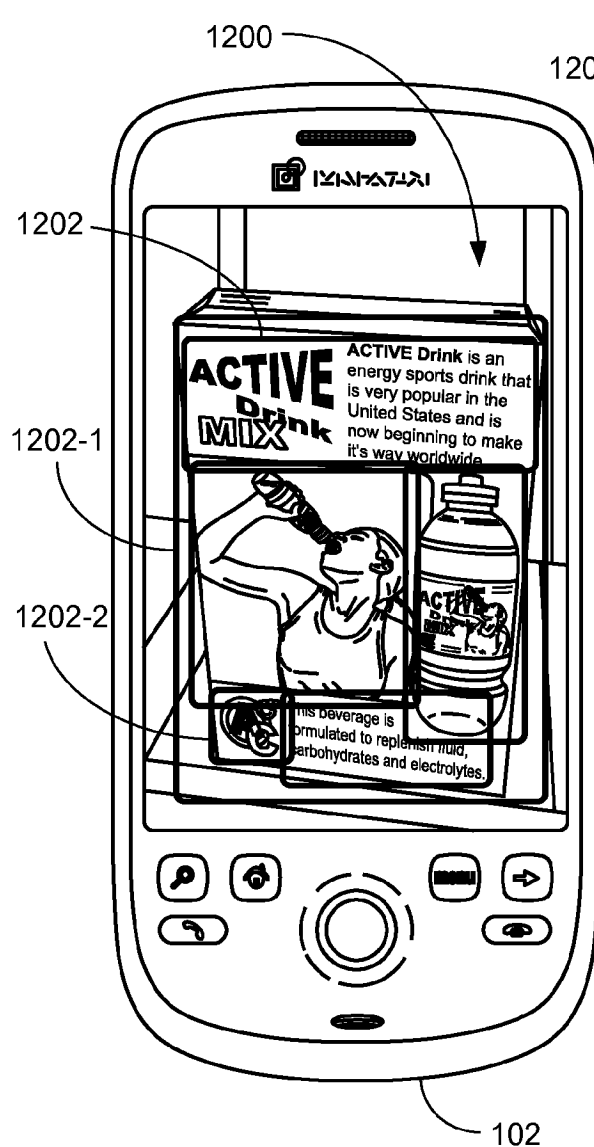


图12A

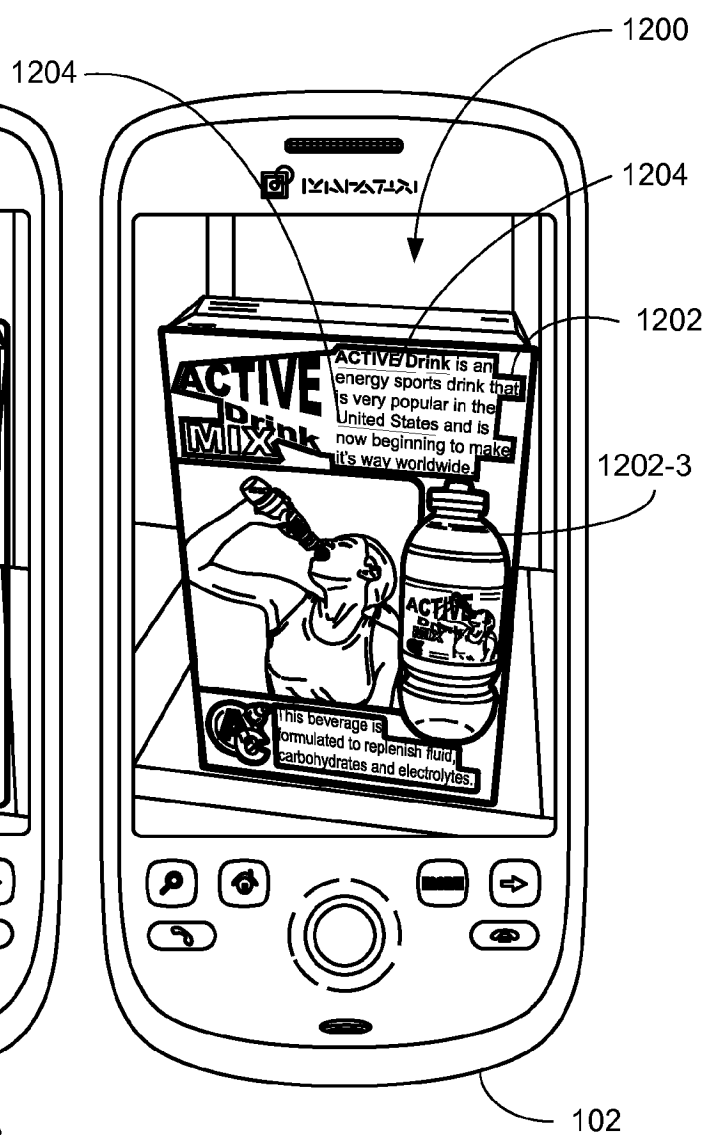


图12B

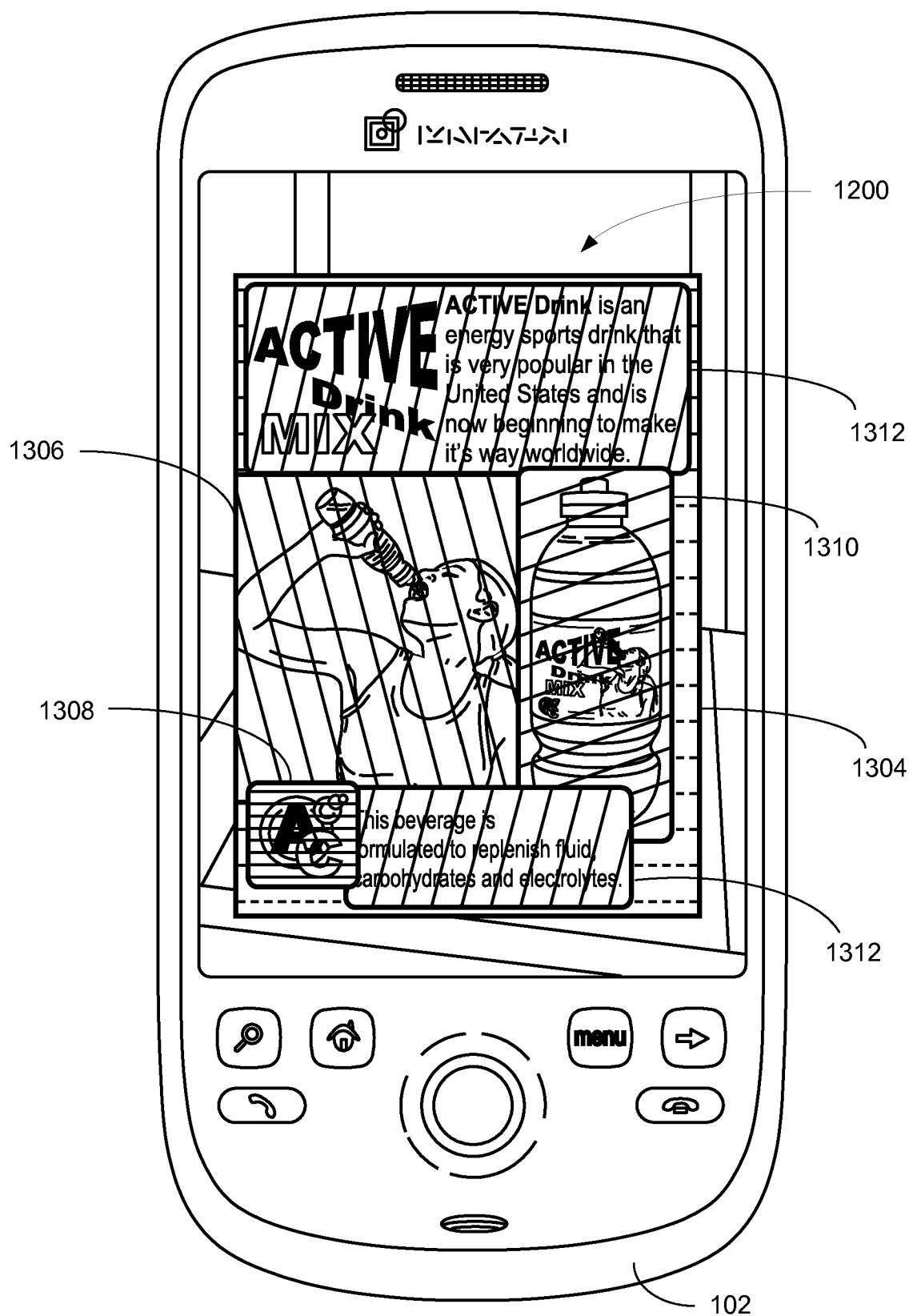


图 13

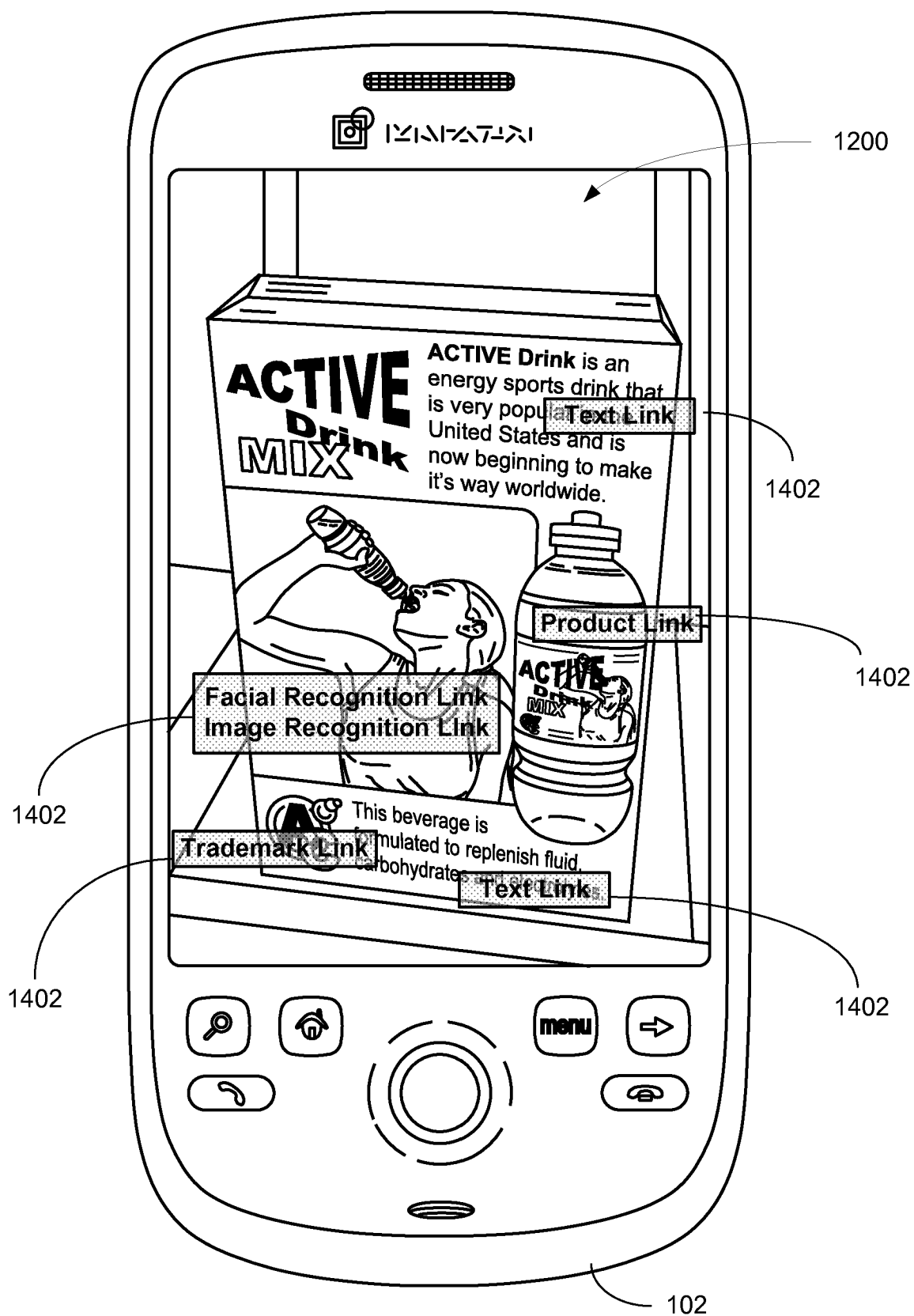


图 14

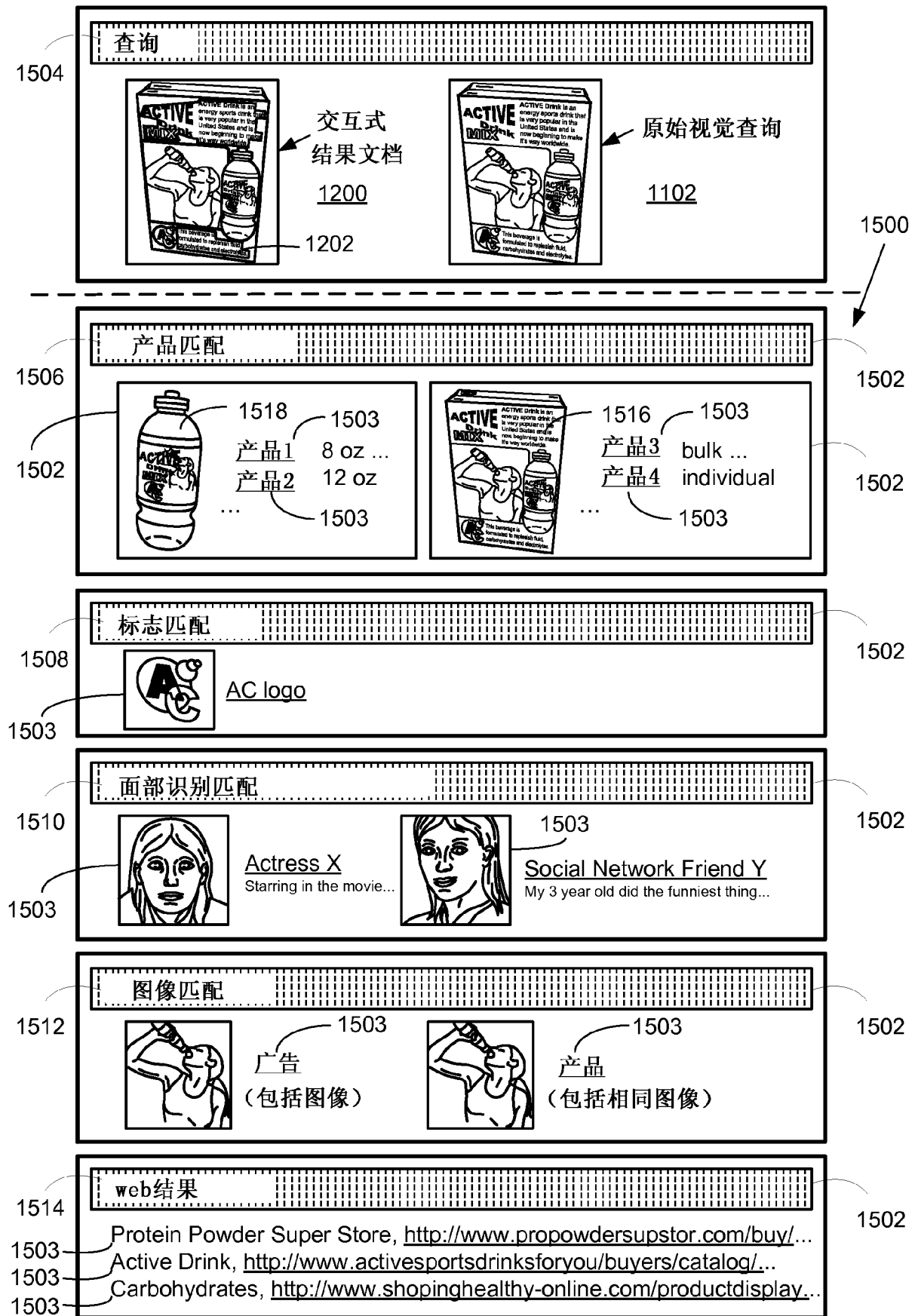


图 15