



(12) 发明专利申请

(10) 申请公布号 CN 103930903 A

(43) 申请公布日 2014. 07. 16

(21) 申请号 201280054713. 5

(22) 申请日 2012. 11. 01

(30) 优先权数据

13/290, 658 2011. 11. 07 US

(85) PCT国际申请进入国家阶段日

2014. 05. 07

(86) PCT国际申请的申请数据

PCT/FI2012/051062 2012. 11. 01

(87) PCT国际申请的公布数据

W02013/068638 EN 2013. 05. 16

(71) 申请人 诺基亚公司

地址 芬兰埃斯波

申请人 利兰斯坦福青年大学托管委员会

(72) 发明人 R·韦旦萨姆 R·格热茨祖克

D·M·陈 S-H·蔡 B·格罗德

(74) 专利代理机构 北京市中咨律师事务所

11247

代理人 张静美 杨晓光

(51) Int. Cl.

G06K 9/62 (2006. 01)

G06F 17/30 (2006. 01)

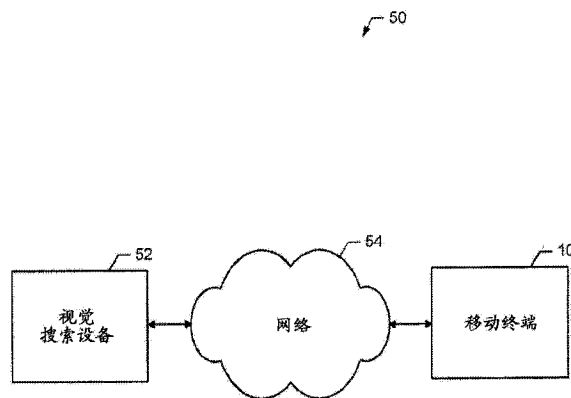
权利要求书2页 说明书11页 附图4页

(54) 发明名称

用于移动视觉搜索的方法和设备

(57) 摘要

本发明提供一种用于 REW 系统的方法、设备及计算机程序产品,所述 REW 系统配置为提供可在移动终端上操作的 MVS。一种示例方法可包括:利用从图像中提取出来的局部特征描述符使多个矢量词残留为至少一个视觉词聚合。所述方法可进一步包括:利用分类感知线性判别分析使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低。所述方法可进一步包括:利用处理器计算至少一个紧凑型图像签名的加权相关性,该图像签名在与候选者列表比较时从聚合后的至少一个矢量词残留二值化得到。所述方法可进一步包括:基于计算出的加权相关性确定候选者的排名列表。



1. 一种方法,其包括:

利用从图像中提取出来的局部特征描述符使多个矢量词残留为至少一个视觉词聚合;

利用分类感知线性判别分析使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低;

利用处理器使得计算至少一个紧凑型图像签名的加权相关性,该图像签名在与候选者列表比较时从聚合后的至少一个矢量词残留二值化得到;以及

基于计算出的加权相关性确定候选者的排名列表。

2. 根据权利要求1所述的方法,其进一步包括:将所述图像表示为一个或多个视觉词的矢量词残留,其中,将所述图像中的各个描述符量化为最接近的视觉词。

3. 根据权利要求1所述的方法,其进一步包括:使得聚合后的至少一个矢量词残留二值化,其中,所述二值化导致了产生紧凑型图像签名。

4. 根据权利要求1所述的方法,其中,基于所述矢量词残留的平均值或中间值中的至少一个值聚合所述矢量词残留。

5. 根据权利要求1所述的方法,其进一步包括:在通过丢弃与视觉词的距离超过预定百分位的特征来形成矢量词残留时,剔除异常特征。

6. 根据权利要求1所述的方法,其进一步包括:将冥次定律应用到聚合后的至少一个矢量词残留。

7. 根据权利要求1所述的方法,其中,所述加权相关性基于匹配似然比而加权。

8. 一种设备,其包括:

至少一个处理器;以及

至少一个存储器,所述存储器包括计算机程序代码,所述至少一个存储器和所述计算机程序代码配置为和所述至少一个处理器一起使所述设备执行至少以下步骤:

利用从图像中提取出来的局部特征描述符使多个矢量词残留为至少一个视觉词聚合;

利用分类感知线性判别分析使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低;

计算至少一个紧凑型图像签名的加权相关性,该图像签名在与候选者列表比较时从聚合后的至少一个矢量词残留二值化得到;以及

基于计算出的加权相关性确定候选者的排名列表。

9. 根据权利要求8所述的设备,其中,包括所述计算机程序代码的所述至少一个存储器进一步配置为:与所述至少一个处理器一起使所述设备将图像表示为一个或多个视觉词的矢量词残留,其中,将图像中的各个描述符量化为最接近的视觉词。

10. 根据权利要求8所述的设备,其中,包括所述计算机程序代码的所述至少一个存储器进一步配置为:与所述至少一个处理器一起使所述设备使得聚合后的至少一个矢量词残留二值化,其中,所述二值化导致了产生紧凑型图像签名。

11. 根据权利要求8所述的设备,其中,基于所述矢量词残留的平均值或中间值中的至少一个值聚合所述矢量词残留。

12. 根据权利要求8所述的设备,其中,包括所述计算机程序代码的所述至少一个存储

器进一步配置为：与所述至少一个处理器一起使所述设备在通过丢弃与视觉词的距离超过预定百分位的特征来形成矢量词残留时，剔除异常特征。

13. 根据权利要求 8 所述的设备，其中，包括所述计算机程序代码的所述至少一个存储器进一步配置为与所述至少一个处理器一起使所述设备将冥次定律应用于聚合后的至少一个矢量词残留。

14. 根据权利要求 8 所述的设备，其中，所述加权相关性基于匹配似然比而加权。

15. 一种计算机程序产品，其包括：

其上存储有程序代码的至少一个计算机可读非暂时性存储介质，所述程序代码在由设备执行时使所述设备执行至少以下步骤：

利用从图像中提取出的局部特征描述符使至少一个视觉词的至少一个矢量词残留聚合，其中，基于所述矢量词残留的平均值或中间值中的至少一个值聚合所述矢量词残留；

利用分类感知线性判别分析使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低；

计算至少一个紧凑型图像签名的加权相关性，该图像签名在与候选者列表比较时从聚合后的至少一个矢量词残留二值化得到；

基于计算出的加权相关性确定候选者的排名列表。

16. 根据权利要求 15 所述的计算机程序产品，其进一步包括：程序代码指令，所述程序代码指令配置为将图像表示为一个或多个视觉词的矢量词残留，其中，将图像中的各个描述符量化为最接近的视觉词。

17. 根据权利要求 15 所述的计算机程序产品，其进一步包括：程序代码指令，所述程序代码指令配置为将聚集后的至少一个矢量词残留二值化，其中，所述二值化导致了产生紧凑型图像签名。

18. 根据权利要求 15 所述的计算机程序产品，其进一步包括：程序代码指令，所述程序代码指令配置为在通过丢弃与视觉词的距离超过预定百分位的特征来形成矢量词残留时，剔除异常特征。

19. 根据权利要求 15 所述的计算机程序产品，其进一步包括：程序代码指令，所述程序代码指令配置为将冥次定律应用到聚合后的至少一个矢量词残留。

20. 根据权利要求 15 所述的计算机程序产品，其中，所述加权相关性基于匹配似然比而加权。

用于移动视觉搜索的方法和设备

技术领域

[0001] 本发明实施例大体上涉及视觉搜索技术,尤其涉及一种使用移动终端促进视觉搜索的方法、设备以及计算机程序产品。

背景技术

[0002] 随着移动终端的性能和处理能力持续发展,移动终端越来越多地用于之前仅为更大更少的移动装置保留的众多服务中。其中一种这类服务可包括基于捕获的图像而进行的视觉搜索和识别。

[0003] 移动视觉搜索 (MVS) 是指一类图像识别服务,在这类服务中,用户可捕获对象的图片以便接收关于该对象的有用信息。MVS 例如可用于识别户外地标、产品范围、酒水标签、印刷文献等。

[0004] 通常, MVS 系统采用收纳了基于视觉的搜索所用的多个图像、捕获媒介、视频等的大型远程数据库。为了搜索大型远程数据库以发现在视觉上与用户生成的查询图像相似的示例,一般使用词汇树 (VT)。VT 允许在查询图像和大型图像数据库之间进行快速比较。通常,需要几千兆字节的随机存取存储器 (RAM) 来表示各种数据结构以及与 VT 相关的图像签名。远程服务器通常用于这种目的,这是因为其具有大量的可用 RAM 并可承受典型视觉搜索系统对大存储器和大存储装置的要求。各个这类在先系统均依赖于大量的存储器和处理能力,以在执行视觉搜索时确保高水平的准确性。

发明内容

[0005] 本发明中提供一种用于紧凑型残留增强视觉矢量 (REVV) 系统的方法、设备以及计算机程序产品,该系统配置为使能设备上 (例如,移动终端) 的 MVS。根据本发明某些实施例的示例 REVV 可配置为:为查询图像形成紧凑型图像签名,然后将紧凑型图像签名与存储在本地数据库中的图像签名进行比较,以产生候选者的排名列表。然后,本发明所述的系统和方法可将候选者的排名列表显示在用户接口上和 / 或检索关于排名首位的候选者的有用信息。

[0006] 在一个实施例中,提供了一种方法,所述方法包括:利用从图像中提取出来的局部特征描述符使多个矢量词残留为至少一个视觉词聚合。该实施例的方法还可包括:利用分类感知线性判别分析使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低。该实施例的方法还可包括:利用处理器计算至少一个紧凑型图像签名的加权相关性,该图像签名在与候选者列表比较时从聚合后的至少一个矢量词残留二值化得到。该实施例的方法还可包括:基于计算出的加权相关性确定候选者的排名列表。

[0007] 在另一实施例中,提供了一种设备,所述设备包括:至少一个处理器;以及至少一个存储器,所述存储器包括计算机程序代码,所述至少一个处理器和所述计算机程序代码配置为和所述至少一个处理器一起使所述设备执行至少以下步骤:利用从图像中提取出来的局部特征描述符为至少一个视觉词至少使多个矢量词残留聚合,其中,所述矢量词残留

基于所述矢量词残留的平均值、中间值等聚合。所述至少一个存储器和计算机程序代码还可配置为与所述至少一个处理器一起使所述设备：利用分类感知线性判别分析使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低。所述至少一个存储器和计算机程序代码还可配置为与所述至少一个处理器一起使所述设备：利用处理器计算至少一个紧凑型图像签名的加权相关性，所述图像签名在与候选者列表比较时可从聚合后的至少一个矢量词残留二值化得到。所述至少一个存储器和计算机程序代码还可配置为与所述至少一个处理器一起使所述设备：基于计算出的加权相关性确定候选者的排名列表。

[0008] 在另一实施例中，可提供一种计算机程序产品，所述计算机程序产品包括：其中存储有计算机可读程序指令的至少一个非暂时性计算机可读存储介质，所述计算机可读程序指令包括配置为利用从图像中提取出的局部特征描述符使多个矢量词残留为至少一个视觉词聚合的程序指令，其中，所述矢量词残留基于所述矢量词残留的平均值、中间值等聚合。所述计算机可读程序指令还可包括配置为利用分类感知线性判别分析使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低的程序指令。所述计算机可读程序指令还可包括配置为利用处理器计算至少一个紧凑型图像签名的加权相关性的程序指令，该图像签名在与候选者列表比较时从聚合后的至少一个矢量词残留二值化得到。所述计算机可读程序指令还可包括配置为基于所述计算出的加权相关性确定候选者的排名列表的程序指令。

[0009] 在另一实施例中，提供了一种设备，所述设备包括：用于利用从图像中提取出来的局部特征描述符使多个矢量词残留为至少一个视觉词聚合的装置。该实施例的设备还可包括：用于利用分类感知线性判别分析使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低的装置。该实施例的设备还可包括：用于利用处理器计算至少一个紧凑型图像签名的加权相关性的装置，该图像签名在与候选者列表比较时从聚合后的至少一个矢量词残留二值化得到。该实施例的设备还可包括：用于基于计算出的加权相关性确定候选者的排名列表的装置。

附图说明

[0010] 由此，在已经用一般性术语描述了本发明的实施例之后，现在将对附图做出参考，所述附图不一定按照比例绘制，在附图中：

[0011] 图 1 举例说明了根据本发明示例实施例的示例视觉搜索设备的示例框图；

[0012] 图 2 是根据本发明示例实施例的示例移动终端的示例示意性框图；

[0013] 图 3 举例说明了根据本发明示例实施例的示例 Voronoi 单元、视觉词或图心、图像特征以及词残留矢量；

[0014] 图 4 举例说明了根据本发明示例实施例的示例用户接口；

[0015] 图 5 举例说明了根据本发明示例实施例的示例视觉搜索系统；以及

[0016] 图 6 举例说明了根据本发明示例实施例的用于视觉搜索的示例方法的流程图。

具体实施方式

[0017] 现在将在下文中参照附图对本发明的示例实施例进行更加充分的描述，在附图中示出了本发明的一些但并非所有的实施例。事实上，本发明的各种实施例可以通过很多不同的形式实现，并不应该被解释为仅限于这里所阐述的实施例。相同的附图标记始终指代

相同的元件。正如这里所使用的那样,术语“数据”“内容”“信息”以及类似的术语可以被可互换地用来指代根据本发明的各个实施例能够被传输、接收和 / 或存储的数据。而且,术语“示范性的”,正如这里所使用的,并不是用来表达任何的定性评价,而是仅仅表示对例子的说明。因此,不应该将任何此类术语的使用看作对本发明实施例的精神和范围的限制。

[0018] 如这里所使用的,术语“电路 (circuitry)”指代:(a) 只有硬件的电路实现方式(例如,以模拟电子线路和 / 或数字电子线路的形式实现);(b) 电路与包含存储在一个或多个计算机可读存储器上的软件和 / 或固件指令的计算机程序产品的组合,所述电路与计算机程序产品协同工作以使一装置执行这里所描述的一个或多个功能;以及 (c) 电路,比如例如微处理器或微处理器的一部分,其需要软件或固件以进行操作,即使软件或固件并不确实存在。

[0019] “电子线路”的这种定义适用于本发明所有使用该术语的情形,包括任何权利要求中使用该术语的情形。作为进一步的例子,正如这里所使用的,术语“电子线路”还包括包含一个或多个处理器和 / 或处理器部分以及随附的软件和 / 或固件的实现。为另一个例子,这里所使用的术语“电子线路”还例如包括用于移动电话的基带集成电路或应用处理器集成电路,或者服务器、蜂窝网络设备、其它网络设备和 / 或其它计算设备中的类似的集成电路。

[0020] 图 1 举例说明了根据本发明示例实施例的利用 REVV 的 MVS 系统的视觉搜索设备 102 的框图, REVV 配置为使用图像或一系列图像(例如,媒体剪辑、视频、视频流等)搜索图像或一系列图像的数据库。图 1 的示例 REVV 有利地配置为通过使用平均值、中间值等类似聚合提供残留聚合来执行 MVS。示例 REVV 进一步配置为通过在矢量量化期间丢弃不稳定的特征来执行异常剔除 (outlier rejection)。通过使用线性判别分析取代主成分分析,示例 REVV 还可执行分类感知降维。示例 REVV 可进一步基于压缩域中的图像签名之间的相关性来执行可区分性加权;有利地,因为这些增强,例如, REVV 实现与 VT 相似的检索 (retrieval) 性能,但利用未压缩的和压缩的倒排索引却使用了比 VT 更少的存储器。

[0021] 要理解,视觉搜索设备 102 被提供为本发明的一个实施例的示例,不应被理解为以任何方式使本发明的范围或精神变窄。关于这一点,本公开的范围除了本发明所举例说明和描述的实施例之外,还包括许多潜在的实施例。同样,虽然图 1 举例说明了用于 MVS 的设备配置的一个示例,但其它配置也可用于实施本发明的实施例。

[0022] 视觉搜索设备 102 可体现为:台式电脑、手提电脑、移动终端、移动电脑、平板电脑、移动电话、移动通信装置、一个或多个服务器、一个或多个网络节点、游戏装置、数字摄像机 / 摄录像机、音频 / 视频播放器、电视装置、无线电接收器、数字视频录像机、定位装置、及其任意组合,等等。在一个示例实施例中,视觉搜索设备 102 可体现为移动终端,比如图 2 所示的移动终端。

[0023] 关于这一点,图 2 举例说明了代表视觉搜索设备 102 的一个实施例的移动终端 10 的框图。然而,应理解,图中举例说明且下文所述的移动终端 10 仅仅是可实施本发明实施例和 / 或从本发明实施例中受益的一类视觉搜索设备 102,因此,不应认为对本发明的范围构成了限制。而移动终端的多个实施例(例如,移动终端 10),且出于示例的目的,下文对这些实施例进行了描述,其它类型的移动终端,比如:移动电话、移动电脑、便携式数字助手(PDA)、寻呼器、手提电脑、台式电脑、游戏装置、电视以及其它类型的电子系统,均可采用本

发明的实施例。

[0024] 如图所示,移动终端 10 可包括与发射机 14 和接收器 16 通信的天线 12(或多根天线 12)。移动终端 10 还可包括处理器 20,该处理器 20 配置为分别从发射机和接收器提供信号和接收信号。例如,处理器 20 还可体现为多种装置,包括:电路、一个或多个带有伴随数字信号处理器的微处理器、一个或多个无伴随数字信号处理器的处理器、一个或多个协处理器、一个或多个多核处理器、一个或多个控制器、处理电路、一个或多个计算机、各种包括例如 ASIC(应用专用集成电路)或 FPGA(现场可编程门阵列)等集成电路的其它处理元件、或其某些组合。相应地,虽然图 2 中举例说明的是一个处理器,但是在某些实施例中,处理器 20 包括多个处理器。由处理器 20 发送和接收的信号可包括:与适用的蜂窝系统的空中接口标准和 / 或任意数量的不同的有线或无线网络技术一致的信令信息,包括但不限于:无线保真(Wi-Fi)和无线局域网(WLAN)技术(比如,电子与电气工程师协会(IEEE)802.11、802.16 等等)。另外,这些信号可包括语音数据、用户生成数据、用户请求数据,等等。关于这一点,移动终端能够根据一个或多个空中接口标准、通信协议、调制类型、存取类型等等运行。更具体地,移动终端 10 能够依照以下各种协议运行:第一代(1G)、第二代(2G)、2.5G、第三代(3G)通信协议、第四代(4G)通信协议、互联网协议多媒体子系统(IMS)通信协议(例如,会话初始化协议(SIP))等等。例如,移动终端能够依照以下运行:2G 无线通信协议 IS-136(时分多址(TDMA))、全球移动通信系统(GSM)、IS-95(码分多址(CDMA))等等。同样,例如,移动终端能够依照以下运行:2.5G 无线通信协议通用无线分组业务(GPRS)、加强型数据 GSM 环境(EDGE)等等。进一步地,例如,移动终端能够依照 3G 无线通信协议运行,比如:通用移动通信系统(UMTS)、码分多址 2000(CDMA2000)、宽带码分多址(WCDMA)、时分-同步码分多址(TD-SCDMA)等等。此外,移动终端能够依照 3.9G 无线通信协议运行,比如:长期演进(LTE)或演进通用陆地无线接入网络(E-UTRAN)等等。此外,例如,移动终端能够依照诸如第四代(4G)无线通信协议等以及未来可产生的相似无线通信协议运行。

[0025] 某些窄带高级移动电话系统(NAMPS)以及全接入通信系统(TACS),移动终端还可从本发明的实施例中获益,正如双模或更高模式的电话(例如,数字/模拟或 TDMA/CDMA/模拟电话)应从中获益一样。此外,移动终端 10 能够根据无线保真(Wi-Fi)或全球微波互连接入(WiMAX)协议运行。

[0026] 要理解,处理器 20 可包括用于执行音频/视频和移动终端 10 的逻辑功能的电路。例如,处理器 20 可包括:数字信号处理器装置、微处理器装置、模数转换器、数模转换器等。可根据这些装置各自的能力将移动终端 10 的控制和信号处理功能分配给这些装置之间。进一步地,处理器可包括运行一个或多个可存储在存储器中的软件程序的功能。例如,处理器 20 能够运行连接程序,比如网络浏览器。连接程序可使移动终端 10 根据无线应用协议(WAP)、超文本传送协议(HTTP)等协议发送和接收网页内容,比如基于位置的内容。移动终端 10 能够使用传输控制协议/互联网协议(TCP/IP)在互联网或其它网络间发送和接收网页内容。

[0027] 移动终端 10 还可包括用户接口,所述用户接口包括:例如,耳机或扬声器 24、振铃器 22、麦克风 26、显示器 28、用户输入接口等,上述可以可操作地耦合到处理器 20。关于这一点,处理器 20 可包括用户接口电路,该用户接口电路配置为控制用户接口的一个或多个元件(比如,例如扬声器 24、振铃器 22、麦克风 26、显示器 28 等)的至少某些功能。处理器

20 和 / 或包括处理器 20 的用户接口电路可配置为 : 通过存储在可访问处理器 20 的存储器 (例如, 易失性存储器 40、非易失性存储器 42 等) 上的计算机程序指令 (例如, 软件和 / 或固件), 来控制用户接口的一个或多个元件的一个或多个功能。尽管未示出, 但是移动终端可包括 : 用于为与移动终端有关的各种电路 (例如, 提供机器振动作为可检测输出的电路) 提供电力的电池。用户输入接口可包括使移动终端接收数据的装置, 比如, 小键盘 30、触摸显示器 (未示出)、操纵杆 (未示出) 和 / 或其它输入装置。在包括小键盘的实施例中, 小键盘可包括数字 (0-9) 及相关的键 (#、*), 和 / 或用于操作移动终端的其它键。

[0028] 移动终端 10 可包括与处理器 20 通信的媒体捕获元件, 比如, 摄像机、视频和 / 或音频模块。媒体捕获元件可包括用于捕获图像、视频和 / 或音频进行视觉搜索、存储、显示或传输的任意装置。例如, 在媒体捕获元件包括摄像机电路 36 的一个示例实施例中, 摄像机电路 36 可包括配置为从捕获的图像中形成数字图像文件的数字摄像机。另外, 摄像机电路 36 的数字摄像机可配置为捕获视频剪辑。同样, 摄像机电路 36 可包括 : 所有硬件, 比如透镜或其它光学分量 ; 以及, 用于从捕获的图像和来自捕获的视频剪辑的数字视频文件中创建数字图像文件所必须的软件。可替代地, 摄像机电路 36 可仅包括观看图像所需的硬件, 而移动终端 10 的存储器装置存储由从捕获的图像中创建数字图像文件所需的软件形式的处理器 20 执行的指令。作为另一替代方案, 可将摄像机电路 36 视野内的一个对象或多个对象显示在移动终端 10 的显示器 28 上以举例说明当前显示的图像视图, 该图像视图可在用户需要的情况下被捕获。同样, 捕获的图像可例如包括由摄像机电路 36 捕获的图像以及存储在图像文件中的图像。再如, 捕获的图像可包括当前由显示器或移动终端 10 的 viewfinder 显示的一个对象或多个对象, 但不必存储在图像文件中。在一个示例实施例中, 摄像机电路 36 可进一步包括处理元件, 比如 : 配置为协助处理器 20 处理图像数据的协同处理器以及用于压缩和 / 或解压缩图像数据的编码器和 / 或解码器。编码器和 / 或解码器可根据例如联合图像专家组 (JPEG) 标准、移动图片专家组 (MPEG) 标准或其它格式来进行编码和 / 或解码。

[0029] 移动终端 10 可包括可存储与移动用户相关的信息元素的存储器, 比如用户身份模块 (SIM) 38、可移除用户身份模块 (R-UIM) 等。除 SIM 之外, 移动终端可包括其它可移除的和 / 或固定的存储器。移动终端 10 可包括其它非暂时性存储器, 比如易失性存储器 40 和 / 或非易失性存储器 42。例如, 易失性存储器 40 可包括随机存取存储器 (RAM), 包括 : 动态和 / 或静态 RAM、片上或片外缓存存储器等。可为嵌入式和 / 或可移除的非易失性存储器 42 可包括, 例如, 只读存储器、闪存存储器、磁存储器装置 (例如, 硬盘、软盘驱动器、磁带等)、光盘驱动器和 / 或媒体、非易失性随机存取存储器 (NVRAM) 等。类似于易失性存储器 40, 非易失性存储器 42 可包括用于暂时存储数据的缓存区域。存储器可存储一个或多个可由移动终端用于执行移动终端的功能的软件程序、指令、信息、数据等。例如, 存储器可包括能够唯一地识别移动终端 10 的国际移动设备标识 (IMEI) 码等标识符。

[0030] 再次参见图 1, 在一个示例实施例中, 视觉搜索设备 102 包括用于执行本发明所述各种功能的各种装置。这些装置可包括处理器 110、存储器 112、通信接口 114、用户接口 116、图像捕获电路 118 和 / 或 REVV 模块 120 中的一个或多个。视觉搜索设备 102 的装置如本发明所述可具体体现为, 例如, 电路、硬件元件 (例如, 适当编程的处理器、组合逻辑电路等)、计算机程序产品 (其包括存储在计算机可读媒体 (例如存储器 112) 中可由合适配

置的处理装置（例如，处理器 110）执行的计算机可读程序指令（例如，软件或固件）、或其某些组合。

[0031] 例如，处理器 110 可具体体现为各种装置，包括：一个或多个带有伴随数字信号处理器的微处理器、一个或多个没有伴随数字信号处理器的处理器、一个或多个协处理器、一个或多个多核处理器、一个或多个控制器、处理电路、一个或多个计算机、各种包括比如集成电路的其它处理元件（例如，ASIC 或 FPGA）或其某些组合。相应地，尽管在图 1 中举例说明的是单个处理器，但是在某些实施例中，处理器 110 包括多个处理器。多个处理器彼此之间可有效通信，并且可共同配置为执行本发明所述的视觉搜索设备 102 的一个或多个功能。多个处理器可具体体现在单各计算装置上或分布在多个共同配置以用作视觉搜索设备 102 的计算装置上。在视觉搜索设备 102 具体体现为移动终端 10 的实施例中，处理器 110 可具体体现为或包括处理器 20。在一个示例实施例中，处理器 110 配置为执行存储在存储器 112 中的指令，或者可接入处理器 110。这些指令在由处理器 110 执行时，可使视觉搜索设备 102 执行本发明所述的一个或多个功能。同样，不论是否由硬件或软件方法或其组合配置，处理器 110 可包括：在相应配置时能够根据本发明的实施例执行操作的实体。因此，例如，在处理器 110 具体体现为 ASIC、FPGA 等时，处理器 110 可包括用于执行本发明所述的一个或多个操作而具体配置的硬件。可选择地，如另一示例，当处理器 110 具体体现为指令的执行器时，比如可存储在存储器 112 中的指令的执行器，那么指令可具体配置处理器以执行本发明所述的一个或多个算法和操作。

[0032] 存储器 112 可包括例如非暂时性存储器，比如易失性存储器、非易失性存储器或其某些组合。尽管在图 1 中举例说明为单个存储器，但是存储器 112 可包括多个存储器。多个存储器可具体表现在单个计算装置上或可分布在多个共同配置以起如视觉搜索设备 102 的作用的计算装置上。在各种示例实施例中，存储器 112 可包括，例如：硬盘、随机存取存储器、缓存存储器、闪存存储器、压缩光盘只读存储器（CD-ROM）、数字通用光盘只读存储器（DVD-ROM）、光碟、配置为存储信息的电路、或其某些组合。在视觉搜索设备 102 具体体现为移动终端 10 的实施例中，存储器 112 可包括易失性存储器 40 和 / 或非易失性存储器 42。存储器 112 可配置为存储信息、数据、应用、指令等以使视觉搜索设备 102 能够执行根据各种示例实施例的各种功能。例如，在至少某些实施例中，存储器 112 配置为缓冲输入数据以便处理器 110 处理。另外或可选择地，在至少某些实施例中，存储器 112 配置为存储程序指令以便处理器 110 执行。存储器 112 可以静态和 / 或动态信息的形式存储信息。存储的信息可包括，例如：用于视觉搜索等的模式。存储的信息可被图像捕获电路 118 和 / 或 REVV 模块 120 在执行其功能的期间存储和 / 或使用。存储器 112 还可配置为存储一个或多个可由 REVV 模块 120 存取的图像和 / 或图像签名的数据库。数据库可利用通信接口 114 基于分配、时间等进行更新。

[0033] 通信接口 114 可具体体现为任意装置或设备，其体现为：电路、硬件、计算机程序产品（其包括存储在计算机可读媒体（例如，存储器 112）中的计算机可读程序指令并可由处理装置（例如，处理器 110）执行），或其配置为接收和 / 或传送数据至 / 从另一计算机装置的组合。例如，通信接口 114 可配置为接收将图像表现在网络上的数据。关于这一点，在实施例中，其中，视觉搜索设备 102 包括：服务器、网络节点等；通信接口 114 可配置为与远程移动终端（例如，远程终端 304）通信以使移动终端和 / 或其用户可接入由视觉搜索设

备 102 提供的视觉搜索功能。在示例实施例中,通信接口 114 至少部分地体现为或以其他方式由处理器 110 控制。关于这一点,通信接口 114 可与处理器 110 通讯,例如经由总线。通信接口 114 可包括,例如:天线、发射机、接收器、收发器和 / 或能够与一个或多个远程计算机装置通信的配套硬件或软件。通信接口 114 可配置为利用任意协议接收和 / 或传送数据,该协议可用于在计算装置间通信。关于这一点,通信接口 114 可配置为利用任意协议接收和 / 或传送数据,该协议可用于在无线网络、有线网络及其某些组合中传送数据,或者类似的视觉搜索设备 102 和一个或多个计算装置可用于通信。此外,通信接口 114 可与存储器 112、用户接口 116、图像捕获电路 118 和 / 或 REVV 模块 120,比如经由总线进行通信。

[0034] 用户接口 116 可与处理器 110 通信以接收用户输入的指示和 / 或以将听觉上的、视觉上的、机器的或其它输出提供给用户。同样地,用户接口 116 可包括,例如:键盘、鼠标、操纵杆、显示器、触屏显示器、麦克风、扬声器、和 / 或其它输入 / 输出机构。在实施例中,其中,视觉搜索设备 102 具体体现为一个或多个服务器,可减少用户接口 116 的各方面,或者甚至可消除用户接口 116。用户接口 116 可与存储器 112、通信接口 114、图像捕获电路 118 和 / 或 REVV 模块 120 比如经由总线通信。

[0035] 图像捕获电路 118 可具体体现为各种装置,比如:电路、硬件、计算机程序产品(其包括存储在计算机可读媒体(例如,存储器 112)中的计算机可读程序指令并可由处理装置(例如,处理器 110)执行)、或其组合,在一个实施例中图像捕获电路 118 具体体现为或以其他方式由处理器 110 控制。在实施例中,其中,图像捕获电路 118 可体现为分离于处理器 110,图像捕获电路 118 可与处理器 110 通信。图像捕获电路 118 可进一步比如经由总线与存储器 112、通信接口 114、用户接口 116、和 / 或 REVV 模块 120 中的一个或多个装置通信。

[0036] 图像捕获电路 118 可包括配置为捕获图像的硬件。关于这一点,图像捕获电路 118 可包括:摄像机镜头、IR 镜头和 / 或其它用于捕获数字图像的光学分量。再如,图像捕获电路 118 可包括:电路、硬件、计算机程序产品、或其某些组合,其配置为指导用表现在视觉搜索设备 102 上或以其他方式与视觉搜索设备 102 可操作地连接的单独的摄像机模块进行图像的捕获。在实施例中,其中,视觉搜索设备 102 具体体现为移动终端 10,图像捕获电路 118 可包括摄像机电路 36。在实施例中,其中,视觉搜索设备 102 具体体现为一个或多个服务器或其它远离移动终端的网络节点,该移动终端配置为将图像或视频提供给视觉搜索设备 102 以使视觉搜索设备 102 能够在图像或视频上执行视觉搜索,可减少图像捕获电路 118 的各方面或甚至可消除图像捕获电路 118。

[0037] REVV 模块 120 可具体体现为各种装置,比如:电路、硬件、计算机程序产品(其包括存储在计算机可读媒体(例如,存储器 112)中的计算机可读程序指令并可由处理装置(例如,处理器 110)执行)、或其某些组合,在一个实施例中 REVV 模块 120 具体体现为或者其他方式由处理器 110 控制。在实施例中,其中,REVV 模块 120 体现为分离于处理器 110,REVV 模块 120 可与处理器 110 通信。REVV 模块 120 可进一步比如经由总线与存储器 112、通信接口 114、用户接口 116、和 / 或图像捕获电路 118 中的一个或多个装置通信。

[0038] REVV 模块 120 可配置为为查询的图像形成紧凑型图像签名,然后,将紧凑型图像签名与比如,图像签名(例如存储在存储器 112 中的图像签名)的数据库进行比较。在某些实施例中,紧凑型图像签名通过将聚合的且维度降低的词残留的集合二值化而生成。

[0039] 在某些示例实施例中,REVV 模块 120 配置为将一个或多个从图像中提取出的局部特征描述符量化至最接近的矢量词。将预定数量(例如 128)的矢量词存储在例如存储器 112 中。然后,局部特征可具有矢量词残留,该矢量词残留可为局部特征描述符和最接近的矢量词之间的差值。然后,矢量词残留可通过以下来聚合:丢弃异常的局部特征异常残留;计算矢量词残留的矢量平均值、中间值等;和/或使用冥次定律调整。

[0040] 在另一示例实施例中,REVV 模块 120 可通过执行线性判别分析(LDA)(例如,考虑分类性能的转换)使矢量词的维度降低,然后可进一步地使矢量词残留二值化。二值化签名之间的汉明间距可使用二进制运算 XOR 和/或 POPCOUNT 运算计算得出。然后,可根据匹配/不匹配的似然比来加权这些间距以进一步地加强 REVV 图像签名的区别能力。

[0041] 在某些示例实施例中,REVV 模块 120 可配置为聚合矢量词残留。例如,使得 $c_1 \dots c_k$ 为 d 维视觉词的集合。在将图像中的各个描述符量化为最接近的视觉词之后,矢量词残留的集合可包围各个视觉词。例如,让 $NN(c_i)$ 表示第 i 个视觉词周围的残留集合。为了聚合该残留,有几种不同的方法是可能的,例如:

[0042] 求和聚合:此处,第 i 个视觉词的聚合后的残留可表示为:

$$[0043] \quad a_i = \sum_{v \in NN(c_i)} v;$$

[0044] 平均值聚合:在某些示例实施例中,残留和由 $NN(c_i)$ 的基数标准化,因此,聚合后的残留变为:

$$[0045] \quad a_i = \frac{1}{|NN(c_i)|} \sum_{v \in NN(c_i)} v;$$

[0046] 中间值聚合:在某些示例实施例中,可沿着各个维度确定中间值。

[0047] $a_i(n) = \text{median}(v(n) : v \in NN(c_i))$;

[0048] 例如,通过使用平均值、中间值等聚合方式,至少一个视觉词的多个矢量词残留可通过利用从图像中提取出的局部特征描述符聚合。在某些示例实施例中, S 可为聚合后的词残留的串联: $s = [a_1 \dots a_k]$ 。然后,图像签名 $\bar{s}$ 可形成为 $\bar{s} = S/\|S\|$ 。为了比较两个标准化的图像签名 $\bar{s}_q$ 和 $\bar{s}_d$,可计算得出其欧几里德距离 $\|\bar{s}_q - \bar{s}_d\|$,比如通过处理器 110 计算得出,或等价地还可计算得出内积 $\langle \bar{s}_q, \bar{s}_d \rangle$ 。

[0049] 在某些示例实施例中,REVV 模块 120 可配置为剔除异常特征。例如,接近两个 Voronoi 单元之间的边界的某些特征减少聚合后的残留的重复性。又如,该特征与图 3 所示 c_1 和 c_3 的 Voronoi 单元之间的界限非常接近。例如,即使很少量的噪音也会使该特征被量化为 c_3 而不是 c_1 ,这会明显改变 $NN(c_i)$ 和 $NN(c_3)$ 的构成,由此改变聚合后的残留 a_1 和 a_3 。

[0050] 因此,REVV 模块 120 可配置为剔除异常特征,例如,通过剔除离视觉词最远的这些特征。可替换地或另外地,也可移除超过预定阈值(例如,百分位)的特征。通过移除超过距离分布上的第 C 个百分位的特征,可移除多数异常特征。在某些示例实施例中,第 C 个百分位水平对于各种视觉词而有所不同,因为距离分布一般不同,所以不同的阈值可用于各个视觉词。

[0051] 在某些示例实施例中, REVV 模块 120 可配置为将冥次定律应用于视觉词残留中。在这些实施例中, 应用了冥次定律, 指数 α 的值在冥次定律中可为 $\alpha = 0.4$ 。

[0052] REVV 模块 120 还可配置为利用分类感知线性判别分析 (LDA) 使用于各个视觉词的聚合后的至少一个矢量词残留的维度降低。例如, 在实际增强检索性能的同时, 使用 LDA 可使图像签名的维度降低一半。因为残留矢量的维度与数据库索引的大小成比例, 所以在对检索性能不产生反作用的情况下需要降低维度。在某些示例实施例中, 为各个视觉词应用不同的 LDA 转换。例如, 为了使类之间的方差和投射在方向 w 上的类之间的方差之间的方差比最大化, 可使用以下方程式:

[0053] S_j = 来自图像 j 的词残留

[0054] J_M = 图像 j_1 和 j_2 匹配

[0055] J_{NM} = 图像 j_1 和 j_2 不匹配

$$[0056] \quad \underset{w}{\text{maximize}} \quad \frac{\sum_{(j_1, j_2) \in J_{NM}} (w^T (S_{j_1} - S_{j_2}))^2}{\sum_{(j_1, j_2) \in J_M} (w^T (S_{j_1} - S_{j_2}))^2}$$

[0057] 为降低维度, 可在某些示例实施例中使用以下解法:

[0058] $R_{NM} w_i = \lambda_i R_M w_i$

[0059] $i = 1, 2, \dots, d_{LDA}$

$$[0060] \quad R_M = \sum_{(j_1, j_2) \in J_M} (S_{j_1} - S_{j_2})(S_{j_1} - S_{j_2})^T$$

$$[0061] \quad R_{NM} = \sum_{(j_1, j_2) \in J_{NM}} (S_{j_1} - S_{j_2})(S_{j_1} - S_{j_2})^T$$

[0062] 在某些示例实施例中, REVV 模块 120 配置为取决于符号将残留矢量词的各个分量二值化为 +1 或 -1。有正负之分的二值化可创建最多仅需要 $k \cdot d_{LDA}$ 位的紧凑型图像签名。

例如, 有正负之分的二值化的另一优势在于可快速地评分计算。内积 $\langle \bar{s}_q, \bar{s}_d \rangle$ 可利用以下表达式来最大地接近:

$$[0063] \quad \frac{1}{\|\bar{s}_q\|_1 \|\bar{s}_d\|_1} \sum_{i: \text{visited by } Q \text{ and } D} c(s_{q,i}^{bin}, s_{d,i}^{bin}),$$

[0064] 其中, $c(s_{q,i}^{bin}, s_{d,i}^{bin})$ 为二进制相关系数、 $H(A, B)$ 为 A 与 B 之间的汉明间距、 $s_{q,i}^{bin}$ 和 $s_{d,i}^{bin}$ 为用于查询的二值化残留和第 i 个视觉词处的数据库图像。在某些示例实施例中, 可通过比如处理器 110 利用二进制算法 XOR 快速计算出汉明间距。

[0065] 在某些示例实施例中, REVV 模块 120 可配置为基于二值化签名之间计算出的相关性实施可区分性加权。示例加权函数可包括:

$$[0066] \quad w(C) = \frac{P(C|match)}{P(C|match) + P(C|non-match)}$$

[0067] 假设 $P(match) = P(non-match)$ 那么 $w(C) = P(match|C)$ 。在某些示例实施例中,

使用这种加权函数可将分数更改为：

$$[0068] \quad \frac{1}{\|s_q\|_2 \|s_d\|_2} \sum_{i \text{ visited by Q and D}} c(s_{q,i}^{bin}, s_{d,i}^{bin})$$

[0069] REVV 模块 120 可进一步配置为基于 REVV 图像签名产生数据库候选者的排名列表。然后,可经由例如用户接口 116 来显示结果。

[0070] 图 4 举例说明了示例用户接口,比如在示例移动终端 10 上运行的用户接口 116,该用户接口举例说明了由例如图像捕获电路 118 捕获的图像。在某些示例实施例中,存储器 112 可包含多个图像的数据库。存储在示例移动终端 10 的存储器 112 中的数据库可将以下特征的非详细列表、构建在局部附近的图像表示为由 GPS 确定的、著名地标的图像等。然后,在比如通过处理器 110 使用比如 viewfinder 中的图像捕获电路 118 的内容来查询数据以检测到低运动周期的示例中,可激活 REVV 模块 120 以执行视觉搜索。进一步地,使用处理器 110、REVV 模块 120 等的视觉搜索设备 102 一旦激活,则可产生地标的名称地址、电话号码,该地标确定为与图像捕获电路 118(例如,图像查询)捕获的地标相匹配。用户接口可包括小型地图,该小型地图可选择,从而可观察其位置。

[0071] 如结合图 1 的实施例所述,移动终端 10 可包括视觉搜索设备 102。然而,视觉搜索设备 102 的部分还可与移动终端 10 分开并通信,例如图像、图像签名等。图 5 举例说明了根据本发明的示例实施例执行视觉搜索的系统 50。系统 50 包括视觉搜索设备 52 和配置为通过网络 54 通信的移动终端 10。例如,视觉搜索设备 52 可包括视觉搜索设备 102 的实施例,其中,视觉搜索设备 52 具体体现为一个或多个服务器、一个或多个网络节点、云计算系统等,且配置为接收由例如 REVV 模块 120 生成的 REVV 图像签名,且进一步配置为在存储在视觉搜索设备 52 上的一个或多个图像上执行低比特率视觉查询。移动终端 10 可包括任意移动终端,该移动终端配置为接入网络 54 并与视觉搜索设备 52 通信,以发送 REVV 图像签名并接收视觉搜索结果。在某些示例实施例中,在匹配的图像未定位在移动终端 10 上的情况下可将 REVV 图像签名发送至示例中的视觉搜索设备 52。网络 54 可包括有线网络、无线网络(例如,蜂窝网络、无线局域网、无线广域网,及其某些组合等)、直接通信链路(例如,蓝牙、机器对机器的通信等)或其组合,并且在一个实施例中包括互联网。

[0072] 图 6 举例说明了由根据本发明的一个实施例的方法、设备和计算机程序产品执行的示例操作的示例流程图。要理解,流程图的各个块及流程图中块的组合可通过各种装置实施,比如,硬件、固件、处理器、电路和/或与软件(包括一个或多个计算机程序指令)的执行相关的其它装置。例如,上述的一个或多个程序可具体表现为计算机程序指令。关于这一点,表现了上述程序的计算机程序指令可通过实施本发明实施例的设备的存储器 112 存储,并由该设备中的处理器 110 执行。要理解,任意这种计算机程序指令可加载在计算机或其它可编程设备(例如,硬件)上以制成机器,从而使产生的计算机或其它可编程设备实施流程块中指定的功能。这些计算机程序指令还可存储在可引导计算机或其它可编程设备以一种特别的方式运行的非暂时性计算机可读存储器中,从而使存储在计算机可读存储器中的指令制成成品,该成品的执行实施了流程块中指定的功能。计算机程序指令还可加载在计算机或其它可编程设备上以在计算机或其它可编程设备上执行一系列的操作以产生计算机实施过程,从而使在计算机或其它可编程设备上执行的指令实施流程块中指定的功

能。同样,在执行时,图 6 的操作将计算机或处理电路转变成配置为执行本发明的示例实施例的特定机器。相应地,图 5 的操作定义了用于配置计算机或处理的算法以执行示例实施例。在某些情况下,在处理器执行图 6 的算法来将通用计算机转换成配置为执行示例实施例的特定机器的情况下,可提供通用计算机。

[0073] 相应地,流程图的块支持用于执行指定功能的装置的组合以及用于执行指定功能的操作的组合。同样要理解,流程图的一个或多个块及在流程图中块的组合,可通过执行指定功能的专用的基于硬件的计算机系统或专用硬件及计算机指令的组合来实施。

[0074] 在某些实施例中,可对本发明中的特定操作进行修改或进一步详述,如下文所述。而且,在某些实施例中还可包括附加的可选操作。应理解,各种修改、可选添加或下述详述可与上述操作单独包括,或与其它本发明所述的特征结合。

[0075] 图 6 举例说明了根据本发明示例实施例的根据示例方法执行 REVVMVS 的流程图。如操作 62 所示,设备 102 可包括用于将捕获的和 / 或以其他方式观察到的图像表示为一个或多个视觉词的矢量词残留的装置,比如处理器 110、REVV 模块 120 等,其中,将图像中的各个描述符量化为最接近的视觉词。如操作 64 所示,设备 102 可包括用于利用从图像中提取出的局部特征描述符为至少一个视觉词将多个矢量词残留聚合的装置,比如处理器 110、REVV 模块 120 等。

[0076] 如操作 66 所示,设备 102 可包括用于利用分类感知线性判别分析将用于各个视觉词的聚合后的至少一个矢量词残留的维度降低的装置,比如处理器 110、REVV 模块 120 等。例如,处理器 110、REVV 模块 120 等可在通过丢弃与视觉词的距离超过已确定的百分位的特征来形成矢量词残留时,剔除异常特征,且 / 或将冥次定律应用于聚合的至少一个矢量词残留。

[0077] 如操作 68 所示,设备 102 可包括用于二值化聚合后的矢量词残留的装置,比如处理器 110、REVV 模块 120 等,其中,二值化导致了产生紧凑型图像签名。如操作 70 所示,设备 102 可包括用于计算至少一个紧凑型图像签名的加权相关性的装置,比如处理器 110、REVV 模块 120 等,该图像签名在与候选者列表比较时从聚合后的至少一个矢量词残留二值化得到。如操作 72 所示,设备 120 可包括用于基于计算出的加权相关性确定候选者的排名列表的装置,比如处理器 110、REVV 模块 120 等。

[0078] 有利地,示例 REVV 模块可利用较小的存储器占用的优势。存储器的减少可使多个图像本地存储,比如存储在移动终端的存储器上。移动终端还可与远程服务器数据通信以接入附加的图像。可替换地或另外地,REVV 模块针对可迅速提取(例如,每次查询需一秒钟)的特征。可替换地或另外地,REVV 模块的紧凑特性实现了有效的增量更新。

[0079] 本发明相关领域的技术人员在了解了前述描述和相关附图中给出的教导的益处之后,将会想到本发明提出的发明的许多修改和其它实施例。因此,要理解,本发明不限于公开的具体实施例,修改例及其它实施例旨在包括在所附权利要求的范围之内。而且,虽然前述描述和相关附图在元件和 / 或功能的特定示例组合的上下文中描述了示例实施例,但是应该理解的是,可在不背离所附权利要求的范围的情况下提供元件和 / 或功能的不同组合的备选实施例。关于这一点,例如,与上文明确描述的那些元件和 / 或功能的不同组合也视为可能在一些所附权利要求中予以阐述。虽然本发明采用了特定术语,但是这些术语仅在通用和描述性意义上来使用而非出于限定之目的。

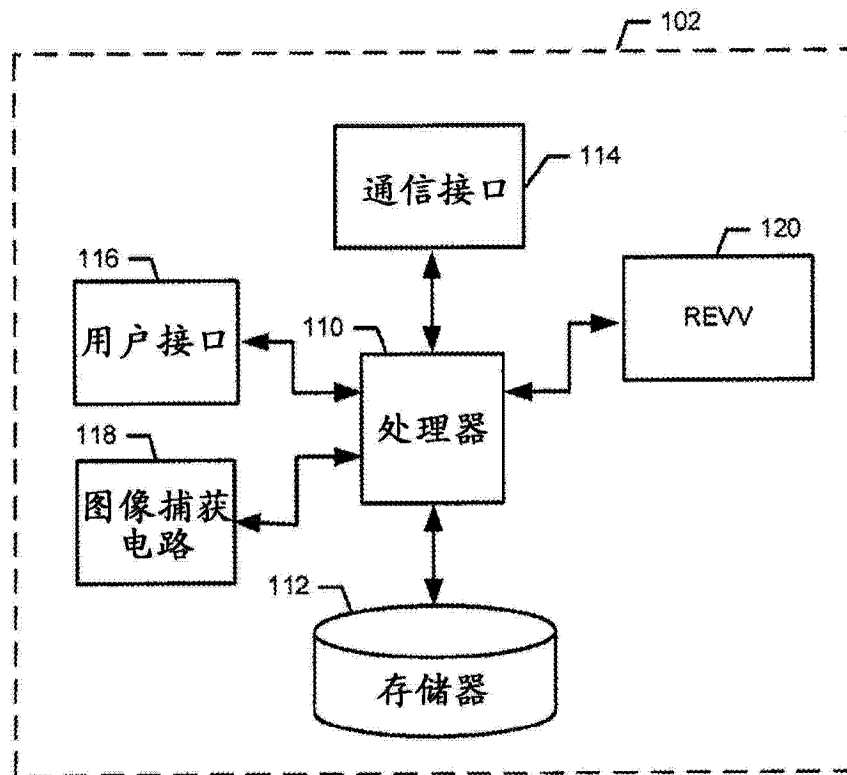


图 1

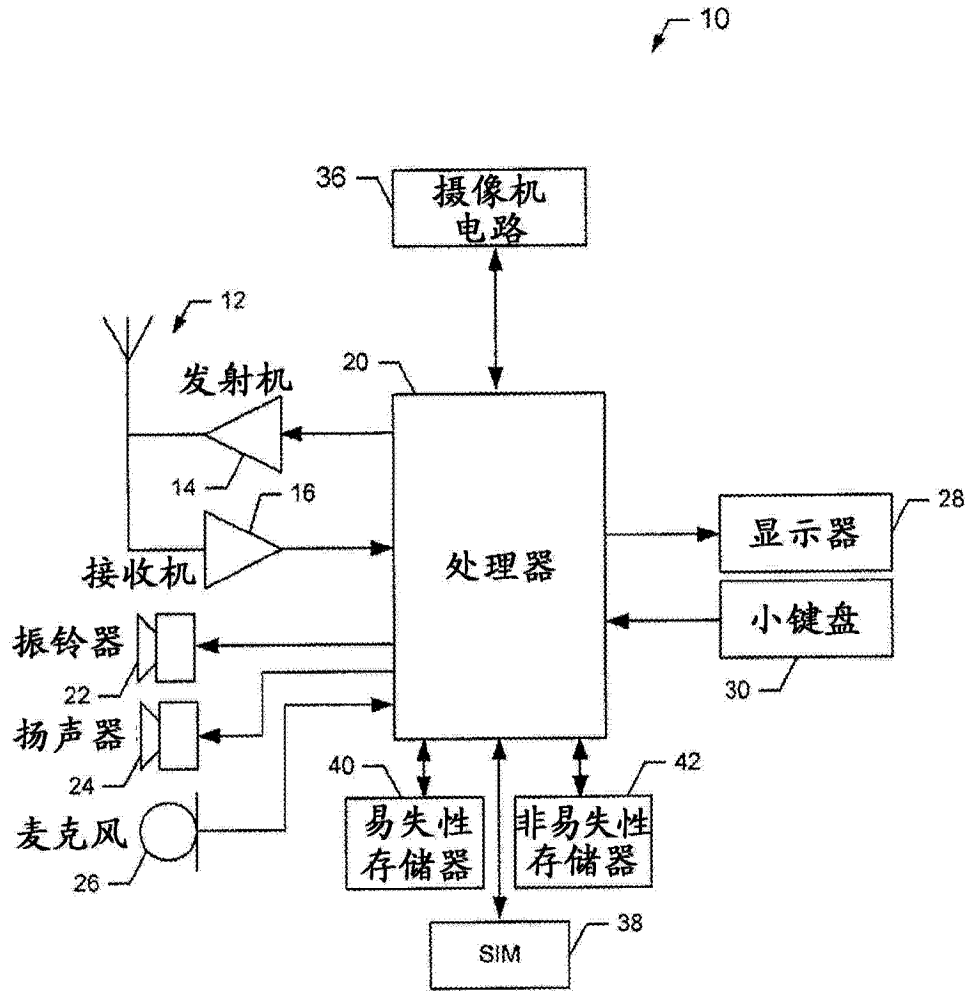


图 2

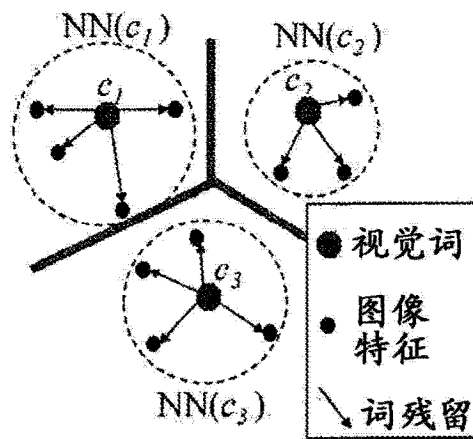


图 3

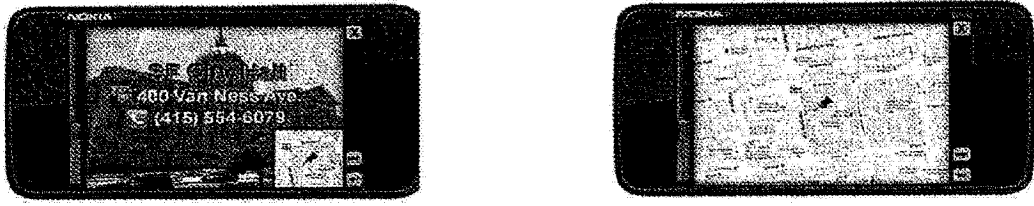


图 4

50

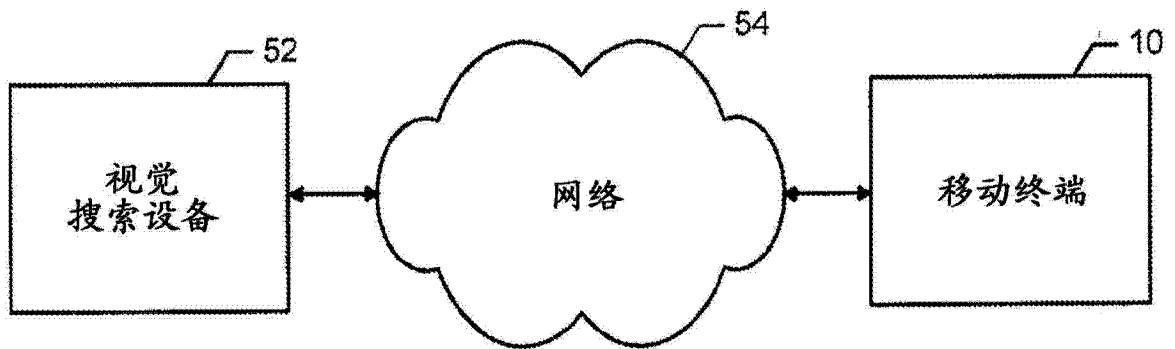


图 5

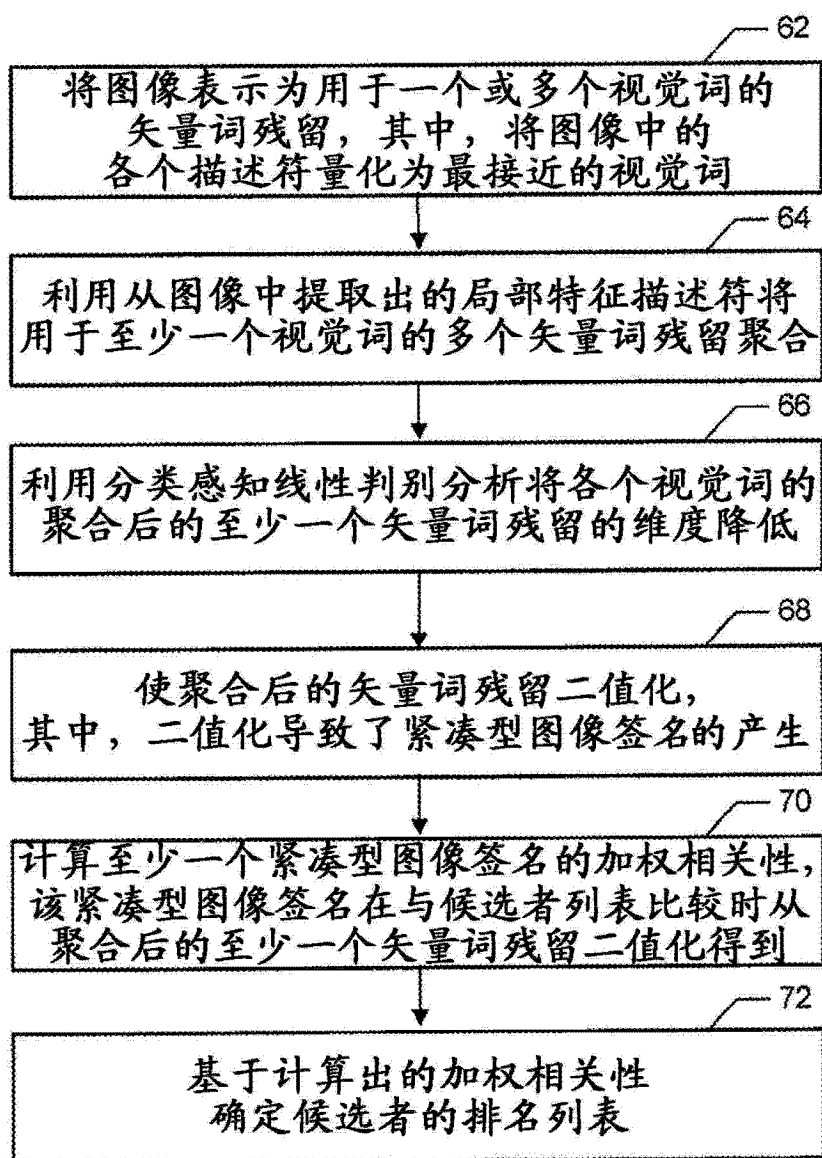


图 6