



(12) 发明专利

(10) 授权公告号 CN 102549571 B

(45) 授权公告日 2015. 11. 25

(21) 申请号 201080030849. 3

(22) 申请日 2010. 05. 14

(30) 优先权数据

12/466, 880 2009. 05. 15 US

(85) PCT国际申请进入国家阶段日

2012. 01. 09

(86) PCT国际申请的申请数据

PCT/US2010/034930 2010. 05. 14

(87) PCT国际申请的公布数据

W02010/132789 EN 2010. 11. 18

(73) 专利权人 谷歌公司

地址 美国加利福尼亚州

(72) 发明人 哈特维希·亚当 张立

(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)

(56) 对比文件

CN 101188021 A , 2008. 05. 28,

CN 101228785 A , 2008. 07. 23,

审查员 权聪敏

权利要求书2页 说明书11页 附图5页

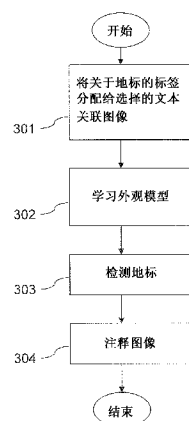
(54) 发明名称

来自数字图片集合的地标

(57) 摘要

公开了用于自动检测数字图像中的地标并且对那些图像进行注释的方法和系统。用于检测并注释数字图像中的地标的方法包括以下步骤：将描述地标的标签自动分配给多个文本关联数字图像中的一个或多个图像，以生成地标标注图像集；从地标标注图像集学习用于地标的外观模型；以及使用外观模型来在新的数字图像中检测地标。该方法还可以包括以下步骤：用描述地标的标签对新的图像进行注释。

300



1. 一种用于检测并注释数字图像中的地标的方法：

(a) 向多个文本关联数字图像中的一个或多个图像自动分配描述地标的标签，以生成地标标注图像集，其中所述地标标注图像集中的图像被算法地确定为包括所述地标，其中所述标签来自与所述一个或多个图像相关联的一个或多个文本，并且其中所述地标识别地理点或地理区域，其中自动分配描述地标的标签的步骤包括：

(i) 从所述多个文本关联数字图像生成地标 n-gram 列表；

(ii) 为 n-gram 集中的每一个地标 n-gram 计算 n-gram 分值，其中所述 n-gram 集是所述地标 n-gram 列表的子集；以及

(iii) 为所述图像中的一个或多个图像分配描述所述地标的所述标签，其中描述所述地标的所述标签基于所述 n-gram 集中的至少一个地标 n-gram；

(b) 从所述地标标注图像集学习用于所述地标的外观模型，其中所述外观模型包括特征向量，所述特征向量包括预确定图像特征集的数字分值；以及

(c) 使用所述外观模型来在新的图像中检测所述地标，

其中所述阶段 (a)-(c) 由至少一个处理器执行。

2. 根据权利要求 1 所述的方法，进一步包括：

(d) 用描述所述地标的所述标签对所述新的图像进行注释。

3. 根据权利要求 1 所述的方法，其中阶段 (a) (i) 包括：

电子地访问所述多个文本关联数字图像；以及

从与所述多个文本关联数字图像中的图像相关联的文本检索所述地标 n-gram 中的至少一个。

4. 根据权利要求 3 所述的方法，其中阶段 (a) (i) 进一步包括：

选择具有至少最小可靠性量度的所述地标 n-gram。

5. 根据权利要求 4 所述的方法，其中所述可靠性量度基于独特作者的数量。

6. 根据权利要求 1 所述的方法，其中阶段 (a) (ii) 包括：

向所述多个文本关联数字图像分配相关权重，其中所述相关权重基于所述多个文本关联数字图像中的图像的元数据的相关性，以及其中所述多个文本关联数字图像包括文本关联图像集；

从所述多个文本关联数字图像生成匹配图像图；以及

将所述地标 n-gram 与所述多个文本关联数字图像中的图像联结，以生成在地标 n-gram 和所述多个文本关联数字图像中的图像之间的联结。

7. 根据权利要求 6 所述的方法，其中阶段 (a) (ii) 进一步包括：

使用所述匹配图像图来为所述多个文本关联数字图像中的每一个图像估计地理可靠性分值。

8. 根据权利要求 6 所述的方法，其中所述 n-gram 分值基于所述匹配图像图。

9. 根据权利要求 8 所述的方法，其中所述 n-gram 分值被计算为所述匹配图像图的内部边的强度和所述匹配图像图的外部边的强度的比率，其中内部边存在于具有至少一个共同地标 n-gram 的图像之间，以及其中外部边存在于没有至少一个共同地标 n-gram 的图像之间。

10. 根据权利要求 7 所述的方法，其中阶段 (a) (ii) 进一步包括：

为所述 n-gram 集的地标 n-gram 计算地理位置差异,其中所述差异基于所述匹配图像图中的具有其 n-gram 集中的所述地标 n-gram 的图像的地理位置;以及

从所述 n-gram 集移除具有超过预确定阈值的地理位置差异的任何地标 n-gram。

11. 根据权利要求 6 所述的方法,其中阶段 (a) (i) 进一步包括:

合并所述 n-gram 集中的两个或更多个地标 n-gram。

12. 根据权利要求 11 所述的方法,其中所述合并至少基于以下的一个:

所述两个或更多个地标 n-gram 的所述分值的相似性,以及

具有联结的地标 n-gram 中的所述两个或更多个地标 n-gram 的图像的重叠。

13. 根据权利要求 6 所述的方法,其中所述元数据包括与以下的至少一个有关的信息:作者,

地理位置,以及

起源时间。

14. 根据权利要求 6 所述的方法,其中所述匹配图中的每一个联结表示在所述多个文本关联数字图像的两个图像之间的匹配特征描述符。

15. 一种用于检测并注释数字图像中的地标的系统,所述系统包括:

第一模块,其被配置成向多个文本关联数字图像中的一个或多个图像分配描述地标的标签,以生成地标标注图像集,其中所述地标标注图像集中的图像被算法地确定为包括所述地标,其中所述标签来自与所述一个或多个图像相关联的一个或多个文本,并且其中所述地标识别地理点或地理区域,其中所述第一模块被进一步配置成:

从所述多个文本关联数字图像生成地标 n-gram 列表;

为 n-gram 集中的每一个地标 n-gram 计算 n-gram 分值,其中所述 n-gram 集是所述地标 n-gram 列表的子集;以及

为所述一个或多个图像中的一个图像分配描述所述地标的所述标签,其中描述所述地标的所述标签基于所述 n-gram 集中的至少一个地标 n-gram;

第二模块,其被配置成从所述地标标注图像集学习用于所述地标的的外观模型,其中所述外观模型包括特征向量,所述特征向量包括预确定图像特征集的数字分值;以及

第三模块,其被配置成使用所述外观模型来在新的图像中检测所述地标。

16. 根据权利要求 15 所述的系统,进一步包括:

第四模块,其被配置成用描述所述地标的所述标签对所述新的图像进行注释。

17. 根据权利要求 15 所述的系统,其中所述第一模块被进一步配置成:

向所述多个文本关联数字图像分配相关权重,其中所述相关权重基于所述多个文本关联数字图像中的图像的元数据的相关性,以及其中所述多个文本关联数字图像包括文本关联图像集;

从所述多个文本关联数字图像生成匹配图像图;以及

将所述地标 n-gram 与所述多个文本关联数字图像中的图像联结,以生成在地标 n-gram 和所述多个文本关联数字图像中的图像之间的联结。

来自数字图片集合的地标

技术领域

[0001] 本发明总体上涉及数字图像集合,更具体地涉及识别大型数字图像集合中的受欢迎地标。

背景技术

[0002] 随着数字图像的使用增加、增大的数字存储容量以及由诸如因特网的数字介质提供的互连性,数量日益增大的人群可访问愈加大型的数字图像库。来自遍及世界的各种位置的具有广泛兴趣的人拍摄各种主题的图片,并且使那些图片可用于其他人例如在因特网上查看。例如,来自世界各地的各种地标和旅游地点的数字图片可以由具有不同的拍摄图片的技能水平的人张贴(post)到web上。这样的图片可以示出来自不同角度、在不同条件下和/或从不同距离拍摄的相同地标。

[0003] 作为受欢迎地标的指示符或到受欢迎地标的指南,可获得的大量这样的图像可以是有用的。为了利用包含在这些大型数字图像库中的信息,必要的是,对所述库进行组织。例如,在诸如Picasa网络相册(来自加州山景城的谷歌公司)的数字图像网站,在高级菜单开始,人们可以向下钻取(drill down)至对其而言图片可用的详细的主题收录。替选地,人们可以能够搜索具有数字图片的一个或多个站点。例如,一些旅游信息网站下载了与已公布的受欢迎旅游地点列表相关联的地标的图像。

[0004] 大部分传统的数字图片组织系统依赖于用户来标注图片。随着大量新的图片被添加到这些数字图像集合,可能不可行的是,用户以会增加那些数字图像集合的有用性的完整和一致的方式人工标记图片。在同样被转让给加州谷歌公司、题为“Automatic Discovery of Popular Landmarks(自动发现受欢迎地标)”的美国专利申请12/119,359中描述了可以从这些大型集合自动提取信息(诸如,最受欢迎旅游目的地)的系统。在申请12/119,359中描述的系统使用处理管线,所述处理管线包括基于地理编码的聚类阶段和基于图像的匹配的视觉特征的聚类阶段。然而,所需要的是,自动发现地标并且对包含地标的图像进行注释的其他方法。

发明内容

[0005] 公开了用于自动检测数字图像中的地标并对那些图像进行注释的方法和系统。在一个实施例中,用于检测并注释数字图像中的地标的方法包括以下步骤:将描述地标的标签自动分配给多个文本关联数字图像中的一个或多个图像。这生成地标标注图像集。可以从地标标注图像集学习用于地标的外观模型。这允许使用外观模型来在新的数字图像中检测地标。该方法还可以包括以下步骤:用描述地标的标签对新的图像进行注释。

[0006] 另一个实施例是用于自动检测并注释数字图像中的地标的系统。该系统具有存储在存储器介质中的至少一个文本关联数字图像集合以及通信地耦接到该介质的至少一个处理器。处理器被配置成将描述地标的标签自动分配给多个文本关联数字图像中的一个或多个图像。这生成地标标注图像集。可以从地标标注图像集学习用于地标的外观模型。这

允许使用外观模型来在新的数字图像中检测地标。

[0007] 在下面参考附图详细描述了本发明的进一步特征和优势, 以及其各种实施例的结构和操作。应注意, 本发明不限于在此描述的特定实施例。在此仅为了说明目的, 提供了这样的实施例。基于包含在此的教导, 另外的实施例对相关领域技术人员将是显而易见的。

附图说明

[0008] 将参考本发明的实施例, 这些实施例的示例可以在附图中图示。这些附图意在说明而非限制。尽管在这些实施例的语境中一般地描述了本发明, 但是应当理解的是, 这并不意在将本发明的范围限制在这些特定实施例。

[0009] 图 1 示出了根据本发明的一个实施例的用于自动检测数字图像中的地标的系统。

[0010] 图 2 示出了根据本发明的一个实施例的图 1 的系统的组件的更多细节。

[0011] 图 3 是根据本发明的一个实施例的用来自动检测数字图像中的地标, 并且对数字图像进行注释的过程。

[0012] 图 4 是根据本发明的一个实施例的用于将关于地标的标签分配给选择的文本关联图像的过程。

[0013] 图 5 是根据本发明的一个实施例的用于基于文本关联图像来生成 n -gram(n 元)列表的过程。

[0014] 图 6 是根据本发明的一个实施例的用于从根据图 4 的过程生成的 n -gram 列表选择 n -gram 集的过程。

具体实施方式

[0015] 虽然在此参考用于特定应用的说明性实施例描述了本发明, 但是应当理解的是, 本发明不限于此。本领域的技术人员使用在此的教导将认识到在本发明范围的另外修改、应用和实施例以及本发明在其中将有显著功用的另外领域。

[0016] 综述

[0017] 本发明包括用于自动识别并分类数字图像中的对象的方法和系统。例如, 本发明的实施例可以基于在因特网上可访问的数字图像集合来识别、分类以及排序最受欢迎的旅游地标。本发明的方法和系统可以使得能够对最受欢迎的旅游位置的最新列表和图像集合进行有效维护。在一些实施例中, 旅游位置的受欢迎度可以基于被用户张贴到因特网上的该位置的图像的数量来估算。

[0018] 许多个人拍摄他们附近地区的环境、他们日常活动中探访的位置以及在他们观光旅行时参观的地点的数字图片。所使用的相机是各种质量和精密水平的。捕捉图像的个人是各种技能水平的。图像从各种角度、在不同光线水平、以不同环境视觉噪音水平、在各种天气条件下等被捕捉。然后将这些图像中的许多张贴在图片共享网站上或通过其他手段使这些图像中的许多数字化可获得。通过诸如因特网的网络使对巨大数字图像集合的访问可获得, 所述数字图像诸如数字图片。

[0019] 通常, 在线张贴图像的用户还例如通过添加一个或多个标签和 / 或题注, 对所张贴的图像进行注释。可以使用标签来对图像命名。还可以将标签分配给图像以分配与图像有关的关键词。例如, 埃菲尔铁塔的图像可能已分配给其有标签“埃菲尔铁塔”、“巴黎”、“法

国”、“欧洲”、“夏天”或被示出在铁塔前摆姿势的人的名字。标签作为在各种粒度水平的组织工具是有价值的：“法国”可以是有用的，以将图像分类在对在法国的地标的搜索下，而仅使“埃菲尔铁塔”成为标签可以将图像排除于对在“巴黎”和 / 或“法国”的地标的搜索之外。尽管在确定包含在那些图像中的地标时，有图像的标签的准确性和有用性的差异，为构建自动地标辨识系统的目的，用户标注图像库是有价值的信息源。

[0020] 其他潜在信息源包括联结文本和图像的各种其他文档和电子源。例如，关于埃菲尔铁塔的杂志文章可以包括其主题的图片。报纸内容、杂志和期刊内容、由个人撰写和 / 或张贴的文章，包括关于各种地标的博客帖子等，通常包括直接束缚于文本描述的图像。具有可辨识的地标相关文本的图像可以被称为地标标注图像。

[0021] 本发明的实施例利用若干类型的关于图像可获得的数据，以获取关于受欢迎地标的信息。例如，地理标签、文本标签、作者信息、时间戳（例如，时间或起源）以及视觉匹配信息是在本发明的实施例中利用的一些类型的信息。该信息中的部分是与每一个图像一起（例如，在与图像相关联的 EXIF 标签中）可获得的。其他信息是用户分配的或算法分配的。当分别取得时，这些数据类型中的每一个可能具有大量弱点。例如，地理位置数据（例如，地理标签）通常基于相机而不是被拍照的地标的位置。并且，在一些情况下，地理位置信息基于诸如城市名称的用户提供的信息，因此可能是不准确的。作者和第三方所提供的文本标签可能没有准确描述地标。每一个图像的作者信息可以基于相机标识符、捕捉图像的人、或将图像上载到网站的人。在诸如在较小区域中存在若干地标、地标看起来相像和 / 或图像质量不足的情况下，视觉匹配信息也可能是错误的。因此，本发明的实施例利用若干类型的可获得信息来获取在数字图像中的较高程度地标检测和准确注释。

[0022] 用于自动地标辨识和注释的系统

[0023] 在图 1 中示出了根据本发明的一个实施例的用于构建注释的受欢迎地标图像数据库的系统 100。系统 100 包括计算机 101、用户接口 102、网络 103 和 104、文本 / 图像文档集合 107、n-gram 集合 108、n-gram 过滤器数据库 109、未注释的图像数据库 110、外观模型数据库 111、注释的图像 112 以及文本 / 图像源 105。本领域技术人员将理解的是，系统 100 可以包括比在上面列出的那些更多、更少或不同的组件和模块，而仍然与本发明一致。

[0024] 计算机 101 可以包括经由通信介质互连的一个或多个计算机、服务器或类似计算设备。例如，计算机 101 可以包括经由一个或多个局域网耦接的一个或多个商用计算服务器，所述局域网诸如以太网、吉比特以太网、WIFI 网络或类似物。计算机 101 包括处理器 121、易失性存储器 122、持久性存储器 123、网络接口 124、数据库接口 125、耦接计算机 101 的模块的通信介质 126 以及非监督的图像注释器模块 127。处理器 121 可以包括一个或多个商用中央处理单元 (CPU)、图形处理器单元 (GPU)、现场可编程门阵列 (FPGA)、数字信号处理器 (DSP) 以及专用集成电路 (ASIC)。处理器 121 控制在计算机 101 内的处理、从计算机 101 接收输入以及将数据输出到计算机 101。例如，非监督的图像注释器模块 127 的处理逻辑可以在处理器 121 上执行。

[0025] 易失性存储器 122 可以包括易失性存储器，诸如动态随机存取存储器 (DRAM)、静态随机存取存储器 (SRAM) 或类似物。易失性存储器 122 可以用来存储配置参数、源数据和模块 127 的处理的中间结果。配置参数可以包括关于文本 / 图像源 105 的连接信息，以及配置例如非监督的图像注释器模块 127 的处理的参数的其他参数。持久性存储器 123 可以

包括一个或多个非易失性存储器设备,诸如磁盘、光盘、闪存、只读存储器 (ROM) 或类似物。持久性存储器 123 可以用于存储用于非监督的图像注释器模块 127 的逻辑指令、配置参数,以及用来存储模块 127 中的处理的中间和其他结果。

[0026] 网络接口 124 可以包括与通过包括网络 103 的网络连接到计算机 101 的实体通信的功能性,所述实体诸如文本 / 图像源 105。例如,网络接口 124 可以包括包含网际协议 (IP) 和超文本传输协议 (HTTP) 处理的处理组件,以使使得计算机 101 能够连接到文本 / 图像源 105 以获取文本和图像信息。例如,HTTP 协议处理机器软件可以被实现为网络接口 124 的一部分。根据本发明的实施例,数据库接口 125 包括将计算机 101 连接到在为地标处理图像时使用的一个或多个数据库的功能性。应当注意的是,术语“数据库”的使用不一定是指数据库管理系统 (DBMS),而是包括任何数据集合。因此,数据库接口 125 可以包括连接到包括一个或多个数据库 107-112 的一个或多个 DBMS 系统的 DBMS 功能性或与每一种类型的数据库 107-112 的类型的数据库通信的处理逻辑。通信介质 126 可以连接计算机 101 的模块,包括模块 121-125 和 127。通信介质 126 可以包括通信设备,诸如 PCI 总线、USB、以太网或类似物。

[0027] 根据本发明的实施例,非监督的图像注释器模块 127 包括以下功能性:识别地标、为所选择的地标生成外观模型;以及对图像进行注释。如下所述,包含在图像中的地标可以基于已与图像相关联的明确标签或通过算法手段来识别。非监督的图像注释器模块 127 的功能性可以以软件、固件、硬件或其任何组合来实现。在一个实施例中,用于非监督的图像注释器模块 127 的功能性的处理逻辑可以以诸如 C、C++、汇编、Java、JavaScript、Perl 或类似物的计算机编程语言或脚本语言来实现。

[0028] 网络 103 可以包括将计算机 101 连接到一个或多个文本 / 图像源 105 的装置。网络 104 可以包括将计算机 101 连接到一个或多个数据库 107-112 的装置。网络 103 和 104 可以包括一个或多个网络介质,包括:外围连接,诸如 USB、火线;或局域网,诸如以太网、WIFI;或广域网,诸如 PSTN 或因特网。在一个实施例中,网络 103 包括因特网,以及网络 104 包括基于以太网的局域网。

[0029] 用户接口 102 可以使用互连机制的任何一个或组合来连接到一个或多个计算机 101,所述互连机制诸如 PCI 总线、IEEE 1394 火线接口、以太网接口、IEEE 802.11 接口或类似物。用户接口 102 允许用户或其他外部实体与计算机 101 交互。在一些实施例中,一个或多个数据库 107-112 还可以通过用户接口 102 交互。图形用户接口、web 接口和应用编程接口中的一个或多个可以被包括在用户接口 130 中。

[0030] 文本 / 图像源 105 可以包括包含地标的图像和相关联的文本(例如,地标标注图像)的各种类型的数字文档集合。在一个实施例中,文本 / 图像源 105 包括具有与题注和标签相关联的图片的一个或多个图片集合。如在此所使用的,题注是指分配给图片的标题。如在此所使用的,标签是指分配给图片的一个或多个词或短语。通常,题注以及标签由图片的作者(例如,图片的创作者、或将图片上载到图片共享网站的人)分配。然而,题注和标签还可以由第三方或自动化工具分配给图片。除非每一个被单独识别,在下面的描述中的术语“标签”包括标签以及题注。

[0031] 文本 / 图像源 105 还可以包括将图像超链接到文档(反之亦然)的超文本文档集合,以及还可以包括报纸库、杂志和期刊库、博客归档、具有数字化书籍的数字图书馆、第三

方注释的图片仓库以及个人和企业网站。例如,旅游和 / 或旅行相关网站、数字旅行指南、城市网站等是一般包括地标的图像和那些地标的描述的一些资源。然而,其中可以取得在一个或多个图像和相关联的文本之间的相互关联的任何数字数据集可以被包括在文本 / 图像源 105 中。

[0032] 文本 / 图像集合 107 是数据库,在一些实施例中,在其中保存在远程文本 / 图像源 105 中原始访问的文本 / 图像数据的本地副本和 / 或经修改的版本,例如以更便利和可靠访问来由非监督的图像注释器 127 处理。例如,由于通过可能是诸如因特网的广域网的网络 103 访问文本 / 图像源 105 中的数据和图像可能涉及较长延时,在计算机 101 中可以存在以下过程(未示出):在本地或本地附加的网络位置中,诸如在文本 / 图像集合 107 中,制作这样的数据和图像的副本。文本 / 图像集合 107 还可以包括已被标注的图像集合,例如 Picasa 网络相册中的用户图片集合和 / 或已根据本发明的教导处理的图像集合。在一些实施例中,文本 / 图像集合 107 可以包括对应于每一个图像的数据结构,其中数据结构包括到文本 / 图像源 105 中的图像和 / 或文档的一个或多个指针,例如以避免必须创建来自文本 / 图像源 105 的图像和 / 或文档的单独副本。

[0033] N-gram 集合 108 是包括 n-gram 集合的数据库。N-gram 可以从与在例如文本 / 图像集合 107 或文本 / 图像源 105 中的图像相关联的题注、标签或文本文档提取。如在此所使用的,n-gram 是一个或多个词的序列。N-gram 的选择可以使用与例如在文本分析中所使用的若干技术中的一个或多个类似的方法来完成。根据本发明的实施例,在下面进一步描述了 n-gram 的选择和提取。

[0034] N-gram 过滤器数据库 109 包括待从 n-gram 集合 108 里过滤的一个或多个 n-gram 列表、和 / 或待应用于 n-gram 集合 108 的一个或多个过滤规则。例如,n-gram 过滤器数据库 109 中的一个列表可以是“坏词列表”,其中在坏词列表中出现的 n-gram 不从文本 / 图像集合 107 或文本 / 图像源 105 提取,并且如果发现他们存在,则从 n-gram 集合 108 移除。另一个列表可以是在图像相关文本中出现太频繁的 n-gram 列表,因此作为地标标识符是很少价值的。诸如“the”和“of”的词可以认为在此类别。另一个列表可以是短语列表,其被已知出现太频繁,因此作为有辨识力的地标标识符不是足够有用的。

[0035] 根据本发明的实施例,未注释的图像数据库 110 包括尚未被注释(例如,标注)的图像。例如,未注释的图像数据库 110 可以包括一个或多个用户上传的未标注的数字图像,以使用本发明的实施例来处理。

[0036] 外观模型数据库 111 保持在此被称为外观模型的辨识模型,其被得到以辨识例如未注释的图像数据库 110 中的图像的图像中的地标。

[0037] 根据本发明的实施例,注释的图像数据库 112 包含被标注的图像。例如,根据本发明的实施例,来自未标注的图像数据库 110 的图像在它们被非监督的图像注释器 127 处理之后被存储在注释的图像数据库 112 中。本领域技术人员将认识到,尽管数据库 107-112 在上面被描述为独立数据库,但是可以以与本发明一致的各种方式来安排和 / 或实现数据库 107-112。

[0038] 图 2 示出了根据本发明的一个实施例的非监督的图像注释器模块 127 的更多细节。在该实施例中,非监督的图像注释器模块 127 包括三个处理模块:地标识别器 201、外观模型生成器 202 以及图像注释器 203。模块 201、202 和 203 可以以软件、固件、硬件或其

组合来实现。在一个实施例中,模块 201-203 使用 C++ 编程语言以软件来实现。在一个实施例中,计算机程序产品可以使包括模块 201-203 的计算机程序逻辑的逻辑记录在计算机可读介质上,所述计算机可读介质诸如硬盘、闪存盘或其他形式的存储。

[0039] 地标识别器模块 201 包括识别文本 / 图像集合 107 和 / 或文本 / 图像源 105 中的地标的功能性。在一个实施例中,地标识别器模块 201 可以将来自文本 / 图像源 105 的图像和相关联的文本用作输入,并且将这样的图像和相关联的文本复制到文本 / 图像集合 107。地标识别器模块 201 还可以在使用以及更新 n-gram 集合 108 的同时对文本 / 图像源 105 中的文本进行分析。N-gram 过滤器数据库 109 还可以在地标识别器模块 201 内的处理中被使用。

[0040] 外观模型生成器 202 包括为例如由地标识别器模块 201 识别的每一个地标生成一个或多个外观模型的功能性。在一个示例中,外观模型生成器 202 可以将文本 / 图像集合 107 中的图像和所识别的地标当作输入,并且为地标中的每一个生成一个或多个外观模型。所生成的外观模型可以被写入外观模型数据库 111。

[0041] 如在此所使用的,外观模型是在自动辨识图像中的某些共同特征时使用的模板。在本发明的一个实施例中,用于辨识地标的外观模型可以包括特征向量,其包括预确定图像特征集的数字分值。图像中的对象辨识和生成特征向量的方法是本领域众所周知的。例如,在 David G. Lowe, " Object recognition from local scale-invariant features," International Conference on Computer Vision, Corfu, Greece (September 1999), pp. 1150-1157 中描述了图像中的对象辨识的方法。除视觉辨识组件外,外观模型还可以包括信息,诸如对应的地标的地理位置信息。例如,外观模型中的特定地标的地理位置信息可以指定地理点和 / 或地理区域。指定地理区域可以减少由于图像的地理位置信息的准确性的差异而创建的不确定性。

[0042] 图像注释器模块 203 包括以下功能性:自动辨识图像中的地标,并且用识别一个或多个对应的地标的信息恰当注释这样的图像。在一个实施例中,图像注释器模块 203 可以使用来自外观模型数据库 111 的外观模型来自动辨识来自未注释的图像数据库 110 的图像中的地标。然后可以根据每一个图像中的所辨识的地标,例如通过关联一个或多个标签来对图像进行注释,以及可以将所注释的图像写入注释的图像数据库 112。

[0043] 用于自动地标辨识和注释的方法

[0044] 图 3 示出了根据本发明的一个实施例的对包括一个或多个受欢迎地标的图像进行注释的过程 300。过程 300 可以例如在非监督的图像注释器模块 127 中实现。过程 300 的步骤 301-304 可以在适当情况下在地标识别器模块 201、外观模型生成器模块 202 以及图像注释器模块 203 中实现。本领域技术人员将理解的是,在此参考过程 300 描述的功能性可以使用模块 201-203 以除在下面描述的方式外的方式来实现。例如,在一个实施例中,地标识别器模块 201、外观模型生成器模块 202 和图像注释器模块 203 每一个可以是独立进程,其在一起实现过程 300。在另一个实施例中,地标识别器模块 201、外观模型生成器模块 202 和图像注释器模块 203 每一个可以是独立线程,其在一起实现过程 300。在又一个实施例中,地标识别器模块 201、外观模型生成器模块 202 和图像注释器模块 203 可以全体实现为单个过程,其实现过程 300。

[0045] 在步骤 301,对图像和与那些图像相关联的文本进行分析以识别地标,具体地为受

欢迎地标。总的来说,受欢迎地标是在诸如文本 / 图像源 105 的所分析的图像 / 文本源中出现最频繁的那些地标。在一个实施例中,到步骤 301 中的处理的输入是在其上正执行过程 300 的一个或多个计算机可访问的一个或多个图像 / 文本源。例如,过程 300 可能正在计算机 101 上执行,以及可以通过网络 103 具有对文本 / 图像源 105 的可访问性。根据一个实施例,来自步骤 301 的输出可以是选择的图像集、在那些图像中的所识别的地标以及相关文本和 n-gram。例如,步骤 301 的输出可以被写入文本 / 图像集合 107。在下面参考图 4-6 进一步描述步骤 301。

[0046] 在步骤 302,得到或学习用于在步骤 301 中识别的地标的一个或多个外观模型。本领域技术人员将认识到,可以使用许多方法中的一个来从作为步骤 301 的结果获取的地标标注图像学习外观模型。根据一个实施例,特定地标的外观模型包括特征向量,其数字上量化被认为包含特定地标的一个或多个图像的一个或多个视觉方面。如前所述,特征向量生成是本领域众所周知的,以及在上面引用的 David G. Lowe, " Object recognition from local scale-invariant features" 中描述了特征向量生成的方法,诸如能够在本发明中使用的。例如,特征向量理想地包括对于许多变化条件相对不变的大量特征,所述变化条件诸如相机距离、相机角度、图像质量、光线条件等。在本发明的一些实施例中,与特定图像相对应的一个或多个外观模型还可以包括图像的非视觉方面,诸如地理位置信息。外观模型可以包括能够在自动辨识在图像中该地标的存在时使用的任何信息,包括特定地标的视觉特征和地理位置信息。

[0047] 在步骤 303,使用在步骤 302 中获取的一个或多个外观模型来检测图像中的对应地标。在一个实施例中,外观模型数据库 111 中的一个或多个外观模型在检测未注释的图像数据库 110 中的对应地标时被使用。例如,可以将来自外观模型数据库 111 的外观模型的特征向量与为正考虑的来自未注释的图像数据库 110 的图像生成的特征向量进行比较。如果特征向量匹配超过预确定的阈值水平,则正被考虑的图像被辨识为包括与匹配的外观模型相对应的地标。对象辨识技术,诸如能够在本发明的实施例中的步骤 303 中使用的,一般是众所周知的。在上面引用的 Lowe, " Object recognition from local scale-invariant features" 中描述了能够在本发明中使用的对象辨识的一个方法。

[0048] 在步骤 304,如果确定正被分析的图像其内具有与例如在步骤 303 的检测中使用的一个或多个外观模型相对应的特定地标,则可以对该图像进行注释。注释的图像和相应的注释可以被写入注释的图像数据库 112。与注释的图像相关联的注释可以包括与发现在该注释的图像中具有匹配的外观模型中的每一个相关联的文本。还预期的是,与注释的图像相关联的注释可以包括基于与对应的外观模型相关联的文本的另外处理的文本或短语。例如,在与对应的外观模型相关联的文本是诸如“大卫雕像”和“罗马”的简单标签的形式的一个实施例中,步骤 304 可以包括生成诸如“在意大利罗马的大卫雕像”、“在意大利罗马韦奇奥宫的大卫雕像”或类似物的句子的另外处理。

[0049] 在图 4 中,进一步详细示出了步骤 301 中涉及的处理。步骤 301 的功能性包括步骤 401-403。在步骤 401,生成描述地标的词或短语的 n-gram 集和 / 或更新现有 n-gram 集。例如,步骤 401 可以将文本 / 图像源 105 当作输入,并且产生 n-gram 集合 108 中的 n-gram 作为输出。在下面关于图 5 提供了步骤 401 的更详细描述,如如何生成描述地标的的一个或多个 n-gram。

[0050] 在步骤 402, 对初步认为对地标确定是有用的 n-gram 集进行评分。例如, 在步骤 402 中所认为的初始 n-gram 集可以是在步骤 401 中从文本 / 图像源 105 得到的 n-gram 集。步骤 402 的处理可以创建 n-gram 集合 108 中的 n-gram 列表。根据各种标准来对 n-gram 进行过滤, 所述各种标准包括对每一个 n-gram 进行评分, 并且仅保留预确定数量的具有最高分值的 n-gram。将 n-gram 分值 $S(k)$ 分配给 n-gram 集合 108 中的 n-gram $N(k)$ 中的每一个。在下面描述了确定 $S(k)$ 的方法。在下面参考图 6 进一步描述了步骤 402 的处理。

[0051] 在步骤 403, 为图像分配来自 n-gram 集合 108 的标签。例如, 对于每一个图像和 n-gram 组合对, 可以分配配对分值。可以定义配对分值, 以使更高值的配对分值意味着强烈相关的图像和 n-gram 对。在一个示例中, 可以为由来自图像 / 文本集合 107 的图像 $I(i)$ 和来自 n-gram 集合 108 的 n-gram $N(k)$ 形成的配对分配由以下定义的配对分值: 在 $I(i)$ 和 $N(k)$ 之间的联结强度 $L(i, k)$ 与 $N(k)$ 的 n-gram 分值的乘积, 即 $L(i, k) * S(k)$ 。在下面描述了确定 $L(i, k)$ 的方法。可以通过聚焦于具有较高配对分值的 n-gram, 并且恰当截取列表, 来生成候选 n-gram 列表。在一个实例中, 当配对分值落入低于列表中的最高配对分值的一半时, 列表可以被截取。以这种方式, 可以为每一个图像分配最相关的 n-gram。

[0052] 图 5 示出了根据上述步骤 401 的在生成 n-gram 集时的处理步骤 501-504。在步骤 501, 例如通过地标识别器模块 201 访问一个或多个文本 / 图像源 105。对文本 / 图像源 105 的访问可以包括通过本地网络或诸如因特网的广域网连接到这样的源。被选择进行处理的文本 / 图像源 105 可以基于各种方法来识别, 所述方法诸如来自用户或操作者的输入、程序组件对网站的自动识别和分类 (例如, web 自动程序对图片库网站的识别) 或针对内容监视的网站或其他库列表。连接到诸如文本 / 图像源 105 的源的方法是众所周知的。在需要的情况下, 本发明的实施方式还应当考虑可能在使用各方所有的图像时涉及的版权、隐私等方面。

[0053] 在步骤 502, 从与文本 / 图像源 105 中的图像相关联的文本检索潜在地标描述符 n-gram 列表。从其中图片与标签和 / 或题注相关联的图片库提取 n-gram 可以包括收集与文本 / 图像源 105 的图片库的图片相关联的标签和 / 或题注集。当图像 / 文本源包括将图像与对应的文本相关联的其他文档和 / 或内容时, 可以使用许多文本分析方法中的一个或多个来提取潜在对应于地标的词语 (标签)。例如, 可以使用本领域众所周知的方法来自动分析与旅游网站中的图像相关联的文本, 所述方法诸如在可用于识别潜在标签的文本上的词频 - 反文档频率 (TF-IDF)。在一个实施例中, 将 TF-IDF 应用于与来自文本 / 图像源 105 的图片库中的图片相关联的标签。

[0054] 可以应用预确定规则来从潜在大量可获得的标签确定是指地标的缩窄和 / 或经过过滤的标签集。例如, 在步骤 503, 可以将一个或多个过滤规则或标准应用于步骤 502 中收集的潜在地标描述符的 n-gram 集。可以应用于潜在地标描述符 n-gram 列表的一个过滤器是坏词过滤器。坏词过滤器包括被预确定为对在地标之间辨别是坏和 / 或无帮助的 n-gram 和短语的列表。被应用的另一个过滤器可以是停止词列表。停止词列表可以包括预期会在标签和 / 或描述符中出现太频繁, 以致它们作为地标描述符不可能是有帮助的 n-gram。诸如 “of”、“the” 和 “and” 的词是可以被包括在停止词列表中的示例 n-gram。可以被应用的另一个过滤器是最小可靠性量度, 诸如最小数量的作者过滤器。可以使用最小数量的作者过滤器来从潜在地标描述符 n-gram 列表移除具有小于预确定数量的独特 (unique) 作者在

其标签中使用那些 n-gram 的任何 n-gram。例如,可以预确定的是,对于待被包括在 n-gram 集合 108 中的任何 n-gram,应当在三个或更多独特作者使用的标签中检测到该 n-gram。

[0055] 在步骤 504,可以将步骤 503 中应用一个或多个规则和 / 或过滤器之后剩下的潜在地标描述符 n-gram 列表写到 n-gram 集合 108 中。由诸如处理步骤 402 的随后处理步骤使用的来自 n-gram 集合 108 的 n-gram 集是已根据如上所述的若干过滤器被过滤的 n-gram 集,因此将仅包括实质上描述地标的 n-gram。

[0056] 图 6 示出了根据一个实施例的图示步骤 402 中涉及的处理的步骤 601-608。在步骤 601,为与在步骤 401 中选择的 n-gram 相关联的图像分配相关权重。在一个实施例中,将与在步骤 401 中选择的 n-gram 相关联的图像复制到文本 / 图像集合 107 中,以及针对那些图像执行权重分配和另外的处理。图像 I(i) 的相关权重 W(i) 是图像 I(i) 与文本 / 图像集合 107 中的其他图像的相关水平的反向量度。例如,如果图像 I(i) 与文本 / 图像集合 107 中的任何其他图像不相关,则为图像 I(i) 分配 1 的相关权重;如果图像 I(i) 与文本 / 图像集合 107 中的 2 个其他图像相关,则为图像 I(i) 和其两个相关图像中的每一个分配 1/3 的相关权重。可以使用预确定的规则或标准集来确定两个图像是否相关。例如,当两个图像由同一作者并且在非常近的地理位置(例如,相互距 1/4 英里内)拍摄时,可以认为该两个图像相关。

[0057] 在步骤 602,从例如文本 / 图像集合 107 中的图像创建匹配图像图。匹配图像图中的节点表示文本 / 图像集合 107 中的图像。匹配图像图中的每一条边表示与两个连接的节点相对应的图像匹配的程度。例如,分配给在图像 I(i) 和 I(j) 之间的边的匹配分值 M(i, j) 可以是基于在图像 I(i) 的特征向量和图像 I(j) 的特征向量之间的匹配来得到的数值。可以为特征向量中的单个特征分配可配置的权重,以及匹配分值 M(i, j) 可以是匹配特征的这样的权重的总和。

[0058] 在步骤 603,在 n-gram 集合 108 中的 n-gram 和文本 / 图像集合 107 中的图像的每一个之间形成联结(被称为图像 - 名称联结)。如果图像的标签包含 n-gram,则图像 - 名称联结可以是被设置为 1 的二进制变量,否则为 0。然而,为了增加结果的鲁棒性,通过在视觉上相似的图像集上求平均值,而不是考虑单个图像来使输出平滑。例如,在图像 I(i) 和 n-gram k 之间的图像 - 名称联结 L(i, k) 可以被定义为:

[0059]

$$L(i, k) = \frac{\sum_{\text{对于具有 n-gram } k \text{ 的所有图像 } j} M(i, j) * W(j)}{\sum_{\text{对于所有图像 } j} M(i, j) * W(j)}$$

[0060] 其中,如上所述,M(i, j) 是图像匹配图中在图像 I(i) 和 I(j) 之间的匹配分值,以及 W(j) 是图像 I(j) 的相关权重。

[0061] 在步骤 604,估计文本 / 图像集合 107 中的每一个图像的地理可靠性。基于具有在相互的预确定距离内的地理位置坐标的图像的视觉一致性的比较,图像 I(i) 的地理可靠性 G(i) 是图像的地理位置信息的准确性的估计。例如,

[0062]

$$G(i) = \frac{\sum_{\text{对于 } n \text{ 个最近图像 } j \text{ 至图像 } i} M(i, j) * W(j)}{\sum_{\text{对于 } n \text{ 个最近图像 } j \text{ 至图像 } i} W(j)}$$

[0063] 其中, n 可以是可配置参数。

[0064] 在步骤 605, 可以可选地为每一个 n-gramN(k) 计算地理差异。例如, N(k) 的地理差异 V(k) 可以被表示为:

[0065]

$$V(k) = EW \left[(loc(i) - EW(loc(i)))^2 \right]$$

[0066] 其中 loc(i) 表示图像 I(i) 的地理位置, 以及 EW 是加权期望。加权期望在为 n-gram 捕捉最显著位置点的差异时是有帮助的。权重可以被计算为 L(i, k) * W(i) * G(i), 即图像 - 名称联结、图像权重和图像的地理可靠性的乘积。随后, 可以将具有大于阈值地理差异的 V(k) 的 n-gram 从 n-gram 集合 108 过滤出。

[0067] 在步骤 606, 使用量度来确定文本 / 图像集合 107 中的每一个 n-gramN(k) 的 n-gram 分值 S(k), 所述量度被设计成捕捉在其标签中具有 n-gramN(k) 的图像之间的内部联结强度, 以及在其标签中具有 n-gramN(k) 的图像和在其标签中没有 n-gramN(k) 的图像之间的外部联结强度。例如, S(k) 可以被表示为:

[0068]

$$S(k) = \frac{\sum_{\text{对于所有图像对 } (i, j)} W(i) * L(i, k) * M(i, j) * L(j, k) * W(j)}{\sum_{\text{对于所有图像对 } (i, j)} W(i) * L(i, k) * M(i, j) * (1 - L(j, k)) * W(j)}$$

[0069] S(k) 越大, 该 n-gramN(k) 越可能是指有意义、视觉上可区分的实体, 因此, 越可能是地标名称。

[0070] 在步骤 607, 在对 n-gram 进行评分之后, 可以可选地执行进一步过滤来识别最受欢迎地标 n-gram。例如, 可以对预确定数量的具有最高 n-gram 分值的 n-gram 的 n-gram 分值求平均值, 来确定阈值平均分。此后, 可以从 n-gram 集合 108 移除除具有高于阈值平均分值的分值的那些 n-gram 外的所有 n-gram。

[0071] 在步骤 608, 合并被认为是指同一地标位置的 n-gram。尽管评分步骤, 以及基于分值的随后过滤一般留下有意义地指地标的 n-gram 列表, 但是是指同一地标的许多 n-gram 可能仍然余留在 n-gram 集合 108 中。由于若干原因, 包括用于同一地标的不同名称、同一名称的不同表述以及子串截取, 可能存在是指同一地标的多个 n-gram。所期望的是, 以有意义的方式将这样的重复 n-gram 合并在一起。为了解决这一点, 在一个示例中, 如果两个 n-gramN(k) 和 N(l) 具有在距相互预确定距离内的分值, 并且如果它们联结到的图像是大幅重叠的, 则将该两个 n-gramN(k) 和 N(l) 合并。可以例如通过以下来确定图像的大幅重叠: 考虑每一个图像 I(i) 和 n-gramN(k) 对的巴特查里亚距离 L(i, k), 并且确定巴特查里亚距离是否高于预确定阈值。巴特查里亚距离的计算是本领域众所周知的。

[0072] 结论

[0073] 模块 127 和 / 或模块 201-203 的处理功能性可以以软件、硬件或其组合来实现。例如, 模块 201 和 203 可以被全部实现为软件模块, 或外观模型生成器模块 202 的部分功能

性可以使用诸如现场可编程门阵列 (FPGA) 的硬件来实现。本领域技术人员将理解的是,非监督的图像注释器模块 127 和 / 或计算机 101 可以包括促进本发明的功能的另外组件和模块。

[0074] 将理解的是,具体实施方式部分而不是发明内容和摘要部分意在用于解释权利要求。发明内容和摘要部分可以阐明发明人考虑的本发明的一个或多个而非所有的示例性实施例,因此发明内容和摘要部分并不意在以任何方式来限制本发明和所附的权利要求。

[0075] 在上面借助于说明指定的功能及其关系的实现的功能构造块来描述了本发明。在此为了便于描述,任意地限定了这些功能构造块的边界。可以限定替选边界,只要所述指定的功能及其关系被适当地执行。

[0076] 特定实施例的前述描述如此充分地揭示了本发明的一般特性,使得其他人可以在不背离本发明的一般思想的情况下,通过应用在本领域的技术内的知识为各种应用而容易地修改和 / 或调整这样的特定实施例,而不用进行过度的实验。因此,基于在此提供的教导和指导,这样的调整和修改意在在所公开的实施例的等同物的含义和范围内。应当理解的是,在此的措词或术语用于描述而非限制的目的,因此本说明书的术语或措词应当由技术人员根据所述教导和指导来解释。

[0077] 本发明的宽度和范围不应当由上述示例性实施例的任何一个限制,而是应当仅根据所附权利要求及其等同物来限定。

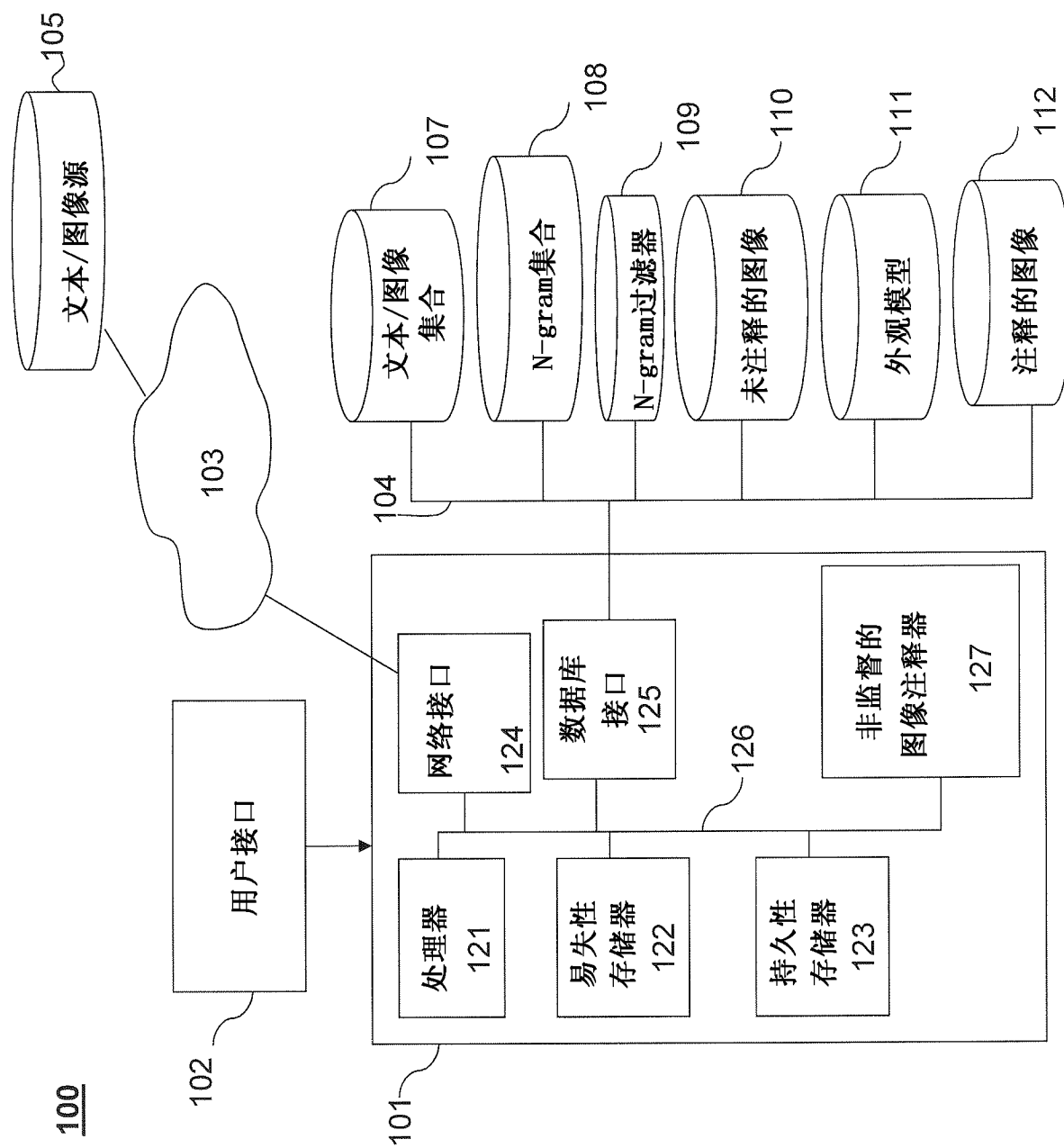


图 1

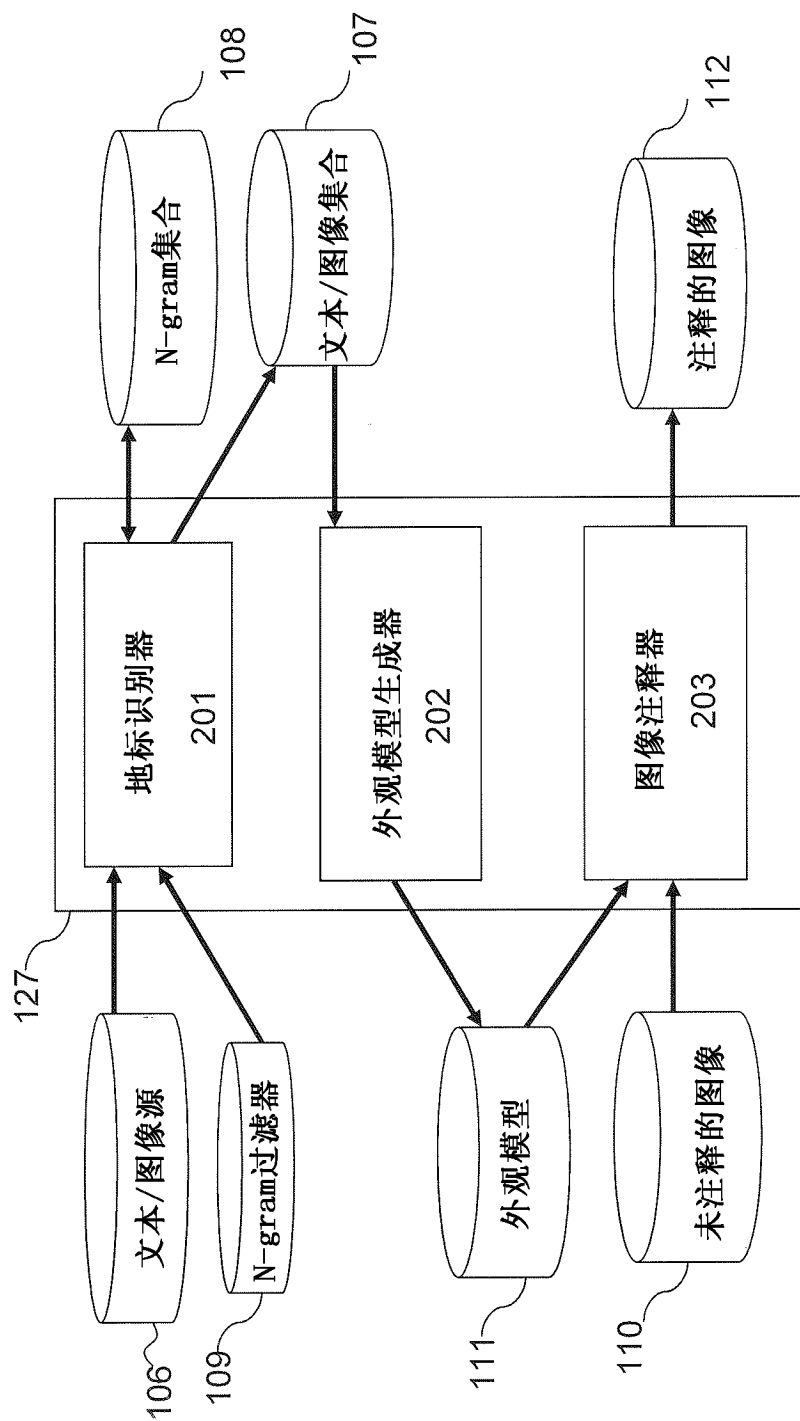


图 2

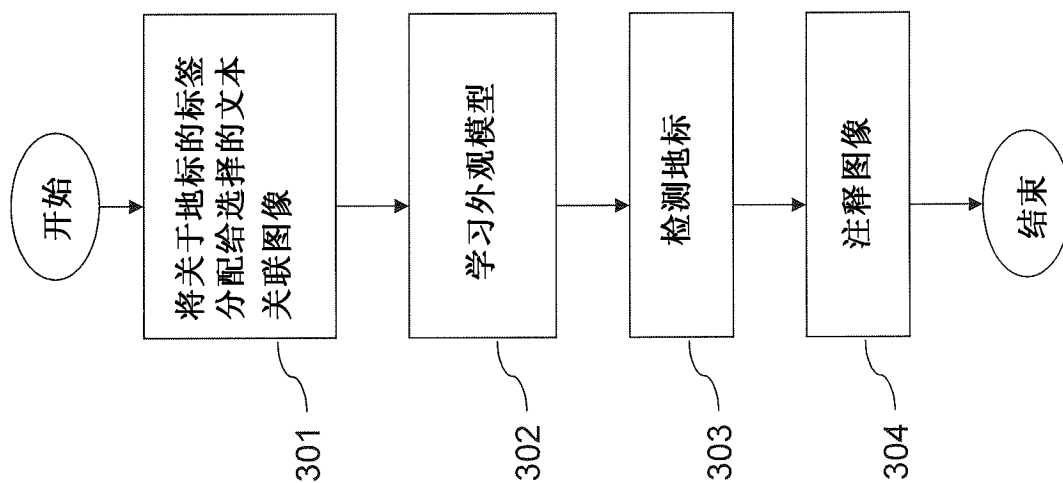
300

图 3

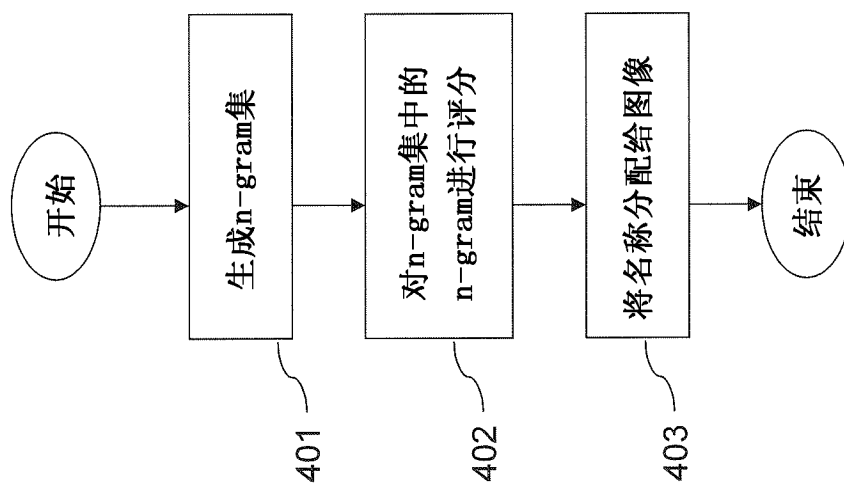
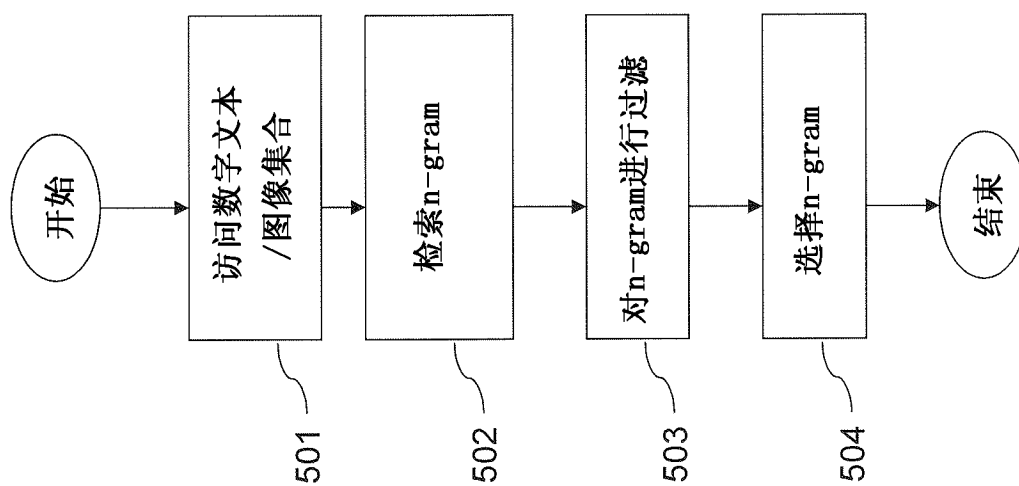
301

图 4



401

图 5

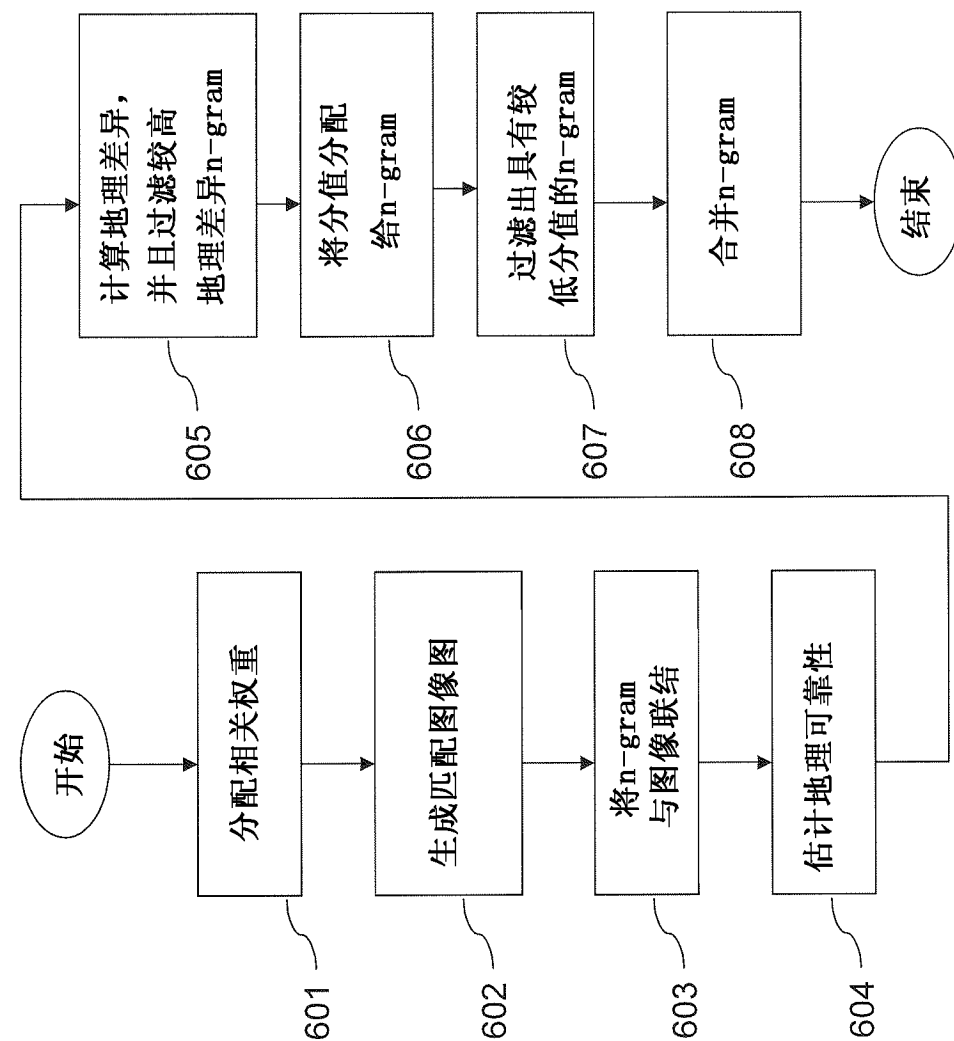


图 6