



(12) 发明专利申请

(10) 申请公布号 CN 103268317 A

(43) 申请公布日 2013. 08. 28

(21) 申请号 201310048527. 7

(22) 申请日 2013. 02. 06

(30) 优先权数据

13/367, 139 2012. 02. 06 US

(71) 申请人 微软公司

地址 美国华盛顿州

(72) 发明人 刘策 迈克尔·鲁宾斯坦

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 王萍 李春晖

(51) Int. Cl.

G06F 17/30 (2006. 01)

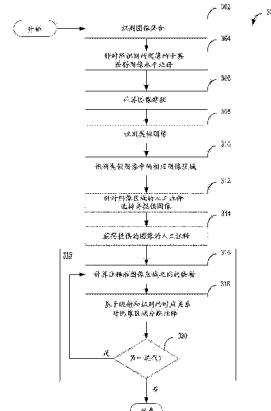
权利要求书2页 说明书20页 附图6页

(54) 发明名称

对图像进行语义注释的系统和方法

(57) 摘要

本公开涉及一种对多个图像中的图像进行语义注释的方法，所述多个图像中的每个图像包括至少一个图像区域，以及一种使得能够至少部分地基于与多个图像中的图像相关联的图像级别注释对所述图像进行基于文本的搜索的系统。上述方法包括：识别包括第一图像和第二图像的至少两个相似图像；识别所述第一图像和所述第二图像中的相应图像区域；以及使用至少一个处理器，通过使用拟合度对所述多个图像中的一个或更多个图像中的图像区域分配注释，所述拟合度指示所分配的注释和所述相应图像区域之间的匹配程度。所述拟合度可以取决于所述多个图像的子集中的每个图像的至少一个注释以及所述第一图像和所述第二图像中的图像区域之间的所识别的对应关系。



1. 一种对多个图像中的图像进行语义注释的方法 (300), 所述多个图像中的每个图像包括至少一个图像区域, 所述方法包括 :

识别包括第一和第二图像的至少两个相似图像 (308) ;

识别所述第一图像和所述第二图像中的相应图像区域 (310) ; 以及

使用至少一个处理器, 通过使用拟合度对所述多个图像中的一个或更多个图像中的图像区域分配注释 (315), 所述拟合度指示所分配的注释和所述相应图像区域之间的匹配程度,

其中, 所述拟合度取决于所述多个图像的子集中的每个图像的至少一个注释以及所述第一图像和所述第二图像中的图像区域之间的所识别的对应关系。

2. 根据权利要求 1 所述的方法 (300), 其中 :

识别相应图像区域 (310) 包括用所述第二图像中的第二图像区域识别所述第一图像中的第一图像区域; 以及

对图像区域分配注释 (315) 包括基于所述拟合度迭代地对图像区域分配注释 (318) ,

其中, 每个迭代包括至少部分基于对所述第二图像区域分配的注释来对所述第一图像区域分配注释。

3. 根据权利要求 2 所述的方法 (300), 其中, 所述第一图像包括与所述第一图像区域相邻的第三图像区域, 以及其中, 每个迭代包括 :

进一步基于对所述第三图像区域分配的注释来对所述第一图像区域分配注释 (318) 。

4. 根据权利要求 2 所述的方法 (300), 其中, 每个迭代包括 :

进一步基于所述注释的空间直方图对所述第一图像区域分配注释 (318) , 其中, 跨所述多个图像中的数个图像计算所述注释的空间直方图。

5. 根据权利要求 1 所述的方法 (300), 还包括 :

至少部分地通过从包含所述多个图像的所述子集中的图像的文档和 / 或网页中的在所述图像附近的文本中获得所述图像的至少一个图像级别注释, 来自动获得 (304) 所述至少一个注释。

6. 一种使得能够至少部分地基于与多个图像中的图像相关联的图像级别注释对所述图像进行基于文本的搜索的系统 (100), 所述系统包括 :

至少一个处理器, 被配置为 :

识别包括第一图像和第二图像的至少两个相似图像 (308) ;

识别所述第一图像和所述第二图像中的相应图像区域 (310) ; 以及

通过使用拟合度对所述多个图像中的一个或更多个图像中的图像区域分配注释 (315), 所述拟合度指示所分配的注释和所述相应图像区域之间的匹配程度,

其中, 所述拟合度取决于所述多个图像的子集中的每个图像的至少一个注释以及所述第一图像和所述第二图像中的图像区域之间的所识别的对应关系。

7. 根据权利要求 6 所述的系统 (100), 其中, 所述至少一个处理器被配置为 :

通过用所述第二图像中的第二图像区域识别所述第一图像中的第一图像区域, 来识别相应图像区域 (310) ; 以及

通过基于所述拟合度迭代地对图像区域分配注释 (316、318), 来对所述图像区域分配注释 (315) ,

其中,每个迭代(316、318)包括至少部分基于对所述第二图像区域分配的注释来对所述第一图像区域分配注释(318)。

8. 根据权利要求6所述的系统(100),其中,每个迭代包括计算所述多个图像中的所述一个或更多个图像中的所述图像区域和注释之间的统计映射(316)。

9. 根据权利要求8所述的系统(100),其中,计算所述统计映射(316)包括估计混合模型的至少一个参数,其中,所述混合模型包括针对注释集合中的每个注释的元素。

10. 根据权利要求6所述的系统(100),其中,所述多个图像的所述子集包括所述多个图像的百分之十或者更少。

对图像进行语义注释的系统和方法

技术领域

[0001] 本申请总体来说涉及图像处理领域，且更具体地涉及一种对多个图像中的图像进行语义注释的方法，以及一种使得能够至少部分地基于与多个图像中的图像相关联的图像级别注释对所述图像进行基于文本的搜索的系统。

背景技术

[0002] 图像检索技术用来帮助用户在海量的图像集中浏览、搜索和检索图像。这些技术使得用户能够在通过因特网可访问的图像和 / 或任意其它图像集中搜索他们寻找的一个或更多个图像。例如，用户可以使用搜索引擎来搜索物体（例如建筑）或者人（例如名人）的图像。为此，用户可以输入诸如“建筑”或者“名人”的搜索询问，以找到用户寻找的图像。

[0003] 搜索引擎可以基于与被搜索的图像相关联的文本注释，响应于用户的搜索询问识别一个或更多个图像。特别地，可以将用户的搜索询问与和被搜索的图像中的每个相关联的一个或更多个文本注释进行比较，并且基于比较的结果向用户呈现搜索结果。例如，如果用户正在使用搜索询问“建筑”来搜索图像，则搜索引擎可以返回使用包括词语“建筑”的文本进行了注释的图像。

[0004] 因此，图像搜索结果的质量取决于与被搜索的图像相关联的注释以及如何识别这些注释。用于图像注释的一些传统技术包含基于图像出现于其中的文档中的数据将注释与该图像相关联。例如，搜索引擎可以从该网页中的信息（诸如图像标签中的信息（例如标题、元数据等）和 / 或围绕网页中的图像的信息）中识别针对出现在该网页中的图像的注释。

发明内容

[0005] 用户可以基于与被搜索的图像相关联的文本注释来搜索图像，这样就可以对被搜索的图像集进行注释。然而，一般以注释对应于包含图像的文档（例如网页）而不是图像本身或者图像内的任意区域的方式，对许多被搜索的图像进行注释。继而，这限制了可以用来搜索图像集中的图像的搜索逻辑，并且限制了搜索引擎可以对该集中的图像编索引的方式。相应地，公开了通过对被搜索的图像的区域分配注释来对图像进行注释的技术。例如，可以对诸如像素或者一组像素的图像区域进行注释。可以至少部分基于其它图像中的相应图像区域，对图像区域分配注释。相应图像区域可以是与被注释的区域相似的图像区域，并且可以基于图像和一个或更多个图像特征之间的相似度度量来被识别。所获得的图像区域注释使得能够使用扩展搜索逻辑、例如通过搜索包含彼此相接的两个特定类型的对象的图像，来搜索图像。

[0006] 还可以在图像搜索之外的情境中应用图像的语义注释。相应地，在一些实施例中，提供一种对多个图像中的图像进行语义注释的方法，所述多个图像中的每个图像包括至少一个图像区域。该方法包括：识别包括第一图像和第二图像的至少两个相似图像；识别所述第一图像和所述第二图像中的相应图像区域；以及使用至少一个处理器，通过使用拟合

度对所述多个图像中的一个或更多个图像中的图像区域分配注释,所述拟合度指示所分配的注释和所述相应图像区域之间的匹配程度。所述拟合度取决于所述多个图像的子集中的每个图像的至少一个注释以及所识别的所述第一图像和所述第二图像中的图像区域之间的对应关系。

[0007] 在另一方面,提供一种使得能够至少部分基于与多个图像中的图像相关联的图像级别注释对所述图像进行基于文本的搜索的系统。该系统包括:至少一个处理器,被配置为:识别包括第一图像和第二图像的至少两个相似图像;识别所述第一图像和所述第二图像中的相应图像区域;以及通过使用拟合度对所述多个图像中的一个或更多个图像中的图像区域分配注释,所述拟合度指示所分配的注释和所述相应图像区域之间的匹配程度。所述拟合度取决于所述多个图像的子集中的每个图像的至少一个注释以及所识别的所述第一图像和所述第二图像中的图像区域之间的对应关系。

[0008] 在又一方面,提供至少一个计算机可读存储介质。该至少一个计算机可读存储介质存储处理器可执行指令,当由至少一个处理器执行时,该处理器可执行指令执行对多个图像中的图像进行语义注释的方法,所述多个图像中的每个图像包括一个或更多个像素。该方法包括:针对所述多个图像的子集中的每个图像获得至少一个图像级别注释;识别包括第一图像和第二图像的至少两个相似图像;识别所述第一图像和所述第二图像中的相应像素;以及使用拟合度对所述多个图像中的一个或更多个图像中的像素分配注释,所述拟合度指示所分配的注释和所述相应像素之间的匹配程度。所述拟合度取决于至少一个获得的图像级别注释以及所识别的所述第一图像和所述第二图像中的像素之间的对应关系。

[0009] 前述内容是本发明的非限制性概述,本发明由所附权利要求限定。

附图说明

[0010] 附图并非意欲按比例进行绘制。在附图中,用相似的附图标记表示在各个图中示出的每个相同或者近似相同的部件。为了清楚起见,没有在每个图中标记每个部件。在附图中:

- [0011] 图1示出了根据本公开的一些实施例的使得能够搜索图像的示例性计算环境。
- [0012] 图2A-2B示出了根据本公开的一些实施例的对图像进行语义注释的说明性示例。
- [0013] 图3是根据本公开的一些实施例的对图像进行语义注释的说明性处理的流程图。
- [0014] 图4A示出了根据本公开的一些实施例的表示相似图像的数据结构。
- [0015] 图4B示出了根据本公开的一些实施例的在相似图像对中识别相应图像区域。
- [0016] 图5是总地示出在实现本公开的各方面时可以使用的计算机系统的示例的框图。

具体实施方式

[0017] 发明人认识并且理解到改进的对图像进行语义注释的技术可以产生改进的用于浏览、搜索和/或检索这些图像的技术。故此,发明人理解,对被搜索的图像的区域进行语义注释,使得对于给定图像可以识别该图像的与特定注释相对应的部分是合乎期望的。例如,对描绘建筑、汽车和行人的图像进行注释,使得将注释“建筑”、“汽车”和“行人”分别与描绘建筑、汽车和行人的图像的部分相关联是合乎期望的。

[0018] 发明人认识到,图像区域的这种语义注释可以被使用来改进已有图像搜索技术,

并且在包括但不限于图像分类、聚类和索引的各种其它应用中使用。例如，在图像搜索的情境下，图像区域的语义注释使得能够使用取决于图像区域注释的搜索询问。例如，可以使用包括具有一个注释的图像区域（例如树）（与具有另一注释的图像区域（例如狗）邻近）的图像的搜索询问。作为另一示例，可以使用预定比例（例如至少 25%、至少 50%、至少 75% 等）的图像区域与特定注释相关联的图像的搜索询问。例如，在图像索引的情境下，图像区域的语义注释可以使得搜索引擎能够改进它们对图像编索引的方式。例如，搜索引擎可以使用图像区域注释对图像编索引，从而细化已有索引方案。许多其它示例对于本领域技术人员将是明显的。

[0019] 本发明的各方面包括通过对某人支付报酬对图像进行手动注释来获得图像区域的注释。然而，发明人理解，对每个图像中的图像区域进行手动注释不可行，因为其耗时并且昂贵。相应地，在一些实施例中，可以基于获得这些手动注释的成本度量和获得其所产生的总体图像注释性能的预期收益，来选择对图像进行手动注释。

[0020] 发明人还认识并且理解到，当可以使用与一些图像相关联的注释来获得其它图像的注释时，可以获得改进的图像注释技术。特别地，当两个图像的图像区域相似时，可以使用与一个图像相关联的注释，将注释与另一图像的区域相关联。特别地，当基于合适的图像相似度标准识别出两个图像区域相似，并且图像区域中的一个与注释相关联时，可以将相同的注释与另一图像区域相关联。例如，一个图像可以描绘汽车和建筑，并且注释“汽车”和“建筑”可以分别与描绘汽车和建筑的图像区域相关联。另一图像可以描绘汽车和加油站，但是可以不与任何注释相关联。在这种情况下，如果发现描绘汽车的两个图像的区域相似，则也可以将另一图像中的描绘汽车的区域与注释“汽车”相关联。

[0021] 发明人还认识到并且理解，最传统的图像注释技术产生图像级别注释。如先前所提及的，一种这种技术涉及根据包含图像的文档（例如网页）中的围绕该图像的文本获得注释。然而，以这种方式获得的注释与作为整体的图像相关联，而不是与图像的任意特定部分相关联。故此，被搜索的图像中的一些可能仅与图像级别注释相关联。例如，上述图像可以与注释“建筑”、“汽车”和“行人”相关联，但是不知道图像的哪些部分描绘建筑、汽车或者行人。

[0022] 传统的图像注释技术的另一缺点是，许多可能被搜索的图像根本不与任何注释相关联，更别说与如上所述的图像区域注释相关联。的确，仅非常小比例的被搜索的图像被包含关于图像的内容的信息的文本（例如陈述“上述图像包含建筑、汽车和行人”的说明性文字）围绕。

[0023] 传统图像注释技术的又一缺点是，在图像级别注释与图像相关联的情况下，许多这种注释可能不准确或者部分不准确，因为这些注释是根据围绕图像的文本获得的，而不是从图像本身得出的。例如，上述图像可能在关于汽车的文章中，故此，图像可能与注释“汽车”、而不与注释“建筑”或者“行人”相关联。然而，图像可能甚至不与注释“汽车”相关联。

[0024] 发明人认识到并且理解，至少部分基于要进行注释的图像中的图像区域之间的识别的对应关系对图像区域分配注释，可以克服上述进行图像注释的传统技术的缺点中的一些。然而，不是每个实施例都解决这些缺点中的每一个，一些实施例可能不解决它们中的任何一个。故此，应当理解，本发明不限于解决上面讨论的进行图像注释的这些传统技术的缺点中的全部或者任意一个。

[0025] 相应地,在一些实施例中,可以进行通过对图像集合中的图像的区域分配注释来对图像进行语义注释的方法。如前所述,可以以一种或者更多种方式使用这种方法,包括获得从该图像集合中浏览、搜索和 / 或检索图像的改进系统。

[0026] 在一些实施例中,可以通过使用指示分配的注释和相应图像区域之间的匹配程度的拟合度,来对一个或更多个图像中的一个或更多个图像区域进行注释。可以对图像区域分配一个或更多个注释。拟合度继而可以取决于大量因素中的任意一个,例如可以取决于识别的要进行注释的图像区域和一个或更多个其它图像区域之间的对应关系。故此,可以基于图像区域与其它图像区域的相似度和 / 或基于与其它图像区域相关联的注释,对图像区域进行注释。

[0027] 在一些实施例中,可以基于拟合度迭代地进行对一个或更多个图像区域的注释分配。在每一次迭代中,可以使用一个或更多个图像级别注释对图像区域分配注释,随后,可以至少部分基于在图像区域和一个或更多个其它图像区域之间识别的对应关系来更新分配。可以以包括、但不限于如下面参考图 3 更详细地描述的方式的大量方式中的任意一种,识别该对应关系。

[0028] 注释可以是任意合适类型的注释。在一些实施例中,注释可以是文本注释,诸如包括一个或更多个字符或数字、一个或更多个词语、一个或更多个词组、一个或更多个句子等的字母数字串。注释可以是图像级别注释,故此,注释可以与作为整体的图像相关联。注释可以是图像区域注释,故此,注释可以与图像的区域相关联。

[0029] 图像区域可以是图像的任意合适部分。图像区域可以是描绘特定物体(例如汽车)的至少一部分的图像的一部分。图像区域还可以描绘多个物体(的至少部分)、人的至少一部分和 / 或图像的任意其它可识别部分的至少一部分。例如,在描绘汽车或者建筑的图像中,图像区域可以是描绘汽车的至少一部分、建筑的至少一部分或者其任意合适的组合的任意合适区域。图像区域可以具有任意合适的大小。在一些实施例中,图像区域可以是像素或者像素组。

[0030] 应当理解,这里描述的本发明的各个方面和概念可以以大量方式中的任意一种来实现,而不限于任何特定实现技术。下面仅出于说明的目的描述具体实现的示例,而这里描述的本发明的各方面不限于这些说明性实现。

[0031] 图 1 示出了本发明的实施例可以操作的非限制性说明性环境 100。出于说明的目的,与使得用户能够搜索图像的系统相结合地描述本发明。然而,应当理解,图像搜索系统仅仅是可以应用对图像进行语义注释的技术的系统的示例,该技术可以以各种其它设置应用于诸如图像分类、聚类和 / 或索引的问题,但不限于此。

[0032] 在该说明性环境中,用户 102 可以通过向在移动设备 104 上执行的软件应用提供搜索询问,来搜索用户 102 寻找的一个或更多个图像。该软件应用可以是任意合适的应用,例如可以是诸如网页浏览器的应用。然而,应当认识到,软件应用不限于是网页浏览器,其可以是能向用户提供在任意合适的图像集合中搜索图像的接口的任意合适的应用。还应当认识到,用户不限于使用移动设备、而是可以使用任意其它合适的计算设备(例如台式计算机、膝上型计算机、平板计算机等)输入搜索询问。

[0033] 搜索询问可以是任意合适的搜索询问。在一些实施例中,搜索询问可以是文本搜索询问,并且可以是包括一个或更多个字符或数字、一个或更多个词语、一个或更多个词

组、一个或更多个句子等的字母数字串。搜索询问可以由用户以任意合适的方式输入，例如可以键入、由用户说出、由用户在一个或更多个选项中选择或者通过其任意合适的组合来输入。

[0034] 可以由服务器 108 经由网络 106 接收搜索询问，服务器 108 可以被配置为使用搜索询问在图像集中搜索一个或更多个图像。网络 106 可以是任意合适的网络，例如其可以包括因特网、内联网、LAN、WAN 和 / 或任意其它有线或无线网络、或者其组合。

[0035] 服务器 108 可以被配置为在任意合适的图像集中搜索一个或更多个图像。例如，服务器 108 可以被配置为在一个或更多个数据库（例如数据库 110 和 / 或数据库 112）中搜索图像。服务器 108 可以被配置为在本地（例如数据库 110）和 / 或远程（例如数据库 112）存储的图像中搜索图像。在一些实施例中，在因特网（或者诸如企业网络的任意其它合适的网络）上搜索图像时，视情况而定，服务器 108 可以被配置为在可存储于多个分布式位置的图像中搜索图像。应当认识到，服务器 108 可以是一个计算设备或者多个计算设备，本发明的各方面在该方面不受限制。

[0036] 不管服务器 108 可以被配置为搜索哪些图像，服务器 108 可以被配置为以任意合适的方式进行搜索。在一些实施例中，服务器 108 可以被配置为通过将用户的搜索询问与和被搜索的图像相关联的一个或更多个注释进行比较，来搜索用户可能正在寻找的图像。可以以任意合适的方式进行这种比较，因为比较用户的搜索询问和图像注释的准确方式不作为对本发明的各方面的限制。

[0037] 不管服务器 108 可以被配置为使用用户 102 提供的搜索询问搜索图像的方式如何，服务器 108 可以被配置为至少向用户 102 呈现搜索结果的子集。可以以任意合适的方式向用户 102 呈现搜索结果，呈现搜索结果的方式不作为对本发明的各方面的限制。

[0038] 在一些实施例中，服务器 108 可以被配置为对服务器 108 可以被配置为搜索的一个或更多个图像进行语义注释。例如，服务器 108 可以被配置为对数据库 110 和 / 或数据库 112 中的一个或更多个图像进行语义注释。作为另一示例，服务器 108 可以被配置为对可以经由网络 106（例如因特网、内联网等）访问的一个或更多个图像进行语义注释。然而，应当认识到，在一些实施例中，用来搜索图像的系统可以与用来对图像进行语义注释的系统不同，本发明的各方面在该方面不受限制。

[0039] 服务器 108 可以被配置为以任意合适的方式对一个或更多个图像进行语义注释。在一些实施例中，服务器 108 可以被配置为对服务器 108 可以被配置为进行语义注释的图像中的图像区域分配一个或更多个注释。故此，服务器 108 可以被配置为使用指示分配的注释和相应图像区域之间的匹配程度的拟合度来分配注释。下面参考图 3 和 4A-4B 对此进行更详细的描述。

[0040] 图 2A 和 2B 示出了对图像进行语义注释的说明性示例。特别地，图 2A 示出了说明性图像 200，图像 200 示出了通过树 204 而与海 206 分离的天空 202。如服务器 108 可以被配置为进行的那样对图像 200 进行语义注释，可以导致对图像 200 中的区域的注释分配。在图 2B 中作为分配 210 示出了一个这种分配，分配 210 对图像 200 中的每个像素分配集合 {“树”、“天空”和“海”} 中的注释。像素集合 212 包括各自分配了注释“天空”的像素。像素集合 214 包括各自分配了注释“树”的像素。像素集合 216 包括各自分配了注释“海”的像素。

[0041] 应当认识到,虽然在所示出的实施例中对图像 200 的每个像素分配了注释,但是对图像进行语义注释不限于对图像中的所有像素进行注释。例如,在一些实施例中,可以对包括多个像素的图像区域进行注释。作为另一示例,可以仅对图像的像素的子集进行注释。应当理解,虽然在所示出的实施例中对每个图像区域分配了单个注释,但是这不限制本发明的各方面,因为可以对图像区域分配一个或更多个注释。这可以以任意合适的方式来完成,例如,这可以通过图像的分层表示来完成。

[0042] 如前所述,服务器 108 可以被配置为基于用户提供的搜索询问与和被搜索的图像相关联的一个或更多个注释来搜索图像。在图 3 中示出了获得这些注释的一种方法,图 3 示出了对图像进行语义注释的说明性处理 300。处理 300 可以由诸如参考图 1 描述的系统 100 的被配置为搜索图像的系统来执行,或者由被配置为对一个或更多个图像进行语义注释的任意其它合适的系统来执行。

[0043] 处理 300 在动作 302 中开始,在动作 302 识别要进行注释的图像集合。要进行注释的图像集合可以是任意合适的图像集合,例如可以是经由网络(例如因特网、内联网等)可访问的图像集合和 / 或存储在一个或更多个数据库中的图像集合。可以以任意合适的方式识别要进行注释的图像集合。在一些实例中,可以手动指定(例如由用户、管理员、在配置文件中等)要进行注释的图像集合。另外或者可替选地,可以自动(例如通过访问一个或更多个网页、一个或更多个文档、存储一个或更多个图像的一个或更多个数据库等)识别要进行注释的图像集合。

[0044] 在动作 302 中识别要进行语义注释的图像集合之后,处理 300 进行到动作 304,在动作 304 中针对所识别的图像的子集中的每个图像获得一个或更多个图像级别注释。可以针对所识别的图像的任意合适的子集中的每个图像获得图像级别注释。例如,可以针对包括所识别的图像的百分之 25 或更少、所识别的图像的百分之 5 或更少、所识别的图像的百分之 1 或更少等的子集,来获得图像级别注释。

[0045] 可以以大量方式中的任意一种来获得图像级别注释。在一些实例中,可以根据与图像相关联的数据获得图像的一个或更多个图像级别注释。与图像相关联的数据可以是任意合适的数据,例如可以包括:包含图像的文档(例如网页、文章、电子邮件中的文本等)中的数据、与图像相关联的元数据(例如图像头中的信息)和 / 或与图像相关联的大量其它类型的数据中的任意一种。在一些实例中,可以自动(例如通过访问一个或更多个网页、一个或更多个文档、存储一个或更多个图像的一个或更多个数据库等)获得图像级别注释。另外或可替选地,可以手动指定一个或更多个图像级别注释。

[0046] 应当认识到,处理 300 不限于仅获得图像级别注释,且可选地在动作 304 中,可以针对所识别的图像的集合中的一个或更多个图像获得一个或更多个图像区域注释。

[0047] 接下来,处理 300 进行到动作 306、308 和 310,其中,可以识别(在动作 302 中识别的图像的)图像区域之间的对应关系。如前所述,可以使用该对应关系来改善对图像区域分配的注释,因为两个图像区域之间的相似性可以指示可以对这两个图像区域分配相同的注释。

[0048] 为了识别图像区域之间的对应关系,处理 300 首先进行到动作 306,其中,根据在动作 302 中获得的图像集合中的图像计算图像特征。然而,应当认识到,在动作 306 中计算的特征可以用于任意合适的目的,而不限于仅用于识别图像区域之间的对应关系。

[0049] 作为动作 306 的一部分,可以针对图像计算大量类型的图像特征中的任意特征。在一些实施例中,可以针对图像中的一个或更多个图像区域计算局部图像特征。例如,可以针对图像中的一个或更多个像素和 / 或图像中的一组或更多组相邻像素,计算局部图像特征。局部图像特征可以指示局部图像结构、局部颜色信息和 / 或任意其它合适类型的信息。可以根据本领域中已知的处理来获得该局部图像特征。例如,可以针对一个或更多个图像区域中的每个计算尺度不变特征变换 (scale-invariant feature transform, SIFT) 特征。作为另一示例,可以针对一个或更多个图像区域中的每个计算方向梯度直方图 (histogram of oriented gradients, HOG) 特征。另外或可替选地,可以作为动作 306 的一部分,针对图像计算全局图像特征 (例如“GIST”特征)。故此,可以针对图像中的一个或更多个图像区域 (例如像素或者像素组) 中的每个,计算多个特征 (例如几十个特征、几百个特征等)。下面,矢量 $D_i(p)$ 可以表示针对第 i 个图像的第 p 个图像区域计算的特征。

[0050] 作为具体的非限制性示例,可以针对图像中的一个或更多个像素中的每个计算 SIFT 和 / 或 HOG 特征。可以使用像素附近的一个或更多个像素集合 (例如该像素三个像素内的像素、该像素七个像素内的像素),来计算该像素的 SIFT 特征,以考虑特征尺度。另外,可以使用矩形小块 (例如 2×2 的小块) 的像素来计算 HOG 特征。

[0051] 如前所述,在动作 306 期间针对每个图像区域计算的特征的数量可能较大。相应地,在一些实施例中,使用包括主分量分析 (principal components analysis, PCA)、加权 PCA、局部线性嵌入的在本领域中已知的任意合适的维降技术和 / 或任意其它线性或非线性维降技术,可以降低与每个图像区域相关联的特征的数量。故此,可以将任意合适数量的特征 (例如 5、10、25、50、75、100 等) 与每个图像区域相关联。

[0052] 接下来,处理 300 进行到动作 308,其中,可以识别一组或更多组相似图像,每个这种组包括来自在动作 302 处识别的图像集合的至少两个图像。在一些实施例中,可以通过识别与在动作 302 中识别的图像集合中的每个图像的一个或更多个相似图像,来识别一组或更多组相似图像。可以使用指示图像对之间的相似程度的相似度度量,来识别一组相似图像。特别地,可以使用任意合适的聚类算法,通过将计算的相似度度量大于预定阈值的任意图像对识别为相似的,来识别相似图像组。

[0053] 在一些实施例中,计算两个图像之间的相似度度量可以包括计算与两个图像中的每个相关联的特征之间的距离。为此,可以使用任意合适的距离函数和任意合适的图像特征 (例如在动作 306 处计算的任意图像特征)。在一些实例中,计算两个图像之间的相似度度量可以包括计算与两个图像中的每个相关联的全局特征之间的欧几里德距离。然而,应当认识到,可以使用任意其它合适的相似度度量。

[0054] 在一些实施例中,将在动作 306 中识别的相似图像组表示为如下数据结构可能是方便的,该数据结构表现了包括表示图像的顶点和表示图像之间的相似度的边的曲线图。在图 4A 中示出了一个这种曲线图,图 4A 示出了包括由边 404 连接的节点 402 和 406 的说明性曲线图 400。边 404 的存在可以指示针对由节点 402 和 404 表示的图像对计算的相似度度量大于预定阈值。应当理解,曲线图中的边不必局限于是对称的,而在一些实例中可以是有方向的边。应当理解,图 4A 所示的曲线图仅仅是说明性的,实际上,曲线图可以包括任意合适数量的节点 (例如至少 100 个、至少 1000 个、至少 10000 个、至少 100000 个、至少一百万个、至少一千万个等),以表示大的图像集合中的图像。

[0055] 接下来,处理 300 进行到动作 310,在动作 310 中可以识别相似图像中的相应图像区域。可以在动作 308 中识别为相似的一个或更多个图像对中识别相应图像区域。例如,可以针对由诸如图 4A 所示的说明性曲线图的曲线图中的连接的顶点表示的任意两个图像,识别相应图像区域。识别相似图像对中的相应图像区域可以包括以另一图像中的一个或更多个区域识别一个图像中的一个或更多个区域。图 4B 示出了识别图像对中的相应图像区域的说明性而非限制性示例。在所示出的示例中,识别出图像 402 的图像区域(例如像素)408 与图像 406 的图像区域(例如像素)410 相对应。

[0056] 可以以任意合适的方式识别两个图像之间的相应图像区域。可以至少部分基于指示图像区域之间的相似程度的目标函数,来识别对应关系。可以使用任意合适的目标函数。在一些实例中,目标函数可以至少部分取决于与图像区域相关联的图像特征。可以使用任意合适的特征,包括但不限于在处理 300 的动作 306 中计算的特征中的任意特征。

[0057] 在一些实施例中,目标函数可以取决于本领域中已知的与图像区域相关联的图像特征之间的距离的大量度量(例如欧几里德距离、 l_1 距离、 l_p 距离等)中的任意一个。另外或者可替选地,目标函数可以包括所谓的“规则化”项,以使目标函数对特征值的小的变化较不敏感。作为具体而非限制性示例,可以使用下面的目标函数(所谓的“SIFT 流”目标函数),来识别图像 I_i 和图像 I_j 之间的相应图像区域:

$$[0058] E(\mathbf{w}) = \sum_{\mathbf{p} \in A_i} |S_i(\mathbf{p}) - S_j(\mathbf{p} + \mathbf{w}(\mathbf{p}))| + \alpha \sum_{\mathbf{p}, \mathbf{q} \in N(\mathbf{p})} |\mathbf{w}(\mathbf{p}) - \mathbf{w}(\mathbf{q})|$$

[0059] 在上面的方程式中, $w(p)$ 表示被识别为与图像 I_i 中的图像区域 p 相对应的图像 I_j 中的区域。此外, A_i 表示图像 I_i 的图像区域的集合(例如在图像区域是单个像素时的情况下图像 I_i 的格子), $N(p)$ 表示(如通过任意合适距离函数所测量的)靠近图像区域 p 的图像区域, α 是用来控制规则化项的效果的调整参数。在一些实施例中,可以将 α 设置为 0 和 1 之间的任意合适的数字,例如可以将 α 设置为 0.1 或 0.25 或 0.5 或者捕捉流场 $w(p)$ 的空间统计特性的任意其它合适的值。

[0060] 可以使用上述目标函数(或者任意其它合适的目标函数)来识别图像 I_i 和图像 I_j 之间的相应图像区域,以获得将图像 I_i 中的图像区域与图像 I_j 中的图像区域相关联的映射 w 。这可以使用包括但不限于置信传播和优化技术的大量接口算法中的任意一个来进行,优化技术包括但不限于梯度下降和期望最大化。

[0061] 接下来,处理 300 进行到可选的动作 312 和 314(如由虚线所指示),其中,可以从一个或更多个人工注释员获得图像区域注释。如先前所述,对大量图像中的图像区域进行手动注释可能是耗时并且昂贵的。然而,在可以使用该资源的实施例中,从一个或更多个人工注释员获得在动作 302 中识别的图像的子集的图像区域注释是有利的。

[0062] 相应地,在一些实施例中,可以将在动作 302 中识别的图像的图像子集(例如少于 2%、少于 1%、少于 0.1%、少于 0.01% 等)提供给一个或更多个人工注释员,从而他们可以将注释分配给每个这种区域中的一个或更多个图像区域。虽然如此,应当认识到,可以选择任意合适数量的图像来进行人工注释。

[0063] 可以以任意合适的方式选择向人工注释员提供的图像的数量。在一些实施例中,可以基于本领域中已知的指示确定的对图像区域的注释分配的准确度的大量因素中的任意一个,来选择图像的数量。这些因素的示例包括但不限于真肯定、假肯定、假否定、真否

定、假警率、漏检率的数量以及诸如精度和检索率的从其得出的量。可以以任意合适的方式获得该量。例如，可以使用具有已知图像区域注释的试验性图像(pilot image)集合或者以任意其它合适的方式，来估计该量。在一些实施例中，可以基于与对图像进行手动注释相关联的金钱成本来选择图像的数量。故此，可以基于手动注释可使用的总体预算来选择图像的数量。

[0064] 在一些实施例中，可以使用依据上述因素中的任意一个的目标函数，来选择图像的数量。目标函数可以以任意合适的方式取决于上述识别的因素或者其它因素中的任意一个。作为具体示例，目标函数可以由下式给出：

$$[0065] F(\rho_t, \rho_1) + \alpha C(\rho_t, \rho_1),$$

[0066] 其中，第一项对应于所谓的 F 度量，描述标记的准确性的特征，由下式给出：

$$[0067] F(\rho_t, \rho_1) = \frac{(\beta^2 + 1) PR}{\beta^2 P + R}$$

[0068] 其中， ρ_t 是具有图像级别注释的图像的百分比， ρ_1 是具有由人工注释员提供的图像区域注释的图像的百分比，P 和 R 分别是相应的精度和检索率。参数 β 可被设置为小于 1 以强调精度，且可被设置为大于 1 以强调检索率。上述目标函数中的第二项是反映根据 ρ_1 获得人工注释的成本的成本函数。最后，可以设置参数 α 以平衡 F 度量的相对重要型和成本函数的相对重要性。

[0069] 不管针对人工注释要提供的图像的数量为何，可以以任意合适的方式选择向人工注释员提供的图像。在一些实例中，可以随机选择向人工注释员提供的图像。然而，在其它实例中，可以至少部分基于在处理 300 的动作 308 中识别的相似图像组来选择图像。例如，在一些实施例中，可以基于表示图像且表示图像之间的相似度的曲线图（例如图 4A 所示的曲线图）来选择图像。这可以以任意合适的方式进行，例如可以至少基于曲线图的结构进行。例如，可以使用页排序型算法来识别要选择的图像。可以使用这种算法，来识别与没有注释和 / 或仅具有部分注释的大图像组相似的图像，并且选择这种图像由一个或更多个人工注释员进行注释。

[0070] 另外或可替选地，可以通过首先将在动作 302 中识别的图像聚类到预定数量的组中，然后选择最接近每个组的中心的图像作为要向人工注释员提供的图像，来识别要选择的图像。在这种情况下，聚类以及图像到图像组的中心的接近程度的确定可以使用指示图像之间的相似度的任意合适的相似度度量来进行，例如，可以使用包括前面讨论的任意相似度度量的任意合适的相似度度量来进行。预定数量的聚类可以是任意合适的数量，在一些实例中，其可以取决于要针对人工注释提供的图像的数量。

[0071] 不管针对人工注释选择图像的方式为何，在动作 312 中，向一个或更多个人工注释员提供所选择的图像。可以以任意合适的方式向人工注释员提供图像，这不限制本发明的各方面。接下来，处理 300 进行到动作 314，其中，可以获得由人工注释员进行的一个或更多个注释。在一些实施例中，可以获得来自人工注释员的附加输入，包括但不限于与由该注释员提供的一个或更多个注释相关联的置信量的指示。可以以任意合适的方式获得注释，这不限制本发明的各方面。

[0072] 接下来，处理 300 进行到包括动作 316 和 318 以及决定块 320 的动作集合 315，其中，可以对在动作 302 中识别的一个或更多个图像区域分配一个或更多个注释。这可以以

任意合适的方式进行,例如,可以如前所述使用指示分配的注释和相应图像区域之间的匹配程度的拟合度来进行。

[0073] 拟合度可以取决于大量因素,包括但不限于与在动作 302 中识别的图像相关联的图像级别注释(例如在动作 302 中获得)、在动作 306 中计算的图像特征中的一个或更多个、在动作 310 中识别的要进行注释的图像区域之间的对应关系以及在可选动作 314 中获得的注释(如果有的话)。虽然如此,应当认识到,上面列出的因素是说明性的,除了或者代替上面列出的因素,拟合度还可以取决于大量其它因素中的任意一个。

[0074] 在一些实施例中,可以迭代地对图像区域分配注释。可以针对每个这种迭代进行动作集合 315。在每个迭代中,如针对动作 316 讨论的,可以获得注释和图像区域之间的映射。映射可以识别可以对特定图像区域分配的一个或更多个注释,并且映射例如可以是如下面进一步详细描述的统计映射。随后,如针对动作 318 讨论的,可以至少部分基于所计算的映射以及在图像区域和一个或更多个其它图像区域之间识别的对应关系(例如在动作 310 中识别的对应关系),来对图像区域分配注释。

[0075] 首先,在动作 316 中,可以获得注释和图像区域之间的映射的初始估计。映射可以是任意合适的注释集合与任意合适的图像区域集合之间的映射。图像区域的集合可以包括在动作 302 中识别的图像的一个或更多个区域。注释集合可以包括在动作 304 中获得的注释中的任意一个和 / 或在动作 314 中获得的注释中的任意一个。虽然如此,应当认识到,本发明的各方面不限于使用在动作 304 和 / 或 314 中获得的注释,而可以使用以任意合适的方式从任意其它合适的源(例如字典、百科全书、任意文档或者文档集的内容等)获得的注释。

[0076] 映射可以是任意合适的类型的映射。在一些实例中,映射可以是对特定注释对应于特定图像区域的似然性分配值的统计映射。在一些实施例中,可以使用生成性概率模型来实现这种映射,生成性概率模型可以用来获得特定注释对应于特定图像区域的概率。下面描述这种生成性概率模型的一个具体示例。虽然如此,应当认识到,映射不限于通过使用生成性概率模型来实现,而可以利用其它方法(例如基于随机森林的辨别方法)。

[0077] 可以利用任意合适的生成性概率模型。例如,在一些实施例中,可以利用大量类型的混合模型中的任意一种。例如,可以利用包括用于注释的集合中的一个或更多个注释的混合元素的混合模型。作为另一示例,可以利用包括用于注释的集合中的每个注释的元素的混合模型,从而如果注释的集合包括 L 个注释,则混合模型可以包括 L 个元素。

[0078] 现在描述生成性概率模型的具体说明性示例。为此,设 $c_i(p)$ 表示将 L 个注释中的一个分配给图像 I_i 中的图像区域 p。在该示例中,生成性概率模型是包括用于 L 个注释中的每个的混合元素的混合模型,其由下式给出:

$$[0079] P(c_i(p); \Theta) = \sum_{l=1}^L \left(\rho_{i,l}(p) \sum_{k=1}^M \pi_{l,k} \mathcal{N}(D_i(p); \mu_{l,k}, \Sigma_{l,k}) \right)$$

$$[0080] + \rho_{i,\epsilon}(p) \mathcal{N}(D_i(p); \mu_\epsilon, \Sigma_\epsilon)$$

[0081] 注意,上述指定混合模型中的每个元素是具有 M 个分量的高斯混合模型。由生成特征 $D_i(p)$ 的第 1 个高斯混合模型的权重 $\rho_{i,1}(p)$,对每个高斯混合模型进行加权。变量 $\pi_{1,k}$, $\mu_{1,k}$ 和 $\Sigma_{1,k}$ 分别是高斯混合模型 1 中的分量 k 的混合权重、平均值和协方差。另外,

上面指定的混合模型包括异常值模型,其中, $\rho_{i,\epsilon}(p)$ 和 $\mu_{i,\epsilon}$, $\Sigma_{i,\epsilon}$ 是图像 I_i 中的每个图像区域 p 的异常值模型的权重、平均值和协方差参数。故此,设 θ_i 表示第 i 个高斯混合模型的参数,则根据下式给出上面的生成性概率模型的所有参数的矢量:

$$[0082] \quad \Theta = (\{\rho_{i,l}\}_{i=1:N, l=1:L}, \{\rho_{i,\epsilon}\}_{i=1:N}, \theta_1, \dots, \theta_L, \theta_\epsilon)$$

[0083] 作为处理 300 的动作 316 的一部分,可以以任意合适的方式获得从注释到图像区域的映射。可以由一个或更多个参数指定映射,并且可以至少部分基于在动作 306 中获得的图像特征和在动作 302 中获得的图像级别注释来计算参数。另外,在一些实施例中,可以基于可在处理 300 的动作 314 中从一个或更多个人工注释员获得的任意图像区域注释来计算参数。

[0084] 可以使用任意合适的技术根据获得的注释和图像特征计算映射的一个或更多个参数。例如,可以使用包括但不限于最大似然法和贝叶斯参数估计技术的大量参数估计技术中的任意一个。在一些实施例中,可以使用期望最大化(EM)算法。例如,可以使用 EM 算法来估计上面指定的混合模型的参数 θ 。在这种情况下,可以获得参数 θ 的初始估计值,并且可以使用一个或更多个迭代来迭代地细化该初始估计值。

[0085] 可以以大量方式中的任意一种获得初始估计值。例如,可以通过使用任意合适的群集算法(例如 K-means)将在动作 306 中获得的图像特征聚类到 L 个簇中来获得初始估计值,并且可以使用任意合适的技术将高斯混合模型拟合到每个簇。可以根据从在动作 302 中识别的图像中选择的随机选择的图像区域将异常值模型初始化。为了解释(account for)部分注释,可以基于图像是与图像区域注释(可能在动作 314 中已获得)、图像级别注释(在动作 302 中获得)相关联、还是不与注释相关联,来对根据图像获得的特征进行加权。从与图像区域注释相关联的图像获得的特征的权重高于从与图像级别注释相关联的图像获得的特征的权重,从与图像级别注释相关联的图像获得的特征的权重又高于从没有进行注释的图像获得的特征的权重。

[0086] 在获得参数 θ 的初始估计值之后,与获得初始估计值的方式无关地,例如可以使用最大似然估计值来细化该估计值。在一些实例中,可以利用修正的最大似然估计值,以促进高斯混合模型之间的对比。在这种情况下,可以根据下式来更新高斯混合模型的均值:

$$[0087] \quad \mu_{l,k}^{(n+1)} = \frac{\sum_{i,p \in \Lambda_i} w_i \gamma_{i,l,k}(p) D_i(p) - \alpha \eta}{\sum_{i,p \in \Lambda_i} w_i \gamma_{i,l,k}(p) - \alpha}$$

[0088] 其中,

$$[0089] \quad \gamma_{i,l,k}(p) = \frac{\rho_{i,l}(p) \pi_{l,k} G_{l,k}(D_i(p))}{\sum_{j=1}^L \rho_{i,j}(p) GMM_j(D_i(p)) + \rho_{i,\epsilon} G_\epsilon(D_i(p))}$$

$$[0090] \quad GMM_l(x) = \sum_{k=1}^M \pi_{l,k} G_{l,k}(x)$$

[0091]

$$G_{l,k}(x) = \mathcal{N}(x; \mu_{l,k}, \Sigma_{l,k})$$

[0092] 以及

$$[0093] \quad \eta = \arg \min_{\mu_{j \neq l,m}^{(n)}} \left\| \mu_{l,k}^{(n)} - \mu_{j,m}^{(n)} \right\|$$

[0094] 其中， α 是可以以任意合适的方式设置的调整参数。

[0095] 虽然如此，应当认识到，可以使用大量其它方式中的任意一种来更新高斯混合模型的参数，在该方面不对本发明的各方面构成限制。

[0096] 相应地，可以根据下式计算可以将注释 1 映射到图像区域 p 的概率：

$$[0097] \quad \rho_{i,l}(\mathbf{p}) = \frac{\text{GMM}_l(D_i(\mathbf{p}))}{\sum_{j=1}^L \text{GMM}_j(D_i(\mathbf{p})) + G_\epsilon(D_i(\mathbf{p}))}.$$

[0098] 不管在动作 316 中计算从注释到图像区域的映射的方式为何，处理 300 接下来进行到动作 318，在动作 318 中可以至少部分基于映射以及在图像区域和一个或更多个其它图像区域之间识别的对应关系（例如在动作 310 中识别的对应关系），对图像区域分配一个或更多个注释。

[0099] 在一些实施例中，可以基于取决于在动作 316 中获得的映射的拟合度和图像区域之间的所识别的对应关系来获得分配。对图像区域分配的注释可以是使拟合度的值最优化的那些注释。例如，在一些实施例中，对图像区域分配的注释可以对应于使用拟合度和最大后验概率准则计算的注释。虽然如此，应当认识到，可以与拟合度一起使用任意其它合适的准则来获得对图像区域的注释分配。

[0100] 如前所述，通过使用拟合度来对图像区域分配注释，可以至少部分基于对一个或更多个其它图像区域（诸如被识别为与特定图像区域对应的图像区域、在包括该特定图像区域的图像中与该特定图像区域相邻的图像区域和 / 或任意其它合适的图像区域）分配的注释，来对该特定图像区域分配注释。

[0101] 可以使用取决于上述因素中的一个或更多个的大量拟合度中的任意一个。由下式给出拟合度的一个具体示例：

$$[0102] \quad E(c) = \sum_{i=1}^N \sum_{\mathbf{p} \in \Lambda_i} \left[\Phi_p(c_i(\mathbf{p})) + \Phi_s(c_i(\mathbf{p})) + \Phi_c(c_i(\mathbf{p})) \right]$$

$$[0103] \quad + \sum_{j \in \mathcal{N}(i)} \Psi_{ext}(c_i(\mathbf{p}), c_j(\mathbf{p} + \mathbf{w}_{ij}(\mathbf{p})))$$

$$[0104] \quad + \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} \Psi_{int}(c_i(\mathbf{p}), c_i(\mathbf{q})) \right]$$

[0105] 根据下式来定义该拟合度中的第一项：

$$[0106] \quad \Phi_p(c_i(\mathbf{p}) = 1) = -\log P(c_i(\mathbf{p}) = 1; \Theta) - \log P_i(1)$$

[0107] 其中，

$$[0108] \quad P_i(l) = \frac{\beta}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \delta[l \in \mathbf{t}_j] + \frac{1-\beta}{Z} \sum_{j \in \mathcal{N}(i)} \sum_{m \in \mathbf{t}_j} h^o(l, m)$$

[0109] 并且

$$[0110] Z = \sum_{j \in N(i)} |t_j|$$

[0111] 其中, β 是可以以任意合适的方式设置的调整参数。

[0112] 应当理解,将第一项定义为 $-\log P(c_i(p) = 1; \Theta)$ 和 $\log P_i(1)$ 之间的差。可以基于在处理 300 的动作 316 中计算的映射来计算该差中的第一项。故此,说明性拟合度取决于计算的映射。特别地,给定参数 θ ,可以作为注释 1 与第 i 个图像中的第 p 个图像区域相关联的概率的对数,来获得该第一项。可以以包括、但不限于针对动作 316 描述的方式的任意合适的方式来获得参数 θ 。

[0113] 差中的第二项(即 $\log(P_i(1))$)是注释的先验概率,其反映了在处理 300 的动作 308 中被识别为相似的图像中的注释的共同出现的频率。特别地,该先验概率中的第一项可以测定在处理 300 的动作 308 中被识别为与第 i 个图像相似的一个或更多个图像中的第 1 个注释的频率。第二项经由矩阵 h^o 反映了共同出现的频率, h^o 是 $L \times L$ 行归一化的注释共同出现矩阵,其可以根据在动作 316 中获得的注释估计值(如果这是动作集合 315 的第一次迭代)或者根据在动作 318 的先前的迭代中获得的注释(如果这不是动作集合 315 的第一次迭代)进行计算。

[0114] 根据下式定义的拟合度中的第二项:

[0115] $\Phi_s(c_i(p) = l) = -\lambda_s \log h_l^s(p)$, 是空间先验概率,其至少部分基于调整参数 λ_s 和使用在处理 300 的动作 302 中识别的图像中的一个或更多个图像计算的第一个注释的空间直方图($h_l^s(p)$),提供关于是否可以对第 i 个图像中的第 p 个图像区域分配第一个注释的指示,可以以任意合适的方式设置参数 λ_s 。

[0116] 根据下式定义的拟合度的第三项:

$$[0117] \Phi_c(c_i(p) = l) = -\lambda_c \log h_{i,l}^c(p),$$

[0118] 是颜色先验概率,其至少部分基于调整参数 λ_c 和第 i 个图像中的第 1 个注释的颜色直方图($h_{i,1}^c(p)$),提供关于是否可以对第 i 个图像中的第 p 个图像区域分配注释 1 的指示,可以以任意合适的方式设置参数 λ_c 。颜色直方图可以以大量方式中的任意一种计算,例如,可以针对每个颜色通道的多个箱(bin)计算。

[0119] 根据下式定义拟合度中的第四项:

$$[0120] \Psi_{int}(c_i(p) = 1_p, c_i(q) = 1_q) = -\lambda_o \log h^o(1_p, 1_q) + \delta [1_p \neq 1_q] \\ \exp(-\lambda_{int} || I_i(p) - I_i(q) ||)$$

[0121] 其提供是否对第 i 个图像内的图像区域对一致地分配注释的指示。故此,当基于拟合度分配注释时,包括该第四项可能暗示可以基于对图像中的一个图像区域分配的注释来对同一图像中的另一图像区域分配注释。参数 λ_o 和 λ_{int} 是可以以任意合适的方式设置的调整参数。

[0122] 根据下式定义拟合度中的第五项:

$$[0123] \Psi_{ext}(c_i(p) = l_p, c_j(r) = l_r) =$$

$$[0124] \delta [l_p \neq l_r] \frac{w_j}{w_i} \exp(-\lambda_{ext} |S_i(p) - S_j(r)|)$$

[0125] 其提供是否对在处理 300 的动作 310 中识别的相应图像区域对一致地分配注释的

指示,其中,例如根据下式,第 j 个图像中的第 r 个图像区域对应于第 i 个图像中的第 p 个图像区域:

$$[0126] \quad r = p + w_{ij}(p)$$

[0127] 直观地,当对相应图像区域分配不同的注释时,第五项可能产生较高的损失(penalty)。这里,参数 λ_{int} 是可以以任意合适的方式设置的调整参数。

[0128] 相应地,上述说明性的拟合度取决于以下参数:

$$[0129] \quad \{c, \Theta, \{\mathbf{h}_l^s\}_{l=1:L}, \{\mathbf{h}_{i,l}^c\}_{i=1:N, l=1:L}, \mathbf{h}^o\}$$

[0130] 其可以使用大量技术中的任意一种以大量方式中的任意一种进行估计。例如,可以使用大量优化和 / 或推断技术中的任意一种,包括但不限于坐标下降、消息传递、置信度传播、最大后验概率估计和迭代条件模式。

[0131] 另外,上述说明性的拟合度取决于以下参数:

$$[0132] \quad \{\alpha, \beta, \lambda_s, \lambda_c, \lambda_o, \lambda_{int}, \lambda_{ext}\}$$

[0133] 其可以以任意大量方式设置,以控制上述各项影响拟合度的方式。这些值可以是任意合适的值,这不对本发明的各方面构成限制。

[0134] 在动作 318 完成之后,处理 300 进行到决定块 320,其中,可以确定是否可以进行动作 316 和 318 的另一迭代。该确定可以以大量方式中的任意一种进行,进行该确定的方式不对本发明的各方面构成限制。如果确定进行另一迭代,则处理 300 经由“是”分支循环回动作 316,重复进行动作 316 和 318。另一方面,如果确定不进行另一迭代,则处理 300 完成。

[0135] 应当认识到,处理 300 仅仅是说明性的,可以对处理 300 进行许多变化。例如,虽然在示出的实施例中,在动作集合 315 中对图像区域分配注释之前获得所有图像级别注释以及可选地获得人工进行的注释,但是在其它实施例中,可以在对图像区域分配注释之后,获得附加注释。这在可能检测到某些图像的注释可以改善注释和分配了注释的图像区域之间的匹配时是有利的。作为另一示例,虽然在示出的实施例中,使用迭代算法来分配注释,但是在其它实施例中,可以使用任意其它非迭代型方法来分配注释。作为又一示例,虽然在说明性实施例中未示出,但是可以使用在执行动作 318 之后获得的注释的估计值来改进在动作 310 中识别相似图像的方式。故此,在一些实施例中,处理 300 可以在动作 318 完成之后循环回动作 310,并且可以进一步基于在动作 318 中分配的注释识别图像是相似的。这可以以任意合适的方式进行。例如,可以将至少共享特定数量的相同图像区域注释的图像识别为相似。仍然有其它变化对于本领域技术人员是明显的。

[0136] 如前所述,包括一个或更多个注释图像区域(例如经由处理 300 或者处理 300 的大量变化中的任意一个进行了注释)的图像可以用于包括但不限于图像搜索的各种应用,由此可以至少部分基于图像区域的注释对图像进行搜索。可以以包括但不限于先前针对图 1 描述的方式的大量方式中的任意一种进行这种搜索。

[0137] 图 5 示出了可以实现本发明的合适的计算系统环境 500 的示例。计算系统环境 500 仅仅是合适的计算环境的一个示例,其不旨在对本发明的使用范围或者功能提出任何限制。也不应当将计算环境 500 解释为具有与在示例性操作环境 500 中示出的部件中的任意一个或者其组合相关的任何依赖性或者要求。

[0138] 本发明可使用大量其它通用或者专用计算系统环境或者配置来操作。适合于本发明使用的公知的计算系统、环境和 / 或配置的示例包括但不限于个人计算机、服务器计算机、手持或者膝上型设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费电子设备、网络 PC、微型计算机、大型计算机、包括上述系统或者设备中的任意一个的分布式计算环境等。

[0139] 该计算环境可以执行诸如程序模块的计算机可执行指令。通常，程序模块包括进行特定任务或者实现特定抽象数据类型的例程、程序、对象、组件、数据结构等。本发明也可以在由通过通信网络链接的远程处理设备执行任务的分布式计算环境中实施。在分布式计算环境中，程序模块可位于包括存储器存储设备的本地和远程计算机存储介质两者中。

[0140] 参考图 5，用于实现本发明的示例性系统包括计算机 510 形式的通用计算设备。计算机 510 的部件可以包括但不限于处理单元 520、系统存储器 530 以及将包括系统存储器的各种系统部件耦合到处理单元 520 的系统总线 521。系统总线 521 可以是包括使用多种总线架构中的任意一种的存储器总线或者存储器控制器、外围总线和本地总线的几种类型的总线结构中的任意一种。作为示例而非限制，这些架构包括工业标准架构 (Industry Standard Architecture, ISA) 总线、微通道架构 (Micro Channel Architecture, MCA) 总线、增强型 ISA(Enhanced ISA, EISA) 总线、视频电子标准协会 (Video Electronics Standards Association, VESA) 本地总线和也作为夹层总线已知的外设组件互连 (Peripheral Component Interconnect, PCI) 总线。

[0141] 计算机 510 一般包括多种计算机可读介质。计算机可读介质可以是计算机 510 可以访问的任意可用介质，其包括易失性和非易失性介质两者、可移除和不可移除介质两者。作为示例而非限制，计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或者其它数据的信息的任意方法或者技术实现的易失性和非易失性、可移除和不可移除介质两者。计算机存储介质包括但不限于 RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字通用盘 (DVD) 或其它光盘存储器、磁带盒、磁带、磁盘存储器或其它磁存储设备、或者可以用来存储希望的信息并且计算机 510 可以访问的任意其它介质。通信介质一般包括诸如载波或者其它传输机制的调制数据信号中的计算机可读指令、数据结构、程序模块或者其它数据，其包括任意信息传递介质。术语“调制数据信号”指的是以对信号中的信息进行编码的方式对其特性中的一个或更多个进行了设置或者改变的信号。作为示例、而非限制，通信介质包括诸如有线网络或直接有线连接的有线介质以及诸如声音、RF、红外线的无线介质和其它无线介质。任意上述内容的组合也包含在计算机可读介质的范围内。

[0142] 系统存储器 530 包括诸如只读存储器 (ROM) 531 和随机存取存储器 (RAM) 532 的易失性和 / 或非易失性存储器形式的计算机存储介质。包含诸如在起动期间帮助在计算机 510 内的元件之间传递信息的基本例程的基本输入 / 输出系统 533(BIOS)，一般存储在 ROM531 中。RAM532 一般包含处理单元 520 立即可访问和 / 或当前正在操作的数据和 / 或程序模块。作为示例而非限制，图 5 示出了操作系统 534、应用程序 535、其它程序模块 536 和程序数据 537。

[0143] 计算机 510 还可以包括其它可移除 / 不可移除、易失性 / 非易失性计算机存储介质。仅仅作为示例，图 5 示出了从不可移除非易失性磁介质中进行读取或者向其进行写入

的硬盘驱动器 541、从可移除非易失性磁盘 552 中进行读取或者向磁盘 552 进行写入的磁盘驱动器 551，以及从诸如 CDROM 的可移除非易失性光盘 556 或其它光介质中进行读取或者向其进行写入的光盘驱动器 555。在本示例性操作环境中可以使用的其它可移除 / 不可移除、易失性 / 非易失性计算机存储介质包括但不限于磁带盒、闪存卡、数字通用盘、数字视频带、固态 RAM、固态 ROM 等。硬盘驱动器 541 一般通过诸如接口 540 的不可移除存储器接口连接到系统总线 521，磁盘驱动器 551 和光盘驱动器 555 一般由诸如接口 550 的可移除存储器接口连接到系统总线 521。

[0144] 上面讨论并且在图 5 中示出的驱动器及其相关联的计算机存储介质，为计算机 510 提供对计算机可读指令、数据结构、程序模块和其它数据的存储。例如，在图 5 中，示出了硬盘驱动器 541 存储操作系统 544、应用程序 545、其它程序模块 546 和程序数据 547。注意，这些组件可以与操作系统 534、应用程序 535、其它程序模块 536 和程序数据 537 相同或者不同。这里对操作系统 544、应用程序 545、其它程序模块 546 和程序数据 547 给予不同的编号，以示出至少它们是不同的副本。用户可以通过诸如键盘 562 和通常称为鼠标、跟踪球或触摸垫的指示设备 561 的输入设备向计算机 510 中输入命令和信息。其它输入设备（未示出）可以包括麦克风、操纵杆、游戏垫、卫星天线、扫描器等。这些和其它输入设备经常通过耦合到系统总线的用户输入接口 560 连接到处理单元 520，但是它们可以由诸如并行端口、游戏端口或者通用串行总线 (USB) 的其它接口和总线结构连接。监视器 591 或者其它类型的显示设备也经由诸如视频接口 590 的接口连接到系统总线 521。除了监视器之外，计算机还可以包括可以通过输出外围设备接口 595 连接的诸如扬声器 597 和打印机 596 的其它外围输出设备。

[0145] 计算机 510 可以使用到诸如远程计算机 580 的一个或更多个远程计算机的逻辑连接在联网环境中工作。远程计算机 580 可以是个人计算机、服务器、路由器、网络 PC、对等设备或者其它公共网络节点，虽然在图 5 中仅示出了存储器设备 581，但是远程计算机 580 一般包括上面相对于计算机 510 描述的元素中的许多或者全部。在图 5 中描绘的逻辑连接包括局域网 (LAN) 571 和广域网 (WAN) 573，但是还可以包括其它网络。这些联网环境在办公室、企业范围计算机网络、内联网和因特网中非常普遍。

[0146] 当在 LAN 联网环境中使用时，计算机 510 通过网络接口或者适配器 570 连接到 LAN571。当在 WAN 联网环境中使用时，计算机 510 一般包括调制解调器 572 或者用于在诸如因特网的 WAN573 上建立通信的其它装置。可以在内部或者外部的调制解调器 572 可以经由用户输入接口 560 或者其它适当的机制连接到系统总线 521。在联网环境中，可以将相对于计算机 510 描绘的程序模块或者其部分存储在远程存储器存储设备中。作为示例而非限制，图 5 示出了如驻留在存储器设备 581 上的远程应用程序 585。应当理解，示出的网络连接是示例性的，可以使用在计算机之间建立通信链接的其它装置。

[0147] 综上，在根据本公开的实施例中，本公开提供了如下方案，但不限于此：

[0148] 1. 一种对多个图像中的图像进行语义注释的方法，所述多个图像中的每个图像包括至少一个图像区域，所述方法包括：

[0149] 识别包括第一和第二图像的至少两个相似图像；

[0150] 识别所述第一图像和所述第二图像中的相应图像区域；以及

[0151] 使用至少一个处理器，通过使用拟合度对所述多个图像中的一个或更多个图像中

的图像区域分配注释,所述拟合度指示所分配的注释和所述相应图像区域之间的匹配程度,

[0152] 其中,所述拟合度取决于所述多个图像的子集中的每个图像的至少一个注释以及所述第一图像和所述第二图像中的图像区域之间的所识别的对应关系。

[0153] 2. 根据方案 1 所述的方法,其中 :

[0154] 识别相应图像区域包括用所述第二图像中的第二图像区域识别所述第一图像中的第一图像区域;以及

[0155] 对图像区域分配注释包括基于所述拟合度迭代地对图像区域分配注释,

[0156] 其中,每个迭代包括至少部分基于对所述第二图像区域分配的注释对所述第一图像区域分配注释。

[0157] 3. 根据方案 2 所述的方法,其中,所述第一图像包括与所述第一图像区域相邻的第三图像区域,以及其中,每个迭代包括:

[0158] 进一步基于对所述第三图像区域分配的注释来对所述第一图像区域分配注释。

[0159] 4. 根据方案 2 所述的方法,其中,每个迭代包括:

[0160] 进一步基于所述注释的空间直方图对所述第一图像区域分配注释,其中,跨所述多个图像中的数个图像计算所述注释的空间直方图。

[0161] 5. 根据方案 1 所述的方法,还包括:

[0162] 至少部分地通过从包含所述多个图像的所述子集中的图像的文档和 / 或网页中的在所述图像附近的文本中获得所述图像的至少一个图像级别注释,来自动获得所述至少一个注释。

[0163] 6. 根据方案 1 所述的方法,进一步包括如下获得所述至少一个注释:

[0164] 将所述多个图像的子集中的至少一个图像提供给用户;以及

[0165] 从用户获得所述至少一个图像的至少一个图像区域注释。

[0166] 7. 根据方案 1 所述的方法,其中识别所述至少两个相似图像包括:

[0167] 针对图像对,计算指示图像之间的相似程度的相似度度量;

[0168] 如果所计算的相似度度量大于预定阈值,将所述图像对识别为相似的。

[0169] 8. 根据方案 1 所述的方法,其中识别所述第一图像和所述第二图像中的相应图像区域包括:

[0170] 基于指示所述图像区域之间的相似程度的目标函数,计算所述第一图像中的图像区域和所述第二图像中的图像区域之间的对应关系。

[0171] 9. 根据方案 1 所述的方法,其中所述第一图像区域是像素。

[0172] 10. 一种使得能够至少部分地基于与多个图像中的图像相关联的图像级别注释对所述图像进行基于文本的搜索的系统,所述系统包括:

[0173] 至少一个处理器,被配置为:

[0174] 识别包括第一图像和第二图像的至少两个相似图像;

[0175] 识别所述第一图像和所述第二图像中的相应图像区域;以及

[0176] 通过使用拟合度对所述多个图像中的一个或更多个图像中的图像区域分配注释,所述拟合度指示所分配的注释和所述相应图像区域之间的匹配程度,

[0177] 其中,所述拟合度取决于所述多个图像的子集中的每个图像的至少一个注释以及

所述第一图像和所述第二图像中的图像区域之间的所识别的对应关系。

[0178] 11. 根据方案 10 所述的系统, 其中, 所述至少一个处理器被配置为 :

[0179] 通过用所述第二图像中的第二图像区域识别所述第一图像中的第一图像区域, 来识别相应图像区域; 以及

[0180] 通过基于所述拟合度迭代地对图像区域分配注释, 来对所述图像区域分配注释,

[0181] 其中, 每个迭代包括至少部分基于对所述第二图像区域分配的注释来对所述第一图像区域分配注释。

[0182] 12. 根据方案 11 所述的系统, 其中, 每个迭代包括计算注释和所述多个图像中的所述一个或更多个图像中的所述图像区域之间的统计映射。

[0183] 13. 根据方案 12 所述的系统, 其中, 计算所述统计映射包括估计混合模型的至少一个参数, 其中, 所述混合模型包括针对注释集合中的每个注释的元素。

[0184] 14. 根据方案 10 所述的系统, 其中, 所述多个图像的所述子集包括所述多个图像的百分之十或者更少。

[0185] 15. 至少一种有形计算机可读存储介质, 存储有处理器可执行的指令, 当该指令被至少一个处理器执行时, 执行用于对多个图像中的图像进行语义注释的方法, 所述多个图像中的每个图像包括一个或更多个像素, 所述方法包括 :

[0186] 获得多个图像的子集中每个图像的至少一个图像级别注释;

[0187] 识别包括第一和第二图像的至少两个相似图像;

[0188] 识别所述第一图像和所述第二图像中的相应像素; 以及

[0189] 使用拟合度对所述多个图像中的一个或更多个图像中的像素分配注释, 所述拟合度指示所分配的注释和所述相应像素之间的匹配程度,

[0190] 其中, 所述拟合度取决于至少一个获得的图像级别注释以及所述第一图像和所述第二图像中的像素之间的所识别的对应关系。

[0191] 16. 根据方案 15 所述的有形计算机可读存储介质, 其中 :

[0192] 识别相应像素包括以所述第二图像中的第二像素识别所述第一图像中的第一像素; 以及

[0193] 对像素分配注释包括基于所述拟合度迭代地对像素分配注释,

[0194] 其中, 每个迭代包括至少部分基于对所述第二像素分配的注释来对所述第一像素分配注释。

[0195] 17. 根据方案 16 所述的有形计算机可读存储介质, 其中所述第一图像包括在所述第一像素附近的第三像素, 以及其中, 每个迭代包括 :

[0196] 进一步基于对所述第三像素分配的注释对所述第一像素分配注释。

[0197] 18. 根据方案 16 所述的有形计算机可读存储介质, 其中, 每个迭代包括计算注释和所述多个图像中的所述一个或更多个图像中的所述图像区域之间的统计映射。

[0198] 19. 根据方案 15 所述的有形计算机可读存储介质, 进一步包括 :

[0199] 从网页上在所述图像附近的文本自动地获得针对所述网页中的图像的至少一个图像级别注释。

[0200] 20. 根据方案 15 所述的有形计算机可读存储介质, 进一步包括通过以下方式获得所述至少一个图像级别注释 :

[0201] 将所述多个图像的子集中的至少一个图像提供给用户；以及

[0202] 从用户获得针对所述至少一个图像的至少一个图像区域注释。

[0203] 虽然如此描述了本发明的至少一个实施例的几个方面，但是应当理解，本领域技术人员容易想到各种变化、变形和改进。

[0204] 这些变化、变形和改进旨在作为本公开的一部分，并且旨在落入本发明的精神和范围内。此外，虽然指出了本发明的优点，但是应当理解，不是每个本发明的实施例包括每个描述的优点。一些实施例可能不实现这里作为优点而描述的一些特征。相应地，前述描述和附图仅仅用作示例。

[0205] 可以以大量方式中的任意一种实现上述本发明的实施例。例如，可以使用硬件、软件或其组合来实现实施例。当以软件实现时，可以在不管设置在单个计算机中还是分布在多个计算机上的任意合适的处理器或者处理器集上执行软件代码。这些处理器可以作为集成电路来实现，其中，一个或更多个处理器在集成电路部件中。虽然如此，处理器可以使用任意合适形式的电路来实现。

[0206] 此外，应当理解，可以以诸如机架安装型计算机、台式计算机、膝上型计算机或者平板计算机的大量形式中的任意一种来实施计算机。另外，可以在通常不被视为计算机、但是具有合适的处理能力的包括个人数字助理 (PDA)、智能电话或者任意其它合适的便携式或固定电子设备的设备中嵌入计算机。

[0207] 此外，计算机可以具有一个或更多个输入和输出设备。除其它之外，可以使用这些设备来呈现用户界面。可以用来提供用户界面的输出设备的示例包括用于输出的可视展示的打印机或显示屏以及用于输出的可听展示的扬声器或其它声音生成设备。可以用于用户界面的输入设备的示例包括键盘以及诸如鼠标、触摸垫和数字化平板的指示设备。作为另一示例，计算机可以通过语音识别或者以其它可听格式接收输入信息。

[0208] 可以由包括如局域网或者诸如企业网或因特网的广域网的任意合适形式的一个或更多个网络，来互连这些计算机。这些网络可以基于任意合适的技术，并且可以根据任意合适的协议工作，并且可以包括无线网络、有线网络或者光纤网络。

[0209] 此外，可以将这里概述的各种方法或处理（例如处理 300）编码为在利用各种操作系统或平台中的任意一种的一个或更多个处理器上可执行的软件。另外，该软件可以使用大量合适的编程语言和 / 或编程或脚本工具中的任意一种编写，还可以被编译为在框架或虚拟机上执行的可执行机器语言代码或中间代码。

[0210] 在这方面，本发明可以作为用一个或更多个程序进行了编码的计算机可读存储介质（或者多个计算机可读介质）（例如计算机存储器、一个或更多个软盘、紧凑盘 (CD)、光盘、数字视频盘 (DVD)、磁带、闪存、现场可编程门阵列或其它半导体设备中的电路配置或者其它有形计算机存储介质）来实施，当在一个或更多个计算机或其它处理器上执行该一个或更多个程序时，该一个或更多个程序执行实现上面讨论的本发明的各种实施例的方法。从前面的示例显而易见，计算机可读存储介质可以将信息保持足够的时间，从而以非临时性的形式提供计算机可执行指令。该计算机可读存储介质可以是可传输的，从而可以将存储在其上的程序加载到一个或更多个不同的计算机或其它处理器上，以实现如上面讨论的本发明的各个方面。如这里所使用的，术语“计算机可读存储介质”仅包含可以被视为产品（即制成品）或机器的计算机可读介质。可替选地或者另外，本发明可以作为计算机可读存

储介质之外的诸如传播信号的计算机可读介质来实施。

[0211] 这里按照通常的意义来使用术语“程序”或“软件”，以指可以用来对计算机或其它处理器进行编程以实现如上面讨论的本发明的各个方面的任意类型的计算机代码或计算机可执行指令的集合。另外，应当理解，根据本实施例的一个方面，在执行时进行本发明的方法的一个或更多个计算机程序不必驻留在单个计算机或处理器上，而可以以模块的形式分布在多个不同的计算机或处理器中，来实现本发明的各个方面。

[0212] 计算机可执行指令可以是由一个或更多个计算机或者其它设备执行的诸如程序模块的许多形式。一般而言，程序模块包括进行特定任务或者实现特定抽象数据类型的例程、程序、对象、组件、数据结构等。通常来说，在各个实施例中，可以按照希望组合或者分布程序模块的功能。

[0213] 此外，可以将数据结构以任意合适的形式存储在计算机可读介质中。为了使说明简单，可以将数据结构示为具有通过数据结构中的位置相关的字段。该关系同样地可以通过对字段的存储分配传递字段之间的关系的计算机可读介质的位置来获得。然而，可以使用任意合适的机制来在数据结构的字段中的信息之间建立关系，包括通过使用在数据元素之间建立关系的指针、标签或其它机制。

[0214] 可以单独、组合或者以在前面描述的实施例中未具体讨论的多种布置使用本发明的各个方面，因此，本发明各方面的应用不限于在前面的描述中叙述或者在附图中示出的细节和部件的布置。例如，可以以任意方式将在一个实施例中描述的方面与在其它实施例中描述的方面组合。

[0215] 此外，本发明可以作为方法来实施，已提供了该方法的示例。可以以任意合适的方式对作为方法的部分执行的动作进行排序。相应地，尽管在说明性实施例中作为连续动作示出，也可以构造按照与所示出的不同的顺序执行动作的实施例，可以包括同时执行一些动作。

[0216] 在权利要求中用于修饰权利要求元素的诸如“第一”、“第二”、“第三”等的序数词的使用，其本身不意味着一个权利要求元素的任何优先级、次序或者顺序先于另一个权利要求元素，或者执行方法的动作的时间顺序，而仅仅用作区分具有特定名称的一个权利要求元素与具有相同名称（除了序数词的使用之外）的另一元素的标签，以区分权利要求元素。

[0217] 此外，这里使用的措辞和术语用于描述的目的，而不应当被视为限制。这里“包括”、“包含”或者“具有”、“含有”、“涉及”及其变体的使用意为涵盖之后列出的项及其等同物以及附加项。

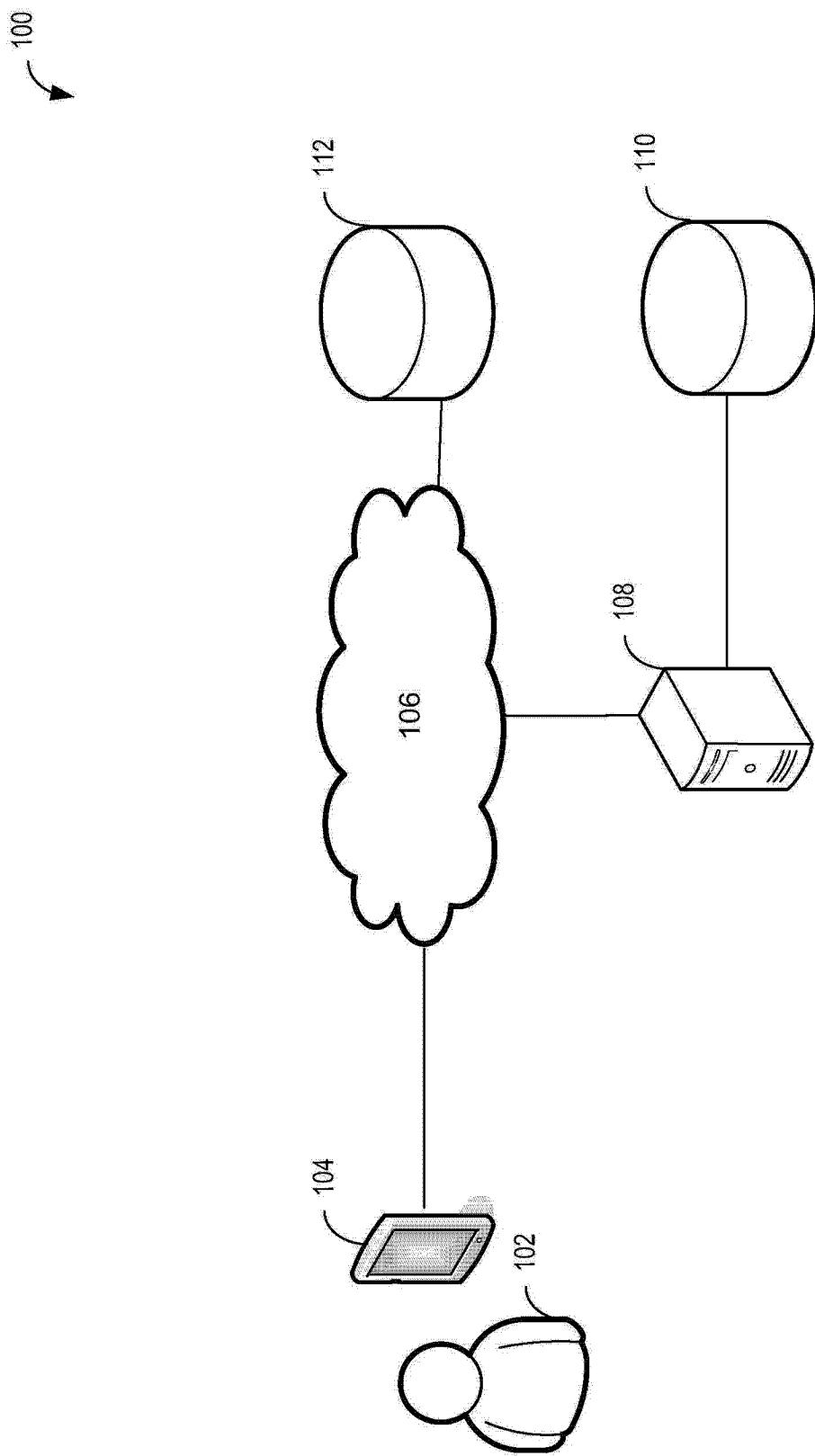


图 1

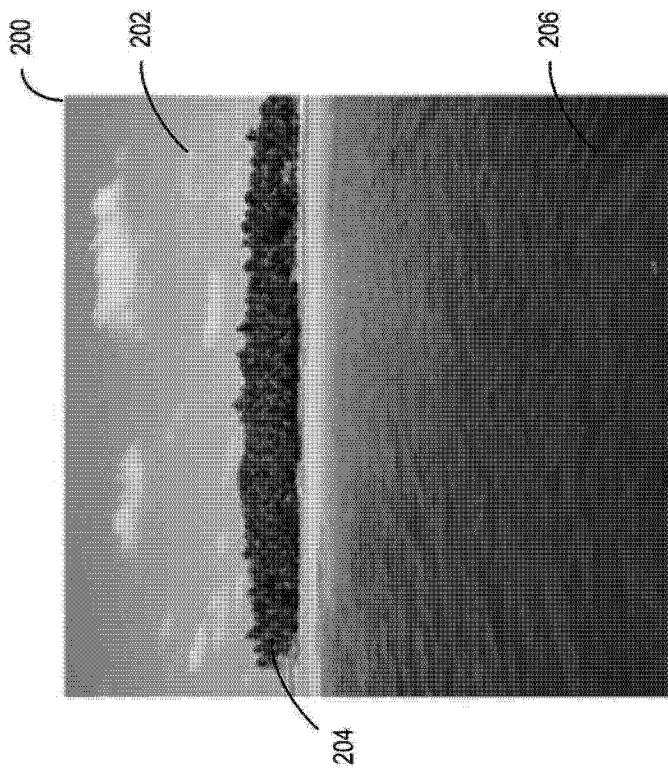


图 2A

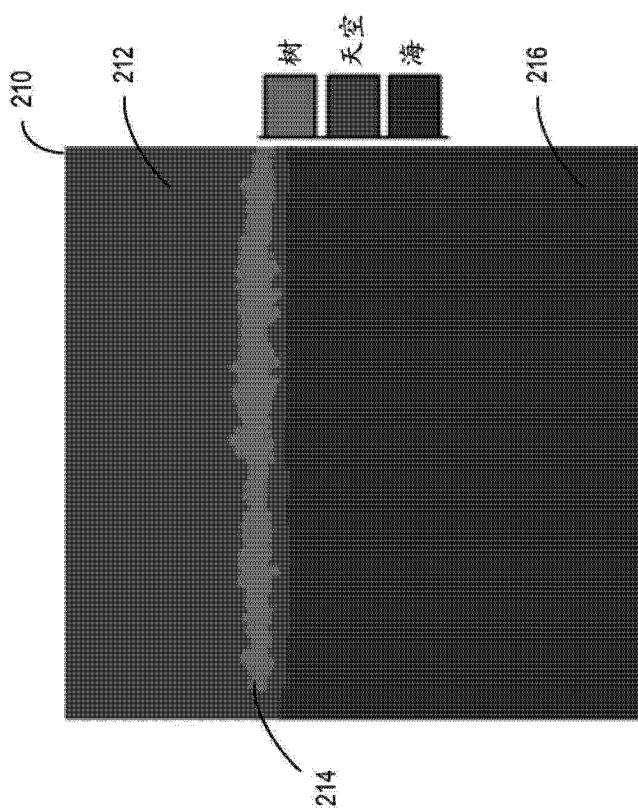


图 2B

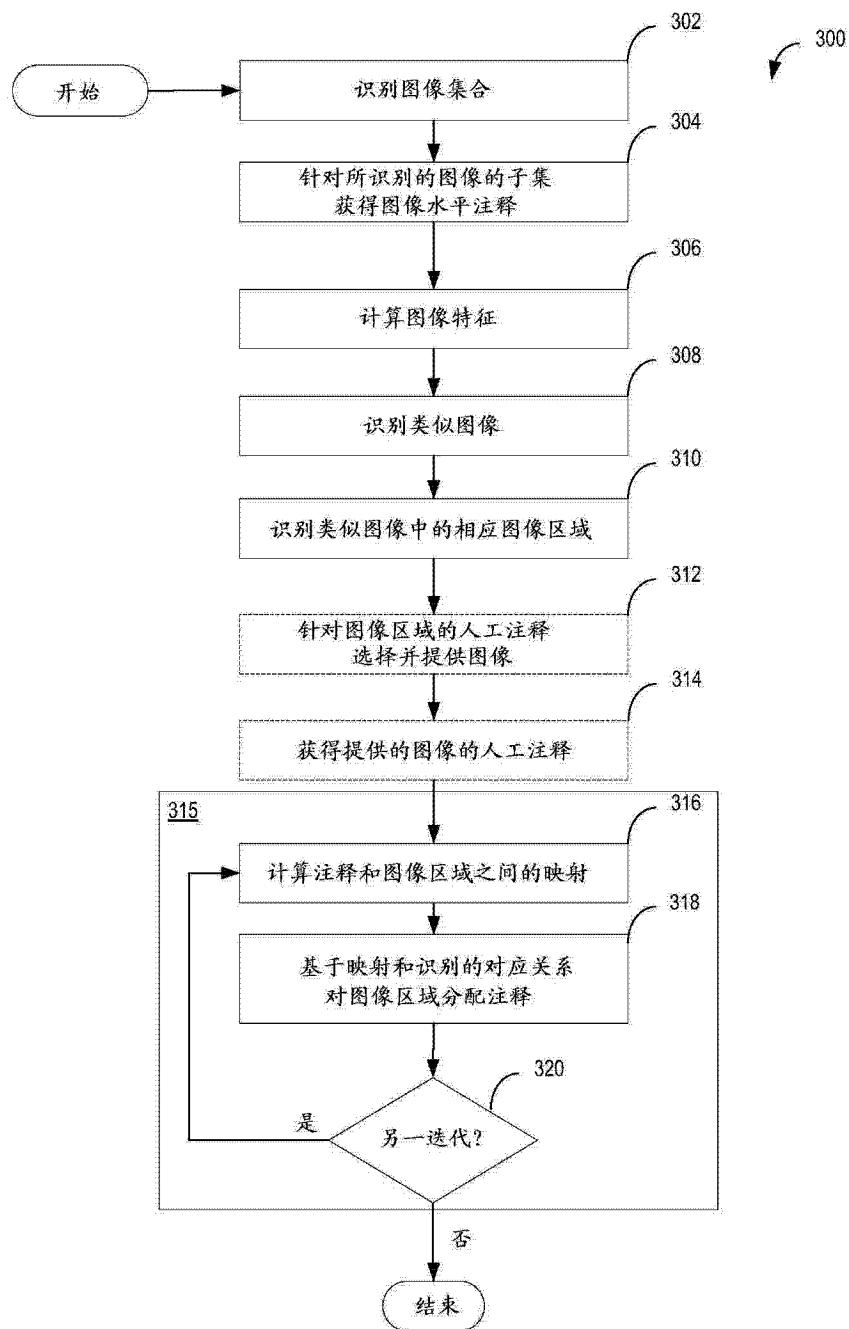


图 3

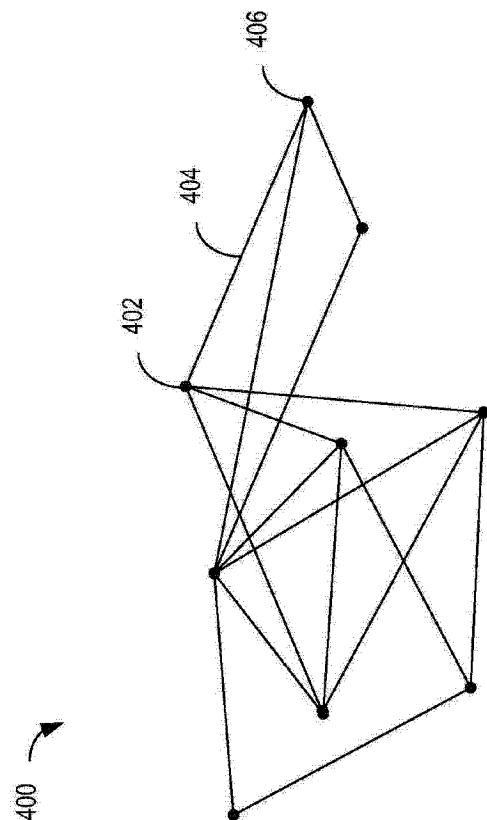


图 4A

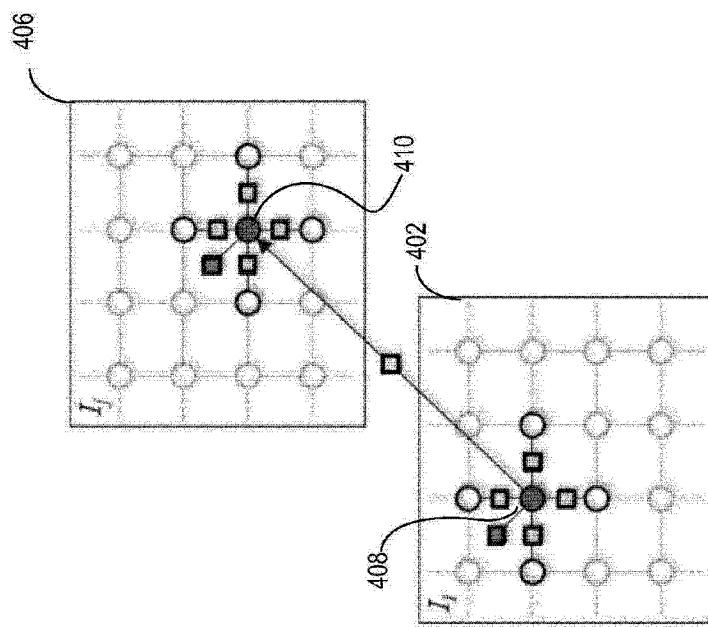


图 4B

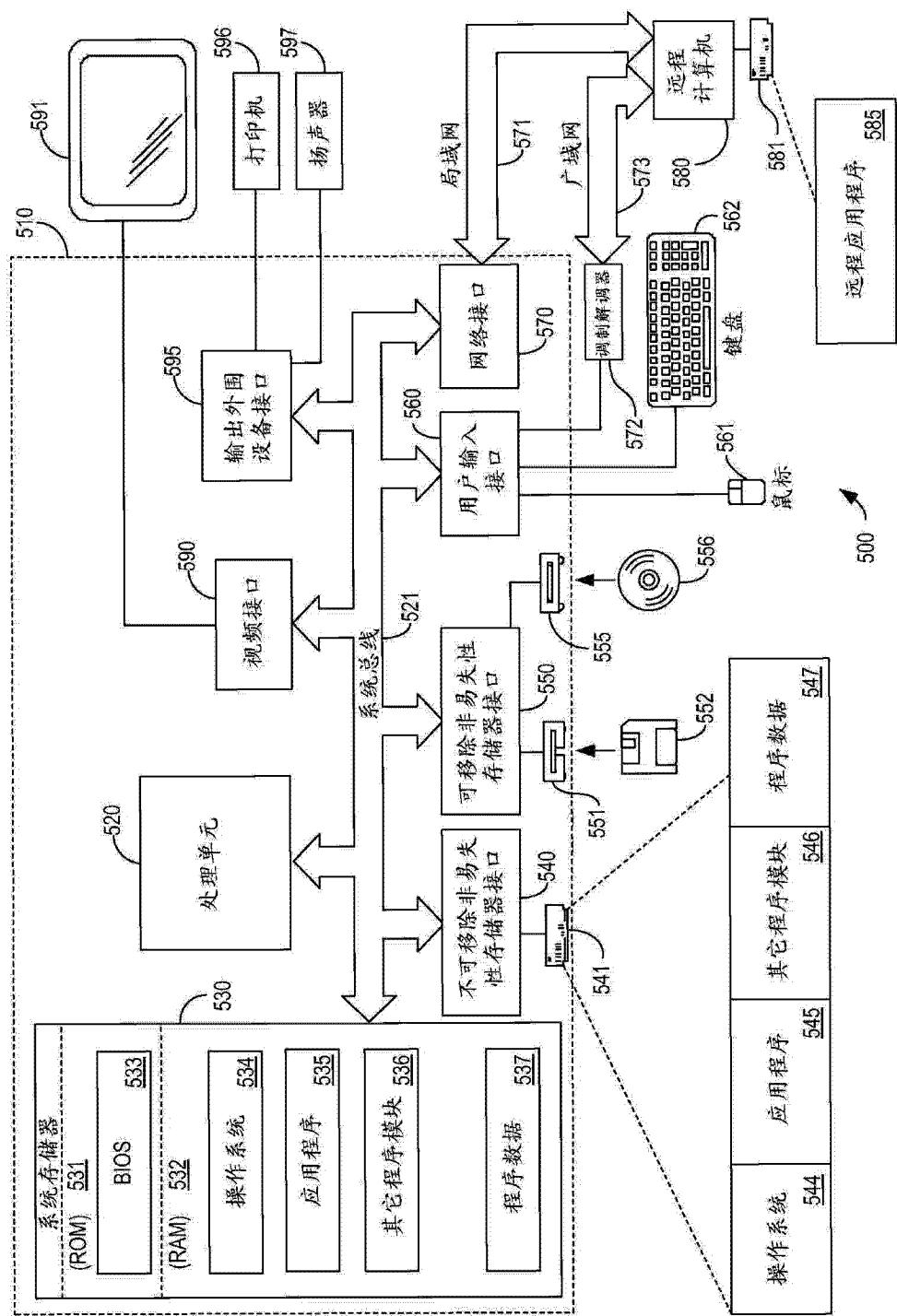


图 5