



(12) 发明专利

(10) 授权公告号 CN 102105901 B

(45) 授权公告日 2013. 03. 06

(21) 申请号 200980129633. X

(22) 申请日 2009. 04. 17

(30) 优先权数据

61/059, 702 2008. 06. 06 US

(85) PCT申请进入国家阶段日

2011. 01. 27

(86) PCT申请的申请数据

PCT/US2009/040975 2009. 04. 17

(87) PCT申请的公布数据

W02009/154861 EN 2009. 12. 23

(73) 专利权人 谷歌公司

地址 美国加利福尼亚州

(72) 发明人 阿米什·马卡迪亚 桑吉夫·库马尔

(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)

(56) 对比文件

W0 2007136861 A2, 2007. 11. 29, 说明书第 0008-0020 段, 第 0025-0034 段, 图 1-3.

W0 2007136861 A2, 2007. 11. 29, 说明书第 0008-0020 段, 第 0025-0034 段, 图 1-3.

JP 2007207031 A, 2007. 08. 16, 全文.

US 20070296993 A1, 2007. 12. 27, 全文.

审查员 王涛

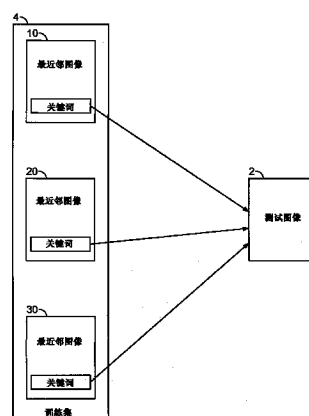
权利要求书 3 页 说明书 11 页 附图 12 页

(54) 发明名称

注释图像

(57) 摘要

用于生成用于自动注释图像的数据的方法、系统和装置, 包括计算机程序产品。在一个方面中, 方法包括接收输入图像, 从图像集合中识别输入图像的一个或多个最近邻图像, 其中一个或多个最近邻图像中的每一个与相应的一个或多个图像标签相关联, 将多个图像标签分配给输入图像, 其中多个图像标签是从与一个或多个最近邻图像相关联的图像标签选择的, 以及将具有分配的多个图像标签的输入图像存储在数据储存库中。在另一个方面中, 方法包括将单个图像标签分配给输入图像, 其中单个图像标签是从与多个排名的最近邻图像相关联的标签选择的。



1. 一种用于注释图像的方法,包括:

在数据处理装置中接收输入图像;

通过所述数据处理装置的操作从存储在计算机可读介质上的数字图像集合中识别所述输入图像的多个最近邻图像,其中所述多个最近邻图像中的每一个与相应的一个或多个图像标签相关联,并且其中识别所述多个最近邻图像包括:

根据与每一个最近邻图像相关联的相应整体图像距离来对所述多个最近邻图像进行排序,每一个整体图像距离表示在所述输入图像和对应的最近邻图像之间的差异程度并且包括特征距离的组合,每一个特征距离表示在与所述输入图像相关联的图像特征和与所述对应的最近邻图像相关联的相应图像特征之间的差异程度;以及

根据在所述数字图像集合中出现的相应频率对一个或多个第一图像标签进行排名,所述一个或多个第一图像标签与具有最小的整体图像距离的最近邻图像相关联;

将多个图像标签分配给所述输入图像,其中所述多个图像标签由所述数据处理装置从与所述多个最近邻图像相关联的所述图像标签选择,并且其中分配所述多个图像标签包括基于对所述一个或多个第一图像标签的所述排名将所述第一图像标签中的至少一个分配给所述输入图像;以及

将具有所分配的所述多个图像标签的所述输入图像存储在数字数据储存库中。

2. 根据权利要求1所述的方法,其中所述输入图像以包括作为元数据的所述多个图像标签的图像文件被存储在所述数字数据储存库中。

3. 根据权利要求1所述的方法,其中所述多个图像标签是从单个最近邻图像选择的。

4. 根据权利要求1所述的方法,

其中识别多个最近邻图像进一步包括:对一个或多个第二图像标签进行排名,其中所述第二图像标签中的每一个与一个或多个剩余的最近邻图像相关联;以及

其中将所述多个图像标签分配给所述输入图像进一步包括基于对所述一个或多个第二图像标签的所述排名将所述第二图像标签中的至少一个分配给所述输入图像。

5. 根据权利要求4所述的方法,其中对所述一个或多个第二图像标签的排名包括根据所述第二图像标签中的每一个与每一个第一图像标签在所述数字图像集合中的共同出现对所述一个或多个第二图像标签进行排序。

6. 根据权利要求4所述的方法,其中对所述一个或多个第二图像标签的排名包括根据所述第二图像标签中的每一个在所述一个或多个剩余的最近邻图像中的局部频率对所述一个或多个第二图像标签进行排序。

7. 根据权利要求1所述的方法,其中与所述输入图像相关联的所述图像特征和与所述对应的最近邻图像相关联的所述相应图像特征每一个包括全局图像特征。

8. 根据权利要求1所述的方法,其中与所述输入图像相关联的所述图像特征和与所述对应的最近邻图像相关联的所述相应图像特征每一个包括局部图像特征。

9. 根据权利要求1所述的方法,进一步包括:

通过计算在所述输入图像和所述对应的最近邻图像之间的所述特征距离的平均值来取得每个整体图像距离。

10. 根据权利要求9所述的方法,其中所述平均值基于来自所述特征距离中的每一个的均等贡献。

11. 根据权利要求 9 所述的方法,其中所述平均值基于来自所述特征距离中的每一个的加权贡献。

12. 根据权利要求 11 所述的方法,进一步包括:

基于所述数字图像集合来计算用于所述特征距离中的每一个的权重,其中所述数字图像集合是包括相似和相异图像对的训练图像组。

13. 根据权利要求 1 所述的方法,进一步包括:

将所述特征距离中的至少一个计算为在所述输入图像的纹理特征和所述对应的最近邻图像的对应的纹理特征之间的差异。

14. 根据权利要求 1 所述的方法,进一步包括:

将所述特征距离中的至少一个计算为在所述输入图像的色彩特征和所述对应的最近邻图像的对应的色彩特征之间的差异。

15. 根据权利要求 1 所述的方法,进一步包括:

缩放所述特征距离;以及

基于所缩放的与所述对应的最近邻图像相关联的特征距离,取得每个整体图像距离。

16. 根据权利要求 12 所述的方法,其中,计算所述权重包括执行逻辑回归。

17. 根据权利要求 16 所述的方法,其中,所述逻辑回归包括 Lasso 回归。

18. 一种用于注释图像的系统,包括:

用于在服务器中接收输入图像的装置;

用于通过所述服务器的操作从存储在计算机可读介质上的数字图像集合中识别所述输入图像的多个最近邻图像的装置,其中所述多个最近邻图像中的每一个与相应的一个或多个图像标签相关联,并且其中用于识别多个最近邻图像的所述装置包括:

用于根据与每一个最近邻图像相关联的相应整体图像距离来对所述多个最近邻图像进行排序的装置,每一个整体图像距离表示在所述输入图像和对应的最近邻图像之间的差异程度并且包括特征距离的组合,每一个特征距离表示在与所述输入图像相关联的图像特征和与所述对应的最近邻图像相关联的相应图像特征之间的差异程度;以及

用于根据在所述数字图像集合中出现的相应频率对一个或多个第一图像标签进行排名的装置,所述一个或多个第一图像标签与具有最小的整体图像距离的最近邻图像相关联;

用于将多个图像标签分配给所述输入图像的装置,其中所述多个图像标签是由所述服务器从与所述多个最近邻图像相关联的所述图像标签选择的,并且其中用于分配所述多个图像标签的装置包括用于基于对所述一个或多个第一图像标签的所述排名将所述第一图像标签中的至少一个分配给所述输入图像的装置;以及

用于将具有所分配的多个图像标签的所述输入图像存储在数字数据储存库中的装置。

19. 根据权利要求 18 所述的系统,其中所述多个图像标签是从单个最近邻图像选择的。

20. 根据权利要求 18 所述的系统,

其中用于识别所述多个最近邻图像的装置进一步包括用于对一个或多个第二图像标签进行排名的装置,其中所述第二图像标签中的每一个与一个或多个剩余的最近邻图像相关联;以及

用于分配所述多个图像标签的装置包括用于基于对所述一个或多个第二图像标签的所述排名将所述第二图像标签中的至少一个分配给所述输入图像的装置。

21. 根据权利要求 20 所述的系统,其中用于对所述一个或多个第二图像标签的排名的装置包括:用于根据所述第二图像标签中的每一个与每一个第一图像标签在所述数字图像集合中的共同出现对所述一个或多个第二图像标签进行排序的装置。

22. 根据权利要求 20 所述的系统,其中用于对所述一个或多个第二图像标签的排名的装置包括:用于根据所述第二图像标签中的每一个在所述一个或多个剩余的最近邻图像中的局部频率对所述一个或多个第二图像标签进行排序的装置。

23. 根据权利要求 18 所述的系统,其中与所述输入图像相关联的所述图像特征和与所述对应的最近邻图像相关联的所述相应图像特征每一个包括全局图像特征。

24. 根据权利要求 18 所述的系统,其中与所述输入图像相关联的所述图像特征和与所述对应的最近邻图像相关联的所述相应图像特征每一个包括局部图像特征。

25. 根据权利要求 18 所述的系统,进一步包括:

用于通过计算在所述输入图像和所述对应的最近邻图像之间的所述特征距离的平均值来取得每个整体图像距离的装置。

26. 根据权利要求 25 所述的系统,其中所述平均值基于来自所述特征距离中的每一个的均等贡献。

27. 根据权利要求 25 所述的系统,其中所述平均值基于来自所述特征距离中的每一个的加权贡献。

28. 根据权利要求 27 所述的系统,进一步包括:

用于基于所述数字图像集合来计算用于所述特征距离中的每一个的权重的装置,其中所述数字图像集合是包括相似和相异图像对的训练图像组。

29. 根据权利要求 18 所述的系统,进一步包括:

用于将所述特征距离中的至少一个计算为在所述输入图像的纹理特征和所述对应的最近邻图像的对应的纹理特征之间的差异的装置。

30. 根据权利要求 18 所述的系统,进一步包括:

用于将所述特征距离中的至少一个计算为在所述输入图像的色彩特征和所述对应的最近邻图像的对应的色彩特征之间的差异的装置。

31. 根据权利要求 18 所述的系统,进一步包括:

用于缩放所述特征距离的装置;以及

用于基于所缩放的与所述对应的最近邻图像相关联的特征距离,取得每个整体图像距离的装置。

32. 根据权利要求 28 所述的系统,其中,用于计算每个所述特征距离的所述权重的装置包括用于执行逻辑回归的装置。

33. 根据权利要求 32 所述的系统,其中,所述逻辑回归包括 Lasso 回归。

注释图像

[0001] 相关申请的交叉引用

[0002] 本申请根据 35 U.S.C. § 119(e) 要求享有于 2008 年 6 月 6 日提交、题为“注释图像”的美国专利申请 No. 61/059,702 的权益,在此通过引用将所述美国专利申请整体合并。

技术领域

[0003] 本说明书涉及图像注释 (annotation)。

背景技术

[0004] 基于文本的图像注释仍然是计算机视觉和信息检索界中的重要的实践以及根本问题。从实践的角度来看,目前的图像搜索解决方案未能将图像内容有效地用于图像搜索。这通常导致有限适用性的搜索结果。

[0005] 给定输入图像,自动图像注释的目标是将反映图像的视觉内容的几个相关文本关键词(也被称为标签(label))分配给图像。可以通过以多种方式中的任何方式将关键词作为元数据存储例如包括图像的数字文件、带有从关键词到图像的链接或引用的数据库、带有联结关键词和图像的数据的 XML 文件或其它中,来将关键词分配给图像(或将关键词与图像相关联)。

[0006] 随着 Web 上和 Web 外迅速增长的图像数据集合,鲁棒的图像搜索和检索正快速成为关键性需求。目前的因特网图像搜索引擎通常采用基于文本的搜索来检索相关图像,而忽略图像内容。利用图像内容来分配更丰富、更相关的关键词集可以允许进一步将这些搜索引擎的快速索引和检索体系结构用于改进的图像搜索。这产生了用有巨大实践利益的相关文本关键词注释图像的问题。

发明内容

[0007] 本说明书描述了与自动注释图像有关的技术。

[0008] 本发明的一个方面涉及一种用于注释图像的方法,包括:在数据处理装置中接收输入图像;通过所述数据处理装置的操作从存储在计算机可读介质上的数字图像集合中识别所述输入图像的多个最近邻图像,其中所述多个最近邻图像中的每一个与相应的一个或多个图像标签相关联,并且其中识别所述多个最近邻图像包括:根据与每一个最近邻图像相关联的相应整体图像距离来对所述多个最近邻图像进行排序,每一个整体图像距离表示在所述输入图像和对应的最近邻图像之间的差异程度并且包括特征距离的组合,每一个特征距离表示在与所述输入图像相关联的图像特征和与所述对应的最近邻图像相关联的相应图像特征之间的差异程度;以及根据在所述数字图像集合中出现的相应频率对一个或多个第一图像标签进行排名,所述一个或多个第一图像标签与具有最小的整体图像距离的最近邻图像相关联;将多个图像标签分配给所述输入图像,其中所述多个图像标签由所述数据处理装置从与所述多个最近邻图像相关联的所述图像标签选择,并且其中分配所述多个图像标签包括基于对所述一个或多个第一图像标签的所述排名将所述第一图像标签中的

至少一个分配给所述输入图像；以及将具有所分配的所述多个图像标签的所述输入图像存储在数字数据储存库中。

[0009] 本发明的另一个方面涉及一种用于注释图像的系统，包括：用于在服务器中接收输入图像的装置；用于通过所述服务器的操作从存储在计算机可读介质上的数字图像集合中识别所述输入图像的多个最近邻图像的装置，其中所述多个最近邻图像中的每一个与相应的一个或多个图像标签相关联，并且其中用于识别多个最近邻图像的所述装置包括：用于根据与每一个最近邻图像相关联的相应整体图像距离来对所述多个最近邻图像进行排序的装置，每一个整体图像距离 (1) 表示在所述输入图像和对应的最近邻图像之间的差异程度，以及 (2) 并且包括特征距离的组合，每一个特征距离表示在与所述输入图像相关联的图像特征和与所述对应的最近邻图像相关联的相应图像特征之间的差异程度；以及用于根据在所述数字图像集合中出现的相应频率对一个或多个第一图像标签进行排名的装置，所述一个或多个第一图像标签与具有最小的整体图像距离的最近邻图像相关联；用于将多个图像标签分配给所述输入图像的装置，其中所述多个图像标签是由所述服务器从与所述一个或多个最近邻图像相关联的所述图像标签选择的，并且其中用于分配所述多个图像标签的装置包括用于基于对所述一个或多个第一图像标签的所述排名将所述第一图像标签中的至少一个分配给所述输入图像的装置；以及用于将具有所分配的多个图像标签的所述输入图像存储在数字数据储存库中的装置。

[0010] 总的来说，可以将在本说明书中描述的主题的一个方面具体化在一种由数据处理装置执行的图像注释的方法中，所述方法包括在数据处理装置中接收输入图像，通过数据处理装置的操作从存储在计算机可读介质上的数字图像集合中识别输入图像的一个或多个最近邻图像，其中一个或多个最近邻图像中的每一个与相应的一个或多个图像标签相关联，将多个图像标签分配给输入图像，其中多个图像标签由数据处理装置从与一个或多个最近邻图像相关联的图像标签选择，以及将输入图像和输入图像与所分配的多个图像标签的关联存储在数据储存库中。

[0011] 在另一个方面中，一种计算机实现的方法包括接收输入图像，从图像集合中识别输入图像的多个排名的最近邻图像，其中排名的最近邻图像中的每一个与相应的一个或多个图像标签相关联，将单个图像标签分配给输入图像，其中单个图像标签从与多个排名的最近邻图像相关联的图像标签选择，以及将具有分配的单个图像标签的输入图像存储在数据储存库中。前述方面的其它实施例包括相对应的系统、装置和计算机程序产品。

[0012] 在另一个方面中，系统包括用于确定一个或多个参考图像是输入图像的最近邻图像的装置，一个或多个参考图像共同地与多个图像标签相关联，以及用于将两个或多个图像标签分配给输入图像的装置，两个或多个图像标签从与一个或多个参考图像相关联的多个图像标签中选择。可以将一个或多个图像中的每一个与多个图像标签相关联。这些和其它实施例可以可选地包括下述特征中的一个或多个。多个图像可以从单个最近邻图像选择。

[0013] 分配多个图像标签可以包括根据与每一个最近邻图像相关联的相应整体图像距离 (whole-image distance) 来对多个最近邻图像进行排序。每一个整体图像距离可以表示在输入图像和对应的最近邻图像之间的差异程度。

[0014] 分配多个图像标签可以进一步包括根据在数字图像集合中相应的出现频率对一

个或多个第一图像标签进行排名,其中一个或多个第一图像标签中的每一个与第一最近邻相关联;以及基于对一个或多个第一图像标签的排名将第一图像标签中的至少一个分配给输入图像。

[0015] 分配多个图像标签还可以包括对一个或多个第二图像标签进行排名,其中第二图像标签中的每一个与一个或多个剩余的最近邻图像相关联,以及基于对一个或多个第二图像标签的排名将第二图像标签中的至少一个分配给输入图像。

[0016] 对一个或多个第二图像标签的排名可以包括根据第二图像标签中的每一个与每一个第一图像标签在数字图像集合中的共同出现对一个或多个第二图像标签进行排序。替选地或另外地,对一个或多个第二图像标签的排名可以包括根据第二图像标签中的每一个在一个或多个剩余的最近邻图像中的局部频率对一个或多个第二图像标签进行排序。

[0017] 整体图像距离可以包括特征距离的组合,每一个特征距离表示在与输入图像相关联的图像特征和与参考图像相关联的相应图像特征之间的差异程度。与输入图像相关联的图像特征和与参考图像相关联的相应图像特征每一个可以包括全局图像特征。替选地或另外地,与输入图像相关联的图像特征和与参考图像相关联的相应图像特征每一个可以包括局部图像特征。

[0018] 整体图像距离可以作为特征距离的平均值而取得。平均值可以基于来自特征距离中的每一个的基本均等的贡献。平均值可以基于来自特征距离中的每一个的加权贡献。可以基于数字图像集合来计算用于特征距离中的每一个的权重,其中数字图像集合是包括相似和相异图像对的训练图像组。

[0019] 特征距离中的至少一个可以被计算为在输入图像的纹理特征和参考图像的对应的纹理特征之间的差异。特征距离中的至少一个可以被计算为在输入图像的色彩特征和参考图像的对应的色彩特征之间的差异。

[0020] 可以实现在本说明书中描述的主题的特定实施例来实现下述优势中的一个或多个。在一些情况下,图像注释技术由最小训练需求表征。在某些实施方式中,将图像注释当作检索问题对待简化了注释过程。

[0021] 在附图和下面的描述中阐述了本发明的一个或多个实施例的细节。本发明的其它特征、方面和优势从描述、附图以及权利要求将变得显而易见。

附图说明

[0022] 图 1 图示了用关键词注释测试图像的示例。

[0023] 图 2A 是提供确定输入图像的最近邻的一般概述的流程图。

[0024] 图 2B 是提供取得复合距离的概述的流程图。

[0025] 图 3 是提供将来自输入图像的最近邻的关键词转送到输入图像的一般概述的流程图。

[0026] 图 4 示出了共同具有至少 4 个关键词的图像对的示例。

[0027] 图 5 示出了共同具有零个关键词的图像对的示例。

[0028] 图 6 和 7 示出了来自不同图像数据集的示例图像。

[0029] 图 8 示出了已注释的图像的示例。

[0030] 图 9 示出了已注释的图像的示例。

[0031] 图 10、11 和 12 示出了分别在三个不同的图像数据集中为多个不同的关键词检索的最先的几个图像的示例。

具体实施方式

[0032] 自动将关键词分配给图像允许检索、索引、组织以及理解大型的图像数据集合。本说明书描述了用于将注释作为检索问题对待的图像注释的技术。所述技术利用低层级图像特征以及基本距离测量的简单组合来找到给定图像的最近邻。然后使用贪婪 (greedy) 标签转送机制来分配关键词。

[0033] 图像注释是艰巨的任务有两个主要的原因：第一，存在像素到谓项 (pixel-to-predicate) 或语义鸿沟 (gap) 问题，其中仅使用例如色彩和纹理的低层级图像特征来提取语义上有意义的实体是困难的。可靠地明确识别成千上万的对象或种类当前是未解决的问题。第二个困难是由于在训练数据中的图像区域和关键词之间的对应性的缺乏而产生。对于每一个图像，可以访问分配给整个图像的关键词，但是图像的哪些区域对应于这些关键词是不知道的。这会妨碍分类器的直接学习，其中每一个关键词被认为是单独的种类。

[0034] 本说明书描述了由最小训练需求表征的技术。针对若干标准数据集以及大型 Web 数据集，所述技术超越了复杂的现有技术的图像注释方法。

[0035] 图 1 图示了用关键词注释测试图像的示例。给定测试图像 2，可以从图像训练集 4 找到它的最近邻（例如，在一些特征空间中用预先指定的距离测量定义的第一最近邻 10、第二最近邻 20 以及第三最近邻 30），并且将与最近邻图像相关联的部分或全部关键词分配给输入测试图像 2。在一些情况下，使用针对全局图像特征定义的简单距离测量执行得比其它注释技术更好。在一些实施方式中，K-最近邻而不只是最近一个用于分配关键词。在多个邻居的情况下，可以使用贪婪方法来将适当的关键词分配给输入图像，进一步增强注释性能。

[0036] K-最近邻方法可以被扩展以合并多个距离测量，其可以通过不同的特征空间定义。组合不同的距离或核心可以在对象识别任务中产生良好性能。将描述组合不同的距离来创建注释方法的两种不同的方式。第一种方式是在适当地缩放每一个距离后计算不同距离的平均值。第二种方式是基于使用已知为 Lasso 的稀疏逻辑回归方法来选择相关距离。对于回归方法，可以使用包含相似和相异图像的训练集。提供给注释任务的典型训练集不直接包含这样的信息。在一些实施方式中，通过从注释训练数据创建标注集来训练 Lasso。即使这样的弱训练的 Lasso 也提供了良好性能。在一些情况下，平均距离技术执行得与噪声 Lasso 技术一样好或比噪声 Lasso 技术更好。

[0037] 现将描述用于图像注释的方法族，其中方法是建立在看起来相似的图像可能共享关键词的前提下。最后，图像注释包括从最近邻转送关键词的过程。使用图像特征来构建领域结构，导致取决于在输入图像和对应的参考图像中的相应特征之间的距离的概念的基本模型。

[0038] 图 2A 是提供确定输入图像的最近邻的一般概述的流程图。在一个或多个计算机上实现的服务器可操作来接收 (200) 数字输入图像。服务器还可以接收从它选择一个或多个参考图像的数字图像集合。可以将数字图像存储在服务器的数据储存库中或其它计算机

可读介质上。服务器然后取得 (202) 在输入图像与从数字图像集合选择的参考图像之间的整体图像距离。整体图像距离表示参考多个图像特征在作为整体的输入图像与作为整体的参考图像之间的差异程度。然后将整体图像距离存储 (204) 在服务器的数字数据储存库中。

[0039] 图像特征可以是全局的 (从整个图像生成) 或局部的 (从感兴趣点生成)。全局图像特征的示例包括色彩和纹理。色彩和纹理是图像表示的两个低层级视觉线索。常见色彩描述符基于像素色彩值的粗直方图。主要由于这些色彩特征的有效性和计算的简单性, 可以在图像匹配和索引方案内利用这些色彩特征。纹理是可以是图像表示的组分的另一个低层级视觉特征。可以利用小波特征捕捉图像纹理。具体地, Gabor 和 Haar 小波在创建稀疏而有识别力的图像特征时非常有效。为了限制个体特征的影响和偏向, 以及为了最大化所提取的信息量, 利用了多个计算色彩和纹理特征的简单和容易的方法。

[0040] 生成来自图像的在三个不同的色彩空间中的特征。这些包括红绿蓝 (RGB)、色相饱和度值 (HSV) 和 CIE 1976 L, a*, b* 值 (LAB) 色彩空间。虽然 RGB 是用于图像捕捉和显示的默认色彩空间, 但是 HSV 和 LAB 两者隔离 RGB 未捕捉的重要外观特性。例如, HSV 色彩空间对照亮在值通道中的色彩的光量进行编码, 以及 LAB 的亮度通道意在反映人类光亮度感。RGB 特征被计算为 RGB 像素值的规范化的三维直方图, 其中每一个通道中 16 个单元 (bin)。类似地, HSV (和 LAB) 特征是在 HSV (和 LAB) 色彩空间中的每通道 16 个单元的直方图。为了确定用于每一个色彩空间的距离测量, 对来自 Corel5K 数据集的人类标注的训练数据评估用于直方图和分布的三个距离测量 (KL- 偏差, L1- 距离以及 L2- 距离)。KL- 偏差是在两个概率分布之间的差异的非交换测量。如果离散随机变量的两个分布是 P1 和 P2, 则 KL 偏差被计算为 $\sum_i (P1[i] * \log(P1[i]/P2[i]))$ 。L₁ 对于 RGB 和 HSV 执行得最好, 而 KL- 偏差被发现适于 LAB 距离。在本公开的剩余部分, RGB 和 HSV 距离暗指 L₁ (Manhattan) 测量, 以及 LAB 距离暗指 KL- 偏差。也可以使用其它距离测量。例如, 在一些情况下, 可以使用余弦距离测量或堆土机距离 (EMD)。

[0041] 图像的纹理可以用 Gabor 和 Haar 小波表示。在本实施方式中, 以三个尺度 (scale) 和四个方向 (orientation) 用 Gabor 小波来过滤每一个图像。十二个响应图像被分成不重叠的区域, 以及来自每一个区域的平均响应幅值被连成特征向量 (在整个文本中该特征被称为“Gabor”)。第二特征捕捉量化的 Gabor 相位。在十二个 Gabor 响应图像的每一个中的 16x16 块上对在每一个响应像素处的相位角取平均。这些平均相位角被量化为 3 位 (8 个值), 并且被连成特征向量 (在整个文本中被称为“GaborQ”)。L₁ 距离用于 Gabor 和 GaborQ 特征。Haar 过滤器是 2x2 边缘过滤器。Haar 小波响应由图像与在三个不同方向 (水平、对角和垂直) 的 Haar 过滤器的块卷积来生成。通过执行与适当子采样的图像的卷积来获取在不同尺度的响应。在将图像重新缩放到尺寸 64x64 像素之后, 通过将 Haar 响应幅值相连来生成 Haar 特征 (该特征正好被称为“Haar”)。正如 Gabor 特征, 还考虑量化版本, 其中 Haar 响应的标记被量化成三个值 (如果响应分别是零、正数或负数, 则 0、1 或 -1)。在本公开的剩余部分, 该量化的特征被称为“HaarQ”。正如 Gabor 特征, L₁ 距离用于 Haar 和 HaarQ 特征。

[0042] 全局图像特征的其它示例包括: “极小图像”, 其是被缩小到非常小的尺寸 (例如, 缩略图) 以及逐像素比较的图像; 2) Gist 变换, 其类似于对可操纵的过滤器的小波变换和

捕捉响应；3) 几何特征的分布，诸如线或其它轮廓的统计；以及 4) 整个图像的梯度方向的直方图。也可以使用其它全局图像特征。例如，前述全局图像特征的距离测量可以包括 L1、L2、KL 偏差、余弦和 EMD。

[0043] 关于局部特征，存在从图像获取特征的两个组成部分：第一，存在“感兴趣点检测”，其中识别图像中的对在图像之间进行匹配或比较将是有用的点或区域的位置。例如，角落是常见的感兴趣点。感兴趣点检测技术的示例包括但不限于边缘检测、斑点检测、脊检测和仿射不变检测。第二个步骤是“特征提取”，其中从感兴趣点生成描述性特征向量。例如，特征向量可以描述在角落的领域中的色彩分布，或特征向量可以描述角落的角度。其它局部描述性特征的示例包括但不限于尺度不变（例如，SIFT 描述符）、旋转不变、梯度幅值、梯度方向和加速鲁棒特征（例如，SURF 描述符）。

[0044] 如上所述，用七个图像特征（例如，3 个色彩直方图和 4 个纹理特征）表示在本实施方式中的每一个图像。在不同的图像中的对应的图像特征之间的距离是“基本距离”。“复合距离”是在图像之间的合并七个特征中的部分或全部的距离测量。在一些实施方式中，复合距离可以包括额外特征。

[0045] 图 2B 是提供取得复合距离的概述的流程图。如前，在一个或多个计算机上实现的服务器接收 (210) 输入图像和参考图像。随后，从输入和参考图像中的每一个提取 (212) 七个整体图像特征。所提取的特征包括三个色彩特征和四个纹理特征。色彩特征包括在 RGB 色彩空间中的图像色彩的直方图、在 HSV 色彩空间中的图像色彩的直方图以及在 LAB 色彩空间中的图像色彩的直方图。纹理特征包括 Gabor 响应的幅值的向量、Gabor 响应的量化相位的向量、Haar 响应的幅值的向量以及 Haar 响应的标记的向量。

[0046] 然后基于从输入和参考图像提取的特征来取得 (214) 基本距离。对于 RGB 色彩空间，基本距离是在 RGB 特征向量中的每一个之间的 L_1 距离。 L_1 距离有时被称为 Manhattan 或城市块距离。类似地，用于 HSV 色彩空间的基本距离是在 HSV 特征向量之间的 L_1 距离。用于 LAB 色彩空间的基本距离是在 LAB 特征向量之间的 KL- 偏差。

[0047] 使用 L_1 距离测量来确定在输入和参考图像的 Gabor、GaborQ、Haar 和 HaarQ 特征之间的基本距离。通过对用于几个不同的距离测量的小型训练集评估每一个特征的性能，并且为每一个特征选择最佳，来确定用于每一个特征的距离测量 (L_1 、KL- 偏差)。除 L_1 和 KL- 偏差外的距离测量也可以用于计算基本距离。例如，可以使用任何 L_p 距离，直方图交集或推土机距离 (EMD)，其是在一些区域上的两个分布之间的差异的数学上的测量。

[0048] 在为每一个整体图像特征获取基本距离之后，对距离进行缩放 (216)。对于七个特征类型中的每一个，从将确保基本距离被限制在 0 和 1 之间，即基本距离被规范化的训练数据确定缩放条件。

[0049] 然后将缩放后的基本距离组合 (218) 成复合距离。例如，简单的基线方法包括对基本距离的线性组合以产生复合距离测量。即，在输入图像和参考图像之间的复合距离是七个基本距离的平均总和。尽管使用了七个特征，但是算法可以与任何数量的特征（包括一）一起容易地工作。

[0050] 在一个实施例中，通过允许每一个基本距离对总计组合距离均等地作出贡献来获取线性组合。该方法被称为联合均等贡献 (JEC)。在另一个实施例中，非均匀地组合基本距离，将优先考虑对于捕捉图像相似性更相关的那些特征，即对基本距离赋予权重。可以使

用稀疏逻辑回归技术 Lasso 来获取用于组合基本距离的权重。也可以利用额外方法。例如, 在一些实施例中, 线性组合基于如在 International Conference on Computer Vision 2007 上 Frome 等人的“Learning Globally-Consistent Local Distance Functions for Shape-Based Image Retrieval and Classification”中所描述的最大裕量 (max-margin) 方法。

[0051] 如果标记的训练数据不可获得, 或如果标签极其嘈杂, 则组合来自不同特征的距离的简单方式是使用 JEC 方法, 其中每一个独立基本距离对总计组合成本或距离均等地作出贡献。使 I_i 为第 i 个图像, 以及假设已提取了所提取的 N 个特征 $F_i = f_i^1, \dots, f_i^N$ (在我们的实例中 $N = 7$)。基本距离 $d_{(i,j)}^k$ 在两个图像 I_i 和 I_j 中的对应的特征 f_i^k 和 f_j^k 之间来计算。 N 个独立基本距离 $d_{(i,j)}^k, k = 1, \dots, N$ 被组合以提供在图像 I_i 和 I_j 之间的综合距离。在将每一个基本距离缩放至落入 0 和 1 之间的 JEC 中, 每一个缩放后的基本距离均等地作出贡献。可以从训练数据经验地确定缩放条件。如果 $\tilde{d}_{(i,j)}^k$ 表示已被适当地缩放的距离, 则在图像 I_i 和 I_j 之间的综合图像距离可以被定义为 $\frac{1}{N} \sum_{k=1}^N \tilde{d}_{(i,j)}^k$ 。该距离是联合均等贡献或简单地 JEC。

[0052] 用来组合特征距离的另一个方法将是识别对于捕捉图像相似性更相关的那些特征。由于不同的色彩 (以及纹理) 特征不是完全独立的, 所以优选的是, 确定哪些色彩 (或纹理) 特征是冗余的。带有 L_1 惩罚的逻辑回归 (即 Lasso) 可以提供用来确定不同特征的相关性的简单方式。

[0053] 为了将逻辑回归应用于特征选择, 图像注释方案应当被变换成可以用于 Lasso 训练的某事物。最后, 新的集 X 被定义, 以及每一个数据点 $x_1 \in X$ 是图像对 (I_i, I_j) 。训练集由 $X = \{x_1 = (I_i, I_j) \mid I_i, I_j \in S, i \neq j\}$ 给出, 其中 S 是输入的所有训练图像集。使 $y_1 \in \{+1, -1\}$ 为附着到每一个训练点 x_1 的标签。如果对 (I_i, I_j) 包含“相似”图像, 则 x_1 被分配标签 $y_1 = 1$, 否则 $y_1 = -1$ 。在 Lasso 中, 通过最小化下述受惩罚的负对数似然来获取最佳权重 $\hat{\omega}$:

$$[0054] \quad \hat{\omega} = \arg \min_{\omega} \sum_{l=1}^L \log(1 + \exp(-\omega^T d_{x_l} y_l)) + \lambda \|\omega\|_1 \quad (1)$$

[0055] 在此 L 是用于训练的图像对的数目, $\|\cdot\|_1$ 是 L_1 范数, d_{x_l} 是包含用于图像对 x_l 的独立基本距离的向量, 以及 λ 是经由交叉验证调整的正的权重参数。给定训练数据 $\{(x_l, y_l)\}$, 可以通过将这转换成约束优化问题来解决等式 (a)。使用 (1) 中计算的权重来线性组合基本距离提供了图像相似性的测量, 因此对结果求反以产生对应的距离。

[0056] 在将前述方案应用于图像注释中的挑战在于创建包含相似和相异图像对的训练集。典型的图像注释数据集不具有该信息, 因为每一个图像仅包含几个文本关键词, 以及不存在在图像之间的相似性 (或相异性) 的概念。在这种背景下, 共享足够关键词的任何图像对是积极训练示例, 以及不带有共有关键词的任何对是消极示例。这样的训练集的质量将取决于在图像对能够被称为“相似的”之前进行匹配所需的关键词的数量。较高阈值将确保更清洁的训练集, 但是减少了积极对的数量。相反, 较低阈值将生成足够的积极对用于以这些对的质量为代价进行训练。在这项工作中, 训练样本从 Core15K 基准的指定训练集获

取。共同具有至少四个关键词的图像被作为用于训练的积极样本对待。图 4 示出了具有至少 4 个关键词的图像对的示例,以及图 5 示出了共同具有零个关键词的图像对的示例。注意,在关键词中的较大重叠不一定转化成更好的图像相似性,暗指训练集本来是嘈杂的。

[0057] 使用 JEC 或 Lasso 来组合基本距离提供了用来计算在图像之间的距离的简单方式。使用这样的复合距离,可能的是,从训练集中的测试集找到图像的 K 个最近邻。随后,然后将关键词从它的最近邻图像分配给测试图像。

[0058] 图 3 是提供将来自输入图像的最近邻的关键词转送到输入图像的一般概述的流程图。在一个或多个计算机上实现的服务器接收 (300) 输入图像,以及在一些实施方式中,参考图像集合。服务器然后从图像集合中识别 (302) 输入图像的一个或多个最近邻图像,其中一个或多个最近邻图像中的每一个与相应的一个或多个图像标签相关联。服务器然后将多个图像标签分配 (304) 给输入图像,其中多个图像标签从与最近邻图像中的一个或多个相关联的图像标签选择。然后将具有分配的多个图像标签的输入图像存储 (306) 在服务器的数字数据储存库中。可以将包含标签的元数据与输入图像一起存储在储存库中或在别处。

[0059] 如下公开了用来将 n 个关键词从训练集中的查询图像 I 的 K 个最近邻转送给查询图像 I 的简单方法。使 $I_i, i \in 1, \dots, K$ 为 I 的在训练集中的根据渐增的距离排序的 K 个最近邻 (即, I_1 是最相似图像)。与 I_1 相关联的关键词的数目由 $|I_1|$ 表示。贪婪标签转送算法的步骤包括:

[0060] 1. 根据关键词的在训练集中的频率对最近邻 I_1 的每一个关键词进行评分。

[0061] 2. 对于 I_1 的 $|I_1|$ 个关键词,将 n 个最高评分的关键词转送给查询 $\tilde{I}$ 。如果 $|I_1| < n$,则我们还需要转送更多关键词,因此前进到步骤 3。

[0062] 3. 根据两个因素来对邻居 I_2 至 I_k 的每一个关键词进行排名:1) 它们的在训练集中的与在步骤 2 中转送的关键词的共同出现,以及 2) 它们的局部频率 (它们作为图像 I_2 至 I_k 的关键词出现的频率)。这两个因素的乘积,在规范化后,提供对这些关键词进行排名所必需的分值。基于该关键词排名,选择最好的 $n - |I_1|$ 个关键词来转送给查询 $\tilde{I}$ 。

[0063] 本质上,标签转送方案应用第一最近邻的所有关键词。如果需要更多关键词,则 (基于两个因素:共同出现和频率) 从邻居 2 至 N 选择它们。

[0064] 总的来说,图像注释方法的所描述的实施方式包括下述步骤。首先,使用复合图像距离 (用 JEC 或 Lasso 计算) 来识别输入图像的最近邻。接着,将期望数量的关键词从最近邻转送给输入图像。

[0065] 图像注释方法的性能是根据不同的图像数据集来评估的。图 6 和 7 示出了来自以下两个不同图像数据集的示例图像:Core15K 集和 ESP 集。图 6 中的图像来自已经成为图像注释界中的事实上的评估基准的 Core15K 数据集。在左边是来自数据集的 25 个随机选择的图像。在右边是两个样本图像和它们的相关联的注释。所述集包含从较大 Core1 CD 集收集的 5000 个图像。所述集从 374 个关键词的词典来注释,其中每一个图像已注释有 1 个关键词和 5 个关键词之间,以及平均起来 3.5 个关键词。在 374 个关键词中,只有 260 个出现在测试集中。

[0066] 图 7 中的图像来自 ESP 图像数据集。在左边是来自数据集的 25 个随机选择的图像。在右边是两个图像和它们的相关联的注释。ESP 集由从 ESP 协作图像标注游戏收集的

21844 个图像组成。ESP 游戏是两个玩家的游戏,其中两个玩家在没有彼此通信的能力的情况下被要求将标签分配给相同的图像。他们一共同具有一个标签,就赞扬他们成功标注图像并且向他们展示下一图像。因此,每次向一对玩家示出图像时获取至多一个标签。随着向更多玩家示出每一个图像,生成了禁忌词列表。不允许游戏的随后玩家在被示出相同图像时分配禁忌词。这些规则确保每一个图像将由许多不同玩家分配许多不同的标签。所使用的所述集包含由 269 个关键词注释的自然景物、人造景物和对象的各种各样的图像。每一个图像注释有至少一个关键词、至多 15 个关键词以及平均起来 4.6 个关键词。

[0067] 使用标签转送将五个关键词分配给每一个图像。在一个实施例, JEC 方案与标签转送算法一起用于将五个关键词分配给 Core15K 数据集中的每一个测试图像。图 8 示出了已注释的图像的示例,即用于多个样本图像的预测关键词与地面实况(例如,人类分配的)关键词的比较。由于人类注释通常包含少于五个关键词,在一些情况下, JEC 预测不在地面实况集中但是正确描述了图像内容的关键词。例如,预测图中的第一图像具有关键词 formation(编队)。可论证地,这是对图像中的飞机的正确描述,尽管它不是人类分配的关键词中的一个。

[0068] 图 9 示出了已注释的图像的示例。使用 JEC 方案和 ESP 图像数据集来注释了图像。尽管使用 JEC 注释方法的预测的关键词没有与人类注释完全重叠,在许多情况下,“不正确的”预测关键词正确地描述了图像。例如,在示出了一个男子坐在充满镶框图片的墙前的沙发上的第四个图像中, JEC 分配的关键词将图像可论证地描述得与通过 ESP 游戏生成的那些一样(或更)准确。

[0069] 在图像注释任务中的挑战是知道要描述图像的内容多少个关键词是必需的。在标签转送阶段期间仅分配 5 个关键词人为地限制了对于图像中的许多图像可以正确调用的关键词的数量。尽管增加分配给图像的关键词的数量可以帮助增加调用(例如,在极端情况下,如果将所有关键词分配给图像数据集中的每一个图像,则对于所有关键词可以确保 100%调用),但是这将导致在精确度上的下降。为了使用注释方法将多于 5 个关键词分配给图像,在标签转送阶段期间所使用的最近邻的数量被建立为看见足够独特关键词所需的最小量。然而,这可以随着调用增加导致在精确度上的降低。这是由于最近邻结构用于标签转送的事实,这对于较小数量的邻居有意义。然而,它随着邻居数量增加而引入更多的错误,这对于分配许多关键词是必然的。

[0070] 将描述性关键词分配给图像允许用户使用基于纯文本的查询来搜索图像。评估图像检索引擎的性能不同于评估注释引擎的性能,因为在检索时,兴趣仅在于与给定关键词相关联的最先几个图像的质量。图 10、11 和 12 示出了分别在三个不同的图像数据集中为多个不同的关键词检索的最先的几个图像的示例。

[0071] 即使对于特别具有挑战性的关键词(例如, cyclist、skull、diagram 以及 tie),最先检索到的图像中的许多也是正确的。并且,许多关键词具有多个含义,通常被称为“词义”。在一些这样的情况下,检索到的图像跨越单词的许多含义。例如,在图 12 中用于关键词 ring 的检索到的图像表示单词“ring”的几个不同含义。

[0072] 所建议的图像注释方法组合在非常简单全局色彩和纹理特征上的基本距离测量。使用这些组合距离来计算的 K-最近邻形成简单贪婪标签转送算法的基础。

[0073] 在本说明书中描述的主题和操作的实施例可以在数字电子电路中,或者在计算机

软件、固件或硬件,包括在本说明书中公开的结构以及其结构等同物中,或者在以上的一个或多个的组合中实现。可以将在本说明书中描述的主题的实施例实现为一个或多个计算机程序,即计算机程序指令的一个或多个模块,其被编码在计算机存储介质上、由数据处理装置执行或控制数据处理装置的操作。替代地或另外地,程序指令可以被编码在非自然生成的传播信号上,所述信号例如机器生成的电的、光学或电磁的信号,其被生成以编码用于传输到适当接收器装置的信息以供数据处理装置执行。计算机存储介质可以是下述或被包括在下述中:计算机可读存储设备、计算机可读存储基片、随机或串行存取存储器阵列或设备、或以上的一个或多个的组合。

[0074] 可以将在本说明书中描述的操作实现为由数据处理装置对存储在一个或多个计算机可读存储设备上或从其它源接收的数据执行的操作。

[0075] 术语“数据处理装置”包含用于处理数据的所有类型的装置、设备以及机器,包括例如可编程处理器、计算机、芯片上的系统或以上的组合。装置可以包括专用逻辑电路,例如 FPGA(场可编程门阵列)或 ASIC(专用集成电路)。除硬件外装置还可以包括创建用于讨论中的计算机程序的执行环境的代码,例如构成处理器固件、协议堆栈、数据库管理系统、操作系统、例如虚拟机的跨平台运行环境、或以上一个或多个的组合的代码。装置和执行环境可以实现各种不同的计算模型基础设施,诸如 web 服务、分布式计算和网格计算基础设施。

[0076] 计算机程序(也称作程序、软件、软件应用、脚本或代码)可以以任何形式的编程语言编写,所述编程语言包括编译或解释语言、说明性或过程性语言,并且其可以以任何形式部署,包括作为独立程序或作为模块、组件、子程序、对象或适于在计算环境中使用的其它单元。计算机程序可以但是不必对应于文件系统中的文件。可以将程序存储在保持其它程序或数据的文件(例如,存储在标记语言文档中的一个或多个脚本)的一部分、专用于讨论中的程序的单个文件或者多个协调文件(例如,存储一个或多个模块、子程序或部分代码的文件)中。可以将计算机程序部署为在一个计算机上或者在位于一个地点或跨多个地点分布并且由通信网络互连的多个计算机上执行。

[0077] 在本说明书中描述的过程和逻辑流可以由执行一个或多个计算机程序的一个或多个可编程处理器执行以通过操作输入数据并且生成输出来执行功能。过程和逻辑流还可以由专用逻辑电路执行,以及装置还可以被实现为专用逻辑电路,专用逻辑电路例如 FPGA(场可编程门阵列)或 ASIC(专用集成电路)。

[0078] 适于执行计算机程序的处理器包括例如通用和专用微处理器两者,以及任何类型的数字计算机的任何一个或多个处理器。通常,处理器将从只读存储器或随机存取存储器或两者接收指令和数据。计算机的主要元件是用于执行或运行指令的处理器和用于存储指令和数据的一个或多个存储器设备。通常,计算机还包括用于存储数据的一个或多个海量存储设备,例如磁盘、磁光盘或光盘,或可操作地耦接以从所述一个或多个海量存储设备接收数据或向所述一个或多个海量存储设备传送数据,或两者。然而,计算机不必具有这样的设备。此外,可以将计算机嵌入另一个设备中,所述设备例如移动电话、个人数字助理(PDA)、移动音频播放器、全球定位系统(GPS)接收器,仅列出一些。适于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质和存储器设备,包括例如:半导体存储器设备,例如 EPROM、EEPROM 和闪存设备;磁盘,例如内部硬盘或可移动盘;

磁光盘 ; 以及 CD-ROM 和 DVD-ROM 盘。处理器和存储器可以由专用逻辑电路补充, 或合并入专用逻辑电路。

[0079] 为了提供与用户的交互, 本说明书中描述的主题的实施例可以在具有下述的计算机上实现 : 用于向用户显示信息的显示设备, 例如 CRT (阴极射线管) 或 LCD (液晶显示) 监视器, 以及用户通过其可以向计算机提供输入的键盘和指示设备, 例如鼠标或跟踪球。也可以使用其它类型的设备来提供与用户的交互 ; 例如, 提供给用户的反馈可以是任何形式的感知反馈, 例如视觉反馈、听觉反馈或触觉反馈 ; 以及可以以任何形式, 包括声学的、话语或触觉的输入, 接收来自用户的输入。

[0080] 本说明书中描述的主题的实施例可以在包括例如作为数据服务器的后端组件、或者包括例如应用服务器的中间件组件、或者包括例如具有用户通过其可以与在本说明书中描述的主题的实施方式交互的图形用户界面或 Web 浏览器的客户端计算机的前端组件、或者一个或多个这样的后端、中间件或前端组件的任何组合的计算系统中实现。系统的组件可以通过例如通信网络的任何形式或介质的数字数据通信互连。通信网络的示例包括局域网 (" LAN ") 和广域网 (" WAN "), 例如因特网。

[0081] 计算系统可以包括客户端和服务端。客户端和服务端通常彼此远离并且典型地通过通信网络交互。客户端和服务端的关系依靠在各个计算机上运行并且彼此具有客户端 - 服务端关系的计算机程序产生。

[0082] 虽然本说明书包含许多具体实施方式细节, 但是这些细节不应当被解释为对本发明或可以主张的内容的范围的限制, 而应当被解释为对具体到本发明的特定实施例的特征的描述。还可以将在本说明书中在分立的实施例的情境中描述的某些特征组合在单个实施例中实现。相反地, 也可以将在单个实施例的情境中描述的各种特征分离地在多个实施例中实现或在任何适当的子组合中实现。此外, 尽管可能在上文将特征描述为在某些组合中起作用, 甚至最初主张如此, 但是可以在一些情况下将来自所主张的组合的一个或多个特征从组合中删去, 并且可以将所主张的组合指向子组合或者子组合的变体。

[0083] 类似地, 虽然在附图中以特定顺序描绘了操作, 但是不应当将这理解为需要以所示的特定顺序或者以连续顺序执行这样的操作、或者需要执行所有图示的操作才能达到期望的结果。在某些情况下, 多任务以及并行处理可以是有利的。此外, 不应当将在上述实施例中的各种系统组件的分离理解为在所有实施例中均需要这样的分离, 而应当理解的是, 通常可以将所描述的程序组件和系统集成到一起成为单个软件产品或封装为多个软件产品。

[0084] 因此, 已经描述了本发明的特定实施例。其它实施例在所附权利要求的范围内。例如, 可以以不同的顺序来执行权利要求中记载的动作并且仍然达到期望的结果。

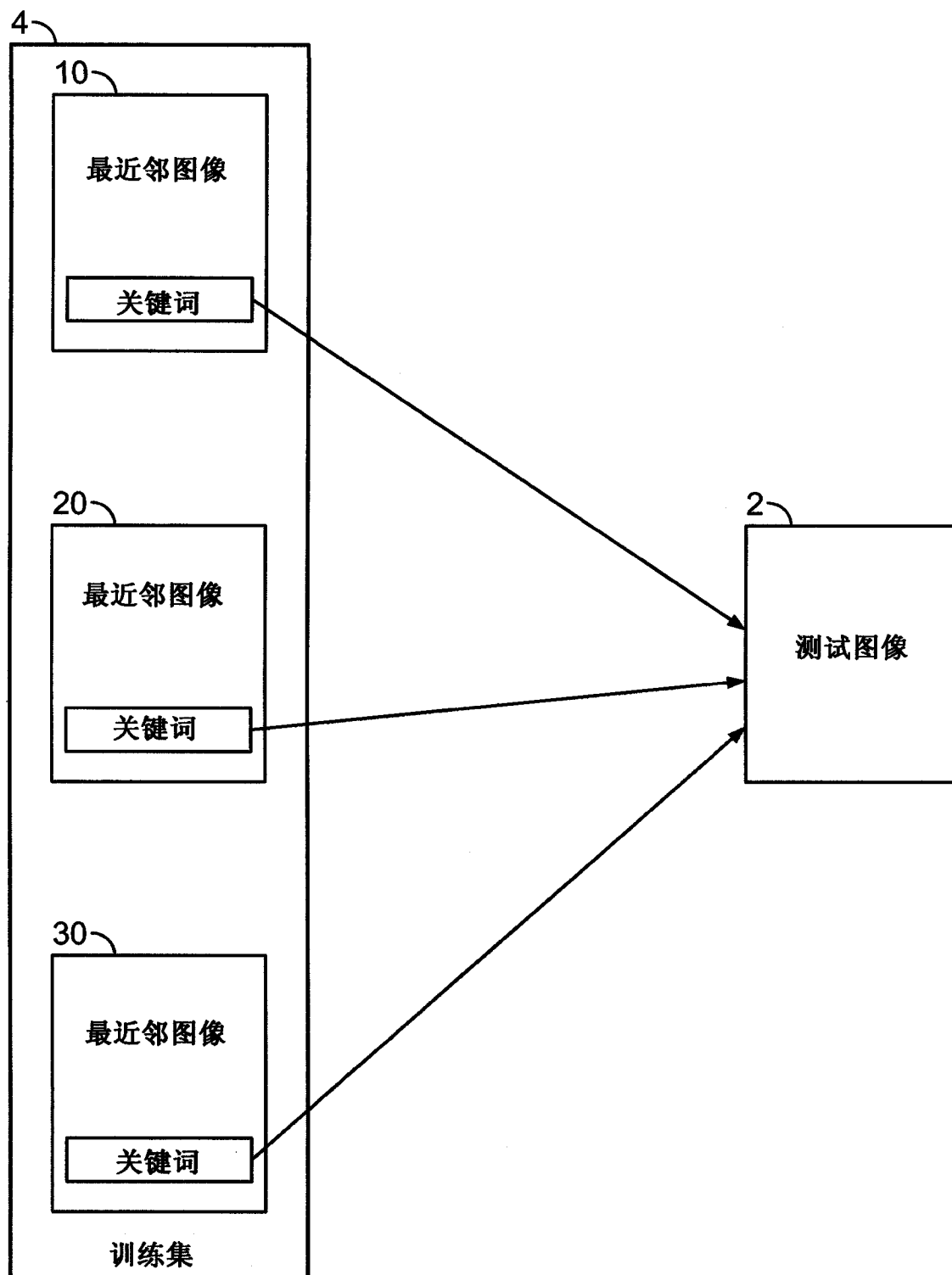


图 1

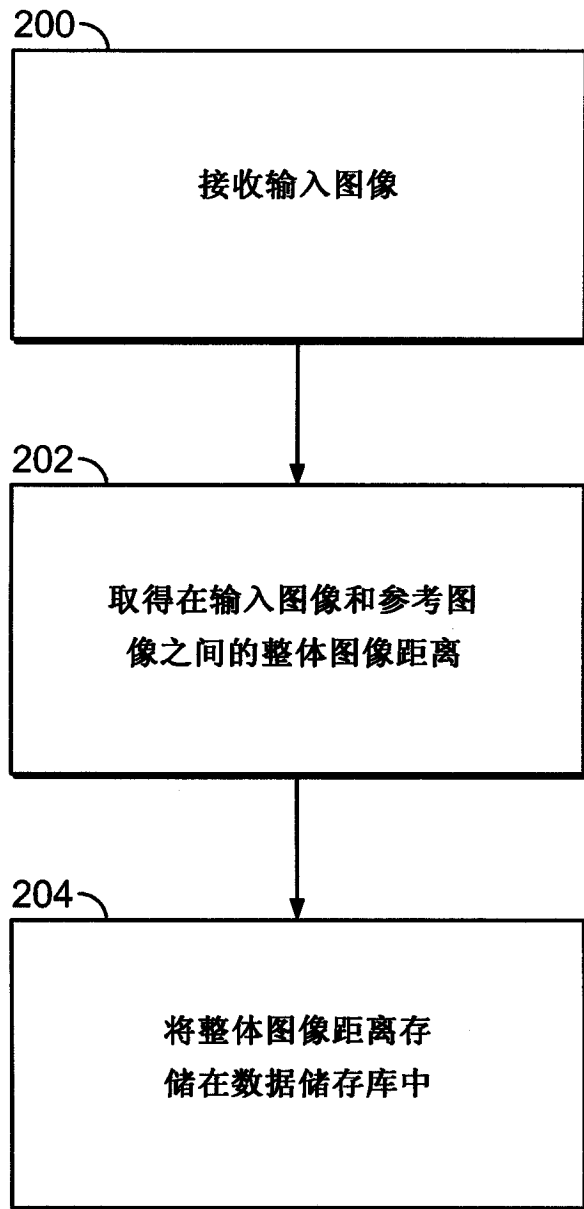


图 2A

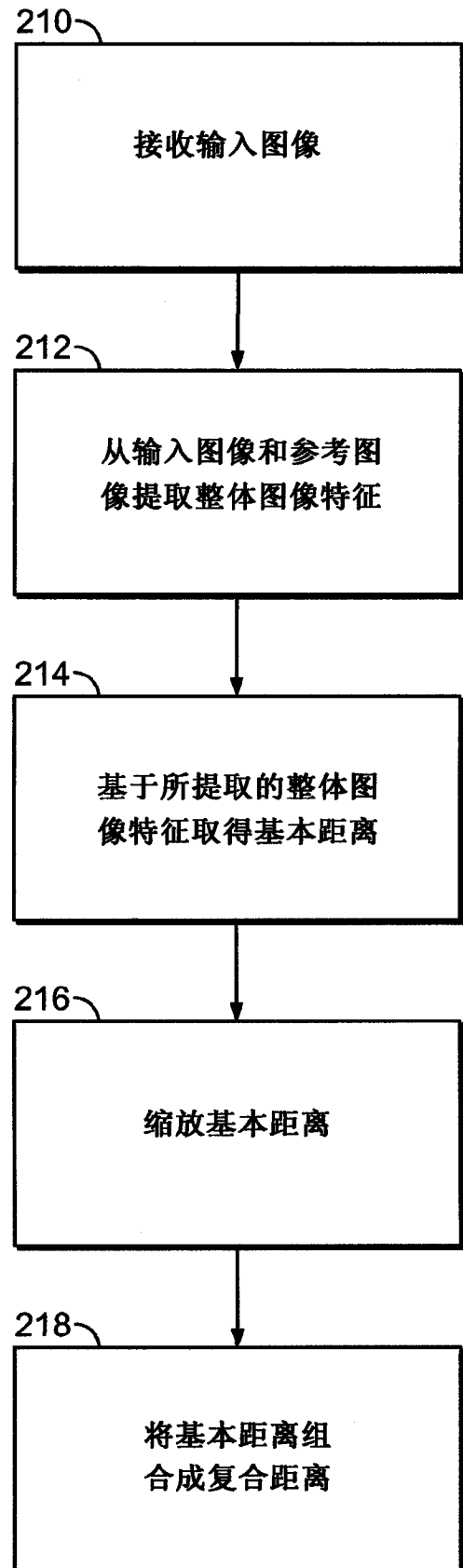


图 2B



图 3



图 4

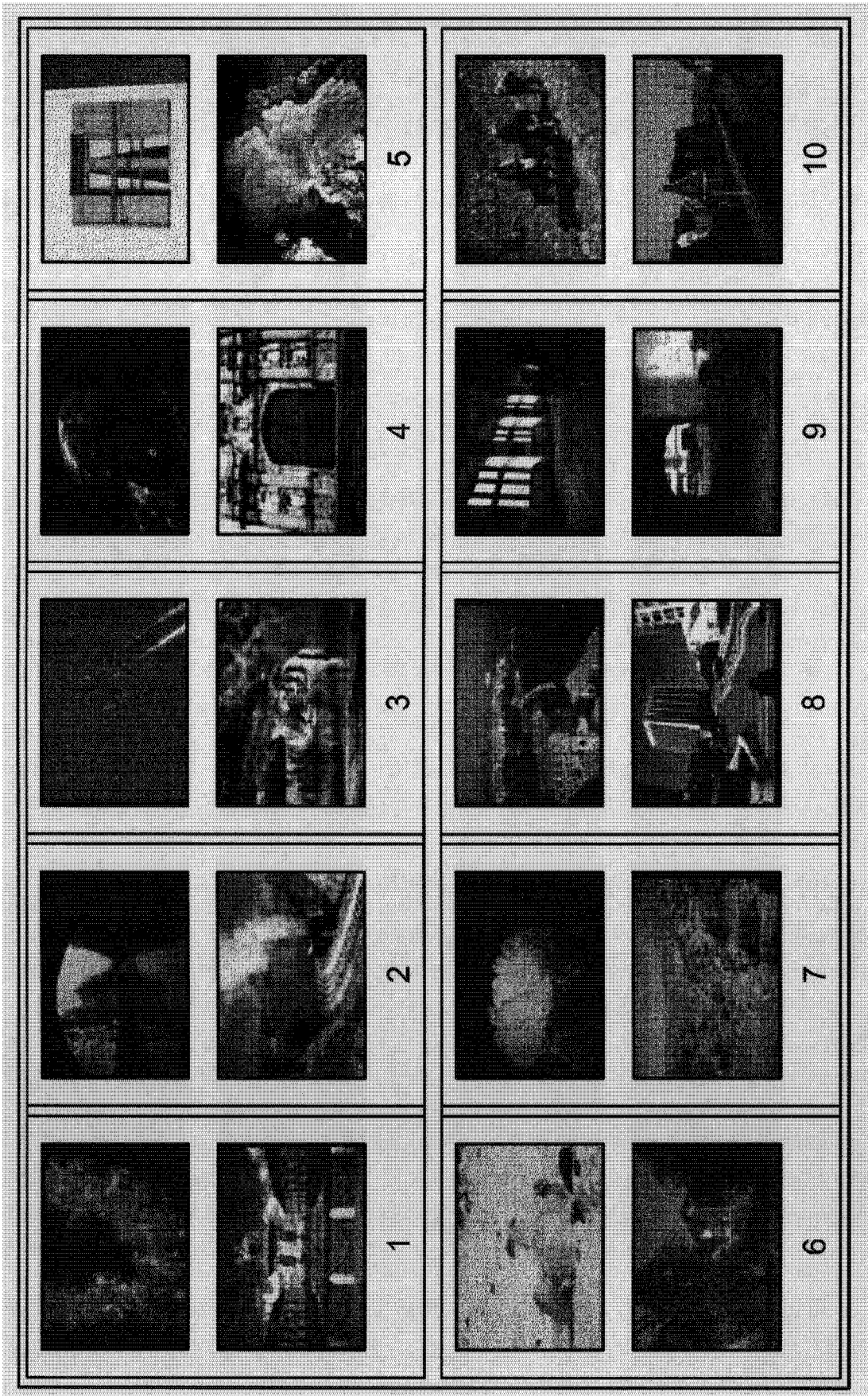


图 5

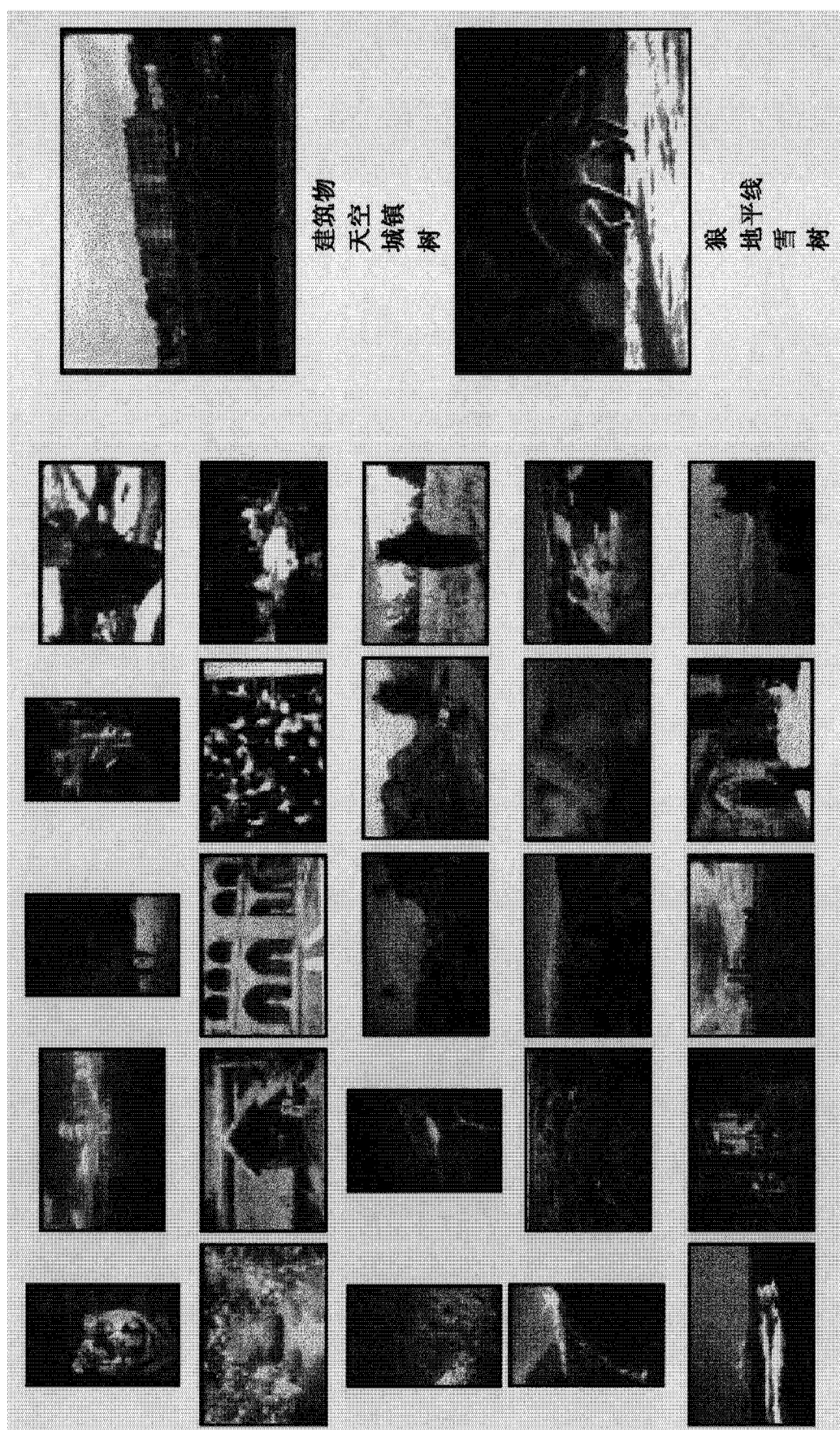




图 7

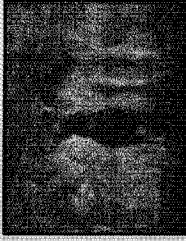
					
预测的关键词	天空、喷射、 飞机、烟雾、编队	草、岩石、 沙、山谷、峡谷	太阳、水、 大海、波浪、鸟	水、树、草、 鹿、白色尾巴	熊、雪、木材、 鹿、白色尾巴
人类注释	天空、喷射、 飞机、烟雾	岩石、沙、 山谷、峡谷	太阳、水、 云、鸟	树、森林、 鹿、白色尾巴	树、雪、 木材、狐狸

图 8

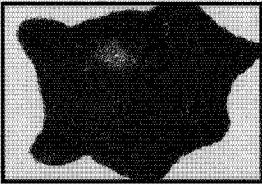
					
预测的 关键词	比基尼、女孩、 草、头发、女人	熊、黑色、棕色、 鼻子、白色	乐队、光线、 男人、音乐、演出	男人、老、 图片、红色、墙壁	云、草、绿色、 小山、红色
人类注释	床、女孩、女人	动物、熊、黑色、 棕色、头、鼻子	乐队、光线、男人、 音乐、红色、车轮	黑色、男人、 老、红色、坐	云、灰色、绿色、 山脉、图片、岩石、 天空、石头

图 9

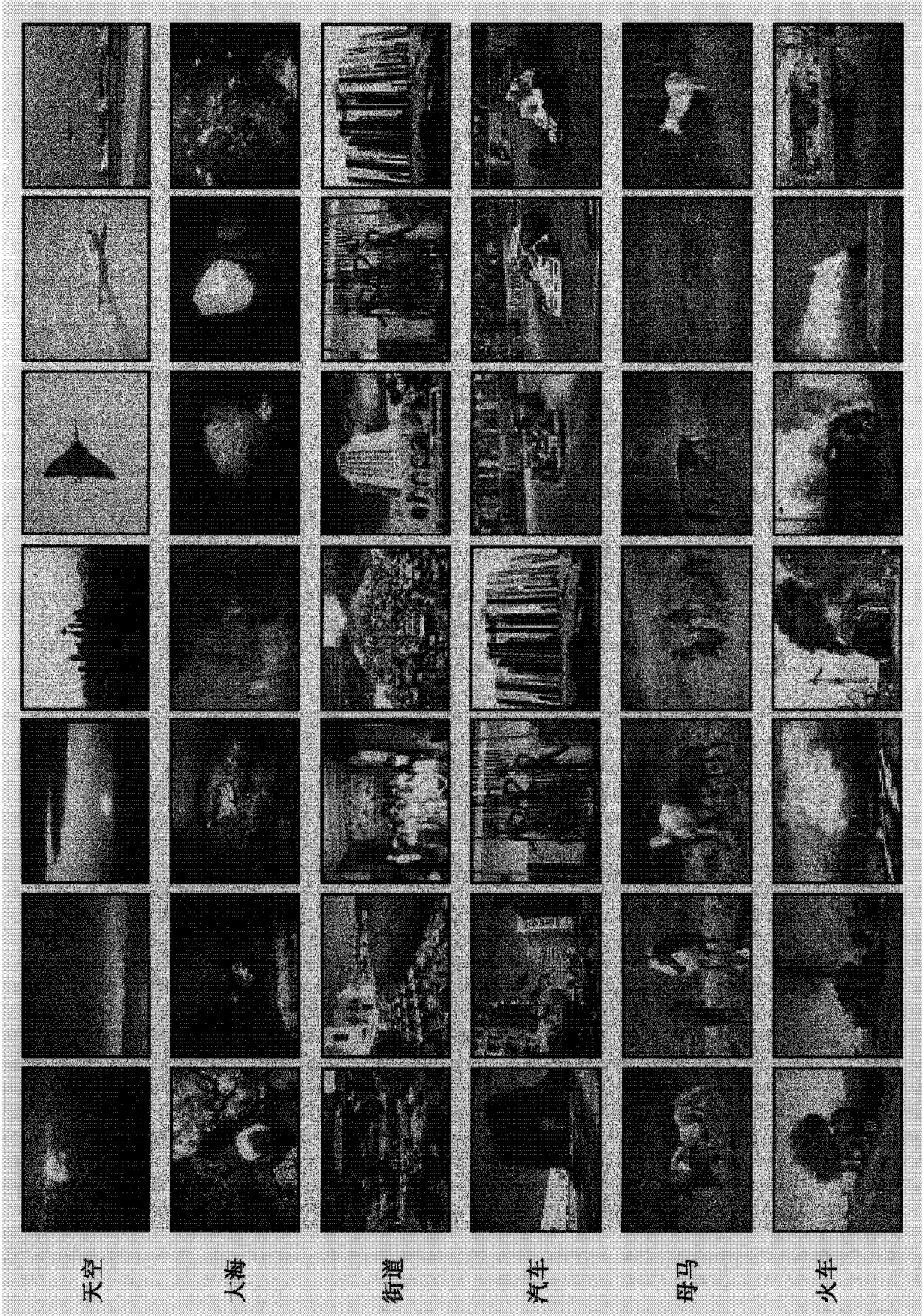


图 10

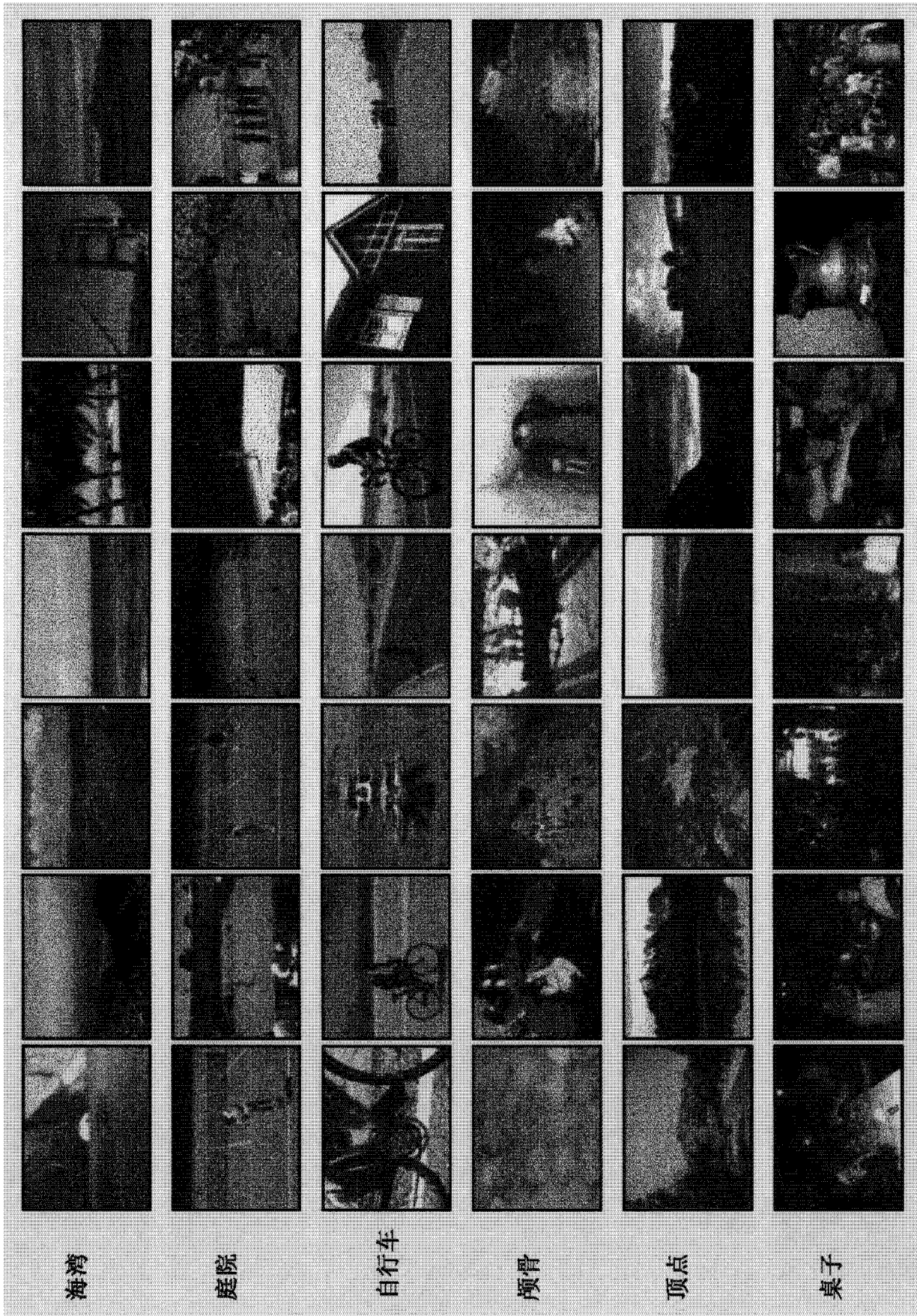


图 11

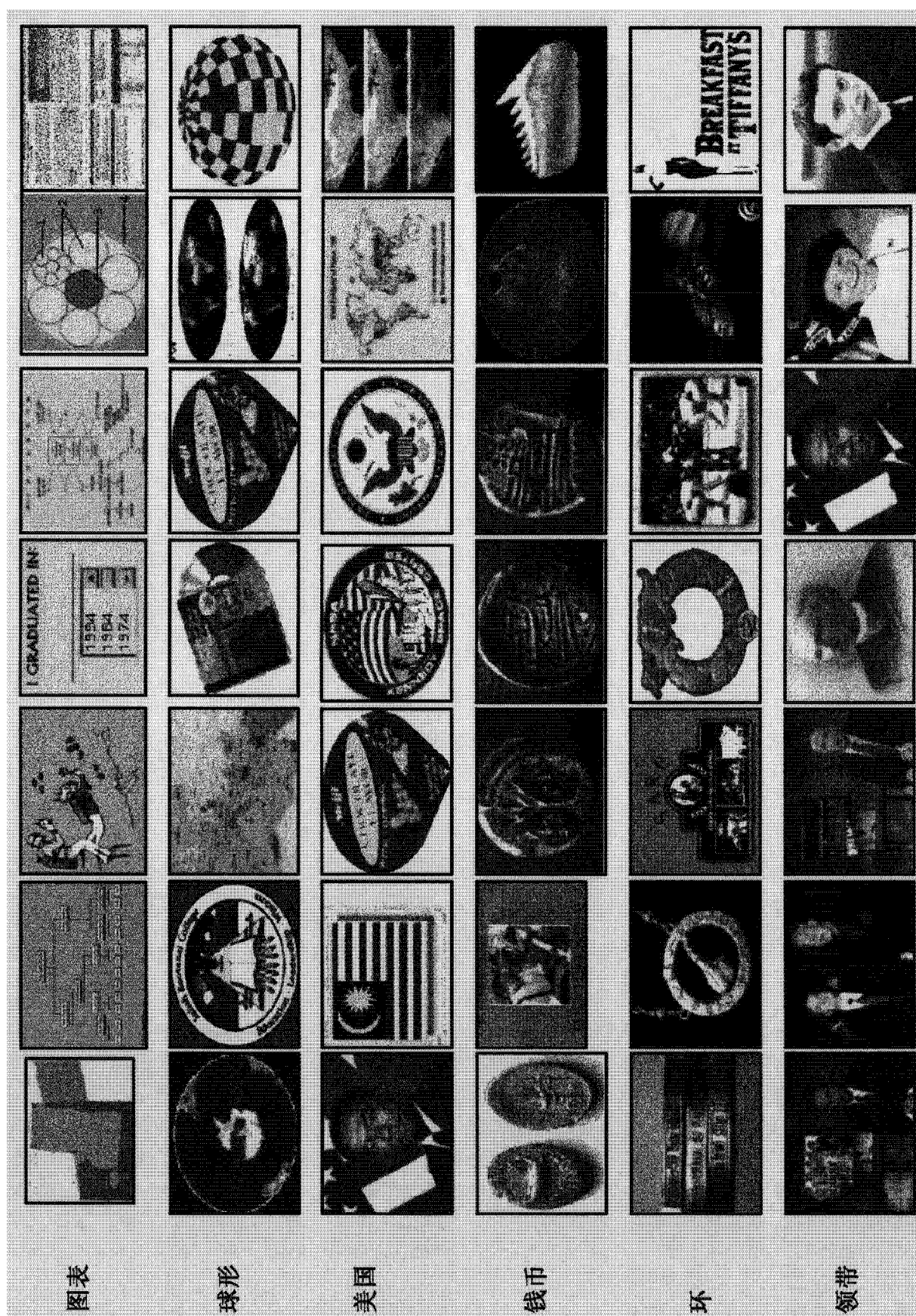


图 12