

Large-Scale Image Retrieval Method Based on Vocabulary Tree

QI Jin, ZHAO Jian, XIE Yu, CHEN Xiao-ning

College of Information Science and Technology, Northwest university, Xi'an, China

1. qijin1117@yahoo.com.cn, 2. zjctec@nwu.edu.cn

Abstract: The techniques for content-based large-scale image retrieval tasks are discussed, and the methods of index-tree structures are analyzed emphatically. Introducing the concept of a vocabulary tree, an improved algorithm for large-scale image retrieval was realized. The vocabulary tree is a data structure based on visual words for indexing images. Each leaf node of the tree is from visual word computed by the K-means clustering algorithm. The query time of system dose not increase linearly as the database size grows. The method can improve the retrieval efficiency and meet two of the major requirements: adaptability and scalability.

Keywords: large-scale image retrieval; index structure; tree structure; vocabulary tree

基于词汇树的大规模图像检索技术

齐锦, 赵健, 陈晓宁, 谢瑜

西北大学 信息科学与技术学院, 西安, 中国, 710069

1. qijin1117@yahoo.com.cn, 2. zjctec@nwu.edu.cn

【摘要】对基于内容的大规模图像检索技术进行了讨论,重点分析了树形索引结构方法,在此基础上引入词汇树的概念,实现了一种基于词汇树的大规模图像检索的改进算法。词汇树是一种高效的基于视觉关键词检索图像的数据结构,其结点由视觉关键词进行K均值聚类形成。利用词汇树结构能够满足大规模图像检索自适应性和可扩展性的要求,系统的检索时间不会随着数据库的规模增大而线性增加,从而实现高效的大规模图像检索。

【关键词】大规模图像检索;索引结构;树形结构;词汇树

1 引言

从庞大的数据库中快速、准确地检索到需要的信息是现代信息科学发展的重要研究内容。目前采用的图像检索技术在进行大规模图像检索时,由于信息量的巨大往往会引起检索效率大幅度下降的情况。词汇树是一种基于视觉关键词检索图像的数据结构,是一种比其他结构检索更高效的数据结构。本文提出的快速检索策略,就是在基于内容的图像检索模型的基础上,改进了传统的图像检索算法,通过引入词汇树结构进行高效检索,提高了大规模图像检索效率。

2 大规模图像检索的主要研究技术

传统的图像检索技术通常采取顺序检索的算法,然而对于大容量、高维数的图像数据来说,这种检索方法在效率上显然不能满足需要。目前,基于内容的大规模图像检索的主要工作集中在识别和描述图像的颜色、纹理、形状、空间关系上,以及图像特征匹配的相似度量上。当图像特征提取和相似度量方法确定以后,检索就变成了在图像数据库中查找与给定图像最相似的图像的过程。于是,构建一个适用于大规模图像数据库的索引结构就成为重要问题。

对大规模图像数据库索引结构的要求通常有以下几点:(1)必须能够处理高维特征。目前现有

资助信息:西北大学研究生重点课程《数值分析》项目基金(09YKC21),陕西省教育厅科技立项项目(112Y029)

的多维索引结构，大多都无法很好的处理高维数据。当数据维数增加时，结点数呈指数性增加会造成其性能变差，导致“维数灾难”。因此，构建索引结构时，应保证其结点数不随维数的增加而呈指数增长。(2) 存储利用率高。由于树形结构的高度能够保持的较低，因此存储利用率高能够减少查询的消耗。例如范围查询，查找与给定对象在一定距离范围内的对象，更易受到存储利用率的影响。倘若存储利用率低，对象则可能分布在更多的结点中，从而增加了查询的消耗。(3) 目录区域的重叠尽可能小。一般来说，重叠则说明这个区域被不止一个目录区域所覆盖。随着数据量增大、数据结构高度增加、数据维数增多，重叠区域会随之增长。通常，重叠将导致多重路径的重复穿越，这给查询操作的性能造成负面影响。(4) 索引结构必须使用相似度量。图像特征通常是以数据点的形式存在于高维特征空间中，相似度通常可以用两点间的欧氏距离来衡量。(5) 能有效的支持各种查询类型，例如范围查询、精确匹配查询、部分匹配查询、K-NN 查询等。(6) 有效处理相似查询。

3 基于树形的大规模图像检索算法

目前广泛采用的基于树形的大规模图像检索算法，在特征提取和相似性度量的基础上，增加了对图像数据库中图像特征的聚类步骤。该步骤是在对数据库中的图像进行特征提取后，对其特征进行分层次聚类，充分利用图像之间的相似性信息，将被检索的图像根据其相似程度形成多个簇。同时，聚类过程中按不同的精度对类的密度分布曲线进行小波光滑化，以形成多个层次的聚类，从而提出建立层次索引，其整体类似于一个树形结构。经过如此处理，样本图像根据索引能够快速检索到匹配的簇，此时仅需要与该簇中包含的图像作特征比较，就能够找到相似的图像，从而在保证检索精度的同时也能够大大提高检索效率。

几种具有代表性的基于树形的索引结构有：

(1) OMNI:

2001 年 Roberto Figueira Santos 等人提出了 OMNI 的概念。提出在全局范围的数据集内选取适当数量的焦点集，然后计算其他的对象到每个焦点之间的距离，构成 Omni-坐标，通过减少距离计算来改善其性能，提高检索速度。

(2) LPC-file

2002 年 Guang-Ho Cha 等人提出了局部极坐标文档 (local polar coordinate file, LPC-file)。其基本思想是把向量空间划分成矩形的单元，每个单元里用局部极坐标来表示数据。检索的时候通过余弦定理计算查询点到单元里具有相同角度和半径的点的最近值，经过初步过滤再精确计算出距离。

(3) NB-Tree

2003 年 Manuel J. Fonseca 等人提出了一种新的索引结构 NB-Tree，这种访问方法通过计算数据点的欧式距离，将多维点映射到一维直线上，并利用 B+-Tree 来存取数据点。使用 B+-Tree 有效的保证顺序查询的简单化。

(4) R-Tree 族

Guttman 在 1984 年提出了 R-Tree 索引结构。R-Tree 是 B+-Tree 的多维扩展，是一种用来访问二维或更高维区域对象组成的空间数据，树是高度平衡的。其叶结点由指向空间数据库中的空间对象，以及包含空间对象的包络组成。R-Tree 的搜索过程从根结点开始，递归调用搜索过程直到叶结点为止，其查询效率很大程度是由覆盖率和重叠率决定的。另一个影响 R-Tree 性能的是数据插入的次序，在一些插入和删除步骤之后重新组织结点将会提高 R-Tree 的效率。在此基础上针对不同空间运算提出了不同改进，形成了一个繁荣的索引树族。

Sellis 为了克服 R-Tree 兄弟结点的重叠问题提出了 R+-Tree，其结点不一定要求至少是半满的，任何非叶结点记录不重叠，叶结点的记录可能重复出现；1990 年，Beckman 和 Kriegel 提出了 R 树的变种 R*-Tree，通过最小化覆盖率改善了 R-Tree，该算法在 R-tree 的插入和分裂算法中引入了强制重新插入机制，使树的结构得到优化；SS-Tree 是 R*-Tree 的改进，是为多维点数据设计的一种索引结构，对区域采用球型包络而不是矩形包络，从而提高了最近邻查询的性能；X-Tree 是 R*-Tree 的变种，是线性数组和层状的 R-Tree 的混合体，通过引入不重叠的分裂和超级结点，减少了最小边界矩形之间的重叠，提高了查询效率。

4 基于词汇树的图像检索算法

目前的图像检索技术在进行大规模图像检索时仍然存在着一些问题：不论哪种图像检索算法，一般都必须满足快速性和准确性的要求，然而当它们被应用到大规模图像系统或图像库中进行检索

时, 这些方法通常就无法满足大规模时出现的一些新要求, 如对算法的自适应性和可扩展性的要求。自适应性是指在现实系统中, 大量的图像被不断的添加到图像数据库中, 这就要求图像检索算法能够快速调整自己, 从而可以更有效的为新图像数据库提供查询服务; 可扩展性主要是指图像检索系统还必须有能力满足多个用户同时查询的能力。自适应性和可扩展性的要求无法满足, 是基于内容的图像检索目前还不能真正应用于大规模图像检索的主要原因, 现有的图像检索算法都没有彻底解决这个问题。本文在对大规模图像检索的特性进行了详细分析的基础上, 提出了在进行大规模图像检索研究中采用词汇树的方法, 使索引结构满足以上要求, 并实现自适应性和可扩展性, 使其更加满足实际的需求。

词汇树的概念是麻省理工学院 (MIT) 计算机科学与人工智能实验室 (Computer Science and Artificial Intelligence Laboratory) 的学者 John J. Lee 等人在 ICCV'07 会议上首次提出的。词汇树是一种基于视觉关键词检索图像的数据结构。利用词汇树的结构可以使得在图像检索算法中自适应性和可扩展性的要求容易满足。

定义 1: 一棵 m 阶词汇树, 它或者为空, 或者满足如下性质:

- (1) 根结点至少包含 2 个子结点。
- (2) 除根结点和失败结点外, 所有结点都至少具有 2 个子结点。
- (3) 通过分层 K -均值聚类形成一个分层的量化, K 为该结点子结点的个数。
- (4) 结点由接近各层聚类中心的描述符构成。

将训练样本图像作为输入可提取到大量的特征, 词汇树的学习使用分层 K -均值法。需要控制的两个参数: 水平数量 L 指定的结果树的高度, 以及分支因子 B 指定的子结点的数量。一个树的叶结点被称为一个视觉关键词。作为传统的词汇包模型, 它代表一个临近特征空间的小局部。该结构的层次布局允许我们快速地量化特征。树的布局如图 1 描述。

词汇树能够有效地将输入图像 I 转换为一系列视觉关键词来表示。通过从输入图像 I 中提取特征, 对 I 中的每个特征 p 在词汇树中找到一个相应的视

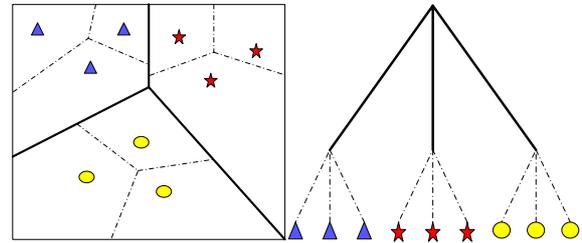


Figure 1. The left image shows a vocabulary tree with nine visual words partition the two-dimensional feature space. The right image shows the structure of this tree.

图 1. 左侧图像显示了用 9 个视觉关键词分割二维特征空间的词汇树; 右侧是该树的结构图。

觉关键词。从树的根结点开始搜索。在每个结点上访问每个子结点, 并且将 p 与每个子结点的中心相比较。转移到中心最接近 p 的子结点上, 并继续搜索, 直到最终到达一个叶结点。叶结点将成为与 p 相对应的视觉关键词。树的每个视觉关键词, 与一个训练样本图像中包含该视觉关键词储存指针的倒文件相关。它也一直在不断地追踪每个视觉关键词出现在每幅图片中的次数。在倒文件中, 每个词条是一组 $[i;c]$, i 是对一幅图像的检索, c 是出现在图像中的视觉关键词具体有多少次的总数。查询图像 q 用一组特征点表示, 与 q 类似的图像, 通过查找与 q 的特征一致的视觉关键词的相关图像, 可以快速被找到。然后, 通过对每个关键词的表决, 为这些图像排序。参加表决的方法如同基于内容的快速图像检索。

词汇树生长的具体步骤如下:

- 1、在词汇树插入新的视觉关键词;
- 2、清除拥挤的叶结点, 为重构做准备;
- 3、将前一步骤取代的视觉关键词聚类, 并插入到词汇树。

每个步骤的行为都由三个参数控制: 容量约束 C 界定的叶结点的最大数量, 邻域大小 S 规定的参与任何重新聚类步骤的最大数目点, 以及缩减因子 R 控制的重新聚类步骤后在叶结点上允许的最大数量的点数。

步骤 1: 通过将视觉关键词放入一个结点, 从而实现由图像提取的特征到树插入一个新结点, 该结点的中心比其他任何兄弟结点更接近特征。如有必要, 可以通过记录从结点中的任何点到聚类中心的最大的距离, 更新结点的数量。同时, 升级相关的金字塔, 从而反应出新结点的插入。如果由于插入

点,使叶结点的数量超过容量约束 C ,则触发算法的第二步,开始释放邻域上的点,将拥挤的结点释放到一个临时的集合 Φ 中。

步骤 2: 释放附近的一些点,使其数量不超过邻域的大小 S 。可以采取一个贪心的做法,首先删除视觉关键词 θ 以触发更新,并继续一个个的删除 θ 的兄弟结点(从离 θ 最近的到最远的)。每当视觉关键词通过这个步骤删除,它的结点也被释放到步骤 1 临时的集合 Φ 中。如果没有兄弟结点剩余,并且释放点的集合 Φ 尚未达到规模 S 时,算法移动到前一步,并开始删除视觉关键词 θ 的双亲和兄弟结点。然而,在这个结点删除的进程中我们忽略了不能释放的结点,其后代由于包含过多结点,以至于将不可避免的造成 Φ 的规模超过 S 。该删除和释放的进程,持续到所有邻域可消除的结点被全部删除。

步骤 3: 对释放的点的集合 Φ 重新聚类。设 X 是所有删除结点中最低的共同祖先(即到达该结点时,删除步骤停止)。假设在删除步骤之后, X 剩余 m 个子女。由于每个结点最多有 B 个子女,我们在集合 Φ 中用 $K=B-m$ 进行 K -均值聚类,以保持分支因子常量 B 贯穿于整个树。缩减因子 R 是在结果分布上指定的限制比率:如果聚类的结果使任何结点有超过 $R*C$ 个子女,我们递归执行更多次的 K -均值聚类,直到制约因子达到满意的值。

通常情况下,我们重新训练与原来那个具有相同高度的一个新的子树,并希望重新训练能够重新分配点,使其更加均匀。然而,有些时候新的子树在同一高度时可能无法提供所期望的缩减因子。在这种情况下,我们允许子树增加一个高度,以获取满足缩减因子所需要的能力。应注意的是,高度的增加只发生在局部。因此,由此产生的词汇树可能有一些分支比其他分支长。

词汇树直接定义了量化,从而检索和量化完全整合,本质上成为一体。通过使用分层 K -均值方法,词汇树能够提供更加高效的训练。大量具有代表性的描述符向量被用于树的无监督训练中。与传统的树形结构不同,当视觉关键词的数量很大时,词汇树允许在次线性时间的关键词查询,而不是扫描全体关键词去寻找匹配的图像,这就大大提高了检索效率。另外,在一个动态环境中,新的图像不断的加入数据库,检索决策由每幅新添加的图像构成,

并且不能延迟。随着数据库规模的扩大,检索过程中识别系统使用事物的词汇树也应该不断成长,以允许系统适应其在环境中遇到的图像,所以词汇树本身就有一定的自适应性。

基于词汇树的大规模图像检索系统,利用视觉关键词进行图像检索,由于视觉关键词很多,线性搜索比较慢,所以将视觉关键词按照 K 均值聚类构成词汇树,通过快速搜索词汇树的结点进行特征匹配,达到对大规模图像数据进行检索的目的。系统的检索时间不会随着数据库的规模增大而线性增加,与传统技术相比在对大规模图像数据库进行检索时速度提高,综合性能更优。我们针对麻省理工学院的 LabelMe 图像库进行实验(该图像库现容量为 6G),对图像的颜色、纹理及形状等视觉内容特征进行提取,构建出的词汇树索引结构,检索效率和准确性均有一定的提高。

5 结束语

大规模图像检索技术发展迅猛,涉及的研究领域广泛,许多问题还有待于进一步探索。将词汇树算法引入大规模图像检索技术的索引结构部分,其检索效率以及检索准确性都有显著的提高。基于词汇树的图像检索技术十分值得进一步研究,并且有可能为下一代互联网和信息服务的规模发展发挥重要作用。同时,该技术的应用能够促进互联网上大规模图像检索的实现,同时进一步提高网络和数据库的信息服务能力。

在之后的研究中,可以通过进一步分析视觉关键词与词汇树结构的内在联系,构建基于视觉关键词的自适应成长的词汇树系统,以增强系统的自适应性 and 可扩展性。并且,通过研究设计更加准确有效的结点增加与删除机制,从而提高系统运算速度。

References (参考文献)

- [1] Roberto Figueira Santos Filho, Agma Traina, Caetano Traina Jr., Christos Faloutsos. Similarity Search without Tears: the OMNI-Family of All-Purpose Access Methods[C].In Proceedings of the 17th International Conference on Data Engineering, Heidelberg, Germany, Apr.2001:623-630.
- [2] Guang-Ho Cha, Xiaoming Zhu, D Petkovic, C W Chung .An Efficient Indexing Method for Nearest Neighbor Searches in High-Dimensional Image Databases[J].IEEE Transactions on Multimedia, Mar.2002, 4 (1):76 - 87.
- [3] Manuel J. Fonscaeca, Joaquim A. Jorge. Indexing High-Dimensional Data for Content-Based Retrieval in Large Databases[C].In Proceedings of the 8th International Conference on Database Systems for Advanced Applications, Kyoto, Japan, Mar.2003:267-274.

- [4] Antonm Guttman. R-trees: A Dynamic Index Structure for Spatial Searching[C].In Proceedings of ACM SIGMOD International Conference on Management of Data, Boston, MA, USA, June.1984:47-57.
- [5] Timos Sellis, Nick Roussopoulos, Christos Faloutsos. The R+-tree: A dynamic Index for Multi-dimensional Objects[C].In Proceedings of the 13th International Conference on Very Large Databases, Brighton, GB, Sept.1987:507-518.
- [6] Norbert Beckman, Hans-peter Begel, Ralf Schneider. The R*-tree: an Efficient and Robust Access Method for Points and Rectangles[C].In Proceedings of ACM SIGMOD International Conference on Management of Data, Atlantic City, NJ, USA, May.1990: 322-331.
- [7] David A. White, R Jain. Similarity Indexing with the SS-tree[C].In Proceedings of the 12th International Conference on Data Engineering, New Orleans, USA, Mar.1996:516-523.
- [8] Stefan Berchtold, Daniel A. Keim, Hans-Peter Kriegel. The X-tree: An Index Structure for High-Dimensional Data[C].In Proceedings of the 22nd International Conference on Very Large Databases, Bombay, India, Sept.1996:28-39.
- [9] Tom Yeh, John Lee, Trevor Darrell. Adaptive Vocabulary Forests for Dynamic Indexing and Category Learning[C].In Proceedings of the 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, Oct.2007:1-8.
- [10] John Lee. Efficient Object Recognition and Image Retrieval for Large-scale Applications[R]. MIT Computer Science and Artificial Intelligence Laboratory, May.2008.
- [11] Svetlana Lazebnik, Cordelia Schmid, Jean Ponce. Beyond Bags of Features:Spatial Pyramid Matching for Recognizing Natural Scene Categories[C].IEEE International Conference on Computer Vision and Pattern Recognition, New York, USA, June. 2006:2169-2178.
- [12] David Chen, Sam S. Tsai, Vijay Chandrasekhar, Gabriel Takacs, Jatinder Singh, and Bernd Girod. Robust Image Retrieval Using Multiview Scalable Vocabulary Trees[C].In Proceedings of Visual Communications and Image Processing Conference, San Jose, California, Jan.2009:360-369.
- [13] David Nister, Henrik Stewenius. Scalable Recognition with a Vocabulary Tree[C].IEEE International Conference on Computer Vision and Pattern Recognition, New York, USA, June. 2006:2161-2168.
- [14] Kate Saenko, Trevor Darrell. Towards Adaptive Object Recognition for Situated Human-computer Interaction[C]. In Proceedings of the 2007 Workshop on Multimodal Interfaces in Semantic Interaction, New York, USA, 2007:43-46.