

基于哈希技术的 BoVW 图像检索

代雪梅^{1,2}, 李生刚¹

(1 陕西师范大学 数学与信息科学学院, 陕西 西安 710062; 2 青海师范大学 数学系, 青海 西宁 810008)

摘 要: 本文提出了一种基于哈希技术的 BoVW 图像检索方法. 选择合适的生成二进制哈希码的哈希算法, 将局部特征点保持相似性地映射为二进制哈希码. 再在二进制哈希码上进行 k-means, 生成视觉词为二进制码的视觉词典. 最后, 用视觉单词的词频向量表示图像内容, 并根据词频向量对图像进行检索. 实验结果表明, 该方法可以缩短视觉词典生成的时间, 占用更少的存储空间.

关键词: 二进制哈希码; 视觉词袋模型; 二进制视觉词典; 图像检索

中图分类号: TP391

文献标识码: A

文章编号: 1000—7180(2016)12—0011—04

Image Retrieval Based on BoVW and Hashing

DAI Xue-mei^{1,2}, LI Sheng-gang¹

(1 College of Mathematics and Information Science, Shaanxi Normal University, Xi'an 710062, China;

2 Department of Mathematics, Qinghai Normal University, Xining 810008, China)

Abstract: Currently, this paper propose an image retrieval method based on BoVW and hashing. Firstly, select an appropriate binary hashing algorithm, which map the similar local feature points into similar binary hash codes. Second, play k-means on the binary hash codes and generate the visual vocabulary with binary visual word. Finally, the image is represented with the word frequency vector of visual words, and image retrieval is achieved based on the word frequency vector. Experimental results on SIFT-1M and Caltech-256 data sets show that the proposed method can reduce the visual vocabulary generation time and take up less storage space.

Key words: binary hashing code; bag of visual word model; binary visual vocabulary; image retrieval

1 引言

随着大数据时代的到来, 互联网图像资源迅猛增长, 如何对大规模图像资源进行快速有效地检索以满足用户需求亟待解决.

针对于大规模数据的检索问题, 哈希技术被广泛用于计算机视觉、机器学习、信息检索等相关领域. 哈希技术^[1]将文档、图像、视频或者其它类型的数据编码为二进制短码, 同时保持原始数据的相似性. 由于二进制码 Hamming 距离计算的高效性和存储空间的优势, 在大规模数据集上基于二进制码可进行非常高效地查询和检索.

将局部特征与哈希技术相结合应用图像检索

中, 本文提出了一种基于哈希技术的 BoVW 图像检索方法.

2 相关工作

局部特征能够表征图像的底层视觉特征, 被大量用于图像内容分析中. 在基于局部特征的图像检索中, 视觉词袋法 (BoVW) 是主流的图像内容表示方法. 视觉词袋法的主要流程包括: (1) 局部特征提取, 一般提取 SIFT 或 SURF 特征; (2) 视觉词典生成, 采用聚类算法对局部特征进行聚类得出; (3) 量化, 将单幅图像的各局部特征量化为视觉单词; (4) 创建直方图, 根据视觉词典创建各图像视觉单词频向量. 得到的词频向量用于表示图像, 作为图像的特

征进行图像检索和识别等工作。

在上述的视觉词袋法的流程中,其关键在于第二步,视觉词典的生成。通常采用 k-means 聚类算法直接对图像的局部特征进行聚类,聚类得出的中心构成视觉词典。但在数据集规模较大时,聚类时间和内存消耗会急剧增加(其时间复杂度为 $o(nk)$, k 是聚类数目, n 为特征点数目)。

Yadong Mu 等^[2]提出了利用哈希技术生成随机化词典的思想,它的基本思想是基于随机映射产生随机视觉词典,并不进行聚类操作。本文的基本思想是用保持相似性的二进制哈希码来代替原空间的特征,进行聚类算法生成视觉词典,即将原始特征变换到 Hamming 空间中,并在 Hamming 空间中再做一次聚类操作,生成相应的视觉词典。这样既利用了二进制哈希码在计算速度和存储空间上的优越性,又找到了一种原特征点与二进制视觉词之间的变换方法。

在无监督的哈希方法中,本文选择 PCA-ITQ^[3]方法,它的基本思想是首先利用 PCAH 对数据进行降维,然后通过迭代学习一个旋转矩阵使得量化误差最小。在 PCA-ITQ 方法中同时考虑了二进制编码的紧致性和二进制量化误差的最小化。

3 基于哈希技术的 BoVW 图像检索

3.1 多索引哈希技术

多索引哈希技术^[4](Multi-Index Hashing, MIH)通过为二进制码建立索引结构,在汉明空间中可以快速地精确的 r 邻域搜索。它的基本思想是将维度为 q 的二进制哈希码 h 划分成 m 个不相交的子串 $h^{(1)}, \dots, h^{(m)}$, 分别为每个子集建立哈希结构。搜索与 h 的 Hamming 距离小于 r 的近邻域 g , 则 g 需要满足在 m 中的某个子串中, h 和 g 之间的 Hamming 距离不超过 r' , 其中 $r' = \lceil r/m \rceil$ 。将在高维 Hamming 空间中 r 邻域的搜索问题转化为了在较低维度空间中进行半径更小的 r' 邻域问题。在较低维空间中进行较小半径的邻域搜索问题,速度更快。利用 MIH 方法为二进制码 g 查找它的 k 近邻域算法总结为算法 1。

算法 1: 给定二进制码 g , 它的 k 近邻查找算法步骤为:

给定查询二进制哈希码 g 的查询子串 $\{g^{(i)}\}_{i=1}^m$, 初始化集合 $N_d = \emptyset$, $0 \leq d \leq q$, 初始化变量 $r' = 0$, $a = 0$, $r = 0$ 且变量需要满足约束 $r = mr' + a$ 。

do

在第 $a+1$ 个子哈希表中查找与 $g^{(a+1)}$ 距离为 r' 的子串, 并将满足条件的子串所属的二进制码串, 作

为 g 候选 K-NN, 将该二进制码加入到候选集中。对候选集中的二进制码, 计算它与 g 的 hamming 距离 d , 并根据其距离将相应的元素加入集合 N_d 中。

$a := a + 1$

If $a \geq m$ then

$a := 0$

$r' := r' + 1$

End if

$r := r + 1$

while($\sum_{d=0}^{r-1} |N_d| < k$)

当循环结束后, 查询图像 g 的 k 个 r 邻域查找完毕。

3.2 基于二进制哈希码的 k-means 算法

YunChao Gong 等^[5]提出了一种基于二进制哈希码的聚类算法, 用于大量图像的分类。它的主要流程为: 首先利用深度学习算法为每张图像提取特征, 然后利用保持相似性的哈希算法将高维的深度特征映射为二进制哈希码, 并用二进制编码作为图像的特征表示进行聚类算法。

基于二进制哈希码的 k-means 算法的目标函数为:

$$\min_{c_j} \sum_i^n \sum_j^k \|x_i - c_j\|_2^2$$

$$c_j \in \{-1, +1\} \quad (1)$$

该目标函数类似于 k-means 算法的目标函数, 只是聚类中心和样本均为离散的二进制值。该优化可以利用 EM 算法来求解, 首先需要将每个二进制哈希码分配到与之最近的聚类中心, 对于当前二进制编码的情形, 可以通过多索引哈希技术(MIH) 加快该处理过程; 然后, 更新二进制聚类中心, 可以通过求解问题(1) 的最优解得到其计算公式。

仅考虑某个聚类中心, 问题(1) 优化问题转化为:

$$\max_{c_j} \sum_i^p \|x_i - c_j\|_2^2 = \sum_i^p \|x_i\|_2^2 + \sum_i^p \|c_j\|_2^2 - \sum_i^p x_i c_j^T \quad (2)$$

由于上式中第一项和第二项均为常数值, 问题(2) 的最小化问题可以转化为等价的极大值问题:

$$\max_{c_j} \sum_i^p x_i c_j^T = \max_{c_j} (\sum_i^p x_i) c_j^T$$

$$c_j \in \{-1, +1\} \quad (3)$$

问题(3) 有解析形式的最优解, 且其最优解为:

$$c_{jk} = \text{sign}(m_{jk}) = \begin{cases} +1 & \text{if } m_{jk} \geq 0 \\ -1 & \text{if } m_{jk} < 0 \end{cases} \quad (4)$$

式中, $m_j = \sum_i^p x_i$ 表示属于聚类中心 c_j 的所有样本 x_i 的连和。

算法 2: 基于二进制哈希码的聚类算法(Bk-means)

给定二进制哈希码样本集 $X = \{x_1, \dots, x_n\}$ 和聚类个数

k 以及最大迭代次数 $\max\text{Iter}$.

从样本集 X 中随机选择 k 个值,构成聚类中心的初始值 $C = \{c_1, \dots, c_k\}$.

for $i = 1, \dots, \max\text{Iter}$

第一步:为每个样本选择与它最近的聚类中心

按照算法 1 的步骤为聚类中心集 C 建立多索引哈希表 M .

for $j = 1, \dots, n$

$\text{idx} = M.\text{Lookup1NN}(x_j)$.

end for

第二步:更新聚类中心 C

for $j = 1, \dots, k$

计算每个聚类中心的均值 m_j .

更新聚类中心 $c_j = \text{sign}(m_j)$.

end for

if $i > 1$

计算本次迭代产生的聚类中心 $C^{(i)}$ 与上次迭代产生的聚类中心 $C^{(i-1)}$ 之间的差别,若集合 $C^{(i)} = C^{(i-1)}$,则退出整个 for 循环,聚类算法结束.

end if

end for

3.3 基于二进制哈希码的 BoVW 图像检索

基于二进制哈希码的 BoVW 图像检索流程图

如图 1 所示.

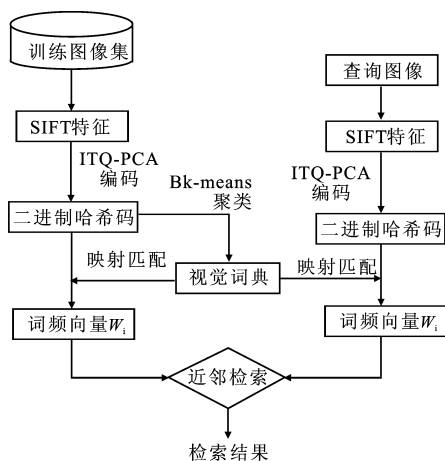


图 1 基于二进制哈希码的 BoVW 图像检索流程

基于二进制哈希码的 BoVW 图像检索分为视觉词典生成和图像检索两个阶段,与传统 BoVW 图像检索方法的区别在于,视觉词典为二进制哈希码以及局部特征点需要变换到汉明空间,转化为二进制哈希码.二进制视觉词典的生成过程,主要包括:

(1)提取图像集中各图像的 SIFT 特征;

(2)从 SIFT 特征点集中选择样本点,训练生成 PCA-ITQ 哈希方法所需参数;

(3)根据(2)得到的参数,利用 PCA-ITQ 算法将 SIFT 特征集转化为二进制哈希码;

(4)根据 3.2 节中基于二进制哈希码的聚类算

法,将二进制哈希码聚成 k 类(k 为事先指定的视觉词典的单词数目);

最后得到的 k 个聚类中心,构成了二进制视觉词典.

利用(2)中的参数将图像集各图像局部特征点转化为二进制哈希码,然后根据二进制视觉词典统计图像集各图像二进制哈希码的词频,生成图像集的词频向量组.

图像检索过程主要包括:提取查询图像的 SIFT 特征;

(1)利用 PCA-ITQ 哈希算法将 SIFT 特征点转化为二进制哈希码;

(2)根据视觉词典统计二进制哈希码的词频,生成查询图像的词频向量组;

(3)计算查询图像词频向量和图像集词频向量的相似度;

(4)进行相似度排序,确定相似图像.

4 实验设置与性能评价

4.1 实验设置

为验证本文方法的有效性,本文在 SIFT-1M^[6]、Caltech-256^[7] 图像集上对本文方法进行了评估. SIFT-1M 是包含了 1 000 000 个 128 维的局部特征点;Caltech-256 图像集是目标分类任务中常用数据集,包含 256 个类别共计 30 608 张图像,其中每个类别中至少包含 80 张图像.

实验硬件配置为双核 3.4 GHz 的 CPU、内存为 16 G 的单机.图像检索性能指标采用 MAP(Mean Average Precision),它表示所有图像的查准率的平均值.其中查准率定义如下:

$$\text{MAP} = \frac{\text{多次图像检索的平均查准率}}{\text{检索次数}} \times 100\%$$

(5)

4.2 实验性能分析

为验证基于二进制哈希码的聚类算法(Bk-means)的有效性,本文在 SIFT-1M 数据集上与传统的 k-means 算法进行比较.首先在 SIFT-1M 数据集中随机选出 10 000 个样本,训练生成 PCA-ITQ 算法所需参数;其次,利用 PCA-ITQ 算法将 SIFT-1M 数据集转化为 64 位的二进制哈希码;然后,利用 3.2 节中基于二进制哈希码的聚类算法进行聚类,主要测试它聚类所用的时间.在 SIFT-1M 数据集上测试在设置不同的聚类个数的情况下, Bk-means 算法和 k-means 算法的所用时间.实验结果

如表 1 所示.

表 1 在数据集 SIFT-1M 上聚类时间对比

	ITQ+Bk-means/s	k-means/s
K=300	909	1 572
K=500	860	2 676
K=800	700	4 145
K=1 000	888	5 253

从表 1 中可以看出,随着聚类个数 K 的增加, k -means 算法的运行时间也随之增加;当聚类个数 $K=1\ 000$ 时,其耗费时间大概是 Bk-means 算法的 6 倍. k -means 算法进行聚类时需要计算 SIFT 特征点与每个聚类中心的距离,时间复杂度为 $o(k)$,总的时间复杂度为 $o(nk)$;但是 Bk-means 算法利用 PCA-ITQ 算法将 SIFT 特征转化为二进制哈希码,并寻找 Hamming 距离最近的聚类中心,时间复杂度降为 $o(1)$,使得 Bk-means 算法的聚类时间并未随着聚类个数 K 的增加而增加,比较适合于大数据集的聚类个数较大的算法. 另外, Bk-means 算法在存储空间上也占有很大的优势,一个 SIFT-1M 中的样本点在原空间为 128 维,经过二进制转化为 64 位二进制值,即仅占 8 个字节,存储空间仅需要原有的十六分之一.

为验证基于二进制聚类算法在图像检索中的有效性,在 Caltech-256 图像集上进行了实验. 首先从 Caltech-256 图像集中选取 20 类,再从每个类别中随机挑选 60 个图像,其中 30 个作为训练图像集,30 个作为测试图像集. 其次,提取训练图像集的 SIFT 特征(600 张图像的特征),组成 SIFT 特征集. 然后,在 SIFT 特征集中随机选择出不同个数的特征点(在实验中选择了 100 000、200 000、300 000、400 000 共四组),分别作为生成视觉词典的聚类样本;对于 Bk-means 算法需要根据特征点训练得到 PCA-ITQ 所需的参数,将特征点转化为 64 位二进制哈希码后进行聚类生成视觉词典;对于 k -means 算法可以直接聚类生成视觉词典. 最后,图像集分别根据不同的视觉词典计算词频,生成相应的词频向量组,并将测试图像集作为查询图像,训练图像集作为数据集进行相似性检索. 在特征点个数 N 不同的情况下,测试基于 Bk-means 的视觉词典生成算法和基于 k -means 的视觉词典生成算法,在视觉词典生成时间和检索结果 MAP 上作比较. 实验结果如表 2 所示.

表 2 不同特征点个数的图像检索 MAP 值和视觉词典生成时间对比

	N	MAP/%	Time/s
k-means	100 000	60.77	405
ITQ+Bk-means	100 000	61.53	236
k-means	200 000	61.00	812
ITQ+Bk-means	200 000	62.66	346
k-means	300 000	62.06	1 035
ITQ+Bk-means	300 000	63.07	490
k-means	400 000	62.72	1 430
ITQ+Bk-means	400 000	64.42	682

从表 2 中可以看出,在不同的特征点 N 的情况下,基于传统 k -means 算法生成视觉词典所用时间是基于二进制哈希码生成视觉词典所用时间的 2 倍,且随着 N 的增加,基于传统 k -means 的视觉生成算法所用时间增长速度大于基于 Bk-means 的视觉词典生成算法. 另外,基于 Bk-means 的图像检索结果要优于传统 k -means 的检索结果的 MAP 相差不大. 因此,先将数据变换为保持相似性的二进制码,再利用二进制码进行 k -means 聚类生成视觉词典比在原数据空间中进行 k -means 生成视觉词典的算法聚类所用时间更短,所用存储空间更少,但是根据这两种方法生成的视觉词典对图像进行特征表示,其检索性能基本一致.

5 结束语

本文提出了一种基于哈希技术的 BoVW 图像检索方法. 首先,为了充分利用二进制哈希码在存储和计算速度上的优越性,选择合适哈希算法将原始空间中的局部特征点保持相似性地映射为二进制哈希码. 然后,针对传统 k -means 算法聚类生成视觉词典时间较长,在二进制哈希码上进行 k -means,生成视觉词为二进制码的视觉词典. 最后,图像用二进制视觉单词的词频向量表示,并根据词频向量对图像进行检索,实验结果有效地验证了本文方法的图像检索精度损失不大,但是在运行时间和存储空间上有较大的优势.

参考文献:

[1] Datar M, Immorlica N, Indyk P, et al. Locality sensitive hashing scheme based on p-stable distributions [C]// Proceedings of the ACM Symposium on Computational Geometry. New York, USA, 2004.

的控制问题,算法通过最小生成树聚类规则来确定合适的簇头数量,节省簇内各节点的传输能量。仿真实验结果表明,本文算法可有效提高无线传感网络能量利用效率,进一步延长了无线网络寿命。

参考文献:

- [1] 谢和平,周海鹰,左德承,等. 无线传感器网络能量优化与建模技术综述[J]. 计算机科学, 2012, 39(10): 15-16.
- [2] Hancke G P, Gungor V C. Guest editorial special section on industrial wireless sensor networks[J]. IEEE Transactions on Industrial Informatics, 2014, 10(1): 762-765.
- [3] 付帅, 马建峰, 李洪涛. 改进的基于分簇无线传感器网络的数据聚合算法[J]. 吉林大学学报:工学版, 2014, 44(4): 1118-1125.
- [4] Kuila P, Jana P K. A novel differential evolution based clustering algorithm for wireless sensor networks[J]. Applied Soft Computing, 2014, 25(c): 414-425.
- [5] AlShawi I S, Yan L, Pan W, et al. Fuzzy chessboard clustering and artificial bee colony routing method for energy-efficient heterogeneous wireless sensor net-

works[J]. International Journal of Communication Systems, 2013, 27(12):3581-3599.

- [6] Mehmood A, Khan S, Shams B, et al. Energy-efficient multi-level and distance-aware clustering mechanism for WSNs[J]. International Journal of Communication Systems, 2013, 28(17): 973-989.
- [7] Yu C B, Yu L, Tan J, et al. DV-Hop localization algorithm in WSN based on weighted of correction in hop distance[J]. Applied Mechanics and Materials, 2013, 32(6): 143-148.
- [8] 彭铎,黎锁平,杨喜娟. 一种能量高效的无线传感网络非均匀分簇路由协议[J]. 传感技术学报, 2014, 27(12):1687-1691.

作者简介:

李凯佳 女, (1979-), 硕士, 讲师. 研究方向为无线网络信息安全及路由协议、数字图像处理.

E-mail: likaijia1979@126.com.

袁凌云 女, (1980-), 博士, 副教授. 研究方向为物联网及无线网路由协议.

俞锐刚 男, (1979-), 博士研究生, 讲师. 研究方向为无线网路.

(上接第 14 页)

- [2] Mu Y D, Sun U, Han T X. Randomized locality sensitive vocabularies for bag-of-features model[C]// Proceedings of the 11th European conference on computer vision conference on Computer vision. Germany, Berlin, Heidelberg, 2010.
- [3] Gong Y, Lazebnik S, Gordo A, et al. Iterative quantization: A Procrustean approach to learning binary codes for large-scale image retrieval[J]. IEEE Transactions on Software Engineering, 2013, 35(12): 2916-2945.
- [4] Norouzi M, Punjani A, Fleet D J. Fast search in hamming space with multi-index hashing [C]// CVPR 2012. USA, Rhode Island, Providence, 2012.
- [5] Gong Y, Pavlowski M, Yang Fei. Web scale photo hash clustering on a single machine[C]// Proc. CVPR 2015. USA, Boston, 2015.
- [6] Lowe D. Distinctive image features from scale-invari-

ant keypoints[J]. Int. J. Comput. Vis., 2004, 60(2): 91-110.

- [7] Lin W, Wang J, Ji R, et al. Supervised hashing with kernels[C]//Proc. CVPR 2012. USA, Rhode Island, Providence, 2012.
- [8] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories[J]. Computer Vision and Image Understanding, 2007, 106(1): 59-70.

作者简介:

代雪梅 女, (1972-), 博士研究生, 副教授. 研究方向为算法设计、格上拓扑学与模糊代数.

E-mail: daixuemei666@tom.com.

李生刚 男, (1959-), 博士, 教授. 研究方向为格上拓扑学和拟阵理论.