

基于多探寻局部敏感哈希和单词映射链投票 的图像检索方法

许 喆 陈福才 李邵梅 李 星

(国家数字交换系统工程技术研究中心 郑州 450002)

摘 要 为解决基于欧氏局部敏感哈希(E2LSH)的视觉词典法存在的内存消耗大、在图像背景明显变化时检索精度不高及增大数据库规模导致检索效率降低的问题,在采用多探寻 LSH 对特征点进行聚类的基础上提出的基于嵌入汉明码的单词映射链投票的图像检索方法。该方法首先采用多单词映射和软量化思想构造单表视觉词典,缩小词典规模以降低内存消耗;然后通过嵌入汉明码生成单词映射链,并提出一种权重赋予函数来增加检索精度;最后对匹配返回的单词映射链进行加权投票完成图像检索。实验结果表明,该方法能有效降低检索的内存消耗,提高检索精度,且适用于大规模数据库条件下的检索处理。

关键词 图像检索,多探寻局部敏感哈希,嵌入汉明码,单词映射链,视觉词典

中图分类号 TP391 **文献标识码** A

Image Retrieval Based on Multi-probe Locality Sensitive Hashing and Word Map Chain Voting

XU Zhe CHEN Fu-cai LI Shao-mei LI Xing

(National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China)

Abstract To solve the problem of high memory cost, low retrieval accuracy when the background is changed obviously and reduced efficiency when the size of the database is increased of the BoVW (bag of visual words) method based on Euclidean locality sensitive hashing (E2LSH), a fast retrieval method based on word map chain voting of Hamming embedding was presented on the basis of clustering the feature points through multi-probe locality sensitive hashing. The method constructs a single-table visual dictionary with multi-word mapping and soft assignment, reduces the size of the dictionary to reduce memory consumption. Then a word map chain is constructed with Hamming embedding, and a weighting function is proposed to increase the retrieval accuracy. A weighted voting of the word map chain from matching results is made to accomplish image retrieval. Experimental results show that this method can effectively reduce memory consumption, improve retrieval accuracy, and is applicable to large scale datasets.

Keywords Image retrieval, Multi-probe locality sensitive hashing, Hamming embedding, Word map chain, Bag of visual words

1 引言

数字多媒体和网络技术的发展使得图像的规模越来越大,近年来对图像进行处理如剪切、旋转等技术变得利于人们使用,从而形成了大量的相似性图像,给图像内容的相似性检索带来巨大挑战。图像的视觉语义特别丰富,一般由高维特征进行表示,如何从这些大规模高维的图像表示特征中查询到相似特征以达到相似图像检索的目的是目前亟待解决的问题。图像特征中的 SIFT^[1]特征及其多种改进对旋转、缩放等变化能保持不变,近年来被广泛用来进行图像和视频的特征表示,其中以 SIFT 特征为基础的视觉词典法(bag of visual words)^[2]由于性能突出而日益受到研究者重视,成为大规模数据集下相似图像检索的主流方法。

视觉词典法将图像信息映射成视觉单词集合,既保存了图像的局部特征又有效地压缩了图像的描述,可以将相对成熟的文本检索技术应用于图像检索之中。虽然视觉词典法以其性能优势近年来得到广泛研究,但它还存在一些开放性问题:一是量化过程中存在严重的信息损失。文献[3]提出汉明嵌入的思想,通过嵌入汉明码的不同位数赋以不同权重,实质是向同一单词内引入距离信息来减少量化信息损失;文献[4]提出了软量化的思想,通过将特征点映射到多个近邻单词,进一步优化量化效果。但这些方法均采用 K-means 算法,由于其本身存在的码本生成不均匀和同义性、歧义性问题而很难取得突破性的效果。二是算法的时间效率与内存消耗的矛盾问题。文献[5]提出的层次 K-means (hierarchical K-means, HKM)算法加快了特征点向视觉单词的分配速度;文献[6, 7]

到稿日期:2013-07-30 返修日期:2013-11-13 本文受国家 863 计划项目(2011AA010603)资助。

许 喆(1989—),男,硕士生,主要研究方向为通信与信息系统,E-mail:806916784@qq.com;陈福才(1974—),男,研究员,硕士生导师,主要研究方向为通信与信息系统;李邵梅(1982—),女,博士,讲师,主要研究方向为通信与信息系统;李 星(1987—),男,硕士,讲师,主要研究方向为通信与信息系统。

利用近似 K-means (approximate K-means, AKM) 算法进一步提高了检索效率; 文献[8]通过引入 LSH (Locality Sensitive Hashing)^[9] 采用多个查询表提高了检索精度, 并使视觉词典的建立更加高效。但这些方法的内存消耗问题仍然没有得到有效解决。三是干扰图像和背景变化图像的查询精度问题。文献[10]通过建模特征点周围的结构提高了性能, 还有一些如文献[11]通过几何验证等后处理技术, 文献[12]通过生成单词词组的方式来改善检索精度。但这些方法都由于引入大量其他信息而降低了效率。

本文拟针对上述问题进行研究, 采用多探寻局部敏感哈希代替 K-means 作为聚类映射方法, 通过引入软量化思想, 使用多探寻的查询策略降低了内存使用量; 其次采用嵌入汉明码的方法对映射到同一单词中的特征点进行标记, 减少了量化过程中的信息损失, 提高检索精度且大大缩小了词典规模^[3]。最后在倒排索引思想的启发下, 对大规模的词典从数据结构存储的方式进行改进, 并使用二进制投票策略代替直方图比对进行相似性检索, 以增强鲁棒性及检索效率。

2 局部敏感哈希基本原理和多探寻实现方案

2.1 局部敏感哈希算法基本原理及参数分析

LSH 算法的基本思想就是用随机的哈希函数值保证相似的数据点以很高的概率发生冲突而能够被检测到。最初的 LSH 应用在汉明空间^[8], 2004 年, 文献[13]将其扩展到欧氏空间, 从此在欧氏空间中直接使用 LSH 成为可能, 这极大地方便了将其扩展到更多的应用之中。欧氏局部敏感哈希 (E2LSH) 继承了原始 LSH 的两个特性: 第一是它很适合于维数高但稀疏的数据点, 如图像特征; 第二是如果数据满足一定的有界增长特性, 它就可以很快地找到精确的近邻。

E2LSH 是基于 p -稳定分布的, 它使用哈希函数将数据库中的矢量点投影到随机的方向 a_i 矢量上, a_i 的每个元素均服从 p -稳定分布 (当 $p=1$ 时是柯西分布; $p=2$ 时是高斯分布)。具有如下性质: 若两个变量都服从 p -稳定分布, 则其线性组合也服从 p -稳定分布。

欧氏局部敏感哈希函数具有如下形式:

$$h_i(v) = \left\lfloor \frac{a_i \cdot v + b}{W} \right\rfloor \quad (1)$$

式中, a_i 与特征点 v 维数相同, b 的取值服从 $[0, w]$ 之间的均匀分布, 目的在于减弱用 W 进行划分时的边界影响, 增强了哈希函数的随机性, W 是投影时的量化宽度。E2LSH 通常会将 k 个哈希函数 ($H = \{h_i\}, \{i=1, \dots, k\}$) 联合使用, 目的是拉大距离近的点与距离远的点分别哈希到相同桶中的概率之差。经哈希函数降维映射后, 再利用主哈希函数和次哈希函数对降维后的向量进行哈希, 主次哈希值均相同的特征点存储在相同的哈希桶中。

鉴于 LSH 将近邻点哈希到同一桶中的聚类特性, 本文拟使用 LSH 代替经典视觉词典法中的 K-means 及其改进算法对图像库中的特征点进行聚类。

表示图像的两个特征点哈希到同一个桶中需要满足:

$$p \left[\left\lfloor \frac{(a \cdot v) + b}{W} \right\rfloor = \left\lfloor \frac{(a \cdot q) + b}{W} \right\rfloor \right] \quad (2)$$

即要求 $|(a \cdot v) - (a \cdot q)| < W$, 并且 $(a \cdot v)$ 和 $(a \cdot q)$ 的值要

落在同一个分段内。可以看出增大量化宽度 W 的值可以提高特征点哈希到同一个桶中的概率, 相应地减少桶的数量, W 也就是本文聚类过程中的聚类尺度。经典局部敏感哈希算法在进行近似近邻查找的过程中建议 W 值设为 4, 以达到将非常相近 (如相差 1) 的相似向量返回的目的。由于图像中特征值千差万别, 在进行聚类压缩的过程中, 同一个桶中的相似特征点不可能控制得非常相近, 否则就会导致类内特征数量过少而单词数量过多从而失去视觉词典法的聚类压缩优势; 而将 W 值设得过大就会导致类内的特征点数量巨大, 它们之间的距离也可能相应变大从而失去聚类的意义, 类内计算返回相似特征点的时间变长, 导致检索结果变差。这样就需要通过训练找到一个合适的 W , 以达到词典规模适中以及查询精确度和效率最优的目的。

2.2 多探寻 LSH 算法描述

经典局部敏感哈希 (E2LSH) 的一个缺点是需要用到很多的表 (L) 来消除随机性, 这样就导致占用大量的内存。文献[14]中提出了多探寻的 LSH 方案, 其通过扩大每个哈希表上的查询范围来减少内存和时间开销, 达到与原有方案相近的查询质量, 大大减少了哈希表的使用数量。

多探寻 LSH 的基本思想是按照一定的探寻策略试探所有可能包含查询点的临近点的哈希桶。根据上节提到的局部敏感哈希函数的性质, 如果一个特征点跟查询点 q 距离近但没有哈希到同一个桶中, 那么它很可能被哈希到查询点 q “附近”的桶中, 这样对 q 进行查找时就要通过查找这些“附近”的桶来增加找到 q 的相似点的机会。

定义哈希置乱向量 (perturbation vector) 为 $\Delta = (\delta_1, \delta_2, \dots, \delta_k)$, 对于查询点 q , 经典 LSH 方法在查找时只需检查 $g(q) = (h_1(q), h_2(q), \dots, h_k(q))$ 对应的哈希桶, 在使用置乱向量之后, 我们将探寻 $g(q) + \Delta$ 所在的桶。如果 LSH 函数中的参数 W 合理地大, 那么相似的特征点就会以较高的概率被哈希到相同或邻近的值, 在实际中所有相似点的哈希值与查询点的哈希值要么相同要么相差 $+1$ 或 -1 , 且不同的哈希值数几乎都在 2 以内^[14], 因此我们把置乱向量 Δ 限制为 $\delta_k \in \{-1, 0, 1\}$ 。本文将多探寻 LSH 策略引入到视觉词典的软量化过程中, 查询时通过将特征点映射到多个单词上, 提高了检索精度并极大地减小了所需表数量 L 的值。

3 基于哈希聚类和单词映射链投票的图像检索

针对传统基于 LSH 视觉词典组的多表查询方法内存消耗大及背景明显变化时检索精度不高的问题, 本文将多探寻 LSH 作为聚类映射算法进行研究, 训练及检索流程如图 1 所示。首先, 通过对训练图像库的特征点进行聚类学习, 找到一个合适的量化宽度 W , 并将哈希过程中的 L 个表融合生成单表视觉词典; 其次采用嵌入汉明码的方法对同一类中的特征点进行标记, 生成单词映射链, 匹配过程通过比较两个特征向量对应汉明码的不同位数, 赋予不同权重; 然后针对 LSH 算法进行聚类的过程具有随机性的问题, 引入软量化思想, 采用多探寻的方式将查询特征点映射到多个单词上, 减少特征点映射到视觉单词的量化损失; 最后通过查询单词的投票策略优化检索性能, 得到最后结果。

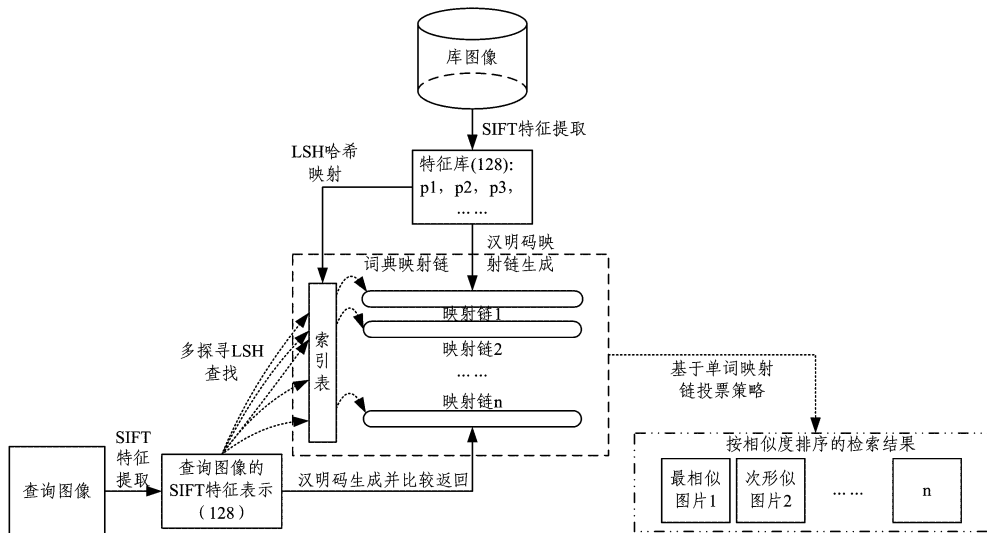


图1 基于哈希聚类 and 单词映射链投票的图像检索流程图

3.1 基于哈希聚类的单表视觉词典生成

针对 K-means 算法量化生成单词过程产生的同义性和歧义性以及采用欧氏 LSH 导致内存消耗大的问题,本文采用多探寻 LSH 作为聚类映射方法,通过引入软量化思想的多探寻查找方式极大地减小所需哈希表 L 的值;其次采用嵌入汉明码的方式进一步缩小了词典规模并提高了检索精度;最后将 L 个哈希表融合生成单表视觉词典。其详细生成过程如下:

(1)特征库生成。为保证实验说服力并方便对比分析,本文仍以图像检索常用的数据库 Oxford5K^[18] 作为图像库,提取图像库中所有图片的 SIFT 特征生成特征库,其中每一个特征都是一个 128 维的高维向量。

(2)视觉单词(即索引值)生成。使用哈希函数族中的多个函数 $g_i = \{H_i\} (i=1, 2, \dots, L)$, 对图像库中的 SIFT 特征点进行哈希映射,每次生成 L 个包含主次哈希值的索引值(本文实验中确定 $L=5$),即能够将每一个特征点映射到不同的几个单词上。

(3)视觉单词(索引值)存储并排序。存储过程为在每次得到视觉单词之后检测所得出的单词是否存在,若存在则直接更新桶映射链,若不存在则存储新的单词,然后更新桶映射链,最后根据哈希值的大小对单词进行排序。

(4)视觉词典生成。将所有 SIFT 特征点进行哈希并存储完成后,去除含数据点太多或太少的视觉单词,最后生成一个单表的视觉词典。

3.2 嵌入汉明码的桶映射链生成

随着图像检索库规模的增大,特征数量迅速增加,如本文采用的 Oxford5K^[18] 图像库提取出的特征就在千万条级别。文献[3]采用嵌入汉明码的方法对图像进行索引,取得了较好的检索精度,但内存消耗较大,难以支撑大规模的图像数据;文献[15]在高维向量检索中提出基于桶映射链的索引存储方法,极大地减少了内存消耗。本文结合这两种思路,提出了嵌入汉明码的桶映射链索引方法,通过嵌入汉明码为同一个单词内的特征点加入距离信息,有效提高了检测精度^[17];并将单词(索引值)存为内存文件而将映射链存为硬盘文件,对图像数据规模的增大有良好的适应性。它对应于单词映射链的

数据结构如图 2 所示。

上述索引数据结构的生成过程如下:

(1)对于图像库中的每幅图像 I ,检测出其所有的 SIFT 特征点,并对其进行哈希映射,找到对应的视觉词;

(2)每个特征点在找到对应的视觉词后,通过与一个 $d \times t$ 的随机矩阵 X 相乘,得到向量 q ,所有特征点计算完成后生成过渡向量集合 Q ,其中 d 为特征点维数, t 为汉明码长度,LEAR 实验室在文献[3]中建议生成的汉明码为 64 位,即占用 8 个字节;

(3)计算向量集合 Q 中每一维的所有值的平均值 m ;将 Q 中的每一个向量 q 与 m 相比较,生成长度为 t 的汉明码;

(4)将汉明码与对应的图片号索引值按图 2 中的结构进行存储。

图像的查询过程类似,得出查询图像的汉明码之后,通过与每个单词链中所含特征点的汉明码相比较,按位数的不同个数进行排序返回,不同位数少的优先返回,赋予较大权重,不同位数多的赋予较小权重,具体的权重赋予策略将在 3.3 节中详细讨论。

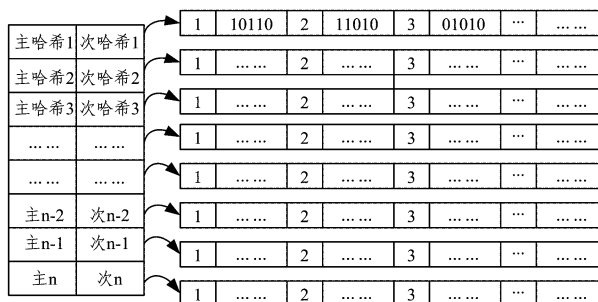


图2 单词(哈希值)映射链结构图

3.3 基于单词映射链投票的图像检索

在使用倒排索引进行文本检索时通过词语或者短语的输入来检索目标文档,单词在文档中出现的频率越高说明该文档关联性越强,但在视觉词典检索中,单词只是统计生成的,并不具有实际意义。单词出现的数量多并不能代表与查询图像的关联性强,在这样一个背景下,用单词的词频向量来代表图片是否具有足够的高层语义支持是计算机视觉领域遗留的兴趣点之一。

其次在图片的相似性查询过程中,对于部分场景重合或背景变化较大的图片,由于其 SIFT 特征总数发生了较大变化,各部分所占的比例也发生较大变化,导致表示图片的直方图向量随之变化。由直方图代表图片进行查询,在一定意义上变成了全局特征查询,即具有图片局部发生变化导致全局特征变化的缺点,导致索引查询的鲁棒性变差。

再次,图像的语义特征丰富,提取出的高维表示特征数量庞大,随着图像检索库规模的增大,特征数量可能达到上亿条,为了取得较好的区分性,单词的规模也随之上升。此时再用所有视觉词组成的词频向量来检索会导致检索效率急剧下降,且内存消耗迅速增长,最终严重影响检索性能。本文提出基于单词映射链投票的方法进行图像检索。

在部分场景相似图片的相似性匹配过程中,虽然背景有一定变化,但相似部分的视觉单词的匹配映射基本不受影响,若按相似单词的个数所占比例赋予一定的权重进行投票,能够得出比直方图匹配更加鲁棒的检索结果。

本文提出一种基于汉明码嵌入的权重赋予策略,按照距离近的点权重大的原则进行计算。

对于视觉词典中的任意单词 $Word$, 定义函数:

$$k f_i(Word) = \frac{\lg(N/n)}{2^r} \quad (3)$$

其中, N 为训练图像库的图像总数, n 为图像库中出现单词 $Word$ 的图像数目, 分子的值由 tf-idf 算法的思想得出, r 为图片 i 在单词映射链中的排列序号。由上述权重赋予函数可以看出, 在单词映射链中排序特别靠后的权重值接近零, 为简化计算, 每个单词只返回前 5 个图片的索引值, 如果待返回的图片索引出现重复, 只保留排序靠前的权重值, 其余舍弃。

图片相似度计算模块由相似的单词进行投票得出, 采用软量化思想, 输入为与查询图片中特征点哈希聚类到相同单词及其邻近的上下两个单词中的特征点所对应的图片集合, 输出为与查询图片相似的图片列表。图片 v_i 和 v_j 的相似度^[16]采用式(4)计算:

$$\text{sim}(v_i, v_j) = \left(\frac{\sum(k f_i \cap k f_j)}{k n_i} + \frac{\sum(k f_i \cap k f_j)}{k n_j} \right) / 2 \quad (4)$$

式中, $k f_i \cap k f_j$ 表示哈希聚类到相同单词的权重值, $\sum(k f_i \cap k f_j)$ 表示两个图片所含相同单词所对应的权重之和, $k n_i$ 和 $k n_j$ 分别表示两个图片的特征点个数。 $\text{sim}(v_i, v_j)$ 测量两个图片用各自特征点个数归一化之后的相似特征点所对应单词的权重之和, 这个值越高表示越相似。根据返回的相似图片集合, 对按照式(4)计算出的相似度进行排序后将大于一定阈值的图片作为相似图片输出。

4 实验结果与性能分析

本文实验数据依然采用牛津大学为图像检索及分类提供的 Oxford5K^[18] 数据库, 其共有 5062 幅图像。其中包含 55 幅标准的查询图像, 每个目标选取 5 幅图像, 涵盖了牛津大学的 11 处标志性建筑, 提取约 16334970 个 SIFT 特征点, 采用不同的 W, L 值来进行聚类检索实验。实验硬件配置为 Core 2. 6GHz×2, 内存 2G 的服务器。

性能评价指标为平均查准率均值 (Mean Average Precision, MAP), 它是图像查全率-查准率曲线所包含面积的均值。其中, N^c 表示查询数据集中相似图像的个数, N_r 表示检测出的相似图像个数, N_c 表示正确检测出的相似图像的个

数, 则查准率 P 和查全率 R 分别为:

$$P = \frac{N_c}{N_r}, R = \frac{N_c}{N^c} \quad (5)$$

在本文方法的实现过程中, W 即聚类尺度取值越大, 则相似特征点哈希聚类到相同单词的概率越大, 单词数量变少, 但是单词映射链中的特征点个数增多, 查询返回的时间变长, 效率降低; 而当桶的个数增多时, 索引内存占用变大。图 3 为 L 值为 1 时哈希桶个数与 W 取值之间的关系。综合考虑算法的精度、效率和占用的内存大小, 本文取 $W=64$, k 的取值沿用文献[8]中对参数 k 的设置, 即 $k=10$ 。

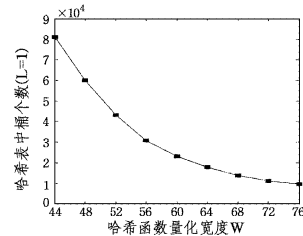


图3 参数 W 与哈希桶个数的关系

由本文单表词典和文献[8]中的多表词典生成过程对比来看, 文献[8]中索引单词的数目随着表数量的增多呈线性增长的趋势, 而为了维持较高的检索精度需要采用的表数量较多。本文采用基于多探测 LSH 的方法大大减少了 L 的数量, 并且对 L 个哈希表进行有效的压缩融合生成单表视觉词典, 主次哈希相同的单词不用重复计数, 进一步减少了索引单词的数量。图 4 反映了在 L 取值不同的情况下词典生成过程中单表和 multis 单词总数的对比关系, 图 5 反映了检索精度随 L 数目的变化情况。由图 5 可知, 本文方法只需较小的 L 值就能得到比较高的检索精度, 在实验时取 $L=5$, 大大减少了单词总量, 对应的索引内存使用量大大减少, 在数据库规模增大时有一定的扩展性, 能够更好地适应大规模数据情况下的检索。

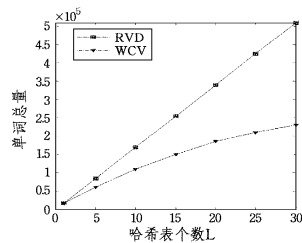


图4 单表与多表哈希桶个数的对比关系

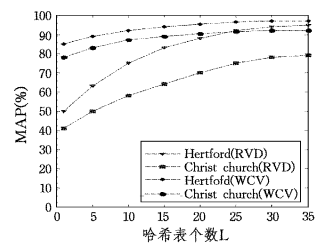


图5 检索精度对比关系

将本文方法 (Word Chain-Exact Vote, WCV) 与经典的基于 AKM 和软分配方法 (AKM+SA) 以及文献[8]中所用的随机化视觉词典组 (RVD) 方法在 Oxford 数据库中对部分图片的检索精度进行比较, 得出在 L 取值较小的情况下本文方法的有效性, 得到 MAP 值, 如表 1 所列。由表 1 结果可得, 在 L 取值为 5 的情况下 RVD 方法的平均查准率均值远低于本文的 WCV 方法, 甚至低于传统的 AKM+SA 方法, 说明在表数量较少的情况下, RVD 的检索精度大大降低; 而由于本文采用了基于嵌入汉明码的权重赋予与单词投票策略, 使得在 RVD 中表数量 L 升至 25 时本文方法与其比较性能依然有所提高。

(下转第 106 页)

Networks[C] // INFOCOM IEEE Conference on Computer Communications Workshops, 2010. San Diego, CA, USA, 2010; 1-6

- [15] Piamrat K, Ksentini A, Viho C, et al. QoE-Aware Admission Control for Multimedia Applications in IEEE 802. 11 Wireless Networks[C]// Vehicular Technology Conference, 2008. VTC 2008 Fall. IEEE 68th. Calgary, BC, Canada, 2008; 1-5
- [16] Bergstra J A, Middelburg C A. ITU-T Recommendation G. 107;

The E-model, A Computational Model for Use in Transmission Planning[S]. 2005

- [17] 胡治国, 张大陆, 张俊生. 一种 VoIP 语音质量评价模型[J]. 计算机科学, 2011, 38(5): 49-53
- [18] Lygizou A, Xergias S, Passas N, et al. rtPS Scheduling with QoE metrics in Joint WiMAX/Satellite Networks[C] // 4th International Conference on Personal Satellite Services (PSATS). Bradford, UK, 2012; 1-8

(上接第 85 页)

表 1 本文方法在 Oxford 数据库下的平均查准率均值

查询目标	AKM+SA	RVD(L=5)	RVD(L=25)	WCV(L=5)
Ashmolean	63.1	61.5	74.2	81.4
Bodleian	54.6	44.7	57.8	70.9
Balliol	59.1	38.1	67.4	74.2
Hertford	77.4	74.8	89.2	90.4
Keble	75.3	62.4	87.1	86.3
平均	65.90	56.32	75.14	80.64

最后, 由于本文采用单词投票策略, 每幅图片在检索时需要查询的单词个数不变, 单词映射链中索引图片数量的增加与整体数据库的增长相比变化较小, 因此本文所用方法在数据库规模增大时, 依然能够有效地进行检索。

结束语 本文采用 E2LSH 作为聚类算法并引入软量化思想生成单表视觉词典, 在此基础上采用嵌入汉明码的方法生成单词映射链, 并提出一种权重赋予函数, 此外引入投票策略进一步优化检索性能。实验结果表明本文方法降低了内存使用量, 提高了检索精度, 且在数据规模大的情况下仍有较强的适用性; 但要指出的是, 从硬盘中读取索引文件需要一定的响应时间, 并不能有效地利用目标区域的上下文信息, 因此, 研究其他更高效的图像检索机制仍十分必要。

参 考 文 献

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [2] Jurie F, Triggs B. Creating efficient codebooks for visual recognition[C]// Proceedings of International Conference on Computer Vision. Beijing, 2005; 604-610
- [3] Jegou H, Douze M, Schmid C. Hamming embedding and weak geometric consistency for large scale image search[C]// IEEE Conference on European Conference on Computer Vision. Heidelberg, 2008; 1-29
- [4] Philbin J, Chum O, Isard M, et al. Lost in quantization; improving particular object retrieval in large scale image databases[C]// IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, 2009; 278-286
- [5] Nister D, Stewenius H. Scalable recognition with a vocabulary tree [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York, 2006; 2161-2168
- [6] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Min-

neapolis, 2007; 1-8

- [7] Cao Yang, Wang Chang-hu, Li Zhi-wei, et al. Spatial-bag-of-features[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010; 3352-3359
- [8] Zhao Yong-we, Li Bi-cheng, Peng Tian-qiang, et al. An Object Retrieval Method Based on Randomized Visual Dictionaries and Query Expansion [J]. Journal of Electronics & Information Technology, 2012, 34(5): 1154-1161
- [9] Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via hashing[C]// Proceedings of the 25th International Conference on Very Large Data Bases (VLDB). Edinburgh, Scotland, 1999; 518-529
- [10] Zhang Bo-lin, Wang Chun-heng, Xiao Bai-hua, et al. Image Representation Using Bag-of-phrases[J]. Acta Automatica Sinica, 2012, 38(1): 46-54
- [11] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, 2007; 1-8
- [12] Sadeghi M A, Farhadi A. Recognition Using Visual Phrases [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, 2011; 20-25
- [13] Datar M, Immorlica N, Indyk P, et al. Locality-sensitive hashing scheme based on p-stable distributions[C]// Proceedings of the Twentieth Annual Symposium on Computational Geometry. Brooklyn, New York, 2004; 253-262
- [14] Lv Q, Josephson W, Wang Z, et al. Multi-probe LSH: efficient indexing for high-dimensional similarity search [C] // Proceedings of the 33rd international conference on Very large data bases, VLDB'07. VLDB Endowment, 2007; 950-961
- [15] Gao Hao-lin, Peng Tian-qiang, Li Bi-cheng, et al. A Fast Retrieval Method Based on Frequent Items Voting of Multi Table and Bucket Map Chain[J]. Journal of Electronics & Information Technology, 2012, 34(11): 2574-2581
- [16] Wu X, Hauptmann A G, Ngo C W. Practical elimination of near-duplicates from web video search[C]// Proceedings of the 15th International Conference on Multimedia. New York, USA; ACM Press, Sep 2007; 218-227
- [17] Jegou H, Douze M, Schmid C. Improving bag-of-features for large scale image search[J]. International Journal of Computer Vision, 2010, 87(3): 316-336
- [18] Robotics research group. Oxford dataset[OL]. http://www.robots.ox.ac.uk/_vgg/data/oxbuildings, 2011-04-16