一种基于视觉词典优化和查询扩展的图像检索方法

柯圣财[1,2] 李弼程[3] 陈 刚[1] 赵永威[1] 魏 晗[1]

摘 要 视觉词典方法 (Bag of visual words, BoVW) 是当前图像检索领域的主流方法, 然而, 传统的视觉词典方法存在计算量大、词典区分性不强以及抗干扰能力差等问题, 难以适应大数据环境. 针对这些问题, 本文提出了一种基于视觉词典优化和查询扩展的图像检索方法. 首先, 利用基于密度的聚类方法对 SIFT 特征进行聚类生成视觉词典, 提高视觉词典的生成效率和质量; 然后, 通过卡方模型分析视觉单词与图像目标的相关性, 去除不包含目标信息的视觉单词, 增强视觉词典的分辨能力; 最后, 采用基于图结构的查询扩展方法对初始检索结果进行重排序. 在 Oxford5K 和 Paris6K 图像集上的实验结果表明, 新方法在一定程度上提高了视觉词典的质量和语义分辨能力, 性能优于当前主流方法.

关键词 视觉词典模型, 密度聚类, 卡方模型, 查询扩展

引用格式 柯圣财, 李弼程, 陈刚, 赵永威, 魏晗. 一种基于视觉词典优化和查询扩展的图像检索方法. 自动化学报, 2018, **44**(1): 99−105

DOI 10.16383/j.aas.2018.c160041

Image Retrieval with Enhanced Visual Dictionary and Query Expansion

KE Sheng-Cai[1,2] LI Bi-Cheng[3] CHEN Gang[1] ZHAO Yong-Wei[1] WEI Han[1]

Abstract The most popular approach in image retrieval is based on the bag of visual-words (BoVW) model. However, there are several fundamental problems that restrict the performance of this method, such as low time efficiency, weak discrimination of visual words and less robustness. So, an image retrieval method with enhanced visual dictionary and query expansion is proposed. Firstly, clustering by fast search and finding density peaks are used to generate a group of visual words. Secondly, non-information words in the dictionary are eliminated by Chi-square model to improve the distinguishing ability of the visual dictionary. Finally, an efficient graph-based visual reranking method is introduced to refine the initial search results. Experimental results of Oxford5K and Paris6K datasets indicate that the expression ability of visual dictionary is effectively improved and the method is superior to the state-of-the-art image retrieval methods in performance.

Key words Bag of visual words (BoVW), clustering based on density, Chi-square model, query expansion

Citation Ke Sheng-Cai, Li Bi-Cheng, Chen Gang, Zhao Yong-Wei, Wei Han. Image retrieval with enhanced visual dictionary and query expansion. *Acta Automatica Sinica*, 2018, **44**(1): 99−105

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的检索以满足用户需求亟待解决. 视觉词典方法 (Bag of visual words, BoVW)[1−3] 通过视觉词典将图像的局部特征量化为词频向量进行检索, 既能利用图像局部信息, 又能达到比局部特征直接检索更快的速度, 成为当前图像检索的主流方法. 但是基于 BoVW 的图像检索方法存在以下问题: 1) 当前生成视觉词典的聚类算法时间效率低、计算量大, 使得 BoVW 难以应用于大规模数据集; 2) 由于聚类算法的局限性和图像背景噪声的存在, 使得视觉词典中存在不包含目标信息的视觉单词, 严重影响视觉词典质量; 3) 没有充分利用初次检索结果中的有用信息, 使得检索效果不理想.

近年来, 研究人员针对这些问题做了许多探索性研究, 如在提高视觉词典生成效率方面: Philbin 等[4] 将 KD-Tree 引入 K-means 中提出近似 K-Means (Approximate K-Means, AKM), 利用 KD-Tree 对聚类中心构建索引目录, 加速寻找最近聚类中心以提高聚类效率. Nister 等[5] 提出了层次 K-means (Hierarchical K-means, HKM), 将时间复杂度降为 O($ndlogk$), 但是该方法忽略了特征维数 d 对聚类效率的影响. 为此, 研究者们提出基于降维的聚类方法, 如主成分分析 (Principal component analysis, PCA)[6]、自组织特征映射 (Self-organizing feature map, SOFM)[7] 等, 主要思路是利用降维算法对高维特征数据进行降维, 再用聚类算法对降维后的特征点进行聚类. 此外, 文献 [8] 通

收稿日期 2016-01-29 录用日期 2016-08-15
Manuscript received January 29, 2016; accepted August 15, 2016
国家自然科学基金 (60872142), 华侨大学科研基金资助
Supported by National Natural Science Foundation of China (60872142) and Scientific Research Funds of Huaqiao University
本文责任编委 刘跃虎
Recommended by Associate Editor LIU Yue-Hu
1. 解放军信息工程大学信息系统工程学院 郑州 450001 2. 75830 部队 广州 510000 3. 华侨大学计算机科学与技术学院 厦门 361021
1. Institute of Information System Engineering, PLA Information Engineering University, Zhengzhou 450001 2. Unit 75830, Guangzhou 510000 3. College of Computer Science and Technology, Huaqiao University, Xiamen 361021

过构造混合概率分布函数来拟合数据集, 但是该方法需要待聚类数据的先验知识, 而且其聚类准确率依赖于密度函数的构造质量.

不包含目标信息的视觉单词类似于文本中的“是”、“的”、“了” 等停用词, 这里称其为 “视觉停用词”, 去除 “视觉停用词” 不仅能缩小词典规模, 还能提高检索准确率. 针对 “视觉停用词” 去除问题, Sivic 等[9] 认为 “视觉停用词” 与其出现的频率存在一定关系, 提出一种基于词频的去除方法. Yuan 等[10] 通过统计视觉短语 (即视觉单词组合) 的出现概率滤除无用信息, Fulkerson 等[11] 则利用信息瓶颈准则滤除一定数量的视觉单词, 但是, 上述方法仅在视觉单词层面考虑如何过滤 “视觉停用词”, 忽略了视觉单词与图像语义概念之间的相互关系.

为利用初次检索结果中的有用信息, 丰富原有查询的信息量, Perd′och 等[12] 提出平均查询扩展策略 (Average query expansion, AQE), 将初始检索结果的图像特征平均值作为新的查询实例, 结合二次检索结果对初次检索得到的图像进行重排序. Shen 等[13] 对查询图像的近邻 (K-nearest neighbors, KNN) 进行多次检索, 对多次检索结果进行重排序得到最终检索结果. Chum 等[14] 则利用查询图像和检索结果中的上下文语义信息提出了自动查询扩展方法, 有效提高了检索准确率. 然而, 现有的查询扩展方法依赖于较高的初始准确率, 在初始准确率较低时, 初始检索结果中的不相关图像会带来负面影响.

综上所述, 为实现更加高效快速的图像检索, 本文提出一种基于视觉词典优化和查询扩展的图像检索方法. 新方法较好地解决了传统方法生成的视觉词典质量差问题, 并有效增强了图像检索性能. 本文剩余部分组织如下: 第 1 节给出了基于视觉词典优化和查询扩展的图像检索方法设计的关键技术, 其中详细介绍了基于密度聚类的视觉词典生成、视觉单词过滤以及基于图结构的查询扩展技术; 第 2 节对本文方法进行了实验验证和性能分析; 最后, 第 3 节为结论.

1 基于视觉词典优化和查询扩展的图像检索

基于视觉词典优化和查询扩展的图像检索方法流程图如图 1 所示. 首先, 提取训练图像的 SIFT (Scale invariant feature transform) 特征, 并利用基于密度的聚类方法对 SIFT 特征进行聚类, 生成视觉词典组; 其次, 通过卡方模型分析视觉单词与目标类别的相关性大小, 同时结合视觉单词词频滤除一定数量的视觉停用词; 然后, 将 SIFT 特征与优化后的视觉词典进行映射匹配, 得到视觉词汇直方图; 最后, 将查询图像的视觉词汇直方图与索引文件进行相似性匹配, 根据初次匹配结果并结合查询扩展策略进行二次或多次检索, 得到最终检索结果.

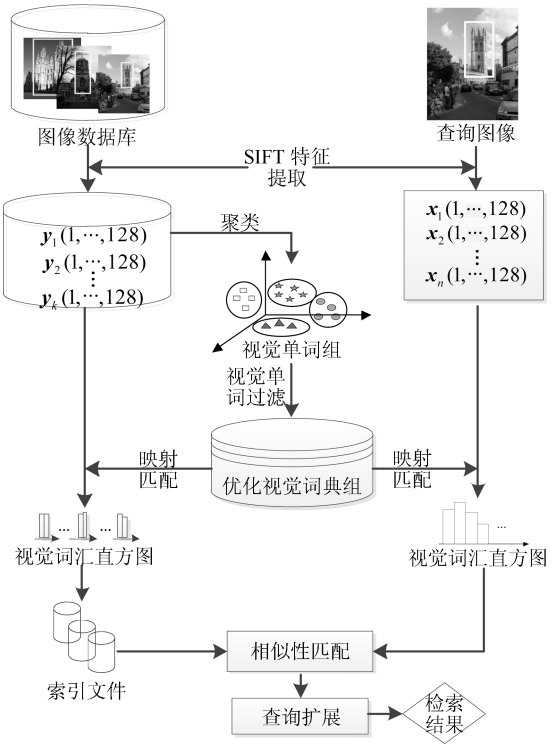


图 1 基于视觉词典优化和查询扩展的图像检索方法流程

Fig. 1 The flow chart of image retrieval based on enhanced visual dictionary and query expansion

1.1 基于密度聚类的视觉词典组

传统的聚类算法需要设计目标函数, 反复迭代计算达到最优, 而文献 [15] 中基于密度的聚类算法 (Density-based clustering, DBC) 通过寻找合适的密度峰值点确定聚类中心, 认为聚类中心同时满足以下 2 个条件: 1) 聚类中心的密度大于临近数据点的密度; 2) 与其他密度更大的数据点距离相距较远. 对待聚类的数据集 $S=\{x_i\}_{i=1}^{N}$ 中数据点 x_i, 定义 ρ_i 表示数据点 x_i 的局部密度:

$$\rho_i=\sum_j \chi(d_{ij}-d_c) \tag{1}$$

其中, d_{ij} 为数据点 x_i 与 x_j 的距离, d_c 是设定的距离阈值, $\chi(x)=\begin{cases}1, & x<0\\ 0, & \text{其他}\end{cases}$, 局部密度 ρ_i 实质是 S 中与 x_i 距离不超过 d_c 的数据点个数. δ_i 表示距 x_i 最近且密度更高点的距离:

$$\delta_i=\begin{cases}\min\limits_{j:\rho_j>\rho_i}(d_{ij}), & \rho_i<\rho_{\max}\\ \max\limits_{j}(d_{ij}), & \rho_i=\rho_{\max}\end{cases} \tag{2}$$

其中, $\rho_{\max}=\max\limits_{i\in S}(\rho_i)$, δ_i 的物理意义是在局部密度大于 ρ_i 的数据点中寻找与 x_i 最近点的距离, 其值

越大, 表示 x_i 与其他高密度点距离越远, 则 x_i 越有可能成为聚类中心; 当 x_i 为密度最大点时, δ_i 等于与 x_i 相距最远点的距离, 该值远远大于其他高密度点的与 δ 值.

为便于选取合适的聚类中心, 定义 $\gamma_i = \rho_i\delta_i$ 作为衡量指标, 显然, 当 γ_i 值越大, 数据点 x_i 越有可能是聚类中心, 因此选取聚类中心时只需对 $\{\gamma_i\}_{i=1}^{N}$ 进行降序排列, 选取前 k 个数据点作为聚类中心即可. 基于密度的聚类算法物理意义清晰, 不需要任何先验信息, 也不用反复迭代计算寻找最优解, 只需设置合适的距离阈值 d_c 即可完成聚类.

1.2　视觉单词过滤

在文本处理中通常根据停用词表过滤文本中的停用词, 然而在 BoVW 中, 视觉单词并不像文本中的单词那样存在确定的实体, 因此无法构造 "视觉停用词" 表, 但是它们之间具有相同的特性: 1) 具有较高的词频; 2) 与目标相关性较小. 针对以上特性, 可以利用卡方模型 (Chi-square model)[16] 统计视觉单词与各目标图像类别之间的相关性, 并结合视觉单词词频信息过滤与目标图像类别无关的视觉单词.

假设视觉单词 w_i 出现的频次独立于图像类别 C_j, 其中 $C_j \in C = \{C_1, C_2, \cdots, C_m\}$, 则视觉单词 w_i 与图像集 C 各图像类别之间的相互关系可由表 1 描述.

表 1　视觉单词 w 与各目标类别统计关系
Table 1　Relation between w and categories of each objective

	C_1	C_2	$\cdots$	C_m	Total
包含 w_i 的图像数目	n_{11}	n_{12}	$\cdots$	n_{1m}	n_{1+}
不包含 w_i 的图像数目	n_{21}	n_{22}	$\cdots$	n_{2m}	n_{2+}
Total	n_{+1}	n_{+2}	$\cdots$	n_{+m}	n_{m+}

其中, n_{1j} 为图像类别 C_j 中包含 w_i 的图像数目, n_{2j} 表示图像类别 C_j 中不包含 w_i 的图像数目, $n_{k+}, k = 1, 2$ 分别表示图像集中包含 w_i 和不包含 w_i 的图像数目, n_{+j} 为图像类别 C_j 中的图像数目, N 为图像集 C 中图像总数目. 则表 1 中视觉单词 w_i 与各图像类别的卡方值为

$$x_i^2 = \sum_{k=1}^{2}\sum_{j=1}^{m}\frac{(N\cdot n_{kj} - n_{k+}\cdot n_{+j})^2}{N\cdot n_{k+}\cdot n_{+j}} \tag{3}$$

卡方值 x_i^2 就代表了 w_i 与各图像类别间统计相关性的大小, 卡方值越大说明视觉单词 w_i 与各图像类别相关性越大; 反之亦然. 考虑到部分视觉单词只出现在很少图像中, 具有较强的类别区分能力, 但是由于其词频较低导致卡方值较小, 因此对卡方值赋予权重如下:

$$\tilde{x}_i^2 = \frac{x_i^2}{tf(w_i)} \tag{4}$$

其中, $f(w_i)$ 为视觉单词 w_i 的词频. 由此, 依据式 (4) 计算各视觉单词加权后的卡方值 $\tilde{x}^2$, 然后过滤 $\tilde{x}^2$ 值较小的视觉停用词.

根据式 (3) 可知计算视觉单词 w_i 的卡方值 χ_i^2 计算复杂度为 O$(2m)$, 其中, m 为图像类别数, 2 对应为 k 分别为 1 和 2 时的累加运算操作, 则去除 "视觉停用词" 的计算复杂度为 O$(2Mm)$, 远小于生成视觉词典的计算复杂度 O(MN), 其中, M 为词典规模, $2m$ 远小于 Oxford5K 数据库的 SIFT 特征数目 N.

1.3　基于图结构的查询扩展

去掉 "视觉停用词" 后, 将图像的 SIFT 特征与优化后的视觉词典进行映射匹配, 得到视觉词汇直方图, 利用图像的视觉词汇直方图进行检索即可得到初始检索结果. 由于图像噪声的存在, 初始检索结果中会存在一些与查询图像无关的检索图像, 因此需要对初始检索结果中的图像进行甄别, 选出与查询图像相关的图像作为新的查询图像, 具体流程如图 2 所示.

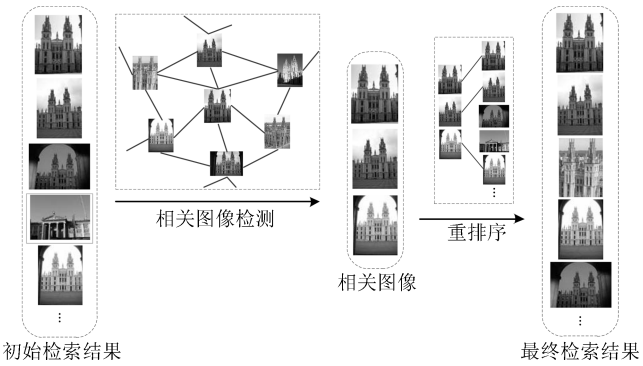


图 2　基于图结构的查询扩展方法流程图
Fig. 2　The flow chart of query expansion based on image structure

在图像集 C 中, 利用图像 i 的视觉词汇直方图 f_i 进行相似性匹配得到 k 近邻图像集 $N_k(i)$, 定义互为对方的 k 近邻图像集中元素的图像对为互相关图像 $R_k(i, i')$:

$$R_k(i, i') = \{(i, i') | i \in N_k(i'), i' \in N_k(i)\} \tag{5}$$

根据式 (5) 构造图 $G = (V, E, W)$, 其中, V 为顶点集, 每一个顶点表示一幅图像, E 是由连接顶点的边组成的集合, W 中的元素为边的权重, 图像 i, i'

之间的连接权重计算如式 (6) 所示:

$$w(i,i')=\begin{cases}\dfrac{|N_k(i)\cap N_k(i')|}{k}, & 若\ (i,i')\in R_k(i,i')\\ 0, & 其他\end{cases} \tag{6}$$

然后, 在图 $G=(V,E,W)$ 中寻找与查询图像相关的密度最大子图 G'[17], 将子图顶点所代表的图像依据与查询图像的相关性大小进行降序排列, 选取前 N_c 幅图像作为新的查询图像, 利用式 (7) 计算扩展查询结果与查询图像的相似性 s_i:

$$s_i=\min\left\{\beta^n\frac{\|f_i-f_n\|_2^2}{\sigma_n^2}|n=1,2,\cdots,N_c\right\} \tag{7}$$

其中, $\beta=0.99$, $\sigma_n^2=\sum_{m=1}^{M}\|f_i-f_n\|_2^2$. 最后, 根据 s_i 的大小进行重排序得到最终检索结果.

2 实验设置与性能评价

2.1 实验设置

为了验证本文方法有效性, 本文在 Oxford5K 图像集[18] 上对本文方法进行了评估, Oxford5K 图像集共包含 5 062 幅图像, 涵盖了牛津大学 11 处标志性建筑, 其中每个目标选取 5 幅图像作为查询图像, 共 55 幅标准查询图像. 此外, 引入 Paris6K 数据库[19] 作为干扰图像, 以验证本文方法在复杂环境下的鲁棒性. 实验硬件配置为内存为 6 GB 的 GPU 设备 GTX Titan 和 Intel Xeon CPU、内存为 16 GB 的服务器. 图像检索性能指标采用平均查询准确率均值 (Mean average precision, MAP) 和查全率−查准率曲线.

2.1.1 实验性能分析

为了分析基于密度聚类算法 (DBC) 中距离阈值参数 d_c 对图像检索 MAP 值的影响. 实验从 Oxford5K 图像集中每类随机选取 50 幅图像, 共计 550 幅图像作为训练图像库, 提取 SIFT 特征后, 在不同距离阈值条件下利用 DBC 进行聚类生成规模 $M=10\,000$ 的视觉词典, 分析距离阈值参数 d_c 对检索 MAP 的影响, 实验结果如图 3 所示.

从图 3 中的 MAP 变化曲线已看出, 距离阈值 $d_c=0.013$ 时, 图像检索准确率达到最高, d_c 设置过大或太小都会降低视觉单词的语义分辨能力. 当离阈值 $d_c>0.013$ 时, 会将距离较远、表达不同图像语义的 SIFT 特征分配到同一个视觉单词, 使得同一视觉单词表达不同的图像语义, 使得检索 MAP 逐渐降低; 而当 $d_c<0.013$ 时, 会将距离较近、表达同一图像语义的 SIFT 特征分到不同的视觉单词, 使得不同视觉单词表达同一图像语义, 导致检索 MAP 值不高.

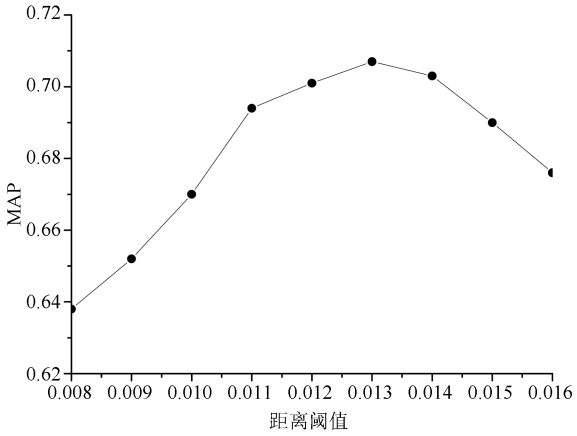


图 3 距离阈值参数 d_c 对图像检索 MAP 值的影响

Fig. 3 The effect of distance threshold on MAP

为了验证基于密度聚类算法的有效性, 设置距离阈值 $d_c=0.013$, 利用 DBC 进行聚类生成不同规模的视觉词典, 分析视觉词典的规模大小对检索 MAP 的影响, 并与 AKM 方法[4] 进行实验对比, 实验结果如图 4 所示:

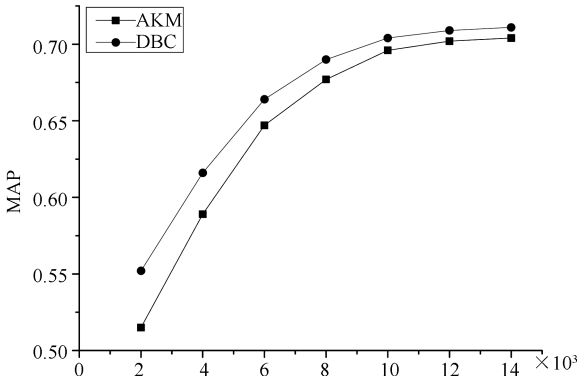


图 4 视觉词典规模对图像检索 MAP 值的影响

Fig. 4 The effect of vocabulary size on MAP

从图 4 可以看出, 当单词数目较小时, 视觉词典的目标分辨能力不强, 随着视觉单词数量不断增加, 其目标分辨能力逐渐增强, MAP 值也逐渐增加, 当词典规模大于 10 K 时, MAP 值增长速度逐渐变慢. 对比 DBC 和 AKM 方法的 MAP 曲线可以看出, DBC 方法的 MAP 值均高于 AKM, 这是因为 AKM 对初始聚类中心的选择敏感且容易陷入局部极值, 而 DBC 的聚类思想不同于基于划分的聚类方法, 既不需要设置初始聚类中心也不用设计目标函数, 而是根据聚类中心具有密度大且与其他高密度点距离较远的特性寻找适合的数据点作为聚类中心, 避免了初值选取对聚类结果的影响, 而且不需要任何先验信息, 只需设置合适的距离阈值 d_c 即可完成聚类.

随后, 为验证卡方模型去除 “视觉停用词” 的有效性, 实验利用 DBC 生成规模 $M=10\,000$ 的视

觉词典, 然后通过卡方模型滤除一定数目的 “视觉停用词”, 并与未去除 “视觉停用词” 的图像检索 MAP 值进行对比, 实验结果如图 5 所示.

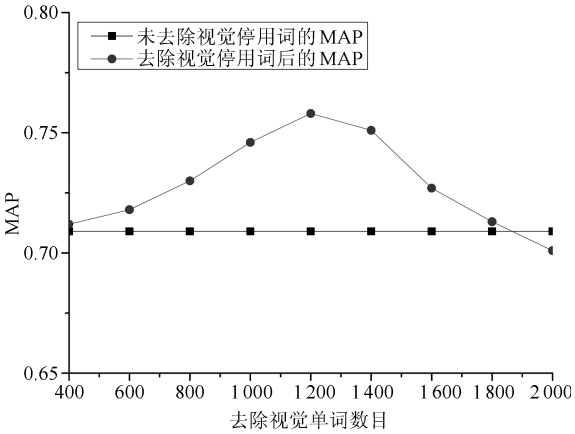


图 5　去除停用词数目对图像检索 MAP 值的影响

Fig. 5　The effect of parameter on MAP

对比图 5 中的 MAP 曲线不难看出, 当去除 “视觉停用词” 数目 $S < 1\,200$ 时, 随着 S 逐渐增加, 视觉词典的目标分辨能力得到有效提高, 并在 $S = 1\,200$ 时 MAP 值达到最大值 75.81 %. 但是, 当 $S > 1\,200$ 时, 随着 “视觉停用词” 滤除数目增加, 使得一些代表性较强的视觉单词被去除, 导致图像检索 MAP 值逐渐降低, 并最终低于未去除 “视觉停用词” 的 MAP 值. 而视觉词典规模 M 发生变化时, 滤除 “视觉停用词” 的最佳数目也会随着变化, 当 M 较小时, 聚类准确率较低, 使得包含目标信息的视觉单词中噪声 SIFT 特征数目较多, “视觉停用词” 数目较少, 因此单词停用率较低; 随着词典规模 M 逐渐变大, 聚类准确率随之增加, 使得包含目标信息的视觉单词中噪声 SIFT 特征数目逐渐减少, “视觉停用词” 的数目逐渐增加, 因此视觉单词停用率逐步增加. 而且不同的图像集中背景噪声均不一样, 因此, 在具体应用时需根据实际情况设置滤除 “视觉停用词” 的数目.

然后, 在词典规模为 10 000, 去除 “视觉停用词” 数目 $S = 1\,200$ 的情况下对查询图像进行检索, 将初始检索结果与平均扩展查询方法 (AQE)[12]、K 近邻重排序方法 (K-nearest neighbors re-ranking, KNNR)[13]、区分扩展查询方法 (Discriminative query expansion, DQE)[20] 和本文方法 (Graph-based query expansion, GBQE) 进行实验对比, 实验结果如表 2 所示. 从表 2 中不难看出, 经过查询扩展后的检索 MAP 值均高于初始检索结果, 说明查询扩展方法能利用初始检索结果中的有用信息, 以此提高检索性能. 其中, AQE 利用初始检索结果的前 k 幅图像的特征平均值作为新的查询实例进行检索, 而 KNNR 方法分别对这 k 幅图像进行扩展查询, 更为有效地利用了扩展图像的细节信息, 但是 AQE 和 KNNR 方法依赖于较高的初始准确率, 没有分析新的查询实例与查询图像之间的相关性. DQE 通过线性支持向量机 (Support vector machine, SVM) 分析扩展项与查询图像的相关性, 并根据相关性大小为其分配权重, 减少无关扩展项的负面影响, 检索性能优于 AQE 和 KNNR 方法, 然而 DQE 仅考虑了查询图像与扩展项的单向相关性, 并没有考虑利用扩展项是否能检索到查询图像. GBQE 方法根据训练图像的互相关图像构建连接图, 定义图像对的 k 近邻中包含相同近邻的数目作为连接权重, 降低了图像中噪声对连接权重的影响, 然后将与查询图像相关的密度最大子图的顶点图像作为扩展项进行扩展查询, 有效去除了无关扩展项对检索结果的影响, 此外, 连接图可离线构造, 减少了在线检索时间, 并可以对新的查询图像进行增量更新. 实验结果表明 GBQE 方法检索性能优于其他方法.

表 2　不同查询扩展方法的图像检索 MAP 值对比 (%)

Table 2　The image retrieval results of different query expansion methods for Oxford5K database (%)

	Initial	AQE	KNNR	DQE	GBQE
All Souls	71.4	79.3	81.8	81.4	83.6
Ashmolean	76.5	81.2	83.1	85.1	87.4
Balliol	73.8	78.4	79.3	80.6	82.5
Bodleian	67.2	70.5	73.4	74.5	74.8
Christ_Church	74.1	78.3	81.5	82.4	83.2
Cornmarket	77.4	82.1	81.8	83.2	84.3
Hertford	85.7	89.2	90.9	91.6	93.2
Keble	86.5	91.6	92.2	93.8	94.4
Magdalen	54.6	61.6	63.8	62.9	63.7
Pitt Rivers	92.4	95.6	95.3	95.1	97.6
Radcliffe cam	74.4	80.8	82.6	84.7	86.1
Average	75.82	80.78	82.34	83.21	84.62

为进一步验证本文方法的性能, 从 Paris6K 数据库中随机选取 1 000 幅图片作为干扰图像, 将本文方法 (EVD + GBQE) 与文献 [20] 中的基于空间特征扩展和区分扩展查询方法 (SPAUG + DQE)、文献 [21] 中的基于上下文近义词和查询扩展图像检索方法 (CSVW + QE) 和文献 [22] 中的基于显著度分析的图像检索方法 (S-sim) 进行实验对比, 实验结果如图 6 所示.

对比图 6 中的数据可知, 采用本文方法 (EVD + GBQE) 较之其他三种方法有更好的表现. S-sim 方法通过对图像显著区域分析, 有效降低了图像背景噪声的不利影响, 由于没有利用初始检索结果对查询图像进行有效扩展, 加入大量干扰图像后其检索性能明显下降; CSVW + QE 方法利用视觉

单词的上下文信息增强单词对图像内容的表达能力, 然而 CSVW + QE 依赖较高的初始查准率, 当无关图像增加时, 其检索性能逐渐下降; SPAUG + DQE 结合视觉单词的上下文信息对局部特征进行扩展, 并根据查询图像与扩展项的相关性大小分配权重, 降低了无关扩展项的不利影响, 使得其抗干扰能力强于 CSVW + QE 和 S-sim, 但是一幅图像包含大量的局部特征, 对局部特征进行扩展的计算和时间开销均较大, 导致实用性不强; EVD + GBQE 则采用无需迭代寻优的聚类方法生成视觉词典, 提高了词典生成效率, 再利用卡方模型滤除不包含目标信息的视觉单词, 增强了词典的语义分辨能力, 然后通过连接图查找与查询图像相关的图像作为扩展项并进行扩展查询, 根据扩展查询结果对初始检索结果重排序, 实验结果表明, EVD + GBQE 在复杂环境下仍具有较好的表现, 实用性更强. 图 7 给出了本文方法在 Oxford5K + Paris6K 数据库上的图像检索结果, 不难看出, 利用本文方法可以将初始检索结果中无关图像剔除, 从而检索得到更多与查询图像相关的图像.

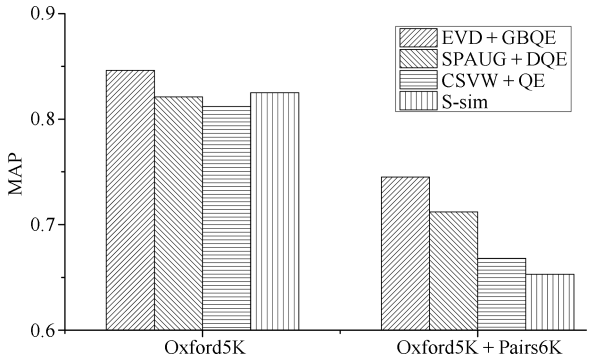


图 6 在 Oxford5K 和 Oxford5K+Paris6K 数据库上的图像检索 MAP 值

Fig. 6 The MAP of different methods for Oxford5K and Oxford5K+Paris6K database

3 结论

本文提出了一种基于视觉词典优化和查询扩展的图像检索方法. 首先, 针对传统视觉词典生成方法效率低问题, 引入基于密度的聚类方法生成视觉词典, 根据聚类中心具有的特性快速寻找适合的数据点作为聚类中心, 避免了迭代寻优过程, 有效提高了词典生成效率; 然后, 利用卡方模型分析视觉单词与图像目标的相关性, 同时结合视觉单词词频滤除不包含目标信息的 "视觉停用词", 提高了视觉词典的质量; 最后, 通过连接图查找与查询图像相关的图像作为扩展项, 并对初始检索结果进行重排序, 降低了初始检索中不相关图像的影响, 提高了图像检索准确率. 实验结果有效地验证了本文方法的图像检索性能优于当前主流方法. 如何将目标空间信息与视觉单词相结合, 增强视觉单词的语义表达能力是本文的下一步研究方向. 此外, 如何通过距离度量的学习使得特征空间的距离更加接近真实的语义距离也是今后亟待解决的问题.

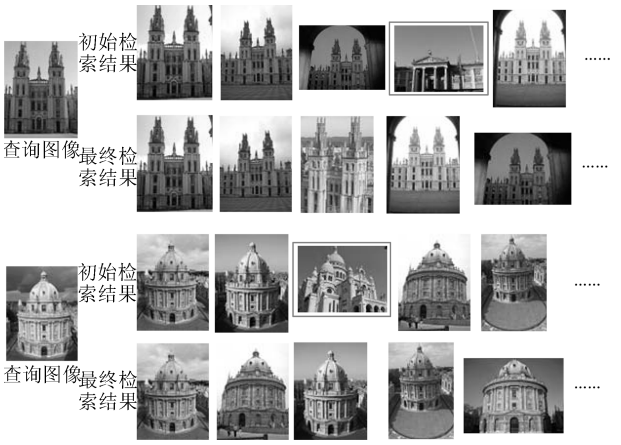


图 7 EVD+GBQE 方法在 Oxford5K+Paris6K 数据库上的检索结果

Fig. 7 The image retrieval results of EVD+GBQE for Oxford5K+Paris6K database

References

1 Chen Y Z, Dick A, Li X, Van Den Hengel A. Spatially aware feature selection and weighting for object retrieval. *Image and Vision Computing*, 2013, **31**(12): 935−948

2 Wang J J Y, Bensmail H, Gao X. Joint learning and weighting of visual vocabulary for bag-of-feature based tissue classification. *Pattern Recognition*, 2013, **46**(12): 3249−3255

3 Cao Y, Wang C H, Li Z W, Zhang L Q, Zhang L. Spatial-bag-of-features. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 3352−3359

4 Philbin J, Chum O, Isard M, Sivic J, Zisserman A. Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007. 1−8

5 Nister D, Stewenius H. Scalable recognition with a vocabulary tree. In: Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 2161−2168

6 Goes J, Zhang T, Arora R, Lerman G. Robust stochastic principal component analysis. In: Proceedings of the 17th International Conference on Artificial Intelligence and Statistics. Reykjavik, Iceland: JMLR, 2014. 266−274

7 Goswami A K, Jain R, Tripathi P. Automatic segmentation of satellite image using self organizing feature map (SOFM) an artificial neural network (ANN) approach. *International Journal of Advanced Research in Computer Science*, 2014, **5**(8): 92−97

8 McLachlan G, Krishnan T. *The EM Algorithm and Extensions* (Second Edition). Hoboken, New Jersey: John Wiley & Sons, 2008.

9 Sivic J, Zisserman A. Video Google: a text retrieval approach to object matching in videos. In: Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003. 1470−1477

10 Yuan J S, Wu Y, Yang M. Discovery of collocation patterns: from visual words to visual phrases. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007. 1−8

11 Fulkerson B, Vedaldi A, Soatto S. Localizing objects with smart dictionaries. In: Proceedings of the 10th European Conference on Computer Vision. Berlin, Heidelberg, Germany: Springer, 2008. 179−192

12 Perd′och M, Chum O, Matas J. Efficient representation of local geometry for large scale object retrieval. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 9−16

13 Shen X H, Lin Z, Brandt J, Avidan S, Wu Y. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 3013−3020

14 Chum O, Philbin J, Sivic J, Isard M, Zisserman A. Total recall: automatic query expansion with a generative feature model for object retrieval. In: Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1−8

15 Rodriguez A, Laio A. Clustering by fast search and find of density peaks. *Science*, 2014, **344**(6191): 1492−1496

16 Kesom K, Poslad S. An enhanced bag-of-visual word vector space model to represent visual content in athletics images. *IEEE Transactions on Multimedia*, 2012, **14**(1): 211−222

17 Zhang S T, Yang M, Cour T, Yu K, Metaxas D N. Query specific rank fusion for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(4): 803−815

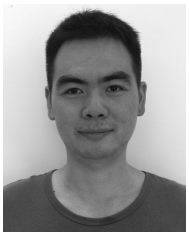
18 Philbin J, Arandjelović R, Zisserman A. Oxford5K dataset [Online], available: http://www.robots.ox.ac.uk/∼vgg/data/oxbuildings/, December, 2015.

19 Philbin J, Zisserman A. Paris6K database [Online], available: http://www.robots.ox.ac.uk/∼vgg/data/parisbuildings/, December, 2015.

20 Arandjelović R, Zisserman A. Three things everyone should know to improve object retrieval. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 2911−2918

21 Xie H T, Zhang Y D, Tan J L, Guo L, Li J T. Contextual query expansion for image retrieval. *IEEE Transactions on Multimedia*, 2014, **16**(4): 1104−1114

22 Gao Y, Shi M J, Tao D C, Xu C. Database saliency for fast image retrieval. *IEEE Transactions on Multimedia*, 2015, **17**(3): 359−369



柯圣财 解放军信息工程大学信息系统工程学院硕士研究生. 解放军 75830 部队助理工程师. 主要研究方向为图像处理和计算机视觉.
E-mail: keshengcai0705@163.com
(**KE Sheng-Cai** Master student at the Institute of Information System Engineering, PLA Information Engineering University, assistant engineer at Unit 65022. His research interest covers image processing and computer vision.)

李弼程 华侨大学计算机科学与技术学院教授. 主要研究方向为文本分析与理解, 语音处理与识别, 图像/视频处理与识别, 信息融合. 本文通信作者.
E-mail: lbclm@163.com
(**LI Bi-Cheng** Professor at the College of Computer Science and Technology, Huaqiao University. His research interest covers text analysis and understanding, speech/image/video processing and recognition, and information fusing. Corresponding author of this paper.)

陈 刚 解放军信息工程大学信息系统工程学院讲师. 主要研究方向为自然语言处理, 图像/视频处理与识别.
E-mail: maplechen111@gmail.com
(**CHEN Gang** Lecturer at the Institute of Information System Engineering, PLA Information Engineering University. His research interest covers natural language processing, image/video processing and recognition.)

赵永威 解放军信息工程大学信息系统工程学院博士研究生. 主要研究方向为图像/视频处理与识别.
E-mail: zhaoyongwei369@163.com
(**ZHAO Yong-Wei** Ph. D. candidate at the Institute of Information System Engineering, PLA Information Engineering University. His research interest covers image/video processing and recognition.)

魏 晗 解放军信息工程大学信息系统工程学院讲师. 主要研究方向为计算机视觉, 图像/视频处理与识别.
E-mail: weihan0627@126.com
(**WEI Han** Lecturer at the Institute of Information System Engineering, PLA Information Engineering University. Her research interest covers computer vision, image/video processing and recognition.)