

Journal of Image
and Graphics

中国图象图形学报



ISSN1006-8961
CN11-3758/TB

2013 **6**
Vol.1 No.

中国科学院遥感应用研究所
中国图象图形学学会主办
北京应用物理与计算数学研究所

中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

2013 年 6 月 第 18 卷 第 6 期(总第 206 期)

目 次

综述

利用块效应特征的 JPEG 图像盲取证研究进展 赵洁, 郭继昌, 张艳(613)

图像处理和编码

可信度双三次插值的 PMD 距离信息和 RGB 图像数据融合 ... 胡良梅, 王竹萌, 高隼, 张旭东, 吴国松(621)

分数阶微分的 CIE $L^*a^*b^*$ 颜色空间边缘检测 汤慧梅, 赵跃进(628)

图像分析和识别

频谱预处理模糊运动方向鉴别的改进算法 孔勇奇, 卢敏, 潘志庚(637)

基于视频动态纹理的火灾检测 邵婧, 王冠香, 郭蔚(647)

医学图像处理

旋转不变的 X 线图像相关度拼接 赵宏, 赵凯, 江春花, 康雁(654)

心肌灌注核磁共振图像的非刚性配准 王建, 潘静薇, 杨新(661)

遥感图像处理

空间约束的无人机影像 SURF 特征点匹配 韩天庆, 赵银娣, 刘善磊, 白杨(669)

支持向量机和水平集的高分辨率遥感图像河流检测 于晓升, 吴成东, 陈东岳, 田子恒(677)

地理信息技术

影响域渐进扩展的居民地增量综合 许俊奎, 武芳(685)

“第 16 届全国图象图形学学术会议”专栏

场景切换视频自适应帧间码率控制 方志军, 高永彬, 舒雷, 袁非牛, 杨勇(692)

适用于 PACS 系统的医学图像近无损压缩 李萍, 蒋慧琴, 杨晓鹏, 刘玉敏(699)

用于视觉词语生成的概率预测器 史淼晶, 徐蕊鑫, 许超(706)

多特征融合的人体目标再识别 范彩霞, 朱虹, 蔺广逢, 罗磊(711)

G^2 连续的低次避障代数样条曲线 陈军(718)

自适应多特征融合的真实感地形快速绘制 曾艳阳, 康凤举, 杨惠珍(724)

第九届和谐人机环境联合学术会议(HHME2013)征稿通知 (730)

Journal of Image and Graphics

(Monthly, Started in 1996)

Vol. 18 No. 6 June 2013

Contents

Review

Survey: blind forensic of JPEG forgeries based on blocking artifacts Zhao Jie, Guo Jichang, Zhang Yan(613)

Image Processing and Coding

PMD distance information and high resolution RGB image data fusion based on bicubic interpolation with credibility
..... Hu Liangmei, Wang Zhumeng, Gao Jun, Zhang Xudong, Wu Guosong(621)

Edge detection in CIE L*a*b* based on fractional differential Tang Huimei, Zhao Yuejin(628)

Image Analysis and Recognition

Motion blur direction detection based on pre-processing of spectrum image Kong Yongqi, Lu Min, Pan Zhigeng(637)

Fire detection based on video dynamic texture Shao Jing, Wang Guanxiang, Guo Wei(647)

Medical Image Processing

X-ray image mosaic based on rotation invariant correlation Zhao Hong, Zhao Kai, Jiang Chunhua, Kang Yan(654)

Non-rigid registration for myocardial perfusion MR image Wang Jian, Pan Jingwei, Yang Xin(661)

Remote Sensing Image Processing

Spatially constrained SURF feature point matching for UAV images Han Tianqing, Zhao Yindi, Liu Shanlei, Bai Yang(669)

Using support vector machine and level set for river detection in high resolution remote sensing image
..... Yu Xiaosheng, Wu Chengdong, Chen Dongyue, Tian Ziheng(677)

Geoinformatics

Expanding the influencing domain gradually for incremental generalization of settlement Xu Junkui, Wu Fang(685)

Issum of the 16th National Conference on Image and Graphics

Adaptive frame-layer rate control in scene change video Fang Zhijun, Gao Yongbin, Shu Lei, Yuan Feiniu, Yang Yong(692)

Near-lossless compression for medical image in PACS Li Ping, Jiang Huiqin, Yang Xiaopeng, Liu Yumin(699)

Probabilistic predictor for fast visual word generation Shi Miaojing, Xu Ruixin, Xu Chao(706)

Person re-identification based on multi-features Fan Caixia, Zhu Hong, Lin Guangfeng, Luo Lei(711)

Construction of low degree algebraic spline curve with G^2 continuity to avoid obstacles Chen Jun(718)

Fast rendering of realistic terrain based on adaptive multiple features fusion Zeng Yanyang, Kang Fengju, Yang Huizhen(724)

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2013)06-0706-05

论文引用格式: 史淼晶, 徐蕊鑫, 许超. 用于视觉词语生成的概率预测器[J]. 中国图象图形学报, 2013, 18(6): 706-710.

用于视觉词语生成的概率预测器

史淼晶¹, 徐蕊鑫², 许超¹

1. 北京大学机器感知与智能教育部重点实验室, 北京 100871; 2. 微软亚洲搜索技术中心, 北京 100080

摘要: 视觉词语的产生是基于字袋模型的图像检索中的重要一环; 根据已知的视觉词典, 查询图像特征被映射到词典中相应的视觉词语。提出一种新的基于空间相关性的快速视觉词语产生算法。统计视觉词典中任意两个词语在数据库中的共生次数, 构建视觉词语共生表。利用共生表, 建立一种新的概率预测器来辅助预测已知词语的近邻词语。将预测器与快速近似最近邻查找算法结合, 在标准图像检索数据库上进行实验测试, 相比较传统的树形搜索算法或哈希算法, 新算法在时间效率上获得明显提高。

关键词: 字袋模型; 空间相关性; 视觉词语共生表; 概率预测器

Probabilistic predictor for fast visual word generation

Shi Miaojing¹, Xu Ruixin², Xu Chao¹

1. Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China;

2. Search Technology Center Asia, Microsoft, Beijing 100080, China

Abstract: Visual word generation is a key observation in obtaining the bag-of-visual-words (BOVW) representation for image retrieval; query image features are mapped to their visual words according to the pre-clustered codebook. In this paper, we propose a novel generation approach based on the spatial correlation of visual words. A visual word co-occurrence table is constructed in the first step. Given the known visual words, a new probabilistic predictor is then presented to accelerate the generation of their neighboring visual words. We combine the co-occurrence table with the fast library for approximate nearest neighbors (FLANN), and test it on the Oxford dataset. Comparisons with representative approaches suggest the efficiency and effectiveness of the new scheme.

Key words: bag-of-visual-words (BOVW); spatial correlation; visual word co-occurrence table; probabilistic predictor

0 引言

基于内容的图像检索系统中, 图像被表示成尺度不变的局部特征的集合。通过将数据库中图像特征聚类与量化, 可以得到视觉词典。则任意一幅新的图像的特征, 都可以根据这本视觉词典映射到相应的视觉词语, 图像被表示为一袋视觉词语, 这就是所谓的字袋模型。其中如何将图像特征映射到相应

的视觉词语也即是所谓的视觉词语产生阶段, 是字袋模型的重要环节, 其时间与精度直接决定了字袋模型的检索效率与效果。

作为一种经典算法, 树形搜索算法被广泛应用于图像检索的视觉词语产生。通过多叉树的索引, 视觉词语的产生理论上最快可以在 $O(\log N)$ 的时间内完成, N 是视觉词典大小。代表算法包括 KD 树和 K 均值树。基于 KD 树搜索算法的相关工作有优先队列法^[1]、随机 KD 森林算法^[2], 分布式随机 KD 森林算

收稿日期: 2010-11-15; 修回日期: 2012-12-14

基金项目: 国家重点基础研究发展计划(973) 基金项目(2011CB302400); 国家自然科学基金项目(60975014, 61121002); 北京市自然科学基金项目(4102024); 深圳基础研究课题(JCYJ20120614152136201)

第一作者简介: 史淼晶(1987—), 男, 北京大学图像处理专业博士研究生, 主要研究方向为图像检索。E-mail: shimj@cis.pku.edu.cn

法^[3]等等;基于K均值树搜索算法的相关工作有Spill-Tree法^[4],以及分级索引的K均值树法^[5]。

Marius 于2009年提出快速近似最近邻查找算法^[6](FLANN),该方法可以根据数据集的分布特点、对映射精度和空间资源消耗的要求来推荐索引类型和检索参数,在高维空间最近邻查找领域得到广泛的使用。文中采用FLANN作为基准算法提供候选树结构:随机KD森林或分级索引K均值树。此外,在高维数据K近邻问题中,局部敏感哈希^[7]因其运行时间与数据维度无关而被广泛采用。

上述算法尽管在查找近似最近邻问题中得到了很好的应用,然而它们并没有考虑图像特征之间的相关性,图像特征之间的映射是独立,随着数据维数的增长,树形搜索的回溯次数将呈指数增长,其搜索效率将大大降低。基于此,许多研究者致力于挖

掘特征之间的相关性来帮助视觉词语的产生以及图像相似度之间的度量^[8-9]。

针对此问题,经过大量的实验发现,自然图像的特征是空间上紧密联系的。如图1所示是从牛津数据库的Radcliffe地标中选出3对共生视觉词语,每对包含两个共生词语(红色与绿色点对),其中红色点表示中心点,黄色椭圆示意它的仿射不变区域。所选的3对共生对在整个数据库中分别共生80,66和171次。这样的共生性表明任何一个特征的出现总能对其近邻特征的出现给予一定的语义暗示。用数学语言来描述这种预期,在整个训练数据库中记录任意两个视觉词语出现的条件概率,并且建立一张视觉词语共生表,以此来表示视觉词语之间的语义相关性^[10],视觉词语共生表被应用于之前的工作^[11]。

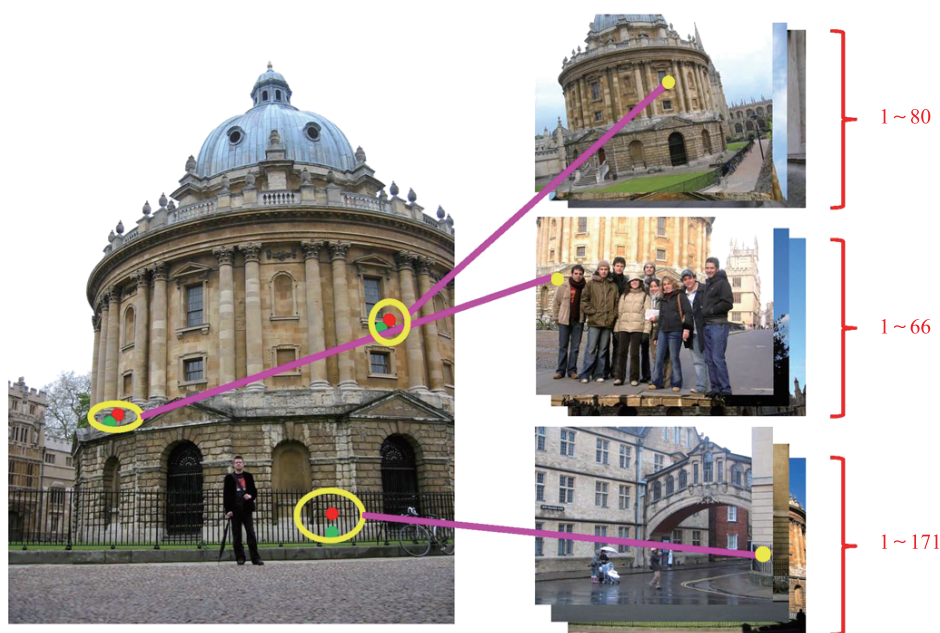


图1 从牛津数据库的Radcliffe地标中选出3对共生视觉词语

Fig.1 An illustration of three pairs of visual words selected from landmark Radcliffe of Oxford dataset

提出一种基于概率的视觉词语预测器:对测试图像,当采样获得一些特征对应的视觉词语后,结合视觉词语共生表,剩余特征的对应视觉词语都可以利用本文概率预测器给出最可能出现的候选视觉词语,然后通过比较欧氏距离得出特征的最近邻视觉词语。

1 视觉词语产生

1.1 视觉词语共现表

如果视觉词语 w_1 对应特征位于视觉词语 w_2 对

应特征的仿射不变区域内,则认为 w_1 与 w_2 是共生的。 $n^i(w_1, w_2)$ 表示它们在第 i 幅图像中共生的次数, $N(w_1, w_2)$ 是它们在整个数据库中共生的次数,即

$$N(w_1, w_2) = \sum_i n^i(w_1, w_2) \quad (1)$$

考虑到特征的仿射不变区域尺寸一般都很小,将所有的区域都扩大五倍,然后记录任意两个视觉词语之间的共生次数并且构成视觉词语共生表。除了共生次数以外,每一个视觉词语在数据库中单独出现

的次数 $N(w)$ 也需要被记录。

1.2 概率预测器

假设已经得到 s 个视觉词语, 构成集合 $S = \{w_0, w_1, \dots, w_{s-1}\}$, 根据 S 以预测这些视觉词语对应特征的近邻特征的视觉词语。前文提到, 对任意一个视觉词语, 在数据库中可能与他们共生的视觉词语记录在共生表中。因为共生是限定在特征的仿射不变区域内的, 所以共生表是很稀疏的。将 S 中的每一个视觉词语对应的共生表中的共生词语集合在一起 $W = \{w_{s0}, w_{s1}, \dots, w_{sv}\}$, 假设有 v 个, 作为近邻特征的候选视觉词语。

利用贝叶斯准则来给出 W 中的最优候选视觉词语 $\hat{w}_s$, 作为最可能与 $w_0, w_1, \dots, w_{s-1}$ 共生的视觉词语。这种可能性在数学中代表了一种最大的后验概率 $p(\hat{w}_s | w_0, \dots, w_{s-1})$, 即

$$\hat{w}_s = \arg \max_{\hat{w}_s \in W} p(\hat{w}_s | w_0, \dots, w_{s-1}) \quad (2)$$

可以用联合概率来计算后验概率, 即

$$p(\hat{w}_s | w_0, \dots, w_{s-1}) = \frac{p(\hat{w}_s, w_0, \dots, w_{s-1})}{p(w_0, \dots, w_{s-1})} \quad (3)$$

$p(\hat{w}_s, w_0, \dots, w_{s-1})$ 表示 $\hat{w}_s, w_0, \dots, w_{s-1}$ 的联合概率, 可以被分解为

$$p(\hat{w}_s, w_0, \dots, w_{s-1}) = p(w_0, \dots, w_{s-1} | \hat{w}_s) p(\hat{w}_s) \quad (4)$$

$p(\hat{w}_s)$ 可以通过 $N(\hat{w}_s)$ 来估计。为了分解式(4), 假设已获得的视觉词语 $w_0, w_1, \dots, w_{s-1}$ 之间是统计独立的, 这样 $p(w_0, \dots, w_{s-1})$ 就可以被近似为 $\prod_{w_i} p(w_i)$, 而 $p(w_0, \dots, w_{s-1} | \hat{w}_s)$ 只依赖于 $\hat{w}_s$, 这样它们都可以被一阶条件概率所近似, 即

$$p(\hat{w}_s, w_0, \dots, w_{s-1}) \approx p(\hat{w}_s) \prod_{w_i \in S} p(w_i | \hat{w}_s) \quad (5)$$

$p(w_i | \hat{w}_s)$ 可以从先验共生次数 $N(w_i, \hat{w}_s)$ 估计得到。式(3)的最大后验概率对应分解形式

$$\begin{aligned} \hat{w}_s &= \arg \max_{\hat{w}_s \in W} p(\hat{w}_s | w_0, \dots, w_{s-1}) \approx \\ &\arg \max_{\hat{w}_s \in W} \frac{p(\hat{w}_s) \prod_{w_i \in S} p(w_i | \hat{w}_s)}{\prod_{w_i \in S} p(w_i)} = \\ &\arg \max_{\hat{w}_s \in W} \frac{N(\hat{w}_s) \prod_{w_i \in S} \frac{N(w_i, \hat{w}_s)}{N(\hat{w}_s)}}{\prod_{w_i \in S} \frac{N(w_i)}{N_T}} \end{aligned} \quad (6)$$

N_T 是所有视觉词语出现的总次数。这里的近似最大化表示 $w_0, w_1, \dots, w_{s-1}$ 与 $\hat{w}_s$ 共生的可能性是最大的, 也即是说由视觉词语 $w_0, w_1, \dots, w_{s-1}, w_s$ 构成的当前局部区域是最可能出现的局部模式。

如果每一个视觉词语的预测仅仅依赖于距离其空间特征最近的视觉词语 w_{s-1} , 则式(6)退化成一阶条件概率的形式, 即

$$\hat{w}_s = \arg \max_{\hat{w}_s \in W} p(\hat{w}_s | w_{s-1}) \quad (7)$$

1.3 近邻组与中心组

在建立共生表的时候统计与每一个中心特征共生的特征, 而对测试图像则统计每一个特征都与哪些中心特征共生。这样方式使得一个特征可以同时属于多个中心特征的仿射不变区域, 从而更多的共生信息可以被利用。

具体算法中共生表建立在每一个中心特征与其共生特征构成的近邻组, 将特征近邻组映射到视觉词语近邻组后建立视觉词语共生表; 在另一方面, 对测试图像则找到每一个特征所属的中心特征区域构成中心组; 中心特征被首先映射到对应的视觉词语。对其共生特征利用概率预测器提供候选视觉词语。须知这样的选择仅仅提供一种可能性。为了产生精确的视觉词语, 通过欧氏距离来比较特征与候选词语之间的距离是必不可少的。

1.4 视觉词语产生步骤

- 1) 给定一幅测试图像, 首先提取 SIFT 特征;
- 2) 随机选择 20% 特征作为中心特征, 并且在他们各自的仿射不变区域内统计他们的近邻特征;
- 3) 将 20% 中心特征用 FLANN 作精确映射;
- 4) 利用概率预测器为那些近邻特征提供候选视觉词语, 比较他们与特征之间的欧氏距离决定最优的视觉词语;
- 5) 如果一些特征不属于任何中心特征的区域, 或者当前计算得到的最佳欧氏距离仍然很大, 就用 FLANN 继续搜索。

相比较树形搜索算法, 本文视觉词语产生算法可以更快地给出那些更可能出现的视觉词语, 从而大大降低比较次数与时间复杂度。

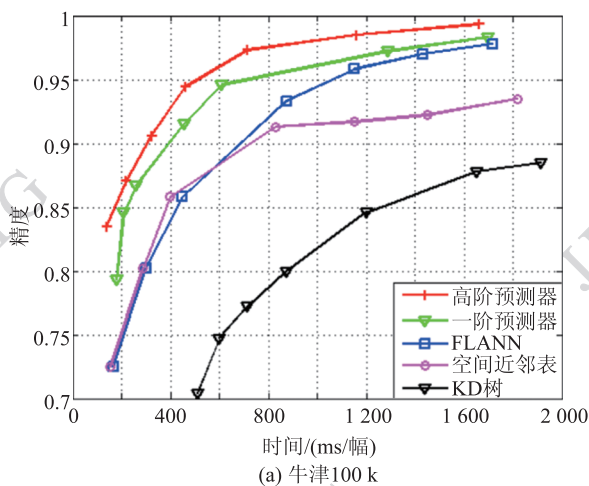
2 实验与讨论

在著名的牛津数据库^[12]上测试本文基于概率

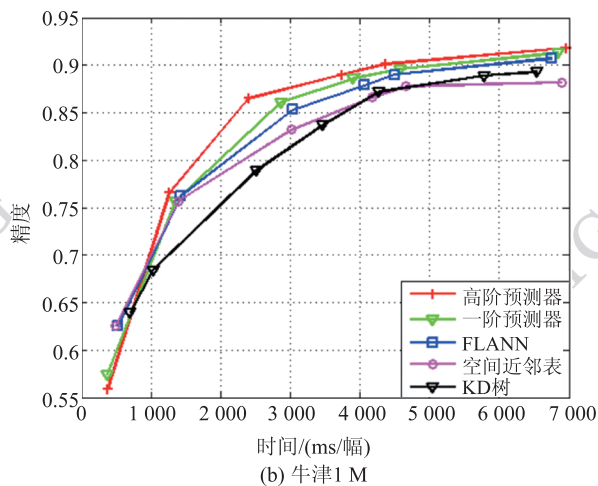
预测器的视觉词语产生算法。将整个数据库 5 061 幅图像分为两部分:一部分包含 90% 的图像用来训练构建视觉词语共生表;另一部分 10% 的图像用以做测试。随机地对数据库进行 10 次划分并且记录平均每幅图像的视觉词语产生时间。设置不同的最大比较次数然后记录相应的预测时间与精度。词典大小分别选择为 100 k 和 1 M。主要比较算法包括 FLANN^[6] 和空间近邻表^[10]。KD 树^[1] 作为基准。

从图 2 中可以看出,本文算法明显地优于 FLANN。尤其在最大比较次数比较小的时候,因为这时候正是概率预测器起作用的时候。可以看出对 100 k 的视觉词典,近似视觉词语产生算法精

度到达 0.95 时,比较高阶概率预测器算法与 FLANN 算法,时间效率提高了约为 142%;其次,在相同时间 500 ms,高阶概率预测器算法精度可以达到 0.95,而 FLANN 精度约为 0.88;对 1 M 的视觉词典,近似视觉词语产生算法精度到达 0.90 时,比较高阶概率预测器算法与 FLANN 算法,时间效率提高了约为 35%;另一方面,在相同时间 4.3 s,高阶概率预测器算法精度可以达到 0.9,而 FLANN 精度约为 0.88。基准算法 KD 树(随机 KD 森林的数目设为 1)。空间近邻表算法在最大比较次数比较低的时候是优于 FLANN 的,随着比较次数的增大,受到其本身稀疏度的限制,产生精度将被限制在一定的数值一下。



(a) 牛津100 k



(b) 牛津1 M

图2 牛津数据库上的视觉词产生的时间精度对应结果

Fig. 2 Visual word generation results on Oxford dataset

从图 2 可以看出,一般来说高阶预测器的效果要明显优于一阶预测器。这就说明对每一个视觉词语来说,加入更多的共生信息确实对其预测产生了积极的效果。然而,在实验中并不意味着加入越多的共生信息效果越好,这一点从 100 k 与 1 M 的实验结果比较中也可以看出。实验结果表明 100 k 的词典效果要优于 1 M 的词典。随着词典大小的增加,共生的偏差也被相应地引入,概率预测器的候选特征词语也相应地不断增加,搜索效率降低。然而这并不意味着词典越小,本文算法越是优越。须知与此同时,传统树形算法的时间复杂度也大大降低,接近其理论值。

3 结 论

基于概率预测器的视觉词语产生算法,通过挖掘视觉词语之间的共生性,建立视觉词语共生表。

利用从训练数据中收集的共生信息之间的条件概率来为视觉词语的产生提供更少更可能出现的候选词语。因而降低了计算的时间复杂度。在牛津数据库上的测试结果表明,相比较其他代表性算法,本文算法能够更加有效而快速地产生视觉词语。

参考文献 (References)

- [1] Beis J, Lowe D. Shape indexing using approximate nearest neighbor search in high-dimensional spaces[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Puerto Rico; IEEE Computer Society, 1997: 1000-1006.
- [2] Aly M, Munich M, Perona P. Distributed kd-trees for retrieval from very large image collections[C]//Proceedings of the 22nd British Machine Vision Conference. Dundee, UK: BMVA, 2011.
- [3] Silpa-Anan C, Hartley C. Optimised kd-trees for fast image de-

- scriptor matching [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska: IEEE Computer Society, 2008:1-8.
- [4] Liu T, Moore A, Gray A, et al. An investigation of practical approximate nearest neighbor algorithms [C]// Proceedings of the 18th Annual Conference on Neural Information Processing Systems. Vancouver, Canada: MIT, 2004: 825-832.
- [5] Nister D, Stewenius H. Scalable recognition with a vocabulary tree [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2006: 2161-2168.
- [6] Muja M, Lowe D. Fast approximate nearest neighbors with automatic algorithm configuration [C]// Proceedings of IEEE Conference on Computer Vision Theory and Applications. Lisbon, Portugal: IEEE Computer Society, 2009: 331-340.
- [7] Chum O, Philbin J, Zisserman A. Near duplicate image detection: min-hash and tf-idf weighting [C]// Proceedings of the 19th British Machine Vision Conference. London, UK: BMVA, 2008: 493-502.
- [8] Jegou H, Douze M, Schmid C. Exploiting descriptor distances for precise image search [R]. Paris: National Institute for Research in Computer Science and Control, 2011.
- [9] Song J, Ma Y, Hu F, et al. Scalable image retrieval based on feature forest [C]// Proceedings of the 10th Asian Conference on Computer Vision. Queenstown, New Zealand: IEEE Computer Society, 2010: 506-515.
- [10] Xu R, Shi M, Geng B, et al. Fast visual word assignment via spatial neighborhood boosting [C]// Proceedings of the 19th International Conference on Multimedia and Expo. Barcelona, Spain: IEEE Computer Society, 2011: 262-270.
- [11] Li Y, Geng B, Zha Z, et al. Query expansion by spatial co-occurrence for image retrieval [C]// Proceedings of the 19th ACM International Conference on Multimedia. New York, USA: ACM, 2011: 1177-1180.
- [12] 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. Minneapolis, Minnesota, USA: IEEE Computer Society, 2007: 1-8.