

6-4-2020

## Image Retrieval Using Weighted Vocabulary Tree with Location Information

Chen Ying

*Key Laboratory of Advanced Control Light Process, Jiangnan University, Wuxi 214000, China;*

Jiayu Guo

*Key Laboratory of Advanced Control Light Process, Jiangnan University, Wuxi 214000, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

---

## Image Retrieval Using Weighted Vocabulary Tree with Location Information

### Abstract

**Abstract:** Aiming at the positional information ignorance problem in image retrieval based on SIFT feature matching, an image retrieval approach using weighted vocabulary tree based on the spatial location information was proposed. *The vocabulary tree is used in the method to quantify SIFT features as the visual words, converting the image match into the visual words' weight vector match. Because only visual words' weight match ignored the mutual position effect, SIFT points' spatial location information was generated, and was classified into the position effect of visual words according to affiliation between SIFT points and visual words. The spatial location information of the visual words was used as the weighting factor of weight vector matches, refining the matching score between features. Similar images were retrieved by similarity sort.* The experimental results show that the algorithm can effectively improve the accuracy of the image retrieval.

### Keywords

vocabulary tree, location information, vector matching, image retrieval

### Recommended Citation

Chen Ying, Guo Jiayu. Image Retrieval Using Weighted Vocabulary Tree with Location Information[J]. Journal of System Simulation, 2017, 29(10): 2353-2360.

# 利用位置信息加权词汇树的图像检索

陈莹, 郭佳宇

(江南大学轻工过程先进控制教育部重点实验室, 江苏 无锡 214000)

**摘要:** 针对基于 SIFT 特征匹配的图像检索方法忽略空间位置信息的缺陷, 提出了基于空间位置信息加权词汇树的图像检索方法。利用分层词汇树将图像 SIFT 特征量化为视觉单词, 将图像匹配转换成视觉单词权值向量匹配。由于单纯的视觉单词权值向量匹配忽略了视觉单词之间的相互位置影响, 生成 SIFT 点的空间位置影响信息, 根据 SIFT 点与视觉单词之间的从属关系, 将 SIFT 点的空间位置影响信息聚类为视觉单词之间的位置影响, 将视觉单词的空间位置信息作为加权系数, 加入到视觉单词的权值匹配中去, 细化特征之间的匹配得分, 通过相似度排序检索相似图像。实验结果表明, 该算法能有效提高图像检索的精确度。

**关键词:** 词汇树; 位置信息; 向量匹配; 图像检索

中图分类号: TP391

文献标识码: A

文章编号: 1004-731X (2017) 10-2353-08

DOI: 10.16182/j.issn1004731x.joss.201710017

## Image Retrieval Using Weighted Vocabulary Tree with Location Information

Chen Ying, Guo Jiayu

(Key Laboratory of Advanced Control Light Process, Jiangnan University, Wuxi 214000, China)

**Abstract:** Aiming at the positional information ignorance problem in image retrieval based on SIFT feature matching, an image retrieval approach using weighted vocabulary tree based on the spatial location information was proposed. The vocabulary tree is used in the method to quantify SIFT features as the visual words, converting the image match into the visual words' weight vector match. Because only visual words' weight match ignored the mutual position effect, SIFT points' spatial location information was generated, and was classified into the position effect of visual words according to affiliation between SIFT points and visual words. The spatial location information of the visual words was used as the weighting factor of weight vector matches, refining the matching score between features. Similar images were retrieved by similarity sort. The experimental results show that the algorithm can effectively improve the accuracy of the image retrieval.

**Keywords:** vocabulary tree; location information; vector matching; image retrieval

## 引言

随着计算机网络的不断快速发展, 数字信息的

数量持续膨胀, 图像已经成为人们获取资源信息的主要手段之一。但是, 图像数据在以几何级数增加, 那么如何建立起高效的图像描述以及检索机制, 俨然已经成了当今国内外研究的焦点。

基于词袋模型<sup>[1]</sup>的图像检索机制是近几年来图像检索领域的主流方法, 该方法将大量训练图像特征(通常是 SIFT 特征)通过 k-means 聚类映射为



收稿日期: 2015-10-13      修回日期: 2016-01-06;  
基金项目: 国家自然科学基金(61104213, 61573168),  
江苏省产学研前瞻性联合研究项目(BY2015019-15);  
作者简介: 陈莹(1976-), 女, 浙江丽水, 博士, 教授,  
研究方向为信息融合、模式识别。

<http://www.china-simulation.com>

• 2353 •

视觉单词的集合, 构成视觉单词词典。然后将测试图像的特征逐个匹配量化到视觉单词词典中, 得到了图像的视觉单词直方图。由于将大量特征量化到视觉单词这个过程时间耗损过大, Nistér 等<sup>[2]</sup>提出了利用分层 k-means 聚类的方法生成视觉单词词汇树, 有效解决了搜索非层次化单词带来的量化过程太慢的问题, 生成的视觉单词由 TF-IDF 模型(词频-逆文档频率)加权<sup>[3]</sup>。Jégou 等<sup>[4]</sup>提出了应用汉明嵌入和弱几何一致性约束技术提高 SIFT 特征与视觉单词的匹配准确率, 从而提高图像检索的准确率。由于视觉单词词汇树的方法忽略了视觉单词的空间位置信息, Qin 等<sup>[5]</sup>构建了上下文视觉单词, 增强视觉单词的区分度, 在分类中得到不错的效果。Tang 等<sup>[6]</sup>利用视觉单词的上下文相似度, 构建了同义词词典来提高图像检索的查全率, 但该方法计算复杂度大, 不适合大规模数据库。Wang 等<sup>[7]</sup>提出了利用 SIFT 特征点在词汇树空间的上下文统计信息, 增强 SIFT 特征的区分性。朱道广等人<sup>[8]</sup>提出利用 SIFT 特征点在词汇树空间相对位置方向信息作为加权系数引入匹配中, 但是使用的位置信息均对视觉单词的区分力比较弱。

在此基础上, 本文将视觉单词的位置信息作为权值引入到视觉单词的匹配中, 细化了特征之间的匹配得分。首先利用分层 k-means 聚类生成视觉单词词汇树, 将待测试图像的特征量化成视觉单词, 再由 TF-IDF 模型进行加权, 在计算图片相似度的时候将视觉单词的位置信息作为权值引入。实验结果表明, 本文方法有效提高了图像检索的精确度。

## 1 基于词汇树的图像检索

### 1.1 词汇树的创建

对图像库中的每幅图像都提取 SIFT 特征, 并且给每幅图像关联一个唯一的标识 ID, 这样就得到了一个特征集合  $F = \{f(i)\}$  以及相应的图像 ID 集合  $\text{imgID} = \{\text{id}(i)\}$ 。然后对特征集合  $F$  进行分层

k-means 聚类。初始时, 在词汇树的第一层上对特征集合  $F$  进行第一次 k-means 聚类, 计算出每个聚类的中心向量  $C_i$ 。类似的, 对新产生的每个聚类再用 k-means 聚类聚成  $k$  个簇集, 不断地重复上述操作直到树的深度达到预先设定的  $L$  值, 每个簇集定义为一个视觉单词, 得到视觉单词词汇树。

### 1.2 图像的权值向量表示

在文本检索中, 单词与文献的相关程度可以用单词的权值来形容, 这个原理同样可以应用到图像检索中来。 $d_j$  表示一幅图像,  $F_i$  表示视觉单词, 用  $t = \sum_{i=1}^L k^l$  表示视觉单词的个数,  $w_{i,j}$  则表示图像  $d_j$  中视觉单词  $F_i$  的权值, 即视觉单词  $F_i$  与图像  $d_j$  的相关程度。所以, 图像  $d_j$  可以用视觉单词的权值向量来表示:

$$d_j = [w_{1,j}, w_{2,j}, \dots, w_{t,j}] \quad (1)$$

那么, 图像库中的所有图像可以用如下矩阵表示:

$$w = \begin{bmatrix} w_{1,1} & w_{2,1} & \dots & w_{t,1} \\ w_{1,2} & w_{2,2} & \dots & w_{t,2} \\ \vdots & \vdots & \dots & \vdots \\ w_{1,N} & w_{2,N} & \dots & w_{t,N} \end{bmatrix} \quad (2)$$

其中权值是仿照 TF-IDF 的原理定义的, 假设  $m_{i,j}$  表示视觉单词  $F_i$  在图像  $d_j$  中出现的次数, 它可以用来衡量视觉单词用来描述图像的好坏程度。假设  $N$  表示图像库中图像的总数,  $n_i$  表示包含视觉单词  $F_i$  的图像数目。逆文档频率定义为  $\text{idf}_i = \lg(N/n_i)$ , 它表示的含义是视觉单词对于区分相似图像和不相似图像的作用大小。所以, 视觉单词  $F_i$  与图像  $d_j$  之间的相关程度  $w_{i,j}$  可以定义为:

$$w_{i,j} = m_{i,j} \times \lg \frac{N}{n_i} \quad (3)$$

### 1.3 相似度计算

通过上述的权值计算, 在图像库中随机选择第  $q(1 \leq q \leq N)$  幅图像作为查询图像, 在不考虑位置信息的情况下, 它与图像库中图像  $d$  的相似度计算公式如下:

$$\text{sim}(q, d) = \sqrt{\sum_{i=1}^t (w_{i,q} - w_{i,d})^2} \quad (4)$$

根据上述公式计算图像之间的差距, 然后依据差距大小进行排序, 合理设置阈值, 差距小于阈值的图像作为检索结果显示。

## 2 基于空间位置信息的加权词汇树

上一节描述的方法充分考虑了视觉单词与图像之间的联系紧密度, 但是却忽略了 SIFT 特征点自身的位置信息。所以考虑到 SIFT 特征并不只有 128 维的描述向量, 包括方向信息、尺度信息和图像空间位置信息, 同时 SIFT 特征点在图像邻域空间中包含了比较丰富的上下文空间信息, 合理地利用这些上下文空间位置信息将有助于进一步提高 SIFT 特征的区分力, 即视觉单词的区分力, 进而提高视觉单词间匹配的准确度。

因此, 本节在上一节的基础上提出一种更为详细的将视觉单词的相互位置影响作为加权系数引入匹配中的方法。此前, 朱道广<sup>[8]</sup>等人也提出过类似方法, 但是他们采用的是依据 SIFT 点的上下文区域构建对应的空间上下文视觉单词分布直方图, 来提高 SIFT 特征的区分度, 然后进行特征匹配。而本文方法则是将 SIFT 特征点之间的空间位置信息转化成视觉单词之间的空间位置信息, 然后将视觉单词位置信息相似度作为对视觉单词间匹配得分的加权系数, 加强了视觉单词的区分力, 细化了相似度计算, 提高匹配精度。

### 2.1 视觉单词空间位置权值计算

在文本检索中, 一个单词的位置信息是指位于该单词周围的文本片段, 比如包含该单词的一条语句或者段落。而一个 SIFT 特征点  $f(i) = \{x(i), u(i), s(i), \theta(i)\}$ ,  $x(i)$  代表 128 维描述向量,  $u(i)$  表示特征点所在图像空间的坐标,  $s(i)$  是指该特征的尺度,  $\theta(i)$  表示主方向。因为在图像中没有类似文本中的语法边界, 所以本文定义图像上一个 SIFT 特征点的空间边界是以  $u(i)$  为中心的半径大小为

$r_c \times s(i)$  的一个圆形区域, 记为  $SCR_{f(i)}$ , 其中  $r_c$  为位置空间的尺度系数, 用于控制空间上下文区域的大小, 可设置为大于 1 的一个常数。由于  $SCR_{f(i)}$  的半径与  $f(i)$  的尺度  $s(i)$  成正比, 因此位置空间区域的相对大小对图像目标的尺度缩放具有一定的鲁棒性。如图 1 所示, 圆点表示 SIFT 特征, 矩形表示每类的聚类中心, 对应于一个视觉单词, 图像中的 SIFT 特征被聚成  $t$  类(图 1 中  $t$  为 3)。  $f_k(1)$ 、 $f_k(2)$  和  $f_k(4)$  对应的视觉单词  $F_k$ ,  $f_m(3)$  和  $f_m(5)$  对应视觉单词  $F_m$ , 其中本文设置尺度系数为 1.5 时, SIFT 特征点  $f_m(3)$  和  $f_m(5)$  为中心分别对应着半径不同的圆形区域。

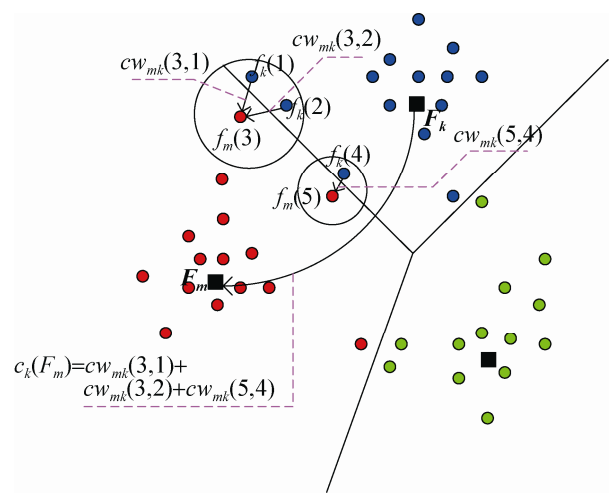


图 1 位置影响示意图  
Fig. 1 Schematic diagram of position effect

考虑到  $SCR_{f(i)}$  内其余 SIFT 点与  $f(i)$  的距离有远有近, 那么其对  $f(i)$  的影响程度也存在不同, 即在  $f(i)$  的上下文特性中的权重也是不同的, 与  $f(i)$  距离较近的 SIFT 点对  $f(i)$  的影响程度较大, 反之则影响程度较小。在图 1 中,  $f_k(1)$ 、 $f_k(2)$  在  $f_k(3)$  所形成的圆形区域内, 说明  $f_k(1)$ 、 $f_k(2)$  对  $f_k(3)$  存在距离的影响, 同理  $f_k(4)$  对  $f_k(5)$  也有影响。为此, 当 SIFT 特征点  $f(i)$  对应的视觉单词是  $F_m$ , 表示为  $f_m(i)$ , SIFT 特征点  $f(j)$  对应的视觉单词是  $F_k$ , 表示为  $f_k(j)$  时, 对于 SIFT 特征点  $f_k(j) \in SCR_{f_m(i)}$ , 定义其到  $f_m(i)$  的相对距离为:

$$d_{mk}(i, j) = \frac{\|u_m(i) - u_k(j)\|_2}{r_c \times s(i)} \quad (5)$$

式中:  $u_m(i)$  和  $u_k(j)$  分别为 SIFT 点  $f_m(i)$ ,  $f_k(j)$  的二维坐标向量;  $\|\bullet\|_2$  为向量的 2-范数。

那么可以定义  $f_k(j)$  在  $f_m(i)$  中的空间上下文特性中的权重为:

$$cw_{mk}(i, j) = \exp\left(\frac{-(d_{mk}(i, j))^2}{2}\right) \quad (6)$$

由此可以看出, 与  $f_m(i)$  距离不同的 SIFT 点对  $f_k(j)$  的空间位置的影响不同, 即它们的权重不同。如图 1 所示,  $f_k(1)$  在  $f_m(3)$  的上下文特性中的权重为  $cw_{mk}(3,1)$ ,  $f_k(2)$  在  $f_m(3)$  的上下文特性中的权重则为  $cw_{mk}(3,2)$ , 两者权重由于距离的不同而不相同。由于不同的 SIFT 点分属于不同的视觉单词, 因此可定义, 视觉单词  $F_k$  在视觉单词  $F_m$  中的上下文特性中的权重为:

$$c_k(F_m) = \sum_{f_m(i)} \sum_{\substack{f_k(j) \in SCR_{f_m(i)} \\ f_k(j)}} cw_{mk}(i, j) \quad (7)$$

其中,  $f_m(i)$  指 SIFT 特征点  $f_i$  对应的视觉单词是  $F_m$ ,  $f_k(j)$  指 SIFT 特征点  $f_j$  对应的视觉单词是  $F_k$ 。

对于视觉单词  $F_m$ , 其余视觉单词对它的空间位置影响为:

$$SC(F_m) = [c_1(F_m), c_2(F_m), \dots, c_t(F_m)] \quad (8)$$

那么对于查询图像  $q$  和图像库中的图像  $d$  的同一个视觉单词  $F_m$ , 它们的空间位置信息的相似度可表示为:

$$\begin{aligned} \text{sim}(SC^q(F_m), SC^d(F_m)) = \\ \sqrt{\sum_{i=1}^t (c_i^q(F_m) - c_i^d(F_m))^2} \quad (9) \end{aligned}$$

## 2.2 基于空间位置加权系数的相似度计算

将上述视觉单词的空间位置信息相似度作为加权系数, 最终得到基于空间位置信息的词汇树相似度查询的表达式为:

$$\text{sim}(q, d) = \sqrt{\sum_{i=1}^t \exp(\text{sim}(SC^q(F_i), SC^d(F_i))) \cdot (w_{i,q} - w_{i,d})^2} \quad (10)$$

综上所述, 将空间位置相似度作为加权系数对相匹配的视觉单词间进行加权, 充分利用了 SIFT 特征点之间的位置信息, 这样, 相似度大的特征之间的匹配得分相对较高, 反之匹配得分相对较低, 提高了特征间匹配的精确度, 从而提高了图像检索的准确率。在本文中, 对于匹配得分设置一个阈值, 当匹配得分低于阈值, 把低于阈值的图像作为检索结果显示出来, 高于阈值的图像则作为不相关图像。

## 3 实验结果及分析

本文实验采用的数据库是 UKbench 图像库<sup>[2]</sup>和 Holidays 图像库<sup>[9]</sup>, 训练图像采用的数据库是 ImageNet-V 图像库<sup>[10]</sup>。

### 3.1 数据库及评价指标说明

UKbench 图像库: 该图像库共包含 10200 幅图像, 每 4 幅图像为一组, 一共有 2550 组。每幅图像依次作为查询图像, 检索结果评价指标为前四幅图像的平均召回率, 用 N-S 来表示(最大值为 4)<sup>[2]</sup>; 平均查准率均值, 用 mAP(mean Average Precision)来表示; 平均查全率-查准率均值曲线。相关定义为:

$$\text{查准率} = \frac{\text{检索出的相关图像}}{\text{检索出的全部图像}} \times 100\% \quad (11)$$

$$\text{查全率} = \frac{\text{检索出的相关图像}}{\text{全部相关图像}} \times 100\% \quad (12)$$

Holidays 图像库: 该图像库共包含 1 491 幅图像, 将其中的 500 幅图像作为查询图像, 每幅查询图像至少有 1~2 幅相似图像, 检索结果评价指标为平均查准率均值, 用 mAP 来表示; 平均查全率-查准率均值曲线。

ImageNet-V 图像库: 该图像库包含 50 000 幅图像, 涉及到不同种类, 随机选取 25 000 幅图像作为训练图像, 用于构建词汇树, 该图像库与前两个图像库无交集, 就是待测试的图像不用于构建词汇树。

### 3.2 检索结果分析比较

在本文中, 设定词汇树的深度  $L$  为 7, k-means 聚类的  $k$  设定为 10。图 2 给出了不同查询图像的

检索结果, 图中只将匹配得分最高的前四幅图像作为检索结果显示出来, 前两幅图像选自于 Holidays 图像库, 后三幅图像选自于 UKbench 图像库。

由第 3 节分析可知, 阈值的设置是影响平均查准率均值和平均查全率均值的一个重要的因素, 如果阈值设置得过小, 那么低于阈值的匹配得分就越少, 即检索出的图像数量就越少, 在之前人为判定的相关图像不变的情况下, 导致平均查准率均值提高, 但平均查全率均值却降低; 如果阈值设置得过大, 那么低于阈值的匹配得分就越多, 即检索出的图像数量越多, 那么依旧是在相关图像数量不变的

情况下, 导致平均查准率均值降低, 但平均查全率均值提高。文献[3]采用的方法是将提取出的图像的 SIFT 特征点量化为视觉单词, 在根据 TF-IDF 模型对视觉单词进行加权, 最后直接对图像进行相似度计算。为了分析阈值对于本文方法和文献[3]方法性能的影响, 图 3 给出了对于 UKbench、Holidays 图像库, 阈值对于平均查准率均的影响比较。当阈值不断增大时, 两种方法的平均查准率均值都在不断下降, 但是本文算法的 mAP 始终高于文献[3]算法, 说明加入了空间位置信息对于提高精确率的有效性。

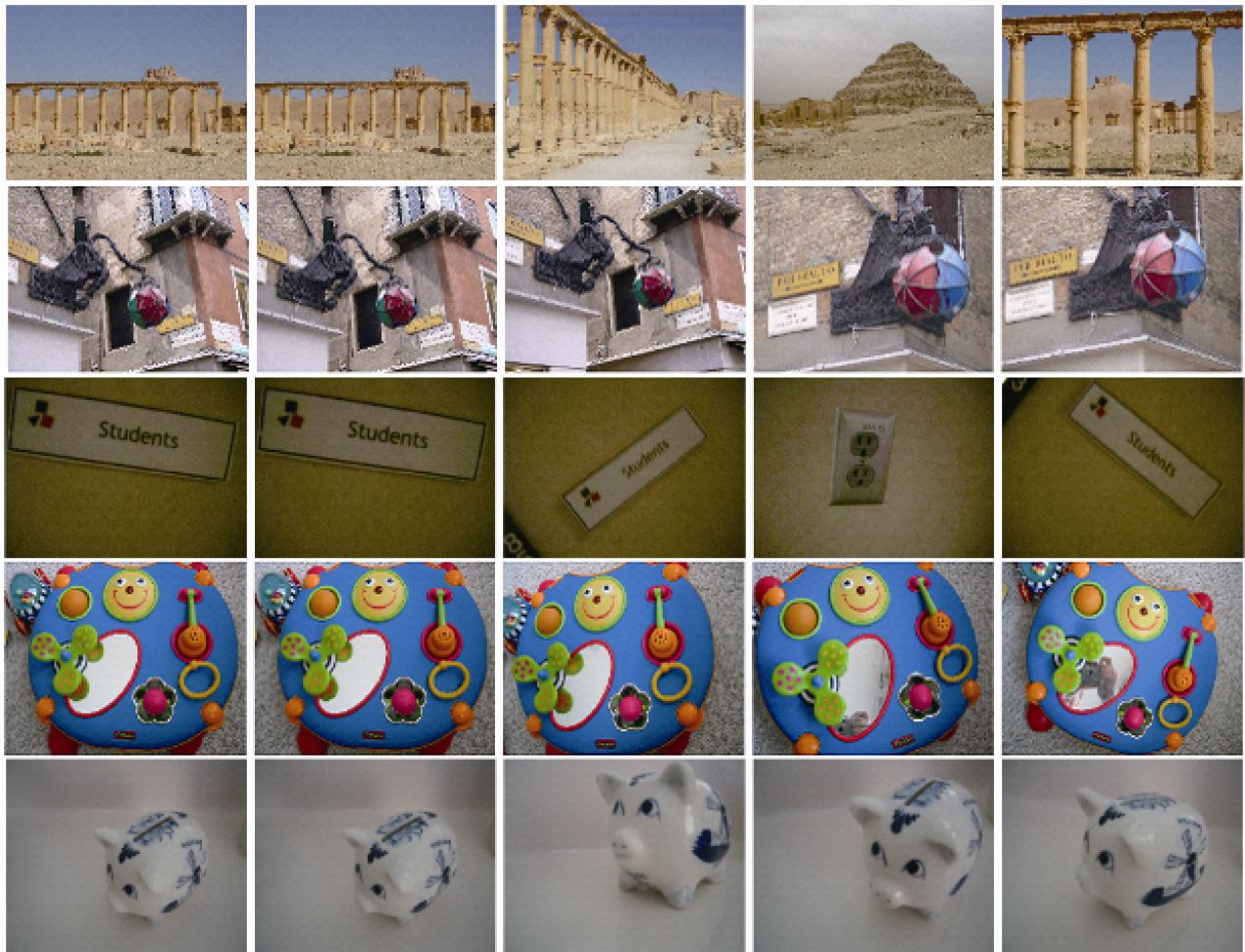
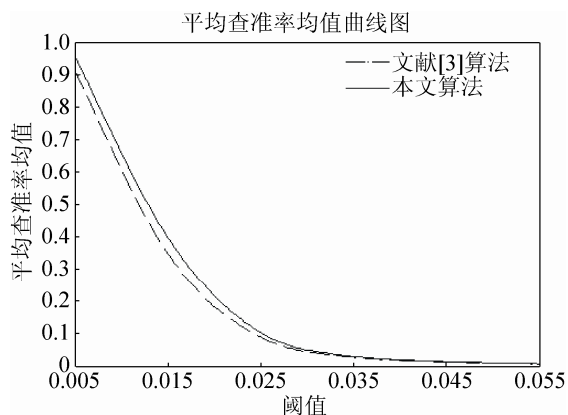


图 2 检索结果前 4 幅图像显示图  
Fig. 2 First four images of the search result

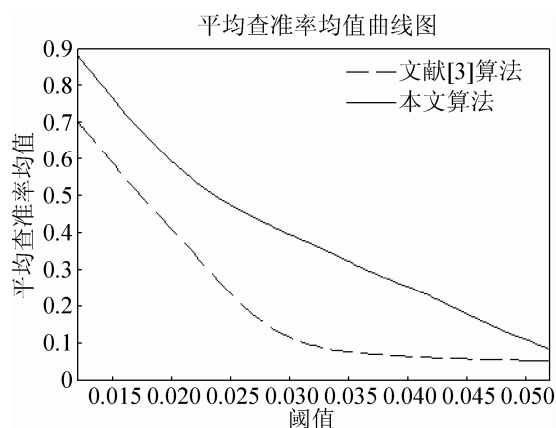
由图 3 可知, 如果想要得到较高的平均查准率均值, 应该选取较小值的阈值, 本文算法中, 当阈值选取过小的时候, 会出现检索结果只有一幅图像

的情况, 即查询图像本身, 这样的检索就无意义了, 所以为了避免这种情况, 阈值不能过小。随着阈值的不断增大, 检索出来的图像不断增多, 在相关图

像数量有限并且固定的情况下,检索出的不相关的图像在不断增加,那么就会导致平均查准率均值不断下降。在本文算法中主要考虑平均查准率均值,所以选取最高的平均查准率均值作为图像库的最佳检索结果,由图可知,对于 UKbench 图像库,最佳的平均查准率均值为 95.84%,对于 Holidays 图像库,最佳的平均查准率均值为 88.22%。



(a) UKbench 图像库

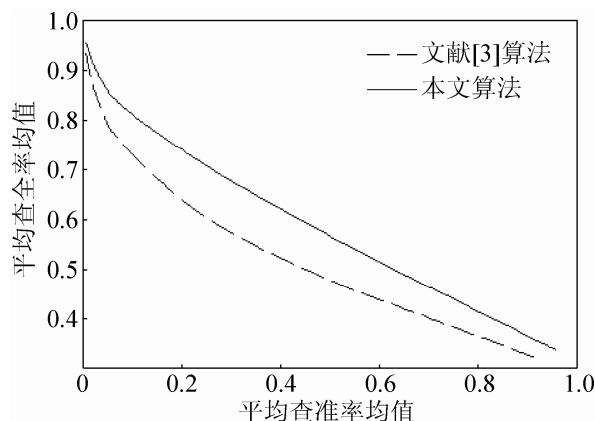


(b) Holidays 图像库

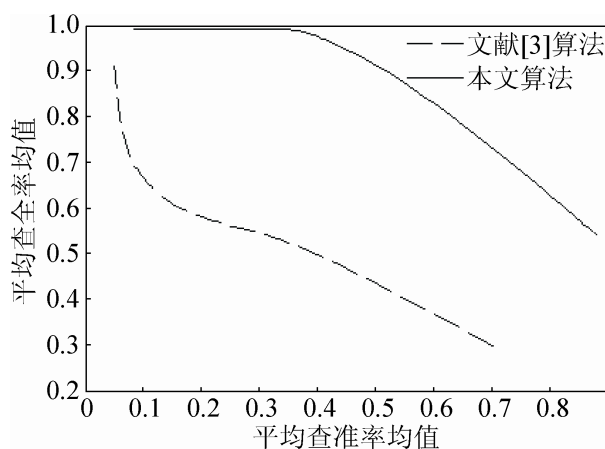
图 3 不同图像库中阈值对 mAP 的影响比较

Fig. 3 Comparison of Thresholds on mAP in Different Image Libraries

图 4 给出了 UKbench、Holidays 图像库的平均查准率-查全率均值的曲线比较,由图可以看出平均查准率均值与平均查全率均值呈反比关系,平均查准率均值提高,则平均查全率均值下降。从图可以看出,本文算法无论在平均查准率均值还是平均查全率均值上都优于文献[3]算法。



(a) UKbench 图像库



(b) Holidays 图像库

图 4 不同图像库 P-R 曲线图比较

Fig. 4 Comparison of P-R curve in different image libraries

此外,为了验证当图像库的规模越来越大时,本文方法的有效性,实验中依次将 100 k, 200 k, 500 k 干扰图像加入到 UKbench 图像库中,依次对 UKbench 图像库中的图像做查询实验,得到结果如图 5 所示。

由图 5 可知,各个算法的平均查准率均值随着干扰图像数量的增加而不断降低,但在不同规模的干扰集下,本文算法的 mAP 值均高于其余算法,验证了本文算法在大规模图像库中的可行性。

表 1 给出了本文算法与其他算法的具体比较结果,无论对于 UKbench 图像库还是 Holidays 图像库,本文算法的平均查准率均值(mAP)明显高于其他算法,对于 UKbench 图像库,前四幅图像的平均召回率(N-S)也略高于其他算法,而且本文算



法在 UKbench 图像库检索图像的平均时间为 0.4 s, 这说明本文算法的可行性。本文合理利用了 SIFT 特征的空间位置信息, 利用位置的影响因素对视觉单词的匹配进行加权, 细化了特征间的匹配得分, 提高了匹配的精确度, 因此 mAP 高于其他算法, 说明本文算法对于提高检索精确度的有效性。

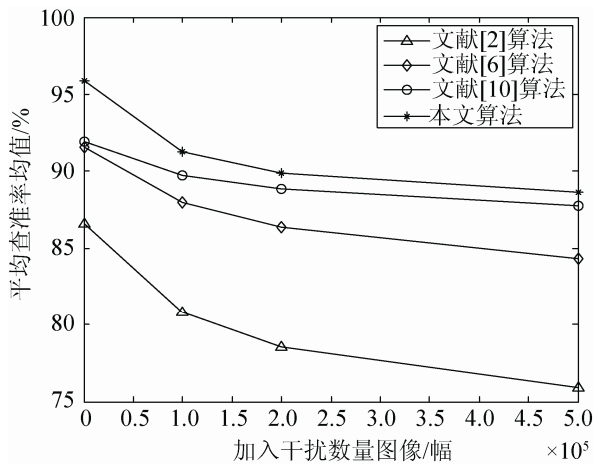


图 5 大规模图像中各算法性能比较

Fig. 5 Comparison of different algorithm performance in large-scale images

表 1 本文算法与其他算法性能比较

Tab. 1 Comparison of different algorithm performance between this algorithm and other algorithms

算法	UKbench		Holidays
	N-S	mAP(%)	mAP(%)
Nistér 等 <sup>[2]</sup>	3.19	76.00	n/a
Jégou 等 <sup>[9]</sup>	3.42	87.80	81.30
陈赟等 <sup>[3]</sup>	n/a	91.27	70.38
Wang 等 <sup>[7]</sup>	3.56	91.70	78.00
朱道广等 <sup>[8]</sup>	n/a	91.95	n/a
本文算法	3.62	95.84	88.22

## 4 结论

本文首先把图像转换成视觉单词的权值向量表示模型, 利用词汇树来解决非层次化视觉单词所带来的量化过程太慢的问题, 从而加快特征的量化过程, 这样图像之间的匹配就转换成视觉单词的权值向量之间的匹配。然而针对 SIFT 特征忽略了空间位置信息这一点, 本文提出了把视觉单词之间的位置影响作为匹配加权系数的算法。由于距离不同

的 SIFT 点之间的相互位置影响各不相同, 所以本文根据 SIFT 点之间的距离远近生成 SIFT 点之间的相互位置影响信息, 由于不同的 SIFT 点分属于不同的视觉单词, 所以将 SIFT 点的位置信息聚类成视觉单词的位置影响, 利用视觉单词位置的影响因素对视觉单词的匹配进行加权, 这样就将视觉单词的空间位置影响也列入考虑范围之中, 使得视觉单词的匹配更加细化。实验表明, 本文算法可以有效提高图像检索的精确度, 并且对于大规模的数据集, 本算法仍具有良好的适用性。

## 参考文献:

- [1] Kesorn K, Poslad S. An enhanced Bag-of-Visual word vector space model to represent visual content in athletics images [J]. IEEE Trans on Multimedia(S1520-9210), 2012, 14(1): 211-222.
- [2] Nistér D, Stewénius H. Scalable recognition with a vocabulary tree [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, June 17-22, 2006. Piscataway, NJ, USA: IEEE, 2006: 2161-2168.
- [3] 陈赟, 沈一帆. 基于词汇树的图片搜索 [J]. 计算机工程, 2010, 36(6): 189-191.(Chen Yun, Shen Yifan. Image search based on vocabulary tree[J]. Computer Engineering, 2010, 36(6): 189-191.)
- [4] Jégou H, Douze M, Schmid C. Improving bag-of-feature for large scale image search [J]. Int'l Journal of Computer Vision(S1573-1405), 2010, 87(3): 316-336.
- [5] Qin Jianzhao, Yung N H C. Scene categorization via contextual visual words [J]. Pattern Recognition ISSN 号 (S0031-3203), 2010, 43(5): 1874-1888.
- [6] Tang Wenbin, Cai Rui, Li Zhiwei, et al. Contextual synonym dictionary for visual object retrieval [C]// Proc of the 19th International Conference on Multi Media. Scottsdale, USA, Nov 28- Dec 1, 2011. Scottsdale, USA: ACM Multimedia, 2011: 503-512.
- [7] Wang Xiaoyu, Yang Ming, Cour T, et al. Contextual weighting for vocabulary tree based image retrieval [C]// Proc of the 13th IEEE International Conference on Computer Vision, Barcelona, Spain, Nov 6-13, 2011. Piscataway, NJ, USA: IEEE, 2011: 209-216.

(下转第 2372 页)