# 图像重排序中与查询相关的图像相似性度量①

王　黎　帅建梅　(中国科学技术大学 自动化系 安徽 合肥 230027)

**摘　要**: 现今的图像搜索引擎主要利用图像周围文本信息为图像排序，根据图像内容重排序可以进一步提高搜索性能。图像相似性的度量对重排序算法的性能至关重要。然而已有的相似性度量没有考虑针对不同的查询，图像的相似性应该不同。提出一种与查询相关的相似性度量方法，将基于全局特征的相似性，基于局部特征的相似性,以及视觉单词同时出现率融合到一个迭代算法中,挖掘出与查询相关的图像信息，计算图像相似性。在 Bing 图像搜索引擎上的实验结果证明本文提出的相似性度量方法优于基于全局特征，局部特征，或它们线性组合的相似性。

**关键词**: 图像重排序; 与查询相关的图像相似性; 视觉单词同时出现率

## Query Dependent Visual Similarity in Image Search Reranking

WANG Li, SHUAI Jian-Mei

(Univerity of Science and Technology of China, Hefei 230027, China)

**Abstract:** Recently image search engines mainly base on associated textual information. Image reranking is an effective approach to refine the initial text-based search result by mining the visual information of the returned images. And the estimation of visual similarity is the fundamental factor in reranking methods. However, the existing similarity measures are independent of the query. This paper proposes a query dependent method by incorporating the global visual similarity, local visual similarity and visual word co-occurrence into an iterative propagation framework. Then it embed the query dependent similarity into random walk rereanking method. The experiments on a collected Live Image dataset demonstrate that the proposed query dependent similarity outperforms the global, local similarity and their linear combination.

**Keywords:** image reranking; query dependent visual similarity; visual words co-occurrence

现今常用的图像搜索引擎，如 Google, 百度, Bing, Yahoo!，主要利用图像周围的文本信息实现图像的搜索和排序，没有考虑图像本身的内容。为了解决基于文本搜索的缺陷，图像搜索的重排序引起很多研究者的关注。重排序，是指在原始搜索结果的基础上，通过挖掘数据内在关系，或者借鉴外部知识和人工干预，对原始搜索结果进行重新排序，使新的序列更能满足用户搜索需求。

目前针对图像的重排序研究重点是从原始搜索结果中挖掘图像内容信息，进而调整原始结果的排列顺序。大致有三类方法：基于分类 [7]，基于聚类[6-8]，以及基于图理论[9-12]。PRF(Pseudo Relevance Feedback) [7]假设大部分正确结果分布在原始搜索结果的前面，错误结果分布在原始搜索结果的后面。依照这个假设，排在原始搜索结果前面的样本被看作正样本，后面的被看作负样本，利用得到的正负样本训练排序函数，然后对原始结果重新排序。Hsu 等人[8]提出信息瓶颈理论(Information Bottleneck)，对原始的搜索结果进行聚类，得到每个类的条件概率(Conditional Probability)，概率大的类被看作与查询相关的结果，概率小的类被看成原始搜索结果中的噪声。重排序分成两步：首先根据条件概率对类排序，

① 基金项目:国家 863 计划资助项目(2006AA01Z449)
收稿时间:2010-03-17;收到修改稿时间:2010-04-23

再对每个类中的结果进行排序。文献[9]根据平滑假设把图理论引入重排序问题，形成随机游走算法(Random Walk)通过不断迭代，更新每个样本相关性得分，最终实现重排序。

以上所有重排序模型中，图像相似性度量是至关重要的。当前计算图像相似性的方法存在两个主要问题。第一个问题是计算相似性时，只考虑全局或局部特征。然而有的查询对应的图像，计算图像相似性时适合使用全局特征，有的却更适于局部特征。比如，文献[10]作者发现对查询“horse”，“bikes”使用全局特征得到的图像相似性更符合用户需求；而查询“zebra”，“car”和“guitar”对应的图像相似性更适合用局部特征计算。在实际应用中，任意给定一个查询，计算图像相似性时很难判定全局特征，局部特征哪个更有效果。因此，如果能同时考虑全局特征和局部特征将会在一定程度上解决这个问题。另外一个问题是目前计算图像相似性的方法都没有考虑查询。也就是说，不管查询是什么，两幅图像的相似性是定值。然而，我们认为判断图像相似性应该依赖查询。如图 1，如果查询是“车”我们认为两幅图像相似，如果查询是“人”或“马路”，这两幅图像不相似。也就是说，两幅图像的相似程度与查询相关。因此，在计算图像相似性时应该考虑查询。

图 1 图像相似性与查询相关

针对传统方法的具体问题，本文提出一种与查询相关的相似性度量方法，把基于全局特征的图像相似性，基于局部特征的图像相似性，以及视觉单词(visual words)同时出现率融合到一个迭代算法中，挖掘出与查询相关的图像部分，计算相似性。

## 1 查询相关的图像相似性

通常，我们通过计算图像视觉特征的相似性来估计图像的相似性。图像视觉特征包含全局特征(如颜色，纹理)和局部特征(如 SIFT(Scale Invariant Feature Transform) )。首先介绍目前普遍采用的基于全局或局部特征计算图像相似性的方法。然后提出与查询相关的相似性度量方法。

### 1.1 基于全局特征的图像相似性

首先提取图像内容的全局特征，然后通过计算相应特征的距离得到图像相似性。通常采用的全局特征包括：颜色(如颜色矩[3]，颜色直方图[1])，边缘(如边缘分布直方图[4])，纹理(如小波变换[2])。将这些特征聚合组成一个特征向量来表示一幅图像，从图像 $I_i$ 提取得到的全局特征串接组成向量 $x_i$。

基于全局特征，采用高斯核计算两幅图像的相似性，具体公式如下

$$G_{ij}=\exp\{-\frac{d(x_i,x_j)^2}{2s^2}\} \quad (1)$$

其中，$x_i$，$x_j$ 分别为从图像，$I_i$，$I_j$ 提取的全局特征向量，$d(x_i,x_j)$ 表示向量 $x_i$，$x_j$ 的欧式距离，$s$ 是高斯核的参数。

根据查询，利用图像周围的文本信息，返回初始结果包含图像 $\{I_1,I_2,...,I_N\}$，用矩阵 $G=[G_{ij}]_{N*N}$ 表示根据全局特征得到这些图像间的相似性。

### 1.2 基于局部特征的图像相似性

基于局部特征的图像相似性是通过考虑图像局部块的相似性来估算图像的相似程度。

计算文本相似性的一种方法是比较文本中的单词在字典上的分布[14,15](如果文本包含字典中的某个字，该字对应的出现次数加一，然后归一化得到分布情况)。受此启发，为了计算基于局部特征的图像相似性，我们把图像看作是视觉单词(visual words)组成的。通过计算图像在视觉字典(visual code)上分布的相似性得到图像相似性。首先，从图片库中提取局部特征描述子。然后，把所有的局部特征描述子进行聚类，每个聚类中心看作一个视觉单词(visual words)，所有的视觉单词形成一个视觉字典(visual code)。最后根据视觉字典，把图像量化为视觉字典上的分布，具体公式如下

$$z_i=[z_{i1},z_{i2},...,z_{iH}]$$
$$z_{ij}=\frac{n_{ij}}{\sum_{h=1}^{H}n_{ih}} \quad (2)$$

$H$ 表示视觉字典中视觉单词的个数，$n_{ij}$ 表示视觉单词 $v_j$ 在图像 $I_i$ 中出现的次数，$z_{ij}$ 表示视觉单词 $v_j$ 在图像 $I_i$ 中出现的频率，$z_i$ 表示图像 $I_i$ 在视觉字典上的

视觉分布。也就是说从图像局部特征层面可以用 $z_i$ 来表示图像 $I_i$。矩阵 $Z=[z_{ij}]_{N*H}$ 其中每一行表示对应的图像在视觉字典上的视觉分布。

相应的，计算图像 $I_i$ 与图像 $I_j$ 基于局部特征的图像相似性 $L_{ij}$，公式如下

$$L_{ij}=z_i z_j^T \tag{3}$$

用矩阵 $L=[L_{ij}]_{N*N}$ 表示根据局部特征得到图像 $\{I_1,I_2,...,I_N\}$ 间的相似性。

### 1.3 查询相关的全局局部相似性传播

基于全局和局部特征的图像相似性从不同粒度上反映了图像视觉的相似性，把两者结合起来可以更全面地反映图像相似性。最直接的方法就是线性结合，比如

$$W=aG+(1-a)L=aG+(1-a)lZZ^T \tag{4}$$

其中 $a$ 是线性结合因子，$l$ 是衰减系数。

虽然线性结合的方法，同时考虑了图像的全局和局部特征，但是没有考虑查询对图像相似性计算的影响。

为了获得与查询相关的相似性，我们需要挖掘和查询相关的图像信息。根据查询，利用文本信息返回初始图像集合。把一幅图像类比作一篇文档，图像可以被看作一系列视觉单词组合而成，所以我们可以从返回的图像集合中得到很多视觉单词。据观察，和查询相关的图像总是以某种形式彼此相似；而和查询无关的图像，它们的不相似各不相同。因此，我们认为和查询相关的视觉单词在返回的图像集合中有更高的同时出现率，因为它们不仅出现在同一幅图像中，而且经常出现在相似图像中。相反地，与查询无关的视觉单词经常被孤立。用视觉单词同时出现的概率作为该视觉单词的权重，得到的图像相似性在一定程度上与查询相关。基于这样的基本假设，我们提出一种与查询相关的相似性度量方法，把基于全局特征的相似性，基于局部特征的相似性，以及视觉单词同时出现率融合到一个迭代算法。

在计算视觉单词同时出现率时，我们采用[13]中提出的方法。根据查询，返回的初始图像集 $\{I_1,I_2,...,I_N\}$，$N$ 表示返回的图像数目。用 $K_{ij}$ 表示视觉单词 $v_i$ 和 $v_j$ 同时出现的概率

$$K_{ij}=\exp\left\{-\frac{\max(\log c(v_i),\log c(v_j))-\log c(v_{i,}v_j))}{\log N-\min(\log c(v_i)-\log c(v_j))}\right\} \tag{5}$$

其中 $c(v_i)$ 表示包含视觉单词 $v_i$ 的图像数目，$c(v_j)$ 表示包含视觉单词 $v_j$ 的图像数目，$c(v_i,v_j)$ 表示同时包含视觉单词 $v_i$，$v_j$ 的图像数目。用矩阵 $K=[K_{ij}]_{H*H}$ 表示视觉单词同时出现概率。如果认为同时出现概率大的视觉单词相似，那么矩阵 $K$ 还可以看作视觉单词相似性矩阵。

下面再重新考虑基于局部特征的图像相似性矩阵 L。在计算 $L_{ij}$ 时，我们默认视觉单词彼此相互独立。然而，从返回的图像搜索结果中，视觉单词同时出现的概率不同。也就是说视觉单词彼此并不完全独立。为了更好地估算基于局部特征的图像相似性，我们应该考虑视觉单词间的关系，公式如下 $L_M=ZKZ^T$

相应地，公式(4)重写作

$$W=aG+(1-a)L_M=aG+(1-a)lZKZ^T \tag{6}$$

由于视觉单词同时出现概率矩阵 $K$ 是从返回的图像结果中挖掘的和查询相关的图像信息，所以通过公式(6)得到的图像相似性，从某种程度上讲，和查询相关。

通过公式(5)得到的视觉单词同时出现概率矩阵 $K$，只考虑了同一幅图像中视觉单词同时出现的情况，并没有考虑相似图像中视觉单词同时出现的情况。因为 $K_{ij}$ 应该反映视觉单词 $v_i$ 和视觉单词 $v_j$ 在返回的图像集合中的依赖关系，所以我们认为计算 $K_{ij}$ 时不仅应该考虑同一幅图像中 $v_i$，$v_j$ 的关系，而且应该考虑在返回的图像集合中不同图像间 $v_i$，$v_j$ 的关系。

与公式(6)相似，我们把视觉单词同时出现概率矩阵 $K$ 优化如下

$$\hat{K}=bK+(1-b)lZ^TWZ \tag{7}$$

其中矩阵 $Z$ 的行表示图像在视觉词典的分布，列表示视觉单词在返回的图像集合上的分布。$Z^TZ$ 表示视觉单词在返回的图像集合上的相似程度。$Z^TWZ$ 通过考虑图像的相似性，调整视觉单词同时出现的概率。由于和查询相关的图像在某种意义上相似，因此在这些图像中的视觉单词同时出现的频率由于图像间的相似性而增强。

得到 $\hat{K}$ 后，我们再将其代入公式(6)，更新 $W$。然后再将 $W$ 代入公式(7)更新 $\hat{K}$。这样循环往复，形成一个迭代算法，公式如下

$$\begin{aligned}\hat{G}&=aG+(1-a)lZKZ^T\\ \hat{K}&=bK+(1-b)lZ^T\hat{G}Z\end{aligned} \tag{8}$$

为了统一表达，我们使用$\hat{G}$代替$W$，表示图像相似矩阵，$\hat{K}$表示视觉单词同时出现概率矩阵，也可以看作视觉单词相似矩阵。通过图像相似性的传播，视觉单词同时出现的概率被增强；通过考虑视觉单词同时出现的概率，图像相似性与查询相关。公式(8)代表两个不同层次(图像，视觉单词)上，相似性的相互迭代过程，这个迭代算法的收敛性证明类似于[16]。最后得到的图像相似矩阵$\hat{G}$和查询相关，并且同时考虑了全局和局部特征。

## 2 随机游走算法

为了评价本文提出的图像相似性度量方法的性能，我们把得到的和查询相关的相似性矩阵应用到目前普遍采用的图像重排序——随机游走算法[9]中。

给定查询，利用图像周围的文本信息，返回的初始结果包含图像$\{I_1, I_2, ..., I_N\}$。把每幅图像看作一个节点，节点间连线的权重用相应图像的相似度来表示，这样构成一个无向图。对图像$\{I_1, I_2, ..., I_N\}$重排序的过程，可以看作是在该无向图上随机游走的过程,具体公式如下

$$r = m\Pr + (1-m)v \quad (9)$$

其中,状态转移概率矩阵$P$是图像相似矩阵$W$通过列归一化得到，向量$r$表示图像$\{I_1, I_2, ..., I_N\}$对应的稳态概率，阻尼向量$v$用图像$\{I_1, I_2, ..., I_N\}$对应的初始排序来表示，$m$表示权衡参数。根据公式(9)随机游走算法不断迭代,最终达到稳定状态，最后根据稳态概率向量$r$的降序对图像重新排序。

## 3 实验分析

### 3.1 实验数据

我们从 Bing Image 搜索引擎查询日志确定 26 个查询，并且为每个查询收集最多 1000 幅图像，最后整理得到包含 24036 幅图像的数据库。对数据库中每幅图像，人工标注它与对应查询的相关程度。本文规定相关性有四个等级：非常相关，相关，有一点相关，完全不相关。为了保证标注的准确性，我们组织三个人同时标注。有异议时，相关性以中间等级为准。

对每幅图像提取 428 维全局特征，其中包括 225 维颜色矩(窗的大小 5*5)，75 维的边缘分布直方图，128 维小波纹理。局部特征采用普遍使用的 SIFT(Scale Invariant Feature Transform)尺度不变特征提取方法。先对每幅图像检测其关键点，然后在这些关键点的局部区域提取 SIFT 特征。

### 3.2 实验结果

为了检验本文提出的图像相似性度量方法的有效性，我们将它与其它图像相似性度量方法进行比较，包括基于全局特征的图像相似性，基于局部特征的图像相似性，基于全局和局部特征线性组合的图像相似性,分别标记为“QueryAware”，“Local”，“Global”，“LinearComb”。根据图像周围的文本信息得到的初始图像排序看作基准，标记为“Text”。然后运用随机游走算法对图像重排序，其中的相似矩阵分别使用以上四种不同的方法得到。实验结果如图 2，从中我们可以发现同时考虑全局和局部特征(如“LinearComb”“QueryAware”)性能优于只使用单一特征(如“Local”，“Global”)。另外，“QueryAware”性能优于“LinearComb”。我们认为主要原因是“QueryAware”考虑了查询相关的图像信息。

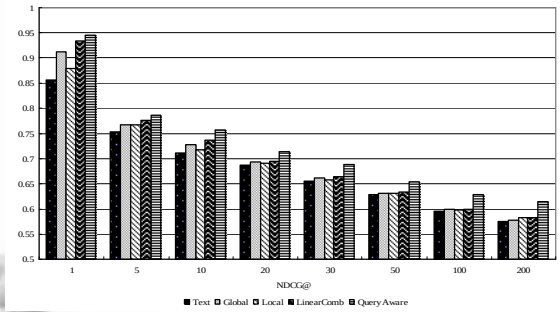


图 2 使用不同的相似性度量方法重排序结果

## 4 结论

本文提出一种与查询相关的相似性度量方法，把基于全局特征的相似性，基于局部特征的相似性，以及视觉单词(visual words)同时出现率融合到一个迭代算法中，挖掘出与查询相关的图像内容计算相似性。通过该算法计算的图像相似性同时考虑了全局特征和局部特征，并且与查询相关。在 Bing 图像搜索引擎上的实验结果证明我们的相似性度量方法有效。

## 参考文献

1 Ma WY, Zhang HJ. Benchmarking of image features

for content-based retrieval. Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems &Computers, Nov 1998. 253–257.

2 Chang T, Kuo CC. Texture analysis and classi cation with tree-structured wavelet transform. In IEEE Transactions on Image Processing, 1993.429–441.

3 Huang J, Kumar SR, Mitra M, Zhu WJ, Zabih R. Image indexing using color correlograms. CVPR, 1997.762.

4 Park DK, Jeon Y. S, Won CS. Efficient use of local edge histogram descriptor. ACM Multimedia, 2000. 51–54.

5 Lowe DG. Object recognition from local scale-invariant features. ICCV, 1999,2:1150–1157.

6 Ben-Haim N, Babenko B, Belongie S. Improving web-based image search via content based clustering. SLAM, New York, 2006.

7 Yan R, Hauptmann E, Jin R. Multimedia search with pseudo-relevance feedback. CIVR, 2003.238–247.

8 Hsu WH, Kennedy LS, Chang SF. Video search reranking via information bottleneck principle. ACM Multimedia, 2006.35–44.

9 Hsu WH, Kennedy LS, Chang SF. Video search reranking through random walk over document-level context graph.ACM Multimedia, 2007.971–980.

10 Zitouni H, Sevil S, Ozkan D, Duygulu P. Re-ranking of web image search results using a graph algorithm. ICPR,Dec. 2008.1–4.

11 Tian X, Yang L, Wang J, Yang Y, Wu X, Hua XS. Bayesian video search reranking. ACM Multi-media, 2008.131–140.

12 Jing Y. Baluja S. Visualrank: Applying pagerank to large-scale image search. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008,30: 1877–1890.

13 Cilibrasi RL. Vitanyi PMB. The google similarity distance. IEEE Transactions on Knowledge and Data Engineering, 2007,19:370–383.

14 Yang J, Jiang YG, Hauptmann AG, Ngo CW. Evaluating bag-of-visual-words representations in scene classi cation. MIR, 2007.197–206.

15 Sparck K Jones. A statistical interpretation of term specificity and its application in retrieval. Document Retrieval Systems, 1988.132–142.

16 Wang XJ, Ma WY, Xue GR, Li X. Multi-model similarity propagation and its application for web image retrieval. ACM Multimedia, 2004.944–951.