图像检索系统中的缩放功能

章进洲
(南京理工大学计算机科学与工程学院 南京 210000)

摘 要 图像检索系统是用户导向的。根据用户意图的不同,检索结果的离散度对用户的体验有着不同的影响。一些情况下,用户希望得到的是"类而不同"的结果。当前以关键字为基础的检索系统并不能很好地捕捉到用户的意图。因此,新的交互内容——缩放比例被引入检索系统,以消除用户的意图与检索结果离散度之间的隔阂,使用户根据自己的意图直接调整检索的结果。首先得到检索系统返回的图像,之后计算图像间的视觉与语义的相似度,再利用层次聚类得到聚类树,最后通过得到用户直接调节的缩放比例,来控制聚类树展开与否。对于每棵展开的子树,选择在原检索结果中拥有最小索引值的节点作为代表。

关键词 图像检索,相关反馈,离散度,层次聚类
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.9.003

Zoom Feature in Image Retrieval System

ZHANG Jin-zhou
(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210000, China)

Abstract Image retrieval systems are user-oriented. Diversity of retrieval results has different effects on users' experiences depending on their intents. Some users may need those different but similar results, which means higher diversity. Nevertheless current retrieval system which is majorly based on query keywords can hardly capture users' intents directly from their query. Thus, a new interactive element, zoom factor, was introduced into retrieval system to bridge the gap between users' intents and the diversity of retrieval results. This enables users to directly control the diversity of results. We first obtained images returned by retrieval system. And then the visual and semantic distances of each other were computed. Hierarchical clustering was then used to form a clustering tree. And finally we controlled the expansion of a sub-tree with users' directly tune of zoom factor. For each expanded sub-tree, the node with the lowest index in the original results was selected as the representative.

Keywords Image retrieval, Relevance Feedback, Diversity, Hierarchical clustering

1 引言

考虑这样一个场景:你在散步时看到一只美丽的小鸟,对它很感兴趣,想了解更多关于这种鸟的信息,却意识到除了见过一眼以外什么都不了解。你于是求助于图像检索系统,希望找到与它有关的图片,得到关于它的信息。你键入了关键字"鸟",但得到的却几乎是某一种类鸟的图片,但你根本不关心这种鸟。在翻了几页后仍然得不到你需要的信息。于是挫败的你感叹:图像检索真难用。

根据信息论的观点,概率大的事件带来的信息量少。在图像检索系统中,得到的很多类似的图片能给我们带来的有用信息是很少的。为了提高信息量,需要增加检索结果的离散度,即呈现更多"类而不同"的结果。

1.1 离散度与用户意图

上述场景中,用户被暴露在了细节(同一类鸟的许多图片)之中。这时候,用户实际需要的是有更高离散度的结果,即更多"类而不同"的结果来帮助他进行粗略的定位。检索系统"难用"的原因在于,它呈现结果的做法是按与查询相关的可能性大小进行排序,这样能得到相关文档的最大期望数量[1]。而当这样的做法与用户的意图相左或用户无法用关键字很好地描述自己的意图时,则得不到预期的结果,体验下降。

在图像检索系统中,用户的意图一般分为3类[2]:浏览者(browser),没有特定意图,只是在浏览图片;查阅者(surfer),有模糊的意图,随着查找意图不断清晰;搜索者(searcher),有清晰的意图,明白自己希望得到的结果。即搜索者知道自己想得到哪张具体的图像;浏览者只是粗略地查看,并不在乎自己看的是什么内容;而查阅者则明白自己想找某类图像,但不清楚具体是哪一类。

可以看到,上述场景中的用户属于查阅者,而检索系统的默认标准显然对搜索者更有利,对查阅者不利。而检索系统仅根据检索的关键字并无法准确地捕捉到用户的意图,因此就无法准确地为用户提供恰当的离散度,以提高他们的体验。

用户的体验与结果的离散度相关。这里的矛盾在于,检

到稿日期:2014-05-23 返修日期:2014-07-29
章进洲(1990—),男,硕士,主要研究方向为信息检索、图像检索,E-mail:jinvii@gmail.com。

索系统需要为不同的用户提供不同的离散度，而它无法准确地定位用户的需求。

1.2 现状与不足

当前的主流检索系统（如 Google、Bing），都使用了关键字提示，以期更准确地把握住用户检索的实际意图。但就结果的离散度而言，依旧不是十分理想。

文献[4]期望通过用户额外的点击来提高检索结果的准确性。文献[5]则试图利用用户的检索与点击历史来推测用户的意图。尽管人们尝试着从反馈中推测用户意图，但这大多用来提高结果的精度，而非广度，即离散度。

另一方面，也有许多工作致力于为用户提供他们所需的离散结果。文献[7-9]都建立了一个新的标准，该标准同时考虑了准确性与离散度，用于对检索结果进行重新排序；另一种方法则对检索结果进行适当聚类[10]，并选出合适的代表，试图通过隐藏相似的信息来提高离散度。我们认为，这些方法的一个重要弊端是试图以一个统一的标准来服务不同的用户，即没有区分不同用户的意图。

因此，我们设计了新的交互因素——缩放，来获取用户的意图，并以此为用户提供他们所需的离散度。为了减轻用户反馈的负担，提供了“所见即所得”的交互方式。

1.3 地图缩放与检索系统

在研究用户的体验与结果离散度的矛盾中，我们从谷歌地图的交互方式中得到灵感。类似地图，我们为检索系统提供了“缩放”的功能，通过“放大”，用户能得到比当前结果更紧凑的检索结果；通过“缩小”能得到比当前结果更离散的检索结果。

“缩放”的好处是可以动态地调整检索结果的离散度，使其与用户检索的语义在同一个层次上。用户尽管并不能准确地用关键字描述自己的意图，但是却能根据调节时的动态反馈选择最接近自己想法的结果，以此来消除用户意图与结果离散度的隔阂。

现在常见的检索系统，无论是文本检索还是图像检索，都停留在传统的交互方式上，即给出搜索目标，返回搜索结果，用户只能在搜索结果中定位，对于查阅者而言，就只能淹没在细节中。因此，我们强调“缩放”功能在检索系统中的重要性。本文就是在图像检索系统中做这样一个尝试，并介绍如何在检索系统中实现“缩放”的功能。

这样的交互方式使用户的意图可以由“缩放比例”这一概念得到，再用它直接调节检索结果的离散度，对用户进行反馈，若用户不满意，可以继续调整，如此反复。该方法的核心思想是直接以“缩放比例”来获取用户的意图，调节离散度算法的参数，用户接受反馈后可继续调整。

2 相关工作

本节将介绍获取用户意图及提高图像检索结果离散度两方面的相关工作。

2.1 获取用户意图

用户意图的获取也称相关反馈（Relevance Feedback），它作为检索系统的特性存在，为用户提供更好的体验。其一般分为 3 类：显示反馈、隐式反馈及盲式反馈[1]。

文献[3]指出了基于内容的图像检索系统（CBIR）中用户高层的概念与底层特征的差距，并利用相关反馈尝试解决这些问题。它让用户对检索返回的结果进行相关性的评价，并根据这些评价动态更新对应图像的底层特征的权重，依据这些新的权重返回新的结果。

文献[18]则将粒子群算法与相关反馈结合，根据用户的反馈引入 w 自适应调整和 Beta 自适应变异的粒子群算法，动态调整图像底层特征的权重，以提高检索的精度。

相关反馈对用户的交互敏感，用户通常在检索的过程中不会过多地提供这些反馈。文献[4]只利用用户额外的一次点击来获取用户的检索意图。它预先定义了能反映用户意图的分类，让用户从其中点击选择。根据用户选择的图片及类别，扩展用户的查询关键字，以此重新检索得到新的结果，并参考用户选择的图片进行重新排序，以提高检索的准确性。文献[5]则是利用了用户的反馈历史来提高其体验。关于相关反馈更全面的综述请参考文献[6]。

相关反馈的矛盾在于，在获取更多用户信息的同时尽量减少用户交互负担。我们认为，为用户提供所见即所得的反馈能减少用户的抵触。

2.2 图像检索中的离散度

图像离散度问题涉及了一个基本问题：如何定义图像间的相似度，以反映人类的语义认知。目前这仍是一个难题。在提高离散度方面，主要有 3 种方法：重新排序、聚类、消除近似图像。

重新排序的一般流程是：

（1）给定一个已知序列；

（2）在剩余的其它结果中寻找一个新的目标，使得某个标准达到最优；

（3）将该目标加入到已有的序列中，重复执行步骤（1），直到序列的数量达到要求。

文献[7]将最大边缘相关性（Maximal Marginal Relevance）作为最优化的标准，即一份文档既与查询相关，同时又与之前选择的文档有最小的相似性。而文献[8]则使用与图像关联的主题丰富程度（Topic Richness）作为标准，即在保证每幅图像与查询相关的同时，保证所有图像涉及的主题数最大。而文献[9]则结合平均精度（Average Precision）创建一个平均离散精度（Average Diverse Precision）进行优化。它们都取得了不错的效果。正如引言中提到的，一个单一的标准是无法满足所有用户的，所以尽管这些方法都有一定的成效，却忽视了用户的意图与期望。

另一种方法是聚类。将检索得到的结果进行聚类，并在每一类中选择一幅图像作为该类的代表。这类方法的研究重点在于如何更好地表达图像间的相似性度量，其有很重要的借鉴意义。文献[10]介绍了如何结合不同的图像相似性度量，并动态地为其赋予权重；同时介绍了 Folding、Maxmin、Reciprocal Election 这 3 种不同的聚类方法。文献[11]结合

---

[1] http://en.wikipedia.org/wiki/Relevance_feedback

图像的 GIST 特征与图像在 ImageNet 的近邻，得到了语义信息，构建了图像的相似度量。文献[12]利用图像搜索的视觉、文本及链接信息做层次聚类，以方便用户浏览。

消除近似图像的一个特点是直接消除了近似的图像。它的本质是查找 k 个最近邻，文献[13]对每个数据对使用了成对检测，文献[14]则结合 LSH 进行快速检测。

聚类与消除最近邻的方法的主要问题在于，聚多少类合适？消除到什么程度合适？不同的用户对这些参数的要求往往是不同的。

3 图像检索中的缩放功能

本文的基本思路是选择一个较为合理的图像相似度，对现有图像检索系统返回的结果执行层次聚类，得到一棵聚类树。之后，根据用户指定的“缩放比例”，从树中选择合适的代表，作为检索的结果。它与传统的提供离散度的方法相比，最大的优点是可以动态进行调整，以满足不同用户或同一用户不同时间的需求。

3.1 图像的相似度

图像的相似度度量是离散度的关键。它直接决定了检索结果与用户的体验是否匹配。尽管人眼能很容易地分辨出两幅图像相似与否，但是人们仍然没有找到一个通用的健壮的算法让计算机达到人工的效果。

这里选择 GIST 描述符[15]作为图像特征，它在文献[9]中对于视觉相似度有较好的结果。计算两幅图像特征的 cosine 距离作为其视觉相似度：

$$d_v=1-\frac{v_a v_b'}{\sqrt{(v_a v_a')(v_b v_b')}}$$

其中，v_a，v_b 分别为图像的特征向量。由于图像的视觉相似度还不能很好地反映人眼的直观感受，因此经常需要结合图像间的语义相似度，以期获得更好的结果。利用与图像相关联的标签信息，求取两幅图像标签的 Jaccard Coefficient，得到语义距离为：

$$d_s=1-J(A,B)=1-\frac{|A\cap B|}{|A\cup B|}$$

其中，A，B 分别表示两幅图像的标签集合，$|X|$ 代表集合 X 的元素个数。一个更好的语义距离的选择可以是谷歌距离[16]。

至此，已经得到两幅图像的视觉与语义相似度，最终求取它们的加权和 d 作为两图像的距离。

$$d=wd_v+(1-w)d_s$$

其中，$w\in[0,1]$为视觉距离所占的权重。

注意到 cosine 距离与 Jaccard 距离均满足距离度量的性质。d 由于是两者线性放缩后的叠加，因此也满足距离度量的性质。

权值 w 反映了用户对视觉信息的重视程度，$w=1$ 表示完全忽略语义信息。该参数原则上与用户的意图直接相关。目前并没有对有效利用该参数进行深入研究，在实验中取定值 $w=0.7$。

3.2 层次聚类

层次聚类也称谱聚类，与扁平式聚类（如 k-means、高斯混合模型聚类）不同，层次聚类得到的结果不是简单的数据与类别的映射关系，而是数据的一个聚类树。之后可以根据标准的不同得到不同的聚类。

层次聚类按层次分解的顺序可以分为：凝聚的（Agglomerative），即自底向上的；分解的（Divisive），即自顶向下的。凝聚算法的基本思想是将每个对象作为一个类，合并两个距离最近的类别以形成新的类，并更新其它类别与新类间的距离，重复该步骤直到只剩下一个类或满足给定条件为止。文献[17]介绍了更多关于层次聚类的内容。

在自底向上方法中，在计算两个类别之间的距离时，有不同的策略。本文采用的是平均策略，即：

$$d(u,v)=\sum_{i\in u,j\in v}\frac{d(u[i],v[j])}{|u|*|v|}$$

其中，$|u|$，$|v|$分别代表类 u，v 的元素个数。

这样，当从检索系统中按用户的关键字取回前 n 幅图像时，就可以以 3.1 节描述的距离，得到 n 幅图形的一个层次聚类，用来最终实现缩放的功能。

聚类树是一棵二叉树，每个叶子节点与数据节点对应，其余节点代表两个子类的聚和。

3.3 缩放功能

通过 3.2 节的层次聚类，得到了一棵聚类树。而“缩放”的功能就是根据用户的直接反馈，从树中选取合适的图像作为代表，返回给用户。

对于得到的树，选择一个阈值 t 来控制聚类树的展开与否。若当前节点 n 的距离即两个子类间的距离小于 t，则不再展开 n，并从 n 的所有叶子节点中选择代表。若 n 的距离大于 t，则递归地对两个叶子节点进行同样的操作。

将阈值 t 用“视距”类比。即对于分辨率（节点的距离）小于视距的类，我们并不想再了解其更多的细节，而只希望得到该类的一个代表。图 1 是一棵通过层次聚类得到的聚类树。增大视距意味着忽略更多细节，即提高结果的离散度，对应“缩小”功能；减小视距意味着观察更多细节，即减小结果的离散度，对应“放大”功能。

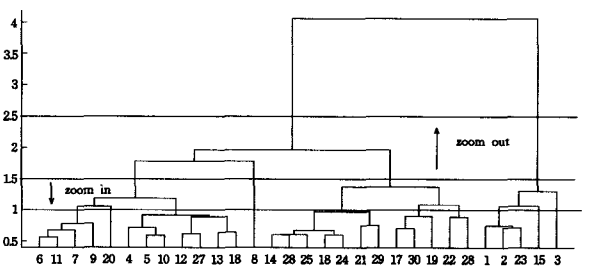

图 1 聚类树与视距的调整

在选取类别的代表时，我们尊重原检索系统的相关度排序，选取该节点下的所有叶子节点中在原检索排序中最靠前的节点。这在后面的算法中得到体现。

我们为用户提供了交互要素：“缩放比例（zoom factor）”。用户在交互界面中可以直接拖动以选取合适的比例。缩放比例与视距 t 间的关系为：

$$t=\text{zoom\_factor} * \text{max\_distance}$$

算法 1 缩放

输入：缩放比例 zoom_factor

输出：最终图像在原结果中的索引值 reps

max_distance=层次树所有类别的最大距离

阈值 t=zoom_factor * max_distance

reps=traversal(根节点 root,t)

算法 2 traversal(遍历)

输入:当前节点 cur_node,停止阈值 t

输出:阈值 t 下,作为代表的图像的索引 reps

IF cur_node. distance > t:

(遍历左右子树)

reps +=traversal(cur_node. left,t)

reps +=traversal(cur_node. right,t)

ELSE:

(不展开该节点,则选择代表图像)

reps=select_representative(cur_node)

同时,我们尊重检索系统返回结果的排序。对于不展开的节点,用如下算法选择代表。

算法 3 select_representative

输入:当前节点 cur_node

输出:当前节点下的作为代表的图像索引 index

indices=cur_node 的所有叶子节点序号

sort(indices)(为索引排序)

index=indices[0](选择最小的序号)

因此,用户可以直接以拖动的形式传递"缩放比例"给系统,而系统便可直接得到视距 t 并以此返回给用户相应的结果,用户再根据结果实时地反馈,如此反复。

4 实验

从文献[9]的标签列表中选择类别离散的 20 个标签作为检索的关键字,分别为:rainbow, wildlife, rabbit, palace, starfish, sailboat, weapon, dolphin, flag, decoration, lion, waterfall, flame, motorcycle, seagull, Olympics, basin, horse, bird, jellyfish,对应的中文含义分别为:彩虹、野生动物、兔子、宫殿、海星、帆船、武器、海豚、旗帜、装饰、狮子、瀑布、火焰、摩托车、海鸥、奥林匹克、盆地、马、鸟、水母。

对于每个检索,从 Flickr 上依据该检索关键字,并按相关性排序,顺序下载前 1000 幅图像,共计 20000 幅图像。对于每幅图像,采用 GIST 的默认参数,得到一个 960 维的 GIST 特征向量作为图像的视觉特征。同时,对于每幅图像,获取对应的标签,并过滤非英文的标签,共有 56193 个独立的标签,平均每幅图像有 11 个标签。

对于每个关键字,首先为用户呈现 Flickr 检索得到的前 16 幅图像,请用户按表 1 为该结果评分。接着,请用户调整缩放比例,整个过程为用户动态呈现 16 幅过滤后的图像。用户认为得到最满意的结果时停下,并对当前结果的满意度评分。同时记录此时的缩放比例。

表 1 数值与满意度对应表

数值	满意度
1	很不满意
2	不满意
3	一般
4	满意
5	很满意

图 2 与图 3 是 30 个用户的实验结果。

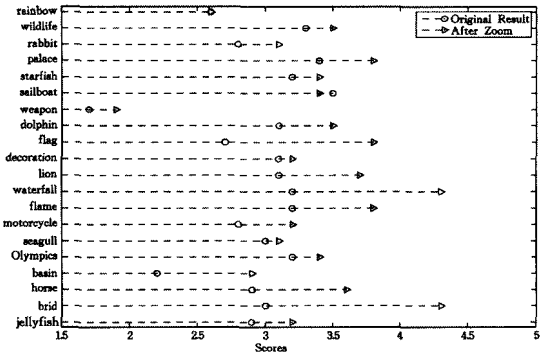

图 2 标签与用户评分(圆点为调整前,三角点为调整后)

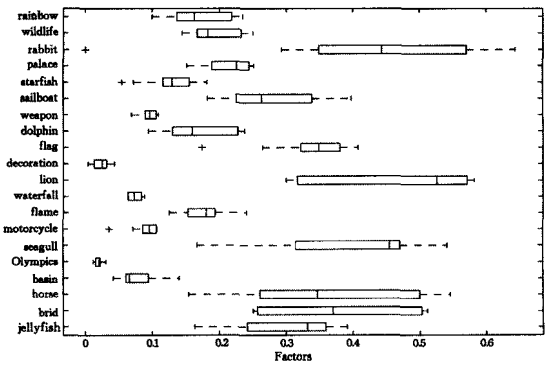

图 3 标签与用户缩放比例

从图 2 中可以看出,通过调整缩放比例,用户对检索结果的满意度有一定程度的提高。满意度有较明显提升的检索类型为方向确定但细节模糊的关键字,如瀑布、鸟。而对于描述已经很清晰的关键字,如帆船、海鸥,则缩放前后没有明显变化。而诸如装饰、奥林匹克等抽象的类别描述同样也没有明显的提升。

可见,"缩放"功能对于查阅者的应用场景相对更有效。因为查阅者通常有一个明确的方向,但不清楚细节。对于搜索者与浏览者作用不是特别大。

同时从图 3 中可以看出,关键字的描述越清晰(如兔子),则缩放比例越大,主要原因是原检索系统(Flickr)得到的结果较为相似;而关键字描述模糊(如武器),则原检索系统得到的结果离散度本身已经较大,因此缩放比例较小。

由于图像结果的离散程度仍旧没有一个很精确的度量标准,且大多时候与用户的意图相关联,即离散度也并非越大越好,因此目前并无法从定量的角度来分析离散度的变化。图 4—图 6 是关键字为"moon"时得到的结果,图 4 为原检索系统返回的图像,图 5、图 6 分别为调整缩放比例时得到的结果(都只取前 16 幅)。可以看到,调整后的结果在拍摄角度、环境、内容的多样性上有一定程度的提升。

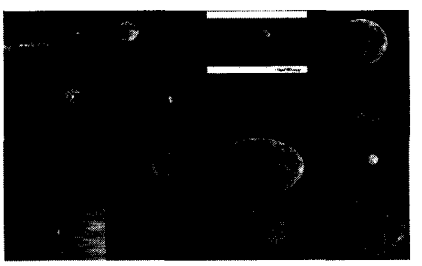

图 4 关键字为"moon"、zoom_factor=0 时的结果

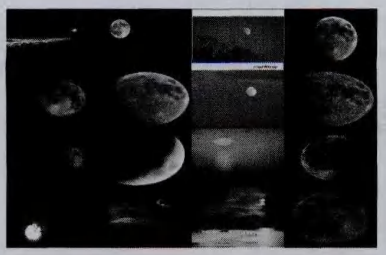

图 5　关键字为“moon”、zoom_factor=0.1936 时的结果

图 6　关键字为“moon”、zoom_factor=0.3059 时的结果

最后，该“缩放”功能依赖于原检索系统结果，如关键字为火焰，得到的结果几乎都是烛火，因此不论如何缩放，得到的结果也都属于烛火的范畴。

结束语　算法的参数决定了算法的应用场景。检索系统是用户导向的，用户的意图是它的一个重要参数。图像检索系统的用户中，查阅者与浏览者和搜索者不同，它只有模糊的意图，本文论述了检索结果的离散度如何帮助具有这类意图的用户。

如何准确获取用户意图十分困难。本文将“缩放”功能引入检索系统中，为用户提供“缩放比例”这一交互要素，使其能动态控制检索结果的离散度，即直接以用户调节缩放比例的方式获取用户的意图，并作用于结果的离散度。

检索系统与用户的交互密切相关，在探索更好算法的同时，我们也应当关注新的交互方式能给系统带来的优势。

参 考 文 献

[1] Chen H, Karger D R. Less is more: probabilistic models for retrieving fewer relevant documents[C]// Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006. Seattle, WA, USA, 2006: 429-436

[2] Ritendra D, Dhiraj J, Li J, et al. Image retrieval[J]. ACM Computing Surveys, 2008, 40(2): 1-60

[3] Rui Y, Huang T S, Ortega M, et al. Relevance feedback: a power tool for interactive content-based image retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998, 8(5): 644-655

[4] Tang X, Liu K, Cui J, et al. IntentSearch: Capturing user intention for one-click internet image search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7), 1342-1353

[5] Hoi C H, Lyu M R. A novel log-based relevance feedback technique in content-based image retrieval[C]// Proceedings of the 12th Annual ACM International Conference on Multimedia, 2004. New York, NY, USA, 2004: 24-31

[6] Zhou X S, Huang T S. Relevance feedback in image retrieval: A comprehensive review[J]. Multimedia systems, 2003, 8(6): 536-544

[7] Carbonell J, Goldstein J. The use of MMR, diversity-based reranking for reordering documents and producing summaries[C]// Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1998. Melbourne, Australia, 1998: 335-336

[8] Song K, Tian Y, Gao W, et al. Diversifying the image retrieval results[C]// Proceedings of the 14th annual ACM international conference on Multimedia, 2006. Santa Barbara, CA, USA, 2006: 707-710

[9] Wang M, Yang K, Hua X S, et al. Towards a relevant and diverse search of social images[J]. IEEE Transactions on Multimedia, 2010, 12(8): 829-842

[10] van Leuken R H, Garcia, et al. Visual diversification of image search results[C]// Proceedings of the 18th international conference on world wide web, 2009. Madrid, Spain, 2009: 341-350

[11] Deselaers T, Ferrari V. Visual and semantic similarity in imagenet[C]// 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, USA, 2011: 1777-1784

[12] Cai D, He X, Li Z, et al. Hierarchical clustering of WWW image search results using visual, textual and link information[C]// Proceedings of the 12th Annual ACM International Conference on Multimedia, 2004. New York, NY, USA, 2004: 952-959

[13] Jaimes A, Chang S F, Loui, et al. Detection of non-identical duplicate consumer photographs [C] // Proceedings of the 2003 Joint Conference of the Fourth International Conference on Multimedia, 2003. Singapore, 2003, 1: 16-20

[14] Fisichella M, Deng F, Nejdl W. Efficient incremental near duplicate detection based on locality sensitive hashing[C]// Database and Expert Systems Applications, 2010. Bilbao, Spain, 2010: 152-166

[15] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. International Journal of Computer Vision, 2001, 42(3), 145-175

[16] Cilibrasi R L, Vitanyi P M. The google similarity distance[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(3), 370-383

[17] 孙吉贵，刘杰，赵连宇，等. 聚类算法研究[J]. 软件学报，2008，19(1)：48-61
Sun Ji-gui, Liu Jie, Zhao Lian-yu, et al. Clustering Algorithms Research[J]. Journal of Software, 2008, 19(1): 48-61

[18] 唐朝霞，章慧，徐冬梅. 一种改进的粒子群算法和相关反馈的图像检索[J]. 计算机科学，2011，38(10)：278-280
Tang Zhao-xia, Zhang Hui, Xu Dong-mei. Image Retrieval Based on Improved PSO Algorithm and Relevance Feedback[J]. Computer Science, 2011, 38(10): 278-280