ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.19, No.10, October 2008, pp.2667−2680
DOI: 10.3724/SP.J.1001.2008.02667

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

一种支持海量跨媒体检索的集成索引结构*

庄 毅[1+], 庄越挺[2], 吴 飞[2]

[1](浙江工商大学 计算机与信息工程学院,浙江 杭州 310018)

[2](浙江大学 计算机科学与技术学院,浙江 杭州 310027)

An Integrated Indexing Structure for Large-Scale Cross-Media Retrieval

ZHUANG Yi[1+], ZHUANG Yue-Ting[2], WU Fei[2]

[1](College of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China)

[2](College of Computer Science, Zhejiang University, Hangzhou 310027, China)

+ Corresponding author: E-mail: zhuangyi@cs.zju.edu.cn

Zhuang Y, Zhuang YT, Wu F. An integrated indexing structure for large-scale cross-media retrieval. *Journal of Software*, 2008,19(10):2667−2680. http://www.jos.org.cn/1000-9825/19/2667.htm

Abstract: This paper proposes a novel integrated indexing structure for the large-scale cross-media retrieval. In the cross-media retrieval, first a cross reference graph (CRG) is generated by hyperlink analysis of the webpages, which supports the cross-media retrieval. Then a refinement process of the CRG is conducted by users' relevance feedbacks. Three steps are made. First, when the user submits a query media object, the candidate media objects are quickly identified by searching the cross reference graph. Then the distance computation of the candidate vectors is conducted to get the answer set. The analysis and experimental results show that the performance of the algorithm is superior to that of sequential scan.

Key words: cross-media, cross reference graph model, media object, integrated index structure

摘 要: 提出一种支持海量跨媒体检索的集成索引结构.该方法首先通过对网页的预处理,分析其中不同模态媒体对象之间的链接关系,生成交叉参照图.然后通过用户相关反馈进行调节.当用户提交一个查询对象时,首先对交叉参照图进行基于索引的快速定位,得到与查询对象相关的候选媒体对象.然后对得到的候选媒体对象进行距离运算,得到结果媒体对象.理论分析和实验表明,该方法较顺序检索具有更好的性能,非常适合海量跨媒体数据检索.

关键词: 跨媒体;交叉参照图模型;媒体对象;集成索引结构

中图法分类号: TP311 **文献标识码**: A

* Supported by the National Natural Science Foundation of China under Grant No.60533090 (国家自然科学基金); the National Science Fund of China for Distinguished Young Scholar under Grant No.60525108 (国家杰出青年基金); the China America Digital Academic Library Project (高等学校中英文图书数字化国际合作计划); the Natural Science Foundation of Zhejiang Province of China under Grant No.Y1080148 (浙江省自然科学基金); the Open Project of Zhejiang Provincial Key Laboratory of Information Network Technology of China (浙江省综合信息网技术重点实验室开放课题)

Received 2007-04-20; Accepted 2007-10-09

随着 Internet 和多媒体技术的不断发展,特别是近几年来,Internet 上多媒体信息的爆炸性增长,基于内容的海量多媒体检索和索引[1]已成为一个热门的研究领域.在这些海量的多媒体信息当中,不同模态媒体对象之间往往存在某种语义相关性,如图 1 所示,“老虎”的图片对应“老虎”的音频和视频等.传统的多媒体检索都是针对单一模态媒体对象,如基于内容的图像[2]、音频[3]和视频[4]检索等.较少有文献系统地研究基于多模态多媒体信息的交叉检索,即通过一种模态的媒体对象检索出另外一种或几种基于相同语义的不同模态的媒体对象.早在 1976 年,McGurk 就已经揭示出了人脑对外界信息的认知需要跨越和综合不同的感官信息,以形成整体性的理解[5].同时认知神经心理学方面的研究也进一步验证了人脑的认知过程呈现出跨媒体的特性[6],即对来自视觉、听觉等不同感官的信息相互刺激、共同作用而产生认知结果.我们将这类检索称为跨媒体检索[7,8],它作为一种新兴的多媒体检索方式正越来越受到国内外学术界的关注.跨媒体也可以看作一种由各种基于相同语义媒体对象构成的复杂媒体类型,显然,对它提取的特征是高维的.而高维相似性检索是一种 CPU 密集性的运算.如何利用索引技术来加快海量跨媒体检索是一个很重要的课题.

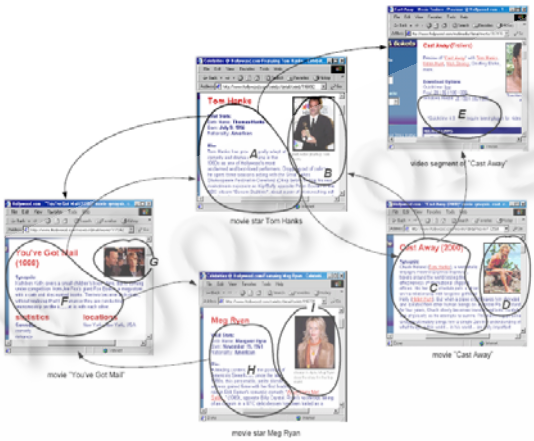


Fig.1 Latent semantic correlation of media objects in the webpages
图 1 网页中不同媒体对象存在的潜在语义关联

本文提出一种基于索引的海量跨媒体检索算法——CIndex,显著地提高了检索效率.首先通过对网页的预处理,提出基于链接分析和相关反馈结合的方式建立交叉参照图模型.然后,提出跨媒体索引的统一表达方式及其查询算法.同时对算法建立代价模型并且进行详细分析,说明各种参数对算法的影响程度.

本文第 1 节回顾相关工作.第 2 节介绍基于链接分析和相关反馈结合的不同模态媒体对象相关性挖掘算法并给出交叉参照图模型.第 3 节介绍跨媒体索引——Cindex.第 4 节给出基于索引的跨媒体检索算法.第 5 节给出查询的代价模型.第 6 节通过实验从不同角度证明该算法优于其他算法.最后是总结和对未来工作作出展望.

1 相关工作

1.1 多媒体检索

自从 20 世纪 90 年代初期以来,基于内容的多媒体检索已经成为一个非常活跃的研究领域.其中最有代表性的是基于内容的图像检索(content-based image retrieval,简称 CBIR),如 QBIC(query by image content)[2],Virage[9],Photobook[10]和 MARS(multimedia analysis and retrieval system)[11]等.基于内容的视频分析与检索系统包括 Carnegie Mellon 大学的 Informedia[4].然而这些检索系统都只是针对单一模态的媒体对象检索.随着互联网及多媒体技术的飞速发展,互联网中的许多不同模态的媒体对象呈现相同的语义特性,文献[12]中提出的 Octopus 是一个具有跨媒体特性的检索系统,它能够实现从一种模态的媒体对象检索出另一种模态的媒体对象的功能.但是该系统尚未考虑索引,对海量数据检索效率不高.文献[13]针对图片的视觉和语义特征,提出了一种混合的索引方式,将两种模态的特征信息采用同一个索引进行表示,但其适用性非常有限,不支持多种模

态媒体对象的索引表达.

1.2 高维索引

跨媒体索引属于高维索引范畴,而高维索引技术经历了 20 多年的研究[14].以 R-tree[15]为代表的树形索引方法适合维数较低的情况,随着维数的增加,其索引的性能往往劣于顺序检索,产生“维数灾难”;之后又提出了以 VA-file[16]为代表的用近似的方法来表示原始对象从而加速高维查找速度的方法.尽管该类方法通过对高维对象进行压缩和近似存储来加速顺序查找速度.然而,数据压缩和量化带来的信息丢失使其首次过滤后的查询精度并不能令人满意.同时,尽管减少了磁盘的 I/O 次数,但由于需要对位串解码同时计算对查询点距离的上界和下界,导致很高的 CPU 运算代价.最近有学者提出通过将高维查询转化为一维查询来进一步提高查询效率,如 iDistance[17]等.iDistance 是基于多参考点的方法,通过引入多参考点并结合聚类的方法有效地过滤了高维数据空间的搜索范围,提高了查询的精度,然而其查询效率在很大程度上取决于参考点的选取,并且依赖于数据聚类和分片.

2 基于链接分析和相关反馈的多模态媒体对象相关性挖掘

为了支持高效的跨媒体检索,提出基于链接分析和相关反馈结合的多模态媒体对象相关性挖掘算法.通过相关性挖掘和分析,得到的交叉参照图模型是进行跨媒体检索的关键所在.

2.1 问题定义

正如前面所介绍的,跨媒体检索[7,8]是指通过一种媒体对象检索出另外一种或几种基于相同语义的不同的媒体对象.如通过提交一张“老虎”的图片,检索出语义为“老虎”的音频或视频.本文通过对多媒体文档(网页)的链接分析建立交叉参照图模型,通过用户的相关反馈来逐步修正交叉参照图,从而实现有效的跨媒体检索.

表 1 给出了不同媒体对象的相似度距离函数,用于基于内容的跨媒体检索.

Table 1 Features of different media objects and their similarity measures

表 1 各种媒体对象的特征和相似度计算函数

Media type	Feature	Similarity function
Image	256 HSV 32-d tamura direction	Euclidean distance
Audio	4 features in compressed domain: (1) Centroid; (2) Rolloff; (3) Spectral flux; (4) RMS	Cosine similarity[3]
Video	Key frames generated by *k*-means algorithm	Multi-View-Based video similarity measure[18]

2.2 交叉参照图模型

多媒体文档是一种逻辑上的文档[1],它是由一些在语义上相关的媒体对象(文本、图像、音频、视频或者图形)所组成的.语义上相近的多媒体文档之间存在着某种关联,这种关联可以用它们之间的链接来表示.多媒体文档的概念在现实生活中存在着许多实例.一个包含了图像和视频的网页就可以看成是一个多媒体文档,并且这种多媒体网页通过其所包含的某个媒体对象的链接指向其他多媒体页面,这种网页目前已经大量存在.其语义框架用来描述多媒体文档及其包含的全部媒体对象以及与其他多媒体文档之间的链接关系.

定义 1. 每个多媒体文档 *MD* 可表示为一个五元组:

$$MD:=\langle DocID,URI,KeywList,ElemSet,LinkSet\rangle,$$

其中,*DocID* 表示文档的标识;*URI* 表示文档的统一资源定位标识;*KeywList* 表示文档的关键字描述;*ElemSet* 表示文档中包含的媒体对象集合;*LinkSet* 表示文档包含的链接集合,包括该文档被其他文档所指向的链接和该文档指向其他文档的链接).*ElemSet* 中的每个媒体对象可以由它的语义特征和底层感知特征来描述,如图像 $Image=\langle ImgID,KeywList,ImgFeature\rangle$.

为了达到跨越多种模态媒体对象统一检索的目的,需要将多媒体文档语义框架中的各种媒体对象之间的高层联系提取出来.本文采用交叉参照图(cross reference graph,简称 CRG)模型来描述媒体对象之间的(语义)相

关性.媒体对象之间的高层语义关联可以通过链接关系表示,如两个媒体对象之间存在超链接,就可以认为它们之间有一定的语义相关性.这种关联与媒体的底层特征无关,并且两个不同模态的媒体对象之间也能够建立关联.如某个多媒体文档中包含了一个图像,该图像通过超链接指向另一个文档中的一个音频对象,虽然音频的听觉感知特征与图像的视觉感知特征差异很大,但是它们之间存在的超链接表示这两个不同模态的媒体对象在语义上有某种关联,通过这种语义上的相关性可以实现不同模态媒体对象之间相互检索的目的(如通过图像检索到音频).由此,交叉参照图模型也可以看作是各种媒体对象之间的交叉参照索引,是指导跨媒体检索的基础.

定义 2. 交叉参照图模型是一个无向图,可形式化地表示为 $CRG=(V,E)$,其中,V 表示媒体对象集;E 表示该图的边集,即两个媒体对象(V_i 和 V_j)之间的相似度或相关度.需要说明的是,当两种媒体对象为同模态时,它们之间的关联称为相似度(similarity),如图 2 中的实线表示;当两种媒体对象为不同模态时,它们之间的关联称为相关度(correlation),如图 2 中的虚线表示.

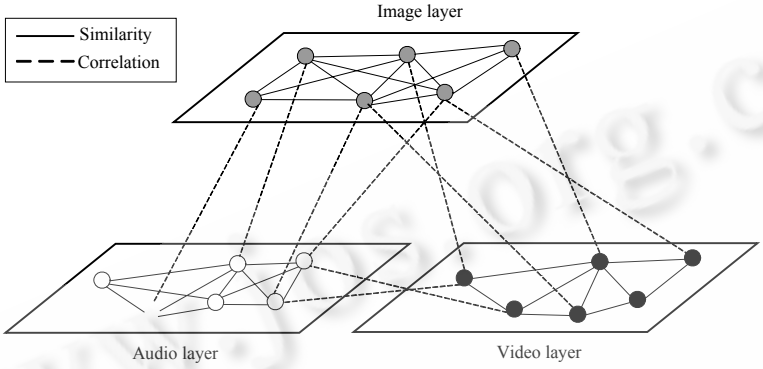


Fig.2 An example of the cross reference graph model

图 2 交叉参照图模型例子

如图 2 所示,交叉参照图模型描述了媒体对象之间潜在的语义联系.在计算媒体对象之间的权重之前,由于互联网中网页存在“噪声”链接信息,如广告栏、菜单条等,这些对于交叉参照图的建立会产生负面的影响.需要预先对网页链接信息进行基于 VIPS(vision-based page segmentation)[19]的过滤,排除网页周围的噪声链接.然后通过如下 3 个先验知识:(1) 属于同一个多媒体文档的媒体对象之间在语义上被认为具有一定的相关性;(2) 被同一个多媒体文档所指的媒体对象在语义上被认为具有一定的相关性;(3) 一个媒体对象被另一个媒体对象所属的多媒体文档所指向,那么它们在语义上被认为具有一定的相关性,据此来初步建立交叉参照.根据对 1 万张实际网页的统计,大约 90%左右的网页所存在的多媒体网页和媒体对象之间的链接关系满足以上 3 个先验知识.结合上面的先验知识和多媒体文档语义框架的描述,可以计算媒体对象之间的权重,这种权重反映了媒体对象之间语义关联的强弱.因此,可以按照算法 1 来计算媒体对象之间的权重.

算法 1. 交叉参照图建立.

输入:e_{ij}:初始的媒体对象 X_i 和 X_j 的权重(设为 0).

输出:CRG:交叉参照图.

1. $\forall X_i \in \Omega, \forall X_j \in \Omega : e_{ij}=0$; /* initialization*/
2. for any two media objects $X_i, X_j \in \Omega$ do
3. if X_i 和 X_j 属于同一个网页 then
4. $e_{ij} \leftarrow e_{ij}+1$;
5. else if X_i 和 X_j 属于被同一个网页所指向或指向同一个网页 then
6. $e_{ij} \leftarrow e_{ij}+1$;
7. else if X_j 被 X_i 所属的网页所指向 then

8. $e_{ij} \leftarrow e_{ij}+1$;
9. end if
10. end for
11. return CRG

通过算法 1 得到的无向边 e_{ij} 的权重反映了媒体对象 X_i 和 X_j 在语义上的关联强弱.基于交叉参照图 CRG 可以实现不同模态媒体对象之间的相互检索.通过上面的方法,已经建立起各种媒体对象的交叉参照图.对于该交叉参照图,还需要通过用户的相关反馈进行逐步调整.

3 跨媒体索引——CIndex

随着媒体对象数据量的快速增加,对应的交叉参照图将变得非常巨大.面对如此庞大的交叉参照图,采用传统的图遍历方法进行相关媒体对象的定位非常低效.如何进行快速而准确的定位是一个很大的挑战.本节提出一种集成的跨媒体索引结构,该方法能够实现对交叉参照图的快速而准确的定位,以实现跨媒体检索的目标.

3.1 预备知识

给定任意媒体库Ω(Ω可以表示为图片、音频或者视频数据库),包含$\|\Omega\|$个该媒体对象 X_i,其中 X_i 可以是一张图片、一段音频例子或者一段视频例子.$\|\Omega\|$ 表示媒体库 Ω 中包含的媒体对象总数,$i \in [1,\|\Omega\|]$ 且 $\forall X_i \in \Omega$.

对任意媒体库 Ω 中的对象进行层次聚类(如 BIRCH[20])得到 T 个类.对于任意一个类 C_j,其中 $j \in [1,T]$.随机选择类中的一个媒体对象作为其类的质心 O_j(不包括类的边缘).这样分别得到了图片、音频和视频媒体对象的层次聚类结果,如图 3 中的虚线圆所示.

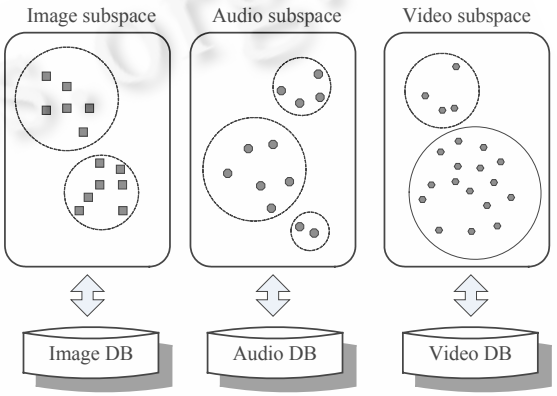


Fig.3 Subspace clustering
图 3 子空间聚类

定义 3(类半径). 对于任意一个类 C_j,其质心对象 O_j 与该类中距离其最远的对象的距离,称为它的类半径,记作 CR_j,其中 $j \in [1,T]$.

给定任意一个类 C_j 及其类半径 CR_j,对应类超球表示为 $\Theta(C_j,CR_j)$.

定义 4(质心距离). 给定任意一个媒体对象 X_i,它的质心距离为到其对应类 C_j 的质心 O_j 的距离,表示为 $\delta(X_i)=d(X_i,O_j)$,其中 $X_i \in C_j$ 且 $j \in [1,T]$.

基于距离的跨媒体索引键值表达的提出基于以下 3 点:第一,高维空间中的同类媒体对象之间的相似性可以通过该对象与某个参考对象来度量和排序;第二,由于距离是一维值,这样可以用一维值来表示高维空间的对象,同时可以使用 B+树来对这些距离数据建立索引;第三,任意两个相似的同类媒体对象具有相似的质心距离,可以有效地过滤高维空间中的不相关同类媒体对象.同模态媒体对象的相似度和不同模态媒体对象的相关度通过线性组合可以构成一个统一的多模态媒体对象的索引键值.

3.2 数据结构

为了支持跨媒体检索,之前已经通过链接分析的方法得到不同模态媒体间的交叉参照图.如图 4 所示,以图片为例,其对应的交叉参照图可以表示成邻接表结构.例如,ID 为 21 的图片,与其语义相关的对应音频对象为 3,9,18 和 26,对应的视频对象为 7 和 39.需要说明的是,每个 ID 下面的数字表示对应的两种模态媒体对象之间的相关度.

假设图 5 表示图片所对应的高维特征空间.对于 ID 为 21 的图片对象来说,图 5(a)中虚线圆包含了与该图片

语义相关的音频对象,图 5(b)中虚线圆包含了与该图片相关的视频对象.因此,图像高维特征空间中的每个数据点(即图片对象)都存在 2 个内嵌子空间.同时,又由于该内嵌子空间中的媒体对象都是语义相关的,可以称为“内嵌相关子空间(embedded correlation subspace,简称 ECS)”.

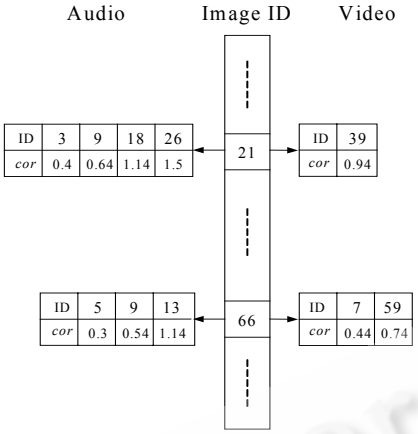
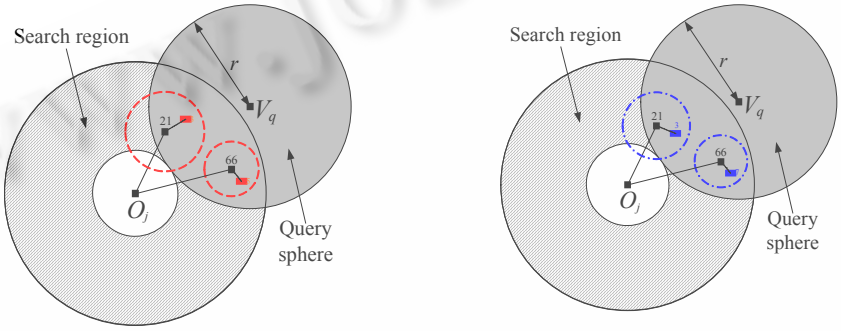


Fig.4 The adjacency graph-based cross reference graph representation

图 4 基于邻接表的交叉参照图表示



(a) Image & Audio (a) 图像和音频

(b) Image & Video (b) 图像和视频

Fig.5 Embedded correlation subspaces in high-dimensional image feature spaces

图 5 高维图片特征空间包含的内嵌子空间

以图片为例,为了实现从图片到音频的跨媒体检索,对应图片 I_i 的索引键值可以表示为

$$key(I_i)=\beta\times\langle d(I_i,O_j),\theta\rangle+\frac{c(I_i,A_k)}{\mathrm{MAX}} \tag{1}$$

其中,$d(I_i,O_j)$表示 I_i 与质心 O_j 的相似距离,$c(I_i,A_k)$表示 I_i 与 A_k 的相关度,$\langle\bullet,\theta\rangle$表示将$\bullet$取到小数点后第$\theta$位.$\beta$为线性放大常数使得 $\langle d(I_i,O_j),\theta\rangle$ 为整数,常数 MAX 使 $c(I_i,A_k)$归一化,这样,相似距离 $d(I_i,O_j)$与相关度 $c(I_i,A_k)$所对应的值域不重叠.

由于图像数据预先通过聚类得到 T 个类,为了将不同类中的图片对象用一个索引键值来表示,可以将式(1)的键值改为式(2)的形式:

$$key(I_i)=\alpha\times CID+\beta\times\langle d(I_i,O_j),\theta\rangle+\frac{c(I_i,A_k)}{\mathrm{MAX}} \tag{2}$$

其中,CID 表示 I_i 对应的类的编号,α为线性扩展常数.

式(2)的索引键值表达实现了图片到音频跨媒体检索的键值统一表达.然而,为了实现从图像到视频的跨媒

体检索,其索引键值可以表达成式(3):

$$key(I_i)=\alpha\times CID+\beta\times\langle d(I_i,O_j),\theta\rangle+\frac{c(I_i,V_w)}{\mathrm{MAX}} \tag{3}$$

为了进一步将式(2)和式(3)对应的索引键值表达成一个统一的索引键值,分别加上两个较大的扩展系数(S_A 和 S_V)即可.因此,综上所述,图像对象 I_i 的统一跨媒体索引可以表示为

$$key(I_i)=\begin{cases}S\_A+\alpha\times CID+\beta\times\langle d(I_i,O_j),\theta\rangle+\dfrac{c(I_i,A_k)}{\mathrm{MAX}}\\ S\_V+\alpha\times CID+\beta\times\langle d(I_i,O_j),\theta\rangle+\dfrac{c(I_i,V_w)}{\mathrm{MAX}}\end{cases} \tag{4}$$

同理,对音频和视频来说,其对应统一跨媒体索引键值可以分别表示为

$$key(A_i)=\begin{cases}S\_I+\alpha\times CID+\beta\times d(A_i,O_j)+\dfrac{c(A_i,I_k)}{\mathrm{MAX}}\\ S\_V+\alpha\times CID+\beta\times d(A_i,O_j)+\dfrac{c(A_i,V_w)}{\mathrm{MAX}}\end{cases} \tag{5}$$

$$key(V_i)=\begin{cases}S\_I+\alpha\times CID+\beta\times\langle d(V_i,O_j),\theta\rangle+\dfrac{c(V_i,I_k)}{\mathrm{MAX}}\\ S\_A+\alpha\times CID+\beta\times\langle d(V_i,O_j),\theta\rangle+\dfrac{c(V_i,A_w)}{\mathrm{MAX}}\end{cases} \tag{6}$$

式(4)~式(6)分别为图片、音频和视频的跨媒体索引键值表达,彼此相互独立,分别对应 3 个独立的索引.为了进一步将它们用一个统一的索引来存储和表示,可得到如式(7)所示的跨媒体检索的统一索引键值表达:

$$key(X_i)=\begin{cases}SCALE\_I+key(I_i), & \text{if } X_i=I_i\\ SCALE\_A+key(A_i), & \text{if } X_i=A_i\\ SCALE\_V+key(V_i), & \text{if } X_i=V_i\end{cases} \tag{7}$$

其中,X_i 表示某种模态的媒体对象,如 X_i 可以是一张图片,也可以是一段音频例子或一段视频例子;$SCALE\_I$, $SCALE\_A$ 和 $SCALE\_V$ 分别为扩展系数,用于线形扩大不同媒体对象的索引键值范围,使其值域互不重叠.

不失一般性,以 CIndex 中的图片索引部分为例,图 6 形象地给出了 2 张图片对应的 4 个跨媒体索引键值(即图像和音频,图像和视频)值域范围在 CIndex 索引叶节点层面的映射.

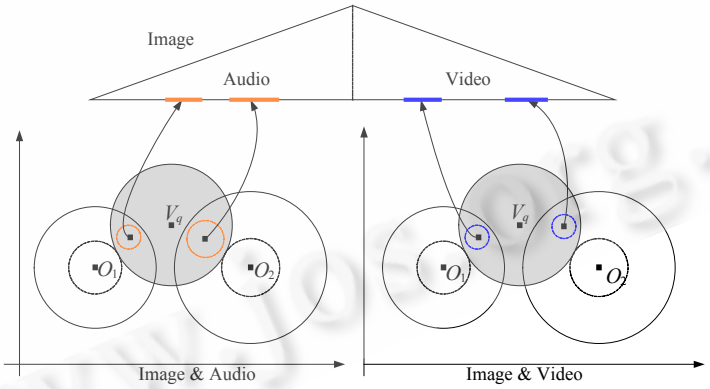


Fig.6 Mapping for the value range of index key of CIndex

图 6 CIndex 中的键值值域范围映射

上面介绍的索引键值表达是对应图 7(a)中的基于二元组的叶节点表示.然而,当用户提交一张图片,需要检索相关的音频和视频时,采用这种方法需要两次访问索引.下面提出的基于三元组的叶节点索引键值表示只需一次访问索引即可得到其他两种语义相关的媒体对象.以图片为例,其索引键值表示为

$$key(I_i)=SCALE\_I+\alpha\times CID+\beta\times\langle d(I_i,O_j),\theta\rangle+\frac{c(I_i,A_k)+c(I_i,V_w)}{\max} \tag{8}$$

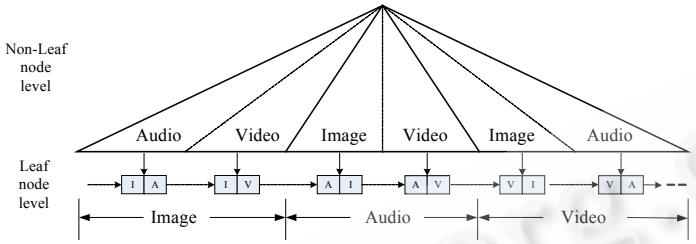
其中,常数 max 使 $c(I_i,A_k)+c(I_i,V_w)$归一化且 max>MAX.

同理,分别得到音频、视频对应的索引键值:

$$key(A_k)=SCALE\_A+\alpha\times CID+\beta\times\langle d(A_k,O_j),\theta\rangle+\frac{c(A_k,I_i)+c(A_k,V_w)}{\max} \tag{9}$$

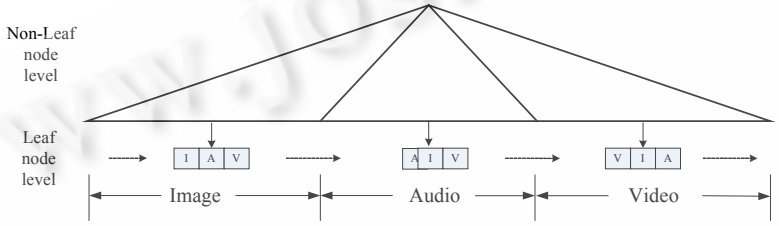
$$key(V_w)=SCALE\_V+\alpha\times CID+\beta\times\langle d(V_w,O_j),\theta\rangle+\frac{c(V_w,I_i)+c(V_w,A_k)}{\max} \tag{10}$$

这样得到的索引如图 7(b)所示.



(a) Two-Tuple-Based representation for index leaf node

(a) 基于二元组的索引叶节点表示



(b) Three-Tuple-Based representation for index leaf node

(b) 基于三元组的索引叶节点表示

Fig.7 Two uniform cross-media indexing structures

图 7 两种跨媒体统一索引结构

3.3 索引生成算法

根据索引叶节点所存储的元素,本节给出两种跨媒体索引结构的表示.即基于二元组叶节点的 CIndex 和基于三元组叶节点的 CIndex.

根据索引叶节点所存储的元素个数的不同,跨媒体统一索引 CIndex 可分为如图 7 所示的两种结构,可分为:1) 基于二元组(图片(I)、音频(A)或视频(V)的两两组合)的索引叶节点表示,如图 7(a)所示;2) 基于三元组(即图片(I)、音频(A)和视频(V)的组合)的索引叶节点表示,如图 7(a)所示.CIndex 索引是一棵平衡的 B+树.

现在从存储和查询代价两方面来比较上述两种索引性能上的差异,首先假设对于某张图片 I_i,与其语义相关的音频和视频对象分别为 n 和 m 个,则

- 存储代价

一般来说,基于三元组叶节点的 CIndex 比二元组的 CIndex 的存储代价要高很多.通过分析可以得到,基于二元组叶节点和基于三元组叶节点的索引存储代价分别为 $O(n+m)$和 $O(n\times m)$.显然,$O(n+m)<<O(n\times m)$.因此,基于二元组的 CIndex 的存储代价要远远小于基于三元组的 CIndex 的存储代价.

- 查询代价

假设基于三元组叶节点得到的索引树高为 H,二元组得到的索引树高为 h.由于 B+树是 CIndex 的基本数据

结构,因此 $H \geq h$.同时由于 B+树查询的代价由两部分构成,即从根节点到叶节点的遍历+叶节点上的遍历.因此可以得到上述两种索引的查询代价分别为 $O(H+m\times n)$和 $O(h+m+n)$.显然,$O(H+m\times n)>>O(h+m+n)$.因此,在相同查询条件下,基于二元组的 CIndex 要优于基于三元组的 CIndex.

综上所述,从理论上可以得到采用基于二元组叶节点的索引无论在存储还是查询方面都要优于基于三元组叶节点的索引.

由于上面的 CIndex 索引包含了 3 种模态的媒体类型,因此可以看作由 3 部分构成,每一部分分别是由与图像、音频或视频对象语义相关的其他 2 种不同模态的媒体对象的组合得到.需要注意的是,它的每个叶节点存储两种媒体对象的 ID.算法 2 表示跨媒体索引创建.以图片为例,假设预先已经得到交叉参照图(CRG),并且对高维图像数据进行了聚类,对于每个类中的图片,通过交叉关联图寻找与其相关的其他模态的媒体对象(第 3~4 行).然后,根据式(7)得到对应媒体对象的索引键值并将其插入 B+树(第 5 行).

算法 2. CIndexBuild(Ω,CRG).

输入: Ω:媒体对象库,CRG:交叉参照图.

输出:*bt*:Cindex.

1. initialize; /*初始化*/
2. **for** each media object $X_i \in \Omega$ do /*X_i 可以表示图片也可以是音频或视频对象*/
3. locate the X_i in G; /*定位媒体对象 X_i 在交叉关联图中的位置*/
4. get the media objects semantically related to X_i; /*通过 G,得到与 X_i 相关的媒体对象*/
5. $bt \leftarrow$***InsertBtree***(***key***(X_i)); /*按照公式(7)得到索引键值并将其插入 B+树*/
6. **end for**
7. **return** *bt*

3.4 索引的可扩展性

第 3.3 节给出的 CIndex 可以支持图像、音频和视频的跨媒体检索.然而,随着多媒体技术的飞速发展,将会出现各种各样的新的媒体对象,如 Flash 动画等.因此需要 CIndex 具有良好的可扩展性,可以支持多种新的媒体对象的跨媒体检索.由于 CIndex 采用基于距离值和线性组合的索引表达机制,因此对于新的媒体对象的引入,具有良好的可扩展性.例如,当有新的媒体对象,如 Flash 动画添加到 CIndex 时,其索引结构可表示为如图 8 所示.图 8 中的阴影部分表示新添加到 CIndex 的部分.这样,该索引变成了可以支持图片、音频、视频和 Flash 动画的新的跨媒体索引结构.

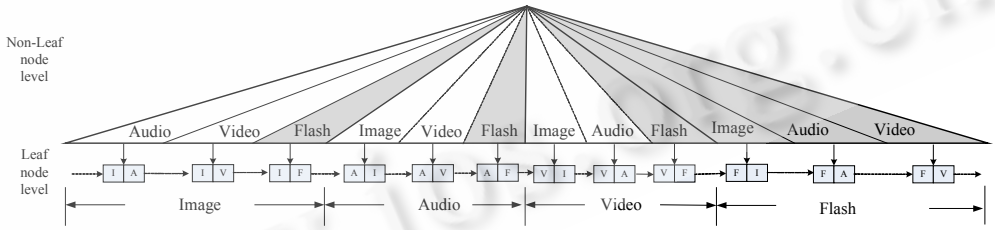


Fig.8 The scability of the CIndex structure

图 8 可扩展的 CIndex 结构

4 查询算法

CIndex 索引能够支持各种媒体对象的跨媒体检索.也就是说,用户的输入可以是图片、音频或视频.以图片为例,当用户提交一张图片例子时,通过交叉关联图(CRG)寻找与其相关的其他模态的媒体对象.然后通过过滤得到的对象,再通过求精来得到.算法 3 为跨媒体查询算法.需要说明的是,查询对象 X_q 中的 X 既可以是图片 I、

音频 A 也可以是视频 V.另外,函数 *Search*()返回的媒体对象已经包括了与提交媒体对象不同模态的媒体对象,从而实现了跨媒体检索.在该函数中,根据例子对象 X_q 的不同,*SCALE_X* 可以是 *SCALE_I*,*SCALE_A* 或 *SCALE_V*.

算法 3. CrossSearch(X_q,r).

输入:例子对象 X_q,查询半径 r /*X_q 可以是图片、音频或视频*/

输出:查询结果 S

1. $S\leftarrow\varnothing$; /*初始化*/
2. for i:=1 to *num* do /**num* 表示需要访问 *num* 次 CIndex 索引*/
3. for j:=1 to T do /**T* 表示总的聚类个数*/
4. if $\Theta(O_j,CR_j)$ dose not intersects $\Theta(X_q,r)$ then
5. next loop;
6. else
7. $S1\leftarrow$***Search***(X_q,r,j);
8. $S\leftarrow S\cup S1$;
9. if $\Theta(O_j,CR_j)$ contains $\Theta(X_q,r)$ then end loop;
10. end if
11. end for
12. end for
13. if user is not satisfied with S then
14. return S; /*返回候选对象*/
15. else
16. Get user's feedback and update S and CRG;

Search(X_q,r,j)

17. $left \leftarrow SCALE\_X + S\_X + \alpha \times CID + \beta \times (d(X_q, O_j) - r)/MCD$;
18. $right \leftarrow SCALE\_X + S\_X + \alpha \times CID + \beta \times CR_j / MCD$;
19. $S\leftarrow$***BRSearch***$[left,right,j]$; /*S 中包括与 X_q 语义相关的不同模态的媒体对象*/
20. for each media object $X_i\in S$
21. if $d(X_q, X_i) > r$ then $S \leftarrow S - X_i$;

/*将 X_i 从候选对象集 S 中删除去的同时,与其相关的其他模态的媒体对象也随之删除*/

22. end for
23. return S; /*返回候选对象*/

5 代价模型

本节给出基于二元组叶节点的 CIndex 索引的查询代价模型.不失一般性,以图片对象为例.令 f 表示 B+树每个节点的平均出度,n 为图片的个数,m_i 为与第 i 张图片相关的音频例子个数,k_i 表示与第 i 张图片相关的视频例子个数.

由于 CIndex 索引树的高度(h)和叶节点个数满足式(11):

$$(f+1)^{h-1}=\frac{1}{f}\sum_{i=1}^{n}(m_i+k_i) \tag{11}$$

求解式(11),得到式(12):

$$h=\left\lceil \frac{\lg\sum_{i=1}^{n}\left(m_i+k_i\right)-\lg f}{\lg\left(f+1\right)} \right\rceil+1 \tag{12}$$

通过分析,可以得到基于二元组叶节点的 CIndex 索引的磁盘块总数为

$$NA_{2-Tuple}=h+\frac{m_i+k_i}{f}=\left\lceil \frac{\lg\sum_{i=1}^{n}\left(m_i+k_i\right)-\lg f}{\lg\left(f+1\right)} \right\rceil+1+\frac{m_i+k_i}{f} \tag{13}$$

从式(13)看出,该索引的查询代价与媒体对象总数成正比,与节点出度成反比.

6 实验结果与分析

为了验证 CIndex 索引方法的有效性,本文通过 5 组实验表明,该算法在提高跨媒体检索性能、降低查询响应时间方面具有较好的性能.我们用 C++语言实现了跨媒体索引及查询算法.该算法采用 B+树作为单维索引结构且索引页大小设为 4 096 字节.所有实验的测试环境为 1 台 CPU 为 Pentium 2GHz,256MB 内存,80G 硬盘的 PC 机.本文采用的测试数据集是从 Internet 上随机下载的 100 000 张图片,2 000 个音频文件和 5 000 个视频文件.按照表 1 的方法提取每种媒体对象的特征并且使用该表中相似距离尺度作为相似匹配的标准.为了客观起见,以下每组实验都运行 100 次,取其均值作为实验结果.

6.1 查准率与查全率比较

在第 1 组实验中,研究查全率与查准率的比较.分别以图片、音频和视频例子作为提交例子,将得到的 3 种查询结果进行统计得到平均查全率和平均查准率.从图 9 可以看出,随着查全率的提高,查准率缓慢下降且本文的检索方法的查准率要高于 Octopus[12].这是因为 Octopus 系统在对网页的处理过程中,没有有效过滤掉一些“噪声”信息,这样导致最终得到的交叉参照图并不能准确地反映不同类型媒体对象的潜在语义关联.

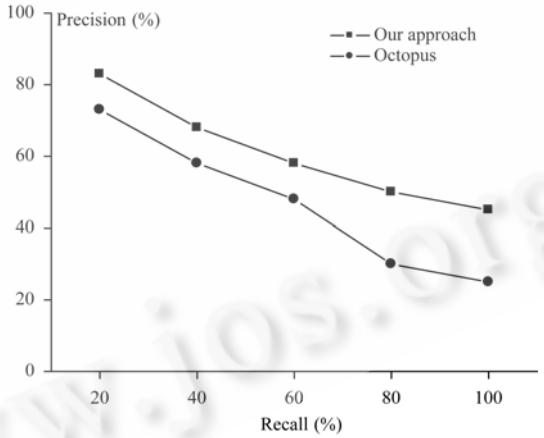
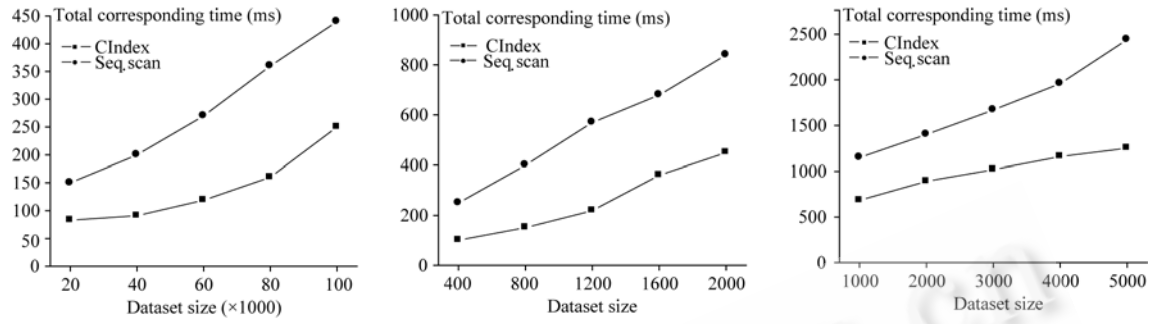


Fig.9 Recall vs. Precision
图 9 查全率与查准率比较

6.2 数据量对查询性能的影响

第 2 组实验研究数据量对跨媒体查询性能的影响.分别以图片、音频和视频例子作为提交例子,将得到的 3 种查询相应时间进行比较.从图 10 可以看出,在查询半径一定的情况下,随着各种媒体对象数据量的增加,检索

的时间也随之增加.基于 CIndex 索引的跨媒体检索性能要高于顺序检索.这是因为对海量交叉参照图的遍历是一个 CPU 密集的运算,而基于 CIndex 的方法可以快速地在交叉参照图找到所需要的媒体对象,因而其查询开销大为减少.同时可以看出,通过视频作为查询例子进行跨媒体检索所需要的时间最长.这是因为,基于内容的视频查询代价要远远高于音频和图像的检索代价.

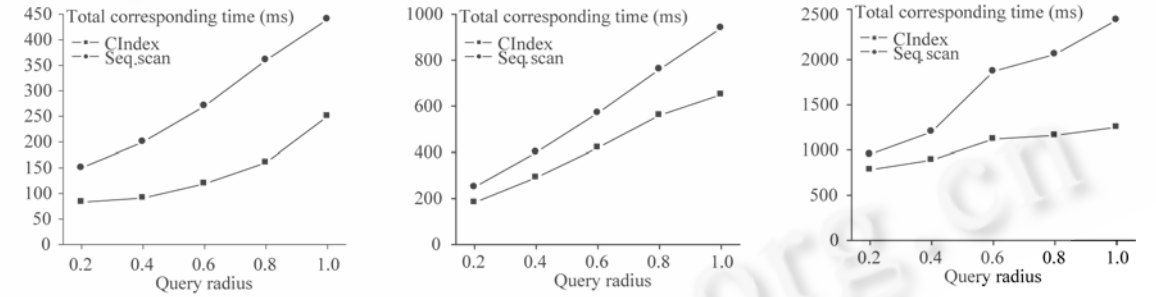


(a) Image (a) 图像 (b) Audio (b) 音频 (c) Video (c) 视频

Fig.10 Effect of data size on query performance
图 10 数据量对查询性能的影响

6.3 查询半径对查询性能的影响

本次实验研究查询半径对跨媒体查询性能的影响.同样分别以 3 种媒体对象作为提交例子.图 11 可以看出,在数据量一定的情况下,随着查询半径的增加,检索时间也随之增加.基于 CIndex 索引的跨媒体检索性能要高于顺序检索.这是因为对海量交叉参照图的遍历是一个 CPU 和 I/O 密集的运算,而基于 CIndex 的方法可以快速地在交叉参照图找到相关的媒体对象,有效地减少了查询开销.



(a) Image (a) 图像 (b) Audio (b) 音频 (c) Video (c) 视频

Fig.11 Effect of query radius on query performance
图 11 查询半径对查询性能的影响

6.4 索引存储代价比较

在本次实验中,研究两种跨媒体索引结构的存储代价.方法 1 采用基于二元组的叶节点索引表示,方法 2 采用基于三元组的叶节点表示.实验采用的测试数据为 100 000 张图片、2 000 个音频例子和 5 000 个视频例子.

从图 11 可以看出,在数据量一定的情况下,基于二元组叶节点方法的索引存储代价大大低于采用三元组的方法.且随着数据量的增加,两者的性能差别越来越大.这是因为,对于相同大小的数据量的媒体对象来说,采用基于三元组的索引叶节点的索引表示比基于二元组的索引的叶节点记录的数据量大为增加.因此其存储代价会有所不同.

6.5 索引更新对查询性能的影响

不失一般性,以图片作为提交例子,研究索引更新对查询性能的影响.本次实验需要进行 2 组实验.在第 1 组实验中,首先插入 80%的数据,然后分 4 次依次插入 5%的数据,每次分别执行范围查询且记录查询的时间;在第 2 组实验中,一次性分 5 种情况(即 80%,85%,90%,95%和 100%的数据量)建立索引且分别执行相同的查询.从图 12,图 13 看出,起初第 1 组实验的方法的查询代价与第 2 组实验一致.随着数据量的增加,两者的性能差异逐步增加.这是因为,对于第 1 组实验的 CIndex 来说,每次插入新的数据会导致聚类的结果较第 2 种方式要差,从而使得两者查询性能差异会随着插入数据的增加而变大.但提高的幅度较为缓慢.因此索引更新对查询性能的影响是可以接受的.

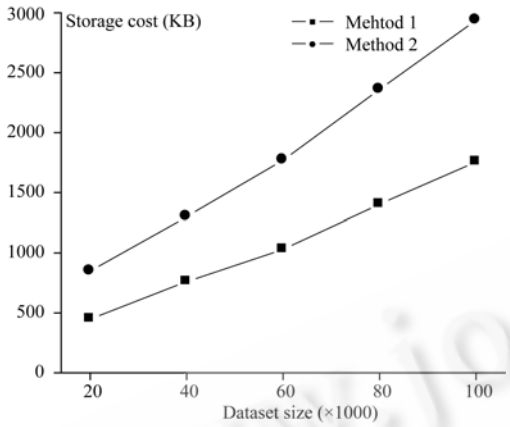


Fig.12 Comparison of index storage cost

图 12 索引存储代价比较

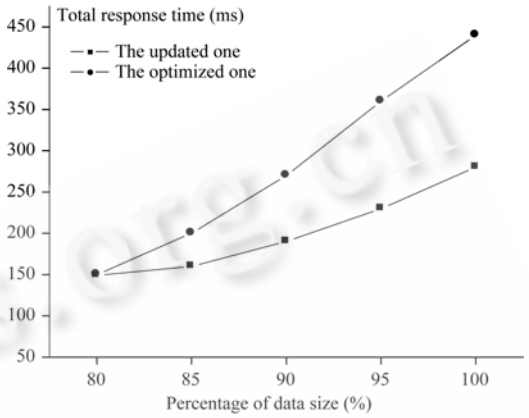


Fig.13 Effect of index update on query performance

图 13 索引更新对查询性能的影响

7 总结与将来工作

本文主要针对海量跨媒体检索提出一种集成的统一索引结构.该算法首先通过对多媒体文档的链接分析得到交叉参照图并且通过用户的相关反馈对其逐步修正和完善.然后提出一种集成的跨媒体索引结构——CIndex.理论和实验结果表明,相比顺序检索而言,本文所提出的海量跨媒体索引算法能够有效提高查询效率,同时具有良好的可扩展性.

海量跨媒体检索仍有大量尚未解决的问题,有很多的研究工作要做.在今后的研究中,将做以下几个方面的工作:1) 研究并行环境下的跨媒体检索,如交叉参照图的分片及并行检索算法等问题;2) 进一步研究通过其他机器学习方法进行媒体对象的相关性挖掘.

References:

[1] Zhuang YT, Pan YH, Wu F. Web-Based Multimedia Analysis and Retrieval. Beijing: Tsinghua University Press, 2002 (in Chinese).

[2] Flicker M, Sawhney H, Niblack W, Ashley J. Query by image and video content: The QBIC system. IEEE Computer, 1995,28(9): 23−32.

[3] Zhao XY, Zhuang YT, Liu JW, Wu F. Audio retrieval with fast relevance feedback based on the constrained fuzzy clustering and stored index table. In: Proc. of the PCM 2002. Berlin, Heidelberg: Springer-Verlag, 2002. 237−244.

[4] Hauptmann A, Jin R, Papernick N, Ng D, Qi YJ, Honghton R, Thornton S. Video retrieval with the informedia digital video library system. In: Proc. of the 10th Text Retrieval Conf.(TREC 2001). Gaithersburg, 2001.

[5] McGurk H, MacDonald J. Hearing lips and seeing voices. Nature, 1976,264:746−748

[6] Calvert GA. Cross-Modal processing in the human brain: Insights from functional neuron imaging studies. Cerebral Cortex, 2001,11(12):1120−1123.

[7] Wu F, Zhang H, Zhuang YT. Learning semantic correlations for cross media retrieval. In: Proc. of the ICIP. 2006. 1465−1468.

[8] Wu F, Yang Y, Zhuang YT, Pan YH. Understanding multimedia document semantics for cross-media retrieval. In: Proc. of the PCM. 2005. 993−1042.

[9] Virage Inc. 2005. http://www.virage.com

[10] Pentland A, Picard RW, Sclarof S. Photobook: Content-Based manipulation of image databases. Int'l Journal of Computer Vision, 1996,18(3):233−254.

[11] Mehrotra S, Rui Y, Chakrabarti K, Ortega M, Huang TS. Multimedia analysis and retrieval system. In: Proc. of the 3rd Int'l Workshop on Multimedia Information Systems. Como, 1997.

[12] Yang J, Li Q, Zhuang Y. OCTOPUS: Aggressive search of multi-modality data using multifaceted knowledge base. In: Proc. of the 11th Int'l Conf. on World Wide Web. Hawaii, 2002. 54−64.

[13] Shen HT, Zhou XF, Cui B. Indexing and integrating multiple features for WWW images. World Wide Web (WWW), 2006,9(3): 343−364.

[14] Böhm C, Berchtold S, Keim D. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. ACM Computing Surveys, 2001,33(3):322−373.

[15] Guttman A. R-tree: A dynamic index structure for spatial searching. In: Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. 1984. 47−54.

[16] Weber R, Schek H, Blott S. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In: Proc. of the 24th Int'l Conf. on Very Large Data Bases. 1998. 194−205.

[17] Jagadish HV, Ooi BC, Tan KL, Yu C, Zhang R. iDistance: An adaptive B^+-tree based indexing method for nearest neighbor search. ACM Trans. on Data Base Systems, 2005,30(2):364−397.

[18] Wu Y, Zhuang YT, Pan YH. Relevance feedback for video retrieval. Journal of Computer Research and Development, 2001,38(5): 546−551 (in Chinese with English abstract).

[19] Cai D, He XF, Wen JR, Ma WY. Block level link analysis. In: Proc. of the SIGIR. Sheffield, 2004. 440−447.

[20] Zhang T, Ramakrishnan R, Livny M. BIRCH: An efficient data clustering method for very large databases. In: Proc. of the ACM SIGMOD'96. 1996. 103−114.

附中文参考文献:

[1] 庄越挺,潘云鹤,吴飞.网上多媒体信息分析与检索.北京:清华大学出版社,2002.

[18] 吴翌,庄越挺,潘云鹤.视频的检索反馈.计算机研究与发展,2001,38(5):546−551.

庄毅(1978－),男,浙江杭州人,博士,讲师,CCF 会员,主要研究领域为多媒体数据库,索引优化.

吴飞(1973－),男,博士,副教授,CCF 高级会员,主要研究领域为多媒体检索,机器学习.

庄越挺(1965－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为多媒体检索,数字图书馆,视频动画.