视频数据库的聚类索引方法

施智平[1] 胡 宏[1] 李清勇[2] 史 俊[1,3] 史忠植[1]

[1](中国科学院计算技术研究所智能信息处理重点实验室 北京 100080)

[2](北京交通大学计算机与信息技术学院 北京 100044)

[3](中国科学院研究生院 北京 100039)

摘 要 理想的视频库组织方法应该把语义相关并且特征相似的视频的特征向量相邻存储.针对大规模视频库的特点,在语义监督下基于低层视觉特征对视频库进行层次聚类划分,当一个聚类中只包含一个语义类别的视频时,为这个聚类建立索引项,每个聚类所包含的原始特征数据在磁盘上连续存储.统计低层特征和高层特征的概率联系,构造Bayes分类器.查询时对用户的查询范例,首先确定最可能的候选聚类,然后在候选聚类范围内查询相似视频片段.实验结果表明,文中的方法不仅提高了检索速度而且提高了检索的语义敏感度.

关键词 视频库;高维索引;视频语义分类;聚类;视频检索

中图法分类号 TP18

Cluster-based Index Method for Video Database

SHI Zhi-Ping[1] HU Hong[1] LI Qing-Yong[2] SHI Jun[1,3] SHI Zhong-Zhi[1]

[1](*Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100080)

[2](*School of Computer and Information Technology, Beijing Jiaotong University, Beijing* 100044)

[3](*Graduate University of Chinese Academy of Sciences, Beijing* 100039)

Abstract A perfect video database organization should be that video feature vectors, which are not only semantic relevant but also their visual features themselves similar, are stored continuously. According to large-scale video database character, the authors hierarchically cluster feature vectors in video database supervised by video semantic classes until every cluster just contains such videos that belong to the same semantic class. The clusters here are called as index clusters. An index entry is created for an index cluster and a Bayes classifier is built with probability relationship between low-level features and the semantic class. For a given query, the first phase computes the distances between the query example and each cluster index and returns the clusters with the smallest distance, here namely candidate clusters; then the second phase retrieves the original feature vectors within the candidate clusters to gain the approximate nearest neighbors. The proposed method speeds up searching and improves retrieval semantic sensitivity.

Keywords video database; high dimensional index; video semantic classification; cluster; video retrieval

---

收稿日期:2005-04-20;修改稿收到日期:2006-09-21.本课题得到国家自然科学基金(60435010,90104021,60675010)、国家"八六三"高技术研究发展计划项目基金(2006AA01Z128)、国家"九七三"重点基础研究发展规划项目基金(2003C6317004)和北京市自然科学基金(4052025)资助.**施智平**,男,1974年生,博士,主要研究方向为基于内容的视觉信息检索、图像理解和机器学习.E-mail:shizp@ics.ict.ac.cn.**胡 宏**,男,1962年生,副研究员,主要研究方向为人工智能、图像模式识别.**李清勇**,男,1979年生,博士,主要研究方向为机器学习、图像理解、视觉信息挖掘.**史 俊**,男,1975年生,硕士研究生,主要研究方向为模式识别、图像理解.**史忠植**,男,1941年生,研究员,博士生导师,主要研究领域为人工智能、机器学习、神经计算、认知科学.

1 引 言

由于数据获取、存储和通信技术的飞速发展，人们可以轻易地获取大量的视频数据，建立大规模视频数据库.为了从视频数据库中查找感兴趣的视频数据，最初的方法是对视频数据按语义理解来分类并且标注关键词，然后按关键词匹配的方法检索.这种方法实现简单，符合人们对信息分类检索的习惯.但是关键词标注需要大量人工来完成，并且具有主观性.而且一般每个类内的数据量还是很大，不便于查找.近年来，研究者提出基于内容的视频检索(Content-Based Video Retrieval，CBVR)技术.CBVR 技术使用视觉特征向量来表示视频内容，视频样本之间的相似度用特征向量的相似距离度量.目前已经提出很多自动提取低层视觉特征的方法.但是视觉特征都是高维向量，对于大规模高维数据库还没有有效的索引机制来保证高效率的检索.而且用户的查询目标通常是具有相同语义的一类视频.由于语义鸿沟的存在，低层视觉特征不能准确反映语义概念.因此基于低层特征的查询结果往往难以满足用户对语义准确率的要求.

为了对两种方法取长补短，现实中的视频数据库中通常存在两种特征索引用于视频检索：(1)语义类别；(2)视觉特征.本文提出一种基于语义监督的特征聚类的索引方法，对视频库的数据兼顾语义相似和视觉特征相似地组织索引结构.我们用高斯混合模型来建模语义类别和视觉特征的分布关系.包含多个语义类别的数据集通过对视觉特征的聚类算法划分为多个子集，如果子集中仍包含多个语义类别，聚类划分迭代进行，直到每个子集只包含一个语义类别的数据.称这种子集为索引聚类，每个聚类内的视觉特征数据存储在连续的磁盘空间，聚类的均值和方差作为聚类的索引项.检索时，首先检索聚类的索引，确定候选聚类，然后读取属于候选聚类内的视觉特征数据，对这些特征数据查询并得到近似的 k 最近邻结果.

2 相关工作

2.1 高维索引技术

执行一次检索的时间主要包括磁盘读取(I/O)时间和距离计算时间.大规模数据库系统中，由于内存容量的限制，大部分数据存储在磁盘上，磁盘读取时间占有支配地位.为了减少磁盘读取时间，高维索引技术可以采用的方法有两种：(1)缩小每个数据特征的长度；(2)缩小需要访问的数据集.

缩小特征长度可以通过降维技术和向量近似(Vector Approximation，VA)的方法来实现.

基于降维的索引方法[1]首先对数据集进行降维处理，降维后的数据维度减少，然后再利用传统的多维索引对降维数据建立索引.降维方法可以在一定程度上克服维度灾难.但是降维操作会丢失检索精度，降维后的数据维度越少，检索精度越差.因此降维的方法一般需要和其他方法结合起来使用.

Weber，Ferhatosmanoglu，Ye 等[2-4]提出了基于向量近似的索引方法，用紧凑的近似向量表示原始数据.这种方法把数据空间划分成 2^b 个超方体形状的胞腔(cell)，用来近似估计原始特征向量.为每一个胞腔分配一个长度为 b 的唯一位串，并用这个位串近似表示落在该胞腔内部的原始特征向量.通过近似向量的过滤，只需访问少量原始向量就可以完成查询，提高了检索效率.这些候选原始向量在数据库中分散存储，读取时需要大量的磁头移动，而磁头移动是机械运动，是 I/O 中最耗时间的操作.要想进一步提高效率，理想的方法是把候选向量存储在连续的磁盘空间中.如果候选向量处于一个或几个胞腔中，每个胞腔中的向量是连续存储的，这样就可以减少磁头移动开销.遗憾的是这些划分方法都是基于维的划分，对于一个 30 维的向量，即使每维划分为两个区间，也会有 2^{30} 个超方体胞腔.这些胞腔大部分是空的，而每个数据向量都分布在不同的胞腔里.

这种缺陷是空间划分方法固有的，但是 Ye[4]的工作说明了高斯混合模型能更好地描述真实图像库的数据分布.

缩小需要访问的数据集可以通过限定语义范围和基于聚类的索引方法来实现.限定语义范围要求用户对视频库的语义标注信息有所了解.基于聚类的划分[5-6]是纯粹的基于数据分布的划分方法，一个聚类中的数据向量可以存储在连续的磁盘空间.这无疑可以提高检索速度，但是所达到的检索准确率取决于聚类算法的性能.

Ferhatosmanoglu 提出一个近似 k-NN 检索的框架[5-6].首先对原始特征向量作 KLT 变换，得到降维后的特征向量.然后对 KLT 变换域的较低维度的向量作 k-means 聚类，用聚类中心作聚类的索引.该文提出了渐进式求精的检索方法，先读入少量

最重要的维数和最近邻的聚类进行查询，返回首轮近似查询结果，随着查询的进行，逐渐读入更多的维数和聚类数据，以提高查询结果的精确性，直至用户满意才停止查询过程. 文献[5-6]的实验表明这种基于聚类的索引方法的性能高于现有的最好的高维索引方法.

2.2　视频语义分类

一般来说，广播视频包含各种视频节目，视频的语义分类就是可以将视频划分到预先定义的语义类别中的技术. 视频风格分类最早是在 1995 年由 Fischer[7] 等人提出，他们进行了视频风格分类领域的首次尝试，对新闻、广告、卡通、网球和赛车节目进行分类. 他们提出了一个分为三步的分类方法，首先提取基本的音频和视觉统计信息，包括视频片段中的场景颜色统计信息、运动、内容模式和声音等属性；第二步，利用这些统计信息导出更抽象的高级电影风格属性；最后映射上述检测得到的风格属性到电影风格类型，用它们的分布识别电影的风格类型.

Chen[8] 提出了一个基于知识的视频内容分类方法，他们在检查了五个视频分类中的许多视频后，形成了知识库中的分类规则. 该方法首先提取一些低级和高级视频特征. 然后将有关视频分类的知识编码成带有置信度的产生式规则，形成规则库，用于视频内容分类系统.

Mittal[9] 把视觉特征的各维的值离散化为小的区间，在训练集上学习各小区间对每个语义类的支持度，导出最小错误率的 Bayes 分类器. 划分小区间的原则是，尽可能在一个小区间里只包含一个语义类的数据样本. 文章的视频语义标注实验报告了非常好的结果. 这种各维分别划分的方法存在一个缺陷，如对图 1 中情况，二维特征空间中有两个语义类 A 和 B，区间 $[x_1, x_2]$ 和 $[y_1, y_2]$ 没有区分语义类 A 和 B 的能力. 把特征矢量分裂为单维的方法弱化了特征矢量的分类能力.

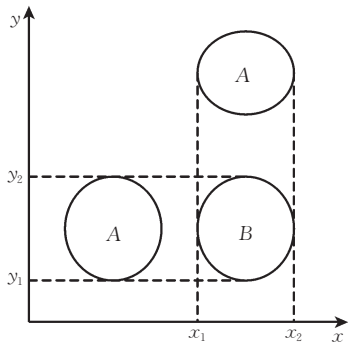
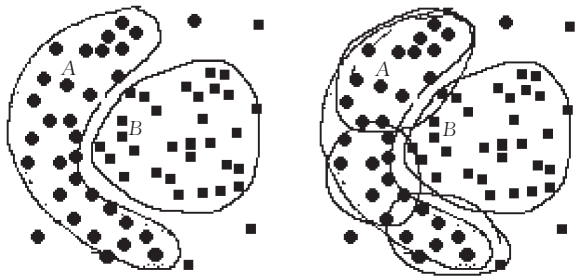


图 1　语义类在特征各维上的投影区间

3　视频数据的语义类别与特征分布

很显然，如果能够得到语义类别和视觉特征之间的联系就可以缩小语义鸿沟的困扰. 人们经常假设整个数据集的特征分布符合高斯混合模型(GMM)[4,11]，每个语义类的特征分布是一个高斯模型. 但是真实数据的分布是复杂的，一个语义类的特征分布有可能是任意形状的. 分析图 2(a)中所示的两个语义类的特征分布，如果各用一个高斯模型来表示两个语义类，A 的高斯模型必然包含大量 B 类样本. 因此本文采用更合理的假设，一个语义类内的样本特征分布仍然是多个分量的高斯混合模型分布. 理论上，高斯混合模型可以表示任意形状的数据分布[4]，而且具有成熟的模型估计算法. 图 2(a)中的 A 类样本特征用 3 个分量的高斯混合模型来表示，就可以避免包含 B 类样本，如图 2(b)所示. 本文采用 EM 算法估计 GMM 的参数.



(a) 两个语义类的样本特征分布　　(b) A 语义类是多个分量的高斯混合分布

图 2　语义类与特征分布示例

4　语义监督的聚类索引方法

我们希望建立这样的索引结构：整个视频数据集划分为许多子集，子集所包含的原始特征数据在磁盘上连续存储，索引文件中为每个子集建立一个索引. 如果子集的平均大小为 d，索引文件的大小就是原始特征数据库的 $1/d$，它比原始特征数据库小很多，可以快速读入内存. 在检索时，首先访问索引文件，计算查询向量和各个子集的相似度，然后只访问相似度最高的几个候选子集的原始特征向量，计算它们和查询向量的相似距离并返回最近邻结果. 因为每个子集中的特征向量是连续存储的，对候选子集的原始特征向量的 I/O 和基于近似向量的方法相比大幅度减少了磁头移动寻址时间，从而可能提供比基于近似向量的方法更好的性能.

那么,应该如何划分视频数据的子集呢?在CBVR系统中,最典型的查询方式是:给定一个查询范例,计算机提取视觉特征进行检索,而用户希望返回语义相关的结果.如果能够估计查询范例属于数据库中的哪个语义类别,就离用户的希望更近一步.采用最小错误率的Bayes分类器,把样本 x 分类到最小错误率的语义类 s_i 中,其分类错误率为

$$E=\int[1-\max_i p(s_i|x)]p(x)\,\mathrm{d}x \tag{1}$$

因为语义类别与特征分布的复杂性,无法精确估计特征到语义的直接隶属关系 $p(s_i|x)$.但是我们可以先对特征空间进行聚类划分,样本特征 x 到聚类 j 的概率关系和聚类 j 与语义类 i 的概率关系可以求得如下:

$$p(c_j|x)=\frac{p(x|c_j)p(c_j)}{p(x)} \tag{2}$$

$$p(s_i|c_j)=\frac{num(i|j)}{num(j)} \tag{3}$$

其中,$num(i|j)$ 表示聚类 j 中含有语义类 i 的样本数,$num(j)$ 表示聚类 j 的总样本数.

聚类集合 $\{c_1,c_2,\cdots,c_j\cdots\}$ 是样本空间的一个划分(集合的划分概念,可以参考离散数据的集合理论),所以有

$$p(s_i|x)=\sum_j p(s_i|c_j)p(c_j|x) \tag{4}$$

则

$$\begin{aligned}E&=\int\left[1-\max_i\sum_j p(s_i|c_j)\frac{p(x|c_j)p(c_j)}{p(x)}\right]p(x)\,\mathrm{d}x\\&=\int\left[p(x)-\max_i\sum_j p(s_i|c_j)p(x|c_j)p(c_j)\right]\mathrm{d}x\\&=\int\left[\sum_j p(x|c_j)p(c_j)-\max_i\sum_j p(s_i|c_j)p(x|c_j)p(c_j)\right]\mathrm{d}x\end{aligned} \tag{5}$$

式(3),(5)表明,当 $\max_i p(s_i|c_j)\to 1$,即一个聚类中只包含一个语义类的样本时,分类错误率最小.聚类算法既要保证能够反映样本特征空间的真实分布,又要保证每个聚类尽可能只包含一种语义的样本.

因此本文的聚类索引算法的基本思路是:视频库中所有的样本(镜头)按层次聚类,如果一个聚类中包含多个语义类的样本,这个聚类的样本作进一步的聚类划分,直到每个聚类的全部(或绝大部分)样本都属于同一个语义类为止,每个聚类建立一条索引.

如果视频库中语义概念是分层次的,生成的聚类索引也是分层次的.例如,视频数据库具有"新闻"和"体育"等语义概念,"体育"下面又分为"足球"、"棒球"等概念.那么聚类过程中,一个聚类 $C1$ 中的样本都是属于体育类的,就生成一条索引,但是体育不是最底层概念,其中样本还可以分为足球和棒球,所以还要继续划分为聚类 $C1_1$ 和 $C1_2$.而聚类 $C1_1$ 和 $C2$ 中分别只包含足球和新闻类样本,这两个语义概念是最底层概念,生成最底层索引,不需要再向下划分了.

5 视频语义分类

对于一个视频样本 x,按最小错误率的Bayes决策准则把它分到适当的语义类中,其语义类别为

$$C(x)=\arg\max_i p(s_i|x) \tag{6}$$

其中 $p(s_i|x)$ 可以根据式(2)~(4)计算.本文强制每个索引聚类中只有一种语义类的样本,所以

$$p(s_i|c_j)=\begin{cases}1, & \text{当 } j\in i(\text{表示聚类 } j \text{ 只包含语义类 } i \text{ 的样本});\\0, & \text{当 } j\notin i(\text{表示聚类 } j \text{ 不包含语义类 } i \text{ 的样本})\end{cases};$$

所以式(4)可以写为

$$p(s_i|x)=\sum_{j\in i}p(c_j|x) \tag{7}$$

为了简化协方差矩阵,可以先对样本空间作KLT变换消除样本向量各维之间的相关性,变换后协方差矩阵为对角阵,大大简化了计算,而且可以得到更精确的高斯聚类.

6 视频检索及其复杂性分析

对于本文的索引方法,有两种检索策略:

(a)先用Bayes方法对查询样本分类,其后的查询限制在概率最高的分类内,只需访问属于相关分类的原始特征向量.记做Class-q.

(b)先对聚类索引做最近邻检索,其后的查询限制在最近的 m 个聚类范围内,只需访问属于这 m 个聚类的原始特征向量.为方便记做Cluster-m,后面的数字 m 表示要访问原始向量的聚类数目.很显然 m 越大,原始数据访问量越大,检索结果越接近精确检索结果,时间开销也越大.

本节下面按照策略(a)讨论算法复杂性,策略(b)的算法复杂性也是类似的.

一次查询的检索时间(Search Time,ST)主要包括从磁盘上读取数据的开销(I/O Time,IT)和相似计算开销(CPU Time,CT).检索时间表示为

$$ST=IT_{index}+CT_{index}+IT_{db}+CT_{db} \quad (8)$$

其中 IT_{index} 是从磁盘读取索引文件的时间开销，CT_{index} 是查询样本对索引聚类隶属概率计算时间和语义类别推测计算时间，IT_{db} 是从磁盘读取数据库候选特征数据的时间，CT_{db} 是候选特征与查询样本特征的相似度计算时间. 每项时间开销都是所涉及的样本数的线性函数 $O(n)$.

设数据库总样本数为 N，索引聚类的平均样本数为 m，数据库语义类别数为 s，虽然各个聚类和语义类的大小不同，但是对各个聚类和语义类访问是等概率事件. 则本文检索算法的平均时间复杂度为

$$ST\sim O_I(N/m)+O_C(N/m)+O_I(N/s)+O_C(N/s) \quad (9)$$

其中下标 I 表示磁盘 I/O 时间，C 表示 CPU 计算时间.

对顺序检索算法而言，因为没有索引结构，其检索时间复杂度为

$$ST\sim O_I(N)+O_C(N) \quad (10)$$

对于大规模高维数据库，由于内存的限制，大部分的数据都存储在磁盘上，检索时间主要受 I/O 次数和 I/O 时间支配. 顺序检索方法中，如果内存不能容纳数据库全部数据，需要多次 I/O 才能完成磁盘数据访问，系统的性能将是无法容忍的. 本文的聚类索引算法中需要访问的磁盘数据为全部数据量的 $(1/(m+1)/s)$，对于大多数实际数据库系统可以一次 I/O 完成访问，检索速度的提高远远大于 m 倍或 s 倍.

VA-file[2] 和 VA$^+$-file[3] 索引方法的检索时间也可以用式(8)表示. 不同的是索引文件的数据量也为 N，而每条索引的位数(bits)要少，也可以一次 I/O完成访问. 索引的计算时间复杂度为 $O_C(N)$，大于本文的方法. 通过近似向量索引过滤，数据库候选访问数据量也大大小于 N，但是其最大的缺点是候选特征的访问是随机 I/O，而本文方法的候选特征访问是顺序 I/O. 顺序 I/O 的效率远远高于随机 I/O，因此本文的方法更具效率优势. 但是本文方法的适用前提是需对数据库进行语义分类，并且返回同语义类的查询结果，VA-file 和 VA$^+$-file 方法则不受此限制.

7　实验结果

7.1　语义分类实验

实验数据包括从电视录制的节目、VCD、DVD 等视频数据. 实验室已经收集了 300h 大约 150GB 的视频文件. 我们从中整理出约 40h 的视频，包括关键帧 40553 幅. 这些关键帧分为 20 个语义类别. 提取这些关键帧的 51 维纹理谱特征[12]，通过本章的聚类索引算法生成 3981 个聚类索引. 随机选取 12 类，每个类随机选取 10 幅关键帧作语义分类实验. 实验结果如表 1.

表 1　视频库语义分类平均准确率

类编号	语义类别	准确率
1	篮球	70
2	曲棍球	80
3	冲浪	90
4	乒乓球	90
5	足球	70
6	赛车	80
7	游泳	90
8	跑步	80
9	文艺节目	100
10	报道节目	70
11	广告	80
12	天气预报	100
avg	平 均	83.3

这个实验数据集极具挑战性，因为包括许多很难区分的语义类，如曲棍球和足球都以绿色场地为主要画面；冲浪和游泳都以水为主要画面；跑步和篮球的场地地面也有很多是类似的. 但是我们的实验仍然给出相当高的准确率.

检索策略 Class-q 中，如果对查询样本 q 的语义分类是正确的，只返回该语义类内的近邻样本，准确率是 100%. 如果语义分类错误，准确率为 0. 即检索的准确性能完全取决于分类正确率.

7.2　视频检索实验

首先进行基于聚类索引结构的快速检索实验. 为了进行比较，我们实现了 VA$^+$ file 的索引算法，对 51 维纹理特征生成 300bits 长的近似向量.

实验系统的特征库存放在服务器的 SQL Server 上，检索程序运行在本地机上，索引文件也存在本地磁盘. 本地机器是 P4 2.4GHz 的 CPU，512MB 内存，开发环境是 VC++6.0. 服务器是 P4 1.0GHz×2CPU，1024MB 内存，原始特征库在 SQL Server 上.

设定 $k=90$，随机做 10 次检索实验，取平均时间，两种索引算法的检索时间比较实验结果如表 2 所示. 因为 VA$^+$ file 方法平均访问 372 条原始特征记录，所以本文方法访问 30 个聚类的原始特征，即 Cluster-30，平均为 358 条记录，原始数据访问量大体相同.

表 2 两种索引方法的查询时间比较

时间开销	IT_{index}/s	CT_{index}/s	IT_{db}/s	CT_{db}/s	ST/s
Cluster-30	0.015	0.015	4.820	0.010	4.860
VA$^+$-file	0.109	1.286	5.450		6.845

表 2 表明,采用 C/S 结构和 SQL Server 数据库服务器方式,查询时间比较长.但是本文的方法比 VA$^+$ file 方法效率更高.尤其是查询索引文件的时间差异很大,因为本文方法是直接距离计算,而 VA$^+$ file 方法需要计算查询范例和近似向量的距离上下界,计算复杂度较高.在原始特征访问和距离计算阶段,两种方法的距离计算方法相同,区别只是 VA$^+$ file 是一条条记录的随机访问方法,而本文方法是逐个聚类访问的方法.

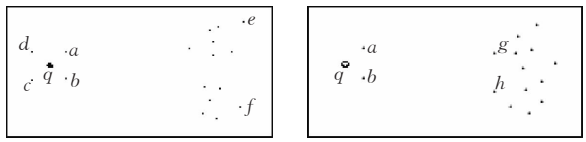
为了评价基于语义监督聚类索引的近似 k-NN 检索的查询质量,采用语义准确率比率和总距离比率两个指标[5-6].

语义准确率比率:设顺序检索的语义准确率为 p_s,近似检索的语义准确率为 p_a,则近似检索的语义准确率

$$P=\frac{p_a}{p_s} \tag{11}$$

显然,P 值越大,检索的语义准确率越高.如果近似检索的语义准确率小于顺序检索的准确率,P 大于 1,否则 P 小于 1.

语义准确率还不足以表现近似检索的近似性能.在图 3 中,与 q 点最近的 2 个点(精确 2-NN)是 a 和 b,近似检索算法可能返回其它结果,但是返回 (c,d) 的结果显然好于返回 (g,h) 的结果.为此使用总距离比率来衡量近似检索的近似效果.



(a) 例 1 (b) 例 2

图 3 对范例样本 q 的 2-NN 检索[6]

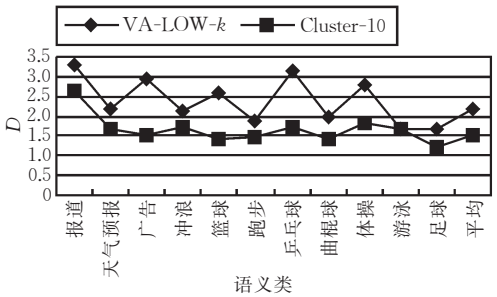
总距离比率:设近似 k-NN 检索算法返回的结果集为 $(a_1,a_2,\cdots,a_k)$,顺序检索的结果集为 $(r_1,r_2,\cdots,r_k)$,所用特征距离函数为 $d_f(q,x)$,总距离比率 D 定义为

$$D=\frac{\sum_{i=1}^{k}d_f(q,a_i)}{\sum_{i=1}^{k}d_f(q,r_i)} \tag{12}$$

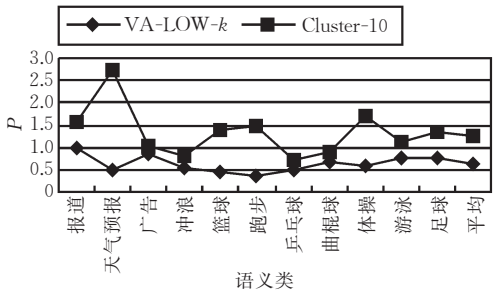
显然,$D\geq 1$,只有近似检索的结果和顺序检索结果即精确结果相同时,等号成立,D 越小,近似结果越接近精确结果.

文献[5,13]提出的 VA-LOW-k 近似算法,是 VA$^+$-File 基础上的近似检索算法.在第一间段得到查询范例与近似向量的距离下界,只保留距离下界的前 k 个结果作为检索的近似结果,访问这 k 个结果的原始特征向量,计算它们与查询向量的准确距离,按距离升序排列输出结果.

随机选 11 个类,每个类随机做三次检索实验取平均值,$k=100$,本文的索引算法取最近的 10 个聚类,记做 Cluster-10,特征距离计算采用欧式距离.两种近似检索算法的比较如图 4 所示.



(a)总距离比率 D 比较



(b) 语义准确率比率 P 比较

图 4 VA-LOW-k 和 Cluster-10 的检索性能比较

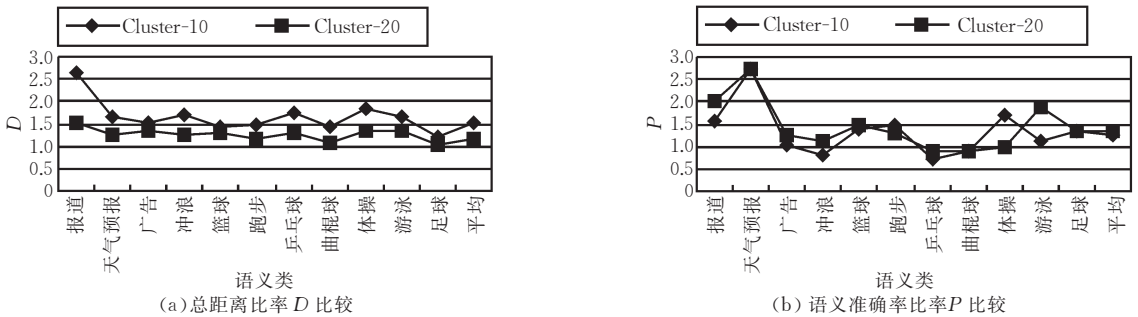
图 4(a)显示,Cluster-10 的 D 值显著小于 VA-LOW-k 的 D 值.表明 Cluster-10 作为近似检索比 VA-LOW-k 方法更接近精确检索.

图 4(b)显示,Cluster-10 的语义准确率比率总是高于 VA-LOW-k 的.Cluster-10 的准确率比率为 1.23,说明其准确率为顺序检索的准确率的 1.23 倍.VA-LOW-k 的准确率比率为 0.58,说明其准确率远低于顺序检索的准确率.一般情况下,近似检索的语义准确率低于精确检索的语义准确率.因为语义监督聚类索引方法在生成索引时考虑了语义信息,因此使得我们的近似检索算法有了超过精确检索的语义准确率.

Cluster-m 策略中 m 越大，原始数据访问量越大，检索结果越接近精确检索结果，时间开销也越大. 为了检验 m 值对检索性能的影响，设 $m=10$ 和 $m=20$，进行比较实验，实验结果如图 5 所示.

图 5(a)显示，Cluster-20 的 D 值小于 Cluster-10 的，更加接近 1. 随着 m 的增大，D 值逐渐趋向于 1，与精确检索的差异越来越小.

图 5(b)显示，Cluster-20 和 Cluster-10 的 P 值基本变化不大，就是说，与查询范例 q 最近的聚类，其语义符合查询语义的概率很高，因此不需要访问太多的聚类，就可以有较高的语义准确率.



(a)总距离比率 D 比较　　(b) 语义准确率比率 P 比较

图 5　Cluster-m 方法当 $m=10$ 和 $m=20$ 时的检索性能比较

8　结　论

视频图像数据库中的数据对象一般都有语义类别(关键词)和视觉特征两种形式的索引. 但是由于语义类别和视觉特征的聚类之间不存在直接对应关系，在视觉特征空间的一个聚类区间中，通常包含多个语义类的视频图像；一个语义类的视频图像可能分布在不连续的视觉特征区间里. 因此在现有的视频图像检索中，两种特征是独立使用的. 用语义特征查找相似的视觉特征以及用视觉特征查找相同语义类的视频对象都是困难的.

本文提出基于语义监督聚类的方法，把视觉特征空间划分为只包含一个语义类样本的小区间(聚类). 以此建立视觉特征到语义类别的概率联系. 这种概率联系可以用于视频图像的自动语义分类. 为每个聚类区间建立索引项，就成为一种高效的索引结构，并且提高查询效率和语义准确率.

参　考　文　献

[1] Kanth K V R, Agrawal D, Singh A. Dimensionality reduction for similarity searching in dynamic databases//Proceedings of the ACM SIGMOD International Conference on Management of Data. Seattle, Washington, 1998: 166-176

[2] Weber R, Schek H, Blott S. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces//Proceedings of the ACM Very Large Data Bases. New York, 1998: 194-205

[3] Ferhatosmanoglu H, Tuncel E, Agrawal D, Abbadi A El. Vector approximation based indexing for non-uniform high dimensional data sets//Proceedings of the 9th ACM International Conference on Information and Knowledge Management. McLean, Virginia, 2000: 202-209

[4] Ye H J, Xu G Y. Fast search in large-scale image database using vector quantization//Proceedings of the International Conference on Image and Video Retrieval, Lecture Notes in Computer Science. Urbana, IL, USA, 2003, LNCS 2728: 458-467

[5] Ferhatosmanoglu H, Tuncel E, Agrawal D, Abbadi A El. Approximate nearest neighbor searching in multimedia databases//Proceedings of the 17th International Conference on Data Engineering. Wanshington, DC, USA, 2001: 503-511

[6] Ferhatosmanoglu H, Tuncel E, Agrawal D, Abbadi A El. High dimensional nearest neighbor searching. Information Systems Journal, 2006, 31(6): 512-540

[7] Fischer S, Lienhart R, Effelsberg W. Automatic recognition of film genres//Proceedings of the ACM Multimedia 95. San Francisco, USA, 1995: 295-304

[8] Chen Y, Wong E K. A knowledge-based approach to video content classification//Proceedings of the SPIE Vol. 4315: Storage and Retrieval for Media Databases, 2001: 292-300

[9] Mittal A, Cheong L F. Addressing the problems of Bayesian network classification of video using high-dimensional features. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(2): 230-244

[10] Sheng Zhou, Xie Shi-Qian, Pan Cheng-Yi. Probability and Statistics. 2nd Edition. Beijing: Higher Education Press, 1989(in Chinese)
(盛　骤，谢式千，潘承毅. 概率论与数理统计. 第 2 版. 北京：高等教育出版社，1989)

[11] Xiang Ri-Hua, Wang Run-Sheng. A range image segmenta-

tion algorithm based on Gaussian mixture model. Journal of Software, 2003, 14(7): 1250-1257(in Chinese)
(向日华,王润生. 一种基于高斯混合模型的距离图像分割算法. 软件学报, 2003, 14(7): 1250-1257)

[12] Shi Zhi-Ping, Hu Hong, Li Qing-Yong, Shi Zhong-Zhi, Duan Chan-Lun. Texture spectrum descriptor based image retrieval. Journal of Software, 2005, 16(6): 1039-1045(in Chinese)
(施智平,胡 宏,李清勇,史忠植,段禅伦. 基于纹理谱描述子的图像检索. 软件学报, 2005, 16(6): 1039-1045)

[13] Weber R, Bohm K. Trading quality for time with nearest-neighbor search//Proceedings of the 7th International Conference on Extending Database Technology. Konstanz, Germany, 2000: 21-35

SHI Zhi-Ping, born in 1974, Ph.D.. His research interests include content-based visual information retrieval, image understanding and machine learning.

HU Hong, born in 1962, associate professor. His research interests include artificial intelligence, pattern recognition.

LI Qing-Yong, born in 1979, Ph. D.. His research interests include machine learning, image understanding and visual information mining.

SHI Jun, born in 1975, master candidate. His research interests include pattern recognition and image understanding.

SHI Zhong-Zhi, born in 1941, professor, Ph. D. supervisor. His research interests include artificial intelligence, machine learning, neural computing and cognitive science.

Background

In video database applications, the amount of data is very large and the dimension of data is very high. So it becomes necessary to support efficient retrieval in CBVR systems. Current approaches can be categorized into two general classes: 1) representative size reduction; 2) retrieved set reduction. The dimensionality reduction and VA-based indexing are examples of representative size reduction. The dimensionality reduction can overcome the curse of dimensionality to a degree. However, dimensionality reduction sacrifices some accuracy. The retrieved set reduction can be achieved by limiting search in some semantics range or by cluster-based indexing approaches. Some semantic video classification techniques are proposed to classify video data into pre-defined semantic classes.

Intuitively, it is reasonable to develop techniques that combine the advantages of both semantics and visual feature index. However, the visual features of some semantically relevant video clips may not be located very close in the visual feature space, or vice versa, the video objects with similar visual features may come from different semantic classes. In this paper, the authors propose a semantics supervised cluster based index approach (briefly as SSCI) to achieve the target. The main character of the technique is that it distinctly improves the search speed and the semantic precision of CBVR.

The paper is supported by the National Science Foundation of China No. 60435010: "Intelligence Computing Model Research Based on Perceptive Learning and Language Cognition". The project want to employ the conclusion of cognitive science and neural biology etc. to improve the intelligent information processing technology. The research group have done some efforts: Proposed a concept of texture pattern equivalent according to texture visual nature to educe more rational texture spectrum descriptor; a task-oriented sparse coding model for pattern classification; a model of attention-guided visual sparse coding; a two-layer feedback vision sparse coding model based on multilayer perceptrons; linguistic expression based image description framework; tolerance relation based information granular space, etc.