

基于语义的跨媒体信息检索技术研究

蔡 平, 王志强, 傅向华

(深圳大学 计算机与软件学院, 广东 深圳 518060)

摘 要: 阐述当前多媒体信息检索技术正从基于内容特征相似性的单一媒体检索发展到基于语义相关性的多种媒体综合检索, 实现跨媒体信息检索. 提出了一种跨媒体信息检索的系统结构, 在分析各种媒体基于语义信息检索的基础上, 设计并实现跨媒体搜索引擎及其查询分解策略和检索结果融合方法等, 实验结果表明: 这种方法能够有效地改善查全率和查准率.

关键词: 多媒体; 跨媒体; 信息检索; 语义挖掘; 检索机制

中图分类号: TP37

文献标识码: A

文章编号: 1000-7180(2010)03-0102-04

Research on the Semantics-based Cross-media Information Retrieval

CAI Ping, WANG Zhi-qiang, FU Xiang-hua

(College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China)

Abstract: Nowadays, the multimedia retrieval technology is moving forward from single media retrieval based on similarity of content feature to multimedia integrated retrieval based on semantic correlation which can be used for cross-media retrieval. In this paper, the architecture of cross-media retrieval is studied. Based on the analysis of semantic information among various media objects with different modalities, we propose and design a cross-media search engine, its query relaxation approach and query result integration method. The experiments' results show a great improvement for increasing recall ratio and precision ratio.

Key words: multimedia; cross-media; information retrieval; semantic mining; retrieval method

1 引言

多媒体信息检索的传统方法是基于文本描述的, 包括各种媒体的属性关键字和人工标注. 但是, 多媒体信息包含丰富的感知特征, 这些感知信息难以用文本反映出来. 于是, 基于内容的多媒体检索技术应运而生, 它们的研究主要集中在通过使用一些低层感知特征对多媒体内容建立索引, 从而进行样例相似度检索.

虽然传统的文本检索以及基于内容的多媒体检索已获得一定成功, 但它们却难以有效地应用到跨媒体信息检索中, 其主要原因是多媒体信息中包含了图像、视频和音频等不同模态的媒体, 且相互间具有语义关联^[1]. 那么, 是否可以跨越不同媒体之间

存在的界限, 使用户只需提交一种媒体的示例就能检索到想要的其他类型的媒体呢? 文中针对各种媒体语义进行了深入分析, 提出跨媒体信息检索系统结构、跨媒体搜索引擎以及查询分解策略和检索结果融合方法, 最后给出了实验结果.

2 跨媒体信息检索系统结构

跨媒体是在多媒体的基础上, 利用各种媒体的形式和特征, 对相同或相关的信息用不同的媒体表达形式进行处理, 由此而产生存储、检索和交换等活动^[2]. 在跨媒体信息环境下, 用户提交一种媒体对象作为查询示例, 检索系统不但可以返回相同种类的相似对象, 而且还能返回不同种类的其他媒体对象, 如利用图像示例检索语义相关的音频或视频片段等.

收稿日期: 2009-02-02; 修回日期: 2009-04-02

基金项目: 广东省自然科学基金项目(07301329); 深圳市科技计划资助项目(200741)

跨媒体信息检索系统需要最大限度地挖掘不同媒体之间相互表达、相互补充的语义关联性和协同效应,通过智能推理重构和新生知识,构建有效存储、管理和检索跨媒体信息的模型和系统。跨媒体信息检索的系统结构如图 1 所示,它包括跨媒体搜索引擎、跨媒体综合处理、数据存储和用户接口等模块。

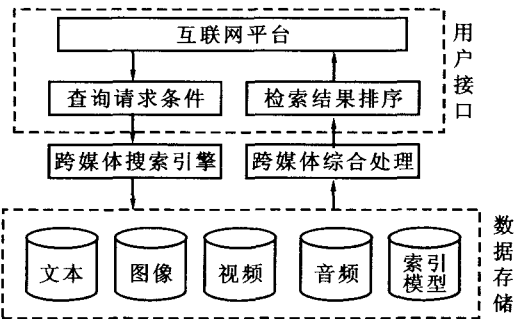


图 1 跨媒体信息检索系统结构

跨媒体搜索引擎模块用于实现跨媒体信息特征提取、识别和融合等功能,实现对不同种类的媒体之间的跨越检索以及对多媒体文档的检索等。跨媒体综合处理模块采用聚类 and 语义归纳等手段实现检索结果。数据存储模块主要存放多媒体实体数据、索引数据以及用户模型等。

用户接口模块要设计成多模态用户界面,它主要包括两部分:一是用户提出的查询条件,用户可能会提交一段文本、一幅或多幅样例图像、一段或多段样例视频和音频等来表达其查询要求;二是用户得到返回结果,可能会获得一段描述的文本信息、一组图像、一段视频或音频等。因此,不管是查询条件还是检索结果,都要求用户通过多模态的人机界面与系统交互。用户可以使用有限自然语言表达查询条件,也可以利用不同媒体的示例组合表达用户的查询条件,系统将利用跨媒体语义网理解用户的查询条件。根据这些语义关联通过跨媒体语义网,就可以找到用户所需的检索信息。

3 主要媒体语义检索技术

人们对多媒体信息的理解是基于语义级的,而目前研究的不同对象相似性匹配则是基于样例或低层特征进行的,两者之间存在的语义鸿沟阻碍了对用户真正感兴趣信息的检索。下面分别讨论不同媒体的语义检索技术。

3.1 图像语义检索

根据图像语义的复杂程度,可以把它分成如表 1 所示的 3 个层次。

表 1 图像语义层次模型

层次	名称	内容
第一层	特征语义	图像中的低层特征,如颜色、纹理、形状等
第二层	对象语义	图像中的对象和对象的空间关系等
第三层	抽象语义	图像的抽象属性,如场景语义、行为语义、情感语义

第一层次是特征语义,通过图像的颜色、纹理和形状等低层特征及其组合提取相关语义描述;第二层次是对象语义,通过识别和推理找出图像中的具体对象及其相互之间的空间关系,然后给出语义表达;第三层次是抽象语义,通过对图像所描述的对象、场景的含义和目标进行推理,得到相关的语义描述。第三层次的语义主要涉及图像的场景语义、行为语义和情感语义。其中,场景语义是指图像中的场景,如草原、天空等;行为语义是指图像中对象的行为及活动,如一场篮球比赛;情感语义是指图像给人带来的主观感受,如高兴、安详等,它们的结合形成了一幅图像所要表达的意义。

从图像中提取语义信息,实际上就是利用先验知识将低层视觉特征映射到高层语义。对于人类来说,这样的知识可以通过日积月累获得。但是,对于语义检索系统,只能是事先提供知识库,而这种知识库生成一般比较困难,通常要利用专家知识或进行大量的试验。由于通过人工干预生成的语义知识库工作量很大,因此可以利用机器学习和数据挖掘方法来尝试自动语义规则的生成方法,也可以利用人工智能方法来实现图像语义特征提取的方法^[3-4],即利用模糊逻辑、神经网络和遗传算法 3 者的融合来提取图像的语义信息,以便提高图像的检索效率和精度。

3.2 视频语义检索

视频语义检索是利用人类思维中高层语义概念来进行视频内容的过滤和检索,即访问视频信息的较为自然方法是通过高层语义概念来进行操纵,而低层特征对用户是不可见。图 2 表示视频的高层语义与低层特征之间的语义鸿沟。

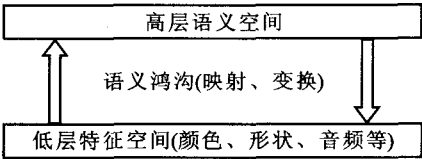


图 2 视频语义鸿沟

低层特征空间包括颜色特征、形状特征和音频特征等,这些特征一般可以自动从视频信息中提取.高层语义空间对应于人们思维中的高级语义概念.低层特征到高层语义空间的变换不是线性变换,这种变换或映射很难用数学方法描述,不容易建立模型,这就是语义鸿沟出现的根本原因.目前,提取视频语义的主要方法有采用模式分类的概率统计法和统计学习法、基于规则推理方法、结合特定领域方法以及基于注释方法、语义模板方法等^[4].

3.3 音频语义检索

音频内容可以划分为 3 个等级,即物理样本级、声学特征级和高层语义级,如表 2 所示.

表 2 音频内容的级别

层次	名称	内容
第一层	物理样本级	采样频率、量化精度、编码方法、时间刻度
第二层	声学特征级	感觉特征:音调、音高、音色、旋律、节奏 物理特征:线性预测系数、对数倒频谱系数
第三层	高层语义级	音乐叙事、音乐对象描述、语音识别文本

物理样本级是以媒体流的形式存在,包括原始音频数据和注册数据,如采样频率、量化精度和编码方法等.用户可以通过时间刻度来检索、浏览和调用音频的样本数据.声学特征级是从音频数据中自动抽取的,它可分为物理特征和感觉特征.感觉特征一般都与某些物理特征之间存在一定的联系,如有些特征表达了用户对音频的感知,可以直接用于检索.有些特征可用于语音的识别或检测,支持更高层的内容表示.高层语义级是音频内容、音频对象的概念级描述.

音频分类是音频语义检索的核心问题,音频自动分类的前提是音频的语义特征能够通过声学特征来表达,它们之间存在一定的映射关系.比较典型的音频分类算法有最小距离法、支持向量机、神经网络、决策树方法和隐马尔可夫模型方法等^[5].

4 跨媒体搜索引擎及检索机制

跨媒体搜索引擎是一种能够同时对多种媒体进行综合检索的搜索系统,它由元搜索引擎和成员搜索引擎组成.其中:成员搜索引擎包括文本搜索引擎、图像搜索引擎、视频搜索引擎和音频搜索引擎等类型,每一类型的成员搜索引擎还可以有多个,如 TE_1 、 TE_2 、 TE_3 或 AE_1 、 AE_2 等,如图 3 所示.

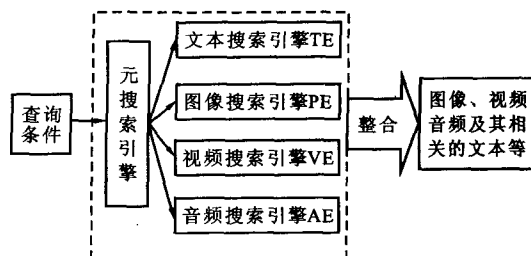


图 3 跨媒体搜索引擎结构

跨媒体搜索引擎的处理过程是将查询条件(如关键词、图像特征和语义等)分解成多个子条件传递给相应媒体的成员搜索引擎,而成员搜索引擎进行检索并返回相关的媒体信息,收集并整合检索出来的各种媒体信息,最终得到包含多种媒体的检索结果并集,即所需要的图像、视频或音频及其相关的文本等.

将用户的查询条件 Q 分解成若干个子条件,即 $q_1, q_2, \dots, q_n$ 等,然后根据其语义进行相关性组合,由于用户不能确定哪种组合条件最适合检索图像、视频或音频,因此在查询条件分解策略中, Q 被分解成一系列查询子条件的集合,例如:

$$\left\{ \begin{array}{l} \varphi, \{q_1, q_2, \dots, q_n\} \\ \{q_1\}, \{q_2, \dots, q_n\} \\ \{q_1, q_3\}, \{q_2, q_4, \dots, q_n\} \\ \{q_2\}, \{q_1, q_3\}, \{q_4, q_5, \dots, q_n\} \\ \dots \end{array} \right.$$

将分解或组合的子条件分别传递给不同媒体的成员搜索引擎进行检索,查询子条件的数量取决于成员搜索引擎的类型数量.例如,某个查询需要三类成员搜索引擎,查询条件将被分解为 3 个独立的子条件.第 1 个子条件集合可能传递给文本搜索引擎,第 2 个子条件集合可能传递给图像搜索引擎,第 3 个子条件集合可能传递给音频搜索引擎.也就是将查询条件 $\{q_2\}, \{q_1, q_3\}, \{q_4, q_5, \dots, q_n\}$ 中的 q_2 传递给文本搜索引擎,而 q_1, q_3 是图像搜索引擎的子条件, $q_4, q_5, \dots, q_n$ 是音频搜索引擎的子条件.通常,子条件中的数量越多,检索结果相对就越多.

查询条件 Q 的检索结果是对所有不同组合的子条件分别进行检索后所得结果的并集,即 $RES(Q) = Res(q_2, TE) \cup Res((q_1, q_3), PE) \cup Res((q_4, q_5, \dots, q_n), AE)$. 其中, $Res(q_2, TE)$ 表示子条件 q_2 通过文本搜索引擎 TE 进行检索后的结果集, $Res((q_1, q_3), PE)$ 表示将子条件 q_1, q_3 发送给图像搜索引擎 PE 后的检索结果集. $Res((q_4, q_5, \dots,$

q_n),AE)表示将子条件 $q_4,q_5,\cdots,q_n$ 发送给音频搜索引擎 AE 后的检索结果集.它们将返回给用户与检索条件相匹配的文本、图像和音频等媒体信息.

检索系统分别得到 $\text{Res}(q_2, \text{TE})$ 、 $\text{Res}((q_1, q_3), \text{PE})$ 和 $\text{Res}((q_4, q_5, \cdots, q_n), \text{AE})$, 然后对其进行整合,获得结果集为 $\text{Res}(q_2, \text{TE}) \cup \text{Res}((q_1, q_3), \text{PE}) \cup \text{Res}((q_4, q_5, \cdots, q_n), \text{AE})$, 最终的检索结果集 $\text{RES}(Q)$ 是所有检索结果集的并集.

5 仿真实验

假设查询子条件的数量为 3,从网页或报纸上随机选取 3 个语义关键词,如“北京,奥运,志愿者”,表 3 是对检索结果作一比较.

表 3 实验结果

组号	文本检索条件	图像检索条件	查全率/%	查准率/%
1	北京,奥运	志愿者	100	100
	北京,志愿者	奥运		
	奥运,志愿者	北京		
2	志愿者	北京,奥运	69	81
	奥运	北京,志愿者		
	北京	奥运,志愿者		
3	北京,奥运,志愿者		33	75

显而易见,用于文本检索的关键词数量越多,返回的相关信息就越多,检索效果就越好.假设以第 1 组信息为参考值,实验结果表明:文中所采取的查询策略和检索方法能够有效地改善查全率,查准率也能够保持在 75% 以上.

6 结束语

在分析各种媒体基于语义信息检索的基础上,给出跨媒体检索系统结构、跨媒体搜索引擎及其查

询分解策略和检索结果融合方法.使得基于低层特征相似性的单一媒体检索发展到基于语义相关性的跨媒体信息检索.该研究领域的发展不仅会创造巨大的社会效益,而且其应用前景也非常广泛,如多媒体数据库、数字图书馆、智能搜索引擎、综合医疗信息系统以及军事指挥系统等领域将得到广泛的应用.

参考文献:

[1] Zhuang Yueting, Yang Yi, Wu Fei. Mining semantic correlation of heterogeneous multimedia data for cross-media retrieval[J]. IEEE Transactions on Multimedia, 2008,10(2): 221-229.

[2] 曹加恒,李晶,蔡朝晖,等.新一代多媒体技术与应用[M]. 武汉:武汉大学出版社,2006: 379-410.

[3] 袁薇,高森.综合语义与颜色特征的图像检索技术[J]. 微电子学与计算机,2006,23(1):36-39.

[4] Ahmet Ekin, Murat Tekalp A, Rajiv Mehrotra. Integrated semantic-syntactic video modeling for search and browsing[J]. IEEE Transactions on Multimedia, 2004,6(6): 839-851.

[5] Umapathy K, Krishnan S, Jimaa S. Multigroup classification of audio signals using time-frequency parameters[J]. IEEE Transactions on Multimedia, 2005,7(2): 308-315.

作者简介:

蔡平女,(1960-),讲师.研究方向为多媒体检索、数据库技术等.

王志强男,(1963-),副教授.研究方向为多媒体与虚拟现实等.

傅向华男,(1976-),副教授.研究方向为信息检索等.

(上接第 101 页)

[2] 徐雅斌,李昕,张晓宇.基于 GPS 的车辆监控系统的设计[J]. 微电子学与计算机,2003,20(9):31-34.

[3] Elliott n Kaplan, Christopher J Hegarty. GPS 原理与应用[M]. 寇艳红,译. 北京:电子工业出版社,2002(8):111-179.

[4] 程佩青. 数字信号处理教程[M]. 2 版. 北京:清华大学出版社,2001:87-132.

[5] Parkinson B W, James J. Global positioning system theory and applications: Volume I[M]. Washington: American

Institute of Aeronautics and Astronautics Inc,1996:395-396.

[6] Tsui J B. Fundamentals of global positioning system receivers: a software approach[M]. 2nd ed. New York: Wiley, 2004: 72-79/258-263.

作者简介:

陈学男,(1983-),硕士研究生.研究方向为无线通信信号处理算法设计.