

●张玉峰 蔡昌许

## 基于语义的图像检索系统研究<sup>\*</sup>

**摘 要** 基于语义的图像检索系统的构建包括以下工作:设计高层语义的描述方式;图像语义提取,即如何从视觉特征映射到高层语义;语义检索系统的实现。基于 MPEG-7 图像语义的 XML 描述和系统检索的实现方法,构建了基于语义的图像检索系统的框架。图 4。参考文献 8。

**关键词** 语义图像检索 图像语义层次 图像语义标注 语义描述

**分类号** G354

**ABSTRACT** In this paper the authors summarize major aspects of the construction of a semantic image retrieval system, and introduce the XML description and system realization method based on MPEG-7 image semantics, which can be used to construct a framework for a semantic image retrieval system. 4 figs. 8 refs.

**KEY WORDS** Semantic image retrieval. Semantic level of images. Semantic markup of images. Semantic description.

**CLASS NUMBER** G354

早期的图像检索技术主要是基于文本的图像检索。20 世纪 90 年代,出现了基于内容的图像检索。它可以直接从图像信息源中获得视觉内容特征,如颜色、纹理、形状等来判断图像之间的相似性。

但是在实际应用中,用户往往事先对所需的图像只存在有关图像描述的对象、事件以及表达的情感等含义上的概念,用户需要的是图像含义的查询,而不是颜色、纹理、形状等特征。这些图像的含义就是图像的高层语义特征,它包含了人们对图像内容的理解,这种理解是无法直接从图像的视觉特征获得的,而要根据人们的认知知识来判断。因而,基于语义的图像检索需要综合应用多媒体技术、人工智能、信息科学、认知科学等多学科知识来实现图像语义特征的提取、表示和检索。

要使图像检索技术真正满足普通用户的需求,必须利用更多的语义信息。一个理想的检索系统应该提供基于内容的检索和基于语义的检索方法。基于语义的图像检索系统的构建包括以下几方面的工作:设计高层语义的描述方式;图像语义提取,即如何从视觉特征映射到高层语义;语义检索系统的实现。其中图像语义的提取是比较困难的工作。下面介绍一种基于 MPEG-7 标准的语义描述和检索系统实现方法。

### 1 图像语义层次及其标注

#### 1.1 图像语义层次

语义就是人们对图像的理解。语义的层次在图像检索中说明了描述和检索的复杂度,从低层到高层,分为特征语义层、对象语义层、空间关系语义层、场景语义层、行为语义层、情感语义层等。对于特征语义,可以通过图像的相似性来检索。对于绝大多数用户来说,他们习惯用自己的理解来描述图像,而不习惯用颜色、纹理等图像特征语义来表示。对于更高的语义层,必须通过一定的知识推理,识别出图像所包括的对象、空间关系,如对象语义层,空间关系语义层。语义层次越高,必须有更高层、更智能的推理判断过程,如场景语义层、行为语义层、情感语义层等,这是一个主观判断的过程。由于各用户的差异性,在检索的推理过程中必须体现个体的差异性。在这一层上,还需要有建立用户知识库的过程,知识库的好坏对检索结果的影响很大。

#### 1.2 图像语义标注

人们对图像的理解所产生的语义,必须通过一定的方式表达出来以供应用系统处理。图像表示方法有多种,其中最直接最简单的就是用文本表示,由于文本方式对于图像检索的局限性,这种表示方法

<sup>\*</sup> 本文为教育部人文社会科学研究博士点基金项目(项目号:03JB870002)的研究成果。

不具有普遍性。有些专家采用了人工智能领域的一些表示方法,如语义网络、数理逻辑、框架方法等来表示知识。本系统采用的是基于 MPEG-7 的图像语义的 XML 表示方法。

### 1.2.1 MPEG-7 和 XML

MPEG-7 标准被命名为“多媒体内容描述接口”(Multimedia Content Description Interface),它可以独立于其他 MPEG 标准使用。它旨在解决对多媒体信息描述的标准问题,并将该描述的内容相互关联,以便实现快速高效检索。MPEG-7 规定了一个用于描述各种类型的多媒体信息的描述符的标准集合,它还定义了其他描述符及结构的标准。

XML(扩展标识语言)是 SGML 的一个子集,它和 HTML(超文本标识语言)都是从 SGML 发展而来的文档形式。XML 需要自己定义数据类型,有两种方式:文档类型定义 DTD 和 Schema。XML 有强大的描述能力、灵活性、可扩展性等优点,应用越来越广泛。现在 XML 已经成为数据描述和交换的标准,因此针对 XML 的半结构化特性,可以实现比传统图像检索更好的检索效果。

### 1.2.2 MPEG-7 标准的图像描述方案

用 XML 描述图像,首先要定义所用到的规则和标注系列。图像描述体系的基础是对象元素,一个对象元素代表图像的一个区域。对象分为物理对象和逻辑对象。物理对象对应于图像的实际对象,逻辑对象是按高级的语义关系组织起来的,每个对象在一个图像描述中有一个惟一的标识符。为了提高检索效率,必须对图片进行分类,一定数量的图片描述为一个 XML 文档。图 1 为图像的 XML 描述示例,包括了对象、动作等语义的描述。

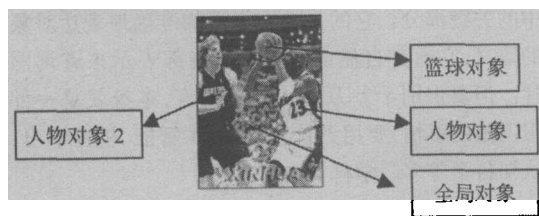


图 1 描述示例图像

该图的语义描述示例如下:

```
<SemanticBase id="player _ obj181" xsi:type="
AgentObjectType">
  <Label href="...">
    <Name>篮球运动员</Name>
  </Label>
```

```
<Definition>
  <FreeTextAnnotation>华盛顿奇才队和达拉斯小
牛队进行比赛中,诺维斯基防守丹</
FreeTextAnnotation>
</Definition>
<Agent xsi:type="PersonType">
  <Name>
    <GivenName></GivenName>
    <FamilyName>乔丹</FamilyName>
  </Name>
</Agent>
</SemanticBase>
<SemanticBase id="player _ obj181a" xsi:type="
AgentObjectType">
  <Label href="...">
    <Name>篮球运动员</Name>
  </Label>
</Definition>
<FreeTextAnnotation>达拉斯小牛队和华盛顿
奇才队进行比赛,诺维斯基防守乔丹</
FreeTextAnnotation>
</Definition>
<Agent xsi:type="PersonType">
  <Name>
    <GivenName></GivenName>
    <FamilyName>诺维斯基</FamilyName>
  </Name>
</Agent>
</SemanticBase>
```

## 2 基于 MPEG-7 标准的语义图像检索系统

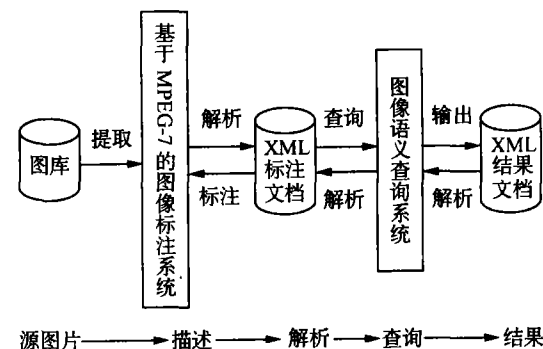


图 2 基于语义的图像检索系统整体结构

本文提出的检索系统是在图像语义标注系统的

基础之上进行的。图像语义标注主要完成图像语义的表示和提取工作,检索系统主要完成应用程序界面设计和对标注结果进行操作(这里指基于 MPEG-7 的图像语义的 XML 标注文档)。图像标注系统把图像语义标注为 XML 文档,图像语义查询系统通过对 XML 文档的操作,实现 XML 文档内容的语义内容与用户检索词的逻辑匹配,从而得到查询结果。基于语义的图像标注和检索系统的整体结构如图 2 所示。

### 2.1 图像查询流程

图像信息的查找过程可分为两步:首先定位查找的文档,然后执行图像语义查询。定位文档通常有以下几种方法。

- (1)通过操作系统中的文件名定位文档。
- (2)利用文档相互索引,这种方法类似于传统的 Web 站点。
- (3)通过数据库索引文档。可以把 XML 文档保存为文件形式,然后通过数据库索引,或者将 XML 也存储在数据库中。也可以使用二进制大对象(Binary Large Object)字段。如果利用关系数据库来索引 XML 文档,就可以使用 SQL 访问它们。

本文采用文件名定位方式查询,通过对 XML 文档的语义解析,查找解析后的 DOM 树结点的内容,从而得到查询结果。

本检索系统的查询流程如图 3 所示,主要包括定位 XML 索引文档,根据查询条件执行图像语义查询,然后浏览图片以确认图片,输出所需图片到 XML 结果文档,以备提高查询效率或供其他应用程序所用。查询过程中,首先对 XML 索引文档进行语义解析,解析方式有如下两种。一种是基于文档对象类型 DOM(Document Object Model)的解析方式,将索引文档解析为树结构,应用程序可以随机和反复对树进行操作,但解析的树是在内存里进行的,比较占用内存,当树结构比较大时,运行速度会比较慢。另一种是基于 SAX(Simple API For XML)的解析方式,是一种事件驱动,通过串行方式来处理文档,与 DOM 方式相比,需要的内存要少得多。它是一种比较简单的解析方式,运行速度比较快。本系统采用的是基于 DOM 的解析方式,然后利用树的遍历算法,找到叶子结点,与输入的关键字进行匹配。如果包含查询关键字,则输出并进行图像浏览;如果是用户需要的图片,则将其加入到 XML 结果文档里面保存;如果没找到,则重新修改查询条件继续查询。

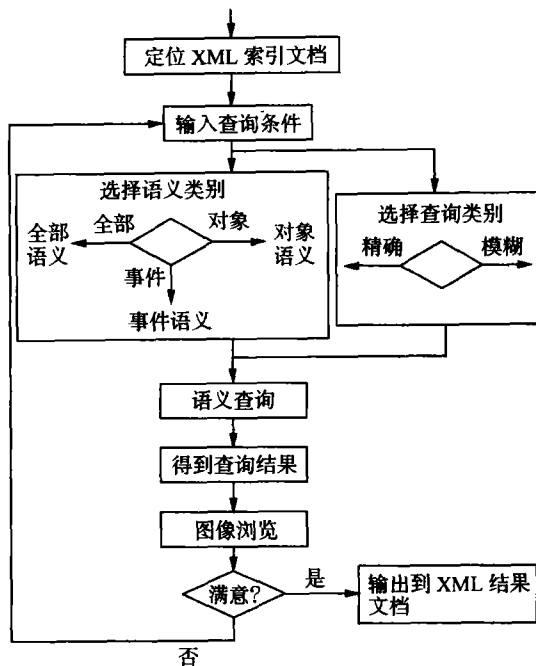


图 3 查询流程

### 2.2 检索系统的交互界面设计

界面设计对一个图像查询系统来说相当重要。它不但可以体现所用到的关键技术,还可以体现人性化的一面,首先要求操作方便,界面友好。对于图像检索的输入界面来说,主要有两种方式。

一种是以关键字输入方式。这种方式比较简单,但由于各用户的差异性,各自对图形的语义描述不一样,或者很难用简单的文字方式描述出来,有时很难查询出用户所需要的结果。现在有很多研究者提出自然语言查询。自然语言是人们在日常生活和工作中使用的语言,它是需求表达和查询处理过程中的关键部分。它的目的在于更加准确地表达对象的语义,它通过自然语言的机器内部表示来查询图片。但要识别千差万别的自然语言,本身就是一件很困难的事情,要用到统计学、人工智能、机器学习、认知科学等多门学科。

另外一种方式是用户提供范例图或者草图,系统首先提取范例图或者草图的语义,然后在图库的语义中进行相似检索而得出结果。但机器很难自动提取所提供的图像语义,需要与用户不断交互,让软件自动调整语义扩展的算法和扩展的深度。通过多次的学习和自适应后,一般情况下软件可以实现较好的检索结果,但有时,返回的结果不能使用户满意。由于每种输入方式都有优缺点,提供多种查询

方式的交互界面是一种可行的解决办法。在检索中要对关键词和高频词等自动建立索引库,从而能从底层结构上提高检索速度。

### 2.3 语义检索系统实例

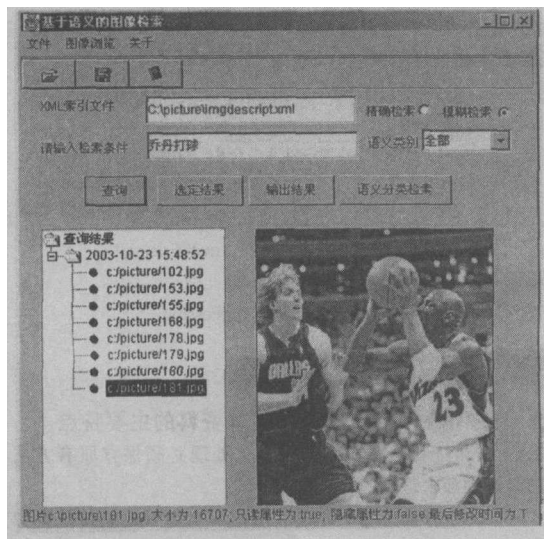


图4 基于 MPEG-7 的语义检索界面

本系统对 1000 张基于 MPEG-7 标准标注过的图片进行检索,采用输入语义关键字方式,根据输入的关键字进行查询。图 4 为本系统的交互界面,它分为查询区、结果区和图形显示区。其中“选定结果”按钮为用户选定满意的图片,“输出结果”按钮则把用户选定的结果输出到 XML 结果文档,为以后提高查询效率和统计查询关键字使用。下拉列表用来选定所要查询的语义类型,可以提高查询效率。状态栏表达所选定图片的基本属性。

通过对 1000 张图片的实验,结果表明,在一定的硬件条件下,只要单个 XML 文档不是太大,查询速度很快,如本实例中 1000 幅图片描述为一个 XML 文档。对一幅图片只要语义描述得详细,检索的效果一般比较好,特别是对于对象、事件、行为等语义的描述与检索。

### 3 展望

为解决多媒体信息的急剧增长与人们对媒体信息需求困难的矛盾,各种多媒体信息检索技术不断出现,从基于文本的图像检索到基于内容的图像检索,再发展到更高语义层的检索,但是很难找到一种

完全满足用户的检索方式。只有将几种方式结合起来,采用多种检索方法和多样化的交互方式,才能更全面、更完美。基于内容的检索和基于语义的检索结合在一起,既满足低层客观的特征语义信息,又能提供复杂多变的高层语义的检索功能。笔者认为在语义检索层面上,最困难的是语义的提取过程。由于海量的图片,不可能对每幅图片进行手工标注,必须采用自动化或者半自动化的标注方式。自动化标注与人工智能密切相关,建立有关知识库是一个关键。

### 参考文献

- 1 Yang Rui, and Thomas S. Huang. Image Retrieval: Past, Present, And Future. International Symposium on Multimedia Information Processing, Dec 1997
- 2 H. Müller, W. Müller, D. McG. Squire, and T. Pun. Performance evaluation in content-based image retrieval: Overview and proposals. Technical Report 99. 05, Computer Vision Group, CUI, University of Geneva, Switzerland, 1999
- 3 Yueting Zhuang, Sharad Mehrotra, Thomas S. Huangy. A Multimedia Information Retrieval Model Based On Semantic And Visual Content. Proceedings of the 5th International ICYCS Conference, Nanjing, China, 1999.6
- 4 王惠锋,金翔宇,孙正兴. 语义图像检索研究进展. 计算机研究与发展, 2002(5)
- 5 万华林,胡宏,史忠植. 信息检索的革命——基于内容的多媒体信息检索. 微电脑世界, 2001(1)
- 6 曾春平. XML 编程从入门到精通. 北京:北京希望电子出版社, 2002
- 7 余兆明,李晓飞,陈来春. MPEG-7 标准机器应用. 北京:北京邮电大学出版社, 2002
- 8 Chih-Cheng Hsu, Wesley W. Chu, Ricky K. Knowledge-Based Approach for Retrieve Images by Content. Taira Knowledge and Data Engineering. August 1, 1996

张玉峰 武汉大学信息资源中心教授,博士生导师。

通讯地址:武汉大学信息管理学院. 邮编 430072。

蔡昌许 武汉大学信息管理学院 2002 级研究生。通讯地址同上。

(来稿时间:2004-02-27)