

XML 驱动的图像语义检索研究

陈久军 肖刚 盛颂恩 高飞

(浙江工业大学信息学院, 杭州 310014)

摘要 本文研究面向语义检索的图像内容描述机制。首先提出图像语义检索整体框架,系统采用XML技术,将图像内容层次描述、图像语义对象自动获取、图像语义相似测度等功能模块加以融合,实现语义层面的图像检索。重点对系统框架中与图像内容描述相关的图像特征分层描述模型、空间位置算子定义、语义对象操作等关键技术进行讨论,并定义相应的XML语义描述框架。检索实验结果表明,该方法具有较好的语义检索性能。

关键词 语义检索 XML描述 分层描述 图像检索

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2007)10-1770-04

XML-driven Image Semantic Retrieval System

CHEN Jiu-jun, XIAO Gang, SHENG Song-en, GAO Fei

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014)

Abstract In this paper we firstly present an XML-driven image semantic retrieval system, which consists of several functional modules, i. e. image content hierarchy modeling, semantic object automatic extraction, and semantic similarity measure. Secondly, the paper concentrates on the image character description model, and we discuss the definitions and operators of image object. Then some related XML semantic descriptors are defined to represent the image content. Experiment results on traffic images show good retrieval quality.

Keywords semantic retrieval, XML descriptor, content hierarchy, image retrieval

1 引言

如何实现图像语义检索,是目前信息检索研究中一个重要的课题。其核心问题在于如何自动获取、有效描述图像语义内容^[1]。目前相关研究包括:对不同层次的图像内容分别进行分析与提取^[2];将低层特征在时空中组合起来构成语义单元^[3];基于语义直接检测事件^[4];Lin^[5]利用混合Bayesian方法实现了医学图像的语义检索。但目前还没有一种完整的基于语义的图像内容描述方案来支持实际的检索过程。

XML是一种用于定义标记语言的数据描述标准,具有简单性、开放性、可扩展性、灵活性和自描述

性等特性。它可以将大量信息组织成为具有确定意义的整体结构。为图像语义内容的有效描述提供了一个全新的革命性的应用框架。

本文通过引入XML描述机制,提出XML驱动的图像语义检索系统。重点研究框架中与图像内容描述相关的图像特征分层描述模型、空间位置算子定义、语义对象操作等关键技术,定义相应的XML语义描述框架,实现语义图像检索。

2 图像内容语义描述

2.1 系统框架

系统框架如图1所示,包含3个模块:模块1,特征提取与语义学习;模块2,图像内容的XML层

基金项目:国家重点基础研究发展计划973项目(2003CB317000);浙江省科技厅项目(2005C31010)

收稿日期:2007-07-05;改回日期:2007-07-25

第一作者简介:陈久军(1977~),男,讲师,2006年获浙江大学计算机应用专业博士学位。主要研究方向为图像分析、机器视觉。

E-mail: rackycj@zjut.edu.cn

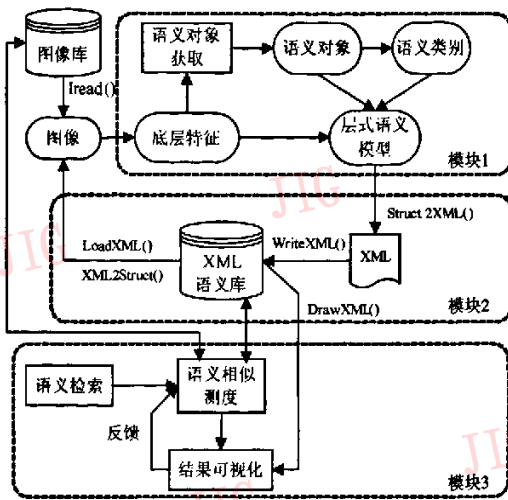


图 1 XML 驱动的图像语义检索框架

Fig.1 XML-driven image semantic retrieval architecture

次描述;模块 3,应用语义测度的检索。

模块 1,图像数据库包括低层特征、语义对象,以及相关的语义类别信息。图像库是可拓展的,可以通过手工标注的方式,添加图像语义对象边界和相应的语义描述;也可通过语义对象识别器实现对未标注图像的自动语义获取与标注。

模块 2,处理图像内容的 XML 描述,有效衔接语义获取模块与检索模块。模块 2 将模块 1 所获取的层式图像内容描述转换为 XML 文件(函数 StructZXML())。利用 XML 的灵活性与强大的语义表达能力,建立关于图像的 XML 语义数据库(函数 WriteXML())。为了增强语义的交互性。XML 语义库中的 XML 文档可以重新读出(函数 LoadXML()),通过 XMLZStrust() 函数转为图像原始标注信息,进行修改与完善。图像语义内容的 XML 描述,为图像语义检索提供简洁的、全面的、分层的、可拓展的图像语义信息。

模块 3,基于所构建的 XML 语义库和图像库,根据需要检索的语义信息,与 XML 语义库中的语义描述进行相似匹配,获取所需语义的图像信息,并将原始图像与标注信息以可视化的方式传递给用户。框架同时引入了反馈机制,以提高检索性能。

2.2 图像内容的层式描述

根据不同的语义粒度,将图像内容分为(1)低层特征,包括颜色、纹理、形状等;(2)元语义数据,包括图像中的对象元素、对象几何特征、空间位置关

系、对象语义映射等;(3)高层语义特征,包括图像场景、行为语义、情感语义等。

定义 1 图像对象集合 定义图像为语义对象的集合:(1)空间对象信息;(2)对象形状几何特征、空间关系;(3)对象语义描述,如式(1)所示:

$$I = \{ \langle O_i : (Polyon(PointSet, Center, Area), Color) \rangle \quad (1)$$

$$R_{O_i}, M_{..} \mid i, j \in (1, n), O_i \in O, S \in MetaSemanticSet \}$$

式(1)表示图像 I 由包含特定属性和相互关系的 n 个语义对象组成,其中, $Polyon(PointSet, Center, Area)$ 定义为对象 O_i 的几何特征,表示对象的边界多边形,它包含 3 个参数: $PointSet$ 是边界多边形特征点的集合; $Center$ 是边界多边形的中心点,用 $Center$ 来构建对象 O_i 的空间位置关系; $Area$ 表示对象的区域面积;

$Color$ 表示对象的颜色增强特征,与低层颜色特征不同,它表示的是对象区域范围内颜色特征;

R_{O_i} 表示对象 O_i 与 O_j 的空间位置关系。本文定义 12 种位置算子、2 种结构算子和 3 种操作代数;

$M_{..}$ 是元语义与对象集合的映射关系:

$$M_{..} = f(MetaSemanticSet, O_i) (i \in (1, n)) \quad (2)$$

其中, $MetaSemanticSet$ 是预定义元语义和附加元语义的集合, O_i 表示图像中的对象。

2.3 空间位置算子

定义对象边界多边形的中心点来表征对象元素。因此,在式(1)的基础上,图像可以进一步定义为空间位置点的集合

$$I = \{ \langle P_1, \dots, P_i, \dots, P_n \rangle, SpatialRelation(P_i, P_j) \} \quad (3)$$

$$P_i = O_i \cdot Center \quad i, j \in (1, n)$$

定义 2 对象空间位置图 假设图像 I 中包含 n 个对象,则 n 个对象的中心点集合可以表示为 $\{ O_i \cdot Center \mid i \in (1, n) \}$,定义由这些点组成的无向图为对象空间位置图,满足:

(1) 对象的中心点对应空间位置图的各个顶点,所有点组成节点集合,如式(3)。

(2) 链接对象间的边定义为对象间的空间关系,所有边构成对象关系集合

$$E = \{ e_1, e_2, \dots, e_j, \dots, e_m \} \quad (4)$$

(3) 空间位置图定义为 P 与 E 的二元函数:

$$G = (P, E) \quad (5)$$

(4) 定义 6 种空间位置关系:左、右、前、后、上、下。将 XY 水平面上的 4 个位置算子组合产生 4 组位置算子,再加上 Inside 和 Outside 两种,共计 12 种

位置算子,如表 1 所示。

表 1 空间对象的位置算子

Tab. 1 Spatial object location operators

| 位置算子 | 表示方法 | 含义 |
|-------------------|------------------|----------------|
| Left(X, Y) | $X > Y$ | X 在 Y 的左侧 |
| Right(X, Y) | $X < Y$ | X 在 Y 的右侧 |
| Up(X, Y) | $X \cup Y$ | X 在 Y 的上面 |
| Below(X, Y) | $X \cap Y$ | X 在 Y 的下面 |
| Front(X, Y) | $X \gg Y$ | X 在 Y 的前面 |
| Back(X, Y) | $X \ll Y$ | X 在 Y 的后侧 |
| RFront(X, Y) | $X \uparrow Y$ | X 在 Y 的右前方 |
| RBack(X, Y) | $X \neg Y$ | X 在 Y 的右后方 |
| LFront(X, Y) | $X \dashv Y$ | X 在 Y 的左前方 |
| LBack(X, Y) | $X \downarrow Y$ | X 在 Y 的左后方 |
| Inside(X, Y) | $X \odot Y$ | X 在 Y 的内侧 |
| Outside(X, Y) | $X \square Y$ | X 在 Y 的外侧 |

定义结构算子 Contain 和 ComposeBY。

定义 3 Contain 算子 $\forall O_i, I$, 当且仅当 $O_i \in I \cdot O$, $\text{Contain}(I, O_i) = \text{ture}$ 。对于给定的语义对象 O_i , 只有在图像的对象集合中包含该对象时, Contain 算子返回 true 值。

定义 4 ComposeBy 算子 该算子描述了对象元素之间的结构组成关系。当对象 O_i 由 O_k, O_l, O_m 3 对象组成时, 可以表示为 $\text{ComposeBy}(O_i, \{O_k, O_l, O_m\}) = \text{true}$ 。

2.4 对象操作代数定义

定义 5 Select 操作代数(\mathcal{E}) 对于给定的图像数据库 I_{db} 和语义对象 O_i , 从 I_{db} 中获取包含 O_i , $i \in (1, n)$ 的所有图像 I_j , $j \in (1, m)$, 构成图像集合 M , 即 $M(I_j) = I_{db}^*(O_i)$, 当且仅当 $\text{Contain}(I_j, O_i) = \text{true}, I_j \in I_{db}, i \in (1, n), j \in (1, m)$ 。

定义 6 SpatialSemantic 操作代数(π) 度量语义对象的空间位置近似程度。

对于给定的语义对象 $O_i, i \in (1, n)$, 其空间关系 $s_k, k \in (1, p)$ 和最小期望相似度 MinSimilarity , 从图像数据库 I_{db} 中获取满足 $Q(O_i, s_k, \text{MinSimilarity})$ 的最相近的图像 N , 分两步定义:

(1) 从图像数据库 I_{db} 中获取包含对象 $O_i, i \in (1, n)$ 的图像集合 M , 即

$$M(I_j) = I_{db}^*(Q \cdot O_i), i \in (1, n), j \in (1, m) \quad (6)$$

(2) 从图像集合 M 中获取符合最小期望相似度的最相近图像 N 。定义 S_j^N 为图像集合 M 中对应

于 I_j 的对象空间关系, 则在 I_j 中与 $O_i, i \in (1, n)$ 相关的对象空间关系集合定义为 $S_{j,o_i}^N, i \in (1, n), j \in (1, m)$, 则 $N = M^*(Q \cdot s_k)$, 当且仅当 $\max_{j \in (1, m)} \{ \max_{k \in (1, p)} (s_k \odot S_{j,o_i}^N) \} \geq \text{MinSimilarity}$ 。

3 XML 语义描述与检索

3.1 XML 语义描述

定义的 XML 文件框架与描述属性, 主要包括颜色属性、形状属性、元语义属性、元语义映射属性和空间属性等, 表示如下:

```

< annotation >
< category > Natural Scene image </ category >
< filename > ... </ filename >
< subject > ... < subject >
< Description / >
< images >
< image >
< object >
< objectname > ... </ objectname >
< features >
< colors > ... </ colors >
< shape > ... </ shape >
< texture > ... </ texture >
</ features >
</ object >
< semantic >
< metasemantic > ... </ metasemantic >
< relation > ... </ relation >
< mapping > ... </ mapping >
</ semantic >
</ image >
</ images >
</ annotation >

< relation >
< left > < o > Tree </ o > < o > House </ o > </ left >
< right >
< o > Forest </ o > < o > House </ o >
< o > Water </ o > < o > House </ o >
</ right >
< up >
< o > Flag </ o > < o > House </ o >
< o > Sky&cloud </ o > < o > House </ o >

```

定义对象间风景图像中所包含的“tree”、“house”、“boat”等对象的空间位置关系如下:

```

</up >
<inside >
  <o > House </o > <o > Forest </o >
  <o > Person </o > <o > Tree </o >
  <o > Boat </o > <o > Water </o >
</inside >
<back > Mountain </back >
<composeby > <o > House </o > <o > Flag </o > <o >
Window </o > </composeby > <contain > <o > House </o >
<o > Flag </o > </contain >
</relation >

```

3.2 语义检索实例

选取 200 张城市场景户外图像。通过语义标注,共得到了 605 个语义对象,主要包括了汽车、行人、交通标记、树木等。将这些信息以 XML 文档形式存放到 XML 语义库。并通过关键词语义检索来获取相关图像。

图 2 和图 3 分别是通过语义检索获取的“car”、“tree”的结果。在图 2 检索中,所返回的 73 幅都包

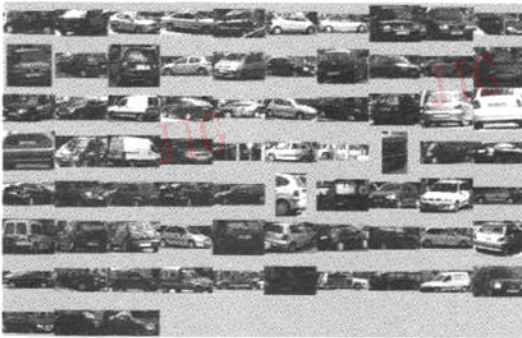


图 2 语义“car”检索结果
Fig.2 Retrieval result of car

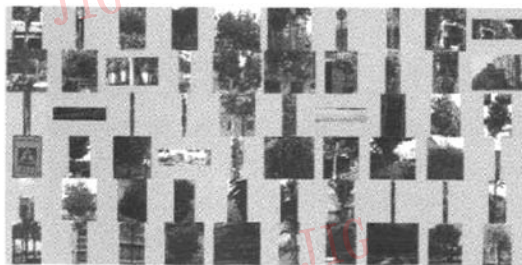


图 3 语义“tree”检索结果
Fig.3 Retrieval result of tree

含了“car”对象;在图 3 的检索中,所获取的 60 幅图像中有 55 幅与“tree”主题相关,检索结果具有较高的精度,良好的语义相关性和可解释性。

4 结 论

本文提出了 XML 驱动的图像语义检索框架,重点研究了图像内容描述相关的图像特征分层描述模型、空间位置算子定义、语义对象操作等关键技术,并定义了相应的 XML 语义描述框架。通过相关检索实验,表明了该框架具有良好的检索性能,检索结果具有较好的语义相关性。

参考文献 (References)

- 1 Zhang Yu-ji. Content-based Vision Information Retrieval [M]. Beijing: Science Publishing House, 2003. [章毓晋著. 基于内容的视觉信息检索[M]. 北京:科学出版社, 2003.]
- 2 Jaines A, Chang S F. Model-based classification of visual information for content-based retrieval[A]. In: Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Database VII. 1999 [C], San Jose, CA, USA, 1999, 3656: 402-414.
- 3 Hanjalic A, Legendijk R L, Biemond J. Video content analysis: from visual features to video semantics[A]. In: Proceedings of the First International Conference on Image and Graphics (ICIG 2000) [C], Tianjin, China, 2000: 1-9.
- 4 Li B X, Sezan I. Event detection and summarization in American football broadcast video[A]. In: Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Database [C], San Jose, CA, USA, 2002, 4676: 202-213.
- 5 Lin C Y, Yin J X, Cao X, et al. A semantic modeling: approach for medical image semantic retrieval using hybrid bayesian networks[A]. In: Proceedings of the ISDA 2006 [C], Jinan, Shandong, China, 2006: 482-487.