

# 一种基于图像高层语义信息的图像检索方法

吴楠 宋方敏

(南京大学软件新技术国家重点实验室, 南京 210093)

**摘要** 由于基于图像高层语义信息的图像检索与传统的图像特征匹配检索相比,在检索的准确性、结果相关性以及降低误检率等方面具有明显的优势,因此高效的图像检索方法应该充分利用图像蕴涵的高层语义信息。为了利用图像的高层语义信息来进行图像检索,在深入研究图像高层语义的低层特征描述的基础上,提出了图像语义的层次划分,并对每个高层语义层提出了语义抽取和检索算法。实验结果表明,该检索算法可以有效地对图像高层语义信息进行提取,并可作为新型高效图像检索系统的一个模型。

**关键词** 基于内容的图像检索 层次化语义模型 图像高层语义信息 信息抽取 图像挖掘

中图法分类号: TP391.3 文献标识码: A 文章编号: 1006-8961(2006)12-1774-07

## An Image Retrieval Method Based on High-level Image Semantic Information

WU Nan, SONG Fang-min

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093)

**Abstract** Comparing with the traditional image retrieval based on features matching, the method based on the contents of high-level semantic in image has distinguished advance in accuracy, content-relativity and anti-interference. An effective image retrieval method should make fully use of the high-level semantics contents of images. Based on the deep researching on the low-level features' description of image's high-level semantics, in this paper a new model on image's semantic levels is presented, and the extraction and retrieval algorithm for every level of image's semantic is established. The experimental results show that the algorithms can make effective extraction on image's high-level semantic information, and they can be used as a model on novel and effective image retrieval systems.

**Keywords** content based image retrieval, stratified semantic model, high-level semantic information in image, information extraction, image mining

## 1 引言

随着多媒体和互联网技术的迅速发展,图像成为一种重要的信息资源,由于其所包含的信息量远超过文字,因此图像的检索技术已经成为一个研究热点,并成为数字化图书馆、多媒体网络信息检索等重大研究项目中的关键技术<sup>[1,2]</sup>。

图像检索就是通过比较待查图像(或关键字)

与图像库中图像的相似程度来找出最符合检索要求的图像的过程。基于内容的图像检索(content based image retrieval, CBIR)因与其他图像检索方法相比具有很多方面的优势而已成为目前图像检索技术的研究主流。

然而,目前CBIR主要只利用了图像的低层视觉特征(如颜色、纹理、形状)进行检索,而对图像的描述以及对图像之间相似性的判断则一般是建立在图像所描述的对象、场景以及行为特征之上<sup>[3]</sup>,

基金项目 江苏省自然科学基金项目(BK2003066)

收稿日期 2005-01-11 改回日期 2005-10-26

第一作者简介 吴楠(1981~),男,现为南京大学硕博连读研究生。主要研究领域为软件新技术与方法学、信息检索与获取技术。

E-mail: macao\_1999@263.net

也就是说,由于检索者进行检索的主要依据是图像的高层语义(如图像所表达的对象、行为、场景或感情色彩),而不是图像的颜色、纹理、形状等细节特征,因此,CBIR 无法真正支持基于语义的图像检索,其关键的局限在于没有利用知识推理系统来进一步挖掘图像的高层语义含义。

基于高层语义信息的图像检索被认为是最具利用价值的图像语义检索方式,因为它允许用户在检索过程中可以用主观感觉来描述图像,这对于图像检索者而言,无疑有助于提高检索效率和准确度。基于高层语义的图像检索需要解决 3 个主要问题<sup>[4]</sup>:一是必须提供高层语义的描述方式;二是必须具有将图像视觉特征映射到高层语义的方法;三是语义检索系统的语义处理方法。本文将从这 3 个方面对图像高层语义的建模、表示以及抽取算法等具体问题进行分析,并提出一种在实际应用中可行的基于高层语义信息的图像检索方法。

## 2 图像层次化语义模型

根据图像中各语义要素间组合的抽象程度,可将图像的语义信息分为 6 个层次,称为图像层次化语义模型<sup>[5-7]</sup>(图 1)。

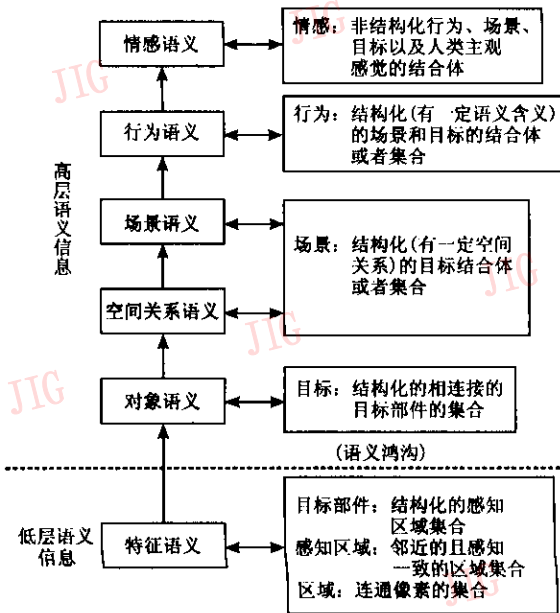


图 1 本文对层次化语义模型的划分

Fig. 1 Image's stratified semantic model

目前 CBIR 所进行的研究工作主要集中在第 1 层即特征语义层,而从语义层次中的第 2 层开始,对

象的概念就进入了语义,其虽比第 1 层次更接近人的抽象思维也更容易被人所描述和接受,然而,由于引入对象语义必须解决许多图像处理中的具体问题,如图像分割、图像中对象的识别等,因而导致语义层次的第 1、2 层之间存在着一个“语义鸿沟 (semantic gap)”<sup>[8,9]</sup>。但近年来图像分割和图像中对象识别技术的发展为逾越“语义鸿沟”提供了解决途径。

本文方法主要处理图 1 中的高层语义信息部分。

## 3 图像高层语义的提取算法与实验结果

图像高层语义提取的基础是图像的视觉特征提取、图像中对象的识别以及基于邻域知识的推理。现分别给出图像高层语义层中的各语义层次的描述方法、提取算法以及相关的实验结果。

### 3.1 对象语义的提取算法与实验结果

由于对象语义层紧邻特征语义层,且对象语义提取结果直接影响到图像更高层次语义的提取,因此对象语义的提取对基于高层语义的图像检索具有基础性的作用。

本文提出了如图 2 所示的对象语义提取算法。算法的大致流程为:

(1) 提取出原始图像中有意义的部分,这将有助于提高语义信息提取的准确度及速度,提取时对于较简单的图像,可以采取一般的色度直方图分割法,对于较复杂的图像,则可利用图像的多维密度函数,通过综合其色度、纹理和颜色等信息来选择恰当的聚类分割算法。

(2) 进行区域内对象的分割和识别,即首先利用分块图像来提取目标区域的各组成部分,同时根据区域连通性准则将包含目标信息的子图像块划分到各自所属的区域中;然后选取连通面积最大的一个连通域,并以它的外接矩形作为最终的目标区域;进而将划分区域后的对象的纹理特征和色度直方图抽取出来在模板匹配时使用,再将划定的区域使用某种轮廓描述算子描述其轮廓;最后使用图像数据库中的可变模板和已保存的对象纹理、色度等信息来对对象的轮廓进行识别。

在算法的具体实现中,选择小波轮廓描述符来对对象进行轮廓化处理及后续的匹配,具体算法参

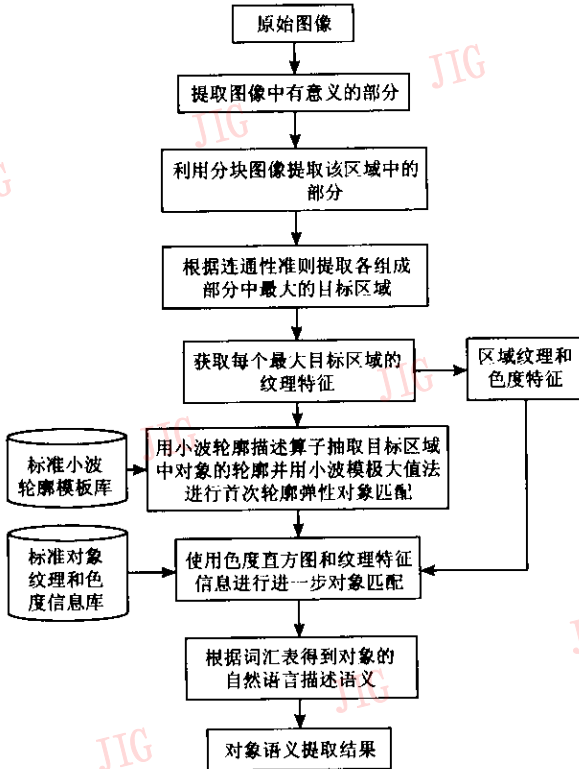


图2 图像对象语义提取流程图

Fig. 2 Flow chart of the extraction algorithm of object semantic in image

见文献 [10] [11]。

进行对象的轮廓描述后,再引入小波模极大值法<sup>[13]</sup>来提取图像中的多尺度目标边缘信息,然后采用多尺度的不变矩作为特征来度量图像中目标形状的相似性。

为了表示小波变换后的每一个多尺度边界边缘的特征,可以使用对平移、旋转和尺度变换不变的矩来进行提取。基于小波模极大值和不变矩的算法参见文献 [6]、[10]。

具体实现时,首先经过上述步骤将对象的轮廓特征限制在一个很小的候选集内,然后将提取出的轮廓内的色度直方图和纹理特征与图像库中的相关对象的色度直方图和纹理特征进行匹配,再配合标准对象库中的词汇表,就可以比较准确地给出图像中各主要对象的匹配结果,并可以根据近似度阈值输出一项或者几项。图 3 是用上述算法对一幅具体图像中各个对象进行提取的过程示意图。

用该算法对一个包含 360 幅自然物体(包括花草、衣物、建筑物、山川、河流等)的图像库进行了对象语义提取实验(全部实验图像选取

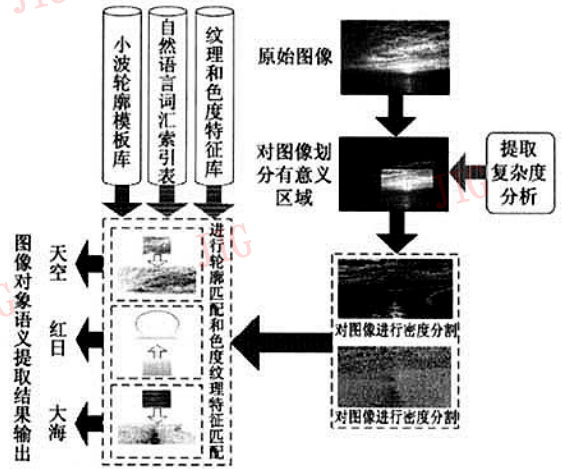


图3 对“海上日出”图像进行对象语义提取

Fig. 3 An example of the extraction of object semantic in a picture

自 CorelDRAW 标准图像库,均为 640 × 480 像素的非压缩 256 色位图文件;实验程序由 Microsoft Visual C++ 6.0 编写,实验用 PC 配置: Intel Pentium 4, 3.06 GHz 超线程处理器, 1GB RAM, 操作系统为 Microsoft Windows Server 2003)。图像中的对象与标准图像库中的模板相比,分别有不同程度的平移、变形、旋转、拉伸、收缩、扭曲、镜像和轮廓小范围突变等变化(图 4)。实验结果如图 5 所示,由图 5 易见,该算法具有较高的效率。

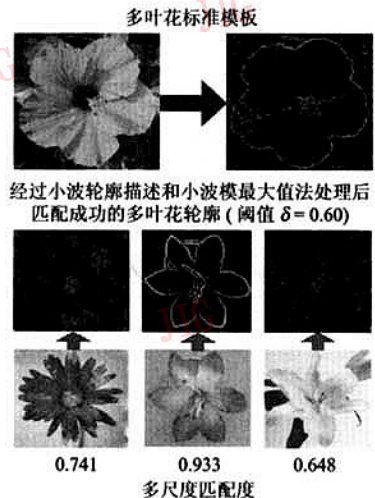
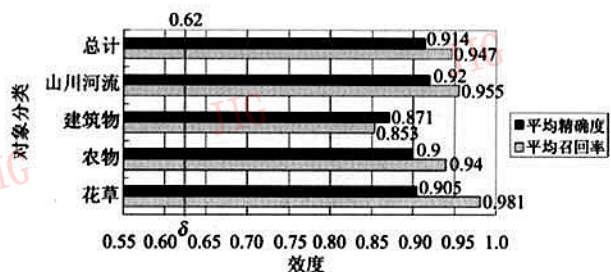


图4 用小波轮廓描述符与小波模极大值法处理后匹配成功的例子

Fig. 4 Some successful matched examples by using the algorithms



5 图像对象语义提取实验结果(匹配度阈值  $\delta = 0.62$ , 中精确度即查准率,召回率即查全率标度 1 即为 100%)

Fig. 5 Experimental results of the extraction of object semantic in image

### 3.2 空间关系语义的提取算法

众所周知,空间关系主要是反映图像中主要对象之间的位置、分布、组合等信息的关系。本文中用于语义提取的空间关系主要涉及图像中各主要对象的空间实体所在的位置以及空间实体在场景中的投影。

为了描述各个对象之间的空间位置关系,本文算法使用 9-相交模型来对所有识别出的对象两两地进行处理<sup>[13]</sup>。该模型可以表达 2 个目标之间的不接触、包含、内部、相等、接触、覆盖、被覆盖、重叠等 8 种位置关系。

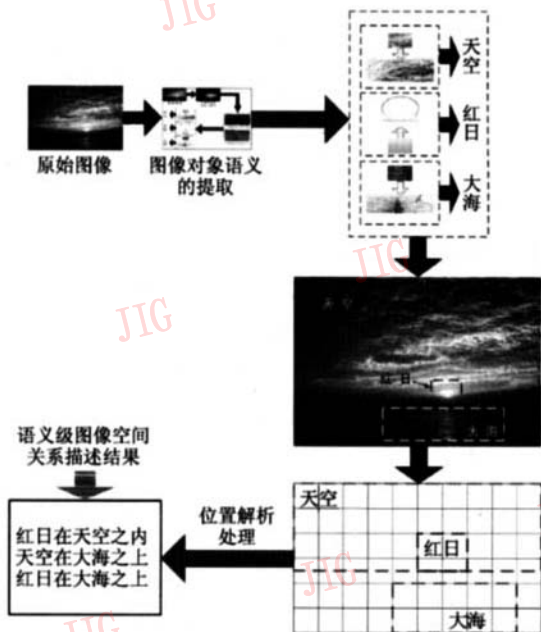
为了对各个对象的区域进行定位,使用最小边缘矩形(minimal bounding rectangle, MBR)法对上述区域作近似估算,其处理得到的一系列矩形,而每个矩形则对应对象轮廓的一条闭合曲线。对象之间的方位关系可通过建立方向关系矩阵来确定。这样,保存最小边缘矩形与各个对象中心的坐标就为空间关系相似度的计算提供了方便。

这里主要考虑对空间关系语义贡献最大的对象空间朝向的提取。空间朝向可以有多种定义方法,但实际情况中,一个目标相对于另一个目标的朝向更加有用。由于参照目标的坐标可以被确定为最小边缘矩形的中心点,所以重点考虑一个目标相对于多个点(参照目标 MBR 的中心点)的朝向。

本文算法表达一个目标相对于多个点的简单朝向的方法是 2D 符号数组<sup>[4]</sup>,它可以表达由 1 个点(如目标的重心)或者由 4 个点(如目标 MBR 的 4 个顶点)代表的目标间的相对朝向。

实践表明,上述方法对于目标结构比较简单的图像可以准确地进行空间关系语义的提取;当图像中目标的结构比较复杂,且含有大量的子集时,为了准确地判定子集之间的关系(特

别是重合的次序),本文算法用 2D-B 串进行对象间复杂空间关系的描述,由于 2D-B 串引入坐标系内对象的相对位置,因此使用 2D-B 串不需要进行子集分割(目标内部切割)就可以获得复杂的空间关系。具体方法参见文献[14]。图像空间关系语义提取流程的例子如图 6 所示。



6 对“海上日出”图像进行空间关系语义提取流程图  
Fig. 6 Flow chart of the extraction algorithm of spatial relationship semantic in image

完成上述步骤后即可直观地得到在 8 个方向(上、下、左、右、左上、左下、右上、右下)上对象所处空间位置的形式描述以及对象之间相邻和交叠的关系。

上述空间关系语义的提取是场景语义提取的重要基础。

### 3.3 场景语义的提取算法与实验结果

空间关系语义层生成的形式描述虽然准确,却不太符合人的习惯,例如“海上日出”一般不描述为“太阳在海水上方”。心理学和认知的研究认为,人们一般将一组特定对象的特定空间关系抽象为一种场景(scene)<sup>[3]</sup>。因此用图像的场景语义进行图像的描述和检索将更接近于人的认识也更加符合检索习惯。

为了区分易混淆场景中的对象组合,应首先对待提取场景语义的图像进行场景分类。例如,一般室外场景可以分为“自然景观(如海滨、山川等)”和“人文场景(如楼群、比赛等)”

两种。通常快速场景识别可以通过对图像的低频分量信息(全局图像特征)的识别来进行<sup>[10]</sup>,并可在对象识别前,通过全局低层特征进行粗略的场景分类。本文算法使用组合区域模板(composite region template)来进行简单场景分类,并取得了较好的效果。分类时,首先将图像的低频空间划分为固定大小的子块,然后对这些子块分别确定其各自的简单场景语义类别,最后根据子块间的组合关系来确定整幅图像的简单场景语义类别。组合区域模板算法详见文献[4]。

要将已提取的对象语义和空间关系语义的形式描述转化为特定的场景描述,最重要的问题就是如何在图像场景数据库中查找出符合形式化空间关系语义的关键词,这一步骤需依赖一种场景查询语言和一个描述场景的词汇表。

下面提出一种场景-对象查询词汇表及其查询算法,该算法显著地提高了图像场景语义提取的效率和效果。

词汇表的结构如图 7 所示,该图显示,在包含各种场景和对象的层次化词汇表的各个关键词之间加入空间位置的关联,词汇表中的元素由下向上是一种抽象的关系。检索是在词汇表中自底向上地进行,并约定如果到达的场景层次不足以包含所有必需的对象及其位置关系,则停止检索,并输出当前最高层次的场景语义。例如,对于一幅大草原上的日出图像,在对象语义层提取出天空、红日、草原 3 个对象;在空间关系语义层提取出红日在天空之内、红日在草原之上及天空在草原之上 3 个空间关系描述公式。而场景语义层则根据这些提取结果,首先根据“红日在天空之内”得到“日出”的场景(图

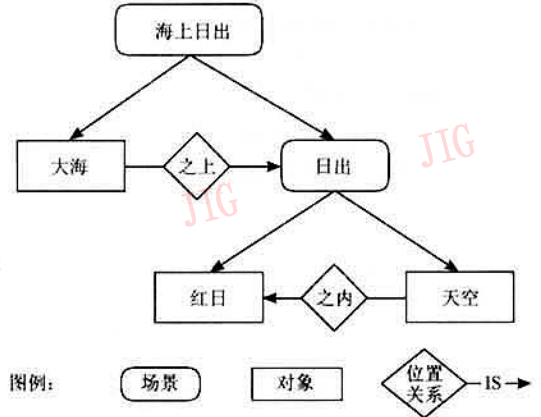


图 7 拓扑词汇表场景词汇数据库

Fig. 7 Database of the topological scenes' words

7),然后根据区域的划分,将“红日在草原之上”和“天空在草原之上”归纳为“(红日’与’天空)在草原之上”,形成“日出在草原之上”的结果,最后再查询数据库,若发现在“日出”的同层次中没有关于对象“草原”的组合,则检索停止,而直接输出当前最高层场景语义“日出”。这种提取方法比较符合人类进行场景判断的习惯,如果需要更加精确的场景语义描述,则可以调整组词方式以输出“草原之上的日出”或者“草原日出”等结果。

利用 VC++ 6.0 结合 Matlab 6.5 的图像处理工具箱实现了上述图像场景语义提取模型,并进行了实验,部分实验结果如图 8 所示。另外,还在一个包含 40 个自然场景和人文场景的场景词汇表中随机选出 5 个特定场景,并为每个场景找了 4~16 幅不同的图像,在对图像对象进行校准后,对每个独立的图像进行了场景语义提取实验,实验结果如图 9 所示。



图 8 图像场景语义提取实验的部分结果及结果评价

g. 8 Some experimental results and their judgments of the extraction of scene semantic in image

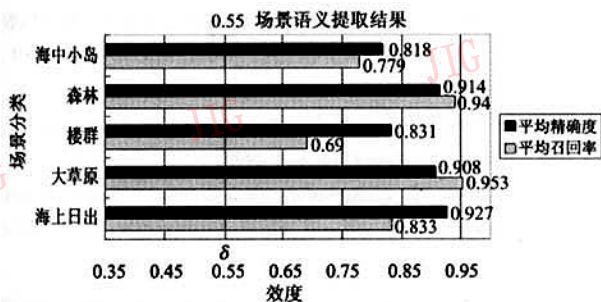


图 9 图像场景语义提取实验结果( 场景匹配度阈值  $\delta=0.55$  )

Fig. 9 Experimental results of the extraction of scene semantic in image

### 3.4 行为语义的提取模型

按照实验心理学的定义,“人类行为(human behavior)”是指人类在一定场景下的特定动作或者状态<sup>[15]</sup>,因此行为语义的提取必须在场景语义提取的基础上完成。由于行为语义涉及到的人物、场景和行为状态是特定的,所以行为语义的一方面提取必须依赖一个较完备的知识库,并且需要知识库的支持系统具有一定的推理能力;另一方面,由于行为的发生主体均认为是人,因此应当在对象语义提取阶段特别注意对图像中关键人物的提取,如有必要可对人的动作进行适当的分类,但人的动作分类算法目前尚不够成熟。下面给出一个基于知识库和推理系统的提取图像行为语义的例子(如图 10 所示)。

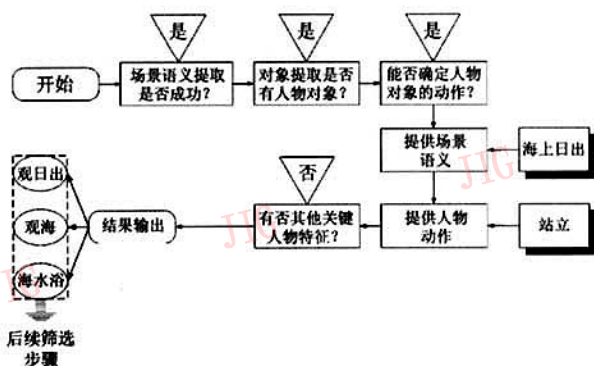


图 10 图像的行为语义提取示例

Fig. 10 An example for the extraction of behavior semantic in image

### 3.5 情感语义提取的分析与探讨

图像的情感语义层处在图像语义层次的顶端,由于该层语义提取的复杂度很大,因此要综合许多领域学科的知识,并且需要构造非常庞大的知识库

和复杂的推理系统。从实验的角度来讲,构造一个具有实际意义的情感语义提取平台的难度是很大的。本文仅对图像情感语义提取中的关键问题进行讨论。

人们普遍认识到,色彩具有唤醒情感的作用<sup>[16]</sup>。由于色彩的象征性可以引发人的联想,使色彩与一定的情绪、情感相联系,因此可以说色彩具有“情绪性”。虽然因为文化背景、个人经历和心理因素的不同而使颜色所引起的情感会有些差异,但是在人的生理构造和色彩的物理特性等因素下,色彩心理还是有很多共性的感受。

从色调上看,一般把色彩分为暖色和冷色,如红、黄、橙色以及倾向于这些色的颜色,给人以温暖、热烈、喜悦等感觉,可称之为暖色;而蓝、绿、白色以及倾向于这些色的色调给人以幽静、清新、淡雅以及消沉、冷峻的一面,故称之为冷色。

除了色调之外,图像色彩中不同的饱和度、色彩的区域和聚集度等参数都可以代表不同的情感。为了研究人类对图像产生的主观情感,必须将人类对颜色的感知和心理与图像特征的表示结合起来,以便寻找与人类感知和心理相关的颜色特征表示方法。

## 4 结 论

基于图像高层语义信息的图像检索是计算机信息处理和检索学科中一个全新的领域。而充分利用图像的高层语义信息则可以最大限度地弥补现有 CBIR 系统的不足。

本文提出了一种基于图像高层语义信息的图像检索方法,其主要包括图像语义的层次化模型、图像的对象语义、空间关系语义、场景语义、行为语义和情感语义的描述模型和检索算法。实验结果表明,用本文方法进行检索可以得到比较满意的提取和检索结果,且检索的准确率和召回率都较传统方法的 45%~60%<sup>[6]</sup>有较大提高。

鉴于现在互联网上的图像信息不断增多,使得网页中图像的检索和网页文档的检索一样,成为一个重要的研究和应用领域。本文提出的方法可以在大部分高层语义层次对图像的高层语义进行较为准确地提取,但如何组合这些相对独立的模块和所提取的相对分散的高层语义信息,并加以合理的归纳和重组,以形成一个对图像内容的较为主观的认识

和理解,这是一个涉及到多个学科和领域的难题。当前人工智能的发展还无法由机器独立完成语义图像检索这一智能的任务,更可行的方法是把人作为系统的一部分,利用人机的交互行为来改进系统的能力<sup>[17]</sup>。一个成功的图像高层语义信息检索系统需要多个研究领域紧密结合,因此计算机视觉、数字图像处理、模式识别与分类技术,包括心理学、认知科学、生物视觉模型等技术的新发展和综合应用,必将推动图像高层语义信息检索的突破性进展。

### 参考文献 (References)

- 1 Eakins J P. Automatic image content retrieval-Are we getting anywhere?[ A ]. In : Proceedings of Third International Conference on Electronic Library and Visual Information Research[ C ], Milton Keynes , England , 1996 : 123 ~ 135.
- 2 Hermes T , Klauck C , KreyB J , et al. Image retrieval for information systems[ A ]. In : Proceedings of IS&T/SPIE ' s Symposium on Electronic Image : Science & Technology[ C ], San Jose , CA , USA , 1995 , 2420 : 394 ~ 405.
- 3 Gregory R L. Visual Psychology[ M ]. Beijing : Beijing Normal University Press , 1986. [ Gregory R L 著. 视觉心理学[ M ]. 彭聘龄 , 杨旻译. 北京 : 北京师范大学出版社 , 1986. ]
- 4 Wang Hui-feng , Sun Zheng-xing , Wang Jian. Semantic image retrieval : review and research[ J ]. Journal of Computer Research and Development , 2002 , 39( 5 ) : 513 ~ 523. [ 王惠锋 , 孙正兴 , 王箭. 语义图像检索研究进展[ J ]. 计算机研究与发展 , 2002 , 39( 5 ) : 513 ~ 523. ]
- 5 Colombo C , DelBimbo A , Pala P. Semantics in visual information retrieval[ J ]. IEEE Multimedia , 1999 , 6( 3 ) : 38 ~ 53.
- 6 Zhang Yu-jin. Visual Information Retrieval Based on Content[ M ]. Beijing : Science Press , 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. ]
- 7 Wang Hui-feng , Sun Zheng-xing. The methods of semantics processing in content-based image retrieval[ J ]. Journal of Image and Graphics , 2001 , 6( 10 ) : 945 ~ 952. [ 王惠锋 , 孙正兴. 基于内容的图象检索中的语义处理方法[ J ]. 中国图象图形学报 , 2001 , 6( 10 ) : 945 ~ 952. ]
- 8 Bimbo A. Visual Information Retrieval[ M ]. San Francisco , CA , USA : Morgan Kaufmann Publishers , 1999.
- 9 Zhao R , Grosky W I. Narrowing the semantic gap-improved text-based Web document retrieval using visual features[ J ]. IEEE Transactions on Multimedia , 2002 , 4( 2 ) : 189 ~ 200.
- 10 Yang Xiang-yin , Zhang Yu-jin. Wavelet boundary descriptors and its application in image querying[ J ]. Chinese Journal of Computers , 1999 , 22( 7 ) : 752 ~ 757. [ 杨翔英 , 章毓晋. 小波轮廓描述符及在图象查询中的应用[ J ]. 计算机学报 , 1999 , 22( 7 ) : 752 ~ 757. ]
- 11 Chuang G , Kuo C. Wavelet descriptor of planner curves : theory and applications[ J ]. IEEE Transactions on Image Processing , 1996 , 5( 1 ) : 56 ~ 70.
- 12 Egenhofer M J , Franzosa R. Point-set topological spatial relations [ J ]. International Journal of Geographical Information Systems , 1991 , 5( 2 ) : 47 ~ 57.
- 13 Egenhofer M J , Franzosa R. On the equivalence of topological relations[ J ]. International Journal of Geographical Information Systems , 1992 , 9( 2 ) : 133 ~ 152.
- 14 Lee S Y , Yang M C , Chen J W. Signature file as a spatial filter for iconic image database[ J ]. Journal of Visual Languages and Computing , 1992 , 3( 3 ) : 373 ~ 397.
- 15 Hao Bao-yuan , Zhang Hou-can , Chen Shu-yong. Experimental Psychology[ M ]. Beijing : Beijing University Press , 1983. [ 郝葆源 , 张厚燊 , 陈舒永著. 实验心理学[ M ]. 北京 : 北京大学出版社 , 1983. ]
- 16 Gu Da-zhi , Fu Shi-shen , Yang Ren-ming. Theory of Color and Graphical Visua[ M ]. Beijing : Science Press , 2000. [ 古大治 , 傅师申 , 杨仁鸣著. 色彩与图形视觉原理[ M ]. 北京 : 科学出版社 , 2000. ]
- 17 Rui Y , Thomas S H , Chang S F. Image retrieval : past , present , and future[ A ]. In : Symposium on Multimedia Information Processing[ C ], Taipei , Taiwan , China , 1997.