

基于目标语义特征的图像检索系统¹

高永英 章毓晋 罗云

(清华大学电子工程系 北京 100084)

摘 要 为克服当前基于内容的图像检索技术中低级特征无法准确全面地描述高级语义的问题, 该文设计和实现了一个基于目标高级语义特征的检索系统. 该系统利用了一个多级图像描述模型将语义特征结合到图像检索技术中. 该图像描述模型通过在不同层次上对图像内容进行分析和描述, 实现了从低级特征到高级语义的过渡. 在此模型的基础上还研究了相应的检索机制和反馈技术. 该系统的检索机制定位于图像中目标的语义内容, 与传统的图像检索系统相比更接近人对图像内容的理解, 从而使检索过程更简便, 检索效率也得到很大提高. 基于目标描述的自适应相关反馈可针对不同用户的不同需求给出相应的检索方案, 从而使检索结果得到优化.

关键词 基于内容检索, 图像语义, 图像描述模型, 目标理解, 相关反馈

中图分类号 TN911.73, TP391.4

1 引 言

基于内容的图像检索是当前图像检索技术的主流, 许多图像检索系统, 如 QBIC, Photobook, VisualSEEK 和 CAFIIR 等, 均采用了基于内容的检索技术. 然而, 现有图像检索系统对图像内容的描述大多采用了传统的低级图像特征, 如颜色, 纹理, 形状等. 由于这些特征与人对图像内容的描述和理解之间存在相当大的差异, 因此很多情况下仅以图像低级特征为检索依据得到的结果不尽人意. 如何将图像语义特征结合到检索中是提高检索系统性能的关键所在. 现有的结合图像语义特征的检索技术大多着眼于检索过程中的某一环节, 或者通过相关反馈来增强人机交互, 获得用户对图像理解的信息^[1], 或者通过分类器对图像中某些区域进行语义标注^[2,3].

针对以上问题, 本文实现了一个基于目标语义特征的图像检索系统. 该系统以多级图像描述模型为核心建立了相应的检索和反馈机制, 并根据该图像描述模型在不同层次上对图像内容进行分析和提取, 实现了渐进式图像语义理解. 该系统采用了基于图像语义的查询条件和检索方式, 使检索过程更方便用户. 本文还给出了一种自适应的反馈机制, 针对不同用户的不同需求给出相应的检索方案, 充分利用了基于语义特征的表达和描述特点. 与传统的图像检索系统相比, 该系统的工作方式更接近人对图像内容的理解, 对自然风景图像的检索实验表明有较好的效果. 下面先简单介绍描述模型和语义提取, 重点将放在系统的结构、机理和效果上.

2 多级图像内容描述模型和语义特征提取

有效地描述图像内容是图像检索技术的关键所在. 传统的图像描述模型仅建立在低级特征的基础之上, 一般以统计数据矢量的形式出现. 实际上, 这些统计数据与人对图像的内容理解存在很大差异. 首先, 人对图像的理解是建立在人类已有知识的基础之上的, 而这些低级特征无法反映这些经验知识. 其次, 图像内容具有“模糊”的特性, 不能仅用简单的特征矢量来描述. 针对以上两点, 我们提出了一种多级的图像内容描述模型, 在不同层次上提取图像的内容信息, 实现一种渐进式的图像内容描述. 该模型的结构可以用图 1 说明.

从图 1 可以看出, 该模型包含 5 层: 原始图像层, 有效区域层, 视觉感知层, 目标表达层以及场景理解层. 虚线以右给出了每一层相应的公式描述. 整个多级模型 (MLM) 可以表示为

$$\text{MLM} = \{\text{OIL}, \text{MRL}, \text{VPL}, \text{ODL}, \text{SUL}\} \quad (1)$$

¹ 2002-02-18 收到, 2003-03-10 改回

国家自然科学基金 (69672029, 60172025) 资助

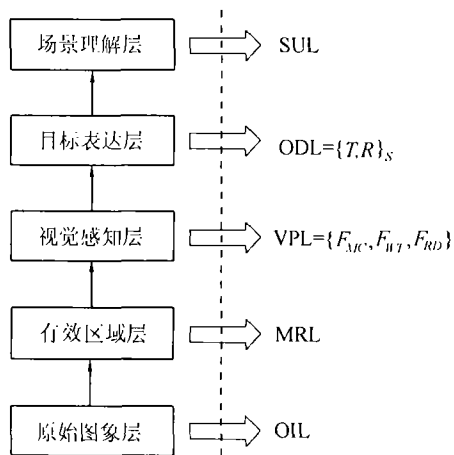


图 1 多级图像描述模型

图 1 还给出了多级模型各层之间的关系。在相邻两层之间，上一层的描述直接来源于下一层。目前在我们的图像检索系统中应用的主要是前 4 层，即从原始图像层直到目标描述层。至于场景理解层，它主要考虑的是一幅图像作为一个整体所体现的语义，目前对于场景理解层的描述仍在研究之中。下面将对有效区域层，视觉感知层和目标描述层的获取和描述给予简单介绍。

2.1 预分割

与通常意义下的分割不同的是，这里无须对图像进行像素级的精细分割，仅提取相对于人类视觉有一定意义的区域^[4]。预分割的结果给出原始图像的标号图像 MRL，作为有效区域层的表述。

2.2 特征提取

在获得了有效区域层的描述后，需要对图像中每一有效区域进行特征提取以获得视觉感知层的描述。在具体应用中，低级特征的选择受到实际需要的影响。考虑到实验所用花卉和风景图像颜色和纹理信息均比较丰富的特点，因此选择了混合颜色特征^[5]和小波包纹理特征^[6]来描述视觉感知层。此外，区域描述也作为视觉感知层描述的一部分，主要是考虑到后续目标描述层的需要。

2.3 目标识别和关系表述

在图 1 给出的多级图像描述模型中，目标描述层是至关重要的一个部分，它是从图像低级特征到高层语义的过渡。具体来说，目标描述层利用视觉感知层提供的信息，通过目标识别和关系表述形成本层的描述。这里如何利用人类的知识并“识别”出图像中每一个有效区域“是什么”是获得目标层描述的关键环节。自动目标识别能填补图像低级特征对图像内容描述的不足，与人工识别相比又可克服主观性影响，提高效率。这里我们采用了迭代方法对训练样本集和测试样本进行逐次比较，最终完成目标识别。整个过程包括动态权重设置，多途径识别以及多条件判决 3 个关键环节^[7]。

关系表述是目标层的另一个组成部分，它可以帮助对同一幅图像给予不同的描述，体现了图像内容的“模糊”特点。关系表述有很多方法，最常用的是空间位置关系表述。给定目标 A 和目标 B 的区域重心坐标 (由视觉感知层的区域描述子提供) $P_A(x_A, y_A)$ 和 $P_B(x_B, y_B)$ ，它们之间的空间关系表述 (Spatial Relationship Representation, SRR) 可以由下式得到：

$$\text{SRR}(A, B) = \text{tg}^{-1}[(y_B - y_A)/(x_B - x_A)] \quad (2)$$

2.4 图像描述单元

综上所述，对图像库中每幅图像均应提供从原始图像层到目标描述层的描述。与通常的特征矢量描述相比，由于不同图像包含的目标个数不同，因此每幅图像内容描述信息的长度是不同的。这里我们提出一种名为“图像描述单元”的结构实现非定长图像描述信息的存储，如图 2 所示。

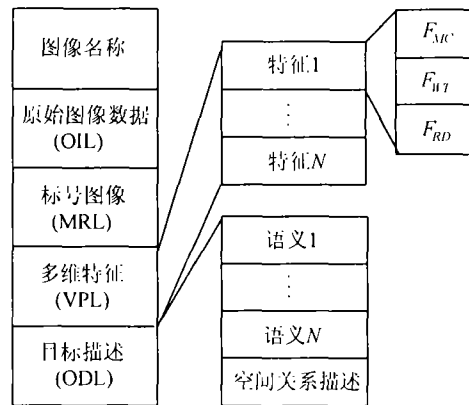


图 2 图像描述单元结构

3 系统框架和模块

我们设计的图像检索系统框图如图 3 所示, 可以分为 5 个模块: 图像库, 检索, 反馈, 用户输入以及系统输出模块。每块完成相对独立的功能, 同时与其他块紧密结合形成一个有机的整体。

在图 3 中, 用户输入与系统输出的功能由用户操作接口来完成, 实现用户查询命令的输入以及系统检索结果的输出。虚线框中的 3 个模块——图像库, 检索和反馈模块是整个系统的主要部分, 他们均建立在同一个基于语义的图像描述模型基础之上。其中, 图像库中存储的是每一幅图像的描述模型的实例。换句话说, 对每一幅图像均根据图像描述模型给出具体的数值化图像内容信息, 这些信息即为检索和反馈的基础。下面对其特点进行介绍。

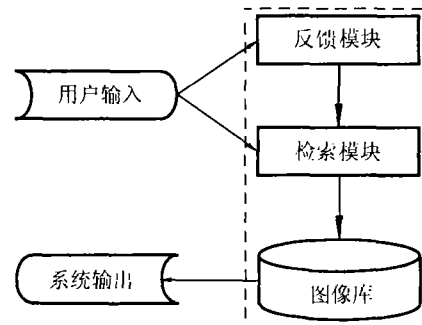


图 3 检索系统框图

3.1 无反馈查询——目标查询和关系查询

在该系统中, 用户可以根据系统提供的信息来选择所希望的图像内容直接进行查询。这种无反馈的信息查询可以分为两类: 一类可称为“目标查询”, 另一类称为“关系查询”。

所谓目标查询, 即用户根据系统提供的目标类别关键字选择所希望的图像内容。在此种情况下, 系统首先在图像库中搜索包含用户所选目标的图像作为候选图像。然后根据目标的归一化面积 (由视觉感知层的区域描述符提供) 来确定系统返回图像的排序。例如用户选择了图像库中第 k 种目标, 则候选图像的排序可根据这些候选图像中隶属于第 k 种目标的区域的归一化面积 A_k 来确定, 这里假设用户希望选择的目标在整幅图像中所占的比例越大越好。对于用户选择了多个目标的情况, 排序要考虑候选图像中所有隶属于用户所选目标范畴的区域的归一化面积的总和。设用户选择了 k_1, k_2, \dots, k_L 共 L 个目标, 则候选图像的相似度可由下式得到:

$$S = \sum_{j=1}^L A_{k_j} \quad (3)$$

关系查询是建立在目标查询的基础之上的。这里主要指目标间的空间位置关系，细节可见文献 [8]。

3.2 有反馈查询——自适应相关反馈

在图像检索中，反馈是实现人机交互的重要手段。相关反馈的引入可以提供给系统更多的信息，对于检索过程具有重要的指导作用^[1,9]。借助前面介绍的多级图像描述模型，我们提出了一种基于目标层的自适应相关反馈方法，在不同的图像描述层次上通过分析用户所选择的相关图像的相似性，给出不同的查询方案。图 4 给出了自适应相关反馈的处理流程。

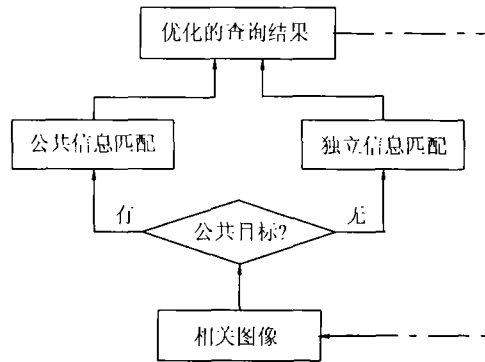


图 4 自适应相关反馈处理

从图 4 可以看出，系统首先对用户选择的相关图像进行目标描述层的相似性分析，然后根据这些图像是否含有相同的目标来选择不同的查询方案。在图 4 中，公共信息匹配用于处理相关图像存在相同目标的情况，而独立信息匹配则用于不存在相同目标的情况。

在用户反馈给系统的相关图像中存在相同目标的情况下，系统首先在图像库里搜索包含相关图像中所有相同目标的图像作为候选图像，然后通过视觉感知层的特征距离度量来决定每一幅候选图像的排序。在计算候选图像的排序时，首先计算其与每一幅相关图像的特征距离，然后取其均值的倒数作为该候选图像的排序值。这里我们将所有的相关图像作为一个整体，反馈的目的不是针对某一幅相关图像，而是提取相关图像之间的共性，因此上述处理方法是合理的。

在用户反馈的相关图像中不存在公共目标的情况下，需要对图像库中每一幅图像计算其与相关图像集中每一幅图像的特征距离，取其最大值作为该图像的排序值。这里有两点需要强调。首先，特征距离的度量是针对目标的，也就是说，对于候选图像和相关图像，对具有相同目标范畴的区域分别计算其特征距离，然后相加。其次，虽然相关图像集中无公共目标，但可以统计每一目标的出现频率，对于出现频率高的目标，在计算特征距离时加大其权值。

从上述讨论可以看出，自适应相关反馈的处理是以目标为基础的，或者说在不同层次的相似性分析中，目标层具有最高优先级。如果相关图像集包含公共目标，则系统选择公共信息匹配，在相关图像集不存在公共目标的情况下系统才会选择独立信息匹配。

3.3 有记忆反馈和无记忆反馈

在图 4 给出的自适应相关反馈处理流程中，当前优化的查询结果可以与相关图像结合进行反馈（见图中点划线），实际应用中这个过程可能需要使用若干次以逐步得到满意的结果。在反馈迭代进行中用户可有两种选择：有时用户多次反馈而其查询要求基本不变；有时用户多次反馈每次查询要求都发生变化。为了使系统能够处理这两种不同情况下的反馈迭代，这里提供了有记忆和无记忆两种反馈迭代的方式。

对于无记忆反馈，每一次迭代都是一个独立的过程，仅需考虑本次用户所选择的相关图像，在此之前迭代过程中的相关图像均不予以考虑。而在有记忆反馈的情况下，相关图像集不但包括本次迭代用户选择的相关图像，还包括在此之前迭代过程中用户选择的相关图像。然而，不同迭代过程中的相关图像在当前反馈中的作用是不同的。神经生物学告诉我们，记忆是一个随时间变化的神经活动^[10]。随着时间的推移，存储在大脑中的信息会逐渐消失。所以这里我们设计权重按一条模拟记忆曲线的折线来变化，如图 5 所示，其中横轴表示迭代的时间，纵轴表示权重，权重到一定间隔后随时间递减。

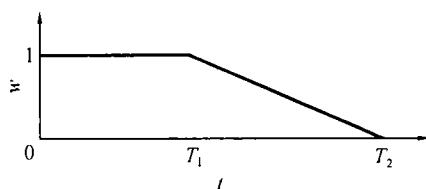


图 5 模拟记忆曲线

3.4 用户界面

用户界面支持用户输入和系统输出的功能。系统在启动后的用户界面如图 6 所示, 可以分为上下两部分。上半部分用于显示检索结果, 但是在启动时系统将先提供一些样本图像, 用户可以直接通过反馈来进行查询(见下节)。下半部分是操作面板, 用户可以通过简单的选择来提

交查询请求。其中用户界面的左下部是系统提供的关键字菜单, 以支持目标查询。在用户界面的右下部给出了 4 个功能按钮: Search, Advanced Search, Feedback With Memory 以及 Feedback Without Memory, 分别对应目标查询, 关系查询, 有记忆查询及无记忆查询 4 种情况。具体功能见下节。

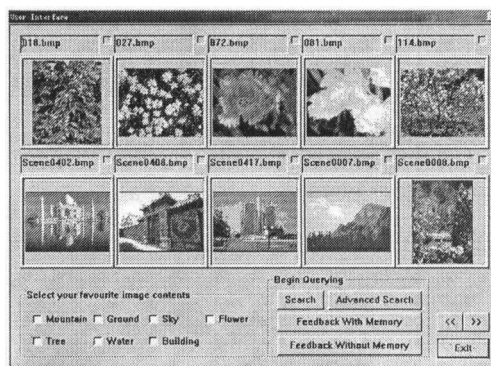


图 6 图像检索系统的用户界面

4 实验和结果

由以上介绍可见, 该系统的设计宗旨是要方便用户, 简化对用户的要求, 首先, 用户无需提供样本图像; 其次, 系统提供了多种查询方法, 并且多种查询方法可以结合使用, 为用户带来极大的方便和灵活性。下面给出几个查询示例。

4.1 目标查询

图 7 给出了一个目标查询的实例。这里用户选择了“Mountain”和“Tree”两个目标, 检索得到的前 10 幅图像显示在图中。从图 7 可以看出, 这 10 幅图像从低级图像特征的角度来看存在较大的差异, 不论用颜色或纹理等特征都很难得到这样的检索结果。但从图像语义的角度来看这些图像都包含了相同的目标范畴, 因此出现在同一检索输出结果中。

4.2 关系查询

在图 7 给出的结果上可借助关系查询进一步缩小检索图像的范围, 获得更有意义的结果。图 8(a) 给出了关系查询的选择界面。系统提供了多种空间关系, 用户可以自由选择^[8]。这里

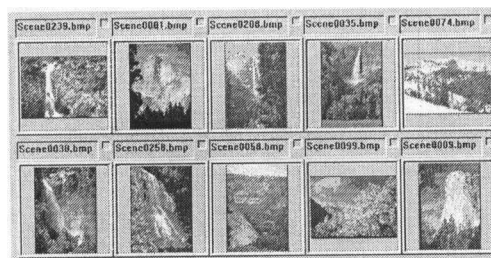


图 7 目标查询结果示例

我们选择了“左 (Mountain) \Leftrightarrow 右 (Tree)”的关系, 查询结果如图 8(b) 所示。

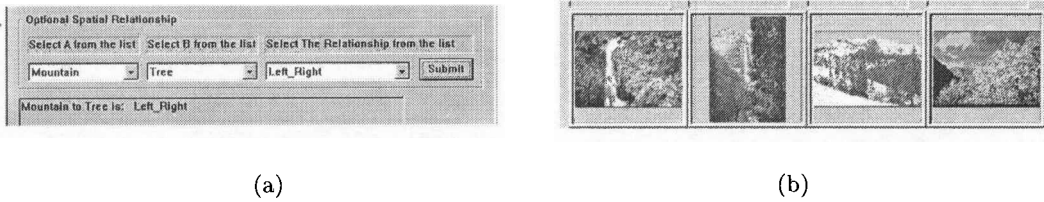


图 8 关系查询示例

4.3 无记忆相关反馈

由图 6 给出的系统启动后所提供的样本图像出发, 用户可直接点击所感兴趣的图像来进行查询。如果用户选择的相关图像是第 1 行的第 4 幅和第 2 行的第 4 幅, 此时可以看出他们无公共目标, 因此系统通过独立信息匹配给出查询结果, 如图 9(a) 所示。需要说明的一点是, 由于这是第一次反馈, 因此选择有记忆或无记忆方式没有差别。从图 9(a) 可以看出, 系统返回的图像也不具有公共目标, 但是它们至少与相关图像集中某一个存在相似性。现在图 9(a) 的基础上, 考虑用户兴趣转移的情况。比如用户的兴趣转移到了“Building”这一目标并选择了第 2 行的第 1 和第 5 幅作为相关图像。这时用户应选择无记忆反馈, 则系统将仅仅对当前的两幅相关图像进行相似性分析, 采用公共信息匹配来优化查询, 结果如图 9(b) 所示。

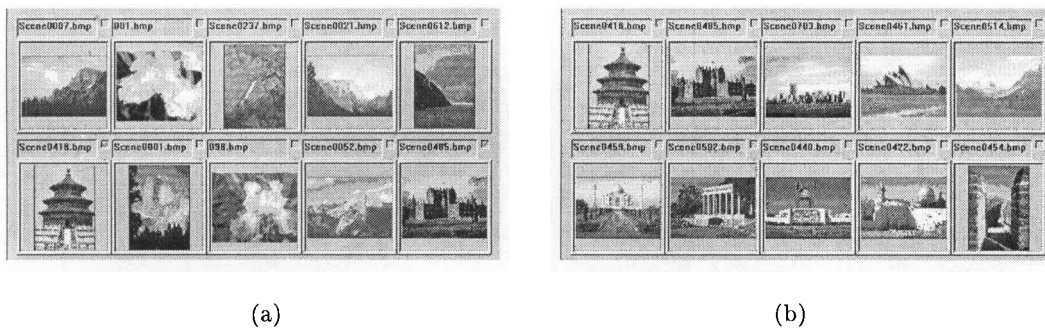


图 9 无记忆相关反馈示例

4.4 有记忆相关反馈

图 10(a) 给出由图 6 出发, 用户选择“Flower”得到的目标查询结果。现在设用户选择了第 1 行的第 5 幅以及第 2 行的第 3 幅和第 5 幅作为相关图像。可以看出, 这 3 幅图像中的花都具有较大的花瓣。系统根据这次反馈得到的检索结果如图 10(b) 所示。为了进一步优化查询结果, 用户又选择了图 10(b) 中第 1 行的第 5 幅和第 2 行的第 5 幅作为相关图像, 并利用有记忆反馈来优化查询。从图 10(c) 的结果可以看出, 前次反馈中所选择的相关图像在本次反馈返回的结果中仍处于靠前的位置, 而新增加进来的图像基本上也具有较大的花瓣。

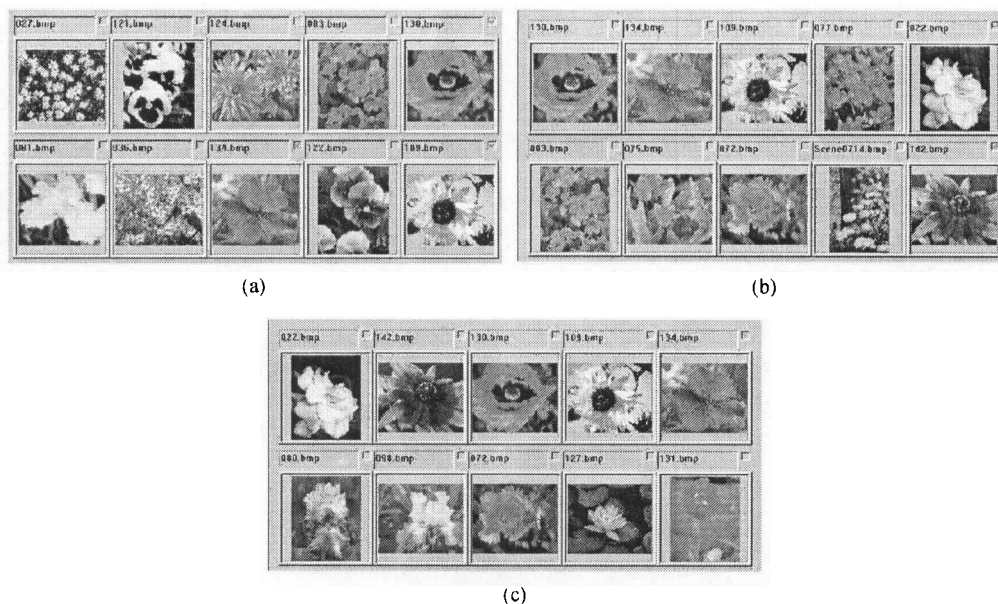


图 10 有记忆相关反馈示例

5 结 论

采用低级特征来描述图像内容的特征级检索在检索效率和交互性方面均有待提高, 如何将图像语义结合到检索中去是解决这一问题的关键。本文实现了一个基于目标语义特征的图像检索系统。该系统以多级图像描述模型为核心, 在不同层次上对图像内容依次进行分析和提取, 实现了图像低级特征与高级语义之间的过渡。在此模型的基础上我们构造了相应的检索机制。该系统的检索机制将检索定位于图像的语义内容, 利用目标层的描述信息将图像语义结合到检索过程中。该系统同时提供了一种自适应的相关反馈功能, 针对不同用户的不同需求给出相应的检索方案, 并可与目标查询或关系查询相结合来使检索结果得到优化。与传统的图像检索系统相比, 该系统更接近人对图像内容的理解, 并提供了良好的人机交互功能, 同时检索的效率也得到很大提高。

参 考 文 献

- [1] Y. Rui, T. S. Huang, S. Mehrotra, Relevance feedback techniques in interactive content-based image retrieval, 1998, SPIE 3312: 25-34.
- [2] D. Z. Hong, J. K. Wu, S. S. Singh, Refining image retrieval based on context-driven method, 1999, SPIE 3656: 581-593.
- [3] A. Jaimes, S. F. Chang, Model-based classification of visual information for content-based retrieval, SPIE 3656: 402-414.
- [4] E. J. Pauwels, G. Frederix, Finding salient regions in images: Nonparametric clustering for image segmentation and grouping, Computer Vision and Image Understanding, 1999, 75(1): 73-85.
- [5] Y. Y. Gao, Y. J. Zhang, Object classification using mixed color feature, Proc. ICASSP, Istanbul, 2000, 4: 2003-2006.
- [6] S. G. Mallat, Multifrequency channel decompositions of images and wavelet models, IEEE Trans. on ASSP, 1989, ASSP-37(12): 2091-2110.
- [7] Y. Y. Gao, Y. J. Zhang, N. S. Merzlyakov, Semantic-based image description model and its implementation in image retrieval, Proc. of ICIG'2000, Tianjin, 2000, 657-660.
- [8] 高永英, 章毓晋, 基于多级描述模型的渐近式图像内容理解, 电子学报, 2001, 29(10): 1376-1380.

- [9] G. Ciocca, R. Schettini, Using a relevance feedback mechanism to improve content-based image retrieval, Proc. of 3rd VISUAL'99, Amsterdam, 1999, 107-114.
- [10] 梅镇彤, 学习和记忆的神经生物学, 上海, 上海科学技术出版社, 1997, 51-53.

IMAGE RETRIEVAL SYSTEM BASED ON SEMANTIC FEATURES OF OBJECTS

Gao Yongying Zhang Yujin Luo Yun

(*Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*)

Abstract Most existing content-based image retrieval systems using low-level features that could not describe high-level semantics thoroughly and accurately. In this paper, a novel system for content-based image retrieval is designed and created, which combines image semantics based on a multi-level model for image description. In this image description model, image contents could be analyzed and represented through different levels and the transition from low-level features to high-level semantics is thus achieved. Corresponding querying mechanism and feedback are also proposed based on this image model. Aiming at object semantics in image, this querying mechanism is much closer to human beings' understanding of image contents so that it provides a convenient and effective querying procedure. The feedback used in the system is a self-adaptive relevance feedback based on object descriptions, it permits to propose different querying schemes according to the different demands raised by various users, and thus optimal results could be refined.

Key words Content-based retrieval, Image semantics, Image description model, Object understanding, Relevance feedback

高永英: 女, 1970年生, 现在美国攻读博士学位, 研究方向为图像检索和理解.

章毓晋: 男, 1950年生, 教授, 研究方向图像工程(图像处理, 分析, 理解及技术应用).

罗云: 女, 1970年生, 现在美国攻读硕士学位, 研究方向为图像处理和模式识别.