

# 基于用户多媒体数据管理模型的个性化图像检索

邱兆文<sup>1,2</sup>,张田文<sup>1</sup>

(1 哈尔滨工业大学计算机科学与技术学院,黑龙江哈尔滨 150001;2 东北林业大学信息与计算机工程学院,黑龙江哈尔滨 150040)

**摘要:** 为了提高基于内容的图像检索系统的性能和减少图像底层视觉特征和高层语义特征之间的鸿沟,本文提出了用户多媒体数据管理模型.包括用户语义模型、用户兴趣模型和用户意图模型.实验证明用户多媒体数据管理模型可成功为用户提供个性化图像检索和个性化推荐服务.

**关键词:** 个性化图像检索;用户语义模型;用户兴趣模型;用户意图模型

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2008) 09-1746-04

## Individuation Image Retrieval Based on User Multimedia Data Management Model

QIU Zhao-wen<sup>1,2</sup>,ZHANG Tian-wen<sup>1</sup>

(1. Institute of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China;

2. Institute of Information and Computer Engineering, Northeast Forestry University, Harbin, Heilongjiang 150040, China)

**Abstract:** In order to improve the retrieval accuracy of content-based image retrieval systems and reducing the 'semantic gap' between the visual features and the richness of human semantics. We proposed an user multimedia data management model. The user multimedia data management model includes user semantic model, user interest model and user intention model. Experiments show that user multimedia data management model can be successfully used in individuation image retrieval and individuation recommendation service.

**Key words:** individuation image retrieval; user semantic model; user interest model; user intention model

### 1 引言

随着 Internet 的快速发展和数码相机、扫描仪等数字设备的普及应用,互联网上的图像库规模在飞速增长.在过去十年间,国内外开发出了一些 CBIR (content-based image retrieval, 基于内容的图像检索) 系统,如 QBIC, Photo-book, Virage, VisualSEEK 等<sup>[1]</sup>,这些 CBIR 系统虽取得了一定的成功,但其检索性能离用户的期望还相差很远<sup>[2]</sup>.

在传统的 CBIR 系统中,图像是按底层视觉特征如颜色、纹理和形状进行索引.而人们趋向用高层语义特征如关键词,文本来解释图像和度量它们的相似度.这就是底层视觉特征和高层语义特征之间的语义鸿沟<sup>[3]</sup>.为了减少语义鸿沟,本文提出了用户多媒体数据管理 3 级模型,即用户语义模型、用户兴趣模型和用户意图模型.

本文从个人用户的角度将对用户有用的图像资源分成三个层次,用户保存的图像、用户曾浏览过的图像以及用户未曾见过但可能对用户有用的图像.通过对用户如何使用(包括:查询、浏览、下载等)这些图像的观察,建立用户多媒体数据管理 3 级模型,即用户语义模型、用户兴趣模型和用户意图模型.通过建立用户多媒体数据管

理 3 级模型,可以快速、方便地帮助用户实现个性化图像检索和个性化推荐功能.因此这将极大的提高人机交互的效率,也使得机器更人性化的为用户服务.

### 2 用户语义模型

#### 2.1 用户语义模型

如图 1 所示,用户语义模型分为三层,从下往上依此是用户特征语义、用户对象语义、用户场景语义、用户行为语义和用户情感语义.用户特征语义描述了用户感兴趣的颜色、纹理和形状等视觉特征.用户对象语义描述了用户感兴趣的对象和事物.用户场景、行为和情感语义则描述了用户感兴趣的场景、行为和情感等.用户语义模型可通过对用户访问过的信息进行自动隐式累积学习得到.

Level 1: 用户特征语义描述了用户感兴趣的颜色、纹理和形状等视觉特征.用户特征语义可通过图像自动直接提取.

Level 2: 用户对象语义描述了用户感兴趣的对象和事物.用户对象语义不能通过图像直接提取,需要先对每类图像建立对象语义模板,通过识别后得到对象语义.

收稿日期:2007-06-11;修回日期:2008-01-18

基金项目:国家自然科学基金(No. 60475011);黑龙江省科技攻关重点项目(No. GB06A107);黑龙江省青年科学基金(No. QC06C003)

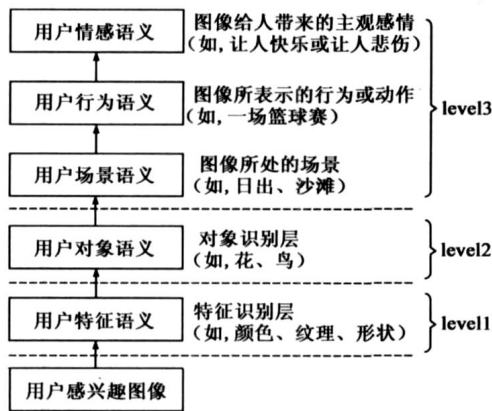


图1 用户语义模型

Level 3: 用户场景、行为和情感语义则描述了用户感兴趣的场景、行为和情感等。这是最高层的语义,需要建立推理规则,通过推理得到。

建立用户语义模型后,系统能支持高级语义查询。例如用户可以提交“快乐”、“悲伤”等高级语义检索相关图像。Level2 和 level3 被认为是高层语义,Level1 和 level2 之间的鸿沟就是语义鸿沟<sup>[4]</sup>。

### 2.2 用户语义模型的建立

先对图像库中所有的图像建立语义索引,提取每幅图像的特征语义、对象语义、场景语义、行为语义和情感语义。通过对用户访问过信息的学习就能得到用户感兴趣的图像,进而可得到用户特征语义、用户对象语义、用户场景语义、用户行为语义和用户情感语义。

在用户语义 3 层模型中,底层的视觉特征语义如颜色、纹理和形状等可通过图像直接计算。其中颜色特征的提取本文采用了基于 HSV 空间的 20 色非均匀量化算法<sup>[5]</sup>。与文献[6]中的 72 色非均匀量化算法和文献[7]中的 36 色非均匀量化算法相比,该算法降低了时间和空间复杂度,易表达用户喜爱的颜色,提高了检索的准确率。

基于 HSV 空间的 20 色非均匀量化算法中 H,S,V 分量的量化见公式(1)。

$$\begin{cases}
 H = \begin{cases} 0, & \text{if } h \in (330, 360] \quad [0, 25] \\
 1, & \text{if } h \in (24, 41] \\
 2, & \text{if } h \in (41, 75] \\
 3, & \text{if } h \in (75, 156] \\
 4, & \text{if } h \in (156, 201] \\
 5, & \text{if } h \in (201, 272] \\
 6, & \text{if } h \in (272, 285] \\
 7, & \text{if } h \in (285, 330] \end{cases} \\
 S = \begin{cases} 0, & \text{if } s \in (0.1, 0.65) \\
 1, & \text{if } s \in (0.65, 1) \end{cases} \\
 V = 0, & \text{if } v \in (0.15, 1] \end{cases} \quad (1)$$

按照以上的量化级,把 H,S,V 3 个分量合成一维特征矢量:

$$l = HQ_s Q_v + SQ_v + V \quad (2)$$

其中  $Q_s$  和  $Q_v$  是 S 分量和 V 分量的量化数,从公式(1)可知  $Q_s = 2, Q_v = 1$ ,于是公式(2)转化为:

$$l = \begin{cases} 3H + S + V, & \text{if } s \in (0.1, 1] \text{ 且 } v \in (0.15, 1] \\
 16, & \text{if } v \in [0, 0.15] \\
 17, & \text{if } s \in [0, 0.1] \text{ 且 } v \in (0.15, 0.65] \\
 18, & \text{if } s \in [0, 0.1] \text{ 且 } v \in (0.65, 0.9] \\
 19, & \text{if } s \in [0, 0.1] \text{ 且 } v \in (0.9, 1] \end{cases} \quad (3)$$

根据公式(3),  $l$  取值范围为  $[0, 1, \dots, 19]$ , 计算  $l$  获得 20 柄的一维直方图。

本文纹理特征的提取采用灰度共生矩阵,选取表示纹理特征的四个统计量:对比度(contrast)、纹理的一致性(uniformity)、像素对灰度的相关性(correlation)和熵(entropy)作为特征向量<sup>[8]</sup>。

在  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  四个方向上提取上述的四个纹理特征,组成 16 维的特征向量。

形状特征的常用描述方法有矩不变量和傅立叶描述子。因目前还没有通用有效的形状特征描述方法,本文只提取图像的颜色特征和纹理特征。

本文融合图像的颜色和纹理特征作为图像的特征向量,用支持向量机为所有的语义类建立语义分类器,来实现图像对象语义的自动提取<sup>[9]</sup>。

图像的场景语义、行为语义和情感语义计算较复杂,需要建立推理规则,目前主要靠手工标注。

### 3 用户兴趣模型

用户兴趣模型描述了该用户近期的兴趣和爱好,这主要是通过统计和学习用户访问过的信息来获得的,需要不断的更新。通过用户兴趣模型可以为用户提供个性化图像检索和个性化推荐服务。

表1 用户动作对用户兴趣的影响

用户动作	对用户兴趣的影响	影响因子
查询	中	1
浏览	中	1
放大显示图像	大	2
点击相关链接	大	2
保存图像	很大	3

对于不同的用户,对象语义有不同的相应权值,用户动作对用户兴趣影响的定义见表1。所有对象语义都对应一个五元组:(用户 ID, 语义代码, 最近访问时间, 权值, 代表图像)。用户兴趣模型的学习是自动隐式的,系统根据用户的访问记录自动调整用户的兴趣模型。用户执行查询、浏览、放大显示图像、点击相关链接、保存图像等动作,该用户相应语义代码的权值会增加。利用

用户兴趣模型检索图像的时候, 优先将权值大的语义类返回给用户. 根据最近邻原则, 用户最近访问的图像类在不久的将来访问的可能性较大. 因此, 在语义类权值相同的情况下, 优先返回用户最近访问的语义类.

#### 4 用户意图模型和用户意图推断

##### 4.1 用户意图模型

用户意图模型, 就是根据用户的当前操作或者前几步操作来描述用户下一步的操作, 或他的最终目的. 用户的意图可用多个层次来描述, 如图 2 所示. 相对底层的意图称为行为意图. 行为意图包括鼠标点击, 键盘输入和其他基本行为. 相对高层的意图称为语义意图, 表达了用户的抽象想法. 比如用户浏览图像最终目的是为了欣赏还是获取相关信息. 可以根据用户的行为意图推断用户的语义意图, 帮助用户完成任务, 为用户提供个性化服务.

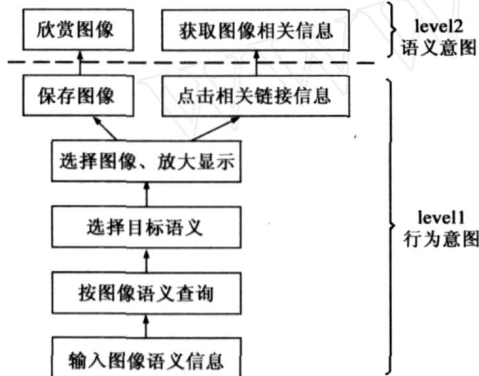


图2 用户意图模型

##### 4.2 用户意图推断

利用用户意图模型可进行用户意图推断, 帮助用户完成任务.

定义 1 在用户意图模型中, 设  $A$  和  $B$  为用户的任意两种行为, 且  $P(B) > 0$ . 则称比值  $P(AB)/p(B)$  为行为  $A$  在行为  $B$  发生的条件下的条件概率, 记作:

$$P(A|B) = P(AB) / p(B)$$

用户意图推断举例:

例, 对某用户进行意图建模后, 用户选择图像并放大显示的概率为  $P(A) = 0.8$ , 放大显示后保存图像的概率为  $P(B|A) = 0.4$ . 则现在该用户选择图像并放大显示后, 保存图像的概率为

$$P(B|A) = \frac{P(AB)}{p(A)}$$

由于  $B \subset A$ , 故  $AB = B$  于是

$$P(B|A) = \frac{P(AB)}{p(A)} = \frac{P(B)}{p(A)} = \frac{0.4}{0.8} = 0.5$$

得知用户下一步保存该图像的概率为 0.5. 所以当该用户选择图像并放大显示后, 就可直接帮助用户自动下载图像.

#### 5 实验与应用

##### 5.1 个性化图像检索系统

本文采用 J2EE 架构和 Agent 分布式处理技术完成了一个个性化图像检索系统 IIRS (Individuation Image Retrieval System, <http://mingcha1.com>). IIRS 主要由图像自动采集 Spider 子系统、图像语义自动标注子系统、用户兴趣学习子系统和人机协同查询子系统组成. 基于用户兴趣模型的个性化图像检索流程如图 3 所示. IIRS 提供了基于示例的图像检索和语义查询两种查询方式.

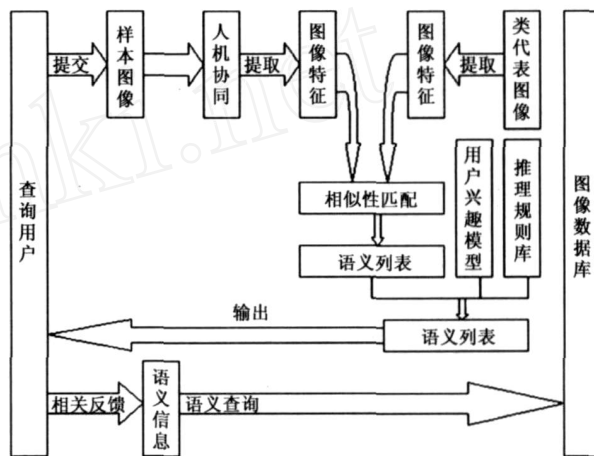


图3 个性化图像检索流程

##### 5.2 个性化图像检索应用

用户每次使用 IIRS 检索图像的时候, 如果输入了用户 ID 和密码, 则根据用户 ID 调用相应的用户多媒体数据管理模型, 为用户提供个性化图像检索服务. 如果用户不输入用户 ID, IIRS 系统根据用户浏览器的 cookie 自动识别出用户的 ID, 然后调用相应的用户多媒体数据管理模型, 提供个性化检索服务.

用户使用 IIRS 系统检索图像的次数越多, IIRS 系统为用户建立的多媒体数据管理模型越完善. 例如用户 user1 喜欢花卉, user1 经常检索和浏览各种鲜花. 即使 user1 不提供查询示例图像, 也不提供任何语义信息, 直接检索图像, 也能得到理想的结果. 检索结果如图 4 所示.



图4 个性化图像检索应用

示,而且用户经常访问的花卉语义权值大,排在前面.如果用户在语义列表中选择“兰花”,结果如图 5 所示,所有“兰花”图像按和用户最感兴趣的“兰花”图像的相似度进行排序.在 IIRS 系统中,不同的用户因用户兴趣模型不同,即使检索相同的语义,得的检索结果也不相同.

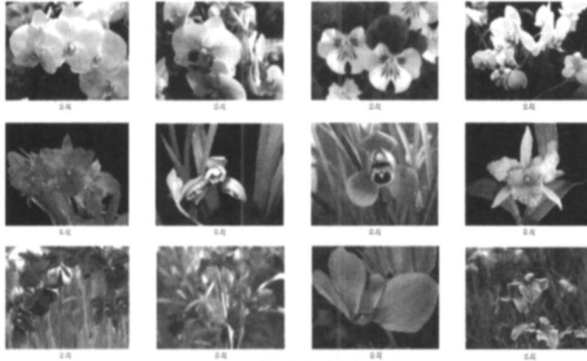


图5 图像按相似度排序

### 5.3 图像语义标注实验

本文实验图像大部分来自 Corel 图像库,将 6000 幅图像人工分成 80 个语义类,包括植物、花卉、动物、风景和建筑等.从 80 个语义类中随机选择 100 幅图像作为测试集.用户提交查询示例图像后,系统将根据已经建立好的对象语义分类器自动识别出图像的语义,并返回语义列表.每页显示 12 个语义类,如果第 1 页的 12 个语义类中有正确的语义类,用户就可直接选择正确的语义类进行检索,我们就认为语义标注结果正确.

图像语义标注的准确率定义如下:

$$\text{准确率} = \frac{\text{语义标注正确的图像}}{\text{所有查询示例测试图像}}$$

在图像语义标注实验中,将使用用户兴趣模型和不使用用户兴趣模型做了对比实验.

#### (1) 实验环境

硬件: CPU 2.66GHz, 内存 1GB. 软件: Windows Advanced Server 2003 + SQL Server 2003 + Eclipse + Tomcat 5.

#### (2) 实验结果与分析

从图 6 可以看出,使用用户兴趣模型进行图像语义标注取了较好的效果,用户提交查询示例图像后,正确

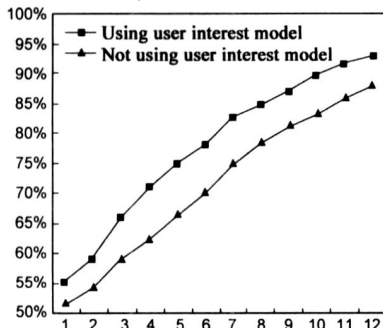


图6 使用用户兴趣模型与不使用用户兴趣模型的图像语义标注准确率对比

图像语义出现在第 1 页前 12 位的概率高达 93%,比不使用用户兴趣模型高出了 6.75%.正确图像语义出现在第 1 页第 1 位的概率只有 55%,这说明在现有的技术条件下还必须借助人机协同技术来提高识别的准确率.要想提高正确图像语义出现在第 1 页第 1 位的概率,必须使用更有效的图像底层视觉特征描述方法和建立更有效的图像分类器.

## 6 结论

(1) 基于 HSV 空间的 20 色非均匀量化算法和传统量化算法相比,特征维数少,计算量小,易表达用户对颜色的喜好,检索效果较好;(2) 用户多媒体数据管理模型可成功用户个性化图像检索和个性化推荐服务;(3) 用户语义模型和用户兴趣模型用于图像语义标注和个性化图像检索,取得了较好效果;(4) 用户意图模型可成功用于用户意图推断,帮助用户完成任务.

### 参考文献:

- [1] Smeulders A, Worring M, et al. Content-based image retrieval at the end of the early years [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12): 1349 - 1380
- [2] 黄祥林, 沈兰荪. 基于内容的图像检索技术研究 [J]. 电子学报, 2002, 30(7): 1065 - 1071.
- [3] Ying Liua, Dengsheng Zhanga. A survey of content-based image retrieval with high-level semantics [J]. Pattern Recognition, 2007(40): 262 - 282.
- [4] 王慧锋, 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): 513522
- [5] 邱兆文, 张田文. 一种新的图像颜色特征提取方法 [J]. 哈尔滨工业大学学报, 2004, 36(12): 1699 - 1701.
- [6] 曹莉华, 柳伟, 李国辉. 基于多种主色调的图像检索算法研究与实现 [J]. 计算机研究与发展, 1999, 36(1): 96 ~ 100.
- [7] Zhang Lei, Lin Fuzong, Zhang Bo. A CBIR method based on color-spatial feature [A]. IEEE Region 10 Annual International Conference 1999 [C]. Cheju, Korea: IEEE, 1999: 166 - 169.
- [8] B. S. Manjunath, W. Y. Ma. Texture features for browsing and retrieval of large image data [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1996, 18(8): 837 - 842.
- [9] Burges C. J. C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998, 2(2): 121 - 167.

### 作者简介:

邱兆文 男, 1974 年生, 副教授, 博士研究生, 研究方向为多媒体信息检索, 三维重建和虚拟现实. Email: qiuzyw@nefu.edu.cn

张田文 男, 1940 年生, 教授, 博士生导师, 研究方向为人工智能和计算机视觉.