

基于高层语义的图像检索算法*

王崇骏⁺, 杨育彬, 陈世福

(南京大学 计算机软件新技术国家重点实验室, 江苏 南京 210093)

Algorithms of High-Level Semantic-Based Image Retrieval

WANG Chong-Jun⁺, YANG Yu-Bin, CHEN Shi-Fu

(National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)

+ Corresponding author: Phn: +86-25-83593163, E-mail: chjwang_con@163.com, <http://www.nju.edu.cn>

Received 2003-07-07; Accepted 2004-02-03

Wang CJ, Yang YB, Chen SF. Algorithms of high-level semantic-based image retrieval. *Journal of Software*, 2004,15(10):1461~1469.

<http://www.jos.org.cn/1000-9825/15/1461.htm>

Abstract: IPSM is an integrated probabilistic image semantic description multi-level model. This model includes input layer, feature layer, semantic layer, synthetical probability layer, probability propagation layer, and semantic mapping layer. Based on the model and characterizing of the image high-level semantic content according to Bayesian theory, SHM (semantic high-level retrieval algorithm) and SRF (high-level semantic relevance feedback) for image retrieval based on high-level semantic content, for user relevance feedback respectively, are designed and implemented. Experimental results indicate that IPSM, SHM and SRF are effective in characterizing image high-level semantic content and can provide sound and robust image retrieval performance.

Key words: SHM; semantic feature extraction; Bayesian statistical learning; semantic thesaurus; relevance feedback

摘要: 利用 Bayes 统计学习和决策理论,建立了一种图像语义综合概率描述模型(image probability semantic model,简称 IPSM)。该模型是一种基于描述性特征建模方法的分层体系结构,由原始图像层、图像特征层、图像语义层、综合概率层、概率传播层和语义映射层 6 个部分组成。并在 IPSM 模型对图像的语义分类特征进行描述和提取的基础上,提出并实现了基于高层语义的图像检索算法(semantic high-level retrieval algorithm,简称 SHM)以及基于高层语义的相关反馈算法(semantic relevance feedback,简称 SRF)。实验结果表明,IPSM 模型及 SHR 和 SRF 两个算法能够有效地对图像的高层语义进行刻画,其图像匹配检索效果良好,并具有稳定的检索性能。

关键词: SHM;语义特征提取;Bayes 统计学习;语义分类辞典;相关反馈

中图法分类号: TP391 文献标识码: A

在基于内容的图像检索中,图像特征索引技术主要研究的是图像的视觉和形象特征,包括颜色及其分布关系、纹理、对象形状以及它们之间的空间关系等方面^[1]。以图像的底层视觉和形象特征为索引对图像进行检索

* Supported by the Natural Science Foundation of Jiangsu Province of China under Grant No.BJ200009 (江苏省自然科学基金)

作者简介: 王崇骏(1975—),男,江苏淮安人,博士,主要研究领域为信息检索,人工智能,图像处理;杨育彬(1977—),男,博士,主要研究领域为数据挖掘,人工智能,信息检索;陈世福(1938—),男,教授,博士生导师,主要研究领域为人工智能,信息检索,图像处理。

具有计算简单、性能稳定的特点,但目前这些特征都有一定的局限性.为了实现更为贴近用户理解能力的自然而简洁的查询方式,并提高图像检索的精度,在 CBIR(content based image retrieval)技术领域进行包含语义的检索方法的研究是十分必要的^[2].因此,近几年的研究逐渐转向基于区域和目标对象的高层图像语义描述特征的提取,尝试由图像的深层次特征推知高层次语义,从而使用高层语义特征计算图像相似程度.如何从图像的深层次特征中自动提取语义特征,成为基于内容的图像检索领域的难题.

目前,已有许多学者在图像高层语义特征提取技术研究方面进行了大量的尝试性工作.G. Sheikholeslami 提出一种 SemQuery 算法^[3],使用不同性质的图像特征进行语义聚类,每个语义聚类包含颜色、形状等各种底层特征子聚类,最终组成自顶向下的图像语义层次聚类结构.在进行图像检索时,提出一种基于多层感知器神经网络模型的特征合并算法 NeuroMerge,实现从用户提交的视觉查询到高层语义的映射,最终得出图像的相似程度.Fung 等人提出了学习图像块语义的方法,将图像的语义分为两个层次^[4]:原始语义和场景语义.前者从图像块的视觉特征得到,后者则从原始语义推出,并将图像的语义类别看成一种“词语”,从而可以使用文本检索算法实现图像相似性匹配.实验表明,该方法能够有效地检索图像的语义信息,如“瀑布”、“雪山”等.Vailaya 等人提出了一种能够有效进行检索的图像内容描述层次结构^[5],其特点在于提出了由图像的底层基本特征到高层语义分类特征之间进行映射的一种可行的数据描述方法,并通过基于 Bayes 概率公式的方法加以实现.文献^[6]提出利用主成分分析法来消除用户在检索过程中提供的正反馈样本在各特征空间中的分布特性的噪声,实现了对特征空间的降维,最终达到在不牺牲检索精度的前提下提高速度,降低存储复杂度.Iqbal 等人提出了“概念分组(perceptual grouping)”的概念,并利用 Bayes 方法设计了图像检索系统 CIRES.该系统能检索“植物”、“树”、“天空”以及“建筑”、“塔”、“桥”等结构性图像语义,具有较好的检索效果^[7].

目前,语义特征一般用文字描述,传统的方法一般用半自动或手动输入的方式提取语义特征.图像的高层语义特征需要人为理解,而且每个用户对图像语义的主观理解又不尽相同,这大大增加了图像检索中高层语义特征提取的难度.研究者普遍认为,在基于内容的图像检索研究中合理地引入人工智能和机器学习技术是解决高层语义难题的有效途径^[8].可以通过嵌入知识表示、推理学习机制和机器学习技术,采用统计的方法自动建立图像多维信息特征的索引结构,自动选取合适的底层特征组合实现图像与其语义内容的对应,实现图像语义的有效检索.

针对以上问题,本文利用 Bayes 统计学习和决策理论,基于图像的语义分类辞典结构建立了一种图像语义综合概率描述模型(image probability semantic model,简称 IPSM).该模型是一种基于描述性特征建模方法的分层体系结构,由原始图像层、图像特征层、图像语义层、综合概率集合、概率传播层和语义映射层 6 部分组成.基于图像语义综合概率描述模型 IPSM,我们提出一种图像高层语义特征提取算法(semantic feature extraction,简称 SFE)^[9].该算法以提取的图像底层特征(如颜色、形状及纹理等)为基础,通过 IPSM 模型和 Bayes 统计决策理论对图像的语义分类特征进行描述和提取.在此基础上实现了基于高层语义的图像检索算法(semantic high-level retrieval algorithm,简称 SHM)以及基于高层语义的相关反馈算法(semantic relevance feedback,简称 SRF).这两个算法已应用于我们研制的基于内容的图像检索系统中.经试用,其效果很好.

1 Bayes 统计决策理论简述

Bayes 学习算法是一种监督学习(supervised learning),是基于参数估计(parameter estimation)的方法^[10].在一般情况下,为减少需要计算的参数个数,并保证从有限数量的训练集中对其进行充分估计,无论是训练阶段的概率分布模型,还是在测试阶段定义的分类函数,都假定特征属性的各分量间相对于决策变量在统计意义上是相对独立的,也就是说,各分量独立地作用于决策变量.采用这种条件独立性假设的 Bayes 方法称为简单 Bayes,或朴素 Bayes(naive Bayes)模型.研究表明,尽管在实践中可能存在变量之间的依赖性,但基于这种条件独立性假设的朴素 Bayes 模型仍然具有相当的实用性,且该方法在存在噪声和无关特征属性的情况下仍很健壮^[11].本文的工作均以服从条件独立性假设的朴素 Bayes 模型框架为基础.

Bayes 统计决策方法进行分类的核心就在于如式(1)所示的 Bayes 定理(Bayes rule):

$$P(C_i | X) = \frac{P(x | C_i) \cdot P(C_i)}{P(X)}, 1 \leq i \leq m \quad (1)$$

其中,条件概率 $P(C_i|X)$ 称为后验概率(*posterior probability*),表示由样本 X 推出其属于分类 C_i 的可能性,条件概率 $P(X|C_i)$ 称为似然性估计(*likelihood*),表示数据样本 X 在分类 C_i 上的一种随机分布模型; $P(C_i)$ 为分类 C_i 的先验概率值, $P(X)$ 为样本 X 的先验概率值,是一个常量,仅具有规范化常数的作用。

对于模式分类问题,定义如式(2)所示的“0-1”型损失函数:

$$\lambda(C_i, C_j) = \begin{cases} 0, & \text{if } i = j \\ 1, & \text{if } i \neq j \end{cases}, \text{其中 } 1 \leq i, j \leq m \quad (2)$$

该式表示当算法分类结果为 C_i ,但正确分类结果为 C_j 时所产生的损失评价.对于数据样本 X ,算法所产生的平均损失见式(3).

$$L(C_i | X) = \sum_{j=1}^m \lambda(C_i, C_j) \cdot P(C_j | X) = \sum_{j \neq i} P(C_j | X) = 1 - P(C_i | X) \quad (3)$$

由式(3)可知,要使分类算法最优,需要满足 $L(C_i|X)$ 最小,使其类属后验概率最大,于是就得出 Bayes 统计决策理论用于分类问题最为常用的 MAP(*maximum a posteriori*)准则,见式(4)^[12].

$$C_i = \delta(X) = \arg \max_{1 \leq i \leq m} \{P(C_i | X)\} \quad (4)$$

根据式(1)中的 Bayes 定理,有

$$P(C | X) = \frac{P(X | C) \cdot P(C)}{P(X)} \quad (5)$$

由于 $P(X)$ 对于所有分类都是常数,所以,只需最大化 $P(X|C_i) \cdot P(C_i)$,式(4)可以进一步转化为式(6)的形式.

$$C_i = \delta(X) = \arg \max_{1 \leq i \leq m} \{P(X | C_i) \cdot P(C_i)\} \quad (6)$$

在上式中,关键在于似然性估计 $P(X|C_i)$ 的计算.根据朴素 Bayes 模型中的条件独立性假设,给定数据样本 X 的各个属性值之间是相互条件独立的,因此 $P(X|C_i)$ 的计算公式见式(7).

$$P(X | C_i) = P(x_1, \dots, x_n | C_i) = \prod_{k=1}^n P(x_k | C_i), 1 \leq i \leq m \quad (7)$$

将其代入式(6)可得式(8),

$$C_i = \delta(X) = \arg \max_{1 \leq i \leq m} \{P(C_i) \cdot \prod_{k=1}^n P(x_k | C_i)\} \quad (8)$$

由式(8)可以定义 Bayes 分类算法所使用的判别函数(*discriminant function*) $g_i(X)$,见式(9).

$$g_i(X) = P(C_i) \cdot \prod_{k=1}^n P(x_k | C_i), 1 \leq i \leq m \quad (9)$$

从而得出如下的 Bayes 决策分类规则,见式(10).

$$\forall j \neq i, g_i(X) > g_j(X) \Rightarrow X \in C_i, 1 \leq i, j \leq m \quad (10)$$

在对数据样本 X 进行决策分类的同时, Bayes 分类算法还可以为每个样本提供其属于该分类(假设为 C_i)的可信度,其大小为类属后验概率值 $P(C_i|X)$.

2 图像语义综合概率描述模型 IPSM

2.1 基本定义

在使用 Bayes 统计学习方法建立图像的高层语义概率描述模型之前,首先给出如下基本定义.

定义 1. 语义辞典(*semantic thesaurus*).图像库集合 Ω 中每幅图像 x 都属于一个图像语义分类框架,其语义分类结果以语义树的形式表示,如图 1 所示.

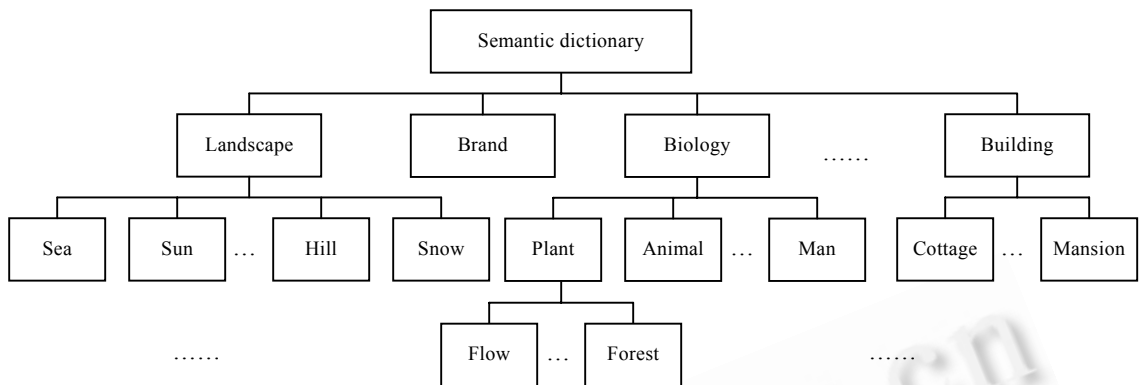


Fig.1 Semantic dictionary of image database

图 1 图像数据库的语义辞典

该语义框架称为图像库的语义辞典,由语义层级(semantic hierarchy)和语义分类标签(semantic classification label)组成.所有处于同一语义层级上的全部语义特征分别构成了图像 x 在该层级上的语义分类标签集合 $\mathbf{L}=\{C_1,\dots,C_i,\dots,C_k\}$,该集合包括该层的所有 k 个语义分类标签.考虑到人类在语义层感觉认知上的模糊性和多样性,对同一幅图像 $x\in C_i$ 可能存在语义 C_j ,使其同时满足 $x\in C_j(i\neq j,1\leq i,j\leq k)$.由于语义辞典为不同的语义层级均定义了对应的语义分类标签集合,因此它在一定程度上可以对图像进行不同精度的语义特性描述.

定义 2. 图像语义的先验概率.参与检索的每幅图像在同一语义层级上分别隶属于该层的语义分类标签集合 \mathbf{L} .按照统计学的观点,图像具有某一语义特征类别可以视为具有特定概率分布的随机事件,以图像语义分类的先验概率(prior probability)的形式来表示,其值为集合 \mathbf{L} 上的一组函数 $\{P(C_1),\dots,P(C_i),\dots,P(C_k)\}$.

定义 3. 图像语义的条件概率密度函数.每幅图像 x 都被视为随机变量 X 的抽样样本,其类属条件概率密度函数(class-conditional probability density function)为 $f_X(x|C_i)(i=1,\dots,k)$.对于图像检索问题,不失一般性,可以采用图像的一组特征向量代替图像的原始像素值矩阵,以减小计算量,即计算出表示图像像素矩阵的 n 种相互独立的底层特征向量(如 mosaic 分块主颜色、颜色共生矩阵等) $Y=\{y^{(1)},y^{(2)},\dots,y^{(n)}\}$ 来代替原始图像.特征向量 Y 具有类属条件概率密度函数 $f_Y(y|C_i)(i=1,\dots,k)$,称为图像语义的条件概率密度函数,用于描述其与语义分类标签集合之间的特定映射关系.

定义 4. 图像语义的后验概率.对于一幅语义分类情况未知的图像 x ,可以根据其底层特征向量推出其具有语义分类标签 C_i (即 $x\in C_i,i=1,\dots,k$)的可能性,用一组概率值集合 $\{P(C_1|x),\dots,P(C_i|x),\dots,P(C_k|x)\}$ 表示,称为图像语义的后验概率(posterior probability).

定义 5. 图像语义的分类.图像语义的分类问题的最终结果是一个 Bayes 决策规则集 $\delta(x):\Omega\rightarrow\mathbf{L}$,该决策集实现图像库中图像 x 与语义分类标签集合之间的函数映射.

2.2 IPSM模型

定义 6. IPSM 模型定义.根据上述的基本定义,提出了一种图像语义综合概率描述模型 IPSM,由原始图像层 X 、图像特征层 Y 、图像语义层 C 、综合概率集合 P 、概率传播层 Q 和语义映射层 A 这 6 个部分组成,其表示形式为

$$\text{IPSM}=M(X,Y,C,P,Q,A).$$

模型中各成分的含义定义如下.

定义 7. 原始图像层 $X=\{x_1,\dots,x_N\}$.

由图像库中所有图像(总数为 N)组成的集合,是 IPSM 模型的输入空间.每幅图像在各个语义层级上分别对应于该层的语义分类标签集合.

定义 8. 图像特征层 $Y=\{y_{i1},\dots,y_{ij},\dots,y_{in}\}_{i=1}^{N_c}$.

其中, N_c 为图像所提取的底层图像视觉特征的种类数目, n 为各种图像特征向量的维数, y_{ij} 为第 i 种图像特征

向量的第 j 维分量. IPISM 模型中所使用的图像视觉特征包括图像的颜色直方图、mosaic 分块主颜色、形状不变矩、颜色共生矩阵等.

定义 9. 图像语义层 $C = \{C_{i1}, \dots, C_{ij}, \dots, C_{ik}\}_{i=1}^{N_h}$.

其中, N_h 为图像库语义词典中的语义层级总数, k 为该语义层级上所有语义分类标签数.

定义 10. 综合概率集合 $P = \{P_{H_1}, \dots, P_{H_t}, \dots, P_{H_{N_h}}\}$.

对于每个语义层级 $H_t (1 \leq t \leq N_h)$, 都存在一个对应的概率集合

$$P_{H_t} = \{P(C_i | y_1), \dots, P(C_i | y_j), \dots, P(C_i | y_{N_c})\}_{i=1}^k,$$

其中, y_j 代表第 j 个图像底层视觉特征向量, $P(C_i | y_j)$ 为图像语义 C_i 的后验概率, 表示图像特征 y_j 导致高层语义 C_i 成立的概率值. 该层实际上是 IPISM 模型的输出空间. 根据 IPISM 模型中不同的语义层级概率集合 P_{H_t} 可以在不同层次上进行基于高层语义的图像检索.

定义 11. 概率传播层 $Q = \{Q_{i1}, \dots, Q_{ij}, \dots, Q_{ik}\}_{i=1}^{N_h}$.

其中, Q_{ij} 表示对于 IPISM 模型的任一输入 $x \in X$ (对应地, 有 $y \in Y$), 在获取其关于第 i 个语义层级中 C_j 语义的条件概率密度函数 $f_X(x | C_i)$ (即 $f_j(y | C_i), i=1, \dots, k$) 的前提下, 产生的语义后验概率计算函数 $f(y, C_j): Y \times C \rightarrow P$. 对应地, 整个概率传播层表现为一组概率计算函数集 $F(x): Y \times C \rightarrow P$.

定义 12. 语义映射层 $A = \{A_{i1}, \dots, A_{ij}, \dots, A_{ik}\}_{i=1}^{N_h}$.

其中, A_{ij} 表示对于 IPISM 模型的任一输入 $x \in X$ (对应地, 有 $y \in Y$), 在获取其关于第 i 个语义层级中的 C_j 语义的后验概率 $P(C_j | x)$ (即 $P(C_j | y), j=1, \dots, k$) 的前提下, 产生的 Bayes 决策函数 $g(P(C_j | x)): P \rightarrow C$. 对应地, 整个语义映射层表现为一组 Bayes 决策函数集 $G(x): P \rightarrow C$. 该层为 IPISM 模型的行为空间, 通过利用综合概率集合 P 完成图像层 X 和图像特征层 Y 到图像语义层 C 之间的映射.

3 基于高层语义的图像检索算法 SHM

3.1 SHM 算法描述

利用我们提出的图像语义综合概率描述模型 IPISM, 并基于 SFE 图像高层语义特征提取算法, 对于图像库中的一幅图像 x , 可以对其进行基于图像高层语义的图像检索. 我们提出一种基于高层语义的图像检索算法 SHM. 该算法根据提取到的图像语义, 按照图 1 所示的语义词典层次逐级进行语义特征的相似性匹配.

基于高层语义的图像检索算法 SHM 具体步骤描述如下.

算法 1. 基于高层语义的图像检索算法 SHM.

Step 1. 提取待检索图像 x 的底层图像特征, 利用图像高层语义特征提取 SFE 算法, 提取其语义分类特征描述, 将其对应于每个语义层级 $H_t (1 \leq t \leq N_h)$, 提取出的语义分类特征集合记为 $F_t = \{(c_1, w_1), \dots, (c_i, w_i), \dots, (c_{n_t}, w_{n_t})\}$, 其中, c_i 为语义分类特征, w_i 为该语义分类特征对应的后验概率值 $P(c_i | x)$, n_t 为集合 F_t 中的语义分类特征的个数. 类似地, 对于图像库中的图像 y , 其对应于各语义层级的语义分类特征集合记为

$$F_{t'} = \{(c'_1, w'_1), (c'_i, w'_i), (c'_{n_{t'}}, w'_{n_{t'}})\}.$$

Step 2. 令 $t = N_h, i = 0$, 图像 x 与图像 y 之间的距离度量为 D_{xy} , 初始值 $D_{xy}^{(0)} = 0$.

Step 3. 令 $i = i + 1, S = F_t \cap F_{t'} = \{c_1, \dots, c_j, \dots, c_q\} (1 \leq j \leq q \leq \min(n_t, n_{t'}))$, 按式(11)计算图像 x 与图像 y 之间的距离度量 $D_{xy}^{(i)}$:

$$D_{xy}^{(i)} = \sqrt{\sum_{k=1}^q (w_k - w'_k)^2} \times \frac{1}{1 - D_{xy}^{(i-1)}} \quad (11)$$

Step 4. 如果满足 $D_{xy}^{(i)} \leq T$ (T 为距离阈值), 则转至 Step 6; 否则, 转至 Step 5.

Step 5. 令 $t = t - 1$; 如果 $t > 0$, 则转至 Step 3; 否则, 转至 Step 5.

Step 6. 令 $D_{xy} = D_{xy}^{(i)}$, 若 $D_{xy} \leq T$, 则图像 x 与图像 y 在图像的语义内容上相似, 其相似度为 $S_{xy} = 1 - D_{xy}^{(i)}$, 将图像 y

加入图像 x 的检索结果集,按其相似度的大小排序输出;否则,图像 x 与图像 y 不相似.

Step 7. 算法结束.

3.2 SHM算法复杂度分析

SHM 算法充分利用了图像库语义辞典中的语义层级结构和语义分类标签信息,实现了逐步求精的图像语义相似性匹配.在最优情况下,只需计算两幅图像在一个语义层级 N_h 上的距离,其复杂度为 $O(n_i)$;在最坏情况下,需要遍历所有语义层级并计算相似性距离,其复杂度为 $O(N_h \cdot n_i)$,即算法的复杂度与语义辞典中语义分类项目的数目成比例,从而使得基于图像语义的图像检索结果更加符合人类对于语义信息逐层近似的认知能力,具有较好的实用前景.

4 基于高层语义的相关反馈算法 SRF

4.1 SRF算法描述

由于图像高层语义特征提取算法 SFE 的性能与其所选用的图像底层特征集有很大关系,对于以 QBE(query by example)方式进行的基于图像高层语义的图像检索,可以根据用户的反馈信息来调整各个语义分量特征的概率函数 $P(c|Y)(c \in C)$,以更好地反映用户的语义查询需求^[12,13].传统的相关反馈算法大都是直接基于图像底层特征的,未涉及图像的高层语义,而图像内容的复杂性限制了这些系统检索的精确度.为此,我们提出并实现了一个基于高层语义的相关反馈算法 SRF,在使用 SHM 检索算法得出图像 x 的语义检索结果图像集后,由用户对检索结果进行评价和标记,将满意度反馈给系统,系统将根据用户的反馈意见调整检索策略,返回更加精确的结果.如此反复,直至得到满意的检索结果为止.

在 SRF 算法中,用户可以提交的反馈信息包括“好”和“差”两种,分别表示被检索图像 y 在语义分类特征意义上符合或不符合待检索图像 x .SRF 算法通过有目的地调整图像所具有的语义分类特征的后验概率,然后再利用 SHM 检索算法对图像进行更加精确的检索.

SRF 算法描述如下.

算法 2. 基于高层语义的相关反馈算法 SRF.

Step 1. 获取用户提交的相关反馈信息,设 $X_r = \{x_1, \dots, x_{n_r}\}$ 是标记为“好”的相关图像集, $Y_r = \{y_1, \dots, y_{n_r}\}$ 是标记为“差”的不相关图像集.对每个语义层级 $H_i (1 \leq i \leq N_h)$,求得图像 x 的语义分类特征集合与图像集 X_r 中所有图像的语义分类特征集合的交集,定义为 $S_r = \{c_1^r, \dots, c_j^r, \dots, c_p^r\} (1 \leq j \leq p)$;同时,求得图像 x 的语义分类特征集合与图像集 Y_r 中所有图像的语义分类特征集合的交集,定义为 $L_{ir} = \{c_1^{ir}, \dots, c_k^{ir}, \dots, c_q^{ir}\} (1 \leq j \leq q)$.

Step 2. 令 $S = S \cap L = \{c_1, \dots, c_t, \dots, c_m\} (1 \leq t \leq m)$, $S = S - S, L = L - S$.

Step 3. 对图像 x 在每个语义层级 H_i 的语义特征集合 $F_i = \{(c_1, w_1), \dots, (c_i, w_i), \dots, (c_{n_i}, w_{n_i})\}$,当 $1 \leq i \leq n_i$ 时,对 F_i 中的 w_i 进行如下调整,见式(12).

$$\begin{cases} w_i = w_i + \alpha \cdot \frac{n_r}{n_{total}} \cdot \left| \frac{1}{m} \sum_{c_j^r \in S_r} w_j^r - w_i \right| - \beta \cdot \frac{n_{ir}}{n_{total}} \cdot \left| \frac{1}{m} \sum_{c_j^{ir} \in L_{ir}} w_j^{ir} - w_i \right|, & \text{if } c_i \in S' \\ w_i = w_i + \alpha \cdot \frac{n_r}{n_{total}} \cdot \left| \frac{1}{p} \sum_{c_j^r \in S_r} w_j^r - w_i \right|, & \text{if } c_i \in S_r \\ w_i = w_i - \beta \cdot \frac{n_{ir}}{n_{total}} \cdot \left| \frac{1}{q} \sum_{c_k^{ir} \in L_{ir}} w_k^{ir} - w_i \right|, & \text{if } c_i \in L_{ir} \end{cases} \quad (12)$$

其中, n 为图像 x 所具有的底层特征种类数, n_r 为图像集 X_r 中图像数目, n_{ir} 为图像集 Y_{ir} 中图像数目, $n_{total} = n_r + n_{ir}$, n_i 为图像 x 在语义层级 H_i 的语义特征数目, α, β 为常系数.

Step 4. 将调整后图像 x 的语义特征集合 F_i 提交给 SHM 算法重新进行检索.

Step 5. 重复执行 Step 1~Step 4,直到用户不再提交反馈信息为止.

Step 6. 算法结束.

4.2 SRF算法复杂度分析

SRF 算法根据用户所做的主观评价,对计算出的语义后验概率进行了重新调整计算,对于用户认为有效的语义特征,通过增大其后验概率值,使其在该语义上逐渐地“靠近”期望图像;与之相反,对于用户认为效果不好的语义特征,通过降低其出现的概率,使其逐渐地“远离”非期望图像,从而达到精确检索的目的.其时间复杂度是 $O(N_h \cdot n_f \cdot n_f)$,其中 n_f 为用户的反馈次数.

5 实验结果与分析

我们在有 5 000 幅通用图像的图像库中对本文提出的 SHM 算法和 SRF 算法进行了实验.按照如图 1 中语义词典所示的图像语义结构,我们选取了相应的分类图像组成训练样本集,对其进行训练,学习其对应的先验概率密度估计.训练图像样本的范围包括高山、建筑物、商标图案、鲜花、汽车等多种类型的图像.

按照图像库中不同种类的图像出现的比例来选取分类训练样本集的数目,最终选定的训练集规模为 $m=1402$ 幅图像.通过使用 LBG 算法对训练集的聚类量化^[9],对应于各个语义层级 $H_i(1 \leq i \leq N_h)$,分别生成 $l=2^{(i-1)}$ 个样本类别中心.

在随后的测试阶段,这些基于图像特征的语义分类特征将被用来对图像库中除训练集之外的其他图像进行基于高层语义分类的检索.我们从剩余的图像中按比例选择了 351 幅图像进行 SHM 算法和 SRF 算法的检索实验.其中,参与进行概率密度估计的图像底层特征包括:RGB 颜色空间的 $8 \times 8 \times 8$ 级颜色直方图,HSV 颜色空间的 $8 \times 3 \times 3$ 级量化 mosaic 分块主颜色特征,基于 Hu 不变矩的图像全局形状特征,颜色共生矩阵纹理特征以及图像能量谱的半径直方图和角度直方图特征.

(1) SHM 算法和 SRF 算法对建筑图像的检索实例与分析

图 2 给出了 SHM 算法和 SRF 算法对建筑图像的基于高层语义的检索实例.

一种好的图像检索算法,不仅要求有令人满意的检索精度,还需要具有很好的召回率指标,其性能要保持稳定^[14].SHM 算法能较好地按照“建筑”的高层语义返回图像检索结果,其检索出的结果图像大部分是建筑图像,具有较高的召回率和稳定的检索性能,如图 2(a)所示.我们利用颜色、形状或纹理等底层视觉特征进行检索,其结果中会检索出大量与原图语义不相关的图像,其召回率指标低,而且对于两幅语义相近的图像,其检索结果之间存在较大的差别,检索的稳定性较差.实验结果表明,如果仅从前 10 幅返回的检索结果来看,SHM 算法的检索精度并不比仅仅利用颜色、形状或纹理等底层视觉特征的检索精度高,这是因为 SHM 算法为了实现逐步求精的图像语义检索,利用了图像库语义词典中的每一个语义层级中的语义分类标签信息,以便使图像检索结果更加符合人类对于语义信息逐层近似的认知能力.这使得 SHM 算法在保证令人满意的检索精度的情况下,大幅度提高了图像检索的召回率,保持了稳定的检索性能.

从图 2(b)可以看出,基于高层语义的相关反馈算法 SRF 可以很好地提高图像检索的精度.用户对图像检索结果进行评价的过程,相当于在图 1 所示的图像库语义词典的语义树结构中向下搜索,越往下其语义描述的精度越高,因此其检索精度也越高.

(2) SRF 的检索结果与分析

图 3 给出了对于几种典型的图像语义分类特征,使用相关反馈算法 SRF 后,其图像检索精度随着用户反馈次数而提高的效果.

在图 3 中,当相关反馈次数为 0 时,表示的是图像语义检索算法 SHM 的图像检索精度.不难发现,对于不同的语义分类特征而言,其检索精度是有差别的.这说明其性能与其选用的图像底层特征集仍存在一定关系,受到该底层特征判别某种具体图像语义能力大小的影响.这也从另一方面表明提取有效的、适合于图像检索任务的图像底层特征的重要性.

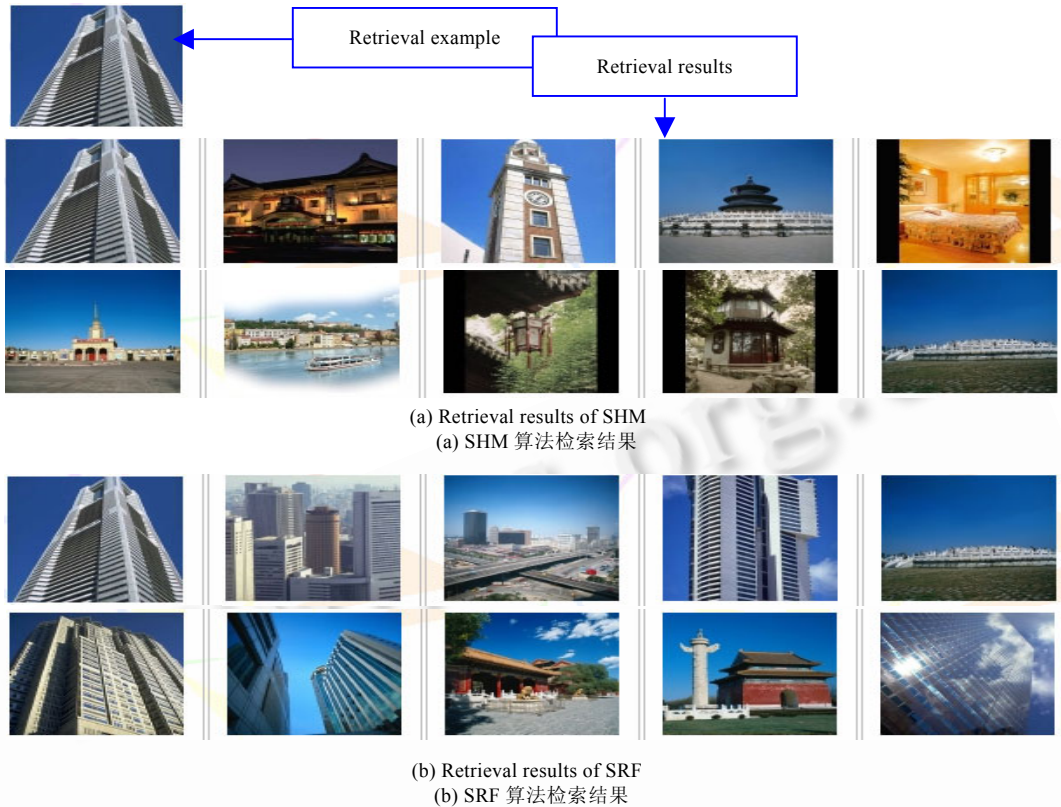


Fig.2 High-Level semantic-based retrieval example of building images
图 2 基于高层语义的建筑图像检索实例

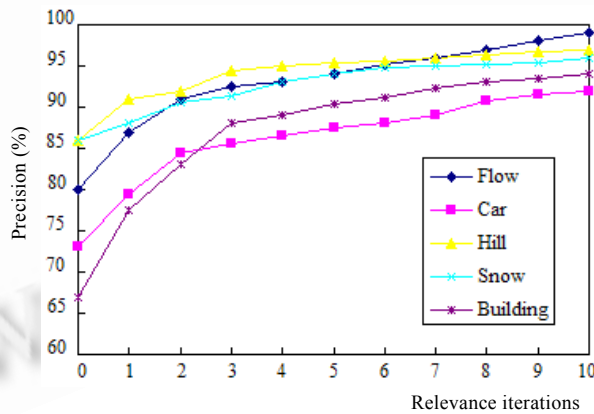


Fig.3 Comparison of image retrieval performance of SRF
图 3 SRF 算法的图像检索性能比较

从图 3 可以看出,用户进行前几轮(一般在前 3~4 轮)的相关反馈时,其检索精度有大幅度的提高.此后,检索精度的提升则渐趋平缓.这实际上也符合用户进行交互的心理学原理,人们往往不希望无休止地与检索系统进行交互.因此,在设计和评价用户反馈算法时,特别要注重其前几轮反馈的性能提升,SRF 算法一般需 3~4 次用户反馈即可达到一定的满意度.

6 结束语

本文利用 Bayes 统计学习和决策理论,对图像的语义分类辞典结构建立了一种图像语义综合概率描述模型 IPSM,并基于此模型提出一种图像高层语义特征提取算法 SFE.该算法同时综合图像的颜色、形状及纹理等底层特征,通过密度估计方法计算图像特征的语义先验概率函数,再根据 Bayes 定理对图像的语义分类特征进行描述和提取.在此基础上实现了基于高层语义的图像检索算法 SHM 以及基于高层语义的相关反馈算法 SRF.实验表明,IPSM 算法,SHR 算法和 SRF 算法能较为有效地对图像的高层语义进行刻画,其图像匹配检索效果良好,并具有稳定的检索性能.

上述的语义提取和检索方法必须对语义分类层次结构预先进行设定,并通过一组训练样本进行 Bayes 统计学习,因此,为提高该方法的实用性和扩展性,研究对语义分类特征层次的增量学习机制将是我们进一步研究的内容.

References:

- [1] Rui Y, Huang TS, Chang SF. Image retrieval, current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 1999,10(1):39~62.
- [2] Colombo C, Bimbo AD, Pala P. Semantics in visual information retrieval. *IEEE Multimedia*, 1999,6(3):38~53.
- [3] Sheikholeslami G, Chang W, Zhang A. Semantic clustering and querying on heterogeneous features for visual data. In: *Proc. of the ACM Multimedia*. Bristol: ACM Press, 1998. 3~12.
- [4] Fung CY, Loe FK. Learning primitive and scene semantics of images for classification and retrieval. In: *Proc. of the ACM Multimedia*, Vol 2. Orlando: ACM Press, 1999. 9~12.
- [5] Vailaya A, Zhong Y, Jaing AK. A hierarchical system for efficient image retrieval. In: *Proc. of the 13th Int'l Conf. on Pattern Recognition*. Washington: IEEE Computer Society, 1996. 356~360.
- [6] Su Z, Ma SP, Zhang HJ. Feature subspaces extraction for content-based image retrieval. *Journal of Software*, 2003,14(2):190~193 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/190.htm>
- [7] Iqbal Q, Aggarwal JK. CIRES: A system for content-based retrieval in digital image libraries. In: *Proc. of the Invited Session on Content Based Image Retrieval: Techniques and Applications*, Int'l Conf. on Control, Automation, Robotics and Vision (ICARCV 2002). Singapore: IEEE Computer Society, 2002. 205~210.
- [8] Koskela M, Laaksonen J, Oja E. Comparison of techniques for content-based image retrieval. In: *Proc. of the 12th Scandinavian Conference on Image Analysis (SCIA 2001)*. 2001. 579~586. <http://www.cis.hut.fi/picsom/scia2001.pdf>
- [9] Yang YB. Research and application on the key techniques of content-based image retrieval [Ph.D. Thesis]. Nanjing: Nanjing University, 2003 (in Chinese with English abstract).
- [10] Vasconcelos N, Lippman A. Bayesian representations and learning mechanisms for content based image retrieval. In: *Proc. of the SPIE: Storage and Retrieval for Media Databases*. 2000. 43~54. <http://citeseer.ist.psu.edu/vasconcelos00bayesian.htm>
- [11] Clark P, Niblett T. The CN2 induction algorithm. *Machine Learning*, 1989,3(4):261~283.
- [12] Lu Y, Hu CH, Zhu XQ. A unified framework for semantics and feature based relevance feedback in image retrieval systems. In: *Proc. of the ACM Multimedia*. ACM Press. 2000. 31~38.
- [13] Laaksonen JT, Koskela JM, Laakso SP, Oja E. Self-Organizing maps as a relevance feedback technique in content-based image retrieval. *Pattern Analysis and Applications*, 2001,4(6):140~152.
- [14] Smith JR. Image retrieval evaluation. In: *IEEE Workshop on Content-based Access of Image and Video Libraries*. 1998. 112~113. <http://www.ee.columbia.edu/~jrsmith/html/pubs/cbaiv198p.pdf>

附中文参考文献:

- [6] 苏中,马少平,张宏江.基于内容图像检索的特征子空间抽取.软件学报,2003,14(2):170~173. <http://www.jos.org.cn/1000-9825/14/190.htm>
- [9] 杨育彬.基于内容的图像检索关键技术及其应用研究[博士学位论文].南京:南京大学,2003.