基于语义分类的文物图像标注研究

乔荣华,周明全,耿国华
(西北大学 可视化技术研究所,陕西 西安 710127)

摘 要:由于图像数据中普遍存在的"语义鸿沟"问题,传统的基于内容的图像检索技术对于数字图书馆中的图像检索往往力不从心。而图像标注能有效地弥补语义的缺失。文中分析了图像语义标注的现状以及存在的问题,提出了基于语义分类的文物语义标注方法。算法首先通过构建一个 Bayes 语义分类器对待标注图像进行语义分类,进而通过在语义类内部建立基于统计的标注模型,实现了图像的语义标注。在针对文物图像进行标注的实验中,该方法获得了较好的标注准确率和效率。

关键词:基于内容的图像检索;语义鸿沟;语义分类;语义标注;贝叶斯语义分类器
中图分类号:TP391.41 **文献标识码**:A **文章编号**:1673-629X(2007)07-0200-04

Research on Semantic Annotation to Cultural Relic Images Based on Semantic Classification

QIAO Rong-hua, ZHOU Ming-quan, GENG Guo-hua
(Institute of Visualization Technology, Northwest University, Xi'an 710127, China)

Abstract: Because of the "semantic gap" which is often encountered in the image data, traditional CBIR technology can't deal with the problem of image retrieval in digital libraries sometimes. Image annotation can counterbalance semantic loss. In this paper the research situation of the problems existed in image semantic annotation is overviewed. An image annotation method based on the semantic classification is put forward. The algorithm classify images which are to be annotated firstly by the Bayes semantic classifier constructed, then the semantic annotation to the images are got by establishing a statistical annotation model within those semantic classes. Experimental results on a large collection of cultural relic images have shown the effectiveness and robustness of proposed algorithms.

Key words: content-based image retrieval; semantic gap; semantic classification; semantic annotation; Bayes semantic classifier

0 引 言

随着多媒体数据量的迅速增长,有效的图像检索成为数字图书馆建设中的众多关键技术之一。基于内容的图像检索(Content Based Image Retrieval ,CBIR)实现的基础是对图像底层特征信息的计算和比较,也即是"视觉相似"[1]。目前,有许多商业产品和实验原型系统被开发出来,如 QB IC, Photobook 和 Virage 等。

然而,由于计算机对图像信息的理解和人对图像信息的理解存在着客观区别,检索系统中就难免存在计算机认为的"视觉相似"和人们所理解的"语义相似"之间的"语义鸿沟(Semantic gap)"。"语义鸿沟"的存在是当前 CBIR 系统还很难被用户广泛接受的主要原因,如何连接"语义鸿沟"是目前 CBIR 系统亟待解决的技术难题。图像标注这个直观的想法能有效地弥补语义的缺失,是克服"语义鸿沟"问题的一个受到长期关注的研究课题。但纯粹的手工标注需要耗费巨大的人力和物力,因此研究人员将统计模型引入到标注研究领域,希望通过机器学习的方法自动标注图像,以支持基于语义的图像检索。

收稿日期:2006-10-01
基金项目:国家自然科学基金(60372072)
作者简介:乔荣华(1982-),女,陕西子洲人,硕士研究生,研究方向为计算机图形图像处理;周明全,教授,博士生导师,研究方向为图像处理与可视化技术;耿国华,教授,博士生导师,研究方向为智能信息处理。

1 相关工作

近年来图像自动标注技术正成为国际上图像检索领域的一个研究热点。图像自动标注技术的出现是为了自动获取图像的语义信息,从而在语义级别上对检索做出支持,许多机器学习方法由于能很好地获取图像特征和文本描述之间的对应关系,因而被引入这一领域[2],如 Mori[3] 等人提出的同现模型(Co-occurrence Model),研究人员通过网格(grid)方式将图像划

分成规则区域,然后将这些区域进行分类,根据不同类别的图像区域和关键词的共生概率来计算图像应该被赋予某一关键词的概率大小。而 Duygulu[4]则提出了一个基于机器翻译的对象识别模型,在该模型中,识别即是一个用关键字标注图像区域的过程。首先,通过 Nomalized Cuts 将图像分割为一些区域,然后提取区域特征,并将这些区域特征进行聚类,形成各个区域类。图像区域类和关键词之间的对应关系将通过 EM 算法得以学习。这个过程类似于从一个对齐的双语文本中学习构造一个词典。另外,Jeon[5]等提出了跨媒体相关模型(Cross - Media Relevance, CMR)。该模型则是通过计算每个关键词和组成待标注图像的图像单元的联合概率,从而作为将这一关键字标注给图像的依据。相关的研究还有 LDA 模型、CRM 模型等等。

目前已经存在很多带有手工标注的多媒体数据集合,如 Corel 图像库(www. corel. com),还有很多博物馆图像库(如 www. thinker. org)等等都带有手工标注的文本信息。这些关键字描述了图像所表达的概念,这为研究人员在语义自动标注上的研究提供了实验数据支持。

2　一个基于 CMR 模型的实验

基于 Duygulu 等人在互联网上公布的数据集(即 eccv_2002 数据集),对 Jeon[5]等提出的 FACMR 模型(fixed annotation - based CMRM)进行了实验。许多研究图像语义标注的研究者采用了该数据集。CMR 模型假定图像中的区域可以用少量的一组 blob 词汇所描述,blob 是通过对图像区域特征的聚类而产生,并从一组已有标注信息的训练图像集统计计算得出每个关键词和 blob 组的联合分布概率,并以此作为将某个关键词标注给图像的依据。通过训练集中的图像构建 blob 词组的过程:

步骤 1:将图像分解为若干区域。

步骤 2:提取各个图像区域底层视觉特征(eccv_2002 数据集对每个区域提取了 36 维的特征向量)。

步骤 3:聚类相似图像区域,得到 500 个区域类(blob),为每个类赋予一个唯一编号作为类标识(blob id)。

至此,图像集中的每一幅图像 I 都可以用一组离散的 blob 词汇来表示:

$$I = \{b_1, b_2, \cdots, b_n\}$$

FACMRM 模型算法的基本思想如下:

对于已有标注信息的训练图像集合 T,CMR 认为 T 中任意一幅图像 J 依据关键词 words 和 blobs 都有双重表示 $J = \{b_1, \cdots, b_m; w_1, \cdots, w_n\}$。其中 $\{b_1, b_2, \cdots, b_m\}$ 表示的是对应于图像区域的 blob 组,而 $\{w_1, w_2, \cdots, w_n\}$ 则表示图像中的关键词。并假定一组关键词 $\{w_1, w_2, \cdots, w_n\}$ 与 blob 词汇组所代表的图像对象之间存在某种关联。通过对训练集的统计分析,构建图像单元和语义关键词之间的关联模型。对于给定的一幅待标注图像,首先对图像进行分割并计算区域特征,然后对应图像区域找出在已有区域类(blob) 中最相近的 blob 组作为图像的表示。最后通过计算该 blob 组与训练集中各个关键词的联合概率,取联合概率最大的 N 个关键词作为图像的标注。

对于待标注图像 I,需要估算训练集中各个关键词 w 在图像中出现的概率 $P(w \mid I)$,假设用于表示 I 的 blob 组为 $\{b_1, b_2, \cdots, b_m\}$,则

$$P(w \mid I) \approx P(w \mid b_1, \cdots, b_m) \tag{1}$$

根据条件概率以及全概率公式,得出:

$$P(w, b_1, \cdots, b_m) = \sum_{J \in T} P(J) P(w, b_1, \cdots, b_m \mid J) \tag{2}$$

假定在选定图像 J 的情况下,在图像中观察到 w 和 $b_1, b_2, \cdots, b_m$ 的事件是相互独立的,这也是由之前确定的图像双重表示模型而得出的,因此公式(2) 等同于:

$$P(w, b_1, \cdots, b_m) = \sum_{J \in T} P(J) P(w \mid J) \prod_{i=1}^{m} P(b_i \mid J) \tag{3}$$

先验概率 $P(J)$ 对于训练集 T 中的所有图像都是唯一的常数。即如果训练集中共有 M 幅图像,则 $P(J) = 1/M$。

而关键词 w 或者某个 blob 在图像 J 中出现的概率可通过下式给出:

$$P(w \mid J) = (1 - \alpha_J) \frac{\#(w, J)}{\mid J \mid} + \alpha_J \frac{\#(w, T)}{\mid T \mid} \tag{4}$$

$$P(b \mid J) = (1 - \beta_J) \frac{\#(b, J)}{\mid J \mid} + \beta_J \frac{\#(b, T)}{\mid T \mid} \tag{5}$$

其中,$\#(w, J)$ 表示关键词 w 在图像 J 中实际出现的次数(通常为 0 或 1),$\#(w, T)$ 表示 w 在训练集 T 中出现的次数。类似地,$\#(b, J)$ 反映了 J 中图像区域被归于 b 的实际次数,$\#(b, T)$ 则是 b 在训练集 T 中出现的总次数。$\mid J \mid$ 代表了图像 J 中出现的关键词或者 blob 的总数目,而 $\mid T \mid$ 代表了整个训练集的大小。α_J 和 β_J 均为平滑参数。

在 VC6. 0 实验环境下,对此标注模型进行了实验,该算法的标注准确率和效率都较低,事实上,这也是目前许多图像语义标注模型中普遍存在的问题。根

据浙江大学彭青松在文献[6]中的观点,将内容一致的图像归类到同一集合,然后对集合中的图像进行统计分析,可以实现高效率的标注。据此针对文物图像实验了在对待标注图像进行语义分类的基础上进行的语义标注模型。

3 基于 Bayes 语义分类器的文物图像标注

3.1 待标注图像的语义分类

目前,利用图像的底层视觉特征进行语义分类是一个极具挑战性的课题,它是 CBIR 中的一项重要的研究内容,也出现了许多的分类算法。如文献[7]提出了使用支持向量机(SVM)学习自然图像的类别,学习到的模型用于自然图像分类和检索。文中对文物图像的语义分类是基于 Bayes 语义分类器[8]的,首先给出文献[8]中的一些基本定义:

定义 1 记 $C=\{c_1,c_2,\cdots,c_m\}$ 为包含 m 个给定语义类型的图像语义类型 $c_i \cap c_j=\varnothing$。图像 $I\in c_i \cap$ 图像 $I\notin c_j$,则称图像 I 归属于语义类 c_i, $i,j=1,2,\cdots,m, i\neq j$。

定义 2 设 $\boldsymbol{x}$ 为图像 I 的任一个特征向量,记 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_n\}$ 为图像 I 所有特征向量的集合。定义 f 为从图像 I 到图像特征集合 $\boldsymbol{X}$ 的映射,即 $f(I)=\boldsymbol{X}$,使得 $I\in c_i \cap I\notin c_j, i,j=1,2,\cdots,m, i\neq j$ 成立。

显然,映射 f 实现了基于语义分类的图像特征选择,其中 $\boldsymbol{X}$ 称做图像 I 的主属性。

记 $P(c_j \mid \boldsymbol{x})$ 为在特征向量为 $\boldsymbol{x}$ 的条件下,图像属于语义类型 c_j 的概率,按照 Bayes 公式:

$$P(c_j \mid \boldsymbol{x})=\frac{P(\boldsymbol{x} \mid c_j)P(c_j)}{P(\boldsymbol{x})} \tag{6}$$

其中,对于给定的一组训练样本图像,若样本总数为 N,语义类型 c_j 中包含的样本个数为 N_j,则记

$$P(c_j)=\frac{N_j}{N} \tag{7}$$

$P(\boldsymbol{x})$ 在贝叶斯分类框架中可以简单地设为被正规化的常量(因为通常认为任何的待分类图像都是平等的,即任意一幅待分类图像所包含的图像特征分布的出现概率与其他任何图像的出现概率都是相等的)。

于是,不难得出判别任意待分类图像的归属类别的准则:

$$\bar{\omega}=\arg\max_{\omega\in C}\{P(\omega \mid \boldsymbol{x})\}=\arg\max_{\omega\in C}\{P(\boldsymbol{x} \mid \omega)P(\omega)\}$$

由此得出的 $\bar{\omega}$ 即为具有最大类属概率的语义类别,也即待分类图像的最终归属类别。

研究表明,人类的视觉内容往往存在一定的偏差。这种偏差可以通过正态分布拟合给予弥补,即对于任一种语义类型 c_j,首先把同样的 Gaussian 内核放入它的所有训练样本的特征向量 $\boldsymbol{X}_i$,然后再把这些 Gaussian 内核累加起来作为条件概率 $P(\boldsymbol{x} \mid c_j)$ 的估计[8]:

$$P(\boldsymbol{x} \mid c_j)=\frac{1}{N_j}\sum_{X_i\in I\cap I\in c_j} G(\boldsymbol{X}-X_j,\sigma) \tag{8}$$

这里,$G(\boldsymbol{X}-X_j,\sigma)$ 是 Gaussian 内核,σ 是模糊度(即标准差),模糊度根据图像质量由用户指定。不同的视觉特征对不同语义的图像有不同的辨识能力。重要的是要从图像特征集合中选择一类或几类特征,使得被选择特征对特定语义类型的图像具有最强的表达能力,如图 1 所示。

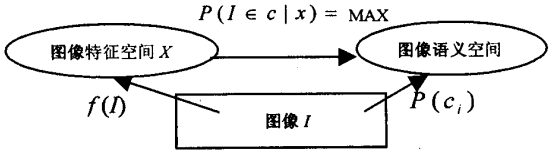

图 1 图像、图像特征和图像语义三者的关系

根据以上给出的贝叶斯分类器的形式化数学模型,可得出分类过程的整体算法流程如下:

(a) 从图像库中选择一部分图像组成训练图像集合。

(b) 对训练样本进行手工语义分类,设定好系统的语义分类器。

(c) 选择对于当前分类问题具有良好"判别能力"的图像低级特征集合,即给定一个语义类型集 C,寻找一个从图像 I 到图像特征向量集合 $\boldsymbol{X}$ 的映射 $f(I)=\boldsymbol{X}$ 使 $\mathrm{MAX}(P(I\in c_i \mid f(I)=\boldsymbol{X}), i=1,2,\cdots,m)$ 成立。分类器中所有的计算都是基于这些特征向量的。

(d) 对训练集合中每一幅图像提取在(b)中所选择的图像特征。

(e) 提取待分类图像特征,用公式(1)计算其后验概率。哪个类别的后验概率最大并且大于一定的阈值,测试图像就属于这个类别。

3.2 在文物图像上的试验

文中利用文物图像进行试验,所采用的文物包括瓦当、秦兵马俑、青铜器等等,如图 2 所示。该方法亦可扩展至其他图像数据库。

据此,设计了一个三类分类器,人工地将训练集合中的文物图像分别归入瓦当、青铜器、玉石三个语义类型中,并给图像都赋予相应的语义标签。通过在测试数据上进行的全面的试验,分别提取了图像的面积、周长、圆形度、二阶力矩等形状特征,以及包括粗糙度、对比度、方向度等的 Tamura 纹理特征,每幅待标注图像作为分类器的输入,根据对实验结果数据的分析,形状

特征对于分类器的区别能力要比其他视觉特征强,直观上用于实验的各类图像在边缘上的特征的区分度较强,因此用形状特征能够比较容易地区别各类。

图 2　实验数据库中的图像

在对待标注图像进行语义分类以后,其所属语义类的语义标签就自动地被图像所继承。然后再利用统计模型在语义类内部进行基于统计的图像标注,这样,图像语义标注的效率就大大地得到了提高。在所做的实验中,基于形状特征的实验达到了 81.5% 的较高的标注准确率,证明了算法的有效性和在文物图像上的适用性。图 3 对于算法在形状特征和纹理特征上各自的标注精度(标注准确率)做了相应的对比。

测试数据	纹理特征	形状特征
训练图像集	52.8%	81.5%
测试图像集	47.4%	71.4%
所有图像集	50.1%	75.0%

图 3　实验结果数据

4　结束语

图像语义标注作为图像检索领域一个较新的研究方向,是减小图像视觉特征和图像语义鸿沟的一种有效手段。基于语义分类的文物图像标注方法在实验中表现出了较好的标注准确率和效率。但图像语义标注在文物图像中的应用还有很多尚未解决的问题,需要做进一步深入的研究。

参考文献:

[1] 张鸿斌,陈　豫. 连接基于内容图像检索技术中的语义鸿沟[J]. 情报理论与实践,2004(2):196－198.

[2] 沈青松. 图像语义标注与检索及在数字图书馆中的应用[D]. 杭州:浙江大学,2006.

[3] Mori Y, Takahashi H, Oka R. Image－to－word transformation based on dividing and vector quantizing images with words [C]// In MISRM'99 First International workshop on multimedia intelligent storage and retrieval management. [s. l.]: [s. n.], 1999.

[4] Duygulu P, Bamard K, de Freitas N, et al. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary[C]// The 7th European Conf on Computer Vision. Copenhagen, Denmark:[s. n.], 2002.

[5] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross－media relevance models[C]// In: Proc of the 26th Annual Int'l ACM SIGIR Conf. New York: ACM Press, 2003:119－126.

[6] 彭青松. 多媒体交叉参照检索和语义自动标注[D]. 杭州:浙江大学,2005.

[7] Fu Y, Wang W, Gao W. Content－Based Natural Image Classification and Retrieval Using SVM[J]. Chinese Journal of Computers, 2003, 26(10):1260－1265.

[8] 许天兵. 一个用语义分类实现的图象检索框架[J]. 计算机工程与应用,2003(2):106－107.

(上接第 166 页)

尽管基于纠错码的叛逆者追踪模型有较高的性能,但提出划分更少数据块个数和需要更少加密密钥的新模型仍需要进一步研究。

参考文献:

[1] Fiat A, Naor M. Broadcast encryption[C]//In Advances in Cryptology－CRYPTO'93 Lecture Notes in Computer Science. [s. l.]:[s. n.], 1994:480－491.

[2] Boneh D, Shaw J. Collusion－secure fingerprinting for digital data[J]. IEEE Trans Inform Theory, 1998, 44:1897－1905.

[3] Chor B, Fiat A, Naor M. Tracing traitors[C]//in Proc. Advances in Cryptology－Crypto '94. Santa Barbara, California: Springer－Verlag, 1994:257－270.

[4] Fiat A, Tassa T. Dynamic traitor tracing[C]//in Proc. Advances in Cryptology－Crypto '99. Santa Barbara, California: Springer－Verlag, 1999:388－397.

[5] Safavi－Naini R, Wang Y. Sequential Traitor Tracing[C]// Proc Crypto 2000. Santa Barbara, California: Springer－Verlag, 2000.

[6] Dwork C, Lotspiech J, Naor M. Digital signets: Self－enforcing protection of digital information[C]//Proceeding of the 28th Annual Symposium on Theory of Computing. [s. l.]: ACM, 1996.

[7] Boneh D, Franklin M. An efficient public key tracing scheme [C]//in Proc Advances in Cryptology－Crypto'99. Santa Barbara, California: Springer－Verlag, 1999:338－353.

[8] van Lint. Introduction to coding theory[M]. Berlin: Springer－Verlag, 1982.