图像检索中语义映射方法综述

李志欣[1,2] 施智平[1] 李志清[1,2] 史忠植[1]
[1] (中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)
[2] (中国科学院研究生院 北京 100049)
(lizhixin@ics.ict.ac.cn)

摘 要 “语义鸿沟”已成为基于内容图像检索的瓶颈,解决这个问题需要建立从图像的低层特征到高层语义的映射.对当前语义映射研究进行了综述,首先给出一个结合语义的图像检索框架,并分析了图像内容的层次模型及图像语义的表示方法;然后根据算法的特点,将现有的语义映射方法和技术分为 4 大类,重点阐述了各类方法提出的思路、模型,并讨论各自的优势和局限性;最后以图像检索实际应用的需要为依据,提出在图像语义检索相关领域的重要课题和研究方向.

关键词 语义映射;基于内容图像检索;语义概念;图像标注;支持向量机;相关反馈
中图法分类号 TP391

A Survey of Semantic Mapping in Image Retrieval

Li Zhixin[1,2] Shi Zhiping[1] Li Zhiqing[1,2] Shi Zhongzhi[1]
[1] (*Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)
[2] (*Graduate University of Chinese Academy of Sciences, Beijing* 100049)

Abstract Semantic gap has become a bottleneck of content-based image retrieval. In order to bridge the gap and improve retrieval accuracy, a map from lower-level visual features to high-level semantics should be formulated. This paper provides a comprehensive survey on semantic mapping. Firstly, an image retrieval framework integrated with high-level semantics is presented. Secondly, image semantic description is introduced in two aspects: image content level-models and semantic representations. Thirdly, as the emphasis of this paper, semantic mapping approaches and techniques are investigated by classifying them into four main categories in terms of their characteristics. Various ideas and models proposed in these approaches are analyzed. In addition, advantages and limitations of each category are discussed. Finally, based on the state-of-the-art technology and the demand from real-world applications, several important issues related to semantic image retrieval are identified and some promising research directions are suggested.

Key words semantic mapping; content-based image retrieval; semantic concept; image annotation; support vector machine; relevance feedback

在信息技术高速发展的今天,各种信息源上数字图像的数量每天都在持续增长,如何对这些图像信息进行有效地组织、访问、存储和检索,已成为近年来的重要课题.图像检索技术自 20 世纪 70 年代

收稿日期:2007-12-19;修回日期:2008-04-14.基金项目:国家自然科学基金重点项目(60435010);国家“八六三”高技术研究发展计划(2006AA01Z128);国家“九七三”重点基础研究发展规划项目(2007CB311004).李志欣,男,1971 年生,博士研究生,讲师,CCF 会员,主要研究方向为图像理解、机器学习、基于内容的视觉信息检索.施智平,男,1974 年生,博士,助理研究员,主要研究方向为图像理解、机器学习、基于内容的视觉信息检索.李志清,男,1975 年生,博士研究生,讲师,CCF 会员,主要研究方向为图像理解、机器学习、视觉信息挖掘.史忠植,男,1941 年生,研究员,博士生导师,IEEE 高级会员,CCF 高级会员,主要研究方向为人工智能、机器学习、神经计算、认知科学.

就成为一个活跃的研究方向,研究者主要从数据库管理和计算机视觉 2 个方面对其进行研究,形成 2 种主流的检索技术:基于文本的图像检索和基于内容的图像检索(content-based image retrieval,CBIR)[1].

基于文本的图像检索主要由数据库技术的研究者提倡和发展,普遍采用的方法是人工标注图像并利用标注文本进行检索.它的最大优点是:如果图像的标注完整适当,会产生较好的检索效果. Google 和 Yahoo 在进行图像检索时都是采用基于文本的方式.但是,这种方法存在 2 个困难:1) 当图像数据库很大时,人工标注的工作量太大;2) 更重要的是,人工标注具有主观性和不确定性(不同的人看相同的图像会有不同的视觉理解),因而不能完全满足用户需求.为了克服基于文本的图像检索的局限性,计算机视觉的研究者提出了 CBIR,它已成为近十几年来研究的主流技术[2-5].著名的 CBIR 系统包括 QBIC[6], Virage[7], NeTra[8], SIMPLIcity[9], Blobworld[10]等.这些 CBIR 系统依靠特征提取和高维索引技术进行图像检索,采用的方法是:系统从每一幅图像中自动提取出若干低层视觉特征(如颜色、纹理、形状等),以高维向量的形式存入数据库,通过比较这些特征的相似度来获得检索结果.这种方法在某些特殊领域得到了很好的应用(如人脸识别[11],商标识别[12]等),因为在这些领域内视觉特征的相似度起了关键的作用.但在大多数情形下,用户习惯于根据图像的语义(如"日落")而不是视觉特征(如"红色或橙色的圆形")来进行查询,而视觉特征相似的图像其语义可能差别很大,这导致大多数 CBIR 系统在进行某些查询时会得到灾难性的结果.为避免这种现象,需要 CBIR 系统具有处理高层语义的能力,即能够获取图像的语义概念并在此基础上进行语义检索[13].

然而,获取图像的高层语义是非常困难的,因为图像的高层语义和低层特征之间没有直接关联,存在巨大的"语义鸿沟".语义鸿沟是指低层特征有限的表达能力与用户丰富的语义表达能力之间的差异.也就是说,从视觉数据中可提取到的信息与用户对同样数据的解释缺乏一致性[3,14].图像的语义通常在一个高层次上描述图像内容,无论提取到的低层特征是什么,都很难用这些特征直接推导出语义.因此,如何建立一个从图像的低层特征到高层语义的映射成为当前研究的热点.

1 结合语义的图像检索框架

用户对图像检索的要求主要体现在 2 个方面[3,13,15]:1) 用户需要对图像中实体的类别或特性进行查询;2) 用户要求搜索的图像既具有相似的低层特征也能表达类似的语义.目前,CBIR 不能实际应用的关键在于它只能提供图像的视觉特征相似度,而用户是根据语义相似度来搜索的;同时,从图像提取的特征有时与实际对象的特征差别是很大的,因为人是在一个三维空间学习和认识世界的,而图像目前只能提供二维的数据.

CBIR 具有较成熟的特征提取方法和高维索引机制,根据用户要求,可以在 CBIR 的基础上结合语义映射的方法来设计一个结合高层语义和低层视觉特征的检索系统,其基本框架如图 1 所示.由图 1 可见,结合语义的图像检索是 CBIR 的扩充,而 CBIR 是结合语义图像检索的基础.在 CBIR 的基础上,采用适当的语义映射方法获取图像中关键的语义信息、建立语义空间,同时提供语义的相似度度量方法,就可以实现图像的语义检索.事实上,目前大多数图像语义检索系统都是在相应的 CBIR 系统的基础上建立的.图像语义检索的最大困难在于缩减语义鸿沟,而这个问题正是通过建立从低层特征到高层语义的映射来解决的,建立这个映射需要解决 3 个主要问题:1) 提取有效的图像全局和局部低层特征;2) 提供图像内容的语义描述方法;3) 利用先验知识和各种学习算法将图像的低层视觉特征映射到高层语义,即提供语义映射的方法.鉴于已有很多文献对图像的低层特征进行了讨论[1,3,14],本文主要分析后 2 个问题.

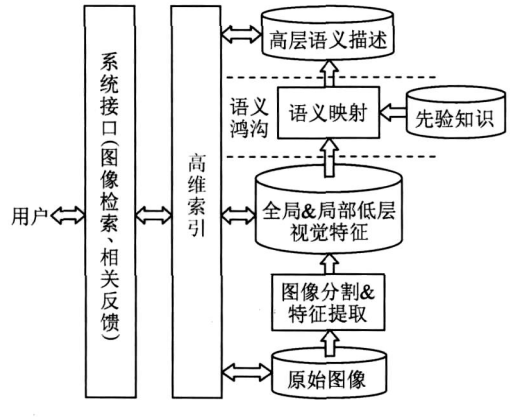

图 1 结合高层语义和低层特征的图像检索框架

2　图像的语义描述

图像内容具有模糊性、复杂性、抽象性等特点,仅仅用低层特征进行描述是远远不够的,需要利用高层语义来对图像的抽象属性进行描述.图像语义大致可分为特征语义、对象语义、场景语义、行为语义和情感语义等[16],用以对不同层次的图像内容进行描述.

2.1　图像内容的层次分析

由于图像检索需要在不同的内容层次上进行,所以可以利用层次模型对图像内容进行分析,相应地获取不同粒度的图像语义,从而逐步地理解图像内容.

Gudivada 等[2]将图像内容分为原始特征层和逻辑特征层 2 个层次.原始特征是指可以根据原始图像数据自动或半自动提取的特征,逻辑特征是指从原始特征通过直接或间接推理而得到的特征.Eakins 等[17]在此基础上将图像内容进一步分成 3 个层次:第一层仍为原始特征层,包括描述图像的视觉特征,如颜色、纹理、形状等,反映的是图像的一些具有客观统计特性的内容,对应于图像的特征语义;第二层为导出属性层,涉及由低层视觉特征推导而得到的属性,用以识别图像中描绘的对象(如"太阳"、"篮球"等),对应于图像的对象语义;第三层是抽象属性层,包括对对象和场景进行更高层的推理而得到的抽象属性(如"日出"、"篮球比赛"等),对应图像的场景语义、行为语义和情感语义等.一般将第一层与第二层之间的差距称为"语义鸿沟",图像检索是否真正使用了语义主要体现在是否获取了第二层的图像内容.目前很多研究者都致力于获取图像中"感兴趣"的对象语义,普遍采用自动或半自动的语义标注方法.

Jaimes 等[18]把图像内容概括成 5 层:区域层、感知区域层、对象部件层、对象层以及场景层.区域是像素的集合,是指图像中分割出来的连通的区域;感知区域是相邻且感知相似的区域的集合;一个或多个感知区域构成具有语义概念的对象部件;对象则可表示为若干关联的对象部件的集合;多个对象构成一个有意义的场景.此外,高永英等[19]也提出了一个包含 5 个层次的多级图像描述模型,依次为原始图像层、有效区域层、视觉感知层、目标层和场景层.该模型在不同层次上对图像内容进行分析,从而实现图像内容的全方位描述和渐进式的图像理解过程.

2.2　图像语义的表示方法

图像语义表示本质上是一种知识的表示,但与一般的知识表示有所不同.首先,图像中包含了大量的语义信息,并且这些信息之间存在着复杂的关系,因此需要一个具有强大的表达能力的方法;其次,由于图像理解的主观性,图像语义的表示方法需要一定的模糊和非精确性,用以支持图像的相似度检索.

MPEG-7致力于制定一个标准化的框架来描述多媒体内容,以便多媒体内容的有效表示和检索[20].为此,MPEG-7标准提出了多媒体内容描述子的概念,用于描述多媒体信息的颜色、纹理、形状等特征.但是,MPEG-7标准只是对内容的描述制定标准,不涉及如何提取和表示这些特征或内容,也没有涉及如何度量特征的相似度,而这些问题正是图像检索中最困难的问题,因此如何有效地描述图像语义仍需要研究者的不断努力.

目前最简单最常用的方法就是采用文本表示,即用文本对图像或图像区域进行解释.同时,可以利用词典(如 WordNet[21])或词汇表将文本表示的相关语义概念联系起来,从而获得一定的模糊匹配能力.文本描述的优点是直观、易处理,且可以表达一些抽象概念;缺点是文本描述自动获取困难,且对于概念之间的复杂关系缺乏足够的表达能力,难以独立完成语义描述的任务.鉴于目前语义映射方法还不成熟的情况下,大部分研究者还是采用词典或词汇表作为语义表示方法.

另一种表示方式是基于人工智能的知识表示方法,如语义网络、框架和框架网、基于本体的表示等.这种方法能够表达较为复杂的关系,并且具备模糊匹配能力,但是还不存在通用的适于各种背景的知识表示模型.例如,Lu 等[22]用一个语义网络结构来表示图像语义,图像库中的每幅图像用不同的关键词和权重来描述,一幅图像对应于一个或多个关键词,一个关键词也对应于一幅或者多幅图像.每个关键词按照一定的权值来描述一幅图像,权值越大,则该关键词越能清晰地描述该幅图像.Mezaris 等[23]使用一个对象本体来定义用户查询的高层语义概念(语义对象),使用一系列中间层描述器和关系识别器来描述对象间的相互关系,并结合相关反馈机制进行检索,具备一定的语义推理和扩充的能力.Town[24]等使用本体描述图像的低层特征和抽象概念及其相互关系,进行简单的语义推理,并在此基础上设计本体查询语言 OQUEL 进行图像检索.此外,Li 等[25]提出用语言变量描述图像语义特征,并

采用遗传算法来获取图像的语义. 该语言变量定义为一个五元组,包括变量值、与变量值对应的模糊集合、论域、语法规则以及语义规则.

3 图像语义映射的方法和技术

考虑一个图像数据库 $D=\{I_1, \cdots, I_p\}$(其中 I_i 为数据库中的图像)和一个语义词汇表 $V=\{w_1, \cdots, w_q\}$(其中 w_j 为语义关键词),则图像语义映射的目标是:给定图像 I_i,能从语义词汇表中找出最适合描述 I_i 的关键词集合 W. 图像语义检索的目标是:给定一个关键词 w_j,能从图像数据库中找出包含概念 w_j 的图像集 I. 实现这 2 个目标都需要通过一个训练集 $T=\{(I_1, w_1), \cdots, (I_D, w_D)\}$进行学习.

我们将当前图像语义映射的方法和技术分为 4 类:1) 图像的分类和聚类;2) 关联图像和语义的建模;3) 利用相关反馈学习图像语义;4) 特殊领域的语义映射方法. 需要注意的是,这些方法和技术的分类并不是相互独立的,相反,它们之间具有紧密的联系. 例如第 2 类和第 3 类方法都在不同程度上应用了第 1 类的方法技术.

3.1 图像的分类和聚类

大多数情况下,获取图像高层语义都需要使用机器学习技术,通过有监督和无监督的学习将图像归并到某种语义类,在一定程度上获得图像的语义标注信息. 机器学习可分为有监督学习和无监督学习. 有监督学习的目标是基于输入数据集合来预测输出数据的度量值(如语义类别标签);无监督学习没有输出度量值,它的目标在于对输入数据进行合理有效的组织或聚类.

3.1.1 基于分类的方法

有监督的分类方法首先通过学习、训练事先给定的经过语义标注的一组样本图像,获得图像语义分类器,然后利用分类器将未标注或未归类的图像归并到某一语义类. 最常用的有监督学习技术有贝叶斯分类器和支持向量机(support vector machine, SVM).

贝叶斯决策理论是模式分类的一个基本方法,使用这个方法进行分类时要求:1) 各个类别的总体概率分布是已知的;2) 要决策分类的类别数是一定的. 而实际应用时这些条件并不一定满足,需要对先验知识进行分析,估计先验概率和类条件概率密度,这正是训练和使用贝叶斯分类器的难点. 早期对贝叶斯分类技术的研究主要是检测简单的语义概念,例如区分图像是户内还是户外[26]、是城市还是自然风景[27]等. 这类方法可以看作是一个有监督学习的过程:首先选择一个图像训练集,由具有目标概念或不具有目标概念的图像组成,利用这个图像集训练一个二类贝叶斯分类器;然后将这个分类器应用到数据库中所有的图像进行概念检测,并判断图像是否具有目标概念. 由于分类器采用"一对所有"的方式(目标概念对所有其他概念)进行训练,称这种语义标注框架为监督 OVA (one vs. all). Carneiro 和 Vasconcelos 等[28-30]对监督 OVA 方法进行改进,采用基于最小错误率的优化准则和统计分类的思想,提出一种监督多类标注方法(supervised multiclass labeling, SML). 其基本思想是:将每一个语义概念定义为一个语义类别,引进一个随机变量 W,其取值范围为$\{1, \cdots, T\}$,使得当且仅当样本 x 具有语义概念 w_i 时 $W=i$(这里 $i \in \{1, \cdots, T\}$). 同时,引进条件概率密度 $P_{X|W}(x|i)$ 作为给定语义类别的低层特征分布,然后利用贝叶斯决策规则推导具有最小错误率的 W 的状态. SML 在训练分类器阶段为每幅图像提取一个特征集,利用多例学习(multiple instance learning, MIL)算法从多幅图像的特征集中学习语义概念,为每个语义概念建立概率模型,并采用期望最大化(expectation-maximization, EM)算法估计模型参数. 于是,在标注阶段可以通过各个分类器推导图像所具有的多个语义概念,同时根据后验概率产生语义标注的自然排序,便于实现语义的相似度检索.

另一类广泛使用的分类技术是 SVM,它具有很强的理论基础,在图像检索中得到了较好的应用. SVM 最初设计为二类分类器,假设有训练集$\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$,其中训练数据$\{x_1, x_2, \cdots, x_n\}$是某个向量空间 $X \subseteq R^d$ 中的向量,而它们给定的标注 $y_i \in \{-1, 1\}$. 训练集中的向量 x_i 分属于 2 个不同的类别:类 I 的 $y=1$,类 II 的 $y=-1$,我们希望在向量空间中找到一个超平面来分离不同类别的数据. 在所有可能的超平面中,最优超平面是唯一的,它使得超平面与各个类最接近的数据点之间的间隔最大,如图 2 所示[31]. 在超平面一侧的数据标为 +1,另一侧标为 -1,"支持向量"是指最接近超平面的训练样本. 为了利用 SVM 学习多个语义概念,需要对每个概念单独进行训练. 例如,Cusano 等[32]将 SVM 进行推广,用以处理多于二类的情况. 选择 7 类语义关键词(天空、大地、雪、建筑物等)进行实验,利用训练得到的多类 SVM 分类器

对图像区域进行分类,从而产生图像的语义标注. Gao 等[33]提出一种分层提升算法来合并特征层次,增进 SVM 图像分类器在高维特征空间中的训练. 该算法将高维多模态异类视觉特征划分为多个低维单模态同类特征子集,每个子集用于表示图像的某个视觉特性. 使用主成分分析的方法为每个特征子集训练一个弱分类器,然后选择最具代表性的特征集,将这些弱分类器合并成为一个优化的分类器,用以预测图像包含的对象或语义概念. Chang 等[34]提出一个基于内容的软标注(content-based soft annotation,CBSA)系统为图像提供语义标注. CBSA 首先选择一个训练图像集对全体分类器进行训练,其中每幅图像具有一个标注(如森林、动物、天空等);然后将全体分类器应用到一幅给定的图像上以获取图像的多个软标注. 该系统使用 2 种学习方法:SVM 和贝叶斯点机,并对这 2 种方法的标注精度进行了比较.

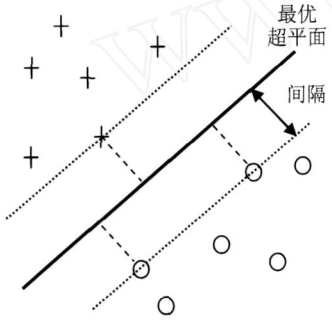

图 2　一个简单的线性 SVM

此外,神经网络和隐马尔可夫模型等机器学习技术也用来对图像的语义概念进行识别和检测. Town 等[35]首先选择 11 类语义概念,然后将大量训练数据(分割区域的低层特征)输入神经网络分类器,从而建立图像低层特征和高层语义的联系;但该方法要求的训练数据量大、计算复杂度高. Li 等[36]提出图像自动语义索引系统(automatic linguistic indexing of pictures,ALIP),该系统使用一个二维多分辨率隐马尔可夫模型捕获给定语义类别的图像特征之间和内部的空间依赖关系,各个语义类别的模型是分别独立学习和存储的. 标注方法是计算查询图像与各个语义类别之间的相似度,然后选择最相似的类别所包含的语义进行标注. 之后,他们又提出了一个实时图像标注系统[37],它继承了 ALIP 的高级学习架构,且建模方法更简单,可以进行统计相似度的实时计算. 作为第一个实时图像标注引擎,该系统对图像检索的实际应用有重大影响.

基于分类的方法将各个语义类别(一个关键词或关键词集合)看作独立的概念,为每个语义类别建立各不相同的分类模型,检索准确率较高. 这类方法存在 2 个问题:1) 需要大量用于训练的图像样本,并要对样本进行细致的人工标注,这是一个繁杂枯燥的工作,而且容易出错;2) 由于采用离线学习的方式,在学习和应用阶段训练集和概念是相对固定的. 如果应用领域发生变化,就需要提供新的样本以保证分类器的效率.

3.1.2　基于聚类的方法

有监督学习存在输出变量指导学习过程,而无监督学习没有输出值,它的任务只是寻找如何将输入数据进行组织和聚类的方法. 图像聚类是典型的无监督学习技术,它根据图像内容将库中图像(或图像区域)聚类到某些有意义的集合. 图像聚类的原理是将图像集分组成为多个聚类,使得位于同一聚类内的图像相似度尽可能大,而位于不同聚类的图像的相似度尽可能小;然后利用统计方法为每个聚类加一个类标签,以获得各个图像聚类中的语义信息.

图像聚类最常用的技术是传统的 k-means 聚类及其变形. Stan 等[38]提出的语义标注系统由 2 个阶段构成:首先应用一个改进的 k-means 聚类算法在低层特征空间中寻找数据的自然模式,该算法使用非欧氏距离的度量公制,以适合人类感知的方式进行设计;然后使用统计学方法对各个聚类的差别进行测量,并由此产生从最重要低层特征到各个聚类使用最频繁的关键字之间的一系列映射规则. 利用这些映射规则可以获取新加入数据库的未标注图像的语义内容. Bilenko 等[39]提出一个由 k-means 算法派生的半监督聚类算法,利用少量的标注数据进行无监督学习. 在该算法中集成了 2 种学习技术:1) 基于约束条件的学习方法通过修改聚类目标函数指导聚类算法,使得训练数据能够以适当的方式分组;2) 基于度量公制的学习方法为各个聚类学习自适应的度量公制,使得各个聚类更符合人类的感知概念. Jin 等[40]将 PCK-means(pair-wise constraints k-means)算法应用到图像的语义标注中. 在学习阶段,使用 PCK-means 算法对图像的分割区域进行聚类;在标注阶段,使用贝叶斯方法计算赋予各个区域聚类的语义概念的后验概率. 这样,对于一个新的图像可以选择最高后验概率的语义概念进行标注.

Chen 等[41]使用 CLUE(CLUster-based rEtrieval of images)的方法来缩减语义鸿沟,尝试检索语义连贯的图像聚类. 基于相似语义的图像倾向于分组到

同一聚类中这个前提,CLUE 使用 NCut (normalized cut)聚类算法[42]将目标图像聚集到不同的语义聚类中,然后根据用户反馈调整相似度度量模型,并显示与用户查询最接近的图像聚类. 这种方法对于流形数据的聚类取得了成功,但 NCut 聚类算法不能产生一个显式的映射函数. Zheng 等[43]提出 LPC (locality preserving clustering)聚类算法,该算法具有非线性频谱聚类算法的数据表达特性,同时能够提供显式的映射函数. 实验结果表明,LPC 聚类算法具有与 NCut 聚类算法相当的精度,且计算效率更高.

基于聚类的方法通常在语义映射的训练阶段使用图像聚类技术,对目标图像进行有意义的分组. 该方法对于手工标注的训练集要求较低,训练数据和语义概念具有可扩展性. 但是严格地说,单纯的图像聚类并不能为一个新的图像获取显式的语义标签,需要与其他技术结合使用来进行图像的自动语义标注,充分发挥其效率,并达到较高的检索精度.

综上所述,图像分类和聚类的研究目的是从低层视觉特征提取图像的语义信息,辅助图像的存储和管理,优化图像索引策略,实现图像快速、有效的检索.

3.2 关联图像和语义的建模

许多研究者通过建立基于学习的关联模型来进行自动语义标注,利用现有的已标注好的图像数据集,使用机器学习技术学习图像的视觉特征和文本关键词的关联;然后将这种关联应用于未标注的图像来预测图像的语义信息,并实现图像的多模态检索. 与图像分类方法为每个语义概念训练一个分类器不同,这类方法只学习一个关联模型并将该模型应用于所有的语义概念.

最早的关联模型是 Mori 等[44]提出的共生模型,这个模型采用 2 个过程:一是将训练集中的每个图像按照统一大小的网格划分为固定大小的图像方块,这些图像方块继承了原图像的所有关键词;另一个是对这些图像方块用向量量化的方法进行聚类,然后根据聚类的图像方块中关键词出现的频度来标注某个图像方块的聚类. 这个模型避免了复杂的图像分割过程,实现比较简单,但所得到的图像标注精度不高. 于是,Duygulu 和 Barnard 等[45-46]提出一种机器翻译模型,用 NCut 聚类算法[42]将图像分割为任意形状的区域,这些区域大致对应于一个对象或对象的一部分;然后依据区域特征将图像区域聚类为量化区域,同时对标注关键词进行聚类. 随之而来的一个自然的假设是:图像的量化区域和某个关键词聚类之间存在某种隐含的一一对应关系. 借助机器翻译的概念,该模型将量化区域和关键词聚类看作是 2 种对等的"语言",于是标注的过程可以看作是一个将图像量化区域翻译为关键词聚类的过程. 该模型采用 EM 算法来估计区域和关键词的联合概率分布,一旦经过学习确定了模型参数,就可以用于标注新的图像. 这类模型具有较高的标注精度,对后来的研究工作起到了很大的促进作用,其缺陷是图像分割的结果会对标注精度造成很大的影响,实施的难度较大.

Blei 等[47-48]使用更复杂的 CORR-LDA (correspondence latent Dirichlet allocation)模型为关键词和图像创建一个基于语言的关联. 该模型首先使用 Dirichlet 分布产生一系列隐藏变量(潜在层面)用以关联文本模态和图像模态,则一幅图像可分解为一系列潜在层面的混合;然后在这些潜在层面中选择一个子集转换为若干基于 LDA 的混合模型,使用高斯分布为图像的区域特征建模,使用多项式分布为标注关键词建模,从而产生图像的语义标注. Monay 等[49]随后提出概率潜在语义分析(probabilistic latent semantic analysis, PLSA)模型,该模型也将图像看作一系列潜在层面的混合,但与前面的模型不同的是,PLSA 考虑了区域和关键词内在的关系,并不认为它们是相互独立的,而且将图像和文本视为 2 种不对等的模态,在学习过程中能针对它们的影响做出不同的变化.

Jeon 等[50]提出的跨媒体相关模型(cross-media relevance model, CMRM)也采用分割区域表示图像,但与翻译模型不同的是,它并不认为图像的关键词和区域之间是一对一的对应关系,而是通过学习关键词和区域的联合概率分布为整幅图像标注若干关键词. Lavrenko 等[51]随后提出类似的连续空间相关模型(continuous-space relevance model, CRM). CRM 与 CMRM 有 2 点重要区别:1) CMRM 是一个离散模型,不能利用连续的特征,使用它进行标注需要对连续的特征进行量化得到离散的词汇表,而 CRM 可以对连续的特征建模;2) CMRM 依赖对特征向量的聚类,标注质量对聚类错误非常敏感,需要预先选择聚类粒度,而 CRM 不依赖于特征向量的聚类且不受聚类粒度问题的困扰. Feng 等[52]在此基础上提出多贝努里相关模型(multiple Bernoulli relevance model, MBRM),该模型利用训练集计算关键词和图像特征的联合概率分布,使用

MBRM 估计关键词概率,使用无参数核密度函数估计图像区域特征的概率,然后使用训练好的模型对测试集进行标注. 与此类模型相关的还有 Jin 等[53]提出的一致语言模型,该模型利用关键词与关键词的关联来增强标注精度.

在以上讨论的方法中,图像都由区域(分割后的不规则区域或划分后的图像方块)来表示. 一旦图像分割完毕,就使用量化方法获取一个有限的量化区域表. 在这些模型中,图像被看作一系列由隐藏变量(潜在层面)产生的关键词和区域的集合,这些潜在层面在图像区域上产生多元分布,在关键词上产生多项式分布. 因此对给定图像进行标注的问题就可以通过学习关键词与图像区域的联合概率分布来解决. 这类方法只需要训练一个模型,具有可扩展的训练进程,对训练样本手工标注的精度要求较低,所需的计算量和工作量较小. 但是,这类方法大多依赖于精确的图像分割,而图像分割至今仍是一个未解决的公开难题,所以实施起来有一定难度;而且这类方法没有显式地将图像语义对应的关键词作为类,也就无法在识别或检索意义上保证语义标注的最优化. 也就是说,它不是基于最小错误率的标注,而是在假定的混合模型中寻找具有最大联合概率的语义标注.

3.3　利用相关反馈学习图像语义

与第 3.1,3.2 节讨论的离线学习技术不同,相关反馈是一种在线学习技术,其基本思想是在检索过程中结合人类感知的主观性,给用户评估检索结果的机会,并依据这些评估学习用户意图,以获取更完善的查询和相似度度量方法. 相关反馈技术的一个典型方案如下[54]:

Step1. 通过图例查询、草图查询等,机器首先提供一个初始的检索结果.

Step2. 用户在现有的显示结果上提供一个判断,决定结果是否(或在多大程度上)与查询要求相关.

Step3. 学习用户的反馈结果,再次给出检索结果,转 Step2 或结束.

相关反馈可以看作是一个学习问题:用户提供查询结果的正例和反例作为训练样本,系统通过学习精练检索结果,其学习策略大致可分为参数调整方法、移动查询点方法和机器学习方法. 利用相关反馈学习用户意图既能在视觉特征层次上进行,也能在语义层次上实现. 但是,在相关反馈背景下的学习具有 3 个重要特征[54]:1) 训练样本少. 用户每次反馈的样本数目通常不会超过 20 个;2) 训练样本不对称. 用户提供的正例和反例数目不均匀;3) 实时要求. 由于有用户的参与,需要快速地得出检索结果. 可见,相关反馈背景下的在线学习方法需要考虑以上种种问题,与离线学习采用的算法相比有其自身的特点. 例如,Tong 等[55]提出 SVM 主动学习算法来处理用户的反馈信息,该算法选择最能表示用户查询需求的图像并迅速地学习相关图像和不相关图像的边界,使得查询结果最大限度地满足用户需求. 解决以上问题的技术细节见文献[54,56-57],本节重点阐述在相关反馈过程中学习图像语义和结合语义进行检索的方法和技术.

早期对相关反馈的研究主要集中于视觉特征层次,但也有研究者对于在相关反馈过程中结合图像的语义信息做了努力. Minka 等[58]提出的 FourEyes 系统和 Cox 等[59]提出的 PicHunter 系统都在学习过程中利用了图像的隐含标注. Tieu 等[60]在一个多于 46000 个特征的特征空间中利用 boosting 技术训练一个分类函数,这些特征表现为稀疏情况且具有高的曲线峰态,被认为能表达图像的语义概念.

为了充分利用基于关键词表示和基于视觉特征表示的图像检索的功能,在相关反馈的背景下,研究者提出了大量集成关键词和视觉特征的方法. 这类集成方法的关键问题是如何结合关键词和视觉特征,使它们在检索中相互补充和协作,从而获取有效和高精度的检索. Lu 等[22]最早提出如图 3 所示的基于相关反馈的集成语义和视觉特征的统一框架. 在图像数据库的顶端构造语义网络,使用机器学习技术学习用户的查询和反馈,以更新语义网络和增进系统性能,并利用 iFind 框架集成图像的语义信息及其低层特征索引. Zhang 等[61]将此框架延伸为基于内容的图像搜索引擎,在相关反馈过程中将关键词标注从已标注的图像传播到未标注图像上. 通过这样的传播过程,越来越多的图像能够获得隐含的语义标注,这样的标注传播过程也能帮助系统积累学习到的知识,从而增进未来的检索性能. 在此基础上,Jing 等[62]提出了一个集成关键词和视觉特征的框架,此框架基于少量已标注图像的视觉特征建立一系列统计模型来表示语义概念,并利用这些模型将关键词传播到未标注的图像. 当更多的图像通过用户的相关反馈获得隐含标注时,这些模型会自动周期性地更新,关键词的统计模型用于积累和记忆用户反馈的知识. 该框架采用 2 种相似度度量方法,结合相关反馈机制分别用于进行关键词查询和图例查询,能有效地利用相关反馈技术学习图像的语义概念. Zhou 等[63]将查询过程和相关反馈机制

进行无缝的连接,提出了一种拟分类算法学习关键词的相似度矩阵,便于关键词的语义分类、词典构造和软查询扩展,从而将图像视觉内容与相应的关键词结合起来,以获得图像的语义信息. He 等[64]提出一个集成长期学习和短期学习的图像检索系统. 长期学习用于隐含地创建一个语义空间,而短期学习用于在该语义空间中学习目标函数并创建分类器. 其基本思想是:假设经过几轮反馈后,与同一个查询关联的图像属于同一个语义类,通过聚集这些结果可以逐步地构建一个语义空间;然后该系统使用单值分解的方法对语义空间降维,在降维后的语义空间中,正例和反例图像不再是线性分离的,此时系统使用 SVM 算法来学习目标函数以检索相关的图像.

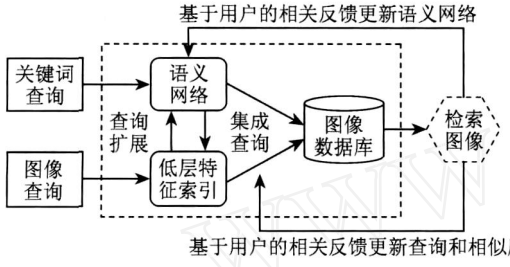

图 3 基于相关反馈的集成语义和视觉特征的统一框架

许多研究者利用相关反馈机制与各种统计学习方法相结合以获取图像语义和提高检索精度. 例如,Wu 等[65]将图像检索看作一个传递学习的问题,使用 Discriminant-EM 算法构造传递学习框架,从有限的标注图像数据中提升分类器的训练. 施智平等[66]提出一种基于贝叶斯分类器的集成视觉特征和语义信息的相关反馈检索方法,将图像库的数据经语义监督的视觉特征聚类算法划分为小的聚类,再以聚类为单位标注正负反馈的实例,最后利用贝叶斯分类器修正相似距离,从而达到较高的检索准确率. Yang 等[67]将图像标注作为在 MIL 框架下的监督学习,提出一种基于不对称 SVM 的多例学习(asymmetrical SVM-based MIL, ASVM-MIL)算法,通过引入错误正例和错误反例的不对称损失函数进行 MIL 设置. 在 MIL 的约束下最大化模式间隔,ASVM-MIL 将 MIL 问题转化为传统的监督学习问题,从而能充分地利用 SVM 强大的学习能力. Djordjevic 等[68]引入新的融合低层特征的策略,将低层特征融合于一个结构化的多特征空间中,并采用相关反馈机制和对象模型协同工作的方法来连接视觉对象和语义信息. 该方法使用 SVM 学习用户的相关反馈信息,同时定义了一个适应回旋内核来处理融合低层特征的多特征空间,能够得到精确度较高的检索效果.

利用相关反馈进行学习的方法在视觉特征层次和语义层次都能获得较好的检索效果,这类方法注重学习用户反馈的知识并建立语义空间模型,进而创建低层特征空间到高层语义空间的映射;然后将待查图像转变为语义空间中的表示形式,从而实现图像在语义空间中的检索. 但是,将当前查询会话的反馈结果与已有的语义信息相结合的过程并不总是清晰的,使用相关反馈训练一个语义检索系统有几个局限性:1) 耗时太长. 因为学习每一个语义概念的统计模型都需要用户指定相当多的图像范例;2) 在一个相关反馈的会话中,要使所有查询都与同一个语义概念相关通常是不现实的. 例如某用户要查询“北京的建筑”,就可能会在查询“北京”的图像和“建筑”的图像之间震荡. 尽管这样,相关反馈的方法具有与人类学习方法类似的思路,是一种很有价值的研究语义映射的方法.

3.4 特殊领域的语义映射方法

客观世界的语义多种多样,第 3.1～3.3 节讨论的都是面向具有普遍意义的图像提出的方法,而不同的领域会采用不同的映射方法,也会需要不同的背景知识. 在艺术美学领域,Colombo 等[69]通过研究如何获取艺术图像的情感语义,建立了一系列映射规则,用以判断艺术图像传达的情感. 首先通过颜色聚类将图像按颜色分割为不同的区域,用一个二值数组表示每个区域中是否存在 8 种基本颜色,并以此作为图像的颜色特征;然后根据 Itten 球模型,得出描述图像颜色的冷暖、和谐度、对比度等对人的感觉的语义描述公式;最后根据这些描述,得到图像的基本情感描述:愉快、紧张、放松、动感等. 王上飞等[70]选用 150 个较大的小波系数作为图像特征来进行图像的情感分类,采用 SVM 算法建立图像内容与其表达的情感语义之间的联系,并自动对未曾评价过的图像进行注释.

Sun 等[71]研究遥感领域的语义描述模型,使用本体和网格技术进行语义检索. Li 等[72]使用一个环境敏感的贝叶斯网络推断遥感图像的语义概念. 在医学领域,Tang 等[73]采用 I-Browse 系统对医学图像进行语义的识别. Ogiela 等[74]使用基于感知的方法对医学图像进行智能语义分析,能够分析图像的重要诊断特征,该方法基于一种特殊的图像描述语言和语法形式. Barb 等[75]使用知识仓库和交流框架管理诊断图像数据库,该框架使用语义的方法描述视觉异常,并提供医学领域的知识提取和交流. 总之,由于不同领域面对的图像各有其独特的特点,针对这些特点采取的语义映射方法也各有其独特性.

4　存在的问题和进一步的研究方向

经过研究者的努力,在图像语义相关的领域已经取得了很大的进展,然而,要提高图像的自动语义标注精度,使图像的语义检索能够实际应用仍有许多问题需要解决.

4.1　图像的低层处理

为了获取图像的语义,需要提取图像的低层特征并进行训练.目前大多数语义映射方法都基于一个前提——具有类似语义信息的图像或图像区域应该具有相似的全局或局部低层特征.而从原始图像的存储结构中可以获得的数据实际上只是一个像素点阵,如何从原始数据中提取有效的低层特征,对图像的语义映射有重要意义.研究者对图像的颜色、纹理、形状等低层特征已经进行了大量研究,但仍存在维数高和提取算法不完善等问题,在提取低层特征的同时也丢失了一些重要的原始数据,要利用这些特征进一步获取图像语义,存在很大困难.图像语义映射需要更多可选的低层特征和更高效的提取算法,如何筛选现有的低层特征,研究更有效的低层视觉特征,克服光照、旋转、遮挡等因素的影响,实现多种图像特征的融合,是一个重要的课题.

此外,图像分割算法的不成熟,大大地影响了高层语义的获取.要获取图像的对象语义需要处理图像的局部特征,一般采用图像分割的算法将原始图像分割为多个区域,分别进行处理.而已有的图像分割算法并不能将所有图像都分割为有意义的区域,存在过度分割等问题.也就是说,图像分割的结果并不一定可靠,这给语义映射的研究带来了很大的困难.寻找快速、鲁棒的图像分割算法,是非常困难但也是非常有价值的研究课题.

4.2　语义映射机制

由于低层特征提取不够完善,而现实世界的语义又是复杂多样的,因此图像的语义映射领域就需要承担很大的工作量.一方面要想办法弥补低层特征提取不完善带来的问题,进行合理的特征选择;另一方面要合理地组织获取的语义信息,以满足用户的检索需求,这都需要一个强大的语义映射机制来实现.目前大多数语义映射方法都需要利用一些现有的先验知识或者用户提供的知识来对机器进行训练,然后采用机器学习的方法来识别检索图像包含的语义概念,虽然每种方法都有各自的优势,但也存在一些局限性.理想的语义映射方法应该是在尽量减少人机交互的前提下,能够让系统具有必要的知识积累和学习能力,从而对新图像产生越来越精确的标注.为实现这个目标,需要研究者在各自的研究领域提出新的思路和方法,并结合有效的学习算法来实现.总之,研究合理的语义映射机制,提高图像语义标注精度,是一个有前途的研究方向.

4.3　语义空间建模

图像的语义空间模型需要满足 3 个要求:1) 能较好地表示客观世界复杂的语义;2) 能支持语义的相似度度量;3) 能有效地存储从图像视觉特征映射得到的语义信息.目前表示图像语义的方法虽然很多,但都不能完全满足以上要求.由于语义之间关系复杂,且语义具有模糊性和多样性等特点,因此有效地表示语义是非常困难的.虽然现有的语义表示方法在某些方面证明是有效的,但仍缺乏一种通用的表示方法,因此建立一个通用的能够广泛认可的语义空间模型极具挑战性.

另外,一个好的语义相似度度量公制[76]也是一个需要研究的课题.当前的相似度度量和聚类、分类研究将注意力集中在一些具有良好数学性质的公制距离度量函数上,例如欧氏距离.然而认知科学认为,人类视觉相似度判断应该是非公制的,也就是说,我们应该尝试一些不满足三角不等式的图像相似度度量算法,例如基于核函数的非线性距离度量、流形结构等.

4.4　性能评估标准

一项技术的发展离不开它的评估标准,一个好的评估标准可以客观地评价提出的新方法,从而引导其走向正确的方向[77].检索性能评估的 2 个通用的准则是效率和效果.目前,图像检索的评估标准主要采用信息检索领域的 2 个评估措施:查准率和查全率.查准率是指在检索所得的图像中与当前查询相关的图像所占的百分率;查全率是指在图像数据库中所有与当前查询相关的图像中,被检索出的图像所占的百分率.虽然这些评估标准能在一定程度上测试系统的性能,但还远远不能令人满意.另外,在图像检索领域还没有提供标准的图像测试集.虽然很多研究者采用 Corel 图像库进行测试,但由于系统需求和查询方式的不同,使用 Corel 图像所得到的结果并不一定能为判断系统性能提供良好的依据[78].没有一个定义清晰的测试和评估标准,要客观地比较和评价不同的系统是非常困难的.

导致难于定义良好的评估标准的一个主要原因在于图像内容感知的主观性,也就是说,对图像内容感知的主观性使得我们不能定义一个客观的评估

标准.如何寻找一个好的方法定义良好的性能评估标准,以指引研究者努力的方向,这本身就是一个值得研究的课题.

5 结束语

为了解决在 CBIR 研究中遇到的"语义鸿沟"的问题,研究者对图像语义映射的研究越来越活跃,也提出了许多语义映射的方法.本文对其中具有代表性的方法进行讨论,并分析它们各自的优势和缺陷,希望对今后的研究有一定的借鉴作用.

过去几年对图像语义映射的研究大都是在 CBIR 的基础上利用机器学习的策略来获取语义,但仍未出现一种令人满意的高效的通用的方法.图像语义映射已成为图像检索领域关注的一个热点,未来的发展需要研究者借鉴计算机视觉、智能科学、信息检索、人机交互等领域的成果及其综合运用,不断引入新的有效的机器学习算法.

参 考 文 献

[1] Rui Y, Huang T S, Chang S F. Image retrieval: current techniques, promising directions, and open issues [J]. Journal of Visual Communication and Image Representation, 1999, 10(1): 39-62

[2] Gudivada V N, Raghavan V V. Content-based image retrieval systems [J]. Computer, 1995, 28(9): 18-22

[3] Smeulders A W M, Worring M, Santini S, *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

[4] Datta R, Li J, Wang J Z. Content-based image retrieval — approaches and trends of the new age [C] //Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, Singapore, 2005: 253-262

[5] Lew M S, Sebe N, Djeraba C, *et al*. Content-based multimedia information retrieval: state of the art and challenges [J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2006, 2(1): 1-19

[6] Flickner M, Sawhney H, Niblack W, *et al*. Query by image and video content: the QBIC system [J]. Computer, 1995, 28(9): 23-32

[7] Gupta A, Jain R. Visual information retrieval [J]. Communications of the ACM, 1997, 40(5): 71-79

[8] Ma W Y, Manjunath B S. NeTra: a toolbox for navigating large image databases [J]. Multimedia Systems, 1999, 7(3): 184-198

[9] Wang J Z, Li J, Wiederhold G. SIMPLIcity: semantics-sensitive integrated matching for picture libraries [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(9): 947-963

[10] Carson C, Belongie S, Greenspan H, *et al*. Blobworld: image segmentation using expectation-maximization and its application to image querying [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(8): 1026-1038

[11] Zhao W, Chellappa R, Phillips P J, *et al*. Face recognition: a literature survey [J]. ACM Computing Surveys, 2003, 35(4): 399-458

[12] Jain A K, Vailaya A. Shape-based retrieval: a case study with trademark image databases [J]. Pattern Recognition, 1998, 31(9): 1369-1390

[13] Vasconcelos N. From pixels to semantic spaces: advances in content-based image retrieval [J]. Computer, 2007, 40(7): 20-26

[14] Liu Y, Zhang D S, Lu G J, *et al*. A survey of content-based image retrieval with high-level semantics [J]. Pattern Recognition, 2007, 40(1): 262-282

[15] Zhu M, Badii A. Semantic-associative visual content labelling and retrieval: a multimodal approach [J]. Signal Processing: Image Communication, 2007, 22(6): 569-582

[16] Wang Huifeng, Sun Zhengxing, Wang Jian. Semantic image retrieval: review and research [J]. Journal of Computer Research and Development, 2002, 39(5): 513-523 (in Chinese)
(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)

[17] Eakins J, Graham M. Content-based image retrieval [R]. Newcastle: University of Northumbria, 1999

[18] Jaimes A, Chang S F. Model-based classification of visual information for content-based retrieval [C] //Proceedings of SPIE, San Jose, 1999, 3656: 402-414

[19] Gao Yongying, Zhang Yujin. Progressive image content understanding based on multi-level image description model [J]. Acta Electronica Sinica, 2001, 29(10): 1376-1380 (in Chinese)
(高永英, 章毓晋. 基于多级描述模型的渐进式图像内容理解[J]. 电子学报, 2001, 29(10): 1376-1380)

[20] Chang S F, Sikora T, Puri A. Overview of the MPEG-7 standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(6): 688-695

[21] Miller G A. WordNet: a lexical database for English [J]. Communications of the ACM, 1995, 38(11): 39-41

[22] Lu Y, Hu C H, Zhu X Q, *et al*. A unified framework of semantics and feature based relevance feedback in image retrieval systems [C] //Proceedings of the 8th ACM International Conference on Multimedia, Los Angeles, 2000: 31-37

[23] Mezaris V, Kompatsiaris I, Strintzis M G. Region-based image retrieval using an object ontology and relevance feedback [J]. EURASIP Journal on Applied Signal Processing, 2004, 6(1): 886-901

[24] Town C, Sinclair D. Language-based querying of image collections on the basis of an extensible ontology [J]. Image and Vision Computing, 2004, 22(3): 251-267

[25] Li Q Y, Hu H, Shi Z Z. Semantic feature extraction using genetic programming in image retrieval [C] //Proceedings of the 17th IEEE International Conference on Pattern Recognition, Cambridge, 2004: 648-651

[26] Luo J B, Savakis A. Indoor vs outdoor classification of consumer photographs using low-level and semantic features [C] //Proceedings of IEEE International Conference on Image Processing, Thessaloniki, 2001: 745-748

[27] Vailaya A, Figueiredo M A T, Jain A K, *et al*. Image classification for content-based indexing [J]. IEEE Transactions on Image Processing, 2001, 10(1): 117-130

[28] Carneiro G, Chan A B, Moreno P J, *et al*. Supervised learning of semantic classes for image annotation and retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3): 394-410

[29] Vasconcelos N. Minimum probability of error image retrieval [J]. IEEE Transactions on Signal Processing, 2004, 52(8): 2322-2336

[30] Rasiwasia N, Moreno P J, Vasconcelos N. Bridging the gap: query by semantic example [J]. IEEE Transactions on Multimedia, 2007, 9(5): 923-938

[31] Goh K S, Chang E, Cheng K T. SVM binary classifier ensembles for image classification [C] //Proceedings of the 10th International Conference on Information and Knowledge Management, Atlanta, 2001: 395-402

[32] Cusano C, Ciocca G, Schettini R. Image annotation using SVM [C] //Proceedings of SPIE, San Jose, 2004, 5304: 330-338

[33] Gao Y L, Fan J P, Xue X Y, *et al*. Automatic image annotation by incorporating feature hierarchy and boosting to scale up SVM classifiers [C] //Proceedings of the 14th ACM International Conference on Multimedia, Santa Barbara, 2006: 901-910

[34] Chang E, Goh K, Sychay G, *et al*. CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(1): 26-38

[35] Town C, Sinclair D. Content-based image retrieval using semantic visual categories [R]. Cambridge: AT&T Laboratories, 2001

[36] Li J, Wang J Z. Automatic linguistic indexing of pictures by a statistical modeling approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(9): 1075-1088

[37] Li J, Wang J Z. Real-time computerized annotation of pictures [C] //Proceedings of the 14th ACM International Conference on Multimedia, Santa Barbara, 2006: 911-920

[38] Stan D, Sethi I K. Mapping low-level image features to semantic concepts [C] //Proceedings of SPIE, San Jose, 2001, 4315: 172-179

[39] Bilenko M, Basu S, Mooney R J. Integrating constraints and metric learning in semi-supervised clustering [C] // Proceedings of the 21st International Conference on Machine Learning, Banff, 2004: 81-88

[40] Jin W J, Shi R, Chua T S. A semi-naïve Bayesian method incorporating clustering with pair-wise constraints for auto image annotation [C] //Proceedings of the 12th ACM International Conference on Multimedia, New York, 2004: 336-339

[41] Chen Y X, Wang J Z, Krovetz R. An unsupervised learning approach to content-based image retrieval [C] //Proceedings of the 7th IEEE International Symposium on Signal Processing and its Applications, Paris, 2003: 197-200

[42] Shi J B, Malik J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905

[43] Zheng X, Cai D, He X F, *et al*. Locality preserving clustering for image database [C] //Proceedings of the 12th ACM International Conference on Multimedia, New York, 2004: 885-891

[44] Mori Y, Takahashi H, Oka R. Image-to-word transformation based on dividing and vector quantizing images with words [OL]. [2007-12-19]. http://citeseer.ist.psu.edu/368129.html

[45] Duygulu P, Barnard K, de Freitas J F G, *et al*. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary [M] //Lecture Notes in Computer Science. Heidelberg: Springer, 2002, 2353: 97-112

[46] Barnard K, Duygulu P, Forsyth D, *et al*. Matching words and pictures [J]. Journal of Machine Learning Research, 2003, 3(2): 1107-1135

[47] Blei D M, Jordan M I. Modeling annotated data [C] // Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, 2003: 127-134

[48] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(1): 993-1022

[49] Monay F, Gatica-Perez D. Modeling semantic aspects for cross-media image indexing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(10): 1802-1817

[50] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models [C] //Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, 2003: 119-126

[51] Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures [C] //Proceedings of the 17th International Conference on Neural Information Processing Systems, Vancouver, 2003: 553-560

[52] Feng S L, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation [C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington D C, 2004: 1002-1009

[53] Jin R, Chai J Y, Si L. Effective automatic image annotation via a coherent language model and active learning [C] // Proceedings of the 12th ACM International Conference on Multimedia, New York, 2004: 892-899

[54] Zhou X S, Huang T S. Relevance feedback in image retrieval: a comprehensive review [J]. Multimedia Systems, 2003, 8(6): 536-544

[55] Tong S, Chang E. Support vector machine active learning for image retrieval [C] //Proceedings of the 9th ACM International Conference on Multimedia, Ottawa, 2001: 107-118

[56] Rui Y, Huang T S, Ortega M, *et al*. Relevance feedback: a power tool for interactive content-based image retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998, 8(5): 644-655

[57] Wu Hong, Lu Hanqing, Ma Songde. A survey of relevance feedback techniques in content-based image retrieval [J]. Chinese Journal of Computers, 2005, 28(12): 1969-1979 (in Chinese)
(吴 洪, 卢汉清, 马颂德. 基于内容图像检索中相关反馈技术的回顾[J]. 计算机学报, 2005, 28(12): 1969-1979)

[58] Minka T P, Picard R W. Interactive learning with a "society of models"[J]. Pattern Recognition, 1997, 30(4): 565-581

[59] Cox I J, Miller M L, Minka T P, *et al*. The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments [J]. IEEE Transactions on Image Processing, 2000, 9(1): 20-37

[60] Tieu K, Viola P. Boosting image retrieval [J]. International Journal of Computer Vision, 2004, 56(1/2): 17-36

[61] Zhang H J, Chen Z, Li M J, *et al*. Relevance feedback and learning in content-based image search [J]. World Wide Web: Internet and Web Information Systems, 2003, 6(2): 131-155

[62] Jing F, Li M J, Zhang H J, *et al*. A unified framework for image retrieval using keyword and visual features [J]. IEEE Transactions on Image Processing, 2005, 14(7): 979-989

[63] Zhou X S, Huang T S. Unifying keywords and visual contents in image retrieval [J]. IEEE Multimedia, 2002, 9(2): 23-33

[64] He X F, King O, Ma W Y, *et al*. Learning a semantic space from user's relevance feedback for image retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(1): 39-48

[65] Wu Y, Tian Q, Huang T S. Discriminant-EM algorithm with application to image retrieval [C] //Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, 2000, I: 222-227

[66] Shi Zhiping, Li Qingyong, Shi Jun, *et al*. Relevance feedback method by integration of visual features and semantics [J]. Journal of Computer-Aided Design & Computer Graphics, 2007, 19(9): 1138-1142 (in Chinese)
(施智平, 李清勇, 史 俊, 等. 集成视觉特征和语义信息的相关反馈方法[J]. 计算机辅助设计与图形学学报, 2007, 19(9): 1138-1142)

[67] Yang C B, Dong M, Hua J. Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning [C] //Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, 2006: 2057-2063

[68] Djordjevic D, Izquierdo E. An object- and user-driven system for semantic-based image annotation and retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(3): 313-323

[69] Colombo C, Del Bimbo A, Pala P. Semantics in visual information retrieval [J]. IEEE Multimedia, 1999, 6(3): 38-53

[70] Wang Shangfei, Chen Enhong, Wang Zuyuan, *et al*. Research of emotion semantic image annotation and retrieval algorithm using support vector machine [J]. Pattern Recognition and Artificial Intelligence, 2004, 17(1): 27-33 (in Chinese)
(王上飞, 陈恩红, 汪祖媛, 等. 基于支持向量机的图像情感语义注释和检索算法的研究[J]. 模式识别与人工智能, 2004, 17(1): 27-33)

[71] Sun H, Li S X, Li W J, *et al*. Semantic-based retrieval of remote sensing images in a grid environment [J]. IEEE Geoscience and Remote Sensing Letters, 2005, 2(4): 440-444

[72] Li Y K, Bretschneider T R. Semantic-sensitive satellite image retrieval [J]. IEEE Transactions on Geoscience and Remote Sensing, 2007, 45(4): 853-860

[73] Tang H L, Hanka R, Ip H H S. Histological image retrieval based on semantic content analysis [J]. IEEE Transactions on Information Technology in Biomedicine, 2003, 7(1): 26-36

[74] Ogiela M R, Tadeusiewicz R. Nonlinear processing and semantic content analysis in medical imaging —a cognitive approach [J]. IEEE Transactions on Instrumentation and Measurement, 2005, 54(6): 2149-2155

[75] Barb A S, Shyu C-R, Sethi Y P. Knowledge representation and sharing using visual semantic modeling for diagnostic medical image databases [J]. IEEE Transactions on Information Technology in Biomedicine, 2005, 9(4): 538-553

[76] Mojsilović A, Rogowitz B E. Semantic metric for image library exploration [J]. IEEE Transactions on Multimedia, 2004, 6(6): 828-838

[77] Shirahatti N V, Barnard K. Evaluating image retrieval [C] //Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, 2005: 955-961

[78] Müller H, Marchand-Maillet S, Pun T. The truth about Corel —evaluation in image retrieval [M] //Lecture Notes in Computer Science. Heidelberg: Springer, 2002, 2383: 38-49