

基于概念级语义的图像描述与识别^{*}

任越美 程显毅 李小燕 谢玉宇

(江苏大学计算机科学与通信工程学院 镇江 212013)

摘要 由于计算机自动提取的图像视觉特征与人所理解的图像内容存在巨大的差异,传统的低层的视觉特征(如颜色、纹理、形状等)CBIR(Content-Based Image Retrieval)系统的检索结果往往不尽如人意。近年来,根据概念级语义(如男孩、高兴、浪漫等)的CBIR引起了研究者的重视。本文对CBIR领域的大量文献进行了深入的分析,从工程角度综述了图像概念级语义的描述模型、概念级语义特征提取和概念级语义图像检索问题的研究进展,并阐述了作者的一些观点。

关键词 图像语义,图像描述与识别,语义鸿沟,相关反馈

Description and Recognition of Image Based on Concept Semantics

REN Yue-mei CHENG Xian-yi LI Xiao-yan XIE Yu-yu

(School of Computer Science and Telecommunications Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract As the huge difference between the visual features of image extracted automatically by computer and the image contents appreciated by people, the search results of traditional CBIR (Content-Based Image Retrieval) system with lower-level visual features (such as color, texture, shape) are not satisfactory. In recent years, CBIR based on concept semantics (such as boys, happy, romantic, etc.) attracts the attention of researchers. In this paper, based on profound research on large numbers of references in the domain of CBIR, a survey of the description model of concept semantics, semantic feature extraction, and the study progress of the CBIR based on concept semantic is made in details from the engineering perspective, and then some of the author's viewpoints are pointed out.

Keywords Image semantic, Description and recognition of image, Semantic gap, Relevance feedback

随着信息时代的飞速发展,基于内容的图像检索技术(Content-Based Image Retrieval, CBIR)的很多成果已转化为应用技术,而且日益增长的应用需求反过来推动其研究向更高的层次发展。CBIR的基本思路^[1]是:从图像中抽取低层的视觉特征(如颜色、纹理、形状等),然后基于这些特征将用户查询的图像与数据库中的图像进行相似程度衡量,图像的相似性判断建立在视觉特征的相似性上。已开发出许多商业产品和实验原型系统,如 QBIC, Photobook 和 Virage 等。

然而,由于计算机自动提取的图像视觉特征与人所理解的图像内容存在巨大的差异,传统的CBIR系统的检索结果往往不尽如人意。人们判断图像的相似性并非建立在图像的颜色、形状等相似性上,而是主要根据图像的含义来判断图像是否符合自己的需要,这时就要使用一些高层次概念(如男孩、高兴、浪漫等),这些高层次概念包含了人对图像所描述的对象、事件以及表达的情感等内容的理解,而这就是图像所具有的语义特征,即是图像的高层语义知识。所以,理想的图像查询模式即是基于图像概念级语义特征的检索,例如用户需要查找“房屋前的桃树”,这就要求系统能够自动识别“房屋”和“桃树”。然而,目前的计算机图像识别技术还不能自动地提取图像的语义特征,主要有来自两个方面的困难:一是提供具有一定概念级语义特征的语义结构描述;二是低层的图像视觉特征映射到高层语义特征之间的“语义鸿沟”^[2,3]问题还没有得到完全的解决,即要有将低层的图像视觉特征映射到

高层语义的提取方法。为此,本文从这些方面对图像语义的模型描述、图像语义的提取方法以及图像语义在图像检索中的应用等研究状况进行归纳和总结,并由此提出基于语义的图像检索的进一步研究方向。

1 图像语义模型

1.1 图像语义模型定义

语义模型是图像语义的直观形象的描述形式。图像语义模型概念^[4]:语义模型是一个对图像特征提取的一个高度抽象的描述,它由图像固有信息和意念信息两部分组成。固有信息包含了图像本身的颜色、纹理和形状等,以及图像中的对象、对象与对象的相对关系;意念信息由场景语义、行为语义和情感语义组成。语义模型具有层次提炼性、网络继承性和自学习适应性等特点;它具有自身的知识库和形式描述语言,是一个由图像信息、计算机和知识概念组成的系统。

1.2 图像语义模型结构描述

语义模型的结构是对语义模型的抽象表达,以直观、形象的方式来描述语义模型的结构。语义模型的研究遵循着由低级向高级发展,由粗糙到精细的过程,由物质特性升华到意识形态,这是语义模型发展的过程。

1.2.1 层次结构

图像语义层次分为三层^[5,6],即底层特征层、对象层和语义概念层。底层特征层主要涉及到基于内容的颜色、形状、纹

^{*}国家自然科学基金(编号:60473039)、江苏省重点科技攻关项目(BE2004093)。任越美 硕士研究生,主要研究方向为模式识别;程显毅 教授,博士生导师,主要研究方向为模式识别、多agent理论与应用。

理等,而对象层主要考虑图像中的对象、对象的空间关系问题,而语义概念层主要解决图像通过场景、行为和情感所表达的意义。

由于图像语义的含义从图像底层的固有性质开始,到对象的空间关系和人为的意识语义,这个纵向的发展过程是图像语义中最主要、最基本的内容。要建立好语义模型,层次结构是基础,最能反映出层次结构特点与图像语义的是 MARKOV 树^[7]。

定义 1 如果图像的语义 S 是可描述的,则 S 是一个有近代的、逆向的、可能包含有其它分支结点的树(图),其树根为 S。

定义 2 假若树中某结点包含有非本分支结点,则将该结点及其子树复制剥离出来,使得该树中任意结点都不包含其它分支中的结点。

定义 3 Θ 运算是一个通过与外界有智能接口、内有形式语言及规则,利用知识库,结合人为反馈信息的运算,它的作用是将信息进行抽象、概括和决策后形成新的信息。

根据定义,可以构造出图像语义的基本模型,如图 1 所示。

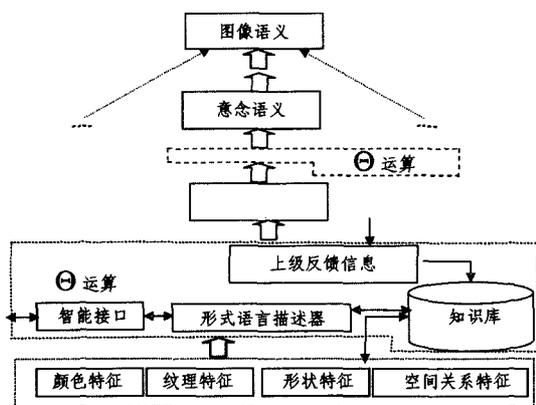


图 1 图像语义模型的层次体系结构

1.2.2 网络结构

图像语义的获取,除了其自有固有信息及对固有信息的提升运算外,还存在某一属性的信息与其它属性的信息相互交叉作用的情况,这样使得语义更具复杂性。在考虑这个问题时,为了能使问题简单化,把结点考虑成一个具有固有信息、反馈信息的 Agent^[8],每个结点都是构成语义描述模型的一个单元。

定义 4 语义网络中的每个单元结点是一个包含有固有属性、 Θ 运算、图运算^[9]的 Agent。

定义 5 语义网络是一种狭隘的网络,它在某些结点上只有单向图运算操作。

语义的网络模型描述并非广义网,对于某一属性的一个结点,在处理与之相关作用的属性结点时,只对所选用的、满足一定相关度的结点进行信息提取相关运算,把语义树抽象成一棵包含闭运算结点 P 的树。根据定义和其自身的特点,语义模型的网络结构描述如图 2 所示。

1.2.3 意念结构

图像语义的意念模型分两个层次。首先,根据图像语义模型的网络结构,对图像进行分类^[10]操作,获取图像最基本的意识信息,然后,根据运算规则,结合人为的反馈信息,来修正或完善图像的语义意念。

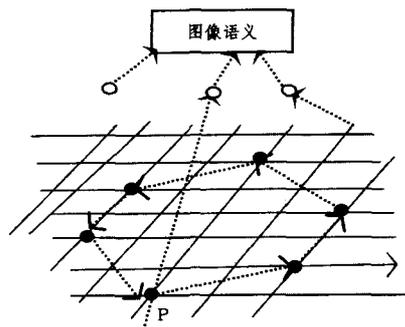


图 2 图像语义模型网络结点图运算示意图

图像语义的基本意念提炼于图像本身,结合语义模型的网络结构,可以构造图像语义,提炼基本意念的示意图,如图 3。

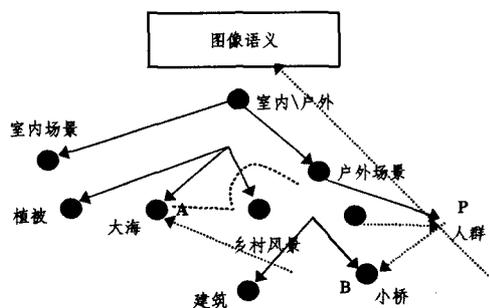


图 3 图像语义意念提炼基本图

根据图 3,对于任意的一个图像 I,首先对它进行分类操作,分类过程中的每一个结点,都是语义意念的一个基本描述。在语义意念树中(图中黑色部分),距离根结点越近,其意念越抽象;距离叶结点越近,其语义意念越具体、越明确。相对于操作结点 P 来说,它与结点 B, A 进行图运算,它是语义网络中的一点。

进行了图运算的结点 P 是一个具有 Θ 运算的 Agent 结点。通过 Θ 运算,结点 P 能够向上一层次提供本次运算后的结果信息,即在整体的图像语义运算中的重要一步,为上一层次的语义运算准备了数据。

2 图像概念级语义特征提取方法

近年来,研究人员已经提出了一些方法来解决语义特征提取时的“语义鸿沟”问题。这些方法大多是一些图像分割的方法,提取图像区域,然后获取其内容以及图像中与之相关的部分。James Z. Wang 等^[11]提出了一个基于图像分割的整体区域匹配相似性的图像检索系统。系统对图像按语义进行分类,如有纹理/无纹理图像。为了检索目的,他们还提出了一系列的基于统计学的图像分类方法。Duygulu 等和 Hoffman 用一种概率方法提取潜在的图像字符集^[12,13]。Zhu 等人^[14]用获得块特征的 codebook 将图像分割成相等大小的块和索引区域。Wang 等^[15]也提出了一种基于 codebook 的方法,借助 codebook 进而对基于区域颜色和纹理特征的统计学特征对图像进行分割。这种方法也试图考虑邻近区域的特性。Pan 等^[16]用一种简单的图像分割技术对图像区域进行分割并提取其特征,每个区域被标为一个 blob-token,并试图找到这些 blob-tokens 之间的联系以及索引图像的相关说明。文献^[17]中, Li 和 Wang 采用了一种统计模型方法,将一个给定

概念的图像看作是描述这个概念的随机过程实例。它们的方法利用 2D 隐马尔柯夫模型计算图像和概念的文本描述之前的相关测度。最后, Mojsilovic 和 Rogowitz^[18] 试图将图像低层特征直接关联到高层语义特征。然而这些方法仍然不能很好地与人类视觉系统进行匹配。为此, Depalov 等^[19] 提出了一种基于感知的语义图像检索方法, 作者用一种新的用于图像分割和分类的语义标注方法, 开发了一种具有一定适应度的、感知的颜色-纹理分割算法。这个算法将人们的感知知识同对特征信号的理解结合起来, 用于把自然图像分割成感知的、有语义的相同区域^[20]; 把得到的宽度特征(颜色、空间特征)作为中间层特征描述, 从而有效地缓解了“语义鸿沟”问题。但是这种描述只在分割是基于感知或语义是相关时才是有意义的, 其检索效果也主要依赖于文献^[20]中的图像分割算法。

Iqbal 等人利用 Bayes 方法设计了图像检索系统 CIRES, 检索“植物”等静态事物具有较好的效果, 但对其它对象检索效果很差^[21]。美国 Illinois Urbana Champaign 大学的 Beckman 研究所提出了一种基于空间变换模型和相关反馈的语义提取算法, 这种算法的空间变换模型比较复杂, 优化过程比较困难^[22]。HP 试验室的 C. Lee 等人提出了聚类分析的方法来进行图像语义的提取^[23]。Sheikholeslami^[24] 提出了基于神经网络模型的 SemQuery 算法, 实现从用户提交的视觉查询到高层语义的映射, 这种算法过多地依赖于领域专家。这些语义提取算法存在语义提取效率低、检索精度不高等缺点。针对上述问题, 文献^[25]利用 Bayes 统计学习和决策理论, 对图像的语义分类词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。王等^[26]提出了一种基于学习的聚类图像语义检索算法, 建立了底层特征和高层语义的映射, 实现了语义的快速提取和图像的精确检索, 它由聚类构造算法、聚类学习和图像检索算法组成。这两种方法都综合采用了颜色距、基于小波的纹理、边缘方向直方图和 HSV 颜色空间中量化到 256 维的颜色直方图等图像底层特征, 具有很好的查全率和查准率。不过, 这两种算法对语义分类层次结构须预先进行设定, 是基于监督学习的, 还需要进一步研究无监督、全自动的语义提取方法。

概括起来, 目前图像语义的提取方法大致可以归纳成三类^[27], 分别是基于知识的语义提取、人工交互语义提取、利用外部信息源的语义生成。它们构成了目前语义提取方法的三条主线, 代表了提取语义的三个方向。

2.1 利用系统知识的语义提取

基于知识的语义提取, 主要特征是需要预先给系统提供必要的领域知识, 如对象模板、图像场景分类器等。依据提取的语义内容和采取的方法, 又可以分为基于区域的方法和全局处理方法。

2.1.1 基于区域的语义特征提取

目前的基于区域的图像检索过程一般是在图像分块或子图像上进行图像分割, 计算图像各个区域之间的相似度, 而后根据一定的区域匹配准则, 计算图像各个区域之间的匹配度, 最终得到图像之间的匹配度。因此, 为了获得较好的图像语义, 这种检索方法重点在于图像分割和区域匹配准则的定义。

现有的基于区域的检索系统主要有: Blobworld^[28], 贝尔实验室的 WALRUS^[29], 意大利 Bologna 大学的 Windsurf^[30] 等, 其中 Blobworld 是通过用户指定查询图像中的一个或几

个区域和相应的特征权重, 进行单个区域匹配, 搜索含有与指定区域相似的区域图像。而 WALRUS, Windsurf 无须用户指定区域, 查询图的所有区域都参加图像相似性度量, 综合两图的所有相似区域对的相似性, 决定两图的相似性。

这些系统还不能完全满足用户的需要, 主要原因在于许多关键技术还不成熟: 复杂内容的图像中的图像分割得到的区域经常不能形成有意义的物体; 区域匹配不能完全符合用户对相似性的理解。人们在观察一幅图像时, 主要关注的是其中的前景对象和主要背景区域, 这构成了图像的主要观察区。图像的语义更多地体现在主要区域的组成上, 而不在于每个区域的细节上。而且图像前景和背景是两个层次的语义, 其在不同的应用中起着不同的作用。据此, 有学者^[31]提出了一种无监督、全自动并具有一定自适应能力的图像主要区域提取方法, 对图像先进行基于子块聚类的图像分割和分割结果优化, 再通过分析不同区域在整个图像中的作用来鉴别图像整体区域分布合理性(即进行区域重要程度分析), 以便据其去除与图像内容关系不大的信息。这样, 既保证了处理速度, 也便于进一步的处理。工作流程如图 4。

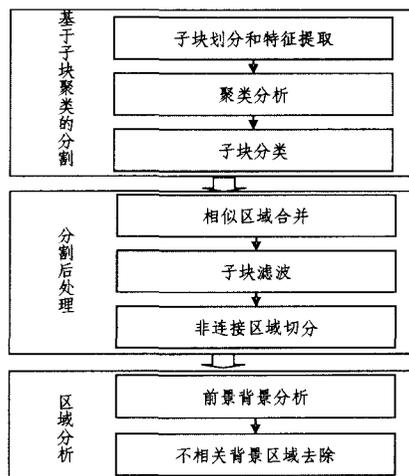


图 4 图像主要区域提取工作流程图

徐等^[32]提出了基于图像主观察区的语义提取方法, 并介绍了一个基于语义的图像检索系统——VisEngine^[27], 该系统采用基于图像主要区域的图像分割方法, 分别提取图像前景、背景和全局的视觉和抽象语义内容, 构造相应的语义模板。接着把这些特征导入到一个面向对象的中间信息结构中, 在此基础上进行多种方式的相似性匹配和检索。而主观察区的获得必须基于图像区域分割, 将图像分割为前景和背景两部分, 然后分别对其建立语义模板, 进而获取语义。

目前对于图像中有意义区域的提取方法还有很多, 主要是基于各种分割算法^[33]来实现的。这些分割算法均没有考虑到对图像语义区域边缘的精确分割, 而且存在着不同的缺点。例如: 主动轮廓模型方法(ACM)^[34]需要良好的初始分割; 基于边缘检测的分割法^[35]只使用了局部信息, 难以保证分割区域内部的颜色一致, 且不能产生连续的闭区域轮廓; 区域生长与合并技术, 该方法常得到不规则的边界和小洞, 而且区域个数严重依赖于初始种子数, 计算量也较大; 基于马尔柯夫随机场(MRF)的聚类^[36]计算量巨大。

同时, 这些分割算法一般试图分割出数个不同的像素点群, 每个点群的特性以一个恒定的颜色值来标记^[37]。而人对图像的理解是基于图像中主要区域的选择来进行的, 并且在

有些情况下,算法最终运行后所得的点群数目需事先说明,对于某些分割算法来说难以实现。

为了克服这些问题,文献[38]基于 RGB 空间提出了一种彩色图像有意义区域提取算法,算法注重人对图像的主观感受,降低了提取结果的冗余信息。通过引入分割精确度控制参数 k ,提高了对语义区域提取的准确性,同时具有较小的运算量和较小的鲁棒性。算法描述如下:

(1)计算图像复杂度分割阈值。

(2)判断像素的像素色度值是否大(小)于阈值。是,则转 3;否,则转 4。

(3)根据定理 1,判断该像素与其周围像素的欧氏距离是否大于 K 。是,则转 4;否,则转 5。

(4)把该像素点置成白色,指针+1,转 2。

(5)该像素点属于图像语义区域,保留该像素。指针+1,转 2。

2.1.2 基于全局特征的语义特征提取

基于全局特征的语义提取主要是利用低频信息(全局图像特征)进行粗的场景分类。这方面的工作是建立在心理学研究基础之上的:在一定的情况下,人类的场景识别可以在不知道任何对象信息的情况下进行。Biederman 等^[39]也提出:独立于局部对象的识别,根据用体积代表的所有显著对象区域空间排列,可以快速地场景识别。文献[40]利用一系列 2 类 Bayes 分类器的组合,通过全局的底层视觉信息来获取简单的环境分类信息,分类器的层次组合对于简单的环境分类具有较高的分辨率。

Yang ZiJun 等人^[41]提出的图像检索系统包含语义分类,此系统使用图像内容分析和图像聚类技术组织图像数据库。在其框架中,分类过程包括粗分类和精细分类。粗分类基于颜色和边缘的图像内容分析过程实现,而精细分类是利用多特征模板的图像聚类法解决的。粗分类过程中,用到了很多的假设和对图像理解的先验知识,如“背景是蓝色的可能是天空或海洋”、“大部分风景图像有较少的几条长且清晰的边缘,且常是水平的,而建筑和组织图像一般会有很多的短边缘”等。通常情况下,这种假设是成立的,可是也有很多图像不满足这些条件。所以,如果用在大型图像数据库中,可能会产生很多错误。在精细分类中,首先选择一些典型的语义相似的图像,计算每个图像的低层特征向量,因为语义相似的图像可能有很不同的低层特征向量,其用平均特征向量代表每个聚类,这些平均向量建立了一组特征模板。在其系统中,同时应用特征向量和特征模板进行语义查询。可是低层特征的选取是否合适及对类的典型图像的主观判断都会影响分类的性能。

区域方法和全局方法的区别是所抽取特征的作用粒度不同,前者的粒度是区域,后者的粒度是整幅图像。区域层的图像语义获取是建立在基于低层特征的图像分割基础之上的,通常抽取的是局部和全局特征。全局特征侧重于整体性和综合性,而局部特征强调区域性和细节性,两者的无缝结合可以提高图像分割的准确度,从而可以反映较真实的图像语义。

2.2 根据用户交互的语义提取

目前通用的完全自动的图像语义处理还存在一些难以逾越的障碍。要在计算机视觉和人工智能现有发展水平上进行概念级的语义处理,必须充分考虑到人的作用,把人当作系统的一个组成部分。人工交互的语义提取,主要体现在图像库预处理和反馈学习两个方面。预处理包括简单的人工标注和

较为合理的结合领域知识的方法等;用户的相关反馈可以起到两个作用,一是系统可以根据用户的反馈逐渐把握用户的需求,得到最终结果;二是使系统建立高层语义与低层视觉特征的对应关系,或修正图像的分类,或修正语义和视觉特征的映射关系。反馈技术的基本思想是^[42]:在检索过程中允许用户对检索结果进行评价和标记,指出结果中哪些是与查询图像相关的,哪些是不相关的,然后将用户标记的相关信息作为训练样本反馈给系统进行学习,以指导下一轮检索,从而使得检索结果更加符合用户的需要。

2.2.1 基于语义向量的相关反馈

基于低级语义特征的反馈主要是基于语义向量模型的,大多数方法可以归入查询点移动或者权值调整。把图像语义内容表示成向量的形式,并将其与图像的低层特征一同处理。这种方法所表达的语义内容是固定的,因为语义特征向量的维数是固定的。

所谓语义向量^[3]就是事先定义一个语义属性组 $A = \{A_1, A_2, A_3, \dots, A_n\}$,其中 A_i 表示一个语义内容,如“女孩”、“风景”、“太阳”等, n 表示语义属性的个数。再为数据库中的每个图像定义语义属性特征向量 $X = \{P_1, P_2, P_3, \dots, P_n\}$,其中 P_i 表示图像含有对应属性 A_i 的概率。例如, $X = \{0, 0, 0, \dots, 1\}$ 表示图像含有属性 A_n ,而不含其它属性。注意, $P_1, P_2, P_3, \dots, P_n$ 的取值可以是布尔型的^[43],也可以是取 $[0, 1]$ 之间的数,称之为关键字的软向量表示^[44-46]。其中 n 的取值取决于系统建立者定义的属性个数,在特定系统中 n 是固定的。例如在文[44]中, $n = 79$;在文[45]中, $n = 15$;在文[46]中, $n = 30$ 。经过这样的向量化处理,图像的语义特征就可以和低层特征一样表示成向量的形式,可以应用加权的欧氏距离作为相似性度量,也可以和低层特征使用相同的相关反馈算法改善检索性能。

2.2.2 基于关键字语义网络的相关反馈

MARS 系统^[47]仅仅利用底层特征的反馈,在通用的检索背景中无法有质的提高。为了克服以上单独根据低层特征的相关反馈方法的缺点,必须考虑实现高层语义概念基础上的相关反馈。Show&Tell 将语言 and 选择设备结合起来^[48],可以给系统提供更丰富的语义信息,它采用了语音识别技术和自然语言理解的方式,将用户的语言转化为文本可理解的概念。文献[49]从机器学习的角度,在每次反馈中对用户标记的“正例”和“反例”样本进行学习,用支持向量机构造出最优分类器,作为模型实现用户反馈的几何示意,并在特征空间中根据该模型对图像进行判别和检索,从而建立了语义与特征空间的联系。文献[50]中介绍的 iFind 系统则是一个应用了语义相关反馈的现实系统。它结合了早期根据关键词的图像检索,在系统中建立了一个关键词与图像相关联的语义网络,对图像库中每幅图像都以不同的关键词和权重加以描述,其中一幅图像可能有一个或多个关键词与之对应,且每幅图像对应的关键词及其权重可以根据用户反馈信息加以调整。这样,随着足够多的反馈和学习训练之后,系统就会建立起一个相对真实的语义空间到特征空间的映射关系。若能在 iFind 中结合基于机器学习的方法,进一步改善将语义信息和视觉特征相结合的相关反馈方法,必将大大提高图像检索系统的性能。

文献[51]中的方法也是基于语义网络的,可是采用的是语义概念而不是低层特征进行标定阶段的工作。在语义标定阶段,包括了三个模块:关键字提取、关键字扩展和关键字加

权。采用电子词汇系统(WorkNet)^[52,53]的信息库来扩充与图像相关的关键词,然后用从字符集关键字集中提取通用字的过滤器对关键字集的结果进行过滤,并建立相应的索引。在查询扩展中,字符集用来帮助用户从给定的查询中多而复杂的结果中找到相关的或准确的结果。

另外,Shneiderman 和 Kang^[54] 提出一个直接注释方法,只注释照片中的人名。通过这种方法,用户可以只在人名列表中选择人名并将其拖动到图像上为其注释。尽管此方法省掉了很多的键入工作,它仍是一个包含很多人工操作(拖动操作)的方法。即使在改进的用户界面中,人工标注还是需要很多的时间和体力劳动,这是用户不喜欢的,所以希望有自动标注技术。

Mori 等^[55] 用一种同时发生(co-occurrence)模型,查询在由规则网格(regular grid)产生的图像区域中同时出现的关键词,并进行自动标注。E. Celebi 和 A. Alpkocak^[56,57] 提出了结合文本和视觉聚类的图像检索和自动标注方法。这种方法的主要前提是被分到同一文本聚类中的图像能够用共同的视觉特征来描述,有效地提高了检索的效率。

基于语义向量的用户反馈方法中的向量对应的语义属性本身是由系统设计者定义的直观的关键字,所以能表达的语义内容比较丰富,可以从中选择最佳注释样本,提高检索速度。可是,又因为此语义内容的固定性,用户不能扩展所要表达的语义信息,使数据库的应用范围受到很大限制,不适合用于大型数据库,同时也存在粒度不均问题。基于关键字语义网络的相关反馈方法,事先不需要为图像注释,其关键字注释是在检索过程中添加和删除的,比较灵活,而且图像不必与每个关键字关联。这样,在增加关键字的同时,并没有增加计算量,不过也存在一些关键字的同义或歧义问题,有待进一步改进,以提高检索性能。

2.3 利用外部信息源的语义生成

基于外部信息的语义提取,是指根据图像来源处的其它信息,来获得与图像相关的相对高层的描述信息。由于多媒体描述方式的普及,目前图像出现在各种场合,如报刊杂志、论文报告、Web 新闻等。利用图像的文件名、其周围的文字信息、文章标题等来获得图像的语义信息,便是基于外部信息的语义提取方式。Sclaroff^[58] 使用了 Latent Semantic Indexing 的方法,从图像相关的文档中生成低维的文本向量,然后在“LSI 语义空间”中匹配。Yahoo, Google, Altavista 等著名的搜索引擎都有了自己的图像检索,其工作方式一般是首先通过关键字检索,或是语义分类浏览进行查找,Altavista 还可以在关键字检索结果的基础上寻找视觉相似的图像。

3 图像语义的应用

3.1 图像语义在医学图像检索中的应用

医学是重要的计算机应用领域,并且近年来医学领域的数字图像信息应用也迅速增长起来^[59]。由于医学图像实体的形状、纹理、颜色、空间关系很难用文字信息确切描述,这些特征也不代表图像真正的语义信息,传统的基于内容的检索往往不能令人满意,于是有人提出一种综合低级特征和高级语义特征的医学图像检索方法^[60],首先提取图像的灰度特征、矩特征和纹理特征,进一步采用遗传算法进行最优特征的选择。由于这些低层特征对图像的描述与人类对图像的描述存在较大差异,直接利用这些特征作为检索依据常常得不到满意的结果,因此需要进一步提取语义特征,将影像报告中医

生给出的关于图像的描述作为语义内容进行相似性检索。检索流程图如图 5 所示。

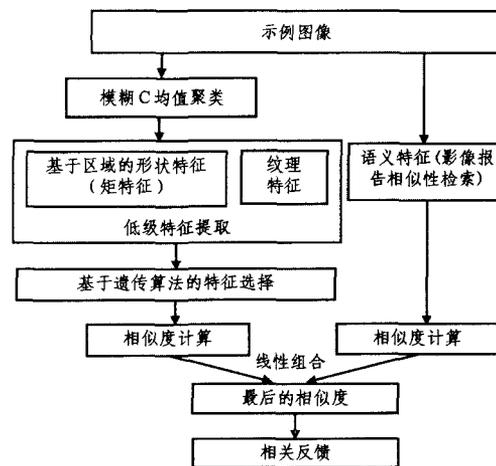


图 5 基于语义的医学图像检索流程

这种方法在一定程度上提高了图像检索的效果,不过检索系统的性能不太令人满意。文献[2]将基于支持向量机分类的语义提取方法应用到肺部医学 CT 图像数据库中,并通过引入人的交互以及相关反馈,即通过选择一些正确或错误的例子作为反馈,进一步提高了检索性能。

3.2 图像语义在 WWW 图像检索中的应用

随着 WWW 的迅速发展,网上资源日益丰富,查找和利用网络上的相关图像资源已成为一种需求,在网络环境下实现对图像快速而有效的检索变得十分必要和迫切^[61]。网络图像的最大特点是嵌入在具有文本内容的 HTML 文档中,并与网页有着紧密的联系。因此,可以根据图像名、图像注释、图像周围文本、图像 URL、图像所在网页 URL 以及标题、图像超链接网页 URL 及标题等相关文本^[62],提取图像的主题词、主体词及其属性词等语义信息,提高基于内容和语义的 WWW 图像检索的准确率。

在利用图像的语义信息对图像进行检索时,首先利用图像低层特征表示图像的内容,并且根据其内容进行分类(如自然图像或合成图像),然后将其存放在相应的位置。当用户进行检索时,可以很方便地进行查询。

图像的语义信息通过计算机自动获取,构成图像的语义特征,存储到图像索引数据库中,同时对这些语义根据字典进行再抽取,得到最能表现该图像的关键词,保存到数据库中。语义检索框图见图 6。

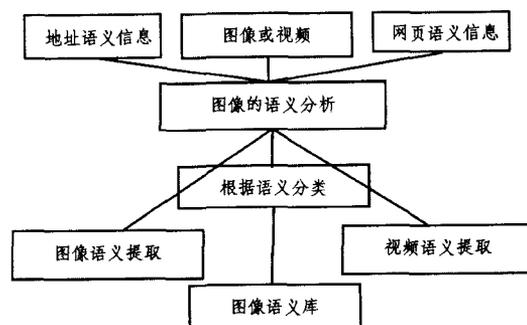


图 6 网络图像的语义检索框图

文献[63]还利用语义特征开发了一个方便实用的基于语义的自然图像检索系统。该系统首先对图像进行分割,将图

像分成不同的物体,然后人工参与,把物体分成不同的类别,利用设计好的分类器对不同类别进行训练,得到各个类别的模型。采用离线方式对数据库中的每一幅图像进行识别,检索效率比较可观。

利用语义的图像搜索是提高检索效率的重要方法。今后的图像搜索应是基于图像语义的描述与基于图像内容理解的图像检索技术及网络技术的综合,而关键的是依赖计算机自动识别图像及标引算法的不断发展,以及更加友好的用户界面的生成。

结束语 在短短的几十年时间里,CBIR 得到了长足的发展,而基于概念级语义的 CBIR 的研究也越发显得重要。它的研究和应用领域涉及了诸多方面,还要与其他领域相关技术密切结合,如计算机视觉、数字图像处理 and 模式识别技术,包括心理学、生物视觉模型等技术的新发展和综合运用。从目前的发展状况来看,基于图像概念级语义的检索还有一些重要的方面需要进一步研究,如综合图像的其它低层特征,如颜色、形状、空间分布等,研究完整的语义检索框架;语义 Web 环境下基于 Ontology 的语义检索也是研究和实现语义检索的一个重点和突破点;将语义特征用于商标检索中,对相关反馈技术及算法进行深入的研究,以提高相关反馈中语义特征的自动标注功能,使检索结果更符合用户的语义要求等。

本文主要对图像语义结构模型描述、语义特征的提取方法进行了总结和评述,并介绍了基于概念级语义特征的图像检索的一些重要应用,为以后进一步进行这方面的工作奠定了理论基础。

参 考 文 献

- [1] 孙志杰,许宏丽.一种图像底层视觉特征到高层语义的映射方法[J].计算机应用,2004,24(12):22-24
- [2] Zhao Rong,Grosky W I. Bridging the semantic gap in image retrieval. Distributed Multimedia Databases: Techniques & Applications,2002
- [3] 耿苑.结合低层特征和高层语义的图像检索系统[D].西北工业大学,2004
- [4] 周学海,李光亚,赵振西,等.基于图像对象语义模型的图像对象的创建[J].计算机研究与发展,1998,35(9):855-858
- [5] 陈剑斌,老松扬,等.基于内容的图像检索的发展最新趋势[J].计算机工程与应用,2002,38(10):47-49
- [6] 张宜.基于内容的图像检索技术研究综述[J].广西广播电视大学学报,2003(12):48-56
- [7] Shaffrey C W,Kingsbury N G. Unsupervised image segmentation via Markov trees and complex wavelets[C]//IEEE 2002 International Conference on Image Processing. Rochester, New York,2002
- [8] 石跃祥,等.基于内容的智能图像检索系统中反馈信息 Agent 的设计[J].计算机工程与应用,2002,38(20):10-20
- [9] Shaffrey C W,Jermyn I H,Kingsbury N G. Psychovisual evaluation of image segmentation algorithms[C]//Proceedings of ACIVS 2002 (Advanced Concepts for Intelligent Vision Systems). Ghent,Belgium,2002
- [10] Chang S F,Sundaram H. Structural and semantic analysis of video[C]. In: ICME 2000, New York,2002
- [11] Wang J Z,Li J,Wiederhold G. SIMPLiCity, Semantics-Sensitive Integrated Matching for Picture Libraries. IEEE Trans. ONPAMI,2001,23(9):57-60
- [12] Duygulu P,Barnard K,Freitas J F G, et al. Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary // European Conference on Computer Vision (ECCV2002). 2002
- [13] Hofmann T. Unsupervised Learning by Probabilistic Latent Semantic Analysis. Machine Learning,2001
- [14] Zhu L, et al. Keyblock: An approach for content-based image retrieval. ACM Multimedia,2000:157-166
- [15] Wang W,Song Y,Zhang A. Semantics retrieval by content and context of image regions // Proc. 15th Int. Conf. Vision Interface, May 2002
- [16] Pan J Y, Yang H J, Duygulu P, et al. Automatic image captioning. ICME,2004
- [17] Li J,Wang J. Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Tr. PAMI,2003,25
- [18] Mojsilović A,Rogowitz B E. Semantic metric for image library exploration. IEEE Tr. MM,2004,6:828-838
- [19] Depalov D,Pappas T,Li Dongge, et al. A Perceptual Approach of Semantic Image Retrieval. Acoustics, Speech and Signal Processing,2006
- [20] Chen J, et al. Adaptive perceptual color-texture image segmentation. IEEE Tr. Im. Proc. ,2005,14:1524-1536
- [21] Iqbal Q, Aggarwal J K. CIREs: A System for Content-based Retrieval in Digital Image Libraries [A]//Proc. of the 7th Int'l Conf. on Control, Automation, Robotics and Vision [C]. 2002: 205-210
- [22] Rui Y,Zhou S X,Huang T S. Efficient Access to Video Content in a Unified Framework[A]//Proc. of the IEEE Int'l Conf. on Multimedia Computing and Systems [C]. 1999:735-740
- [23] Lee C, Ma W Y, Zhang H J. Information Embedding Based on Users Relevance Feedback for Image Retrieval [R]. Technical Report. HPLabs, 1998
- [24] Sheikholeslami G, Chang W, Zhang Aidong. SemQuery Semantic Clustering and Querying on Heterogeneous Features for Visual Data[J]. IEEE Trans on Knowledge and Data Engineering, 2002,14(5):988-1002
- [25] 王崇骏,杨育彬,陈世福.基于高层语义的图像检索算法[J].软件学报,2004,15(10):1461-1468
- [26] 王天江,田刚.基于学习聚类的图像语义检索算法[J].计算机工程与科学,2007,29(3):49-51
- [27] 王惠锋.基于语义的图像检索系统及其关键技术究[D].南京大学硕士学位论文.2002:6-8
- [28] Carson C, Thomas M, Belongie S, et al. Blobworld: A System for Region-Based Image Indexing and Retrieval // Third Int. Conf. on Visual Information Systems. June 1999:509-516
- [29] Nstsev R R. WALRUS: A similarity retrieval algorithm for image databases. SIGMOD, Philadelphia, PA, 1999
- [30] Ardizzone S, Bartolini I, Patella M. Windsurf: Region-based Image Retrieval Using wavelets // IWOS. 99, Florence, Italy, Sept. 1999:167-173
- [31] 王惠锋,孙正兴.一种面向图像语义的主要区域提取方法[J].中国图象图形学报,2003,8(1):27-32
- [32] 徐晓刚,彭彬彬,孙正兴.一个基于语义的图像检索系统: VisEngine[J].计算机工程,2004,30(4):51-53
- [33] 王晓峰. CBIR 系统中的图像语义分割技术[J].计算机工程与应用,2002,21:114-117
- [34] Zhu S C, Yuille A. Region competition: unifying snakes region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans. Pattern Analysis Mach. Intel, 1996,18(9):884-900
- [35] Farag A A, Delp E J. Edge linking by sequential search. Pattern Recognition, 1995,28(5):611-633
- [36] Urago S, Zerubia J, Berthod M. A Markovian model for contour grouping. Pattern Recognition, 1995,28(5):683-693
- [37] 章毓晋.基于内容的视觉信息检索[M].科学出版社,2003:256-290
- [38] 王辅之,陈亚军.一种基于语义的图像有意义区域提取算法[J].计算机与数字工程,2006,34(3):4-6
- [39] Biederman I. On the Semantics of a Glance at a Scene. Perceptual Organization, Hillsdale, NJ: Erlbaum, 1981:213-253
- [40] Vailaya A, Figueiredo M A T, Jain A K, et al. Image Classification for Content-based Indexing. IEEE Transactions on Image Processing, 2001,10(1):117-130
- [41] Yang Zijun, Kuo C C J, et al. A semantic classification and composite indexing approach to robust image retrieval // Proc. of IEEE Int. Conf. Image Processing. vol1, Kobe, Japan, Oct. 1999:

- 134-138
- [42] 吴洪,卢汉青,马颂得. 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, 12: 53-57
- [43] Wu Hong, Li Mingjing, Zhang Hongjiang, et al. Improving Image Retrieval with Semantic Classification Using Relevance Feedback. VDB, 2002; 327-329
- [44] Zhang C, Chen T. An Active Learning Framework for Content Based Information Retrieval. Technical Report. CMU-AMP-01-04
- [45] Sychay G, Chang E. Effective image annotation via active learning // IEEE International Conference on Multimedia. 2002
- [46] Zhou Xiang Sean, Huang T S. Unifying Keywords and Contents for Image Retrieval // International Workshop on Content-based Multimedia Indexing. Italy, September 2001
- [47] Rui Y, Huang T S. A Novel Relevance Feedback Technique in Image Retrieval // Proceeding of the 7th ACM International Conference on Multimedia. ACM Press, 1999; 67-70
- [48] Srihari R K, Zhang Z F. Show & Tell: A Semi-automated Image Annotation System. IEEE Multimedia, 2000, 7(3): 61-71
- [49] 张磊, 林福宗, 张钺. 基于支持向量机的相关图像检索算法[J]. 清华大学学报(自然科学版), 2002, 42(1): 80-83
- [50] 朱兴全, 张宏江, 刘文印. 等. iFind: 一个结合语义和视觉特征的图像相关反馈检索系统[J]. 计算机学报, 2002, 25(7): 681-689
- [51] Kuo Chin-Hwa, Chou Tzu-Chuan, Tsao Nai-Lung, et al. Can-Find-a semantic image indexing and retrieval system. Circuits and Systems, 2003, 2: 644-647
- [52] Aslandogan Y A. Using Semantic Contents and WordNet in Image Retrieval // Proceeding of ACM SIGIR. 1999; 286-295
- [53] WordNet. <http://www.cogsci.princeton.edu/d>, 2002
- [54] Shneiderman B, Kang H. Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos // Proc. International Conference Information Visualization. London, England, 2000
- [55] Mori Y, Takahashi H, Oka R. Image-to-word transformation based on dividing and vector quantizing images with words // First International Workshop on Multimedia Intelligent Storage and Retrieval MAlanagement. 1999
- [56] Celebi E, Alpkocak A. Combining textual and visual clusters for semantic image retrieval and auto-annotation Intergration of Knowledge. Semantics and Digital Media Technology, 2006; 219-225
- [57] Celebi E, Alpkocak A. Semantic image retrieval and auto-annotation by converting keyword space to image space // Multimedia Modelling Conference Proceedings, 2006 12th International Jan.
- [58] Sclaroff S, Cascia M L, Taycher L. Unifying Textual and Visual Cues for Content-based Image Retrieval on the World Wide Web. Computer Vision and Image Understanding, 1999, 75(1): 86-98
- [59] 刘怡, 陶欣. 肺部医学 CT 图像数据库中语义对象的提取[J]. 河南师范大学学报(自然科学版), 2006, 34(3): 171-173
- [60] 邵虹, 崔文成, 张继武, 等. 低级特征和语义特征结合的医学图像检索方法[J]. 中国图象图形学报, 2004, 9(2): 220-224
- [61] 网络环境下的图像检索技术. http://diy.ccidnet.com/pub/article/c28_a78270_p1.html
- [62] 张量, 詹国华, 袁侦明. 基于 Web 的图像搜索[J]. 计算机, 2002, 5: 30-35
- [63] 刘晶. 基于语义的自然图像检索[D]. 西北工业大学, 2006; 23-35

(上接第 201 页)

像这样的结果和我们得出的结论相反, 其实不然。由于 ST 华光的价格非常低, 造成其投机性倾向的两个原因凸现出来。一是尽管企业的业绩不好, 只要它的价格低, 跌也跌不到哪里去; 二是因为价格低上涨相对容易, 涨 1 元钱就上涨了 50%。所以这种符合投机特征股票的价格上涨是投机的结果, 不能说它就具有投资价值。相反, 从投资角度来看, ST 华光的投资风险是巨大的。陕西金叶和山大华特作为较不适宜投资的股票, 它的市场表现和原因与 ST 华光相似。

从 Vague 集相似度量的股票选择结果看, 歌华有线是最具有投资价值的股票, 事实证明的确如此。从一个较长的时间(从 2005 年至今)角度看, 它的价格处于长期的上升通道之中便说明了这一点。由于其价格已经比较高, 是 ST 华光的 7 倍多, 短期内股民对它的追涨意愿下降, 从 2006 年第四季度到 2007 年第一季度看, 其涨幅比 ST 华光的涨幅小一些也是合理的。有鉴于此, 基于 Vague 集相似度量的股票选择是有效的。

对于较适宜投资的电广传媒、浙江东日、博瑞传媒和东方明珠 4 只股票, 验证的结果也很符合我们的分析。这 4 只股票满足了上涨的所有特征, 公司的业绩好, 市盈率低, 价格适中, 购买力强, 所以上涨的幅度较大, 证明了我们股票选择的有效性。

同 ST 华光一样, 交大南洋和西藏圣地是我们选择不适宜投资的股票, 它们的市场表现也是一样, 它们的上涨是因为大盘快速拉升所产生的普涨结果。

结束语 每股收益、每股净资产、每股公积金、每股未分配利润、每股经营现金流、主营收入同比增长率、净资产收益

率、股东权益比率是影响股票是否具有内在投资价值的 8 个关键指标因素, 根据上述关键性指标构建股票期望 Vague 集, 计算选择股票的 Vague 集并与期望 Vague 集之间的相似度量性比较, 进而划分股票投资指标等级体系, 以此确定股票的投资性是非常有效的, 对投资者的理性选择很有价值。尽管影响股票市场的因素很多, 宏观因素包括政策法规、利率调整、国际金融市场的变化、重要人物的讲话等等、微观因素包括行业发展、区域经济, 公司发展等等, 但理性的股票投资选择是以公司内部的财务发展状况做出的。从企业业绩的角度来看, 基于 Vague 集相似度量的股票选择方法, 无疑对投资者控制风险增加收益是很有价值的。

参 考 文 献

- [1] 郝奕, 张强. 基于 Vague 集和属性综合评价的股票投资价值分析方法. 中国管理科学, 2005, 13(2): 15-21
- [2] Gau Wen-lung, Buehrer D J. Vague sets. IEEE Transactions on Systems, Man and Cybernetics, 1993, 23(2): 610-614
- [3] Zhu Zhen-guo, Wang Guo-yin. Similarity Measure of Vague Set // 2007 International Conference on Artificial Intelligence and Pattern Recognition (AIPR-07). Orlando, 2007; 495-499
- [4] Pindyck R S, Rubinfeld D L. Econometric Models and Economic Forecasts, 4th edition. the McGraw-Hill companies Inc, 1998
- [5] 施东晖. 上海股票市场风险性实证研究. 经济研究, 1996(10): 44-48
- [6] 常健. 我国上市公司业绩决定机制实证分析. 管理世界, 2003(5): 109-115
- [7] 戚海峰. 完善我国资本市场的方略. 金融研究, 2001(4): 9-12