

中图法分类号: TP391 文献标志码: A 文章编号: 1006-8961(2011)10-1866-10

论文索引信息: 贾宇, 冯志勇, 陈祉宏, 于永新. 颜色语义特征描述提取及其在图像分类中的应用 [J]. 中国图象图形学报, 2011, 16(10): 1866-1875

颜色语义特征描述提取及其在图像分类中的应用

贾宇, 冯志勇, 陈祉宏, 于永新

(天津大学计算机科学与技术学院, 天津 300072)

摘要: 为了弥补图像底层特征到高层语义之间的语义鸿沟, 提出一种颜色语义特征的构建方法以建立新的语义映射来提高图像分类准确率。通过提取底层颜色特征, 构建包含颜色概念的语义网络, 建立了颜色语义特征三元组, 利用机器学习分类算法进行图像分类。实验结果表明, 利用文章提出的新方法构建的语义特征向量进行图像分类, 不仅可以取得优秀的分类结果, 同时对不同的分类算法具有鲁棒性。

关键词: 语义鸿沟; 语义网络; 图像分类; 语义特征

Extraction of color semantic feature and application in image classification

Jia Yu, Feng Zhiyong, Chen Zhihong, Yu Yongxin

(School of Computer Science & Technology, Tianjin University, Tianjin 300072 China)

Abstract: Aiming at overcoming the semantic gap between low-level features and high-level semantic concepts of images and to improve the precision of image classification, we propose an approach to construct a color semantic feature to map low-level features to semantically meaningful categories. Color features are extracted first and then a semantic network containing a set of color concepts and related objects is build. Finally, a color semantic 3-tuple is being constructed. The proposed method can not only get better results in image classification, but is also more robust.

Keywords: semantic gap; semantic network; image classification; semantic feature

0 引言

随着现代电子计算机技术及 Internet 网络的迅速发展, 庞大的、多样的图像信息不断出现。各行各业对图像的使用越来越广泛, 进而推动了图像信息资源管理研究进一步发展。现在的图像广泛的分布在互联网上, 缺少有效的组织, 难以达到资源共享的目的。因此, 如何利用如此众多的图像信息并从中找出自己需要的信息, 是对图像信息查询技术提出的重大挑战。从 20 世纪 90 年代初期开始, 基于图像的内容(如颜色、形状、纹理和空间关系等特征)

来检索图像的技术(CBIR)应运而生, 具体应用如大型的图像检索系统 QBIC^[1], Visualseek^[2], Virage^[3], Photobook^[4]等。CBIR 克服了传统方法的不足, 融合了图像识别和数据库技术, 从而提供了更有效的检索手段。然而图像的数字化表征, 并不能很好地揭示其内在的本质, 深层的语义内容得不到很好的反映。单纯的底层数字特征已经没有办法为图像的有效管理提供更好的支持。为此, 本文提出一种新颖的高层语义特征构造方法, 通过底层的颜色直方图特征和预先定义的颜色概念, 结合 ConceptNet^[5]构建颜色概念和对象概念关联的语义网络, 根据图像的标注信息生成一个颜色语义三元组特征向量。

收稿日期: 2009-12-15; 修回日期: 2010-11-11

基金项目: 天津市科技支撑计划项目(08ZCKFGX00700)。

第一作者简介: 贾宇(1985—), 男, 天津大学计算机应用技术专业在读硕士研究生, 主要研究方向为语义图像分类及检索。E-mail: jiaYu7_7@yahoo.com.cn。

最后通过实验表明,加入了语义特征后分类的效果明显优于普通的底层颜色特征。本文同时与文献[6]中 Chen 等人提出的方法进行实验对比,展现了优秀的结果,并在不同的分类算法上具有很好的鲁棒性。

1 相关工作

Eakins^[7]根据用户需求,将图像查询分为 3 个语义层次,如图 1 所示。其中第 1 层是特征语义,即简单的利用图像底层数字特征(如颜色、纹理和形状)进行图像分类和检索,CBIR 技术就基于这个层次。而第 2 层是对象和场景语义,主要利用图像的导出特征结合逻辑推理进行对象和场景的分类与检索,目前主要利用机器学习的方法进行图像对象语义层的分类。第 3 层则是行为和情感语义,通过更高层的抽象,实现对图像表达的行为信息和情感信息的分类,这一层的研究目前较少。其中特征语义到高层语义的差别被称为“语义鸿沟”(semantic gap)。

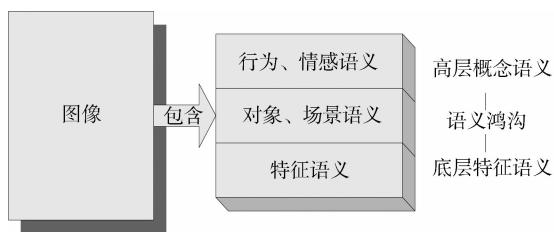


图 1 图像语义层次模型

Fig. 1 Image semantic hierarchical model

为了弥补这一语义鸿沟,目前的研究主要是针对第 2 层对象和场景语义,通过一些底层特征或底层特征的组合,利用机器学习方法进行图像分类。其中,Szumner 和 Picard^[8]首先考虑了户内/户外图像的分类问题,使用基于子块的颜色直方图图像特征及 k-NN 分类方法在两类分类中取得了 90.3% 的优秀结果。Vailaya 等人^[9]基于全局图像试图通过加入约束条件从底层特征中捕获高层语义特征,并使用 k-NN 分类器进行图像分类。Shen 等人^[10]认为图像分类应该采用多种视觉特征的组合(CMVF),通过组合各个特征构建特征向量来提高分类有效性,并使用 SVM、k-NN 和 GMM 等分类方法验证。Kang 等人^[11]选择 MPEG-7 提供的特征描述符,并引入一个联合的神经网络进行针对运动类图像的分类。Li 等人^[12]在图像分块后对每个区域

进行文本标注,提出一种概率框架模型用来自动的学习场景概念,与他人的区别在于他们使用了 codewords 进行无指导学习。

另外,部分研究也从特征语义层入手,通过对底层特征进行处理,或者加入标注,从而构建语义信息,完成图像分类。其中,Chang 等人^[13]利用图像全局特征并以一定的置信度为每幅图像生成了一系列语义标记,通过手工标注为训练集加入标注信息,并使用 SVM 分类器进行分类。Nezamabadi-pour 等人^[14]采用同样的方法,通过将颜色直方图和边缘直方图等视觉特征与自定义的语义特征组合构建特征向量的方法,使用模糊 k-NN 分类算法进行分类。Serrano 等人^[15]则提出一种语义信息构造方法,通过使用 Bayesian Networks 整合底层特征来产生用于分类的语义信息特征,并用该特征进行户内/户外图像分类。Paek 等人^[16]提出一种丰富户内/户外图像分类的方法,通过在特征向量中包含用来描述场景概念的文本信息,结合颜色和纹理特征来共同构建特征向量用于分类,并在 1 300 幅图像中取得了 86% 的分类准确率。而 Fredembach 等人^[17]则提出特征区域的思想,特征区域是围绕图像区域、位置和形状的几何特征,利用 PCA 方法计算,将结果存入特征区域数据库,最终利用特征区域匹配进行图像分类。Baharudin 等人^[18]提出一种图像特征构建方法,通过对 JPEG 图像文件编码并 DCT 变换,与底层颜色直方图特征等组合构建一个高维特征向量,使用 ANN 和 SVM 进行图像分类,并测试对比了各种用于图像特征的效果及两类分类算法的性能。Chow 等人^[19]提出一种基于树形结构化特征集合的图像分类方法,其中图像内容通过一棵 2-level 树表达,树根节点代表全局特征,而树叶节点表示不同区域,该方法构建的特征融合了图像全局和局部特征。本文提出的方法也是基于特征语义层去解决语义鸿沟问题,但是与上述文献所提方法不同之处在于,本文方法通过对颜色特征的语义分析,创建了新的颜色语义特征分量,这种三元组分量融合了图像底层特征的数字特性和高层语义的概念特性,很好地表达了特征语义。

2 语义特征建模及图像分类

图 2 展示了本文提出方法的整个流程。为了解决语义鸿沟问题,本文通过对底层颜色特征提取和

为颜色语义建模,显式的将图像的颜色语义表达出来,并将图像的特征表达为“底层特征+语义特征”形式的特征向量。最后利用分类算法进行分类。

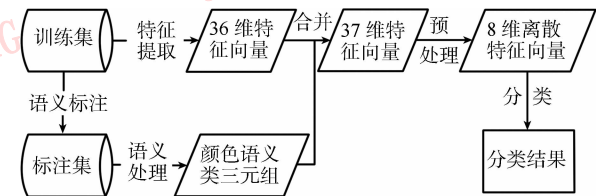


图 2 总体结构图
Fig. 2 Overall structure chart

2.1 底层特征提取

颜色是图像的一个重要特征,是图像分类中具有广泛应用的特征之一。相对几何特征而言,颜色特征稳定性好,对大小、方向等均不太敏感,并且具有彩色不变性(color constancy)^[20]等特点,因此,在许多情况下,颜色成为描述一幅图像最简便且十分有效的特征,在图像分类中最早得到应用。纹理特征是对局部区域中像素之间关系的一种度量,通常被看做图像的局部性质。形状特征也是图像处理中常用的一个重要特征,但是由于物体形状的自动获取很困难,形状特征的应用往往局限于容易识别物体图像的比较。而本文所提方法试图构建的语义特征往往表达了图像某种主观的容易理解的性质,相较纹理和形状等其他底层特征而言,颜色所带来的视觉直观刺激更明显,所表达的语义更容易被理解和提取。因此,只采用颜色特征作为构建语义特征的底层数字特征。

常用的数字图像一般是基于 RGB 颜色空间的,考虑到 HSV 颜色空间更接近于人们对颜色的主观认识,因此 HSV 空间下的颜色直方图特征^[21]作为底层数字特征被选用。通过计算提取颜色直方图特征,并将其量化为 36 柱,计算结果为 36 维特征向量,并归一化至[0,1]区间,得到向量 f_i ,其中 i 为1~36。

2.2 定义颜色语义类

对颜色语义的表达,反映到人类的主观行为上就是为颜色命名,从而方便理解和分类。文献[22-23]提出颜色命名表来规范颜色的语义。具体做法是针对 HSV 颜色空间,定义了不同的命名规则。其中针对主色度(HP),定义了 11 种词汇表达;针对次色度(HS)定义了 4 种词汇表达;针对色饱和度(S)定义了 5 种词汇表达;针对亮度(L)定义了 7 种表达;针对消色差(A)定义了 3 种表达,具体见表 1。

表 1 颜色命名表^[22]
Tab.1 Color naming syntax^[22]

主色度	次色度	饱和度	亮度	消色差
红	红	灰	黑	黑
橙	棕	温和	较暗	灰
棕	黄	中等	暗	白
黄	绿	强烈	中等	
绿		鲜艳	亮	
蓝			很亮	
紫			白	
粉				
米黄				
洋红				
橄榄绿				

表 1 对所有的颜色概念命名,为理解颜色语义提供了一个很好的标准。但是,由于实际中人理解图像时对颜色饱和度的关注远远弱于对色度和亮度的关注,本文定义了颜色语义类三元组 $T = [HP, HS, L]$,三元组中的每一元对应表 2 中的每一列。通过这样的划分将图像所表达的颜色语义划分为主色度、次色度和亮度 3 种。

表 2 颜色语义类表
Tab.2 Color semantic class naming syntax

主色度	次色度	亮度
红	红	黑
橙	棕	暗
棕	黄	灰
黄	绿	亮
绿	蓝	白
蓝	紫	
紫	粉	
粉		
米黄		
洋红		
橄榄绿		

三元组中 3 个元素的组合共有 $11 \times 7 \times 5 = 385$ 种。作者认为人类看到的物体和场景所包含的颜色语义信息可以通过这些颜色名称的组合来有效表达。例如,图 3 所示的一幅海滨图像,人类观察的顺序可能首先是沙滩和大海,其次是人。即大面积区域会率先进入人的眼球,因此,取沙滩和大海对应的

颜色作为颜色类的主要颜色和次要颜色。而大海和沙滩对应的颜色,通过常识可以知道是蓝和黄,所以对应的三元组可能是[蓝,黄,亮]。这种颜色与对象的关联知识通过常识颜色语义网络来获取。常识颜色语义网络是一系列对象概念和其颜色属性的关联,比如人看到大海就会联系到蓝,而看到树叶就会联系到绿等。



图 3 海滨图像
Fig. 3 Beach image

本文通过构建一个表达颜色和对象概念关联的语义网络,将颜色概念和对象概念关联起来从而得到对象的颜色语义属性。ConceptNet 是 MIT 媒体实验室发明的基于常识(common-sense)的大型语义网络。该网络中每个节点对应一个常识概念,每条边对应两个概念之间的某种语义关系。ConceptNet 的底层知识库是由 300 000 个节点接受 1 600 000 个断言构成的。它的关系本体由 20 种关系构成,知识库则通过互联网用户贡献的知识构建。本文构建的语义网络是通过获取与表 2 中颜色名相关的节点概念,将形如“对象-颜色”的映射存入数据库,最终压缩过滤 ConceptNet 中的语义概念形成一个只与颜色概念有关的精简网络结构,称为 ColorSemNet,结构如图 4 所示。

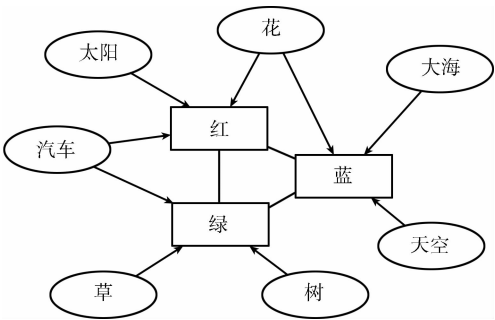


图 4 ColorSemNet 示意图
Fig. 4 ColorSemNet model

在 ColorSemNet 网络结构中,矩形标记表示颜色概念,椭圆标记表示对象概念,有向边表示对象可能具有的颜色概念。比如概念“太阳”指向颜色“红”表明太阳具有颜色属性红,“花”指向“红”和“蓝”表明花可能具有颜色属性集合[红,蓝]。有些概念具有多重可能颜色,因此可以包含颜色集作为候选颜色。本文就是通过构建这样一个概念网络来表达常识中对象所表达的颜色语义信息。

2.3 生成颜色语义三元组

结合 2.2 节提到的颜色语义类的定义和 ColorSemNet 语义网络,本节具体讲述如何构造颜色语义特征。

首先,利用图像标注信息及关键词构造语义特征。形式自由的标注词以及不统一的词汇表示为特征构造带来很大的障碍。本文通过使用 WordNet^[24]对所有关键词进行过滤,实现图像标注词与 ColorSemNet 概念节点关键词的匹配,为后续的特征构造操作提供了一致的词汇概念。

然后,如图 5 所示,构造一个图像的标注词集合到 ColorSemNet 的映射,定义一个映射表来存储图像中包含的不同颜色所占据的权重。匹配过程的具体做法是:如果 ColorSemNet 中包含这个关键词概念,那么该关键词对应的颜色属性被添加到映射表中作为键,并将值 +1。说明这个标注词所表达的概念在 ColorSemNet 中已经有定义,可以直接得到其颜色属性。如果不包括该词,对于这种异常情况,或者表明 ColorSemNet 不够完善,没有包含足够多的概念颜色信息,或者表明该标注词可能非常晦涩少见,并非常识概念。但是,这个标注词的影响不能被忽视,所以引入 unknown 添加到映射表中作为辅助表示。Unknown 弥补了语义网络的规模不足,表达了知识处理中对未知性的处理。比如表 3 中的图像(a)对应的语义计算结果,由于“太阳伞”在 ColorSemNet 中无定义,因此加入 unknown 表达这种知识的缺失。最终这个映射表将包含图像中由标注词表达的所有颜色的权值,将这些权值排序输出,取权值最大的两个颜色概念作为三元组的 HP 和 HS,取最大的亮度值作为三元组的 L。当然,在匹配过程中可能发生的另外一种异常情况就是一幅图像所包含的标注词太少。对于一些包含少量对象的简单场景图像,比如表 3 中的图像(b),标注词可能只有“花”和“树叶”。通过 ColorSemNet 获取的颜色属性只有“粉”和“绿”无法构建颜色语义三元组。为了

避免三元组属性缺失,本文引入 none 作为辅助元添加到三元组中。在标注词数量少于 3 个时,使用 none 来填充空余位置以保证三元组的完整。相比 unknown 表示的语义未知性和不确定性,none 在语义上表达了非重要性,即 none 表达了图像所包含的常识中不重视的区域或部分。

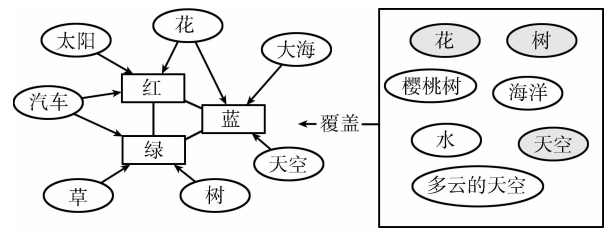


图 5 标注词到 ColorSemNet 的映射

Fig. 5 Mapping from annotation to ColorSemNet

表 3 颜色语义类示例

Tab. 3 Color semantic class example

图片	标注词	颜色语义三元组
	天空 海滩 太阳伞 人 海洋	[蓝, unknown, 白]
(a)		
	花 树叶	[粉, 绿, none]
(b)		

以上一系列步骤完成了对颜色语义三元组的构造,而整个创建过程等同于模拟人类理解一幅图像的过程。对于其中可能遇到的两种异常情况,通过引入 unknown 和 none 作为辅助元,解决了异常处理问题。最终构造出的颜色语义特征用 T 表示,构造算法的伪代码如下:

Input:

1) 图像标注的关键词列表: $wlist[w_1, w_2, \dots]$, 其中 w_i 是 $wlist$ 中的第 i 个关键词;

2) 语义网络中的关键词-颜色映射: $wmap < word, color >$, 其中 $word$ 是 ColorSemNet 中的圆形节点, $color$ 是矩形节点,

Foreach word w in $wlist$

利用 wordnet 计算 $base(w)$

If $wmap$ 包含 w

Then 从 $wmap[w]$ 得到 $colorlist$, 并定义 $colormap$

Foreach color c in $colorlist$

If $colormap$ 包含 c

Then $colormap[c]++$

Else 把 c 放到 $colormap$ 中, 并令 $colormap[c] = 1$

End Foreach

Else 令 $colormap[unknown]++$

End Foreach

If $colormap$ 的容量小于 3

Then 把 none 加入 $tuple$

Else 按照不同的颜色排序 $colormap$, 并将前 3 个 colors 加入 $tuple$, 并定义为 HP, HS and L

Output: 三元组 [HP, HS, L]

在图像语义特征构造完成之后,需要进行图像分类。主要过程分为两步:1) 特征的预处理;2) 利用分类算法分类。

2.4 特征预处理

加入颜色语义三元组后所构建的特征向量共包含 37 维特征,其中 36 维为颜色直方图特征,1 维是颜色语义三元组。由于颜色直方图特征中部分属性的无关性会影响分类效果,为了剔除这些影响,本文在分类前对特征进行了属性选择和离散化。

属性选择的一种方法就是选择一个属性子集,属性各自与类属性有较大关联,但几乎没有内部关联。两个名词性属性 A 和 B 之间的关系可用对称不定性来衡量

$$U(A, B) = 2.0 \times \frac{H(A) + H(B) - H(A, B)}{H(A) + H(B)} \tag{1}$$

式中, H 是熵函数。熵是以每个属性值的概率为基础的; $H(A, B)$ 是 A 和 B 的联合熵,由 A 和 B 的所有组合值的联合概率计算出来。对称不定性总是在 0 和 1 之间。本文采用基于相关性的特征选择方法^[25],该方法通过式(2)决定一个属性集的优良性。

$$\frac{\sum_j U(A_j, C)}{\sqrt{\sum_i \sum_j U(A_i, A_j)}} \tag{2}$$

式中, C 是类属性, i 和 j 包括属性集里的所有属性。使用这个标准来寻找特征子集,需要在属性空间进行搜索。本文采用最佳优先搜索(best-first search)方法来搜索整个特征空间,设置连续无改进情况的节点数目作为搜索停止条件。通过上述算法选定的

属性子集 $\mathbf{Subset} = \{f_5, f_{10}, f_{13}, f_{23}, f_{25}, f_{30}, f_{34}, T, c\}$ 进行分类,其中 f_i 是选择的特征属性, T 是颜色语义三元组, c 是类标签。颜色语义三元组属性在经过属性选择后仍然得以保留,这也证明了该特征具有很好的相关性和非重复性。

由于颜色直方图特征是连续的数值属性,而采用的分类器只能处理名词性属性,所以需要将属性子集离散化。本文使用基于熵的有指导离散方法^[26],该方法主要通过递归的分裂区间直至终止条件,选取分裂点同决策树的构建相似,选信息值最小的分裂点。从理论上讲,最小的信息值绝对不会出现在两个同属一个类的实例之间,接受或者拒绝一个分裂使用 MDL 原理见式(3)。结论是使信息增益 G 超出某个定值时进行分裂,这个定值由实例数量 N 、类别数量 k 、实例集的熵 $\mathcal{E}$ 、每个子区间内实例集的熵 $\mathcal{E}_1$ 和 $\mathcal{E}_2$,以及每个子区间内的类别数量 k_1 和 k_2 决定。

$$G > \frac{\log_2(N-1)}{N} + \frac{\log_2(3^k-2) - k\mathcal{E} + k_1\mathcal{E}_1 + k_2\mathcal{E}_2}{N} \quad (3)$$

式中,第 1 部分是描述分裂点所需的信息,第 2 部分是传输那些类别分别对应于前面和后面的子区间所需的一个修正量。

经过上述的特征选择和离散化操作后,最终得到了可以用于分类的属性子集。

2.5 使用 WAODE 分类

由于本文提出的语义特征与标注词相关联,而与图像的颜色直方图特征无明显依赖关系,因此可以认为特征向量中的直方图部分与语义三元组部分是相互独立的。所以可以采用基于朴素贝叶斯的 WAODE 算法对图像分类。

WAODE 算法是 Jiang 等人^[27]提出的一种基于 AODE 的改进算法,AODE 算法为朴素贝叶斯中的每个属性建立了一个共同的父亲节点,弱化了这种独立性假设。在 AODE 中,每个树增强型朴素贝叶斯(TAN)的节点被等同的对待。而现实中的许多应用表明,不同的属性在分类过程中可能起到不同的作用。因此,WAODE 为每个节点加入了权重,从而体现这种特征属性之间的区别。

WAODE 算法通过式(4)对实例进行分类

$$\arg \max_{c \in C} \left(\frac{\sum_{i=1}^n W_i P(a_i, c) \prod_{j=1, j \neq i}^n P(a_j | a_i, c)}{\sum_{i=1}^n W_i} \right) \quad (4)$$

式中, W_i 表示 TAN 中的属性 A_i 的权重。
在分类过程中,特征属性的重要性可以用该属性与类标签之间的交互信息来度量,交互信息 $I_p(A_i; C)$ 可以由式(5)计算

$$I_p(X; Y) = \sum_{x, y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \quad (5)$$

式中, X 和 Y 是属性变量。WAODE 用到的基础概率 $P(a_i, c)$ 和 $P(a_j | a_i, c)$ 使用 M 估计来计算,具体如下:

$$P(a_i, c) = \frac{F(a_i, c) + \frac{1.0}{(v_i \times k)}}{N + 1.0} \quad (6)$$

$$P(a_j | a_i, c) = \frac{F(a_j, a_i, c) + \frac{1.0}{v_j}}{F(a_i, c) + 1.0} \quad (7)$$

式中, $F(\cdot)$ 是属性组合在训练数据集中出现的频率, N 是训练数据实例数, v_i 是根属性 A_i 的取值数目, v_j 是叶属性 A_j 的取值数目, k 是类标签数目。

最终实验使用 WAODE 对经过预处理后的特征子集 $\mathbf{Subset} = \{f_5, f_{10}, f_{13}, f_{23}, f_{25}, f_{30}, f_{34}, T, c\}$ 进行分类,其中 f_i 和 T 是离散属性, c 是类标签。

3 实验结果

本文设计了 3 组实验来验证颜色语义特征的优秀性。图像库来源于 Wang 等人^[28]的 1 000 幅图像,预分为 10 类。每个类包含 100 幅图像,并且每幅图像进行了手工关键词标注,具体分类情况如表 4 所示。

表 4 图像分类主题
Tab.4 Image classification themes

编号	主题
1	建筑物
2	海滩
3	公共汽车
4	恐龙
5	大象
6	花
7	马
8	雪山
9	食物
10	非洲人民

3.1 分类准确率实验

为了证明引入颜色语义三元组特征在图像分类中能提高分类准确率,本文设计了分类准确率实验。

本实验在 WEKA^[29] 下进行,分类器选择 WAODE 方法,实验样本集为图像库全部 1 000 幅图像,特征向量即为文章第 2 部分构建的 *Subset* 特征子集。实验采用十折交叉验证,迭代 10 次测试,查准率 (precision) 和查全率 (recall) 为

$$\begin{aligned} \text{查准率} &= \frac{\text{检索出的相关图像数}}{\text{检索出的图像总数}} \times 100\% \\ \text{查全率} &= \frac{\text{检索出的相关图像数}}{\text{系统中的图像总数}} \times 100\% \end{aligned} \tag{8}$$

分类结果如图 6 所示,平均分类准确率达到 91.7%,“海滩”和“雪山”类分类结果不够理想。图 7 显示了分类正确和错误的示例图像,其中 (a) (b) 为分类正确的情况,而 (c) (d) 为分类错误的情况。可以看到,本是“公共汽车”类与“雪山”类,却被错误划分为“建筑物”类和“海滩”类。分析原因,图 7(c) 中的“公共汽车”在标注词的选择上是“白色的公共汽车”,与常见的公共汽车的颜色 (黄色或者红色) 违背。根据这样的标注词构造的颜色语义三元组可能是 [白, 蓝, unknown]。而“公共汽车”类图像普遍包含“公共汽车”、“天空”和“道路”,对应的颜色语义三元组是 [黄, 蓝, 灰]。“建筑物”类图像普遍包含“建筑物”、“天空”和“道路”之类的词,对应的颜色语义三元组是 [白, 蓝, 灰]。在特征空间中,该特征与本来隶属的“公共汽车”类相距较远,而与“建筑物”类相距较近,引起误差。同理,图 7(d) 中的“雪山”被误分为“海滩”,主要原因也是标注词

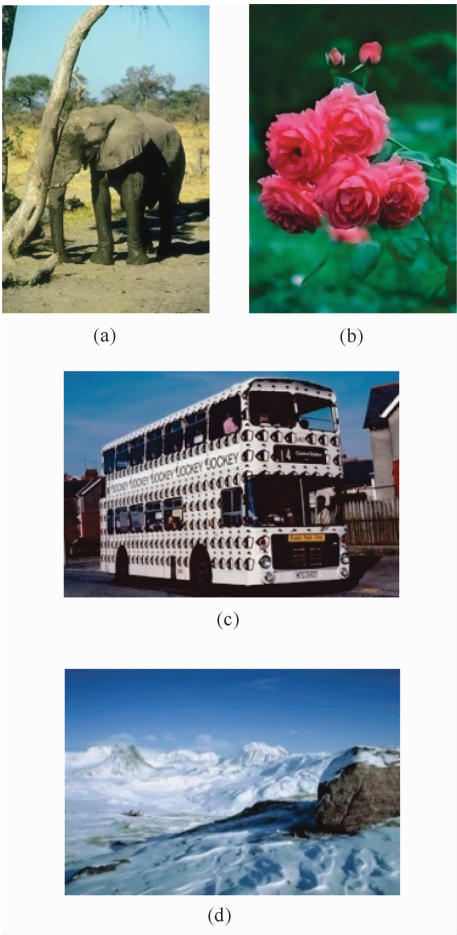


图 7 分类正确和错误示例

Fig. 7 True and false example

“天空”、“雪”和“岩石”产生的颜色语义三元组与“海滩”类的对应特征相近似导致。改进这种分类误差将是未来要做的工作。

图 8 显示了与图 7 包含的 4 幅图像对应的查准率-查全年曲线。相比传统的只使用颜色特征来分类的方法,本文提出的颜色语义特征在同样的分类算法中具有明显优势。

3.2 分类准确率与样本数的敏感性实验

为了观察语义特征的分类性能随样本数的增多而产生的变化趋势,本文还设计了分类准确率与样本数的敏感性实验。

本实验设定与 3.1 节相同的实验环境,样本集从图像库子集 200 幅图像递增到 1 000 幅,分类特征分别为语义特征向量 (37 维) 与传统的底层特征向量 (36 维)。图 9 显示了在分类过程中随着样本数增多的一个平均准确率曲线图。

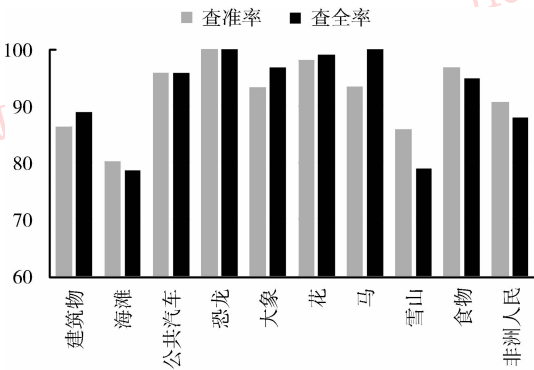


图 6 分类结果

Fig. 6 Classification result

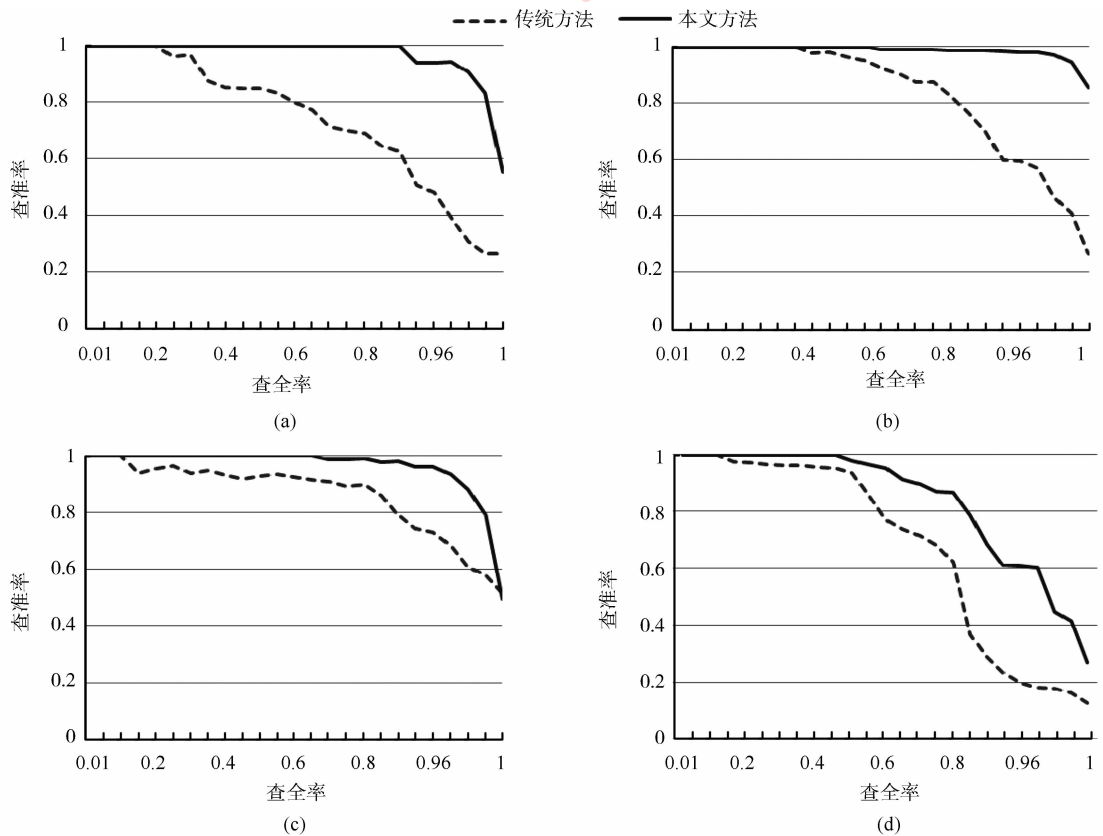


图 8 查准率-查全率曲线
Fig. 8 Precision-recall curve

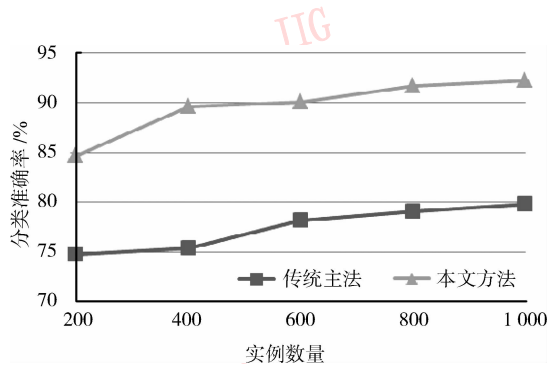


图 9 新旧方法准确率对比图
Fig. 9 Comparison with traditional method

由对比可见,本文颜色语义特征构建方法,比传统只使用底层特征进行分类,在特征维度上增加了 1 维,但是准确率却大幅度提升。

与此同时还进行了另一个对比实验,实验环境设置与文献[6]相同,基于同样的图像库来进行实验。如图 10 所示,随着测试样例的减少,训练样例的增多,本文方法在分类时有着优良的性能。

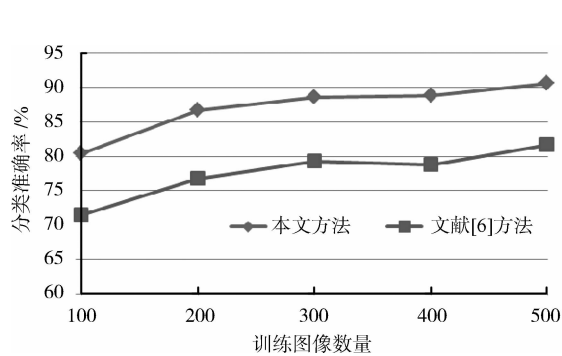


图 10 与文献[6]的实验对比图
Fig. 10 Comparison with reference [6]

3.3 特征鲁棒性实验

为了证明本文提出的特征构建方法产生的语义特征是有效的并且适合多种分类算法(对分类算法无依赖),本文还设计了特征鲁棒性实验。

实验环境仍然选用 WEKA,样本集从图像库子集 200 幅递增到 1 000 幅,分类算法采用了 WAODE、SVM、朴素贝叶斯、PART、J48 决策树和随机子空间共 6 种学习算法。结果如图 11 所示,对于所有的分类器均表现

出优秀的分类准确率。从而证明了本文提出的语义特征在分类算法层面的无依赖性。

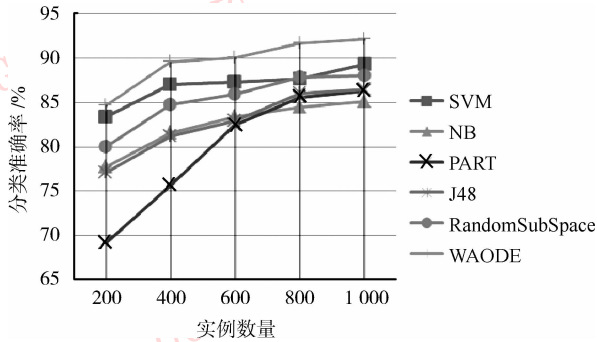


图 11 不同分类算法下的分类结果

Fig. 11 Classification result on different classifier

4 结 论

首先提出一种语义特征构造方法,结合图像的 HSV 空间颜色直方图特征以及标注信息,利用 ConceptNet 和 WordNet 构建了颜色语义网络 ColorSemNet;然后生成了一种包含语义特性的颜色语义三元组;最后利用 WEKA 提供的多种分类器进行图像分类实验。实验结果表明,本文提出的方法所构建的语义特征不仅可以在小型图像库中取得优秀的分类结果,而且对于不同类型的分类器算法在分类性能上差距不大且性能卓越,具有鲁棒性。当然鉴于目前实验图像库的规模不足以及特征选取的单一,下一步工作将在大型图像库上进行分类实验,同时扩充底层特征向量,将纹理和形状等底层特征结合进来,赋予其特征语义,构建不仅能表达颜色同时还能表达形状和纹理的语义特征。完善语义网络,改善分类性能。

参考文献 (References)

- [1] Faloutsos C, Barber R, Flickner M, et al. Efficient and effective querying by image content[J]. Journal of Intelligent Information Systems, 1994, 3(3-4): 231-262.
- [2] Smith J R, Chang S F. Visualseek: a fully automated content-based image query system[C]//Proceedings of the Fourth ACM International Conference on Multimedia. Boston, Massachusetts, USA: ACM, 1997: 87-98.
- [3] Hampapur A, Gupta A, Horowitz B, et al. Virage video engine [C]//Proc. SPIE: Storage and Retrieval for Image and Video Databases V. San Jose, CA, USA: SPIE, 1997: 188-197.
- [4] Pentland A, Picard R W, Sclaroff S. Photobook: content-based manipulation of image databases [J]. International Journal of Computer Vision, 1996, 18(3): 233-254.
- [5] Liu H, Singh P. ConceptNet—a practical commonsense reasoning tool-kit[J]. BT Technology Journal, 2004, 22: 211-226.
- [6] Chen Y, Wang J Z. Image categorization by learning and reasoning with regions [J]. Journal of Machine Learning Research, 2004, 5: 913-939.
- [7] Eakins J P. Automatic image content retrieval—are we getting anywhere? [C]//Proceedings of Third International Conference on Electronic Library and Visual Information Research. Milton, Keynes: De Montfort University, 1996: 123-135.
- [8] Szummer M, Picard R W. Indoor-outdoor image classification [C]//IEEE International Workshop on Content-Based Access of Image and Video Databases. Bombay, India: CAIVD, 1998: 42-51.
- [9] Vailaya A, Jain A, Zhang H. On image classification: city vs. landscapes [J]. Pattern Recognition, 1998, 31(12): 1921-1935.
- [10] Shen J, Shepherd J, Ngu A H H. Semantic-sensitive classification for large image libraries [C]//International Multimedia Modelling Conference. Melbourne, Australia: MMM, 2005: 340-345.
- [11] Kang S, Park S. A fusion neural network classifier for image classification[J]. Pattern Recognition Letters, 2009, 30(9): 789-793.
- [12] Li F F, Perona P. A Bayesian hierarchical model for learning natural scene categories[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: CVPR, 2005: 524-531.
- [13] 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 Special Issue on Conceptual and Dynamical Aspects of Multimedia Content Description, 2003, 13(1): 26-38.
- [14] Nezamabadi-pour H, Kabir E. Concept learning by fuzzy k-NN classification and relevance feedback for efficient image retrieval [J]. Expert Systems with Applications: an International Journal, 2009, 36(3): 5948-5954.
- [15] Serrano N, Savakis A E, Luo J. Improved scene classification using efficient low-level feature and semantic cues[J]. Pattern Recognition, 2004, 37(9): 1773-1784.
- [16] Paek S, Sable C L, Hatzivassiloglou V, et al. Integration of visual and text based approaches for the content labeling and classification of photographs [C]//ACM SIGIR '99 Workshop Multimedia Indexing Retrieval. Berkeley, CA, USA: SIGIR, 1999: 15-19.
- [17] Fredembach C, Schröder M, Süssstrunk S. Eigenregions for image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(12): 1645-1649.
- [18] Baharudin B, Qahwaji R, Jiang J, et al. Combining image

- features for image classification[C]//International Conference on Intelligent Advanced Systems. Kuala Lumpur: ICIAS, 2007: 268-272.
- [19] Chow T W S, Rahman M K M. A new image classification technique using tree-structured regional features [J]. Neurocomputing, 2007, 70(4-6): 1040-1050.
- [20] Wandell B A. The synthesis and analysis of color image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1987, 9(1): 2-13.
- [21] Manjunath B S, Ohm J R, Vasudevan V V, et al. Color and texture descriptors [J]. IEEE Trans. Circ. Syst. Video Technol., 1998, 11(6): 703-715.
- [22] Depalov D, Pappas T, Li Dongge, et al. A perceptual approach for semantic image retrieval[C]//Acoustics, Speech and Signal Processing. Toulouse: ICASSP, 2006: 14-19.
- [23] Mojsilović A. A computational model for color naming and describing color composition of images[J]. IEEE Transactions on Image Processing, 2005, 14(5): 690-699.
- [24] Miller G A, Beckwith R, Fellbaum C, et al. WordNet: an on-line lexical database [J]. International Journal of Lexicography, 1990(3): 235-244.
- [25] Hall M. Correlation-based feature subset selection for machine learning[D]. Hamilton, New Zealand: University of Waikato, 1999.
- [26] Fayyad U M, Irani K B. Multi-interval discretization of continuous-valued attributes for classification learning [C]//International Joint Conference on Artificial Intelligence. Chambéry, France: IJCAI, 1993: 1022-1027.
- [27] Jiang L X, Zhang H. Weightily averaged one-dependence estimators [C]//Pacific Rim International Conference on Artificial Intelligence. Guilin, China: PRICAI, 2006: 970-974.
- [28] Wang J, 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.
- [29] Hall M, Frank E, Holmes G, et al. The WEKA data mining software: an update[C]//Sigkdd Explorations. New York, NY, USA: ACM, 2009: 10-18.