基于多尺度上下文语义信息的图像场景分类算法

张瑞杰[1],李弼程[2],魏福山[1]
(1.解放军信息工程大学四院,河南郑州 450000;2.解放军信息工程大学信息系统工程学院,河南郑州 450000)

摘 要: 传统视觉词典模型没有考虑图像的多尺度和上下文语义共生关系.本文提出一种基于多尺度上下文语义信息的图像场景分类算法.首先,对图像进行多尺度分解,从多个尺度提取不同粒度的视觉信息;其次利用基于密度的自适应选择算法确定最优概率潜在语义分析模型主题数;然后,结合 Markov 随机场共同挖掘图像块的上下文语义共生信息,得到图像的多尺度直方图表示;最后结合支持向量机实现场景分类.实验结果表明,本文算法能有效利用图像的多尺度和上下文语义信息,提高视觉单词的语义准确性,从而改善场景分类性能.

关键词: 场景分类; 多尺度信息; 概率潜在语义分析; 自适应主题数; 上下文语义信息
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2014)04-0646-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2014.04.004

Image Scene Classification Based on Multi-Scale and Contextual Semantic Information

ZHANG Rui-jie[1], LI Bi-cheng[2], WEI Fu-shan[1]
(1. *The Fouth Institute of Information Engineering University of PLA, Zhengzhou, Henan* 450000, *China*;
2. *Institute of Information System Engineering, Information Engineering University of PLA, Zhengzhou, Henan* 450000, *China*)

Abstract: The conventional BoVW neglects image multi-scale and contextual semantic co-occurrence information. This paper proposes an image scene classification algorithm based on multi-scale and contextual semantic information. Firstly, Images are decomposed into variant scales and diverse visual details are extracted from different scale layers. Secondly, a density-based adaptive selection method is employed to choose the best topic number of probabilistic latent semantic analysis model. Then, Markov random field are combined to mine the contextual semantic co-occurrence information, thus to obtain a multi-scale histogram as the image representation. Finally, the support vector machine classifier is utilized to perform scene classification. The experimental results demonstrate that our algorithm can effectively utilize the multi-scale and contextual semantic information of images and improve image scene classification performance.

Key words: scene classification; multi-scale information; probabilistic latent semantic analysis; adaptive topic number; contextual semantic information

1 引言

图像场景分类是将图像标记为不同语义类别的过程,它根据整幅图像或图像的某个局部区域中包含的语义内容,利用给定的一组语义类别对图像数据库进行自动标注,如城市、森林和沙滩等.场景分类将整幅图像看作一个整体,根据图像的语义信息构造场景内容,并不涉及图像中的具体目标.因此,场景分类能够很好地支持基于语义的图像分析和检索[1],同时也可以为目标识别等更高层次的图像理解提供有效的上下文语义信息.目前,图像场景分类技术得到了研究者的广泛关注.

早期的场景分类方法大都根据图像的全局特征来描述场景内容,如 Vailaya 等[2]提取图像的边缘方向一致性矢量和空间颜色矩来描述场景内容,并训练二类贝叶斯分类器将旅游图像分为 Indoor 和 Outdoor 两种场景.Rachid 等[3]融合图像的纹理和颜色特征实现足球场景的分类.近年来,视觉词典模型(Bag of Visual Words, BoVW)逐渐成为场景分类的主流方法[4~6].该方法的基本思想是首先定义图像块的不同语义概念(如天空、岩石等),称其为视觉单词(Visual Words);然后统计图像中视觉单词出现的频次作为图像的场景内容表示;最后结合机器学习方法训练识别图像场景类别.但是,这种简

收稿日期:2013-04-27;修回日期:2013-09-29;责任编辑:蓝红杰
基金项目:国家自然科学基金(No.60872142)

单的 BoVW 表示并没有考虑视觉单词与其周围图像块在语义空间中的共生性,因此并不能有效揭示图像块之间的上下文语义关系.而对于图像数据,相邻图像块之间存在着较强的上下文语义共生关系.

鉴于此,Lazebnik 等[7]利用空间金字塔(Spatial Pyramid Matching,SPM)模型将图像在空间上进行不同层次的划分,然后提取视觉单词的空间分布信息完成场景分类.Bosch 和 Feifei 等[8,9]分别利用概率潜在语义分析(Probabilistic Latent Semantic Analysis,pLSA)模型和潜在狄利克雷分布(Latent Dirichlet Allocation,LDA)模型分析得到图像的主题或者潜在语义.Emrah 等[10]将 pLSA 模型和空间金字塔方法相结合,先对图像进行空间金字塔划分,再分别在图像各层上利用 pLSA 模型挖掘潜在语义,提出基于 SPM-pLSA 模型的图像场景分类算法.刘硕研等[11]提出一种基于上下文语义信息的图像块视觉单词生成算法,利用 pLSA 模型和 Markov 随机场共同挖掘图像块的上下文语义共生信息,为图像块定义更准确的视觉单词.但是,上述方法均未考虑图像的多尺度信息,空间金字塔方法虽然按照多尺度特征汇总的方式融合了局部特征的空间分布信息,但这只是在视觉词典模型的后期阶段实施的.图像的视觉信息往往分布在多个尺度中,而多尺度特性作为图像的一个特有属性,对于获得有效的图像场景表示具有十分重要的意义.另外,语义主题模型中主题数的确定对于模型性能具有重要影响,但现有方法大都通过人工预设固定值的方式确定模型主题数,没有一种通用的自适应选择方法.

针对上述问题,本文在刘硕研等工作的基础上,对图像进行多尺度分解,从图像的多个尺度提取不同粒度的视觉信息,构建图像的多尺度视觉单词分布直方图表示;同时采用基于密度的自适应选择算法确定最优 pLSA 模型主题数,提出了一种基于多尺度和上下文语义信息的图像场景分类算法.首先,对图像进行多尺度分解,然后分别在各个尺度上利用改进的 pLSA 模型和 Markov 随机场挖掘图像块之间的上下文语义共生关系,构建图像不同尺度下的视觉单词分布直方图;最后加权连接各个尺度上的视觉单词分布直方图,得到图像的多尺度直方图表示,再结合支持向量机实现场景分类.实验结果表明,本文算法构造的图像多尺度直方图表示具有较好的分类性能.

2　相关理论

2.1　图像多尺度分解

尺度空间[12]是在图像处理和建模过程中引入一个尺度参数,通过这个参数的连续变化,获得原始图像数据的多尺度空间表示.Lindeberg 等在文献[13]中已证明在一般性的假设条件下,高斯核是尺度空间变换的唯一线性平滑核,并且满足平移不变性、半群结构、非增局部极值、尺度不变性和旋转不变性等重要性质.因此,本文选取高斯函数对图像进行多尺度变换.首先,采用高斯滤波器对图像进行平滑处理,假设输入图像为 $I(x,y)$,则平滑后图像 $L(x,y,\sigma_n)$ 由输入图像 $I(x,y)$ 与高斯函数 $G(x,y,\sigma_n)$ 卷积得到.

$$L(x,y,\sigma_n)=I(x,y)*G(x,y,\sigma_n) \tag{1}$$

其中,$*$ 表示卷积操作,高斯函数定义为

$$G(x,y,\sigma_n)=\frac{1}{2\pi\sigma_n^2}e^{-(x^2+y^2)/2\sigma_n^2} \tag{2}$$

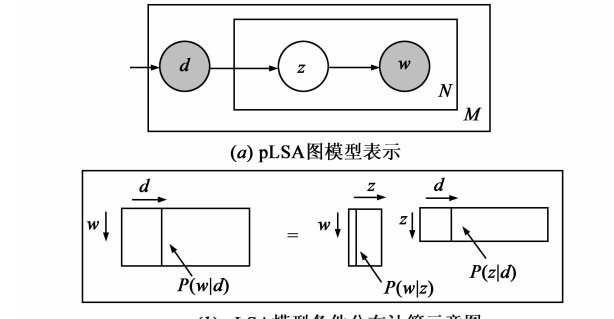
式中,$G(x,y,\sigma_n)$ 是以 σ_n 为尺度因子的高斯滤波函数.σ_n 大小决定图像的平滑程度,大尺度对应图像的概貌特征,小尺度对应图像的细节特征.本文中,尺度因子的递增过程如式(3)所示.

$$\sigma_n=\sigma\times 2^{n/s},\ n=0,1,2,\cdots,s+2 \tag{3}$$

其中,σ 取值 0.98;s 取值为 1,这样每个图像对应的金字塔有 4 层.原始图像按照式(1)卷积后,它的精细程度逐渐被平滑,图像中的一些细节和噪声也被消除和抑制,从而使得部分区域的视觉特征发生变化.因此,通过对原始图像的多尺度变换和分析,能够捕获各个图像块在不同尺度上的本质特性.

2.2　pLSA 模型原理

pLSA 模型[14]是 Hoffman 针对潜在语义分析提出的一种生成模型,它的主要思想是分析文档集中单词的共生性,最早用于文本处理,其模型原理如图 1 所示.



(a) pLSA图模型表示

(b) pLSA模型条件分布计算示意图

图1　pLSA模型原理示意图

图中利用黑框来表示模型 M 个文档和每个文档中 N 个单词的重复生成过程,其中实心部分 d 和 w 是实际观测到的变量,空心部分 Z 代表需要通过模型预测的未知变量,即文档中的潜在语义.给定一个文档集合 $D=\{w_1,w_2,\cdots,w_M\}$,文档中的单词取自词汇表 $\{1,2,\cdots,V\}$,则可以得到文档和单词的共现频率矩阵 $\boldsymbol{N}=[n(d_i,w_j)]$,其中 $n(d_i,w_j)$ 统计文档 d_i 中单词 w_j 出现的个数.假定由文档中的潜在语义主题 z 生成了单词,那

么整个文档的生成过程如下:

(1)选择一篇文档,其概率表示为 $P(d)$;

(2)选择一个隐含的主题 z,使得 $P(z|d)$满足多项式分布;

(3)在已知主题的条件下,单词 w 出现的条件概率 $P(w|z)$满足多项式分布.

由上述生成过程,单词和文档共同出现的联合分布表示为:

$$P(w,d)=\sum_{z\in Z}P(w,d,z)=P(d)\sum_{z\in Z}P(w|z)P(z|d) \tag{4}$$

由于 $P(w,d)=P(d)P(w|d)$,则根据式(4),可将 $P(w|d)$写为

$$P(w|d)=\sum_{z\in Z}P(w|z)P(z|d) \tag{5}$$

根据图 1,式(5)可认为是一个矩阵分解的过程,即每个文档中的单词分布由潜在语义主题 z 的凸组合构成,其中权重 $P(w|z)$统计了已知主题中单词的条件分布,$P(w|z)$与具体的文档无关.通过 EM 算法迭代如下的最大对数似然函数,就可以很容易地估计 pLSA 模型中的条件分布 $P(w|z)$和 $P(z|d)$.

$$L=\log P(\boldsymbol{D},\boldsymbol{W})=\sum_{d\in D}\sum_{w\in W}n(w,d)\log P(w,d) \tag{6}$$

类似地,对于图像数据,通过训练 pLSA 模型可以得到潜在语义在视觉单词上的分布概率 $P(w|z)$,假设有视觉单词 w_1 和 w_2,根据文[11],定义 $P(w_1,w_2)$为视觉单词 w_1 和 w_2 的语义共生概率,则

$$\begin{aligned}P(w_1,w_2)&=P(w_1)\sum_{t=1}^{T}P(w_2|z_t)P(z_t|w_1)\\&=\sum_{t=1}^{T}P(w_2|z_t)P(w_1|z_t)P(z_t)\\&=P(z)\sum_{t=1}^{T}P(w_2|z_t)P(w_1|z_t)\end{aligned} \tag{7}$$

其中,$P(z)$为语义主题的先验分布概率,T 是语义主题数.由式(7)可看出,视觉单词 w_1 和 w_2 之间的语义共生概率 $P(w_1,w_2)$可由潜在语义在视觉单词上的分布概率 $P(w|z)$求得.

2.3 Markov 随机场模型

Markov 随机场模型[11]考虑每个像元关于它的邻近像元的条件分布,能够有效地描述图像的空间上下文统计特性.因此,可以利用 Markov 随机场对二维网格上某邻域中像元间的关系进行建模.Markov 随机场理论指出:随机场在像元 i 处取相应状态值的概率仅与其邻域 $N(i)$的状态有关,表示为 $P(w_i|w_{N(i)})$,其中 $N(i)=\{s=1,2,\cdots,i-1,i+1,\cdots,N\}$是某邻域系统下格点 i 的所有邻域位置.根据 Hamersley-Clifford 定理[15],设图像块 $X=\{x_i\}$对应的视觉单词为 $W=\{w_i\}$,则视觉单词 $\{w_i\}$的上下文语义共生概率可表示为:

$$P(w_i|w_{N(i)})=\frac{\exp(\beta\sum_{j\in N(i)}P(w_i,w_j))}{\sum_{w_i}\exp(\beta\sum_{j\in N(i)}P(w_i,w_j))} \tag{8}$$

其中,β 是控制邻域间作用强度的参数.上式表明,图像块 x_i 对应的语义概念与其邻域区域语义概念的相关度越高,则 x_i 标记为该语义概念的可能性越大.因此,Markov 随机场理论可以为视觉单词的生成加入上下文语义约束信息,从而成为描述视觉单词上下文语义关系的有力工具.

3 基于多尺度上下文语义信息的图像场景分类

3.1 基于密度的最优 pLSA 模型主题数选择

在现有的主题模型研究中,主题数大多由人工设定为一个固定值[10,11].但在实际应用中,各图像场景类的繁简意义各不相同,所包含的潜在语义数量必定也不一样.因此需要设计一种自适应主题数选择算法,能够根据场景内容的繁简程度自动选择最优主题数,以更好地拟合场景类的语义内容.

目前这方面的研究工作还相对较少.2008 年,曹娟等[16]针对文本分析中主题数的确定问题,提出并证明了主题之间的相似度最小时模型最优的理论,并在此基础上提出基于密度的 LDA 模型主题数自适应选择算法,用相对少的迭代自动找出最优主题结构.考虑到 pLSA 模型是 LDA 模型的特例,本文将曹的理论引入图像场景分类领域,提出基于密度的最优 pLSA 模型主题数选择算法,以确定最优 pLSA 模型.下面我们先定义一些基本概念,再详细介绍算法具体流程.

定义 1 主题之间的相关性:设主题集合为 $Z=\{z_1,z_2,\cdots,z_T\}$,图像集为 $D=\{d_1,d_2,\cdots,d_N\}$,则主题 z_1 和 z_2 之间的相关性 $corr(z_i,z_j)$定义为

$$corr(z_i,z_j)=\frac{\sum_{n=1}^{N}P(z_i\mid d_n)\cdot P(z_j\mid d_n)}{\sqrt{\sum_{n=1}^{N}(P(z_i\mid d_n))^2\cdot\sum_{n=1}^{N}(P(z_j\mid d_n))^2}} \tag{9}$$

其中,$corr(z_i,z_j)$越小,主题之间越独立.

定义 2 所有主题之间的平均相似度:所有主题之间的平均相似度定义为:

$$avg\_corr=\frac{\sum_{i=1}^{T-1}\sum_{j=i+1}^{T}corr(z_i,z_j)}{T\times(T-1)/2} \tag{10}$$

其中,T 为总的主题数,$corr(z_i,z_j)$为主题 z_1 和 z_2 之间的相关性.根据文[16],当所有主题之间的平均相似度最小时,主题结构最稳定,对应的主题模型性能最优.

定义 3　主题密度:对给定的主题 z 和距离 r,以 z 为中心,r 为半径画圆,根据式(9)分别计算 z 与其它主题之间的相似度,相似度取值落在圆内的主题数称为 z 基于 r 的密度,记做 $Density(z,r)$.

定义 4　模型基数:给定一个主题模型 Z 和正整数 m,模型中密度小于或等于 m 的主题数称为该模型的基数,记做 $Cardinality(Z,m)$.

定义 5　参考样本:对于主题分布中的一个点 p、距离半径 r 和阈值 n,如果满足 $Density(p,r)\leq n$,则称 p 代表的视觉单词为主题 z 的一个参考样本.

在以上定义的基础上,基于密度的最优 *pLSA* 模型主题数选择算法可描述为如下过程:

(1)设定初始主题数 K,根据设定的 K 值,对图像集以随机抽样的方式结合 2.2 节中的 pLSA 模型参数估计算法,得到主题向量 $P(z_i|d_j)(i=1,2,\cdots,T;j=1,2,\cdots,N)$,作为初始化向量;

(2)利用 $P(z_i|d_j)$计算所有主题之间的相似度矩阵$[corr(z_i,z_j)]$和平均相似度 avg_corr.令 $r=avg\_corr$,根据定义 3,计算所有主题的密度 $Density(z_i,r)|_{i=1}^{T}$;另设 $m=0$,计算该主题模型 Z 的基数 $C_n=Cardinality(Z,m)$;

(3)根据式(11)更新 K 值,再利用新的 K 值重新估计 pLSA 模型参数;

$$K_{n+1}=K_n+f(r)\times(K_n-C_n) \tag{11}$$

其中,$f(r)$表示 r 的变化方向,当 r 值减小时,$f_{n+1}(r)=-1\times f_n(r)$;当 r 值增加时,$f_{n+1}(r)=f_n(r)$;$f_0(r)=-1$.

当 $f_n(r)=-1$ 时,将主题从小到大按照密度排列,并将前 C_n 个主题看作参考样本,初始化下次 *pLSA* 模型参数估计的主题向量,反之则采用从训练集中抽样的方式初始化主题向量.反复执行步骤(2)和(3),直到平均相似度 r 和参数 K 同时收敛.

根据式(11),K 收敛的条件是 $\mathrm{argmin}(K_n-C_n)$.由模型基数 C 的定义可知,C 随着平均相似度的减小而增加,且 $C_n\leq K_n$.当 r 达到最小时,C 达到最大.因此,可以保证 r 和参数 K 同时收敛.

综上,基于密度的最优 pLSA 模型主题数选择算法通过不断迭代以最小化主题之间的平均相似度寻找模型的最优主题数.由文[16]知,此时的 pLSA 模型性能也是最优的.因此,通过基于密度的自适应主题数选择算法,可以在相对少的迭代情况下自动找到最优的 pLSA 模型主题结构,改善模型性能.

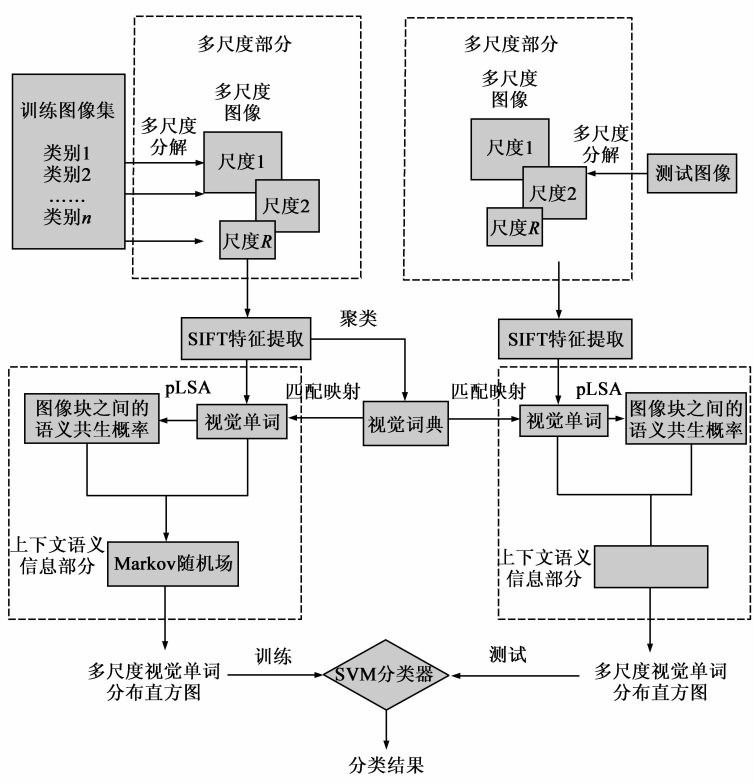


图2　基于多尺度上下文语义信息的场景分类算法流程图

3.2　基于多尺度上下文语义信息的图像场景分类算法流程

基于多尺度上下文语义信息的图像场景分类算法流程如图 2 所示.它包括训练和测试两个模块,每个模块又包含多尺度和上下文语义信息两部分.多尺度部分通过高斯滤波和降采样,将场景图像分解为包含多个不同尺度的高斯金字塔;上下文语义信息部分则在高斯金字塔的各层图像上,利用改进的 pLSA 模型和 Markov 随机场共同挖掘图像块的上下文语义共生信息,提取语义更准确的视觉单词,构建更有效的场景内容表示.

对于训练图像集 I^{Train},首先对其进行多尺度分解,然后使用 K-Means 算法聚类 I^{Train} 中各层图像的 SIFT 特征集,将每个聚类中心看作一个视觉单词,最后可生成各层图像的视觉词典 $W_r(r=1,2,\cdots,R)$.

对于测试图像 I^{Test},首先将其多尺度分解为 $L^{Test}=\{L_1^{Test},L_2^{Test},\cdots,L_r^{Test},\cdots,L_R^{Test}\}$,设第 r 层图像 L_r 包含图像块$\{x_1^r,x_2^r,\cdots,x_{D_r}^r\}$,$D_r$ 为 L_r 中包含的图像块总数;依据第 r 层图像的视觉词典 W_r,计算图像块$\{x_d^r\}(d=1,2,\cdots,D_r)$的 SIFT 特征与 W_r 中每个视觉单词的欧氏距离,选择最近邻的视觉单词来定义该图像块;这样就得到图像块$\{x_d^r\}(d=1,2,\cdots,D_r)$对应的视觉单词集$\{w_1^r,$

$w_2^r, \cdots, w_d^r, \cdots, w_{D_r}^r\}$($w_d^r \in W_r$).然后结合改进的 pLSA 模型和 Markov 随机场挖掘图像块的上下文信息.算法的具体流程如下所示:

(1)初始化:输入图像块$\{x_i^r\}$和视觉词典$W_r = \{w_d^r\}$,设置最大迭代次数 T 和阈值ε;

(2)利用改进的 pLSA 模型根据式(7)计算视觉单词 w_i^r 和 w_j^r 的语义共生概率 $P(w_i^r, w_j^r)$;

(3)利用 Markov 随机场模型根据式(8)计算视觉单词的上下文语义共生概率 $P(w_i^r \mid w_{N_r(i)}^r)$.

(4)根据 $Dis_{new}^2(x_i^r, w_k^r) = Dis^2(x_i^r, w_k^r)/P(w_i^r = k \mid w_{N_r(i)}^r)$重新计算图像块与视觉单词的距离,$Dis_{new}^2(x_i^r, w_k^r)$中包含有图像块 w_i^r 的类别及上下文语义信息.若用 z_i^r 表示更新后的视觉单词,则 $z_i^r = \underset{1 \le k \le D_r}{\operatorname{argmin}} Dis_{new}^2(x_i^r, w_k^r)$;且本次修改后视觉单词的数量更新为 $Num^r = \sum_i Num_i^r$,其中

$$Num_i^r = \begin{cases} 0, z_i^r(t) = z_i^r(t+1) \\ 1, \text{else} \end{cases} \tag{12}$$

(5)迭代步骤(4),如果 $t > T$ 或 $\max_t \| Num^r(t) - Num^r(t+1) \| < \varepsilon$,则 x_i^r 对应的视觉单词为 $z_i^r = \underset{1 \le k \le D_r}{\operatorname{argmin}} Dis_{new}^2(x_i^r, w_k^r)$;否则 $t = t+1$,转至(4).

经过上述步骤,第 r 层图像的视觉词典为$Z_r = \{z_i^r\}$($i = 1,2,\cdots,Num^r$).依次对图像的多个尺度层 L_r($r = 1,2,\cdots,R$)构建各自的视觉词典,就得到整幅图像的多尺度视觉词典 $Z = \{Z_1, Z_2, \cdots, Z_r, \cdots, Z_R\}$.图像的每一层 L_r 依据视觉词典 Z 生成直方图H_r,从而得到直方图集合 $H = \{H_1, H_2, \cdots, H_r, \cdots, H_R\}$.$H$ 中各项对应不同尺度的图像且维数相同,加权连接各尺度的直方图,得到一组更"长"的直方图表示,作为最终的场景内容表示.设加权权重为 $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_r, \cdots, \alpha_R\}$,则图像的多尺度直方图表示为

$$H_{\text{multiscale}} = \alpha_1 \cdot H_1 + \alpha_2 \cdot H_2 + \cdots + \alpha_r \cdot H_r + \cdots + \alpha_R \cdot H_R \tag{13}$$

多尺度直方图 $H_{\text{multi-scale}}$包含了不同尺度下图像场景的视觉信息,与传统的视觉单词分布直方图相比,具有更强的描述能力,因此可以辅助提高场景分类性能.

4 实验结果与分析

4.1 实验设置

本文实验数据库采用 OT 库、FP 库和 LSP 库[1].其中 OT 库包含 8 类场景,分别是 Forest, Mountain, Open Country, Coast, Highway, City, Tall Building 和 Street;FP 库在 OT 库的基础上增加了 Surburb, Bedroom, Living Room, Kitchen 和 Office 共 5 类场景;LSP 库又对 FP 库进行扩充,增加了 Store 和 Industrial 场景,共包含 15 类自然场景.实验中,训练样本集从每类场景图像中随机选取 100 幅图像组成,测试样本集则在剩余图像中每类随机选取 100 幅得到.

实验采用稠密 SIFT 特征[11],视觉词典容量设为 1000;分类器分别采用径向基 SVM 和 K 近邻($k = 5$).SVM 采用"一对多"的策略,核函数为 $K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2)(\gamma > 0)$,参数 γ 通过训练集上的 5-fold 交叉验证得到.

4.2 实验结果分析

本文采用正确率[1]来综合评价场景分类性能.正确率是对各类场景召回率的平均值,召回率和正确率的定义如下所示.

$$\text{召回率} = \frac{\text{被正确分类的图像数}}{\text{该类图像总数}} \times 100\%$$

$$\text{正确率} = \frac{\sum_{i=1}^{C}}{C} \times 100\%$$

其中,C 是场景类别总数.

为验证本文算法的有效性,本节设计了两组实验.首先,验证基于密度的最优 pLSA 模型主题数选择算法的有效性;然后采用自适应主题数选择方法设置 pLSA 模型的最优主题数,在此基础上进一步测试基于多尺度上下文语义信息的场景分类算法性能.

4.2.1 基于密度的最优 pLSA 模型主题数选择算法有效性分析

基于密度的最优 pLSA 模型主题数选择算法是基于"当主题结构的平均相似度最小时,对应的模型最优"这一理论的,它通过最小化主题之间的平均相似度,迭代找出模型的最优主题数.图 3 比较了在不同主题数 K 下,语义主题之间的平均相似度与场景分类正确率的变化曲线.

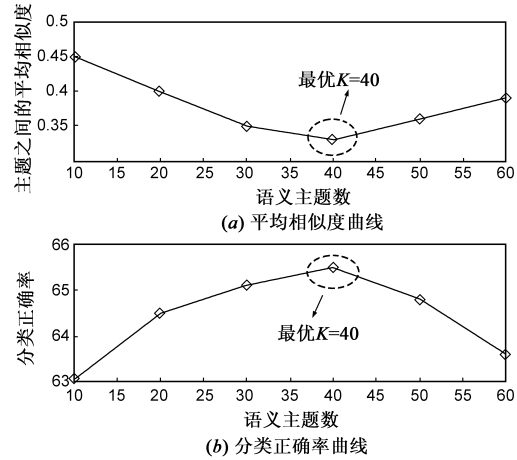


图3 平均相似度曲线和分类正确率曲线在不同K值下的结果对照

从图中可以看出,两条曲线的极值点对应着相同的 K 值.这说明当语义主题之间的平均相似度最小时,模型性能达到最优.因此,通过最小化主题之间的平均相似度搜索最优模型结构和主题数的方法是可行的,从而验证了基于密度的最优 pLSA 模型主题数选择算法的有效性.

4.2.2 基于多尺度上下文语义信息的图像场景分类算法性能分析

为验证本文算法性能,将它与现有的几种典型场景分类算法进行比较,包括 Lazebnik 的空间金字塔(简称 SPM)方法[7]、Feifei 等的变形 LDA 模型(简称 V-LDA)方法[9]、Emrah 等的空间金字塔和 pLSA 模型相结合(简称 SPM-pLSA)的方法[10]、刘硕研等的 pLSA 模型和 Markov 随机场相结合(简称 pLSA + MRF)的方法[11].这些算法的具体参数设置分别为:SPM 中空间金字塔设定为 3 层,分别是 1×1、2×2 和 4×4;V-LDA:LDA 模型主题数设定为 30;SPM-pLSA:空间金字塔设定为 3 层,分别是 1×1、2×2 和 4×4,pLSA 模型主题数设定为 30;pLSA + MRF:pLSA 模型主题数设定为 30,更新视觉单词的算法中 $T=100$,$\varepsilon=0.001$;本文算法:多尺度加权权重 $\alpha_1=0.4$,$\alpha_2=0.3$,$\alpha_3=0.2$,$\alpha_4=0.1$;更新视觉单词的算法中 $T=100$,$\varepsilon=0.001$,pLSA 模型初始主题数设定为 30.

表 1　三种图像库上几种算法的分类平均正确率对比

分类算法	分类器	图像库		
		OT 库	FP 库	LSP 库
SPM	SVM	63.3%	61.6%	58.8%
	KNN	65.1%	58.3%	53.5%
V-LDA	SVM	73.5%	68.7%	64.6%
	KNN	71.2%	66.5%	60.2%
SPM-pLSA	SVM	82.5%	75.4%	69.8%
	KNN	78.8%	70.3%	65.1%
pLSA + MRF	SVM	86.6%	83.2%	79.6%
	KNN	85.1%	80.4%	77.5%
本文算法	SVM	88.4%	84.5%	81.3%
	KNN	83.2%	81.6%	79.8%

从表 1 结果可以看出,本文算法的平均分类正确率要优于其它几种方法,在类别较少的任务(如 OT 库)中性能尤为显著,且 SVM 分类器下的平均正确率要高于 KNN 分类器.对于 OT 数据库,在利用 SVM 分类器的情况下,本文算法取得了最高分类正确率 88.4%,相比 pLSA + MRF 方法,正确率提高了 1.8 个百分点,说明图像的多尺度信息对于改善场景分类性能具有积极的作用;而 pLSA + MRF 方法相比于 SPM 方法、V-LDA 方法和 SPM-pLSA 方法,正确率分别提高了 23.3、13.1 和 4.1 个百分点,说明上下文语义信息在构建更为准确的场景类主题描述过程中具有重要的辅助作用.SPM 方法和 V-LDA 方法则由于没有考虑图像的多尺度和上下文语义信息,在分类正确率上明显低于其它几种方法.这进一步验证了多尺度信息和上下文语义信息对于提高分类性能的有效性.

此外,图 4 还给出了 LSP 库中 15 类自然场景分别采用 SPM 方法、V-LDA 方法、SPM-pLSA 方法、pLSA + MRF 方法以及本文方法的准确率对比.

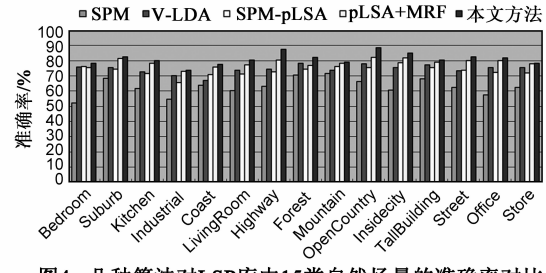


图4　几种算法对LSP库中15类自然场景的准确率对比

从图 4 中可以看出,本文方法在分类性能上较之其它几种方法均有所提高,尤其对于 Highway 和 OpenCountry 场景,分类性能提高幅度最大.这主要因为 Highway 场景与天空、树林等图像区域具有较高的语义共生概率,OpenCountry 场景与草地、天空等图像区域具有较高的语义共生概率;并且 Highway 和 OpenCountry 场景中有的对象结构尺度较大,有的对象结构尺度较小,即场景的视觉信息分布在多个尺度中.本文方法在考虑图像视觉特征相似性的同时还考虑了图像区域的多尺度特性和语义概念之间的共生性,这进一步增加了局部特征所携带的上下文信息,从而可以生成更准确的视觉单词,提高该类场景的分类性能.但对于 Industrial 场景,由于其上下文语义共生特性并不明显;因此,本文算法对其分类性能提高幅度较小,需要进一步挖掘图像的其它信息才能有效改善该类场景的分类性能.

5　结论

本文提出了一种基于多尺度上下文语义信息的图像场景分类算法.首先对图像进行多尺度分解,从图像的多个尺度提取视觉信息,再利用基于密度的自适应选择方法确定 pLSA 模型最优主题数;然后,结合改进的 pLSA 模型和 Markov 随机场共同挖掘图像块的上下文语义共生信息,构造出更有意义的视觉单词;最后,加权连接多个尺度的视觉单词分布直方图得到图像的多尺度直方图表示,进而结合 SVM 分类器实现场景分类.实验结果表明,本文算法能有效利用图像的多尺度和上下文语义信息,提高视觉单词的语义准确性,改善

图像场景分类性能.

参考文献

[1] 解文杰.基于中层语义表示的图像场景分类研究[D].北京:北京交通大学,2011.
Xie Wenjie. Research on middle semantic representation based image scene classification[D]. Beijing: Beijing Jiaotong University, 2011. (in Chinese)

[2] Vailaya A, Figueiredo M, Jain A, et al. Content-based hierarchical classification of vacation images[A]. IEEE International Conference on Multimedia Computing and Systems[C]. Washington: IEEE Computer Society, 1999. 518 – 523.

[3] Rachid Benmokhtar, Benoit Huet, et al. Low-level feature fusion models for soccer scene classification[A]. IEEE International Conference Multimedia and Expo[C]. Hannover: IEEE Press, 2008: 1329 – 1332.

[4] R Marée, P Denis, L Wehenkel, et al. Incremental indexing and distributed image search using shared randomized vocabularies [A]. Proc of MIR10[C]. New York: ACM Press, 2010. 91 – 100.

[5] J C Van Gemert, C J Veenman, A W M Smeulders, et al. Visual word ambiguity[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 7(32): 1271 – 1283.

[6] Jing-yan Wang, Yong-ping Li, Ying Zhang, et al. Bag-of-features based medical image retrieval via multiple assignment and visual words weighting[J]. IEEE Transactions on Medical Imaging, 2011, 30(11): 1996 – 2011.

[7] Lazebnik S, Schmid C, Ponce J. Beyond bag of features: spatial pyramid matching for recognizing natural scene categories[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. New York: IEEE Computer Society, 2006. 2169 – 2178.

[8] Bosch A, Munoz X. Object andscene classification: what does a supervised approach provide us? [A]. International Conference on Pattern Recognition[C]. Hong Kong: IEEE Press, 2006. 773 – 777.

[9] Fei-fei Li, Perona P. Abayesian hierarchical model for learning natural scene categories[A]. IEEE International Conference on Computer Vision and Pattern Recognition[C]. San Diego: IEEE Computer Society, 2005. 524 – 531.

[10] Emrah E, Nafiz A. Sceneclassification using spatial pyramid of latent topics[A]. 20th International Conference on Pattern Recognition[C]. Istanbul: IEEE Press, 2010. 3603 – 3606.

[11] 刘硕研,须德,冯松鹤,刘镝,裘正定.一种基于上下文语义信息的图像视觉单词生成算法[J].电子学报,2010,38(5):1156 – 1161.
Liu S Y, Xu D, Feng S H, Liu D, Qiu Z D. A novel visual words definition algorithm of image patch based on contextual semantic information[J]. Acta Electronica Sinica, 2010, 38(5): 1156 – 1161. (in Chinese)

[12] Lindeberg T. Scale-Space Theory in Computer Vision[M]. Germany: Springer, 1994.

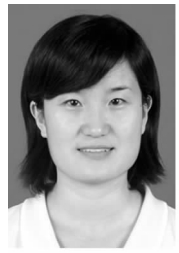
[13] Lindeberg T. Scale-space theory: a basic tool for analyzing structures at different scales[J]. Journal of Applied Statistics, 1994, 21(2): 224 – 270.

[14] Hoffmann T. Probabilisticlatent semantic analysis[A]. Uncertainty in Artificial Intelligence[C]. Stockholm: Morgan Kaufmann, 1999. 289 – 296.

[15] 傅兴玉,尤红建,付琨.基于改进 Markov 随机场的高分辨率 SAR 图像建筑物分割算法[J].电子学报,2012,40(6):1141 – 1147.
Fu X Y, You H J, Fu K. Building segmentation from high-resolution SAR images based on improved markov random field [J]. Acta Electronica Sinica, 2012, 40(6): 1141 – 1147. (in Chinese)

[16] 曹娟,张勇东,李锦涛,唐胜.一种基于密度的自适应最优 LDA 模型选择方法[J].计算机学报,2008,31(10):1780 – 1787.
Cao J, Zhang Y D, Li J T, Tang S. A method of adaptively selecting best LDA model based on density[J]. Chinese Journal of Computers, 2008, 31(10): 1780 – 1787. (in Chinese)

作者简介



张瑞杰 女,1984 年生于河南郑州,博士,解放军信息工程大学四院讲师,主要研究方向为网络舆情分析、图像场景分类与检索等.
E-mail: rjz_wonder@163.com

李弼程 男,1971 年生于湖南衡阳,博士生导师,解放军信息工程大学信息系统工程学院教授,主要研究方向为智能信息处理、数据挖掘等.
E-mail: lbclm@163.com

魏福山 男,1983 年生于甘肃武威,博士,解放军信息工程大学四院讲师,主要研究方向为密码学、信息安全等.
E-mail: weifs831020@163.com