基于全局优化策略的场景分类算法*

金泰松[1] 李玲玲[2] 李翠华[1]

[1](厦门大学 信息科学与技术学院 厦门 361005)
[2](郑州航空工业管理学院 计算机科学与应用系 郑州 450015)

摘 要 提出一种基于全局优化策略的场景分类算法. 该算法基于整幅图像提取全局场景特征——空间包络特征. 从图像块中提取视觉单词,且定义隐变量表示该视觉单词语义,然后引入隐状态结构图描述整幅图像的视觉单词上下文;在场景分类策略上,构造由相容函数组成的目标函数,其中相容函数度量全局场景特征、隐变量与场景类别标记的相容度,通过求解目标函数的全局最优解推断图像的场景类别标记. 在标准场景图像库上的对比实验表明该算法优于当前有代表性的场景分类算法.

关键词 图像解析,场景分类,函数优化,视觉单词
中图法分类号 TP 391

Scene Classification Based on Global Optimized Framework

JIN Tai-Song[1], LI Ling-Ling[2], LI Cui-Hua[1]
[1](*School of Information Science and Technology, Xiamen University, Xiamen* 361005)
[2](*Department of Computer Science and Application, Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou* 450015)

ABSTRACT

A scene classification algorithm based on global optimized framework is proposed. Firstly, the global scene feature named spatial envelop is obtained from the whole image, the visual word of each image block is extracted, and latent variable is defined to represent the semantic feature of the extracted visual word. Secondly, the structure graph of latent state is introduced to represent the context of visual words. In respect to scene classification strategy, objective function consisting of different potential functions is constructed in which potential functions are defined to measure the relevance of the variables including global scene feature, latent variables and scene category. Finally, the scene category of the image is determined when the global optimized solution of objective function is obtained. The experiments on the standard dataset demonstrate that the proposed algorithm achieves better results than the state-of-the-art algorithms.

* 国家自然科学基金项目(No. 41171341)、航空科学基金项目(No. 20125168001)、教育部新世纪优秀人才支持计划项目(No. NCET-09-0126)、教育部博士点基金项目(No. 20110121110020)、河南省科技创新人才杰出青年项目(No. 114100510006)、福建省自然科学基金项目(No. 2011J01365)、郑州市科技创新人才培育计划项目(No. 10PTGG342-1)资助
收稿日期:2012-06-13;修回日期:2012-12-24
作者简介 金泰松(通讯作者),男,1978 年生,博士,讲师,主要研究方向为计算机视觉、遥感影像分析. E-mail: jintaisong@xmu.edu.cn. 李玲玲,女,1973 年生,副教授,博士后,主要研究方向为计算机视觉、图像处理. 李翠华,男,1960 年生,教授,博士生导师,主要研究方向为计算机视觉、模式识别.

Key Words　Image Analysis, Scene Classification, Function Optimization, Visual Word

1　引　言

图像场景分类在计算机视觉、模式识别领域有广泛的应用,特别是 2006 年召开的首次场景理解研讨会上明确提出"场景分类是图像理解的一个新的有前途的研究方向",且随后召开的场景理解研讨会再次确定场景分类的重要地位后,现已成为研究热点. 近来,基于中层语义特征的方法受到研究人员的广泛关注. 该类方法大多以视觉单词为基础,实现图像内容的视觉词包表示,其上下文的提取和利用方面得到人们的重视[1-4]. 这类方法利用视觉单词的上下文信息能部分解决不同视觉单词表示相同语义的问题,场景分类的性能亟待提高.

由于每幅图像由多个视觉单词表示,使得图像在某种程度上具备文本的特点,因此文本分析中主题分析模型也被用到场景分类中来,通过引入潜在主题维将图像的高维视觉词包表示映射到低维的主题表示[5-7].

以视觉词包和主题模型为代表的场景分类方法大多依赖于视觉词典的构建,对于复杂场景图像,图像内容的复杂多样及噪声直接影响视觉词典的生成效率和表示,且特征维数高及对初始值敏感等缺点,降低了方法的性能和适用价值[8],因此融合全局特征与视觉词包的图像内容表述方法受到研究人员的关注[9]. 此外,少数研究者尝试跳脱视觉词包模型,使用生物启发特征(Biologically Inspired Feature, BIF),并通过流形学习降低特征维数[10].

隐条件随机场、隐马尔科夫随机场及各种优化手段的出现为解决该问题提供新的思路,且优化策略已广泛应用到视觉的不同领域[11-12],但到目前为止在场景分类领域还没有应用的相关报道. 本文基于优化理论的成功及融合全局特征与视觉词包表示进行图像内容表述的优点[9]提出一种基于优化策略的场景分类算法. 该方法引入隐变量,并通过隐状态结构图描述视觉单词的语义上下文,隐变量的使用避免训练图像语义对象的标记过程,只需标定每幅图像的场景类别,提高训练效率. 同时,为弥补视觉单词导致的图像信息丢失,使用基于整个图像的全局场景特征——空间包络特征[13-14]对图像进行整体描述,然后定义度量全局特征、视觉单词、隐状态与场景类别相容度的相容函数作为全局优化的目标函数,图像场景分类结果依据目标函数的全局最优解确定. 同以往工作相比,本文算法使用优化策略将图像内容表述和场景分类有机地融合成一个整体,且通过改变全局优化函数中不同相容函数的定义,可较方便地对模型进行扩展,这为人们研究场景分类问题提供一种新的解决方案. 在标准图像集上进行的实验证明了本文算法的有效性.

2　图像的观测数据提取

为得到场景图像的语义描述,需要建立高层场景语义描述与低层视觉特征之间的联系. 基于视觉词包的方法通过视觉单词分布对图像内容进行描述. 基于整个图像提取的全局场景特征对图像进行整体描述. 视觉单词和空间包络特征分别从图像的局部和全局的角度对图像内容进行描述,已有研究证明[9]将全局特征与视觉词包表示进行融合的优点,因此本文选取"视觉单词"和全局空间包络特征作为从图像中提取的观测数据.

2.1　空间包络特征

心理学实验已经证明: 在一定条件下,人类进行场景识别能在不知道任何语义对象的条件下进行,且表明基于五维向量(导航性、开阔性、崎岖性、延伸性、粗糙性)组成的空间包络特征对自然图像做一个粗分类[13-14]. 该特征是 Oliva 等[13]提出的描述环境场景的全局场景特征,能够实现图像特征和场景语义之间的映射. 本文将空间包络特征作为描述图像场景的全局特征. 具体提取方法是先将图像变换到灰度空间,然后输入 Gabor 滤波器,按照文献[13]方法得到图像的空间包络特征.

2.2　基于 SIFT 特征的视觉单词提取

与全局场景特征相比,局部特征对遮挡和空间变化有较好的鲁棒性,能有效描述图像内容. 在场景分类中用的最多的局部特征是 SIFT 特征,该特征对特征点尺度、位置、旋转和光照等因素的变化不敏感,较适合对场景图像进行描述. 提取该特征通常包含两个步骤: 特征点检测和特征表达. 现有 SIFT 特征提取根据特征点检测不同主要分为两类: 1) 检测不同图像尺度空间的极值点作为特征点进行特征描述[15]; 2) 将图像进行均匀分块形成固定网格,每次将网格蛇行移动 N(通常为 8 或 16)个像素,由网格的位置来确定特征点,进行特征描述[16]. 场景分类领域中进行的相关实验已证实,尽可能采用较大数量图像特征点进行图像的各个区域特征描述,能获

得较高的分类准确率[17]. 只采用极值点进行特征描述忽略较多场景图像的内容细节,因此,本文采用对图像进行均匀分块的 SIFT 特征提取方法. 首先对图像进行网状分割,网格采样间隔为 8 个像素,然后在每个网格采样点提取其周围 16×16 像素的区域进行特征描述,将其划分为 4×4 个子块,分别在每个子块上计算 8 个方向的梯度直方图,最后产生的 SIFT 特征向量就有 $16 \times 8 = 128$ 维,对应图像块大小是 16×16 像素. 然后使用 K 均值聚类算法对训练图像集中所有图像块提取的 SIFT 特征进行聚类,每个聚类中心对应一个视觉单词,生成一个由视觉单词组成的视觉词典[3].

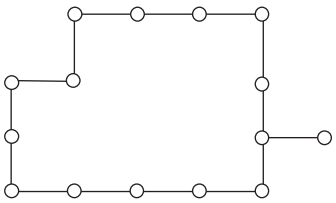
对一幅图像,计算每个图像块的 SIFT 特征与视觉词典中不同视觉单词对应 SIFT 特征的欧氏距离,以距离最小的视觉单词表示该图像块,整个图像表示为 $\boldsymbol{x}' = (x_1', \cdots, x_i', \cdots, x_n')$,其中,$n$ 为图像中图像块的个数,x_i' 是第 i 个图像块的视觉单词标号. 定义隐变量作为元素的向量 $\boldsymbol{r} = (r_1, \cdots, r_i, \cdots, r_n)$,其中,$r_i \in R$ 是第 i 个图像块对应的隐变量,表示第 i 个视觉单词的语义; R 是隐状态集合,表示视觉单词所有可能的语义. 定义图像场景类别集 $L = \{1, \cdots, i, \cdots, p\}$,其中,$i$ 表示场景类别标号,共有 p 个场景类别.

现有视觉词包表示大多忽略视觉单词之间的语义联系,基于语义对象的场景分类方法需要提取图像的语义对象,然而语义对象识别本身就是一个待解决的问题,再加上图像分割自身的缺陷,这类方法对复杂场景的分类能力不是很理想. 本文注意到场景图像中不同语义对象之间具有一定的依赖关系. 对一幅街道场景图像,小汽车常和公路一起出现,而语义对象的依赖性可表示为不同视觉单词上下文. 隐变量表示方法的出现,使得现有方法不必直接提取语义对象,上下文信息通过隐变量上下文关系表示,本文定义该图为隐状态结构图.

2.3 隐状态结构图的构造

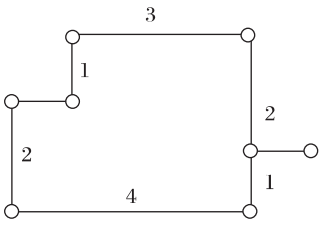
隐状态结构图 $G(V, E)$ 由隐变量顶点集 V 和边集 E 组成. 图像中每个图像块为一个隐变量顶点,共有 n 个顶点. 本文根据图像块空间位置(图像块的相邻图像块共有四个,分别是上相邻、下相邻、左相邻和右相邻图像块)的不同,比较图像块与四个相邻图像块视觉单词同异的准则定义图的边. 如果相邻两个图像块的视觉单词不相同,连接该图像块与相邻图像块的隐变量顶点作为一条边; 否则,不连接两个顶点. 不断连接隐状态结构图中不同相邻顶点定义图的边集.

隐状态结构图可能包含环路,可能属于子图,也可能图本身是一个环路结构. 为消除图的环路,对图中的每个环路结构,本文构造环路结构子图(该子图保留环路结构图中的分支顶点、转角顶点和端点顶点作为顶点集,连接环路结构中直线通路的顶点作为边集,其边的权定义为直线通路上顶点的个数减一),然后对环路结构子图计算最小生成树. 图 1 是隐状态结构图的各种子图.



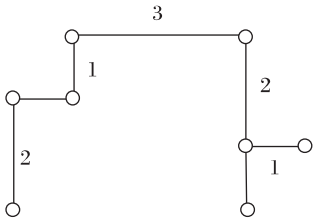
(a) 环路结构图

(a) Graph of loop structure



(b) 环路结构子图

(b) Subgraph of loop structure



(c) 环路结构子图的最小生成树

(c) Minimum spanning tree for subgraph of loop structure

图 1 隐状态结构图的各种子图

Fig. 1 Subgraphs of latent state structure graph

3 场景分类的全局优化模型

本文提出的基于优化策略的场景分类算法框架见图 2. 采用优化策略的关键是设计合适的目标函数,通过求解函数的全局最优解进行场景分类. 本文采用的目标函数由度量观测数据、隐变量状态和场景类别标记相容度的相容函数组成,表示隐状态结构、观测数据和场景类别之间的符合程度.

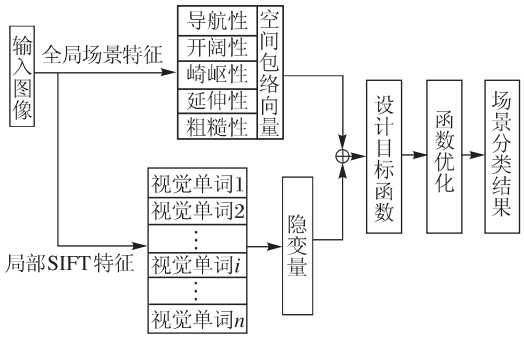


图 2　模型框架

Fig. 2　Framework of the proposed model

3.1　全局优化模型的目标函数

目标函数包括 3 个相容函数(Potential Function): 即定义在观测空间和隐状态空间的相容函数度量输入图像与隐变量的相容度; 定义在隐状态空间和场景类别标记空间的相容函数度量隐变量与场景类别的相容度; 定义在观测空间和场景类别标记空间的相容函数度量输入图像与场景类别的相容度.

定义 $\boldsymbol{x}=\boldsymbol{x}'\cup\boldsymbol{x}''$ 是输入的观测向量,其中 $\boldsymbol{x}'$ 是局部观察向量,是以图像块的视觉单词标号作为元素的向量,$\boldsymbol{x}''$ 是全局空间包络特征向量,$l\in L$ 是场景类别标记,则全局优化模型的目标函数定义为

$$f(\boldsymbol{x},\boldsymbol{r},l;w)=f_1(\boldsymbol{x}',\boldsymbol{r};\alpha)+f_2(\boldsymbol{r},l;\beta,\gamma)+f_3(\boldsymbol{x}'',l;\delta),$$

其中 $w=\{\alpha,\beta,\gamma,\delta\}$ 是模型参数; 相容函数 $f_1(\boldsymbol{x}',\boldsymbol{r};\alpha)$ 表示局部观测向量 $\boldsymbol{x}'$ 与以隐变量作为元素的向量 $\boldsymbol{r}$ 的相容度,相容函数 $f_2(\boldsymbol{r},l;\gamma)$ 表示 $\boldsymbol{r}$ 与场景类别 l 的相容度,相容函数 $f_3(\boldsymbol{x}'',l;\delta)$ 表示全局空间包络五维向量与 l 的相容度.

接下来定义相容函数. 首先定义示性函数 $I_{\{\cdot\}}$,当 $\{\cdot\}$ 中条件满足时,该函数取值为 1; 否则,取值为 0.

相容函数

$$f_1(\boldsymbol{x}',\boldsymbol{r};\alpha)=\sum_{i\in V}\sum_{a\in R}\alpha_a\cdot I_{\{r_i=a\}}\cdot\boldsymbol{x}'_i,$$

相容度由参数 α 度量,其中 α_a 表示隐状态为 a 的隐变量与第 i 个图像块对应视觉单词的相容度; 相容函数

$$f_2(\boldsymbol{r},l;\beta,\gamma)=f'_2(r,l;\beta)+f''_2(r,l;\gamma),$$

其中,

$$f'_2(\boldsymbol{r},l;\beta)=\sum_{i\in V}\sum_{a\in L}\sum_{b\in R}\beta_{a,b}\cdot I_{\{l=a\}}\cdot I_{\{r_i=b\}}$$

表示单个隐变量与场景类别的相容度,由参数 β 度量,其中 $\beta_{a,b}$ 表示隐状态为 b 的隐变量与场景类别标记为 a 的相容度;

$$f''_2(\boldsymbol{r},l;\gamma)=\sum_{(i,j)\in E}\sum_{a\in L}\sum_{b\in R}\sum_{c\in R}\gamma_{a,b,c}\cdot I_{\{l=a\}}\cdot I_{\{r_i=b\}}\cdot I_{\{r_j=c\}}$$

表示隐变量对与场景类别的相容度,由参数 γ 度量,其中 $\gamma_{a,b,c}$ 表示两个隐状态分别为 b 和 c 的隐变量对与场景类别标记为 a 的相容性; 相容函数

$$f_3(\boldsymbol{x}'',l;\delta)=\sum_{a\in L}\delta_a\cdot I_{\{l=a\}}\cdot\boldsymbol{x}'',$$

相容度由参数 δ 度量,其中 δ_a 表示场景类别为 a 与空间包络特征的相容度.

图像的场景分类转化为全局优化问题:

$$l^*=\arg\max_{l\in L}\max_{\boldsymbol{r}}f(\boldsymbol{x},\boldsymbol{r},l;w).$$

对给定的场景类别标记 l,使用最小生成树算法消除隐状态结构图的环路结构后,目标函数的最优解可通过置信度传播方法[18]计算. 该算法通过在不同顶点之间进行消息的迭代传递,当传输的消息趋于稳定时,顶点的置信度趋于稳定,迭代终止. 通过枚举所有可能的场景类别标记获目标函数的全局最优解,将具有全局最优解的场景类别标记推断为图像的类别标记.

3.2　模型的参数估计

存在 M 个训练图像 $\{(\boldsymbol{x}^{(m)},l^{(m)})\}_{m=1}^{M}$ 的图像库,模型中需要估计的参数由 4 部分组成,即 $w=\{\alpha,\beta,\gamma,\delta\}$. 由于隐变量的存在,参数估计是包含隐变量的约束优化问题[19]:

$$\min_w\left(\frac{1}{2}\|w\|^2+C\sum_{n=1}^{M}\xi_n\right),$$

$$\text{s.t.}\ \max_{\boldsymbol{r}}f(\boldsymbol{x}^{(n)},\boldsymbol{r},l^{(n)})-\max_{l\in L}\max_{\boldsymbol{r}}f(\boldsymbol{x}^{(n)},\boldsymbol{r},l)\geqslant\Delta(l-l^{(n)})-\xi_n,\forall n,\forall l,$$

其中,$\Delta(\cdot)$ 是 0–1 损失函数,它控制着训练精度; C 是惩罚因子. 使用拉格朗日乘法将约束优化转化为无约束优化问题:

$$\min_w\left(\frac{1}{2}\|w\|^2+C\sum_{n=1}^{M}(P^n-Q^n)\right),$$

$$P^n=\max_{l\in L}\max_{\boldsymbol{r}}(\Delta(l-l^{(n)})+f(\boldsymbol{x}^{(n)},\boldsymbol{r},l)),$$

$$Q^n=\max_{\boldsymbol{r}}f(\boldsymbol{x}^{(n)},\boldsymbol{r},l^{(n)}),$$

其中,$C\sum_{n=1}^{M}(P^n-Q^n)$ 是非凸函数,最优化求解是非凸优化问题. 本文采用非凸束优化方法计算该函数的全局最优解[12,20]. 该方法通过迭代一系列二次函数逼近非凸函数,在每个迭代步根据函数在 w 处的次梯度构造新的二次函数逼近原函数,因此该方法的关键是计算函数在点 w 处的次梯度,即 $\partial_w P^n$ 和 $\partial_w Q^n$.

从 $P^n=\max_{l\in L}\max_{\boldsymbol{r}}(\Delta(l-l^{(n)})+f(\boldsymbol{x}^{(n)},\boldsymbol{r},l))$ 定义可知: 求解最大值的函数表达式由两部分组成:

$f(x^{(n)},r,l)$ 和损失函数 $\Delta(l-l^{(n)})$；其中 $f(x^{(n)},r,l)$ 的最大值由置信度传播算法获得，$\Delta(l-l^{(n)})$ 可通过列举所有可能的场景类别标记计算，从而得到 P^n 关于参数 w 的表达式，然后 P^n 在点 w 处的次梯度. 同理，可计算出 $\partial_w Q^n$.

3.3 基于优化策略的场景分类方法

本文提出的场景分类算法分为训练和测试两个部分.

训练步骤如下.

step 1 对每幅训练图像进行网状分割，网格采样间隔为 8 个像素，在每个网格采样点提取其周围 16 × 16 像素的区域进行 SIFT 特征描述，得到 SIFT 特征向量集；然后使用 K 均值算法对 SIFT 特征进行聚类. 所有的聚类中心得到视觉词典，其每个聚类中心是一个视觉单词.

step 2 计算每个图像块的 SIFT 特征与视觉词典不同视觉单词对应 SIFT 特征的欧氏距离，以距离最小的视觉单词标号表示该图像块，整幅图像用视觉单词标号向量表示.

step 3 采用 2.3 节方法构造图像的隐状态结构图，然后提取图像的全局空间包络特征，在整个训练图像库上，使用 3.2 节方法进行参数估计.

测试步骤如下.

对一幅测试图像，采用训练 step 2，step 3 提取图像的空间包络特征、视觉单词标号向量及隐状态结构，然后对给定场景类别标记使用置信度传播方法[18]计算函数的局部最优解，再通过枚举所有杨景类别标记计算目标函数的全局最优解，将具有全局最优解的场景类别标记推断为图像的场景类别.

4 实验与结果分析

本文选择文献[4]使用的场景图像库作为基准图像库，简称该图像库为 LSP 图像库. 该图像库总共有4 485 张图像，是目前为止最大的自然场景图像库，包括森林、海岸等自然场景图像及客厅、办公室等人造场景图像，每类场景大约有 200 ~ 400 幅图像，共 15 个类别(图像的大小为 256 × 256). 实验对图像库图像进行 5 次随机划分，生成相应的训练集和测试集，每次从各类场景图像随机抽取 100 幅，共 1 500 幅图像作为训练集样本，图像库中剩余图像作为测试集样本，分别计算每次划分的分类准确率，5 次划分得到分类准确率的均值作为算法最终的平均分类准确率.

影响场景分类性能表现的重要参数是视觉词典的容量. 实验中隐状态数设定为 32，然后选择不同的视觉词典容量，分析平均分类准确率随着视觉词典容量的变化趋势. 图 3 是采用不同视觉词典容量场景分类算法的分类准确率.

从图 3 看出，在视觉词典容量从小到大的变化过程中，分类准确率在开始阶段明显增加，增加到某个值后，随着视觉词典容量的继续增加，分类准确率变化不大. 这是因为视觉词典容量小时，不同视觉特征可能被表述为相同的视觉单词，产生过聚类，通过提高视觉词典容量能提高场景分类性能；然而过大的视觉词典容量可能导致每个图像块被分配不同的视觉单词，并不能提高场景分类的性能，而且它将引起图像表示的高维现象，增加算法的计算复杂性. 因此本文选取的视觉词典容量是 600.

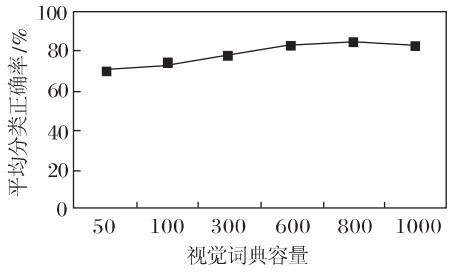


图 3 不同容量视觉词典下的场景分类结果

Fig. 3 Scene classification results of different sizes of visual dictionary

目标函数由相容函数组成，不同相容函数组合可构造不同的目标函数. 为研究相容函数在场景分类中所起的作用，本文实验 4 种基于优化策略的场景分类算法，它们区别在目标函数的定义上，其它都相同. 其上下文无关模型的目标函数是 $f_1(x',r;\alpha)+f_2'(r,l;\beta)$；上下文相关模型的目标函数 $f_1(x',r;\alpha)+f_2(r,l;\beta,\gamma)$；全局模型的目标函数是 $f_3(x'',l;\delta)$；本文模型的目标函数是 $f(x,r,l;w)$. 基于优化模型的 4 种场景分类算法的平均分类准确率为：上下文无关模型为 68%，上下文相关模型为 72%，全局模型为 54%，本文模型为 85%.

从平均分类准确率看出，基于上下文相关模型场景分类算法的平均分类准确率高于基于上下文无关模型和基于全局模型的场景分类算法，基于本文模型的场景分类算法取得最高图像的场景类别，基于本文模型的场景分类算法融合全局特征、视觉单词上下文，能更好地区分不同类别的场景图像.

下面给出一个场景分类实例. 图 4 是一幅城市场景图像，图中包含建筑、公路和小汽车等语义对

象. 本文定义

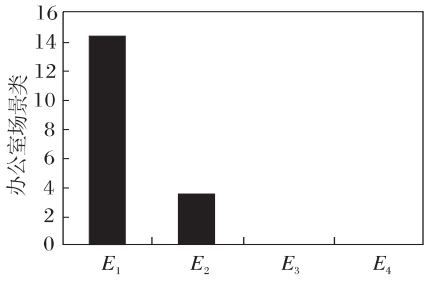
$$E_1 = f_1(\boldsymbol{x}', \boldsymbol{r}; \alpha),\ E_2 = f_2'(\boldsymbol{r}, l; \beta),$$

$$E_3 = f_2''(\boldsymbol{r}, l; \gamma),\ E_4 = f_3(\boldsymbol{x}'', l; \delta);$$

则基于 4 种模型的场景分类方法的分类结果与对应 E_2, E_3, E_4, E_1 的取值见图 5.

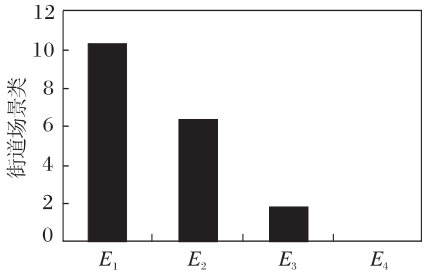
图 4　一幅城市场景图像

Fig. 4　An inside city scene image



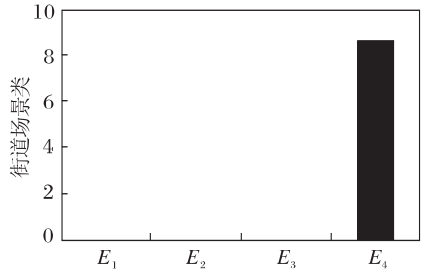
(a) 上下文无关模型

(a) Context-free model



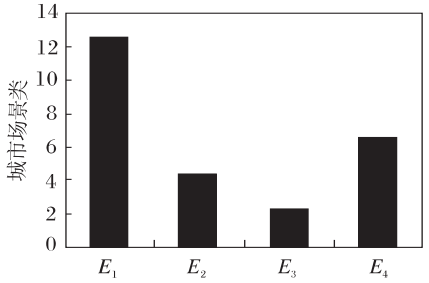
(b) 上下文相关模型

(b) Context dependent model



(c) 全局模型

(c) Global model



(d) 本文模型

(d) The proposed model

图 5　分类结果和相容函数的取值

Fig. 5　Classification results and the values of potential functions

从图 5 看出,基于上下文无关模型的场景分类算法将城市场景图像错分为办公室场景类; 基于上下文相关模型的场景分类算法将城市场景错分为街道场景类; 基于本文模型的场景分类算法正确分类该图像为声调场景类; 且从相容函数的取值情况可知,与 E_2, E_3, E_4 相比,E_1 对场景分类取较大作用.

图 6 是基于本文模型场景分类算法的隐状态分布直方图. 从图中看出: 目标函数的全局最优解中隐变量的取值情况中,不同隐状态主要分布在 5、16 和 23 维上,这 3 维与城市场景图像中的 3 种语义对象相符合,本文认为这较好地描述了场景图像的内容.

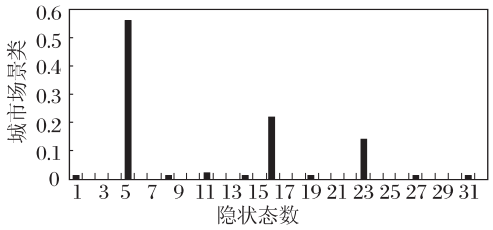


图 6　本文模型的隐状态分布直方图

Fig. 6　Hidden state distribution histogram of the proposed model

此外,通过对比实验比较本文算法和几种代表性场景分类算法的分类准确率,包括基于视觉单词上下文的场景分类算法[3]、基于金字塔匹配的场景分类算法[4]、基于全局空间包络特征的场景分类算法[13]和基于 pLSA 主题模型的场景分类算法[21].

文献[3]根据 SIFT 特征提取视觉单词,通过不同层次信息构建视觉单词上下文表示,使用支持向量机进行场景分类. 文献[4]将特征空间的金字塔匹配引入图像空间,提出一种基于金字塔匹配的场景分类算法. 文献[13]利用全局空间包络特征表示场景图像内容,使用支持向量机进行场景分类. 文献

[21] 利用 pLSA 主题模型表示场景图像内容，使用 K 近邻分类器进行场景分类.

这 4 种算法中，文献 [13] 是基于全局特征的场景分类方法，其它 3 种算法分别从视觉单词上下文、视觉单词的金字塔模型和基于视觉单词分布生成语义主题的角度得到场景内容的图像表示，利用现有分类器进行场景分类. 与这 4 种算法不同的是本文将函数优化引入场景分类，通过设计目标函数，并求解函数的全局最优解实现场景分类. 不同算法在基准图像库上的平均分类准确率如下: 文献 [3] 为 83%、文献 [4] 为 81%、文献 [13] 为 56%、文献 [21] 为 58%，本文算法为 85%，这表明将函数优化应用于场景分类是可行的.

本文所有实验在 CPU 为 Pentuim Ⅳ 3.2GB，内存 1GB 的台式机上进行，程序使用 VC++6 编程实现. 本文算法对每幅图像的平均训练花费是 4.3s，主要用在视觉词典生成和参数估计步骤，每幅测试图像的平均处理时间是 0.2s，主要花费在优化计算上.

5 结 束 语

本文提出一种基于优化策略的场景分类算法. 该算法通过构造目标函数，并计算函数全局最优解推断图像的场景类别. 在现有标准图像库上进行的实验表明: 本文算法取得较好的场景分类效果，且由于参数估计避免大量手工标注图像局部语义的过程，因此具有较强的实用价值.

致 谢 感谢李丕范在实验数据处理和分析方面所做的工作.

参 考 文 献

[1] Jiang Yue, Wang Runsheng, Wang Cheng. Scene Classification with Context Pyramid Features. Journal of Computer-Aided Design & Computer Graphics, 2010, 22(8): 1366 - 1373 (in Chinese)
(江 悦，王润生，王 程. 采用上下文金字塔特征的场景分类. 计算机辅助设计与图形学学报, 2010, 22(8): 1366 - 1373)

[2] Wang Yushi, Gao Wen. Kernel-Based Image Classification Using the Context of Visual Words. Journal of Image and Graphics, 2010, 15(4): 607 - 616 (in Chinese)
(王宇石，高 文. 用基于视觉单词上下文的核函数对图像分类. 中国图象图形学报, 2010, 15(4): 607 - 616)

[3] Qin Jianzhao, Yung N H C. Scene Categorization via Contextual Visual Words. Pattern Recognition, 2010, 43(5): 1874 - 1888

[4] Lazebnik S, Schmid C, Ponce J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA, 2006, Ⅱ: 2169 - 2178

[5] Bosch A, Muñoz, Zisserman A. Scene Classification Using a Hybrid Generative/Discriminative Approach. IEEE Trans on Pattern Analysis and Machine Intelligence, 2008, 30(4): 712 - 727

[6] Rasiwasia N, Vasconcelos N M. Scene Classification with Low-Dimensional Semantic Spaces and Weak Supervision // Proc of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, USA, 2008: 1 - 6

[7] Bosch A, Muñoz X, Marti R. A Review: Which is the Best Way to Organize/Classify Images by Content? Images and Vision Computing, 2007, 25(6): 778 - 791

[8] Yang Dan, Li Bo, Zhao Hong. An Adaptive Algorithm for Robust Visual Codebook Generation and Its Natural Scene Categorization Application. Journal of Electronics & Information Technology, 2010, 32(9): 2139 - 2144 (in Chinese)
(杨 丹，李 博，赵 红. 鲁棒视觉词汇本的自适应构造与自然场景分类应用. 电子与信息学报, 2010, 32(9): 2139 - 2144)

[9] Jiang Y, Chen J, Wang R S. Fusing Local and Global Information for Scene Classification. Optical Engineering, 2010, 49(4): 1 - 10

[10] Song D J, Tao D C. Biologically Inspired Feature Manifold for Scene Classification. IEEE Trans on Image Processing, 2010, 19(1): 174 - 184

[11] Xiao Chunxia, Liu Shu, Lin Chengchun, *et al*. A Global Space-Time Optimization Framework for Video Completion. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(9): 1204 - 1211
(肖春霞，刘 舒，林成春，等. 基于时空全局优化的视频修复. 计算机辅助设计与图形学学报, 2008, 20(9): 1204 - 1211)

[12] Wang Y, Mori G. A Discriminative Latent Model of Object Classes and Attributes // Proc of the 11th European Conference on Computer Vision, Crete, Greece, 2010: 155 - 168

[13] Oliva A, Torralba A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. International Journal of Computer Vision, 2001, 42(3): 145 - 175

[14] Greene M R, Oliva A. Recognition of Natural Scenes from Global Properties: Seeing the Forest without Representing the Trees. Cognitive Psychology, 2009, 58(2): 137 - 176

[15] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 2004, 60(2): 91 - 110

[16] Jurie F, Triggs B. Creating Efficient Codebooks for Visual Recognition // Proc of 10th IEEE International Conference on Computer Vision, 2005, Ⅰ: 604 - 610

[17] Li F F, Pieriro P. A Bayesian Hierarchical Model for Learning Natural Scene Categories // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA, 2005, Ⅱ: 524 - 531

[18] Felzenszwalb P F, Huttenlocher D P. Efficient Belief Propagation for Early Vision // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington DC, USA, 2004: 261 - 268

[19] Felzenszwalb P F, McAllester D, Ramanan D. A Discriminatively Trained, Multiscale, Deformable Part Model // Proc of IEEE Conference on Computer Vision and Pattern Recognition. Alaska, USA, 2008: 1 - 8

[20] Do T M T, Artières T. Large Margin Training for Hidden Markov Models with Partially Observed States // Proc of the 26th International Conference on Machine Learning. New York, USA, 2009: 265 - 272

[21] Bosch A, Zisserman A, Muñoz A. Scene Classification via pLSA // Proc of the 9th European Conference on Computer Vision. Graz, Austria, 2006: 517 - 530