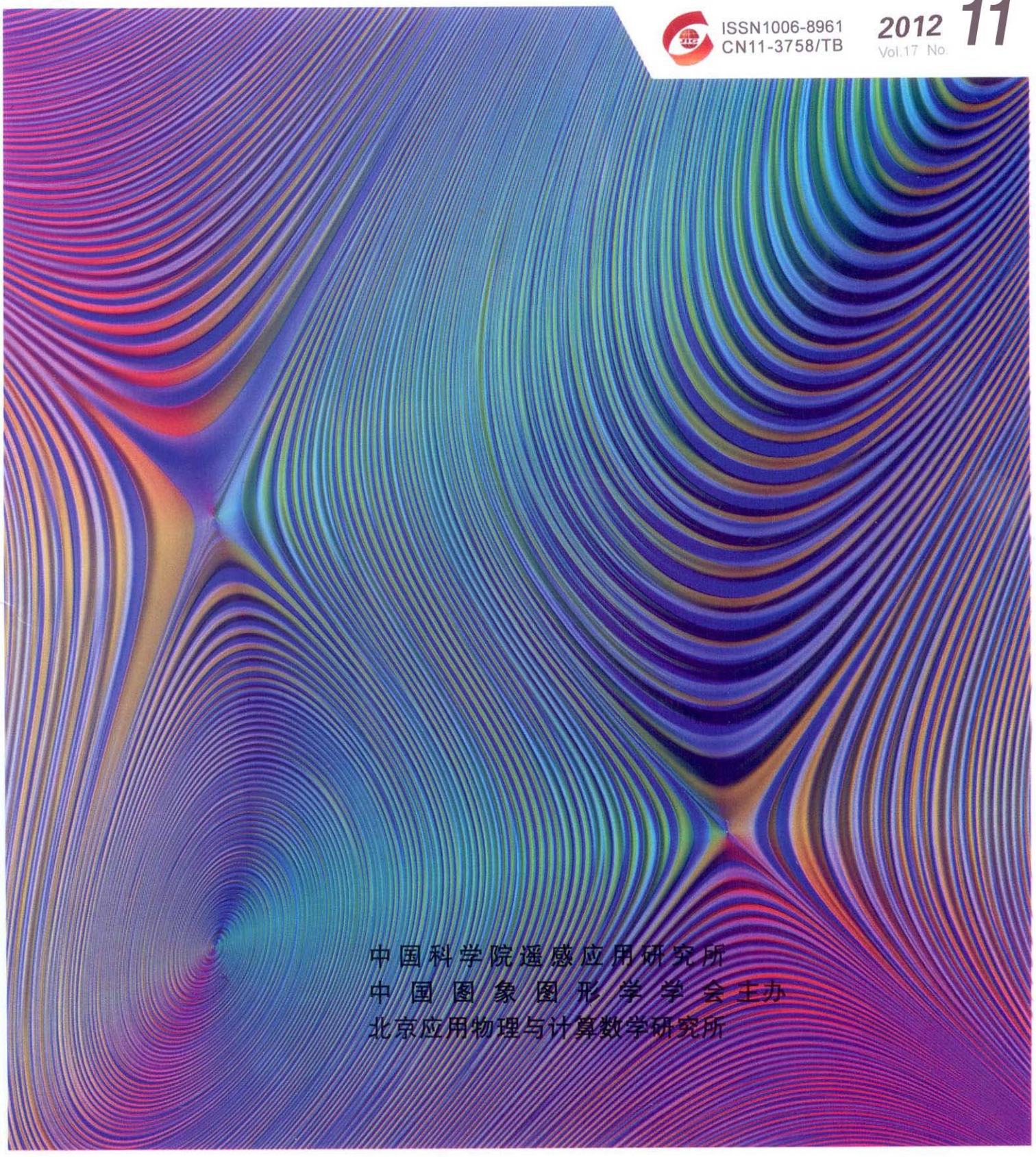
# Journal of Image and Graphics

# 中国图象图形学报

ISSN1006-8961
CN11-3758/TB

2012 11
Vol.17 No.



中国科学院遥感应用研究所
中国图象图形学学会主办
北京应用物理与计算数学研究所

中国精品科技期刊

# 中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

**2012 年 11 月　第 17 卷　第 11 期(总第 199 期)**

## 目　次

**中国图象图形学报**

刊名题字：宋　健　　月刊(1996 年创刊)　　第 17 卷　第 11 期　　2012 年 11 月 16 日出版

| | |
|---|---|
| **主管单位** | 中国科学院 |
| **主　　办** | 中国科学院遥感应用研究所<br>中国图象图形学学会<br>北京应用物理与计算数学研究所 |
| **主　　编** | 李小文 |
| **编辑出版** | 《中国图象图形学报》编辑出版委员会<br>北京 9718 信箱　邮编 100101<br>电子信箱：jig@ irsa. ac. cn<br>电话：010-64807995　010-82614429<br>网　址：www. cjig. cn |
| **印刷装订** | 北京北林印刷厂 |
| **广告经营许可证** | 京朝工商广字第 0346 号 |
| **总 发 行** | 北京报刊发行局 |
| **订　　购** | 全国各地邮局 |
| **国外发行** | 中国国际图书贸易总公司<br>(中国国际书店)<br>(北京 399 信箱　邮编 100044) |

| | |
|---|---|
| **Superintended by** | Chinese Academy of Sciences |
| **Sponsored by** | Institute of Remote Sensing Application, CAS<br>China Society of Image and Graphics<br>Institute of Applied Physics and Computational Mathematics |
| **Chief editor** | LI Xiaowen |
| **Editor, Publisher** | Editorial and Publishing Board of Journal of Image and Graphics<br>(P. O. Box 9718, Beijing 100101, China)<br>E-mail: jig@ irsa. ac. cn |
| **Distributed by** | Beijing Bureau for Distribution of Newspapers and Journals |
| **Domestic** | All Local Post Offices in China |
| **Foreign** | China International Book Trading Corporation<br>(P. O. Box 399, Beijing 100044, China) |
| **Printed by** | Beijing Beilin Printing House |

ISSN 1006-8961　CN11-3758/TB　CODE ZTTXFZ　国内邮发代号：82-831　国外发行代号：M1406　国内定价：45.00 元

第 17 卷 第 11 期
2012 年 11 月

中国图象图形学报
Journal of Image and Graphics

Vol. 17, No. 11
Nov., 2012

**中图法分类号**: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2012)11-1409-09

**论文引用格式**: 段菲,章毓晋. 有监督子空间建模和稀疏表示的场景分类[J]. 中国图象图形学报,2012,17(11):1409-1417.

# 有监督子空间建模和稀疏表示的场景分类

段菲, 章毓晋

清华大学电子工程系, 北京 100084

**摘 要**: 提出了一种基于有监督子空间建模和稀疏表示的场景分类算法。该算法将采用非监督方式求取所有场景类别公共字典的稀疏编码模型分解为一系列各目标函数相互独立的多目标优化问题,实现了各类别字典的有监督学习。在所有类别的字典学习完毕后,再以各子空间和的基集来对每幅图像中所有局部特征进行协同编码,并借助空间金字塔表示(SPR)和特征各维最大汇总(max pooling)构成最终图像的全局特征表示。为对算法的有效性进行验证,在 4 个常用的场景图像库上进行了分类实验,结果表明该算法比采用非监督字典学习的方法在性能上有了显著提升。

**关键词**: 稀疏表示;字典学习;空间金字塔匹配;场景分类;子空间建模

# Scene categorization via supervised subspace modeling and sparse representation

Duan Fei, Zhang Yujin

*Department of Electronic Engineering, Tsinghua University, Beijing* 100084, *China*

**Abstract**: In this paper, we present a new scene categorization algorithm based on supervised subspace modeling and sparse representation. The proposed method implements supervised dictionary learning via decomposing the unsupervised sparse dictionary learning model into a group of independent optimization problems. After learning the dictionaries of all categories, we aggregate them to form a global dictionary and encode each local feature of an image based on it. After using spatial pyramid representation and max pooling of local features' coding vectors, the final holistic feature depicting a scene image can be retrived. Comprehensive experimental results on four popular benchmark scene datasets show that our method achieves very promising result compared to existing state-of-the-art techniques.

**Key words**: sparse representation; dictionary learning; spatial pyramid matching; scene categorization; subspace modeling

## 0 引 言

场景分类的研究目标是依据给定图像的视觉信息,自动赋予该图像一个能反映其全局信息的语义标签,其主要难点在于场景图像具有较大的类内方差和较小的类间方差,即由于光照、遮挡、视角等因素的影响,相同类别的场景中所含的对象可能存在巨大差异,而不同场景类别之间又有可能包含大量相似的对象。加之不同对象之间又往往有着极为复杂的交互关系,因此采用何种特征来概括图像的全局内容就显得至关重要。

近年来人们提出了多种图像内容表示方法[1-6]。其中,最有影响的模型当属词袋模型[6](BoW)。该模型的主要思想是若将从图像的局部块(local patch)中所提取的描述符(通常是 SIFT[7] 特征,视

**收稿日期**:2012-03-13;**修回日期**:2012-05-14

**基金项目**:国家自然科学基金项目(NNSF:61171118);教育部高等学校博士学科点专项科研基金项目(SRFDP-20110002110057)

**第一作者简介**:段菲(1979— ),男,现为清华大学信息与通信工程专业博士研究生,主要研究方向为图像处理和机器学习。

E-mail: duanf@163.com

作视觉单词)与文本分析领域的自然语言词汇类比,而将整幅图像与文本文档类比,则图像可被看作是一个由不考虑空间信息的无序局部特征所构成的集合。一般说来,BoW 由下列4个模块构成:局部特征提取字典学习、局部特征编码和特征汇总。该表示方法介于传统的底层特征(如颜色直方图、纹理特征等)和语义表示方法(关键词标注)之间,为缩短语义鸿沟搭建了一个中间层。

传统的 BoW 模型采用非监督方式(通常选用 K-Means 算法)对训练集中所有图像的局部特征的一个子集进行聚类,其目标是通过最小化重建误差来保留原始特征中最充分的信息。使用非监督算法有几个缺陷:首先,这种方法没有充分利用训练集中的类别标签信息;其次,K-Means 算法所追求的是误差平方和准则下的最优,因此它倾向于将聚类中心选择在密度较高的那些区域,故采用该方法生成的字典对那些位于原始特征空间中低密度区域的样本缺乏足够的描述能力;再次,使用 K-Means 算法需要事先依据经验指定聚类的类别数 K。若类别数过小,会对数据造成欠分割,降低字典的区分能力。若类别数过大,又会造成样本空间过分割。

在局部特征编码环节,传统的方法往往采用基于最近邻(NN)匹配的向量量化(VQ),这种方法的缺陷是会造成较大的量化误差,而若字典中词汇过完备,则又易受随机扰动的影响给编码带来不确定性。针对该问题,Gemert 等人[8]提出了软量化(soft assignment)的概念,该方法将 VQ 中的最近邻匹配改造为 K 近邻(KNN)匹配,依据从字典中找到的 K 近邻词汇到待量化特征的距离自适应地计算 K 个权值,最终以这 K 个近邻词汇的线性组合来表示待量化的特征。该方法与 VQ 相比能够显著减少量化误差,但 K 的选择对量化结果具有显著影响。

针对该问题,Yang 等人[9]将稀疏编码[10](sparse coding)引入 BoW 模型。该算法与 VQ 和软量化相比,能够从预先学习好的字典中在满足一定稀疏度要求的前提下自适应地筛选出一组能够较好描述待量化特征的字典基元。同时从该模型中还可学习到与编码方法最为匹配的过完备字典。此外,由于有了过完备和冗余表示理论[11]的支持,在基集过完备的情况下,字典长度对系统性能的影响并不像 K-Means 等算法那样敏感。在特征汇总阶段,Yang 等人[9]在空间金字塔表示(SPR[3])框架下提出特征维最大汇总(max pooling),进一步提高了汇总特征对于局部空间平移变换的稳健性,同时有更合理的生物学解释。Boureau 等人[12]在 Yang 等人[9]所提出的框架下对各种模块(即字典学习模块和特征汇总模块)的不同组合方案进行了系统的试验评价。他们得出以下结论:稀疏编码优于软量化和 VQ;特征各维最大汇总优于传统的平均汇总(average pooling)。这些结论的产生,将 BoW 模型的性能提升至了一个全新的水平,但上述方法也有着明显的不足。首先,由于稀疏编码的数值求解方法(如 K-SVD[13]等)的计算复杂度较高,在面对大规模分类问题时,可学习到的字典长度会受到显著制约;其次,当数据集中的类别数发生变化或某些类别中出现新增样本而需要对字典进行更新时,前述方法只能重新建立字典;再次,由于受内存容量的制约,当数据集中的类别或样本总数较大时,通过随机采样参与到字典学习的局部特征数目将会成比例降低,势必导致字典对数据集中各类别描述能力产生不均衡现象。

针对上述问题,Qin 等人[14]和 Wojcikiewicz 等人[15]提出对每个类别使用 K-Means 算法(在一个或多个尺度上)单独学习一个字典,然后将各类别字典进行串接,形成适合于描述所关心问题域中所有类别样本的全局字典,之后再使用该全局字典对所有样本进行编码。在这两种方法中,字典学习和编码采用的是传统的 K-Means 和 VQ。在对局部特征编码完毕后,这两种方法均采用特征平均汇总的方式,显然这些环节均有较大的改进余地;此外 Qin 等人[14]和 Wojcikiewicz 等人[15]并未对各类子字典串接这一做法的理论可行性进行分析。

基于上述文献中存在的不足,提出一种基于有监督子空间建模和稀疏表示的场景图像表示方法。与 Wright 等人[16]类似,这里所做的基本假设是当选用 SIFT 描述符来刻画图像的局部特征时,每个类别的样本都位于一个低维的线性子空间中,而问题域的所有类别则位于一个更高维的外围空间(ambient space)中。本文的主要贡献可归纳为以下5点:1)利用稀疏编码算法针对各类别单独学习一组过完备基,隐式地将类别标签融入了字典学习中,并在一定程度上提供了一个学习大规模字典的可行方案;2)由于将全局字典的学习分解为对若干子空间基的独立学习,使得训练集中更多的局部特征可参与到字典学习中,有利于捕捉不同子空间中同类对象的细微差异,由此提升对各子空间的描述能力;3)严格证明了将各子空间的基进行组合的做法本质上是利用子空间的和来表示问题域中各类别子空间的外围

空间。本文是首次给出此类做法可行性证明的文献;4)提出了一个以重建误差为度量的过完备基长度选取的自适应算法;5)在编码阶段,基于 Zhang 等人[17]提出的协同表示优于稀疏表示的结论,将各图像的局部特征在组合后的过完备基上进行编码,并结合 SPR[13]特征各维最大汇总(max pooling)方法,将局部特征的稀疏编码向量整合为一个刻画图像完整内容的全局向量。最后采用 SVM + HIK[18](直方图相交核)来对各样本进行训练和测试。为了对本文算法的有效性进行验证,在 4 个常用的场景图像库 Scene-8[19]、Scene-13[20]、Scene-15[2]和 UIUC-Sports8[21]上进行了分类实验,结果表明本文算法与许多当前表现最优异的算法相比,性能有了显著提升。

## 1　数学符号及相关概念

设有训练集 $\{(y_i, x_i)\}_{i=1}^{N}$,其中 $y_i \in \{1,2,\cdots,C\}$ 为第 $i$ 个训练样本的类别标签,$C$ 为数据集中类别总数,而 $x_i \in \mathbf{R}^{d\times 1}$ 表示第 $i$ 个训练样本的特征。

### 1.1　向量量化

向量量化实际上要求解如下的优化问题:

$$\min_{\boldsymbol{B}} \sum_{i=1}^{N} \min_{j=1,\cdots,K} \| \boldsymbol{x}_i - \boldsymbol{b}_j \|_2^2 \tag{1}$$

式中,$\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_K] \in \mathbf{R}^{d\times K}$ 为用 K-Means 算法求得的 $K$ 个聚类中心。

K-Means 算法可表述为优化模型

$$\min_{S,K} \sum_{i=1}^{N} \| \boldsymbol{x}_i - \boldsymbol{B}\boldsymbol{s}_i \|_2^2$$
$$\text{s. t. } \| \boldsymbol{s}_i \|_0 = 1,\ \| \boldsymbol{s}_i \|_1 = 1, \boldsymbol{s}_i \geqslant 0, i = 1,\cdots,N \tag{2}$$

式中,$\boldsymbol{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_K] \in \mathbf{R}^{K\times N}$ 表示训练样本在基集 $\boldsymbol{B}$ 下的编码向量(coding vector),每列对应一个样本;$\| \boldsymbol{s}_i \|_0$ 为 0-范数,表示向量 $\boldsymbol{s}_i$ 中非零分量的个数,$\boldsymbol{s}_i \geqslant 0$ 表示编码向量 $\boldsymbol{s}_i$ 的所有分量均为非负。这样 K-Means 和 VQ 就统一在同一框架下。在 VQ 的训练阶段,依据式(2)求得 $\boldsymbol{B}$ 和 $\boldsymbol{S}$ 的最优解;而在编码阶段,则固定 $\boldsymbol{B}$ 和 $\boldsymbol{X}$ 来求 $\boldsymbol{S}$ 的最优解。显然上述约束过于严格,很容易产生较大的量化误差。

### 1.2　稀疏编码

如果对式(2)的约束条件进行放松,在重建误差项后附加一个可调节编码向量稀疏度的正则项,则问题就转化为

$$\min_{\boldsymbol{B},\boldsymbol{S}} \| \boldsymbol{X} - \boldsymbol{B}\boldsymbol{S} \|_{\mathrm{F}}^2 + \lambda \| \boldsymbol{S} \|_1$$
$$\text{s. t. } \| \boldsymbol{b}_j \|_2 \leqslant 1, \forall j \tag{3}$$

式中,正则项系数 $\lambda$ 用于在重建误差和编码系数的稀疏度之间进行权衡。可以看出,K-Means 算法实际是式(3)的几个极端稀疏的例子。式(3)对于 $\boldsymbol{B}$ 和 $\boldsymbol{S}$ 来说虽然是非凸的,但若将二者之一固定时,则式(3)对于另一决策变量来说便为凸优化问题。具体说来,当固定 $\boldsymbol{S}$ 时,式(3)为一个带约束项的最小二乘问题;而当固定 $\boldsymbol{B}$ 时,式(3)转化为一个带 $l_1$-正则项的无约束优化问题,因此可采用 $\boldsymbol{B}$、$\boldsymbol{S}$ 交替优化的策略来求解。

采用这种模型的原因在于[9]:首先稀疏编码与 VQ 相比可以获得更小的重建误差;其次稀疏编码对于数据具有良好的自适应性;最后,图像统计学[10]的研究表明自然图像本身即为稀疏信号。

### 1.3　空间金字塔表示模型

传统的词袋模型的另一个致命缺陷是由于它将图像视为局部特征的无序集合,即对局部特征之间的空间关系完全不予考虑,也严重制约了这种描述方案的表达能力。为了克服这一问题,Lazebnik 等人[3]提出了称为空间金字塔表示(SPR)的扩展模型。该方法的思路是在不同空间尺度($l = 0, 1, 2$)上将图像用网格划分为 $2^l \times 2^l$ 个图像块,然后为每个图像块(共 $2^4 + 2^2 + 2^0 = 21$ 个)计算一个视觉词汇直方图,最后再将各图像块的直方图串接,形成对图像的最终全局描述特征。如今,SPR 已成为当前几乎所有表现最优异系统[3,9,12,22]的组成模块。

### 1.4　特征汇总

当来自一幅图像的所有局部特征编码完毕后,传统的 BoW 模型直接将这些局部特征编码取平均,作为图像的最终表示,但这种汇总方式并未考虑局部特征的空间关系,而且无法融入更多的信息[12]。Yang 等人[9]在空间金字塔表示框架下将特征各维最大汇总引入特征汇总环节中,通过实验验证了这种汇总方式与平均汇总方式相比有着明显的优势,而且这种汇总方式有合理的生物学解释[23]。Boureau 等人[24]对各种特征汇总方式进行了深入的理论分析,得出特征各维最大汇总特别适合区分非常稀疏的特征的结论。

## 2　本文算法

### 2.1　基本假设

如前所述,本文算法所做的基本假设是图像虽

为高维信号,但每个类别的样本均存在于一个低维的线性子空间中,而所有类别的样本的并集则位于一个维数高于这些低维子空间的外围空间中(在 2.4 节中将证明该外围空间实际为各类别子空间的和空间)。在场景图像这类较为复杂的研究对象中,不同场景中往往含有相同或相似的结构,这也就意味着容纳不同类别的线性子空间的交集中除了零向量外,还有大量非零的公共向量(局部特征)。以“海滩”类和“森林”类为例,在这两个类别的场景中,起主导作用的对象虽有较大差别(一个主要以水面、沙滩等对象为主,另一个则以植被为主),但这两个类别中通常都包含了天空。这与其他在对子空间建模时追求不同子空间为不相交集的文献有着根本的区别。

对各子空间建模时利用稀疏编码作为工具,目的是为与编码方法相适应而学习高质量的过完备基。基的过完备性是稀疏编码取得可靠结果的基本前提。本算法的基本考虑是如果所学习到的基集对其所张成的线性子空间具有足够的描述能力,则由于各类别样本均来自相应的线性子空间,在理想情况下对每类样本的编码时,编码向量中的非零分量应集中在与该样本最为匹配的子空间的基的一个子集上。用图 1 作一形象的说明,图 1 的上部表示来自场景类“海滩”和“森林”的图像中的局部特征集,假设来自“海滩”和“森林”的样本分别位于由 $\boldsymbol{B}_1$ 和 $\boldsymbol{B}_C$ 所张成的低维线性子空间 $\boldsymbol{V}_1$ 和 $\boldsymbol{V}_c$ 中,即

$$\boldsymbol{V}_i = \{\boldsymbol{B}_i\boldsymbol{x} \mid \boldsymbol{x} \in \mathbf{R}^{l_i}, \boldsymbol{B}_i = [\boldsymbol{b}_1^{(i)}, \boldsymbol{b}_2^{(i)}, \cdots, \boldsymbol{b}_{l_i}^{(i)}]\}, i = 1, \cdots, C \tag{4}$$

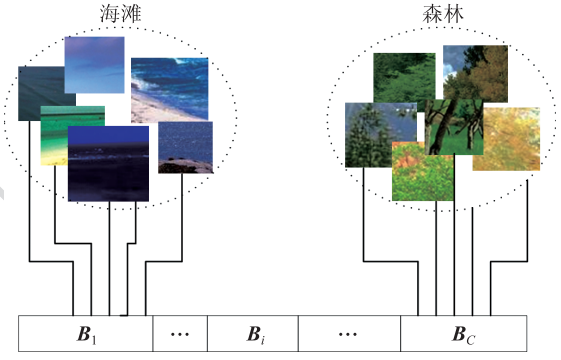



图 1 理想情况下的编码

Fig. 1 Encoding in the ideal case

式中,$\boldsymbol{B}_i = [\boldsymbol{b}_1^{(i)}, \boldsymbol{b}_2^{(i)}, \cdots, \boldsymbol{b}_{l_i}^{(i)}] \in \mathbf{R}^{d \times l_i}$ 表示第 $i$ 类样本所在子空间的基,$\boldsymbol{b}_j^{(i)} \in \mathbf{R}^{d \times 1}$ 为第 $i$ 个子空间中第 $j$ 个基向量。假如 $\boldsymbol{B}_1$ 和 $\boldsymbol{B}_C$ 是完备的,则当用基集 $\boldsymbol{B} = [\boldsymbol{B}_1, \boldsymbol{B}_2, \cdots, \boldsymbol{B}_C]$ 来线性表示来自“海滩”类的样本时,只要系数满足一定稀疏度要求,非零项完全可以只集中在 $\boldsymbol{B}_1$ 和 $\boldsymbol{B}_C$ 的某些基向量上;而当编码时,之所以依据基集 $\boldsymbol{B}$ 而非在每个子空间的基集,是由于从训练集上学习到的基集无法保证其完备性。利用其他子空间的基集作为补充,则有可能增强基集的完备性(这一点在人脸识别文献[16]中也得到了验证),这是由场景图像自身的复杂性所决定的。例如测试集中完全有可能包含有训练集的场景中未包含的对象,而这些对象有可能出现在其他场景中,这样各个子空间的基之间就形成了很好的互补关系。

## 2.2 字典学习

图 2 展示了本文所提出的过完备基学习算法流程。首先从每个类别的训练集中随机抽取若干图像块,以 SIFT 作为描述符,构成字典学习的训练集。设来自第 $i$ 类训练集中场景图像的局部特征共有 $N_i$ 个,即 $\boldsymbol{X}_i = (\boldsymbol{x}_1^{(i)}, \cdots, \boldsymbol{x}_{N_i}^{(i)}) \in \mathbf{R}^{d \times N_i}$,则问题中所有类别的训练集可表示为 $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots \boldsymbol{X}_C] \in \mathbf{R}^{d \times N}$,式中 $N = \sum_{i=1}^{C} N_i$ 为各类训练样本总和,每个局部特征的维数均为 $d$。记特征 $x_j^{(i)}$ 经编码后的特征为 $\boldsymbol{s}_j^{(i)} \in \mathbf{R}^{l \times 1}$,式中 $l = \sum_{i=1}^{C} l_i$ 表示各子空间基向量的总数,并以 $\boldsymbol{S}_i = [\boldsymbol{s}_1^{(i)}, \boldsymbol{s}_2^{(i)}, \cdots, \boldsymbol{s}_{N_1}^{(i)}] \in \mathbf{R}^{l \times N_1}$ 表示第 $i$ 类训练集中样本的编码向量。可用如下映射来描述样本特征与其编码向量之间的映射关系:$f_C: \boldsymbol{X}_i \rightarrow \boldsymbol{S}_i, \forall i$。

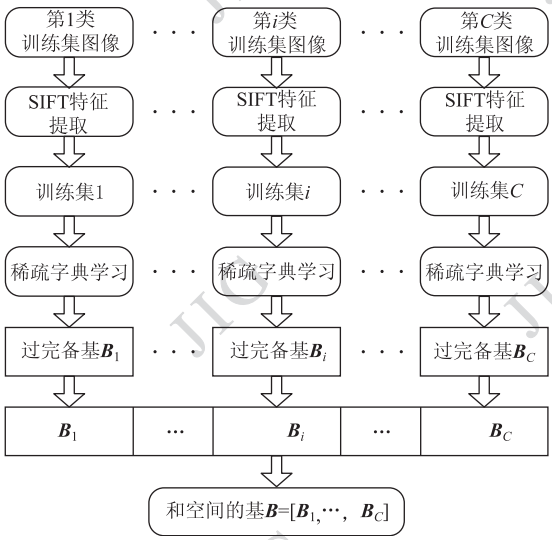



图 2 本文提出的和空间字典学习算法

Fig. 2 The proposed dictionary learning of the sum of multiple subspaces

传统的利用稀疏编码学习字典的算法是通过求解下列模型来实现

$$\begin{aligned}&\min_{\boldsymbol{B},\boldsymbol{S}} \| \boldsymbol{X}-\boldsymbol{B}\boldsymbol{S} \|_{\mathrm{F}}^{2}+\lambda \| \boldsymbol{S} \|_{1}\\ &\text{s. t.} \quad \| \boldsymbol{B}_{\cdot j} \|_{2}\leqslant 1,\forall j=1,\cdots,l\end{aligned} \tag{5}$$

而本算法直接相当于对式(5)按训练集样本的所属类别分解为一个各目标函数独立的多目标优化问题

$$\left.\begin{aligned}&\min_{\boldsymbol{B},\boldsymbol{S}} \| \boldsymbol{X}-\boldsymbol{B}\boldsymbol{S} \|_{\mathrm{F}}^{2}+\lambda \| \boldsymbol{S} \|_{1}\\ &\text{s. t.} \quad \| \boldsymbol{B}_{\cdot j} \|_{2}\leqslant 1,j=1,\cdots,l\end{aligned}\right\}\Rightarrow$$

$$\begin{cases}\min\limits_{B_1,S_1} \| \boldsymbol{X}_1-\boldsymbol{B}_1\boldsymbol{S}_1 \|_{\mathrm{F}}^{2}+\lambda_1 \| \boldsymbol{S}_1 \|_{1}\\ \text{s. t.} \quad \| \boldsymbol{b}_j^{(1)} \|_{2}\leqslant 1,j=1,2,\cdots,l_1\\ \quad\vdots\\ \min\limits_{B_i,S_i} \| \boldsymbol{X}_i-\boldsymbol{B}_i\boldsymbol{S}_i \|_{\mathrm{F}}^{2}+\lambda_i \| \boldsymbol{S}_i \|_{1}\\ \text{s. t.} \quad \| \boldsymbol{b}_j^{(i)} \|_{2}\leqslant 1,j=1,2,\cdots,l_i\\ \quad\vdots\\ \min\limits_{B_C,S_C} \| \boldsymbol{X}_C-\boldsymbol{B}_C\boldsymbol{S}_C \|_{\mathrm{F}}^{2}+\lambda_C \| \boldsymbol{S}_C \|_{1}\\ \text{s. t.} \quad \| \boldsymbol{b}_j^{(c)} \|_{2}\leqslant 1,j=1,2,\cdots,l_C\end{cases} \tag{6}$$

可以看出,由于式(6)中各子目标函数是彼此独立的,这种算法很易于分布式和并行实现。在传统的利用式(5)求解所有子空间的公共基时,由于受内存、运行时间等因素的制约,需要从所有类别的训练集中对各图像的局部特征进行均匀随机抽样。这样在内存不变的情况下,随着问题规模的增大,从每类训练图像集中获得的局部特征将随类别数及每个类别中训练图像的个数成比例减少,相应引发的问题则是学习到的公共基对各类的描述能力分布不均。如有些场景比较复杂,其中涉及的对象种类较多,要完善描述需要大量采样;而有些场景则相对简单,只需要相对较少的局部特征就能学习到高质量的基集。而且在用式(5)进行字典求解过程中,由于是对 $\boldsymbol{B}$ 和 $\boldsymbol{S}$ 交替优化求解的,每轮迭代需要求解两个子模型,当需要求解的基的规模很大时,运算时间将大大增加,甚至在既定硬件条件下不可计算。本算法则通过对式(5)的分解,降低了子问题求解的空间复杂度。由于每个子空间的基是单独求解的,因此可以为每个类别增大训练基中局部特征的采样密度,从而增强基集的描述能力。在实践环节往往需要面对每类基集规模的设定问题,这一点可通过衡量重建误差的大小来控制,即使用基集规模自适应求取算法,子空间 $V_i$ 基集规模自适应学习算法其流程如下:

输入:训练数据 $\boldsymbol{X}_i=[\boldsymbol{x}_1^{(i)},\boldsymbol{x}_2^{(i)},\cdots,\boldsymbol{x}_{N_i}^{(i)}]$,正则项系数 $\lambda_i$,最大基集规模 $L$,重建误差阈值 $T$。

输出:字典 $\boldsymbol{B}_i=[\boldsymbol{b}_1^{(i)},\boldsymbol{b}_2^{(i)},\cdots,\boldsymbol{b}_l^{(i)}]$

1)随机初始化训练样本系数编码矩阵 $\boldsymbol{S}_i=[\boldsymbol{s}_1^{(i)},\boldsymbol{s}_2^{(i)},\cdots,\boldsymbol{s}_{N_i}^{(i)}]$ 为 $\boldsymbol{S}_i^{(0)}$,初始化基集规模 $l\leftarrow l_0$,迭代次数 $k\leftarrow 0$。

2)$k\leftarrow k+1$,求解带约束的最小二乘问题

$$\begin{aligned}&\boldsymbol{B}_i^{(k)}=\arg\min_{\boldsymbol{B}} \| \boldsymbol{X}_i-\boldsymbol{B}\boldsymbol{S}_i \|_{\mathrm{F}}^{2}\\ &\text{s. t.} \quad \| \boldsymbol{b}_j \|_{2}\leqslant 1,\forall j\in\{1,2,\cdots,l\}\end{aligned}$$

3)计算重建误差 $error=\| \boldsymbol{X}_i-\boldsymbol{B}_i^{(k)}\boldsymbol{S}_i^{(k)} \|_{\mathrm{F}}$,若 $error\leqslant T$,则跳至步骤 4);否则求解下列带 $l_1$-正则项的无约束最小二乘问题

$$\boldsymbol{S}_i^{(k)}=\arg\min_{S} \| \boldsymbol{X}_i-\boldsymbol{B}_i^{(k)}\boldsymbol{S} \|_{\mathrm{F}}^{2}+\lambda_i \| \boldsymbol{S} \|_{1},$$

并跳至步骤 2)。

4)返回子空间 $\boldsymbol{V}_i$ 的基集 $\boldsymbol{B}_i=[\boldsymbol{b}_1^{(i)},\boldsymbol{b}_2^{(i)},\cdots,\boldsymbol{b}_l^{(i)}]$。

其中求解步骤 2)带约束的最小二乘问题时,可将其变换至对偶空间来求解,这样做的好处是可将决策变量的数目由 $d\times l$ 减少至 $l$,同时可以获得解析解。

### 2.3　图像表示与分类

在利用式(6)求解出问题域中各子空间($\boldsymbol{V}_1,\cdots,\boldsymbol{V}_C$)的基后,利用各子空间基的互补关系及协同表示优于稀疏表示的结论[17]将其串接,构成新基

$$\begin{aligned}&\boldsymbol{B}=[\boldsymbol{B}_1,\boldsymbol{B}_2,\cdots,\boldsymbol{B}_C]=\\ &[\boldsymbol{b}_1^{(1)},\cdots,\boldsymbol{b}_{l_1}^{(1)};\boldsymbol{b}_1^{(2)},\cdots,\boldsymbol{b}_{l_2}^{(2)};\cdots;\boldsymbol{b}_1^{(C)},\cdots,\boldsymbol{b}_{l_C}^{(C)}]\end{aligned} \tag{7}$$

然后利用下列模型

$$\boldsymbol{S}^{*}=\arg\min_{S} \| \boldsymbol{X}-\boldsymbol{B}\boldsymbol{S} \|_{\mathrm{F}}^{2}+\lambda \| \boldsymbol{S} \|_{1} \tag{8}$$

求解出所有样本的编码向量,式中 $l_1$ 正则项前的系数 $\lambda$ 用于在编码向量稀疏度和重建误差进行权衡。按照模型式(6)的分解方式,可以为每个子空间学习一组规模较大的基。但在各子空间的基集合并之后,面临的新问题则是面对规模较大的基集 $\boldsymbol{B}$,如何高效地进行编码。为此,采用了 Ng 等人[25]提出的非常高效的特征符号搜索(feature-sign search)算法。该算法的主要思路是通过维护一个特征的潜在非零项的索引及其该维特征符号的活动集(active set)来消除由 $l_1$ 正则项带来的编码向量取绝对值的问题。通过不断猜测各维的正负号,将问题转换为一个存在解析解的无约束二次规划(QP)问题,来检验上一轮迭代中符号猜测结果的正确性,并进行更新。在实验中发现,该算法的时间复杂度随样本规

模增长的趋势低于 $O(n)$，非常适合融入本文算法框架中。

为了获得对图像的最终表示，还需将每幅图像的局部特征进行汇总。仿照文献[9]的方式，采用了 2.3 节中介绍的空间金字塔表示[3]框架和特征各维最大汇总来融入局部特征的空间信息。每幅图像编码、特征汇总完毕之后，再利用训练集学习出一个 SVM 分类器。由于空间金字塔表示方法形成的向量是经过归一化的，可将其视为直方图[3]。最适于计算直方图类型特征的核函数为 $\chi^2$ 核和直方图相交核[18](HIK)，其中 $\chi^2$ 核的计算复杂度为 $O(n^3)$，因此在训练 SVM 分类器时，采用了时间复杂度相对较低的直方图相交核(时间复杂度为 $O(n^2)$)。

整个算法的编码、汇总及分类器训练和测试的流程如图 3 所示。

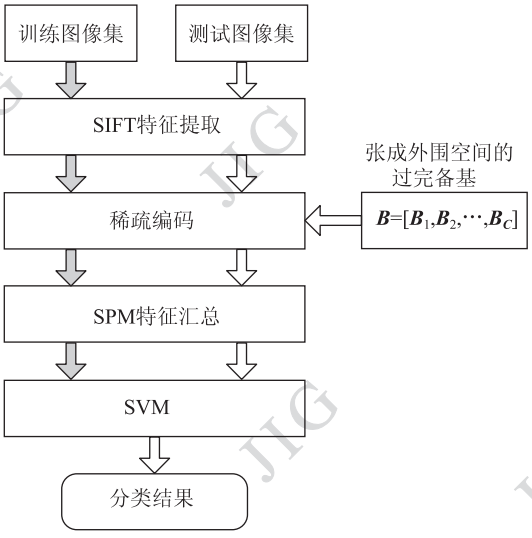



图 3 图像表示方法及分类流程

Fig. 3 The proposed image representation method and the pipeline of classification

### 2.4 理论依据

按照图 2 所示的流程学习到的过完备基 $\boldsymbol{B}$ 实际上张成的是所有类别对应子空间并集的一个超子空间。这是在各类文献中首次对类似的基组合方式做出的严格论述。

首先介绍以下引理[26]：

**引理 1** 假设有线性子空间 $\boldsymbol{V}_1$ 和 $\boldsymbol{V}_2$，其中 $\boldsymbol{V}_i=\{\boldsymbol{B}_i\boldsymbol{x}\mid\boldsymbol{x}\in\mathbf{R}^{n_i}\}$，$i=1,2$，$\boldsymbol{B}_1=[\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{n_1}]$，$\boldsymbol{B}_2=[\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n_2}]$，则二者的和空间 $\boldsymbol{V}_1+\boldsymbol{V}_2=\{\boldsymbol{B}_1\boldsymbol{x}+\boldsymbol{B}_2\boldsymbol{y}\mid\boldsymbol{x}\in\mathbf{R}^{n_1},\boldsymbol{y}\in\mathbf{R}^{n_2}\}$，$i=1,2$ 为包含 $\boldsymbol{V}_1\cup\boldsymbol{V}_2$ 的最小的线性子空间。

由于子空间的并一般来说不再是子空间(因其中元素对线性运算可能不封闭，除非参与并运算的子空间中有一个为其余子空间的超集)，因此将 $\boldsymbol{B}$ 解释为子空间的和更为合理。为此介绍下列定理，并给出相应的证明过程。

**定理 1** 假设 $\boldsymbol{B}_1=[\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{n_1}]$ 和 $\boldsymbol{B}^2=[\boldsymbol{v}_1,\cdots,v_{n_2}]$ 为子空间 $\boldsymbol{V}_1$ 和 $\boldsymbol{V}_2$ 的过完备基，则将两个子空间的过完备基按照式(9)的方式组合后，就成为和空间 $\boldsymbol{V}_1+\boldsymbol{V}_2$ 的一组过完备基。

$$\boldsymbol{B}=[\underbrace{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{n_1}}_{\boldsymbol{B}_1},\underbrace{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n_2}}_{\boldsymbol{B}_2}] \tag{9}$$

**证明**：设 $\dim\boldsymbol{V}_1=s$，$\dim\boldsymbol{V}_2=t$，$\dim(\boldsymbol{V}_1\cap\boldsymbol{V}_2)=r$，令 $\boldsymbol{S}_1=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r]$ 为子空间 $\boldsymbol{V}_1\cap\boldsymbol{V}_2$ 的一个基。由于 $\boldsymbol{S}_1$ 中的基向量线性无关，因而可按如下方式将其扩展为子空间 $\boldsymbol{V}_1$ 的基 $\boldsymbol{S}_2=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_{s-r}]$。由于任意的 $\boldsymbol{\alpha}_i$ 均可由 $\boldsymbol{B}_1=[\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{n_1}]$ 线性表出，故可知 $\mathrm{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_{s-r})\subseteq\mathrm{span}(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{n_1})$。同理，可将 $\boldsymbol{S}_1$ 扩展为子空间 $\boldsymbol{V}_2$ 的一组基 $\boldsymbol{S}_3=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_{t-r}]$，可知 $\mathrm{span}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_{t-r})\subseteq\mathrm{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n_2})$。令 $\boldsymbol{S}_4=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r,\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_{s-r},\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_{t-r}]$，则由于 $\boldsymbol{S}_4$ 为和空间 $\boldsymbol{V}_1+\boldsymbol{V}_2$ 的一组基[26]，且 $\mathrm{span}(\boldsymbol{S}_4)\subseteq\mathrm{span}(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{n_1},\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n_2})$，故可知 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{n_1},\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n_2})$ 为和空间 $\boldsymbol{V}_1+\boldsymbol{V}_2$ 的一组过完备基。

若参与基的组合的子空间数目为多个，仿照定理 1 的证明方法，还可得到如下推论。

**推论** 若有 $n$ 个线性子空间，$\boldsymbol{V}_i$，$i=1,\cdots,n$，每个子空间的过完备基为 $\boldsymbol{B}_i=[\boldsymbol{b}_1^{(i)},\cdots,\boldsymbol{b}_{l_i}^{(i)}]$，$i=1,\cdots,n$ 则各子空间的基按照式(10)组合后，就成为和子空间 $\sum_{i=1}^{n}\boldsymbol{V}_i$ 的一组过完备基。

$$\boldsymbol{B}=[\underbrace{\boldsymbol{b}_1^{(1)},\cdots,\boldsymbol{b}_{l_1}^{(1)}}_{\boldsymbol{B}_1},\cdots,\underbrace{\boldsymbol{b}_1^{(i)},\cdots,\boldsymbol{b}_{l_i}^{(i)}}_{\boldsymbol{B}_i},\cdots,\underbrace{\boldsymbol{b}_1^{(C)},\cdots,\boldsymbol{b}_{l_C}^{(C)}}_{\boldsymbol{B}_C}] \tag{10}$$

## 3 实验结果与分析

为了对本文算法的性能进行定量评价，在 4 个最常用的标准场景库上进行了分类实验，它们分别为 Scene-8[19]、Scene-13[20]、Scene-15[2]、UIUC-Sports8[21]，其中前 3 个数据集依次为包含关系，即 Scene-8 为 Scene-13 的子集，而 Scene-13 又为 Scene-15 的子集。这 3 个库所存在的包含关系非常利于

体现本文算法适合在线学习的优点。

在实验中,遵循了与前述其他方法一样的条件设置,事先将所有彩色图像都转为灰度图像。为了减小数据规模,对所有图像在保持纵横比的条件下将所有图像缩放到了 300×300 像素以内。对每幅图像依光栅扫描顺序以步长 8 个像素等间隔从每幅图像提取若干 16×16 像素的子图像,并以 SIFT[7] 描述符(128 维)作为每个图像块的特征。分别将每个类别的字典长度分别取为 128、256 和 512 共 3 组参数。在最后的特征汇总阶段,采用了特征各维最大汇总的方式。所采用的分类器为 SVM(软件包为台湾大学林智仁博士编写的 LibSVM[27]),使用的核函数为直方图相交核[18]。为了保证结果的客观性,在每个库上均独立进行了 10 次训练集和测试集的随机划分(各类别对应子空间的基均从训练集中学习得到),并将平均准确率作为最后的评价指标。

### 3.1　Scene-8 数据集

Scene-8[19] 场景库中共包含 2 688 幅图像,共 8 个类别,各类别名称及相应的样本总数如下:coast(360 幅)、forest(328 幅)、mountain(374 幅)、open country(410 幅)、highway(260 幅)、inner city(308 幅)、tall building(356 幅)、streets(292 幅)。

按照各文献中通行的实验设置,在每次试验中随机抽取 100 幅图像作为训练集,而将其余图像作为测试集。各类算法的最优结果比较如表 1 所示。

**表 1　各类算法在 Scene-8 库上的最佳平均准确率**

**Table 1　Performance comparison on Scene-8 dataset**

| 算法 | 平均准确率/% |
|---|---|
| Bosch[28] | 86.65 |
| Bosch[29] | 87.80 |
| Qin[14] | 88.81 |
| 本文算法(128) | 89.19 |
| 本文算法(256) | 89.40 |
| 本文算法(512) | 90.37 |

从上表可以看出,随着各类字典长度的不断增加,本算法的性能一直呈上升趋势,与所列出的其他当前优秀系统相比,本文算法体现出明显的优势。

### 3.2　Scene-13 数据集

Scene-13[20] 场景库在 Scene-8 的基础上新增了 5 个类别,共 3 759 幅图像。各类别名称及相应的样本总数如下:coast(360 幅)、forest(328 幅)、mountain(374 幅)、open country(410 幅)、highway(260 幅)、inner city(308 幅)、tall building(356 幅)、streets(292 幅)、bedroom(216 幅)、kitchen(210 幅)、living room(289 幅)、office(215 幅)、suburb(241 幅)。由于该数据集为在 Scene-8 的基础上扩展而来,因此在学习各类子空间的基时,只需对新增的 5 个类别单独进行即可。在该数据集的分类实验中,训练集和测试集的划分方法同 3.1 节。各类算法的最优结果比较如表 2 所示。

**表 2　各类算法在 Scene-13 库上的最佳平均准确率**

**Table 2　Performance comparison on Scene-13 dataset**

| 算法 | 平均准确率/% |
|---|---|
| Bosch[28] | 73.40 |
| Bosch[29] | 85.90 |
| Qin[14] | 85.05 |
| 本文算法(128) | 86.24 |
| 本文算法(256) | 87.05 |
| 本文算法(512) | 87.80 |

### 3.3　Scene-15 数据集

Scene-15[2] 场景库在 Scene-13 的基础上新增了 2 个类别,共 4 485 幅图像。各类别名称及相应的样本总数如下:coast(360 幅)、forest(328 幅)、mountain(374 幅)、open country(410 幅)、highway(260 幅)、inner city(308 幅)、tall building(356 幅)、streets(292 幅)、bedroom(216 幅)、kitchen(210 幅)、living room(289 幅)、office(215 幅)、suburb(241 幅)、store(315 幅)、industrial(311 幅)。同样,由于该数据集在 Scene-13 基础上扩展而来,只需单独学习新增的 2 个类别的基即可。该数据集的分类实验中,训练集和测试集的划分方法同 3.1 节。各类算法的最优结果比较如表 3 所示。

### 3.4　UIUC-Sports8 数据集

UIUC-Sports8[21] 也是场景分类中的一个常用数据集,其中共含 8 个运动场景类别的 1 578 幅图像。各类别名称及相应的样本总数如下:badminton(200 幅)、bocce(137 幅)、croquet(236 幅)、polo(182 幅)、rock climbing(194 幅)、rowing(250 幅)、sailing(190 幅)、snowboarding(190 幅)。为了与其他算法进行比较,在该库的每次分类试验中,从每类随机选取 70 幅图像作为训练集,然后再从剩余图像中随机选取 60 幅图作为测试集。各类算法的最优结果比

较如表4所示。

**表3 各类算法在Scene-15库上的最佳平均准确率**

**Table 3 Performance comparison on Scene-15 dataset**

| 算法 | 平均准确率/% |
|---|---|
| ScSPM+HIK核[9] | 82.40 |
| KSPM[3] | 81.40 |
| CENTRIST[1] | 83.88 |
| HIK+OCSVM[31] | 84.00 |
| Dai[30] | 75.70 |
| 本文算法(128) | 82.42 |
| 本文算法(256) | 84.33 |
| 本文算法(512) | 84.41 |

**表4 各类算法在UIUC-Sports8上的最佳平均准确率**

**Table 4 Performance comparison on UIUC-Sports8 dataset**

| 算法 | 平均准确率/% |
|---|---|
| ScSPM[9] | 82.74 |
| HIK+OCSVM[31] | 83.54 |
| Fei-Fei[21] | 73.40 |
| 本文算法(128) | 83.25 |
| 本文算法(256) | 84.20 |
| 本文算法(512) | 86.11 |

从表1—表4的实验结果可以看出,本算法由于在字典学习过程中融入了训练样本的标签信息,增强了所学习到字典的鉴别性;同时由于对各子空间单独建模,能够保证所学习到的字典对不同子空间描述的均衡性,并使得每个字典基元都具备了类别信息,因而本算法在4个场景库上均表现出明显的优势。随着字典长度的增加,分类准确率仍一直呈现出上升的趋势。此外,由于采取了对各子空间分别建模的方式,当数据集的类别或数目发生变化时,只需对数据集的新增子集进行学习或更新已有子集对应的字典,而无需更新其他类别的字典,因而表现出良好的可伸缩性。

## 4 结 论

基于有监督子空间建模和稀疏表示,提出了一种新的场景图像表示方法。通过把采用非监督方式的稀疏编码来求取全局字典的模型分解为各目标函数相互独立的多目标优化问题,将全局字典的学习问题转化为对每个类别对应子空间的基的有监督学习问题。待所有类别的字典(基)学习完毕后,再将其组合,形成能够描述所有子空间之和的基。这样就形成一个具有显著结构性的过完备字典。文中还对此类做法的理论依据和可行性给出了严格的证明。在此基础上,选用了一种快速的稀疏编码方法来依据全局字典对每幅图像的所有局部特征进行编码,再经由空间金字塔匹配和特征各维最大汇总形成作为最终图像表示的全局特征。在4个常用的标准场景库上的分类结果表明本算法在分类准确率上与所列方法相比表现出明显的优势。此外,本文算法还具备了在线学习算法的特点,与传统的对所有子空间联合建模的方法相比,表现出更强的可伸缩性。在今后的工作中,将研究如何利用本文学习到字典的结构性来提高特征编码的质量。

## 参考文献(References)


[1] Wu J X, Rehg J M. Centrist: A visual descriptor for scene categorization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1389-1501.

[2] Li F F, Perona P. A Bayesian hierarchical model for learning natural scene categories [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005, 524-531.

[3] Lazebnik S, Schmid C, Ponce J. Beyond bag of features: spatial pyramid matching for recognizing natural scene categories [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York, NY, USA: IEEE, 2006, 2169-2178.

[4] Sigian C, Itti L. Gist: a mobile robots application of context-based vision in outdoor environment [C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005, 1063-1069.

[5] Sigian C, Itti L. Rapid biologically-inspired scene classification using features shared with visual attention [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2), 300-312.

[6] Sivic J, Zisserman A. Video google: a text retrieval approach to object matching in videos [C]//Proceedings of IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003, 1470-1477.

[7] Lowe D. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2), 91-110.

[8] Gemert J, Geusebroek J, Veenman C, et al. Kernel codebooks

for scene categorization [C]//Proceedings of Europe Conference on Computer Vision. Marseille, France: Springer, 2008, 696-709.

[9] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Florida, USA:IEEE,2009, 1794-1801.

[10] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. Nature, 1996, 381(6583), 607-609.

[11] Lewicki M, Sejnowski T. Learning overcomplete representations [J]. Neural computation, 2000, 12(2):337-365.

[12] Boureau Y L, Bach F, LeCun Y, et al. Learning mid-level features for recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA:IEEE, 2010, 2559-2566.

[13] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation [J]. IEEE Transactions on Signal Processing, 2006, 54(11), 4311-4322.

[14] Qin J, Yung N. Scene categorization with multiscale category-specific visual words [J]. Optical Engineering, 2009, 48(4): 047203(1-13).

[15] Wojcikiewicz W, Binder A, Kawanabe M. Enhancing image classification with class-wise clustered vocabularies [C]//Proceedings of IEEE International Conference on Pattern Recognition. Istanbul, Turkey:IEEE, 2010, 1060-1063.

[16] Wright J, Yang A Y, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.

[17] Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: which helps face recognition? [C]//Proceedings of IEEE International Conference on Computer Vision. Colorado Springs. USA:IEEE, 2011, 471-478.

[18] Barla A, Odone F, Verri A. Histogram intersection kernel for image classification [C]//Proceedings of IEEE International Conference on Image Processing. Nice, France: IEEE, 2003, 513-516.

[19] Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelop [J]. International Journal of Computer Vision, 2001, 42(3):145-175.

[20] Vogel J, Schiele B. Natural scene retrieval based on a semantic modeling step [C]//Proceedings of International Conference on Image and Video Retrieval. Dublin, Ireland:Springer, 2004.

[21] Li L, Li F F. What, where and who? Classifying events by scene and object recognition [C]//Proceedings of IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007.

[22] Everingham M, Gool L V, Williams C, et al. The PASCAL visual object classes challenge 2008(VOC2008)[C]//Europe Conference on Computer Vision Workshop Symposium. Marseille, France:Springer, 2008.

[23] Serre T, Wolf L, Poggio T. Object recognition with features inspired by visual cortex [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA:IEEE, 2005, 994-1000.

[24] Boureau J, Ponce J, LeCun Y. A theoretical analysis of feature pooling in visual recognition [C]//Proceedings of International Conference on Machine Learning, Haifa, Israel: Omnipress, 2010, 111-118.

[25] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms [C]//Proceedings of Annual Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2006, 801-808.

[26] Axler S. Linear Algebra Done Right[M]. 2nd ed. New York: Springer, 2000:33-34.

[27] Chang C C, Lin C J. Libsvm: A library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 1-27.

[28] Bosch A, Zisserman A, Mouz X. Scene classification via pLSA [C]//Proceedings of Europe Conference on Computer Vision. Graz, Austria:Springer, 2006, 517-530.

[29] Bosch A, Zisserman A, Mouz X. Scene classification using a hybrid generative/discriminative approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(4): 712-727.

[30] Dai D, Yang W, Wu T. Three-layer spatial coding for image classification [C]//Proceedings of IEEE International Conference on Pattern Recognition. Istanbul, Turkey: IEEE, 2010, 613-616.