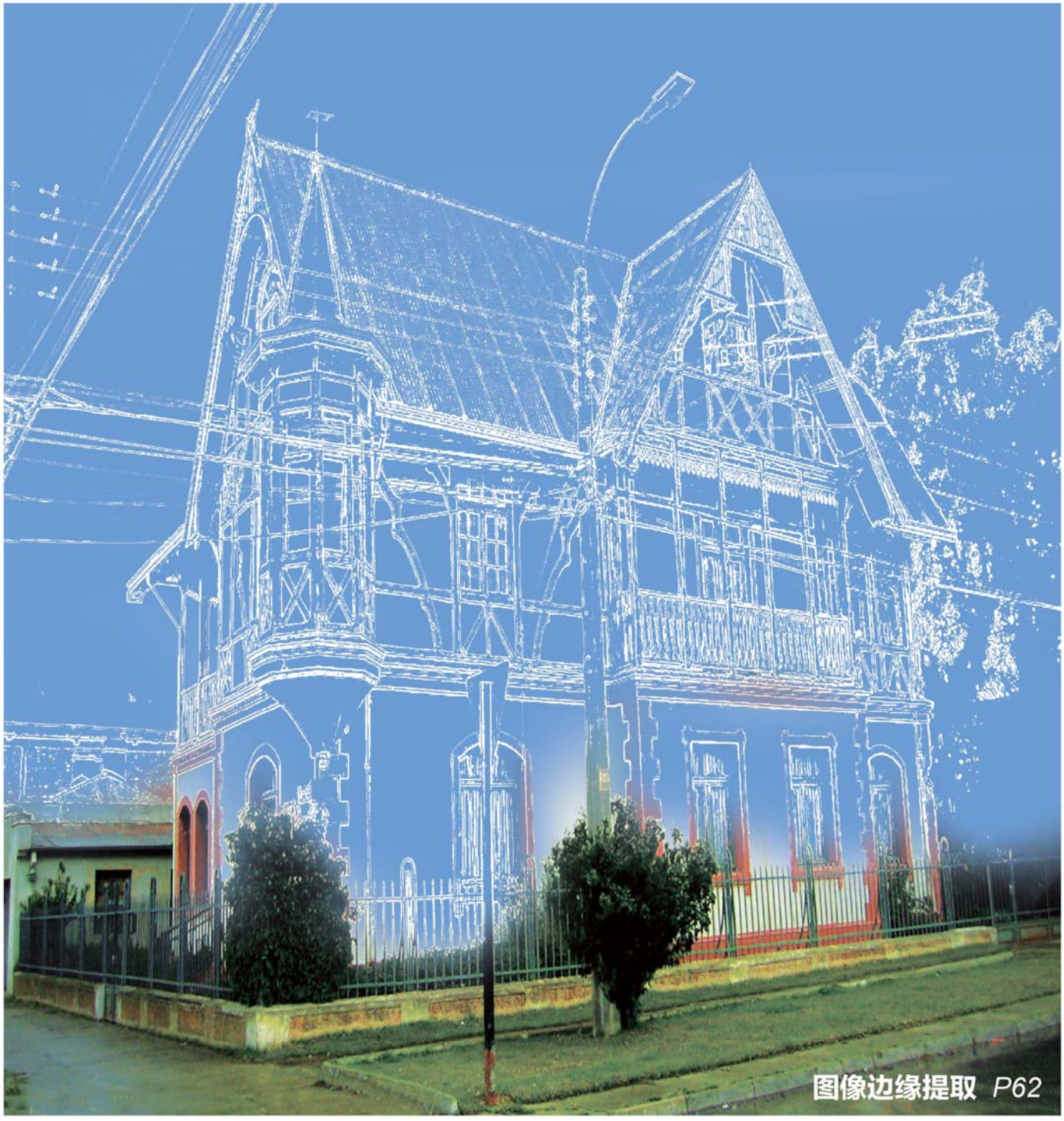
JIG
JOURNAL OF IMAGE AND GRAPHICS
主办：中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2014
01
VOL.19
ISSN1006-8961
CN11-3758/TB



图像边缘提取 *P62*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第19卷第1期（总第213期）
2014年1月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊

主管单位 中国科学院
主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
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
邮政信箱 北京9718信箱
邮　　编 100101
电子信箱 jig@irsa.ac.cn
电　　话 010-64807995
网　　址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
P.O.Box 9718, Beijing, P.R.China
Zip code 100101
E-mail jig@irsa.ac.cn
Telephone 010-64807995
Website www.cjig.cn

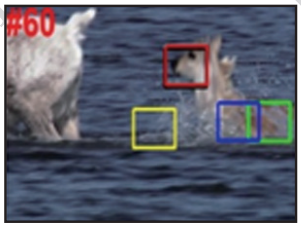
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 50.00元

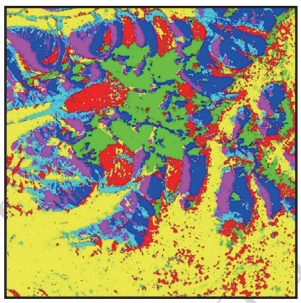
2014年1月 第19卷 第1期（总第213期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次



基于L_2范数最小化的实时目标跟踪(第36页)

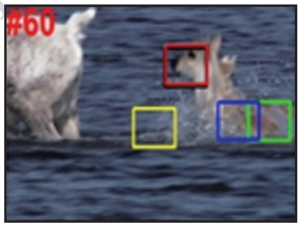
数字地球渲染中单精度浮点数误差改正(第119页)



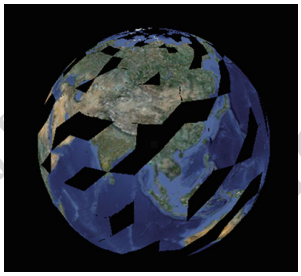
全极化合成孔径雷达影像地形纠正及其在雪冰制图中的应用(第150页)

Volume 19, Number 1
Published January 16, 2014

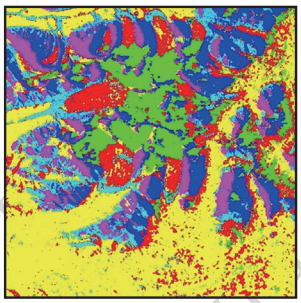
CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



Real-time object tracking based on L_2-norm minimization(P36)



Correction method of single-precision floating-point error for digita Earth rendering(P119)



Topographic correction of polarimetric synthetic aperture radar and its application in snow and ice mapping(P150)

E-mail: jig@irsa.ac.cn
Website: www.cjig.cn
Tel: 010-64807995

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号：TP391　　**文献标识码**：A　　**文章编号**：1006-8961(2014)01-0109-11

论文引用格式：季海峰，高隽，郑鹏，王婧．多尺度空间判别性概率潜在语义分析的场景分类[J]．中国图象图形学报，2014，19(1)：109-118.
[DOI：10.11834/jig.20140114]

多尺度空间判别性概率潜在语义分析的场景分类

季海峰，高隽，郑鹏，王婧
合肥工业大学计算机与信息学院，合肥　230009

摘　要：目的　传统潜在语义分析(LSA)方法无法获得场景目标空间分布信息和潜在主题的判别信息。**方法**　针对这一问题提出了一种基于多尺度空间判别性概率潜在语义分析(PLSA)的场景分类方法。首先通过空间金字塔方法对图像进行空间多尺度划分获得图像空间信息，结合 PLSA 模型获得每个局部块的潜在语义信息；然后串接每个特定局部块中的语义信息得到图像多尺度空间潜在语义信息；最后结合提出的权值学习方法来学习不同图像主题间的判别信息，从而得到图像的多尺度空间判别性潜在语义信息，并将学习到的权值信息嵌入支持向量机(Support Vector Machine，SVM)分类器中完成图像的场景分类。**结果**　在常用的 3 个场景图像库(Scene-13、Scene-15 和 Caltech-101)上的实验结果表明，本文方法平均分类精度比现有许多 state-of-art 方法均优。**结论**　充分说明了空间信息和判别性信息在场景分类中的重要性，并进一步验证了其有效性和鲁棒性。
关键词：概率潜在语义分析；空间金字塔；判别信息；场景分类

Scene classification based on multi-scale spatial discriminative probabilistic latent semantic analysis

Ji Haifeng，Gao Jun，Zheng Peng，Wang Jing
College of Computer and Information，Hefei University of Technology，Hefei　230009，*China*

Abstract：**Objective** Due to the problem that traditional latent semantic analysis (LSA) method is unable to obtain spatial distribution information of objects and discriminative information of latent topic. **Method** We propose a scene classification approach based on multi-scale spatial discriminative probabilistic latent semantic analysis (PLSA). First，it decomposes images in multiple scales using a spatial pyramid approach to obtain spatial distribution information for images. Then，the PLSA model is used to extract the latent semantic information of each local block. Next，the latent semantic features of all local blocks are concatenated with different weights to produce the multi-scale spatial latent semantic information of image. Finally，we exploit weight learning method to learn the discriminative information between different image topics and get multi-scale spatial discriminative latent semantic information of image. Afterwards，the weight information is integrated into the support vector machine (SVM) classifier to perform image classification. **Result** Experimental results on the common three scene image datasets，viz. Scene-13，Scene-15 and Caltech-101，demonstrate that our method performs much better than the existing state-of-the-art approaches. **Conclusion** Which demonstrate the importance of spatial information and discriminative information in image classification and further verify the effectiveness and robustness of our approach.
Key words：probabilistic latent semantic analysis；spatial pyramid；discriminative information；scene classification

收稿日期：2013-05-29；**修回日期**：2013-07-04
基金项目：国家自然科学基金项目(61005007)；国家高技术研究发展计划(863)基金项目(2012AA011005)
第一作者简介：季海峰(1986—　)，男，合肥工业大学计算机应用专业硕士研究生，主要研究方向为图像理解、模式识别、计算机视觉。E-mail：jihaifeng086@126.com

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 1, Jan. 2014

0 引 言

随着高速互联网时代的到来,数字彩色图像的获取和存储变得更加容易,人们接触到的图像数据也以前所未有的速度增长。面对海量的图像数据,如何使计算机模拟人类对图像的理解认知机理,自动把图像按照人们理解的方式分类到不同的语义类别就成为一个关键问题。场景图像分类不仅包含人们对一幅图像的总体认识,而且还为进一步识别出图像中的其他内容提供了基础。因此,图像场景分类成为当前计算机视觉和多媒体信息处理领域的热点问题。

根据场景图像描述方式的不同,场景图像分类方法可以分为基于全局特征和基于中层语义信息两大类。早期的场景分类方法通过提取图像的颜色、纹理、形状等全局底层特征后再利用特征矢量表示场景内容。如 Vailaya 等人[1]提出的结合图像低级特征和二元贝叶斯分类器的多级框架;Szummer 等人[2]采用综合颜色、纹理和频率信息的方法来提取场景特征;Oliva 等人[3]提出空间包络面方法表示图像内容。虽然这些方法在一定条件下能够很好地表示图像,但是由于其都是基于图像的底层特征,无法解决图像分类中的"语义鸿沟"问题[4],即底层视觉特征和高层语义特征之间的不统一性。为了解决这些问题,近年来人们提出了基于图像中层语义信息来表述场景,如基于"词包模型"(BOW)[5-6]方法以及在 BOW 基础上利用概率潜在语义分析(PLSA)[7]和潜在狄雷克雷分布(LDA)[8]等主题分析模型来找出图像最可能属于的主题或者潜在语义分布,从而完成图像场景分类。如 Sivic 等人[9]在稀疏特征上利用 PLSA 模型来进行目标识别,Bosh 等人[10]提出的基于 PLSA 的场景图像分类方法。唐颖军等人[11]在 LDA 基础上进行改进提出了基于类主题空间的潜在狄里克雷分布(CTS-LDA)用来实现自然图像场景分类。然而,基于 BOW 描述的主题分析模型描述场景图像中视觉词汇出现的总体频次,无法获得场景中各目标空间分布信息。为了获取目标空间分布信息,Lazebnik 等人[12]提出了一种空间金字塔匹配方法,将图像在不同尺度下分割成多个子区域,然后对每个子区域中的视觉词汇进行直方图统计和串联,从而得到场景图像的空间分布表达。典型的例子还有 Yang 等人[13]提出的一种基于稀疏编码的空间金字塔匹配方法。Harada 等人[14]提出的判别性空间金字塔图像分类方法。亓晓振等人[15]提出了一种基于稀疏编码的多核学习图像分类方法。通过采用对图像进行空间金字塔多划分方式为特征加入空间信息限制来提高场景图像分类精度,以及杨昭等人[16]提出的一种基于稠密网格的局部 Gist 特征描述,利用空间金字塔结构加入空间信息。上述方法要么只考虑了图像的潜在语义信息,要么只考虑了图像的空间信息,都没有很好地将图像潜在语义信息与空间信息进行有效结合,除此以外,传统的潜在语义方法也没有考虑主题之间的判别性信息。

针对现有方法的不足,在综合以上研究的基础上,提出了一种多尺度空间概率潜在语义分析算法(MS^2-PLSA)并在此基础上进行改善得到本文多尺度空间判别性概率潜在语义分析方法(MS^2D-PLSA)。该方法通过空间金字塔思想对场景图像进行空间分层和局部区域分块划分获得图像局部块之间的空间关系,接着利用 PLSA 模型对每个局部块进行潜在语义挖掘以获得其潜在语义信息分布。之后再通过加权将不同尺度上汇总的潜在语义信息进行串接得到图像最终的特征描述。该特征不仅考虑到图像的局部潜在语义信息,而且考虑到了图像空间信息。在上述多尺度空间概率潜在语义分析算法步骤的基础上再通过权值学习方法来学习主题之间的判别性信息,由此得到的图像权值主题在分类时使得相同类别图像之间的分类间隔尽可能小,不同类别间的分类间隔尽可能大。从而进一步提高场景分类精度。通过在 3 个常用的数据库上的实验结果表明本文算法具有较高的分类性能和很好的鲁棒性。

1 基于 PLSA 图像潜在语义提取

PLSA 模型是由 Hofmann[7]提出的一种用于文本检索的概率生成模型。借鉴 PLSA 主题模型来对图像进行分析[10],基于 PLSA 的图像潜在语义提取表述如下:

假设有训练图像库 $\boldsymbol{D}=[d_1,\cdots,d_N]$,其中 N 表示图像个数,d_j 表示图像库中第 j 幅图像。由 K-means 算法[17]聚类而成的视觉词词汇 $\boldsymbol{W}=[\boldsymbol{w}_1,\cdots,\boldsymbol{w}_M]\in\mathbf{R}^{T\times M}$,其中 T 表示特征维数,M 表示视觉词汇个数,图像区域特征最近邻向量

量化(NN-VQ)[6,12-13,17]形成的共生矩阵 $\boldsymbol{Q}$,其中 $Q_{ij}=n(\boldsymbol{w}_i,d_j)\in M\times N$ 表示视觉词汇 $\boldsymbol{w}_i$ 在图像 d_j 中出现的频次。用 $\boldsymbol{Z}=[z_1,\cdots,z_K]$ 表示潜在语义(主题)集合,K 为常数,为经验确定的主题个数。PLSA 假设"图像—视觉词汇"之间是条件独立的并且潜在语义在图像和视觉词上分布也是条件独立的,在这个前提下,可以用式(1)表示"图像—视觉词汇"条件概率分布,即

$$P(\boldsymbol{w}_i,d_j)=P(d_j)\sum_{k=1}^{K}P(\boldsymbol{w}_i|z_k)P(z_k|d_j) \quad (1)$$

式中,$P(d_j)$ 表示第 j 个图像概率,$P(w_i|z_k)$ 为潜在语义在视觉词汇上分布概率,$P(z_k|d_j)$ 为图像中的潜在语义分布概率。PLSA 的图模型[10]表示如图 1 所示。使用 PLSA 模型可以将一个图像表示为一个对应于主题分布的 K 维向量,这等价于图 2 中所示的矩阵分解[10]。PLSA 模型使用 EM 算法[7,10,17]求解参数,通过极大化以下似然函数来进行模型的参数设计,即

$$L=\sum_{i=1}^{M}\sum_{j=1}^{N}n(\boldsymbol{w}_i,d_j)\log P(\boldsymbol{w}_i|d_j)\propto \sum_{i=1}^{M}\sum_{j=1}^{N}n(\boldsymbol{w}_i,d_j)\log\sum_{k=1}^{K}P(\boldsymbol{w}_i|z_k)P(z_k|d_j) \quad (2)$$

式中,EM 算法[7,10,17]的两个步骤表示如下:

E 步:利用前面估计的模型参数,计算给定观察对 $(\boldsymbol{w}_i,d_j)$ 时潜在主题 z_k 的条件概率分布,即

$$P(z_k|\boldsymbol{w}_i,d_j)=\frac{P(z_k|d_j)P(\boldsymbol{w}_i|z_k)}{\sum_{\ell=1}^{K}P(z_\ell|d_j)P(\boldsymbol{w}_i|z_\ell)} \quad (3)$$

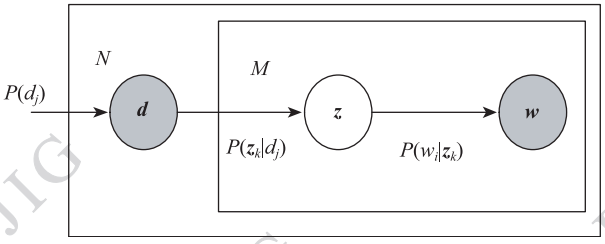


图 1　PLSA 图解模型

Fig. 1　Graph model of PLSA

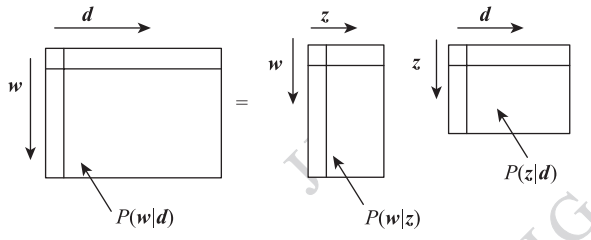


图 2　PLSA 模型中的矩阵分解

Fig. 2　Matrix decomposition in PLSA model

M 步:利用新的期望值 $P(z_k|\boldsymbol{w}_i,d_j)$ 更新参数 $P(\boldsymbol{w}_i|z_k)$ 和 $P(z_k|\boldsymbol{d}_j)$,即

$$P(\boldsymbol{w}_i|z_k)=\frac{\sum_{j=1}^{N}n(\boldsymbol{w}_i,d_j)P(z_k|\boldsymbol{w}_i,d_j)}{\sum_{m=1}^{M}\sum_{j=1}^{N}n(\boldsymbol{w}_m,d_j)P(z_k|\boldsymbol{w}_m,d_j)} \quad (4)$$

$$P(z_k|d_j)=\frac{\sum_{i=1}^{M}n(\boldsymbol{w}_i,d_j)P(z_k|\boldsymbol{w}_i,d_j)}{\sum_{i=1}^{M}n(\boldsymbol{w}_i,d_j)} \quad (5)$$

通过 EM 算法迭代优化得到参数 $P(z|d)$ 和 $P(\boldsymbol{w}|z)$。最终得到的 $P(z|d)$ 描述了图像的特征在潜在语义上的分布情况,通过它来表示图像最终特征。对于测试图像,训练阶段得到的 $P(\boldsymbol{w}|z)$ 保持不变利用使用 folding-in 方法[7]可以快速地推导出测试图像的概率潜在语义 $P(z|d_{\text{test}})$。

2　本文方法

2.1　架构总述

图 3 为本文分类方法的整体框架示意图。系统主要分为 3 个阶段:视觉字典学习、训练和测试。在视觉字典学习阶段,随机地从图像库中选取若干图像进行稠密 SIFT[18]特征提取,然后利用 K-means 算法对提取的特征进行聚类,所有的聚类中心构成视觉字典,每个聚类中心就是一个视觉单词。在训练阶段,通过提出的图像多尺度空间潜在语义模型学习图像多尺度空间语义信息,然后结合权值学习方法学习该主题的判别信息,之后利用所得到权值信息训练 SVM 分类器模型从而得到最优的 SVM 分类器模型。在测试阶段,首先结合训练阶段得到的最优 PLSA 模型参数 $P(\boldsymbol{w}|z)$ 及测试图像的每个局部区域最近邻向量量化形成的共生矩阵,使用 folding-in 方法[7]来计算测试图像的空间潜在语义信息分布,再联合训练阶段生成的最优 SVM 对该潜在语义信息分布向量进行分类,从而实现测试图像分类。

2.2　图像多尺度空间 PLSA 算法

为了利用图像的空间潜在语义信息,构造图像多尺度空间 PLSA 模型。通过对图像空间进行多尺度分割得到局部块后,利用 PLSA 模型对每个局部子区域进行潜在语义学习得到该子区域的潜在语义信息,再将每个局部区域的潜在语义特征加权拼接得到图像的最终特征。模型如图 4 所示。图像空

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 1, Jan. 2014

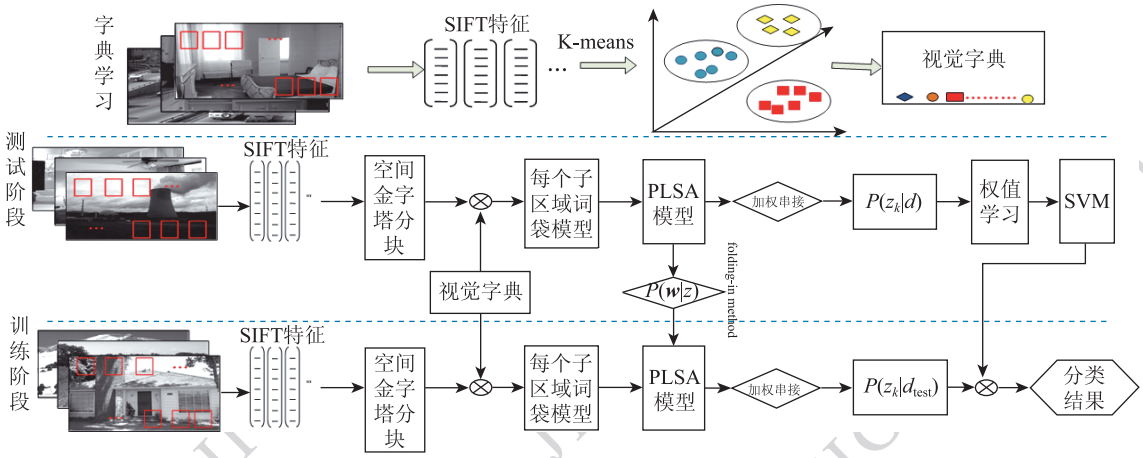


图 3　本文的总体框架图

Fig. 3　The overall framework of this paper

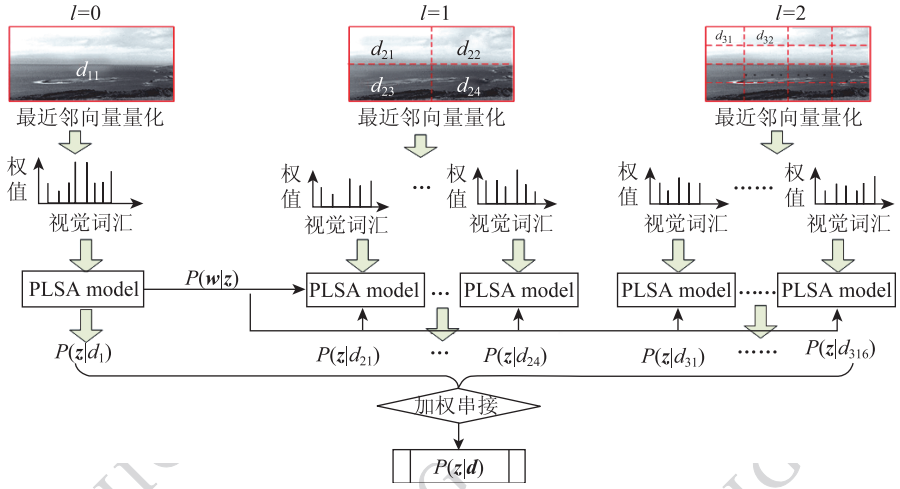


图 4　图像多尺度空间 PLSA 模型

Fig. 4　Multi-scale spatial PLSA model of image

间金字塔的构建是依据 Lazebnik 等人[12]提出的空间金字思想将图像先分层，然后在每层上将图像分割成若干个子区域(块)，每层上子区域数与该层的关系为 $L=2^{d'l}$，$l=0,\cdots,n$，其中 L 表示的是在对 l 层上的子区域数，d'是图像维数，构造一个 3 层空间金字塔模型，$l=0$ 表示是原图像，$l=1$ 对图像进行 2×2 分块划分，$l=2$ 对图像进行 4×4 分块划分则一共可以得到 21 图像块。由此得到的图像区域包含了图像中目标大小和空间关系，因此具有更好的鲁棒性。

在金字塔分割的基础上，利用 NN-VQ 方法来对每层的每个局部块中的特征进行直方图统计，形成每个局部块的共生矩阵 $\boldsymbol{Q}$，然后利用 PLSA 模型来对每层的每个局部区域块进行潜在语义学习，第 0 层潜在语义学习得到的 $P(\boldsymbol{w}|z)$保持不变，而对第 1 层和第 2 层的各子区域利用 folding-in 方法，将 EM 算法中第 M 步的 $P(\boldsymbol{w}|z)$保持不变，不断地更新第 M 步和 E 步中的其他参数使得式(2)中似然函数值达到最大，进而得到每局部块的潜在语义信息 $P(z|d)$。假设潜在语义主题个数为 K_1，则每个局部块 d_{ij}可以产生一个 K_1 维特征向量$[P(z_1|d_{ij}),\cdots,P(z_{K_1}|d_{ij})]$，将同层每个块内学习得到 $P(z|d)$串接形成每层的潜在语义概率分布，最后对各分割层次进行加权串接以形成最终的图像多尺度空间潜在语义概率分布。

测试图像生成多尺度空间潜在语义信息过程与训练图像类似，只是在后期的局部块潜在语义学习阶段要利用训练阶段得到的 $P(\boldsymbol{w}|z)$。这里，$P(\boldsymbol{w}|z)$实际上就是局部特征的潜在语义模型，保持 $P(\boldsymbol{w}|z)$不变，对测试图像的第 0 层、第 1 层和第 2

层每个图像局部块，利用 EM 算法迭代直至收敛，从而得到测试图像在每层的每个局部块中的潜在语义分布 $P(z|\boldsymbol{d})$，然后利用训练阶段对每个局部块内潜在语义处理方法得到测试图像最终的多尺度空间潜在语义分布。

加权拼接中的系数采用文献[13]中的设置方法，即 $l=0$ 层的系数设置为 1/2，第 l 层的系数设置为 $1/2^{L'-l+1}$，其中 L'表示金字塔总层次数，此处 $l=1,\cdots,n$，l 表示的特征所处的层次。故 3 层金字塔结构的加权系数为[0.25,0.25,0.5]。

2.3　多尺度空间潜在语义的判别性学习

大多数基于 PLSA 的图像分类方法认为所有的潜在语义信息对图像分类具有相同的贡献度，即所有的潜在语义信息在分类时应该具有相同的权重。然而，实际上，一些潜在语义之间可能更加相关，对分类具有更多的判别信息。依据大间隔最近邻分类思想[19]，将在上述图像多尺度空间潜在语义学习的基础上分析不同图像多尺度空间主题间的判别性信息，并且通过训练图像库来进行训练。通过权值学习方法得到的权值来表示不同多尺度空间潜在主题间的判别信息。

假设 $d_i,d_j\in\boldsymbol{D}$，及其分别对应的多尺度空间潜在语义 $P(z|d_i)$，$P(z|d_j)$，其权值学习公式表示为

$$dist_{\boldsymbol{\omega}_{k'}}(d_i,d_j)=\left\{\sum_{k'=1}^{K'}(\boldsymbol{\omega}_{k'}\|P(z_{k'}|d_i)-P(z_{k'}|d_j)\|^2\right\}^{1/2} \tag{6}$$

式中，$dist_{\boldsymbol{\omega}_{k'}}(d_i,d_j)$表示图像 d_i 和 d_j 之间的权值欧氏距离。$K'=21\times K$，K 是选择的主题个数。

针对来自不同类别的样本式(6)学习到的权值 $\boldsymbol{\omega}_{k'}$有助于增加潜在主题之间判别性，能够使得来自不同类别样本间类间距尽可能大，同类别样本间类间距尽可能小。因此学习的权值满足如下的约束

$$dist_{\boldsymbol{\omega}_{k'}}(d_i,d_j)>dist_{\boldsymbol{\omega}_{k'}}(d_i,d_{l'})\quad \forall(i,j,l')\in\boldsymbol{G} \tag{7}$$

式中，三元集合 $\boldsymbol{G}=\{(i,j,l')_\gamma|d_i$ 与 d_j 具有相同类别标签，与 $d_{l'}$具有不同的类别标签 $\forall d_i,d_j,d_{l'}$属于训练图像库$\}$，然而，式(7)不能同时满足训练图像中的所有约束，为了解决这个问题引入松弛变量 $\xi_\gamma,\gamma=1,\cdots,|\boldsymbol{G}|$，$|\boldsymbol{G}|$是三元数据集的长度，则式(7)可以改写为

$$dist_{\boldsymbol{\omega}_{k'}}(d_i,d_{l'})^2-dist_{\boldsymbol{\omega}_{k'}}(d_i,d_{l'})^2\geqslant 1-\xi_\gamma \tag{8}$$

对于式(8)的求解可以通过近似求解如下优化问题

$$\begin{aligned}&\min_{\boldsymbol{\omega},\xi_{ijl'(i,j)\in S'}}\sum dist_{\boldsymbol{\omega}_{k'}}(d_i,d_j)^2+C\sum_\gamma\xi_\gamma\\&\text{s.t. }\ dist_{\boldsymbol{\omega}_{k'}}(d_i,d_{l'})^2-dist_{\boldsymbol{\omega}_{k'}}(d_i,d_j)^2\geqslant 1-\xi_\gamma\\&\xi_r\geqslant 0,\boldsymbol{\omega}_{k'}>0,k'=1,2,\cdots,K'\end{aligned} \tag{9}$$

式中，$\boldsymbol{S}'=\{(i,j)|d_i$ 与 d_j 具有相同的类别标签，$\forall d_i,d_j$ 属于训练图像集$\}$，C 是正则约束项，式(9)的优化可以利用标准的目标优化软件来求解[20]。最终得到的权值是由图像多尺度空间主题表征学习生成的，因此权值具有更多的判别信息，能够大大的提高分类准确率。将最终学习到的权值嵌入到 SVM 分类器中，进而改善分类器的分类性能。

3　实验结果与分析

为了验证本文算法的有效性和鲁棒性，将在 3 个常用数据集 Scene-13[21]、Scene-15[12] 和 Caltech-101[22]测试其分类精度，并与其他 state-of-the-art 图像场景分类方法进行比较。

3.1　实验配置

Scene-13 图像数据集共包括了室内场景（如厨房、起居室等）、室外场景（如城市、街道等）等 13 类自然场景 3 859 幅场景图像[21]，Scene-15 则是在 Scene-13 的基础上另增加工厂和商店两个类别，因此共包含 15 个类别的 4 485 幅场景图像[12]。每个类别包含 200～400 幅图像，平均图像大小为 300×250 像素。图 5 给出了 Scene-13 和 Scene-15 每类场景的部分示例图像。为公平比较，参照文献[12-13]，在每次随机实验中，随机抽取各类别 100 幅图像作为训练集，而将剩余图像作为测试集。

Caltech-101 图像数据集[22]（部分示例图片如图 6 所示）包含了如车辆、飞机、花等，以及一些从 Google 搜索引擎中得到的图像共 101 个类别 9 146 幅图像。每类包含的图像数目从 31～800 不等，图像尺寸大小约为 300×300 像素。为了和以前的方法比较[12]，从每个类别中随机选择了 15 或 30 幅图像进行训练，剩下的图像作为测试。

在实验中，遵循与前述其他方法一样的条件设置。舍弃图像颜色信息，事先将图像库中彩色图像均转换为灰度图像。同时，为了减少运算量，在保持纵横比不变的条件下，将所有图像的尺寸缩放到了 300×300 像素以内。采用稠密 SIFT 特征[18]提取算

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 1, Jan. 2014

图 5　Scene13 和 Scene15 数据集部分示例图片

Fig. 5　Part of sample images for Scene-13 and Scene-15 datasets

图 6　Caltech-101 数据集部分示例图片

Fig. 6　Part of sample images for Caltech-101 dataset

法来进行图像局部特征描述;提取 SIFT 特征的图像块为 16×16 像素,步长为 8 像素。

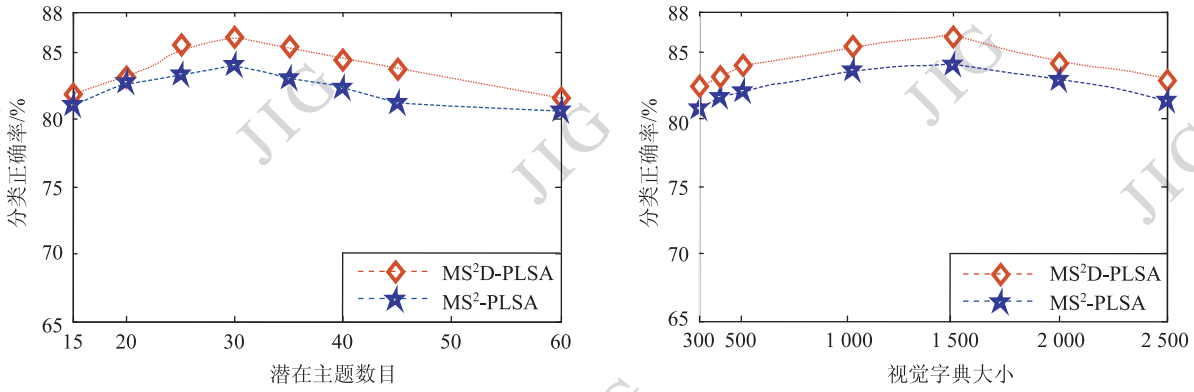
SVM 分类器采用了一对多(one-vs-all)的方法来处理多类问题。此外,将传统 SVM 分类器中使用的 RBF 核函数($k_{rbf}=\exp\left(\frac{-dist(d_i,d_j)}{\gamma}\right),\gamma>0$)中的距离度量函数 $dist(d_i,d_j)$ 用式(6)中给出的 $dist_{\omega_k}(d_i,d_j)$ 替代。使得学习到的图像多尺度空间判别性潜在语义在分类时实现来自相同类别间图像分类间隔尽可能小,不同类别间图像分类间隔尽可能大,从而进一步提高分类器分类精度。参数通过交叉验证方法进行确定。与此同时为了保证结果的客观性,在每个库上独立进行了 5 次随机实验,并将平均分类准确率和标准方差作为评价指标。所有程序都在 Linux 操作系统,12 G 内存,Matlab 2012a 环境下运行。

3.2　实验参数选择

通过对本文方法的分析可以发现在本文方法中有 2 个自由参数:1) 视觉字典的大小;2) 主题的个数。为了探究这 2 个参数对分类精度的影响,将多尺度空间概率潜在语义分析算法(MS^2-PLSA)和在此基础上改进的多尺度空间判别性概率潜在语义分析算法(MS^{2}D-PLSA)算法即本文方法在 Scene-13 数据集上进行了研究,结果如图 7 所示,发现当主题个数为 30 视觉字典大小为 1 500 时本文方法分类精度达到最大。因此,接下来的实验中将 30 和 1 500 分别作为主题个数和视觉字典大小的缺省值。此外,该图也验证了通过在图像多尺度空间潜在主题上进行判别性学习的 MS^{2}D-PLSA 方法比单纯 MS^2-PLSA 分类精度更高,说明了通过本文方法学习的主题具有较高的判别信息,能够显著提高样本在分类时彼此之间区分性,从而提高场景图像的分类精度。

3.3　空间金字塔分层对分类精度影响

为了说明多尺度空间金字塔匹配对提升分类精度的重要性,将本文方法与 MS^2-PLSA,基于最近邻向量量化的空间金字塔匹配(NN-VQ + SPM) 以及基于稀疏编码的空间金字塔(ScSPM)方法[13] 分别在图像单层划分和空间金字塔划分上进行了比较实验,结果如表 1 所示。由表 1 可知,对图像每层进行多尺度划分粒度越细分类精度越高,联合各层形



(a) 潜在主题数目对分类精度的影响　　(b) 视觉字典大小对分类精度的影响

图 7　潜在主题数目和视觉字典大小的变化对本文方法和 MS^2-PLSA 分类精度的影响

Fig. 7　Impacts on classification accuracy of our method and MS^2-PLSA with various numbers of latent topics and size of visual dictionary

成的多尺度空间金字塔划分能够得到更高的分类精度，表明了多尺度空间金字塔划分能够提高分类精度。同时还发现无论是在单层还是空间金字塔上，MS^2D-PLSA 方法都比 MS^2-PLSA、NN-VQ + SPM 和 ScSPM 分类精度高。例如在 Scene-13 数据集上当 $l=(0,1)$ 时 MS^2D-PLSA 分类精度比 MS^2-PLSA，NN-VQ + SPM 和 ScSPM 分别高 1.9%，5.1%，0.3%，当 $l=(0,1,2)$ 时 MS^2D-PLSA 分类精度比 MS^2-PLSA，NN-VQ + SPM 和 ScSPM 分别高 2.1%，5.9%，1.1%。

表 1　不同方法在单层分块与空间金字塔分块上的分类精度比较

Table 1　Comparison on classification accuracy of single level division and spatial pyramid division in different methods (Scene-13)

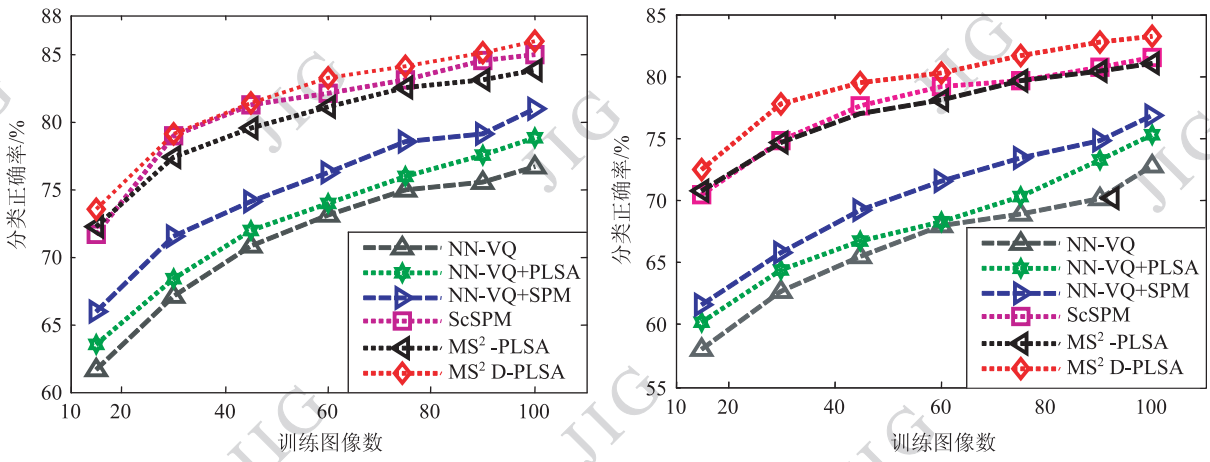
数据集	方法	单层划分 /%			空间金字塔 /%	
		$l=0$	$l=1$	$l=2$	$l=(0,1)$	$l=(0,1,2)$
Scene-13	ScSPM	76.14 ± 0.71	82.62 ± 0.31	82.89 ± 0.45	83.89 ± 0.81	84.94 ± 0.85
	NN-VQ + SPM *	75.23 ± 0.63	77.15 ± 0.71	78.45 ± 0.64	79.06 ± 0.82	80.16 ± 0.63
	MS^2-PLSA	77.43 ± 0.37	80.64 ± 0.31	81.71 ± 0.52	82.24 ± 0.86	83.96 ± 0.53
	MS^2D-PLSA	78.51 ± 0.45	82.73 ± 0.43	83.27 ± 0.68	84.15 ± 0.71	86.07 ± 0.72
Scene-15	ScSPM	71.74 ± 0.62	78.21 ± 0.74	78.73 ± 0.57	79.43 ± 0.79	81.46 ± 0.94
	NN-VQ + SPM *	71.24 ± 0.32	72.43 ± 0.58	73.99 ± 0.58	74.21 ± 0.74	76.01 ± 0.68
	MS^2-PLSA	72.45 ± 0.54	78.62 ± 0.32	79.83 ± 0.57	80.56 ± 0.37	81.47 ± 0.65
	MS^2D-PLSA	73.24 ± 0.63	80.37 ± 0.54	81.25 ± 0.63	82.01 ± 0.81	83.27 ± 0.72

注：* 表示 NN - VQ + SPM 类似 Lazebnik[12] 中 KSPM，此处没有利用直方图交叉核而是将图像空间金字塔划分后每个局部区域进行 NN - VQ 然后串接每个局部块得到图像最终表征。

3.4　本文方法与其他 state-of-the-art 方法的比较

图 8 表示的是通过改变训练图片数来比较本文方法和其他相关方法在 Scene-13 和 Scene-15 两个图像数据集上的分类精度比较，通过图 8(a)(b)可以发现基于多尺度空间概率潜在语义分析(MS^2-PLSA)比基于稀疏编码的空间金字塔方法分类性能略低，比没有在局部块中进行潜在学习的 NN-VQ + SPM 分类精度高。但是在 MS^2-PLSA 基础上进行判别性学习得到的多尺度空间判别性概率潜在语义分析方法比 MS^2-PLSA 和 ScSPM 分类性能均高。说明了该方法能够更加准确地表示图像信息，具有很好的鲁棒性。

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 1, Jan. 2014



(a) Scene-13数据集　　(b) Scene-15数据集

图 8　不同方法在训练图像数变化时的分类精度比较

Fig. 8　Comparison on classification accuracy when varying number of training images in different methods

为了分析本文方法在这两个数据集上不同类别的分类正确率，给出了本文方法在 Scene-13 和 Scene-15 两个数据集上分类生成的混淆矩阵(confusion matrix)，如图 9 和图 10 所示。混淆矩阵对角线上的元素表示本文方法在各类别上的平均分类识别率。在第 i 行和第 j 列的元素表示的是将类别 j 的图像误分类到 i 类别上的百分比。

图 11 比较了本文方法与 MS^2-PLSA 方法在具体的一幅场景图像进行单层划分和空间金字塔划分后在每层上的计算时间。通过图 11 可以发现，无论是在单层划分还是在图像进行空间金字塔划分上，MS^{2}D-PLSA 计算时间都比没有进行判别性学习的 MS^2-PLSA 时间长。说明了在 MS^2-PLSA 上进行判别性学习增加了计算时间的开销。

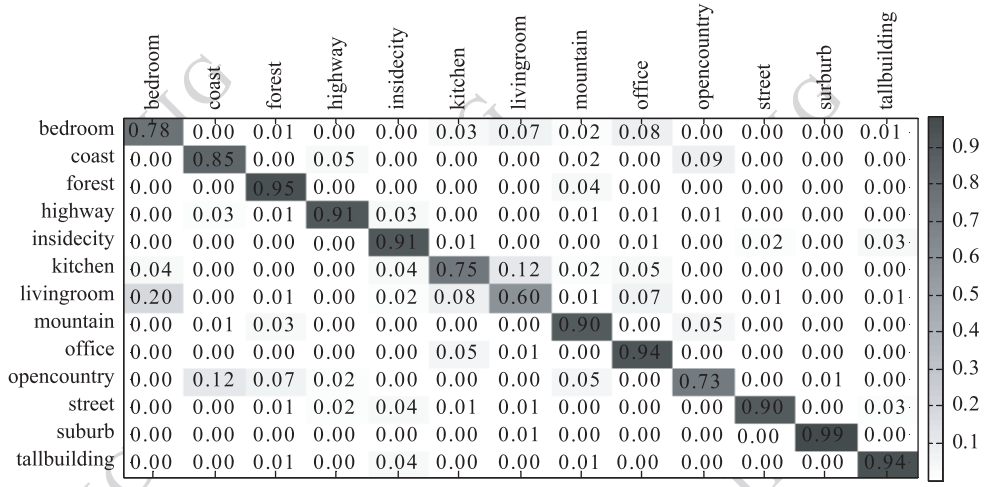


图 9　本文方法在 Scene13 数据集上分类产生的混淆矩阵(分类准确率为 86.07%)

Fig. 9　Confusion matrix generated on Scene-13 datasets in our method (classification accuracy is 86.07%)

表 2 和表 3 分别给出了本文方法和其他性能较好方法在 Scene-13、Scene-15、caltech-101 的分类正确率比较。由表 2 可知本文方法在 Scene-13 数据集上分类率达到了 86.07%，在 Scene-15 数据集上达到了 83.27%，比其他常用的 7 种图像分类算法(空间金字塔匹配核(KSPM)[13]、朴素贝叶斯最近邻(NBNN)[23]、基于相关性视觉词编码模型图像分类方法[24]、基于稀疏编码的空间金字塔匹配(ScSPM)[12]方法、核字典(KC)[25]、概率潜在语义分析(PLSA)[6]、基于最近邻向量量化的空间金字塔(NN－VQ＋SPM)方法)性能均优。由表 3 可看出，在 Caltech-101 数据集上，当训练图像数为 15 时，

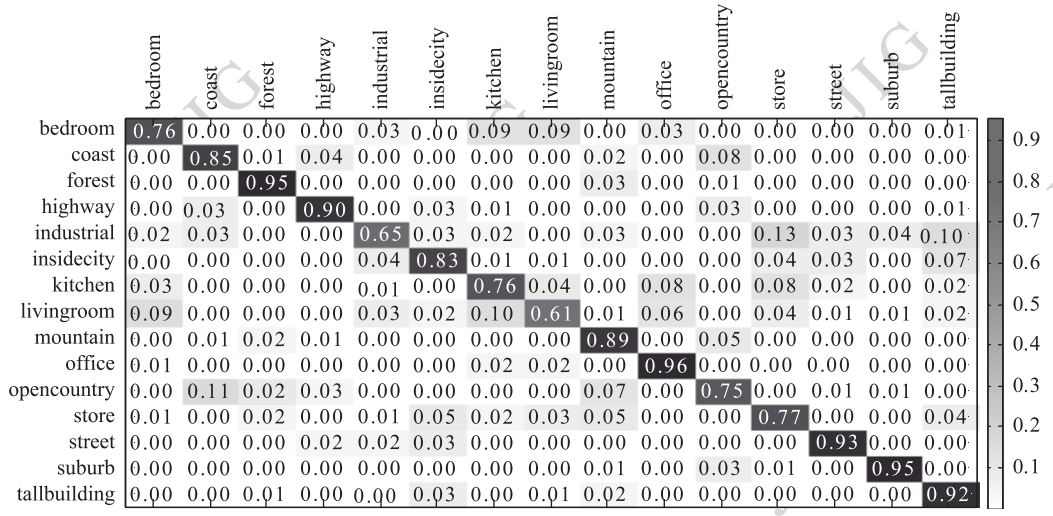


图 10　本文方法在 Scene-15 数据集上分类产生的混淆矩阵(分类准确率为 83.27%)

Fig. 10　Confusion matrix generated on Scene-15 datasets in our method (classification accuracy is 83.27%)

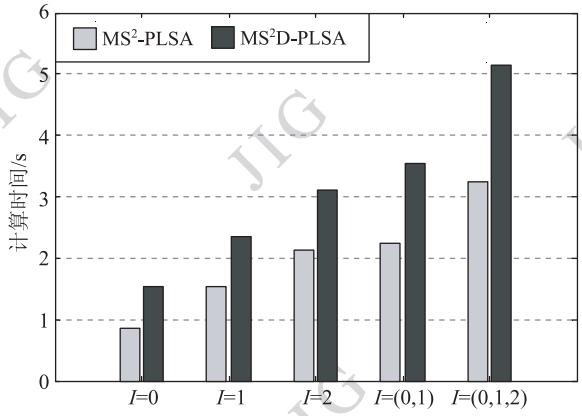


图 11　本文方法和 MS^2-PLSA 在单层划分和空间金字塔划分上的计算时间对比

Fig. 11　Comparison on computing time of single level division and spatial pyramid division in our method and MS^2-PLSA

表 2　本文方法和其他分类方法之间比较

Table 2　Comparison of our method with other methods

/%

方法	Scene-13 数据集	Scene-15 数据集
KSPM[12]	—	81.40 ± 0.50
NBNN[23]	—	77.00
Huang[24] *	84.56 ± 0.68	82.19 ± 0.72
KC[25]	—	76.67 ± 0.39
PLSA[10] *	76.35 ± 0.82	72.74 ± 0.57
ScSPM[13] *	83.15 ± 0.58	81.26 ± 0.59
NN-VQ + SPM	80.16 ± 0.63	76.01 ± 0.68
MS^2-PLSA	84.43 ± 0.53	82.05 ± 0.65
本文方法	86.07 ± 0.71	83.27 ± 0.85

注：* 由于本文的对比实验中参数的设置问题，对比实验结果数据与原始作者数据之间可能有所差异。

表 3　本文方法和其他方法分类精度比较(Caltech-101)

Table 3　Comparison on classification accuracy between our method and other methods(Caltech-101)

/%

方法	训练图像个数	
	15	30
KSPM[12]	56.4	64.6
NBNN[23]	65.0	70.4
Huang[24]	66.9	74.3
KC[25]	—	64.1 ± 1.18
PLSA[10] *	37.5 ± 0.48	42.75 ± 0.87
ScSPM[13]	67.0	73.2
NN - VQ + SPM	47.7 ± 0.45	57.4 ± 0.72
MS^2 - PLSA	66.3 ± 0.56	72.9 ± 0.63
本文方法	68.1 ± 0.71	74.1 ± 0.54

本文分类精度比 MS^2-PLSA 和 ScSPM 分别高出 1.8% ,1.1% ,而当训练图像数为 30 时,其分类性能比 MS^2-PLSA 和 ScSPM 分别提升近 1.2% ,1.0% ;与其他各方法比较,本文方法分类性能均有明显提高。实验结果很好地说明了 MS^2D-PLSA 的有效性和鲁棒性。

4　结　论

提出了一种基于多尺度空间判别性潜在语义分析的图像分类方法。该方法采用空间金字塔结构来对图像进行空间划分,综合考虑了图像的不同层次,不同局部区域块对图像精度的影响;然后利用 PLSA 模型来对每个局部区域进行潜在语义学习,并将每层

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 1, Jan. 2014

学习到的潜在语义信息进行串接得到图像的多尺度空间潜在语义信息分布,该特征能够更加准确的表示原图像信息;最后结合提出的权值学习方法来学习主题间的判别信息,通过改变 SVM 中的度量函数,将学习到的权值信息嵌入到 SVM 中使得图像在分类时来自同类别图像间的分类间隔尽可能小,不同类别图像间的分类间隔尽可能大。从而提高图像的分类性能。通过在 3 个常用的数据集上的实验表明该方法相比较其他现有方法具有很好的分类效果,明显提升了场景图像分类性能。验证了本文方法的有效性和鲁棒性。

参考文献(References)

[1] Vailaya A, Figueiredo M A T, Jain A K, et al. Image classification for content-based indexing[J]. IEEE Transactions on Image Processing, 2001, 10(1): 117-130.

[2] Szummer M, Picard R W. Indoor-outdoor image classification [C]// Proceedings of IEEE International Workshop on Content-Based Access of Image and Video Database. Bombay, India: IEEE Press, 1998: 42-51.

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

[4] 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. Washington DC, USA: IEEE Press, 2010: 2559-2566.

[5] Hao J, Jie X. Improved bags-of-words algorithm for scene recognition[C]// Proceedings of the 2nd International Conference on Signal Processing Systems. Dalian, China: IEEE Press, 2010, 2: V2-279 - V2-282.

[6] Wang Z, Feng J, Yan S, et al. Linear distance coding for image classification[J]. IEEE Transactions on Image Procession, 2013, 22(2):537-548.

[7] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis[J]. Journal of Machine Learning Research, 2001, 42(1-2): 177-196.

[8] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3: 993-1022.

[9] Sivic J, Russell B C, Efros A A, et al. Discovering objects and their locations in images[C]// Proceedings of IEEE International Conference on Computer Vision. Beijing, China: IEEE Press, 2005: 370-377.

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

[11] Tang Y J, Xu D, Jie W J, et al. A novel image scene classification method based on category topic simplex[J]. Journal of Image and Graphics, 2010,15(7):1067-1073. [唐颖军,须德,解文杰,等. 一种基于类主题空间的图像场景分类方法[J]. 中国图象图形学报,2010,15(7):1067-1073.

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

[13] 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. Miami. USA: IEEE Press, 2009: 1794-1801.

[14] Harada T, Ushiku Y, Yamashita Y, et al. Discriminative spatial pyramid[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE Press, 2011: 1617-1624.

[15] Qi X Z, Wang Q. An image classification approach based on sparse coding and multiple kernel learning[J]. Acta Electronic Sinica, 2012, 40(4): 773-779. [亓晓振,王庆. 一种基于稀疏编码的多核学习图像分类方法[J]. 电子学报, 2012, 40(4): 773-779.]

[16] Yang Z, Gao J, Xie Z, et al. Scene categorization of local Gist feature match kernel[J]. Journal of Image and Graphics, 2013, 18(3):264-270. [杨昭,高隽,谢昭,等. 局部 Gist 特征匹配核的场景分类[J]. 中国图象图形学报,2013,18(3):264-270.]

[17] Bishop C M. Pattern Recognition and Machine Learning[M]. New York: Springer, 2006.

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

[19] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research, 2009, 10: 207-244.

[20] Grant M, Boyd S, Ye Y. CVX: Matlab Software for Disciplined Convex Programming, version1. 21 [EB/OL]. (2012-08-16) [2013-04- 25]. http://cvxr. com/cvx.

[21] 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, USA: IEEE Press, 2005: 524-531.

[22] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop. Washington DC, USA.: IEEE Press, 2004: 178-178.

[23] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE Press, 2008: 1-8.

[24] Huang Y, Huang K, Wang C, et al. Exploring relations of visual codes for image classification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE Press, 2011: 1649-1656.

[25] Jan C. van G, Jan-Mark G, Cor J. V, et al. Kernel codebooks for scene categorization[C] // Proceedings of Conference on European Conference on Computer Vision. Marseille, France, Springer Berlin Heidelberg Press, 2008: 696-709.