



中国图形文字报

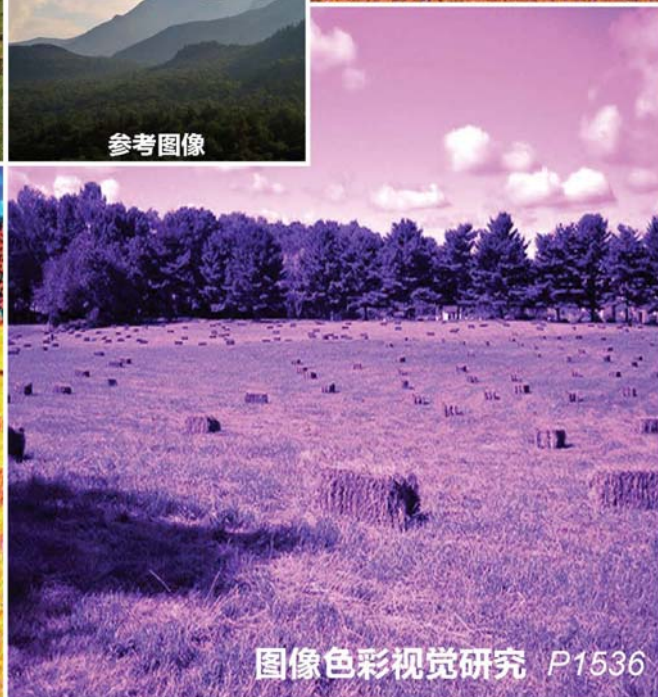
11

ISSN1006-8961
CN11-3758 /TB



原始图像

参考图像



图像色彩视觉研究 P1536

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）



第18卷第11期（总第211期）

2013年11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

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

2013 all rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

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 010-82614429

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 010-82614429

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

综述

- 多媒体技术研究: 2012——多媒体数据索引与检索技术研究进展
中国计算机学会多媒体专业委员会 1383

图像处理和编码

- 热扩散和整体变分模型自适应调整的图像放大模型
王相海, 艾新南 1398
- 椒盐图像的方向加权均值滤波算法
李佐勇, 汤可宗, 胡锦涛, 林亚明 1407
- 3维自适应最小误差阈值分割法
刘金, 金炜东 1416
- 结构约束和样本稀疏表示的图像修复
康佳伦, 唐向宏, 任澍 1425
- 结合小波变换和自适应分块的多聚焦图像快速融合
刘羽, 汪增福 1435

图像分析和识别

- 飞行时间3维相机的多视角散乱点云优化配准
张旭东, 吴国松, 胡良梅, 王竹萌, 邱维巍 1445
- 采用独立色素浓度分布分割皮损图像
徐舒畅 1452
- 基于CUDA的并行加速渲染算法
刘镇, 郝冬宁, 梅向东 1457
- 连续最大流图像分割模型及算法
杨晓艺, 王小欢, 宋锦萍 1462
- 词袋特征压缩算法比较研究
杨赛, 赵春霞 1468

图像理解和计算机视觉

- 融合对象性和视觉显著度的单目图像2D转3D
袁红星, 吴少群, 朱仁祥, 安鹏 1478

计算机图形学

- 多风格融合的复杂森林场景自适应可视化
董天阳, 夏佳佳, 范菁 1486

遥感图像处理

- 保留结构细节的遥感海表温度图像的采样与重构
崔雪森, 杨胜龙, 伍玉梅, 戴阳, 范秀梅 1497
- 组合四相机拼接影像的颜色平衡
任超锋, 姚娜, 林宗坚 1504
- 复杂背景下多样水体遥感自动解译
夏列钢, 沈占锋, 李均力, 骆剑承 1513

“第十届全国智能CAD与数字娱乐会议”专栏

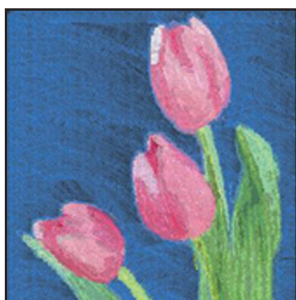
- 3D树木建模技术研究进展
谭云兰, 贾金原, 张晨, 李光耀 1520
- 网格分割的3维网格模型非盲水印算法
杜顺, 詹永照, 王新宇 1529
- 基于放缩系数和均值的多参数颜色迁移
王世亮, 徐岗, 储清林, 胡维华, 汪国昭 1536
- 基于参考图像的色粉画模拟
郑智斌, 张岩, 孙正兴, 杨克微 1542



融合对象性和视觉显著度的单目图像2D转3D(第1478页)



网格分割的3维网格模型非盲水印算法(第1529页)



基于参考图像的色粉画模拟(第1542页)

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Single-view image 2D-to-3D conversion based on objectness and visual saliency(P1478)



Non-blind watermarking algorithm for 3D mesh models based on mesh segmentation(P1529)



Color transfer with multiple parameters by combining the scaling and mean values(P1542)

Review

- Researches on multimedia technology, 2012
Multimedia data indexing and retrieval technology: a review
CCF Multimedia Technology Committee 1383

Image Processing and Coding

- Image magnification model based on adaptive adjustment of thermal diffusion and TV model
Wang Xianghai, Ai Xinnan 1398
Directional weighted mean filter for image with salt & pepper noise
Li Zuoyong, Tang Kezong, Hu Jinmei, Lin Yaming 1407
Three-dimensional adaptive minimum error thresholding segmentation algorithm
Liu Jin, Jin Weidong 1416
Image inpainting by structural constraints and sample sparse representation
Kang Jialun, Tang Xianghong, Ren Shu 1425
Multi-focus image fusion based on wavelet transform and adaptive block
Liu Yu, Wang Zengfu 1435

Image Analysis and Recognition

- Multi-view scattered point cloud optimization registration using a TOF camera
Zhang Xudong, Wu Guosong, Hu Liangmei, Wang Zhumeng, Di Weiwei 1445
Skin lesion segmentation by using independent pigment concentration distribution
Xu Shuchang 1452
Based on the parallel rendering algorithm in CUDA
Liu Zhen, Hao Dongning, Mei Xiangdong 1457
Image segmentation model and algorithm based on continuous max-flow approach
Yang Xiaoyi, Wang Xiaohuan, Song Jinping 1462
Comparative research on compression algorithms of bag-of-features
Yang Sai, Zhao Chunxia 1468

Image Understanding and Computer Vision

- Single-view image 2D-to-3D conversion based on objectness and visual saliency
Yuan Hongxing, Wu Shaoqun, Zhu Renxiang, An Peng 1478

Computer Graphics

- Adaptive visualization of multi-style composition for complex forest scene
Dong Tianyang, Xia Jiajia, Fan Jing 1486

Remote Sensing Image Processing

- Structural detail-reserving method of sampling and reconstruction of remote sensing SST image
Cui Xuesen, Yang Shenglong, Wu Yumei, Dai Yang, Fan Xiumei 1497
Color balancing of the stitched images of a combined four-head camera
Ren Chaofeng, Yao Na, Lin Zongjian 1504
Automatic interpretation of diverse water bodies from remotely sensed images in complex background
Xia Liegang, Shen Zhanfeng, Li Junli, Luo Jiancheng 1513

Issum of CIDE' 2013

- Survey on Virtual 3D Tree Modeling Technologies
Tan Yunlan, Jia Jinyuan, Zhang Chen, Li Guangyao 1520
Non-blind watermarking algorithm for 3D mesh models based on mesh segmentation
Du Shun, Zhan Yongzhao, Wang Xinyu 1529
Color transfer with multiple parameters by combining the scaling and mean values
Wang Shiliang, Xu Gang, Chu Qinglin, Hu Weihua, Wang Guozhao 1536
Image-based pastel simulation
Zheng Zhibin, Zhang Yan, Sun Zhengxing, Yang Kewei 1542

中图法分类号: TP301.6 文献标识码: A 文章编号: 1006-8961(2013)11-1468-10

论文引用格式: 杨赛, 赵春霞. 词袋特征压缩算法比较研究[J]. 中国图象图形学报, 2013, 18(11): 1468-1477. [DOI: 11.11834/jig.20131111]

词袋特征压缩算法比较研究

杨赛, 赵春霞

南京理工大学计算机科学与工程学院, 南京 210094

摘要: 针对目前词袋(BoF)特征压缩算法忽略编码矢量之间空间关系的问题, 本文给出了压缩算法与金字塔模型相配合的图像分类步骤。同时以多个公开图像数据集为实验对象, 对典型词袋特征压缩算法的性能进行比较性研究报道。实验结果表明, 压缩算法对于视觉单词数目以及编码方法具有良好的鲁棒性; 其中基于子空间方法的压缩算法在高层图像特征空间中的分类性能最优, 在多个图像数据集上的分类性能最优, 时间开销最小。

关键词: 词袋特征; 主题模型; 子空间方法; 字典学习; 图像分类

Comparative research on compression algorithms of bag-of-features

Yang Sai, Zhao Chunxia

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract: Against the problem that compression algorithms for bag-of-features (BoF) ignore the spatial relationships of coded vectors, we propose a fusing algorithm of compression algorithms and spatial pyramid model in this paper. Meanwhile, we carried out a set of comparative experiments on several public image datasets. The experimental results show that compression algorithms are robust to visual word numbers and pooling methods of coded vectors. Otherwise, compression algorithms based on subspace methods have achieved best classification performances in the high-level feature space, and best accuracies and smallest time cost in multiple image datasets.

Key words: bag-of-features; topic model; subspace method; dictionary learning; image classification

0 引言

近年来, 词袋(BoF)模型在图像分类研究中得到了广泛关注和显著发展。它的基本思想是离线训练视觉词典, 按照某种编码规则将图像中每个局部描述矢量映射到视觉词典中某个或者某几个视觉单词上, 使用编码矢量的聚集特性表征图像。目前已经出现了大量关于 BoF 模型的改进工作。其中, 针对硬分配编码产生量化误差以及软分配编码忽略局部描述矢量之间势场关系这些问题, Yang 等学者^[1]

提出稀疏编码方法, 它通过解凸优化问题求取局部描述矢量关于视觉单词向量的稀疏表示, 使用多个视觉单词向量的线性组合表征局部矢量, 大大提高了编码方法的性能。针对 BoF 模型丢弃编码矢量之间空间关系的缺点, Lazebnik 等学者^[2]提出空间金字塔模型, 在第 l 层, 图像被粗糙地划分为 4^l 个子区域, 对每个子区域分别建立 BoF 模型, 然后串联起来作为图像的第 l 层表示, 图像之间的相似度为相对应的每层直方图之间匹配核函数的加权和。稀疏编码方法与空间金字塔模型相结合是 BoF 模型发展中的重要里程碑, 成为目前广泛使用的图像特

收稿日期: 2013-01-31; 修回日期: 2013-05-28

基金项目: 国家自然科学基金重大研究计划项目(90820306); 国家自然科学基金面上项目(61272220); 国家青年科学基金项目(61103059); 国家青年科学基金项目(61101197)

第一作者简介: 杨赛(1981—), 女, 南京理工大学模式识别与智能系统专业博士研究生, 主要研究方向为计算机视觉与机器学习, E-mail: yangsai6868@163.com

征表示框架。

然而,对于一幅中等尺寸的图像而言,需要抽取的局部特征数目数以百计,加之稀疏编码方法需要解决 L_1 范数的优化问题,因此对局部描述矢量进行稀疏编码的运算时间很长,时间复杂度非常高。另外,空间金字塔模型每层特征的维数随层次增加,而随之关于视觉单词数目线性增长,这些都严重制约了 BoF 模型在实时性要求非常高的视觉任务中的广泛应用,例如,图像检索^[3]、目标检测^[4]、物体分割^[5]、视觉跟踪^[6]、机器人导航^[7]等。上述问题都可以通过某种压缩算法将图像从稀疏的高维词典空间映射到一个低维的向量空间进行解决。文献[8]使用硬分配编码取代稀疏编码,然后使用监督字典学习算法对 BoF 特征进行压缩从而弥补编码方法带来的不足;文献[9]利用概率隐含语义分析模型(pLSA)得到 BoF 特征在隐含主题空间中的分布,然后使用判别式学习方法对图像进行分类;文献[10]利用蒙特卡洛方法中的 Gibbs 抽样近似估算隐含狄利克雷分配(LDA)模型参数,获取图像的主题向量表示,克服了 pLSA 模型不适用于大规模图像数据集的缺点;文献[11]使用 Fisher 线性判别方法(FLDA)对 BoF 特征进行变换,克服了主题模型反复迭代的缺陷。但是这些关于压缩算法的研究工作还存在以下遗漏:1)在构建 BoF 特征时使用硬分配编码方法与平均统计值,还没有出现关于其他编码及聚集方法如何影响压缩算法性能的研究报道;2)只是对图像的 0 层特征进行压缩,没有将压缩算法与金字塔模型相结合。为此,给出了压缩算法与金字塔模型相配合用于图像分类的详细描述以及压缩算法的时间复杂度分析。并且选取 15 类场景数据集和 MSRcv2 物体数据集为实验对象,验证参数、编码及聚集方法如何影响压缩算法的性能,以及压缩算法在高层空间金字塔特征中的分类性能。在 15 类场景、SUN2012、MSRcv2、Caltech101、PASCAL2007 多个公开测试图像数据集上对压缩算法的正确性以及快速性指标进行比较性分析。

1 词袋模型描述

已知一个图像数据集 $D = \{d_1, d_2, \dots, d_N\}$, 图像经过局部特征提取及描述后,图像被表示为 $X_t = \{x_1, x_2, \dots, x_T\}$, $x_t (t=1, 2, \dots, T)$ 表示第 t 个描述矢

量,利用 K 均值算法对训练图像的所有局部描述矢量聚类得到视觉词典 V ,第 k 个视觉单词表示为 v_k 。为了缩小底层视觉特征与高层语义特征之间的“语义鸿沟”,BoF 模型使用编码方法将局部描述矢量映射到视觉单词向量所张成的语义空间中。最常用的实现方法为硬分配^[12],它选择距离最近的视觉单词表征局部描述矢量。该方法会产生很大的量化误差,直观的改进方法为软分配^[13]方法。但是软分配方法使用所有视觉单词的加权和表征局部描述矢量,忽略了局部描述矢量之间的势场关系。而稀疏编码方法^[11]使用有限个视觉单词向量稀疏表示局部描述矢量,克服了上述缺陷,大大提高了编码性能。各种编码方法的数学表达式为:

1) 硬分配方法

$$u_{tk} = \begin{cases} 1 & j = \arg \min_{j=1, \dots, K} \|x_t - v_k\|_2^2 \\ 0 & \text{其他} \end{cases} \quad (1)$$

2) 软分配方法

$$u_{tk} = \frac{\exp(-\beta \|x_t - v_k\|_2^2)}{\sum_{k=1}^K \exp(-\beta \|x_t - v_k\|_2^2)} \quad (2)$$

3) 稀疏编码

$$u_t = \arg \min_{u_t \in \mathbb{R}^k} \|x_t - Vu_t\|_2^2 + \lambda \|u_t\|_1 \quad (3)$$

式中, u_{tk} 为第 t 个编码矢量的第 k 维值, β 为光滑参数, u_t 为第 t 个编码矢量, λ 为稀疏正则系数。

图像中的局部描述矢量经过编码操作后,需要计算编码矢量的聚集特征才能得到图像内容的紧致表示。而越来越多的相关研究工作表明空间聚集算法具有更优描述性能,其中最具代表性的工作为空间金字塔模型。它在图像的两个坐标方向进行 2 的指数倍划分,记 $l=0, 1, \dots, L$ 为金字塔层次,金字塔总层数为 $L+1$ 。在图像的第 l 层,在水平方向上把图像分成 2^l 块,在竖直方向将图像划分成 2^l 块,图像最终被划分成 $2^l \times 2^l = 4^l$ 块。图 1 为对图像进行金字塔划分的 3 层空间模型,假设第 i 个子区域内有 T_i 个编码矢量,此区域内编码矢量的总和统计值 B_{ia} 和平均统计值 B_{is} 计算公式分别为^[14]

$$B_{ia} = \sum_{t=1}^{T_i} u_t \quad (4)$$

$$B_{is} = \frac{1}{T_i} \sum_{t=1}^{T_i} u_t \quad (5)$$

最近的研究表明最大统计值 B_{im} 的性能要明显优于平均统计值^[15],最大统计值 B_{im} 的数学表达式为

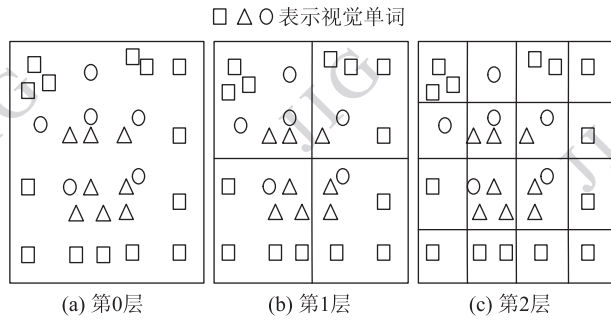


图1 3层空间金字塔模型

Fig. 1 Spatial pyramid model of three levels

$$\mathbf{B}_{im} = \max_{t=1,2,\dots,T_i} \mathbf{u}_t \quad (6)$$

将每个子区域内的特征串联起来得到图像第 l 层特征 $\mathbf{B}_l = [\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_l^T]$, 图像之间的核函数为

$$K(d_i, d_j) = \sum_{l=1}^L \frac{1}{4^{l-1}} K_l(\mathbf{B}_{il}, \mathbf{B}_{jl}) \quad (7)$$

式中, K_l 为核函数, $\mathbf{B}_{il}$ 和 $\mathbf{B}_{jl}$ 分别为第 i, j 幅图像的第 l 层词袋特征。基于支持向量机(SVM)分类的目标函数为

$$\begin{aligned} \max \left(\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(d_i, d_j) \right) \\ \text{s. t.} \quad \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C \end{aligned} \quad (8)$$

式中, C 为惩罚系数, y_i, y_j 为图像对应的类别标号, α_i 为与每个图像样本对应的 Lagrange 乘子。则对测试样本的分类决策函数为

$$f(x) = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i K(d_i, d) + b^* \right) \quad (9)$$

式中, b^* 为分类阈值, l 为支持向量的数目。

2 压缩算法

2.1 算法描述

2.1.1 主题模型

1) 概率隐含语义分析

概率隐含语义分析模型^[16-17]引入隐含变量 $z_m \in \{z_1, z_2, \dots, z_M\}$, 假设 $p(d_i)$ 表示图像在数据集出现的概率, $p(\mathbf{v}_k | z_m)$ 表示主题 z_m 下第 k 个视觉单词的条件概率, $p(z_m | d_i)$ 表示图像中主题的概率分布, 则图像与视觉单词的联合概率分布 $p(d_i, \mathbf{v}_k)$ 为

$$p(d_i, \mathbf{v}_k) = p(d_i) p(\mathbf{v}_k | d_i) = \sum_{m=1}^M p(\mathbf{v}_k | z_m) p(z_m) p(d_i | z_m) \quad (10)$$

根据 Jensen 不等式和概率小于 1 的性质, N 幅图像似然函数的下界为

$$\begin{aligned} L = \sum_{i=1}^N \sum_{k=1}^K \mathbf{B}_i \log p(d_i, \mathbf{v}_k) > \sum_{i=1}^N \sum_{k=1}^K \mathbf{B}_i \\ \sum_{m=1}^M p(z_m | d_i, \mathbf{v}_k) \log(p(\mathbf{v}_k | z_m) p(z_m) p(d_i | z_m)) \\ \text{s. t.} \quad \sum p(v | z) = 1, \sum p(z) = 1, \sum p(d | z) = 1 \end{aligned} \quad (11)$$

利用拉格朗日方法, 式(11)中的优化目标转化为

$$\begin{aligned} L = \sum_{i=1}^N \sum_{k=1}^K \mathbf{B}_i \sum_{m=1}^M p(z_m | d_i, \mathbf{v}_k) \\ \log(p(\mathbf{v}_k | z_m) p(z_m) p(d_i | z_m)) + \\ \sum_{m=1}^M \lambda_m \left[\sum_{i=1}^N p(d_i | z_m) - 1 \right] + \\ \sum_{m=1}^M \eta_m \left[\sum_{k=1}^K p(\mathbf{v}_k | z_m) - 1 \right] + \varepsilon \left[\sum_{m=1}^M p(z_m) - 1 \right] \end{aligned} \quad (12)$$

式中, $\mathbf{B}_i$ 为图像第 0 层的 BoF 特征, $\lambda_m, \eta_m, \varepsilon$ 为拉格朗日乘子。由于式(12)存在对数内部的加运算, 没有闭合解析式, 需要借助于 EM 算法进行求解, 其中 E 步骤的计算公式为

$$\begin{aligned} p(v_k | z_m) &\propto \sum_d \mathbf{B}_i p(z_m | d_i, \mathbf{v}_k) \\ p(d_i | z_m) &\propto \sum_v \mathbf{B}_i p(z_m | d_i, \mathbf{v}_k) \\ p(z_m) &\propto \sum_d \sum_v \mathbf{B}_i p(z_m | d_i, \mathbf{v}_k) \end{aligned} \quad (13)$$

已知变量 d_i 和 $\mathbf{v}_k$ 关于主题 z_m 的条件概率, 计算关于隐含变量 z_m 后验概率的 M 步骤为

$$p(z_m | d_i, \mathbf{v}_k) = \frac{p(\mathbf{v}_k | z_m) p(z_m) p(d_i | z_m)}{\sum_m p(\mathbf{v}_k | z_m) p(z_m) p(d_i | z_m)} \quad (14)$$

交替运行 E 步骤和 M 步骤直到满足某一收敛条件时就可以得到 BoF 特征的压缩特征。

2) 隐含狄利克雷分配模型

隐含狄利克雷分配模型^[18]是一种产生式 3 层概率模型, 它将图像数据集中的每幅图像与 M 个主题 z 的多项式分布 τ 相对应, 每个主题又与 K 个视觉单词的多项分布 ψ 相对应, 它所代表的概率表示为

$$p(\tau, z, v | \alpha, \beta) = p(\tau | \alpha) \prod_{n=1}^{N_d} p(z_n | \tau) p(v_n | z_n, \beta)$$

式中, α 和 β 分别为 τ 和 ψ 的先验分布参数, N_d 为图像中单词的总数目, z_n 为第 n 个单词对应的主题, v_n 为第 n 个单词, 数据集 $\mathbf{D}$ 的似然函数为

$$p(\mathbf{D} | \alpha, \beta) = \prod_{d=1}^M \int p(\tau_d | \alpha) \times$$

$$\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} | \tau_d) p(v_{dn} | z_{dn}, \beta) d\tau_d \quad (15)$$

式中, d 表示第 d 个图像, d_n 为第 d 个图像的第 n 个单词。采用 Gibbs 抽样算法对式 (15) 进行求解, 逐个处理图像集合里的每个词符, 在已知其他词符主题分布的条件下, 当前词符属于每个主题的条件分布估算公式为

$$\phi(z_k = m | z_{-k}, v_k, d_i, \cdot) \propto \frac{C_{vm}^{VM} + \beta}{\sum_{v=1}^V C_{vm}^{VM} + V\beta} \frac{C_{dm}^{DM} + \alpha}{\sum_{m=1}^M C_{dm}^{DM} + M\alpha} \quad (16)$$

式中, v_k 和 d_i 分别表示每个词符的词汇索引和图像索引, z_k 表示它所对应的主题变量, $z_k = m$ 表示把主题 m 赋给词符 k 作为主题, z_{-k} 表示除当前词符外的其他所有词符的主题赋值, $\cdot$ 表示其他所有的已知或可见的信息, 如其他所有词符的词汇索引 v_k 和图像索引 d_i , 以及超参数 α 和 β , C_{vm}^{VM} 、 C_{dm}^{DM} 分别是 $V \times M$ 、 $D \times M$ 维的整数矩阵, C_{vm}^{VM} 为除当前词符 k 外单词 v 被赋予主题 m 的次数, C_{dm}^{DM} 表示除当前词符 k 外图像 d 中的词符被赋予主题 m 的次数^[9]。Gibbs 抽样算法求解 LDA 模型的详细描述如下^[19]：

1) k 从 1 循环到 N_d , 将图像中的主题 z_k 初始化为 1 到 M 之间的某个随机整数, 此为 Markov 链的初始状态;

2) k 从 1 循环到 N_d , 根据式 (16) 将词汇分配给主题, 获取 Markov 链的下一个状态;

3) 迭代步骤 2) 足够次数以后, 认为 Markov 链接近目标分布, 将 z_k 的当前值记录下来, 则 BoF 特征的第 m 维压缩特征计算公式为

$$P(z_m | d) = \frac{C_{dm}^{DM} + \alpha}{\sum_{m=1}^M C_{dm}^{DM} + M\alpha} \quad (17)$$

2.1.2 子空间方法

1) 主成分分析

主成分分析^[20] (PCA) 算法的思想来源于 KL 变换, 目的是通过线性变换寻找一组最优的标准正交向量基, 用它们的线性组合来重建原样本, 并使重建后的样本和原样本在均方意义下重构误差最小。假设 N 幅图像样本的 0 层 BoF 特征为 $B_1, B_2, \dots, B_N$, 其样本协方差矩阵为

$$\Sigma = \sum_{i=1}^N (B_i - \mu)(B_i - \mu)^T = BB^T \quad (18)$$

式中, $B = (B_1 - \mu, B_2 - \mu, \dots, B_N - \mu)$, μ 为样本均值。借助于奇异值分解定理间接地求解协方差矩阵 Σ 的特征向量, 选取 M 个最大特征值对应的特征向量组成变换矩阵 W_{PCA} , BoF 特征的压缩特征表示为 $W_{PCA}B$ 。

2) Fisher 线性判别分析

Fisher 线性判别分析^[21] 的目的是寻找一个最优鉴别矢量集, 使得映射后的图像样本的类间散布矩阵和类内散布矩阵的比值最大。假设图像样本的 0 层 BoF 特征 B 的类内散布矩阵与类间散布矩阵分别表示为 S_b, S_w , 线性映射后图像特征的类内散布矩阵为 $D_b = \Phi^T S_b \Phi$, 类间散布矩阵为 $D_w = \Phi^T S_w \Phi$, 则 Fisher 线性判别分析对应的优化准则函数为

$$J_F(\Phi) = \frac{\Phi^T S_b \Phi}{\Phi^T S_w \Phi} \quad (19)$$

对式 (19) 中 Φ 的各分量求偏导数并令其为零, 得到 $\sigma = \frac{\Phi^T S_b \Phi}{\Phi^T S_w \Phi}$, 则变换矢量的求解最终转化为求解方程的特征向量问题, 即

$$S_b \Phi = \sigma S_w \Phi \quad (20)$$

由于 BoF 模型中视觉单词数目往往很大, 样本的数目远远小于特征的维数, S_w 为奇异矩阵, 变换矢量 Φ 不能直接从式 (20) 中求解。本文采用 Fisherfaces 算法^[22] 解决这个问题, 首先利用 PCA 算法对 BoF 特征降维, 在此基础上作 Fisher 线性判别, 最后得到的变换矩阵为 $\Phi_F = \Phi_{PCA} \Phi_{FLDA}$, BoF 特征的压缩特征表示为 $\Phi_F B$ 。

2.1.3 字典学习算法

假设 $B = [B_1, \dots, B_N]$ 表示图像样本在 0 层下的 BoF 特征, 使用 $W = [w_1, w_2, \dots, w_M]$ 来表示字典向量集合, M 为字典中原子向量数目, w_m ($m = 1, 2, \dots, M$) 表示第 m 个原子向量, α_i 表示相应的稀疏向量, 字典学习算法的目标函数为^[23]

$$\min \sum_{i=1}^N \left(\frac{1}{N} \|B_i - W\alpha_i\|^2 + \gamma \|\alpha_i\|_1 \right) \quad \text{s. t.} \quad \|w_m\|^2 \leq \rho, \forall m = 1, 2, \dots, M \quad (21)$$

式中, γ 表示稀疏正则系数, 对式 (21) 使用分块协调下降算法进行求解, 固定字典 W , 式 (21) 转化为一个关于稀疏向量 α_i 的基于 L_1 范数的最小平方凸优化问题, 可以利用特征符号搜索算法^[24] 进行求解, 具体描述见附录; 固定稀疏向量 α_i , 式 (21) 转化为关于 W 的带二次约束最小平方凸优化问题, 求解方

法具体描述见附录。通过交替求解上述两个问题直到满足某一收敛的终止条件就可以得到 BoF 特征的压缩特征 α_i , 将这种使用所有类别训练图像样本进行字典学习获取压缩特征的方法记为通用字典学习算法 (GDL)。使用文献[25]中的方法构建监督字典学习算法 (SDL) 获取压缩特征, 即根据式 (21) 分别利用每类别的训练图像样本学习字典, 每个类别字典合并为监督字典, 然后求取图像 BoF 特征关于监督字典的稀疏表示。

2.2 基于压缩算法的图像分类步骤及分析

在 2.1 节中, 以图像的 0 层 BoF 特征为例介绍了压缩算法, 其与金字塔模型相配合用于图像分类任务中的步骤详细描述如下:

1) 提取图像局部特征, 训练图像的矢量集合记为 $X_T = \{X_1, X_2, \dots, X_N\}$;

2) 对 X_T 进行 K 均值聚类, 得到视觉词典 $V = \{v_1, v_2, \dots, v_K\}$;

3) 根据式 (1) (2) 或者 (3) 对 X 进行编码, 根据式 (4) (5) 或式 (6) 得到图像第 l 层的特征 B_l ;

4) 根据 2.1 节中描述的压缩算法步骤对训练图像的每层特征 B_l 进行压缩并串联起来, 代入式 (7) 中计算图像之间的核矩阵, 对 SVM 进行训练;

5) 根据压缩算法步骤对测试图像的每层特征 B_l 进行压缩并串联起来, 根据式 (9) 进行分类判别。

压缩算法的时间复杂度大小严重影响算法的实施, 主要取决于图像样本的数目 N 、BoF 特征的维数 K 、压缩特征维数 M 、迭代次数 L 、图像的类别数 C 、每个类别中图像样本数目 C_N 。对于 pLSA 算法, E 步骤的时间复杂度为 $O(NK)$, M 步骤的时间复杂度为 $O(K^3)$, 则迭代 L 次的时间复杂度为 $O(LNK)$ 与 $O(LK^3)$ 之和。LDA 算法中获取一个马尔可夫链的时间复杂度为 $O(NM^K)$, 则迭代 L 次的时间复杂度为 $O(LNM^K)$ 。PCA 算法和 FLDA 算法计算协方差矩阵的时间复杂度为 $O(NK)$, 而对 $K \times K$ 矩阵进行特征分析的时间复杂度为 $O(K^3)$, 则总的时间复杂度为 $O(NK)$ 与 $O(K^3)$ 之和。在字典学习算法中, 稀疏表示的时间复杂度为 $O(N^3)$, 更新字典时计算 $M \times M$ 逆矩阵的时间复杂度为 $O(M^3)$, 则迭代 L 次的 GDL 算法的时间复杂度为 $O(LN^3)$ 与 $O(LM^3)$ 之和; 当图像的类别数目为 C 时, SDL 算法的时间复

杂度为 $O(CLC_N^3)$ 与 $O(CLM^3)$ 之和。由此可见, 基于主题模型的压缩算法的时间复杂度最高, 基于字典学习的压缩算法次之, 基于子空间方法的压缩算法的复杂度最低。

3 实验

3.1 图像数据集

选取场景和一般物体分类任务中的经典图像数据集作为实验对象, 分别为: 1) 15 类场景数据集^[2], 它涵盖自然与人工场景共 15 种类别, 每种类别图像的数目不等, 在 200 ~ 400, 图像的平均尺寸为 250×300 , 全部为灰度图像。根据测试协议, 随机抽取每个类别中的 100 幅图像为训练图像, 剩余图像为测试图像, 同时在训练数据集中随机选取 100 幅图像作为验证集。2) SUN2012 数据集^[26] 是迄今类别数目最多的场景数据集, 共有 908 个场景类别, 图像数目总计 131 072, 全部为彩色图片, 尺寸大约为 250×300 。选取其中 50 个与人类生活息息相关的场景类别作为实验对象, 根据测试协议, 随机选取每个类别中的 50 幅图像为训练图像, 剩余图像为测试图像, 同时在训练数据集中随机选取 100 幅图像作为验证集。3) MSRCv2 物体数据集^[27] 共有 15 类物体, 每类物体有 30 幅图像, 图像的尺寸在 300×200 左右。根据测试协议, 选择其中的 9 类物体作为实验对象, 随机抽取每类中的 15 幅图像为训练图像, 剩余图像为测试图像, 同时在训练数据集中随机选取 30 幅图像作为验证集。4) Caltech101 数据集^[28] 是目前使用最为广泛的一般物体分类算法标准测试数据集, 由 101 类物体以及一个背景类别组成, 每个类别有 40 ~ 800 幅图像, 总计 9 146 幅图像, 图像的大小在 300×200 左右。根据测试协议, 随机抽取每个类别中的 30 幅图像为训练图像, 剩余图像为测试图像, 同时在训练数据集中随机选取 200 幅图像作为验证集。5) PASCAL2007 数据集^[29] 共有 20 类物体, 图片的数目总计 9 943。此数据集是为视觉物体识别挑战赛的参赛者准备的数据, 图像中的物体有丰富的视角、外观、尺度变化, 也有物体被遮挡的情况, 物体所处的场景复杂多变, 是一个很有挑战性的数据集。根据测试协议, 使用主办方事先准备的训练验证数据集训练分类器, 使用主办方事先准备的测试数据集测试算法的性能。每个数据集的示例图像见图 2。



(a) 15类场景数据集



(b) SUN2012数据集



(c) MSRCv2物体数据集



(d) Caltech101数据集



(e) PASCAL2007数据集

图2 数据集的示例图片

Fig. 2 Example images of datasets

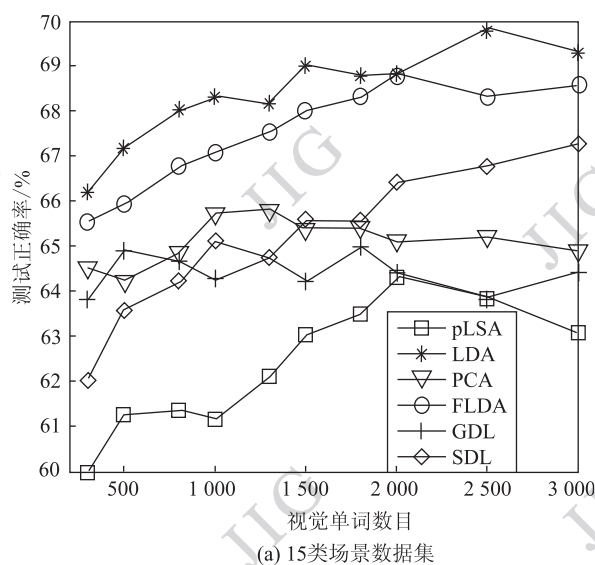
3.2 实验方法及结果分析

实验所使用计算机的性能为:CPU 为 2.1 GHz, 内存为 6 GB, 实验平台为 Matlab 7.9 软件, 使用 M 语言和 C 语言混合编程。pLSA、LDA、PCA、FLDA、GDL、SDL 分别表示基于概率隐含语义分析、隐含狄利克雷分配模型、主成分分析、Fisher 线性判别分析、通用字典学习、监督字典学习的 BoF 特征压缩算法。对图像进行划格, 图像块的大小为 16×16 , 采样步长为 8 个像素, 使用 SIFT 算子描述图像块的视觉特征, 使用 K 均值算法构造视觉词典, 使用 SVM 分类器, 选用线性核函数, 惩罚系数被设为 4, 多类分类选用一对一策略。

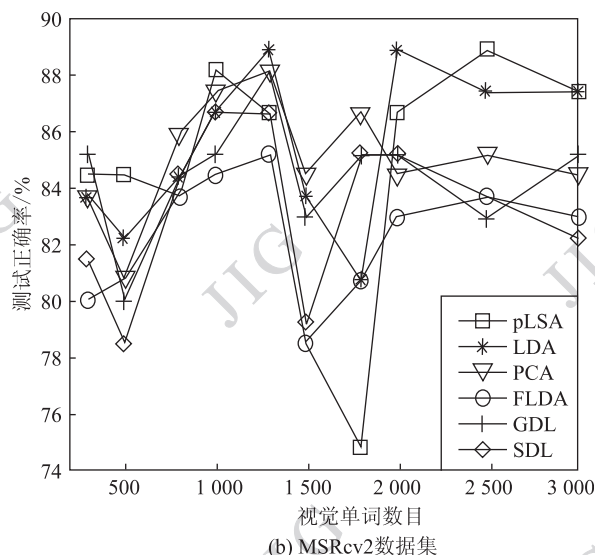
3.2.1 参数

视觉单词数目以及压缩特征维数是 BoF 特征压缩算法中的重要参数。验证参数对算法性能影响

的实验使用 15 类场景数据集和 MSRCv2 数据集, 视觉单词的数目分别为 300、500、800、1 000、1 300、1 500、1 800、2 000、2 500、3 000, 根据式 (1) 对局部描述矢量进行编码, 根据式 (4) 计算编码矢量的聚集特征, 在验证集上得到每个视觉单词数目下性能最优的压缩特征维数, 使用 SVM 对图像进行分类, 视觉单词数目对算法的影响曲线见图 3。根据图 3 的实验结果, 选择使压缩算法性能达到最优的视觉单词数目, 使用硬分配和总和统计方法计算每幅图像的 BoF 特征, 压缩特征维数分别为 20、30、40、50、60、70、80、90、100、120, 使用 SVM 对图像进行分类, 压缩特征维数对算法的影响曲线见图 4。图中的实验结果显示, 不同视觉单词数目下的分类性能之



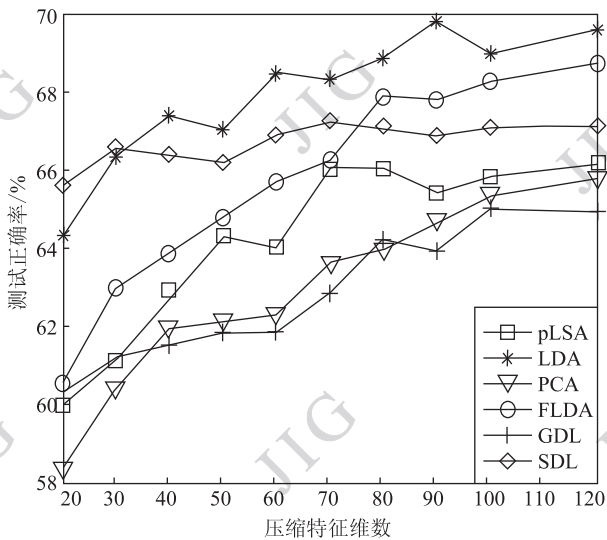
(a) 15类场景数据集



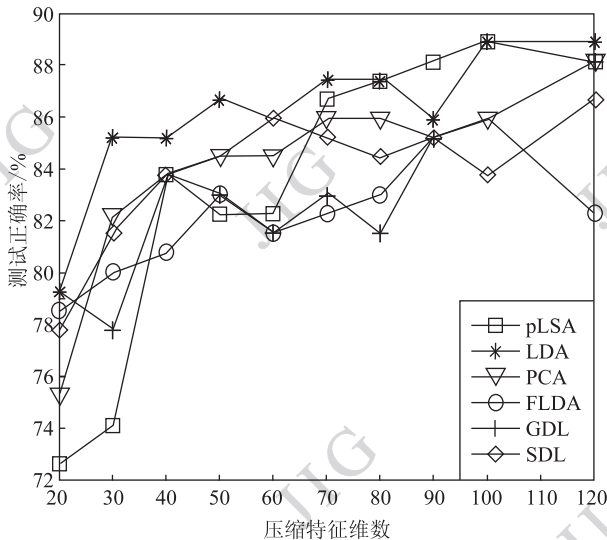
(b) MSRCv2数据集

图3 视觉单词数目性能影响曲线

Fig. 3 Performance curves of different visual word numbers



(a) 15类场景数据集



(b) MSRCv2数据集

图4 压缩特征维数性能影响曲线

Fig. 4 Performance curves of different wmpression dimensions

差要小于不同压缩特征维数下的分类性能之差,由此可见,压缩特征维数对压缩算法性能的影响更大。

3.2.2 编码和聚集方法

本节实验仍然使用15类场景数据集和MSRCv2数据集,视觉单词的数目分别为1500和1300,根据式(1)一(3)分别对局部描述矢量进行编码,根据式(4)计算编码矢量的总和统计值,需要说明的是,对于基于软分配编码的BoF特征,主题模型对其进行取整预处理操作。在验证集上得到压缩算法最优性能下的特征维数,使用SVM进行分类。不同编码方法对应的分类性能见表1,由表中的结果可知,相比于硬分配编码(Hard),软分配编码(Soft)和稀疏编码(Sparse)并没有普遍提高压缩算法的分类性能,多

数情况下反而降低了压缩算法的分类性能。并且除LDA算法中各种编码算法之间的性能之差高达7个百分点以外,其他压缩算法中各种编码方法之间的性能之差在1~3个百分点左右。验证聚集方法对压缩算法性能影响的实验使用硬分配方法对局部描述矢量进行编码,根据式(4)、式(6)分别计算编码矢量的总和统计值、最大统计值,在验证集上得到压缩算法最优性能下的特征维数,使用SVM进行分类。不同聚集方法对应的分类性能见表2,由表中的结果可知,除LDA算法外,相比于总和统计值,最大化统计值提高了压缩算法的分类性能,各种聚集算法之间的性能之差在3~5个百分点。由此可见,相比于编码方法,聚集方法对压缩算法性能的影响更大。

表1 不同编码方法对应的压缩算法分类性能

Table 1 Compression algorithms' performances of different coding method							
数据集	编码方法	pLSA	LDA	PCA	FLDA	GDL	SDL
15类场景	Hard	63.02	69.01	65.36	68.01	64.22	65.56
	Soft	63.79	62.91	64.19	66.53	64.92	66.73
	Sparse	65.67	68.94	64.62	66.63	62.91	66.09
MSRCv2	Hard	86.67	88.89	88.15	85.19	88.15	86.67
	Soft	88.15	84.44	86.67	83.70	84.44	84.44
	Sparse	88.89	84.44	84.44	85.93	85.19	83.70

表2 不同聚集方法对应的压缩算法分类性能

Table 2 Compression algorithms' performances of different pooling method							
数据集	聚集方法	pLSA	LDA	PCA	FLDA	GDL	SDL
15类场景	总和	63.02	69.01	65.36	68.01	64.22	65.56
	最大	65.33	65.53	69.31	69.55	69.35	68.91
MSRCv2	总和	86.67	88.89	88.15	85.19	88.15	86.67
	最大	87.41	80.00	91.11	87.41	85.19	91.86

3.2.3 空间金子塔

实验仍然使用15类场景数据集和MSRCv2数据集,视觉单词的数目分别为1500和1300,根据式(1)对局部描述矢量进行编码,为了验证各种压缩算法对图像高层特征性能的影响,将图像划分为4格,分别计算每个方格中编码矢量的最大统计值,然后将其串联起来作为图像的特征表示,在验证集上

得到压缩算法最优性能下的维数,使用 SVM 在此特征空间中对图像进行分类,各种压缩算法对应的分类性能见表3,由表中的结果可知,基于子空间方法的压缩算法在高层图像特征空间中的分类性能最优。

表3 压缩算法对图像第1层特征的性能影响

Table 3 Impact on images'1st level features of compression algorithms						
数据集	pLSA	LDA	PCA	FLDA	GDL	SDL
15 类场景	64.59	65.06	71.47	72.73	69.93	70.47
MSRCv2	82.96	81.48	86.67	87.41	84.44	85.93

3.2.4 压缩算法的性能对比

为了实现压缩算法的性能指标对比,实验使用3.1中描述的所有图像数据集,图像为3层金字塔模型,使用硬分配和最大统计值,在验证集上得到压缩特征维数的最优值,使用 SVM 分类,压缩算法的训练和测试正确率见表4,表4中还列出了获取压缩特征的运行时间。由表4的结果可知,基于子空间的压缩算法在所有图像数据集上都取得了最优的

表4 压缩算法性能对比

Table 4 Performances comparison of compression algorithms							
数据集	性能指标	pLSA	LDA	PCA	FLDA	GDL	SDL
15 类场景	训练	67.93	68.87	100	92.20	100	100
	测试	67.10	69.18	74.57	75.85	72.73	74.67
	时间/s	3698	352.2	15.79	15.37	66.19	21.58
SUN2012	训练	39.63	42.00	100	99.37	99.95	99.95
	测试	36.51	37.57	43.28	44.25	42.91	42.22
	时间/s	4509	492.0	24.56	32.54	124.7	39.24
MSRCv2	训练	85.93	84.44	100	99.26	100	100
	测试	85.19	83.70	91.85	92.59	88.15	92.59
	时间/s	834.2	8.26	0.56	0.58	2.36	1.89
Caltech101	训练	38.61	39.64	100	100	100	100
	测试	36.16	37.06	53.74	54.39	51.33	53.74
	时间/s	5814	500.8	61.09	70.8	47.71	162.1
PASCAL 2007	训练	39.42	42.51	58.47	55.95	58.52	46.81
	测试	37.12	40.23	47.48	49.25	47.57	47.60
	时间/s	8022	792.7	77.11	230.5	36.25	31.42

分类性能,基于字典学习的压缩算法的分类性能次之,基于主题模型的压缩算法的分类性能最差。在提取压缩特征的时间方面,基于主题模型的运行时间最长,除在 PASCAL2007 数据集上,基于字典学习的压缩算法运行时间少于基于子空间的压缩算法,在其他数据集上,基于子空间的压缩算法的时间开销最小。但是,基于子空间和字典学习的压缩算法的训练正确率却远远高于测试正确率,说明这两类压缩算法中存在过学习现象。

4 结 论

将已经出现的词袋特征压缩算法分为基于主题模型、基于子空间方法、基于字典学习算法3类进行详细介绍,给出了压缩算法与金字塔模型相配合用于图像分类的详细描述以及压缩算法的时间复杂度分析。以多个场景和一般物体分类图像数据集为实验对象,对压缩算法中的典型模型进行比较性仿真实验,实验结果表明,相比与视觉单词数目,压缩特征维数对算法性能的影响更大;相比与聚集算法,压缩算法对编码方法具有更好的鲁棒性;在所有压缩算法中,基于子空间方法的压缩算法在高层图像特征空间中的分类性能最优,在多个图像数据集上的分类性能最优,并且时间复杂度最低,但是这种算法存在过学习现象。因此,解决过学习问题以及挖掘更优的子空间方法用于 BoF 特征压缩是今后工作努力的方向。

参考文献 (References)

- [1] Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//Proceedings of the 22nd International Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE Computer Society, 2009: 1794-1801.
- [2] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[C]//Proceedings of the 19th International Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Computer Society, 2006: 2169-2178.
- [3] Eitz M, Hildebrand K, Boubekeur T, et al. Sketch-based image retrieval: benchmark and bag-of-features descriptors[J]. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(11): 1624-1636.
- [4] Vedali A, Gulshan V, Varma M, et al. Multiple kernels of ob-

- ject detection[C]//Proceedings of the 12th International Conference on Computer Vision. Tokyo, Japan: IEEE Computer Society, 2009: 606-613.
- [5] Signaraju D, Vidal R. Using global bag of features models in random fields for joint categorization and segmentation of objects[C]//Proceedings of the 24th International Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE Computer Society, 2011: 2313-2319.
- [6] Yang F, Lu H, Zhang W, et al. Visual tracking via bag of features[J]. Image Processing, 2012, 6(2): 115-128.
- [7] Nicosevici T, Garcia R. Automatic visual bag-of-words for online robot navigation and mapping[J]. IEEE Transactions on Robotics, 2013, 28(4): 886-898.
- [8] Ji R R, Yao H X, Liu W. Task-dependent visual-codebook compression[J]. IEEE Transactions on Image Processing, 2012; 21(4): 2282-2293.
- [9] Bosch A, Zisserman A, Munoz X. Scene classification using a hybrid generative/discriminative approach[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(4): 712-728.
- [10] Yang S, Zhao C X. The image classification algorithm based on latent dirichlet allocation model[J]. Computer Engineering, 2012, 38(14): 181-183. [杨赛, 赵春霞. 基于隐含狄利克雷分配模型的图像分类算法[J]. 计算机工程, 2012, 38(14): 181-183.]
- [11] Yang S, Zhao C X. A Fusing algorithm of bag-of-features and Fisher linear discriminative analysis in image classification[C]//Proceedings of the 2nd International Conference on Information Theory and Technology. Wuhan, China: IEEE Computer Society, 2012: 380-383.
- [12] Jiang Y G, Ngo C W, Yang J. Towards optimal bag-of-features for object categorization and semantic video retrieval[C]//Proceedings of the 6th ACM International Conference on Image and Video Retrieval. New York, USA: ACM, 2007: 494-501.
- [13] Wang Z, Feng J, Yan S, et al. Linear distance coding for image classification[J]. IEEE Transactions on Image Processing, 2013, 22(2): 537-548.
- [14] Liu L Q, Wang L, Liu X W. In defense of soft-assignment coding[C]//Proceedings of the 13th International Conference on Computer Vision. Barcelona, Spain: IEEE Computer Society, 2011: 2486-2493.
- [15] Thierault C, Thome N, Cord M. Extended coding and pooling in the HMAX model[J]. IEEE Transactions on Image Processing, 2013, 22(2): 764-777.
- [16] Liu Y Q, Yao H Y. Load-forecasting model based on normalized Gaussian pLSA collaborative filtering[J]. Control Theory & Applications, 2008, 25(5): 929-932. [刘粤钳, 姚红玉. 基于标准高斯 pLSA 协同过滤的用电量预测模型[J]. 控制理论与应用, 2008, 25(5): 929-932.]
- [17] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis[J]. Machine Learning, 2001, 42(1): 177-196.
- [18] Blei D M, Andrew Y N, Michael I J. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3(1): 993-1022.
- [19] Shi J, Fan M, Li W L. Topic analysis based on LDA model[J]. Acta Automatica Sinica, 2009, 35(12): 1587-1594. [石晶, 范猛, 李万龙. 基于 LDA 模型的主题分析[J]. 自动化学报, 2009, 35(12): 1587-1594.]
- [20] Ian J. Principal Component Analysis[M]. New York: Springer-Verlag, 1986.
- [21] Duda R, Hart P, Stork D. Pattern Classification[M]. New York: Wiley, 2001.
- [22] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720.
- [23] Guo J F, Li X. Structured dictionary learning based on group sparsity[J]. Journal of Image and Graphics, 2012, 17(11): 1347-1352. [郭景峰, 李贤. 基于群稀疏的结构化字典学习[J]. 中国图象图形学报, 2012, 17(11): 1347-1352.]
- [24] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithm[C]//Proceedings of the 20th Annual Conference on Neural Information Processing Systems. Vancouver B. C., Canada: MIT Press, 2006: 801-808.
- [25] Yang M, Zhang L, Yang J, et al. Metaface learning for sparse representation based face recognition[C]//Proceedings of the 17th International Conference on Image Processing. Hong Kong, China: IEEE Computer Society, 2010: 1601-1604.
- [26] Xiao J X, Hays J, Ehinger K A, et al. SUN database: large-scale scene recognition from abbey to zoo[C]//Proceedings of the 23rd International Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE Computer Society, 2010: 3485-3492.
- [27] Zhang Y, Chen T. Efficient kernels for identifying unbounded-order spatial features[C]//Proceedings of the International Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE Computer Society, 2009: 1762-1769.
- [28] Maji S, Berg A C, Malik J. Efficient classification for additive kernel SVMs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 66-77.
- [29] Everingham M, Gool L V, Williams C, et al. The PASCAL visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

附录 A

1) 稀疏系数的求解

设 $\Gamma(\alpha_i, W) = \sum_{i=1}^N \frac{1}{N} \|B_i - W\alpha_i\|^2$, 固定字典 W 时, 利用特征符号搜索算法求解稀疏系数 α_i 的步骤描述如下:

(1) 初始化 α_i, δ , 活动集合为空集合, 其中 δ 的取值为 $\{-1, 0, 1\}$, 表示 $\text{sign}(\alpha_i)$ 的函数值;

(2) 从 α_i 的零系数开始, 选择 $j = \arg\max_j \left| \frac{\partial \Gamma(\alpha_i)}{\partial \alpha_{ij}} \right|$, 当且仅当此值能够显著改变目标函数值时才将其放入到活动集合中, 即:

如果 $\left| \frac{\partial \Gamma(\alpha_i)}{\partial \alpha_{ij}} \right| > \gamma$, 将 δ 的值置为 -1, 并将 j 放入活动集合中;

如果 $\left| \frac{\partial \Gamma(\alpha_i)}{\partial \alpha_{ij}} \right| < -\gamma$, 将 δ 的值置为 1, 并将 j 放入活动集合中;

(3) 令 $\hat{W}$ 为 W 的子矩阵, 它只包含活动集合中元素对应的列向量; 令 $\hat{\alpha}_i$ 和 $\hat{\delta}$ 分别为 α_i 和 δ 的子向量, 只包含活动集合中元素对应的维度, 因此原问题被转化为一个无约束的二次优化问题:

$$\min \Gamma(\hat{\alpha}_i, \hat{W}) + \gamma \hat{\delta}^T \hat{\alpha}_i \quad (A1)$$

式(A1)关于 $\hat{\alpha}_i$ 的解析式为

$$\hat{\alpha}_{i_{\text{new}}} = W^T \hat{W}^{-1} \left(W^T \alpha_i - \frac{1}{2} \gamma \hat{\delta} \right)$$

在 $\hat{\alpha}_i$ 到 $\hat{\alpha}_{i_{\text{new}}}$ 之间进行离散线性搜索, 寻找每一个使符号发生变更的点, 更新 $\hat{\alpha}_i$ 到与之最近的点。移出活动集合中 $\hat{\alpha}_i$ 零系数对应的元素, 并更新 δ : $= \text{sign}(\alpha_i)$ 。

(4) 检查是否已经达到最优

①非零系数的最优条件为

$$\left| \frac{\partial \Gamma(\alpha_i)}{\partial \alpha_{ij}} \right| + \gamma \text{sign}(\alpha_{ij}) = 0, \forall \alpha_{ij} \neq 0$$

如果条件①不满足, 说明当前解不是一个稳定解, 转到步骤(3), 继续内循环, 否则检查条件②。

②零系数的最优条件为

$$\left| \frac{\partial \Gamma(\alpha_i)}{\partial \alpha_{ij}} \right| \leq \gamma, \forall \alpha_{ij} = 0$$

如果条件②不满足, 转到步骤(2), 否则返回 α_i 为所求的解。

2) 字典 W 的求解

当稀疏系数矩阵 α 固定时, 字典 W 的求解被转化为以下问题, 即

$$\min \|B - W\alpha\|^2$$

$$\text{s. t. } \sum_{k=1}^K w_{mk}^2 \leq \rho, \forall m = 1, 2, \dots, M \quad (A2)$$

上式为带有约束的最小二乘问题, 使用拉格朗日对偶法是解决此问题最有效的方法, 首先考虑拉格朗日函数为

$$L(W, \xi) = \|B - W\alpha\|^2 + \sum_{m=1}^M \xi_m \left(\sum_{k=1}^K w_{mk}^2 - \rho \right) \quad (A3)$$

式中, ξ_m 为对偶变量, 以 W 为变量求拉格朗日函数式(A3)的最小值, 可以得到拉格朗日对偶式为

$$W(\xi) = \text{tr}(B^T B - B\alpha^T(\alpha\alpha^T + \Lambda) + (B\alpha^T)^T - \rho\Lambda) \quad (A4)$$

式中, ξ 为对偶矢量, 即 $\xi = [\xi_1, \xi_2, \dots, \xi_M]$, $\Lambda = \text{diag}(\xi)$, 并且关于 ξ 的梯度和 Hessian 阵为

$$\frac{\partial W(\xi)}{\partial \xi_i} = \|B\alpha^T(\alpha\alpha^T + \Lambda)^{-1}e_i\|^2 - C$$

$$\frac{\partial^2 W(\xi)}{\partial \xi_i \partial \xi_j} = -2((\alpha\alpha^T + \Lambda)^{-1}(B\alpha^T)^T B\alpha^T (\alpha\alpha^T + \Lambda)^{-1})_{ij}$$

式中, $e_i \in \mathbf{R}^M$ 表示单位向量, 利用牛顿法求解对偶式(A4), 最后可得到字典 W 的解析式为

$$W^T = (\alpha\alpha^T + \Lambda)^{-1}(B\alpha^T)^T \quad (A5)$$