

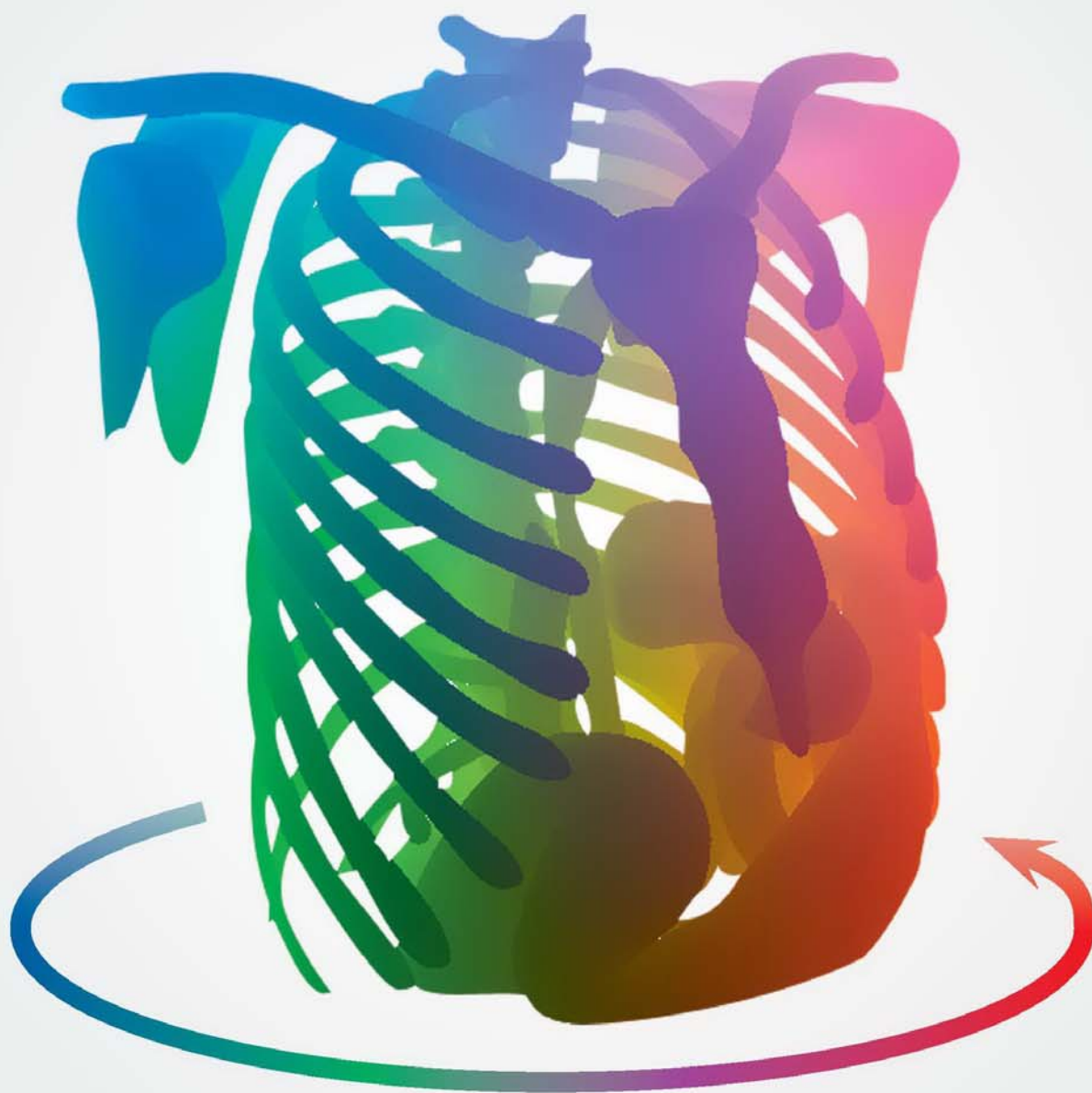
JOURNAL OF IMAGE AND GRAPHICS

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

2015
02
VOL.20

ISSN1006-8961
CN11-3758/TB



等值面光线跟踪 P193



第20卷第2期（总第226期）

2015年2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

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

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@radi.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



多尺度活动网格在云场景仿真中的应用(第0202页)



自然环境视频中基于显著鲁棒轨迹的行为识别(第0245页)



虚拟手抓持力觉生成算法真实性的评价(第0280页)

图像处理和编码

引入神经网络中间神经元的快速小波图像压缩

张海涛, 张永霖 0159

拟正态分布扩散的图像平滑

周先春, 汪美玲, 周林锋 0169

可选特征的快速分形图像编码

袁宗文, 鲁业频, 杨汉生 0177

图像分析和识别

融合多尺度码本的全局编码图像分类

董振宇, 赵杰煜, 祝军 0183

计算机图形学

结合K-D树和Shell的快速动态等值面光线跟踪法

罗月童, 石放放, 张伟, 朱会国 0193

多尺度活动网格在云场景仿真中的应用

范晓磊, 张立民, 张建廷, 徐涛 0202

第十届和谐人机环境联合会议专栏

低资源语言的无监督语音关键词检测技术综述

杨鹏, 谢磊, 张艳宁 0211

快速离散Curvelet变换域的图像融合

杨勇, 董松, 黄淑英 0219

高效视频编码中变换跳过模式的快速选择

王宁, 张永飞, 樊锐 0229

主动学习的多标签图像在线分类

徐美香, 孙福明, 李豪杰 0237

自然环境视频中基于显著鲁棒轨迹的行为识别

易云, 王瀚漓 0245

特征变换和数据集分块的行人比对

沈洋, 林巍晓, 殷惠清 0254

移动手持环境下触摸式目标选择技术对比

辛义忠, 李岩, 袁伟强, 郑谦 0264

虚拟环境的用户意图捕获

程成, 赵东坡, 卢保安 0271

虚拟手抓持力觉生成算法真实性的评价

杨文珍, 许艳, 颜传武, 吴新丽, 张昊, 孟闯 0280

交互式乐器演奏的六自由度力觉渲染方法

童号, 史有姣, 王党校, 张玉茹 0288

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Multi-scale moving grids and its application in clouds scene simulation(P0202)



Action recognition from unconstrained videos via salient and robust trajectory (P0245)



Validation the Authenticity of Virtual fingers Grasping Haptic Algorithm(P0280)

Image Processing and Coding

- Fast wavelet image compression introducing neural network of relay neuron
Zhang Haitao, Zhang Yonglin 0159
- Image smoothing algorithm based on matching normal distribution diffusion
Zhou Xianchun, Wang Meiling, Zhou Linfeng 0169
- Optional features fast fractal image coding algorithm
Yuan Zongwen, Lu Yepin, Yang Hansheng 0177

Image Analysis and Recognition

- Image classification based on global coding combined with multi-scale codebook
Dong Zhenyu, Zhao Jieyu, Zhu Jun 0183

Computer Graphics

- Fast isosurface ray tracing method by combining K-D tree and Shell
Luo Yuetong, Shi Fangfang, Zhang Wei, Zhu Huiguo 0193
- Multi-scale moving grids and its application in clouds scene simulation
Fan Xiaolei, Zhang Limin, Zhang Jianting, Xu Tao 0202

Column of HHME'2014

- Survey on unsupervised spoken term detection for low-resource languages
Yang Peng, Xie Lei, Zhang Yanning 0211
- Image fusion based on fast discrete Curvelet transform
Yang Yong, Tong Song, Huang Shuying 0219
- Fast transform skip mode decision for high efficiency video coding
Wang Ning, Zhang Yongfei, Fan Rui 0229
- Online multi-label image classification with active learning
Xu Meixiang, Sun Fuming, Li haojie 0237
- Action recognition from unconstrained videos via salient and robust trajectory
Yi Yun, Wang Hanli 0245
- Pedestrian re-identification based on feature transformation and dataset classification
Shen Yang, Lin Weiyao, Yin Huiqing 0254
- Comparison of target selection techniques in mobile handheld environment
Xin Yizhong, Li Yan, Yuan Weiqiang, Zheng Qian 0264
- Intent understanding for virtual environments
Cheng Cheng, Zhao Dongpo, Lu Baoan 0271
- Validation the authenticity of virtual fingers grasping haptic algorithm
Yang Wenzhen, Xu Yan, Yan Chuanwu, Wu Xinli, Zhang Hao, Meng Chuang 0280
- Interactive simulation of music instrument playing by a six degree-of-freedom haptic rendering approach
Tong Hao, Shi Youjiao, Wang Dangxiao, Zhang Yuru 0288

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2015)02-0183-10

论文引用格式: Dong Z Y, Zhao J Y, Zhu J. Image classification based on global coding combined with multi-scale codebook [J]. Journal of Image and Graphics, 2015, 20(2): 0183-0192. [董振宇, 赵杰煜, 祝军. 融合多尺度码本的全局编码图像分类[J]. 中国图象图形学报, 2015, 20(2): 0183-0192.] [DOI:10.11834/jig.20150204]

融合多尺度码本的全局编码图像分类

董振宇, 赵杰煜, 祝军

宁波大学信息科学与工程学院, 宁波 315211

摘要: 目的 词袋模型在图像分类领域中的分类效果主要受限于局部特征的量化误差。针对这一点, 提出一种融合多尺度码本的全局编码图像分类方法, 有效减少特征量化误差。**方法** 通过使用多尺度特征密集采样, 构建多尺度码本, 使码本具备一种层次结构, 通过充分利用图像特征的流形结构, 计算码本全局信息, 实现全局编码。通过本文方法得到的编码系数比较平滑和准确。最后使用多路径方法, 分别将不同尺度的特征表示进行级联, 得到最终的图像特征表示。这种特征表示具备了一定程度上的尺度不变性。**结果** 在 UIUC-8 和 Caltech-101 两个常用的标准图像数据集上进行测试, 分类准确率分别达到 88.0% 和 83.2%。**结论** 实验结果表明, 相比于基于固定尺度码本的局部编码方法, 本文方法在分类识别率方面有了显著提升。

关键词: 图像分类; 特征编码; 多尺度码本; 全局编码

Image classification based on global coding combined with multi-scale codebook

Dong Zhenyu, Zhao Jieyu, Zhu Jun

College of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: **Objective** The performance of the Bag-of-Words model in the field of image classification is limited mainly by the quantization error of the local feature. To reduce the quantization error of the local feature effectively, an image classification method based on global coding combined with multi-scale codebook is proposed. **Method** A global coding is implemented by utilizing fully the manifold structure of the image features and by computing the global information of the codebook. The coding coefficients obtained by the method are relatively smooth and accurate. Furthermore, a multi-path method is designed to integrate all feature representations to describe the image. To a certain extent, this method can achieve the scale invariance of feature representations. **Conclusion** Several experiments are conducted on two commonly used benchmark data sets, namely, UIUC-8 and Caltech-101, and the average classification accuracy rates reach up to 88.0% and 83.2%, respectively. **Result** Experimental results show that the proposed method improves the performance significantly compared with the fixed-scale locality-constrained coding methods.

Key words: image classification; feature coding; multi-scale codebook; global coding

收稿日期: 2014-08-11; 修回日期: 2014-10-16

基金项目: 科技部国际科技合作专项(2013DFG12810); 国家“十二五”科技支撑计划项目(2012BAF12B11); 浙江省国际科技合作专项(2013C24027)

第一作者简介: 董振宇(1990—), 男, 2012 年于宁波大学获计算机应用技术专业硕士学位, 主要研究方向为图像图形与机器学习。

E-mail: dzzy020727@163.com

Supported by: Key Projects in the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (2012BAF12B11)

0 引言

词袋模型(BoW),是一种基于图像块的特征学习模型,主要由局部特征提取^[1]、码本构建^[2-4]、特征编码^[5-7]、特征汇总^[8-10]4个环节构成。该模型将图像的局部特征量化表示为一系列无序的视觉单词,然后统计图像中所有视觉单词的分布来形成图像表示。词袋模型已经由大量实验证明了它在图像分类中的可用性和高效性,被广泛地应用于图像识别^[11]、场景分类^[12-13]、人体姿态估计^[14]等视觉领域。

传统词袋模型并没有过多地考虑局部特征描述子之间的空间关系,这种缺陷造成图像内容中的空间位置信息大量丢失。为了弥补这一缺陷,Lazebnik^[11]提出了空间金字塔表示模型(SPM),一定程度上扩展了图像内容的空间位置信息,增强了特征表示的描述能力,从而提升了词袋模型的性能。

除了图像空间位置信息的扩展,码本构建和特征编码对分类性能有着重要影响,也存在着一些不足。视觉码本的构建大多采用无监督聚类算法,如K均值、K奇异值分解(KSVD^[4])、流形学习(manifold learning)、均值漂移(mean shift)等算法。现有研究更多地集中在给定视觉码本的情况下,如何对局部特征描述子进行有效的量化编码和汇总,减少特征量化误差。

在编码过程中,最简单的编码方式是矢量量化编码(VQ),即一个特征描述子只能投影到码本中的一个视觉单词。当一个局部特征与多个视觉单词具有相似性时,特征量化误差会比较大。

因此,Gemert等人^[2, 15]提出了一种软量化编码(soft-assignment coding)的编码思想来解决这一问题,通过将局部描述子映射到码本中的所有单词,按照相似性程度来调整权值,共同构成该特征的编码系数,其中编码系数代表一个特征描述子到码本中所有视觉词汇的相似性程度。Yang等人在文献[5]中提出一种线性稀疏编码(ScSPM),一定程度上减少了量化误差。这种编码方式在码本构建和特征编码阶段比较复杂,速度较慢。Yu在文献[6]提出了局部线性编码(LLC),在稀疏编码的过程中控制编码的局部性,进一步提升了分类性能,Yu认为局部性比稀疏性更重要,局部性一定能保证稀疏,而稀疏

不一定能保证有局部性。Liu在文献[7]中提出使用局部软量化(LSC)方法,将局部性编码和软量化编码相结合,在保证分类准确率的同时,使编码方式更为简洁高效。

现有研究方案大多是对局部性编码方法的改进,没有关注图像特征中存在的流形结构,也没有重视码本结构的改善,造成特征表示不够平滑,量化误差较大。

提出一种融合多尺度码本的全局编码方法来弥补这一缺陷。首先对图像进行多尺度特征密集采样,形成多尺度码本,然后计算码本的全局结构信息,将不同尺度的特征分别用多尺度码本进行特征编码,在编码过程中加入全局信息,实现全局编码。最后通过多路径方法将不同的特征表示按照一定的方式级联,得到最终的全局图像特征表示。在UIUC-8和Caltech-101两个常用的数据集上进行的实验结果表明,本文方法可以有效减少特征量化误差,提高分类准确率。

1 相关工作

特征量化编码是在给定码本的情况下,确定特征编码系数的过程。对于一个 d 维图像特征 $\mathbf{x} \in \mathbf{R}^{d \times 1}$,其量化形式可表示为

$$\begin{aligned}\mathbf{x} &\approx \mathbf{B}\mathbf{u} \\ \mathbf{B} &= [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M] \\ \mathbf{u} &= [u_1, u_2, \dots, u_M]^T\end{aligned}$$

式中, $\mathbf{b}_i$ 为码本所包含的 d 维码本项, $\mathbf{B} \in \mathbf{R}^{d \times M}$ 为包含 M 个码本项的码本, u_i 为针对码本项 $\mathbf{b}_i$ 进行编码得到的编码系数, $\mathbf{u}$ 为所求的编码系数向量,即采用 $\mathbf{B}\mathbf{u}$ 来近似表示特征 $\mathbf{x}$ 。

传统特征编码的方法是采用VQ编码,量化误差较大。针对这一点,Gemert提出了一种软量化编码方案^[2, 15],能够更加精准地对特征进行表述,即

$$u_{ij} = \frac{\exp(-\beta \|\mathbf{x}_i - \mathbf{b}_j\|^2)}{\sum_{k=1}^n \exp(-\beta \|\mathbf{x}_i - \mathbf{b}_k\|^2)} \quad (1)$$

如式(1)所示,一个数据点可以投影到所有码本项,以此来弥补VQ编码单一映射的量化误差。然而它忽略了局部特征潜在的流形结构,从而导致编码过程出现较大量化误差。实践证明,这种编码方式不能真正提高分类效果。

在 ScSPM^[5] 中, 稀疏编码 (sparse coding) 被用于图像特征编码。Yang 等人认为, 一个数据点应该映射到少数几个码本项。它的目标函数如下:

$$\begin{aligned} \min_{U, B} \sum_{i=1}^n (\| \mathbf{x}_i - \mathbf{B} \mathbf{u}_i \|^2 + \lambda \| \mathbf{u}_i \|) \\ \text{s. t. } \| \mathbf{b}_k \| \leq 1, \forall k = 1, 2, \dots, K \end{aligned} \quad (2)$$

式中, $\mathbf{u}_i$ 为 $\mathbf{x}_i$ 所对应的编码系数向量。ScSPM 取消了一个特征只能投影到一个视觉单词的限制。然而它依然没有考虑到图像特征的流形结构, 相似的数据点可能投影到距离很远且完全不同的几个码本项上。

Wang^[6] 提出了一种局部线性编码 (LLC) 的编码方式, 将待量化的特征表示成 K 近邻视觉单词的自适应加权线性组合, 求解目标函数来最小化量化误差, 即

$$\begin{aligned} \min_{U, B} \sum_{i=1}^n (\| \mathbf{x}_i - \mathbf{B} \mathbf{u}_i \|^2 + \lambda \| \mathbf{d}_i \cdot \mathbf{u}_i \|^2) \\ \text{s. t. } \| \mathbf{b}_k \| \leq 1, \forall k = 1, 2, \dots, K \\ \mathbf{1}^T \mathbf{u}_i = 1 \\ d_{ij} = \exp\left(\frac{\text{dist}(\mathbf{x}_i, \mathbf{b}_j)}{\sigma}\right) \end{aligned} \quad (3)$$

式中, dist 函数表示向量间欧氏距离的平方。考虑到图像特征的流形结构, LLC 在保证局部性的基础上进行线性编码, 效率比 ScSPM 要好。

Liu^[7] 将局部性编码和软量化编码相结合, 提出一种局部软量化编码 (LSC) 编码方法。LSC 以局部相关性的形式表示编码系数。

局部相关性程度定义为

$$\begin{aligned} p_{ij} &= \frac{\exp(-\beta \hat{d}(\mathbf{x}_i, \mathbf{b}_j))}{\sum_{l=1}^n \exp(-\beta \hat{d}(\mathbf{x}_i, \mathbf{b}_l))} \\ \hat{d}(\mathbf{x}_i, \mathbf{b}_j) &= \begin{cases} \text{dist}(\mathbf{x}_i, \mathbf{b}_j) & \mathbf{b}_j \in KNN(\mathbf{b}_i) \\ 0 & \text{其他} \end{cases} \end{aligned} \quad (4)$$

式中, KNN 函数表示与目标向量欧氏距离最近的 K 个向量的集合。

Qin^[13] 提出了一种多路径编码的思想, 对图像提取不同尺度的特征描述子, 分别对其进行编码和汇总, 使目标特征表示更加线性可分。然而 Qin 在码本构建阶段将所有尺度的特征进行聚类, 这样会导致某些尺度的码本项大量丢失, 码本缺少一种层次结构, 导致码本的描述能力下降, 从而影响分类效果。

Duan^[16] 在 Qin 的基础上提出使用 KSVD 方法训练码本, 在码本构建阶段分别形成不同尺度的码本, 码本过程在各自的码本上进行, 计算效率和分类效果有所提升。

以上几种编码方式都没有充分利用多尺度特征的层次结构以及流形结构, 导致编码过程中出现较大量化误差, 从而导致分类效果下降。

考虑到这一缺陷, 提出一种融合多尺度码本的全局编码方法, 使码本具有层次结构, 使特征编码更为平滑, 有效减少特征量化误差, 提高分类效果。

2 多尺度全局编码

在常见的局部性编码方式中, K 近邻值的大小往往对分类效果有较大影响。Wang 在文献[6]中提到, 小幅增大 K 近邻值, 可以使特征编码相对平滑, 有效减少特征量化误差。Liu 在文献[7]中指出, 图像特征存在着流形结构, 如图 1 所示, 假设图像特征维数为 3, 可以将特征空间想象成 3 维空间, XYZ 坐标轴分别表示空间中的 3 维坐标轴。如果两点间欧氏距离足够近, 就可以近似等同于流形距离, 如果两点间距离较远, 就会导致流形距离计算出现偏差。随着 K 近邻值的增大, 一方面由于有更多的编码项, 可以有效减少量化误差, 另一方面由于数据点与码本项之间距离变远, 导致流形距离计算出现偏差, 产生量化误差。

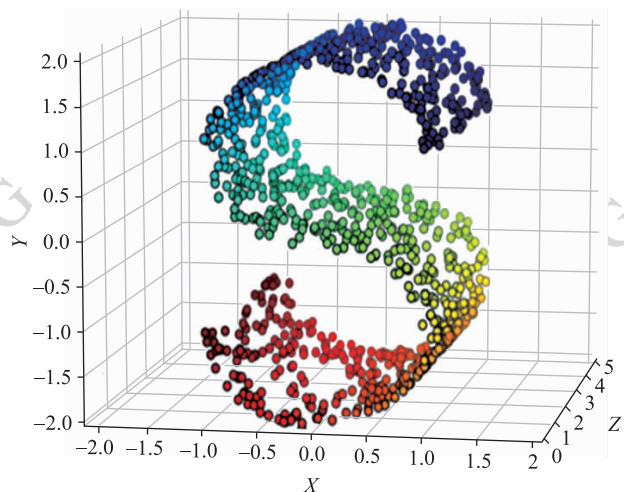
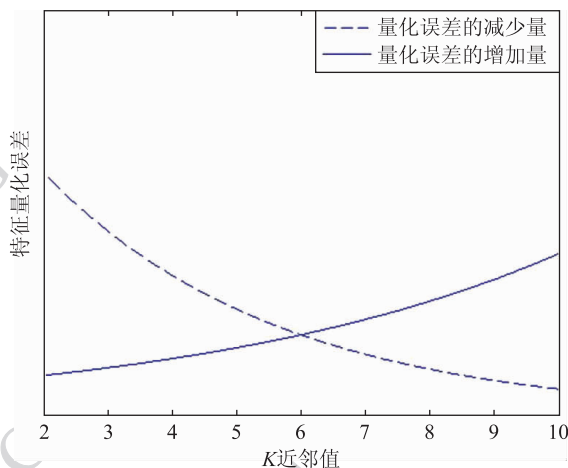


图 1 图像特征的流形结构

Fig. 1 The Manifold structure of image features

图 2 为 K 值增大造成的影响示意图, 虚线表示

因编码项增加造成的量化误差减少,而实线表示因流形距离计算出现偏差造成的量化误差增加。当 K 值较小时,量化误差的减少量大于增加量,因此总体误差下降,而当 K 值增大到一定程度后,量化误差的减少量小于增加量,总体误差上升。这也正好解释了 Wang 和 Liu 的结论:局部性编码只有找到一个合适的 K 值才能有比较好的效果。

图2 K 值增大造成的影响Fig. 2 The impact of the increase of K value

受到 Qin 混合码本的影响以及 Liu 的流形计算的思路,提出一种多尺度全局编码的方法,利用多尺度码本,有效抑制由于流形结构造成的计算偏差,而利用全局编码,使更多的码本参与编码,从而减少局部特征的量化误差。

2.1 多尺度码本构建

现实中的图像包含复杂多变的高维信息,这些信息往往随着尺度、视距和场景的变化而变化。单一尺度的码本只能捕捉到同一尺度图像块的无序特征,存在较大几率丢失掉目标的结构信息以及图像块之间的上下文信息,导致码本对图像整体目标的描述能力下降,造成特征量化编码产生偏差。

本文提出使用多尺度码本,将多个不同尺度的码本进行互补。假定采用 L 种不同大小的尺度,分别对图像按不同尺度进行密集特征采样,形成 L 级尺度不同的特征集,通过对各个尺度的特征集进行聚类,完成码本构建,形成 L 级代表不同尺度特征的码本 $B_1, B_2, \dots, B_L$,最后将所有码本拼接起来,形成多尺度码本 B 。

分别从总体效果与实际效果两方面来描述多尺度码本。图3为单尺度码本与多尺度码本的总体效

果对比示意图。图3(a)中的白色码本项为常规尺度码本项,而图3(b)则在图3(a)的基础上增加了其他尺度码本项,分别用灰色和黑色来表示。假定 K 值固定为3,单尺度码本在编码过程中数据点离码本项较远,有可能导致流形距离计算出现偏差,而造成量化误差,而多尺度码本则可以利用所有的码本项进行编码,码本较为密集,编码距离比较近,从而抑制量化误差。

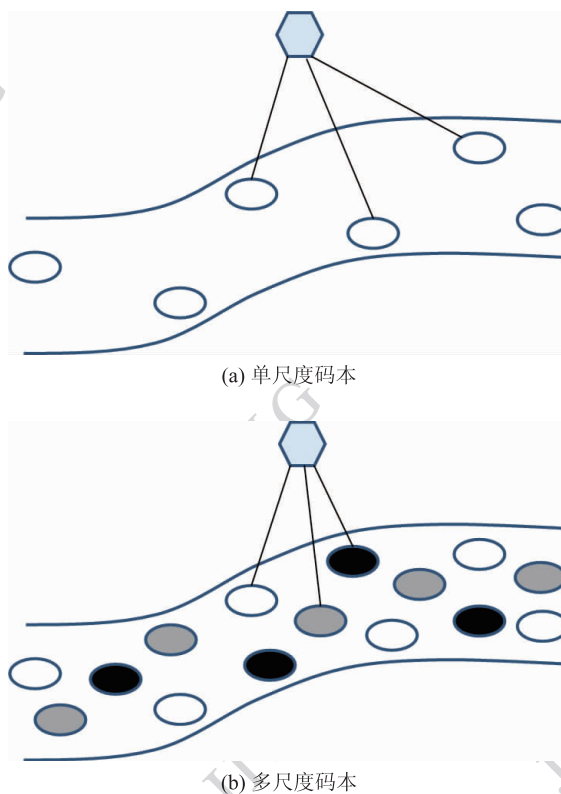


图3 单尺度码本与多尺度码本的总体效果对比

Fig. 3 The overall effect comparison of single-scale codebook and multi-scale codebook

图4为单尺度码本与多尺度码本的实际效果对比示意图。图4(a)表示为单尺度码本编码过程,小型尺度的码本包含了目标特征的大量细节信息,目标特征指的是图像块的形状特征,类似地,中等尺度和大型尺度的码本则分别包含了目标特征的中层信息和结构信息。图4(a)下方的图像块代表待编码特征,红色、黄色以及蓝色虚线分别指示仅用小型、中等和大型尺度码本得到的码本项,省略号表示码本中的其他码本项。对于此待编码图像块,背景、脚和球具有代表性。不难发现,小型尺度码本项中只有背景图像块相似程度较高,中等尺度中只有脚的相似性较高,而大型尺度中只有球的相似性较高。

仅用单尺度码本往往只能得到局部最优码本项,而全局最优码本项往往隐藏在各个不同尺度码本中。仅用单尺度码本描述,可能会造成描述偏差。图4(b)为多尺度码本编码过程示意图,将不同尺度码本进行级联得到多尺度码本,通过多尺度码本则可以在所有码本中寻找最合适的码本项,图4(b)中实线所指示的码本项为全局最优的。这种多尺度码本能够在一定尺度范围内包含目标各个层次的信息,使码本具有一定的层次结构,能够更加精准地描述目标对象。

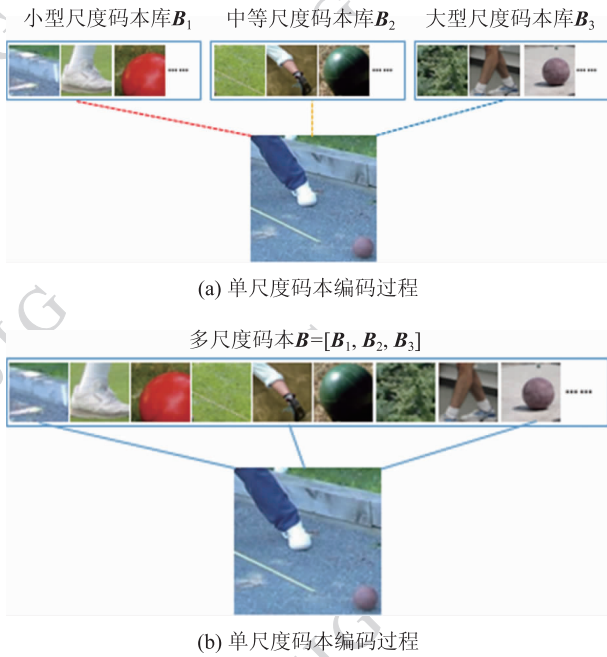


图4 单尺度码本与多尺度码本的实际效果对比
Fig. 4 The actual effect comparison of single-scale codebook and multi-scale codebook

2.2 全局编码

考虑到图像特征的流形结构,欧氏距离只能近似的表示一个点到附近 K 个点的流形距离。当 K 值取得过小的时候,会造成特征表示的不平滑,造成一定量化误差,而当 K 取值过大的时候,又会导致流形距离计算出现偏差,同样会造成量化误差。因此局部性编码方法的编码效果很大程度上依赖于 K 值的选取。此外由于局部性编码 K 值一般取得较小,导致码本中只有极少数码本项参与编码,这就极大地浪费了多尺度码本的性能,造成特征编码不平滑。

全局编码的核心在于通过码本全局信息的计算,使原本散乱无关的码本项彼此相关。数据点能够通过少数几个点间接映射到码本中的其他码本

项,从而模拟出一种全局相关性。

全局编码算法框架主要由3个模块组成:码本全局信息计算、全局编码、多路径编码。

2.2.1 码本全局信息计算

采用非线性近似估计的方法来获得码本之间的全局相关性。

高斯核方法经常被用于流行学习,选用它用于计算码本项之间的局部相关性,即

$$W(i, j) = \begin{cases} \exp\left(-\frac{\|b_i - b_j\|^2}{\sigma}\right) & b_j \in KNN(b_i) \\ 0 & \text{其他} \end{cases} \quad (5)$$

$$P = D^{-1}W \quad (6)$$

式中, D 是一个对角矩阵, $D_{ii} = \sum_j W_{ij}$, 归一化操作使得 $\sum_j P_{ij} = 1$, $P_{ij} = p(b_j | b_i)$ 表示码本项 b_j 归属于 b_i 的后验概率。如果 P_{ij} 不为零,那么表示 $b_j \in KNN(b_i)$ 。如果 $b_j \notin KNN(b_i)$, 无法直接表达两点的相关性,但可以通过 t 步迭代^[17]的方式间接地计算两点相关性。

$$p^{(t)}(b_k | b_i) = \sum_{l=1}^M p(b_k | b_l) p^{(t-1)}(b_l | b_i) \quad (7)$$

式中, M 为码本项个数, $p^{(t)}(b_k | b_i)$ 表示码本项 b_k 通过 t 步迭代间接归属于 b_i 的概率。

可以发现,矩阵 P^t 表示码本项之间 t 步相关性,这种相关性主要取决于 t 的大小。当 $t=1$ 时, P^t 就退化为码本项间的局部相关性,当 $t>1$ 时, P^t 被认为是一种 t 步非局部相关性。而对于码本项之间的任意两点,近似表达一种全局相关性为

$$S = \frac{1}{t} \sum_{e=0}^{t-1} P^e \quad (8)$$

式中, S 为全局相关矩阵, S_{ij} 表示 t 步内任意码本项 b_j 与 b_i 的相关性。

2.2.2 局部编码和全局编码

完成码本全局信息计算后,需要对特征描述子进行局部编码。编码方式可以是任意的局部编码方法,选用 LSC 方法作为局部编码方法,即

$$u_{ij} = p(b_j | x_i) = \begin{cases} \frac{1}{Z} \exp(-\beta \|x_i - b_j\|^2) & b_j \in KNN(x_i) \\ 0 & \text{其他} \end{cases} \quad (9)$$

式中, Z 是一个归一化常数, b_j 为多尺度码本 B 中一

个码本项, u_i 即为对应数据点 x_i 编码后得到的局部编码系数。

对于数据点 x_i 的一个非邻近码本项 b_k , 可以通过全局信息以迭代的方法间接得到两者之间的相关性, 即

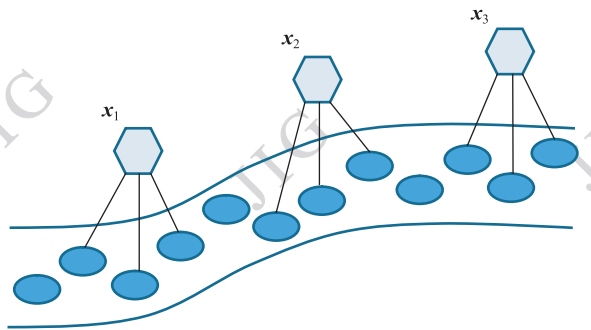
$$g_{ik} = \sum_{l=1}^M p^{(< i)}(b_k | b_l) p(b_l | x_i) = \sum_{l=1}^M S_{lk} u_{il} \quad (10)$$

式(10)可以等价地表示为

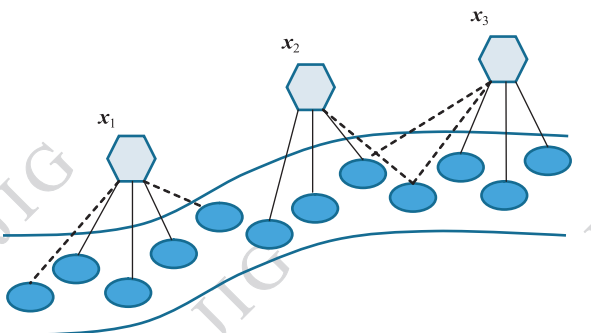
$$\mathbf{g}_i = \mathbf{S}^T \mathbf{u}_i \quad (11)$$

式中, $\mathbf{g}_i$ 就是最终得到的全局编码系数。

全局编码充分利用了多尺度码本的流形结构, 使编码过程变得更加平滑。图5为局部编码与全局编码的结果对比示意图, 带下划线的编码系数为局



(a) 局部编码



(b) 全局编码

图5 局部编码与全局编码

Fig. 5 Local coding and global coding

部编码中编码系数的相应位置, 带虚线的码本项为通过全局编码间接相关码本项。相对于局部编码, 全局编码拥有更多的码本项参与编码, 编码系数过渡得更加自然, 能够更好地表达数据点, 同时不会导致流形距离计算出现偏差。

3 系统实现

本节将详细介绍本文方法在图像分类系统的实现过程, 总体系统流程如图6所示。

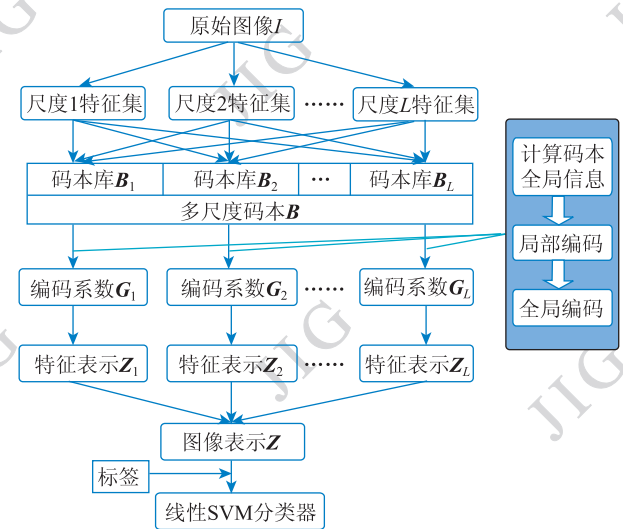


图6 本文方法的流程图

Fig. 6 The flowchart of our method

具体实验步骤如下:

1) 多尺度特征提取。获取图像的多尺度特征之前, 首先需要确定尺度的个数以及各个尺度的大小。假设已按从小到大的顺序划分了 L 级尺度, 接下来分别需要对每一个尺度进行特征提取。局部特征采用密集采样的 d 维特征描述子, 即按照光栅扫描顺序, 以固定步长像素等间隔地从每幅图像中提取各个尺度像素块的特征, 形成各自特征库 $C_1, C_2, \dots, C_L \in \mathbb{R}^d$ 。

2) 多尺度码本构建。本文提出采用多尺度码本, 即将特征库 $C_1, C_2, \dots, C_L$ 各自采用聚类算法进行聚类, 形成各自的码本 $B_1, B_2, \dots, B_L$, 假定码本大小分别为 $m_1, m_2, \dots, m_L$, 将这些码本级联, 构建多尺度码本 $B = [B_1, B_2, \dots, B_L] \in \mathbb{R}^{d \times m}$, $m = m_1 + m_2 + \dots + m_L$ 。值得一提的是, 在聚类过程中, 考虑到码本构建的计算复杂度, 结合文献[3]的结论, 采用了较为普遍的 K 均值聚类算法。针对码本的各种优化, 其

他一些无监督聚类算法,如 KSVD^[4]、均值漂移等方法,同样可以用于码本构建。

3)全局编码。采用一种非线性近似估计的方法,通过计算码本项间的局部相关性逐渐过渡到全局相关性,在编码阶段加入全局信息,实现全局编码。

计算局部相关矩阵 W ,即

$$W(i,j) = \begin{cases} \exp\left(-\frac{\|b_i - b_j\|^2}{\sigma}\right) & b_j \in KNN(b_i) \\ 0 & \text{其他} \end{cases}$$

对 W 进行归一化操作,即

$$D_{ii} = \sum_j W_{ij} \\ P = D^{-1}W$$

计算全局相关矩阵

$$S = \frac{1}{t} \sum_{e=0}^{t-1} P^e$$

计算局部编码

$$u_{ij} = p(b_j | x_i) = \begin{cases} \frac{1}{Z} \exp(-\beta \|x_i - b_j\|^2) & b_j \in KNN(x_i) \\ 0 & \text{其他} \end{cases}$$

计算全局编码

$$g_i = S^T u_i$$

4)多路径编码。采用多路径编码方式对多尺度特征进行特征量化编码。对一个单一尺度的特征描述子 $x_i \in C_l^d (l=1,2,\dots,L)$,以全局编码方法计算其编码系数 $g_i^{(c_l)}$ 。同理,对于其他所有尺度的特征描述子 $x_i = \{C_1^d, \dots, C_L^d\}$ 进行处理,可以得到其他尺度的编码系数 $g_i^{(c_1)}, \dots, g_i^{(c_L)}$,完成编码。

5)特征汇总。对应不同的尺度大小,一幅图像分别有数量不等的图像块,对一个图像块进行特征编码可以得到一个编码系数。设 $G^{(c_l)}$ 为该图像中尺度 l 下的编码系数集合,对该编码系数矩阵使用特征汇总函数 $\mathcal{F}$ 进行处理,即 $z = \mathcal{F}(G)$,得到该图像的特征表示 z 。常见汇总方式有均值汇总(average pooling),以及最大值汇总(max pooling)。由于 max pooling 能够一定程度上过滤掉噪声,选用它来得到特征的显著性表示。

$$z_j^{(c_l)} = \max \{ |G_{1j}^{(c_l)}|, |G_{2j}^{(c_l)}|, \dots, |G_{M_j}^{(c_l)}| \}$$

式中, $z_j^{(c_l)}$ 是 $z^{(c_l)}$ 中的第 j 个元素, M 为该图像在尺度 l 下图像块的个数,而 $z^{(c_l)}$ 为该图像在尺度 l 下的特征表示,将各个尺度下的特征表示进行级联,即可

得到最终的图像表示,即

$$z = [z^{(c_1)}, z^{(c_2)}, \dots, z^{(c_L)}]$$

4 实验及结果分析

为验证本文方法的有效性,在 UIUC-8 和 Caltech-101 两种不同的图像数据集上对分类效果进行了测试,并将它与目前主流算法作比较。

在特征提取阶段,以固定步长 6 像素等间隔地从每幅图像中提取各个尺度的图像块。对于每个局部图像块,选用 128 维的 SIFT^[1] 描述符作为其特征。在码本学习阶段,为了简化设定,每个尺度的字典大小统一设定为 1 024。在分类器设计阶段,采用台湾大学林智仁博士等开发的 SVM 软件包 LIBLINEAR^[18] 进行特征的训练和测试。

4.1 影响因素分析

4.1.1 尺度级数

图像的每个层次的特征分别集中在大小不同的尺度上。尺度越多,图像特征所包含的目标信息越丰富,所消耗的计算资源也越多。如何选取合适的尺度级数,在保证分类效果显著提高的同时,尽可能减少计算消耗,是需要考虑的问题之一。

在这个问题上,做了一些相关测试,如图 7 所示,在初始阶段,随着尺度级数的增加,分类准确率近似线性增长,证明本文方法对于分类效果有较大提升。保持尺度级数继续增加,当尺度级数超过 3 时,分类准确率的提升开始减弱,当尺度级数超过 5 时,准确率的提升已经比较微弱了。这表明,最佳的尺度级数选择在 3~5。为了节约计算资源,本文选用 3 级尺度,分别为 16×16 , 24×24 和 36×36 。

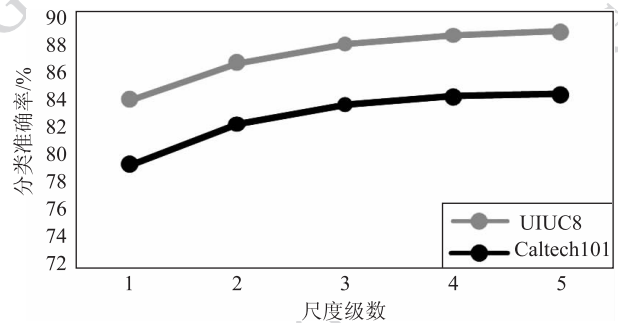


图7 尺度级数对于图像分类准确率的影响

Fig. 7 The impact of the number of scales on accuracy rate of image classification

4.1.2 K 近邻值的大小

采用 LSC 编码方案作为局部性编码方案, K 近邻值的选取对于分类效果的影响比较明显。选择合适的 K 值, 使分类准确率达到最大值, 也是本文所需要考虑的问题之一。针对这一问题, 使用不同的 K 值分别局部编码和全局编码进行比较, 图 8 为不同的 K 值对两种编码方式的测试结果。

可以发现, 局部编码对 K 值的变化更加敏感, 而全局编码对 K 值的变化相对鲁棒。当 K 值相对较小的时候, 局部编码效果并不理想, 而进行全局编码后分类效果有较大提升。当值达到最优值后, 全局编码的作用开始下降, 当 K 值过大时, 全局编码对分类效果已不再有提升。分析其原因, 可能是过大的 K 值会导致局部编码的计算出现偏差, 从而影响到全局编码。在不清楚最优值的情况下, 一个较小的 K 值可以更好地发挥全局编码的优势。

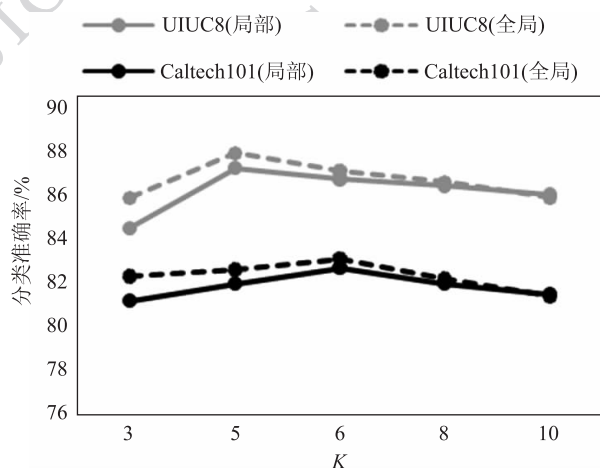


图 8 不同 K 近邻值对于图像分类准确率的影响

Fig. 8 The impact of the value of K on accuracy rate of image classification

4.1.3 多尺度码本

Duan^[16]使用的多路径编码方式都是建立在各自固定尺度的码本上, 分类效果受限于码本的描述能力。本文提出多尺度码本, 使码本具备了一定的层次结构, 并结合全局编码方式, 进一步改善了多尺度模型。在 UIUC-8 数据集上进行测试, 发现多尺度码本能有效提升模型的识别能力, 表明多尺度码本效果显著。实验结果如表 1 所示。

表 1 模型有无多尺度码本的实验对比

Table 1 Experimental comparison of whether the multi-scale codebook exists in the model

	Duan ^[16]	本文无多尺度码本	本文有多尺度码本
准确率/%	85.49 ± 1.17	86.5 ± 0.8	88.0 ± 0.8

4.1.4 复杂度分析

利用多尺度码本以及全局编码来减少特征量化误差, 提高分类准确率, 相对于固定尺度局部编码方法, 本文方法需要增加额外的计算资源。

在多尺度码本方面, 本文方法需要对不同尺度的特征进行处理, 与传统单尺度方法比较, 计算资源消耗比较大, 为单尺度方法的 L 倍, 在单线程状态下分类效率较低。同时由于各个尺度之间的工作相互独立, 在多线程状态下, 可以通过并行计算完成多尺度特征处理, 提高分类效率。

全局编码的工作量主要集中在码本全局信息计算以及全局编码本身这两块, 其中码本全局信息计算为特征编码的前期处理工作, 只需要执行一次, 其时间复杂度为 $O(n^3)$, n 表示码本的大小。全局编码本身的工作是对局部编码系数的转换, 每进行一次编码就执行一次, 其时间复杂度为 $O(n^2)$ 。

4.2 数据集测试

4.2.1 运动姿势估计

使用 UIUC-8^[19]数据集来测试本算法在运动姿势估计中的效果, UIUC-8 中共包含 8 个类别的 1 579 幅运动姿势图像。

根据 UIUC-8 的实验设置^[19], 分别在每一类随机抽取 70 幅图像作为训练样本, 在剩余图片中继续随机抽取 60 幅作为测试样本。

将 K 值设定为 5, 进行 30 次随机测试实验。实验结果如表 2 所示, 不难发现, 本文提出的多尺度码本在融合 LSC 编码方式后, 分类准确率取得了较大的提升。实验结果表明, 相对于 Duan^[16]提出的多路径方法, 本文方法的提升十分显著。

4.2.2 目标识别

使用 Caltech-101^[23]数据集来测试本文方法在目标识别中的效果, Caltech-101 中共包含 101 类目标图像以及 1 类背景图像, 每个类别包含 40 ~ 800 幅左右的图像, 总计为 9 144 幅场景图像。

表2 不同算法在 UIUC-8 上的准确率比较

Table 2 Comparison of accuracy rates of different algorithm on UIUC-8 database

		/%	
算法	平均准确率	算法	平均准确率
OB ^[19]	76.3	LSC ^[7]	82.3
SIFT + GGM ^[20]	73.4	ScSPM ^[5]	82.7
HMP ^[21]	85.7	Duan ^[16]	85.49
HP ^[22]	87.2	本文方法	88.0

根据 Caltech-101 的实验设定, 分别在每一类随机抽取 30 幅图像作为训练样本, 而剩余图像中继续随机抽取 50 幅作为测试样本。将 K 值设定为 6, 进行 30 次随机测试实验, 取平均值作为测试结果。实验结果如表 3 所示。实验表明, 相对于单一尺度的 LSC 局部编码方式, 本文方法优势明显。

表3 不同算法在 Caltech-101 上的准确率比较

Table 3 Comparison of accuracy rates of different algorithm on Caltech-101 database

		/%	
算法	平均准确率	算法	平均准确率
SIFT + T ^[24]	67.7	HSC ^[25]	74.0
Local NBNN ^[26]	71.9	HMP ^[21]	76.8
LC-KSVD ^[27]	73.6	MHMP ^[28]	82.5
LLC ^[6]	73.4	LSC ^[7]	78.6
Ask locals ^[9]	77.1	本文	83.2

为了进一步评估本文方法的有效性, 继续在 Caltech-101 上进行后续实验。根据文献[6]中的测试方案, 严格控制训练集的数量, 其余设定不变。训练集大小从每类 5 幅图像逐渐增长到 30 幅图像, 剩余图像全部作为测试样本。表 4 为各分类算法在 Caltech-101 数据集上分类效果的比较结果。

可以看出本文方法在严格控制训练数量的情况下仍然具有很好的性能, 在训练集大小仅为 5 的情况下, 仍有 60.40% 的分类准确率。随着训练集大小的缓慢增加, 分类效果呈线性增长, 当训练集大小增长到 30 时, 准确率达到 83.2%, 增长幅度达到了 22.8%。这表明, 本文方法的学习效率比较高, 在训练集不足的情况下可以得到较高的分类准确率, 在训练集足够的情况下, 分类效果更为显著。

表4 不同训练集大小在 Caltech-101 上的准确率比较

Table 4 Comparison of accuracy rates of different size of training set on Caltech-101 database

算法	训练集大小						/%
	5	10	15	20	25	30	
Local NBNN ^[26]	—	—	66.1	—	—	71.9	
LLC ^[6]	51.2	60.0	65.4	67.7	70.2	73.4	
LSGC ^[29]	54.0	63.9	68.7	71.6	73.7	75.1	
CRBM ^[30]	56.7	66.4	70.7	73.5	75.4	77.4	
本文	60.4	71.9	76.5	79.4	81.2	83.2	

5 结 论

根据不同尺度码本之间存在的互补性, 提出了一种融合多尺度码本的全局编码的图像分类方法。多尺度码本的构建, 实现了图像特征的多尺度扩展, 增强了码本的描述能力。通过全局编码, 使特征表示更加平滑准确, 有利于分类器进行区分, 从而提高分类准确率。实验结果表明, 相对于单尺度局部特征编码方式, 本文方法在分类性能上得到了显著提升。尽管如此, 编码效率仍有待改善。因此, 在多路径方法中加入并行技术, 提升编码效率, 也是下一步研究计划的重点。

参考文献 (References)

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International journal of computer vision, 2004, 60(2): 91-110.
- [2] Gemert J C V, Geusebroek J M, Veenman C J. Kernel codebooks for scene categorization [C]// Proceedings of the 10th European Conference on Computer Vision. Marseille, France: Springer, 2008: 696-709.
- [3] Jurie F, Triggs B. Creating efficient codebooks for visual recognition [C]// Proceedings of the 10th International Conference on Computer Vision. Beijing, China: IEEE, 2005: 604-610.
- [4] 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.
- [5] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009: 1794-1801.
- [6] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification [C]// Proceedings of the IEEE Computer

- Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA; IEEE, 2010; 3360-3367.
- [7] Liu L, Wang L, Liu X. In defense of soft-assignment coding [C]// Proceedings of the 13th International Conference on Computer Vision. Barcelona, Spain; IEEE, 2011; 2486-2493.
- [8] Feng J, Ni B, Tian Q, et al. Geometric l p-norm feature pooling for image classification [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Providence, USA; IEEE, 2011; 2609-2704.
- [9] Boureau Y L, Le Roux N, Bach F, et al. Ask the locals: multi-way local pooling for image recognition [C]// Proceedings of the 13th International Conference on Computer Vision. Barcelona, Spain; IEEE, 2011; 2651-2658.
- [10] Perronnin F, Sánchez J, Mensink T. Improving the fisher kernel for large-scale image classification [C]// Proceedings of the 11th European Conference on Computer Vision. Heraklion, Greece; Springer, 2010; 143-156.
- [11] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA; IEEE, 2006, II; 2169-2178.
- [12] Li F F, Perona P. A bayesian hierarchical model for learning natural scene categories [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA; IEEE, 2005; 524-531.
- [13] Qin J, Yung N H. Scene categorization with multiscale category-specific visual words [J]. Optical Engineering, 2009, 48(4): 1-13.
- [14] Agarwal A, Triggs B. Recovering 3D human pose from monocular images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(1): 44-58.
- [15] van Gemert J C, Veenman C J, Smeulders A W, et al. Visual word ambiguity [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(7): 1271-1283.
- [16] Duan F, Zhang Y J. Multi-scale sparse representation based scene categorization [J]. Application Research of Computers, 2012, 29(10): 3938-3941. [段菲, 章毓晋. 基于多尺度稀疏表示的场景分类 [J]. 计算机应用研究, 2012, 29(10): 3938-3941.]
- [17] Jaakkola M S T, Szummer M. Partially labeled classification with Markov random walks [J]. Advances in Neural Information Processing Systems, 2002, 14: 945-952.
- [18] Fan R E, Chang K W, Hsieh C J, et al. LIBLINEAR: a library for large linear classification [J]. The Journal of Machine Learning Research, 2008, 9: 1871-1874.
- [19] Li L J, Su H, Li F F, et al. Object bank: a high-level image representation for scene classification & semantic feature sparsification [C]// Proceedings of the 24th Annual Conference on Neural Information Processing Systems. Vancouver, Canada; Curran Associates, 2010; 1378-1386.
- [20] Li L J, Li F F. What, where and who? classifying events by scene and object recognition [C]// Proceedings of the 11th International Conference on Computer Vision. Rio de Janeiro, Brazil; IEEE, 2007; 1-8.
- [21] Bo L, Ren X, Fox D. Hierarchical matching pursuit for image classification: architecture and fast algorithms [C]// Proceedings of 25th Annual Conference on Neural Information Processing Systems. Granada, Spain; Curran Associates, 2011; 2115-2123.
- [22] Zheng Y, Jiang Y G, Xue X. Learning hybrid part filters for scene recognition [C]// Proceedings of the 12th European Conference on Computer Vision. Florence, Italy; Springer, 2012; 172-185.
- [23] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories [J]. Computer Vision and Image Understanding, 2007, 106(1): 59-70.
- [24] Coates A, Ng A Y. The importance of encoding versus training with sparse coding and vector quantization [C]// Proceedings of The 28th International Conference on Machine Learning. Bellevue, USA; Omnipress, 2011; 921-928.
- [25] Yu K, Lin Y, Lafferty J. Learning image representations from the pixel level via hierarchical sparse coding [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA; IEEE, 2011; 1713-1720.
- [26] McCann S, Lowe D G. Local naive bayes nearest neighbor for image classification [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Providence, USA; IEEE, 2012; 3650-3656.
- [27] Jiang Z, Lin Z, Davis L S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA; IEEE, 2011; 1697-1704.
- [28] Bo L, Ren X, Fox D. Multipath sparse coding using hierarchical matching pursuit [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Portland, USA; IEEE, 2013; 660-667.
- [29] Shaban A, Rabiee H R, Farajtabar M, et al. From local similarity to global coding: an application to image classification [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Portland, USA; IEEE, 2013; 2794-2801.
- [30] Sohn K, Jung D Y, Lee H, et al. Efficient learning of sparse, distributed, convolutional feature representations for object recognition [C]// Proceedings of the 13th International Conference on Computer Vision. Barcelona, Spain; IEEE, 2011; 2643-2650.