

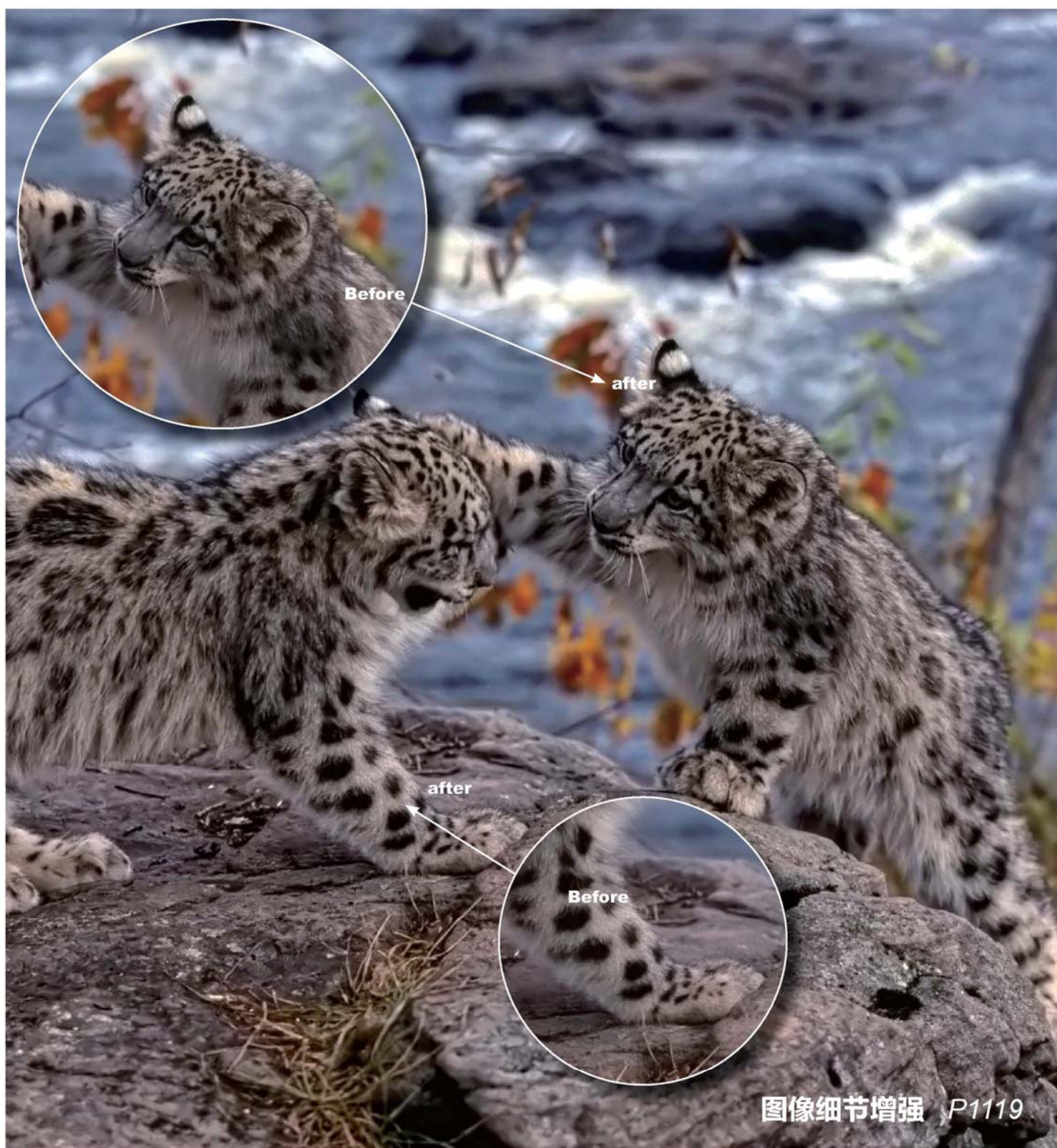


主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2016
09
VOL.21

ISSN1006-8961
CN11-3758/TB



图像细节增强 P1119



第21卷第9期（总第245期）
2016年9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
中文核心期刊

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
中国图象图形学会
北京应用物理与计算数学研究所

主 编 顾行发
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@radi.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

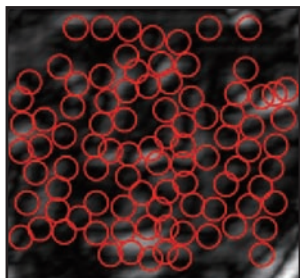
Title inscription: Song Jian Monthly, Started in 1996

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

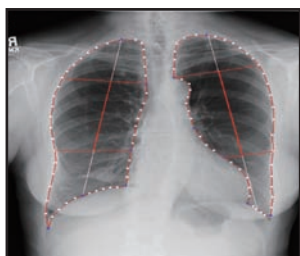
Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
P.O.Box 9718, Beijing, P.R.China
Zip code 100101
E-mail jig@radi.ac.cn
Telephone 010-64807995
Website www.cjig.cn

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 60.00元



基于频率簇模型的人脸识别
(第1166页)



胸部解剖结构回归模型的虚拟双能量X线减影方法(第1247页)



参考1维光谱差异的区域生长种子点选取方法(第1256页)

图像处理和编码

融合梯度信息的改进引导滤波

谢伟, 周玉钦, 游敏 1119

用于屏幕图像编码的索引图快速预测算法

陈规胜, 宋传鸣, 王相海, 刘丹 1127

图像分析和识别

哈希编码结合空间金字塔的图像分类

彭天强, 栗芳 1138

自上而下注意力分割的细粒度图像分类

冯语姝, 王子磊 1147

结合类内和类间距离的可能聚类分割算法

刘璐, 吴成茂 1155

基于频率簇模型的人脸识别

袁姮, 王志宏, 姜文涛 1166

图像理解和计算机视觉

不同池化模型的卷积神经网络学习性能研究

刘万军, 梁雪剑, 曲海成 1178

惰性随机游走视觉显著性检测算法

李波, 卢春园, 金连宝, 冷成财 1191

自适应邻域相关性的背景建模

万剑, 洪明坚, 赵晨丘 1202

实时鲁棒的特征点匹配算法

陈天华, 王福龙 1213

暗通道先验图像去雾的大气光校验和光晕消除

赵锦威, 沈逸云, 刘春晓, 欧阳毅 1221

计算机图形学

定点容量限制质心Power图生成

郑利平, 郜文灿, 李尚林, 曹力 1229

医学图像处理

多相期增强CT中肝内血管的自动分割

袁戎, 石姝玥, 谢庆国 1238

胸部解剖结构回归模型的虚拟双能量X线减影方法

陈胜, 张茗屋 1247

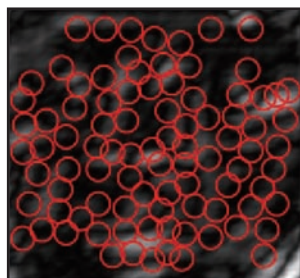
遥感图像处理

参考1维光谱差异的区域生长种子点选取方法

李修霞, 荆林海, 李慧, 唐韵玮, 戈文艳 1256

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Face recognition method based on frequency cluster (P1166)



Virtual dual-energy subtraction method for X-ray radiographs by using regression model based on chest anatomical structure(P1247)



Seed extraction method for seeded region growing based on one-dimensional spectral differences(P1256)

Image Processing and Coding

- Improved guided image filtering integrated with gradient information
Xie Wei, Zhou Yuqin, You Min 1119
- Fast prediction algorithm of index maps for screen image coding
Chen Guisheng, Song Chuanming, Wang Xianghai, Liu Dan 1127

Image Analysis and Recognition

- Image classification algorithm based on hash codes and space pyramid
Peng Tianqiang, Li Fang 1138
- Fine-grained image categorization with segmentation based on top-down attention map
Feng Yushan, Wang Zilei 1147
- Possibilistic clustering segmentation algorithm based on intra-class and inter-class distance
Liu Lu, Wu Chengmao 1155
- Face recognition method based on frequency cluster
Yuan Heng, Wang Zhihong, Jiang Wentao 1166

Image Understanding and Computer Vision

- Learning performance of convolutional neural networks with different pooling models
Liu Wanjun, Liang Xuejian, Qu Haicheng 1178
- Saliency detection based on lazy random walk
Li Bo, Lu Chunyuan, Jin Lianbao, Leng Chengcai 1191
- Background modeling based on adaptive neighborhood correlation
Wan Jian, Hong Mingjian, Zhao Chenqiu 1202
- Real-time robust feature-point matching algorithm
Chen Tianhua, Wang Fulong 1213
- Dark channel prior-based image dehazing with atmospheric light validation and halo elimination
Zhao Jinwei, Shen Yiyun, Liu Chunxiao, Ouyang Yi 1221

Computer Graphics

- Generation method for a centroidal capacity constrained Power diagram with fixed sites
Zheng Liping, Gao Wencan, Li Shanglin, Cao Li 1229

Medical Image Processing

- Automatic hepatic vessel segmentation algorithm in multi-phase contrast-enhanced CT
Yuan Rong, Shi Shuyue, Xie Qingguo 1238
- Virtual dual-energy subtraction method for X-ray radiographs by using regression model based on chest anatomical structure
Chen Sheng, Zhang Mingwu 1247

Remote Sensing Image Processing

- Seed extraction method for seeded region growing based on one-dimensional spectral differences
Li Xiuxia, Jing Linhai, Li Hui, Tang Yunwei, Ge Wenyan 1256

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2016)09-1138-09

论文引用格式: Peng T Q, Li F. Image classification algorithm based on hash codes and space pyramid[J]. Journal of Image and Graphics, 2016, 21(9): 1138-1146. [彭天强, 栗芳. 哈希编码结合空间金字塔的图像分类[J]. 中国图象图形学报, 2016, 21(9): 1138-1146.] [DOI: 10. 11834/jig. 20160903]

哈希编码结合空间金字塔的图像分类

彭天强¹, 栗芳²

1. 河南工程学院计算机学院, 郑州 451191; 2. 河南图像识别工程技术中心, 郑州 450002

摘要: 目的 稀疏编码是当前广泛使用的一种图像表示方法, 针对稀疏编码及其改进算法计算过程复杂、费时等问题, 提出一种哈希编码结合空间金字塔的图像分类算法。方法 首先, 提取图像的局部特征点, 构成局部特征点描述集。其次, 学习自编码哈希函数, 将局部特征点表示为二进制哈希编码。然后, 在二进制哈希编码的基础上进行 K 均值聚类生成二进制视觉词典。最后, 结合空间金字塔模型, 将图像表示为空间金字塔直方图向量, 并应用于图像分类。结果 在常用的 Caltech-101 和 Scene-15 数据集上进行实验验证, 并和目前与稀疏编码相关的算法进行实验对比。与稀疏编码相关的算法相比, 本文算法词典学习时间缩短了 50%, 在线编码速度提高了 1.3 ~ 12.4 倍, 分类正确率提高了 1% ~ 5%。结论 提出了一种哈希编码结合空间金字塔的图像分类算法, 利用哈希编码代替稀疏编码对局部特征点进行编码, 并结合空间金字塔模型用于图像分类。实验结果表明, 本文算法词典学习时间更短、编码速度更快, 适用于在线词典学习和应用。

关键词: 哈希编码; 空间金字塔匹配模型; 稀疏编码; 二进制 K 均值聚类; 图像分类

Image classification algorithm based on hash codes and space pyramid

Peng Tianqiang¹, Li Fang²

1. Department of Computer Science and Engineering, Henan Institute of Engineering, Zhengzhou 451191, China;

2. Henan Image Recognition Engineering Center, Zhengzhou 450002, China

Abstract: **Objective** Sparse coding is widely used to represent images. However, this method and its improved algorithms require complex computation and long running times, among other drawbacks. An image classification algorithm, based on hash codes and space pyramids, is proposed to solve these issues. **Method** The algorithm consists of four steps. First, extract local feature points from the images. Second, learn binary auto-encoder hashing functions, which map the local feature points into hash codes. Third, perform binary k-means cluster on the binary hash codes and generate the binary visual vocabularies. Finally, combine with a spatial pyramid matching model, and represent the image by the histogram vector of the space pyramid, which is used for image classification. **Result** In order to verify the efficiency of the proposed algorithm, we used two common datasets, Caltech-101 and Scene-15. The results were compared with state-of-the-art sparse coding algorithms, which showed the time of learning vocabularies of our method was 50% left, the online encoder speed was increased 1.3 ~ 12.4 times, and the classification accuracy increased 1% ~ 5%. We also compared the classification performance of different hash encode methods, such as PCA-ITQ and KSH. **Conclusion** In this paper, a novel image classification algo-

收稿日期: 2016-02-19; 修回日期: 2016-05-09

基金项目: 国家自然科学基金项目(61301232)

第一作者简介: 彭天强(1978—), 男, 副教授, 2008 年于解放军信息工程大学获信息科学专业博士学位, 主要研究方向为多媒体信息检索、模式识别。E-mail: ptq_drumboy@163.com

Supported by: National Natural Science Foundation of China(61301232)

rithm was proposed. The proposed algorithm, encoded local feature points by hash codes rather than sparse coding, and image classification was achieved with a spatial pyramid matching model. Experimental results showed that the proposed algorithm had faster learning vocabulary and encoder speed, and could be used for online vocabulary learning and other online applications.

Key words: hash codes; spatial pyramid matching model; sparse coding; binary K-means cluster; image classification

0 引言

图像分类是计算机视觉领域中的一个重要问题,在目标识别、图像检索和视觉监控等领域中均有着广泛应用。视觉词袋(BoVW)模型^[1]被广泛应用于图像分类中,但 BoVW 模型未考虑局部特征之间的空间关系,这严重限制了该方法的图像表示能力。文献[2]提出了一种基于空间金字塔匹配(SPM)模型的 BoVW,首先将图像在空间上进行不同尺度的划分,然后在每个尺度上将各个图像子块表示为直方图向量,最后将每个尺度的向量进行组合,得到图像的空间金字塔向量表示。

早期 BoVW 模型主要采用硬分配策略,选择词典中与局部特征最近的视觉单词为量化结果,但是这种量化方式造成信息损失较大。文献[3]提出了一种软分配策略,让局部特征与多个视觉词汇相关联,降低量化误差,提高编码能力。为了提高编码能力,Yang 等人^[4]采用稀疏编码(SC)方法对图像局部特征进行量化,将局部特征表示为过完备词典的稀疏表示。通过对稀疏编码进一步深入研究,Yu 等人^[5]在图像稀疏表示加入局部性,并提出了局部约束线性编码(LLC)以保证相近的局部特征具有相近的编码,提高了局部特征的表示精度。文献[6]在稀疏编码的基础上,加入稀疏系数非负性约束并加入保持局部相似性信息的优化目标,提出了拉普拉斯非负稀疏编码(LNSC),在图像分类中取得了不错的效果。文献[7]在 LLC 基础上进行改进,提出了非负稀疏局部线性编码(NSLLC)。文献[8]在 NSLLC 基础上,引入 Fisher 判别约束准则,提出了基于 Fisher 判别约束的非负稀疏局部线性编码模型,提高了图像的分类性能。以上这些与稀疏编码相关的算法均以局部特征点重构误差最小化为目标,且后续改进算法加入了保持局部相似性信息,即相似的局部特征点应具有相近的稀疏编码。这些与稀疏编码相关的算法,虽然提高了对局部特征的表示精

度,但需要反复迭代求得完备词典,计算过程复杂、费时。

针对与稀疏编码相关算法计算过程复杂的问题,提出了利用哈希编码代替稀疏编码对局部特征点进行编码的思想。局部敏感哈希^[9]可以将相似的对象映射为相似的编码,特别地,自编码哈希^[10]的哈希函数的构造目标是哈希编码与原样本点的重构误差最小,与 SC^[4]的优化目标一致。由于哈希编码不具有稀疏性,利用空间金字塔视觉词袋模型表示图像,提出了一种哈希编码结合空间金字塔的图像分类算法。首先提取图像的局部特征点,构成局部特征点描述集;其次,学习自编码哈希函数,将局部特征点表示为二进制哈希编码;然后,在二进制哈希码的基础上进行 K 均值聚类生成二进制视觉词典;最后,结合空间金字塔模型,将图像表示为空间金字塔直方图向量,并应用于图像分类。本文方法将局部特征点进行哈希编码,提高了编码速度和精度,并考虑局部特征之间的空间信息,将图像表示为空间金字塔直方图向量,在图像分类中达到了不错的分类效果。

1 相关工作

1.1 稀疏编码

稀疏编码^[4]将图像的局部特征表示为过完备词典的几个基向量的线性组合。给定局部 SIFT (scale-invariant feature transform) 特征点构成的集合 $X = [x_1, \dots, x_M] \in \mathbf{R}^{D \times M}$, 需要求解过完备词典 V 和局部特征点的稀疏表示系数 $u_m, m = 1, 2, \dots, M$, 其目标为用过完备词典的基向量重构局部特征集, 优化约束为

$$\min_{U, V} \sum_{m=1}^M \|x_m - u_m V\|^2 + \lambda \|u_m\| \quad (1)$$

$$\text{s. t. } \|v_k\| \leq 1, k = 1, 2, \dots, K$$

式中,过完备词典 $V = [v_1, \dots, v_K]^T$ 中基向量的个数 $K \gg D$ 。利用这种方式得到的编码,具有稀疏性但不

能保证相似的局部特征点具有相似的稀疏编码。

局部线性编码^[5],在稀疏编码的目标函数上增加了编码的局部性,优化约束为

$$\min_{U,V} \sum_{m=1}^M \|x_m - u_m V\|^2 + \lambda |d_m \odot u_m|$$

$$\text{s. t. } \mathbf{1}^T u_m = 1, m = 1, 2, \dots, K \quad (2)$$

式中, $d_m \in \mathbf{R}^K$ 是特征点 x_m 与词典 V 的每一个基向量之间的距离,用公式表示为

$$d_m = \exp\left(\frac{\text{dist}(x_m, V)}{\sigma}\right)$$

$$\text{dist}(x_m, V) = [\text{dist}(x_m, v_1), \dots, \text{dist}(x_m, v_K)]^T \quad (3)$$

式中, σ 表示局部性适应权重, $d_m \odot u_m$ 是由向量 d_m 和 u_m 对应分量相乘得到,体现了编码的局部性。利用这种方式得到的编码,具有稀疏性且能够保证相似的局部特征点具有相似的稀疏编码。

文献[6-8]在稀疏编码和局部线性编码上的改进,其优化目标与它们都基本相同,只是加上了编码表示系数 $u_{mk} \geq 0$ 的约束。

1.2 哈希编码

原始数据空间中的两个相邻数据点通过相同的映射后,这两个数据点在新的数据空间中仍然相邻的概率很大,而不相邻的数据点被映射为相同值的概率很小,满足这种性质的映射称为哈希函数。在不同的度量下,哈希函数的构造方法是不同的,特别地,在 Cosine 度量下,对应的哈希函数为 $h(x) = \text{sgn}(W \cdot x)$, 其中 W 是服从高斯分布的随机向量,这种哈希函数也称为二进制哈希函数。二进制哈希函数将高维数据映射到汉明空间中,为每一个对象产生二进制哈希编码,且保证在原空间中相近的对象,具有相近的二进制哈希编码。

二进制哈希方法可以分为无监督的哈希算法、半监督的哈希算法和监督的哈希算法。无监督的哈希方法不考虑数据的标签信息,包括 Isotropic hashing^[11]、谱哈希(SH)^[12]、PCA-ITQ^[13] 和自编码哈希(BA)^[10]等;半监督的哈希方法考虑部分的相似性信息,包括半监督哈希(SSH)^[14];监督的哈希方法利用数据集的标签信息或者相似性点对信息作为监督信息,包括 BRE Binary Reconstructive Embedding^[15]、监督的核哈希(KSH)^[16]等。上述的这些二进制哈希算法的目标是构造出能够较好地保持特征

在原空间中的相似性信息且能够生成紧致哈希码的哈希函数。

BA^[10]构造的哈希函数,不仅能够保持原始空间数据间的相似性,且能够保证哈希编码与原始数据之间的重构误差最小。给定哈希函数 $h: x \rightarrow z$ 将向量 $x \in \mathbf{R}^D$ 映射到低维汉明空间 $z \in \{0, 1\}^L$ 中的 L 维二进制哈希码 $L < D$ 。解码函数 $f: z \rightarrow x$ 将二进制哈希码 z 重新映射到 $\mathbf{R}^D$ 空间以重构向量 x 。BA 的优化目标函数为

$$E(h, f) = \sum_{n=1}^N \|x_n - f(h(x_n))\|^2$$

$$\text{s. t. } h(x_n) \in \{0, 1\}^L \quad (4)$$

引入中间变量,上述问题的最优解等价于

$$\min_{h, f, z} \sum_{n=1}^N \|x_n - f(z_n)\|^2 + \mu \|z_n - h(x_n)\|^2$$

$$\text{s. t. } z_n \in \{0, 1\}^L, n = 1, \dots, N \quad (5)$$

从优化目标中可以看出,BA 学习到哈希函数,可以保证数据的二进制哈希编码与数据的重构误差最小,由于哈希函数的特性,它同时也能保证相近的数据具有相近的二进制哈希码。

二进制哈希编码也具有空间上的优势,例如给定原始数据大小 $N = 1\,000\,000$,每个数据的维数 $D = 300$,映射为二进制编码的长度为 $L = 32$,则在数据在原始空间中需要占用 2.4 GB 的存储空间,而其相应的二进制编码只需要占用 4 MB 的存储空间。因此,将局部特征点进行哈希编码,可以大大降低存储空间,适用于大规模图像集的后续处理(例如聚类)。

由于不同的哈希算法可以构造出不同的哈希编码,在无监督的哈希算法中 PCA-ITQ 的性能较优,在监督的哈希算法中, KSH 的性能较优,在实验部分,将 BA 的哈希编码分别与 PCA-ITQ、KSH 的哈希编码在图像分类中的性能作对比。

1.3 空间金字塔视觉词袋模型

空间金字塔视觉词袋模型^[17]将金字塔匹配原理^[18]与视觉词袋模型相结合,首先将图像在不同尺度 $l = 0, 1, \dots, L$ 上进行划分为 $2^l \times 2^l$ 个子区域,对于每一个子区域,分别计算所有视觉单词在该区域出现的次数,构成该区域的视觉词频直方图向量。然后将一个尺度下所有子区域的词频直方图向量按顺序组成一个向量,即

$$\mathbf{H}_X^l = [\mathbf{H}_X^{l,1}, \dots, \mathbf{H}_X^{l,k}, \dots, \mathbf{H}_X^{l,4^l}] \quad (6)$$

式中, $\mathbf{H}_X^l$ 表示图像 X 在第 l 尺度下的词频向量, $\mathbf{H}_X^{l,k}$ 表示图像在第 l 尺度的第 k 个子区域的词频向量。

对于两幅图像 X 和 Y , 它们在第 l 尺度下的相似性可用直方图交叉函数来计算, 即

$$I(\mathbf{H}_X^l, \mathbf{H}_Y^l) = \sum_{k=1}^{4^l} \sum_{m=1}^M \min(\mathbf{H}_X^{l,k}(m), \mathbf{H}_Y^{l,k}(m)) \quad (7)$$

式中, $\mathbf{H}_X^{l,k}(m)$ 表示图像 X 的第 l 尺度中第 k 个子区域的直方图向量的第 m 维的值。

根据金字塔匹配原理可知, 区域划分越稠密, 匹配精度越高, 不同尺度需要设置不同的权重。对所有尺度的子区域进行计算, 得到空间金字塔匹配核, 即

$$K_{\text{SPM}} = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (I^l - I^{l+1}) \quad (8)$$

式中, I^l 是式(7)中定义的第 l 尺度下两幅图像的相似性。在本文中利用特征点的空间信息, 用空间金字塔直方图向量表示图像。

2 基于哈希编码和空间金字塔的图像分类算法

为了应用局部特征的空间信息、提高局部特征点的编码速度和精度, 提出了一种哈希编码结合空间金字塔的图像分类算法。

2.1 基于二进制哈希编码的 K 均值聚类算法

为了适用于大规模图像的聚类, Gong 等人^[19]提出了基于二进制哈希编码的 K 均值聚类算法。基于二进制哈希码的 K-means 算法的目标函数为

$$\min_{c_j} \sum_{i=1}^n \sum_{j=1}^k \|x_i - c_j\|^2 \quad \text{s. t. } c_j \in \{-1, 1\} \quad (9)$$

该目标函数类似于 K-means 算法的目标函数, 只是聚类中心和样本均为离散的二进制值。该优化可以利用 EM 算法来求解, 首先需要将每个二进制哈希码分配到与之最近的聚类中心, 对于二进制哈希码的最近邻问题, 可以通过多索引哈希技术 (MIH)^[20] 快速地进行精确的 r 邻域搜索。其次, 更新二进制聚类中心, 可以通过求解问题式(9)的最优解得到其计算公式。

仅考虑某个聚类中心 c_j , 假设有 p 个样本点 $x_1, x_2, \dots, x_p$ 与该聚类中心最近, 问题式(9)优化问题转化为

$$\min_{c_j} \sum_{i=1}^p \|x_i - c_j\|_2^2 = \sum_{i=1}^p \|x_i\|_2^2 + \sum_{i=1}^p \|c_j\|_2^2 - \sum_{i=1}^p x_i c_j^T \quad (10)$$

由于式(10)中第1项和第2项均为常数值, 问题式(10)的最小化问题可以转化为等价的极大值问题, 即

$$\max_{c_j} \sum_{i=1}^p x_i c_j^T = \max_{c_j} (\sum_{i=1}^p x_i) c_j^T \quad \text{s. t. } c_j \in \{-1, 1\} \quad (11)$$

问题式(11)有解析形式的最优解, 且其最优解为

$$c_{jk} = \text{sgn}(m_{jk}) = \begin{cases} 1 & m_{jk} \geq 0 \\ -1 & m_{jk} < 0 \end{cases} \quad (12)$$

式中, $m_j = \sum_{i=1}^p x_i$ 表示属于聚类中心 c_j 的所有样本 x_i 的连和。具体过程为:

算法1: 基于二进制哈希码的聚类算法 (Bk-means)

给定二进制哈希码样本集 $X = \{x_1, \dots, x_n\}$ 和聚类个数 k 以及最大迭代次数 maxIter 。

从样本集 X 中随机选择 k 个值, 构成聚类中心的初始值 $C = \{c_1, \dots, c_k\}$ 。

for $i = 1, \dots, \text{maxIter}$

for $j = 1, \dots, n$ // 为每个样本选择与它最近的聚类中心

$\text{idx} = M.\text{Lookup1NN}(x_j)$ // 按照 MIH 技术为聚类中心集

C 建立多索引哈希表 M 。

end for

for $j = 1, \dots, k$ // 更新聚类中心 C

计算每个聚类中心的均值 m_j

更新聚类中心 $c_j = \text{sgn}(m_j)$

end for

if $i > 1$

计算本次迭代产生的聚类中心 C^i 与上次迭代产生的聚类中心 C^{i-1} 之间的差别, 若集合 $C^i = C^{i-1}$, 则退出整个 for 循环, 聚类算法结束。

end if

end for

2.2 基于哈希编码和空间金字塔的图像表示

BA 生成的哈希编码与稀疏编码具有类似的特性, 因此利用哈希编码表示图像的局部描述子是合理的。基于哈希编码和空间金字塔的图像表示分为

视觉词典构造和向量表示两步。采用哈希编码表示局部特征描述子,并在哈希编码上进行二进制 K 均值聚类生成视觉词典。在构建视觉词典后,结合空间金字塔模型,将图像表示为空间金字塔直方图向量。

基于哈希编码的视觉词典构造过程的主要步骤:

1) 提取训练图像集的 SIFT 特征,得到了包含 M 个 SIFT 特征点的图像局部特征库 $R = \{r_1, r_2, \dots, r_i, \dots, r_M\}$, 其中每个点 r_i 是一个 128 维的 SIFT 特征向量;

2) 指定二进制哈希编码的位数 K , 利用图像特征库 R 根据 1.2 节中 BA 算法, 学习得到 K 个哈希函数 $H = \{h_1, h_2, \dots, h_K\}$;

3) 利用 K 个哈希函数将特征库 R 映射为二进制哈希编码集 $B = \{b_1, b_2, \dots, b_i, \dots, b_M\}$, 每个哈希编码包含 K 位。由于自编码哈希的特性, 将原始空间中距离相近的特征点映射为相近的二进制哈希码, 且每个二进制哈希编码与特征点之间的重构误差最小;

4) 给定聚类数目 C , 根据 2.1 节中的算法对二进制哈希编码集 B 进行 BK-means 聚类, 得到的聚类中心构成了视觉词典, 记为 $D = \{v_1, v_2, \dots, v_C\}$ 。

从上述过程中可以看出, 基于哈希编码的视觉词典的构造过程中需要学习哈希函数, 将局部特征点表示为二进制哈希码, 由于二进制码占内存较少, 该方法适用于存储大规模的哈希编码, 并在此基础上进行聚类。在二进制哈希编码上进行 BK-means 聚类, 在每次迭代中只需要查表就可以得到与之最近的聚类中心, 其时间复杂度为 $O(M)$, 其中 M 为训练特征库中特征点的个数, 而传统 K-means 聚类算法在此步骤中的时间复杂度为 $O(KM)$, 其中 K 为聚类中心个数。基于哈希编码生成的视觉词典, 大大减少了视觉词典的生成时间, 适用于大规模图像集的视觉词典的生成。

基于哈希编码生成视觉词典后, 结合空间金字塔模型, 将图像进行向量表示, 主要步骤为:

1) 提取图像的局部特征;

2) 利用已学习到的 BA 哈希函数 H , 将局部特征映射为二进制哈希编码;

3) 将每个二进制哈希编码按照查表的方式映射为视觉词典 D 中的视觉单词;

4) 按照空间金字塔匹配模型的思想, 设置图像的尺度, 计算每个尺度下各个子区域的视觉词直方图向量, 最后将各尺度下所有子区域的直方图向量进行组合, 得到图像的空间金字塔直方图向量。

这样, 就得到了图像的空间金字塔直方图向量表示。

2.3 基于哈希编码和空间金字塔的图像分类算法

将基于哈希编码和空间金字塔的图像表示, 应用于图像分类, 其流程图如图 1 所示。

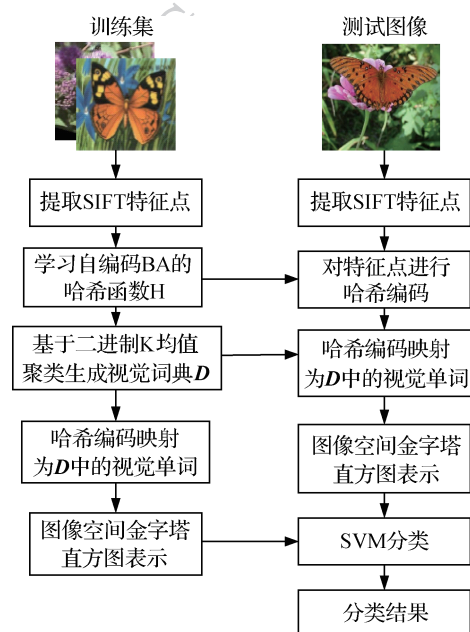


图 1 基于本文算法的图像分类流程图

Fig. 1 The flow chart of image classification based on our method

在训练阶段, 利用训练集学习哈希编码、构造视觉词典并训练分类器, 首先提取训练图像集中各图像的 SIFT 特征作为训练向量集, 然后利用 2.2 节中的方法学习哈希函数 H , 构造视觉词典 D , 并取空间金字塔尺度 $l=2$, 将图像表示为一个 $21 \times C$ 维的空间金字塔直方图向量; 最后利用图像的向量表示训练 SVM 分类器。

在测试阶段, 给定测试图像, 首先提取测试图像的 SIFT 特征; 然后利用训练阶段学习到的哈希函数将局部特征点进行哈希编码, 并映射为视觉词典 D 中的视觉单词, 将图像进行分层分区域表示为 $21 \times C$ 维的空间金字塔直方图向量; 最后将测试图像的向量表示输入训练好的 SVM 分类器进行分类。

3 实验结果与分析

3.1 实验设置

为了评估本文算法的正确性和有效性,进行了实验验证。实验硬件配置为内存为 8GB、i3 3.4 GHz CPU 的 PC,软件编程环境为 Win7 旗舰版,Matlab 2012。

选择在 2 个常用图像分类数据集上进行实验,分别是 Caltech-101^[21] 和 Scene-15^[22]。这个两个数据集是图像分类性能评估的常用图像库。Caltech-101 图像库,包括了 101 类,每类中包括 31~800 幅图像,共 9 144 幅图像。在该数据集中,从每一类中挑选出 30 个样本作为训练图像,其余的图像构成测试集。在训练集上提取 SIFT 特征点,并随机挑选出二十万个特征点作为图像特征库,用于学习哈希函数以及构建视觉词典。Scene-15 图像库,包括了 15 类,每类中包括 200~400 幅图像,共 4 485 幅图像。在该数据集中,从每一类中挑选出 30 个样本作为训练图像,其余的图像构成测试集。在训练集上提取 SIFT 特征点,并随机挑选出 20 万个特征点作为图像特征库,用于学习哈希函数以及构建视觉词典。

为了验证本文算法在图像分类中的性能,首先,将本文方法的分类精度、词典构造时间和在线编码时间,分别与基于稀疏编码的 ScSPM 算法^[4]和 LLC 算法^[5]、以及在这两个算法下的改进算法 NSLLC 算法^[7]、LNNSC 算法^[6]、Fisher + NSLLC 算法^[8]作比较。实验采用线性 SVM 分类器,采用“一对多”策略,对每个类别分别训练各自的 SVM 分类器。在这几种算法中,ScSPM 算法、LLC 算法和 Fisher + NSLLC 算法利用作者提供的源代码,并对 NSLLC 算法、LNNSC 算法方法按照其算法流程进行了实现,以便与本文算法进行性能比较。

其次,为了验证不同哈希编码算法对分类性能的影响,将本文算法中的 BA^[10] 哈希编码分别替换为无监督的 PCA-ITQ^[13] 的哈希编码和监督的 KSH^[16] 的哈希编码,并比较了这 3 种哈希编码在图像分类中的性能。

3.2 与稀疏编码相关算法的分类性能对比

将本文算法中哈希编码的长度设置为 $K = 32$,分别与稀疏编码相关的方法的性能作对比。

表 1 给出了本文方法和与稀疏编码相关的方法

在两个数据集上分类精度对比值。从表 1 中可以看出,在两个数据集上,ScSPM 算法分类正确率最低,因为它只考虑了编码的稀疏性;LLC 算法的分类正确率略高于 ScSPM 算法,因为它不仅考虑了编码的稀疏性,而且还考虑了编码的局部性;NSLLC 算法的分类正确率高于 LLC 算法,因为它在 LLC 算法基础上,加入表示系数的非负性约束,得到了更好地编码表示;Fisher + NSLLC 算法在 NSLLC 的基础上,加入了 Fisher 判别约束准则,因此进一步提高了图像的分类正确率。LNNSC 算法的分类正确率比较高,主要是它在稀疏编码的基础上,加入了保持局部相似性信息的优化目标,得到了更好的编码。本文算法的分类精度略高于 LNNSC 算法,因为本文利用哈希编码表示局部描述子,较好地保持了局部相似性信息,且在哈希编码的基础上又做了一次聚类,得到了区分能力更强的视觉词,并利用空间金字塔直方图表示图像,应用了局部特征点间的空间信息。

表 1 在两个数据集上不同算法的分类精度

Table 1 The classification accuracy of different methods compare on the two datasets

算法	Caltech101/%	Scene-15/%
ScSPM	73.2	70.8
LLC	73.4	71.2
LNNSC	75.4	75.1
NSLLC	73.8	72.7
Fisher + NSLLC	74.5	75.2
本文	75.8	76.3

表 2 给出了本文方法和稀疏编码相关方法在数据集 Caltech-101 的词典构造时间和在线编码时间的对比值。本文方法的词典构造时间包括 BA 哈希函数学习时间和基于二进制哈希编码的聚类时间;而其他方法的词典构造时间包括过完备词典的学习时间。本文方法的在线编码时间是将局部特征表示为哈希码的时间,其他方法的在线编码时间是将局部特征表示为稀疏码的时间。从表 2 中可以看出,本文方法的词典构造时间最短,而其他与稀疏编码相关的方法词典构造时间较长,且它们的构造时间相差不大,因为与稀疏编码相关的方法均需要求解方程以得到过完备词典,只是它们的目标函数或约束条件略有不同。本文方法的在线编码时间最短,因为局部特征点的哈希编码只需要与相应的哈希函数作内积,而局部特征点的稀疏编码需要求解方程。

虽然本文方法的词典构造时间与其他方法相比较短,但是仍旧用时较长,仅适合于离线训练,在 3.3 节考虑用其他哈希方法进行编码。

表 2 在 Caltech-101 上不同算法的时间对比
Table 2 The time of different methods compare on the Caltech-101 dataset

算法	词典构造时间/s	在线编码时间/s
ScSPM	23 176	1.120 8
LLC	27 678	0.137 9
LNNSC	27 174	0.152 1
NSLLC	27 810	0.128 1
Fisher + NSLLC	27 857	0.145 6
本文	13 265	0.096 4

3.3 不同哈希编码方法下图像分类性能对比

BA 构造的哈希函数虽然满足局部相似性和重构误差最小,但是它的哈希函数学习时间太长,严重影响了其效率,不适用于在线训练。考虑用其他哈希方法来替换本文的 BA 哈希编码,并分析它们的性能。在无监督的哈希算法中 PCA-ITQ 的性能较优,在监督的哈希算法中, KSH 的性能较优,因此比较在 PCA-ITQ 哈希编码、BA 哈希编码和 KSH 哈希编码下,它们的分类正确率和词典生成时间、局部特征在线编码时间对比。

图 2 给出了在 Scene-15 数据集下,取 BK-means 聚类的聚类个数 $C = 1\ 024$,不同编码位数对图像识别正确率的影响。当编码位数较小时 ($K = 8$),3 种编码方法构建的视觉词典的图像分类正确率都比较

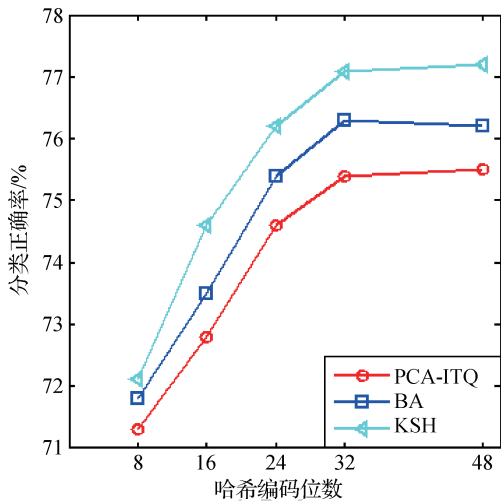


图 2 不同哈希编码位数下图像分类正确率对比
Fig. 2 The classification accuracy compared based on different hash code bits

低,随着编码位数的增加,其相应的图像分类正确率也随之增加。这主要因为二进制哈希的特性,随着编码长度的增加,其编码能够更好地保持局部相似性。但从图 2 中也可以看出,当编码位数从 32 位增加到 48 位时,图像分类正确率增加不大,这说明了 32 位的编码就能够很好地描述原始特征点,若使用更加长的编码,仅仅是增加了计算复杂度。

表 3 和表 4 分别给出了 3 种哈希编码方法在两个数据集上在哈希编码长度为 $K = 32$, BK-means 的聚类个数 $C = 1\ 024$ 下,3 种哈希方法的视觉词典生成时间和分类正确率对比。从表 3 和表 4 中可以看出,词典构造时间是 PCA-ITQ 方法用时最少,BA 方法用时最长, KSH 也用时较长,因为词典构造时间包括哈希函数学习时间和 BK-means 聚类时间,而聚类时间它们相差不大,而在哈希函数学习的过程中,PCA-ITQ 方法是一种无监督的学习方法,速度最快;BA 虽然是无监督的学习算法但是需要反复迭代求解哈希函数,用时最长;KSH 是监督的方法,需要利用标签信息且需要求解核矩阵,因此用时也比较长。但是三者的在线编码时间相差不大,因为构造出哈希函数之后,只需要相应的内积操作就可以得到编码。KSH 的分类精度最高,BA 次之,PCA-ITQ 方法最低,因为 KSH 使用了图像的标签信息,而 BA 和 PCA-ITQ 是无监督的哈希方法。

表 3 在数据集 Caltech101 上 3 种哈希方法对比
Table 3 The three hash methods compare on Caltech101

编码方法	词典构造时间/s	在线编码时间/s	分类精度/%
PCA-ITQ	25	0.089 93	74.6
BA	13 265	0.097 4	75.8
KSH	1 069	0.092 5	76.3

表 4 在数据集 Scene-15 上 3 种哈希方法对比
Table 4 The three hash methods compare on Scene-15

编码方法	词典构造时间/s	在线编码时间/s	分类精度/%
PCA-ITQ	48	0.092 5	75.4
BA	15 203	0.108 6	76.3
KSH	1 396	0.101 3	77.1

4 结 论

提出了一种哈希编码结合空间金字塔的图像分类算法,用哈希编码表示局部描述子,较好地保持了

局部相似性信息,且在哈希编码的基础上又做了一次聚类,得到了区分能力更强的视觉词典,最后结合空间金字塔模型,将图像表示为空间金字塔直方图。在常用的 Caltech-101 和 Scene-15 图像分类数据集上的实验结果表明,本文算法词典学习时间更短、编码速度更快且具有较高的分类正确率,均优于现有与稀疏编码相关的方法。

在本文中考虑了不同哈希编码方法对图像分类性能的影响。在实际应用中,可以根据需要,选择无监督的哈希方法或者监督的哈希方法对局部特征点进行哈希编码,PCA-ITQ 哈希编码词典生成时间最少,适用于在线视觉词典学习。下一步考虑利用深度网络学习的哈希函数对局部特征进行编码,看是否能够进一步提高分类精度。同时,也可以考虑将本文算法应用于相似图像的检索中,分析其检索性能。

参考文献 (References)

- [1] Sivic J, Zisserman A. Video Google: a text retrieval approach to object matching in videos[C]//Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003, 2: 1470-1477. [DOI: 10.1109/ICCV.2003.1238663]
- [2] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C]//Proceedings of 2006 IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006: 2169-2178. [DOI: 10.1109/CVPR.2006.68]
- [3] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching[C]//Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA: IEEE, 2007: 1-8. [DOI: 10.1109/CVPR.2007.383172]
- [4] Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009: 1794-1801. [DOI: 10.1109/CVPR.2009.5206757]
- [5] Wang J J, Yang J C, Yu K, et al. Locality-constrained linear coding for image classification[C]//Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010: 3360-3367. [DOI: 10.1109/CVPR.2010.5540018]
- [6] Li Q Q, Cao G. Image classification based on Laplacian non-negative sparse coding[J]. Computer Engineering, 2013, 39(11): 240-244. [李钱钱, 曹国. 基于拉普拉斯非负稀疏编码的图像分类[J]. 计算机工程, 2013, 39(11): 240-244.] [DOI: 10.3969/j.issn.1000-3428.2013.11.054]
- [7] Zhuang L S, Gao H Y, Liu C, et al. Nonnegative sparse locally linear coding[J]. Journal of Software, 2011, 22(S2): 89-95. [庄连生, 高浩渊, 刘超, 等. 非负稀疏局部线性编码[J]. 软件学报, 2011, 22(S2): 89-95.]
- [8] Zhang R J, Wei F S. Image scene classification based on fisher discriminative analysis and sparse coding[J]. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(5): 808-814. [张瑞杰, 魏福山. 结合 Fisher 判别分析和稀疏编码的图像场景分类[J]. 计算机辅助设计与图形学学报, 2015, 27(5): 808-814.] [DOI: 10.3969/j.issn.1003-9775.2015.05.005]
- [9] Datar M, Immorlica N, Indyk P, et al. Locality-sensitive hashing scheme based on p-stable distributions[C]//20th Annual Symposium on Computational Geometry. New York, USA: ACM, 2004: 253-262. [DOI: 10.1145/997817.997857]
- [10] Carreira-Perpiñán M A, Raziperchikolaei R. Hashing with binary autoencoders[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 557-566. [DOI: 10.1109/CVPR.2015.7298654]
- [11] Kong W H, Li W J. Isotropic hashing[C]//Proceedings of Advances in Neural Information Processing Systems 25. Nevada, USA: NIPS, 2012: 1646-1654.
- [12] Weiss Y, Torralba A, Fergus R. Spectral hashing[C]//Proceedings of Advances in Neural Information Processing Systems 21. Vancouver BC, Canada: NIPS, 2008: 1753-1760.
- [13] Gong Y C, Lazebnik S, Gordo A, et al. Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2916-2929. [DOI: 10.1109/TPAMI.2012.193]
- [14] Wang J, Kumar S, Chang S F. Semi-supervised hashing for large-scale search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(12): 2393-2406. [DOI: 10.1109/TPAMI.2012.48]
- [15] Kulis B, Darrell T. Learning to hash with binary reconstructive embeddings[C]//Proceedings of Advances in Neural Information Processing Systems 21. Vancouver BC, Canada: NIPS, 2008: 1042-1050.
- [16] Lin W, Wang J, Ji R R, et al. Supervised hashing with kernels[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2012: 2074-2081. [DOI: 10.1109/CVPR.2012.6247912]
- [17] Sheng H D, Duan H C, Kong C. Spatial pyramid Bag-of-words model for image classification based on semantic phrases[J]. Journal of Chinese Computer Systems, 2015, 36(4): 877-881. [生海迪, 段会川, 孔超. 基于语义短语的空间金字塔词袋

- 模型图像分类方法[J]. 小型微型计算机系统, 2015, 36(4): 877-881.]
- [18] Zhang L B, Wang C H, Xiao B H, et al. Image representation using Bag-of-phrases[J]. Acta Automatica Sinica, 2012, 38(1): 46-54. [张琳波, 王春恒, 肖柏华, 等. 基于 Bag-of-phrases 的图像表示方法[J]. 自动化学报, 2012, 38(1): 46-54.] [DOI: 10.3724/SP.J.1004.2012.00046]
- [19] Gong Y C, Pawlowski M, Yang F, et al. Web scale photo hash clustering on a single machine[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 19-27. [DOI: 10.1109/CVPR.2015.7298596]
- [20] Norouzi M, Punjani A, Fleet D J. Fast search in hamming space with multi-index hashing[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2012: 3108-3115. [DOI: 10.1109/CVPR.2012.6248043]
- [21] Griffin G, Holub A, Perona P. Caltech-256 object category dataset[R]. Technical Report 7694, Pasadena, CA, USA: California Institute of Technology, 2007.
- [22] Guo W J, Zhang Y Q, Liu Y J. Image category learning based on feature words and shape model[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(10): 1467-1475. [郭文静, 张艳秋, 刘永进. 基于特征词及形状模型的图像类别学习[J]. 计算机辅助设计与图形学学报, 2013, 25(10): 1467-1475.]