

# 基于低秩稀疏分解与协作表示的图像分类算法

张旭<sup>1,2</sup> 蒋建国<sup>1</sup> 洪日昌<sup>1</sup> 杜跃<sup>3</sup>

(合肥工业大学计算机与信息学院 合肥 230009)<sup>1</sup> (陆军军官学院计算机教研室 合肥 230031)<sup>2</sup>  
(陆军军官学院控制工程教研室 合肥 230031)<sup>3</sup>

**摘要** 目前,大部分图像分类算法为了获取较高的性能均需要充分的训练学习过程,然而在实际应用中,往往存在训练样本不足及过拟合等问题。为了避免上述问题出现,在朴素贝叶斯最近邻分类算法的原理框架下,基于非负稀疏编码、低秩稀疏分解以及协作表示提出一种非参数学习的图像分类算法。首先,基于非负稀疏编码和最大值汇聚操作表示图像信息,并构建具有低秩性质的同类训练图像集的局部特征矩阵;其次,采用低秩稀疏分解结合别类标签信息构建两类视觉词典以充分利用同类图像的相关性和差异性;最后基于协作表示表征测试图像并进行分类决策,实验结果验证了所提算法的有效性。

**关键词** 图像分类,视觉词袋,稀疏编码,低秩稀疏分解,协作表示

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.7.014

## Image Classification Algorithm Based on Low Rank and Sparse Decomposition and Collaborative Representation

ZHANG Xu<sup>1,2</sup> JIANG Jian-guo<sup>1</sup> HONG Ri-chang<sup>1</sup> DU Yue<sup>3</sup>

(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)<sup>1</sup>

(Department of Computer, Army Officer Academy of PLA, Hefei 230031, China)<sup>2</sup>

(Department of Control Engineering, Army Officer Academy of PLA, Hefei 230031, China)<sup>3</sup>

**Abstract** Currently, in order to achieve high performance, most image classification methods require adequate training and learning process. However, problems such as scarcity of training samples and overfitting of parameters are often encountered. To avoid these problems, we presented a non-parameter learning algorithm under the framework of Naive-Bayes Nearest-Neighbor (NBNN), where non-negative sparse coding, low rank and sparse decomposition and collaboration representation are jointly employed. Firstly, non-negative sparse coding combined with max pooling is introduced to represent images, and local feature matrices of similar training image sets with low-rank characteristic are generated. Secondly, two kinds of visual dictionary with category labels are constructed by leveraging low rank and sparse decomposition to make full use of the correlation and diversity of images with the same category label. Lastly, test images are represented based on collaboration representation for classification. Experimental results demonstrate effectiveness of the proposed algorithm.

**Keywords** Image classification, Bag of visual words, Sparse coding, Low rank and sparse decomposition, Collaborative representation

## 1 引言

作为计算机视觉和模式分类的基础性研究问题之一,图像分类在图像检索<sup>[1]</sup>、图像标注<sup>[2]</sup>等领域具有关键性的作用,因此吸引着众多研究者的关注。近年来,随着人工智能和机器学习的发展,图像分类技术也取得了快速的发展。一般而言,这些算法可以归纳为两大类<sup>[3]</sup>,第一类是基于分类器学习的,这类算法需要参数学习的过程;第二类是基于非参数学习的,相比于前类方法,这类算法不需要参数学习的过程,分类

决策直接取决于分类数据并能避免过拟合等问题。

近年来,视觉词袋(BoVW)模型<sup>[4]</sup>被广泛使用并在图像分类任务中取得了非常不错的性能。基于BoVW模型的图像分类算法主要包括以下步骤:1)特征提取与描述;2)构建视觉词典;3)特征编码(Encoding)与汇聚(Pooling);4)分类器学习。构建视觉词典和特征编码与汇聚是上述过程的重要组成部分,很大程度上决定了分类的性能,但上述两个过程存在视觉词典语义信息丢失和量化误差等局限性。

为了解决上述问题,Yang等<sup>[5]</sup>基于空间金字塔匹配核提

到稿日期:2015-04-28 返修日期:2015-07-02 本文受国家自然科学基金项目(61172164),安徽省自然科学基金项目(1508085QF114,1608085QF144)资助。

张旭(1981-),男,博士生,讲师,主要研究方向为图像分类、物体识别,E-mail:zhangxu21cn@163.com(通信作者);蒋建国(1955-),男,教授,主要研究方向为数字图像分析与处理、分布式智能系统和DSP技术及应用;洪日昌(1981-),男,博士,教授,主要研究方向为多媒体信息处理、信息检索;杜跃(1982-),女,硕士,讲师,主要研究方向为数据挖掘、数字图像处理。

出一种 ScSPM 分类算法,该算法通过稀疏编码,借助最小二乘优化问题利用部分视觉单词对各局部特征进行重构,最后通过最大值汇聚操作获取图像的最终表示,但重构和最大值汇聚操作导致了一定程度的信息损失,尽管基于稀疏编码和局部约束线性编码的重构编码方法取得了较好的性能。

稀疏编码理论在人脸识别领域也取得显著性的成功。Wright 等<sup>[6]</sup>提出一种基于稀疏表示的人脸识别算法(SRC)。然而,Zhang 等人<sup>[7]</sup>认为 SRC 算法具有有效性的原因并非在于  $\ell^1$  正则化稀疏表示,而是协作表示(CRC)即采用所有类别的全部训练样本表示测试图像。SRC 和 CRC 直接使用原始训练图像集作为视觉词典,然而当训练样本存在受损、干扰、遮挡、形变等情况时,算法均缺乏稳健性,分类识别性能较差。而现实条件下获取的自然图像几乎不可避免地受到噪声、遮挡、形变等因素的干扰,实验证明若采用这类受污染的图像直接作为视觉词典将会降低算法的性能。

为了解决上述问题,Chen 等<sup>[8]</sup>基于低秩稀疏分解将同类训练样本所构成的矩阵分解为一个低秩矩阵噪声矩阵并将噪声从原训练样本中直接丢弃,以提高人脸识别的鲁棒性和正确率。Zhang 等在文献<sup>[9]</sup>中对隶属于同一类别的图像进行低秩稀疏分解,得到局部特征矩阵的低秩部分和稀疏部分,并将这两部分直接组合构建成视觉词典,结合 SVM 分类器算法取得了令人满意的结果。然而,上述算法并没有完全展现低秩部分和稀疏部分对分类性能的影响,并且该算法需要参数学习的过程,在实际情况中往往存在训练样本不足及过拟合等问题。

实际应用中获取的自然图像更具复杂性,更易受诸如光照变化、尺度变换、遮挡、背景信息等因素的干扰,并且同类图像的目标物体也表现出明显的差异性。尽管如此,但同类别图像的同质性导致目标信息和背景信息表现出一定的相似性和相关性。特别是在提取图像局部特征的过程中如 SIFT,为了获得更优的表示能力,往往采用密集且重叠的方式采样图像块。因此同类训练图像局部特征所构成的矩阵具有明显的低秩性。通过低秩稀疏分解,该矩阵亦可分解为低秩矩阵和稀疏矩阵两部分,其中低秩矩阵表征了同类图像集的相关性和结构性信息,而稀疏矩阵表征了图像集的差异性,在图像分类过程中,充分利用同类图像间的相关性和差异性将会提高算法的性能。

另一方面,文献<sup>[3]</sup>提出一种朴素贝叶斯最近邻图像分类算法(NBNN),其作者认为基于非参数的最近邻分类的性能在实际应用中被低估了,导致其性能下降的原因一方面是算法中的特征量化引起量化误差,带来了信息损失;另一方面是算法采用图像与图像之间的距离进行分类决策。NBNN 算法不是采用传统图像与图像之间的距离度量方式,而在特征空间中度量图像与类别之间的距离,从而避免了特征量化,具有更好的泛化能力。

据上所述,本文在朴素贝叶斯最近邻和低秩稀疏分解的理论框架下,基于非负稀疏编码和协作表示提出一种非参数的图像分类算法。本文主要的创新点有:1)对同类别的训练样本,基于非负稀疏编码构建具有低秩性的表示矩阵;2)充分利用类别标签和同类图像间的相关性和差异性,基于低秩稀疏分解的低秩矩阵和稀疏矩阵构建视觉词典;3)采用协作表示的方式表征测试图像,基于朴素贝叶斯最近邻的原理对图像进行分类。

## 2 相关工作

视觉词袋模型已经广泛应用于图像分类领域并取得了一定成功,近年来,针对该模型的各种改进算法被提出,算法研究的重点主要包括字典学习、特征编码与汇聚、分类器学习等。在众多研究中,Lazebnik 等<sup>[10]</sup>提出空间金字塔匹配核(SPM),表征图像时包含了局部特征的空间信息,分类正确率有了大幅度的提高。Yang 等<sup>[5]</sup>在 SPM 的基础上提出一种 ScSPM 的分类算法,将稀疏编码用于图像分类,通过建立带约束的最小二乘优化问题来完成,仅基于线性 SVM 分类器,该算法便在多个通用图像分类数据库上表现出优异的性能。Wang 等<sup>[11]</sup>增加了约束的条件,提出局部约束线性编码(LLC),在保证分类正确率的同时进一步提高了算法的运行速度。

Wright 等<sup>[6]</sup>提出一种基于稀疏表示的人脸识别算法(SRC),算法通过求解  $\ell^1$  最小化正则项,基于最小化重构误差对测试图像进行分类并取得显著性的成功。Liu 等<sup>[12]</sup>提出一种图像的非负稀疏表示方式,并将该方法应用于人脸识别和图像分类任务中。另一方面,Zhou 等<sup>[13]</sup>在 K-SVD 的基础上增加标签一致性约束,学习更简洁和更具区分性能的视觉词典。Zhang 等<sup>[7]</sup>指出 SRC 算法的优异性并非源于对编码系数的稀疏性约束而是采用全部训练图像协作表示测试图像,由于具有解析解,CRC 算法运行速度更快。

然而,在实际应用中,特别是自然图像,无论是训练图像还是测试图像均易受到遮挡、形变、光照及其他噪声污染的影响,此时, SRC 和 CRC 算法的分类性能均受到很大程度的影响。Chen 等<sup>[8]</sup>在人脸识别中使用低秩稀疏分解技术去除训练图像中的噪声信息以解决上述问题,算法将训练图像所构成的视觉词典分解为低秩部分和稀疏部分,并将稀疏部分所对应的噪声信息直接丢弃。

低秩稀疏分解考虑如何从包含稀疏噪声的数据中恢复数据的低秩结构,该方法在信号处理和机器学习等领域已经获得了广泛的应用,如背景建模<sup>[14]</sup>、目标跟踪<sup>[15]</sup>、子空间分割<sup>[16]</sup>等。目前低秩表示模型也广泛用于图像分类与物体识别任务中,如 Jiang 等<sup>[17]</sup>基于局域受限低秩表示编码提出一种图像分类方法,Zhang<sup>[18]</sup>将低秩稀疏分解和拉普拉斯组稀疏编码相结合用于图像分类,均取得了不错的性能。加速近邻梯度法(APG)和增广拉格朗日乘子算法(ALM)<sup>[19]</sup>可以有效求解上述问题。

## 3 低秩稀疏表示

### 3.1 非负稀疏编码

假设  $X$  表示图像的某类局部特征集合,记为  $X=[x_1, x_2, \dots, x_N]$  ( $x_i \in \mathbb{R}^{d \times 1}, i=1, \dots, N$ ),如 128 维的 SIFT 特征。在 BoVW 模型中,K-均值聚类被广泛地应用于构建视觉词典,每个聚类中心对应一个视觉单词,表示为  $V=[v_1, v_2, \dots, v_k] \in \mathbb{R}^{d \times k}$ ,常采用矢量量化进行编码,求解如式(1)所示的目标函数最小二乘问题。

$$\min_{U, V} \sum_{i=1}^N \|x_i - u_i v\|^2 \quad (1)$$

Subject to  $Card(u_i)=1, |u_i|=1, u_i \geq 0, \forall i$

其中,  $Card(u_i)=1$  表示一个基约束条件,表示  $u_i$  中仅有一个

编码系数不为零,且其值为 1。矢量量化方法中的约束条件过于严格,导致编码过程中特征信息丢失。Yang 等<sup>[5]</sup>通过稀疏编码松弛了上述约束条件,在优化函数中使用  $u_i$  的 1-范数替换原始的约束条件,保证了编码系数的稀疏性。

$$\begin{aligned} \min_{U, V} \sum_{i=1}^N \|x_i - u_i V\|^2 + \lambda \|u_i\|_1 \\ \text{Subject to } \|v_k\|^2 \leq 1, \forall k \end{aligned} \quad (2)$$

其中,  $\lambda$  为正则化权值,  $\lambda$  越大系数的程度就越大;  $\|\cdot\|_1$  表示 1-范数。约束条件  $\|v_k\|^2 \leq 1, \forall k$  可以避免平凡解。

非负稀疏编码可以有效地解决上述问题,其保留着标准稀疏编码的优点,并且与人类视觉神经机能更具一致性。与标准稀疏编码相比,非负稀疏编码增加了编码系数非负性的约束条件,其优化的目标函数如式(3)所示:

$$\begin{aligned} \min_{U, V} \sum_{i=1}^N \|x_i - u_i V\|^2 + \lambda \|u_i\|_1 \\ \text{Subject to } \|v_k\|^2 \leq 1, u_i \geq 0, \forall k, i \end{aligned} \quad (3)$$

上述目标函数的组合优化问题不是一个关于系数  $U$  和字典  $V$  的凸集,但当固定两个变量中的任一个并将其看为常量而只考虑另一个变量时,目标函数可看成关于该变量的凸函数,具体求解过程可参考文献[12]。

### 3.2 结构不一致性低秩稀疏编码

给定满足特定条件的矩阵  $D$ , 该技术将其分解为低秩矩阵  $A$  和稀疏矩阵  $E$ , 目标函数如式(4)所示:

$$\begin{aligned} \min_{A, E} \text{rank}(A) + \beta \|E\|_0 \\ \text{Subject to } D = A + E \end{aligned} \quad (4)$$

其中,  $\text{rank}(\cdot)$  表示矩阵秩函数,  $\|\cdot\|_0$  表示矩阵的 0-范数, 即矩阵非零元素的个数。  $\beta$  为平衡参数, 用于调节低秩项和稀疏项。秩函数在矩阵谱范数单位球上的凸包络是矩阵的核范数, 而 0-范数在无穷范数单位球上的凸包络是其 1-范数。因此上式可以转化为式(5):

$$\begin{aligned} \min_{A, E} \|A\|_* + \beta \|E\|_1 \\ \text{Subject to } D = A + E \end{aligned} \quad (5)$$

其中,  $\|\cdot\|_*$  表示矩阵的核范数(即矩阵所有奇异值的和),  $\|\cdot\|_1$  表示矩阵的 1-范数, 增广拉格朗日乘子算法可以有效地解决该问题。

为了增强不同类别对应低秩矩阵之间的一致性, 进一步提高视觉词典的判别性能, 采用 Chen 等<sup>[8]</sup>所述方法, 在目标函数(5)后增加一判别项, 改进后的目标函数如式(6)所示:

$$\begin{aligned} \min_{A, E} \sum_{i=1}^N \{ \|A_i\|_* + \beta \|E_i\|_1 \} + \eta \sum_{j \neq i} \|A_j^T A_i\|_F \\ \text{Subject to } D_i = A_i + E_i \end{aligned} \quad (6)$$

其中,  $\eta$  为惩罚因子, 为了求解上式, 根据  $\|A_j^T A_i\|_F^2 \leq \|A_j\|_F^2 \|A_i\|_F^2$ , 式(6)被松弛为:

$$\begin{aligned} \min_{A_i, E_i} \|A_i\|_* + \beta \|E_i\|_1 + \eta' \|A_i\|_F^2 \\ \text{Subject to } D_i = A_i + E_i \end{aligned} \quad (7)$$

优化过程中,  $\eta' = \eta \sum_{j \neq i} \|A_j\|_F^2$  可被认为是常量, 通过扩展的增广拉格朗日乘子算法可求解  $A_i$  和  $E_i$ , 具体过程可参照文献[8]。

## 4 本文算法

在基于 BoVW 的模型中, 图像最终被表示为具有固定长度的向量, 如果以图像的代表向量为列, 将同类图像的局部特

征表示为矩阵, 则该矩阵具有低秩性。经过低秩稀疏分解之后, 得到低秩部分和稀疏部分, 其中前者表征了同类图像间的相关性和一致性, 后者表征了同类图像间的差异性和多样性。基于此, 本文在 NBNN 原理的框架下, 基于非负稀疏编码、低秩稀疏分解及协助表示提出一种新的非参数学习图像分类算法, 具体过程描述如下。

输入: 带有类别信息的  $N$  类训练样本以及测试样本  $X = [x_1, x_2, \dots, x_n]$ , 各类参数  $\lambda, \beta, \eta, \gamma, \alpha$

输出: 测试图像的分类信息

步骤 1 采用非负稀疏编码表示训练图像和测试图像。

$$\begin{aligned} \min_{U, V} \sum_{i=1}^N \|x - uV\|^2 + \lambda \|u\|_1 \\ \text{Subject to } \|v\|^2 \leq 1, u \geq 0 \end{aligned}$$

$U = [U_1, U_2, \dots, U_N]$  和  $y$  分别表示训练图像和测试图像的代表向量。

步骤 2 采用式(6)对  $U$  进行低秩稀疏分解:

For  $i=1:N$  do

$$\begin{aligned} \min_{A, E} \sum_{i=1}^N \{ \|A_i\|_* + \beta \|E_i\|_1 \} + \eta \sum_{j \neq i} \|A_j^T A_i\|_F \\ \text{Subject to } U_i = A_i + E_i \end{aligned}$$

End for

并将低秩部分和稀疏部分表示为:  $A = [A_1, A_2, \dots, A_N]$  和  $E = [E_1, E_2, \dots, E_N]$ 。

步骤 3 根据类别标签信息, 以  $A$  作为视觉词典协作表示测试图像  $y_i$ , 并计算编码系数。

$$\begin{aligned} \min_{w_A} \|y_i - A w_A\|_2^2 + \gamma \|w_A\|_2 \\ \hat{w}_A = |w_A| / \|w_A\|_1 \end{aligned}$$

For  $i=1:N$  do

$$\hat{c}_A(i) = \sum_i \hat{w}_{A_i}$$

End for

步骤 4 根据类别标签信息, 以  $E$  作为视觉词典协作表示测试图像  $y_i$ , 并计算编码系数。

$$\begin{aligned} \min_{w_E} \|y_i - E w_E\|_2^2 + \gamma \|w_E\|_2 \\ \hat{w}_E = |w_E| / \|w_E\|_1 \end{aligned}$$

For  $i=1:N$  do

$$\hat{c}_E(i) = \sum_i \hat{w}_{E_i}$$

End for

$$\text{density}(q) = \arg \max_i (\alpha \hat{c}_A(i) + (1 - \alpha) \hat{c}_E(i))$$

## 5 实验结果与分析

### 5.1 实验配置

为了验证本文所提算法的有效性, 选择在 3 个图像分类通用图像库上进行分类识别, 即 Scene-15 图像库、Caltech-101 图像库以及 Caltech-256 图像库。本文所有实验均使用 2-范数归一化的 128 维 SIFT 描述子表示图像局部特征, 采用稠密重叠方式提取图像块的大小为  $16 \times 16$ , 步长为 8。在 Scene-15 及 Caltech 101 中非负稀疏编码过程的视觉词典大小为 1024, 而 Caltech 256 图像库中视觉词典的大小设置为 2048。为了获得可靠稳定的实验结果, 采用图像分类算法的一般性做法, 每组实验分别执行 5 次, 每次重新随机选择训练图像和测试图像, 模型评价指标为图像集中所有类别图像的平均分类正确率和方差。

本文算法包含了 5 个主要的参数, 其中  $\lambda$  在非负稀疏编

码中用于控制编码系数的稀疏性,一般情况 $\lambda$ 越大,所获得的编码系数越稀疏,实际经验中系数的稀疏性在10%左右效果较优,本实验统一设置 $\lambda=0.15$ ;  $\gamma$ 为协作表示的正则化参数,实验中统一取 $\gamma$ 的值为0.001。低秩稀疏分解过程主要包括参数 $\beta$ 和 $\eta$ ,其中 $\beta$ 用于平衡低秩部分和稀疏部分,随着 $\beta$ 的增加稀疏矩阵越稀疏; $\eta$ 是一个正的平衡因子,用于平衡低秩矩阵部分的不一致性; $\alpha$ 为分类决策中的权重系数。为了获取较优的分类性能,算法结合实际经验,根据不同的图像库分别调整上述3个参数的取值。特别地,与基于参数学习的分类方法不同,非参数化的分类算法不需要训练样本训练分类器,但是为了便于比较与表述,仍使用训练样本表示带有类别标签的参考图像,使用测试图像表示待分类图像。

## 5.2 Scene-15 dataset

Scene-15 图像集包含 15 个类别共计 4485 幅图像,每个类别大概有 200~400 幅图像。该图像通常用于场景识别,如办公室、卧室、厨房、客厅和商店等室内场景,以及海岸、森林、公路、山脉、原野等室外场景。图 1 给出了 Scene-15 图像集的示例图像。按照通常做法,每类随机选择 100 幅图像作为训练图像,其余的作为测试图像。低秩稀疏分解过程中取 $\beta=0.02$ , $\eta=1e^{-5}$ ,权重系数 $\alpha=0.8$ 。



图 1 Scene-15 图像集示例图像

实验结果如表 1 所列,从中可以看出,本文算法取得了与 ScSPM 相当的分类正确率,低于其他基于分类器学习的算法(如 LLC, KSPM 和 LScSPM 算法),但优于除了 NBNN5 以外的其他非参数学习算法(如 NBNN, Local NBNN 及 NBINN+NIMBLE 算法)。NBNN5 算法在表示图像信息时,不仅使用了 SIFT 特征描述子,而且还结合了亮度、颜色、形状以及自相似描述子,因此更具有有效性和鲁棒性,尽管该算法获取的分类性能较高,但其复杂性较高,运行速度较慢。本文采用非负稀疏编码和最大值汇聚操作表征图像信息,能够保留更多有用的信息,有效减少量化损失。此外,算法基于低秩稀疏分解构建视觉词典,并采用协作表示的方式对测试图像进行编码,充分利用了训练图像间的相关性和差异性。LScSPM<sup>[20]</sup> 提出拉普拉斯稀疏编码方式,结合分类器学习的过程取得了非常理想的性能,但该算法需要充分训练学习的过程。

表 1 Scene-15 图像集上分类性能比较

算法	分类正确率	是否需要参数学习
KSPM <sup>[10]</sup>	81.40±0.50	Yes
ScSPM <sup>[5]</sup>	80.28±0.93	Yes
LLC <sup>[11]</sup>	81.50±0.47	Yes
LScSPM <sup>[20]</sup>	89.75±0.50	Yes
NBNN <sup>[3]</sup>	75.00±3.30	No
Local NBNN <sup>[21]</sup>	79.28±2.34	No
NBINN+NIMBLE <sup>[22]</sup>	78.23±1.00	No
NBNN5 <sup>[3]</sup>	82.73±1.81	No
本文算法	80.31±3.11	No

## 5.3 Caltech-101 dataset

Caltech-101 图像集包含了 101 类图像,如动物、家具、车辆、花朵等类别,每类图像包含 30~800 幅图像。该图像库具有较大的类内差异,是图像分类中最常用的数据库之一,示例图像如图 2 所示。按照通常的实验设置,分别从每类中选取 15 幅和 30 幅作为训练图像,15 幅作为测试图像。实验中, $\beta=0.02$ , $\eta=1e^{-5}$ ,权重系数 $\alpha=0.6$ 。



图 2 Caltech-101 图像集示例图像

表 2 给出了本文算法和其他一些常用算法在该图像集上的相关结果。当训练图像数为 15 时,本文算法优于 NBNN 和 Local NBNN 等非参数学习分类算法和部分基于分类器学习的分类算法如 KSPM, SVM-KNN 及 ScSPM 等。特别地,当训练图像数为 15 时,本文算法比 LLC 算法提高了将近 2 个百分点,而当训练图像数为 30 时,算法分类正确率低于 LLC 和 ScSPM 算法,但仍优于其他非参数学习分类算法。值得注意的是在不同训练图像数目下,LR-Sc+SPM 算法均取得了非常不错的性能,该算法采用非负稀疏编码和最大值汇聚表征图像,有效地减少了信息损失,并通过低秩稀疏分解构建更具判别性能的视觉词典,这与本文算法思想一致,也从侧面证明了本文算法的有效性,但算法直接将低秩部分和稀疏部分相组合构建视觉词典,并没有分析两部分对分类性能的影响,且额外的训练学习过程进一步提升了算法的性能。

表 2 Caltech-101 图像集上分类性能比较

算法	15 幅训练图像	30 幅训练图像	是否需要参数学习
KSPM <sup>[10]</sup>	56.40	64.40±0.80	Yes
KSPM <sup>[23]</sup>	—	64.14±1.18	Yes
SVM-KNN <sup>[24]</sup>	59.10±0.60	66.20±0.50	Yes
ScSPM <sup>[5]</sup>	67.00±0.45	73.20±0.54	Yes
LLC <sup>[11]</sup>	65.43	73.44	Yes
LR-Sc+SPM <sup>[9]</sup>	69.58±0.97	75.68±0.89	Yes
NBNN <sup>[3]</sup>	65.00±1.14	70.40	No
Local NBNN <sup>[21]</sup>	66.1±1.17	71.9±0.6	No
本文算法	67.11±2.14	72.16±2.42	No

## 5.4 Caltech-256 dataset

Caltech-256 图像集包含 256 类共 29780 幅图像,其中每类别包含至少 80 幅图像。与 Caltech-101 图像集相比,由于类间差异性大以及图像自身更具多样性和复杂性,该图像集更具挑战性,示例图像如图 3 所示。实验分别从每类图像中随机选择 15 幅和 30 幅图像作为训练图像,30 幅作为测试图像。当训练样本为 15 时, $\beta=0.02$ , $\eta=1e^{-3}$ ;训练样本为 30 时, $\beta=0.01$ , $\eta=1e^{-5}$ ;权重系数 $\alpha=0.6$ 。

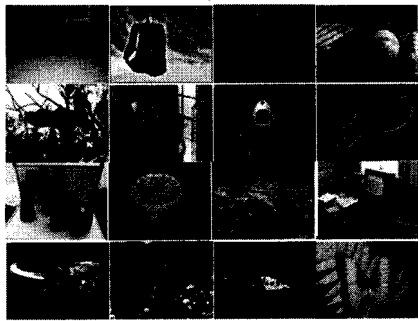


图3 Caltech-256 图像集示例图像

Caltech-256 图像集上的实验结果如果表 3 所列。从中可以发现,一方面,除 LR-Sc+SPM(训练样本为 15 时)外,本文算法的分类正确率优于其他基于分类器学习的算法;另一方面,除 Local NBNN(训练样本为 30 时)外,本文算法优于其他非参数学习算法。由于减小了图像表示过程中信息的损失,并且采用低秩矩阵分解的方法充分利用同类图像集的相关性和差异性,因此本算法对这类差异较大的图像的分类效果较理想。在训练图像个数为 15 时,ScSPM 和 LLC 算法的分类正确率分别为 27.73% 和 27.74%,而本算法较两者而言正确率提高了约 6%;当训练图像个数为 30 时,本算法正确率较 ScSPM 和 LLC 算法分别提高了近 3% 和 5%,尽管与 NBNN 算法的分类正确率相当,但由于协作表示具有解析解并且不需要最近邻搜索,因此本文算法运行速度更快。

表3 Caltech-256 图像集上分类性能比较

算法	15 幅训练图像	30 幅训练图像	是否需要参数学习
KSPM <sup>[10]</sup>	—	34.10	Yes
KCSPM <sup>[23]</sup>	—	27.17±0.46	Yes
ScSPM <sup>[5]</sup>	27.73±0.51	34.02±0.35	Yes
LLC <sup>[11]</sup>	27.74±0.32	32.07±0.24	Yes
LR-Sc+SPM <sup>[9]</sup>	35.31±0.70	—	Yes
NBNN <sup>[3]</sup>	30.5	37.00	No
Local NBNN <sup>[21]</sup>	33.5±0.9	40.1±0.1	No
本文算法	33.73±1.34	37.38±1.16	No

进一步可以发现,相比基于参数学习的分类算法,非参数学习分类算法在 Caltech-256 图像集上获得了较好的分类正确性,这也体现了非参数学习分类算法的优点:不需要参数学习的过程,避免了参数的过拟合等问题,并且具有更好的泛化性能,能够更合理地应用于类别较多的分类任务中。

### 5.5 讨论与分析

本文算法对同类训练图像的特征矩阵进行低秩稀疏分解,获得低秩矩阵和稀疏矩阵,分别将其作为视觉词典协作表示测试图像,并通过不同的权重系数对测试图像进行分类决策。低秩矩阵一定程度上表征了类别图像的结构性和相关性,稀疏矩阵一定程度上反映了类别图像间的差异性,由于两类词典具有不同的物理含义,算法通过不同的权重系数反映其对分类决策的影响。在 Scene-15 和 Caltech-101 图像集上衡量两类视觉词典对最终分类性能的影响,选择权重系数  $\alpha$  从 0 到 1,步长为 0.1,其中 Caltech-101 图像集中训练样本个数为 30。

实验结果如图 4 所示,图像分类正确率随着  $\alpha$  的变化而变化,反映出两类视觉词典对分类性能的影响。特别地,当  $\alpha$

取 0 时,在 Scene-15 和 Caltech-101 图像集上的分类正确率分别为 29.14% 和 19.74%;当  $\alpha$  取 1 时,在 Scene-15 和 Caltech-101 图像集上的分类正确率分别为 65.27% 和 42.18%,由于同类训练图像被表示为 BoVW 模型的特征向量,因此基于低秩矩阵构建的视觉词典对分类决策的影响较大。

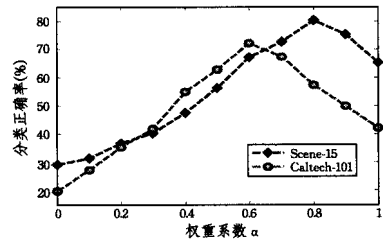


图4 权重系数对分类性能的影响

为了进一步衡量本算法,分别在 Scene-15 和 Caltech-101 图像集中随机逐次选择 2 到 10 个图像作为训练图像,选择 20 个图像作为测试图像。实验给出了本文算法、NBNN、Sc-SPM 和 LLC 算法在不同训练样本数目下的分类正确率,实验结果如图 5 和图 6 所示。在 Scene-15 图像集中,当训练样本数目较少(小于 4)时,ScSPM 算法性能最优,相同情况也出现在 Caltech-101 图像集中;然而随着训练样本数目的增加,本文算法逐渐表现出较优的性能,在 Caltech-101 图像集中这种现象更加明显。

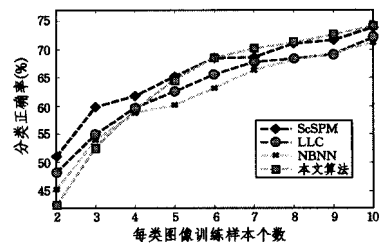


图5 Scene-15 图像集上实验结果对比

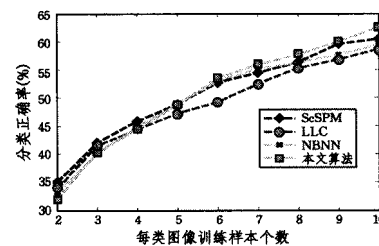


图6 Caltech-101 图像集上实验结果对比

在上述实验的各算法中,ScSPM 和 LLC 算法通过建立带约束的最小二乘优化来完成部分视觉单词对各局部特征的重构,尽管稀疏编码避免了矢量化带来的误差,但最大值汇聚操作将会带来信息损失。特别是 LLC 算法,在重构编码的过程中增加了局部性约束,即仅仅通过局部少量的视觉单词重构局部特征,由于视觉单词数目小于局部特征的维数,该类重构通常是欠定问题,因此不可避免地出现重构误差。NBNN 算法基于图像-类别距离直接对训练图像进行分类识别,算法简洁明了,但没有充分训练图像集的结构性和相关性,并且对测试图像的每个局部特征搜索最近邻的开销较高。本文算法有效地解决了上述问题,不仅充分利用了同类图像集间的相关性和差异性来构建视觉词典,并在此基础上采用

协作表示对图像进行分类决策,运行速度更快。特别地,本文算法继承了非参数学习算法的优点,具有更好的泛化性能,能够避免分类器训练过程中的问题。

**结束语** 本文基于非负稀疏编、低秩稀疏分解和协作表示等技术,提出一种非参数学习的图像分类方法。算法采用非负稀疏编码结合最大值汇聚操作表示具有低秩性质的局部特征矩阵,结合类别信息利用低秩稀疏分解的低秩矩阵和稀疏矩阵分别构建视觉词典,在 NBNN 算法原理的框架下,采用协作表示表征测试图像并进行分类决策,实验结果验证了算法的有效性。但本算法结合类别信息构建视觉词典缺乏学习的过程,并在基于两类视觉词典的编码系数进行分类决策时不能够自适应地调整权重系数,这也是后期算法需要进一步研究的问题。

### 参 考 文 献

- [1] Hong R, Tang J, Tan H K, et al. Beyond search; event-driven summarization for web videos [J]. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2011,7(4):702-715
- [2] Hong R, Wang M, Gao Y, et al. Image annotation by multiple-instance learning with discriminative feature mapping and selection[J]. *IEEE Transactions on Cybernetics*, 2014, 44(5): 669-680
- [3] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification[C]//*IEEE Conference on Computer Vision and Pattern Recognition*, 2008 (CVPR 2008). IEEE, 2008;1-8
- [4] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos[C]//*Ninth IEEE International Conference on Computer Vision*, 2003. IEEE, 2003;1470-1477
- [5] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//*IEEE Conference on Computer Vision and Pattern Recognition*, 2009(CVPR 2009). IEEE, 2009;1794-1801
- [6] Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition [J]. *Proceedings of the IEEE*, 2010,98(6):1031-1044
- [7] Zhang D, Yang M, Feng X. Sparse representation or collaborative representation: Which helps face recognition? [C]//*2011 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2011;471-478
- [8] Chen C F, Wei C P, Wang Y C F. Low-rank matrix recovery with structural incoherence for robust face recognition [C]//*2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012;2618-2625
- [9] Zhang C, Liu J, Liang C, et al. Image classification by non-negative sparse coding, correlation constrained low-rank and sparse decomposition[J]. *Computer Vision and Image Understanding*, 2014, 123;14-22
- [10] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]//*2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2006, 2;2169-2178
- [11] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification[C]//*2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010;3360-3367
- [12] Liu Y, Wu F, Zhang Z, et al. Sparse representation using non-negative curds and whey[C]//*2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010;3578-3585
- [13] Jiang Z, Lin Z, Davis L S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD[C]//*2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011;1697-1704
- [14] Zhou X, Yang C, Yu W. Moving object detection by detecting contiguous outliers in the low-rank representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013,35(3);597-610
- [15] Zhang T, Liu S, Ahuja N, et al. Robust visual tracking via consistent low-rank sparse learning [J]. *International Journal of Computer Vision*, 2014,111(2);171-190
- [16] Zhang X, Sun F, Liu G, et al. Fast low-rank subspace segmentation [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014,26(5);1293-1297
- [17] Jiang Z, Guo P, Peng L. Locality-Constrained Low-Rank Coding for Image Classification[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2014;2780-2786
- [18] Zhang L, Ma C. Low-rank decomposition and Laplacian group sparse coding for image classification [J]. *Neurocomputing*, 2014,135;339-347
- [19] Shen Y, Wen Z, Zhang Y. Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization [J]. *Optimization Methods and Software*, 2014, 29(2): 239-263
- [20] Gao S, Tsang I W H, Chia L T, et al. Local features are not lonely-Laplacian sparse coding for image classification [C]//*2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010;3555-3561
- [21] McCann S, Lowe D G. Local naive bayes nearest neighbor for image classification[C]//*2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012;3650-3656
- [22] Timofte R, Van Gool L. Iterative nearest neighbors for classification and dimensionality reduction[C]//*2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012;2456-2463
- [23] Van Gemert J C, Veenman C J, Smeulders A W M, et al. Visual word ambiguity[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010,32(7);1271-1283
- [24] Zhang H, Berg A C, Maire M, et al. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition [C]//*2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2006, 2;2126-2136