

基于核拉普拉斯稀疏编码的图像分类

张立和*, 潘磊, 刘涛, 马臣

(大连理工大学 信息与通信工程学院, 辽宁 大连 116024)

摘要: 使用稀疏编码解决计算机视觉问题可以取得良好的效果. 然而, 以往的稀疏编码都是在原始特征空间进行. 受核方法可以获得特征的高维非线性映射的启发, 扩展了拉普拉斯稀疏编码(LSc), 提出了核拉普拉斯稀疏编码(KLSc), 它可以降低特征量化误差, 增强稀疏编码的性能. 在3个标准数据集上的实验结果表明, 所提出的基于KLSc的图像分类算法具有良好的分类效果, 分类正确率优于LSc.

关键词: 图像分类; 稀疏编码; 拉普拉斯稀疏编码; 核方法; 空间金字塔匹配(SPM)

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.7511/dllgxb201502011

0 引言

近年来, 词包(bag of words, BoW)模型^[1]由于在尺度、平移和旋转等方面具有很强的鲁棒性而备受关注. 研究者提出很多扩展 BoW 模型, 例如, 模拟码字或描述子同现的生成算法^[2]; 替代标准无监督聚类算法的字典学习方法^[3]; 模拟局部特征空间分布的空间金字塔匹配核^[4]等. 在这些扩展中, 受 Grauman 和 Darrell 提出的空间金字塔匹配(spatial pyramid matching, SPM)的启发, Lazebnik 等提出的 SPM 取得了很大的成功^[4].

由于简单易用, SPM 方法在实际应用中非常有效, 它是当前分类系统的重要组成部分. 尽管 SPM 方法很流行, 但人们发现它只有与非线性核结合才能取得好的性能, 如交叉核和卡方检验核, 这就需要复杂的计算和大量的存储空间. 认识到这一点, Bosch 等用随机树^[5]代替支持向量机(support vector machine, SVM)来进行快速训练和检测. 后来, Maji 等提出构建直方图交集核 SVM 更有效^[6], 然而, 其只对经过预先训练的非线性 SVM 有效. 而实际应用中通常包含成千上万的训练特征, 都预先训练是不现实的. 因此, 线

性核 SVM 更受青睐, 因为其训练和测试速度更快, 需要的存储空间更小.

Yang 等采用稀疏编码技术进行字典生成和特征量化用于 BoW 模型, 并结合 SPM 构成图像的特征表示^[7], 在图像分类中取得了很好的效果. 该方法给每个局部特征分配优化的权重(即编码), 使特征量化精度有了很大提高. 然而, 该方法忽视了特征之间的相互关系, 可能导致相似特征的编码差异很大. 为此, Gao 等^[8]提出加入一个拉普拉斯正则化项来描述局部特征间的相关性, 使特征表示更具鲁棒性, 称为拉普拉斯稀疏编码(LSc). Kavukcuoglu 等提出空间一致性稀疏编码用于局部特征探测^[9]. 他们通过对稀疏表示加入空间一致性约束, 可以获取有鉴别力的特征. Mairal 等成功将稀疏编码应用于产生字典一个子集响应的所有特征^[10]. 通过这种方法, 可以获得特征间的相似性.

现有 BoW 模型的特征量化方法都是在原始信号空间进行的. 联想到核方法可以将信号映射到高维空间^[11], 使得信号在高维空间里更倾向于可线性分离, 从而可以更容易得到信号的稀疏表

示,且降低特征量化误差. 本文将核方法用于拉普拉斯稀疏编码中,提出核拉普拉斯稀疏编码(KLSc)并将其应用到图像分类中.

1 核拉普拉斯稀疏编码的构成

1.1 拉普拉斯稀疏编码

对于一般的稀疏编码,其任务是在字典 $U \in \mathbb{R}^{d \times n}$ 上得到局部特征的稀疏表示,同时最小化重建误差. 其目标函数如下:

$$\begin{aligned} \min_{U, v} \|x - Uv\|^2 + \lambda \|v\|_1 \\ \text{s. t. } \|u_i\|^2 \leq 1 \end{aligned} \quad (1)$$

式中: x 为 d 维的局部特征; U 为含 n 个基的字典, $U = (u_1 \ u_2 \ \cdots \ u_n)$; v 为 x 的编码; λ 为参数. 式(1)优化方程中第一项为重建误差,第二项的作用是控制编码 v 的稀疏性. 根据经验, λ 越大编码越稀疏.

由于字典的过完备性,局部特征之间的微小变化可能导致它们的编码差异很大. 也就是说,稀疏编码对局部描述子的变化非常敏感、不具鲁棒性;另一方面,稀疏编码对局部描述子独立处理,没有利用它们的相互关系. 为了描述局部特征间的关系,降低稀疏编码对局部描述子变化的敏感性, Gao 等提出了拉普拉斯稀疏编码^[8]. 通过在最优优化问题(1)中引入拉普拉斯正则化项,使相似特征编码的一致性保留下来. 这个方法提高了局部描述子的表示效果,使对局部描述子的表达更具鲁棒性,从而提升了图像分类性能. 其优化目标形式如下:

$$\begin{aligned} \min_{U, v} \|X - UV\|_F^2 + \lambda \sum_i \|v_i\|_1 + \\ \frac{\beta}{2} \sum_{i,j} \|v_i - v_j\|^2 W_{ij} = \end{aligned} \quad (2)$$

$$\min_{U, v} \|X - UV\|_F^2 + \lambda \sum_i \|v_i\|_1 + \beta \text{tr}(VLV^T)$$

$$\text{s. t. } \|u_m\|^2 \leq 1$$

式中: $X = (x_1 \ x_2 \ \cdots \ x_N)$, 为局部特征的集合. $V = (v_1 \ v_2 \ \cdots \ v_N)$, 为编码的集合. W 为相似矩阵,用来表达局部特征间的相似关系,可以通过

直方图交集来度量,即 $W(H, S) = \sum_{i=1}^p \min(H_i,$

$S_i)$, 这里 p 为两个直方图的维数. 拉普拉斯矩阵 $L = D - W$, 其中 D 为对角矩阵: $d_{ii} = \sum_j W_{ij}$.

应用式(2)进行稀疏编码需要两个步骤:

(1) 随机选择部分样本作为训练集,用这些样本训练字典,并保存训练样本的编码;(2) 获得字典后,利用下式对每个局部特征进行编码:

$$\min_{v, v_i} \|x - Uv\|^2 + \lambda \|v\|_1 + \beta \sum_i \|v - v_i\|^2 W_i$$

其中 v_i 为第一步保存下来的训练样本的稀疏编码, W_i 为被编码特征与训练样本的相似度.

1.2 核拉普拉斯稀疏编码

受核方法^[11]的启发,假设存在一个特征映射函数 $\phi: \mathbb{R}^d \rightarrow \mathbb{R}^k (d < k)$, 将特征和字典映射到高维空间,即 $x \rightarrow \phi(x)$, $U = (u_1 \ u_2 \ \cdots \ u_n) \rightarrow \mathcal{U} = (\phi(u_1) \ \phi(u_2) \ \cdots \ \phi(u_n))$. 将映射到高维空间的特征和字典应用到拉普拉斯稀疏编码中,得到核拉普拉斯稀疏编码:

$$\min_{v, v_i} \|\phi(x) - \mathcal{U}v\|^2 + \lambda \|v\|_1 + \beta \sum_i \|v - v_i\|^2 W_i$$

由于高斯核在许多研究^[11-12]中都取得了优异的性能,本文算法也采用高斯核: $\kappa(x_1, x_2) = \exp(-\gamma \|x_1 - x_2\|^2)$.

核拉普拉斯稀疏编码中的目标函数对字典 $\mathcal{U}$ 和编码 v 不同时为凸,不能同时进行优化. 因此,对字典 $\mathcal{U}$ 和编码 v 进行交替优化.

在映射空间稀疏编码:当固定字典 $\mathcal{U}$ 时,核拉普拉斯稀疏编码优化问题可以整理成

$$\begin{aligned} \min_{v, v_i} \|\phi(x) - \mathcal{U}v\|^2 + \lambda \|v\|_1 + \beta \sum_i \|v - v_i\|^2 W_i = \\ \kappa(x, x) + v^T \kappa_{UV} v - 2v^T \kappa_U(x) + \lambda \|v\|_1 + \\ \beta \sum_i \|v - v_i\|^2 W_i \end{aligned} \quad (3)$$

式中: $\kappa(x, x) = 1$; κ_{UV} 是 $n \times n$ 矩阵,元素 $(\kappa_{UV})_{ij} = \kappa(u_i, u_j)$; $\kappa_U(x)$ 是 $n \times 1$ 矢量,元素 $(\kappa_U(x))_i = \kappa(u_i, x)$. 除了 κ_{UV} 、 $\kappa_U(x)$ 这两个关于核的定义,其余部分与传统稀疏编码一样,可以很容易地通过 feature-sign 算法来求解^[13].

更新字典:当固定编码 v 时,学习字典 $\mathcal{U}$. 假设随机取样 N 个样本来训练字典,然后重写目标函数如下:

$$\begin{aligned}
 f(\mathbf{u}) = & \frac{1}{N} \sum_{i=1}^N [\|\phi(\mathbf{x}_i) - \mathbf{u} \mathbf{v}_i\|^2 + \lambda \|\mathbf{v}_i\|_1 + \\
 & \beta \sum_i \|\mathbf{v} - \mathbf{v}_i\|^2 W_i] = \\
 & \frac{1}{N} \sum_{i=1}^N \left[1 + \sum_{s=1}^k \sum_{t=1}^k v_{i,s} v_{i,t} \kappa(\mathbf{u}_s, \mathbf{u}_t) - \right. \\
 & \left. 2 \sum_{s=1}^k v_{i,s} \kappa(\mathbf{u}_s, \mathbf{x}_i) \right] + \\
 & \frac{1}{N} \sum_{i=1}^N \left[\lambda \|\mathbf{v}_i\|_1 + \beta \sum_i \|\mathbf{v} - \mathbf{v}_i\|^2 W_i \right]
 \end{aligned} \quad (4)$$

其中 s, t 用来索引字典中的列, 采用传统方法 (如随机梯度下降法) 优化字典是非常具有挑战性的, 因此, 本文采用交替优化 $\mathbf{u}$ 的每一列的方法来进行字典训练. 求 $f(\mathbf{u})$ 对 $\mathbf{u}_m$ 的导数, $\mathbf{u}_m$ 是需要更新的字典的第 m 列:

$$\begin{aligned}
 \frac{\partial f}{\partial \mathbf{u}_m} = & \frac{-4\gamma}{N} \sum_{i=1}^N \left[\sum_{t=1}^k v_{i,m} v_{i,t} \kappa(\mathbf{u}_m, \mathbf{u}_t) (\mathbf{u}_m - \mathbf{u}_t) \right] + \\
 & \frac{4\gamma}{N} \sum_{i=1}^N [v_{i,m} \kappa(\mathbf{u}_m, \mathbf{x}_i) (\mathbf{u}_m - \mathbf{x}_i)]
 \end{aligned} \quad (5)$$

为了获得优化的 $\mathbf{u}_m$, 设 $\partial f / \partial \mathbf{u}_m = \mathbf{0}$. 然而, 由于式 (5) 中存在项 $\kappa(\mathbf{u}_m, \cdot)$, 求解很困难, 因此, 采用近似解来代替精确解, 在第 n 次迭代中使用第 $n-1$ 次结果来计算核函数一项. 用 $\mathbf{u}_{m,n}$ 表示第 n 次迭代的稀疏编码, 则目标函数变为

$$\begin{aligned}
 \frac{\partial f}{\partial \mathbf{u}_{m,n}} \cong & \frac{-4\gamma}{N} \sum_{i=1}^N \left[\sum_{t=1}^k v_{i,m} v_{i,t} \kappa(\mathbf{u}_{m,n-1}, \mathbf{u}_t) \times (\mathbf{u}_{m,n} - \right. \\
 & \left. \mathbf{u}_t) \right] + \frac{4\gamma}{N} \sum_{i=1}^N [v_{i,m} \kappa(\mathbf{u}_{m,n-1}, \mathbf{x}_i) (\mathbf{u}_{m,n} - \\
 & \mathbf{x}_i)] = \mathbf{0}
 \end{aligned} \quad (6)$$

当保持其余列不变时, 只求解第 m 列 $\mathbf{u}_m$ 变为一个容易求解的线性计算. 另外, 采用 k 均值算法初始化字典.

1.3 基于 SPM 的图像表示

对每幅图像进行稀疏编码后, 本文用最大值合并方法来表示图像. 最大值合并方法就是用对基的最大响应值来表示图像区域. 具体的, 假设一个图像区域提取了 m 个局部特征, 即有 m 个稀疏编码, 且字典的大小为 k , 经最大值合并后这一图像区域用一个 k 维的矢量 $\mathbf{y}$ 来表示, 则矢量 $\mathbf{y}$ 的第 i 个元素 y_i 等于这一图像区域内所有稀疏编码第 i 个元素的最大值, 即

$$y_i = \max \{ |v_{1i}|, |v_{2i}|, \dots, |v_{mi}| \} \quad (7)$$

为了保持图像的空间信息, 本文采用了 SPM 方法把图像分成合适的子区域, 在各个子区域里计算局部特征的直方图, 然后把这些直方图串联成向量, 作为图像的最终表示. 一般的, 采用 $2^l \times 2^l$ 的子区域, 其中 $l = 0, 1, 2$. 当仅取 $l = 0$ 时, SPM 退化为 BoW.

2 实 验

2.1 参数设置

局部特征描述子是图像表示的基本因素. 本文采用在图像识别领域具有优越性能且被广泛应用的 SIFT^[14] 特征. 为和以前的工作^[7-8] 保持一致, 用稠密网格方法来提取特征区域, 并采用相同的参数设置进行 SIFT 提取. 根据经验, 采用小的网格提取特征会得到好的表示效果, 但是也会导致更多的区域块和更高的计算复杂度. 为与其他方法公平比较, 选择步进为 8 pixel, 块尺寸为 16 pixel $\times$ 16 pixel; 设置图像最大尺寸为 300 pixel $\times$ 300 pixel; 字典基的个数为 1 024, 且每个数据集都随机选择 1.2×10^5 个局部特征来训练字典; 稀疏编码时, 设置参数 $\lambda = 0.4, \beta = 0.2$; 高斯核函数中的 γ 设为 $1/64$. 对每个数据集来说, 所有数据都进行 6 次独立分类测试取平均以得到可靠的结果, 且训练图像都是随机选择. 对于 SPM, 本文用 $1 \times 1, 2 \times 2, 4 \times 4$ 的 3 层. 由于线性 SVM 在速度和性能方面的优势, 本文用它来进行分类.

2.2 Caltech 256 数据集

Caltech 256 数据集是被用于场景分类测试的一个大型图像数据库, 它包含 256 类, 共 29 780 幅图像. 这个数据集很有挑战性, 类内变化较大, 而且目标也未必落在图像中心附近. 为公平比较, 本文采用与文献^[15] 相同的设置, 每类取 30 幅图像作为训练集, 其余图像作为测试集. 实验结果表明, 本文所提出的方法可以获得较好的分类正确率 (见表 1).

2.3 Scene 15 数据集

Scene 15 数据集包含 15 类, 共 4 485 幅图像. 为公平比较, 本文采用与文献^[7] 相同的设置, 在每类中随机选择 100 幅图像作为训练集, 其余图像作为测试集. 本文方法的平均分类正确率达

到 90.49%(见表 2). 图 1 列出了混淆矩阵, 给出了每类的分类正确率和误分类率. 在混淆矩阵中, 第 i 行 j 列的数值为第 i 类误分为第 j 类的百分比, 对角线上的数值为每类的分类正确率.

表 1 Caltech 256 数据集上的分类性能

Tab. 1 Classification performance on the Caltech 256 dataset

方法	平均分类正确率/%
KSPM ^[4]	27.17±0.46
ScSPM ^[7]	34.02±0.35
KSRSPM ^[15]	35.67±0.10
LScSPM ^[8]	35.74±0.10
KLScSPM	37.13±0.51

2.4 UIUC-Sport 数据集

UIUC-Sport 数据集包含 8 类运动场景图像,

共 1 792 幅, 每类 137~250 幅. 根据 Wu 等的工作^[17], 每类随机选择 70 幅作为训练集, 再从剩余图像中每类随机选择 60 幅作为测试集, 结果见表 3. 图 2 给出了 UIUC-Sport 图像库的混淆矩阵.

表 2 Scene 15 数据集上的分类性能

Tab. 2 Classification performance on the Scene 15 dataset

方法	平均分类正确率/%
KSPM ^[4]	81.40±0.50
ScSPM ^[7]	80.28±0.93
Liu ^[16]	82.70±0.39
KSRSPM ^[15]	83.68±0.61
HIK+OCSVM ^[17]	84.12±0.52
LScSPM ^[8]	89.75±0.50
KLScSPM	90.49±0.62

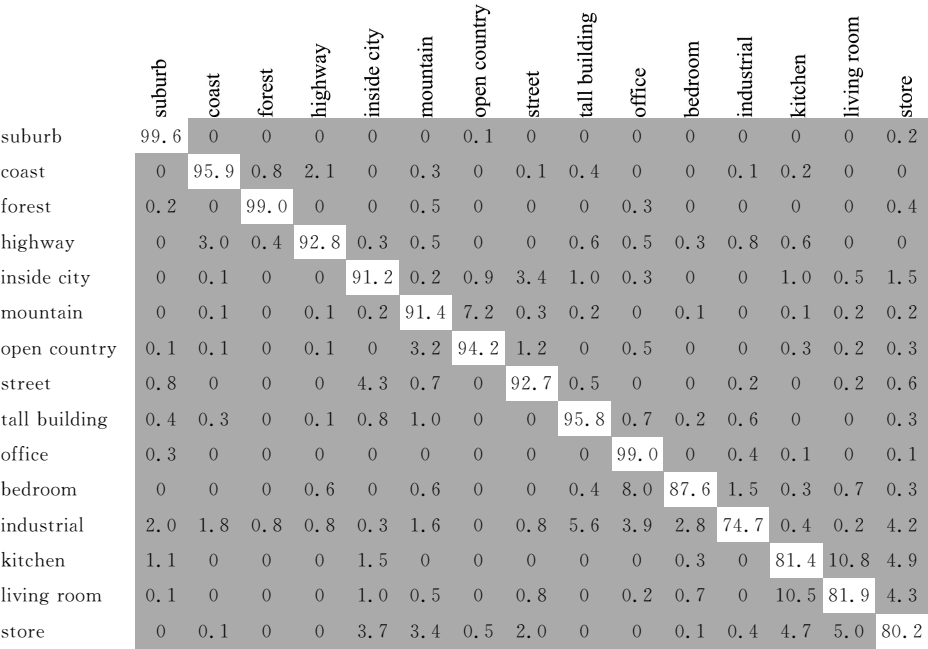


图 1 Scene 15 数据集的混淆矩阵(%)

Fig. 1 Confusion matrix on the Scene 15 dataset (%)

表 3 UIUC-Sport 数据集上的分类性能

Tab. 3 Classification performance on the UIUC-Sport dataset

方法	平均分类正确率/%
Liu ^[16]	82.29±1.84
ScSPM ^[7]	82.74±1.46
HIK+OCSVM ^[17]	83.54±1.13
KSRSPM ^[15]	84.92±0.78
LScSPM ^[8]	85.31±0.51
KLScSPM	86.32±0.57

2.5 平均量化误差

平均量化误差 $AverQE$ 是用来衡量特征量化的一个重要指标, $AverQE = \frac{1}{N} \sum_{i=1}^N \| \mathbf{x}_i - \mathbf{U} \mathbf{v}_i \|_F^2$. 小的量化误差意味着在特征量化过程中较少的信息丢失, 这有利于提高分类性能. 本文计算了 KLScSPM、LScSPM、KSRSPM 和 ScSPM 4 种方

法的平均量化误差来进行比较,如表 4 所示.从中可以看出本文的方法进一步降低了平均量化误差,这也是 KLS_{Sc}SPM 分类效果优于 LScSPM 和 KSRSPM 的原因.

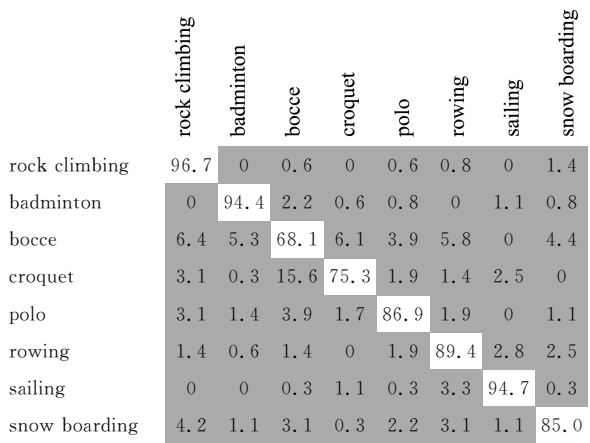


图 2 UIUC-Sport 数据集的混淆矩阵(%)
Fig. 2 Confusion matrix on the UIUC-Sport dataset (%)

表 4 平均量化误差

Tab. 4 Average quantization error

方法	平均量化误差		
	Caltech 256	Scene 15	UIUC-Sport
ScSPM ^[7]	0.916 4	0.868 1	0.886 4
LScSPM ^[8]	0.116 1	0.180 2	0.124 5
KSRSPM ^[15]	0.057 2	0.096 3	0.094 0
KLS _{Sc} SPM	0.050 1	0.064 7	0.062 6

3 结 语

本文提出了一个新的编码方法——核拉普拉斯稀疏编码,即把核方法应用到拉普拉斯稀疏编码中,隐式地把特征映射到高维空间,并在高维空间进行稀疏编码.在 Caltech 256、Scene 15 和 UIUC-Sport 数据库上的实验表明,KLS_{Sc}SPM 方法达到了良好的分类效果.

参考文献:

[1] Joachims T. Text categorization with support vector machines: learning with many relevant features [C] // **Machine Learning: ECML-98**. Berlin: Springer Berlin Heidelberg, 1998:137-142.

[2] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification [C] //

26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR. Piscataway: IEEE, 2008.

[3] Jurie F, Triggs B. Creating efficient codebooks for visual recognition [C] // **10th IEEE International Conference on Computer Vision, 2005. ICCV 2005 (Volume:1)**. Piscataway:IEEE, 2005:604-610.

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

[5] Bosch A, Zisserman A, Muñoz X. Image classification using random forests and ferns [C] // **2007 IEEE 11th International Conference on Computer Vision, ICCV**. Piscataway:IEEE, 2007.

[6] Maji S, Berg A C, Malik J. Classification using intersection kernel support vector machine is efficient [C] // **26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR**. Piscataway: IEEE, 2008.

[7] YANG Jian-chao, YU Kai, GONG Yi-hong, *et al.* Linear spatial pyramid matching using sparse coding for image classification [C] // **2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009**. Piscataway:IEEE, 2009:1794-1801.

[8] GAO Sheng-hua, Tsang I, Chia L, *et al.* Local features are not lonely—Laplacian sparse coding for image classification [C] // **2010 IEEE Conference on Computer Vision and Pattern Recognition, CVPR**. Piscataway:IEEE, 2010:3555-3561.

[9] Kavukcuoglu K, Ranzato M, Fergus R, *et al.* Learning invariant features through topographic filter maps [C] // **2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009**. Piscataway:IEEE, 2009:1605-1612.

[10] Mairal J, Bach F, Ponce J, *et al.* Non-local sparse models for image restoration [C] // **2009 IEEE 12th International Conference on Computer Vision**.

- Piscataway:IEEE, 2009;2272-2279.
- [11] Scholkopf B, Smola A J, Muller K-R. Kernel principal component analysis [J]. **Lecture Notes in Computer Science**, 1997, **1327**:583.
- [12] Bo L, Sminchisescu C. Efficient match kernels between sets of features for visual recognition [C] // **Advances in Neural Information Processing Systems 22 — Proceedings of the 2009 Conference**. New York:Curran Associates Inc. , 2009:135-143.
- [13] Lee H, Battle A, Raina R. Efficient sparse coding algorithms [C] // **Advances in Neural Information Processing Systems 19 — Proceedings of the 2006 Conference**. Canada: Neural Information Processing System Foundation, 2007:801-808.
- [14] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. **International Journal of Computer Vision**, 2004, **60**(2):91-110.
- [15] GAO Sheng-hua, Tsang I W-H, Chia L-T. Kernel sparse representation for image classification and face recognition [C] // **Computer Vision, ECCV 2010 — 11th European Conference on Computer Vision, Proceedings**. Heidelberg: Springer Verlag, 2010:1-14.
- [16] LIU Ling-qiao, WANG Lei, LIU Xin-wang. In defense of soft-assignment coding [C] // **2011 International Conference on Computer Vision, ICCV 2011**. Piscataway:IEEE, 2011:2486-2493.
- [17] WU Jian-xin, Rehg J M. Beyond the Euclidean distance:Creating effective visual codebooks using the histogram intersection kernel [C] // **2009 IEEE 12th International Conference on Computer Vision, ICCV 2009**. Piscataway:IEEE, 2009:630-637.

Kernel Laplacian sparse coding for image classification

ZHANG Li-he*, PAN Lei, LIU Tao, MA Chen

(School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: Sparse coding can achieve good performance in some computer vision problems. However, past sparse coding was implemented in the original feature space. Kernel method can acquire high dimensional nonlinear mapping characteristics. Inspired by it, the Laplacian sparse coding (LSc) is extended, and the kernel Laplacian sparse coding (KLSc) is proposed. It can reduce the feature quantization error and enhance the sparse coding performance. Experimental results of three standard datasets show that the proposed image classification algorithm based on KLSc has good classification effect, and the correct classification rate is better than that of LSc method.

Key words: image classification; sparse coding; Laplacian sparse coding; kernel method; spatial pyramid matching (SPM)