

拉普拉斯稀疏编码的图像检索算法

王瑞霞 彭国华 郑红婵

(西北工业大学理学院 西安 710129)

摘要 由于稀疏编码中过完备的码本和独立的编码处理,在编码过程中图像块相似性信息被丢失。为了保留块与块之间的相似性信息,提出了拉普拉斯稀疏编码的图像检索算法。给定编码特征的初始稀疏码,计算拉普拉斯矩阵,将相似性合并到稀疏目标函数,结合特征符号搜索算法和黄金分割线搜索算法,逐个更新每个稀疏编码系数。实验表明,拉普拉斯稀疏编码增强了稀疏编码的鲁棒性,与 SPM 模型算法相比,拉普拉斯稀疏编码的图像检索算法有较高的准确率。

关键词 稀疏编码,图像检索,码本,拉普拉斯矩阵,相似度矩阵

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.08.058

Image Retrieval Algorithm Based on Laplacian Sparse Coding

WANG Rui-xia PENG Guo-hua ZHENG Hong-chan

(School of Natural and Applied Sciences, Northwestern Polytechnical University, Xi'an 710129, China)

Abstract Due to the overcomplete codebook and the independent coding processing, the similarity of the image is lost between block and block to be encoded. To preserve such similarity information, we proposed the image retrieval algorithm based on Laplacian sparse coding. Given initial sparse coding and calculating the Laplacian matrix, similarity preserving term was incorporated into the objective of sparse coding. We used the feature-sign search algorithm and the golden section line search algorithm to update one by one each coefficient of sparse coding. The experiments show that Laplacian sparse coding can enhance the robustness of sparse coding. Compared with the improved SPM model, the new image retrieval algorithm better improves the retrieval accuracy.

Keywords Sparse coding, Image retrieval, Codebook, Laplacian matrix, Similarity matrix

1 引言

稀疏编码是一种新兴的信号处理技术,由于它可理解的理论知识引起了越来越多的研究人员的兴趣^[1,2]。在许多信号处理的应用中(如信号重构、复原,信号分解,信号去噪等等),它都有优越的性能。在计算机视觉领域,信号研究通常是图像特征或者是图像实例,稀疏编码通常被作为一个有效的技术如特征和图像的重建和相似性度量。除此之外,稀疏编码也被成功应用在许多计算机视觉,如图像分类、人脸识别、图像复原、图像标注等等。

BOVW 用图像的局部特征直方图表示图像^[3],对特征的空间平移不变具有稳定性,但它只用视觉词典中的一个视觉词表示一个图像块,不仅损失了图像块本身的多层语义信息,在特征直方图的量化过程中也损失了图像块与块之间的互信息,忽略了特征的空间排列信息,无法捕获目标的形状和位置。其中,特征量化过程可以看作是分配每个特征到这些码本,有硬分配和软分配两种。硬分配是每个特征只有一个码词,会引起严重的信息缺失,尤其当某个特征和几个码词都比

较相近时是非常不适合的。软分配由 J. Philbin 等^[4]提出,每个视觉特征分配几个最近邻的码词,保留了图像的局部信息。然而,对每个码词分配权重和为每个视觉特征分配码词数并不是一件容易的事。

为了克服特征量化过程损失图像的局部信息的缺点,S. Lazebnik 等^[5]通过研究局部区域的空间信息,扩充 BOW 模型到空间金字塔匹配(SPM)模型。Jianchao Yang^[6]使用稀疏编码对 SPM 模型进行了扩展,应用到图像分类中表现出极好的性能。给定一个码本和编码的特征,稀疏编码的目的是从码本中选择尽可能少的基函数来线性地重构这个特征,同时要保证重构误差尽可能的小,通常码本是过完备的,这就意味着基数比编码特征的维数大得多,码本的过完备特性保证了重构系数的稀疏性。但是,码本的过完备也引起了编码特征的局部信息丢失,在这样的过完备码本下相似特征可能被编码成完全不同的稀疏码,从而导致稀疏编码的不稳定性。

在图像检索过程中,稀疏编码不稳定,直接影响检索结果的准确度。为了提高稀疏编码的鲁棒性,我们提出了基于拉普拉斯稀疏编码的图像检索算法,通过为稀疏编码的公式增

到稿日期:2013-09-24 返修日期:2014-01-31 本文受国家自然科学基金(61070233)资助。

王瑞霞(1984-),女,博士生,主要研究方向为基于内容的图像检索,E-mail:wangruixia921@163.com;彭国华(1962-),男,教授,博士生导师,主要研究方向为计算机图形学、计算机辅助几何处理、图像处理、计算机视觉;郑红婵(1971-),女,教授,主要研究方向为计算机辅助几何设计、计算机图形学、计算机图形图像处理。

加一个额外的局部保留项,该算法能够训练更多有识别力的稀疏码,并且保留稀疏编码过程中特征的局部性和相似性。将我们之前所得到的编码特征的稀疏码作为初始码,定义拉普拉斯矩阵,将特征与特征之间的相似性合并到稀疏目标函数,并结合特征符号搜索算法和黄金分割线搜索算法,逐个更新每个特征的稀疏编码系数,利用最大池函数最终得到整个图像数据库的稀疏码,利用所得的稀疏码库对输入的查询图像进行检索。在相似度计算上,每一种相似度计算都有其优缺点,因此本文将多种相似度算法组合。实验证明,本文提出的图像检索算法不仅提高了稀疏编码的鲁棒性,并且与改进的SPM模型下的稀疏编码检索算法相比,其检索的有效率得到一定的提高。

2 基础知识

2.1 稀疏编码公式

设 $X=[X_1, X_2, \dots, X_N] \in R^{d \times N}$ 是一幅图像的特征, X_i 是局部特征描述子, $B=[B_1, B_2, \dots, B_M] \in R^{d \times M}$ 是码本,由图像库中所有图像的局部描述子通过聚类得到的 M 个聚类中心, B_i 是码本中的第 i 个码词。每个局部描述子 X_i 通过码本 B 进行编码生成一个 M 维的向量,用于生成图像的最终表示。

稀疏编码的公式定义如下:

$$\min_{B, C} \sum_{i=1}^N \|X_i - BC_i\|^2 + \lambda \|C_i\| \quad (1)$$

满足条件: $\|B_i\| \leq 1, \forall i=1, 2, \dots, M^{[6]}$ 。

式(1)中的第一项是重构误差,第二项用来控制稀疏码 C_i 的稀疏性, λ 是用来平衡稀疏性和重构误差的权衡因子。

通常码本 B 是过完备的,这个特性保证了重构系数的稀疏性。但是,码本的过完备损失了编码特征的局部信息,使得相似的特征可能被编码成完全不同的稀疏码,从而导致稀疏编码的不稳定性^[7]。

2.2 拉普拉斯稀疏编码公式

在稀疏编码中,每个特征独立进行编码。正如上面提到的,由于过完备和过饱和的码本,相似特征可能会编码成完全不同的稀疏码,这可能引起编码特征局部信息的损失。为了保留这个局部信息,在稀疏编码目标函数中引入了局部信息项,定义拉普拉斯矩阵,建立了拉普拉斯稀疏编码公式^[8]。

设 $X=[X_1, X_2, \dots, X_N] \in R^{d \times N}$ 是一幅图像的特征, X_i 是局部特征描述子,相应于这些特征的相似矩阵记为 W , W 的元素 W_{ij} 是度量一个特征点对 (X_i, X_j) 的相似性。定义度矩阵 D 为一个对角矩阵, D_{ii} 表示所有与特征 X_i 相关的相似度的总和,即 $D_{ii} = \sum_{j=1}^N W_{ij}$ 。如果两个特征是相似的,这个特征对的编码也应该是相似的,相似特征对的稀疏码之间应该有一定的距离。因此,我们引入特征对之间相应稀疏码的距离和,并且以特征对之间的相似度作为它们的权重,构建新的稀疏编码公式。

定义如下^[8]:

$$\min_{C_1, C_2, \dots, C_n} \sum_{i=1}^N \|X_i - BC_i\|^2 + \lambda \sum_{i=1}^N \|C_i\|_1 + \frac{\beta}{2} \sum_{i, j=1}^N \|C_i - C_j\|^2 W_{ij} \quad (2)$$

其中, F 范数是向量的 2 范数。

定义拉普拉斯矩阵为 $L = D - W$, 式(2)可以改写成式

(3), 得到拉普拉斯稀疏编码公式如下^[8]:

$$\min_{C_i} \sum_{i=1}^N \|X_i - BC_i\|^2 + \lambda \sum_{i=1}^N \|C_i\|_1 + \beta r (CLC^T) \quad (3)$$

其中, $C = [C_1, C_2, \dots, C_N]$ 。

当码本 B 固定时,式(3)是凸的,有解析解。因此可以用它的最优化取代整个稀疏码矩阵 C 的最优化,我们逐个优化每个 C_i 直到整个 C 收敛。为了优化 C_i , 固定所有剩下的稀疏码 C_j , 我们重新写了关于 C_i 的目标函数式(3)^[8]:

$$\min_{C_i} J(C_i) + \lambda \|C_i\|_1, \forall i \quad (4)$$

其中:

$$J(C_i) = \|X_i - BC_i\|^2 + \beta (C_i^T (CL_i) + (CL_i)^T C_i - C_i^T L_{ii} C_i) \quad (5)$$

L_i 是 L 的第 i 列,我们使用特征符号搜索算法^[7] 求解 C_i , 该算法能够快速以最小误差求解出最优的 C_i 。为了加快稀疏码的收敛,我们用空间金字塔稀疏编码进行初始化,一旦完成了 C_i 的最优化, C 将被相应地更新。根据我们的观察, C 仅迭代几次就会收敛。

定义矩阵 C_{-i} 是去掉矩阵 C 的第 i 列的子矩阵,定义向量 $L_{i,-i}$ 是去掉 L_i 第 i 项的子向量, I 是单位矩阵,一些定义如下^[8]:

$$\Gamma_{C_i} = \frac{\partial J(C_i)}{\partial C_i} = 2(B^T BC_i - B^T X_i + \beta C_{-i} L_{i,-i} + \beta L_{ii} C_i) \\ \Gamma_{C_i C_i} = \frac{\partial^2 J(C_i)}{\partial C_i^2} = 2(B^T B + \beta L_{ii} I) \quad (6)$$

2.3 特征符号搜索算法

特征符号搜索算法步骤:

(1) 输入码本 B , X 的第 i 个特征 X_i , 初始化稀疏码 C 以及 β, λ , 计算矩阵 L 。

(2) 初始化 $\theta = \text{sign}(C_i)$, $\text{active set} = \text{find}(C_i \neq 0)$ 。

(3) 从 C_i 的零系数中选出 Γ_{C_i} 绝对值的最大值,记 $s = \arg \max_r |\Gamma_{C_i}^r|$, 其中 $\Gamma_{C_i}^r$ 是 Γ_{C_i} 的第 r 个分量。

如果 $\Gamma_{C_i}^s > \lambda$, 则 $\theta_s = -1$, $\text{active set} = \{s\} \cup \text{active set}$

如果 $\Gamma_{C_i}^s < -\lambda$, 则 $\theta_s = 1$, $\text{active set} = \{s\} \cup \text{active set}$

其中, θ_s 是 θ 的第 s 个分量。

(4) 设 $\hat{B}, \hat{\Gamma}_{C_i C_i}$ 分别是 B 和 $\Gamma_{C_i C_i}$ 的子矩阵,由相应于 active set 的列向量组成,设 $\hat{C}_i, \hat{\theta}$ 是 C_i 和 θ 的子向量,由相应于 active set 的分量组成。计算产生的无约束二次最优化问题的解析解: $\min_{\hat{C}_i} J(\hat{C}_i) + \lambda \hat{\theta}^T \hat{C}_i$, 得到更新的 \hat{C}_i 。

$$\hat{C}_{i, \text{new}} = (\hat{\Gamma}_{C_i C_i})^{-1} (2 \hat{B}^T X_i - 2\beta (C_{-i} L_{i,-i})_{\text{active set}} - \lambda \hat{\theta}) \quad (7)$$

在一个封闭的线段上从 \hat{C}_i 到 $\hat{C}_{i, \text{new}}$ 利用黄金分割线搜索算法执行离散的直线搜索,对 $\hat{C}_{i, \text{new}}$ 和任意改变符号的系数的所有点计算式(4)的目标值,更新 \hat{C}_i 和 C_i 中相应的项,得到更新后的 C_i , 从 active set 中移除零系数并且更新 $\theta = \text{sign}(C_i)$ 。

(5) 检查最优化条件:

(a) 对非零系数的最优化条件: $\Gamma_{C_i}^r + \lambda \text{sign}(C_i^r) = 0, \forall C_i^r \neq 0$ 。若(a)不满足,返回(4);否则检查条件(b)。

(b) 对零系数的最优化条件: $|\Gamma_{C_i}^r| \leq \lambda, \forall C_i^r = 0$ 。若(b)不满足,返回(3);否则返回更新后的 C_i , 并更新到 C 中。

3 拉普拉斯稀疏编码的图像检索算法

在稀疏编码中,局部特征是被分开编码的,不考虑局部特征之间的相互影响,并且稀疏编码可能变化很大,即使对非常相似的特征也可能编码成完全不同的稀疏码,这将严重影响图像编码的质量,从而导致检索结果的准确率很低。本文通过引入拉普拉斯局部保留项到稀疏编码的目标函数中,得到拉普拉斯稀疏编码公式,不仅保留了图像的局部特征,使得相似特征有相似的编码,而且提高了图像编码的鲁棒性。

3.1 相似度矩阵的计算

对于相似度矩阵 W 的计算,我们使用 KNN 算法来构建(本文 $k=5$),采用余弦相似度公式计算两个特征 X_i 和 X_j 之间的相似性,通过测量两个向量内积空间夹角的余弦值来度量它们之间的相似性。在计算过程中,余弦相似度通常用于两个向量的夹角小于等于 90° ,因此余弦相似度的值为 0 到 1 之间。

3.2 稀疏编码的最大池函数

为了与上面内容一致,我们假设一幅图像有 N 个局部特征,码本是 M 个。利用最大池函数提取出显著特征点,每幅图像最终将表示成一个 M 维列向量:

$$y = [y_1, y_2, \dots, y_M]^T$$

其中, $y_i = \max(c_{i1}, c_{i2}, \dots, c_{iN})$, y_i 表示 y 的第 i 项, c_{ij} 是稀疏码矩阵 C 的第 i 行。

我们使用最大池函数表示图像,其优越性已经通过实验和理论得到证明^[9,10]。

3.3 参数的设定

在拉普拉斯稀疏编码公式中,最重要的两个参数是相似性约束权值 β 和控制稀疏码稀疏性权值 λ ,实验表明要获得更好的性能, β 取值范围在 $0.1 \sim 0.4$, λ 取值范围在 $0.2 \sim 0.5$ 。本文使用 Caltech101 数据库^[11] 进行检索,取 $\lambda = 0.4$, $\beta = 0.2$ ^[8]。

3.4 评价准则

图像检索一般采用查全率和查准率作为评价标准。查准率定义为检索出的图像中相关图像的数目占检索结果中所有图像数目的比例;查全率定义为检索出的相关图像的数目占数据库中所有相关的图像数目的比例^[12]。通常情况下,查准率高了,查全率就会下降;而固定查全率,就会相对降低查准率。因此,我们采用综合性能指标来判断检索算法的性能。

综合性能指标^[13]定义如下:

$$\text{综合性能指标} = \frac{2 * \text{查全率} * \text{查准率}}{\text{查全率} + \text{查准率}} \quad (8)$$

综合性能指标将查准率和查全率合并成一个指标,只有在查全率和查准率都提高时,综合性能指标才会提高。以降低其中一个指标为代价来提高另一个指标,反而会使综合性能指标值下降。因此,综合性能指标比单纯的查全率和查准率有更好的评价性能。

4 实验结果

本文使用 Caltech101 数据库,该数据库包含 101 类和 1 个背景类,一共是 102 类,9144 幅图片,图片大小不一。以由

查全率和查准率确定的综合性能指标作为评价准则,与我们之前所做的改进 SPM 稀疏编码图像检索算法比较。从实验结果可以看出,拉普拉斯稀疏编码算法比 SPM 算法有更好的检索结果。图 1 和表 1 中给出了部分实验结果。

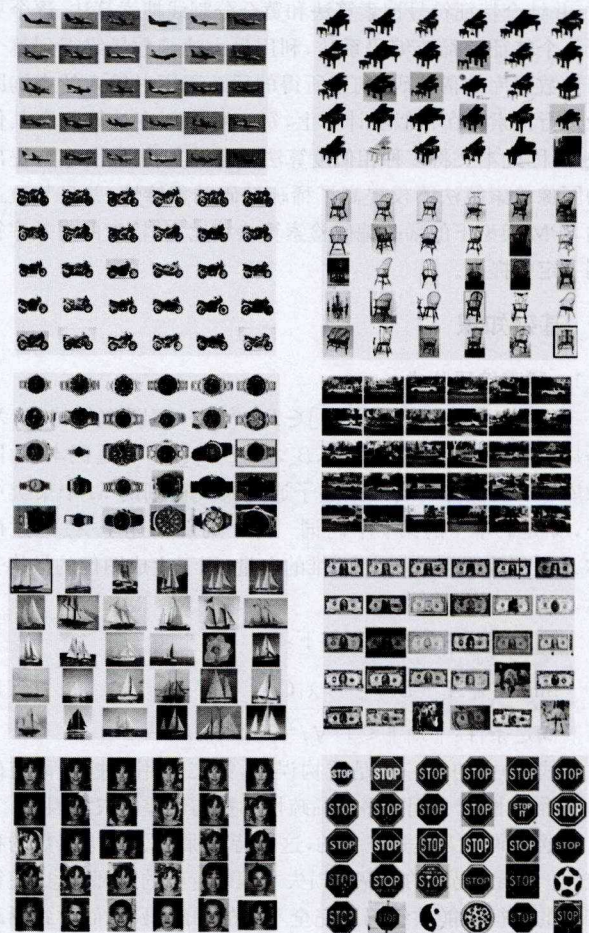


图 1 本文算法实验结果图

表 1 本文算法和改进 SPM 算法综合指标对比

	检索结果数	准确数	库中图片数	查准率 P	查全率 R	本文算法综合指标	SPM 算法综合指标
airplanes	58	47	52	81.03%	90.38%	0.8545	0.8208
grandpiano	108	86	99	79.63%	86.87%	0.8309	0.7609
motorbikes	563	481	798	85.44%	60.28%	0.7069	0.5795
windsorchair	53	44	56	83.02%	78.57%	0.8073	0.8167
watch	207	179	239	86.47%	74.90%	0.8027	0.7484
carside	119	103	123	86.55%	83.74%	0.8512	0.8088
schooner	69	58	63	84.06%	92.06%	0.8788	0.8346
dollarbill	55	49	52	89.09%	94.23%	0.9159	0.8514
Faces_easy	39	31	42	79.49%	73.81%	0.7654	0.7433
stop_sign	67	56	64	83.58%	87.50%	0.8550	0.8134

结束语 本文针对一般的稀疏编码算法在编码过程中损失了图像的相似性信息的缺点,提出了一种基于拉普拉斯矩阵的稀疏编码图像检索新算法。该算法将图像之间的相似性合并到稀疏编码的目标函数中,使得图像在进行编码的同时,也可以获得相似性信息,使相似的图像块具有相似的编码系数,并且采用特征符号搜索算法和线搜索算法寻找最优码。实验结果表明,该算法在图像检索上获得了较好的结果,可以有效地得到正确的检索图像。

(下转第 285 页)

- [6] Yang J, Yu K, Gong Y, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA; IEEE Computer Society, 2009; 1794-1801
- [7] Gao S, Tsang I, Chia L, et al. Local Features Are Not Lonely-Laplacian Sparse Coding for Image Classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA; IEEE Computer Society, 2010; 3555-3561
- [8] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA; IEEE Computer Society, 2010; 3360-3367
- [9] Yu K, Zhang T, Gong Y. Nonlinear Learning Using Local Coordinate Coding[C]//Proceedings of Advances in Neural Information Processing System, 2009. Vancouver, British Columbia, Canada; Springer, 2009
- [10] Liu Ling-qiao, Wang Lei, Liu Xin-wang. In Defense of Soft-assignment Coding[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011. Colorado Springs, CO, USA; IEEE Computer Society, 2011; 2486-2493
- [11] Huang Y, Huang K, Yu Y, et al. Salient Coding for Image Classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011. Colorado Springs, CO, USA; IEEE Computer Society, 2011; 1753-1760
- [12] Lazebnik S, Schmid C, Ponce J. Beyond bags of features; Spatial pyramid matching for recognizing natural scene categories[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006. New York, NY, USA; IEEE Computer Society, 2006; 2169-2178
- [13] Avila S, Thome N, Cord M, et al. Bossa: Extended Bow Formalism for Image Classification[C]//Proceedings of International Conference on Image Processing, 2011. Brussels, Belgium; IEEE Computer Society, 2011; 2909-2912
- [14] Harada T, Ushiku Y, Yamashita Y, et al. Discriminative Spatial Pyramid[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011. Colorado Springs, CO, USA; IEEE Computer Society, 2011; 1617-1624
- [15] Cao Yang, Wang Chang-hu, Li Zhi-wei, et al. Spatial-Bag-of-Features[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010. San Francisco, CA, USA; IEEE Computer Society, 2010; 3352-3359
- [16] Malinowski M, Fritz M. Learnable Pooling Regions for Image Classification[C]//Proceedings of CoRR, 2013
- [17] Jia Yang-qing, Huang Chang, Darrell T. Beyond Spatial Pyramids; Receptive Field Learning for Pooled Image Features[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012. Providence, RI, USA; IEEE Computer Society, 2012; 3370-3377
- [18] McCann S, Lowe D G. Local Naive Bayes Nearest Neighbor for image classification[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2012; 3650-3656
- [19] Lowe D. Distinctive Image Features from Scale-invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2); 91-110
- [20] Van De Sande, Gevers K E A T, Snoek C G M. Evaluating color descriptors for object and scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32; 1582-1596
- [21] Hui Bin, Tang Xu-sheng, Luo Hai-bo, et al. SDF Matched Filter Based on Gabor Wavelet Transform for Face Recognition[J]. Information and Control, 2008, 37(5); 633-636

(上接第 280 页)

参 考 文 献

- [1] Donoho D. For most large underdetermined systems of linear equations the minimal l^1 -norm solution is also the sparsest solution[J]. Comm. Pure and Applied Math, 2006, 59(6); 797-829
- [2] Donoho D. For most large underdetermined systems of linear equations, the minimal l^1 -norm near-solution approximates the sparsest near-solution [R]. The United States, California; Department of Statistics Stanford University, August 2004
- [3] Sivic J, zisserman A. Video google: a text retrieval approach to object matching in videos[C]//IEEE International Conference on Computer Vision (ICCV). 2003; 1470-1477
- [4] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching [C] // Proc. IEEE Conf. Computer Vision and Pattern Recognition. 2007
- [5] Lazebnik S, Schmid C, Ponce J. Beyond bags of features; spatial pyramid matching for recognizing natural scene categories[C]//Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2006
- [6] Yang Jian-chao, Yu Kai, Gong Yi-hong, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2009; 1794-1801
- [7] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[C]//Proc. Advances in Neural Information Processing Systems 19. 2006; 801-808
- [8] Gao Sheng-hua, Tsang I W-H, Chia L-T. Laplacian sparse coding, hypergraph laplacian sparse coding, and applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1); 92-104
- [9] Boureau Y-L, Ponce J, LeCun Y. A Theoretical analysis of feature pooling in visual recognition[C]//Proceedings of the 27'th International Conference on machine Learning. Haifa, Israel, 2010
- [10] Boureau Y-L, Bach F, LeCun Y, et al. Learning mid-level features for recognition [C] // IEEE Conf. Computer Vision and Pattern Recognition, 2010
- [11] L Fei-fei, Fergus R, Perona P. Learning generative visual models from few training examples; an incremental bayesian approach tested on 101 object categories[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop. 2004; 178
- [12] 王保平, 赵静, 苏建康, 等. 基于局部颜色-空间特征的图像检索方法研究[J]. 计算机应用与软件, 2012, 29(5); 28-31, 34
- [13] 兰冲. 基于统计规则的中文分词研究[D]. 西安: 西安电子科技大学, 2011