基于局部特征的图像分类方法

曹 健[1]，魏 星[2]，李海生[1]，蔡 强[1]

(1. 北京工商大学计算机与信息工程学院 北京 海淀区 100048；2. 北京科技大学计算机与通信工程学院 北京 海淀区 100083)

【摘要】为了有效地组织、管理和浏览大规模的图像资源，提出了一种利用局部特征进行图像分类的方法。通过深入分析和比较常见的局部特征，选用合适的局部特征构建视觉单词库。这些视觉单词具有很好的平移、旋转、尺度不变性，并对噪声有一定的抵抗能力。借鉴文本分类领域的向量空间模型进行图像的表示，并设计出了相应的分类算法。标准图像库上的实验结果表明，该方法在图像分类中有效，有较高的实用价值。

关 键 词 凝聚聚类； 分类器； 图像分类； 局部特征； 视觉单词

中图分类号 TP391.41 **文献标志码** A doi:10.3969/j.issn.1001-0548.2017.01.011

Image Classification Methods Based on Local Features

CAO Jian[1], WEI Xing[2], LI Hai-sheng[1], and CAI Qiang[1]

(1. School of Computer and Information Engineering, Beijing Technology and Business University Haidian Beijing 100048;
2. School of Computer and Communication Engineering, University of Science and Technology Beijing Haidian Beijing 100083)

Abstract In order to organize, manage and browse large-scale image databases effectively, an image classification algorithm based on local features is proposed. After analyzing of several fashionable local features at present, we choose the suitable features to construct the visual vocabulary. These visual words are invariant to image scale and rotation, and are shown robust to addition of noise and changes in 3D viewpoint. We also describe two approaches to represent objects using these visual words. As baselines for comparison, some additional classification systems also have been implemented. The performance analysis on the obtained experimental results demonstrates that the proposed methods are effective and highly valuable in practice.

Key words agglomerative clustering; classifier; image classification; local features; visual word

随着数字技术和通信技术的迅速发展，人们在网络上接触到越来越多的图像信息，需要通过计算机对图像进行自动分类处理。作为人工智能的重要应用领域，图像分类的主流方法是通过对选定的图像集(人工标注)进行学习，训练出合适的分类器，并利用分类器对未知图像进行分类决策[1]。

图像目标的特征提取是图像分类中的关键技术，对最终的分类效果具有决定性的影响。常见的一类方法是将图像目标作为一个整体，从大量正样本中学习并提取其整体特征，如面积、周长等，然后采用机器学习或者规则函数进行处理。这种方法有一些无法避免的缺点：1) 预处理和图像分割的好坏极易影响分类效果；2) 需要长时间的学习和大量已标注的图像；3) 由于没有专门捕捉图像目标的局部信息，当目标外观发生较大变化时，容易造成整体特征突变，进而导致分类方法失效。

心理学研究表明，人类的视觉系统可以将看到的场景进行分解，对这些小块的信息及其相互间的关系进行处理，从而分类识别。根据这一理论，在机器视觉领域出现了相对整体特征而言的局部特征，其含有的局部信息可以对图像目标进行多语义层次的描述。最近几年，许多研究者[2-5]不断提出了一些新的局部特征，并在大量的工程应用中验证了其性能相对优越、适用范围比较广泛。于是出现了尝试将局部特征技术用于图像拼接、图像检索和图像分类的文献[6-8]。

本文分析了常用的特征提取、图像处理和分类器技术，提出了一种有效地利用局部特征的图像分类方法。实验结果显示，该方法稳定性、查准率和查全率跟国内外前沿接近，甚至稍好。

收稿日期：2014－06－26；修回日期：2016－06－03

基金项目：国家自然科学基金(61402023)；北京市教委科研计划(SQKM201610011010)；北京市自然科学基金(4162019)；北京市科技计划(Z161100001616004)

作者简介：曹健(1982－)，男，博士，副教授，主要从事图像处理、机器学习、模式识别等方面的研究.

1 图像局部特征提取

1.1 底层局部特征提取

一般将底层局部特征的提取过程分为特征点检测和特征区域描述两步。常用的特征点检测算子有SUSAN检测算子、Harris-Laplace检测算子、Hessian-Laplace检测算子和DoG(difference of gaussian)检测算子[4]等。常用的特征区域描述子有SC(shape context)描述子、GH(geometric histogram)描述子、SIFT(scale invariant feature transform)描述子和GLOH(gradient location orientation histograms)描述子[5]等。

本文对这些特征区域检测算法以及描述子进行了研究分析，发现通过DoG检测算子和GLOH描述子提取的特征具有如下特性：1) 不变性，在图像大小、平移、旋转发生变化，甚至光照改变之后依然稳定；2) 区分性，产生大量包含丰富内容信息的特征，很适合图像分类；3) 高效性，特征提取速度比较快，并且已出现了一些优化的匹配算法。所以，选用DoG检测算子和GLOH描述子提取出的局部特征比较适合于对绝大多数图像内容的描述，进而可以实现准确、稳健的图像分类。

DoG算子用高斯函数作为卷积核，通过与相邻尺度以及相邻位置的点的对比，得到图像中极值点所处的位置和对应的尺度，并在这些候选点中筛选掉对比度较低以及处于边缘的特征点，提取出稳定的特征点，如图1所示。

a. 建筑与行人

b. 汽车与植物

图1 用DoG算子检测图像特征点示意图

GLOH描述子是对SIFT描述思想的改进和发展。通过主分量分析(principal component analysis, PCA)进行降维，最终得到一个128维的向量，在最大程度保留原始数据的同时减少了后续应用的计算时间。

1.2 构建视觉单词库

从众多图像中提取的底层局部特征规模非常大，而且都有或多或少的差别。如图2所示，针对同一种目标的相同部位提取的相似特征，也会存在一些细微的差别。这些“模板”描述得过于具体，虽然可以对某一个体进行精确匹配，但不适于对一类图像目标的分类识别。需要像自然语言中的单词一样，抓住一类事物的共性，即针对众多关于个体样本的“具体描述”进一步抽象出“概念”。

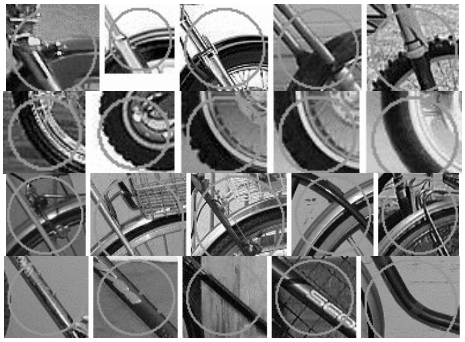

图2 同类目标的局部特征示意图

相近的局部特征经过优化组合之后可以形成“原型”特征，即视觉单词。大量的视觉单词就组成了视觉单词库，又称为视觉词表(visual vocabulary)或码书(codebook)[9-10]。国内一些研究人员将其应用于各种图像的处理和识别过程中，其中对局部特征进行聚类是构造视觉单词的一种有效的途径[11-12]，最常用的是划分方法中的*k*-平均值(*k*-means)算法、*k*-中心点(*k*-medoids)算法和层次方法中的凝聚聚类。划分方法虽然简单便捷，但是它经常以局部最优结束，而且必须实现给出簇的数目，对“噪声”和孤立点数据非常敏感。所以，本文选用RNN (reciprocal nearest neighbor)凝聚聚类算法[13]构建视觉单词库，这种方法先是将每个局部特征作为一个簇，然后将相似度最大的原子簇合并，直至达到某个希望的簇的数目。簇间相似度是通过计算平均相似度(一个簇中所有对象和另一簇所有对象之间的相似度的平均)得到：

$$\mathrm{sim}(C_1,C_2)=\frac{1}{|C_1||C_2|}\sum_{p_1\in C_1}\sum_{p_2\in C_2}\mathrm{sim}(p_1,p_2) \qquad (1)$$

相似度的度量采用的是欧式距离。用每个簇的重心

(簇的所有特征向量的平均值)来代表整个簇。

2 图像的表示与描述

2.1 基于CHI统计量的视觉单词筛选

χ^2 统计量(CHI)被引入信息论，用来衡量某个特征 t_i 和某个类别 C_j 之间的相关联程度，即 χ^2 统计值越高，表明该特征和这个类的相关性越大。本文利用CHI统计量来衡量视觉单词 t_i 和类别 C_j 的相关性，进而实现对视觉单词库的筛选优化。

如果令训练样本总数为N，数量为A的样本包含 t_i 且属于 C_j 类；数量为B的样本包含 t_i 但不属于 C_j 类；数量为C的样本不包含 t_i 但属于 C_j 类；数量为D的样本既不包含 t_i 也不属于 C_j 类。视觉单词 t_i 对 C_j 的CHI值为：

$$\chi^2(t_i,C_j)=\frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \tag{2}$$

如果是针对多分类的情况，筛选视觉单词可以通过两个途径实现：

1) 先计算出每个视觉单词 t_i 对于每个类别的统计值，然后求取所有训练样本中的最大值为：

$$\chi^2_{\mathrm{MAX}}(t_i)=\max_{j=1}^{m}\{\chi^2(t_i,C_j)\} \tag{3}$$

式中，m代表类别的数量。根据给定阈值筛除计算结果低的视觉单词，存留下来的特征作为最终使用的视觉单词库。

2) 计算出每个视觉单词 t_i 对于每个类别的平均值为：

$$\chi^2_{\mathrm{AVG}}(t_i)=\sum_{j=1}^{m}P(C_j)\chi^2(t_i,C_j) \tag{4}$$

然后用这个平均值来衡量该视觉单词与各个类别的相关程度。不过，实验表明求取平均值的效果不如求取最大值，所以本文暂时采用方法1)进行视觉单词的筛选。

2.2 图像的向量空间模型表示

“向量空间模型”(vector space model, VSM)，又称特征包模型或词袋模型，是在20世纪70年代初提出的，早期主要用在自然语言处理领域，尤其是信息检索和文本分类。2004年，该模型逐渐被引入到了图像识别领域。

对于向量空间模型，特征项是最小的不可分的语义单元。对于图像，它可以是任意分割程度上的子区域。所以，一幅图像可以认为是特征项(视觉单词)组成的集合，表示为 $\mathrm{Image}=I(t_1,t_2,\cdots,t_n)$，其中 t_k 是特征项，$1\leqslant k\leqslant n$。每一特征项(视觉单词) t_k 都有一个权重 w_k (依据一定的原则，如语义的重要程度)。

故根据特征项(视觉单词)及其对应的权重，可以将图像 I 可表示为：

$$I=I(t_1,w_1;t_2,w_2;\cdots;t_n,w_n) \tag{5}$$

简记为：

$$I=I(w_1,w_2,\cdots,w_n) \tag{6}$$

如果采用视觉单词作为向量空间模型中的特征项，图像表示的问题转化为求取每个视觉单词的权重问题。最为简单的是采用布尔权重(Boolean weighting)，其基本思想是，如果图像中出现过该视觉单词，那么该视觉单词的权重为1，否则为0。

由于布尔权重的表示方法没有体现视觉单词在图像中的作用程度，因而在实际应用中0、1值逐渐地被更精确的视觉单词的频率所代替，即是绝对词频(term frequency，TF)方法——使用视觉单词在图像中出现的频度 tf_{ij} 作为权重。

3 分类器设计

朴素贝叶斯分类器进行图像分类的基本思想是利用视觉单词和类别的联合概率，估计给定目标图像的类别概率。该模型假定特征向量的各个分量间对于决策变量是相对独立的，即目标图像是基于视觉单词的一元模型，当前视觉单词的出现依赖于图像类别但不依赖于其他视觉单词。

假设图像库中的训练样本分为m类，C_i 是类别标记，$1\leqslant i\leqslant m$。进行分类时，图像I被标记为 C_i，当且仅当：

$$P(C_i\mid I)>P(C_j\mid I)\qquad 1\leqslant j\leqslant m, i\neq j \tag{7}$$

根据概率理论中的贝叶斯公式可知，$P(A\mid B)=[P(A)P(B\mid A)]/P(B)$。应用该式，有：

$$P(C_i\mid I)=\frac{P(C_i)P(I\mid C_i)}{P(I)} \tag{8}$$

式中，$P(C_i)$ 为 C_i 类图像的出现概率，计算比较简单。如果训练集里的图像分为m类，而各个类别的样本数目相同，则 $P(C_i)$ 可以取 $1/m$。$P(I\mid C_i)$ 和 $P(I)$ 的计算根据视觉单词权重计算方法的不同，可以分为两种模型。

在多元伯努利模型(multi-variate Bernouli model)中，特征向量的每个分量采用布尔值，即一幅图像I的每个视觉单词采用布尔权重，故 $P(I\mid C_i)$ 和 $P(I)$ 的计算分别为：

$$P(I\mid C_i)=\prod_{t_k\in I}P(t_k\mid C_i) \tag{9}$$

$$P(I)=\sum_i\left[P(C_i)\prod_{t_k\in I}P(t_k\mid C_i)\right] \tag{10}$$

因此有：

$$P(C_i\mid I)=\frac{P(C_i)\prod_{t_k\in I}P(t_k\mid C_i)}{\sum_i\left[P(C_i)\prod_{t_k\in I}P(t_k\mid C_i)\right]} \tag{11}$$

式中，$P(t_k\mid C_i)$ 是对 C_i 类图像中视觉单词 t_k 出现的条件概率的拉普拉斯估计，有：

$$P(t_k\mid C_i)=\frac{1+N(t_k,C_i)}{M+N(C_i)} \tag{12}$$

式中，$N(t_k,C_i)$ 是训练集中含有特征 t_k 且属于 C_i 类的样本数；$N(C_i)$ 是训练集中 C_i 类样本的数目；M 是类别的数量。

在多项式模型(multinomial model)中，特征向量的每个分量采用绝对词频，即若一幅图像*I*的每个视觉单词采用其出现的频度为权重，则图像*I*属于 C_i 类的概率为：

$$P(C_i\mid I)=\frac{P(C_i)\prod_{t_k\in I}P(t_k\mid C_i)^{\mathrm{TF}(t_k,I)}}{\sum_i\left[P(C_i)\prod_{t_k\in I}P(t_k\mid C_i)^{\mathrm{TF}(t_k,I)}\right]} \tag{13}$$

式中，$\mathrm{TF}(t_k,I)$ 是图像*I*中视觉单词 t_k 出现的频度；$P(t_k\mid C_i)$ 是对 C_i 类图像中视觉单词 t_k 出现的条件概率的拉普拉斯估计，有：

$$P(t_k\mid C_i)=\frac{1+\mathrm{TF}(t_k,C_i)}{|I|+\sum_{t_k\in I}\mathrm{TF}(t_k,C_i)} \tag{14}$$

式中，$\mathrm{TF}(t_k,C_i)$ 是 C_i 类图像中视觉单词 t_k 出现的频度；$|I|$ 为特征分量的总数，为图像表示中所包含的不同视觉单词的总数目，即视觉单词库的规模。

4 实验与结果分析

4.1 实验环境和实验数据

为了验证本文提出的基于局部特征的图像分类方法的有效性，进行了相关实验。实验数据选自Caltech101图像库，该图像库是由加州理工学院的Li等创建，每类目标有40～800幅图像，大小约300×200像素。该图像库的优点在于：图像大小和目标相对位置大体相同，不需要花时间去裁剪图像就能进行实验；图像的杂乱或遮挡部分很少，分类算法可以依赖于目标图像的显著特征；对每幅图像都进行了注释，每个注释包括目标位置的边界盒以及人工描绘的目标轮廓两种信息。

实验选用8类图像分别统计在二分类问题上的实验结果。这8类图像目标分别为汽车、自行车、人、马、花卉、沙发、显示器和建筑物，挑选正负样本各100幅作为训练集样本，各25幅作为测试集样本，并挑选出40～100个正样本(已标注出目标轮廓)用以构造视觉单词库。训练集与测试集相互独立，即两者不含有同一幅图像。

4.2 性能评估方法

ROC(receiver operating characteristics)曲线线图的*Y*轴和*X*轴分别是评价指标TPr(true positive rate)和FPr(false positive rate)，其中，TPr和FPr的计算公式为：

$$\mathrm{TPr}=\frac{\mathrm{TP}}{\mathrm{TP+FN}},\quad \mathrm{FPr}=\frac{\mathrm{FP}}{\mathrm{FP+TN}} \tag{15}$$

ROC空间对样本在类别间的分布不敏感，基于该曲线的相等错误率(equal error rate, EER)即是选取 $\mathrm{TPr}=1-\mathrm{FPr}$ 时的值，可以直观反映分类算法的效果。式中，各个参数的含义如表1所示。

表1 图像分类算法输出结果

算法对二者关系的判断	图像与类别的实际关系	
	属于	不属于
标记为YES	TP	FP
标记为NO	FN	TN

根据表1可以得到查全率(Recall)和查准率(Precision)的计算公式为：

$$\mathrm{Recall}=\mathrm{TPr},\quad \mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{16}$$

对于图像分类，查准率和查全率是一对相互矛盾的物理量，提高查准率往往要牺牲一定的查全率，反之亦然。为更全面地反映分类算法的性能，本文采用RPC曲线图。

4.3 实验分析

为了验证利用聚类算法构造视觉单词这一途径的有效性，将凝聚聚类算法与划分方法中的*k*-平均值和*k*-中心点聚类应用于同一样本集，并比较最终的分类效果。该实验从60幅图片(汽车图像)中共提取出19 127个局部特征用以构造视觉单词库，对小汽车图像和建筑物图像进行分类测试，得到单词库规模为200～1 800之间的正确率，该评估指标是在相等错误率(EER)下的分类效果，对图像进行向量空间模型表示时用的是布尔权重。

如图3所示，由于凝聚聚类算法得到的簇相对紧致，总体比划分方法中的两种聚类算法性能好。关于视觉单词库的规模，在200～800之间随着视觉单词数量的增加，分类效果得到了明显的改善，在800

以上RNN凝聚聚类算法相对稳定，k-平均值和k-中心点方法则会出现波动，这是因为划分方法经常以局部最优结束。

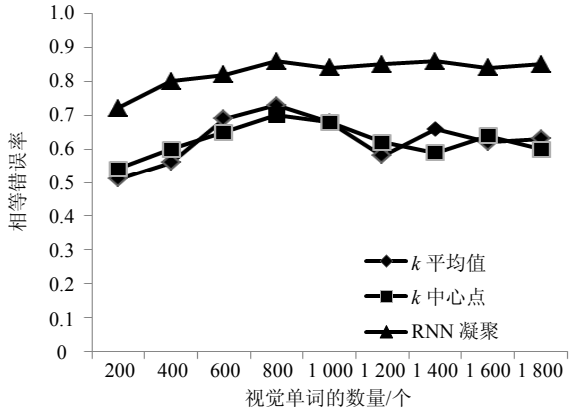

图3　不同视觉单词库构造方法的性能

采用不同的特征权重类型对分类的最终效果也会有较大的影响，本文对布尔、绝对词频(TF)两种特征权重计算方法进行实验对比。采用朴素贝叶斯分类器对8类图像分别进行二分类实验，求取每次分类的查准率和查全率。由于样本在所有类别中分布均匀，计算出的宏平均查准率和查全率等于微平均查准率和查全率。如图4的RPC曲线所示，TF权重较布尔权重效果好。这是由于用0、1代表该视觉单词是否在图像中出现，无法体现视觉单词在图像分类中的作用程度，因而效果不如更精确的TF方法。

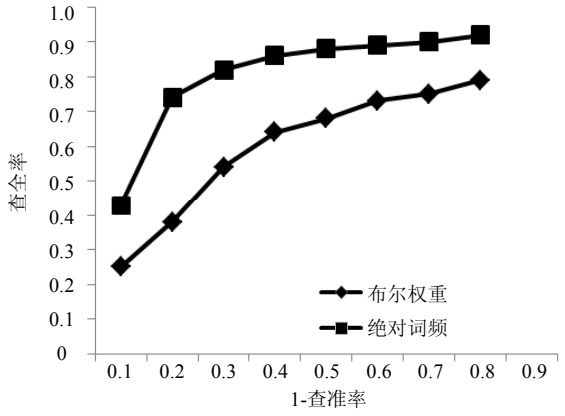

图4　采用不同特征权重的分类效果

为了更为直观地比较本文的方法和其他相近算法[12~14]的性能差异，本文进行一些仿真实验对比。为了与相关算法保持一致，实验所用的摩托车和小汽车(背面视图)两类图像都选自Caltech图像库，算法的正确率是在相等错误率(EER)时计算所得的。从表2可以看出，与其他算法的最佳效果相比，本文的方法在没有经过特征选择优化之前，性能是中等偏上的；经过基于CHI统计量的视觉单词筛选后，性能得到了一定的提升，仅比其他算法中的一种稍逊。总体来看正确率较高，一方面证明了其可行性，另一方面也说明了在特征优化上可能还有较大的改进空间。

表2　与相近算法的实验对比

数据集种类	摩托车/%	小汽车背面/%
Fergus算法	93.3	90.3
Song算法	—	94.2
Bilen算法	98.5	98.3
Leibe算法	94.0	93.9
本文算法 (特征优化之前)	93.7	94.0
本文算法 (特征优化之后)	94.8	95.3

5　结 束 语

本文提出了一种基于局部特征的图像分类方法，通过凝聚聚类将大量的局部特征进一步构造成视觉单词，并引入信息论中的CHI统计量进行视觉单词的筛选。同时，借鉴文本分类领域的向量空间模型进行图像表示，不仅能够描述出图像的关键内容，而且具有很好的平移、旋转、尺度、亮度不变性。根据视觉单词权重的不同计算方法，本文设计出了相应的朴素贝叶斯分类器，并在标准图像库上进行了实验分析，结果证明了该方法的有效性和健壮性。

本文的方法在视觉单词库的构造过程中，没有考虑视觉单词之间的空间关系，使用的局部特征类型也比较单一。下一步工作将考虑建立局部特征的空间关系模型，并将多种局部特征甚至整体特征结合起来，相信会有更好的实用价值和发展前景。

参 考 文 献

[1] CAO Jian, MAO Dian-hui, CAI Qiang, et al. A review of object representation based on local features[J]. Journal of Zhejiang University-Science C (Computers & Electronics), 2013, 14 (7): 495-504.

[2] 葛娟, 曹伟国, 周炜, 等. 一种颜色仿射变换下的局部特征描述子[J]. 计算机辅助设计与图形学学报, 2013, 25(1): 26-33.
GE Juan, CAO Wei-guo, ZHOU Wei, et al. A local feature descriptor under color affine transformation[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(1): 26-33.

[3] 葛琦, 韦志辉, 肖亮, 等. 基于局部特征的自适应快速图像分割模型[J]. 计算机研究与发展, 2013, 50(4): 815-822.
GE Qi, WEI Zhi-hui, XIAO Liang, et al. Adaptive fast image segmentation model based on local feature[J]. Journal of Computer Research and Development, 2013, 50(4):

815-822.

[4] LOWE D. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[5] MIKOLAJCZYK K, SCHMID C. A performance evaluation of local descriptors[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.

[6] LI L J, SU H, LIM Y, et al. Object bank: an object-level image representation for high-level visual recognition[J]. International Journal of Computer Vision, 2014, 107(1): 20-39.

[7] 丁建睿, 黄剑华, 刘家锋, 等. 局部特征与多示例学习结合的超声图像分类方法[J]. 自动化学报, 2013, 39(6): 861-867.
DING Jian-rui, HUANG Jian-hua, LIU Jia-feng, et al. Combining local features and multi-instance learning for ultrasound image classification[J]. Acta Automatica Sinica, 2013, 39(6): 861-867.

[8] CHOI J Y, RO Y M, PLATANIOTIS K N. Color local texture features for color face recognition[J]. IEEE Transactions on Image Processing, 2012, 21(3): 1366-1380.

[9] NAKAYAMA H, HARADA T, KUNIYOSHI Y. Global Gaussian approach for scene categorization using information geometry[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2010: 2336-2343.

[10] REDONDO C C, LOPEZ S R, ACEVEDO R J, et al. SURFing the point clouds: Selective 3D spatial pyramids for category-level object recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2012: 3458-3465.

[11] 韩冰, 杨辰, 高新波. 融合显著信息的 LDA 极光图像分类[J]. 软件学报, 2013, 24(11): 2758-2766.
HAN Bing, YANG Chen, GAO Xin-bo. Aurora image classification based on lda combining with saliency information[J]. Journal of Software, 2013, 24(11): 2758-2766.

[12] 宋相法, 焦李成. 基于稀疏编码和集成学习的多示例多标记图像分类方法[J]. 电子与信息学报, 2013, 35(3): 622-626.
SONG Xiang-fa, JIAO Li-cheng. A multi-instance multi-label image classification method based on sparse coding and ensemble learning[J]. Journal of Electronics & Information Technology, 2013, 35(3): 622-626.

[13] LEIBE B, LEONARDIS A, SCHIELE B. Robust object detection with interleaved categorization and segmentation [J]. International Journal of Computer Vision, 2008, 77(1-3): 259-289.

[14] BILEN H, NAMBOODIRI V P, VAN GOOL L J. Classification with global, local and shared features[C]// Pattern Recognition (Lecture Notes in Computer Science). Berlin: Springer, 2012.

编 辑 黄 莘