鲁棒视觉词汇本的自适应构造与自然场景分类应用

杨　丹[①]　　李　博[②]　　赵　红[①]
[①](重庆大学软件学院　重庆　400030)
[②](重庆大学计算机学院　重庆　400030)

摘　要：该文提出了一种视觉词汇本的优化构造策略。首先引入条件数定量评估海量低层特征的稳定性，排除病态特征，筛选稳定的鲁棒视觉特征；通过分析聚类和降维的内在联系，构造了具有聚类结构的视觉特征自适应降维算法；进而利用低维聚类结构信息中的邻域支持度，自适应选取最佳的初始视觉词汇，同时选择 Sil 指标作为目标函数，从而改进流行的 LBG 词汇本生成算法敏感于初始点的随机选取，并只能得到局部最优等不足。新的视觉词汇本生成算法具有聚类和降维的统一计算功能、良好的鲁棒性和自适应优化等特性。基于概率潜在语义分析技术将该文的视觉词汇本应用于自然场景分类，在 13 类场景图像库上取得了 73.46%的平均分类率。

关键词：模式识别；自然场景分类；视觉词汇本；条件数；
中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1009-5896(2010)09-2139-06
DOI: 10.3724/SP.J.1146.2009.01323

An Adaptive Algorithm for Robust Visual Codebook Generation and Its Natural Scene Categorization Application

Yang Dan[①]　　Li Bo[②]　　Zhao Hong[①]
[①](*School of Software Engineering*, *Chongqing University*, *Chongqing* 400030, *China*)
[②](*College of Computer Science*, *Chongqing University*, *Chongqing* 400030, *China*)

Abstract: This paper describes a novel optimization framework for visual codebook generation. Firstly, the Condition Number (CN) is applied to evaluate the stability of initial visual features, and the well conditioned features are preserved by eliminating the bad ones. At the mean time, an adaptive algorithm to generate low-dimensional visual words is proposed by studying the relationship between clustering and dimension-reducing. In order to overcome the popular LBG codebook design algorithm suffers from local optimality and is sensitive to the initial solution, a parameter called neighborhood-support for each feature is calculated according to clustering structure, which is used to select initial visual words adaptively. Finally, the rational distortion function is redefined using Silhouette. Compared with traditional algorithm, the presented algorithm has excellent properties at simultaneous clustering and dimension reduction, good robustness and adaptive optimization. A good performance (73.46% classification rate) of application this method to 13-Scene classification is obtained by using Probabilistic Latent Semantic Analysis (PLSA).

Key words: Pattern recognition; Natural scene categorization; Visual codebook; Condition Number (CN)

1　引言

自然场景分类是机器视觉、模式识别、多媒体信息管理等领域的热点问题。传统方法通常是将色彩、纹理和形状等图像低层特征直接与监督学习方法结合的“先对象再场景”识别模式。近年来，为克服低层视觉特征与高层语义之间的“语义鸿沟”，避免识别过程大量的人工标注，基于中间层语义特征的场景建模方法得到了广泛关注。这类方法的核心在于中间层语义特征的定义、提取和描述。文本分析中的主题模型被成功应用于语义场景分类。该模型用视觉词袋(Bag-of-visual Words，BoW)的方式描述图像，再用概率潜在语义分析(Probabilistic Latent Semantic Analysis，PLSA)或潜在狄利克雷分布(Latent Dirichlet Allocation，LDA)等主题分析模型提取图像蕴含的主题，从而根据图像在主题空间的概率分布实现语义场景分类。BoW将图像投影到生成的视觉词汇本(visual codebook)中，为图像低维结构化描述、语义分析等提供了新的研究思路，在机器人导航[1]、Web图像搜索[2]、语义建模[3,4]、场景分类[5−8]等领域取得了良好的应用效果。

2009-10-12 收到，2010-04-30 改回
国家自然科学基金(60975015)，教育部博士点基金(20090191110023)和重庆市科技攻关项目(CSTC2009AC2057)资助课题
通信作者：李博　boli.cqu@gmail.com

制约图像 BoW 模型应用效果的关键在于如何构建高效的视觉词汇本。Sungho[9]，Yang[10]等研究了基于分类信息的视觉词汇本生成算法；Farquhar等[11]基于最大化后验概率构造了面向图像类的视觉词汇本；Moosmann 等[12]用随机森林聚类算法构造了具有良好判别力的视觉词汇本；Yu-gang 等[13]评估了影响视觉词汇本性能的因素，包括特征检测、词汇本大小、加权策略等。这些成果从不同方面提出了视觉词汇本的优化，但仍存在以下不足：(1)特征整合仍是在海量数据中进行，大量不稳定和噪声特征影响了词汇本的构造效率和表征性能；(2)现有的特征整合往往直接在高维空间中进行，计算复杂度高；(3)视觉词汇量化过程中，初始代表点往往是随机选取，使得算法对初始值较为敏感，降低了算法的性能和适用价值。

针对以上不足，本文根据视觉词汇本构造的“特征提取、表达与整合”3 个阶段分别提出了优化策略：(1)运用数值分析中的条件数 (Condition Number, CN) 理论定量评估低层特征的稳定性，筛选病态的干扰特征，获得良态的鲁棒视觉特征；(2)通过分析高维空间中的降维与特征整合的内在联系，提出了具有聚类功能的自适应局部线性嵌入算法，获得具有聚类结构信息的视觉特征的低维表示；(3)针对 LBG(Linde-Buzo-Gray)词汇本构造方法[14]只能得到局部最优和初始类代表点随机选取的不足，采用 Silhouette 指标改进原目标函数，然后，基于前一过程得到低维聚类结构信息，通过类内样本点提供的支持度自适应决定初始类中心，从而自适应生成视觉词汇本。最后采用 PLSA 技术[6]将本文的视觉词汇本应用于自然场景分类。

2 基于条件数的视觉特征稳定性评估

在视觉词汇本生成阶段，训练图像的海量性决定了特征的海量性，同时初始特征往往受到噪声、位移、光照等因素影响，包含了大量不稳定特征，现有的研究往往基于视觉的方式定性分析它们的可靠性。然而在很多情况下，这种方法无法克服由于噪声和图像的病态带来的误差影响。在矩阵论中，矩阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 内的元素 a_{ij} 往往带有误差 δ_{ij}，扰动矩阵 $\boldsymbol{\delta}=(\delta_{ij})_{n\times n}$ 的存在对计算 $\boldsymbol{A}$ 的特征值产生的影响程度是用条件数来刻画的，这种扰动正对应了图像处理中的噪声影响。因此条件数为定量分析图像噪声的影响程度提供了新途径。特征 x 对于变换 H 的条件数 C_H 定义为[15]

$$C_H(x)=\max_H \limsup_{\delta\to 0}\limits_{\|\eta\|\le\delta} \|\Delta\theta\|/\|\eta\| \tag{1}$$

η 为图像的噪声，$\Delta\theta$ 为噪声所引起的模型变换误差，||*||表示向量的 2-范数。实际应用中，条件数可由下式逼近：

$$C_H(x)=\left\|(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{M}\right\| \tag{2}$$

$\boldsymbol{A}$ 和 $\boldsymbol{M}$ 分别是 $N\times 2$ 和 $N\times N$ 的矩阵，$\boldsymbol{A}$ 中的元素由以点 x 为中心的一个矩形区域内的所有像素的水平和竖直两个方向上的梯度组成。权重矩阵 $\boldsymbol{M}$ 是一个对角阵，$\boldsymbol{M}_{ii}$ 的值对应 $\boldsymbol{A}$ 中第 i 行元素在以 x 为中心的高斯窗口中的加权系数。对任意的 $\boldsymbol{M}$，都有 $C_{\mathrm{Trans}}\le C_{\mathrm{RST}}\le C_{\mathrm{Affine}}$，其中 C_{Trans}，C_{RST} 和 C_{Affine} 分别表示平移变换、旋转+尺度+平移变换和仿射变换下的条件数。该式说明，如果图像特征对于平移变换模型来说是病态的，那么它对 RST 和仿射变换也是病态的[15]。基于此性质，只需采用计算简单的平移变换的条件数来定量分析特征的稳定性，如果特征点的条件数的下界大于给定阈值，则该点为病态点，予以排除。

通过条件数对初始特征提取过程中的各种病态影响的定量分析，删除不稳定特征，有效地克服因噪声、图像变化等因素引起的误差影响，降低海量数据的处理难度，从而提高了后续视觉词汇本生成的效率和鲁棒性。

3 具有聚类结构信息的低维视觉特征表示

为提高视觉特征的表征性能，往往对特征设计较高维数的描述子，如 128 维的 SIFT 描述子。然而训练样本的海量性加之特征描述向量的高维性，加剧了算法的计算负荷。同时，高维的特征向量中，存在弱相关、不相关、冗余的特征分量。有效进行维数约简，将为后续视觉特征整合提供简洁高效的训练样本。局部线性嵌入(Locally Linear Embedding, LLE)[16]基于局部线性假设，旨在保持原高维数据的局部几何结构条件下获得数据的低维表示。该算法的特征降维与特征聚类有着内在联系：基于特征分量不同的表征能力，聚类结果反映为同类特征在特征分量上具有大致相同的区分度，不同类特征在特征分量上存在较大差异的区分度。所以对聚类后的特征进行降维，提取最具有类间区分度的特征分量，变的更加合理高效；而 LLE 获得的低维特征，保留了高维数据原有的结构信息，根据区分度最好的特征分量能更有效的细化聚类结果。

LLE 算法对每个样本点都取 k 个近邻点构成局部重建区域，这样设定的局部区域往往不是对样本分布的最佳分割。本文提出了聚类与降维统一框架下的近邻参数自适应选择 LLE 算法(Adaptive-LLE，ALLE)，新的算法旨在根据数据的真实分布，为每个样本点设计最佳的近邻搜索空间，自适应选

取邻近点及其个数 k，从而准确构建权值重建矩阵。与 Jing 等[17]提出的邻域收缩和邻域膨胀的两阶段自适应局部线性嵌入相比，本文算法简单易行，具有良好的聚类功能，更适合于视觉词汇本的应用。

本文 ALLE 算法首先用仿射传播聚类(Affinity Propagation Clustering，APC)[18]获得高维数据的聚类结构。由 N 个样本特征构成的 $N\times N$ 的相似度矩阵为 $\boldsymbol{S}=(s(i,j))$，$s(i,k)=-\|x_i-x_k\|^2, i\neq k$。算法开始时将所有的特征都视为潜在的聚类中心，其作为类代表点的可能性由偏向参数 $p=s(k,k)$ 度量，可将所有样本点的 p 设置为相似度的均值。基于消息传递的思想，算法按式(3)–式(5)不断在数据间交换聚类信息：

$$r(i,k)=s(i,k)-\max_{k',k'\neq k}\{a(i,k')+s(i,k')\} \tag{3}$$

$$a(i,k)=\min\left\{0,r(k,k)+\sum_{i',i'\notin\{i,k\}}\max\{0,r(i',k')\}\right\},\quad i\neq k \tag{4}$$

$$a(k,k)=\sum_{i',i'\neq k}\max(0,r(i',k)) \tag{5}$$

为避免迭代中的数值振荡，引入阻尼因子 $\lambda,\lambda\in[0,1)$，实验中设置为 0.5。设当前迭代次数为 t，$r(i,k)$ 和 $a(i,k)$ 的更新结果由当前迭代值和上步迭代值加权得到：$r^{(t)}=(1-\lambda)\cdot r^{(t)}+\lambda\cdot r^{(t-1)}$，$a^{(t)}=(1-\lambda)\cdot a^{(t)}+\lambda\cdot a^{(t-1)}$。特征 k 搜集的证据越强(即 $r(i,k)$ 与 $a(i,k)$ 越大)，其作为最终聚类中心的可能性就越大。将各特征点 x_i 按 $x_k:\arg\max_k(a(i,k)+r(i,k))$ 分配给最近的类中心 x_k 所属的类，获得 m 个簇。

将聚类结果以及初始相似度矩阵作为 LLE 降维的输入，对每个特征仅在其所在的类中学习一个局部重建权值矩阵，产生最适合的重建矩阵。新算法在高维空间中的误差函数重新定义为

$$\min\varepsilon(\boldsymbol{W})=\sum_{i=1}^{N}\left|x_i-\sum_{j=1}^{|C_k|}w_{ij}x_{ij}\right|^2 \tag{6}$$

其中 $|C_k|$ 表示 x_i 所属的类的样本特征个数，这里将权重 $w_{ij}(i=1,2,\cdots,N)$ 存储在 $N\times N$ 的稀疏矩阵 $\boldsymbol{W}$ 中，当 x_j 属于 x_i 所在的类时，$\boldsymbol{W}_{ij}=w_{ij}$，否则，$\boldsymbol{W}_{ij}=0$。在低维空间，重构误差函数定义为

$$\min\varepsilon(\boldsymbol{Y})=\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{L}_{ij}y_i^{\mathrm{T}}y_j \tag{7}$$

损失矩阵 $\boldsymbol{L}=(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{W})$ 是一个 $N\times N$ 的对称矩阵，要使损失函数值达到最小，则取 $\boldsymbol{Y}$ 为 $\boldsymbol{L}$ 的最小 d 个非零特征值所对应的特征向量，通常将最小的第 $2\sim d+1$ 个特征值所对应的特征向量作为寻求的低维嵌入。

通过聚类和降维的统一计算过程，得到了具有聚类结构信息的低维视觉特征表示，这些聚类中心可视为后续视觉词汇本量化的初始词汇。

4　视觉词汇本的自适应生成算法 ALBG

在特征整合阶段，流行的 LBG 算法[14]首先从训练样本 $\boldsymbol{X}=\{x_n:n=1,2,\cdots,N\}$ 中随机选取 M 个初始点，产生初始词汇本 $\boldsymbol{V}=\{v_m:m=1,2,\cdots,M\}$，$M\ll N$，然后将训练样本分配到最近的视觉词汇 v_m 代表的第 m 个类中，$\forall x_n\in\boldsymbol{X}$：$\arg\min_m\{E:E=\|x_n-v_m\|^2\}$，计算所有类的平均离差 $D=\frac{1}{J\times M}\sum_{m=1}^{M}\sum_{j=1}^{J}E$，$J$ 是每个类的特征个数。迭代达到收敛条件则计算每个类的中心作为最终视觉词汇。LBG 算法产生词汇本具有两个固有局限：(1)局部最优：采用近邻准则搜索聚类结构和以类内的平均离差作为最优收敛准则，决定了 LBG 算法只能得到局部最优；(2)对初始类代表点选取敏感：LBG 算法需要随机选取初始聚类中心作为初始词汇本，迭代速度较慢，难以得到高质量的词汇本。

针对以上不足，本文提出自适应选择初始中心的视觉词汇本生成算法(Adaptive–LBG，ALBG)。首先采用 Sil 指标作为迭代目标函数，以达到全局最优的聚类效果，平均 Sil 指标越大表示聚类质量越好。用 $a(x)$ 表示类 $C_i,(i=1,2,\cdots,M)$ 中的样本 x 与类内其他样本的平均不相似度；$d(x,C_j)$ 表示 x 到另一个类 C_j 的所有样本的平均相似度，记 $b(x)=\min\{d(x,C_j)\}$, $x\in C_i, j=1,2\cdots M, j\neq i$，则：$\mathrm{Sil}(x)=(b(x)-a(x))/\max\{a(x),b(x)\}$。

然后利用 ALLE 得到的信息量 r 和 a，定义特征 x_i 的邻域支持度为：$NS(i)=r(i,i)+a(i,i)$，其值越大说明该点作为类代表点的适合程度越大。因此利用降维过程得到的邻域支持度，自适应选取 M 个支持度最大的特征作为初始聚类中心，通过迭代得到平均 Sil 值最优的聚类结构，计算每类的中心生成最终词汇本 $\boldsymbol{V}$。出现频率过高和过低的单词往往具有较低的信息量，可定义在样本集中出现频率高于和低于设定阈值的视觉词汇为停止词，也可将一定百分比的高频和低频词汇定义为停止词，从视觉词汇本中删除，进一步降低视觉词汇本维数。本文词汇本生成算法 ALBG 流程如图 1 所示。

5　自然场景分类实验与分析

为验证本文算法构造视觉词汇本的高效性，在13类场景图像库[15]进行分类实验。该图像库包括13类自然场景的3759张图片：C_1：Highway，C_2：Inside City，C_3：Tall Building，C_4：Street，C_5：Suburb

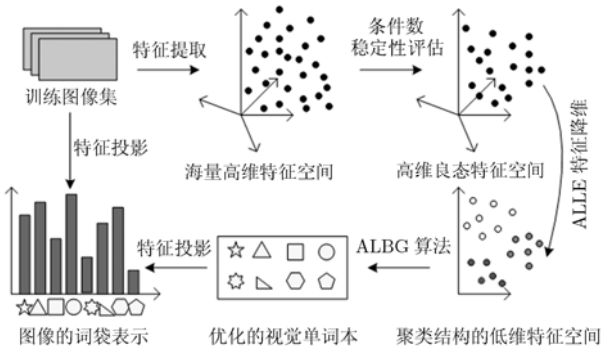


图1 本文词汇本生成算法ALBG流程图

Residence，C_6：Forest，C_7：Coast，C_8：Mountain，C_9：Open Country，C_{10}：Bedroom，C_{11}：Kitchen，C_{12}：Living room，C_{13}：Office。从每类场景图像中随机选择100幅图像作为训练集，剩余图像作为测试集。运用本文算法在训练图像集上构建视觉词汇本，并训练每个场景类的PLSA模型，最后用SVM分类器进行场景分类。

表1是用DOG算子检测出的每类场景图的平均初始特征点个数，与本文条件数筛选的稳定特征个数的对比。每类场景的平均特征个数等于在该类所有图像上提取的特征点个数总和，除以该类包含的图像数。通过本文条件数评估，剔除了大量病态不稳定特征，整个场景图像的平均特征点从1277个减少到539个，降低了初始特征的海量性，提高了样本特征的噪声鲁棒性和稳定性。

表1 条件数筛选后的平均稳定特征个数比较

场景类	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	All
初始特征	1226	1277	1436	1163	1524	1326	1125	1405	1212	1392	1076	1328	1113	1277
稳定特征	**623**	**658**	**504**	**499**	**647**	**721**	**356**	**819**	**430**	**576**	**308**	**537**	**328**	**539**

聚类表现是评估降维算法性能的一个重要准则。为验证本文ALLE算法的自适应性和降维效果，本文在ORL标准人脸库上进行降维实验，用Sil指标、类别数和正确聚类率量化评价聚类结果。从ORL人脸库中随机选取10个人，每个人取10张图像，首先对112×92的高维图像用不同参数下的LLE算法和本文的ALLE算法降维，然后进行聚类。从表2可以看出：原始LLE算法的降维效果受制于近邻参数 k 的选取，而本文的ALLE算法实现了参数的自适应选取，能够得到最优的降维效果，即最高的Sil值0.980和准确的聚类数10，以及最高正确聚类率80.83%。虽然用原始LLE算法在 $k=15$ 时能够得到最高的Sil值0.989，但是此时获得错误的聚类数7。

表 2 不同算法在 ORL 人脸库上的降维效果对比

不同算法	Sil值	类别数	分类率(%)
原始数据	0.838	10	100
LLE(k=8)	0.934	3	30
LLE(k=15)	0.989	7	49.17
LLE(k=30)	0.787	6	55.83
本文ALLE	**0.980**	**10**	**80.83**

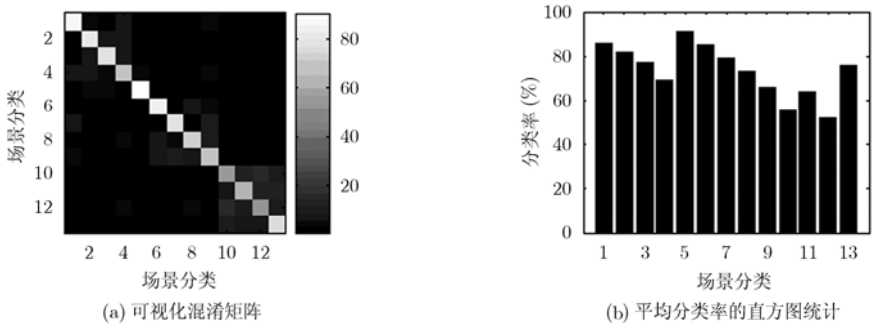
表3是将本文的视觉词汇本优化算法与原始算法在运行时间和应用效果方面的对比。在特征提取阶段，本文采用条件数筛选了大量病态特征，在特征降维阶段，本文的ALLE同时考虑了特征空间的聚类结构信息，得到了更加准确的低维特征描述，在视觉词汇的量化阶段，本文ALBG算法不依赖于初始值的随机选取，从而提高了单词本构造的效率(124813 s vs. 180627 s)，提高了视觉词汇本对分类图像的表征性能，在13场景类识别中获得了更高的分类准确率(73.5% vs. 69.3%)。

表 3 本文算法与原始算法的实现效率和分类性能对比

对比方面	原始算法	本文算法
特征选择	DOG	**DOG + CN**
维数约简	LLE	**ALLE**
词汇量化	LBG	**ALBG**
运行效率(s)	180,627	**124,813**
分类率(%)	69.3	**73.5**

表 4 给出了本文方法在 13-场景类上的分类混淆矩阵。混淆矩阵的第 i 行 j 列表示第 i 类图像被分为第 j 类图像的比例，对角线上元素的值代表了每类场景的分类准确率。图 2(a)是可视化的混淆矩阵，不同的灰度代表不同的分类率，灰度值越大代表的分类率越高。图 2(b)图是对 13 类场景的平均分类率的直方图统计，整体场景分类的平均准确率73.4615%即是各类场景的分类率的平均值。

表 5 比较了本文方法与文献[5–8]的 4 种代表性方法在 13 类场景图像上的特征维数、词汇本大小、



(a) 可视化混淆矩阵　　(b) 平均分类率的直方图统计

图 2　可视化混淆矩阵及场景平均分类率的直方图统计

表 4 13-类场景分类的混淆矩阵

分类率(%)	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}
C_1	**86**	5	0	6	0	0	0	0	3	0	0	0	0
C_2	0	**82**	8	7	0	0	0	0	0	1	1	1	0
C_3	0	7	**77**	8	2	0	0	0	1	2	0	2	1
C_4	8	8	4	**69**	4	0	0	0	3	2	0	2	0
C_5	0	3	3	0	**91**	2	0	0	1	0	0	0	0
C_6	0	2	0	0	2	**85**	0	6	5	0	0	0	0
C_7	8	0	0	0	0	0	**79**	2	11	0	0	0	0
C_8	2	0	2	3	0	7	4	**73**	9	0	0	0	0
C_9	4	0	0	2	0	8	10	8	**66**	2	0	0	0
C_{10}	1	0	2	2	0	0	0	0	0	**55**	13	17	10
C_{11}	0	2	0	1	0	0	0	0	0	7	**64**	12	14
C_{12}	0	1	2	3	2	0	0	3	0	17	11	**52**	9
C_{13}	0	0	1	0	0	0	0	0	0	10	6	7	**76**

表5 本文方法与其他方法的13类场景分类比较

对比算法	特征维数	词汇数	主题数	分类率(%)
Rasiwasia[5]	64	/	13	72.7
Bosch[6]	128	1500	25	73.4
Fei-Fei[7]	128	174	40	65.2
Lazebnik[8]	256	200	60	74.7
本文ALBG	**64**	**500**	**40**	**73.46**

潜在主题个数和平均分类率。实验中对整个图像库进行 10 次随机划分，生成相应的训练集和测试集图像，将 10 次划分得到的分类率的均值作为最终的平均分类准确率。综合考虑建模的准确性和实现的效率，本文通过实验将视觉词汇本的大小取为 500，潜在主题数目为 40，从而训练 PLSA 模型，最后用 SVM 分类器进行分类识别，得到高于 Fei-Fei et al.[7]的分类准确率(73.5% vs. 65.2%)，同时在最低的特征维数 64-dim 下，本文方法的分类率与其他方法接近，从而验证了本文算法构造的视觉词汇本的有效性。

6　结束语

已有研究结果表明 Bag-of-Words 是一种有效的图像表示方法，而视觉词汇本的构造、优化成为这一问题的关键环节。本文从特征稳定性筛选、特征低维描述、词汇本自适应生成方面提出了相应的优化策略。改进的视觉词汇本生成算法具有聚类和降维的统一计算功能、良好的鲁棒性和自适应优化等特性，通过充分的对比实验验证了本文算法的高效性，并在 13-类场景图像库上取得了 73.46%的平均分类率。在研究中发现，视觉词汇本的大小，以及潜在主题的个数对分类识别会产生一定影响，进一步分析这两个因素与分类结果间的数值关系，建立一种参数的自适应优化选择算法，将是提高视觉词汇本性能的又一研究方面。

参 考 文 献

[1] Cummins M and Newman P. FAB-MAP: Probabilistic localization and mapping in the space of appearance[J]. *The International Journal of Robotics Research*, 2008, 27(6): 647–665.

[2] Zhong W, Qifa K, Michael I, and Jian S. Bundling features for large-scale partial-duplicate web image search[C]. IEEE Conference on Computer Vision and Pattern Recognition, Miami, 2009: 25–32.

[3] 李志欣, 施智平, 李志清, 史忠植. 图像检索中语义映射方法综述[J]. 计算机辅助设计与图形学学报, 2008, 20(8): 1085–1096.
Li Zhi-xin, Shi Zhi-ping, Li Zhi-qing, and Shi Zhong-zhi. A survey of semantic mapping in image retrieval[J]. *Journal of Computer Aided Design & Computer Graphics*, 2008, 20(8): 1085–1096.

[4] 石跃祥, 朱东辉, 蔡自兴, Benhabib B. 图像语义特征的抽取方法及其应用[J]. 计算机工程, 2007, 33(19): 177–179.
Shi Yue-xiang, Zhu Dong-hui, Cai Zi-xing, and Benhabib B. Extraction of image semantic attributes and its application[J]. *Computer Engineering*, 2007, 33(19): 177–179.

[5] Rasiwasia N and Vasconcelos N. Scene classification with low-dimensional semantic spaces and weak supervision[C]. IEEE Conference on Computer Vision and Pattern Recognition, Alaska, 2008: 1–6.

[6] Bosch A, Zisserman A, and Munoz X. Scene classification via pLSA [C]. European Conference on Computer Vision, Austria, 2006: 517–530.

[7] Li Fei-fei and Perona P. A Bayesian hierarchical model for learning natural scene categories[C]. IEEE Conference on Computer Vision and Pattern Recognition, San Diego, 2005: 524–531.

[8] Lazebnik S, Schmid C, and Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]. IEEE Conference on Computer Vision and Pattern Recognition, New York, 2006, 2: 2169–2178.

[9] Kim S and Kweon I S. Simultaneous classification and visual word selection using entropy-based minimum description length[C]. IEEE International Conference of Pattern Recognition, Hong Kong, 2006: 650–653.

[10] Liu Yang, Rong Jin, Sukthankar R, and Jurie F. Unifying discriminative visual codebook generation with classifier training for object category recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, Alaska, 2008: 1–8.

[11] Farquhar J, Szedmak S, Meng H, and Taylor J S. Improving bags-of-keypoints image categorization[R]. Tech report, University of Southampton, 2005.

[12] Moosmann F, Triggs B, and Jurie F. Fast discriminative visual codebooks using randomized clustering forests[C]. In Neural Information Processing Systems, Vancouver, 2006: 985–992.

[13] Jiang Yu-gang, Chong-Wah N, and Yang Jun. Towards optimal bag-of-features for object categorization and semantic video retrieval[C]. ACM International Conference on Image and Video Retrieval, New York, 2007: 494–501.

[14] Linde Y, Buzo A, and Gray R M. An algorithm for vector quantizer design[J]. *IEEE Transactions on Communications*, 1980, 28(1): 84–95.

[15] Kenney C, Manjunath B S, and Zuliani M. A condition number for point matching with application to registration and post-registration error estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(11): 1437–1454.

[16] 马瑞，王家廙，宋亦旭. 基于局部线性嵌入(LLE)非线性降维的多流形学习[J]. 清华大学学报(自然科学版), 2008, 48(4): 582–585.
Ma Rui, Wang Jia-xin, and Song Yi-xu. Multi-manifold learning using locally linear embedding (LLE) nonlinear dimensionality reduction[J]. *Journal of Tsinghua University* (*Science and Technology*), 2008, 48(4): 582–585.

[17] Wang Jing, Zhang Zhen-yue, and Zha Hong-yuan. Adaptive manifold learning[C]. Advances in Neural Information Processing Systems, Cambridge, 2005: 1473–1480.

[18] Frey B J and Dueck D. Clustering by passing messages between data points[J]. *Science*, 2007, 315: 972–976.

杨 丹： 男，1962 年生，教授，博士生导师，研究方向为模式识别与人工智能、图像处理、计算机视觉与机器学习、数据挖掘、企业信息化与商务智能等.

李 博： 男，1982 年生，博士生，研究方向为模式识别、机器视觉、图像处理.

赵 红： 男，1986 年生，硕士生，研究方向为模式识别、机器视觉、图像处理.