基于低尺度词袋模型的图像快速分类方法

肖 哲，秦志光，丁 熠，蓝 天，于 跃

(电子科技大学信息与软件工程学院 成都 611731)

【摘要】提出一种新的框架用于改进传统词袋模型效率较低的问题。该方法建立在通过小波变换获取的低尺度图像表示上，利用在低尺度图像上提取单尺度的SIFT特征，建立低尺度视觉词典。由于大幅度减少了图像初始特征维数，该方法可以快速建立视觉词典，并且有效地降低后续图像分类所花费的时间。通过对Caltech101数据集全部8 677张图像的分类测试显示，该方法可以在保证分类性能的同时，有效地提升基于传统词袋模型的图像分类效率。实验结果表明，该方法可以全面提升金字塔匹配的词袋模型分类性能和分类效率，普遍用于传统词袋模型及其衍生方法。

关 键 词 词袋模型； 计算机视觉； 图像分类； 尺度不变特征转换； 小波变换

中图分类号 TP391.4 **文献标志码** A doi:10.3969/j.issn.1001-0548.2016.06.021

Efficient Method for Image Classification Based on Low-Scale Bag of Word Model

XIAO Zhe, QIN Zhi-guang, DING Yi, LAN Tian, and YU Yue

(School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 611731)

Abstract This paper proposes a new framework to improve the efficiency of visual bag-of-words model in large scale image classification. The method is based on the low scale image representation obtained by wavelet transform, and the low scale visual dictionary is built by extracting the SIFT features on the low scale image. Since the feature dimension is reduced, the method can quickly generate the visual dictionary and minimize the time of image classification process. The results of comparison experiments on the 8 677 images of Caltech 101 show that the proposed method can effectively improve the classification performance and efficiency of the traditional visual bag-of-words model and the Pyramid-BOW model.

Key words bag-of-words; computer vision; image classification; scale invariant feture transform; wavelet transform

近年来随着图像数量与日俱增，如何对海量的图像资料进行快速准确的检索、分类、识别，从中挖掘出用户所需的关键信息，逐渐成为计算机视觉领域的重要研究课题。随着越来越多的科研工作者致力于相关领域的研究，出现了新的图像分类方法，然而这些新方法大多只追求分类结果的准确率，并未考虑到实际应用中的高效性需求，往往在处理少量实验数据时效果优越，而当数据量增长到一定程度时就会出现效率低下、甚至难以运算的情况。

词袋模型(bag of word)最初被应用于文本分类领域[1]，文献[2-3]将其引入计算机视觉领域，并广泛应用于基于内容的图像分类中[4-7]。该方法通过对图像的视觉特征进行聚类获得视觉特征词典，利用视觉词典中的单词或词组在图像中出现的频率作为图像表示，进而对图像进行分类。视觉词袋模型(bag of visual word)的提出，一定程度上缓解了图像特征维数巨大、局部特征不统一难以训练的问题，但是在实际应用中，其分类效率仍有待进一步优化。文献[8]指出视觉词典所需解决的两个问题，一个是词典的简化，去除词典中没有区分力的无意义单词；另一个是需要确定一个合理的词典分辨率，往往分辨率越粗的词典分类准确性越差，而分辨率太细的词典又容易受到噪音干扰；文献[9-10]分别采用稀疏编码(sparse coding)和局部线性编码词典(locality-constrained linear coding，LLC)对特征进行量化，以尽可能简化视觉词典；文献[11]中对视觉词袋模型中

收稿日期：2015－07－20；修回日期：2015－11－25

基金项目：国家自然科学基金广东联合基金(U1401257)；国家自然科学基金青年基金(6130090)；四川省科技计划(2014JY0172)；中央高校基本科研业务费专项基金(ZYGX2013J080)；

作者简介：肖哲(1983－)，男，博士生，主要从事计算机视觉与模式识别、医学图像处理方面的研究.

的特征编码和池化方法进行了回顾和评估，并通过大量实验得出结论，在不同的应用中应使用不同的编码和池化方法。近年来，词袋模型的研究更多聚焦于解决实际应用中所遇到的各种问题，文献[12]提出一种关联直方图的词袋表示方法，通过将图像的全局直方图分解为目标及其领域的关联直方图来解决图像分类中的多目标问题；文献[13]提出了一种时空能量袋模型识别动态场景；文献[14]则着手于通过视觉词袋模型重建图像；而视觉词袋模型在医学图像分类中的应用也得到了越来越多的关注[15-18]。

尽管视觉词袋模型发展至今对图像分类性能有了质的提升，但是其计算效率仍难以达到海量图像快速分类的需求，有鉴于此，本文提出了一种基于小波变换的低维视觉词袋模型快速构建方法。该方法利用小波变换获得图像的低尺度表示，再通过均匀采样方式获得单一尺度下SIFT特征，以构建视觉词典。经过对Caltech 101数据集中101个类别8 677张图像的分类实验，验证了该方法可以在保证分类准确率的前提下，大幅度地提升传统词袋模型的计算效率。

1 视觉词袋模型

视觉词袋模型的基本原理是将一幅图像视作若干视觉单词的集合，利用每个视觉单词的出现频率来对图像进行描述，其基本结构如图1所示。

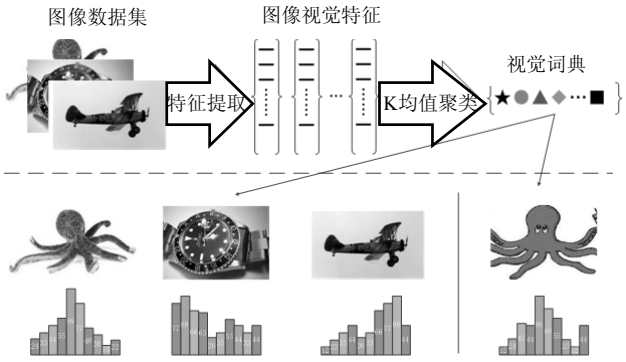
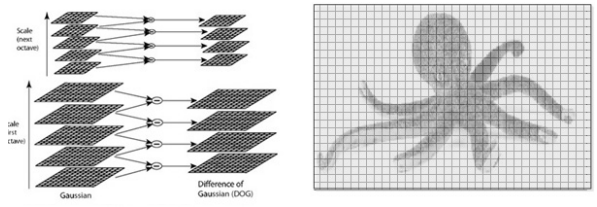


图1 视觉词袋模型的基本结构

1.1 提取视觉特征

视觉词袋模型中首先需要提取图像的底层特征，SIFT[19,-20]是最常用的图像局部特征描述符，该特征具有旋转不变性和尺度不变性，同时具有一定的光照不变性。SIFT特征提取包括采样点检测与特征区域描述两个部分，传统SIFT特征通常需要建立图像的金字塔模型，在多尺度高斯差分空间中检测极值点，在这些极值点所在的不同尺度空间上进行特征提取，从而获得较为稳定的特征；但是这些特征并不都是必需的，由于采样点只集中于少数灰度变化敏感区域，通过传统SIFT检测方法所提取到的特征存在大量重复和不必要特征，同时也丢失了许多可能有助于区分目标的背景信息；此外，在医学、遥感等灰度变化不明显的图像中，经常因为不能检测到足够多的极值点而无法提取局部特征。鉴于此，本文采用均匀采样的方式提取图像的单尺度SIFT特征，其检测点采样方式如图2所示。



多尺度空间SIFT特征检测　　基于均匀采样的单尺度SIFT

图2 基于均匀采样的单尺度SIFT

该方法在传统SIFT上经过简化，取消了向下搜索极值的步骤，通过均匀采样的方式，按照提前给定的采样窗口尺寸和间隔距离，在均匀分布的图像块上提取单尺度SIFT特征。不仅可以较为完整地提取图像各个区域的局部特征，同时也在一定程度上考虑到图像块的空间位置关系，有利于通过局部特征来描述全局图像。经过反复实验表明，均匀采样方式提取的单尺度SIFT特征不仅更为全面地描述了图像的细节特征，同时也极大地减少了计算复杂度，具有更好的鲁棒性和高效性。

1.2 生成视觉词典

每一幅图像中的每一个采样点都会生成一个视觉特征，在词袋模型中，需要将这些数量众多的视觉特征进行聚类，合并相似度较高的视觉特征，最终获得一定数量的聚类中心，作为视觉单词生成视觉词典。聚类中心的数目、聚类算法性能对分类结果有着直接影响。*K*-means算法是最常用的硬聚类算法之一，首先随机划定*K*个初始质心作为种子节点，然后计算每个特征向量到质心的距离，每次循环中将每个特征向量划归到最近的质心，将划归到同一个质心的特征向量视作一个簇，对每个簇计算其聚类中心作为新的种子节点，重复上述步骤直到聚类中心不再改变，最终所获得的聚类中心即为图像视觉单词。

1.3 获得图像表示

由于现实应用中图像尺寸不可能完全相同，所提取的特征数量往往差异巨大，因此需要对图像进行量化的描述，利用前述步骤生成的视觉词典，按照每个视觉单词在图像中出现的频率，将图像描述

为视觉单词直方图。通常采用最近邻查找方法，将图像中的每个特征映射到与之距离最近的视觉单词，然后统计整幅图像中每个视觉单词出现的频率，生成视觉单词直方图作为该图像的词袋模型特征表达式。

2　本文提出的方法

图像的初始特征维数过高是限制视觉词袋模型性能的最大瓶颈，单纯的减少特征采样点数量来提升分类效率，又会因为细节特征丢失而导致分类准确率下降。小波变换是一种多尺度的图像分析方法，可以在不同尺度下对图像进行分析处理，从而有效地捕获图像的细节特征，因此被称为图像处理中的显微镜。利用小波对图像进行尺度变换，可以在不损失局部细节特征的前提下获得图像的低维度特征表示，然后通过在小波低频系数上采用均匀采样的方式，提取单尺度SIFT特征，进一步减少视觉词袋模型中参与计算的局部特征数量，提升分类算法性能。其运算框架如图3所示。

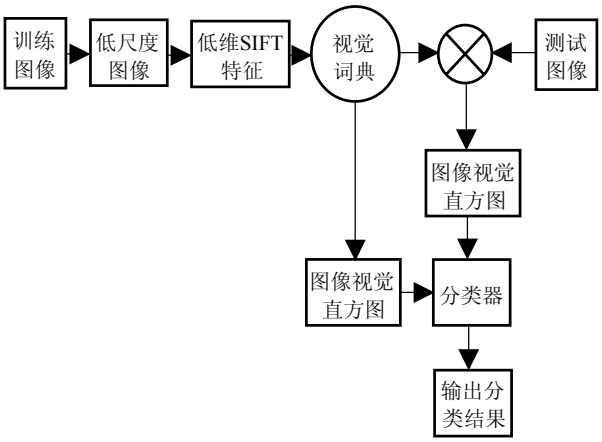


图3　本文的方法结构

1) 利用小波变换对样本图像进行降维，根据图像的尺寸自动确定小波变换层数，最终获得分辨率相近的低维图像。

2) 在低维图像上采用SIFT特征描述子进行均匀采样，根据图像当前尺寸确定采样窗口大小，获得图像的低尺度视觉特征。

3) 利用*K*-means算法对训练集中图像的低尺度视觉特征进行聚类，获得的聚类中心作为视觉单词生成视觉特征词典。

4) 利用视觉特征词典对图像进行量化描述，生成视觉特征直方图。

5) 利用训练集的直方图特征训练SVM分类器，继而对测试图像进行分类识别。

3　实验结果及分析

为验证本文方法的有效性，实验通过图像分类测试与传统词袋模型、空间金字塔词袋模型进行了比较。

3.1　实验数据及环境

实验数据源自加利福尼亚理工学院的Caltech 101数据集，该数据集(除背景类外)共有101个类别，8 677幅图像，每个类别包含30～800张样本图像，每幅图像分辨率从100×300～300×300不等。如图4所示，该数据集具有数据量大、种类多、对象内变化多样等特点，是国际上应用最为广泛的图像分类测试数据集之一。

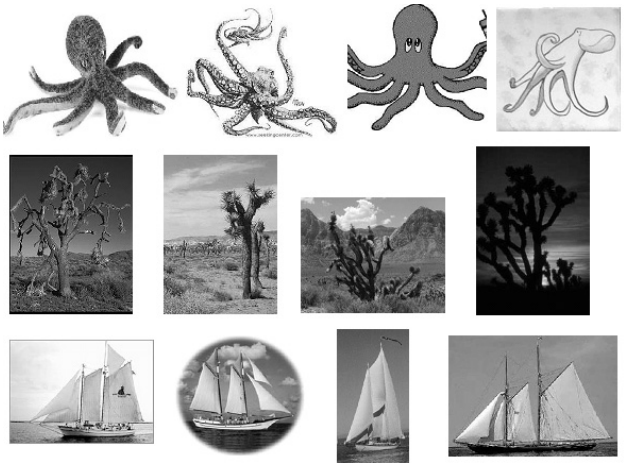


图4　部分实验数据

为显示数据的可靠性，实验将对101类8 677张图像进行完全分类测试，并通过随机抽样等比例分组进行3次交叉验证。所有实验均在Intel酷睿i7处理器、8G内存的台式机，以及Windows7.0系统和MATLAB R2014a的实验环境下进行。

3.2　基于rbf-SVM的图像分类

图像分类中常用的分类器是支持向量机(support vector machine，SVM)，其核心思想是将低维空间的分类问题映射到高维特征空间，通过构造一个超平面来解决非线性分类问题。本文图像分类实验选取LibSVM提供的rbf-SVM分类器。

3.3　与传统词袋模型的对比

通过随机抽取的方式将原始数据集中的每类图像划分为大致均等的*A*、*B*、*C*三组，每组图像数量约为总数的1/3。每次选取其中两组作为训练集，剩余一组作为测试集，进行3次交叉验证，结果如表1所示。

从表1可以看出，本文方法的运行时间仅为传统方法的1/3。而在分类准确率方面，仅有*C*组略微降

低，其余*A*、*B*两组均有所提升，相对保持稳定。实验结果表明，本文的方法可以在保证分类准确率的前提下，极大地提升图像分类效率。

表1 与传统词袋模型对比结果

	方法	运行时间/s	准确率/%
*A*组	传统词袋模型	13 042	50.56
	本文的方法	4 579	51.31
*B*组	传统词袋模型	12 505	52.06
	本文的方法	4 135	53.56
*C*组	传统词袋模型	12 432	52.79
	本文的方法	4 091	51.13

3.4 与空间金字塔词袋模型的对比

为进一步验证本文的方法具有普遍有效性，实验在空间金字塔词袋模型的基础上利用本文的方法进行优化，通过3次交叉验证，结果如表2所示。

表2 与空间金字塔词袋模型对比结果

	方法	运行时间/s	准确率/%
*A*组	空间金字塔词袋模型	18 044	59.23
	本文的方法	6 512	61.39
*B*组	空间金字塔词袋模型	17 512	60.41
	本文的方法	6 544	63.87
*C*组	空间金字塔词袋模型	17 721	59.62
	本文的方法	6 132	61.14

从表2可以看出，本文的方法结合空间金字塔词袋模型，在只增加较少计算时间的情况下，分类准确率相比传统算法有较显著的提高。实验结果表明，本文的方法对于传统词袋模型以外的衍生方法也能够产生积极的效果，可以较大程度地提升此类方法的计算效率。

4 结 束 语

本文提出了一种基于低尺度词袋模型的图像快速分类方法，利用小波降维结合单尺度SIFT特征，极大地减少了词袋模型的初始特征维数，在保证分类性能的前提下，大幅度地提升了计算效率。实验充分验证了该方法可普遍适用于词袋模型及其衍生方法的运算性能的改进。但是，研究中也发现低尺度词袋模型一方面更好的聚焦于低维特征，另一方面也损失一些高维特征，在提升一些类别的分类准确率的同时，也造成了一些类别的分类准确率下降。因此，研究不同尺度下不同特征之间的互补性，将是下一步研究工作的重点。

参 考 文 献

[1] JOACHIMS T. Text categorization with support vector machines: Learning with many relevant features[M]. Heidelberg, Berlin: Springer, 1998.

[2] SIVIC J, ZISSERMAN A. Video google: a text retrieval approach to object matching in videos[C]//Ninth IEEE International Conference on Proceedings of the Computer Vision. Washington D C, USA: IEEE Computer Society, 2003.

[3] FEI-FEI L, PERONA P. A bayesian hierarchical model for learning natural scene categories[C]//Proceedings of the Computer Vision and Pattern Recognition. San Diego, CA USA: IEEE Computer Society, 2005: 524-531.

[4] FERGUS R, FEI-FEI L, PERONA P, et al. Learning object categories from Google's image search[C]//Tenth IEEE International Conference on Proceedings of the Computer Vision. Washington D C, USA: IEEE Computer Society, 2005: 1816-1823.

[5] SUDDERTH E B, TORRALBA A, FREEMAN W T, et al. Learning hierarchical models of scenes, objects, and parts[C]//Tenth IEEE International Conference on Proceedings of the Computer Vision. Washington D C, USA: IEEE Computer Society, 2005: 1331-1338.

[6] RAMESH B, XIANG C, LEE T H. Shape classification using invariant features and contextual information in the bag-of-words model[J]. Pattern Recognition, 2015, 48(3): 894-906.

[7] KHAN R, BARAT C, MUSELET D, et al. Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model[J]. Computer Vision and Image Understanding, 2015, 132: 102-112.

[8] LEI W. Toward a discriminative codebook: Codeword selection across multi-resolution[C]//IEEE Conference on Proceedings of the Computer Vision and Pattern Recognition. Minneapolis, Minnesota, USA: IEEE Computer Society, 2007: 1-8.

[9] JIANCHAO Y, KAI Y, YIHONG G, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//IEEE Conference on Proceedings of the Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE Computer Society, 2009: 1794-1801.

[10] JINJUN W, JIANCHAO Y, KAI Y, et al. Locality-constrained linear coding for image classification[C]//IEEE Conference on Proceedings of the Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE Computer Society, 2010: 3360-3367.

[11] WANG C, HUANG K. How to use bag-of-words model better for image classification[J]. Image and Vision Computing, 2015, 38: 65-74.

[12] GANDHI A, ALAHARI K, JAWAHAR C V. Decomposing bag of words histograms[C]//IEEE International Conference on Computer Vision. Sydney, NSW, Australia: IEEE Computer Society, 2013: 305-312.

[13] FEICHTENHOFER C, PINZ A, WILDES R P. Bags of spacetime energies for dynamic scene recognition[C]//

IEEE Conference on Proceedings of the Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE Computer Society, 2014: 2681-2688.

[14] KATO H, HARADA T. Image reconstruction from bag-of-visual-words[C]//IEEE Conference on Proceedings of the Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE Computer Society, 2014: 955-962.

[15] SADEK I, SIDIB D, MERIAUDEAU F. Automatic discrimination of color retinal images using the bag of words approach[C]//Proc SPIE 9414, Medical Imaging 2015. Orlando, USA: SPIE, 2015: 94141J-8.

[16] CONG Y, WANG S, LIU J, et al. Deep sparse feature selection for computer aided endoscopy diagnosis[J]. Pattern Recognition, 2015, 48(3): 907-917.

[17] SHEN L, LIN J, WU S, et al. HEp-2 image classification using intensity order pooling based features and bag of words[J]. Pattern Recognition, 2014, 47(7): 2419-2427.

[18] BROMURI S, ZUFFEREY D, HENNEBERT J, et al. Multi-label classification of chronically ill patients with bag of words and supervised dimensionality reduction algorithms[J]. Journal of Biomedical Informatics, 2014, 51: 165-175.

[19] LOWE D G. Object recognition from local scale-invariant features[C]//Seventh IEEE International Conference on Proceedings of the Computer Vision. Fort Collins, Colorado, USA: IEEE Computer Society, 1999, 2: 1150-1157.

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

编 辑 黄 莘