文章编号：2095-4980(2014)05-0726-05

一种基于词袋模型的图像分类方法

杨晓敏，严斌宇*，李康丽，苏冰山

(四川大学 电子信息学院，四川 成都 610064)

摘 要：采用词袋模型(BoW)对图像进行分类，并针对传统词袋模型存在的不足进行了改进，提出了一种特征软量化的方式。软赋值量化通过将局部显著特征量化(SIFT)为与其距离最近的若干个视觉单词，并对其进行加权，由此保存特征空间中的距离信息，从而解决硬赋值量化造成的特征空间信息损失问题。通过在 Caltech 101 数据库进行实验，验证了本文方法的有效性，实验结果表明，该方法能够大幅度提高图像分类的性能。

关键词：词袋模型；视觉词典；图像分类

中图分类号：TN911.73；TP391.4 **文献标识码：**A **doi：**10.11805/TKYDA201405.0726

An image classification method based on bag of words model

YANG Xiao-min，YAN Bin-yu*，LI Kang-li，SU Bing-shan

(College of Electronics and Information Engineering，Sichuan University，Chengdu Sichuan 610064，China)

Abstract：The Bag of Words(BoW) model is applied to object classification. An improved algorithm based on soft quantification is proposed. This algorithm quantifies the Scale-Invariant Feature Transform (SIFT) descriptor into several nearest visual words and performs weighting on them, which can conserve the space information. Therefore, it can avoid the information loss of eigen space caused by hard quantification. The experiments are carried out on Caltech 101 database. Experimental results show that the proposed method performs better than traditional method on image classification.

Key words：Bag of Words model；visual words；image classification

近年来，图像数量的大量激增给图像识别、检索以及分类等问题带来了巨大的挑战。如何在浩瀚的数据中准确获得用户所需信息并进行处理，成为该领域亟待解决的问题之一。词袋模型最初应用于文档处理领域，将文档表示成顺序无关的关键词的组合，通过统计文档中关键词出现的频率来进行匹配。近几年来，计算机视觉领域的研究者们成功地将该模型的思想移植到图像处理领域[1-2]，词袋模型(BoW)将图像库看成文档库，将一幅图像看作一篇文档。提取图像特征后，用其生成“视觉单词”，即生成字典，统计每幅图像的视觉单词出现频率，即可完成图像的词袋描述。为了进一步提高该模型的性能，使其在图像识别和分类领域得到更好的应用，研究者们一直致力于对模型的实现过程进行改进和优化[3-8]。

传统局部显著特征的量化就是计算每个局部显著特征与词典中所有视觉单词的最小距离，并将该局部特征用距离它最近的视觉单词进行量化[9-10]。这种方式下局部显著特征的量化仅仅是局部特征之间真实距离的一种粗糙的近似，局部特征在量化过程之中会损失空间信息。处于量化边界两侧的点，虽然实际距离很近，但是量化后其距离就会较大；而处于量化边界同侧的点，虽然实际距离可能很远，但是量化后距离很近。实际上，图像的噪声、场景光照变化、遮挡等因素，导致同一个物体某一区域在不同的图像中，局部显著特征会存在较大差异。在量化时，有可能被量化为不同的视觉单词。因此这种硬赋值方式的量化就会造成检索的错误。本文在传统词袋模型基础上，针对局部显著特征的量化算法进行研究，建立新的量化方式，保存局部显著特征的空间信息，以解决量化赋值带来的信息损失问题。实验结果表明本文方法能够提高图像检索的性能。

收稿日期：2014-07-08；**修回日期：**2014-08-13

基金项目：中国科学院数字地球重点实验室开放基金资助项目(No.2012LDE016)；武汉大学测绘遥感信息工程国家重点实验室开放基金(No.12R03)；高等学校博士学科点专项科研基金资助项目(20130181120005)；四川省科技支撑计划资助项目(No.2014GZ0005)；博士后基金资助项目(No.2014M552357)；南京邮电大学江苏省图像处理与图像通信重点实验室开放基金资助项目(No.LBEK2013001).

***通信作者：**严斌宇 email:yby@scu.edu.cn

本文针对传统的词袋模型忽略图像空间位置特征的问题，对图像中的局部特征量化为其与其最相似的 K 个视觉单词，并根据相似程度进行加权，从而保持局部特征的空间信息。对图像采用基于软量化方式的视觉词典直方图表示，将其代入支持向量机(Support Vector Machine，SVM)分类器进行分类，提高了分类性能。

1　图像的词袋表示

词袋模型的基本原理是将文档看作一个装满了词语的袋子，忽略了词语间的顺序和句法，词袋模型认为每个词的出现都是独立的，并不依赖于其他词的出现。词袋模型需要建立一个词典来包含所有有意义的词，每篇文档就可以表示成词典中单词的直方图。将词袋模型引入到图像检索领域，即将图像看作是一篇文档，将图像中大量的局部特征量化为有限个视觉单词，每幅图像表示为这些视觉单词的直方图。应用词袋模型进行图像分类主要包括特征提取和描述、词典生成以及特征的量化、训练分类器 4 个步骤，如图 1 所示。

1) 特征提取和描述。特征提取和描述主要任务是从图像中提取具有代表性的局部特征，作为图像的描述。传统方法主要采用 SIFT 描述子来进行图像的描述。本文主要采用稠密采样的方式，以固定大小窗口，按照一定的步长遍历整幅图像，将窗口覆盖区域作为一个特征区域，利用描述子进行 SIFT 描述。

2) 词典生成。视觉词典生成的本质主要是适当地划分整个特征空间，将落入同一划分区间的特征向量视为同一个视觉单词可以表达的范围。主要采用 K-means 聚类将 SIFT 特征聚成若干类，每一个类的类中心即视觉单词。所有的视觉单词构成视觉词典，视觉词典的大小即为视觉单词的数量。

3) 特征量化。在进行特征量化时，对于给定的每个图像局部特征，用最近邻查找的方法在视觉词典中查找距离该特征向量最近的聚类中心，所得的即是对应该图像局部特征的视觉单词。然后统计图像中视觉单词出现的次数，得到图像视觉单词直方图，特征量化是对图像的简化表示。

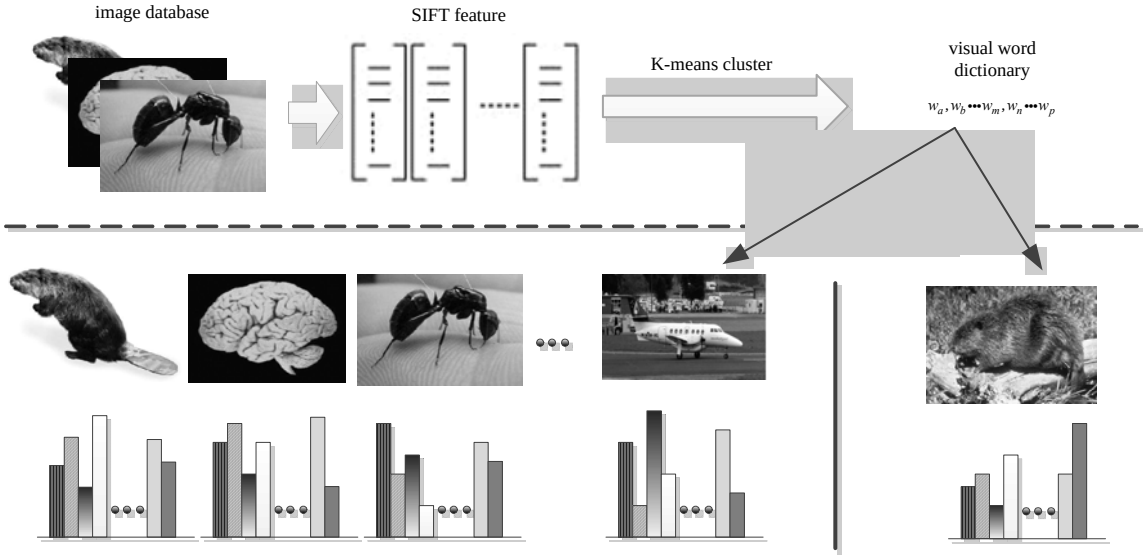


Fig.1 Illustration of Bag of Word model

图 1 词袋表示模型示意图

4) 训练分类器。支持向量机是较常用且实现较为简单的分类器之一。其核心思想通过在特征空间中寻找最优分类超平面，从而对空间中的不同特征进行分类。SVM 求解最优超平面问题可以等价于求解如下方程：

$$\min_{w,\xi}\{\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\xi_i\} \tag{1}$$

约束条件为：$y_i(\boldsymbol{w}\bullet\boldsymbol{x}_i - b) \geqslant 1-\xi_i$，$\xi_i \geqslant 0$，任意的 $i=1,2,\cdots,n$。其中 $\boldsymbol{w}$ 为与超平面的法向量；C 为惩罚因子；ξ_i 为松弛向量。本文采用 SVM 进行分类，选用径向基核函数。

2　基于软赋值的局部显著特征量化

传统局部显著特征的量化就是计算每个局部显著特征与词典中所有视觉单词的最小距离，并将该局部特征用距离它最近的视觉单词进行量化。这种方式下局部显著特征的量化仅仅是局部特征之间真实距离的一种粗糙

的近似，局部特征在量化过程之中会损失空间信息。如图 2 所示，图中 A,B,C,D,E 为构造好的视觉单词，1,2,3,4 为局部显著特征。在传统硬赋值量化方式下，特征 3 和特征 4 量化为[A:1],[C:1]，尽管它们距离很近，但是经过硬赋值量化后它们将属于不同的视觉单词。此外，在硬赋值方式下，特征 1,2,3 同时被量化为视觉单词[A:1]，无法记录特征 2–3 的距离比 1–3 距离近的这种空间信息。处于量化边界两侧的点，虽然实际距离很近，但是量化后其距离就会较大；而处于量化边界同侧的点，虽然实际距离可能很远，但是量化后距离很近。实际上，由于图像的噪声、场景光照变化、遮挡等因素，同一个物体某一区域在不同的图像中，局部显著特征会存在较大差异。在量化时，有可能被量化为不同的视觉单词。因此这种硬赋值方式的量化就会造成检索的错误。硬赋值方式与软赋值方式示意图如图 3 所示。

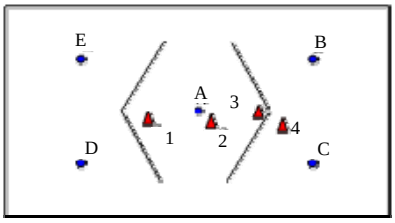


Fig.2 Quantitative examples of SIFT features
图 2 局部特征量化示例

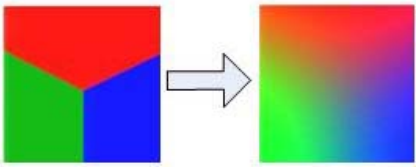


Fig.3 Relationship between hard and soft assignment quantification
图 3 硬赋值量化与软赋值量化之间的关系

针对这一问题，本课题拟采用软赋值方式对局部特征进行赋值。即将局部显著特征量化为距离其最近的 n 个视觉单词，并为每一个视觉单词进行加权，权值如式(2)所示：

$$w=\exp\left(-\frac{d^2}{2\sigma^2}\right) \tag{2}$$

式中：d 为局部显著特征与视觉单词在特征空间中的距离；参数 σ 为设定参数。软赋值量化通过将局部显著特征量化为与其距离最近的 n 个视觉单词，并对其进行加权，由此保存特征空间中的距离信息，从而解决硬赋值量化造成的特征空间信息损失问题。可以看出，采用软赋值方式，局部特征 1,2,3,4 将被赋值为与之距离最近的 n 个视觉单词(若 n=3)，分别为：

$$1\rightarrow\begin{bmatrix}\mathrm{A}:\exp\left(-\frac{d_{1,\mathrm{A}}{}^2}{2\sigma^2}\right)\\ \mathrm{B}:\exp\left(-\frac{d_{1,\mathrm{D}}{}^2}{2\sigma^2}\right)\\ \mathrm{C}:\exp\left(-\frac{d_{1,\mathrm{D}}{}^2}{2\sigma^2}\right)\end{bmatrix},2\rightarrow\begin{bmatrix}\mathrm{A}:\exp\left(-\frac{d_{2,\mathrm{A}}{}^2}{2\sigma^2}\right)\\ \mathrm{B}:\exp\left(-\frac{d_{2,\mathrm{B}}{}^2}{2\sigma^2}\right)\\ \mathrm{C}:\exp\left(-\frac{d_{2,\mathrm{C}}{}^2}{2\sigma^2}\right)\end{bmatrix},3\rightarrow\begin{bmatrix}\mathrm{A}:\exp\left(-\frac{d_{3,\mathrm{A}}{}^2}{2\sigma^2}\right)\\ \mathrm{B}:\exp\left(-\frac{d_{3,\mathrm{B}}{}^2}{2\sigma^2}\right)\\ \mathrm{C}:\exp\left(-\frac{d_{3,\mathrm{C}}{}^2}{2\sigma^2}\right)\end{bmatrix},4\rightarrow\begin{bmatrix}\mathrm{A}:\exp\left(-\frac{d_{4,\mathrm{A}}{}^2}{2\sigma^2}\right)\\ \mathrm{B}:\exp\left(-\frac{d_{4,\mathrm{B}}{}^2}{2\sigma^2}\right)\\ \mathrm{C}:\exp\left(-\frac{d_{4,\mathrm{C}}{}^2}{2\sigma^2}\right)\end{bmatrix}$$

在软赋值量化的方式下，特征 3 与特征 4 不再量化为不同的视觉单词，同时也可以记录特征 2–3 的距离比 1–3 距离近的这种空间信息，解决了硬赋值量化方式下的信息损失问题。硬赋值量化与软赋值量化之间的关系如图 3 所示。

3 本文算法流程

训练过程：

1) 提取训练样本，采用稠密采样的方式提取图片的 SIFT 特征。

2) 对上一步提取出的所有 SIFT 特征，采用 K-means 方式进行聚类，得到若干个聚类中心矢量，即为视觉单词。

3) 对每一幅训练图中的 SIFT 特征进行量化，然后将每幅图采用软量化方式表示成视觉单词的直方图，然后用直方图表示训练图像。

4) 采用 SVM 对训练样本进行训练。单独训练每个类的分类模型，每类的训练样本包括正负样本。正样本为包含这类对象的图像视觉单词直方图，负样本随机选取不包含这类对象的图像视觉单词直方图。正负样本个数相等。

分类过程：

1) 对于测试图像，采用稠密采样的方式提取图片的 SIFT 特征。

2) 对测试图像的 SIFT 特征进行软量化，计算测试图像的视觉单词直方图。

3) 用训练好的 SVM 分类器进行分类，得到分类结果。

4　实验结果与分析

在本文的研究中，采用图像分类和识别任务中较为经典的数据库 Caltech 101 进行实验验证。Caltech 101 数据库具有图像数据庞大、图像种类多、对象类内变化多样等特点，具有一定的代表性。该图像数据集分为 101 类，共 9 146 幅视觉物体图像，包括动物、交通工具、花等类别的物体，具有明显形状变异，如图 4 所示。每个类别包含的图像数量从 40 到 80 不等，每幅图像大约 300×200 像素，属于中等分辨率。

Fig.4 Some samples of the Caltech 101 image database
图 4 部分 Caltech 101 数据库图像

随机抽取 10 类进行实验，然后再分别选取 10,15,20,25 幅图像作为训练数据，其余的则作为测试数据。其中软量化中，紧邻个数 n=10。将全部训练图像稠密提取 SIFT 描述子，图 5 为部分样本稠密提取 SIFT 描述子示意图。用这些描述子构造长度为 500 的码书(使用 K-means 聚类，其中 k=1 000)。所有程序都在 Windows XP 操作系统，2 GB 内存，Matlab 7.0 环境下运行，SVM 分类器采用台湾大学林智仁开发的 LIBSVM。对已有的图像集，选取其中质量不等的字符图像进行训练。每一类图像取 10 个训练样本，分别对参数 C 和 δ 取不同的值进行测试，最后发现当取 C =100, δ= 0.3 或 δ = 0.4 时效果较好，实验结果如表 1 所示。

从表 1 中可以看出，随着每类训练样本数的增加，本文算法与传统词袋算法的分类性能都得到了提高。从总体上看，本文算法的分类准确率高于传统词袋算法。此外，本文还针对视觉单词软量化时近邻的个数选择进行了分析，分别选取近邻个数 n=5,10,15,20，试验结果如图 6 所示。结果表明，分类正确率会随着近邻个数的增长而增长，但到达一定数量时就会下降。

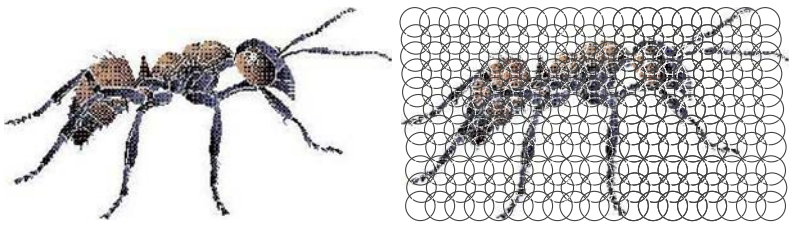


Fig.5 Illustration of extracting dense SIFT descriptor
图 5 样本稠密提取 SIFT 描述子示意图

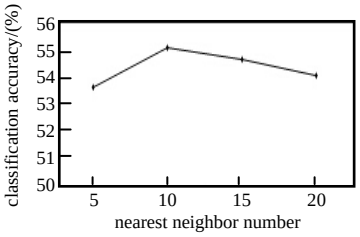


Fig.6 Effect of the nearest neighbor numbers on classification accuracy
图 6 近邻个数对分类准确率的影响

表 1 本文方法与传统词袋算法的准确率比较

Table1 Comparison of accuracies

training sample numbers	10	15	20	25	30
accuracy of the proposed algorithm/(%)	53.2	53.8	54.1	54.3	54.6
accuracy of the traditional bag of word model/(%)	50.6	51.3	51.9	52.2	52.7

5 结论

本文对传统的词袋模型进行了改进。传统词袋模型下，由于量化误差、噪声的缘故，特征空间中距离相近的特征有可能被量化成不同的视觉单词，导致系统分类准确率下降。为此本文提出了一种软量化的方式，将局部显著特征量化为与其距离最近的若干个视觉单词，并对其进行加权，由此保存特征空间中的距离信息，从而解决硬赋值量化造成的特征空间信息损失问题。实验证明，本文算法可以提高图像分类的准确性，且与现有方法相比具有优越性。

参考文献：

[1] WU L,HOI S C H,YU N H. Semantics-preserving bag-of-words models and applications[J]. IEEE Transactions on Image Processing, 2010,19(7):1908–1920.

[2] 谢昭,高隽. 基于高斯统计模型的场景分类及约束机制新方法[J]. 电子学报, 2009,37(4):733–738. (XIE Zhao,GAO Jun. A novel method for scene categorization with constraint mechanism based on Gaussian statistical model[J]. Acta Electronica Sinica, 2009,37(4):733–738.)

[3] Chum O,Philbin J,Sivic J,et al. Total recall: automatic query expansion with a generative feature model for object retrieval[C]// Proceedings of IEEE Computer Vision and Pattern Recognition(CVPR), 2009. [S.l.]:IEEE, 2009.

[4] Perdoch M,Chum O,Matas J. Efficient representation of local geometry for large scale object retrieval[C]// Proceedings of IEEE Computer Vision and Pattern Recognition(CVPR), 2009. [S.l.]:IEEE, 2009.

[5] WANG G,Forsyth D. Object image retrieval by exploiting online knowledge resources[C]// Proceedings of IEEE Computer Vision and Pattern Recognition(CVPR), 2008. [S.l.]:IEEE, 2008.

[6] 刘硕研,须德,冯松鹤,等. 一种基于上下文语义信息的图像块视觉单词生成算法[J]. 电子学报, 2010,38(5):1156–1161. (LIU Shuo-yan,XU De,FENG Song-he,et al. A novel visual words definition algorithm of image patch based on contextual semantic information[J]. Acta Electronica Sinica, 2010,38(5):1156–1161.)

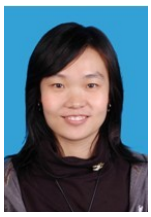
[7] 贾世杰,孔祥维,付海燕,等. 基于互补特征与类描述的商品图像自动分类[J]. 电子与信息学, 2010,32(10):2294–2300. (JIA Shi-jie,KONG Xiang-wei,FU Hai-yan,et al. Auto classification of product images based on complementary features and class descriptor[J]. Journal of Electronics & Information Technology, 2010,32(10):2294–2300.)

[8] LI F F,Liva A O. Fifteen scene categories[DB/OL]. [2014-07-08]. http://www-cvr.ai.uiuc.edu/poncegrp/data/.

[9] 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.)

[10] 吕清秀,李弼程,高毫林. 基于距离度量学习的 DCT 域 JPEG 图像检索[J]. 太赫兹科学与电子信息学报, 2014,12(1):112–118. (LV Qing-xiu,LI Bi-cheng,GAO Hao-lin. JPEG images retrieval in DCT domain based on Distance Metric Learning[J]. Journal of Terahertz Science and Electronic Information Technology, 2014,12(1):112–118.)

作者简介：



杨晓敏(1980–)，女，四川省广安市人，博士，副教授，主要研究方向计算机视觉.email: arielyang@scu.edu.cn.

严斌宇(1975–)，男，河北省保定市人，博士，讲师，主要研究方向计算机视觉.email: yby@scu.edu.cn.

李康丽(1990–)，女，河南省南阳市人，在读硕士研究生，主要研究方向为计算机视觉.

苏冰山(1989–)，男，河南省焦作市人，在读硕士研究生，主要研究方向为计算机视觉.