第37卷 第19期 计 算 机 工 程 2011年10月
Vol.37 No.19 Computer Engineering October 2011

·图形图像处理· 文章编号：1000—3428(2011)19—0204—03 文献标识码：A 中图分类号：TP306

词包模型中视觉单词歧义性分析

刘扬闻，霍 宏，方 涛

(上海交通大学图像处理与模式识别研究所，上海 200240)

摘 要：传统词包(BOW)模型中的视觉单词是通过无监督聚类图像块的特征向量得到的，没有考虑视觉单词的语义信息和语义性质。为解决该问题，提出一种基于文本分类的视觉单词歧义性分析方法。利用传统BOW模型生成初始视觉单词词汇表，使用文档频率、χ^2分布和信息增益这3种文本分类方法分析单词语义性质，剔除具有低类别信息的歧义性单词，并采用支持向量机分类器实现图像分类。实验结果表明，该方法具有较高的分类精度。

关键词：图像分类；视觉单词；文本分类；支持向量机；词包模型

Visual Words Ambiguity Analysis in BOW Model

LIU Yang-wen, HUO Hong, FANG Tao

(Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai 200240, China)

【Abstract】Visual words in the traditional Bag of Word(BOW) model can be gotten by an unsupervised method of clustering the visual features. But one critical limitation of existing BOW is not concerned with the semantic natures of visual words. This paper proposes a visual words ambiguity analysis method based on text categorization. The codebook is generated by the BOW model. There are three ways of analysis——document frequency, χ^2 distribution and information gains, and then they reduce the low information visual words after analyzing. It gets optimized visual words, the histogram formed by the frequency of visual words is used in image categorization task by the Support Vector Machine(SVM) classifier. Experimental results show that this method has higher classification accuracy.

【Key words】image classification; visual words; text classification; Support Vector Machine(SVM); Bag of Word(BOW) model

DOI: 10.3969/j.issn.1000-3428.2011.19.067

1 概述

随着计算机多媒体技术的发展，图像分类已成为计算机视觉领域和多媒体信息处理领域的重点研究课题。图像分类有广泛的应用领域，例如图像理解和目标识别，甚至应用于基于图像内容检索等领域。虽然现阶段图像分类算法有很多种，但是对于不同图像数据库和复杂的图像，各个图像分类算法的效果仍然不理想。因此，图像分类仍是未来计算机视觉的基础问题和热门问题。

根据图像特征描述方式划分当前图像分类的算法，主要分为2类：基于全局特征的描述和基于局部特征的描述。图像分类的早期方法主要集中于图像的基本视觉特性这类全局特征描述场景，但是在多类图像分类算法中，全局特征却存在很大的局限性。近年来，用于场景分类的图像特征还是集中于基于对局部特征的描述。

现阶段文本分类算法与上文所提到的图像分类算法是基于不同的思想，并没有共同之处。但是近年来，为了能够将文本分类的方法用于图像分类，国际上有学者采用文本分类的方法提出基于词包(Bag of Word, BOW)模型[1]的方式描述图像内容。在BOW模型的基础上文献[2]采用概率潜在语义分析(Probabilistic Latent Semantic Analysis, PLSA)[3]模型分析潜在语义性质，完成图像分类。视觉单词形成主要集中于无监督聚类算法，但是由于低维特征与高维语义之间存在语义鸿沟问题，具有同种语义信息的图像特征并不一定会分布于同个视觉单词中，所以，传统BOW模型没有考虑到视觉单词的语义信息和语义性质。BOW模型模拟了文本分类的形式，因此，用于解决文本分类的方法也同样可以用来解决图像分类。

本文将采用文本分类的方法，分析视觉单词的语义性质，即歧义性、同义性和停用词，优化视觉单词的组成，在此基础上改善图像分类性能。

2 视觉单词生成算法

在图像分类中，采用的BOW模型的分类方法很好地模拟了文本分类的过程，但是与文本分类的最大不同之处在于文本中的实际单词和图像中的视觉单词。现有大多数研究都集中在如何能够更好地描述视觉单词，而描述视觉单词的研究主要集中于视觉单词所承载的语义信息，而并非研究类似文本中单词的语义性质。

虽然BOW模型模拟文本分类的方法，但是其所生成的视觉单词的语义性质与文本中的实际单词的语义性质会不同。本文实验结果表明视觉单词的语义性质与文本中的实际单词的语义性质是不同的，通过对视觉单词语义性的分析，消除歧义性的影响，图像分类效果有了较好的提升。因此，分析视觉单词的语义性质对图像分类的影响也是本文的研究重点。

图1给出传统BOW模型和本文改进的模型系统流程框图。本文算法输入的是4类遥感图像。利用K-means聚类算

基金项目：国家“973”计划基金资助项目(2006CB701303)；国家自然科学基金资助项目(41071256)

作者简介：刘扬闻(1985－)，男，硕士研究生，主研方向：图像分类，图像处理；霍 宏，讲师、博士；方 涛，教授、博士生导师

收稿日期：2011-04-26 **E-mail：**liuyangwen@sjtu.edu.cn

法得到初始的视觉单词，然后分析这些初始的视觉单词的语义性质，重新得到新的视觉单词用以分类。在初始的视觉单词构建过程中，本文将每幅图像均匀划分成若干个大小相同的图像块，提取每个图像块的特征。视觉单词形成过程是通过使用 K-means 聚类算法，训练图像集上的所有图像块的特征，每一个聚类中心被定义为一个视觉单词，从而生成由 N 个视觉单词所组成的视觉单词词汇表。计算训练图像中图像块的特征与词汇表中的每个视觉单词所对应的欧氏距离，与其中某个视觉单词距离最近的图像块被记录下来。将图像中的图像块重复上面过程，于是形成一组视觉单词频率统计直方图，利用这组统计直方图代表该幅图像的特征。

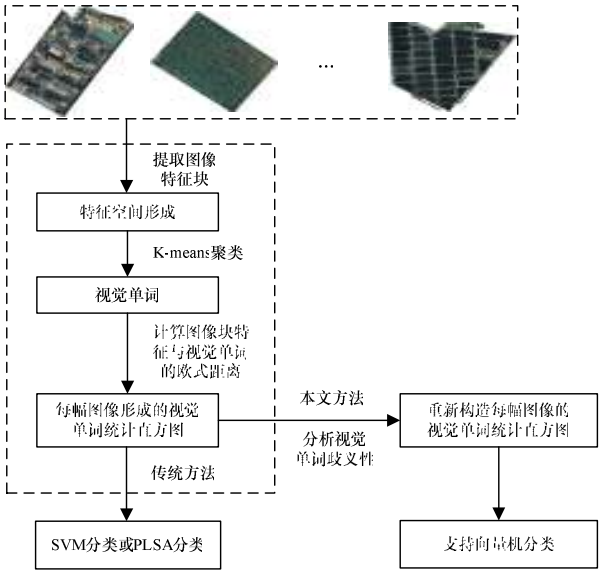


图 1　基于语义性质分析的视觉单词生成流程

3　视觉单词语义性质分析及提取方法

在文本分类的过程中，会对单词的语义性质进行分析，如歧义性、同义性和停用词。文献[4]将视觉单词中出现频次最多的定义为停用词，通过实验分析表明，与文本中出现的停用词不同，此时用在 BOW 模型的停用词是不存在的。

语义性质的另一个特点是同义性，具有同义性的视觉单词对于图像分类来说也是没有影响的，因为本文定义的视觉单词同义性是不同的视觉单词对于同类图像的贡献是相同，即具有同义性的视觉单词只会影响视觉单词词汇表容量大小，而不会改善最终的分类结果。

在文本分析中，语义性质还有一个特点就是歧义性，视觉单词的歧义性概念是同一个视觉单词具有不同的类别信息，即同一个视觉单词对于不同类别的贡献是相同的或相似的。这样视觉单词的歧义性对于图像分类效果的影响会很大。所以，分析视觉单词的歧义性对于图像分类是有必要的。

3.1　视觉单词歧义性分析

为了研究视觉单词的歧义性，本文首先考虑视觉单词频率统计直方图中所能反映的性质。图 2 分别展示 2 幅由不同类别经过标准化后得到的统计直方图，该 4 组统计直方图是按照上文所提到 BOW 模型的传统算法所获得。本文使用了 120 个聚类中心作为视觉单词，其他不同容量的视觉单词词汇表也有类似性质。

从图 2 可以看到，一些视觉单词的波峰或波谷同时出现在不同的类别中，例如图 2 的点 1、点 2 和点 3、点 4。视觉单词歧义性的定义是同一视觉单词对不同类别的贡献是相同的。这里仅简单将“贡献相同”定义为视觉单词在不同类别的出现统计频数是相同或相近。不同视觉单词对于同一类别图像的贡献大小是不同的，虽说几乎所有的视觉单词都会出现在该类的视觉单词统计图中，但是出现的概率是不同的，由此可见，视觉单词的歧义性可能存在差异性，即图 2 中的点 1、点 2 和点 3、点 4 所代表的不同的歧义性视觉单词对最后的分类结果造成不同的影响。

本文定义 2 种不同的歧义性视觉单词：第 1 类歧义性视觉单词，对不同类别信息拥有相同或相似的贡献并且这种对不同类别信息贡献较大，在视觉单词统计图中处于较高波峰处，如图 2 中的点 1、点 2 所示；第 2 类歧义性视觉单词，同样对不同类别信息拥有相同或相似的贡献，但是却对不同类别信息的贡献较小，在视觉单词统计图中处于较低的波谷处，如图 2 中的点 3、点 4 所示。

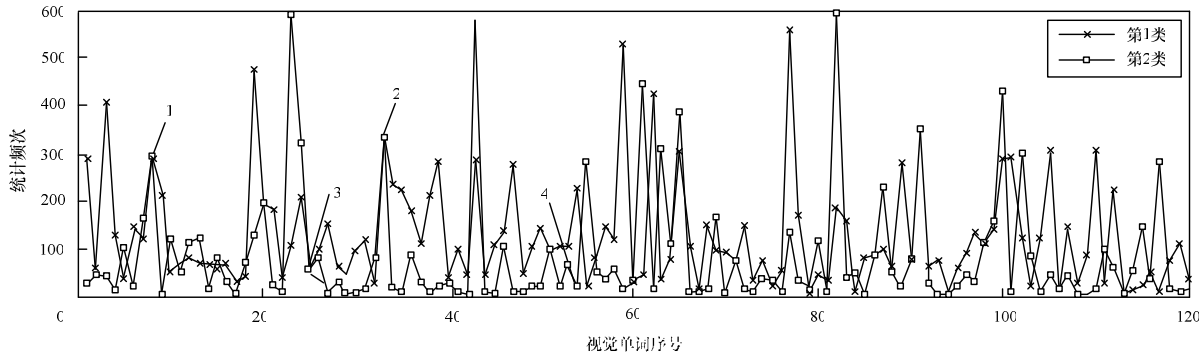


图 2　视觉单词统计频次

3.2　歧义性视觉单词的提取方法

传统文本中存在的单词歧义性大多是从语义内容上定义的，与文本的类别关系不大。而在传统 BOW 模型中的视觉单词的语义内容是不存在的，一般需要人工标注视觉单词的语义，而在形成视觉单词频率统计直方图的过程中，不同图像形成的直方图是具有类别信息的，这样便很容易了解到视觉单词与类别信息之间的关系。例如，同一个视觉单词 w，经过直方图的分析，会同时出现在不同类别的直方图中，对于不同类别的贡献相同或相近，这样的视觉单词就是一个具有歧义性的视觉单词。本文将采用 4 种方法定义视觉单词的歧义性。

3.2.1　文档频率

文档频率(Document Frequency, DF)是文本分类中最基本的方法，统计每个单词在训练文档中出现的频次。在歧义性视觉单词提取过程中，在计算视觉单词的文档频率时分别计算在不同类别信息下发生的文档频率。但是，文档频率也有

其局限性，因为只从概率的角度计算视觉单词发生的概率，并没有从信息论的角度考虑视觉单词与类别信息之间的可靠关系。

3.2.2 χ^2 分布

χ^2 分布(CHI)是一种常用的测量 2 个随机变量的独立性的方法，本文讨论的是视觉单词 w 和图像类别 c_i 之间的相关性。χ^2 分布的定义如下：

$$\chi^2(w,c_i)=N(AD-BC)^2/(A+C)(B+D)(A+B)(C+D) \quad (1)$$

其中，N 表示视觉单词 w 发生总频数；A 表示该视觉单词 w 是属于类别 c_i 的发生频数；B 表示该视觉单词 w 中属于非类别 c_i 的发生频数；C 表示在属于该类别 c_i 全部视觉单词中，除了视觉单词 w，剩下的视觉单词发生频数；D 表示既不属于视觉单词 w 也不属于类别 c_i 的视觉单词发生频数。$\chi^2(w,c_i)$ 值越大说明视觉单词 w 与类别 c_i 相关性越强；相反地，$\chi^2(w,c_i)$ 值越小说明视觉单词 w 与类别 c_i 相关性越低。

3.2.3 信息增益

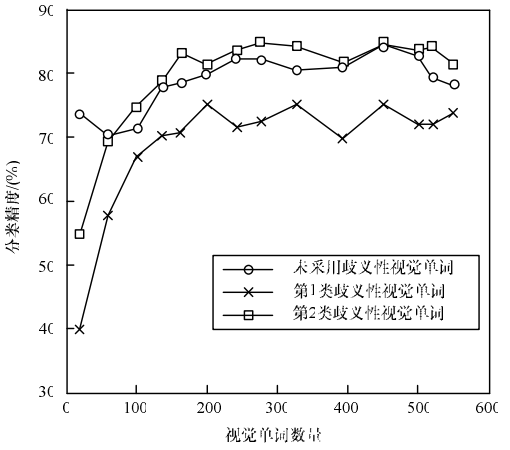
信息增益(Information Gain, IG)可以用于视觉单词歧义性检测。在信息增益提取歧义性视觉单词时，衡量的是每个视觉单词 w 能够为不同类别带来多少信息，如果带来的信息相似则可以划分为具有歧义性的视觉单词。信息增益是定义在熵的基础上，视觉单词 w 能给类别 c_i 带来信息增益，即类别自身所具有的原本的熵与在该视觉单词已发生情况的下条件熵的差值，信息增益定义如下：

$$I_{\mathrm{IG}}(w,c_i)=-\sum_{c_i\in\{0,1\}}p(c_i)\lg p(c_i)+\sum_{w\in\{0,1\}}p(w)\sum_{c_i\in\{0,1\}}p(c_i|w)\lg p(c_i|w) \quad (2)$$

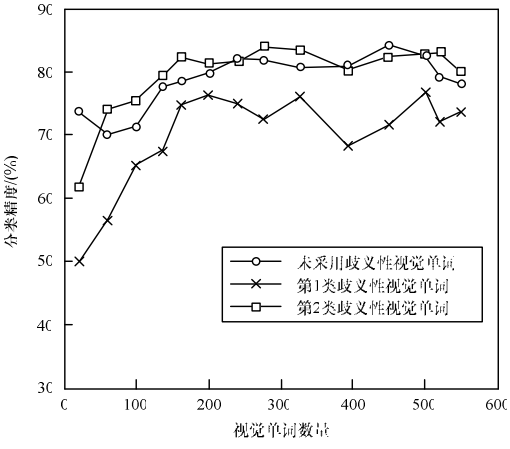
4 实验及结果分析

为了研究视觉单词歧义性对 BOW 模型分类效果的影响，实验采用 4 类遥感图像分别是果园、池塘、房屋和耕地，总共 1 882 幅图像，每类包括 400 幅~500 幅图像。整个图像数据库被随机分为 1 130 幅训练图像和 752 幅测试图像 2 个部分，采用径向基函数作为 SVM 分类器的核函数。利用已得到的视觉单词频率统计直方图表示每幅图像，分析视觉单词的歧义性，剔除直方图中包含歧义性的视觉单词，将优化后直方图作为 SVM 分类器[5]的输入特征进行训练，最后利用训练好的模型对测试集分类。

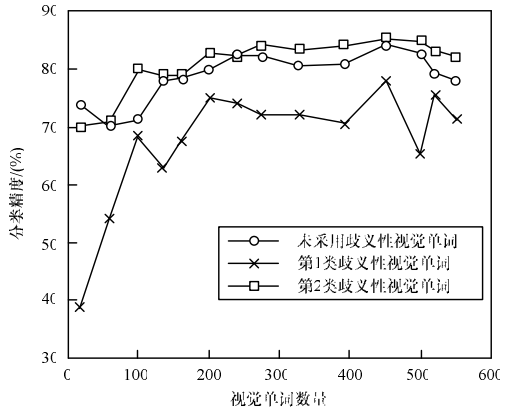
本文首先分析第 1 类歧义性视觉单词和第 2 类歧义性视觉单词对分类效果的影响，采用视觉单词词汇表容量从 20~550 之间，14 种容量大小不同的词汇表，并且将第 1 类歧义性视觉单词和第 2 类歧义性视觉单词分开作为比较，分别采用上文提到的 3 种分析歧义性的方法：文档频率，χ^2 分布和信息增益。如图 3 所示，采用容量不同的词汇表对分类性能有一定影响，当视觉单词数目过小时，分类精度非常低，但随着视觉单词数目增加，分类精度会变高，并且分类精度的稳定性趋于平缓。从 3 种分析歧义性方法可以看出，分类精度依次从第 2 类歧义性视觉单词、未采用歧义性和第 1 类歧义性递减。去除第 2 类歧义性视觉单词后会对分类结果有积极的影响；反之，去除第 1 类歧义性视觉单词后对分类结果有负面的影响。第 1 类歧义性视觉单词虽然存在歧义性，但是对分类效果仍有积极的作用。去除第 2 类歧义性视觉单词会对分类效果有较好的改善。为了验证本文 3 种视觉单词歧义性分析的方法对图像分类的影响，利用文档频率、χ^2 分布、信息增益这 3 种方法对第 2 类歧义性视觉单词的分类效果进行比较。从图 3 可以看出，信息增益和文档频率方法在词汇表容量较小时分类效果变化较大，但从总体上看 χ^2 分布和文档频率方法优于信息增益方法，而且这 3 种方法好于传统方法。



(a)DF 方法



(b)IG 方法



(c)CHI 方法

图 3 视觉单词歧义性对分类精度的影响

5 结束语

本文提出基于文本分析方法解决 BOW 模型中的语义性质问题，首先通过传统的 BOW 模型生成视觉单词，对这些视觉单词分析语义性质，剔除具有低类别信息的歧义性单词，从而达到提高分类精度的效果。通过实验分析表明该方法能改善分类精度，然而本文提出的语义性质是针对反应类别信息，与文本中的语义性质还是有一定差别。今后将进一

(下转第 209 页)