

基于 Bag of Words 的干果图像分类研究

施明登^{1,2}, 周鹏^{1,3*}, 白铁成¹ (1. 新疆塔里木大学信息工程学院, 新疆阿拉尔 843300; 2. 浙江大学计算机科学与技术学院, 浙江杭州 310027; 3. 郑州航空工业管理学院电子通信工程系, 河南郑州 450015)

摘要 针对干果图像信息量大、分类精度低和耗时多的特点, 提出利用 Bag of Words 模型提取图片的代表特征, 并采用朴素贝叶斯分类器指导特征矩阵分类。结果表明, 图像分类精度能达到 80%, 分类处理时间约为 2 s。通过增加学习样本来进一步提高分类精度, 将 Bag of Words 应用于干果图像识别和分类是可行的。

关键词 图像分类; 词袋模型; 朴素贝叶斯分类器

中图分类号 S126; TP391 文献标识码 A 文章编号 0517-6611(2014)29-10381-03

Research on Digital Dried Fruit Image Classification Based on Bag-of-Words Model

SHI Ming-deng^{1,2}, ZHOU Peng^{1,3*}, BAI Tie-cheng¹ (1. College of Information Engineering, Tarim University, Alar, Xinjiang 843300; 2. Computer Science College, Zhejiang University, Hangzhou, Zhejiang 310027; 3. Department of Electronic and Communication Engineering, Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou, Henan 450015)

Abstract According to the characteristics of digital dried fruit image classification which have lots of information, weaken classification accuracy and more time-consuming, it is put forward to extract image representation using the Bag-of-Words model and to classify the feature matrix with Naïve Bayes Classifier. The results showed that the accuracy was over 80%, the treatment time was 2 seconds. By increasing the learning samples to further improve the classification accuracy, the Bag of Words applied to the dried fruit image recognition and classification is feasible.

Key words Image classification; Bag-of-words model; Naïve Bayes Classifier

物体识别分类是当前国内外计算机视觉领域一个活跃的研究方向, 物体识别和分类的本质就是建立一个能够识别出图像中感兴趣物体类别的计算系统, 在现实生活中有着广泛的应用需求, 具有相当高的应用价值和研究意义^[1]。在现实世界中, 人们用人眼视觉进行图像中物体的认知和识别, 通过人眼捕获外界传来的视觉信息, 然后将这些信息交给大脑进行处理, 大脑进行分析之后, 将会告诉人们图像的内容。而这对于计算机来讲却是一项很复杂的任务, 轻微的物体外观的改变也会导致计算机的分析认知过程产生偏差, 从而给出错误的结果。因此, 如何使计算机识别最大限度接近甚至超越人眼识别的精确度, 仍需要进行进一步的探索。

目前对干果图像进行识别和分类的方法主要有两种。一种是根据图像的颜色、纹理、边缘和光谱等底层的特征进行识别和分类^[2], 这种方法容易受光照度、尺寸变换和角度等因素影响, 识别误差较大。近年来, 建立在提取图像局部特征之上图像分类算法得到越来越多的研究^[3], 与前一种方法相比, 利用局部特征进行图像描述时, 每幅图像被分解成很多独立的局部区域, 每个区域对应一个特征向量。这些局部区域的特征具有各种不变性(旋转不变性、尺度不变性、仿射不变性、灰度不变性等)^[4-5], 为图像的识别和分类提供一种统计意义的图像内容表述。其中典型的代表是 Bag of Words 算法, 它具有较高的识别和分类效率, 近年来受到越来越多研究者的关注^[6]。

为此, 笔者针对干果图像信息量大、分类精度低和耗时多的特点, 提出利用 Bag of Words 模型提取图片的代表特征,

并采用朴素贝叶斯分类器指导特征矩阵分类, 对 Bag of Words 算法及其在干果识别上的应用进行了深入的研究。

1 算法实现

Bag of Words 算法源于自然语言处理和信息检索, 在文本分类任务中取得了很大成效^[7]。与文本相似, 图像可以被视为一些与位置信息无关的局部特征的集合, 这些局部特征的地位类似于文本中的单词, 这里称作“视觉单词”, 视觉单词的集合称为“视觉词典”; 图像的灰度、梯度等低级特征通过视觉单词形成了中级语义表示, Bag of Words 模型就可以类比地应用于图像识别分类中^[8]。

Bag of Words 算法一般分为 4 步: 第 1 步, 从图像中提取出局部区域, 将图像表示成局部区域的集合; 第 2 步, 对这些提取出来的局部区域进行描述, 将每一个局部区域用一个特征向量来描述, 为了获得好的分类效果, 这些特征向量需要具备不同程度的不变性, 如旋转、缩放、平移等不变性; 第 3 步, 利用上一步得到的大量的特征向量, 抽取其中有代表性的向量作为单词, 形成词典; 第 4 步, 当图像的局部区域和某一单词的相似性超过某一阈值, 就认为图像中包含这个单词。最终对图像中的单词进行统计, 把图像表示成单词的分布, 设计并训练一个分类器, 利用图像中单词的分布进行分类。其基本流程见图 1。

1.1 图像的特征提取和表示 常用的特征提取方法有密集采样、随机采样和基于特征点的采集。密集采样是对图像中每个像素点进行处理, 这样做能获取图像中的大多数信息。但是随着科技的发展, 人们常用的摄像产品的像素越来越高, 拍出来的图像也越来越大, 这样就会造成所获取的信息量过大, 会带来存储及运算速度上的问题, 所以这种方法在实际应用中并不可取。随机采样是图像中任意选取一部分图像块来作为图像的特征, 虽然当采样数量较小时, 可以解决存储和运算上的问题, 但同时也会造成识别效果较差。该

基金项目 国家自然科学基金项目(61162018); 国家自然科学基金项目(F010408); 新疆农业信息化研究中心重点项目(TSAI201401)。

作者简介 施明登(1977-), 男, 福建三明人, 讲师, 在读博士, 从事计算机视觉、嵌入式系统研究。* 通讯作者, 教授。

收稿日期 2014-09-01

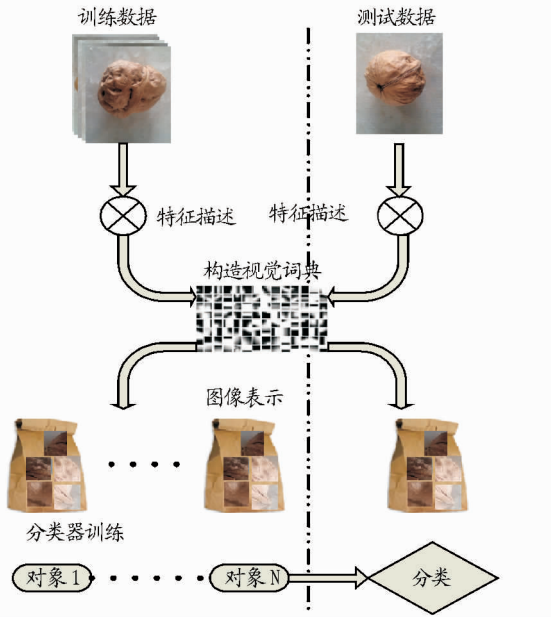


图1 Bag of Words 算法基本流程示意

研究运用基于特征点的采集方法,采用 SIFT 算法对图像进行特征点提取。

SIFT 算法一般分为 4 步:①检测尺度空间的极值点;②定位极值点;③确定每个极值点的方向参数;④生成极值点描述子。SIFT 算法在尺度空间中使用了—种图像金字塔结构,其中包括高斯金字塔和高斯残差金字塔两个部分。高斯残差金字塔由对应相邻的高斯金字塔中的两个图像尺度层相减获得。金字塔由多级组成,每级包含多个图像尺度层,每层之间的 σ 值相差 k 倍。每级的底层由下一级中对应的尺度层通过系数为 2 的抽样操作获得。高斯残差是尺度规格化 Laplacian 算子的一种近似,因而 SIFT 算法直接选取高斯残差金字塔中在局部区域内获得极值的像素点为特征点。具体的 SIFT 特征点提取过程可参见文献[5]。图 2 是测试中所使用的原图和 SIFT 效果图。



图2 原图(a)和 SIFT 效果图(b)

1.2 视觉词典创建 用 SIFT 等特征提取方法可以从一个图像中检测并提取到大量的兴趣点。从一个图像中通常可以得到数百个甚至上千个兴趣点,其是由图像本身的内容以及兴趣点检测算子的敏感度等决定的。然后,图像分类算法还要对这些大量的兴趣点进行处理,减少要处理的兴趣点的数量。因为过多的兴趣点会降低词袋模型的学习效率,将兴趣点聚类为单词是词袋模型的一个必要步骤,聚类中心的数目、聚类效果的好坏对分类正确率有直接影响。该研究采用 K -均值算法进行聚类,然后通过统计视觉单词出现的频

率和多个单词共同出现的情况获得视觉单词的直方图。 K -均值算法聚类过程可参见文献[9]。图 3 是测试过程中得到的核桃类直方图。

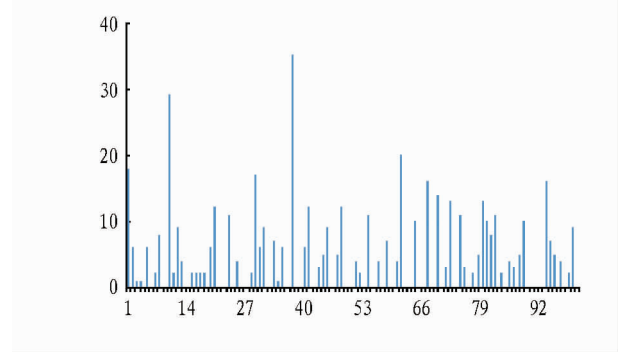


图3 测试过程得到的核桃类直方图

1.3 分类器设计 在应用聚类方法生成视觉单词之后,每个图像将会由一个单词袋来表示,下一步就涉及如何训练分类器,对图像类别进行区分。因为 Bag of Words 模型是源自于文档处理领域,因此文档处理领域所应用的一些分类方法同样也适用于图像处理领域。该研究采用具有代表性的 Naïve 贝叶斯模型为生成模型。

Naïve 贝叶斯分类器的最优解使得后验概率 $p(c|w)$ 最大,其数学描述如下:

$$c^* = \operatorname{argmax}_c(p(c|w)) = \operatorname{argmax}_c(p(c) \prod_i w_i)$$

计算类 $p(w|c)$ 的时候,将该类所有训练图片 $p(w|c)$ 值的平均值作为该类的 $p(w|c)$ 。其中,模型变量及函数定义如下: V :视觉词典的大小,也就是 K -均值聚类的聚类中心个数; N :单幅图像提取的 SIFT 特征点数; w_i 一个 SIFT 特征点可表示为一个 V 维向量,若该点以视觉词典中第 k 个视觉单词表示,则 w_i 是第 k 维为 1,其余为 0 的 V 维向量; w :单幅图片可使用其 N 个 SIFT 点来表示,因此 $w = \sum_{i=1}^N w_i$ 用来唯一表示单幅图片; c :图像类别; $p(c)$:类别 c 的先验概率; $p(w|c)$:类别 c 的 w 概率分布; $p(c|w)$:图像 w 为类别 c 的概率。

2 验证试验结果与分析

为验证基于 Bag of Words 算法对干果识别和分类的有效性,选取核桃、沙漠果、无花果、巴旦木和灰枣 5 类干果图像数据进行试验,每类干果分别用 5 幅图片数据用于训练,3 幅图片数据用于测试。开发环境:Windows 7, Visio Studio 2010, C++ , OpenCV2.4.3。

分类正确率为检测到的正确分类的样本个数占总样本数的比率,公式为:分类正确率 = 正确的目标类样本数/目标类样本数。

试验中首先用 SIFT 算法对训练图像进行特征提取和描述,得到图像的描述符;然后对得到的所有特征点和描述符用 K -均值聚类量化得到类心;接着用得到的类心通过计算其与特征点描述符的距离来处理每幅图像得到向量柱形图(频数表);最后用向量柱形图来表示这幅图片,最后用 Naïve 贝叶斯分类器对图像进行分类。其中 $w = \{w_1, w_2, \dots, w_{100}\}$, $c = \{c_1, c_2, c_3, c_4, c_5\}$ 。

从每类中各取 1 幅测试,测试结果统计如表 1 所示。此次测试将无花果错误分类至核桃类。

表 1 单幅测试中图像 w 为类别 c 的概率

$p(c w)$	灰枣	无花果	巴旦木	沙漠果	核桃
c_1	2.9E+106	2.9E+124	1.2E+149	1.1E+148	1.4E+145
c_2	1.0E+98	2.8E+126	3.3E+153	7.7E+151	2.7E+145
c_3	6.9E+93	2.1E+125	3.9E+160	3.7E+150	2.8E+145
c_4	5.4E+97	1.3E+125	5.8E+158	9.4E+152	1.8E+146
c_5	3.0E+102	2.9E+127	4.6E+153	1.1E+151	2.8E+147

15 幅全部测试后,测试结果统计如表 2 所示。每幅分辨率为 $300 * 250$,测试耗时 2 s,总体正确率为 80%。

通过对测试中获得的结果数据进行分析发现,该算法对灰枣、巴旦木和沙漠果的识别效果较好,而无花果和核桃的图像比较相似,测试误差较大。

表 2 测试统计结果

$p(c w)$	第 1 张	第 2 张	第 3 张	正确率//%
c_1	✓	✓	✓	100.0
c_2	✗	✓	✓	100.0
c_3	✓	✓	✓	100.0
c_4	✓	✓	✓	66.7
c_5	✓	✗	✗	33.3

3 结语

在图像分类任务中,在词袋模型的视觉词典的学习和改

(上接第 10370 页)

景,特别是分子生物学与微电子学、光电子学、微细加工技术及纳米技术等新学科、新技术结合,使得生物传感器的研究开发已成为世界科技发展的新热点,形成 21 世纪新兴的高技术产业的重要组成部分,具有重要的战略意义。

生物传感器主要应用在以下主要领域:①食品工业。生物传感器在食品分析中的应用包括食品成分、食品添加剂、有害毒物及食品鲜度等的测定分析。在食品工业中,葡萄糖的含量是衡量水果成熟度和贮藏寿命的一个重要指标。已开发的酶电极型生物传感器可用来分析白酒、苹果汁、果酱和蜂蜜中的葡萄糖;其他糖类,如果糖及啤酒、麦芽汁中的麦芽糖,也有成熟的测定传感器。另外,还可以将传感器广泛应用于食品添加剂的分析、农药残留量分析、食品鲜度的检测和微生物和毒素的检验等^[5]。②环境监测。生物传感器已广泛应用于水环境监测(如生化需氧量和硝酸根离子)和大气环境监测(如 SO_2 和 NO)等。③发酵工业。在各种生物

过程中通常需要大量的样本,这是因为一个图像类别内部的个体本身就存在着差异,为了学到足够的类别知识,需要涵盖这些有差异的类内个体,所以学习一个类别需要成百甚至成千的训练样本。在仅有少量样本的条件下,分类算法要保持良好的分类性能,最有效的方法就是充分利用已经学习好的源类别图像的知识,将其恰当地迁移到当前的目标类别图像的学习中。

参考文献

- [1] 王冰. 基于词袋模型的图像分类方法研究[D]. 合肥:中国科学技术大学,2008.
- [2] 田渊栋. 基于特征组合的一般物体识别相关算法研究[D]. 上海:上海交通大学,2007.
- [3] ZITOVA B, FLUSSER J. Image registration methods: A survey[J]. Image and Vision Computing, 2003, 21(11): 977-1000.
- [4] 郭立君, 刘曦, 赵杰煜, 等. 基于改进局部特征分布的图像分类方法[J]. 模式识别与人工智能, 2011, 24(3): 368-375.
- [5] LOWE D G. Distinctive image features from scale invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [6] 黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述[J]. 计算机学报, 2014, 37(6): 1225-1240.
- [7] 苏金树, 张博锋, 徐昕. 基于机器学习的文本分类技术研究进展[J]. 软件学报, 2006, 17(9): 1848-1859.
- [8] 赵春晖, 王莹, KANEKO M. 一种基于词袋模型的图像优化分类方法[J]. 电子与信息学报, 2012(9): 2064-2070.
- [9] 张白妮, 骆嘉伟, 汤德佑. 动态的 K -均值聚类算法在图像检索中的应用[J]. 计算机工程与设计, 2004, 25(10): 1843-1846.

传感器中,微生物传感器具有成本低、设备简单、不受发酵液混浊程度的限制、可能消除发酵过程中干扰物质的干扰等特点。因此,在发酵工业中广泛地采用微生物传感器作为一种有效的测量工具。④医学。医学领域的生物传感器发挥着越来越大的作用。生物传感技术不仅为基础医学研究及临床诊断提供了一种快速简便的新型方法,而且因为其专一、灵敏、响应快等特点,在军事医学方面也具有广阔的应用前景。

参考文献

- [1] 邱城, 余耀斌, 吴雪莲. 电感耦合等离子体发射光谱法分析青稞中微量元素的研究[J]. 中国农业科技导报, 2012, 13(3): 94-100.
- [2] 张红梅, 王俊. 电子鼻传感器阵列优化及其在小麦储藏年限检测中的应用[J]. 农业工程学报, 2006, 12(22): 164-167.
- [3] 邹小波, 赵杰文. 电子鼻快速检测谷物霉变的研究[J]. 农业工程学报, 2004, 20(4): 121-124.
- [4] RIDER T H, PETROVICK M S, NARGI F E. A B cell-based sensor for rapid identification of pathogens[J]. Science, 2003, 301(5630): 213-215.
- [5] 朱将伟, 柴春彦, 刘国艳. 己烯雌酚的过氧化氢标记及己烯雌酚的免疫传感器[J]. 食品科学, 2006, 27(5): 57-60.