

文章编号 :1007 - 6492( 2000 )04 - 0062 - 04

# 粮仓害虫的特征提取与分类的研究

徐 昉,邱道尹,沈宪章

( 郑州工业大学电气信息工程学院,河南 郑州 450002 )

摘 要 :运用图像处理和模式识别的研究方法,对粮仓中两种主要害虫玉米象和谷蠹进行了特征提取和分类.根据其形态学特征提取了面积、复杂度和周长,进一步设计了最近邻分类器,实现了玉米象和谷蠹的正确分类,识别率达 95% 以上.该系统为粮仓害虫在线监测系统的实现打下了良好的基础,同时也为粮仓害虫种类的识别提供了一种新方法.

关键词 :储粮害虫 ;图像处理 ;模式识别 ;特征提取 ;分类器

中图分类号 :TP 752 文献标识码 :A

## 0 引言

有关资料显示,全世界每年至少有 5% 的粮食被害虫糟踏,如果人力、物力和技术跟不上,可能会达到 20% ~ 30%<sup>[1]</sup>.我国粮食总储量约为 9500 亿斤,国库粮食损失约为 0.2%<sup>[2]</sup>.因此许多科研人员致力于粮仓害虫的检测、识别和杀虫的研究,新的测虫、杀虫方法不断出现.

在检测害虫方面,目前国外主要是通过对仓储害虫信息素的结构分析与鉴定,进行人工合成害虫信息素,进而设计基于害虫信息素的诱捕器,进行诱捕和捕杀害虫.而国内对害虫信息素的研究起步较晚,大部分粮库现场仍采用简单的粮食取样器和食物诱捕器.对于基于害虫性信息素的诱捕器,每种只能提供检测捕杀一种或几种害虫<sup>[3]</sup>,因为每种害虫的生物特性不同,性信息素当然也不同,从而造成了这类诱捕器的性能单一性,应用上的局限性.再者,当前还有一个重要问题就是:仓虫的密度和种类呈上升趋势,这也给研究信息素的诱捕器带来了一定的困难.国外有一种名为 WB Probe 的害虫监测器也仅提供同时监测 3 种仓库甲虫<sup>[4]</sup>.能否寻找一种实时的、高效的,具有学习功能并能同时监测多种储粮害虫的粮仓监控仪器,是粮食储藏中的重要课题之一.

随着计算机技术的飞速发展,计算机视觉及图像处理与识别技术在农业中也取得了卓有成效

的应用<sup>[5~7]</sup>.为此,提出了基于图像识别的检测害虫的新方法,并在此方法的基础上开发了粮仓害虫检测系统.本文的研究内容正是粮仓害虫监测系统的一个重要组成部分,利用图像处理和模式识别实现了粮仓害虫的特征提取与分类.

## 1 图像预处理

图 1 为粮仓害虫特征提取与分类的系统总框图.其中,图像预处理过程主要包括图像增强、图像去噪、图像二值化以及二值图像的形态学处理(如开运算和闭运算等),为特征提取做好准备工作.

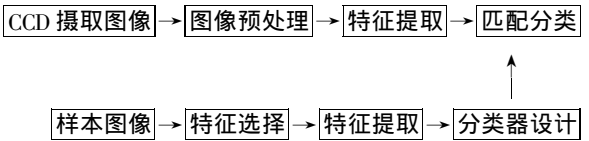


图 1 害虫分类系统框图

下面简要介绍图像预处理的各个环节:

(1) 图像滤除噪声就是滤除图像中的一些随机噪声.因为图像噪声一般都是高频成分,故可以采用低通滤波,其思路是保留图像空间的低频成分而减少图像的高频成分,但是滤波后使图像变得模糊.本文采用中值滤波以滤除噪声,中值滤波的原理是将区域中所有值按大小进行排序,并将排序后位于中间的像素值赋予中心像素.

(2) 图像增强.其目的是为了突出图像的边缘信息,加强图像的轮廓特征,以便于人眼和机器

收稿日期 2000 - 06 - 15,修订日期 2000 - 08 - 14

基金项目 :中科院模式识别国家重点实验室资助项目

作者简介:徐 昉(1977 - )男,安徽省宿州市人,郑州工业大学硕士研究生.

万方数据

的识别.

图像增强处理有两种方法 :空间域法和频率域法.本文所采用的是对比度拉伸增强图像 ,是一种空间域的方法.对比度拉伸的原理是 :假如原图目标像素的大部分灰度等级在小区间 $[r_1,r_2]$ 内 ,这时使用如下的变换关系 ,即

当  $r_1 \leq r \leq r_2$  时 ,

$$s = \frac{s_2 - s_1}{r_2 - r_1}(r - r_1) + s_1 ;$$

当  $r < r_1$  时 ,  $s = s_1 ;$

当  $r > r_2$  时 ,  $s = s_2 ;$

其中 , $r$  为灰度变量 ; $s$  为扩展后的灰度变量 ; $r_1 , r_2$  为原图像中要进行对比度扩展的范围 ; $s_1 , s_2$  表示扩展后对应的新值.

这种分段的线性变换 ,把原图 $[r_1,r_2]$ 区间内的灰度扩展了 ,但把原图在 $[s_1,r_1]$ 及 $[r_2,s_2]$ 区间外的灰度压缩了 ,达到了对比度增强的目的.

(3) 图像的二值化 ,其目的是将目标与背景很好的区分开来变成目标与背景的黑白图 ,以利于下一步的目标特征提取.图像二值化的方法很多 ,文献 [8-9] 总结了 50 多种阈值选取方法 ,而文献 [10,11] 则对其中一些图像分割方法进行了性能评价.考虑到算法的实时性 ,我们采用的是简单统计法 ,其阈值  $t$  的计算公式见参考文献 [6].

(4) 经过二值化以后的图像还可能有孤立点噪声 ,另外分割后的目标也有可能具有小孔洞和小的裂缝 ,这时可借助于形态学处理的方法对分割后的二值图进行开运算或闭运算.开运算使目标轮廓光滑 ,并去掉了毛刺和孤立点 ,钝化了角.闭运算则填平了小沟 ,弥合了孔洞和裂缝.

## 2 特征提取

为了完成害虫识别的任务 ,首先要提取对识别有效的图像特征 ,即提取能够反应害虫的各种属性的特征 ,并将这些特征构成一个  $n$  维空间 ,这样所有害虫种类都抽象为特征空间中的点.特征提取是一个比较复杂的问题 ,需要对识别对象即害虫的各种属性有充分的理解.下面对粮仓中两种主要害虫玉米象和谷蠹给予简要分析 :

玉米象成虫体长 2.3~5 mm ,褐色至暗褐色 ,头部啄显著 ,呈象鼻状 ;谷蠹成虫体长 2~3 mm ,长圆筒形 ,暗赤褐色 ,略有光泽.经过分析和直观的观察 ,我们提取了以下几个特征 :目标的面积、目标的周长、目标的复杂度.具体数据见表 1 与表 2.

表 1 谷蠹样本特征数据表

样本	面积	周长	复杂度
样本 1	876	113	14.576
样本 2	892	123	16.941
样本 3	876	129	17.934
样本 4	837	115	15.805
样本 5	891	120	16.164
样本 6	735	109	16.161
样本 7	893	123	16.941
样本 8	934	114	15.914
样本 9	1019	121	15.368
样本 10	865	121	16.926

表 2 玉米象样本特征数据表

样本	面积	周长	复杂度
样本 1	1543	196	24.896
样本 2	1131	154	20.969
样本 3	1272	161	20.378
样本 4	1525	180	21.245
样本 5	1465	184	23.109
样本 6	1402	177	22.346
样本 7	1275	158	19.576
样本 8	1097	156	22.184
样本 9	1343	175	22.803
样本 10	1545	187	22.634

上述数据均是在相同的条件下获得的 ,目标的面积对数字图像而言即目标区域所包含的像素个数 ,设目标像素的灰度值为“ 1 ”,则对于  $N \times M$  大小的数字图像 ,目标面积为

$$s = \sum_{i=1}^{N-1} \sum_{j=1}^{M-1} (i,j) ,$$

目标的周长简单定义为该目标区域边界像素的个数.具体计算是将预处理得到的二值图像的内部点掏空后 ,计算所剩下的像素个数作为目标的周长.

目标的复杂度定义为目标周长的平方除以目标的面积.

## 3 特征选择与数据归一化

特征选择即从所提取的特征中筛选出被认为对分类有效的特征.其实在模式识别中 ,经常碰到的问题之一就是 Bellman 所谓的“ 维数灾难 ”问题.早在 1978 年 ,Jain 和 Waller 就已经证明了样本个数与最优特征维数之间的关系 ,当样本个数固定时 ,减少分类错误率一个显而易见的办法就是增加新的独立特征 ,但是增加特征超过一定限度会导致分类器性能变坏 ,而且特征维数的增加也意味着计算量和所要存储的信息量的猛增 ,最终表

现为维数组合爆炸和分类器失灵.因此,并不是提取的特征越多越好,只要能正确地将样本分类,特征的维数越少越好.

本文提取 3 个特征:面积、周长和复杂度.很显然,面积与复杂度以及周长与复杂度都是相关的.从直观的角度来讲,面积最能反映玉米象和谷蠹之间的差别,但是也有一些大的谷蠹和小的玉米象的面积可能十分接近;从玉米象和谷蠹的各自身体形态来看,复杂度不失为一个很好的特征,即使某两个样本(玉米象和谷蠹)的面积相近,但其复杂度仍然相差较大,因为复杂度是反映害虫本身形态的一个特征,这一点从所得的实验数据也可以分析出来.因此,为了计算的简便,选取面积和复杂度作为区分玉米象和谷蠹的特征.

数据归一化处理:首先,分别计算出两种害虫的 10 个样本特征的均值向量为:玉米象定为  $w_1$  类,  $w_1 = (1359, 800, 22, 019)$ ;谷蠹定为  $w_2$  类,  $w_2 = (881, 904, 16, 279)$ .以玉米象为基准进行数据归一化,结果为:  $w_1 = (1, 1)$ ,  $w_2 = (0.65, 0.74)$ .

#### 4 分类器设计

模式识别的分类问题就是根据待识客体的特征向量及其他约束条件将其分到某个类别中去.为了把待识模式分配到各自的模式类中去,必须设计出一套分类判别规则.分类器的设计方法很多,本文选用的是一种较为常用的非线性分类器——最近邻分类器.最近邻分类的思想是:若待识模式  $x$  与样本  $x^k$  之间的距离最小,而  $x^k \in w_i$ , 则决策  $x$  属于  $w_i$  类.将  $x$  与  $w_i$  类之间的距离定义为:

$$D_i(x) = \min\{\|x - x^k\|^2\}, x^k \in w_i,$$

则最近邻方法的决策规则为:

若  $D_i(x) < D_j(x)$ ,  $j = 1, 2, \dots, C$ ,  $j \neq i$ ; 则  $x \in w_i$  类.在这里,  $w_1 = (1, 1)$ ,  $w_2 = (0.648, 0.739)$ .

例如:在相同的条件下任意给出几幅待识图像(已知害虫为玉米象或者谷蠹),经过预处理、特征提取和选择后,得到特征向量为:  $x^1 = (893, 16, 942)$ ,  $x^2 = (735, 16, 164)$ ,  $x^3 = (1465, 23, 110)$ ,  $x^4 = (1275, 19, 580)$ .经数据归一化处理后为:  $x^1 = (0.657, 0.769)$ ,  $x^2 = (0.541, 0.734)$ ,  $x^3 = (1.077, 1.049)$ ,  $x^4 = (0.937, 0.889)$ .

进行最近邻分类:  $D_1(x^1) = 0.1710$ ,  $D_2(x^1) = 0.0098$ , 显然  $D_1(x^1) > D_2(x^1)$ , 由决策规则判决为  $x^1 \in w_2$ , 即为谷蠹;  $D_1(x^2) = 0.2814$ ,  $D_2(x^2) = 0.0001$ , 显然  $D_1(x^2) > D_2(x^2)$ , 由决策规则判

决为  $x^2 \in w_2$ , 即为谷蠹;

$D_1(x^3) = 0.0083$ ,  $D_2(x^3) = 0.2845$ , 显然  $D_2(x^3) > D_1(x^3)$ , 由决策规则判决为  $x_3 \in w_1$ , 即为玉米象;  $D_1(x^4) = 0.0163$ ,  $D_2(x^4) = 0.1060$ , 显然  $D_2(x^4) > D_1(x^4)$ , 由决策规则判决为  $x_4 \in w_1$ , 即为玉米象.

对上述 20 个样本图像,最近邻分类器可以很好地完成分类任务,其正确识别率为 100%.现在另外给出几个待识图像,并且先验知识为待识目标既不是玉米象也不是谷蠹,而是赤拟谷盗和扁谷盗.待识目标特征向量  $x^5 = (1520, 16, 216)$ ;  $x^6 = (1523, 16, 809)$ ;  $x^7 = (573, 18, 157)$ ;  $x^8 = (565, 20, 263)$ ; 以玉米象为基准的归一化特征向量为:  $x^5 = (1.118, 0.737)$ ;  $x^6 = (1.120, 0.760)$ ;  $x^7 = (0.421, 0.821)$ ;  $x^8 = (0.416, 0.917)$ .

由数据来看,如果再使用原来的分类器,显然是不行的.比如  $D_1(x^5) = 0.083$ ,  $D_2(x^5) = 0.223$ , 决策为该害虫为玉米象;  $D_1(x^7) = 0.367$ ,  $D_2(x^7) = 0.057$ , 决策为谷蠹,显然是错误的.因此,需要改造分类器,其措施是加阈值限制.经过大量数据分析得出,待识对象若是玉米象或谷蠹,最后得到的  $D_i(x^k)$  的最小值一般都小于 0.05, 因此设定阈值  $T$  为 0.05, 分类器的决策规则修改为:若  $D_i(x) < D_j(x)$ ,  $j = 1, 2, \dots, C$ ,  $j \neq i$ ; 而且  $D_i(x) < 0.05$ , 则  $x \in w_i$  类.在这里,  $w_1 = (1, 1)$  为玉米象,  $w_2 = (0.648, 0.739)$  为谷蠹; 若  $D_i(x) > 0.05$ , 则判别为其它类.

利用修改后的分类判别规则,在上述 20 个样本的基础上添加 5 只赤拟谷盗和扁谷盗,则修改后分类器的正确识别率达到了 95%, 误识率仅为 5%.

上述所有实验数据及实验结果均在相同的实验条件下获得,实验条件如下:摄像头焦距为 7 cm,自然光照射,目标背景为白色,实验对象为玉米象和谷蠹各 10 只,其它害虫(赤拟谷盗和扁谷盗)5 只.

#### 5 结论与展望

这种基于图像处理的害虫检测方法,经试验证明是可行的,对粮仓两种主要害虫玉米象和谷蠹的正确识别率达到了 95% 以上,误识率仅为 5%.该方法不仅扩展了图像识别的应用领域,而且为实现粮虫的在线检测与识别提供了一种新的思路,可以省去传统的人工取样、过筛及肉眼识别

等大量工作,提高了劳动效率.该方法还有助于缩短粮仓的害虫检测周期,准确地预报粮仓虫害的状况,使工作人员能够及时地采取措施.

另外,还进一步研究提取一些其它有利的特征,争取能够识别国家储粮技术规范所规定的十余种主要害虫.设计出基于神经网络具有监督学习功能的非线性分类器,争取能够正确、迅速地识别出各种储粮主要害虫.

参考文献：

[ 1 ] 陈启宗,黄建国.仓库昆虫图册[ M ].北京:科学出版社,1985.

[ 2 ] 万拯群.当前我国科学保粮问题之我见[ J ].粮食储藏,1999,28( 1 ):39-43.

[ 3 ] JAVER A ,BORDEN J H ,PIERCE J H D ,et al. Evaluation of pheromonebaited traps for monitoring of cucujid and tenebrionid beetles in stored grain[ J ].J Econ Entomol ,1990 ,83( 1 ):268-272.

[ 4 ] 赵 奇,田本志.国外用人工合成信息素防治仓储

害虫的研究[ J ].粮食储藏,1999,14( 2 ):52-55.

[ 5 ] 李庆中,汪懋华.农业生物模式识别中的计算机视觉技术[ J ].中国图象图形学学报,1999,4( 7 ):610-616.

[ 6 ] JIMENZ A R ,JANI A K ,CERES R ,et al , Automatic fruit recognition :a survey and new results using range/attenuation images[ J ]. Pattern Recognition ,1999 ,32 :1718-1736.

[ 7 ] MICHAEL A Mullen. Development of a pheromone trap for monitoring Tribolium castaneum[ J ].J Stored Prod Res ,1992,28( 4 ):245-249.

[ 8 ] 吴一全,朱兆达.图像处理中阈值选取方法 30 年( 1962-1992 )的进展( 一 ) [ J ].数据采集与处理,1993,8( 3 ):193-201.

[ 9 ] 吴一全,朱兆达.图像处理中阈值选取方法 30 年( 1962-1992 )的进展( 二 ) [ J ].数据采集与处理,1993,8( 4 ):268-278.

[ 10 ] 刘文萍,吴立德.图像分割中阈值选取方法比较研究[ J ].模式识别与人工智能,1997,10( 3 ):271-277.

[ 11 ] 罗惠韬,章 晋.一个图像分割评价实例及讨论[ J ].数据采集与处理,1997,11( 1 ):18-22.

Research of Feature Extraction and Classification about Stored Grain Insects

XU Fang , QIU Dao - yin , SHEN Xian - zhang

( College of Electrical & Information Engineering ,Zhengzhou University of Technology ,Zhengzhou 450002 ,China )

**Abstract** :This paper tries to use the image processing and pattern recognition technology to extract features of Sitophilus zeamais and Rhizopertha dominica and classify them in the stored grain. According to the morphological feature , we extract area , complexity and perimeter , furthermore , we design the nearest neighbor classifier and fulfill correct classification between Sitophilus zeamais and Rhizopertha dominica and its the recognition rate is up to 95 percent. The execution of this system sets a good and solid foundation for the on - line monitoring system. At the same time , it brings about a new method for species recognition of stored grain insects .

**Key words** :stored grain insects ; image processing ; pattern recognition ; feature extraction ; classifier