

棉花异性纤维图像特征提取*

刘双喜 张 馨 郑文秀 王金星
(山东农业大学机械与电子工程学院, 泰安 271018)

【摘要】 针对棉花加工过程中存在的异性纤维,采用机器视觉技术,通过图像处理方法提取异性纤维目标,采集异性纤维特征数据,应用一种改进型粗糙集理论,进行异性纤维图像目标特征向量的提取,得到有效的特征向量。最后采用决策树理论,利用提取的特征向量进行识别,实验表明,所提取的特征向量对于识别棉花异性纤维是有效的,识别率达到95%。

关键词: 棉花 异性纤维 特征提取 粗糙集 决策树
中图分类号: TP274+.3 文献标识码: A 文章编号: 1000-1298(2010)03-0158-05

Image Feature Extraction of Cotton Foreign Fibre

Liu Shuangxi Zhang Xin Zheng Wenxiu Wang Jinxing
(College of Mechanical and Electronic Engineering, Shandong Agricultural University, Taian 271018, China)

Abstract

For the existence of foreign fibre during cotton processing, machine vision technology and image processing were used in order to not only extract foreign fibre goals, but also collect characteristic data of foreign fibre. Decision tree theory and feature vectors extracted were used to recognize foreign fibre after eigenvectors of aimed foreign images were effectively extracted through an improved version of rough set theory. Experimental results showed that feature vectors extracted from the image of foreign fibre for the identification of cotton foreign fibre was effective and the recognition rate reached more than 95%.

Key words Cotton, Foreign fibre, Feature extraction, Rough set, Decision tree

引言

棉花中的异性纤维是指在原棉生产、加工和流通过程中混入棉花中的对棉花及其制品质量有严重影响的非棉纤维和有色纤维,如化学纤维、毛发、丝、麻、塑料绳、染色线等。异性纤维混入棉花后,不但在纺纱加工过程中难以清除,而且会造成细纱断头,降低工作效率。织布染色后,会在布面出现各种色点,严重影响布面外观质量,最终影响产品的内销及出口。

国内外关于棉花异性纤维剔除系统的研究,主要分为光学检测系统和传感器检测系统两种^[1]。光学检测系统主要是利用高速工业 CCD 相机,借助

光敏元件对棉层加工过程中的棉层成像,然后对图像进行处理得到结果。传感器检测系统是利用电容式传感器和微电子技术进行信息检测,对棉层加工过程中的棉层进行检测得到结果。随着精准化生产要求的不断提高,棉花异性纤维检测设备不但要求能够剔除棉花中的异性纤维,而且要求能够识别出其中各种棉花异性纤维的种类^[2]。这一领域的研究还是空白。

本文通过对棉花异性纤维图像特征进行研究,提出采用改进粗糙集理论作为棉花异性纤维特征的提取方法,并且提取出力矩、似圆度、占空比和长宽比作为棉花异性纤维的识别特征。

收稿日期: 2009-04-09 修回日期: 2009-06-10
* “十一五”国家科技支撑计划资助项目(2006BAD11A14-3)
作者简介: 刘双喜,讲师,主要从事图像处理和模式识别研究,E-mail: lentree@sdau.edu.cn
通讯作者: 王金星,副教授,主要从事现代农业装备研究,E-mail: jinxingw@sdau.edu.cn

1 研究对象

选择黑色塑料片(每个样本 50 mg)、有色线(以丝绸作为代表,每个样本 100 mg)、麻绳(每个样本 100 mg)、头发(每个样本 10 mg)、红色丙纶丝(每个样本 50 mg)和羽毛(每个样本 20 mg) 6 种典型异性纤维作为研究对象,每种异性纤维共生成 10 个样本。采集图片时选择了足量的不含异性纤维皮棉,利用皮棉开松机对皮棉进行充分开松,形成 80 cm 宽、2 mm 厚的不间断均匀棉层,并在开松过程中逐次放入上述异性纤维样本,并将开松后含有异性纤维的棉层保存下来,利用 DALASA 线扫描相机对这些棉层样本进行图像扫描,硬件环境如图 1(结构简图)所示,每个样本大约生成 400 张4 000 × 500 大小的 24 bits 位图图像,以这些图片作为研究对象。原始图像示例如图 2 所示。

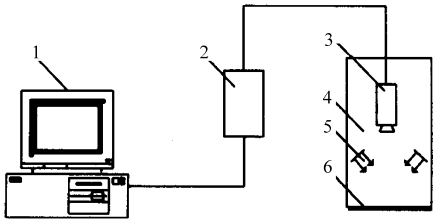


图 1 图像采集设备结构简图

Fig. 1 Structural diagram of image acquisition equipment
1. 计算机 2. 图像采集卡 3. CCD 相机 4. 光照箱 5. 光源
6. 棉层



图 2 原始羽毛图像
Fig. 2 Original image of feather

2 特征提取

图像特征提取是进行图像识别的关键环节。目前常用的图像特征包括形状、色彩、角点、纹理和力

矩等^[3]。棉花异性纤维由于种类繁多,颜色多样。需要对图像进行前期处理,包括图像滤波和彩色目标提取。

2.1 图像前期处理

图像前期首先完成图像滤波处理。由于图像采集环境干扰因素较少,试验采用了 3 × 3 的中值滤波技术对图像进行处理。

然后进行图像彩色目标提取,即提取出棉花异性纤维目标。此处采用边缘轮廓模版法进行彩色目标提取^[4]。具体实施步骤如下:首先对图像进行自适应阈值分割技术,得到二值化图像;在提取出的二值图像基础上,进行轮廓提取,采用掏空内部点的方法进行;在提取出异性纤维轮廓的基础上,采用邻域搜索法再提取出每个异性纤维轮廓目标;最后通过彩色原始图像与二值化轮廓图像进行比对,提取出彩色异性纤维目标。

通过以上几个步骤,得到棉花异性纤维的轮廓图像和彩色目标图像,羽毛效果图如图 3、4 所示。所有样本中提取出的彩色目标数目如表 1 所示。

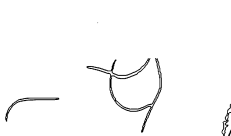


图 3 羽毛轮廓图像
Fig. 3 Outline image
of feather

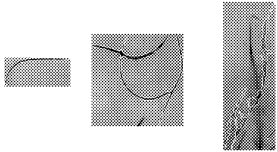


图 4 羽毛彩色目标图像
Fig. 4 Colorful object
image of feather

2.2 异性纤维特征提取

在进行图像前期处理过程中,可以得到大量异性纤维的数据信息,包括异性纤维轮廓周长、轮廓面积、轮廓角点、目标图像长度和宽度、彩色异性纤维目标图像的色彩信息(以 RGB 模型表示)等。由于数据量大,很难发现可以利用的特征,在此提出一种基于改进型粗糙集理论的数据分析方法,用以提取有效异性纤维特征。

表 1 各种异性纤维数目
Tab.1 Number of all foreign fibres

样本	头发	麻绳	羽毛	蓝色丝绸	红色丝绸	丙纶丝	塑料片
数目	20	1 568	1 478	589	605	27	231

2.2.1 棉花异性纤维粗糙集理论

粗糙集(RS)理论是由波兰学者 Pawlak Z 在 1982 年^[5]提出的。1991 年 Pawlak Z 系统全面地阐述了 RS 理论,奠定了严密的数学基础。粗糙集理论的出发点在于:根据目前已有的对给定问题的知识将问题的论域进行划分;然后对划分后的每一个组成部分确定其对某一概念的支持程度,即肯定支持此概念、肯定

不支持此概念和可能支持此概念。基本粗糙集包括两方面的内容,即 RS 知识表示方法和粗糙集约简^[6]。

2.2.2 异性纤维 RS 知识表示方法

基于 rough sets 理论的知识发现,主要是借助于信息表这样一种有效的数据表知识表达方式。信息表知识表达系统的基本成分是研究对象的集合。关于这些对象的知识是通过指定对象的属性(特征)

和它们的属性值(特征值)来描述的。一般地,一个信息表知识表达系统 s 可以表示为一个四元组 $s = (U, R_1, V, f)$ 。这里, U 是对象的集合,也称为论域; R_1 是属性的非空有限集合; $V = \cup \{a \mid a \in A\}$, V 表示属性 a 的值域; $f: U \times R' \rightarrow V$ 是一个信息函数,它指定 U 中每一个对象 x 的属性值。在粗集理论中,知识被认为是一种分类能力,即根据事物的特征差别将其分门别类的能力。分类过程中相差不大的个体被归于同一类,它们的关系就是不可分辨关系,即等价关系。其定义为:对于每个属性子集 $B' \in R_1$,本文定义一个不可分辨关系(indiscernibility relation) $IND(B')$,即显然, $IND(B')$ 是一个等价关系(满足自反性、对称性和传递性)。

粗糙集理论的不确定性是建立在上、下近似的概念上的。给定知识表达系统 $s = (U, R_1, V, f)$,对于每个子集 $X \in U$ 和不可分辨关系 B' , X 的上近似(upper approximation)和下近似(lower approximation)分别定义如下: X 的上近似 $B'(X) = \cup \{Y_i \mid Y_i \in U \mid IND(B'), Y_i \cap X \neq \emptyset\}$; X 的下近似 $B'(\underline{X}) = \cup \{Y_i \mid Y_i \in U \mid IND(B'), Y_i \subseteq X\}$ 。

集合 $BNB(X) = B' - (X) - B' - (\underline{X})$ 称为 X 的 B' 边界(boundary); $POSB(X) = B' - (\underline{X})$ 称为 X 的 B' 正域(positive region); $NEGB(X) = U - B'(X)$ 称为 X 的 B' 负域(negative region)。 $B'(X)$ 是根据知识 B' (属性子集 B')、 U 中所有一能归入集合 X 的元素构成的集合,即所有包含于 X 的基本集 Y_i 的并; $B'(\underline{X})$ 是根据知识 B' 、 U 中所有一能和可能归入集合 X 的元素构成的集合,即所有与 X 的交不为空集的基本集 Y_i 的并; $BNB(X)$ 是根据知识 B' 、 U 中既不能肯定归入 X ,又不能肯定归入集合 X 的元素构成的集合;正域 $POSB(X)$ 是根据知识 B' 、 U 中所有能确定一定归入 X 的元素的集合;负域 $NEGB(X)$ 是根据知识 B' 、 U 中所有不能确定一定归入 X 的元素的集合;边界域 $BNB(X)$ 是某种意义上论域上的不确定域,边界域中的元素既不能肯定地属于集合 X ,也不能肯定地属于 X 。

集合的不精确性正是由于边界域的存在而引起的。集合的边界域越大,其精确性则越低。如果 $BNB(X)$ 是空集,则 X 关于 R_1 是清晰的(crisp),称集合 X 为精确集;如果 $BNB(X)$ 不是空集,则称集合 X 为关于 R_1 的粗糙集(rough sets)。

2.2.3 RS 约简理论

RS 约简理论是粗糙集的核心内容。其主要思想是:保持相对分类能力不变的条件下,删除冗余的、不必要的属性或属性值,达到知识简化的目的。知识库中的知识(属性)并不是同等重要的,甚至其

中某些知识是冗余的。所谓知识约简,就是在保持知识库分类能力不变的条件下,删除其中不相关或不重要的知识。

令 R_1 为由特征集定义的等价关系族。对于任一属性 $a \in R_1$,如果 $IND(R_1) \neq IND(R_1 - \{a\})$,则称 a 为 R_1 中不可省略的(indispensable);否则 a 为 R_1 中可省略的(dispensable)。 R_1 中所有不可省略关系的集合称为 R_1 的核,记为 $CORE(R_1)$,即 $CORE(R_1) = \cap RED(R_1)$ 。这里 $RED(R_1)$ 是 R_1 的所有简化族。核中的属性是影响分类的重要属性。

综上所述,粗糙集理论的基本框架可归纳为:以不可分辨关系划分所研究论域的知识,形成知识表达系统;利用上、下近似集逼近描述对象,通过知识约简,从而获得最简知识。

2.2.4 棉花异性纤维的粗糙集知识表示表

将提取出来的所有数据加入到知识表示表中,形成棉花异性纤维知识表。知识表示表包括的异纤特征有周长 L ,面积 S ,异纤宽度,异纤长度,异纤三通道色彩均值 R 均值、 G 均值和 B 均值以及异纤轮廓的角点数目(Num)。

2.2.5 棉花异性纤维的粗糙集知识约简

根据约简和核的概念,粗糙集理论提供了分析多余属性的方法,对知识的处理通过对决策表中属性值的处理实现。提出一种改进型 RS 约简方法。具体步骤如下:删除重复的实例;组合或者删除多余的属性列;组合或者删除每个实例多余的属性值;求出最小约简;根据最小约简,求出逻辑规则。

从知识表示表中可以发现,不存在重复实例,每种异性纤维互相独立;单纯的面积、周长、目标长度、宽度都不具有分类属性特征,在此对其进行组合,发现可以提取出有效的特征属性值,分别是力矩、占空比、似圆度和外观比;色彩信息属性值较低,可以作为局部小区域高属性值;删除角点属性列。

通过组合发现:轮廓力矩与面积和周长平方之比有直接关系,所以提出采用面积 S 与周长 L 平方之比作为新的分类属性,力矩定义为 $d_m = S/L^2$ 。力矩分布图如图 5 所示,其中 tfi 代表头发样本,blsi 代表丙纶丝样本,bti 代表布条样本,msi 代表麻绳样本,slbi 代表塑料片样本,ysxi 代表有色线样本,cdi 代表磁带样本,jmi 代表羽毛样本, i 为样本数。

从图 5 中可以看出:头发的力矩分布在 0.04 ~ 0.06 之间。丙纶丝、麻绳、有色线、鸡毛和布条的力矩分布在 0.06 ~ 0.15 之间,磁带和塑料片的力矩分布在 0.15 ~ 0.225 之间。分布图呈现三段式分布,对于每一个等级内部的分类可以通过其他特征进行区分,分别对应丝类、条类和片类 3 个等级。力矩可

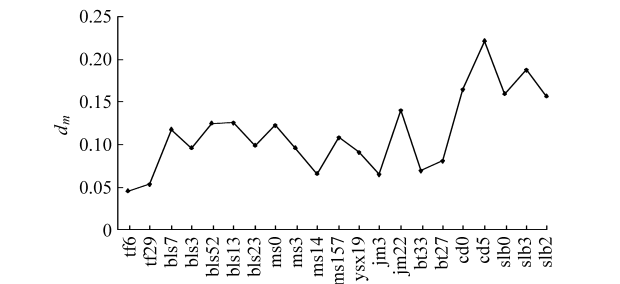


图5 轮廓力矩分布图

Fig.5 Map of contour moments distribution

以作为分类核的首要属性。

占空比 Z 反映目标区域的充实度,反映了目标在提取出的图像内部的占空大小,体现的是目标的肥瘦度。似圆度 S_s 反映的是目标的圆形程度,观察

异性纤维图片可以发现,片类异性纤维,如塑料布、纸片等的圆形度最好;条类异性纤维,如布条、麻绳等的圆形度较差;而丝类的异性纤维,如头发等的圆形度最差。外观比 W 反映的是目标轮廓的矩形程度,由于异性纤维形状的复杂性,从外观上很难区分异性纤维的种类,外观比在全局不具有分类属性,只可以作为低属性特征。 R 、 G 、 B 均值代表异性纤维的色彩信息,其在全局不具有分类属性,但是在局部具有一定的区分度,只能作为低属性进行分类。角点数目属于低层次的图像处理,由于棉花异性纤维图像的复杂性,在相同搜寻距离内得到的角点均值进行分类的属性,所以删除角点数一列。得到的以各属性中心点为代表的异性纤维属性特征表如表2所示。

表2 异性纤维属性特征
Tab.2 Feature of all foreign fibres

异性纤维样本	轮廓力矩 d_m	外观比 W	占空比 Z	似圆度 S_s	R 均值	G 均值	B 均值
头发	0.020 6	2.921 7	0.060 0	386	204	206	192
麻绳	0.115 0	3.987 4	0.189 6	269	195	207	203
羽毛	0.075 0	5.635 4	0.158 9	153	121	116	98
红色丙纶丝	0.075 8	3.489 0	0.456 8	226	214	207	198
蓝色丝绸	0.098 0	5.980 7	0.198 6	257	203	222	225
红色丝绸	0.101 5	3.370 4	0.175 0	248	205	226	227
黑色塑料布	0.168 9	4.509 3	0.780 0	16	208	224	226

3 分类方法设计

从特征属性表可以得出,属性集合 $R' = \{d_m, \{W, Z, S_s, \{R, G, B\}\}\}$,其中 d_m 作为首要属性值,可以将异性纤维分为片类、丝类和条类; $\{W, Z, S_s\}$ 属性值可以作为每一类内部进行的首要属性值,也是全局分类属性的第2层; $\{R, G, B\}$ 属性值可以作为细节属性,用以缩小各个分类属性的上下限,以达到更加准确的分类结果。由此提出基于决策树的分类思想,结合粗糙集数据表中的异性纤维特征进行。通过大量实验进行优化组合得到图6所示的分类流程图。“其他”是不符合异性纤维特征的个别异物。

4 实验结果

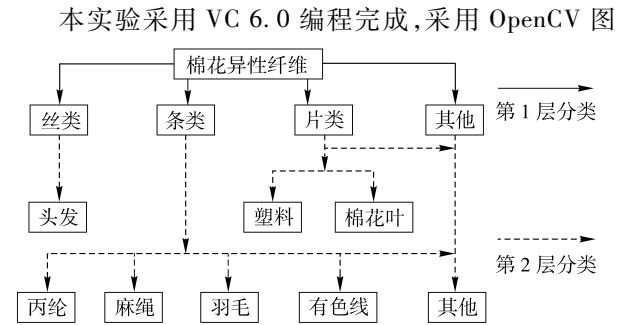


图6 决策树识别图

Fig.6 Recognition map of decision tree

像开源处理库进行处理,分类结果如表3所示。

表3 识别结果
Tab.3 Results of recognition

异性纤维	实际数目	识别数	识别率/%
头发	20	20	100
麻绳	1 568	1 483	94.6
羽毛	1 478	1 429	96.7
红色丝绸	589	573	97.3
蓝色丝绸	605	591	97.6
丙纶丝	27	26	96.3
塑料片	231	230	99.5

从表3可以看出:提取出来的特征向量进行识别是有效的,异性纤维的平均识别率达到了95%以上,对于头发和塑料片的识别率更高,主要是由于头发和塑料片的特征向量聚集度较高,与别的异性纤维之间交集范围较小所致。由于棉花异性纤维种类繁多,样式各异,对于其他异性纤维的识别工作还有待于进一步研究。

5 结束语

本文提出使用粗糙集理论作为棉花异性纤维特征量向量的提取方法,得到了以异性纤维的轮廓力矩作为棉花异性纤维分类的标准,结合决策树方法,

完成了对棉花异性纤维的分类工作。通过对棉花异性纤维样本图像进行实验,分类平均准确率可达到 95%。实验结果表明,该特征向量和分类标准是有效准确的。

参 考 文 献

- 1 金守峰,张慧,冯涛. 基于计算机视觉的棉花异性纤维检测识别算法[J]. 中国棉花加工,2008(2):23~24.
Jin Shoufeng, Zhang Hui, Feng Tao. An inspecting identification algorithm for foreign fibers based on machine vision[J]. China Cotton Processing, 2008(2):23~24. (in Chinese)
- 2 李玉军. 棉花异性纤维在线检测与剔除系统研究与实现[D]. 西安:西安理工大学,2005.
Li Yujun. Studying and realization of a cotton isomerism fibre online measuring and eliminating system[D]. Xi'an: Xi'an University of Technology, 2005. (in Chinese)
- 3 林开颜,吴军辉,徐立鸿. 彩色图像分割方法综述[J]. 中国图像图形学报,2005,10(1):1~9.
Lin Kaiyan, Wu Junhui, Xu Lihong. A survey on color image segmentation techniques[J]. Journal of Image and Graphics, 2005,10(1):1~9. (in Chinese)
- 4 郑文秀,刘双喜,王金星,等. 棉花异性纤维图像分割方法研究与实现[J]. 中国棉花加工,2008(4):24~26.
Zheng Wenxiu, Liu Shuangxi, Wang Jinxing, et al. Study and realization of image segmentation on the cotton foreign fibers [J]. China Cotton Processing, 2008(4):24~26. (in Chinese)
- 5 吴成东,许可,韩中华,等. 基于粗糙集和决策树的数据挖掘方法[J]. 东北大学学报:自然科学版,2006,29(5):481~484.
Wu Chengdong, Xu Ke, Han Zhonghua, et al. Approach to data mining based on rough sets and decision tree[J]. Journal of Northeastern University: Natural Science, 2006, 29(5):481~484. (in Chinese)
- 6 吴成东,许可,王欣,等. 软计算方法在数据挖掘中的应用[J]. 计算机测量与控制,2005(3):294~297.
Wu Chengdong, Xu Ke, Wang Xin, et al. Soft computing and its application in data mining[J]. Computer Measurement and Control, 2005(3):294~297. (in Chinese)
- 7 李庆中,张漫,汪懋华. 基于遗传神经网络的苹果颜色实时分级方法[J]. 中国图像图形学报,2000,5(9):779~783.
Li Qingzhong, Zhang Man, Wang Maohua. Real-time apple color grading based on genetic neural network[J]. Journal of Image and Graphics, 2000, 5(9):779~783. (in Chinese)
- 8 杨文柱,李道亮,魏新华,等. 棉花异性纤维图像分割方法[J]. 农业机械学报,2009,40(3):193~198.
Yang Wenzhu, Li Daoliang, Wei Xinhua, et al. Realization of image segmentation on the cotton foreign fibers [J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(3):193~198. (in Chinese)
- 9 王茜蓓,彭中,刘莉. 一种基于自适应阈值的图像分割算法[J]. 北京理工大学学报,2003,23(4):521~524.
Wang Qianqian, Peng Zhong, Liu Li. An adaptive method of image segmentation[J]. Transactions of Beijing Institute of Technology, 2003, 23(4):521~524. (in Chinese)
- 10 Ying Xu, Victor Olman, Histon Smith, et al. A segmentation algorithm for noisy images: design and evaluation[J]. Pattern Recognition Letters, 1998, 19(13):1213~1224.
- 11 王长军,朱善安. 基于 Mean Shift 的目标平移与旋转跟踪[J]. 中国图像图形学报,2007,12(8):1367~1371.
Wang Changjun, Zhu Shan'an. Mean Shift based targets' rotation and translation tracking[J]. Journal of Image and Graphics, 2007, 12(8):1367~1371. (in Chinese)
- 12 沈明霞,李秀智,姬长英. 水果品质检测中的模糊阈值分割方法[J]. 农业机械学报,2003,34(5):113~115.
Shen Mingxia, Li Xiuzhi, Ji Changying. Fuzzy threshold segmentation method in quality detection of fruit[J]. Transactions of the Chinese Society for Agricultural Machinery, 2003, 34(5):113~115. (in Chinese)
- 13 Remco C Veltkamp. Shape matching: similarity measures and algorithms[C]//SMI 2001 International Conference on Shape Modeling and Applications, 2001:188~197.
- 14 孙仲康,沈振康. 数字图像处理及其应用[M]. 北京:国防工业出版社,1985.
- 15 Marr D. Theory of edge detection[M]. Proc. Roy. Soc., 1980, B27:187~217.
- 16 李赵红,侯建军,宋伟. 基于等级结构的二值文本图像认证水印算法[J]. 自动化学报,2008,34(8):841~848.
Li Zhao hong, Hou Jianjun, Song Wei. Binary document image authentication watermarking technique based on hierarchical structure[J]. Acta Automatica Sinica, 2008, 34(8):841~848. (in Chinese)
- 17 郭俊先,应义斌. 皮棉中杂质检测技术与检出装备的研究进展[J]. 农业机械学报,2008,39(7):107~113,106.
Guo Junxian, Ying Yibin. Progress on detecting technique and sorter of raw cotton foreign matters[J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(7):107~113, 106. (in Chinese)