

DOI: 10.7652/xjtuxb201409001

光纤缺陷实时检测与分类方法研究

刘小勇, 郑琨

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 针对人工检测高速运动光纤表面缺陷的效率低、准确性差、难以实时检测等问题,设计实现了一套基于机器视觉的快速光纤缺陷检测系统。定义光纤缺陷,建立分类数据库和分类标准;设计全方位数据采集系统自动连续获取光纤表面图像信息,输入工控机进行处理;提取光纤目标区域,获得光纤缺陷形态学特征数据;针对光纤缺陷特点和 AdaBoost 分类器的优缺点,设计了一种改进的基于形态特征的 AdaBoost 级联分类器用于光纤缺陷检测与分类,实现了光纤质量的实时监控。最后,将改进算法与标准 AdaBoost 算法在实际工业环境下进行对照实验,实验数据表明,改进算法的准确率达到 99% 以上,同时能大幅减少检测耗时,证明了所设计的检测系统具有很好的实时性和准确性。

关键词: 光纤缺陷;快速检测;形态特征;AdaBoost 级联分类器

中图分类号: TP274 **文献标志码:** A **文章编号:** 0253-987X(2014)09-0001-05

Research on Optical Fiber Defect Real-Time Detection and Classification

LIU Xiaoyong, ZHENG Kun

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A real-time optical fiber defect detection system based on machine vision is designed to solve the problem that manual method is inefficient and imprecise in detecting fast moving fiber. Fiber defects are defined to establish a database and criterions for classification. Continuous fiber images captured automatically by 3 industrial cameras are transmitted to IPC for classification, and the IPC obtains morphological characteristics data of defects in fiber target after image processing. An advanced AdaBoost cascade classifier based on morphological characteristics is designed to analyze the defect image features. The advanced AdaBoost method and the classical AdaBoost method are tested under industry condition, and detection results show that the detection accuracy of the system with advanced AdaBoost is over 99% but the system with advanced AdaBoost takes much less time. It can be concluded that the proposed detection method has good real-time performance and a high detecting accuracy.

Keywords: optical fiber defect; fast detection; morphological characteristics; AdaBoost cascade classifier

光纤制导弹所用线包在绕制之前必须进行光纤表面缺陷检测,以保证其完整性。目前通常依靠人工检测:检查者观察采集系统拍摄到的光线表面图像,人工判定是否存在缺陷。检查者需要连续工

作,容易疲倦,导致检查效率低下,且人工检测难以保证标准一致,产品质量不稳定。线包绕制过程中,光纤行进速率通常为 1~1.5 m/s,人眼在这样的速度下很难看清光纤上的缺陷,人工检测只能通过降

低绕制速度提高检测准确率,但这样不利于实现光纤的全自动绕制。因此,设计一种非接触式的、实时可靠、成本低、自动化程度高的光纤缺陷检测分类方法具有重要的现实意义,而机器视觉检测技术为该问题提供了可行的解决方案。

机器视觉检测技术利用图像采集系统获取图像,采用图像处理算法提取图像中的关键信息,运用各种分类识别算法检测出目标。提取信息的方法主要分为两类:一类是频域分析法,贡玉南等^[1]设计 Gabor 滤波器检测织物疵点,此类方法适合具有一定规律纹理的图片;另一类是分析空间域检测图像中的形态学信息法,胡涛等^[2]采用联通区域获取 PCB 板目标区域,采用轮廓匹配算法识别 PCB 缺陷,这种方法适合标准形状目标中缺陷的检测。分类识别算法有支持向量机、神经网络分类器、主成分分析等算法。这些方法多面临计算量大或者训练周期较长的问题,故不适应实时性要求高的系统。

缺陷检测的问题关键在于目标信息的提取和识别算法的设计,要根据实际问题的特点设计合适的方法。本文中需要检测的光纤缺陷主要有黑斑、表面损伤和线径异常 3 种,这些缺陷会造成光纤断裂和绕制不均匀。黑斑是由于光纤加强层制造过程中掺入杂质所致,呈现不规则斑点形状;表面损伤是由于加强层制造完成后受到破坏所致,严重的情况会损伤到光纤纤芯,呈现不规则边缘凹陷和凸起;线径异常是由于加强层过厚或者过薄造成,异常部位的直径大于或者小于光纤标准直径。根据这些缺陷的特点,提取了缺陷目标形态特征,设计了级联的 AdaBoost 分类器对其进行分类。AdaBoost 是一种迭代算法,可以通过组合多个弱分类器形成高精度的强分类器,对弱分类器的精度要求低。对于系统实时性要求高的特点,本文设计了一种改进的 AdaBoost 算法,用形态学特征值代替 AdaBoost 常用的矩形特征值,大幅减少特征值数量从而降低分类器计算量,减少分类器耗时,同时级联多个分类器来减少平均检测时间,最终满足系统实时性要求。

1 光纤检测系统硬件结构

光纤缺陷检测系统由图像采集系统、速度检测模块和工控机组成。数据采集系统结构图如图 1 所示,包括光源、高放大倍数的镜头和面阵工业相机。光纤速度检测模块由基于 PCI 总线的编码器、正交脉冲计数卡及编码器组成。

3 个工业相机安装在垂直于光纤的同一个平面

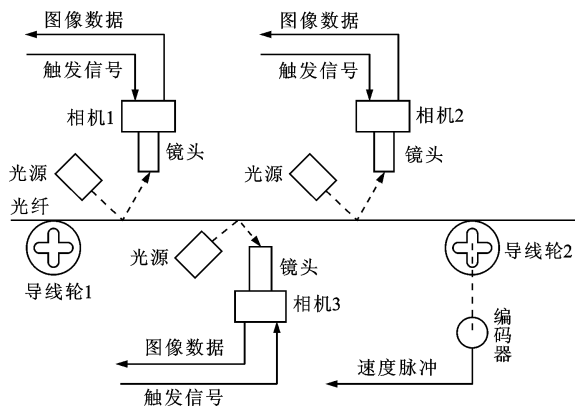


图 1 光纤缺陷检测图像采集系统示意图

上,互成 120° 角,同时采集 3 路图像,以便获取完整的光纤表面信息。系统工作时,光纤在导轮上运动,带动编码器旋转,编码器输出信号进入脉冲计数卡,工控机通过该卡板获得光纤位置和速度信息,光纤运行一帧图像实际长度后,板卡输出相机触发信号,工业相机获得触发信号抓取图像并传回给工控机,工控机读取数据进行缺陷检测。如果检测到缺陷,工控机将图像数据和缺陷信息保存到数据库中,并且显示缺陷信息到人机界面上。缺陷信息包括缺陷类型、缺陷程度、缺陷在光纤的位置、光纤速度和光纤实测线径等数据。

2 缺陷检测与分类算法

2.1 检测算法性能需求分析

系统要求能够识别出大于 $0.2\text{ mm} \times 0.2\text{ mm}$ 的缺陷,最低分辨率应满足一个像素对应 0.04 mm^2 。为便于精确检测,保留一定的冗余量,一个像素对应 $4 \times 10^{-4}\text{ mm}^2$,即最小的黑斑缺陷对应 100 个像素。采用分辨率为 640×480 的相机,其水平视场的极限为 12.8 mm 。光纤的运动速度最大为 1.5 m/s ,为保证拍摄到光纤表面所有缺陷,图像帧率为 139。系统采用 3 个相机同时工作,1 s 内需要处理的图像数量为 417 帧,每帧图像的处理时间为 2.4 ms 。缺陷检测分类算法既要保证检测分类的准确性,又要满足系统实时性的要求。

2.2 目标区域提取

采集到的图片普遍存在边缘模糊、对比度偏弱、目标区域难以提取等问题,因此采用方形模板的中值滤波器对图像进行滤波,并进行对比度线性展宽和 γ 校正相处理,从而提高光纤图像的清晰度和对比度;然后采用最大类间差法^[3]选择分割阈值来区分背景,提取目标区域,同时对图像进行闭运算等一

系列形态学处理,消除目标边缘毛刺内部细小孔洞,并根据目标区域的面积、长度等信息最终找到图片中的光纤目标。

2.3 基于矩形特征的 AdaBoost 分类器检测算法

(1) AdaBoost 算法^[4]通过组合多个弱分类器构造出高精度强分类器,其基本原理是:对样本集进行训练,若样本没有被当前弱分类器正确分类,则在训练下一个弱分类器时该样本被选入训练集的概率增加;若被正确分类,则概率减小。算法能够聚焦于更多信息的样本上,从而提高分类的准确度。

假设 X 表示样本空间, Y 表示样本标识集合, $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 表示训练样本集合, $x_i \in X, y_i \in Y, i = 1, 2, \dots, N$ 。AdaBoost 算法如下^[5]:

{初始化样本的权值,

$$\forall (x_i, y_i) \in S, D_1(x_i, y_i) = \frac{1}{N}$$

for $t = 1, 2, \dots, T$ (T 表示迭代过程的轮数)

①根据样本权值 D_t , 调用弱学习算法获得弱分类器 $h_t: X \rightarrow Y$;

$$\textcircled{2} \text{ 计算 } \epsilon_t = \sum_{\forall y_i \neq h_t(x_i)} D_t(x_i, y_i),$$

$$\text{如果 } \epsilon_t < 0.5, a_t = \frac{1}{2} \ln[(1 - \epsilon_t) / \epsilon_t],$$

如果 $\epsilon_t \geq 0.5$, 终止算法;

③计算样本权值

$$D_{t+1}(x_i, y_i) = \frac{D_t(x_i, y_i) e^{-a_t y_i h_t(x_i)}}{Z_t}, Z_t \text{ 是归一}$$

化系数,使 $\sum_{i=1}^N D_t(x_i, y_i) = 1$, 获得强分类器 $H(x)$

$$= \text{sign}\left(\sum_{t=1}^T a_t h_t(x)\right);$$

a_t 表示第 t 轮产生的弱分类器 $h_t(x)$ 的权重,由第 t 轮分类错误的样本权值和 ϵ_t 决定, ϵ_t 越小表示 $h_t(x)$ 越重要,权重越大。}

Freund 等首先从理论上推导出了 $H(x)$ 的训练误差边界^[5-6]

$$\begin{aligned} \epsilon &= \frac{1}{N} \sum_{i=1}^N [H(x_i) \neq y_i] \leq \\ &\frac{1}{N} \sum_{i=1}^N \exp(-y_i f(x_i)) = \prod_{i=1}^T z_i \end{aligned} \quad (1)$$

由式(1)可知,增加迭代次数和弱分类器数量,可以降低总体分类误差上限,提高强分类器对样本的分类精度。

(2) AdaBoost 分类器通常采用矩形特征^[7-8],常

用的 5 个矩形特征模板如图 2 所示。



图 2 AdaBoost 分类器常用的矩形特征模板

矩形特征值为黑色区域的像素灰度值的和与白色区域的像素灰度值的和的差值。假设图像大小为 $W \times H$, 特征模板大小为 $w \times h$, 那么样本图像产生的特征数量可由下式计算

$$XY \left(W+1-w \frac{X+1}{2} \right) \left(H+1-h \frac{Y+1}{2} \right)$$

式中: $X = \lfloor \frac{W}{w} \rfloor, Y = \lfloor \frac{H}{h} \rfloor$, $\lfloor \cdot \rfloor$ 表示向下取整。

对于图 2 所示的 5 种特征模板, 20×20 的图像有 88 460 种矩形特征, 矩形特征的数量远远超出了图像的像素数。

(3) Adaboost 分类器具有很强的泛化能力, 通过增加弱分类器个数和迭代次数能够达到很高的准确性, 但是 Adaboost 分类器也有不足之处: 训练周期长, 样本数据多, 需要对大量特征值多轮迭代生成多个弱分类器, 耗费大量时间; 分类耗时长, 获取特征值计算量大, 大量的特征值导致强分类器中包含大量弱分类器, 分类计算复杂, 难以实现实时检测。

2.4 改进的 AdaBoost 级联分类器

本文根据实际情况, 提出用具有实际物理意义的形态特征值代替常用的矩形特征值的 AdaBoost 级联分类器方法。该方法可以大幅减少特征值数量, 减少了分类器的计算量, 满足了系统的实时性要求, 并且所有的特征值在缺陷上有对应的含义, 操作人员可以在人机界面输入参数来干预分类过程。

(1) AdaBoost 算法中需要大量训练样本, 样本的数量和质量直接关系到算法的性能。本文的样本数据全部采用实际环境中不同角度拍摄的光纤图像, 以确保训练样本和实际检测样本的一致性。

(2) 光纤图像通过处理可以获得带有实际物理意义的特征值。对所有需要检测的光纤缺陷进行粒子分析, 获取粒子面积、粒子周长、孔的个数、孔的面积、粒子的边界信息等数据。这些数据作为分类器的特征值能够准确描述光纤和缺陷的形态特征。

(3) 采用阈值分类作为弱分类方法。训练弱分类器就是根据当前特征和当前权重分布, 确定对应的最优阈值 θ^* , 使得这个弱分类器对所有训练样本的分类误差最低。采用从累积直方图中查找阈值 θ 的方法^[9] 求解 θ^* , 具体操作步骤是: 计算所有训练

样本的特征值,统计每个特征值对应下的缺陷图片占总样本数目的比例和正常图片占总样本数目的比例,进一步获取缺陷图片累积数量和正常图片累积数量占图片总数的比例,如图3所示, $f(x)$ 代表特征值, S 代表图片累积数量占图片总数的比例,在特征值阈值 θ 下, S_f 是判断为缺陷的图片数占样本总数的比例, S_n 是判断为正常图片的数目占样本总数的比例。

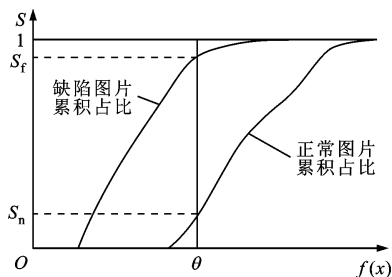


图3 最优阈值计算方法

该阈值分类器的分类误差为

$$e = 1 - S_f + S_n \quad (2)$$

计算所有的特征值 $f(x)$ 对应的分类误差 e ,找出最小分类误差所对应的阈值,该阈值就是最优分类阈值 θ^* ,利用这个最优阈值生成一个弱分类器。

(4)通过 AdaBoost 分类器级联研究实际光纤样本图片发现,在所有图片中缺陷图片所占比例较小,低于10%,具有明显的正负样本分布不平衡、不对称的特点,这样的特点适合采用级联分类器^[10]。在图4所示的级联分类器中,每个分类器的分类对象输入是前一级分类器的未识别对象输出,已经识别的对象不再进入下一级分类器,各级分类器通过输入不同的特征向量和内部运算对分类对象进行分类。

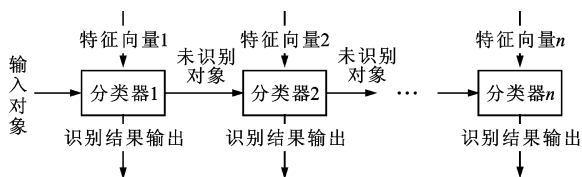


图4 级联分类器结构

假定第 i 个分类器的识别率、拒识率和误识率分别用 r_i 、 δ_i 和 m_i 来表示,则整个系统的识别率

$$R_n = r_1 + r_2(1 - r_1 - m_1) + \dots + r_n(1 - r_1 - m_1) \dots (1 - r_{n-1} - m_{n-1}) = \sum_{i=1}^n r_i \prod_{j=1}^{i-1} \delta_j \quad (3)$$

整个系统的拒识率

$$\Delta_n = \prod_{j=1}^n \delta_j \quad (4)$$

整个系统的误识率

$$M_n = \sum_{i=1}^n m_i \prod_{j=1}^{i-1} \delta_j \quad (5)$$

式中: $\delta_j = 1 - r_j - m_j$ 。

由式(3)知,增加分类器级数,可以提高识别率。根据光纤缺陷的特点,设计了如图5所示的 AdaBoost 3级分类器。第1级分类器算法简单,采用目标粒子数据作为特征值,可以过滤大部分正常图像,大大减轻了2、3级分类器的负荷,减少了平均检测时间。第2级分类器采用边缘信息作为分类器特征值,第3级分类器采用线径大小作为分类器特征值。2、3级分类器特征值数量增多,包含较多弱分类器,可对缺陷进行精确划分。

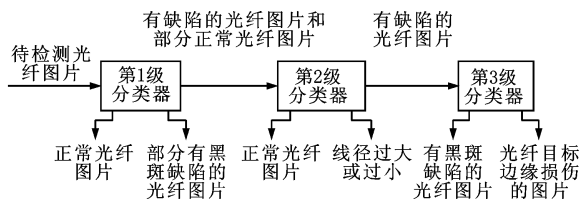


图5 光纤缺陷检测3级分类器

针对 AdaBoost 分类器只能进行正负样本分类的问题,本设计将多分类问题转化为多个二分类问题,每1级分类器中有2个并联的 AdaBoost 分类器。

3 测试结果

本文设计的光纤缺陷检测系统运行平台为 Intel i7-2600 处理器,主频 3.4 GHz、8 GB 内存的工控机。机器视觉图像方法在光纤缺陷检测中鲜有应用,作为对照,本文在该平台采用常用的矩形特征值的标准 AdaBoost 算法进行测试。系统要求检出大于 0.2 mm×0.2 mm 的黑斑缺陷,宽度大于 0.2 mm 的外表损伤缺陷,超出光纤直径 5% 的线径异常缺陷。

2种方法经长时间循环测试,获取了大量实验数据,如表1和表2所示,由于训练样本和实际使用环境中的检测样本一致性较高,缺陷形态种类完善,所以2种算法对3种缺陷的检测率都达到了97%以上,改进算法达到99%以上。改进的 AdaBoost 算法对外表损伤缺陷有相对较多的误检是由于光纤上的粘结剂碎屑干扰了检测。标准算法对黑斑缺陷和外表损伤缺陷的检测率高于改进算法,并且有更低的误检率,但是对线径异常缺陷的检测率低于改进算法。这是由于线径异常缺陷较难在矩形特征中体现出来,而采用形态学特征很容易检测这样的缺陷。

表 1 改进的 AdaBoost 分类器检测结果统计

缺陷类型	总数量	检出数量	误检数量	检测率 /%	误检率 /%
黑斑缺陷	5 000	4 973	0	99.46	0.00
外表损伤	2 000	1 982	13	99.10	0.65
线径异常	500	499	1	99.60	0.20

表 2 标准 AdaBoost 分类器检测结果统计

缺陷类型	总数量	检出数量	误检数量	检测率 /%	误检率 /%
黑斑缺陷	5 000	4 986	2	99.72	0.04
外表损伤	2 000	1 992	4	99.60	0.20
线径异常	500	485	0	97.00	0.00

2 种算法进行图像处理与缺陷检测的耗时如表 3 和表 4 所示。改进的 AdaBosst 算法处理与检测正常光纤图像的平均耗时为 1.28 ms,检测有缺陷光纤图像耗时 3.16 ms。在处理与缺陷检测算法中,采用队列结构用于平均图像的处理和检测时间,能够保证平均处理时间在 2.4 ms 以内,在光纤以 1.5 m/s 的速率绕制过程中检测没有丢失数据,满足系统的实时性要求。标准 AdaBoost 算法由于不针对具体特征,使用分类器级联效果不明显,正常光纤和有缺陷光纤检测耗时都为 16 ms 以上,不能满足系统实时性的要求。

表 3 改进 AdaBosst 算法耗时 ms

测试项目	算法			
	测试 1	测试 2	测试 3	平均值
图像预处理	1.03	1.11	1.12	1.09
检测正常光纤图像	1.26	1.24	1.33	1.28
检测有缺陷的光纤图像	3.11	3.18	3.20	3.16

表 4 标准 AdaBoost 算法耗时 ms

测试项目	算法			
	测试 1	测试 2	测试 3	平均值
图像预处理	0.95	0.94	0.94	0.94
检测正常光纤图像	15.78	16.43	16.28	16.16
检测有缺陷的光纤图像	16.31	17.18	16.73	16.74

4 结 论

AdaBoost 分类算法通过组合多个结构简单、准确性要求不高的弱分类器构造出任意精度的强分类器。本文首次尝试将此分类器应用于光纤缺陷检测系统,根据实际光纤缺陷的特点,针对通常采用的矩形特征值数量多、计算量大的问题,提出了采用光纤缺陷形态数据作为特征值的改进方法,大大减少了分类器的运算量和检测耗时,并设计了 AdaBoost 级联分类器以提高检测精度,并减少平均检测时间。与标准 Adaboost 算法对比,改进算法能够达到与标准算法相同的检测精度,同时大大缩短了检测时间,满足了系统的实时性要求。经实际测试,改进算法的光纤缺陷检测系统工作稳定、准确率高,较好地实现了对快速运动光纤缺陷的实时检测。

参考文献:

[1] 贡玉南, 华建兴, 黄秀宝. 基于匹配 Gabor 滤波器的规则纹理缺陷检测方法 [J]. 中国图象图形学报, 2001, 6(7): 14-18.
GONG Yunan, HUA Jianxing, HUANG Xiubao. Regular texture defect detection based on matched Gabor filters [J]. Journal of Image and Graphics, 2001, 6(7): 14-18.

[2] 胡涛. 基于轮廓对比的 PCB 裸板缺陷检测算法研究 [D]. 武汉: 华中科技大学, 2009.

[3] OTSU N. A threshold selection method from gray-level histograms [J]. Automatica, 1975, 11(1): 23-27.

[4] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.

[5] LIENHART R, KURANOV A, PISAREVSKY V. Empirical analysis of detection cascades of boosted classifiers for rapid object detection [C]// Proceedings of the 25th German Pattern Recognition Symposium. Berlin, Germany: Springer, 2003: 297-304.

[6] SCHAPIRE R E, SINGER Y. Improved boosting algorithms using confidence-rated predictions [J]. Machine Learning, 1999, 37(3): 297-336.

[7] PAUL V, MICHAEL J. Robust real-time object detection, TR CRL 2001-01 [R]. Cambridge, UK: Cambridge Research Laboratory, 2001.

(下转第 18 页)