Ultrasonics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Ultrasonics

journal homepage: www.elsevier.com/locate/ultras

Feature selection for neural network based defect classification of ceramic components using high frequency ultrasound

Manasa Kesharaju *, Romesh Nagarajah

Swinburne University of Technology, Faculty of Engineering & Industrial Sciences, Melbourne, Victoria 3122, Australia
Defence Materials Technology Centre (DMTC LTD), Melbourne, Victoria 3122, Australia

ARTICLE INFO

Article history:
Received 1 April 2015
Accepted 30 May 2015
Available online xxxx

Keywords:
Feature extraction
Genetic algorithms
Artificial neural networks
Principal Component Analysis

ABSTRACT

The motivation for this research stems from a need for providing a non-destructive testing method capable of detecting and locating any defects and microstructural variations within armour ceramic components before issuing them to the soldiers who rely on them for their survival. The development of an automated ultrasonic inspection based classification system would make possible the checking of each ceramic component and immediately alert the operator about the presence of defects. Generally, in many classification problems a choice of features or dimensionality reduction is significant and simultaneously very difficult, as a substantial computational effort is required to evaluate possible feature subsets. In this research, a combination of artificial neural networks and genetic algorithms are used to optimize the feature subset used in classification of various defects in reaction-sintered silicon carbide ceramic components. Initially wavelet based feature extraction is implemented from the region of interest. An Artificial Neural Network classifier is employed to evaluate the performance of these features. Genetic Algorithm based feature selection is performed. Principal Component Analysis is a popular technique used for feature selection and is compared with the genetic algorithm based technique in terms of classification accuracy and selection of optimal number of features. The experimental results confirm that features identified by Principal Component Analysis lead to improved performance in terms of classification percentage with 96% than Genetic algorithm with 94%.

1. Introduction

Ultrasonic inspection of reaction-sintered silicon carbide (RSSC) ceramics is difficult due to their high density, variations in grain boundary compositions, thickness variation and influence of microstructure variations. Currently used X-radiography does not lend itself to differentiating between defects in many regions of the material [1]. Moreover, discontinuities can be miniscule, numerous and widely dispersed that it is impractical to resolve them individually [1]. Porosity, density variation, presence of free silicon metal and fatigue in ceramics are examples of such defects [2]. Hence, there is a need for developing an on-line inspection system that would be far more cost effective than present methods and, moreover, assist manufacturers in checking the location of high density areas defects and enable real time quality control. The processing of ultrasonic raw signals and classification of defective signals obtained from the investigated ceramic components is a complex task. In order to develop automated online inspection system, several artificial intelligence approaches were used to identify and select signal features, that are then used in a Neural Networks (NN) based signal classification system to classify ultrasonic signals and identify defects within the armour ceramic components. Feature extraction referred in this case represents distinctive information as inputs into the classifier. Moreover, feature extraction addresses the problem of finding the most compact and informative set of features, to improve the efficiency or data storage and processing [3].

Machine learning systems perform two main functions, feature extraction and classification [4,5]. Over the last decade, extensive research has taken place on the development of efficient and reliable methods for the selection of features in the design of machine learning systems, where features constitute inputs to a classifier [6–9]. Selecting the most representative features as inputs to the classifier presents another challenge as the performance of the learning algorithm over the test set often depends on detecting the relevant features and discarding irrelevant features [10].

The success of various approaches mainly depends on the search strategy in the feature selection process. Different

* Corresponding author at: Swinburne University of Technology, Faculty of Engineering & Industrial Sciences, Melbourne, Australia. Tel.: +61 3 92143613.
E-mail address: mkesharaju@swin.edu.au (M. Kesharaju).

http://dx.doi.org/10.1016/j.ultras.2015.05.027
0041-624X/© 2015 Elsevier B.V. All rights reserved.

approaches use different ways to generate subsets and progress the search process. In sequential forward search (SFS), the features are successively added to an empty set [11,12]. Conversely, in sequential backward search (SBS), a full set of features are used and during the search process the redundant features are successively removed [7,13–15]. Though this sequential strategy is fast and easy to implement, it is affected by the nesting effect [16], which signifies that once a feature is added or removed it cannot be removed or added later. In order to overcome this effect, a modified random search technique is investigated in this research. In addition, these search algorithms involves partial search and computational complexity and hence it will be difficult in finding near optimal solution [16].

Optimization has a significant role in engineering design, especially when multiple parameters are involved. The problem of dimensionality reduction is well suited to be configured as an optimization problem [9,17]. Genetic Algorithms (GAs), are a type of randomized population-based stochastic search technique that offer an effective approach in finding near optimal solutions to complicated optimization problems [18]. GA is applicable to feature selection in the current research study, as it involves an exponential search space. The innovative work of Siedlecki and Sklansky validates the evidence for GAs advantage compared to classical search algorithms [19]. Subsequently, many studies showing advantages of using GAs as a feature selection technique have been published [7–9,18,20–26]. Principal Component Analysis (PCA) is another effective data pre-processing technique used to identify linear dependencies among attributes of a data set. It compresses the attribute space by identifying the strongest pattern in the data [27]. The PCA technique also known as Karhunen–Loeve transform has been explored by several researchers [21,28–32] for image and signal processing.

In this research, a GA based feature selection algorithm is proposed in addressing the problem in supervised learning that involves identifying the relevant or useful features in a dataset and providing only that subset of features to the learning algorithm for superior defect classification in ceramic components. The experimental results obtained using the proposed algorithm is then compared with the results obtained by using the feature transformation technique (PCA).

2. Methods

The silicon carbide ceramic samples used in the current study were supplied by Australian Defence Apparel (Melbourne, Australia). The percentage composition of SiC is 88%, as there is about 12% of residual silicon in these products. The pulse echo ultrasonic technique was used to inspect three double-curved, ceramic tiles of 300 mm in length and 7.5 ± 0.5 mm in thickness. A delay line contact transducer of 10 MHz frequency, 12.7 mm element diameter was chosen for scanning the defective ceramic tiles. The air gap between the specimen and probe was eliminated by applying thick lubricant on the surface of the specimen. Different defects such as porosity, free silicon and un-sintered material are generated in the ceramic tiles during and after the manufacturing process. The location of these defects was recorded using the X-ray technique. Ultrasonic signals were acquired at a sampling frequency of 100 MHz and each of the A-scan signals consisted of 2000 data points. As existing practice in the industry involves classifying each captured A-scan ultrasonic signal, gating is necessary for reducing the size of the data. Hence, a gating technique was applied to each of the signals, that checks and positions the time-gating on digitally captured A-scan image. This is a type of dimension reduction that makes it feasible to classify each echo.

2.1. Data set

The data sets used for the experiments were created by scanning two ceramic tiles with various defects. To investigate the effectiveness of the proposed feature selection method, experimental data was collected from both defect and defect-free regions of the ceramic tiles along with their individual locations marked precisely by means of a calibration method. A grid pattern with accurate increments of 10 mm was drawn across the surface of ceramic components and a contact probe was placed manually on each of these grid points in the calibration procedure. Each of these grid points was considered a test points and numbered accordingly. Due to the rough curved surface as well as presence of high density areas in ceramic components, the direct contact transducers with frequencies (5 MHz and 10 MHz) were unable to find small flaws, porosity and provide better resolution. Therefore, a delay line contact transducer of 10 MHz frequency has been chosen for scanning the defective ceramic components as it provides excellent near surface resolution compared to the normal contact transducers [33]. The contact testing set-up is shown in Fig. 1.

The training of the network was done using a dataset of examples known as 'training data'. The training dataset consisted of 132 ultrasonic signals obtained from both defect (includes porosity, un-sintered silicon, black spots and cracks) and defect-free regions. Table 1 presents the dataset used for training the network.

To ensure that the network did not over-fit; a cross validation procedure was followed by presenting the network with a validation dataset that consisted of 67 signals. The process of training and validating was repeated until the validation error decreased and the weights of the network were saved at this time. Later, a testing dataset consisting of 400 signals was presented to the network to evaluate the network performance. The function '*divideind*' available in the MATLAB signal processing toolbox was applied to the input matrix that divides the training, validation and testing datasets.

2.2. Wavelet feature extraction

The method developed as described in this paper includes a multi-domain approach using both time and frequency based information including wavelet domain features as inputs to a classifier. Statistical features from the time-domain (kurtosis, mean, front surface echo amplitude, first back surface echo amplitude, energy of samples and peak value) are considered for classification

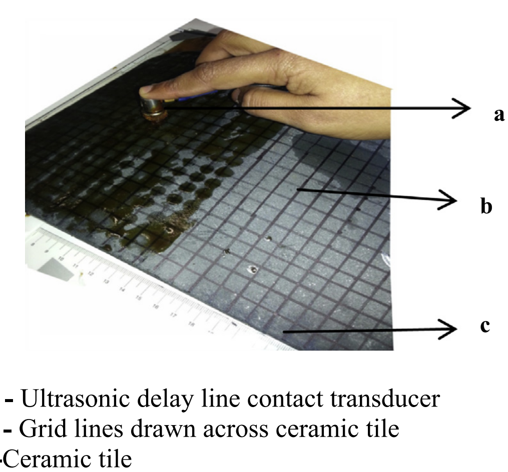

Fig. 1. Contact testing of a defect tile using delay line transducer.