01 Jan 1991

Pattern Recognition for Nondestructive Evaluation

S. Morris

P. O'Rorke

William E. Bond
Missouri University of Science and Technology, bondw@mst.edu

M. M. Amirfathi

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/comsci_facwork/205

Recommended Citation

S. Morris et al., "Pattern Recognition for Nondestructive Evaluation," *Proceedings of the IEEE Aerospace Applications Conference, 1991*, Institute of Electrical and Electronics Engineers (IEEE), Jan 1991.
The definitive version is available at https://doi.org/10.1109/AERO.1991.154534

Pattern Recognition for Nondestructive Evaluation*

M. M. Amirfathi
Douglas Aircraft Company
Long Beach, CA 90846
amirfathi@mdcgwy.mdc.com

S. Morris
P. O'Rorke
University of California - Irvine
Irvine, CA 92717

W. E. Bond
D. C. St. Clair**
McDonnell Douglas Research Laboratories
St. Louis, MO 63166

Abstract

This paper outlines the issues involved in automating nondestructive evaluation techniques. Nondestructive evaluation techniques are used to inspect a variety of parts during manufacturing and service. Currently, humans analyze the output obtained from test techniques by looking for features which indicate that a defect is located in the material. This evaluation is dependent on both the experience and alertness of the technician performing the test. Automation of these processes should improve the consistency of results and enhance the testing of more complex materials. Machine learning and pattern recognition techniques are being investigated to automate the process.

1.0 Introduction

Nondestructive evaluation (NDE) techniques are used to inspect a variety of parts during manufacturing and service. A wide variety of industries apply NDE to their processes, including aircraft manufacturers who use NDE techniques during the production of aircraft. Airline operators also utilize NDE techniques to maintain and inspect their fleets in service. Current NDE technologies are being challenged by demands for increased accuracy, increased speed, and increased reliability, but lower costs.

The growing use of advanced composite materials in current and future aircraft and the inspection requirements of aging aircraft fleets require inspection capabilities that often exceed current technology. The problems of stress, corrosion, and fatigue cracking are receiving increased attention as more aircraft continue in service beyond their normal design life. As requirements on minimum detectable crack sizes become more stringent, it becomes harder to discriminate the crack indication from noise. Other influences, such as crack orientation, corrosion, grain size, transducer coupling, and improper calibration, can cause false indications.

Douglas Aircraft Company (DAC) is exploring the use of artificial intelligence (AI) to automate the NDE process. This research is being conducted in conjunction with McDonnell Douglas Research Laboratories in St. Louis and with two universities: University of California-Irvine and University of Missouri-Rolla. Research has focused on the application of machine learning techniques to the construction and maintenance of knowledge-based systems which are capable of evaluating the readings from nondestructive tests that have been performed on aircraft components.

This paper describes the preliminary results obtained from this research. Section two describes the NDE test environment and outlines areas which need improvement. Sections three and four

---

* Export Authority: 22 CFR 125.4(b) (13).

** D. C. St. Clair is professor of computer science at University of Missouri-Rolla, Graduate Engineering Center and also visiting scientist at McDonnell Douglas Research Laboratories in St. Louis.

TH0361-6/91/0000-0001$01.00

describe the application of a symbolic machine learning algorithm, ID3, to the NDE problem. ID3 has been used by Douglas Aircraft to classify defects in sets of standard NDE reference blocks. Preliminary results are presented in section five. Based on these preliminary results, a need for an improved method of distinguishing features in the test waveforms is identified. Section six outlines a feature extraction approach from pattern recognition, called scale-space filtering, which can be used to preprocess data for input into a classification algorithm such as ID3.

2.0 Nondestructive Evaluation

While several NDE techniques are used for evaluating aircraft components, one of the techniques most often used by aircraft manufacturers and operators is the ultrasonic test. Ultrasonic inspection methods are used extensively during the fabrication and in-service periods of an aircraft's life cycle.

Ultrasonic testing utilizes ultrasonic sound waves in the 1 MHz to 25 MHz frequency range to measure the thickness of a material or to examine the internal structure of a material for possible defects such as voids, delaminations, and cracks. By transmitting a sound wave through the material and examining the amount of sound energy that is transmitted or reflected, it is possible to make determinations about the internal structure of the material.[1,2]

To produce a sound wave in a test piece, a transmitter applies high frequency electrical pulses to a "piezoelectric" crystal. Piezoelectricity refers to a reversible phenomenon whereby a crystal, when vibrated, produces an electric current; or conversely, when an electric current is applied to the crystal, the crystal vibrates. When energized with electrical pulses, the crystal transforms the electrical energy into mechanical vibrations and transmits the vibrations through a coupling medium, such as water or oil, into the test material. These pulsed vibrations propagate through the object with a velocity that depends on the density and elasticity of the test material. They are modified by the geometry of the medium and by intervals of discontinuity within the material. When sound waves strike a discontinuity in the material, most of the energy is reflected. These reflections may then be picked up by a second crystal or transducer, or the emitting crystal can be used to pick up the reflected signal.

Ultrasonic waveforms can be displayed using three different formats, A-scan, B-scan, and C-scan. A pulse-echo A-scan displays reflected sound amplitude as a function of time (depth) for a particular location on the test specimen. A typical defect-free wave form will show high amplitudes for the front and back surfaces of the sample and low amplitudes in the interior. A waveform indicating a defect will have additional high amplitude indications between the front and back surfaces. Figure 1 shows an A-scan representation of an aluminum block with no defect, while Figure 2 shows the A-scan results of the same type of aluminum block in which there is a defect. Both samples were taken from a standard set of reference blocks known as Alcoa Series B, or Hitt Block. These blocks are used as a reference standard in NDE tests. The B-scan display is a cross sectional view of the test specimen showing the depth of any flaw indications. The C-scan display gives a plan-view of the sample and any defects but does not indicate the depth of the defects.

2.1 Shortcomings of Current Operational Approach

Most in-service ultrasonic tests are currently being done manually using the A-scan display of the ultrasonic signal. Typically, a technician scans the probe over the part being tested, and observes the resulting wave pattern on a CRT. Based on the technician's level of training and experience in conducting the particular test, the technician will make a subjective decision as to whether the part is defective or good.

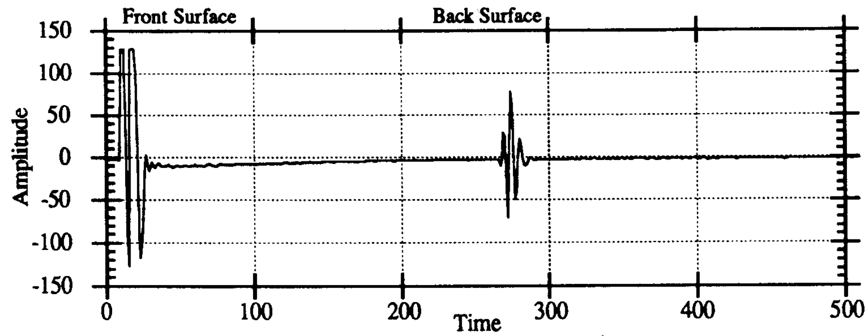

Figure 1. Typical A-scan representation of no-defect

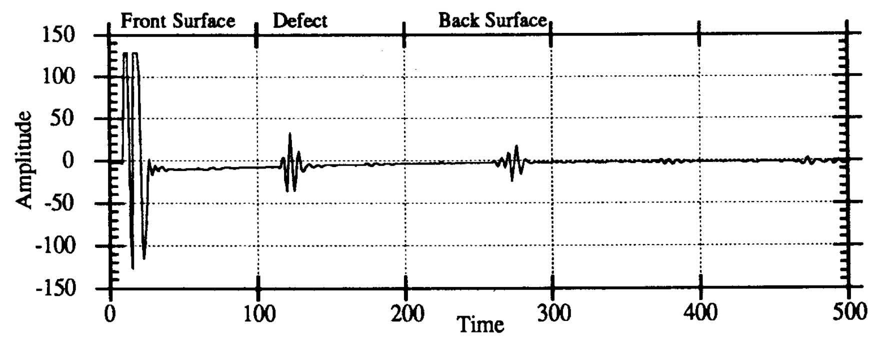

Figure 2. Typical A-scan representation of defect

Manual NDE methods have several deficiencies which limit their capabilities on advanced materials. The following outlines some of these problems:

1. Analysis of test results is subjective; it depends on the operator's judgement and state of alertness.

2. It is not always possible to duplicate test results gathered from the same test specimen, due to the fact that NDE technicians often manually alter the setup parameters on different tests.

3. The minimum detectable defect size depends on the initial setup parameters and can vary for different tests. The initial calibration of the instrument affects the amplitude to noise ratio of the wave pattern, which is the major factor in determining the minimum detectable defect size.

4. All relevant features of the wave pattern associated with each test case are not known.

5. It is not easy to extract all the relevant features (known and unknown) from the wave pattern.

Often, the weakest link in the inspection process is the human. Experience, training, and fatigue can significantly impact inspection reliability. Many inaccurate inspections result from faulty instrument calibrations, incorrect probe selection, or inaccurate interpretation of inspection results. The human factor, when combined with variations in instrumentation, contribute to a lack of consistency in inspection results and interpretation. Recent incidents such as the Sioux City DC-10 crash on July 19, 1989, the Aloha Airline 737's loss of a top section of fuselage on April 28, 1988, and the United 747 cargo door incident show the catastrophic results of undetected defects.

3.0 An Artificial Intelligence Approach to NDE

One of the major problems in NDE is the consistent and correct evaluation of test patterns. To automate the classification of defects from the analysis of NDE test patterns, a knowledge-based system is desired which will take test patterns as input and produce a classification as output. The basic idea is represented in Figure 3.

Input Pattern → | Knowledge-based System | → Classification

Figure 3. Basic automated system for NDE

The input pattern would be obtained from NDE test equipment. The knowledge-based system would evaluate the input pattern and report defect depth and size or the fact that no defect was present. The knowledge-based system would have knowledge about NDE and the ability to perform some type of "reasoning" over this knowledge in order to determine whether or not a defect was present. A fundamental issue is the design, implementation, and maintenance of the knowledge base.

One area of AI which suggests a solution to these problems is machine learning. The knowledge base could be constructed by a machine learning algorithm which is shown the waveforms resulting from both good and defective parts. The algorithm would then learn to distinguish between the patterns of good and bad parts. This knowledge would be used by the knowledge-based system to classify test patterns which it has never seen. If the algorithm has successfully learned the training data, it should then be able to classify new test data accurately. Any changes in the knowledge base necessitated by changes in test equipment or in the geometry of parts being classified could easily be implemented by rerunning the machine learning algorithm.

Symbolic inductive machine learning algorithms are a possible solution to this problem. These algorithms learn concepts by examining a set of training instances and then constructing a set of rules to describe the domain being analyzed. Two basic types of symbolic inductive algorithms can be found in the literature: decision-tree classifiers,[3-6] and rule-set builders.[7,8] Decision-tree classifier systems use training instances to build "optimal" decision trees. Rule-set builders construct a set of rules to describe the training instances observed. A decision-tree classifier was chosen for this research because of two reasons: 1) Results are output as a decision-tree which can be examined to determine which features are considered "most significant" by the machine learning algorithm; 2) The execution time required to produce the decision-tree is relatively small.

3.1 Decision-Tree Classifier Systems

Decision-tree classifier systems take training examples as input and produce one or more classification trees as output. This type of algorithm is exemplified by Quinlan's ID3.[4] To ID3, each training instance, t, is a tuple of attribute values and a corresponding classification value, viz.

$$t = (a_{1h}, a_{2i}, a_{3j}, \ldots, a_{nm}, c_j).$$

The value a_{1h}, represents the h-th value of the first attribute A_1, etc. The attributes A_i are taken from a predetermined description space,

$$D = \{A_1, A_2, ..., A_n\}$$

where each A_i represents a discrete-valued attribute. The value c_j is the associated classification taken from the set of all classifications **C**.

ID3 represents learned concepts in the form of decision trees. Figure 4 illustrates a typical tree. Each path from the root of the tree to a leaf denotes a concept. These concepts can also be thought of as rules. For example, the highlighted path in the decision tree in Figure 4 is analogous to the rule:

$$A_1(a_{11}) \wedge A_4(a_{43}) \wedge A_2(a_{22}) \Rightarrow C(N_4)$$

or equivalently:

If the value of attribute A_1 is a_{11}, and
the value of attribute A_4 is a_{43}, and
the value of attribute A_2 is a_{22},

then all classifications $c_j \in N_4$ may be assumed.

The set, $N_4 \subseteq C$, is the set of one or more conclusions (classifications), $c_j \in C$ which occur at this node. The expression $C(N_4)$ indicates that all classifications $c_j \in N_4$ may be asserted. Ideally, each path in the tree would terminate with a leaf node containing a single conclusion ($|N_i| = 1$ for all i). Such a classification tree would be able to completely differentiate each of the concepts it has learned.[9]

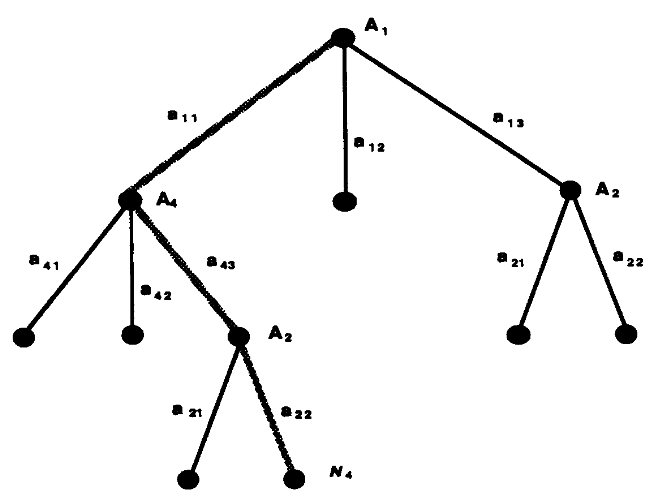

Figure 4. Decision tree showing $A_1(a_{11}) \wedge A_4(a_{43}) \wedge A_2(a_{22}) \Rightarrow C(N_4)$

During construction of the tree, the decision-tree classifier must determine the "best" attribute to be used to expand the tree at each node. It must also determine when no further attributes should be added to a path of the tree. Quinlan uses an entropy measurement to select the best attribute. This attribute will expand the tree width while minimizing tree depth. Mingers has experimented with the use of other attribute selection metrics.[10]

Induction of decision trees may be incremental or nonincremental. ID3 represents nonincremental induction, since all training instances are processed at one time and the decision tree created. At this point, the learning process is considered completed. In incremental induction, learning is performed each time the decision tree is used to classify a new instance. Two incremental versions of ID3 have been developed: ID4 by Schlimmer and Fisher,[5] and ID5 by Utgoff.[6]

4.0 Application of Machine Learning to NDE

It appears that NDE problems naturally lend themselves to characterization by machine learning techniques. It is desired to have the machine learning technique identify patterns and the classifications which describe those patterns. Then, given a new pattern, the resulting program should characterize the pattern and indicate what classification best represents that pattern. There are several factors which simplify/complicate the characterization of NDE results by machine learning techniques: 1) The testing setup is precisely defined, thus filtering out many environmental conditions; 2) Preprocessing of data is required to facilitate the determination of features; and 3) Machine learning techniques require that features be in a specific representation. Each of these will be discussed in more detail below.

Current NDE processing specifications precisely define the specific conditions under which a test should be performed. These specifications include the type of equipment, type of probe, equipment settings, and location of probe. In addition, a calibration sample is often provided to align and verify operation of the test equipment. Thus, a precise testing environment is defined. This control is intended to minimize the effects of varying environmental conditions as much as possible. However, the human operators do not always follow the processing specifications correctly. Since varying conditions would introduce "noise" into the feature data and in some cases could be so pronounced that they might influence learning techniques to concentrate on "incorrect features," this environmental control should contribute positively to the success of machine learning techniques in NDE. Automation of the calibration and set-up process would control correct processing of specifications for each test and also improve the test results and reduce the chances of mixing noise with the feature data.

Machine learning techniques attempt to build a classifier system that will accurately allow the program to distinguish between classifications, given input data. The machine learning algorithms take a set of given input features and follow a procedure to determine which subset of input features do the "best" job of predicting the correct category. Each technique also determines how and in what order the features should be combined. Thus, given the identified features, the machine learning algorithm constructs a "best attempt" to predict classifications. The results obtained, however, are greatly influenced by how well the chosen features accurately reflect the classifications. As an example, suppose it is desired to learn the difference between tall ($L/W > 1$) and wide ($L/W < 1$) boxes. If we calculate the ratio L/W for each box and use that as an attribute to a machine learning algorithm to predict the classifications tall and wide, the results will be straightforward and very accurate. However, if the simple attributes L and W had been used without preprocessing, the results would not have been as encouraging. In the same sense, it is anticipated that raw NDE data will have to be preprocessed to provide consistent, accurate results. Humans analyze the waveforms resulting from NDE by looking for shifts in the waveforms and unusual shapes appearing in unexpected locations. Preprocessing may require several transformations.

In addition to preprocessing to highlight features, it is also sometimes necessary to preprocess NDE data so that it can be used by a particular machine learning technique. Raw ultrasonic data is usually represented in terms of several hundred continuous-valued attributes. This raw data must be preprocessed and grouped into discrete classes. The "discretization" process then assigns a symbolic value to each class. This approach prevents the development of overspecialized rules. Thus it is anticipated that preprocessing will be required to accommodate the machine learning techniques that are used.

5.0 Preliminary Results

To evaluate the effectiveness of using the ID3 algorithm, data was obtained from standard NDE test blocks which are used to calibrate test equipment. These blocks are carefully manufactured to have a defect at a known depth and a specific location in the block. Several readings were taken from each block in positions where no defect was present and also in positions where a defect was known to exist. Figures 1 and 2 represent typical waveforms that were obtained for no-defect and defect conditions.

Preprocessing was used to convert the continuous-valued waveforms into equivalent discrete symbolic values. The preprocessed waveforms were used as training instances for the ID3 algorithm. The input data is a set of tuples which represent the amplitude of the waveform at a specified time (amplitude-at-time-t). The initial results of applying the ID3 algorithm to the preprocessed data were encouraging. Decision-trees produced by the ID3 algorithm indicated that the ID3 algorithm focused on the features that are also used by humans to evaluate waveforms: 1) The amplitude of the energy which represents the back surface (defects result in a decrease in the amplitude); and 2) The presence of significant vertical deflections in the waveform between the front and back surfaces (defects reflect a significant amount of energy). However, when these decision-trees were used to classify waveforms that were obtained at a later time, the decision-trees failed to correctly classify some of the waveforms. Two conditions contributed to this situation.

First, when the test data was obtained, one of the set up parameters, the zero offset, was altered. As a result, the training data set had a different zero offset than the testing data set. Second, it was apparent that there can be a slight shift in amplitude and location of some of the features associated with wave patterns taken from an identical test specimen. These fluctuations and shifts in the wave patterns were sufficient enough for the test data to often be misclassified.

This representation of data as amplitude-at-time-t is also inadequate for another reason. Parts can have a continuum of defect sizes and a continuum of defect locations. Clearly, the problem is to look at an ultrasound reading and "see" the defect - see where it is and how big it is. This is currently what the human operators are doing - a nontrivial task to say the least! The waveform should be represented in terms of the location of peaks and their size (not amplitudes at single times).

Furthermore, in the hope of keeping the classification algorithm as simple as possible, it is desirable to use a qualitative description of these peaks, rather than using all the information on a peak that is available in the signal. For example, a peak could be described as beginning at approximately time t1, ending at approximately time t2, and having a size which corresponds to the maximum amplitude in this time interval. To address these issues a more robust method of preprocessing is proposed.

6.0 Feature Extraction Using the Scale-Space Method

A common task in machine vision is the identification of the "edges" present in an image. Consider a measurement of light intensity along a one dimensional, linear cut through the image. An edge in this intensity data occurs when the first derivative of the intensity reaches a local extremum. At such an inflection point in the data, the second derivative goes to zero, and "a zero-crossing" is said to occur. Graphically, an edge is located at the point on the "side of a hill" in the data at which the slope is a local extremum.

The identification of edges is useful in machine vision because they provide a basis for parsing an image into meaningful parts. However, a problem of scale occurs: should all edges be considered, regardless of the size of intensity variation between them, or should small variations be filtered out, resulting in a scene described only in terms of "major" edges? Computational efficiency suggests the latter course - parse the image into significant regions based on the identification of edges of a certain strength. Thus the image is described in terms of primitives called "edges." These edges provide a qualitative description of the intensity data. Any technique that provides such a description can be applied to any one dimensional, quasi-continuous set of data. One particular technique for obtaining qualitative descriptions, called the scale-space method, was introduced by Witkin.[11] This method has been used to extract qualitative descriptions of NDE waveforms.

A common technique in machine vision for extracting features from data is to perform a discrete convolution of the data set with a Gaussian filter. The amount of smoothing produced by the filter depends on the width of the Gaussian function used. Each smoothing results in a certain degree of fine-grained features being obliterated, while coarser features survive.

One problem that occurs in using Gaussian filters is that a degree of smoothing that works well for certain segments of the data might also obliterate what are considered significant (interesting) features in other segments of the data. The particular Gaussian filter used, in essence, selects a scale of detail to preserve in the data. The problem is that no fixed finite set of scales suffice to provide good descriptions of a variety of data. It is this "scale problem" that is addressed by the scale-space method.

The scale-space method works by smoothing the data at many different scales and plotting the zero-crossings as a function of their location and the scale at which they are produced. The scale dimension is represented by the parameter σ, which is the standard deviation of the Gaussian filter used, and gives a measure of the width of the Gaussian. Figure 5 shows the scale-space plot for the data from Figure 2. The plotted zero-crossings can be arranged into lines rising vertically through the scale-space. Each line, called a "contour," is associated with a single inflection point as it survives across successive degrees of smoothing. In general, contours vanish in pairs corresponding to two inflection points approaching and meeting each other as the common peak that they bound is smoothed out. These two contours can be traced down to the time axis where they identify the two inflection points in the unsmoothed data. Thus, pairs of contours allow the identification of a feature, and the max σ to which the contours survive give an indication of how difficult it was to smooth out that feature.

In Figure 5, in addition to the three robust contours corresponding to the front surface, defect, and back surface of the block (the three contours on the left that survive above $\sigma = 43$), there is also an undesired long-lived contour on the far right. This contour is generated by inflection points in a region of the data that contains relatively small variations in amplitude. On the other hand, the relatively large amplitude peak associated with the defect does not survive as long; its inflection points are smoothed out by $\sigma = 50$. This phenomenon was found to be disappointing. It was hoped that the smoothing operations would eventually filter out all the small peaks in a way that, above some σ threshold, only the three peaks that have "significant" amplitudes in the original data would still have contours.

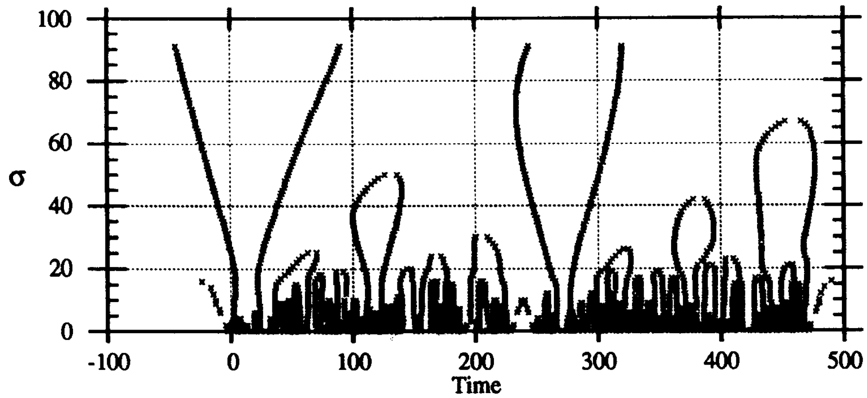

Figure 5. Scale-space plot for data from figure 2

Upon examining the second derivative of the smoothed data at different σs, we noticed that the zero-crossings for the significant peaks invariable had greater slopes, by about an order of magnitude, than the slopes for all other contours. Based on this, we introduce a test on the slope at the zero-crossings that filters out most of the zero-crossings associated with insignificant features.

The test is: for a given scale σ, if the absolute value of the slope of a zero-crossing is less than 0.005 of the highest absolute slope of all the zero-crossings, then that zero-crossing is filtered out. The value 0.005 was chosen (by hand) as working well over as many of the blocks as possible. Figure 6 shows the zero-crossings for the Figure 2 data that survived this test.

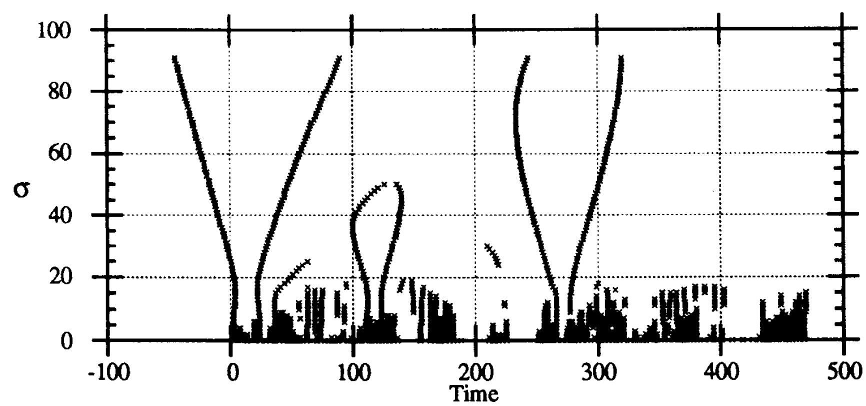

Figure 6. Scale-space plot with zero crossing test

6.1 The Extracted Qualitative Descriptions

The filtered scale-space technique lets us identify time bounds for significant features. For each block, a σ value was chosen to sample the surviving contours. These contours are traced back to the time axis to give bounding times for each feature. Then the maximum amplitude of the unsmoothed signal within each pair of time bounds is determined. Thus, each significant peak in the waveform is represented as a triple (t1, t2, amplitude). Three significant features were found in samples with a defect: the front surface, the defect, and the back surface. Two significant features were found in samples without a defect: the front surface and the back surface. These features can be used as input to a machine learning algorithm.

7.0 Conclusion

This paper has outlined how nondestructive evaluation is performed on materials and how machine learning and pattern recognition techniques could be used to automate the process. Preliminary results indicate that it is possible to utilize machine learning techniques to build a system that can identify materials which have defects. In particular, it is encouraging that the machine learning techniques focus on features that are also used by humans to identify defects. However, ultrasonic waveforms are complex and the results are subject to environmental conditions. For this reason it is necessary to develop preprocessing methods that are tolerant of noise and focus on the qualitative features of an ultrasonic waveform. Scale-space methods seem to provide a means of extracting such features. The extracted features will be used as input to a machine learning algorithm.

Acknowledgments

The authors gratefully acknowledge M. R. Collingwood and D. J. Hagemaier of Douglas Aircraft Company for providing NDE domain knowledge and test data. Thanks to Deepak Kulkarni at NASA Ames for informing us about the scale-space method, and to Kiriakos Kutulakos at the Univ. of Wisconsin, Madison, for the scale-space code he developed at NASA Ames.

References

1. J. P. Gallagher, F. J. Giessler, A. P. Berens, and R. M. Engle, **USAF Damage Tolerant Design Handbook: Guidelines for the Analysis and Design of Damage Tolerant Aircraft Structures**, Wright-Patterson Air Force Base, Air Force Report No. AFWAL-TR-82-3073, 1984.

2. R. C. McMaster, "Ultrasonic Test Principles," **Nondestructive Testing Handbook**, Vol. II, Section 43, 1959.

3. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, **Classification and Regression Trees**, Wadsworth and Brooks/Cole, 1984.

4. J. R. Quinlan, "Induction of Decision Trees," **Machine Learning**, Vol. 1, 1986, (pp. 81-106).

5. J. C. Schlimmer and D. Fisher, "A Case Study of Incremental Concept Induction," **Proceedings of the Fifth National Conference on Artificial Intelligence**, Vol. 1, Morgan Kaufmann Publishers, Inc., 1986, (pp. 496-501).

6. P. E. Utgoff, "Incremental Induction of Decision Trees," **Machine Learning**, Kluwer Academic Publishers, Vol. 4, No. 2, 1989, (pp. 161-186).

7. R. S. Michalski, I. Mozetic, J. Hong, and N. Lavrac, "The Multipurpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains," **Proceedings of the Fifth National Conference on Artificial Intelligence**, Morgan Kaufmann, 1986, (pp. 1041-1045).

8. P. Clark and T. Niblett, "The CN2 Induction Algorithm," **Machine Learning**, Vol. 3, No. 4, 1989, (pp. 319-342).

9. D. C. St. Clair, W. E. Bond, and C. L. Sabharwal, "Measuring Concept Strength in Classification Trees," **Proceedings of the Fifth International Symposium on Methodologies for Intelligent Systems: Poster Session Program,** October 1990.

10. J. Mingers, "An Empirical Comparison of Selection Measures for Decision-Tree Induction." **Machine Learning**, Vol. 3, No. 4, 1989, (pp. 319-342).

11. A. Witkin, "Scale-space Filtering," **Proceedings of the 8th International Joint Conference on Artificial Intelligence**, Karlsruhe, West Germany, 1983, (pp. 1019-1022).