

Sensor Validation in Non-Destructive Evaluation using Clustering

Viswanath Avasarala*^a, Jose R. Celaya^b, Kai Goebel^c, Neil Eklund^c

^aCollege of IST, Pennsylvania State University, U. Park, PA-16801

^bDecision Sciences and Engineering Systems Department, Rensselaer Polytechnic Institute, Troy, NY-12180

^cComputing and Decision Sciences, GE Global Research Center, Niskayuna, New York-12309

ABSTRACT

Non-destructive evaluation (NDE) techniques for condition monitoring in remote solid structures have evolved vastly in the last few years. Algorithms for estimation of sensor integrity and for noise correction form a crucial aspect of NDE. This paper presents a sensor validation approach that verifies sensor integrity, identifies and corrects noise effects and selects the best possible array of sensors for multi-sensor fusion. The proposed methodology uses a novel change detection algorithm for noise correction and a clustering algorithm to isolate useful signal information from the sensor data. It was used for sensor selection in a NDE field study, where multiple sensors were used to examine a solid structure. The methodology achieved 97% accuracy in the experiments, indicating its efficacy.

Keywords: Clustering, Sensor Validation, Cusum method.

1. INTRODUCTION

Non-destructive evaluation (NDE) techniques for condition monitoring in remote solid structures have evolved vastly in the last few years¹. A typical condition monitoring system used for defect identification/classification is shown in Figure 1². Condition monitoring systems involve feature extraction from the sensor data and pattern recognition algorithms that operate on the extracted features to characterize the state of the system. To maximally exploit sensor information, sensor information is transformed into various domains, where different features are extracted with the potential of providing (or at least contributing) to the desired characteristics. Examples are frequency domain features like spectral analysis and time domain features like mean value, kurtosis etc. Performance of pattern recognition algorithms is contingent on quality of sensor data and the number of training samples available. In many practical applications, NDE sensors are noise prone, faulty and are not significantly more reliable than the systems being monitored³. For most field studies, the number of training examples available is also very limited. Due to these conditions, the probability of an erroneous diagnosis of system state by the condition monitoring system could be unacceptably large. To minimize the risk of incorrect diagnosis, NDE techniques typically rely on sensor validation techniques to detect and isolate faulty sensors and multi-sensor fusion to correctly estimate variable values despite faulty measurements. The sensor validation techniques usually rely on redundant measurements to identify abnormal sensor behavior^{4,5,6}. Techniques that rely on redundant sensor measurements or multi-sensor fusion may not be effective in an environment where a majority of sensors record poor quality measurements. In these situations, development of specialized algorithms for estimation of sensor integrity and for noise correction is essential to the success of NDE. This paper presents a sensor validation approach for high noise NDE sensors that verifies sensor integrity, identifies and corrects noise effects and selects the best possible sensors for higher level processing.

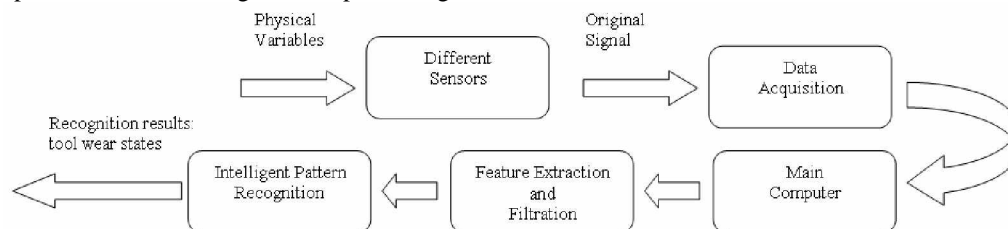


Figure 1: Schematic Representation of a tool condition monitoring system

This paper is organized as follows. Section 2 describes in greater detail, the problem for which the algorithms have been developed. Section 3 describes the algorithms that were used for identifying the sensor with least noise and best defect signature. Results obtained from our automated sensor selection mechanism are described in section 4.

2. PROBLEM DESCRIPTION

A NDE field study was conducted to differentiate between two types of defects, occurring in a solid structure⁷. Multiple, heterogeneous NDE sensors were employed to examine the solid structure. Analysts marked areas of potential defects in the solid structure as *regions of interest (ROI)*, using *c-scan* visualization that exhibits signal information from all the sensors. Time domain features like mean value and kurtosis of the amplitudes recorded were then extracted from the sensor readings. The feature values used in the final classifier algorithm were averaged across all available sensors. However, it was noticed using *b-scan* visualization that noise effects dominated the ROI readings of a large percentage of sensors and consequently, features extracted from these readings have little discrimination capability. The noise effects were due to a. malfunctioning sensors or b. other elements of the solid structure that dwarfed the defect. For example, Figure 2 shows both a noise-free and a noisy *b-scan*. If the sensor readings are noise free (figure 2.a), then the adjacent area is signal free whereas the ROI has a clear defect reflection. However, when the signals are present both in the adjacent area and ROI (figure 2.b), it is clear that the recorded signals are not related to the defect but are noise. Due to these noise effects, it was difficult to find more than one sensor that recorded noise-free good quality defect signatures. Therefore, the features used for pattern recognition were based on measurements of only one sensor. A key challenge was to identify the best sensor for use in the higher-level algorithms. Initially, a manual approach was adopted. A human operator individually looked at the *b-scan* visualizations of each sensor that recorded the reflected signals of the defect. The ROI was manually adjusted to remove the noise effects from it (see Figure 3). From the various sensors available, the human operator chose the sensor with the best defect signature as “best sensor” for use in the higher level processing. If more than one sensor had equally good reflections, then the operator arbitrarily chose one of them as the “best sensor”. However, this procedure was time consuming and prone to human error. Differences in judgment also lead to variations in results between different analysts.

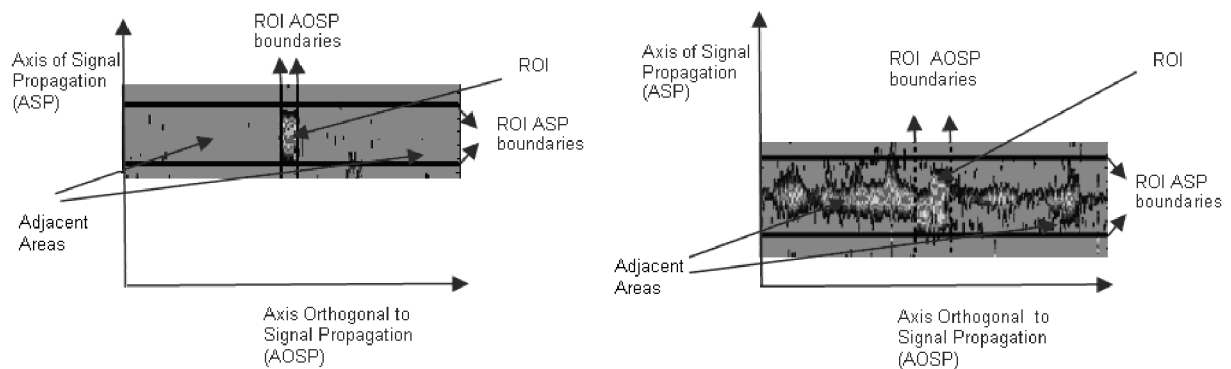


Figure 2: B-scans of Defect Areas. 2(a) (b-scan on the left) shows a ROI, where the adjacent area is noise free. 2(b) (b-scan on the right) shows a ROI, where the adjacent area is noisy.

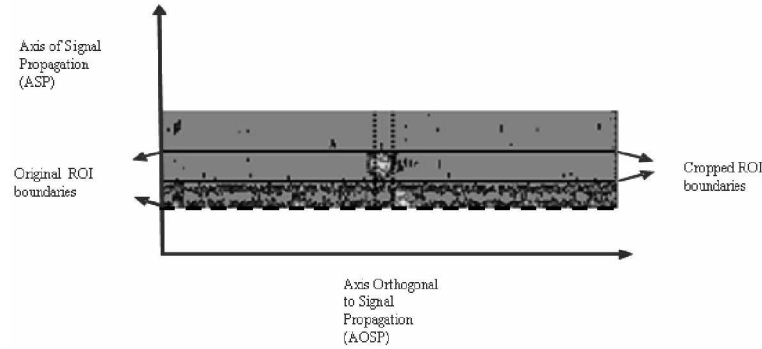


Figure 3: B-scan showing a ROI that is cropped to reduce noise effects

To avoid these problems, we devised an automated approach for noise reduction and sensor validation. Rectification for noise was based on a specially devised change detection algorithm which detects patterns across the region of interest and its adjoining areas, called as “adjacent areas” in the solid structure. The algorithm identifies patterns that do not show a significant change from the ROI to the adjacent area and classifies them as noise. After noise was identified and rectified for, the sensor validation algorithm selects the sensor with the best defect signature. To that end, a clustering algorithm was used to detect and mark the signature of the defect in the ROI. All sensors were ranked using a scheme, based on cluster properties, like cluster area. The highest-ranking sensor was selected for feature extraction. This technique is explained in greater detail in the next section.

3. SENSOR VALIDATION ALGORITHM

The automated sensor selection is a three-stage process. Stage one corresponds to noise removal. In this stage, ROI boundaries are cropped and new boundaries are drawn to remove noise effects inside the ROI. As shown in Figure 3, continuous patterns of noise were observed along only one direction, i.e., in the direction perpendicular to the direction of the signal propagation (called AOSP: Axis Orthogonal to Signal Propagation, henceforth). Hence ROI boundaries were cropped only in the direction of signal propagation (called ASP: Axis of Signal Propagation, henceforth). The ROI AOSP boundaries were not altered. The procedure is as follows:

For some sensor i ,

- Let the initial ROI ASP boundaries based on c-scan visualization be $initial_y$ and $final_y$
- For each y belonging to the interval $[initial_y, final_y]$, verify if the vector of amplitudes Y in the b-scan at y , corresponds to defect signature or noise (see section 3.1). If vector of amplitudes Y , corresponds to noise, label y as noisy.
- Calculate the longest continuous interval between $[initial_y, final_y]$, which has fewer than “ $maxiumAllowed$ ” noisy ASP values. Let the interval boundaries be $non-noisy_initial_y$ and $non_noisy_final_y$.
- Change the ROI ASP boundaries to $non-noisy_initial_y$ and $non_noisy_final_y$.

The second stage deals with cluster analysis. After changing the boundaries of the ROI, cluster analysis algorithms are used to isolate the cluster corresponding to defect signature in the ROI (see section 3.2). Stage three corresponds to assigning ranks to the various sensors, based on cluster properties. A score based on the cluster properties is calculated as follows:

$$SCORE = 0.1*x + 0.2*y + 0.7*z \text{ where}$$

x = area of the largest cluster in the ROI

y = percentage of ROI area that is occupied by clusters

z = energy content of the largest cluster.

For each sensor type, the sensor with the largest score is selected as the “best sensor”.

3.1 Noise Identification Algorithm

Noise identification algorithm uses a change detection scheme to determine whether signals recorded in the ROI at each particular coordinate of the AOSP axis corresponds to noise or not. The change detection algorithm is similar to Page's cusum method⁸. The algorithm finds a set of lines that can be drawn through a given data series, such that at no point, the residual is greater than the prescribed residual-limit. Residual at a point is the difference between the observed value and the predicted value (obtained from the regression line) at that point. Before the curve fitting phase, the data are smoothed using a kernel filter to reduce the effect of outliers. The noise reduction algorithm works as follows:

For each value of ASP axis y belonging to ROI

- Form the vector Z and vector X , the set of amplitudes recorded at this particular value of ASP-axis and their corresponding AOSP coordinates respectively.
- use the change detection algorithm to fit a set of lines through this data.
- If ROI at ASP coordinate y has a strong defect signature and is noise free, then the slope of the lines between adjacent area and ROI will be large in magnitude. Otherwise, the slope of the line between the adjacent area and ROI will not be significantly different from the slopes of the remaining lines. Based on this idea, a statistical test was formulated as follows:

H_0 : Signals in ROI at ASP coordinate y correspond to noise (the slope of line between adjacent area and ROI is not significantly larger in magnitude compared to the other regression lines).

H_1 : Signals in ROI at ASP coordinate y correspond to defect signature (slope of line between adjacent area and ROI is significantly larger, compared to slope of other regression lines).

It is assumed that the slopes are normally distributed under the null hypothesis. If the null hypothesis cannot be rejected for a given p-value, then the signal information at co-ordinate y in the ROI is regarded as noise. Else, it is regarded as defect reflection.

The procedure is illustrated using the following artificial data. Let the amplitude signals recorded at a particular ASP value be

$$Z = [1 \ 1 \ 1 \ 2 \ 3 \ 3 \ 3 \ 4 \ 5 \ 5 \ 30 \ 30 \ 30 \ 30 \ 5 \ 5 \ 4 \ 3 \ 3 \ 3 \ 2 \ 1 \ 1 \ 1] \text{ and}$$

their corresponding AOSP coordinates be

$$X = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13 \ 14 \ 15 \ 16 \ 17 \ 18 \ 19 \ 20 \ 21 \ 22 \ 23 \ 24].$$

The coordinates in italics belong to the ROI. If kernel smoothing is neglected and maximum allowable residual is taken as 0.0, the line fitting algorithm gives the set of points in X that have to be joined by straight lines such that the given residual constraint is met (see Figure 4). For the above inputs, the change detection algorithm gives the output as follows:

$$R = [1\&3, 3\&5 \ 5\&7, 7\&9, 9\&10, 10\&11, 11\&14, 14\&15, 15\&16, 16\&18, 18\&20 \ 20\&22, 22\&24].$$

The slopes of the regression line segments are

$$S = [0, 1, 0, 1, 0 \ 25, 0, -25, 0, -1, 0, -1, 0].$$

The slopes of line segments that exist between context-ROI are ± 25 . The p-value for the test under the normality assumptions is zero and hence null hypothesis is rejected at any acceptable level of significance.

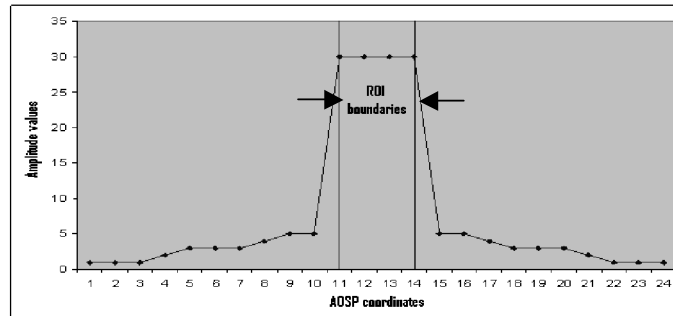


Figure 4: Results of change detection algorithm for sample data given in section 3.1

3.2 Cluster Identification Algorithm

The sensor validation algorithm implemented was very sensitive to the areas of the clusters formed by the cluster identification algorithm. Also, the number of clusters of the defect signature in the ROI is not known *a priori*. To address these issues, we developed a simple clustering algorithm that exploits the low dimensionality of the sensor data. The algorithm uses two-design parameters a) `minAmplitudeThreshold` and b) `minClusterLength`. The `minAmplitudeThreshold` is the minimum signal amplitude in the ROI that can be considered for assignment to a cluster. Amplitudes less than `minAmplitudeThreshold` are considered noise. The `minClusterLength` parameter denotes the minimum AOSP width that a cluster needs to possess, at each ASP. Clusters, whose width is smaller than `minClusterLength`, are regarded as noise. The input to the cluster identification function is the matrix `matrixROI` which contains amplitude information in ROI. Let the initial ASP boundaries of the ROI be `initial_AOSP`, `final_AOSP`. The algorithm is described in detail in the below table.

<pre> noOfClusters = 0; for i = initial_AOSP to final_AOSP a. Let the set X be the ASP coordinates at the which matrixROI(i, ASP) > minAmplitudeThreshold b. Split X into sets X₁ ... X_n such that ASPs values at X_j are continuous. (look at Figure 5 for explanation). c. for k = 1 to n c.1) is cardinality of X_k > minClusterLength? If yes, go to step c.1.1), else continue to step c. c.1.1) belongsToPreviousCluster = 0; c.1.2) For j = 1 to noOfClusters c.1.2.1) can X_k be assigned to cluster j? (A set X_k can be assigned to cluster j, if there is an overlap between its ASP range and the ASP boundaries of cluster j at AOSP = i- 1) . If yes, go to step c.1.2.2, else continue (go to step c.1.2). c.1.2.2)%X_k belongs to cluster j belongsToPreviousCluster = 1; currentCluster = j; break loop c.1.2. end %for all clusters j c.1.3) if (lbelongsToPreviousCluster) %initialize a new cluster currentCluster = noOfClusters +1; noOfClusters = noOfClusters + 1; end% the if condition c.1.4) Save the cluster ASP boundaries of cluster, currentCluster, at AOSP i; end %loop for sets X_k. end % loop for all AOSPs i </pre>

Table 1: Cluster Identification Algorithm

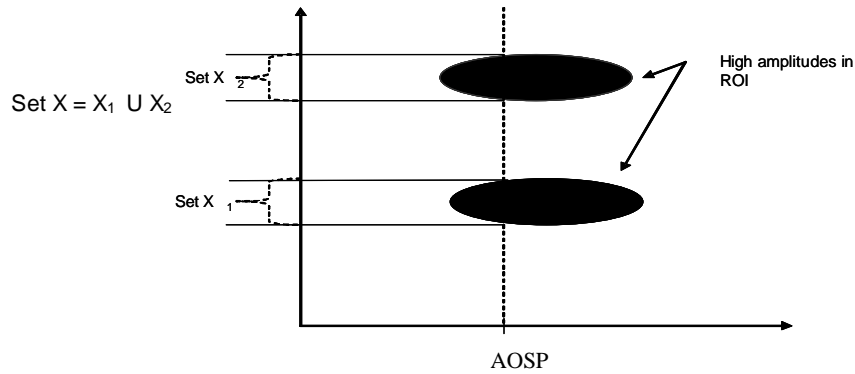


Figure 5: Illustration of formation of clusters using cluster identification algorithm

This algorithm was very effective in currently identifying the various clusters in the ROI and their corresponding boundaries. The results of the cluster identification algorithm on a sample b-scan are shown in Figure 6, where four different signal clusters (demarcated by white boundaries) are identified by the algorithm.

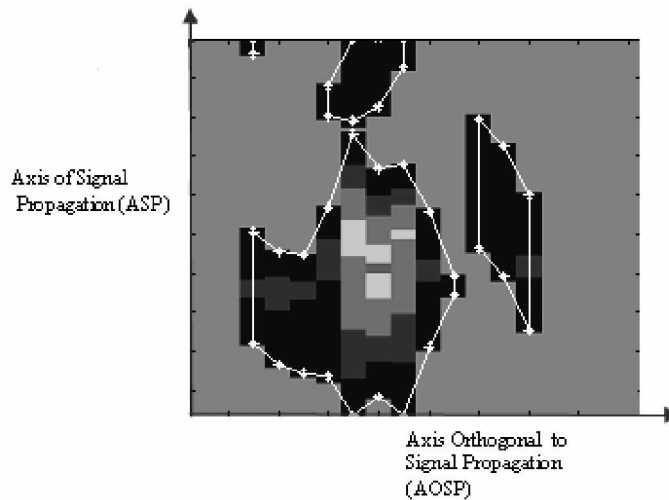


Figure 6: Clusters in the ROI identified by the cluster algorithm for a sample b-scan

4. RESULTS

The parameters used in our experiments are shown in Table 2. We tested our procedure on 157 cases. For these cases, the operators were also asked to manually select the best sensor as described in section 2. We found that in 58.4% of the cases, results from the manual selection and from the automated selection coincided. Post verification of the remaining cases showed that in 38.5% of the remaining cases, the automated selection procedure gave satisfactory results. These cases pertained to the situation where the operator arbitrarily chooses the “best sensor” from the multiple sensors with clean defect signatures in their ROI. In these cases, the automated sensor selection results were ascertained to be equally good, as compared to the manual selection. However, in 3.1% of the cases, the manual selection was better than the automated sensor selection procedure. For example, Figure 7 shows a B-scan where the ROI is clearly noisy (since signals are also present in context), but the noise identification algorithm failed to identify the signal information as noise. This occurred because, the statistical test for noise identification algorithm was too lenient and sometimes mislabeled noisy ROIs as clean. In principle, this can be avoided by

making the statistical test for noise identification more strict i.e., decreasing the level of significance for the statistical test in the noise identification algorithm. Various levels of significance were attempted to find the optimal level of significance for the test. It was found that more stringent parameters had the adverse impact of aggressively marking clean regions as noisy and decreased the algorithms performance even further. However, 3.1% error rate was deemed satisfactory for an automated system with zero human input, especially when the condition monitoring system is used for online testing where real time performance is also critical.

Parameter	Value
maximumAllowed (in noise identification algorithm)	12 ASP units
Residual-limit (for change-detection algorithm)	4
Epanechnikov kernel smoothing parameter	5
p-value (for noise identification hypothesis test)	0.82
amplitudeThreshold (for cluster identification algorithm)	27 DB
minClusterLength (for cluster identification algorithm)	200 ASP units

Table 2: Parameters used in the experiments

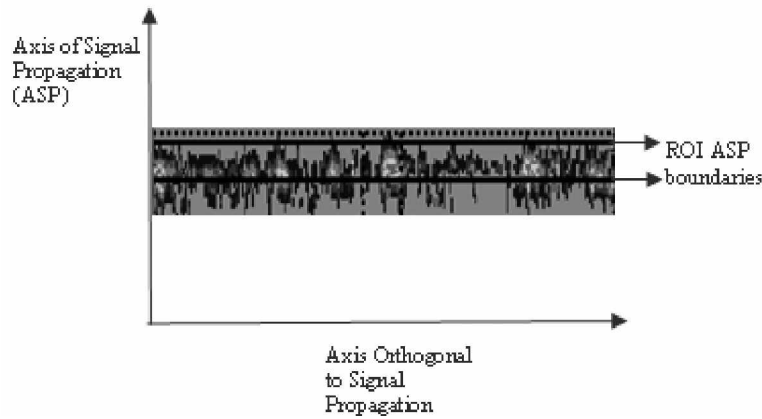


Figure 7: Example of a B-scan where ROI cropping is too lenient

5. DISCUSSION

The 97% accuracy achieved during our experiments support the validity of the proposed algorithms. In this experiment, only one sensor was selected for feature extraction, due to lack of sensing resources. When adequate sensing resources are available, the algorithm can be used for multi-sensor selection, by selecting the desired number of highly ranked sensors from the available resources. We are also investigating the use of this algorithm for selection of cases that are used to train the pattern recognition algorithm. For example, a few cases in the experiments had zero sensors that had clean defect signatures. If features extracted from these cases are used for training, the performance of the pattern recognition might be adversely affected. To automatically remove these cases from the training set, we are currently experimenting on using a unit step function, based on the score of the best sensor as the input variable. All cases where the score of the best sensor is below a prescribed threshold value are considered ineligible for use in the training set.

6. REFERENCES

1. X.E.Gros, "Applications of NDT data fusion", Kluwer Academic Publications, 2001.
2. P. Fu, A. Hope, "Fuzzy Logic Data Fusion For Condition Monitoring", Chapter in *Applications of NDT data fusion*, Edited by X.E Gros., Kluwer Academic Publishers, pp. 59-89 , 2001.

3. S. Alag, A. M. Agogino, M. Morjaria, "A methodology for intelligent sensor measurement, validation, fusion, and fault detection for equipment monitoring and diagnostics", *AI EDAM* 15(4): 307-320, 2001.
4. E. Eryurek, and E. Turkcan. "Neural Networks for Sensor Validation and Plantwide Monitoring", *Nuclear Europe Worldscan*, vol. 12, no. 1-2, pp. 72-74, 1992.
5. C. Lee, "Sensor value validation based on systematic exploration of the sensor redundancy for fault diagnosis KBS", *IEEE Transactions on Systems, Man, and Cybernetics*, 24(4): 594-605, 1994.
6. P. Ibarquengoytia, L. E. Sucar, and S. Vadera, "Any Time Probabilistic Reasoning for Sensor Validation", *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pp 266-273, 1998.
7. K. Goebel, W. Yan, N. Eklund, X. Hu, V. Avasarala, J.R. Celaya, "Defect classification of highly noisy NDE data using classifier ensembles", *SPIE conference on Smart Structures and Materials and NDE for Health Monitoring and Diagnostics*, Feb 26th – Mar 2nd, 2006, San Diego, California USA.
8. E.S. Page, "Continuous Inspection Schemes", *Biometrika* 41, 100-114, 1954.