Proc. Vth International Workshop, Advances in Signal Processing for Non Destructive Evaluation of Materials
Québec City (Canada),2-4 Aug. 2005. © X. Maldague ed., É. du CAO (2006), ISBN 2-9809199-0-X

The Use of Locally Weighted Regression for the Data Fusion with Dempster-Shafer Theory

by Z. Liu, D. S. Forsyth, S. M. Safizadeh, M.Genest, C. Mandache, and A. Fahr

Structures, Materials Performance Laboratory, Institute for Aerospace Research, National Research Council Canada, Montreal Road 1200, Ottawa, Ontario, K1A0R6, Canada

Abstract

The Dempster-Shafer (DS) theory provides an efficient framework to implement multi-sensor data fusion. Both the flexibility and the difficulty consist in defining the probability mass function. The fusion result is a discrete value or a label, which is determined by the corresponding maximum probability values. However, in some applications a continuous result is expected. In this paper, a scheme based on DS reasoning and locally weighted regression is proposed to fuse the data obtained from aircraft corrosion damage inspections. The proposed approach implements a pairwise regression that is optimized by the DS method when multiple inputs are involved. Experimental results on the fusion of conventional eddy current and pulsed eddy current data for the application of aircraft corrosion quantification are presented.

Keywords: Dempster-Shafer theory, data fusion, corrosion quantification, local weighted regression, classification

1. Introduction

The Dempster-Shafer theory has been used to fuse the data from multiple sensing modalities in various applications. The uncertainty of the measurements is modeled with the probability mass function, which defines a value between zero and one (basic probability assignment) to indicate the degree of support for a proposition [1]. In the framework provided by DS theory, the frame of discernment (θ) consists of 2^{θ} singletons, which are based on θ mutually exclusive and exhaustive propositions. In practical applications, such proposition is expressed as a discrete value or pre-defined class number to describe certain condition or status. However, not all the applications fit into this framework. For example, the quantification of corrosion damage in aircraft lap joints employs a continuous value to represent the material loss. The material loss by layer serves as one of the corrosion metrics for structural integrity and life prediction analysis [3]. Usually, the quantifying procedure is implemented by using the calibration results represented by a calibration curve. Only the first-layer corrosion can be characterized with the calibration method. Meanwhile, the problem of superimposed corrosion between multiple layers can not be solved.

The idea of fusing multiple nondestructive inspection (NDI) data is to characterize different corrosion with integrated features from multiple sources [4, 5]. The fusion operation should optimally map those features to the target outputs. In this paper, a locally weighted regression (LWR) is used with the DS approach to estimate the material loss in aircraft lap joints due to corrosion damage. In the proposed scheme, the NDI measurements are first discriminated with trained classifiers. The classification results are further combined based on the DS rule. Finally, the outputs are quantified with the LWR method for the thickness estimation. The result is compared with the "ground truth" data from the teardown inspection of a naturally corroded specimen from a service-retired aircraft. The feasibility of the proposed approach is demonstrated in this paper.

2. The approach

2.1 The general data fusion procedure

The flowchart in *Figure 1* depicts the procedure of the data fusion scheme. The estimation is carried out based on the previous available knowledge or data, which is named as the training data. The optimal number of clusters for the training data is found by using fuzzy k-means clustering and cluster validation indexes [6, 7]. Proper classifiers are assigned to the eddy current (ET) and pulsed eddy current (P-ET) data based on the cross-validation test results. The labelled training data is used to train the specific classifiers for ET and P-ET inputs respectively. The classified results determine the "data class" while the labelled X-ray data set defines the "information class". The probability mass function is defined based on the relation between these two classes.

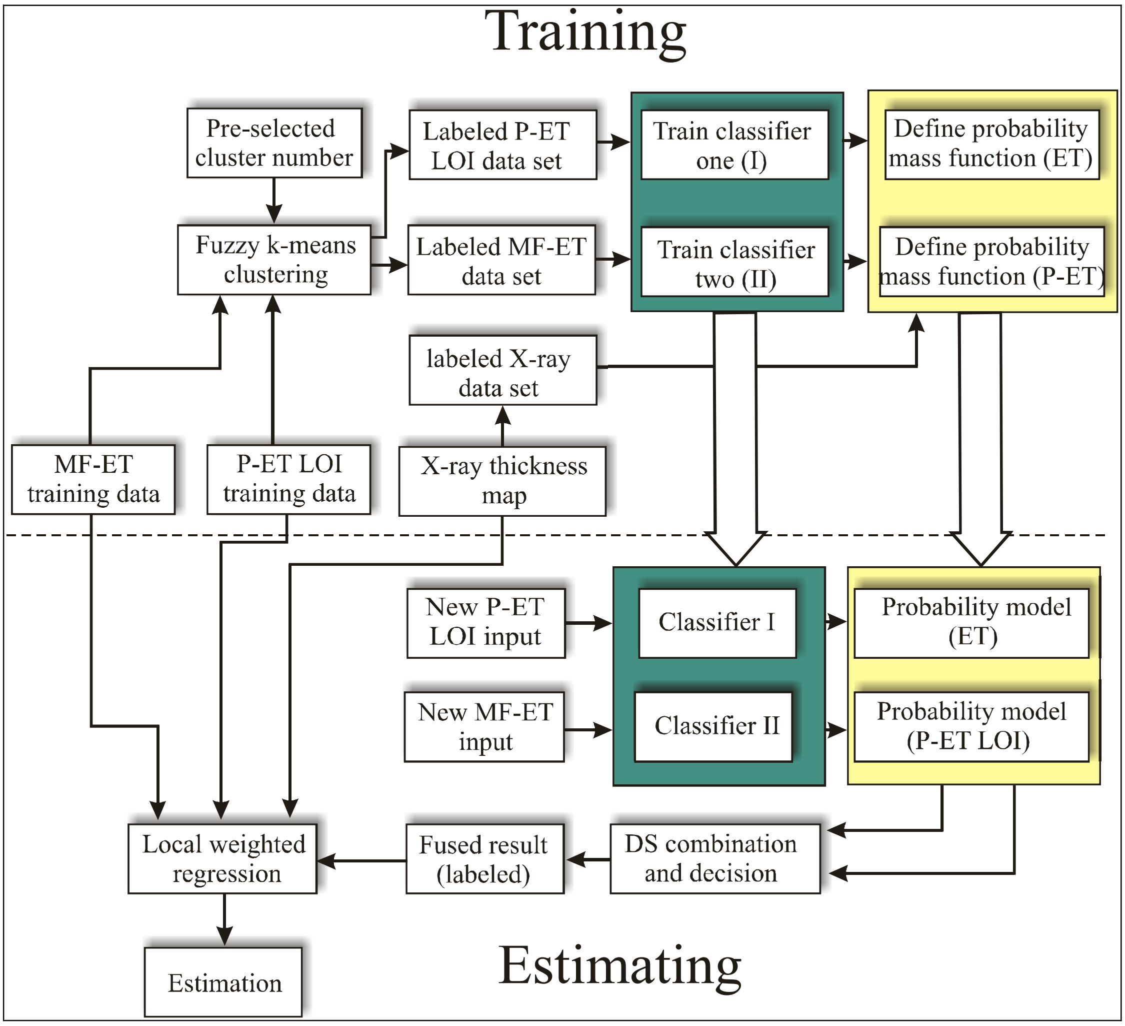


Figure 1. The data fusion procedure.

When a prediction is carried out on a new NDI measurement, the trained classifiers are applied to assign the input a pre-defined data class number. The probability mass value is derived from the probability model established during the training process. The DS combination rule fuses the probability values and a decision is made on the maximum mass output. The regression is implemented with only the corresponding data points defined by this information class. Therefore, a continuous estimation result together with a belief value can be achieved.

2.2 Definition of probability mass function

Assuming that C_i ($i = 1,2,..., N$) represents the information class (corrosion types) and $\bar{x}$ is the vector of the measurement values, the mass value can be defined as the probability of being certain information class based on the statistical information from available training data sets, i.e. $m_s(C_i) = p(C_i|\bar{x})$ [8]. Herein, s indicates the different data sources ($s = 1,2,..., S$). The input $\bar{x}$ is mapped to data class d_j ($j = 1,2,..., M$) by a classification operation. Thus, the basic probability assignment (BPA) is defined as:

$$m_s(C_i) = p(C_i|d_j)p_s(d_j|C_i) \tag{1}$$

The second value $p_s(d_j|C_i)$ is regarded as a measurement of the capability of each data source for discriminating the information classes. According to DS theory, the BPA values must be normalized to meet the requirement $\sum_i m_s(C_i) = 1$ before the updating operation is applied.

2.3 Locally weighted regression

The local regression is estimating the value using information pertaining only to a neighborhood of the input query [9]. Supposing that the variable $y \in R$ represents the material thickness and the NDI measurement vector is $\bar{x} \in R^m$, we need to find the mapping function $f : R^m \to R$. That is:

$$y_p = f(\bar{x}_p) + \varepsilon_p \quad (p = 1,2,..., n) \tag{2}$$

Herein ε_p is a random variable. Given a query point $\bar{x}_q$, to obtain the regression applicable to this query point, the following cost function is minimized:

$$J = \sum_{p=1}^{n} \left(\beta^T \mathbf{x}_{p,1} - y_p\right)^2 K\left(\frac{|\bar{x}_q - \bar{x}_p|}{h}\right) \tag{3}$$

where, $K(\cdot)$ is a weight function and h is the bandwidth. $\mathbf{X}$ is matrix of data samples with $[\bar{x}_p, 1]^T$ in the p[th] row and $\mathbf{y} = [y_1, y_2, \cdots, y_n]^T$. The solution to the above cost function is [9]:

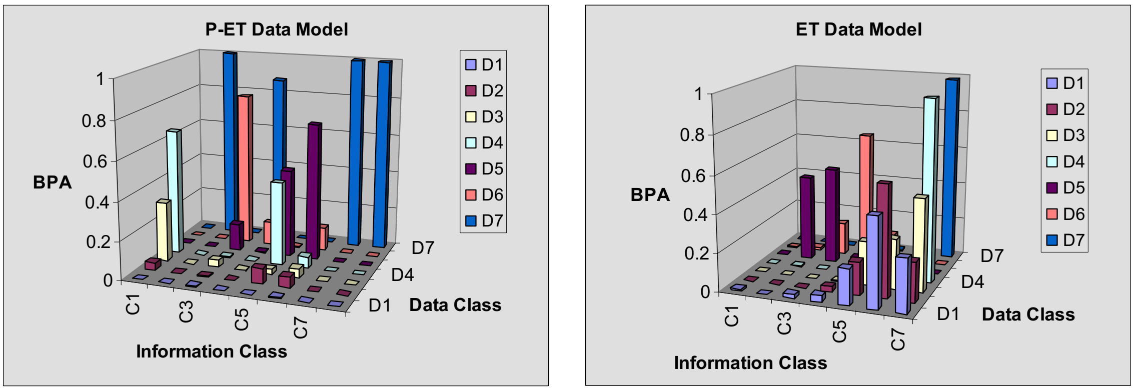
$$\hat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v} \tag{4}$$

where $\mathbf{Z} = \mathbf{WX}$ and $\mathbf{v} = \mathbf{Wy}$. $\mathbf{W}$ is a diagonal matrix with the i[th] diagonal element $w_{ii} = \sqrt{K(|\bar{x}_p, \bar{x}_q|/h)}$. The estimation at $\bar{x}_q$ is then given by : $\hat{y}(\bar{x}_p) = \hat{\beta}\bar{x}_p$.

3. Experimental results

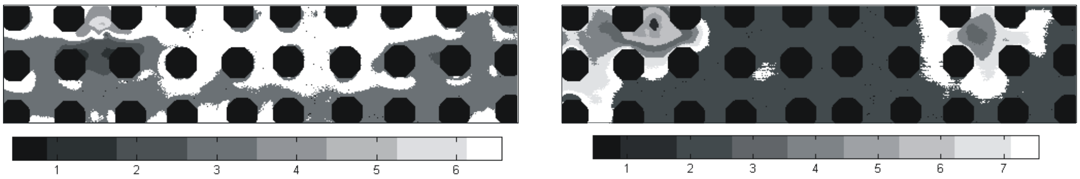
Data sets from the inspection of a service-retired Boeing 727 aircraft are used in the experiment. A two-layer lap joint cut out from below the cargo floor was inspected by the multi-frequency eddy current testing at 5.5kHz, 8kHz, 17kHz, and 30kHz frequencies and the pulsed eddy current testing. The P-ET lift-off-intersection (LOI) scan is extracted and used for analysis [11]. The ground truth data is obtained by using digital X-ray mapping technique on each layer. Data from two sections are used for training and testing respectively. The ET data obtained at 17kHz and 30kHz together with P-ET LOI are used for the first layer thickness estimation.

The ET and P-ET data are clustered by applying the fuzzy k-means clustering algorithm. The initial cluster number is set from 2 to 14. The clustering partition index, separation index, Xie and Beni's index, and Dunn' s index are considered to select a proper cluster number for NDI data. Consequently, ET and P-ET data are clustered into 7 and 8 groups. Therefore, the X-ray data is segmented into seven parts based on the percentage of material loss. The clustered/labelled NDI data is used to train classifiers. To find an efficient classifier, the cross-validation test is applied to several candidate classifiers. The one with the smallest error is selected. In the experiment, the nearest mean classifier is employed for classifying ET and P-ET data. According to 2.2, the probability functions are defined for ET and P-ET data respectively as shown in *Figure 2*.



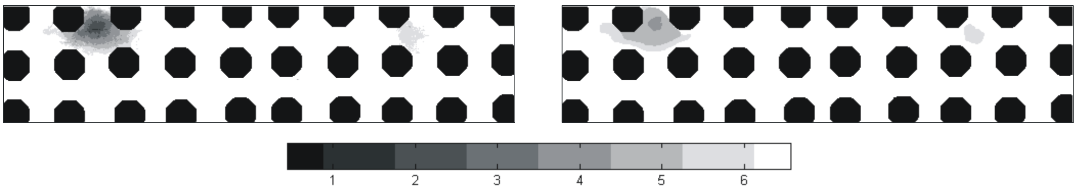
(a) (b)

Figure 2. Probability model of NDI data: (a) pulsed eddy current and (b) eddy current (17kHz and 30kHz).



(a) (b)

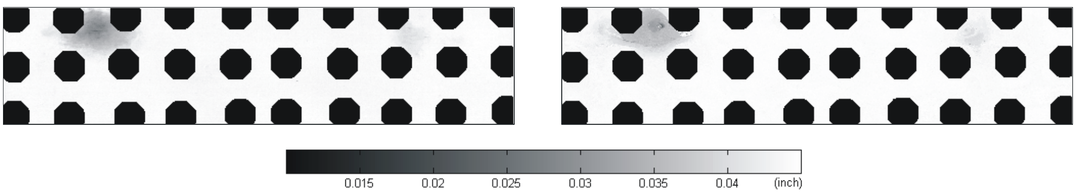
Figure 3. The classification results of (a) eddy current (17kHz and 30kHz) and (b) pulsed eddy current LOI scan.



(a) (b)

Figure 4. (a) The segmentation of X-ray thickness map and (b) the fused result of Figure 2(a) and (b).

The classification results are from ET scan (17kHz and 30kHz) and P-ET LOI scan are shown in *Figure 3* (a) and (b), respectively. *Figure 4* (a) shows the segmented X-ray thickness map. The corresponding definition of the classes is given in *Table 1*. The fused result of *Figure 3* (a) and (b) is presented in *Figure 4* (b) after a morphological processing. The X-ray thickness map and the regression result can be found in *Figure 5* (a) and (b), respectively.



(a) (b)

Figure 5. (a) The X-ray thickness map and (b) the estimated result obtained by DS fusion and LWR.

Table 1. The definition of information classes in term of material loss (corrosion type).

Corrosion class	C1	C2	C3	C4	C5	C6	C7
Thickness range (inch)	0.024 – 0.027	0.027 – 0.030	0.030 – 0.033	0.033 – 0.036	0.036 – 0.039	0.039 – 0.042	0.042 – 0.045

Table 2. The evaluation of the fusion result.

	RMSE (E-3)	CORR	PSNR	DE	MI
ET 17kHz	1.4585	0.9992	15.0381	1.0021	0.7150
ET 30kHz	1.5017	0.9991	15.0475	0.9657	0.7157
P-ET LOI	2.4106	0.9977	16.4591	0.6286	0.8170
Fusion result	0.8151	0.9997	36.0737	0.0673	0.7939

The fusion result is compared with the X-ray thickness reference in terms of a number of image comparison metrics listed in *Table 2*, i.e. root mean square error (RMSE), cross-correlation (CORR), peak signal-to-noise ratio (PSNR), difference entropy (DE), and mutual information (MI) [10]. The DS-based fusion followed by LWR process achieves the best result. The output of DS fusion as shown in *Figure 6* can be used as an indication of the degree to which we can trust the result.

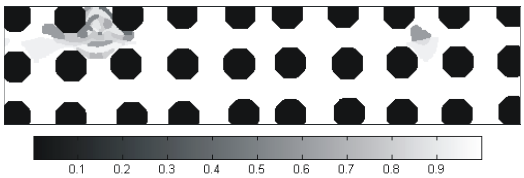


Figure 6. The DS fusion output.

4. Discussion

The result obtained in this experiment for the first-layer thickness estimation is better than the results reported previously [5]. The fusion algorithm developed here can also be applied to the deep-layer thickness estimation. However, the capability of P-ET technique for revealing and discriminating deeper layer corrosion has not been fully explored in this work. Only LOI scan, which is a single feature point in the time domain, is employed for the analysis. Another study indicates that the LOI is not fixed and changes with the presence of different corrosion damages [11]. Therefore, using single LOI point scan might not be an optimal solution.

In the proposed scheme, the classification error of the selected classifiers is not considered in the probability model. This value indicates how much we can rely on the results of a specific classifier and the probability of a measurement belongs to certain class. On one hand, an iterative classification scheme might be helpful to reduce the classification error; on the other hand, this error can be considered when the probability model is built.

The classification-based approach as presented in [12] provides a promising solution for identifying corrosion at different layers because it is a typical classification application. How to train a classifier efficiently with limited training data still remains an unsolved problem, but the classification-based approach may lead to further improvement when new data is available. Similarly, the proposed data fusion scheme also provides a mechanism for a possible improvement.

The quantification analysis relies on artificially prepared "damages" on a calibration specimen. The calibration curve is determined by limited points. The performance of the proposed fusion algorithm can be further improved in two aspects: the accuracy of the classifier and the probability model. However, the performance does not increase with the accumulation of data because the quality of the data cannot be assured. The updating mechanisms for selecting good data and improving the fusion analysis by using the accumulated historical data would therefore make for an interesting topic for future work.

5. Summary

In this paper, a data fusion scheme based on Dempster-Shafer theory and locally weighted regression is proposed. The measurement value is optimally classified by fusing the classification results with the DS combination mechanism. The final estimation is achieved by locally weighted regression of the pre-classified results. The data fusion method achieves a better estimation of lap joint thickness than the one obtained by calibration. For our future work, the use of multiple P-ET slices around LOI point is considered. The data fusion algorithm may be applied to these P-ET images for characterizing deep-level corrosion. This will also help to identify the contribution of the eddy current technique for the thickness estimation. Another way to improve the method is to develop a mechanism to select good data and update the model and classifiers for an improved estimation.

REFERENCES

[1] HAMID (R.). – *An Experimental Data Fusion Model for Multisensor System*. PhD. Dissertation, New Mexico State University 1989.
[2] HALL (D.). - *Mathematical Techniques in Multisensor Data Fusion*. Norwood, MA, USA, Artech House, 1992.
[3] LIU (Z.), FORSYTH (D.S.), and KOMOROWSKI (J.P.). – *Fusion of Multimodal NDI Images for Aircraft Corrosion Detection and Quantification*. Blum (R.S.) and Liu (Z.) ed. Multi-sensor Image Fusion and its Applications, p. 375-404, CRC Press, 2005.
[4] FAHR (A.), FORSYTH (D.S.), and CHAPMAN (C.E.). – *Survey of Nondestructive Evaluation (NDE) Techniques for Corrosion in Aging Aircraft*. National Research Council Canada Technical Report, LTR-ST-2238, Oct. 1999.
[5] LIU (Z.), FORSYTH (D.S.), LEPINE (B.A.), SAFIZADEH (M.S.), and FAHR (A.). – *Quantifying Aircraft Hidden Corrosion by Using Multi-modal NDI*. Thompson (D.) and Chimenti (D.) ed. Review of Progress in Quantitative NDE, Vol.23, p. 1355 – 1362, Green Bay , Wisconsin, 2003.
[6] BENSAID (A.M.), HALL (L.O.), BEZDEK (J.C.), CLARKE (L.P.), SILBIGER (M.L.), ARRINGTON (J.A.), MURTAGH (R.F.). - *Validity-Guided (re)Clustering with Applications to Image Segmentation*. IEEE Transactions on Fuzzy Systems 4:112—123, 1996.
[7] XIE (X.L.), BENI (G.A.) - *Validity Measure for Fuzzy Clustering*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(8):841 – 847, 1998
[8] FORSYTH (D.S.), LIU (Z.), KOMOROWSKI (J.P.), and PEELER (D.). – *An Application of NDI Data Fusion to Aging Aircraft Structures*. 6^{th} Joint FAA/DoD/NASA Conference on Aging Aircraft , San Francisco, CA, USA., Sep 2002.
[9] ATKESON (C.G.), MOORE (A.W.), and SCHAAL (S.). - *Locally Weighted Learning*. Artificial Intelligence Review, 11:11-73, 1997.
[10] XUE (Z.), BLUM (R.S.), and LI (Y.). – *Fusion of Visual and IR Images for Concealed Weapon Detection*. Proc. ISIF, p. 1198 – 1205, 2002.
[11] LEFEBVRE (J.H.V.) and DUBOIS (J.M.S.). – *Lift-off Point of Intercept (LOI) Behavior*. Thompson (D.) and Chimenti (D.) ed. Review of Progress in Quantitative NDE, Vol. 24, p. 523 - 530, 2004.
[12] SAFIZADEH (M.S.), LIU (Z.), MANDACHE (C.), FORSYTH (D.S.), and FAHR (A.). – *Intelligent Pulsed Eddy Current Method for Detection and Classification of Hidden Corrosion*. V International Workshop – Advances in Signal Processing for NDE of Materials, August 2005.