11th European Conference on Non-Destructive Testing (ECNDT 2014), October 6-10, 2014, Prague, Czech Republic

Improving the reliability of NDT inspection through information fusion: applications in X-ray and ultrasound modalities

Ahmad OSMAN [1], Valérie Kaftandjian [2], Ulf Hassler [1]

[1] Fraunhofer Development Center X-ray Technologies EZRT, Flugplatzstraße 75,
90768 Fuerth, Germany
Phone: ++49 911 58061 - 7642, Fax: +49 911 58061-7599; e-mail: ahmad.osman@iis.fraunhofer.de,
ulf.hassler@iis.fraunhofer.de
[2] National Institute of Applied Science INSA-Lyon; Université de Lyon, 25 bis Avenue Capelle, F-69621
Villeurbanne, France; E-mail: valerie.kaftandjian@insa-lyon.fr

Abstract
In this contribution we present a classification method based on the evidence theory. The classification method is compared to the state of the art support vector machine classifier on an industrial radioscopic data and 3D CT data of aluminium castings as well as 3D ultrasonic data of composite materials. The reported experimental results reveal the robustness of the proposed method and its advantages as well as disadvantages.

Keywords: X-ray, Computed-Tomography, ultrasound, Sampling Phased Array, Aluminium Casting, CFRP, Fusion Theory, defects classification.

1. Introduction

The improvement of the inspection reliability is a critical issue in Non-Destructive Testing (NDT), whether in the context of manual or automated results evaluation. Optimizing the reliability of inspection corresponds to increasing the percentage of true defects detections while minimizing the percentage of false alarms. The Receiver Operating Characteristics (ROC) curves plotted for different decision thresholds and the area under the curve allow comparing several classifiers and / or select an optimal threshold. The term classifier refers here to the distinction between true defects and false alarms.

Independently of the NDT technique being used, when the measured data are processed automatically, the first phase is a critical step of segmentation, where the smallest possible defects must be separated from the background noise. The segmentation of ultrasound data is particularly difficult due to different reasons among which is the strong influence of speckle noise on the data quality. In radioscopy, false alarms are also common when it comes to detecting very low contrast defects. In the case of Computed Tomography (CT), the problem is even more critical because of the frequent presence of artefacts that can be mistakenly marked as foreground during thresholding. Consequently, for all these techniques, after segmentation, the regions (i.e., objects) marked as foreground without necessarily being defective is usually high. For this reason, the determination of the object type (default true/false alarm) is required and is the second critical phase of data evaluation.

Our presented study falls in the context of classification of foreground regions after segmentation. The proposed approach for this phase uses measured characteristics of the segmented objects, and merging the information from these features.

In the following text, we briefly explain the method already published in [1] and report the results for different modalities: classical X-ray radioscopy of Aluminum castings [1], 3D CT of Aluminum castings [2] and 3D ultrasound data of Carbon Fiber Reinforced Polymers (CFRP) obtained by the method of "sampling phased array" [3]. Our classification method is compared with the state of the art Support Vector Machines (SVM) classifier [2].

2. Data fusion classification method

The Data Fusion Classification method is based on the Dempster-Shafer (DS) [4-5] theory also known as evidence theory, which is an extension of the classical Bayesian probability theory. In DS theory, the set of assumptions considered (or frame of discernment) Ω is a set of mutually exclusive and exhaustive hypotheses: in our case it is true hypotheses class H_1 = True Defect (TD) and H_2 = False Defect (FD). In a probabilistic framework, any piece of information would be distributed between these two hypotheses. However the DS theory supports the consideration of a union hypothesis. For example, if a source of information cannot decide if an object is TD or FD, it has hesitance. This is allowed by introducing a third hypothesis, which is the union of both hypotheses H_1 or H_2 to represent hesitation or ignorance ($H_3 = H_1 \cup H_2$). This difference DS theory and probability theory is fundamental.

Once the frame of discernment Ω is defined, any information from a measurement must be represented by a value called "mass", "piece of evidence" or "basic belief assignment" that can be assigned to a single hypothesis or union of hypotheses. Thus, the workspace is no longer Ω, it is rather the 2^{Ω} which is composed of all single hypotheses and all their possible unions. The set 2^{Ω} includes all proposals for which the source can bring a piece of evidence. Obtaining the distribution of masses values $m(A_i)$, $0 \leq m\ (A_i) \leq 1$ with $A_i \in 2^{\Omega}$, is an important step because it represents all the knowledge related to the application, the accuracy as well as the uncertainty associated with the measurement.

The power set in our application is composed of the following three hypotheses: H_1 = True Defect (TD), H_2 = False Defect (FD) and the union of the previous two hypotheses representing the ignorance class $H_3 = H_1 \cup H_2$.

To build the mass function, an original method was proposed in [1] that requires a learning dataset in order to obtain regions of confidence each characterized by its mass value. Briefly described, the method works as follows: from a set of segmented objects, a list of features is subtracted. Each feature is considered as a source of information. For a certain feature, its spatial distribution for all the segmented objects is considered to build the histograms repartition of true and false defects. For each interval of the histogram, the proportion of true defects is directly equal to the mass of the hypothesis H_1, while the rest is assigned to ignorance H_3. Similar intervals in terms of proportion of true defects values are grouped into regions and fuzzy transitions is defined between regions to avoid abrupt transitions. Once all the features are transformed into masses, they can be merged using the combination rule of Dempster [4-5]. The decision threshold is then applied to the final mass of the hypothesis H_1.

The complete procedure is summarized in the following figure, and more detail are given in [1].

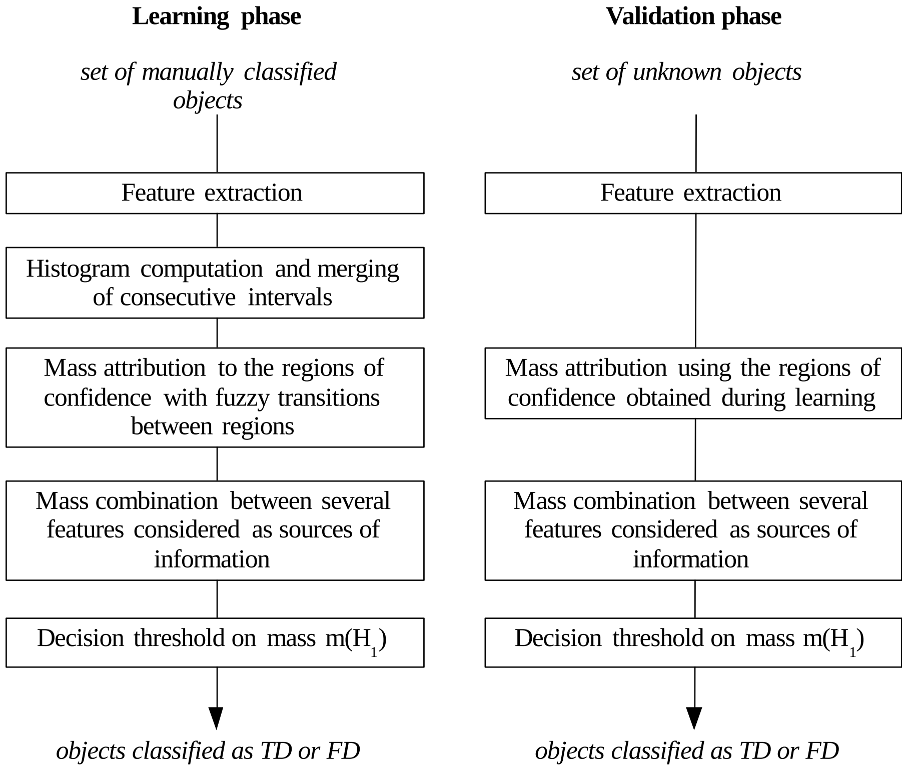

Figure 1. Flow diagram of the proposed Data Fusion Classification method.

3. Performance measures

To evaluate the performance of a source in discriminating between two classes the following terms are necessary to introduce: let P be the total of positives (true defects), N be the total of negatives (false alarms), TP be the total number of positives correctly classified, TN be the total number of negatives correctly classified, FN be the total number of positives incorrectly classified and FP be the total number of negatives incorrectly classified.

To measure the performance of a source or combination of sources after fusion, a threshold on the mass m (H_1) (decision threshold S) is applied. Objects with mass value above the decision threshold are considered defects, if not they are classified as false alarms. The performances of the sources are then evaluated by computing the following percentages.

- True Defects classification rate:

$$\mathrm{PTD} = \frac{\mathrm{TP}}{\mathrm{P}}$$

- False Defects detection rate:

$$\mathrm{PFD} = \frac{\mathrm{TN}}{\mathrm{N}}$$

4. Experimental results

4.1 First case: application to X-ray aluminium castings

The radioscopic database is derived from an industrial application of castings (fig.2). After segmentation of the X-ray images, 11 features are measured (area, contrast, etc ...). The objects are sorted manually and the decision of the expert pro object is regarded as the true decision. The database consists of objects 361: 231 are real defects (oxides, porosity, gas cavities) and false alarm 130 (see figure 2). The database is divided into two parts, one dedicated to the learning phase (65 FD and 115 TD) and the other for the validation phase (65 FD and 116 TD).

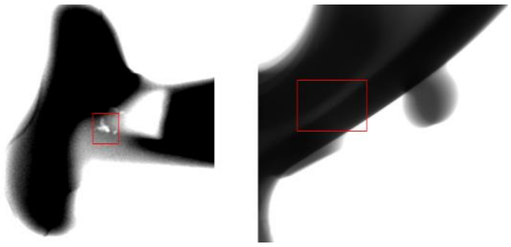

Figure 2. Inside the left side red rectangle appears a true defect and on the right side red rectangle appears an artefact caused by the structure of the inspected part.

The classification results are reported in the table 1 below for the different possible combinations: mean mass, optimal DS combinations with only two features, ISAR industrial system currently used and SVM [6, 7].

Source	Learning		Testing	
	PFD	PTD	PFD	PTD
Mean Mass	1	0.991	0.974	0.953
DS(MaxElongation & InOutContrast)	0.938	1	0.982	0.841
DS(Depth2Thickness & InOutContrast)	0.98	1	0.964	0.846
ISAR	0.723	0.974	0.723	0.982
SVM	1	0.982	0.969697	0.965

Table 1. Performances of the different classifiers in true of PTD ratio and PFD ratio obtained with the DS fusion method, statistical combination i.e. mean mass, ISAR and SVM.

4.2 Second case: application to 3D CT of aluminium castings

Here we dispose of a series of 3D CT volumes of aluminum casting (see Figure 3). The segmentation of the volumes gave 442 objects which were manually classified by an expert with: 44 TD and 398 FD. As it can be remarked, the database is very unbalanced in terms of number of each respective class. For each detected object, 30 characteristics are measured.

The database is divided into two parts: learning (FD 200, 26 TD) and validation (FD 198, 18 TD). Classification results are given in Table 2 [2, 7].

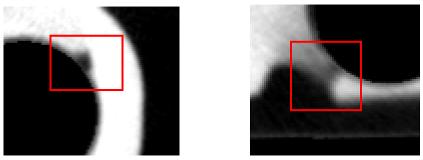

Figure 3. Part of a slice view extracted from a 3D-CT volume: on the left side a true defect (surface defect) appears as a darker area and on the right side a false alarm (reconstruction artefact) appears as a darker area as well.

Source	Learning process		Testing process	
	PFD	PTD	PFD	PTD
Mean mass	0.99	0.96	0.87	1
DS (features 14 & 18)	1	0.8	0.99	0.61
SVM			0.97	0.94

Table 2. Performances of the classifiers obtained on the CT dataset.

4.3 Third case: application to classify data of inspected composites by means of ultrasonic SPA method

In this case, the classifiers are applied to classify the objects detected after segmenting 3D SPA volumes of CFRP composites [3, 8]. Here the total amount of objects is 419 divided into: a learning dataset (164 FD and 48 TD) and a validation dataset (164 FD and 43 TD). Each object was described by means of 35 geometrical and intensity based features. The classification results are reported in the table 3 [3, 8].

Source	Learning process		Testing process	
	PFD	PTD	PFD	PTD
DSF(E_{xy}, CS_{Max})	0.958	0.975	1	0.932
SVM			0.976	0.971

Table 3. Performances of the classifiers obtained on 3D ultrasonic SPA volumes.

4. Conclusion

We have presented a method for classifying objects as true defects or false alarms through the combination of the measured characteristics on the segmented objects. The key point of our method is to translate the feature values into mass or basic belief assignment associated with the hypothesis "true defect." These mass functions obtained for the different features repartitions are actually merged. The major interest of the theory of Dempster-Shafer fusion lies in the fact that it has the ability to model the hesitation or ignorance between several hypotheses. Thus, the ignorance of a source, if it is combined with another source which has a belief that the truth lies in the hypothesis "true defect", helps to increase the final mass value assigned to the "true default" hypothesis. It is this aspect that allowed by using the DS theory to improve the reliability of the inspection by obtaining higher defect and false alarms classification ratio. Several cases of industrial applications have been presented in NDT 2D and 3D X-ray as well as in 3D ultrasound. As it was seen, the same method of classification is effective in all cases. However the SVM classifier remained better if the two classification rate (TD and FD) are simultaneously considered. DS classifier requires a balanced database as its performance can drop when the number of objects is too low (as in the case 2 of the TD), while the SVM is more robust towards an imbalance between classes. However, the SVM works with all measured characteristics measured (thirty in the reported second case) while the optimal DS combinations uses only two characteristics. It can thus accelerate the step of extracting features and keeps maintaining optimal performances.

References

1. A. Osman, V. Kaftandjian, and U. Hassler. “Improvement of X-ray castings inspection reliability by using Dempster-Shafer data fusion theory”. Pattern Recognition Letters, Vol. 2, Issue 32, pp: 160-180, 2011.
2. A. Osman, V. Kaftandjian, and U. Hassler. “Automatic classification of 3D segmented CT data using data fusion and support vector machine”. Journal of Electronic Imaging of Spie, Vol. 21, May 2012.
3. A. Osman, V. Kaftandjian, U. Hassler, and J. Hornegger. “An automated data processing method dedicated to 3D ultrasonic non-destructive testing of composite pieces”. International Symposium on Ultrasound in the Control of Industrial Processes, Madrid Spain, 8-20 April 2012.
4. A. Dempster, "Upper and lower probabilities induced by multivalued mapping", Annals of Mathematical Statistics, 38, pp. 325-339, 1967.
5. G. Shafer, "A mathematical theory of evidence", Princeton University Press, Princeton, pp. 297, 1976.
6. A. Osman, V. Kaftandjian, U. Hassler, Application of data fusion theory and support vector machine to X-ray castings inspection. Proceedings of 10th European Conference on Non-Destructive Testing (ECNDT), 7-11 June 2010, Moscow Russia.
7. A. Osman, U. Hassler, V. Kaftandjian, Interest of data fusion for improvement of classification in X-ray inspection, Digital Industrial Radiology and Computerized Tomography DIR2011, Berlin, June 2011
8. A. Osman, Automated evaluation of three dimensional ultrasonic datasets. Phd thesis, university of Erlangen-Nuremberg and INSA-Lyon, June 2013.