Interest of Data Fusion for Improvement of Classification in X-ray Inspection

Ahmad OSMAN[*], Ulf HASSLER[*], Valerie KAFTANDJIAN[**]

[*]Fraunhofer Development Center X-ray Technology EZRT, Dr.-Mack-Str. 81, 90762 Fuerth, Germany
[**] National Institute of Applied Sciences INSA- Lyon, Bat. St. Exupery, 20 Avenue Albert EINSTEIN, 69621 Villeurbanne, France

Abstract. Dempster-Shafer (DS) evidence theory is developed as an attempt to overcome the limitation of conventional probability theory by handling uncertain, imprecise and incomplete information. In this paper we present a classification system based on this fusion theory. The performance of this system is evaluated on 2D and 3D X-ray data. Obtained results are very promising and encourage us to use this system in other applications, namely for ultrasound data classification.

1. Introduction

X-ray inspection is a traditional non destructive testing method used to thoroughly test industrial parts, such as aluminium castings in the automotive sector. Safety specifications and quality control task are the main focus of the inspection process. Digital image processing, computational intelligence and hardware progress allowed automating this task. While the detection of true defects is the objective, one main difficulty in X-ray inspection is the detection of false alarms (or false defects), especially if very small and low contrasted defects have to be detected. Therefore, reducing the rejection ratio of good parts without risking missing true defects is a serious challenge.

The automatic detection and recognition of defects requires computerized image processing, image analysis, and decision process. The image processing step is critical to detect potential defects. During the image analysis step, features are extracted to be further used in classification between true defects TD and false defects FD. Our intervention is in the classification step, where we developed a specific approach based on data fusion theory to combine different sources of information together in order to improve the true classification rates of true defects and false alarms. To make fusion between different sources possible, a transition from the source space into a common space called "mass values" space takes place. We present a completely automatic mass value attribution procedure with no need for expert supervision. Performances of the classifier are quantified in terms of correct classification rates respectively for true defects and false alarms, and also using Receiver Operating Characteristics curves (ROC). The developed classification system, called Data Fusion Classifier (DFC), was used to classify defects from segmented 2D radiographic images and 3D-CT volumes where it gave in both cases highly precise and reliable decisions.

In the next section, the data fusion theory is briefed, and then section three describes the classification algorithm. Afterwards applications and results are resumed in section four. Discussions and future improvement are briefed in section five.

2. Data Fusion

Data fusion finds its useful application in multisource contexts, when data are uncertain and imprecise and where there is possibility of missing data. In decision making process, the application of data fusion aims at building a more pertinent decision. The general concern is to successfully classify an element x into a class C_i of the frame of discernment Ω, using information f_j provided by the source j.

2.1 Principe

The most used models in data fusion are the probability approach, the possibility approach and the Evidence theory (Dempster-Shafer theory). These theories of uncertainty represent the imperfections of information through a confidence measure. For each class C_i of Ω, a value $M(x \in C_i / f_j(x))$ represents the confidence accorded to the fact that an element x belongs to a class C_i, using f_j information provided by the source j. The data fusion process involves three steps illustrated in figure 1.

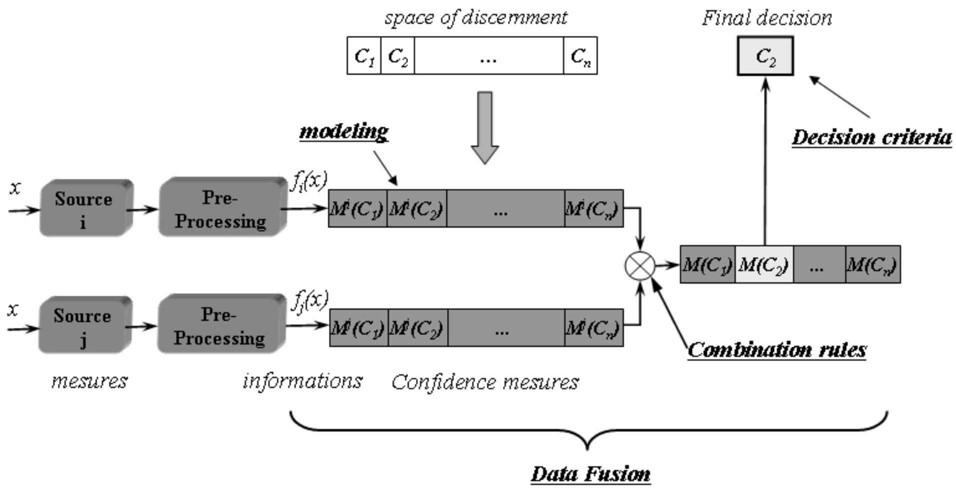

Figure 1. Data fusion process steps: knowledge modeling, confidence combination and decision [1].

The aim of pre-processing is to extract the parameters of x. The selected parameters must be pertinent regarding the objectives. In some applications, like casting defects, expert choice can serve us to select the important parameters. The classes, which are provided by the frame of discernment, must correspond to the needs of the application and depending on the type of the problem these classes can be exclusive (one true and others false) or not (intermediate symbols OR).

A measure $M(C_i)$ allows us to quantify the confidence in the hypothesis that an element x is in class C_i of Ω. Confidence measures proposed by each theory fit into a common mathematical framework but are distinguished by their ability to model some nuances of language (probability, plausibility, belief, necessity, credibility). In the next step we combine the confidence measures of various sources for obtaining a single set of confidence measures. Finally, depending on the decision criteria, we choose the class which has the highest confidence.

In the next section, the Demspter-Shafer theory (also called evidence theory) is described.

2.2 Evidence theory

Dempster-Shafer (DS) theory, [2] and [3], is a general framework which offers more flexibility than probability theory. It is suitable to reason with uncertainty and allows to distinguish between uncertainty and imprecision. This is achieved in particular by making it possible to handle composite hypotheses. DS theory is also suitable for combining information from different sources. For a given frame of discernment Ω (a set of classes or hypotheses), the particularity of DS theory is that any source of information can give a piece of evidence on any subset of Ω which can be a simple hypothesis H_i or a union of simple hypotheses. Furthermore, theses hypotheses are not necessarily exclusive, e.g. {friend, enemy, neutral}; Hence, 2^Ω represents the working space for the application being considered since it consists of all propositions for which the information sources can provide evidence. Information sources can distribute mass values on the subsets of the frame of discernment

$$\begin{aligned} m : 2^\Omega &\to [0,1] \\ A_i &\mapsto m(A_i) \end{aligned}$$

is called mass distribution if and only if:

$$\begin{aligned} &m(\Phi) = 0 \\ &\sum_{A \subset \Omega} m(A) = 1 \end{aligned}$$

The derivation of the mass distribution is the most crucial step, since it represents the knowledge on the actual application, as well as uncertainty incorporated in the selected information source [4]:

$$0 \le m(A) \le 1$$

As already mentioned, an information source assigns mass values to any hypothesis of Ω. That is, if an information source cannot distinguish between two propositions A_i and A_j, it assigns a mass value to the set including both propositions ($A_i \cup A_j$).

2.3 Fusion process

The combination rule of Dempster provides a mathematical relation to combine measures of confidence (called mass values) from different sources of information. Mass distributions m_1, m_2 from two different sources are combined with Dempster's orthogonal rule (conjunctive combination). The result is a new distribution, $m = m_1 \oplus m_2$, which carries the joint information provided by the two sources:

$$m(x \in C) = (m_1 \oplus m_2)(x \in C) = \frac{\sum_{A_i \cap A_j = C} m_1(x \in A_i) \cdot m_2(x \in A_j)}{1 - K}$$

$$K = \sum_{A_i \cap A_j = \Phi} m_1(x \in A_i) \cdot m_2(x \in A_j)$$

where K is the measure of conflict between sources and it is introduced in this equation as a normalization factor. The larger K, the more the sources are conflicting and the less sense their combination. As a consequence some authors, Smets in particular [5], require the use of the Dempster combination rule without normalisation. Indeed, when K increases, the fused mass increases although it is not related to an increase of confidence. For this reason, when the normalized rule is used, the conflict K must be included in the decision criteria.

In case of M different information sources B_1, B_2 ...$B_{M,}$ the DS rule is:

$$m(x \in C) = \frac{\sum_{B_1 \cap B_2 \ldots B_N = C} m_1(x \in B_1) \cdot m_2(x \in B_2) \cdots m_N(x \in B_M)}{1 - K}$$

$$\text{where } K = \sum_{B_1 \cap B_2 \ldots B_N = \Phi} m_1(x \in B_1) \cdot m_2(x \in B_2) \cdots m_N(x \in B_M) < 1$$

It is possible to show, through successive combination of mass values iterations, that the operator of Dempster strength the certainty [6]. In other words, when the actions of two separate sensors lead to prefer one hypothesis, the trust granted in this hypothesis will be larger after merging, which is not necessarily the case with Bayes probability.

3. Classification algorithm: Data Fusion Classification (DFC)

The issue is to classify each detected object in one of the two classes: defect (TD) or false defect (FD). In the frame of Dempster-Shafer theory, this corresponds to three hypotheses:
H_1: this object is a defect
H_2: this object is a false alarm
$H_3 = H_1 \cup H_2$: ignorance

Detected objects are described via a list of geometric and grey level based features. Each measured feature is considered as a source of information, and the combination of two or more features is expected to improve the classification performance. In order to be combined, features values must be translated into mass values for each hypothesis of the frame of discernment.

We have developed an original method for automatic attribution of mass values to features measured on a detected object. The implemented method uses each feature histogram as a source of information. The histogram is used to divide the feature's space into regions and specify for each one of these regions a corresponding mass value to the H_1 hypothesis. The complement to one is affected to the H_3 hypothesis, and no mass is assigned to H_2 in order to avoid any conflict during combination.

A smooth transition between regions is obtained through the use of fuzzy sets, such as introduced in [4]. The method can be summarized as follows:

3.1 Learning on a set of known objects

- Features extraction and histogram computation
- From features to regions: computation of true defect proportion p(i) in each histogram interval i. A region is defined by a stable true defect proportion (via a criterion on the derivative of p(i))
- Masses attribution : $m(H_1)$ = p(region), $m(H_3) = 1 - m(H_1)$; $m(H_2) = 0$
- Different masses combination rules: normalised Demspter orthogonal rule (pair wise, three or all features together) and statistical rules (mean and median masses).
- Selection of best sources (successful sources) relatively to an original external inspection system or satisfying an input demand on overall detection rate.

3.2 Validation on a set of unknown objects

- Features extraction using selected sources only
- Attribution of masses
- Masses combination
- Performance test

A more detailed description of the method is given in [7].

3.3 Performance measures

In order to measure the performance of a source, a threshold S is applied on the mass value $m(H_1)$ for each feature, and after each combination. The objects whose mass $m(H_1)$ is above the threshold are considered as true defects, and the others as false. The classification results is then compared to the true decision given by the expert, and the following rates are computed :

- Correct decisions rate (PCD):

$$PCD = \frac{\text{number of true defects correctly classified} + \text{false defects correctly classified}}{\text{total number of true defects and false defects}}$$

- True Defects classification rate (PTD):

$$PTD = \frac{\text{number of true defects correctly classified}}{\text{total number of true defects}}$$

- False Defects classification rate (PFD):

$$PFD = \frac{\text{number of false defects correctly classified}}{\text{total number of false defects}}$$

- Overall classification rate (R):

$$R = \frac{a \cdot PCD + b \cdot PTD + c \cdot PFD}{a + b + c}$$

The rate R was introduced in such a way to give the user the possibility to attribute more importance to the correct classification of either TD or FD. Usually more importance is given to TD detection. Thus, the overall rate R is computed with a=c=1, and b=5.

4. Applications and results

We tested the proposed classification algorithm DFC on 2D radiography dataset and on 3D computed tomography (CT) dataset. In the first case, the DFC system is compared with the Intelligent System for Automated Radioscopy (ISAR), developed by EZRT Fraunhofer and used for radioscopic quality control in the production of castings. Information about the ISAR system can be found in the reference [8].
In this paper, we resume the obtained results. For detailed explanations about each application, please refer to papers [9] and [10].

4.1 Castings radiography dataset

This dataset is extracted from industrial radioscopic images of castings. After segmentation, detect objects are characterized by an array of eleven features (such as area, contrast, volume, etc...). Each object is classified manually into TD and FD by a radiological expert, his decisions are considered as the true decisions.
The dataset is formed of 361 objects. It contains 231 true defects TD including oxides, gas voids and porosities and 130 false defects FD (see figure 2).
The dataset is divided into two parts, a learning database (formed of 65 FD and 115 TD), and a testing dataset (formed of 65 FD and 116 TD).

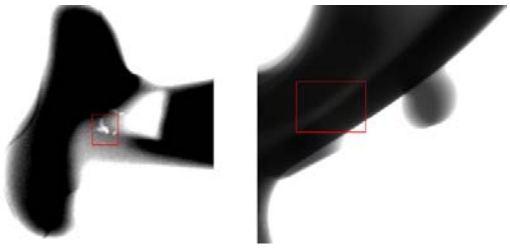

Figure 2. The true or false defects appear as a brighter zone in the red rectangle. On the left side appears a true defect and on the right side appears an artefact caused by the structure of the inspected part.

First the learning process takes place on the learning dataset. Then a selection of the best sources which give the highest overall classification rate R. Afterwards the testing dataset is classified using the selected best sources. Their performance on the learning and testing datasets are compared to the ISAR's performance. Results are resumed in the following table 1 ordered by decreasing overall classification rate R obtained on the testing dataset.

Source	Learning process			Testing process		
	R	PFD	PTD	R	PFD	PTD
Mean Mass	0.992	1	0.991	**0.97**	0.974	0.953
DFC(MaxElongation & InOutContrast)	0.988	0.938	1	**0.955**	0.982	0.841
DFC(Depth2Thickness & InOutContrast)	0.997	0.98	1	**0.941**	0.964	0.846
ISAR	0.925	0.723	0.974	**0.932**	0.723	0.982

Table 1. Achieved overall rates R performances on 2D castings radiography by classification using ISAR and data fusion classifiers.

4.2 Castings 3D CT dataset

The 3D CT dataset is for aluminium castings. It contains 442 object (or potential defects) classified by an expert between true and false defects (see figure 3). Only 44 objects are true defects and the remaining 398 objects are false defects, therefore this dataset is very unbalanced. Existing defects are from different types: porosities, cracks and voids. False defects are due to noise, CT artefacts and the segmentation of structural elements as potential defects. For automatic classification purpose, a total number of 30 features are measured on each object. These features represent the input sources of information for DFC system to automatically classify the entry object as true or false defect.

For the learning and testing processes, the complete dataset is divided into:
- Learning dataset: 226 potential defects consisted of: 200 false alarms and 26 true defects.
- Testing dataset: 216 potential defects consisted of: 198 false defects and 18 true defects.

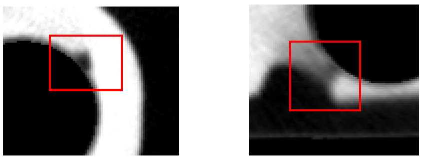

Figure 3. Part of a slice view extracted from a 3D CT volume: on the left side a true defect (surface defect) appears as a darker area and on the right side a false alarm (reconstruction artefact) appears, as a darker area as well.

Results for DFC system are presented in table 2 ordered by decreasing R.

Source	Learning process			Testing process		
	R	PFD	PTD	R	PFD	PTD
Median Mass	0.99	0.97	1	**0.97**	0.89	1
DFC(features 15 & 23)	0.88	1	0.84	**0.63**	0.99	0.5

Table 2. Achieved overall rates R performances on castings 3D CT data by classification using data fusion classifiers.

5. Discussion and perspectives

First of all, the interest of this classification approach is that it is completely "transparent", i.e. it is not a black box such as neural network techniques for example. Classification results obtained are very good in comparison to the actual industrial system as was proved on 2D data (no actual system for 3D data). This is especially noticeable on the false alarms classification rate which is greatly improved with data fusion system. Concerning the combination rule, actually mean or median mass appear better that DFC combination, may be because all the features are used, while with Dempster Shafer rule, using all the features decrease the performance. A complete analysis of the Dempster rule and its behaviour with the number of combined sources is out of scope of this paper. It is worth noting that the mass attribution method (which can be compared to a confidence measure) is a very useful tool to translate any feature into a common space allowing their combination (either using a statistical rule, or the orthogonal sum of Dempster).
Our present work is to apply a similar method to other 3D data, namely from ultrasonic testing for composite materials.

References

[1] Dupuis. O., "Fusion entre les données ultrasonores et les images de radioscopie a haute résolution: Application au contrôle de cordon de soudure". Rapport de thèse: Institut Nationale des Sciences Appliquées de Lyon, 2000.
[2] Dempster A., "Upper and lower probabilities induced by multivalued mapping". Annals of Mathematical Statistics. 38, pp. 325–339, 1967.
[3] Shafer, G., "A Mathematical Theory of Evidence". Princeton University Press, Princeton. p. 297, 1976.

[4] Kaftandjian V. et al., "Uncertainty modeling using Dempster-Shafer theory for improving detection of weld defects". Pattern Recognition Letters, Volume 24, pp. 547-564, 2003.
[5] Smets P., "The combination of evidence in the Transferable Belief Model", IEEE trans. Pattern Analysis& Machine Intelligence, vol 12, n°5,pp. 447-458, 1990.
[6] Gacogne L., "Eléments de logique floue", Paris : Hermes. ISBN 2-86601-618-1, 1997.
[7] Osman A. et al., "Improvement of X-ray castings inspection reliability by using Dempster–Shafer data fusion theory" .Pattern Recognition Letters 32, pp. 168–180, 2011.
[8] Fuchs T., Hassler U., Huetten U., and Wenzel T., "A new system for fully automatic inspection of digital flat-panel detector radiographs of aluminium castings". Proceedings of 9th European Conference on Non-Destructive Testing (ECNDT), Sept. 25 – 29, 2006, Berlin Germany.
[9] Osman A. et al., "APPLICATION OF DATA FUSION THEORY AND SUPPORT VECTOR MACHINE TO X-RAY CASTINGS INSPECTION". In: 10th European Conference on Non-Destructive Testing, Moscow 2010, June 7-11.
[10] Osman A et al., "Automatic classification of 3D segmented CT data using data fusion and support vector machine". 10th International Conference on Quality Control by Artificial Vision QCAV'2011, 28-30 June 2011, Saint-Etienne France.