Improvement of X-ray castings inspection reliability by using Dempster-Shafer data fusion theory
Ahmad Osman, Valerie Kaftandjian, Ulf Hassler

To cite this version:

HAL Id: hal-00878528
https://hal.archives-ouvertes.fr/hal-00878528
Submitted on 13 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Improvement of X-ray castings inspection reliability by using Dempster-Shafer data fusion theory

Ahmad Osman\textsuperscript{1,2}, Valérie Kaftandjian\textsuperscript{3*}, Ulf Hassler\textsuperscript{1}

\textsuperscript{1}Fraunhofer Development Center X-ray Technologies, A Cooperative Department of IZFP Saarbrücken and IIS 8Erlangen, Dr-Mack-Str. 81, 90762 Fürth, Germany.
\textsuperscript{2}National Institute of Applied Sciences INSA- Lyon, Non Destructive Testing using Ionising Radiations laboratory 10CNDRI, Bat. St. Exupery, 25 Avenue Capelle, 69621 Villeurbanne, France.

Abstract:
The aim of this work is to improve the classification of defects in X-ray inspection by developing a new method based on Dempster-Shafer data fusion theory where measured features on the detected objects are considered as information sources. From the histogram of features values on a learning database of manually classified objects, an automatic procedure is proposed to define a set of mass functions for each feature. The spatial repartition of features is divided into regions of confidence with corresponding mass functions. A smooth transition between regions is ensured by using fuzzy membership functions. The whole process is carried out without any expert intervention. Validation takes place on a testing database. Data fusion leads to a significant improvement of classification performances with respect to the actual system.

Key words: X-ray imaging, castings inspection, true defects, false alarms, features extraction, data fusion, confidence levels, mass value, Dempster-Shafer theory.

1. Introduction:

X-ray inspection is increasingly used as a tool for non destructive testing of industrial parts, such as aluminium castings in the automotive sector. In this field, new materials or processes are developed for cost and weight optimization. As those materials are operated near their load limit, it is necessary to check each produced part thoroughly. One main difficulty in X-ray castings inspection is the detection of false alarms (or false defects), especially if very small and low contrasted defects have to be detected. Therefore, efforts have to be made to reduce the rejection rate of good parts without risking to miss true defects. Once the X-ray images are processed for defect detection, the problem can be reduced to a classification problem, in order to discriminate the true defects (TD) from the false defects (FD).

*Corresponding author: Tel. +33 4 7243 8061. Email: valerie.kaftandjian@insa-lyon.fr (V. Kaftandjian).
Several approaches have been proposed for the classification of true defects and false defects, using X-ray inspection. Filtering techniques adapted to specific defect shapes such as bidirectional template matching techniques were used by (Newman and Jain, 1994) for finding gross casting defects. The method is attractive but limited to specific defect types and time consuming.

Neural networks were used by (Lawson and Parker, 1994) for the automated detection of discontinuities in X-ray images of welds. A big advantage of this approach is to learn the neural net on a limited number of defect samples. In another work (Bonser and Lawson, 1998) a defect enhancement technique is introduced using a laws filter matched for round defects and Kirsch oriented filter matched for longitudinal defects was reported. The authors concluded that the neural net is more sensitive to fine defects but gives a high false alarm rate, whereas the filtering approach is more reliable and faster but less sensitive to small defects. Additional investigations of neural networks are described in (Dobrazanski et al., 2005) for flaws in aluminium castings. The optimum network type was selected among multilayer perceptron (MLP), linear network, radial base functions network (RBF), generalized regression neural network (GRNN) and probabilistic neural network (PNN). The best results were obtained employing the PNN and MLP networks with performances of 90% and 81% respectively on the test vectors, and 85% and 94.5% for the validation vectors. Performance is here given by the classification correctness of different defect types, but the false alarms rate is not mentioned.

In (Hernandez et al., 2004) the extracted features of each potential flaw were analysed and assigned to one of the following classes: ‘defect’ or ‘regular structure’ by using a Neuro-Fuzzy method named ANFIS (Jang, 1993) which is used for constructing and training a fuzzy inference model that best classifies new data. The advantage of Neuro-Fuzzy systems is the combination of both properties: non linear learning based on numerical data and handling uncertainties in data. However the fuzzy model identification is very sensitive to which attributes are presented and class distribution in the training data. The developed system was used in casting and welds defects detection. The best performance was obtained for aluminium castings with 95% of true defect classification (57 out of 6860 instances of defect class) and less than 1% (199 non defects classified as defects out of 22876 available) of false alarms rate. However the number of true defects on which the method is validated is small.

Statistical pattern recognition was used in (Mery and Filbert, 2002) to classify castings defects. In statistical pattern recognition the classification is performed using the concept of similarity: similar
patterns are assigned to the same class (Newman and Jain, 1994). Five classifiers were used by
Mery: Linear classifier, Threshold classifier, Nearest neighbour classifier, Mahalanobis classifier
and Bayes classifier. For each detected region, 71 features were extracted to characterise the poten-
tial flaws. The selection of the relevant features is made using sequential forward selection. Accord-
ing to the authors, best performances were achieved with 100% of true defect classification and
910% of false alarms rate for the Bayes classifier, and 94.7% of true defects classification and 4.1%
of false alarms respectively for the threshold classifier. However, the method is validated on only
8176 defects.

A flexible inspection system called ISAR (Intelligent System for Automated Radioscopy) is presen-
ted in (Fuchs, 2006). Defect classification is here performed by individual evaluation of character-
istic features of the segmented defect area, such as size, contrast, elongation. A special type of de-
fect classification was implemented by (Wenzel, 2003). Fourier coefficients are computed in order
to get translation- and rotation-invariant defect features. Those features can be assigned to a certain
group, each representing a certain proper structure of the image. By computing the minimum dis-
tance of an actual defect region to each of the groups, artefacts generated at regular object structures
may be recognized. Our study aims to improve the ISAR system performance, especially concern-
ing the false alarm rate, by introducing data fusion of objects features.

Data fusion theory was already applied by the authors in building a classification system, where the
so-called confidence levels are assigned by a source to each detected object. The fusion between
confidence levels (or mass functions) is based on Dempster-Shafer theory (DS) and the final de-
cision on the sample acceptability was done by a threshold in the confidence level. This approach
was used in improving the detection of weld defects in (Kaftandjian et al, 2003), and in castings in-
spection (Lecomte et al., 2006). The obtained results showed a more precise and reliable decision
by using the data fusion approach. However, the supervision of the expert was necessary to assign
the confidence levels.

In the present work, a new mass value attribution procedure is presented, which is completely auto-
matic, so that the expert supervision is no more necessary. Moreover, eleven features were com-
cuted for each segmented structure in the casting (instead of four features in the previous study).
Each feature is considered as a source of information. By using data fusion, the aim is to build an
automatic classification system. To be able to fuse different sources, a common space where values
are comparable is necessary. An automatic method is introduced to convert from the space of fea-
ture values into the mass value space.
The rest of this paper is organized as follows. Section 2 introduces Dempster-Shafer data fusion theory. Sections 3 and 4 present the proposed classification method and the performance measures. Afterwards experimental results are presented in section 5, discussed in section 6 before concluding.

2. Dempster-Shafer data fusion theory

Dempster-Shafer (DS) evidence theory was developed as an attempt to overcome the limitation of conventional probability theory by handling uncertain, imprecise and incomplete information (Dempster, 1967; Shafer, 1976). It is also suited for combining information from different sources. In DS theory, a fixed set of N mutually exclusive and exhaustive elements, called the frame of discernment, is defined that is symbolized by \( \Theta = \{H_1, H_2, \ldots, H_N\} \). The frame of discernment \( \Theta \) defines the working space for the application being considered since it consists of all propositions for which the information sources can provide evidence through using the so-called mass values. Information sources can also distribute mass values on subsets of the frame of discernment, \( A_i \in 2^\Theta \). Here, \( A_i \) designates a single hypothesis \( H_i \) or union of simple hypotheses (composite hypotheses).

An information source assigns mass values only to those hypotheses, for which it has direct evidence. That is, if an information source can not distinguish between two propositions \( A_i \) and \( A_j \), it assigns a mass value to the set including both propositions \( (A_i \cup A_j) \). This point is precisely the reason for us to choose the DS theory because it reflects the hesitation between two hypotheses.

Obtaining the mass distribution or function \( m(A_i) \) \( (0 \leq m(A_i) \leq 1) \) is the most important step since it represents the knowledge about the current application as well as the uncertainty and imprecision incorporated in the selected information source. The mass distribution for all the hypotheses has to fulfil the following conditions:

\[
\sum_{A_i \in 2^\Theta} m(A_i) = 1
\]

Mass distributions \( m_1, m_2 \) from two different information sources are combined with Dempster’s orthogonal rule. The result is a new distribution, \( m = m_1 \oplus m_2 \), which carries the joint information provided by the two sources:

\[
m(A_i) = (1 - \mathbf{K})^{-1} \times \sum_{A_p \cap A_q = A_i} m_1(A_p)m_2(A_q)
\]
where

$$K = \sum_{A_i \cap A_j \neq \emptyset} m_i(A_i) m_j(A_j)$$

142 $K$ is often interpreted as a measure of conflict between the two sources and introduced in Equation 143(2) as a normalization factor. The larger the $K$, the more conflicting are the sources and the less 144sense makes their combination. As a consequence some authors, Smets in particular (Smets, 1990), 145require the use of the Dempster combination rule without normalisation. However, in this case, the 146relation (1) is no more verified. In our case, the normalised rule is used, and the $K$ value is taken 147into account in the decision process.

148In case of $N$ different information sources $B_1, B_2...B_N$, the DS rule is:

$$m(A) = \frac{\sum_{B_1 \cap B_2 \cap ... \cap B_N \subseteq A} m_i(B_1) m_2(B_2) ... m_N(B_N)}{1 - K}$$ (3)

150where

$$K = \sum_{B_1 \cap B_2 \cap ... \cap B_N \subseteq \emptyset} m_i(B_1) m_2(B_2) ... m_N(B_N) < 1$$

152From a mass distribution, two functions can be evaluated that characterize the uncertainty about the 153hypothesis $A_i$. The belief function $Bel$ measures the minimum uncertainty value about $A_i$ whereas 154plausibility $Pls$ reflects the maximum uncertainty value about this hypothesis. These two measures 155span an uncertainty interval $[Bel(A_i), Pls(A_i)]$, which is called “belief interval”. The length of this 156interval gives a measurement of imprecision about the uncertainty value. Belief and plausibility 157functions are defined from the set of symbolic set of classes $2^\Theta$ to the unit interval $[0, 1]$:

$$Bel(A_i) = \sum_{A_j \subseteq A_i} m(A_j)$$ (4)

$$Pls(A_i) = \sum_{A_j \cap A_i = \emptyset} m(A_j)$$ (5)

161These measures have been sometimes referred to as lower and upper probability functions.

162If only two hypotheses are used, the credibility of each single hypothesis equals its mass value, and 163the credibility of the combined hypotheses (which represents the ignorance value) equals 1. This is 164the case for us when considering our frame of discernment: hypothesis $H_1$ (“this object is a true 165defect TD”), hypothesis $H_2$ (“this object is not a true defect”, i.e. it is a false defect FD), and the 166ignorance is represented by the combined hypothesis ($H_1 \cup H_2 = H_3$).
Once data are combined, a decision rule must be applied. In our case, a threshold is applied on the mass value corresponding to the $H_1$ hypothesis. In the modelling step, specific care is taken in order to avoid conflicts between sources.

3. Proposed approach

The proposed new approach has the goal to automatically classify a detected object into either a TD or a FD, without expert supervision. This is done by analysing the features values extracted from each detected object, assigning to it a confidence level (or mass) and then combine mass values obtained from different features.

In the following, the term “object” refers to the result of the segmentation process, each object being potentially either a true or a false defect. Both the ISAR system and our new classifier are based on the same segmentation stage, so that the pre-processing does not influence the results (pre-processing is explained in section 5.1).

The proposed approach consists of a learning stage, and a classification stage, which are detailed hereafter.

The learning is done on a population of known and manually classified objects. For those objects, a set of features is computed, and the histogram of each feature is built. From these histograms, regions of confidence are automatically defined with fuzzy transitions and masses are attributed, depending on the true defects proportion in the region. The complete procedure is detailed below.

After this learning stage, for each object, a set of mass values is computed for each of its features. It is then possible to consider each feature as a source of information about the set of hypotheses, and combine them. Final decision is obtained by a threshold applied on the combined mass associated to the hypothesis “defect” ($H_1$).

3.1 Learning

3.1.1 Estimation of true defect proportion

For a population of known detected objects (TD and FD), the set of values taken by a feature can be represented in the form of a spatial repartition. The histogram is used to divide this spatial repartition into several regions of confidence. To illustrate the method, we use an individual normal dis-
Distributed random variable \( x \) and two sets: set \( A \) representing the class of true defects TD, and set \( B \) representing the class of false defects FD (see figure 1). An instance of \( x \) can be classified into \( A \) or \( B \). \( x \) can take any value of the codomain of a defect feature. This example will be used for illustration in each step.

The histogram of feature distribution is plotted in figure 2 where a number of 100 intervals was chosen in the feature values range in order to have a precise estimation.

Let \( I \) be the set of intervals forming the feature's histogram. For each interval \( i \in I \), the percentage of instances of the class \( A \) present in this region is calculated using the following function:

\[
P_{A,B}(i) = \frac{h_A(i)}{h_A(i) + h_B(i)}
\]

\( h_A(i) \) represents the number of instances of \( A \) inside \( i \).

\( h_B(i) \) represents the number of instances of \( B \) inside \( i \).

Elaboration of regions of confidence and associated mass functions

At first, each interval \( i \) is considered as a region, and a set of mass functions is attributed to this interval:

\[
m(H_1) = P_{A,B}(i)
\]

\[
m(H_2) = 0
\]

\[
m(H_3) = 1 - P_{A,B}(i)
\]

\( m(H_1) \) is the mass assigned to the hypothesis: instance of \( A \) or TD.

\( m(H_2) \) is the mass assigned to the hypothesis ignorance (instance of \( A \) or \( B \)).

Secondly, subsequent intervals are congregated to form a region of confidence. To be able to merge two intervals, the variation of the function \( P_{A,B} \) is computed. Let \( \Delta P_{A,B}(i) \) be the obtained function (see figure 3).
If the difference between the $P_{A,B}$ values for two adjacent regions is less than a fixed threshold $T$ [first constraint], these two regions are merged and they will have in this case the same mass values:

$$\Delta P_{A,B} = P_A - P_B < T$$

If $P_{i+1} < P_i - \Delta P_{A,B}$, Region $i$ is merged with region $i+1$.

Applying this method shows that the first constraint is not completely effective alone because some regions do not contain enough points to be considered as significant. For this reason a second constraint on the number of points existing in each region is imposed: a region should contain at least a certain percentage of points [second constraint] to be considered as having enough significance. Applying the merging procedure and ensuring enough points in each region, the obtained result is shown in Figure 4, where three regions of confidence are found. The $m(H_i)$ is specified for each region.

### 3.1.3 Definition of fuzzy transitions between regions

It can be observed in figure 4 that the mass value from one region to another may change greatly. Therefore it is important to be able to represent the transition between two regions of the graph in a continuous manner. The theory of fuzzy sets is used to solve this issue, by building membership functions which allow the continuous passage from one region to another (Kaufandjian et al., 2003, 2005).

Let $f_k$ be a source of information used to classify an object in class A or B according to its feature $x$, and $R_1, R_2, \ldots, R_s$ be the set of regions of confidence. $k=1\ldots N$ is the number of features considered as information sources.

Each object has a set of degrees of membership $\mu_i$ ($i=1\ldots s$) to each region $i$ with:

$$\sum_{i=1}^{s} \mu_i(x) = 1$$

Figure 5 illustrates the fuzzy transitions obtained for the regions of figure 4. A fuzzy set is defined and a membership function (classical trapezoidal shape) is built for each region. The slope of the membership function is chosen to be proportional to the difference between the mass values of the adjacent regions.
3.2 Classification

3.2.1 Mass values attribution

The final mass attributed by a source $f_k$ to the object from its feature $x$ is calculated by weighting the mass values of each region $i$ by its degree of membership $\mu_i$:

$$m(\text{object } \in H_1) = m(x / f_k) = \sum_{i=1}^{s} \mu_i(x) \cdot m(R_i)$$  \hspace{1cm} (6)

The final mass function obtained for the regions of figure 4 is represented in figure 5. The mass assigned to hypothesis $H_3$ is the complement to 1 of the mass assigned to $H_1$.

Afterwards the fusion process of different sources takes place.

3.2.2 Data fusion

The Dempster-Shafer rule is used to combine each two sources among $N$ (see eq. 2), three most optimistic sources for each object and the $N$ sources (see eq. 3).

Statistical fusion methods are also used by introducing: the mean mass value and the median mass value.

For an object, the mean mass is the mean value of all the mass values corresponding to this object. It gives a sort of equal influence to the different opinions.

The median mass is the median value of all the mass values corresponding to an object. It allows to introduce a majority decision for the cases where one or few sources are not in accordance with the others.

3.2.3 Decision

To classify an object using the information source $f_k$, a threshold $S$ is applied on its mass value $m(H_1)$ (which represents here the credibility of the hypothesis as said in section 2).

The object is classified as:
303- a defect if \( m(H_i) \geq S \).
304- unknown (defect or not) if \( m(H_i) < S \).
305
3064. Performance measurements and choice of optimal sources
307
308Classification rates are introduced to measure the performance of the different combinations of
309information sources. These measures are calculated for different thresholds \( S \) applied on the mass
310values. For a certain threshold \( S \), an object with mass \( m(H_i) \) is correctly classified if the source
311decision is in accordance with the human decision, this means:
312- the object is a true defect (human decision) and its corresponding mass value is higher or equal to
313the threshold \( m(H_i) \geq S \).
314- or it is a false defect (human decision) and its mass is lower than the threshold \( m(H_i) < S \).
315
316We introduce four measures:
317
318- Correct decisions rate (PCD):
319
\[
PCD = \frac{\text{number of true defects correctly classified} + \text{false defects correctly classified}}{\text{total number of true defects and false defects}}
\]
320
321- True Defects detection rate (PTD):
322
\[
PTD = \frac{\text{number of true defects correctly classified}}{\text{total number of true defects}}
\]
323
324- False Defects detection rate (PFD):
325
\[
PFD = \frac{\text{number of false defects correctly classified}}{\text{total number of false defects}}
\]
326
327
328
329
330- Overall detection rate \( R \):
331
\[
R = \frac{a \cdot PCD + b \cdot PTD + c \cdot PFD}{a + b + c}
\]
332
The rate $R$ was introduced in such a way to give the user the possibility to attribute more import-
ance to the correct classification of either TD or FD. Usually more importance is given to TD detec-
tion.

Using these measures allows comparing the final performance of the method with the existing in-
spection system ISAR introduced in the next section.

5. Experimental results

This section is devoted to present the practical application of our new method. The pre-processing
stage and features are described first, then the data base is presented and analysis of the results after
data fusion is finally detailed.

5.1 Pre-processing stage and selected features

ISAR is a fully automatic system for radioscopic quality control currently in use in the production
of castings. Within the database used in the present paper, evaluation is done by using a reference
image, because thus a higher sensitivity to small defects is possible (Fuchs et al, 2006). The data
processing can be separated into five stages: registration, calibration, image processing, fault
segmentation, and quality assessment.

• Registration
Each inspected part that is processed during inline testing is not necessarily measured at exactly the
same position. Due to mechanical instabilities, long-term changes of the moulds and slight
differences between the moulds in use, an automatic registration step is needed, where the current
test image is registrated to the appropriate reference image using an affine transformation.

• Calibration
In order to allow for quantitative measurement of the depth of faults, i.e. to extract their length in
the direction of the penetrating radiation, a physical calibration object is evaluated. The necessary
characteristic line is created using a step-wedge made out of the same material as the part and with
typically 10 to 15 steps covering the maximum thickness of the objects to be tested.

• Image processing
The reference image is subtracted from the current image. Thereby, all structures that do not appear
within the fault-free reference part will become clearly visible, that is to say potential defects. In
the case of improper registration results however, edges due to the object structure may remain in the
subtraction image, which is one cause of false defects.

• Segmentation
The segmentation step is to assemble the single suspicious pixels into well defined fault regions
appropriately. The analysis is based on the subtraction image. As result of this operation, we will
have for each detected region several features that describe this object and which will be used for
defect classification. Each detected object in this stage is classified by ISAR as true defect or false
defect, and our new classification system starts with the same features.

• Quality assessment
Quality criteria are applied by ISAR to the identified defects, which are defined by the end user for
final decision quality of the part (e.g. acceptance based on size or shape of defects). This post-
processing was not included in our system.
The features which are selected are described as following:

- **Area [mm²]**: is defined as the number of pixels of the detected object multiplied by the size of the pixel defined by the acquisition geometry.
- **Depth [mm]**: the main objective of using the length image is to precise the size of the defect in the X-ray penetration direction. Depth is defined as the mean value of the pixels of the detected object.
- **Volume [mm³]**: is calculated using the area occupied by this object multiplied by its depth.
- **MaxDepth [mm]**: is the highest value inside the defect region.
- **DepthDev**: is equal to standard deviation of the defect’s depth in the defect region.
- **InOutContrast [mm]**: is also measured on the length image and is equal to
  \[ \text{InOutContrast} = |\text{mean(defect region)} - \text{mean(surrounding region)}| \]
- **MaxElongation [mm]**: is defined as the diameter of the minimum covering circle of the detected object.
- **Diff1stDev**: This feature represents the difference, inside the defect region, of the mean gradient value within the reference image to the mean gradient value within the current image.
- **InOutContrastGV**: is the same as InOutContrast but in grey values instead of mm.
- **Thickness [mm]**: is the thickness of the part inside the defect region.
- **Depth2Thickness**: is the ratio between the depth of the object and the thickness of the part in the corresponding region.

### 5.2 Data base description

In this work a database is extracted from industrial images. The detected objects, also called potential defects, are classified manually into TD and FD.

The database is formed of 597 objects. It contains 382 true defects including oxides, gas voids and porosities and 215 false defects.

The database was divided into two parts:

- Learning database: formed of 115 false defects and of 243 true defects.
- Testing database: formed of 100 false defects and 139 true defects.

### 5.3 Application: Learning and classification

Spatial repartitions of four features (Area, Depth, InOutcontrast and Thickness) corresponding to the objects of the learning database are presented in figure 6. All the other features present their own spatial repartition with some overlapping between TD and FD classes. The Area graph exhibits the
better discrimination between classes with true defects mainly having small areas, although some false defects also.

First the estimations of the features histograms take place. Figure 7 illustrates the histograms of the four features represented in figure 6. As already mentioned, overlapping between TD and FD classes is important.

Using the feature's histogram the regions of confidence are built. The results for the four previous features are shown in figure 8 where different number of confidence regions are identified. The values on the right correspond to the mass values \( m(H_i) \) attributed for each point falling into the corresponding region. In addition, membership functions were built for each feature (see figure 9).

Using the estimated regions of confidence, their corresponding mass and membership functions, the single mass values of the learning database objects are computed (eq. 6). Mass functions obtained for the four previous features are also shown in figure 9.

Afterwards the fusion process takes place. As an example, the histogram of mass values obtained from the combination of Area and InOutContrast is presented in figure 10. It can be observed that the translation from feature values to mass values results into an easier separation between TD and FD. Moreover, the fact that the mass values are included between 0 and 1 facilitates the interpretation.

Subsequently, the classification of the learning database objects using different thresholds on the mass values given by each source (single and combined sources) is performed.

**5.4 Performance results: Learning database**

The performance measures for ISAR on the learning database are: PTD=0.978, PFD=0.723 and R=0.928 (with \( a=c=1, b=5 \)). This choice of parameters a, b and c allows to give more importance to the classification of true defects, which is an industrial requirement.

Individual features performance after masses attribution is shown in Table 1. As it can be noticed, some of the features are better than the others (Area in particular, as was visible from its histogram), nevertheless, all of them will be kept for fusion due to the fact that they could bring some information.
ISAR has, in the tested configuration an excellent performance in the detection of true defects but it has a false detection rate of about 28%. After the data fusion process, PTD, PFD and R are measured for all the single and combined sources (total of 70 sources). Several combined sources give better classification rates compared to the original ISAR decision (see figure 11). All these combinations present better PTD, PFD and R than ISAR. Grouped into a set \( L \), they are then used as sources of information to classify the objects of the testing database.

The best classification performance as measured by the overall rate R is obtained by the DS combination of the features Area and Depth with 96%, for a threshold on the mass values \( S=0.8 \), instead of 92% for ISAR. Table 2 gives the results obtained for PTD and PFD of this combination which are all higher than ISAR's measures.

When considering the performance of true defects classification only, the best source is the DS combination of Area and Thickness with PTD of 99%, for the threshold \( S=0.9 \), instead of 97% for ISAR. Here also PFD and R are both higher than ISAR's measures (see table 3).

The best performance for false defects classification only is obtained by the mean mass with a PFD of 98%, for the threshold \( S=0.7 \), which is 28% better than ISAR's PFD. However, PTD is in this case lower than ISAR's PTD which is not acceptable for this application (see table 4), therefore the mean mass is not presented in figure 11. Nevertheless, this source will be considered for testing.

5.5 Validation: Testing database

The testing database is formed of 139 true defects and 100 false defects.

ISAR shows similar performance on the testing database with: PTD=0.980, PFD=0.725 and R=0.930.

After the learning step, the set of sources \( L \) are used to automatically attribute mass values for the new objects of the testing database. Afterwards data fusion and classification take place.

All combined sources improve with an important percentage the classification of false defects and also are equal or even exceed ISAR performance on true defects detection (see figure 11 (b)). In the learning database, the optimal overall source was the fusion between Area and Depth. On the testing database, this combination presents a rate R of 96.37% (for the threshold \( S=0.8 \)), instead of 93.07%
for ISAR, with a very small difference from the optimal overall source on the testing database which is the fusion between Area and MaxElongation with R=96.86% (see table 5).

The optimal combination for the detection of true defects in the learning database (DS combination of Area and Thickness) is still the best combination for the detection of true defects in the testing database with a PTD of 0.99% for the threshold S=0.9 (see table 6). The improvement in the detection of true defects is 1% relatively to ISAR, and 12% in the detection of false defects. The optimal combination for the detection of false defects in case of testing database is still the mean mass value for the threshold S=0.7 (see table 7). The improvement in the detection of false defects is 25.8% relatively to ISAR, but a loss of almost 4.5% in the percentage of true defects detection occurs.

To illustrate two cases where ISAR fails, while data fusion gives the good result, two objects are chosen. Figure 12a shows a false defect classified as a true one, and figure 12b shows a true defect classified as a false (the object appears brighter than the background inside the rectangular zone).

6. Discussion

The first point to emphasize is that the results obtained during the testing phase confirm the robustness of the method. All sources selected during learning keep their high performance during testing.

In our previous work (Kaftandjian et al., 2003) only 4 features were selected as best candidates to discriminate between TD and FD. This selection of features (contrast to noise ratio, area, elongation and position) was done by the expert based on his knowledge of X-ray inspection. The drawback was to require the expert's supervision during learning. Here the approach is completely automated and all the features are considered without a priori information. The obtained performance cannot be compared with (Kaftandjian et al., 2003) due to the fact that the application was not the same.

When considering which features give the best classification rates, one can be surprised that the contrast feature is not among the best ones, although contrast is usually considered as having the main physical meaning. Indeed it is impossible to justify a posteriori why some features were better than the others when combined. This is due to the fact that our knowledge is only valid for TD, and the classification of FD cannot be based on a physical basis. This is precisely the difficulty of FD detection and the reason why even the expert find them difficult to classify. Therefore an important advantage of our method is that it adapts itself to all the objective data that are measured on the true
and false defects, without any a priori knowledge.

Another fact to mention is that the developed method is general in the sense that it can work for any inspection method. The translation from the feature space to the mass values space allows to combine any information sources independently of their origin (ultrasonic data, radiographic data...).

7. Conclusion

A new implemented method able to automatically detect regions of confidence from a feature’s histogram and to affect degrees of confidence (mass values) to these regions was presented. The automatic learning process is based on a classified database. The method consists in translating a feature value into a confidence or mass value. These mass values are used within the fusion process. The method was applied on a learning database and validated on a testing database for a casting application. As could be shown, several feature combinations are exceeding the current ISAR decision. Using the data fusion theory, while conserving at least the 98% percentage of classification of TD given by ISAR, an improvement from 72% to 92% (see table 4) in the percentage of FD classification was achieved. An improvement of 3.8% of the overall rate R is obtained.

Some combinations can optimize even better either the TD or the FD classification. A PTD rate of 99% and PFD of 98.3% was obtained separately. Further work is needed to find an optimal decision rule which would associate the best TD and FD classification rates.

In a future work, a study of the robustness of this method for another application will be done. Further investigation will also be conducted to measure the influence of the constraints on the derivative variation and on the percentage of points inside the regions of confidence on the results.

Acknowledgements

This work was co-financed by the European Union and the Free State of Bavaria within the project ZeLuR.

References


Wenzel T. and Hanke R., "Texture analysis for classification of cast defects in x-ray images". 2nd workshop "NDT in progress" International Meeting of NDT Experts, October 6-8, 2003, Prague, Czech Republic.
Fig. 1:
Two normal distributions where a star point represents a member of the class A and a cross point represents a member of class B.

Fig. 2:
Feature's histogram of class A (full line) and class B (dotted line).

Fig. 3:
Illustration of $p_i(A)$ (full line) and $p_i(B)$ (dotted line) corresponding to the histogram of figure 2.

Fig. 4:
Regions of confidence after merging and their corresponding mass values $m(H_j)$.

Fig. 5:
Membership functions corresponding to the regions of figure 4 (full line) and mass functions (dotted line).

Fig. 6:
Spatial repartition of True and False defects for Area (high left), Depth (high right), InOutContrast (low left) and Thickness (low right).

Fig. 7:
Histograms for Area (high left), Depth (high right), InOutContrast (low left) and Thickness (low right).

Fig. 8:
Regions of confidence and corresponding mass values $m(H_i)$ for Area (high left), Depth (high right), InOutContrast (low left) and Thickness (low right).

Fig. 9:
Membership functions (solid line) and mass functions (dotted line) corresponding to the confidence regions of Area (high left), Depth (high right), InOutContrast (low left) and Thickness (low right).

Fig. 10:
Histogram of mass values obtained from the combination of Area and InOutContrast. The majority of true defects have mass values close to one.

Fig. 11:
Classification rates obtained on the learning database (a) and on the testing database (b). The selected sources are those giving a better overall classification rate than ISAR, and the same set of sources are still better on the testing database.
False defect classified as a true one by ISAR (a) and true defect classified as false (b). In those two cases, data fusion give the good result. The true or false defects appear as a brighter zone in the rectangle.

667

**Table 1:**
Performances of single features after mass attribution, in the learning database. Features are classified with respect to the overall rate R

671

**Table 2:**
Combination giving the highest overall rate R in the learning database.

674

**Table 3:**
Combination giving the highest true defects detection rate PTD in the learning database.

677

**Table 4:**
Combination giving the highest false defects detection rate PFD in the learning database.

680

**Table 5:**
Combination giving the highest overall rate R in the testing database.

683

684

**Table 6:**
Combination giving the highest true defects detection rate PTD in the testing database.

687

**Table 7:**
Combination giving the highest false defects detection rate PFD in the testing database.

690

691