How features extraction can complement defect detection : the concept of Defect Confidence Index

Gwenaële Lecomte [1], Valérie Kaftandjian [1], Laurence Chatellier [2], Pierre Wagner [2], Roch Menand [3]

[1] CNDRI, Laboratory for NDT using ionising radiations, INSA-Lyon,
Batiment Saint Exupéry, 20 avenue A. Einstein, F-69621Villeurbanne, France,
Phone: +33 (0)4 72 43 80 61, Fax +33 (0)4 72 43 88 22;
e-mail: gwenaele.lecomte@insa-lyon.fr, valerie.kaftandjian@insa-lyon.fr,

[2] EDF R&D,
6 quai Watier, 78401 Chatou cedex
e-mail : laurence.chatellier@edf.fr, pierre.wagner@edf.fr

[3] EDF CEIDRE
Centre d'Expertise et d'Inspection dans les Domaines de la Réalisation et de l'Exploitation
Bâtiment Bohr, 2 rue Ampère, 93206 Saint-Denis Cedex 1
e-mail: roch.menand@edf.fr

Abstract
We have investigated how to quantify the presence of a defect by measuring its geometrical and grey-level features. From a selection of best features, we define a Defect Confidence Index (DCI) varying continuously between 0 (no confidence = this is not a defect) and 1 (total confidence = this is a defect). Two case studies are chosen for illustration. The first concerns automatic defect detection for castings inspection. The problem is to separate true defects and false alarms after image processing. The DCI is helpful in order to measure the inspection performance and find the best decision threshold. The second case concerns radiographic films where defects must be detected visually by operators. We selected several features on simulated radiograms using not only the grey-level variations due to the defect, but also the background. We define three DCI and compare it with the visibility of the operator.

Keywords: features extraction, image analysis, radioscopy, radiography, detectability, visibility, confidence level

1. Introduction

For most applications, the smallest defects to be detected on radiographic films or radioscopic images present a low contrast. On the one hand, for radiographic films where defects are detected visually by operators, the performance of defect detection is known to be very efficient but the factors which affects the visibility are not well studied. The human detection process is subjective and not always reproducible, especially for defects barely visible. On the other hand, in the case of automatic detection of defects either on digital radioscopy or digitised films, the choice of image processing is a compromise between the detection of true indications and false alarms, and thus, although reproducible, the detection of smallest details is difficult.
These difficulties of detection are linked to the defect itself (size, nature), the sample (thickness variation of the piece, material, etc…), and the parameters of image or film acquisition (energy range, unsharpness, etc…).
For both cases, there is a need to quantify the presence of a defect from its image or appearance on the film. The aim is not to study why the human eye will detect this

defect or not, but more to find a relationship between objective measurements on the optical density variation and manual defect visibility. It is expected that these objective measurements will also help to discriminate between true defects and false alarms from a statistical analysis after image processing and feature extraction.

The goal of this study is to identify and evaluate different parameters on digital images in order to find the best descriptors of the defect presence. To quantify the defect presence instead of describing it by its parameters, we introduce the concept of Defect Confidence Index (DCI) varying continuously from 0 (no confidence in the fact that the indication is a defect), to 1 (total confidence).
Various DCI can be computed from the features extracted on the image, depending on the application. Two applications are presented. The first one focuses on the improvement of inspection performance from a better distinction between true defects and false alarms. The second one consists in determining the defect detectability from the simulated film, i.e. to determine whether the defect indication can be seen or not by controllers. One originality of the work is to take into account the presence of an image background gradient as a factor influencing the visibility of the defect.

2. Identification of the essential parameters

- This paragraph presents the selection of the most influential parameters both for the determination of expert visibility, but also for the automatic image processing. The idea is to identify the most important features of the defect, and to take into account the fact that the image can present a background gradient or not (due to the normal thickness variations of the sample).

The parameters are computed on different images defined below :

- the initial or **raw image *I***, corresponding to the real digitised or simulated radiogram or radioscopic image of the sample under test (thus potentially with a defect),
- the **reference image *Iref***, corresponding to the sample without defect,
- the **defect image *Idef***, corresponding to the optical density or grey-level variations introduced by the presence of the defect (*Idef* = *I* – *Iref* if the defects are brighter than the background, or *Iref* – *I* if the defects are darker)
- the **contrast image *Ic*** is computed as the ratio between the defect image and the reference. $Ic = |I - Iref| / Iref$

The different images are illustrated in figure 1 with an example of more dense defects (darker spots).
The image *Iref* can be acquired on a reference sample, or simulated if the CAD file of the sample is available, or obtained by filtering the initial image *I*. A morphological opening (respectively closing) of adequate size is usually convenient for filtering images with brighter defects (respectively darker). The great advantage of computing *Iref* by filtering the raw image *I* is that no registration step is required.

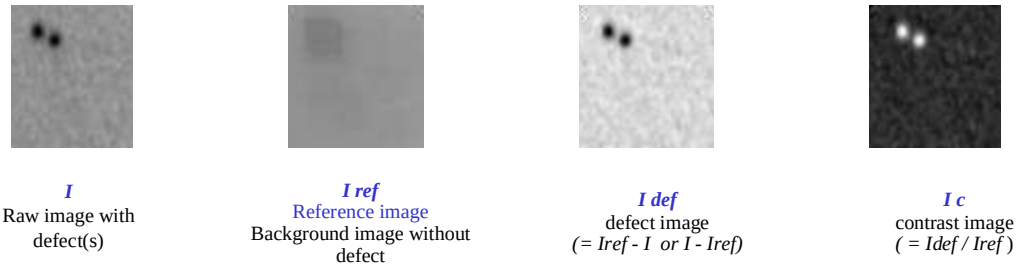

Figure 1 Example of a radioscopic image with darker defects : from left to right : initial image I, background image without defect, Iref, defect image Idef , contrast image Ic. The Iref image was obtained by a morphological closing of image I.

2.1 Defect segmentation

Before computing any parameter, it is necessary to define the "defect area" and its neighbourhood.
In order to consider only the most significant part of the defect in terms of grey-level variation with respect to the background, we use the contrast image *Ic*. All pixels whose *Ic* value is greater than 95% of the maximum value are marked as belonging to the defect (segmentation). The mask obtained after defect segmentation is denoted *Def95*.
Another approach would be to apply a predefined threshold on the image *Idef*, obtained by subtracting *Iref* from the image *I*. When *Iref* is obtained by a morphological opening or closing, this adaptive thresholding process is known as the top hat transform [1]. The advantage of our method is to obtain an automatic segmentation based on a fixed relative grey-level variation of the defects with respect to the reference background.

The area selected for the defect neighbourhood is obtained by a 9 pixels-dilation minus a 6 pixels-dilation of *Def95* (this area corresponds to a 3 pixels width band around the defect). The size of the neighbourhood is thus linked to the defect size.
Figure 2 illustrates a defect mask *Def95* and its neighbourhood *NeighDef95*.

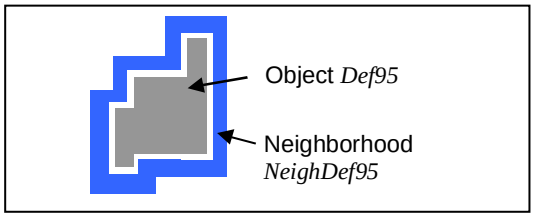

Figure 2 Defect mask and its neighbourhood

2.2. Features definition

We can separate the features in three types :

2.2.1 Defect geometrical features

- Dimensions (length, width, perimeter, area)
- Shape (elongation, compactness),
- Position, and distance to a reference point in the sample,

☞ these features are obtained from *Def95*

2.2.2 Defect grey-level features

- Average grey level value *NGdef*,
- Contrast with the neighbourhood *C95*,
- Contrast to Noise Ratio CNR (where the noise corresponds to the standard deviation of the grey-levels in the neighbourhood)
  - ☞ *NGdef* is measured on *Idef* using the pixels defined by *Def95* (defect area),
  - ☞ *C95* is computed as the difference of average grey-level between the defect area and its neighbourhood in the initial image *I*, divided by *Iref*.

2.2.3 Background grey-level features

In order to quantify and analyse the influence of the image gradient on the visibility, some descriptors have been developed to evaluate :

- the background gradient
- the variations of the background gradient (based on the analysis of the variance of the grey-levels).

Estimation of the gradient in the neighbourhood of the defect
We can select the neighbourhood evaluated from the reference image *Iref* (in order to be free from the defect) or the neighbourhood in the initial image *I*. The gradient is estimated as the maximal grey level difference in the neighbourhood *dNG* (an increase of the background gradient induces an increase of the grey level difference).

Estimation of the variation of the background gradient
The variations of the background gradient have an effect on the visibility. Indeed, for the same background variation (same difference of optical density in the background), a defect is more easily detected on a background having a constant gradient rather than on a background with a gradient variation (like in the presence of a weld for example). The figure 3 illustrates this feature. The second defect will be more easily detected than the first one. Thus, it is essential to introduce a parameter characterising these variations of background gradient to define a confidence level.

In order to measure this variation, we compute the variance image. The variance image *varI* is defined as follows : for each pixel (x,y) of *varI*, we associate the variance of the image *I* evaluated on an 9x9 area centered on (x,y).

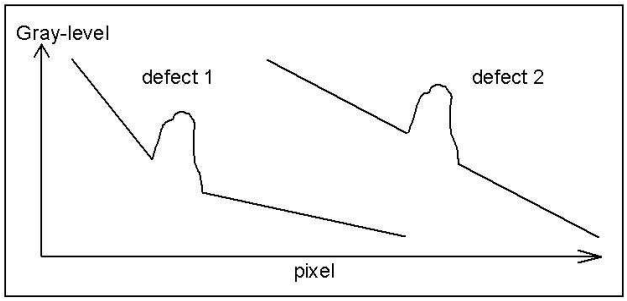

Figure 3 Two defects with the same background gradient but with a variation of gradient (defect 1) or with a constant gradient (defect 2)

The figure 4 below presents, for one of the studied cases, the variance image of the reference image *Iref* and of the raw image *I,* and the vertical grey-level profiles in the initial image (crossing the defect) and in the reference image. The change in the background gradient is clearly visible in the variance image of Iref. The subtraction of the variance image of I and Iref respectively is also shown for illustrating the interest of this process for defect enhancement.

- the rupture of the gradient on the reference image *Iref* induces a rupture of the values on the associated variance image,
- the area of the defect corresponds to the area where the differences between the two variance images are important.

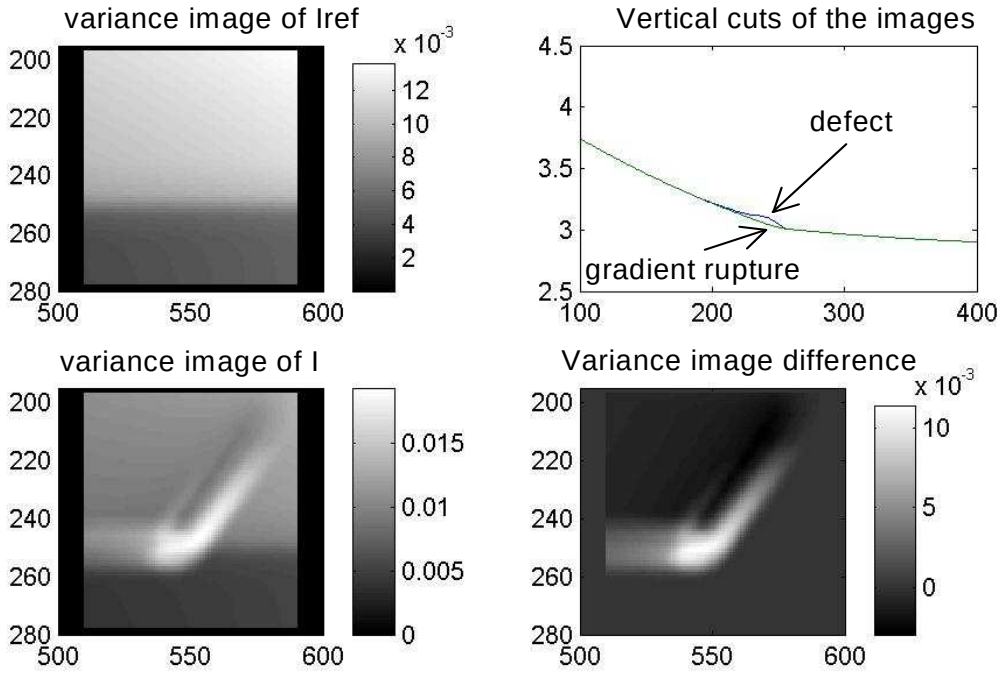

Figure 4 left : Variance image of the reference image and the raw image; right : grey-level profiles in the image I (crossing the defect) and Iref, and difference of the two variance images, enhancing the defect.

The variance of the variance image is then characteristic of the variation of the background gradient.

3. Definition of a Defect Confidence Index

3.1 DCI based on a statistical analysis of the grey-levels

For a sample without defect, the statistical fluctuations of X-ray transmission induce a grey-level variation which is known to follow a gaussian distribution. If μ and σ denote respectively the mean and standard deviation of the distribution, this means that 95% of the levels are included in the range $[\mu - 2\sigma, \mu + 2\sigma]$.

We introduce a Defect Confidence Index based on the gaussian distribution measured on a sample without defect . This DCI is aimed at quantifying to what extent the grey-level of the object is far from the mean value of an homogeneous thickness, as

compared to the standard deviation. The Defect Confidence Index represents thus the confidence level assigned to the fact that the object is potentially a defect. The term « *Measure* » refers to the information available on the object, namely its grey-level as far as radiography is concerned.

$$DCI = 1\text{-}exp(-(Measure-\mu)^2 / (\alpha.2.\sigma^2)) \qquad (1)$$

α is a weighting factor which is used to adapt the DCI distribution to the reliability of the method itself, i.e., the confidence we have on the method.

The Figure 5 illustrates the DCI obtained with a weighting factor $\alpha = 1$ or 2.

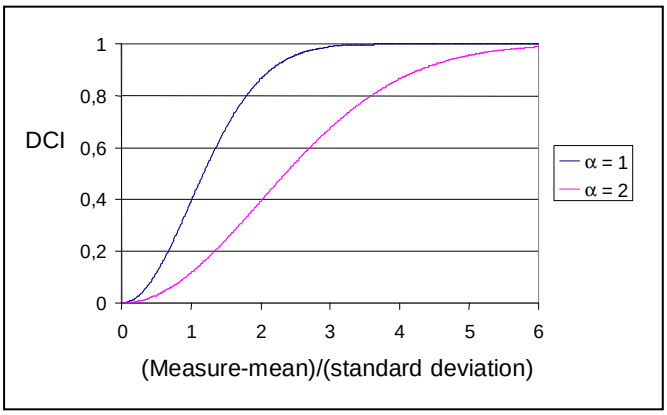

Figure 5 DCI obtained using the relation (1) with a weighting factor of 1 or 2.

For $\alpha = 1$, one can see that the DCI obtained is maximal for a deviation to the mean greater than 3σ (this is thus the initial gaussian distribution). $\alpha = 2$ corresponds to a more cautious modelling, as a confidence level of about 70% is assigned for a deviation of 3σ.
The complement to 1 of the DCI refers to an ignorance level. This reflects the fact that a low deviation (measure – mean) with respect to the noise could either be a small defect indication or a normal measurement.

3.2 DCI based on the normalisation of any selected parameter

For other configurations, when a feature behaviour cannot be easily linked to a parametric representation like the gaussian distribution for example, a simple normalisation relation can be applied.

DCI 1 : Case of a parameter for which a high value indicates a higher confidence in the defect presence

$$DCI1(parameter\ P) = Pvalue / P_{max} \qquad (2)$$

where P_{max} is the highest possible value of the parameter P.

DCI 2 : Case of a parameter for which a low value indicates a higher confidence in the defect presence

$$DCI2(\text{parameter } Q) = 1 - Q/Q_{max} \quad (3)$$

where Q_{max} is the highest possible value of the parameter Q.

4. Application 1 : use of DCI for automatic defect detection

The following results were obtained in the frame of the QUME[1] EU funded project whose aim was to improve the inspection performance of cast parts by combining radioscopic imaging, X-ray spectrometry and vibration analysis. Our study here deals with the processing and combination of data coming from different images acquired by radioscopy, and especially how we defined a DCI associated to each detected object. The data fusion approach with the other NDT modalities was presented in [2].

4.1 Acquisition data

171 samples were studied, from two production batches, among which 30 samples are said not acceptable by the manufacturer (Stampal), in accordance with ASTM standard, representing 62 defects. Different defect types have been investigated, mainly shrinkage cavities, gas cavities, with various gravity levels. 1 inclusion and 2 cold junctions were also present. Those defects were reported by a manual X-ray investigation (where the right projection angle and right X-ray energy are manually chosen for each sample). The case of cold junctions is particular, as it is well known that this defect is very hard to detect by X-rays (this is why the choice of the right angle is so important).
The sample is examined by radioscopy, using a microfocus X-ray tube and an image intensifier. Four orientations are selected to put the emphasis on the critical areas of the sample (see figure 6). The central part is partly visible in the four views.

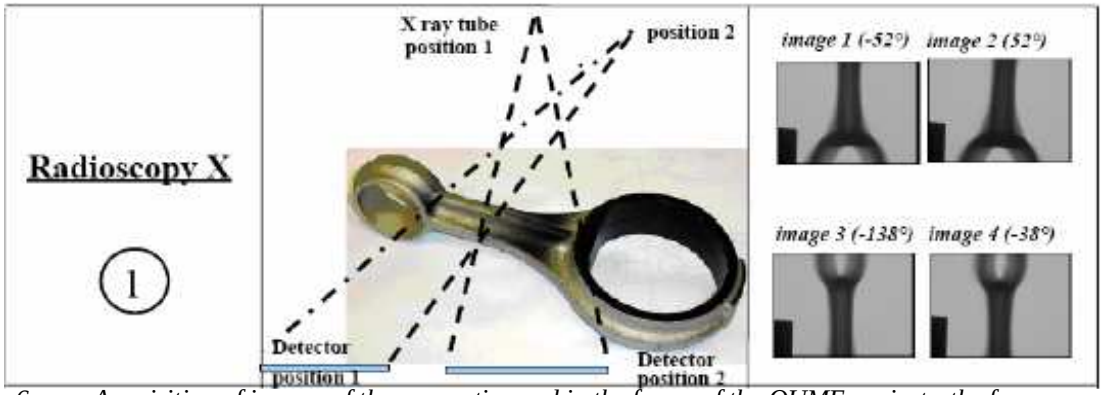

Figure 6 Acquisition of images of the connecting rod in the frame of the QUME project : the four angles were selected for a best inspection of the critical zones.

[1] QUME is the acronym for the GROWTH RTD project entitled: On-line process and Quality optimisation for the manufacturing of cast MEtallic parts, G1RD-2000-00444. Partners : InnospeXion ApS (DK), Carl Bro A/S (DK), RISOE (DK), Univ. Liverpool (UK), Monition Ltd (UK), ISQ (PT), Stampal Spa (IT), IFG (DE), INSA-CNDRI (FR)

The segmentation is done automatically by adaptive thresholding after noise reduction [3]. The reference image *Iref* is obtained by morphological opening. An example of raw and processed image is shown on figure 7. A true defect is detected, but also two false alarms.

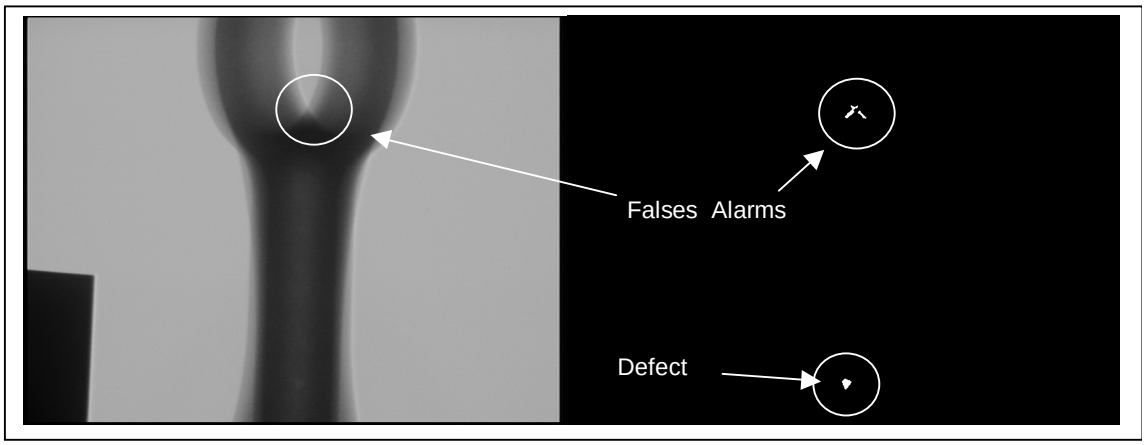

Figure 7 Raw and binary image after automatic segmentation.

4.2 Features extraction

The Figure 8 below shows the features of the detected objects (area versus CNR, and average grey-level versus Distance Dt), where one can see that a direct discrimination of true defects and false alarms is not trivial. The distance Dt is the distance from the object to a reference point chosen in the sample, and computed in the piece coordinate system. Although this distance is specific to this application, the method can be extended to any particular point in a given sample. The originality here is to compute this distance not on the projected image but in the piece coordinate system, which allows us afterwards to combine several radioscopic views after object matching.

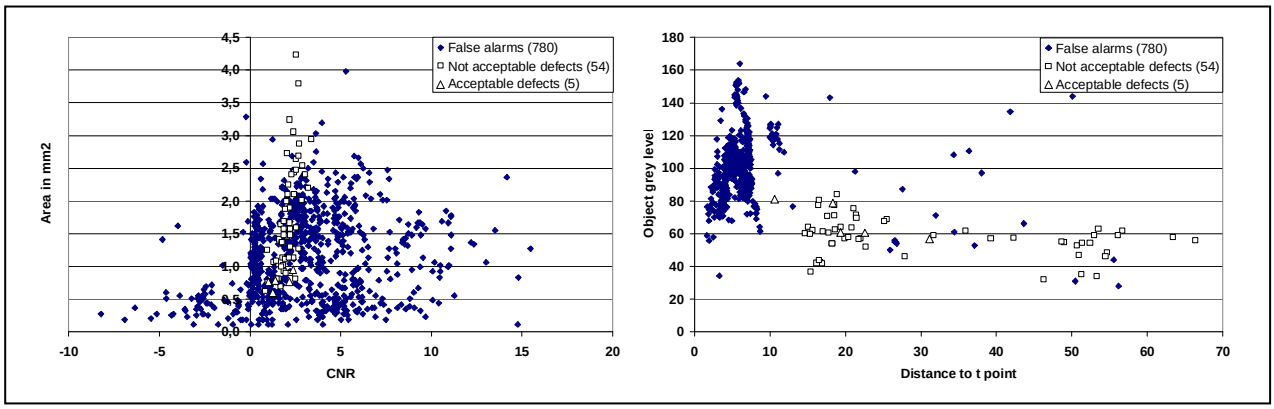

Figure 8 Features representation of the detected objects : (left) area versus CNR and (right) object grey level versus the distance Dt.

4.3 DCI values

Each detected object is assigned a DCI value according to relation (1), where the Measure is the mean grey level of the pixels constituting the object. μ and σ are respectively the mean and standard deviation of the grey-levels of the object neighborhood. Thus, the relation (1) can be expressed using the Contrast to Noise Ratio (CNR) of the object and its neighbourhood (relation (4)). A weighting factor of 1 has been selected, which means that a CNR above 3 yields a confidence index of 1

$$DCI = 1 - \exp(\frac{-CNR^2}{2}) \qquad (4)$$

In the context of the QUME project, the manufacturer has an acceptability level according to the size of the detected objects. Objects with an area inferior to 1 mm^2 in the radioscopic image are not considered critical. It was also observed (see the previous Figure 8) that some false alarms were systematically featured by a grey-level too high, a CNR too high, and a small distance Dt.
To summarize this knowledge (which is specific to the application), any detected object was assigned a DCI=0 if :

- its area is inferior to 1 mm^2,
- its CNR is greater or equal to 4,
- and/or its grey level is superior to GL_0,
- and/or its distance Dt is inferior to 10 mm.

The Figure 9 shows the DCI versus CNR graph, where it appears that true defects get a confidence level in the defect hypothesis (although not very high for some of them if low contrasted) and most false alarms get a null DCI, as well as acceptable defects. Thus, the knowledge we have from the features is well translated in the DCI values.

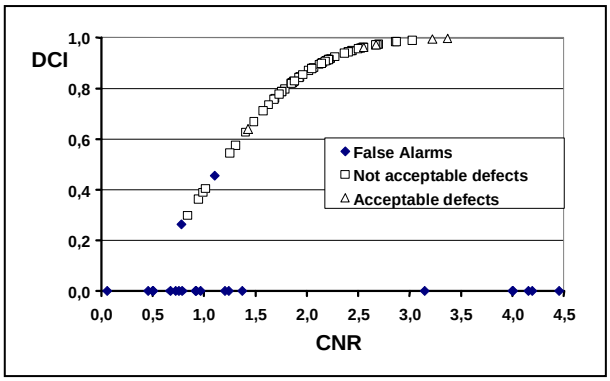

Figure 9 Representation of the DCI versus the CNR, for the objects of Figure 8.

4.4 Final decision on sample acceptability

In this study, four images per sample were acquired, in such a way that the critical zones of the sample were inspected in several views. Hence, a matching procedure was implemented in order to check if an object was detected on more that one image.
Among 839 detected objects, 591 object couples were correctly matched.
The DCI values were combined using the orthogonal sum of Dempster [4][5] so that a unique DCI value is computed for each object. As expected, when an object is detected

in several images, its confidence level increases. More details about this combination can be found in [2].
The following figure 10 shows an example of three fusion results : case 1 reflects the case of a middle confident defect in view a, which is detected with a better confidence in view b, and thus, the combination results in a perfect confidence at the end; case 2 and 3 show also an improvement, although the confidences are already high in each view.

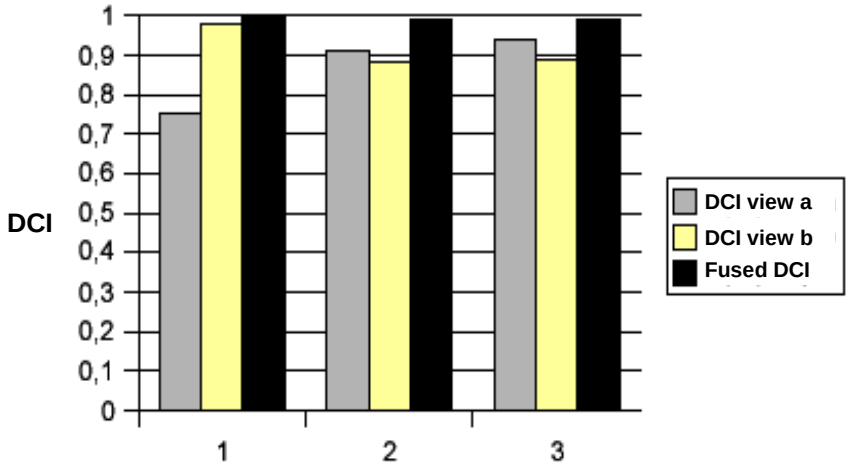

Figure 10 Fusion of the DCI for three different defects matched in two views.

For the sample acceptability, it was decided to apply a decision criterion on the highest DCI value among all the detected objects in the sample. 31 samples have a non null DCI value (Figure 11). Among them, 27 are said unacceptable by the manufacturer, two samples have acceptable defects and two samples are without defect. If a DCI threshold of 0.6 is defined, only defectuous samples are rejected, but one of them is said acceptable by the manufacturer.

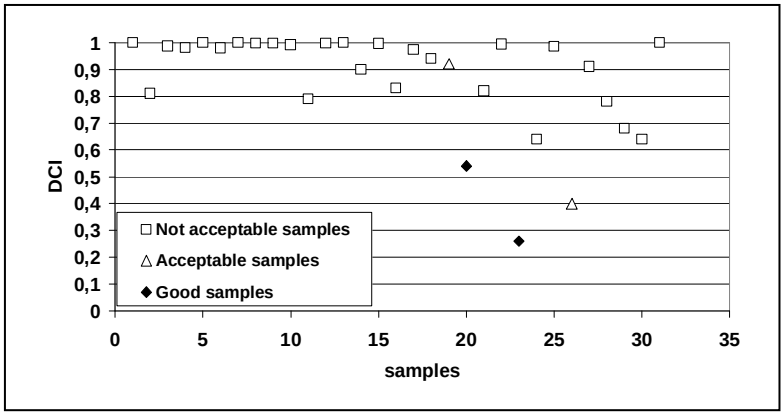

Figure 11 Plot of the 31 samples having a DCI ≠ 0 , labeled with the end user acceptability

The other 140 samples have a DCI = 0. Among these samples three have unacceptable defects (cold junction and inclusion), which were not detected by the automatic radioscopic control as expected.
With the decision threshold of 0.6 on the DCI, the global performance obtained on the 171 samples is thus a probability of detection of 90 % (=27/30) and probability of false

rejects of 0.7 % (=1/140). However, the only false rejected sample contains a defect, which means that our automatic inspection is slightly more severe than the manufacturer control.
The DCI is thus a useful tool to summarize all the knowledge included in several parameters, and translate this information into a comprehensive unique parameter. The decision making can be more easy and objective using this DCI concept.

5. Application 2 : use of DCI to compare with human visibility

For this study, we used simulated data obtained with the EDF computer code MODERATO.
A first study was done to determine, from the simulated film, whether the defect indication can be seen by controllers qualified according to COFREND Level II in radiography [6]. A detectability criterion was defined, based on defect grey level features in a predefined area of 1,6 mm^2. This criterion is valid only if the background of the defect indication is relatively smooth and its use until now has always led to results in agreement with the expert diagnosis.
The present study was carried out in order to evaluate the influence of important background variations on the visibility and to adapt the detectability criterion to these background variations (such as the case of a weld for example). We used a set of data composed of images with more important and penalising background gradients and with simulated defects (notches). These gradients characterise the thickness variations of the irradiated component part. It is worth noting that those simulations correspond to real cases of inspection, for which we have an expert report concerning each defect, with the indication “visible”, “not visible”, “limit of visibility”.

In the following picture, the results of visibility are presented with a full symbol if the defect is detected and with a void symbol if the defect is barely visible (in limit of visibility). Indeed, for this set of data, all the defects were detected by the expert.
The scale of the values refers to the optical density of the simulated radiogram.

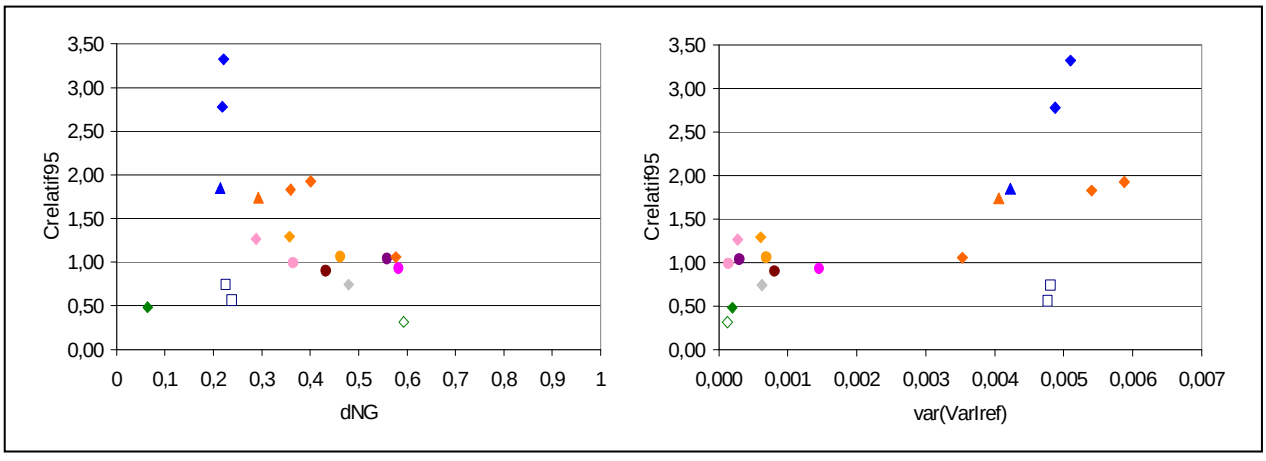

Figure 12 Representation of measured features, contrast Crelatif95 versus image gradient dNG and contrast Crelatif95 versus variations of the background gradient var(VarIref).

At this step, we define three confidence indexes, based on the three selected features (contrast, background gradient and variation of the background gradient).

- **DCI 1 : Defect confidence index characteristic of the defect contrast**

$$DCI1(Crelatif95) = Crelatif95 / Crelatif95_{max} \quad (5)$$

with - *Crelatif95* : contrast (*Idef/Iref*) evaluated in the heart of defect area (*Def95*)
- $Crelatif95_{max}$: maximal contrast

- **DCI 2 : Defect confidence index characteristic of the reference image gradient**

$$DCI2(dNG) = 1 - dNG \quad (6)$$

with *dNG*: : background gradient in the reference image (maximal difference measured in the outer neighbourhood *NeighDef95*)

- **DCI 3 : Defect confidence index characteristic of the gradient variations**

$$DCI3(Var(VarIref)) = 1 - \sqrt{\frac{\mathrm{var}(VarIref)}{\mathrm{var}(VarIref)_{max}}} \quad (7)$$

with - *var(VarIref)* : reference image (Iref) variance
- $var(VarIref)_{max}$: maximal variance

The three DCI obtained for each defect are summed up on the figure13 below.

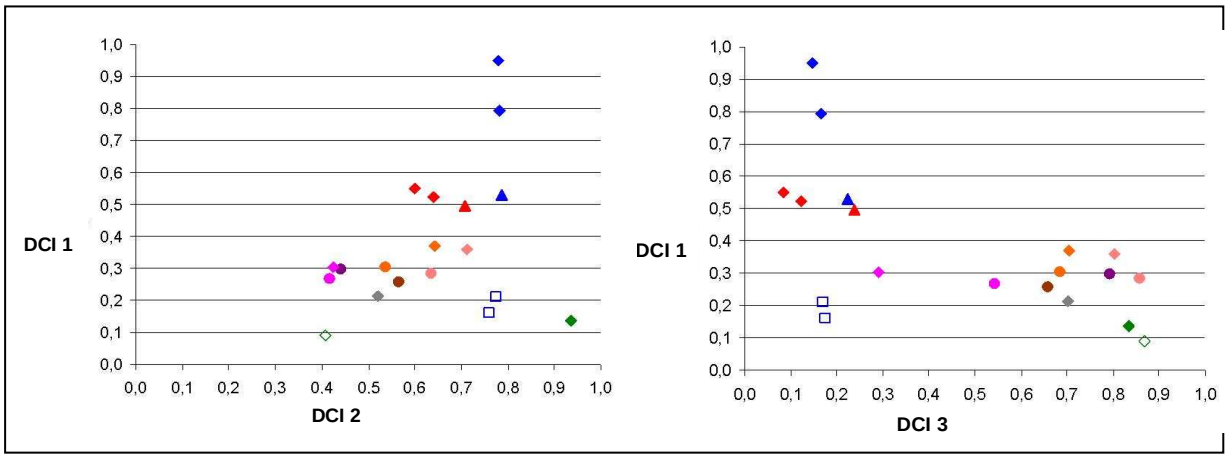

Figure 13 Defect Confidence Index DCI1 versus DCI2 (left) and DCI1 versus DCI3 (right).

We can note that the first index takes into account only the grey level values in an area considered as the heart of the defect but does not use a minimal area as in the criterion used previously (evaluation in a 1.6mm² area). Indeed, the segmentation process selecting the pixels if their contrast is greater than 95% gives a consistent set of values, but the area is not fixed. In a future study, this could be adjusted by including the area of the defect in the DCI computation.

The application of the DCI to the simulated images is coherent with the corresponding expert interpretation. The more the DCI increases, the more chance to be seen by the

operator. The cases in limit of visibility correspond to a low contrast (DCI1 is small, near or less than 0.2), together with either :

- a low gradient (thus high DCI2 = 0.8) but in the presence of a change of gradient (thus a low DCI3 ~0.2),
- a high gradient (thus a low DCI2 = 0.4) but a constant one (thus a high DCI3, near 0.9),

The three DCI can be combined using the Dempster orthogonal sum, in order to get one unique parameter : the result is shown on figure 14 (left). The drawback is that an object having three average values of each DCI get the same fused DCI than another having two high and one low DCI. This tends to mask the differences between a defect in limit of visibility with respect to a visible defect.

Considering the cases in limit of visibility (each having a low DCI1, and either a low DCI2 or DCI3), we decided to combine DCI 1 with respectively DCI2 and DCI3. The result is shown in figure 14 (right). The result is better in the sense that we can now distinguish the visible defects from the limit ones (at least one of the fused DCI is lower than 0.5). Of course, more cases are needed to draw a conclusion about the reliability of such a rule.

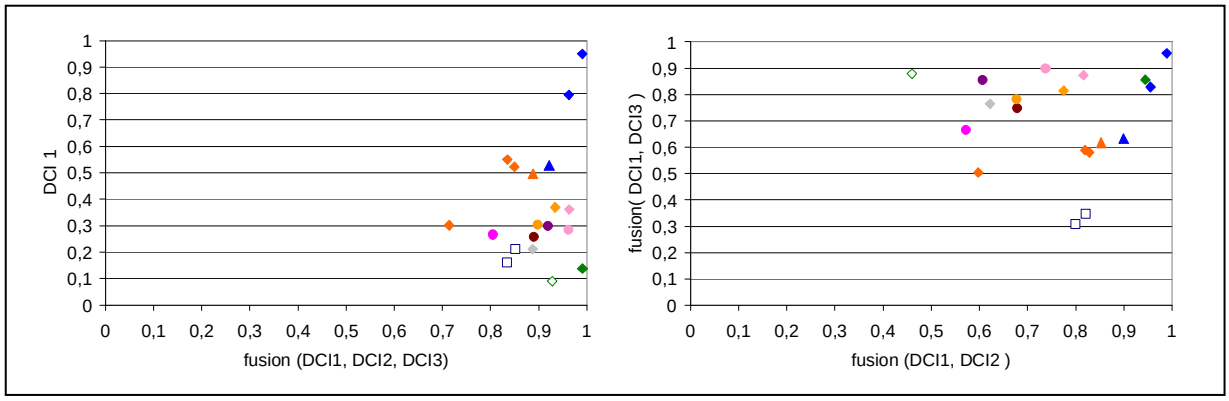

Figure 14 Representation of the DCI with previous data of Figure 12 (left) : DCI1 versus the Fused DCI, (right) fused (DCI1, DCI3) versus (DCI1, DCI2).

Three descriptors have been defined. They are representative of the defect contrast, the background gradient (or slope) and gradient variations of the image (the image of a weld could be an example of representation of the change of gradient). We developed three DCI based on these characteristics. Even if the selected parameters are quite natural (contrast, background gradient), it was not obvious to elaborate representative descriptors of these characteristics (variance, etc…).

These first results are promising, and the following works are under progress :

- consolidation of these results by adding more defect configurations. This work consists in the following points :

  - consolidation of the expert diagnosis by adding more evaluations in order to include the reliability of each individual inspector or laboratory (subjectivity in the detection mechanism),

  - increasing the number of configurations (variability of the case studies).

- taking into account the film class, which means to integrate and take into account the noise in the simulated radiograph. These developments are linked to the evolutions of Moderato that are already underway (see the presentation of A. Schumm in the simulation session).

6. Conclusion

In this paper we have presented the main parameters involved in the defect appearance on a radiograph. We tried to develop a generic method to compute a Defect Confidence Index from one or several parameters, in order to quantify the defect presence.

It was found that for radioscopic images, the contrast to noise ratio is the essential parameter to define a DCI, while some specific features can help to sort false alarms (mainly for those appearing systematically due to normal structures of the samples). The example of a casting sample was chosen to illustrate the approach.
For radiographic films, the optical density contrast between the defect and its neighbourhood is usually a good feature, but for some cases, the optical density variance is the only parameter to explain or justify why a defect was said visible or in limit of visibility by the controller. The combination of several parameters is thus necessary in this case. The validation of the relationship between the proposed DCI and the defect visibility by experts is now under progress.

References

1. J. Serra, Image analysis and mathematical morphology, London : Academic Press, 1982, 610 p.
2. G. Lecomte, V. Kaftandjian, E. Cendre, Combination of information from several X-ray images for improving defect detection performances – Application to castings inspection. Proc. 9th ECNDT, Berlin, 25-29 sept. 2006, pp. 16.
3. G. Lecomte, Analyse d'images radioscopiques et fusion d'informations multimodales pour l'amélioration du contrôle de pièces de fonderie., Thèse INSA de Lyon, Décembre 2005, pp. 150.
4. A. Dempster, Upper and lower probabilities induced by multivalued mapping, Annals of mathematical statistics, 38, p.325-339, 1967.
5. Shafer G., A mathematical theory of evidence, Princeton university Press, Princeton, 297 p., 1976.

6. A. Schumm, O. Bremnes, B. Chassignole. Numerical Simulation of Radiographic Inspections : fast and realistic even for thick components. Proc. 16th WCNDT, Montreal, Canada, 2004.