Probabilistic crack detection
in digitized radiographs of weld inspection

O. Alekseychuk, U. Zscherpel
(BAM Berlin)

1 Introduction

This work addresses the task of crack-like defect detection in digitized radiographs from a probabilistic and informational point of view. The scope of this work is limited to the following:

- The task consists in the development of an algorithm for the automated detection of indications of crack-like objects in radiographic images of welding seams acquired during in-service inspection of steel pipes.
- Only one projection is used at a time.
- Only longitudinal cracks are considered as objects of interest. Defects of other types are ignored.
- Real cracks and crack-like objects (like, e.g. narrow undercuts) are not distinguished.
- No physical models of object fracture or crack propagation is used.

In the considered images the objects of interest are usually of low contrast and sharpness and located on a non-homogeneous and noisy background (Fig.1). This is due to the physical properties of the radiographic method. The extreme low SNR (about 0dB) of real crack indications makes it difficult or complete impossible to use conventional detection algorithms.

The detection task is further complicated by the inability of human operators to express their experience about crack features in a formal way. So instead of guessing possible and useful features, we try to solve the task of detection using only two most obvious features of the crack-like objects: rapid changes of intensity (contrast) in the direction orthogonal to the indication direction and elongated shape of the discontinuity.

2 Estimation of local indication properties

A drawback of many conventional detection algorithms is decision making about object presence or absence using only the local image area. Such detection decisions are very influenced by the noise which leads to high rates of object missing or false alarms. Therefore in the proposed approach the local
Figure 1: A typical radiographic image of welding seam (top) and its high pass filtered version (bottom). The true crack indication is located in boxes 32-39 below the welding seam.
detection is not performed. Instead, the probability of presence of crack like indication (as an analog measure) is calculated.

Why probability? Because neither simple indication contrast nor estimation of SNR are adequate. This is good seen if we consider two indication profiles: one is an ideal without noise, an other with presence of noise (Fig. 2).

In the first case the estimated indication contrast equals to the real indication contrast $h$. The situation is different in case of noise presence: the estimated contrast is $z$, but the interesting question is: *a defect of which intensity (severity) $h$ has caused estimation $z$?* So we want to develop some measure which relates the estimated contrast $z$ and the real one $h$. This measure

- has to be sensitive to $z$ if the noise estimation is low (because even small changes of measured signal intensity have to be considered as meaningful) and on the opposite: if the noise is high, then the same changes of estimation can not be considered as real signal changes, but probably only noise fluctuations;

- has not to grow too high if the underlying image exhibits high signal to noise ratio. On practice this means that if, for example, one measurement shows SNR=10 and an other SNR=100, then we can already assert the presence of the signal from the first measurement and the second measurement does not make us 10 times more sure of that. This demonstrates why the raw SNR value is not adequate.

The probability of presence of the sought object with intensity $h$ given the estimation $z$ ($P(h|z)$) exhibits the required properties and has a clear meaning (it is theoretically based). Unfortunately $P(h|z)$ can not be calculated directly. But the Bayesian composite probability formula can be used:

$$P(z)P(h|z) = P(h)P(z|h) \quad \Rightarrow \quad P(h|z) = \frac{P(h)P(z|h)}{P(z)}, \quad (1)$$

where $P(z)$ is the *a priori* probability that the estimated indication intensity takes value $z$, $P(h)$ is the *a priori* probability of occurrence of a crack indication with the intensity $h$ and $P(z|h)$ is the *a posteriori* probability that the estimated indication intensity takes the value $z$ with the condition of a true indication presence with intensity $h$. The value of $P(z|h)$ can be found from
the image data, but the value of $P(h)$ cannot be calculated from the same image area as $P(z|h)$ and must be supplied externally.

2.1 Calculation of $P(z|h)$

Before $P(z|h)$ can be calculated the value of $z$ (the estimation of indication contrast) has to be found. $z$ is defined as the difference between grey values estimations of background $g_{bkg}$ and indication itself $g_{ind}$ (Fig.3):

$$z(x) = g_{bkg}^*(x) - g_{ind}^*(x).$$

For the grey values estimation a piecewise linear background model is assumed. According to this model the background changes in the local area can be described by a linear function plus added random noise $n_0$:

$$g_{bkg}(x) = a_{bkg} + b_{bkg}x + n_0.$$ 

The parameters $a_{bkg}$ and $b_{bkg}$ of this linear model can be found using the MSD (minimal square deviation) method. Then the estimation of background grey value is:

$$g_{bkg}^*(x) = a_{bkg} + b_{bkg}x.$$ 

The same reasoning is valid for the estimation of indication grey value $g_{ind}^*$. The window size $w_{ind}$, which is used for parameter estimation of the linear model, defines the spatial resolution of the local operator and must be supplied externally.

The estimation $z$, calculated in this way, is influenced by noise. The standard deviation of $z$ is a function of the number of pixels which are used for estimation of the indication and the background grey values:

$$\sigma_z = \sigma_n \sqrt{\frac{w_{bkg} + w_{ind}}{w_{bkg}w_{ind}}}.$$
where $\sigma_n$ is the standard deviation of noise estimated on a per-pixel basis. For the purpose of reducing of $\sigma_z$ the size $w_{bkg}$ of the windows for background estimation can be chosen to be several times bigger than $w_{ind}$.

As the estimation $z$ is influenced by the noise in many image pixels, the assumption about a normal distribution of $z$ with center in $h$ can be made. So the expression for $P(z|h)$ can be written:

$$P(z|h) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{(h-z)^2}{2\sigma_z^2}}.$$

### 2.2 Calculation of $P(h|z)$

$P(h|z)$ can be calculated using Eq.1. Before it can be done a assumption about the a priori probability $P(h)$ has to be made. In general case $P(h)$ can be a continuous distribution $\rho(h)$ and is unknown. Because of this, the common practice is to assume $P(h) = const$ (i.e. $P(h)$ does not depend of $h$). Then the most probable $h$ can be calculated:

$$h^*(z) = \arg\max_h P(h|z) = \arg\max_h P(z|h).$$

Obviously the assumption of $P(h) = const$ is very arbitrary and is just forced by difficulties in determination of distribution of this a priori probability. Instead of this “very arbitrary” assumption, an other, less arbitrary, assumption about statistical properties of indication intensity is made. The binary model “no objects OR object of intensity $h$” is used. The object intensity $h$ and its a priory probability $P(h)$ are supplied as the input parameters.

### 3 Detection as optimization task

Even if we succeed in the estimation of the a posteriori probability of indication presence $P(h|z)$, it is estimated in the local area and therefore very influenced by the noise. To suppress this influence the elongated shape of the sought objects can be utilized: some integral characteristic which is calculated across the full indication length has to be developed. The purpose of this characteristic is to characterize the probability of presence of such an indication, it is called estimation function $f$.

The crucial point here is, that the task to be solved is not the precise estimation of the probability of indication presence, but the task of the detection of the indication. This means, that the coordinates of the image points, over which the estimation function has to be calculated, are unknown and have to be found. The estimation function takes its maximum if it is calculated exactly on the indication points. Thus the task of indication detection can be reformulated as a much more better explored task of finding the maximum of the function:

$$object = \arg\max f(p_1, p_2, \ldots, p_i, \ldots, p_n),$$

where $p_i$ are pixel coordinates and $n$ is the indication length.
It is necessary to put some restrictions on the arguments of the estimation function in order to keep the task of the function optimization adequate to the task of detection of crack indications. This is, for example, to use the preferred direction of crack indication or to require the connectivity of the involved points (the points have to form a continuous curve), etc. So some spatial 2D model of the indication has to be developed.

We have studied different estimation functions and found it useful to classify the approaches to the design of the estimation function into the following two main categories.

3.1 Complex estimation functions

Because the estimation function must allow to distinguish between good or not good hypothesis, one intuitively want to include as much as possible object features in the estimation function formula. These can include such features as the variation of the object intensity along its length, indication average and local curvature, etc.

Despite of the restrictions on the indication shape, the computational complexity of search for the maximum of such an estimation function becomes very high even for just good experimental evaluation, not to speak about practical applicability. This dictates the usage of strong heuristics to reduce computational complexity to the acceptable levels [1, 2].

The heuristics are intuitively based presumptions about the behavior of the estimation function. Application of heuristics allows not only to reduce the search complexity, but the optimality of the final solution is usually sacrificed too. It is important, that the result of optimization is unpredictable and nobody can say how far is the solution, found in this way, from the really optimal one. This is the reason why this approach is not used in this work.

3.2 Separable estimation functions

If the evaluation function \( f = f(p_1, p_2, ..., p_i, ..., p_n) \) can be represented as a sum of terms, each of which depends on only a few variables then the so-called multistage optimization process (serial dynamic programming) can be applied [3, 4].

In general, this method can be described as a procedure in which the optimization is carried out separately for each variable \( p_i \). Memory and time requirements for optimization of each term is proportional to \( P^n \) (\( P \) - number of different states which variable \( p_i \) can take, \( n \) - number of independent variables in the term). This way the dynamic programming formulation helps to define analytically the requirements for the estimation function to be easily optimizable.

This leads to an other approach: if the estimation function can be represented as a sum of simple terms, then its maximum can be found strictly (without any heuristics) and with an acceptable computational complexity.
The drawback is that holding the requirement to be representable as a sum of simple terms, makes it difficult to incorporate certain features in the estimation function. Nevertheless, we try to use this approach to the construction of the estimation function and see what is possible to reach this way. Our motivation is: if some feature cannot be included in the estimation function, then we know, at least, that it is not used. But if some feature is included, then we know that it will be fully utilized (since the optimization procedure is strict).

3.3 The path estimation function used in this work

We calculate the probability of the hypothesis $H_1$ that an indication with predefined intensity $h$ is present across the connected set of pixels $p_1...p_n$: $P(H_1(p_1,..,p_n))$. This probability can be expressed via conditional probabilities in each point of the path:

$$P(H_1(p_1,..,p_n)) = P(H_1(p_1,..,p_{n-1}))P(H_1(p_n)|H_1(p_1,..,p_{n-1})) =$$

$$... = P(H_1(p_1)|H_1 \text{ prior}) \prod_{i=2}^{n} P(H_1(p_i)|H_1(p_1,..,p_{i-1})), \quad (2)$$

where $H_1 \text{ prior}$ is the a priori hypothesis for the indication presence with intensity $h$. According to the used Bayesian paradigm the a posteriori conditional probability in each point depends on the a priori probability (Eq.1) and the a priori probability can be derived from the a posteriori probability in the previous path point:

$$P_1 \text{ prior}(p_i) = P_{\text{trans}}P_1(p_{i-1}).$$

The logarithm from Eq.2 converts the product on the right side to a sum. It is a sum of the terms, each of which depends only on a few variables. So the maximum of this function can be found without application of any heuristics.

But Eq.2 has a drawback: the a posteriori probability of object presence $P(H_1(p_1,..,p_n))$ decreases monotonically with the length of the indication. It is clear why: the conditional probabilities in the product are always less than unity.

To avoid this the a posteriori $P(H_1(p_i))$ can be considered not alone, but compared with the a priori probability $P(H_1 \text{ prior})$. This can be elegantly expressed in terms of entropy and a posteriori information gain [5]:

$$dI_1(p_i) = I_1(p_i) - I_1 \text{ prior} = (-\log P_1 \text{ prior}) - (-\log P_1(p_i)) = \log \frac{P_1(p_i)}{P_1 \text{ prior}}.$$

The cumulative information gain along the path serves as estimation of path “goodness”:

$$f(p_1, p_2, .., p_n) = \sum_{i=1}^{n} dI_1(p_i) = \sum_{i=1}^{n} (\log P_1(p_i) - \log P_1 \text{ prior}). \quad (3)$$
Returning to the notations of Section 2:

\[ P_i(p_i) = \frac{P_i(h)P(z(p_i)|h)}{P(z(p_i))}, \]

\[ P(z(p_i)) = P_i(h)P(z(p_i)|h) + (1 - P_i(h))P(z(p_i)|0), \]

\[ P_i(h) = P_1 prior(p_i) = P_{trans} P_1(p_{i-1}), \]

\[ P_1 prior = P(h), \]

where \( P_i(h) \) is the a priori probability of hypothesis of presence for the point \( p_i \) whereas \( P(h) \) is the a priori probability of the same hypothesis but for the stand-alone point (not influenced by the estimations of indication presence in the neighbour points). Then Eq.3 converts to:

\[ f(p_1, p_2, ..., p_n) = \sum_{i=1}^{n} [\log P(h|z(p_i)) - \log P(h)] = \]

\[ \sum_{i=1}^{n} [\log [(P_i(h)P(z_i|h) + (1 - P_i(h))P(z_i|0))] - \log P(z_i|h)]. \]  \( (4) \)

The estimation function built in this way exhibits the required properties:

- It is theoretically based and has a physical meaning.
- It can increase or decrease depending on the relation between intensity of the sought object \( h \) and the estimated contrast \( z \).
- The speed of these changes depends not only on the difference between \( h \) and \( z \), but on the estimated noise level too.
- It can be represented as a sum of simple terms, so it can be easily optimized.

The values \( h, P(h), P_{trans} \) must be supplied externally and act as control parameters for the algorithm.

4 Experimental evaluation

The described approach is realized by a C program as a computational engine with a simple API. The following input parameters must be provided externally: \( L_{sign} \) (significant indication length, longitudinal resolution), \( w_{ind\ min} \) and \( w_{ind\ max} \) (minimal and maximal indication width, transversal resolution), \( h/\sigma \) (minimal SNR of objects to be detected), \( P(h) \) (a priori probability of presence of an object with minimal SNR) and \( P_{trans} \) (probability of object prolongation). For the \( L_{sign}, w_{ind\ min}, w_{ind\ max}, P(h) \) and \( P_{trans} \) some reasonable defaults can be found which work for all images in the given class. The algorithm sensitivity in each separate case can be controlled by the adjusting of the minimal SNR \( (h/\sigma) \) of the objects to be detected.

The performance of the detection algorithm is tested on synthetic images as far as on the digitized real radiographs from the in-service inspection of welds of austenitic steel pipes. Some examples are given on Fig.4 and Fig.5.
Figure 4: a) An example of a simulated image (crack indication on a nonuniform background (left), SNR = 0 dB); b) Source image (a) after subtraction of background and histogram stretching (for illustration purposes only); c) Output of the developed algorithm (using source image (a))

As far as the positions of real defects are known (in case of real images destructive testing of samples was made) the correct detection and false alarm rates for the developed algorithm can be found. The results can be represented by the Receiver Operation Characteristic (ROC plot). A comparison of the performance of the developed algorithm and a group of trained inspectors (which have analyzed the same image set) is given on Fig. 5.

5 Summary

The presented approach addresses the task of detection of elongated low intensity objects from the probabilistic point of view and demonstrates a detection quality comparable to the trained human inspectors (Fig. 5). The algorithm key features: Bayesian estimation is applied; an integral full-length characteristic is used for estimation of probability of presence of crack-like indications; the characteristic is theoretically based, exhibits desirable behavior on changing signals and has a low computational complexity of optimization. For the most of the algorithm’s control parameters some reasonable defaults exist for each class of application tasks. The parameters can be always adapted for a new application. The sensitivity of the algorithm can be adjusted by changing of a single parameter. The computing time on one 270 MHz MIPS CPU is from 10 min till 2 h per 20 MPix image. Other applications of the algorithm are not excluded, i.e.: roads and rivers detection, angiography, etc.

References

Figure 5: An example of the real radiograph of welding seam (a) and the detection results with the following parameters: b) $L_{\text{sign}} = 128$, $w_{\text{ind min}} = 2$, $w_{\text{ind max}} = 6$, $P(h) = 0.25$, $P_{\text{trans}} = 0.80$, $h = 3.0\sigma$; c) $h = 2.5\sigma$; d) $h = 2.0\sigma$. 
Figure 6: Receiver Operation Characteristic (ROC) of the developed crack detection algorithm with the comparison to the performance of human inspectors. MPA data set of 99 images. “+” - human results (different inspectors), “◊” - developed algorithm (different sensitivity).


