3



# INTENSIFICATION OF RESEARCH IN PRIORITY AREAS (IRPA)

# GRANT (PROJET NO. 03-02-05-2190 EA005)

## COMPREHENSIVE REPORT



**Project title:** Automatic detection of multiple defects in weld radiography

**Project leader:** Dr. Mani Maran a/l Ratnam
School of Mechanical Engineering
Universiti Sains Malaysia,
Engineering Campus,
14300 Nibong Tebal,
Penang.

**Co-researchers:** Assoc. Prof. Dr. Lim Chee Peng
School of Electrical & Electronic Engineering,
Universiti Sains Malaysia,
Engineering Campus,
14300 Nibong Tebal,
Penang.

Dr. Zahurin Samad
School of Mechanical Engineering
Universiti Sains Malaysia,
Engineering Campus,
14300 Nibong Tebal,
Penang.

**Duration of project:** 1 September 2002 – 31 August 2005

# Automatic defect classification in weld radiography from simulated data using MLP network

T.Y.Lim[a], M.M.Ratnam[a,*], M.A.Khalid[b]

[a] School of Mechanical Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia.

[b] Malaysian Institute for Nuclear Technology Research, Bangi, 43000 Kajang, Malaysia.

* corresponding author (e-mail:mmaran@eng.usm.my)

## Abstract

An effective weld defect classification algorithm has been developed by using a large image database of simulated defects. Twenty-five shape descriptors used for the classification were studied and an optimal set of nine descriptors with highest discriminative capability was selected using a statistical approach. A multi-layer perceptron (MLP) network was trained using shape parameters extracted from the simulated images of weld defects. By testing on 60 unknown simulated defects, the optimized set of nine shape descriptors gave the highest classification accuracy of 100%. Defect classification on 49 real defects from digitized radiographs produced maximum overall classification accuracy of 97.96%.

*Keywords*: Weld radiography, automatic defect classification.

## 1. Introduction

Radiography technique (RT) for non-destructive testing (NDT) of welds has evolved rapidly in the past decades and has become an established technology in the field of weld inspection. RT is widely used as an inspection tool for detecting flaws inside welded structures, such as pressure vessels, structural members and power plants [1]. Most radiographic exposures and film interpretations are still carried out manually [2]. However, human interpretation of weld defects is tedious, subjective and dependent upon the experience and knowledge of the inspector.

Much effort has been made recently to converge RT into an automated process of radiographic interpretation. With the advancement of digital image processing and computer architecture, the automated interpretation of weld radiographs is made possible with a system consisting of film digitization, pre-processing, defect detection and defect classification. The automatic interpretation of radiographs using digital image processing reduces human involvement, thus making the inspection more reliable and faster.

Basically, automatic defect interpretation consists of 5 stages: image digitization, pre-processing, weld extraction, defect segmentation and defect classification. A digitized radiographic image, however, is often corrupted by non-uniform illumination, noise and low-contrast [3]. Due to the degraded quality of a radiographic image, image pre-processing is normally carried out as an initial stage of defect detection. This may include noise elimination, contrast enhancement and shading correction. Poor quality radiographic images have led to the development of various automatic defect detection algorithms that focus on extracting defects using various image segmentation methods [4-11]. In addition, the fast growth of artificial intelligence methods such as artificial neural network (ANN) and fuzzy techniques has become an important component to assist and enhance the task of defect detection in poor quality images.

Weld extraction is usually the first step in the development of automated inspection of weld radiographs. Several techniques of weld extraction are available in the literature, such as weld extraction methodology based on the observation that the intensity plot of a weld profile looks more like Gaussian than the other objects in the image [4], extraction of features from line image of welds for use in training a multi-layer perceptron (MLP) neural network [5] and extraction of multiple curved welds from one radiographic image using fuzzy *K-NN* and *C*-means methods [6].

The problems related to poor quality images have been addressed by several authors. For instance, Kehoe et al. [7] introduced the sigma-norm contrast enhancement and mean gradient edge detector to improve the quality of the image. Bonser and Lawson [8] developed filtering and 'window' based variance operator for segmentation of suspected defect areas inside the weld region of partially competed welds. Murakami [9] proposed a local arithmetic operation to a limited region and was followed by thresholding methods. Various image processing techniques were used by other authors to process and segment the weld images [10-11].

Research in weld defect detection has also been carried out actively in the past. For instance, Lashkia [3] proposed a detection algorithm based on fuzzy reasoning to detect low-contrast defects using local image characteristics, such as spatial contrast, spatial variance and distance between two contrast regions. Liao and Li [12] developed welding flaws detection based on the fitted line profiles of a weld image and successfully detected 93.33% of various defects from linear welds. Jacobsen et al. [14] extracted parameters from the intensity profile for the network with five techniques, namely morphological filter, derivatives of Gaussian filter, Gaussian Weighted Image Moment Vector-Operator (GWIMV) filter, Fast Fourier Transform (FFT) filter and Wavelet transform.

The past research in automatic weld radiography focused mainly on enhancing poor quality weld images and detecting defects from the radiographs, whilst the development of automatic defect classification systems is limited. Some of the research on automatic weld defect classification are reviewed briefly.

Jacobsen et al. [14] use several features to differentiate between crack, undercut and absence of defect. Perner et al. [15] introduced a framework for distinguishing classification methods, namely neural nets and decision tree. Their system only classified crack, undercut and absence of defects. Feher [16] presented a simple classification of 3 types of defects, namely porosity, undercut and incomplete penetration by simple rules. However, the system was unable to detect small defects that did not differ much from the background. Aoki et al. [17] used 10 parameters to characterize the defects and classified them into five possible classes using a MLP neural network. Their investigation on 27 defects showed that 25 defects were classified correctly with a success rate of 92.6%. In their work, 35 training samples and 27 test samples were used. These are, however, too small for practical applications. Wang and Liao [18] used a set of parameters to classify six possible defects and obtained the highest accuracy of 92%. In their work 108 data sets were used for training while 12 were used for testing. However, the 12 test samples used for classifying 6 types of defect are considered small and the success rate for individual defect was not reported.

The performance of a weld defect classification system often depends on the set of radiographic images used. Different authors use different sets of characterization parameters and different sets of radiographic images. Besides that, final appearance of defects extracted very much depends on the image segmentation method used. Thus, comparison of the effectiveness of various classification techniques is not possible unless it is based on a standard set of 'ideal' weld defects. As stated by Wang and Liao [18], there is a need to establish a benchmark image set in order to perform the comparison between the performances of various classification techniques.

In this work, an efficient classification system using a large number of simulated images of weld defects that are considered as 'ideal' defects is developed. A set of shape descriptors was defined and the number of descriptors required in the classification was optimized using a statistical approach. The classification is divided into two parts. Firstly, a multi-layer perceptron (MLP) neural network was trained using features extracted from the simulated defects and the network was used to classify a set of simulated defects. Secondly, classification was carried out using features extracted from real defects while the training was done using the same simulated data set. The classification accuracy of individual defect type was also evaluated.

## 2. The simulation process

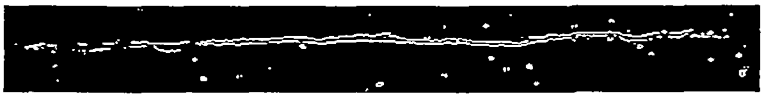
There are two main advantages of using a large database of simulated images to develop the classification task. Firstly, simulated images are created manually by imitating the appearance of real radiographs and, therefore, a large number of images could be created to simulate the variation of shapes in real defects. Secondly, different real radiographs have different intensity variations. Thus, for a given set of images a particular shape feature may give certain accuracy in defect classification while the same feature may give different accuracy in a different set of images due to the intensity variation. Simulated images overcome this shortcoming by providing a basis to test and select the appropriate features for the classification.

Before generating the simulated images, real radiographs were digitized using an *X*-ray film scanner (*Cobrascan CX-612-T*) in 12-bit resolution and saved in *TIFF* image format without compression. The original radiographs were a collection of reference radiographs previously evaluated by qualified inspectors. These radiographs were used as the basis for creating simulated weld defects. The simulation process is divided into four stages: (i) Generation of simulated defect images, (ii) shape definition, (iii) shape parameter optimization and (iv) classification evaluation.

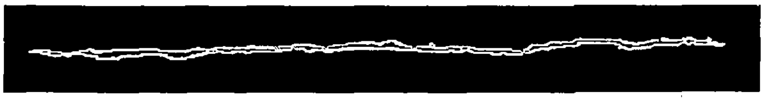
*2.1 Generation of simulated defect images*

Six types of defects were considered in creating the simulated defect images. These are longitudinal crack, incomplete penetration, porosity, cavity, slag inclusion and transverse crack. To create defects having high similarity to the real defects, the real image for each defect type was used and was subjected to edge detection process using the Sobel edge detector. Sobel edge detector uses two convolution kernels, one to detect changes in vertical contrast and another to detect horizontal contrast. Due to noise in the real image, the resulting edges are irregular and discontinuous. These edges were touched up manually by linking the broken edges and eliminating pixels due to noise. A total of 300 simulated defects, 50 images for each defect type, were generated for training the neural network. Additional 10 images for each defect type were generated for testing the accuracy of classification. Figures 1(a)-(d) show an example of a real image of crack, the edge of the defect, the edge after touching up and the simulated defect image. Figures 2(a)-(f) show samples of simulated defects generated for various types of defects.

(a) Real image of crack



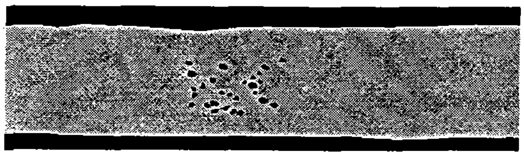
(b) Image after edge detection with Sobel operator



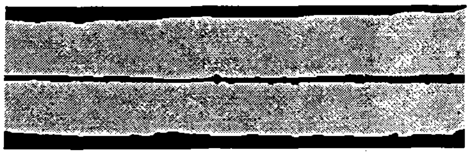
(c) Image after noise elimination and edge linking

(d) The simulated defect image

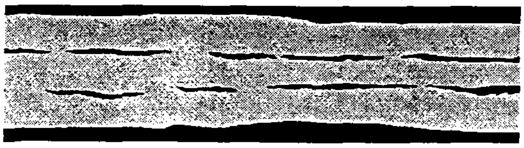
Figure 1. Generation of simulated defects.



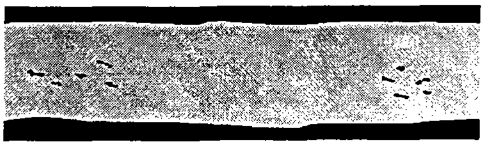
(a) Porosity



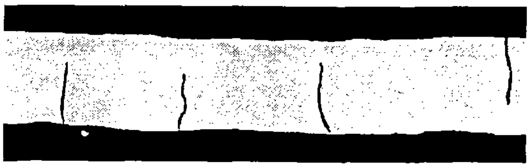
(b) Incomplete penetration



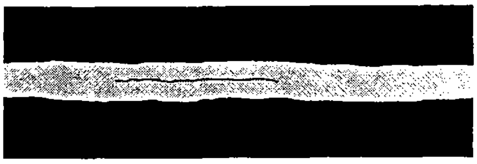
(c) Slag inclusion



(d) Cavity



(e) Transverse crack



(f) Crack

Figure 2. Simulated images for various defect types

*2.2 Shape definition*

Shape recognition and classification are fundamental problems in many applications of machine vision [19], especially in vision-based inspection. In weld defect detection, a radiograph interpreter will inspect the defect shape, geometry and its orientation to recognize the defect. The shape or outline of a defect is particularly important because a certain defect type may have a different shape compared to other defects. Incomplete penetration, for instance, is defined as dark continuous or intermittent line in the middle of the weld. In contrast, porosity is defined as dark shadows of rounded contour.

The main purpose of shape descriptors is to measure geometric attributes of the shapes that can be used to classify and recognize an object. Shapes can be represented using various types of descriptors and techniques. Different types of methods that characterize a shape can be viewed from different context, such as description based on boundaries and region, local and global shape characters, statistical or syntactic object description, object reconstruction ability or incomplete shape recognition ability [20]. A summary of common shape description techniques is given in Ref. [21]. A thorough discussion of shape analysis and recognition can be found in Refs. [20-22].

In this study, 25 common shape descriptors were defined in order to evaluate their capability in discriminating different types of weld defect shapes. Most of the descriptors use area, perimeter, diameter, width and length in describing shapes that are simple attributes of shape boundary. On the other hand, feature like curvature investigates the overall localized properties of shape boundary [23]. Other features such as standard deviation, skewness and kurtosis of radius (see Ref. [19] for details) tend to characterize the variability of radius data. A list of the shape descriptors used in this study is given in Table 1.

Table 1. Shape descriptors

| Shape descriptors | | |
|---|---|---|
| 1) Form | 7) Elongation | 18) Diameter ratio |
| 2) Area to perimeter ratio | 8) Convexity | 19) Curl |
| 3) Roundness | 9) Solidity | 20) Orientation |
| 4) Aspect ratio1 (Maximum diameter/ Minimum diameter) | 10) Compactness | 21) Standard deviation of radius [20] |
| 5) Aspect ratio2 (Minimum diameter/ Mean diameter) | 11) Extent | 22) Skewness of radius [20] |
| 6) Aspect ratio3 (Maximum diameter/ mean diameter) | 12) Heywound diameter | 23) Kurtosis of radius [21] |
| | 13) Major to minor axis length | 24) Horizontal length to area ratio [17] |
| | 14) Miu [18] | 25) Vertical length to area ratio [17] |
| | 15) Sigma [18] | |
| | 16) Circularity [18] | |
| | 17) Curvature [24] | |

In order to choose a combination of optimal shape descriptors that provide the highest capability to classify the weld defects, a multiple comparison procedure was carried out and is discussed in the following section.

*2.3 Shape parameter optimization*

A multiple comparison procedure was employed to compare the mean differences among each type of defect for every shape descriptor. The statistical test was conducted to determine if there are significant differences between each of the defect group for a particular shape descriptor. A typical procedure to perform this test is as follows:

a) Select the statistical hypothesis:

   $H_0 : \mu_i = \mu_j$ for all classes *i* and *j* where $i \neq j$. This hypothesis indicates that all pairs of groups have equal means.

   $H_1 : \mu_i \neq \mu_j$ if at least one pair of means is different.

b) Select the significance level α.

c) Determine the critical value from *F*-test. This is the one tailed 'Analysis of Variance' (ANOVA) *F*-test to find out whether data from several groups have a common mean. Rejection of $H_0$ leads to the multiple comparison test to determine specific differences among those groups.

d) Determine the critical value from the multiple comparison procedure.

e) Evaluate the results and draw conclusion.

The purpose of multiple comparison procedure is to determine where mean differences lie among a set of means in a one-way ANOVA because the rejection of $H_0$ in $F$-test does not indicate which of the means are different [24]. Multiple comparison procedures compare two means to determine if they are statistically different. With every type of shape descriptor, 6 groups of defects were tested to determine if they are significantly different from others. The discriminative criteria of each descriptor is based on the concept that the smaller the number of joining groups in the test (which means most of the means are significantly different), the better is the descriptor in classifying different types of defects.

Figure 3 shows an illustration of the multiple comparison procedure where $\bar{y}_i$ is the mean of the individual defect type. The lines underneath indicate the mean values of defects that are insignificant among others, and thus combine the defects into several groups. Each group indicates that is are no significant difference in mean values among the group members. From Figure 3, four groups of defects have different means with an overlapping result for $\bar{y}_5$. In the current work, two overlapping result are considered as one joining group. Therefore, the shape descriptor (mean) in Figure 3 is considered to have the ability to identify three main groups of defects.

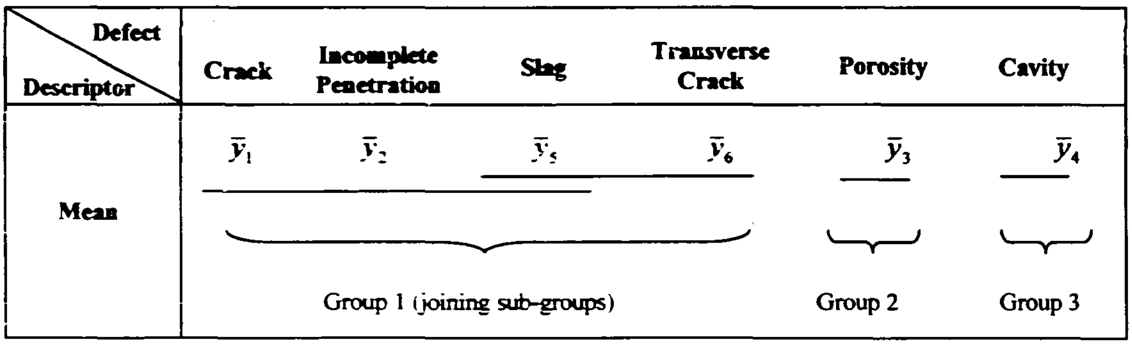



Figure 3. Example of multiple comparison result.

By using this criterion, a comparison among all 25 shape descriptors was carried out and the optimal groups were identified. There are several multiple comparison procedures that can be adopted depending on how error rates are controlled. In this study five types of comparison procedures were used, namely Tukey HSD, Tukey LSD, Scheffe, Bonferonni and Dun-Sidak [24-25] and the result of each test was compared among each other. The numerical values of data were normalized linearly before performing the mean comparison. The assumptions made in order to satisfy the multiple comparison procedure (as well as the $F$-test) are as follows: (a) the samples are randomly and independently selected, (b) the populations are normally distributed, and (c) the populations all have the same variance.

The defects from the simulated images were extracted in order to obtain the numerical data for every shape descriptor. The data were then used to perform the multiple comparison tests. The various comparison procedures were found to produce the same results. The result for one of the comparison procedure is shown in Table 2. Among the 25 shape descriptors, the highest discrimination capability produced by the tests was 5 groups using 3 shape descriptors, followed by 4 groups using 7 shape descriptors. Figures 4 and 5 show examples of the distribution of the shape descriptors *Solidity* and *Circularity* that have 5 and 2 joining groups respectively. From these figures, it can be concluded that *Solidity* has better ability to classify defects compared to *Circularity*.

Table 2. Results of multiple comparison procedure

| **Number of groups** | **Shape descriptors** |
|---|---|
| 5 | Solidity, Extent, Standard deviation of radius distribution |
| 4 | Compactness, Heywound diameter, Vertical length to area ratio, Major to Minor Axis Length, Curvature, Orientation |
| 3 | Form, Area to perimeter ratio, Roundness, Aspect ratio 2, Aspect ratio 3, Elongation, Horizontal Length to area ratio, Eccentricity, |
| 2 | Aspect ratio 1, Circularity, Curl, Kurtosis of radius distribution |
| 1 | Convexity, Miu, Sigmas, Diameter ratio, Skewness of radius distribution |

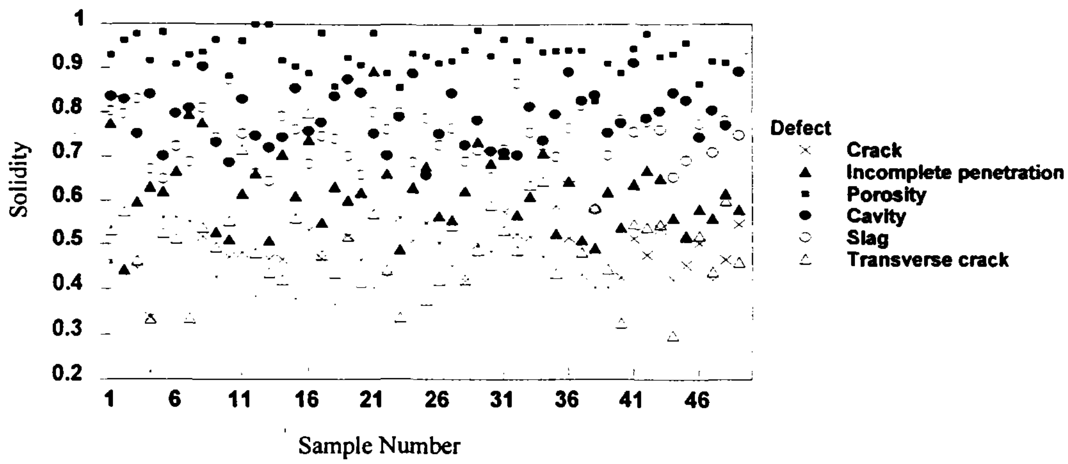



Figure 4. Distribution of *Solidity*

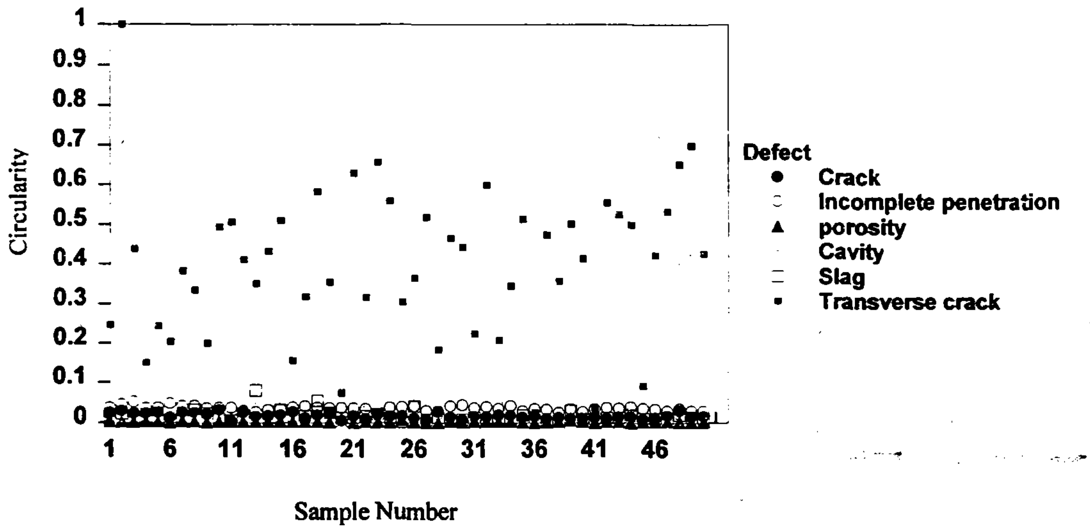



Figure 5. Distribution of *Circularity*

*2.4 Classification of defects using MLP network*

A *multi-layer perceptron* (MLP) model using *back-propagation* (BP) algorithm was used for learning and classifying the defects. MLP networks are gaining popularity in classification task due to their flexibility, robustness and high computational rates. The MLP model is shown in Figure 6.

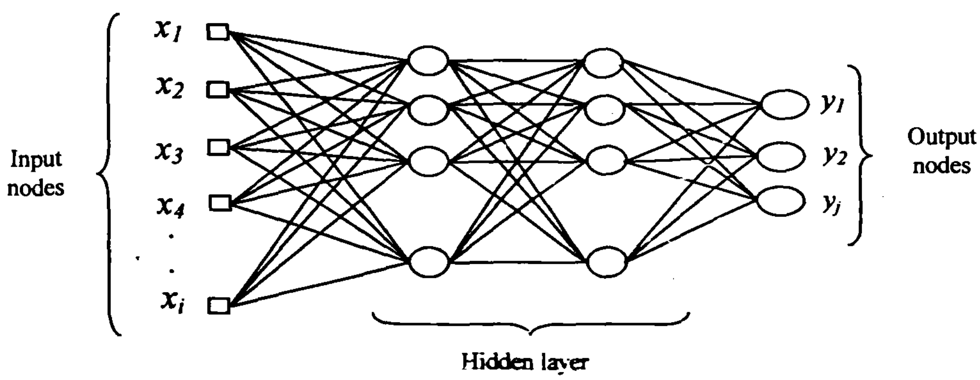



Figure 6. A schematic of a MLP model

The general model of MLP consists of a number of nodes arranged in multiple layers with connections between the nodes in the adjacent layers by weights. The model consists of an input layer that accepts the input variables used in the classification, hidden layers and an output layer. A summation of each neuron $j$ in the hidden layer by its input nodes $x_i$ after multiplying the connection weights $w_{ij}$ gives the output $y_j$ as a function of the sum, that is,

$$y_i = f\left(\sum w_{ji} x_i\right) \tag{1}$$

where $f$ is the sigmoidal or hyperbolic tangent transfer function. Using the BP training algorithm, the weights are minimized based on the squared differences between the actual and desired output values in the output neurons given by,

$$E = \frac{1}{2}\sum_j \left(d_j - y_j\right)^2 \tag{2}$$

where $y_i$ is the actual output of the neuron and $d_j$ is the desired output of neuron $j$. During classification, input data is fed into the network and the classification is performed by assigning a class number to a pixel or segment using the numerical values computed at the output layer. The weight $w_{ij}$ is updated with an increment $\Delta w_{ji}$ and the error $E$ is reduced until an acceptable value of $E$ is reached. The output node that gives the highest value is set to 1 whereas the others are set to 0 in order to obtain the output class vector.

## 3. Results and discussion

From the statistical analysis, the 25 shape descriptors were divided according to their classification ability (see Table 2). To determine the suitable input features for training the MLP network, 3 sets of analyses were carried out using 3 features (5 groups), 9 features (5 and 4 groups), and 17 features (5, 4 and 3 groups) as given in Table 2. In each set, 50 simulated images were used for each defect type, giving a total of 300 images for 6 defect types for training. For testing the classification accuracy 60 images were used (10 images for each defect type). The output layer consists of 6 nodes corresponding to the 6 types of defects. The learning rate and error rate were set, respectively, to 1.0 and 0.

The results of comparison of the classification accuracies of the 3 sets of descriptors (at epoch value 100) are shown in Figure 7. The 3 features set gave the lowest overall accuracy compared to the other two. For the set containing 9 features, the accuracy is generally higher than that for the 17 features set. The comparison between the 9 features set and 17 features set shows that increase in the number of features will not necessarily increase the classification accuracy at the same number of hidden nodes. In addition, a larger features set will increase the processing time while resulting in similar results. Thus, in the current research 9 features set were selected as the optimized set of input features to train the network and classify the defects. The classification accuracy using simulated defects for various numbers of hidden nodes and epoch numbers for the 9 features set is shown in Table 3. The maximum accuracy (100%) was found to occur when 24 hidden nodes were used.

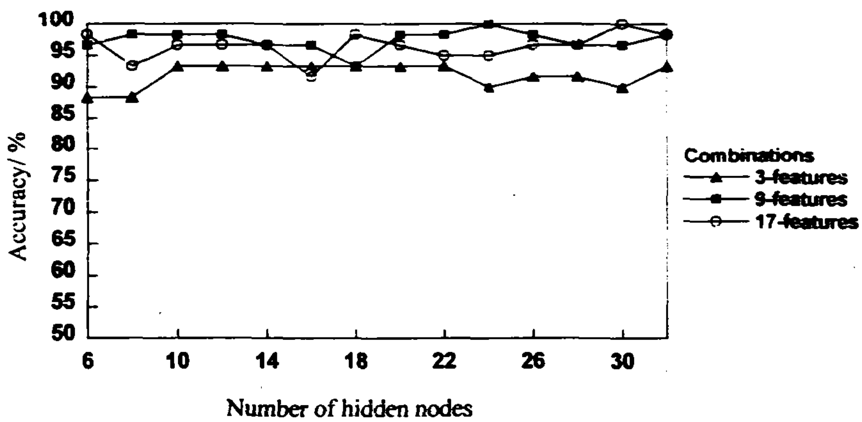



Figure 7. Classification accuracy of different shape descriptors set.

Table 3. Classification accuracy of simulated defects.

| Number of Nodes | Number of epochs | Accuracy/ % | | | Number of Nodes | Number of epochs | Accuracy/ % | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 3 features | 9 features | 17 features | | | 3 features | 9 features | 17 features |
| | 20 | 88.333 | 98.333 | 96.667 | | 20 | 90 | 98.333 | 96.667 |
| | 40 | 90 | 96.667 | 98.333 | | 40 | 91.667 | 98.333 | 95 |
| 6 | 60 | 88.333 | 96.667 | 98.333 | 20 | 60 | 93.333 | 98.333 | 96.667 |
| | 80 | 88.333 | 96.667 | 98.333 | | 80 | 93.333 | 98.333 | 96.667 |
| | 100 | 88.333 | 96.667 | 98.333 | | 100 | 93.333 | 98.333 | 96.667 |
| | 20 | 88.333 | 98.333 | 93.333 | | 20 | 88.333 | 100 | 96.667 |
| | 40 | 88.333 | 98.333 | 93.333 | | 40 | 91.667 | 100 | 96.667 |
| 8 | 60 | 88.333 | 98.333 | 93.333 | 22 | 60 | 93.333 | 98.333 | 95 |
| | 80 | 86.667 | 98.333 | 93.333 | | 80 | 95 | 98.333 | 95 |
| | 100 | 88.333 | 98.333 | 93.333 | | 100 | 93.333 | 98.333 | 95 |
| | 20 | 90 | 100 | 96.667 | | 20 | 90 | 100 | 98.333 |
| | 40 | 95 | 98.333 | 96.667 | | 40 | 91.667 | 100 | 96.667 |
| 10 | 60 | 93.333 | 98.333 | 96.667 | 24 | 60 | 93.333 | 100 | 96.667 |
| | 80 | 93.333 | 98.333 | 96.667 | | 80 | 91.667 | 100 | 96.667 |
| | 100 | 93.333 | 98.333 | 96.667 | | 100 | 90 | 100 | 95 |
| | 20 | 88.333 | 100 | 96.667 | | 20 | 91.667 | 98.333 | 98.333 |
| | 40 | 93.333 | 98.333 | 96.667 | | 40 | 93.333 | 98.333 | 98.333 |
| 12 | 60 | 99.333 | 98.333 | 96.667 | 26 | 60 | 93.333 | 98.333 | 96.667 |
| | 80 | 93.333 | 98.333 | 96.667 | | 80 | 93.333 | 98.333 | 96.667 |
| | 100 | 93.333 | 98.333 | 96.667 | | 100 | 91.667 | 98.333 | 96.667 |
| | 20 | 88.333 | 100 | 96.667 | | 20 | 90 | 98.333 | 96.667 |
| | 40 | 91.667 | 100 | 96.567 | | 40 | 91.657 | 96.667 | 96.667 |
| 14 | 60 | 93.333 | 96.667 | 96.667 | 28 | 60 | 91.667 | 96.667 | 96.667 |
| | 80 | 93.333 | 96.667 | 96.667 | | 80 | 91.667 | 96.667 | 96.667 |
| | 100 | 93.333 | 96.667 | 96.667 | | 100 | 91.667 | 96.667 | 96.667 |
| | 20 | 88.333 | 96.667 | 93.333 | | 20 | 91.667 | 100 | 96.667 |
| | 40 | 88.333 | 96.667 | 91.667 | | 40 | 91.667 | 98.333 | 98.333 |
| 16 | 60 | 88.333 | 96.667 | 91.667 | 30 | 60 | 91.667 | 98.333 | 100 |
| | 80 | 93.333 | 96.667 | 91.667 | | 80 | 91.667 | 98.333 | 100 |
| | 100 | 93.333 | 96.667 | 91.667 | | 100 | 90 | 96.667 | 100 |
| | 20 | 86.667 | 95 | 98.333 | | 20 | 86.667 | 100 | 98.333 |
| | 40 | 90 | 95 | 92.333 | | 40 | 90 | 98.333 | 98.333 |
| 18 | 60 | 91.667 | 95 | 98.333 | 32 | 60 | 91.667 | 93.333 | 98.333 |
| | 80 | 91.667 | 93.333 | 98.333 | | 80 | 93.333 | 98.333 | 98.333 |
| | 100 | 93.333 | 93.333 | 98.333 | | 100 | 93.333 | 98.333 | 98.333 |

The classification accuracies attainable using the shape descriptors proposed by Aoki and Suga [17] and Wang and Liao [18] were also investigated in the simulation study. Only the shape parameter sets defined by both authors as shown in Table 4 were used, whereas the defect images were same as those used in the current study.

Table 4. Different combination set of shape descriptors.

| Authors | Details of shape descriptors | Number of shape descriptors |
|---|---|---|
| Aoki & Suga [17] | Ratio between horizontal & perpendicular length, Vertical length to area ratio, Horizontal length to area ratio, Complexity, Formal Coefficient, Heywound diameter. | 6 |
| Wang & Liao [18] | Miu, Sigma, Circularity, Compactness, Major axis, Width and length, Elongation, Heywound diameter. | 9 |
| Current work | Solidity, Extent, Standard deviation of radius distribution, Compactness, Heywound diameter, Vertical length to area ratio, Major to Minor Axis Length, Curvature, Orientation | 9 |

Figure 8 shows a comparison of the classification accuracy between the three sets of shape descriptors. The optimized set of shape descriptors used in the current work gave a maximum classification accuracy of 100% using 24 nodes in the hidden layer. This classification accuracy is slightly higher than that obtained using the descriptor sets proposed by the other authors. Although the use of the optimized set of shape descriptors in the current work did not produce a significant improvement in the results compared to the other authors' work, this part of the study demonstrates the possibility of comparing different classification algorithms using the set of simulated images generated.

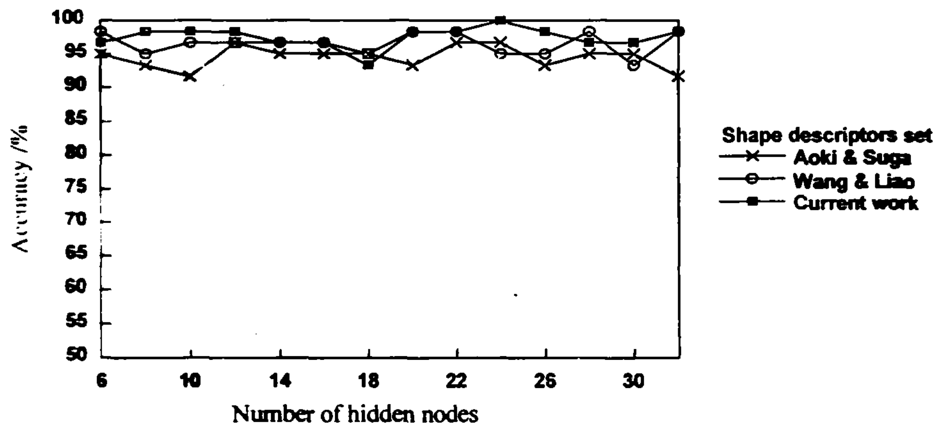



Figure 8. Comparison of classification accuracy with different shape descriptors from various authors.

### *3.1 Application to real images*

In order to investigate the usefulness of the simulation study, real images of defects were used for the classification while the training was conducted using the same data extracted from the simulated images. A series of pre-processing procedures using contrast enhancement and noise suppression were carried out to enhance the images. To extract the weld defects, the Background Subtraction Method (BSM) and Automatic Thresholding were applied [16-17]. The background was estimated using the second order polynomial given by [22]:

$$B(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy \qquad (3)$$

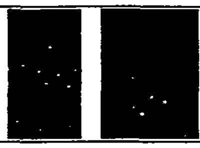
where $x$ and $y$ are pixel coordinates and $a_0 \ldots\ldots a_5$ are fitted constant. The automatic thresholding method was employed where the images were subdivided into a grid of smaller rectangular sub-regions and the pixel having the highest gray value in each sub-region was located. A total of 81 pixels from 9x9 sub-regions were selected and the information obtained, i.e. gray value and coordinates, were used to determine the coefficients in equation (3). Subsequently, the background image was estimated by substituting each coordinates into $B(x,y)$.

The results of applying the BSM method for porosity defect are shown in Figure 9. The original image was preprocessed using histogram equalization to enhance image contrast and noise was eliminated by using median filter. The background image was estimated and subtracted from the enhanced image. After the binarization, noise was removed and the shape descriptors were extracted from the resulting images. These descriptors were then used in the classification. A total of 49 defects from the real images were used for testing the classification accuracy. Some of the real images are shown in Table 5.

Original image | Image after enhancement | Background image | Image after background subtraction | After binarization

Figure 9. Result of background subtraction.

Table 5. Real images of radiographs

| Defect | Image | Defect | Image |
|---|---|---|---|
| Crack |  | Slag inclusion |  |
| Cavity |  | Transverse crack |  |
| Incomplete penetration |  | Porosity |  |

The highest classification accuracy achieved by the neural network using features extracted from the real weld images is 97.96% (Figure 10). This value is slightly lower than the accuracy achieved for classifying simulated images (100%). This is expected because real images usually contain much more irregularities and noise than the simulated images of 'ideal' defects.

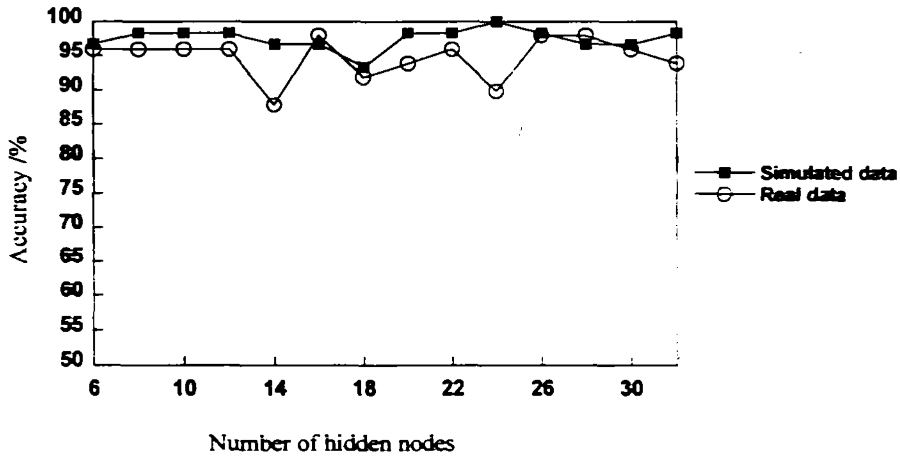



Figure 10. Classification accuracy of simulated and real defects.

The classification accuracy of individual defect type is shown in Figure 11. For individual comparison, it was found that the accuracy varies with the number of nodes. The highest average accuracy is 100% for slag inclusion and the lowest is for incomplete penetration (89.29%). In some cases, incomplete penetration was misclassified as crack. As a matter of fact, only the feature *Extent* shows a significant difference from multiple comparison tests while the rest of the features show joining groups for incomplete penetration and crack. In other words, among the 9 features selected only one feature, i.e. *Extent*, offers the highest ability to differentiate between these two defects.

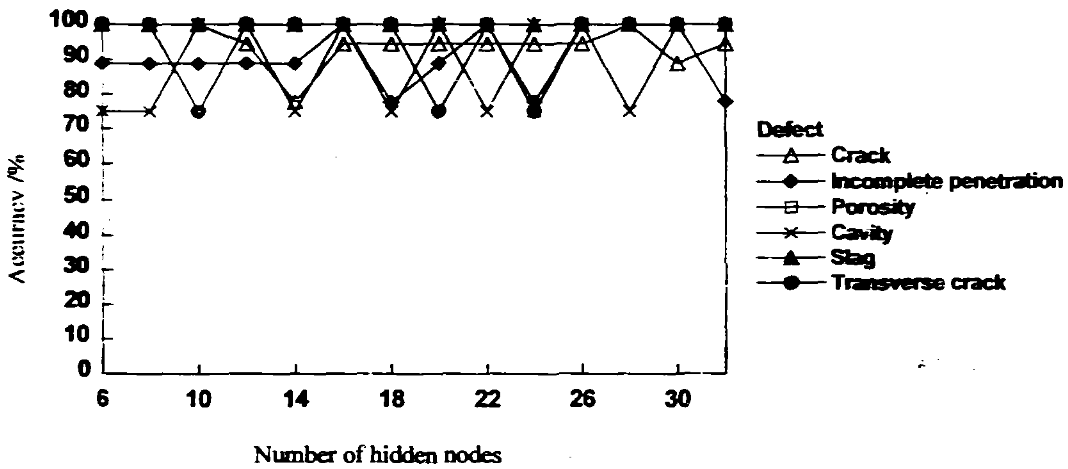



Figure 11. Classification accuracy of individual defects.

## 5. Conclusion

An effective weld defect classification method by using shape descriptors extracted from simulated defects has been developed and demonstrated by classifying real welding defects. The general shape descriptors, which are used to characterize weld defects, were studied and an optimized set of shape descriptors than can effectively classify the weld defects have been identified using a multiple comparison procedure. In order to test the effectiveness of the optimized shape descriptors in classifying defects, shape descriptors extracted from the simulated defects were used to train the network. The simulation study produced a maximum classification accuracy of 100%.

Classification of real defects using the simulated training samples gave the highest overall accuracy of 97.96%. Hence, the proposed defect classification method overcomes the problem of limited real defect samples for classification using neural network. As demonstrated in this work, the simulated defects can also be used as ideal or benchmark images to compare the performance of various classification techniques proposed by other authors.

## 6. Acknowledgements

The authors would like to thank Universiti Sains Malaysia and the Ministry of Science, Technology & Environment for the offer of the IRPA Grant that has enabled this research to be carried out.

## 7. References

[1] Liao TW, Li YM. An automated radiographic NDT system for weld inspection: Part II: Flaw detection. NDT& E Int. 1998; 31(3):183-192.

[2] Fucsok F, Scharmach M. Human factors: The NDE reliability of routine radiographic film evaluation. Proceedings of 15th World Conference on Non-Destructive Testing 2000 (www.ndt.net/article/wcndt00/papers/idn740/idn740.htm).

[3] Lashkia V. Defect detection in X-ray images using fuzzy reasoning. Image and Vision Computing 2001;19:261-269.

[4] Liao TW. Ni JW. An automated radiographic NDT system for weld inspection: Part I – Flaw detection. NDT& E International 1996; 29(3):157-162.

[5] Liao TW, Tang K. Automated extraction of welds from digitized radiographic images based on mlp neural networks. Applied Intelligence, 1997, 11: 197-218.

[6] Liao TW, Li DM, Li YM. Extraction of welds from radiographic images with fuzzy classifiers. Information Science 2000; 126:21-42.

[7] Kehoe A, Parker GA, LeBlanc A. Image processing for automatic defect detection in industrial radiographic images. Image Processing and its Applications – Third International Conference 1989:202 -206.

[8] Murakami K, Image processing for non-destructive testing. Welding International 1990; 4(2): 144-149.

[9] Bonser G, Lawson, SW. Defect detection in partially completed SAW and TIG welds using online radioscopy and image processing. Proceedings of SPIE: Process Control and Sensors for Manufacturing 1998; 3399: 231-239.

[10] Kaftandijan V. Automatic detection and characteristics of aluminum weld defects: comparison between radiography, radioscopy and human interpretation. In: Proceedings of the 7th European Conferences on Non Destructive Testing, Copenhagen, May 26-29 1998; 1179-1186.

[11] Palenchika RM, Alekseichuk AV. Flaw detection in radiographic images by structure-adaptive binary segmentation. In: Proceedings of International Symposium on Computerized Tomography for Industrial Applications and Image Processing in Radiology. Berlin, March 15 – 17 1999; DGZfP Proceedings, 221-232.

[12] Liao TW, Li YM, An automated radiographic NDT system, for weld inspection: part II. Flaw Detection, NDT and E International 1998, 31(3): 183-192.

[13] Liao TW, Li D-M, Li Y-M. Detection of welding flaws from radiographic images with fuzzy clustering methods. Fuzzy Sets and Systems 1999; 108:145-158.

[14] Jacobsen C, Zscherpel U. Automated evaluation of digitized radiograph with neuronal methods. In: Proceedings of Computed Tomography for Industrial Application and Image Processing in Radiology, Berlin, March 15-17, 1999; 141-152

[15] Perner P, Zscherpel U, Jacobsen C. A comparison between neural network and decision trees based on data from industrial radiographic testing. Pattern Recognition Letters 2001; 22:47-54.

[16] Feher Z. Computer aided processing of industrial radiographs. Periodica Polytechnica Ser. El. Eng. 2001; 44(3-4):241-248.

[17] Aoki K, Suga Y. Application of artificial neural network to discrimination of defect type in automatic radiographic testing of welds. ISIJ International 1999; 39(10):1081-1087.

[18] Wang G, Liao TW. Automatic identification of different types of welding defects in radiographic images. NDT&E International 2002; 35: 519-528.

[19] Xu K, Luxmoore AR, Deravi F. Comparison of shape features for the classification of wear particles. Engineering Application of Artificial Intelligence 1997;10(5):485-493.

[20] Sonka M, Hilavac V, Boyle R, *Image Processing, Analysis, and Machine Vision*, Second Edition. PWS Publishing, (USA); 1998:228-289.

[21] Metre BM, Kankahalli MS, Lee WF. Shape measures for content based retrieval: A comparison, information processing & management, Vol. 33(3), p319-337, 1997.

[22] Russ JC. *The Image Processing Handbook* 1999. 3rd edition, CRC Press, USA, pp.509-574.

[23] Williams, D. J. and Shah, M. A fast algorithm for active contours and curvature estimation, *CVGIP: Image Understanding* 1992; 55:14-26.

[24] Weber DC, Skillings JH. *A First Course in the Design of Experiments: A linear Model Approach* 2000:281-306, CRC Press, USA.

[25] Montgomery DC. *Design and analysis of experiments* 1996:95-117, John Wiley & Sons, USA.

# Automatic Label Removal from Digitized Weld Radiographs

Soo Say Leong, Zahurin Samad, Mani Maran Ratnam[a]

School of Mechanical Engineering, Universiti Sains Malaysia,

Engineering Campus, Sri Ampangan,

14300 Nibong Tebal, Penang, Malaysia.

[a]*Corresponding author*: Tel.: +6-04-5937788. Fax: +6-04-5941025

*e-mail*: mmaran@eng.usm.my

Mohd Ashhar Khalid

Intelligent Systems Group,

Malaysia Institute of Nuclear Technology Research (MINT),

43650 Bangi, Selangor, Malaysia.

**Abstract**

This paper presents a methodology to remove labels automatically from digitized weld radiographs as part of the automatic weld defect detection process. An algorithm was developed to detect and remove labels printed onto weld radiographs before weld extraction algorithm or defect detection algorithm is applied. Normality test was used to determine if the intensity profile parallel to the weld contains label pixels. Thresholding followed by region filling operations were carried out to remove the labels. The algorithm was tested on 50 weld radiographs with labels and the labels on 90% of these images were successfully removed.

*Keywords*: Weld radiography, label removal.