

## Feature Extraction and Optimisation for X-ray Weld Image Classification

Ying YIN      Gui Y. TIAN

School of Electrical, Electronic and Computing Engineering,

Merz Court, Newcastle University,

NE1 7RU, Newcastle Upon Tyne, UK

Tel: +44(0) 191 222 5639

Email: [ying.yin1@ncl.ac.uk](mailto:ying.yin1@ncl.ac.uk) [g.y.tian@ncl.ac.uk](mailto:g.y.tian@ncl.ac.uk)

### Abstract

Computer aided image analysis systems for radiographic inspection (X-ray or gamma ray) are among the most commonly used Non-destructive Evaluation (NDE) methods. The accuracy of these systems is very much depending on the selected features which are extracted from weld defect images. In this paper, we firstly introduce a computer aided image analysis system for X-ray image inspection and evaluation. Then, a feature optimisation approach including features extracted based on geometrical shape, edge chain code (ECC) and geometric moment invariants (GMI), feature comparison and feature selection is proposed to get the best features for classification. 7 shape geometry features, 7 ECC features and 4 GMI features are extracted and tested separately. After feature optimisation, 13 features are selected and kept. Finally, a feed-forward back-propagation neural network is implemented for the purpose of defect classification. The experimental results have proved that the new feature extraction and optimisation approach successfully improves the weld defect identification accuracy.

**Keywords:** Non-destructive evaluation, Feature extraction, Feature optimization, defect classification

### 1. Introduction

The non-destructive testing (NDT) and non-destructive evaluation (NDE) techniques have a wide range of applications in industries via Visual inspection, Radiography, Ultrasonic, Eddy current <sup>[1]</sup>, Acoustic emission and Infrared thermograph. Radiographic methods of inspection have been widely used to evaluate the integrity of materials and equipment. The defects in weld joints are commonly caused by porosity, slag inclusion, cracking, undercut, burn through, lack of fusion, etc. <sup>[2]</sup>. Computer aided image analysis systems developed for inspecting defects automatically consist of three major parts: image processing, feature extraction and pattern recognition. Feature extraction is the operation to extract various image features for classification or interpretation of meaningful objects from images and is considered as one of the most important parts. It can be defined as a function of one or more measurements, each of which specify some quantifiable property of an object, and are computed such that it quantifies some significant characteristics of the object <sup>[3]</sup>.

In real time manual radiography analysis, the interpreters normally judge film based on the shape and the intensity of each defect. A lot of research on feature extraction for X-ray weld defect images has been carried out in the past few decades, but most of them are more focussed on shape geometry features. Shafeek et al. <sup>[4]</sup> proposed an 8-neighbourhoods edge chain code (ECC) algorithm to identify the contours of defects. These ECC parameters are used to calculate area, perimeter, width and height of defects by creating short lines between each two successive boundary pixels. In the image processing parts of Jimenex and Jain's paper <sup>[5]</sup>, colour and shape analysis methods are implemented for automatic fruit recognition. Liao et al. have used a set of parameters to identify six different types of defects and obtain

the highest accuracy of 92%. In their work, 108 data sets are investigated and only 12 of them are applied for testing [6]. Nafaa et al. [7] have used a bunch of geometrical features plus two geometric moment invariants (GMI) to construct a set of weld defect descriptors in X-ray images and developed a classifier based on a multiplayer feed forward neural network. Silva et al. [8] present some other popular geometric features for identification such as: Position, Ratio of aspect, Ratio between the width and the area, Roundness, Angle, Area/rectangle and Rectangle or 'box' in their work.

This paper presents an approach to extract and optimise features based on the geometrical shape of the defects, ECC of the defects and GMI of the defect images. In this paper, the structure of the computer aided image analysis system is described. Then, the feature extraction approach is presented. In addition, the performance of the features is demonstrated and experimental testing using ANNs provide a way to compare and discuss the efficiency of the system using these features. Furthermore, some of the high efficiency features will be selected for defect classification.

## 2. Computer Aided Image Analysis System for X-ray Weld Images

In the computer aided image analysis system, image processing intends to enhance image quality, remove image noise, and obtain the region of interest. Discriminating and independent feature extraction is the key to the pattern recognition being successful. Features are the individual measurable characteristics of the pattern being observed. Pattern recognition aims to classify patterns based on advanced defined feature extracted. In our system, there are two stages: training stage and identification stage (see fig 1).

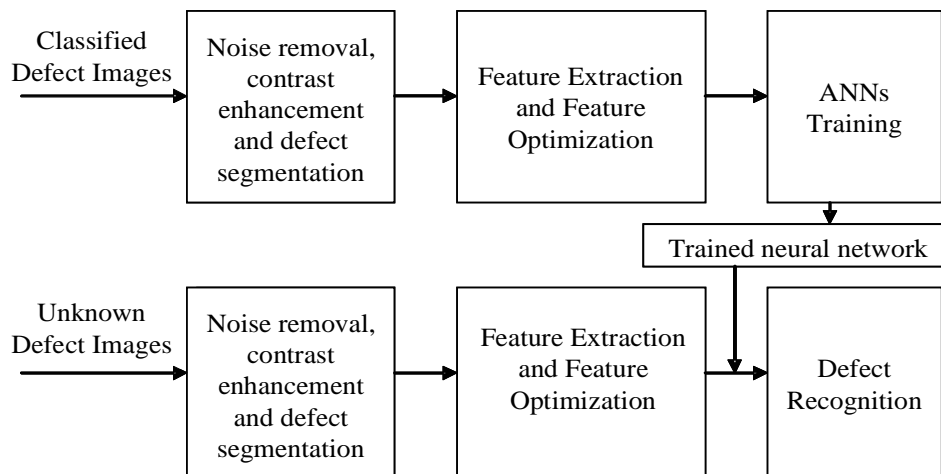


Figure 1. A framework for a computer aided image analysis system

## 3. Digital Image Processing

The images digitised from the X-ray inspection system are normally in very poor condition with low contrast and a high level of noise. For this reason, image pre-processing techniques such as contrast enhancement and noise removal are implemented first. We present several approaches including: local histogram equalization, adaptive morphological filter and eigenvector based image enhancement for image pre-processing and an innovative Principal Component Analysis (PCA) based image segmentation is developed.

## 4. Feature extraction

### 4.1 Shape Geometrical Feature Extraction

The 7 most commonly used shape geometry features extracted from a weld defect are defined and described in Table 1.

**Table 1: Shape features and their physics explanation**

Features	Equation	Feature Description
Anisometry	$Ani = \frac{MajorAL}{MinorAL}$	<i>Ani</i> depends on the direction of principal axis of the defect. Its value is proportional to defect lengthening.
Circular Rate	$CR = \frac{P^2}{Area}$	When the defect is circular, the value of feature <i>CR</i> is around 1.
Ratio between Width and Area	$WA = \frac{MinorAL}{Area}$	<i>WA</i> is very sensitive for discrimination between fusion defects and cracks.
Compactness	$Comp = \frac{4\pi Area}{P^2}$	<i>Comp</i> has little value for sharp defects (crack, lack of fusion) and big values for spherical defects (Porosity, Slag Inclusion, etc.).
Rectangularity	$Rect = \frac{Area}{WBB \times HBB}$	<i>Rect</i> is closely equal to 1 for a rectangular shape defect.
Elongation	$Elong = \frac{WBB}{HBB}$	<i>Elong</i> characterises longitudinal defects, such as cracks.
AP	$AP = \frac{Area}{P}$	The long and narrow defects and irregular defects have a small value of feature <i>AP</i> .

Where *MajorAL* is the larger axis of the ellipse of the area equivalent to the defect and *MinorAL* is the smaller axis. *P* is the perimeter of the defect shape while *Area* is the actual number of pixels in this region. *WBB* and *HBB* are the width and the height of the smaller rectangle that encloses the defect respectively. The performance of 7 shape geometrical features for different types of defect can be seen in figure 3.

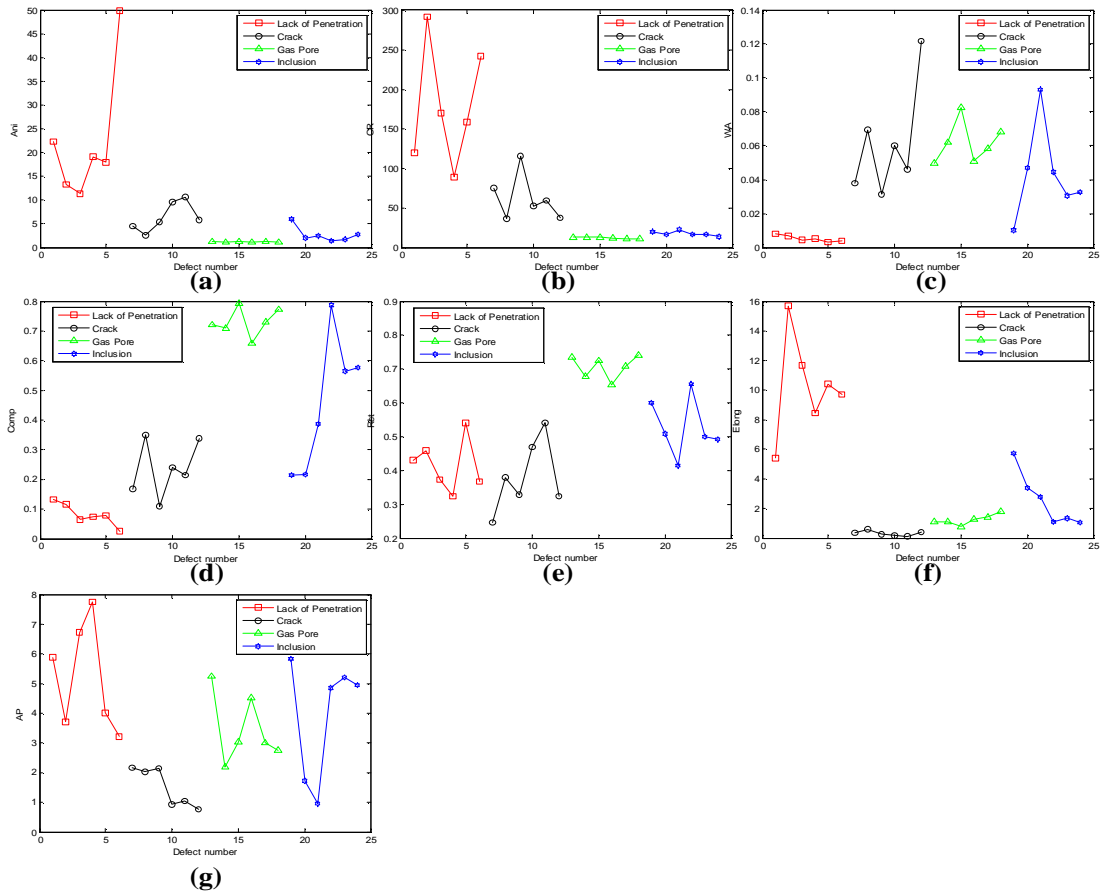
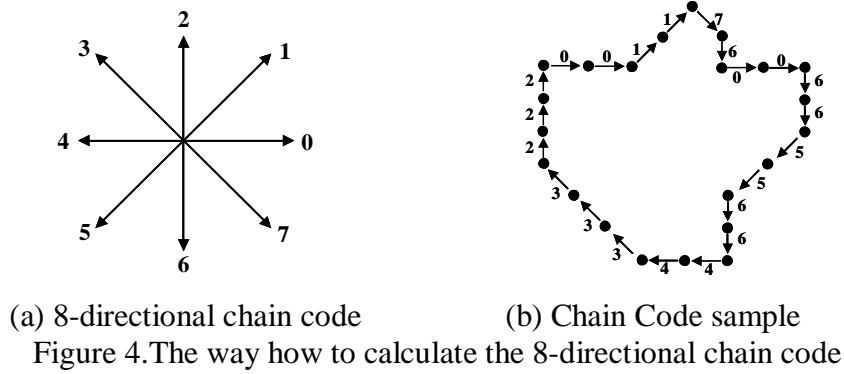


Figure 3. The performance of 7 shape geometrical features; (a) *Ani*; (b) *CR*; (c) *WA*; (d) *Comp*; (e) *Rect*; (f) *Elong*; (g) *AP*

## 4.2 Edge Chain Code Features

Edge chain code is a representation that consists of series of numbers. The numbers represent direction of the next pixel that can be used to represent shape and input format for numerous shape analysis algorithms. Since introduced by Freeman<sup>[9]</sup>, works on chain code in representing images and describing and recognizing shapes have been given consideration as an alternative method. Actually, chain code is a list of codes ranging from 0 to 7 in a clockwise direction. These codes represent the direction of the next pixel connected in a 3×3 window as shown in Figure 4 (a). For example, in figure 4 (b), we start at the first edge on the top left and go clockwise around the edge. We add the code for each edge to a list, which becomes our chain code: 001176006655664433222.



Here is a way to describe distribution ( $X_1, \dots, X_N$ ) using statistical properties called moments. Based on the understanding of moments and considering the defect edge chain code as a distribution, the 6 new features shown in Table 2 can be defined and described to analyse the defect's edge:

**Table 2: Chain code features and their physics explanations**

Feature	Equation	Physics Description
mean_1 <sup>st</sup>	$\bar{X} = \frac{1}{N} \sum_{j=1}^N X_j$	Estimates that value around which central clustering occurs
var_2 <sup>nd</sup>	$\overline{Var}(X_1 \dots X_N) = \frac{1}{N-1} \sum_{j=1}^N (X_j - \bar{X})^2$	Having characterised a distribution's central value, we next characterizes its "width" or "variability" around that value.
SD_2 <sup>nd</sup>	$\sigma(X_1 \dots X_N) = \sqrt{\overline{Var}(X_1 \dots X_N)}$	var_2 <sup>nd</sup> 's square root
ADev_2 <sup>nd</sup>	$ADev(X_1 \dots X_N) = \frac{1}{N} \sum_{j=1}^N  X_j - \bar{X} $	A more robust estimator of width
skew_3 <sup>rd</sup>	$Skew(X_1 \dots X_N) = \frac{1}{N} \sum_{j=1}^N \left[ \frac{X_j - \bar{X}}{\sigma} \right]^3$	Characterises the degree of asymmetry of a distribution around its mean.
Kurt_4 <sup>th</sup>	$Kurt(X_1 \dots X_N) = \left\{ \frac{1}{N} \sum_{j=1}^N \left[ \frac{X_j - \bar{X}}{\sigma} \right]^4 \right\} - 3$	Measures the relative peakedness or flatness of a distribution.

The 6 features above aim to analyse the chain code as a distribution. On the other hand, the code for each pixel represents the direction changing from the last pixel. So, all those codes together represent the change for the entire edge. It is possible to evaluate whether the edge is smooth or not by measuring the average chain code change. Because different types of welding defect have totally different edge roughness appearance, for example Crack defects

have a much rougher edge compared with Lack of penetration. The chain code change from one pixel to the next pixel can be obtained by equation 1:

$$\Delta r_i = |r_{i+1} - r_i| \quad (1)$$

Then, the average change ratio can be calculated by equation 2:

$$R_v = \frac{\sum_{i=0}^{N-1} \Delta r_i}{N-1} \quad (2)$$

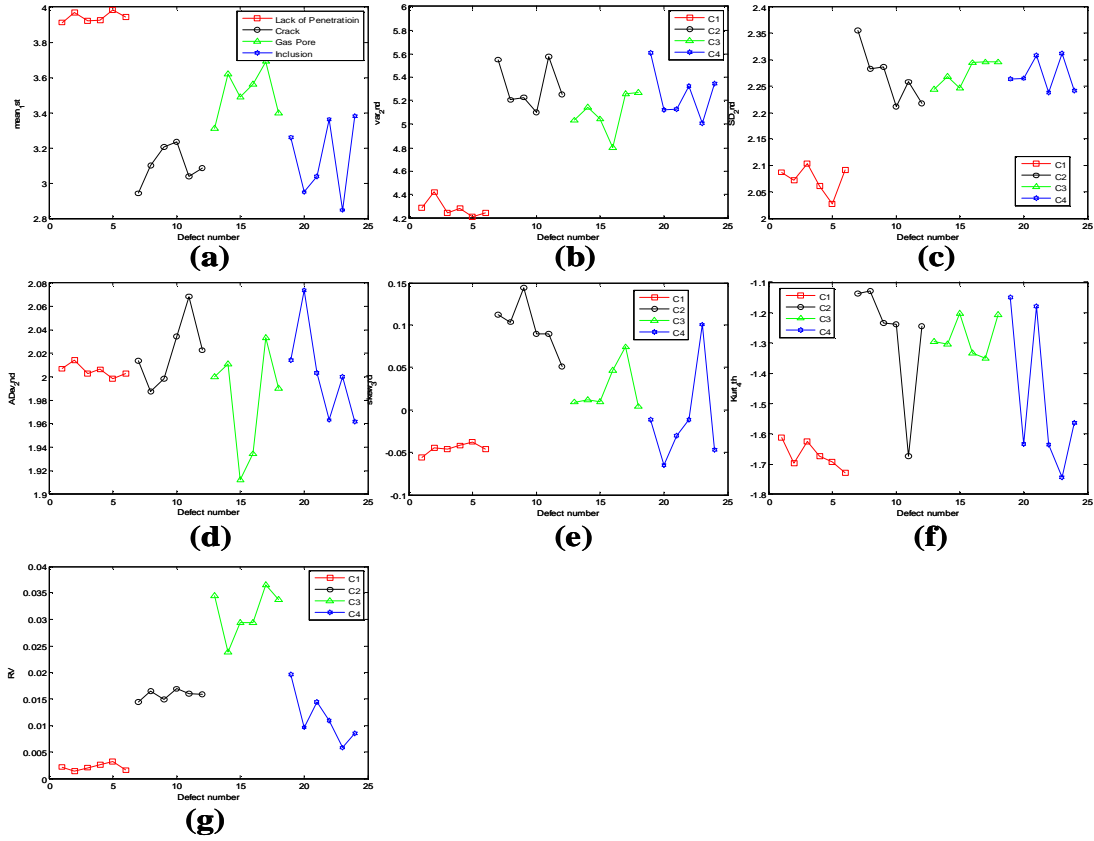


Figure 5: The performance of Edge Chain Code features

### 4.3 Geometric Moment Invariants

Hu <sup>[10]</sup> introduced the use of moment invariants as features for pattern recognition. These features generated are Rotation Scale Translation (RST)-invariant have been extensively used. Rizon <sup>[11]</sup> has provide that GMI can produce a set of feature vectors and get more than 90% identification accuracy for object detection. Two-dimensional moments of a digitally sampled  $M \times M$  image that has intensity function  $f(x, y)$ ,  $(x, y=1, 2, 3, \dots, M)$  is given as,

$$m_{pq} = \sum_{x=1}^{x=M} \sum_{y=1}^{y=M} (x)^p \cdot (y)^q f(x, y) \quad (3)$$

$p, q = 1, 2, 3, \dots$ . Thus the central moments  $m_{pq}$  or  $\mu_{pq}$  can be computed:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \text{and} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q f(x, y) \quad (4)$$

When a scaling normalisation is applied the central moments change as,

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \gamma = [(p + q) / 2] + 1 \quad (5)$$

In terms of the central moments, the four moments are given as below:

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \quad \text{and} \quad \phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \end{aligned} \quad (6)$$

Based on the 4 moments listed in eq.6, 4 features can be extracted and compared in figure 10 and 11 below:

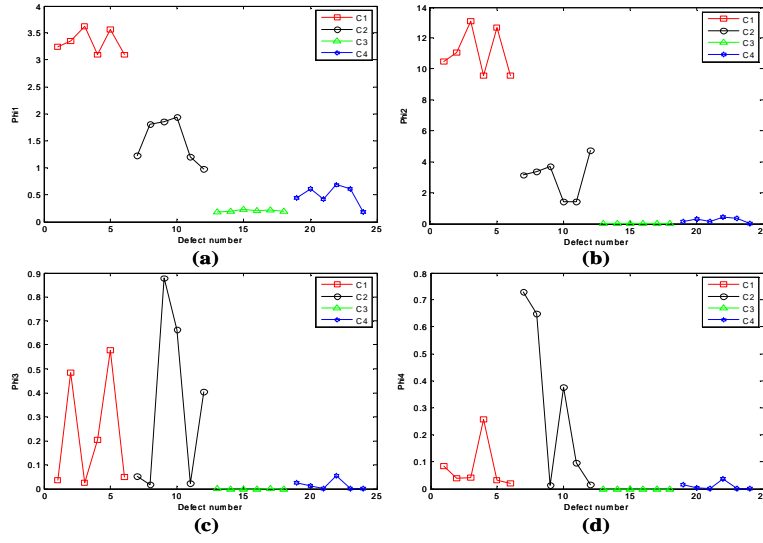


Figure 6. The performance of feature  $\phi_1$ ,  $\phi_2$ ,  $\phi_3$  and  $\phi_4$

## 5. Experimental Testing

### 5.1 Testing Results Using 7 Shape Geometrical Features

For the first stage, only shape descriptors are used for testing, because they have been commonly used and are considered as an effective way to classify different types of welding defect. The results can be seen in Table 3:

**Table 3. Identification accuracy rate using 7 geometrical features**

Defect	Number of Samples	Successfully Identified	Wrongly Identified	Identification Accuracy
C1	27	24	3	88.89%
C2	20	11	9	55.00%
C3	58	56	2	96.55%
C4	54	46	8	85.19%
Over all	86.17%			

Generally, we can get 86.17% of defects, which means 137 out of 159 samples are identified using these 7 geometrical descriptors. Each group of defects have more than 85% accuracy except C2. Crack defects have a long, narrow and rectangular shape. They can be somehow wrongly classified into C1. Using ECC features can improve the identification ability, because of the significant difference on the edge roughness for C1 and C2 defects.

### 3.2 Testing results using 7 chain code features

The results of classification using the ECC features can be seen in table 4.

**Table 4. Testing results using features individually**

Feature name	Identification accuracy
Mean_1 <sup>st</sup>	62.26%
Var_2 <sup>nd</sup>	37.74%
SD_2 <sup>nd</sup>	27.04%
ADev_2 <sup>nd</sup>	8.18%
Skew_3 <sup>rd</sup>	49.69%
Kurt_4 <sup>th</sup>	42.77%
RV	67.30%

**Table 5. Testing results using feature combination**

Feature combination	Identification Accuracy
Mean_1 <sup>st</sup> + Skew_3 <sup>rd</sup> + Kurt_4 <sup>th</sup> + RV	77.99%
Mean_1 <sup>st</sup> + RV	71.70%
Skew_3 <sup>rd</sup> + Kurt_4 <sup>th</sup>	40.88%
Mean_1 <sup>st</sup> + Skew_3 <sup>rd</sup>	65.40%
Kurt_4 <sup>th</sup> + RV	57.23%

Clearly, Var\_2<sup>nd</sup>, SD\_2<sup>nd</sup> and ADev\_2<sup>nd</sup> don't present good identification accuracy as we expect. So before we move to next step, we first abandon these three features. Identification accuracy using different combination of ECC features can be seen in table 5. We select 4 ECC features for further investigation and test results can be found in table 6.

**Table 6. Identification accuracy using 7 geometrical features and 4 chain code features**

Defect	Number of Samples	Successfully Identified	Wrongly Identified	Identification Accuracy
C1	27	24	3	88.89%
C2	20	17	3	85.5%
C3	58	54	4	93.10%
C4	54	45	9	83.33%
<b>Over all</b>	<b>88.05%</b>			

From Table 6 we can see that, the identification accuracy for C2 has been improved significantly, from 55% to 85.5% by the help of chain code features. However, the overall classification rate has not been increased a lot, from 86.17% to 88.05% due to the small number of C2 sample.

### 3.3 Testing results using 4 Geometric Moment Invariants features

Classification using the 4 geometric moment invariants features individually and the identification accuracy can be seen in table 7.

**Table 7. Testing results using features individually**

Feature name	Identification accuracy	Feature combination	Identification accuracy
$\phi_1$	77.99%	$\phi_1$ and $\phi_2$	84.28%
$\phi_2$	83.01%	$\phi_1$ and $\phi_3$	62.26%
$\phi_3$	59.12%	$\phi_2$ and $\phi_4$	67.92%
$\phi_4$	61.64%	$\phi_3$ and $\phi_4$	61.64%

Clearly,  $\phi_3$  and  $\phi_4$  did not present good identification accuracy as we expect. However, we tried to find the results after combine these 4 features in several different ways. The results prove that the invariants  $\phi_1$  and  $\phi_2$  work well for defect classification respectively and together. By using 11 features including 7 shape geometrical features and 4 edge chain code features, we successfully classify 88.05% of samples which means 140 out of 150 defect samples can be successfully identified. After studying geometric moment invariant features,

we have two more features: the invariants  $\phi_1$  and  $\phi_2$ , which are considered to improve overall identification accuracy. The test results can be found in table 9.

**Table 9. Identification accuracy using 7 geometrical features and 4 chain code features;**

Defect	Number of Samples	Successfully Identified	Wrongly Identified	Identification Accuracy
C1	27	26	1	96.29%
C2	20	17	3	85.5%
C3	58	57	1	98.28%
C4	54	49	5	90.74%
<b>Over all</b>	<b>93.71%</b>			

#### 4. Conclusion

In this proposed method, 7 geometrical features, 7 ECC features and 4 GMI features have been extracted from X-ray weld defect images. All 17 features have been studied and discussed. Based on the comparison, 4 edge chain code features and 2 geometric moment invariants features are selected. According to experimental testing, 86.17% of defect samples can be classified using only geometrical features. After feature optimisation, 4 ECC features and 2 GMI features are kept. The identification accuracy can be improved to 88.05% and 93.71% step by step. All test results prove that the defect identification ability can be increased using the proposed feature extraction and optimisation method.

#### References

- [1] GuiYun T, Sophian A, Taylor D, Rudlin J. Wavelet-based PCA defect classification and quantification for pulsed eddy current NDT. Science, Measurement and Technology, IEE Proceedings-Volume 152, Issue 4, 8 July 2005 Page(s):141 – 148.
- [2] Alaknanda R S Anand and Pradeep Kumar. Flaw detection in radiographic weld images using morphological approach. NDT & E International, Volume 39, Issue 1, January 2006, Pages 29-33.
- [3] Kenneth R. Castleman. Digital Image Processing. Prentice Hall. 1996.
- [4] Shafeek HI, Gadelmawla ES, Abdel-Shafy AA and Elewa IM. Assessment of welding defects for gas pipeline radiographs using computer vision. NDT & E International, Volume 37, Issue 4, June 2004, Pages 291-299.
- [5] Jimenex AR and Jain AK, Ceres R, Pons JL. Automatic fruit recognition: a survey and new results using Range/Attenuation images. Pattern Recognition 32 (1999) 1719-1736
- [6] Gang W and Liao TW. Automatic identification of different types of welding defects in radiographic images. NDT & E International, Volume 35, Issue 8, December 2002, Pages 519-528.
- [7] Nafaa N, Latifa H, Djemel Z. Statistical tools for weld defect evaluation in radiographic testing. European Conference on Non-destructive Testing 2006 Proceedings, Mo.2.4.4.
- [8] Romeu R. da Silva, Marcio H.S. Siqueira, Marcos Paulo Vieira de Souza, João M.A. Rebello and Luiz P. Calôba. Estimated accuracy of classification of defects detected in welded joints by radiographic tests. NDT & E International, Volume 38, Issue 5, July 2005, Pages 335-343.
- [9] Herbert F. Computer Processing of Line-drawing Images. Computing Surveys, 6(1):57-97, March 1974.
- [10] Hu M K. Visual pattern recognition by moments invariants. IRE Trans. Information Theory, 8: 179-87.
- [11] Mohamed R, Haniza Y, Puteh S. Object detection using geometric invariant moment. American Journal of Applied Sciences 2 (6): 1876-1878, 2006.