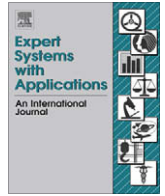




Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Multiclass defect detection and classification in weld radiographic images using geometric and texture features

Ioannis Valavanis^a, Dimitrios Kosmopoulos^{b,*}^aZenon Automation Technologies, 15354 Glyka Nera, Greece^bComputational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", 15310 Aghia Paraskevi, Greece

ARTICLE INFO

Keywords:

Radiography
Welds
Defects
Segmentation
Classification
Texture
Geometrical features

ABSTRACT

In this paper, a method for the detection and classification of defects in weld radiographs is presented. The method has been applied for detecting and discriminating discontinuities in the weld images that may correspond to false alarms or defects such as worm holes, porosity, linear slag inclusion, gas pores, lack of fusion or crack. A set of 43 descriptors corresponding to texture measurements and geometrical features is extracted for each segmented object and given as input to a classifier. The classifier is trained to classify each of the objects it into one of the defect classes or characterize it as non-defect. Three fold cross validation was utilized and experimental results are reported for three different classifiers (Support Vector Machine, Neural Network, *k*-NN).

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The inspection of welds is a very important task for assuring safety and reliability in several industrial sectors, e.g., ship and aircraft industry. For this purpose Non-Destructive Testing (NDT) techniques have been employed to test a material for surface or internal flaws without interfering in any way with its suitability for service. Such methods are the acoustic emission, magnetic particle inspection, eddy current, ultrasonic testing, thermal inspection and several others (Anuncia & Saravanan, 2006). These techniques are based on the observation that weld defects cause some sort of discontinuity to the test signal, which allows for recognition. However, each method is appropriate only for specific types of defects.

On the contrary, radiography (X-rays or sometimes gamma rays) seems to be the most effective method and the experts are able to identify most types of defects in the images produced by this method. The method is based on the fact that the defective areas absorb more energy and thus the defects appear darker in the image (Hayes, 1997).

Some of the most common weld defects that can be identified in the radiographic images are the worm holes (worm-like cavities), slag inclusion (slag or other foreign matter entrapped during welding), linear porosity (linear cavities due to entrapped gas), gas pores (spherical cavities due to entrapped gas), lack of fusion (lack

of union between weld and parent metal) or crack (discontinuity by fracture in the metal). Examples views of each of those defects are given in Fig. 1.

The interpretation of weld radiographs even by experienced inspectors can, however, be subjective and time-consuming and some types of defects could remain undetected due to lack of time. Even the same persons may give contradicting evaluations for the same images, depending on their mental states (Nockemann, Heidt, & Thomsen, 1991). Thus, several researchers have tried to automate the inspection process by employing image processing and pattern recognition methods. The goal of such methods is to give consistent, objective and reliable results.

In the related literature to our knowledge the features that are used do not combine all available sources of information (intensity, geometry and texture) to solve the general multi-class defect detection and classification problem. This is obviously due to lack of appropriately designed features that can capture the attributes of each defect type. Our work tries to fill this gap without sacrificing performance. The innovation of the proposed paper regards the combination of the following:

- Multimodal feature definition using intensity texture and geometric characteristics, thus capturing all visual attributes.
- Feature selection to limit processing to those features that are actually important for each different class avoiding information redundancy.
- Defect detection and classification into seven different classes (including segmented non-defects) comparing some state of the art multi-class classification methods (Support Vector Machines and Neural Networks).

* Corresponding author.

E-mail addresses: ivalavan@zenon.gr (I. Valavanis), dkosmo@iit.demokritos.gr (D. Kosmopoulos).


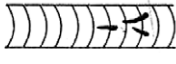



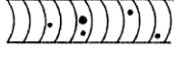

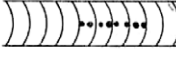

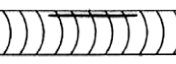

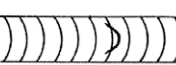

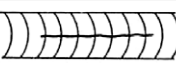

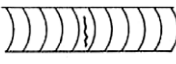


Description	Cross-section of weld	Radiogram
Worm hole		
Linear Slag Inclusion		
Gas Pore		
Porosity (Linear)		
Lack of side-wall fusion - (lack of root fusion)		
Lack of inter-run fusion		
Longitudinal Crack		
Transverse Crack		
Radiating Cracks		

Fig. 1. Types of weld defects in cross section and how it is expected to appear on a radiographic image: worm hole, linear slag inclusion, gas pore, linear porosity, lack of fusion, various types of cracks. Source: From British Standard, 1998.

In the next section the related work is presented. Then the weld radiographic images and the methodology for defect detection (segmentation) and classification are presented in Sections 3 and 4, respectively. In Section 5 the experimental results are reported and Section 6 concludes this paper.

2. Related work

The potential benefits of a fully automated system for weld defect detection and evaluation has motivated a lot of research in the related field. More or less the systems presented so far follow the process described in Fig. 2. After digital image acquisition only a region of interest (ROI) is further processed. Some preprocessing may take place like noise reduction (e.g., gaussian or median filtering) and contrast enhancement (e.g., histogram equalization) to assist the defect segmentation.

Then segmentation of regions that may represent defects is done. Several methods can be used to segment the defect, i.e., detect the defect on the image, ranging from simple segmentation methods (e.g., thresholding) to more advanced methods that combine background subtraction, as explained in the following. In Nacereddine, Tridi, Hamami, and Ziou (2006), Otsu's, Niblack's and Sauvola histogram threshold methods are comparatively assessed in the weld defect detection and Sauvola method is proposed. Sofia and Redouane (2002) describe a defect segmentation method based on a watershed algorithm and morphological operations (erosion and dilation). Carrasco and Mery (2004) propose a set of robust image processing techniques (bottom hat filter, morphological operators and watershed algorithm)

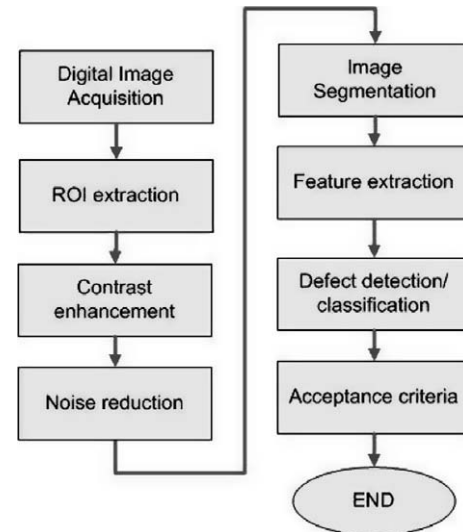


Fig. 2. General process flow of radiography-based automated weld defect recognition systems.

that are appropriately combined to segment defects. Background subtraction methods construct a reference defect-free image (Sood, Blakeley, & Rebuffel, 2006) (known as golden image) and subtract it from the original image to detect defects, however the creation of this golden image is normally non-trivial.

As soon as the defects are segmented using one of the previous methods, features can be extracted and then given as input to classifiers to detect possible defects and eventually to identify the exact defect type. Optionally the defect dimensions are compared to some acceptance criteria defined by experts or international standards and a decision is taken on the acceptability of the monitored weld. In the following we are going to focus on feature extraction and classification methods.

2.1. Feature definition

Several types of features have been used in literature. Among the simplest ones are the geometric features. In Da Silva, Siqueira, De Souza, Rebello, and Caloba (2005) features such as position, aspect ratio, roundness, area/rectangle, rectangle ratio are used. In Wang and Liao (2002) are defined features like distance from the center of weld, mean radius, standard deviation of radius, circularity, compactness, major axis, width, length, elongation, heywood diameter. Similar features can be found in several other works like (Warren Liao, 2008; Liao, 2003). In Shafeek, Gadelmawla, Abdel-Shafy, and Elewa (2004) similar features like form factor, rectangularity factor, location and orientation are used.

The simple intensity features that are commonly used are the average and the standard deviation of the intensity values (see e.g., Wang & Liao, 2002). In Liao and Li (1998) the peak intensity value and the mean square error between each pixel and a Gaussian defect model are used.

Another class of features are the moment-based features which provide information about shape and intensity at the same time. For example in Nacereddine et al. (2006) the Hu moments are employed to model the defects. However, such features are very costly and are well known for their information redundancy.

The texture is another source of features, since it can provide very useful visual cues. For example in Mery and Berti (2003), the Haralick texture features and texture features based on the

Gabor functions are evaluated. Such types of features require much processing power and should be used with caution especially in large images.

It is clear that redundancy of information is expected when large number of features is used. In such cases increasing the feature number does not provide better performance and the system efficiency drops, especially for high resolution images such as the radiographic images. The problem has been identified in the literature. For example in Hernandez et al. (2006) a Self Organizing Map is used for dimensionality reduction. In Nacereddine et al. (2006) and Vilar et al. (2009) the simple PCA was used. In Liao (2003) the features are selected based on correlation. In Mery and Berti (2003) the Sequential Forward Selection method is used (Jain et al., 1997). In Warren Liao (2009) an ant colony optimization method is used.

We differ from the aforementioned literature because we use texture and geometrical features as well, allowing for better modeling of various defects. We also use a feature selection method, which gives an estimate about the expected performance allowing for a reasonable trade-off between performance and efficiency.

2.2. Classification

The classification is a processing step, which is decoupled from the feature definition and extraction. Any classifier in the related literature could be used for the classification of defects and the field would obviously benefit from more advanced classification methods. The main taxonomies here are the methods that treat the task as binary (defect, non-defect) and the ones that treat it as a multi-class problem.

Example of the binary problem is (Hernandez et al., 2006), which uses a fuzzy inference system represented as a Neural Network. In Mery and Berti (2003) the polynomial, Mahalanobis and nearest neighbour classifiers were used. In Lashkia (2001) a fuzzy algorithm method is used for detection.

An example of the multi-class classification is given in Liao (2003), where the authors propose a fuzzy k -nearest neighbour for six types of defects. Wang and Liao (2002) used a multi-layer perceptron Neural Network and a fuzzy expert system for the classification of welding defect types for six types of defects. Da Silva et al. (2005) used a two-layer neural network to identify six defect types. In Mirapeix, Garcia-Allende, Cobo, Conde, and Lpez-Higuera (2007) an artificial Neural Network is used to discriminate between three classes but only qualitative results are presented. In Vilar, Zapata, and Ruiz (2009) an artificial Neural Network is used to discriminate between five defects and false positives.

It is a common phenomenon that some defect classes are more frequent than others and that can favour the most populated classes during classification. This problem has been highlighted in several works. Solutions like random duplication (e.g., Da Silva et al., 2005) or other practices (Warren Liao, 2008) are shown to be of merit.

We enhance the previous classification approaches firstly by using well defined feature vector to solve the multi-class classification problem. We also consider the best practices in training to avoid misclassification of minority classes and we apply the resampling technique of 3-fold cross validation. We also compare the state of the art Support Vector Machine and Neural Network multi-class classifiers. Also in contrast to most works presented in the literature we included in our tests the *false alarms* class as part of the multi-class classification problem because it is not possible in a fully automated system as the one presented here to segment perfectly defects from non-defects, even if an initial binary classification step is assumed.

3. Defect detection

3.1. Local thresholding

Defect detection was done in two sequential steps: (1) Sauvola local thresholding and (2) a graph-based segmentation method. Sauvola is a popular local thresholding method that has been successfully used in several image processing applications, e.g., in Optical Characterization Recognition (OCR). It was used in Nacereddine, Zelmat, Belaifa, and Tridi (2005) for weld defect detection and it was shown that it performs better than the Otsu global thresholding and Niblack local thresholding method. The formula of the threshold T that is compared with the gray value of the pixel centered in a $N \times N$ window is:

$$T = M[1 + K(s/R - 1)] \quad (1)$$

where M and s are the mean and standard deviation in the window, R is the maximum standard deviation of gray level in all windows in the image and K is a parameter defined by the user. N should be sufficiently small to preserve the local details but it must be large enough to remove the noise. After local thresholding, a morphological filtering has to be used to remove a number of single dots in the resulted thresholded image and get a smoothed binary image. This image contains suspicious objects in "white", while other areas belong to the black background. Suspicious objects correspond either to defects or false alarms.

3.2. Segmentation

The next step is to use the regions extracted by preprocessing and to extract segments that correspond to the defects or false positives. Here, we use the graph-based segmentation method proposed in Felzenszwalb and Huttenlocher (2004). This method is able to capture perceptually distinct regions, even though their interior is characterized by large variability, by considering global image characteristics. This is a very desirable feature, since such cases are very common in weld defect images.

According to graph-based segmentation algorithms, an image is represented as an undirected graph $G = (V, E)$, where each pixel p_i has a corresponding vertex $v_i \in V$ and $(v_i, v_j) \in E$ are the edges corresponding to neighbouring vertices. The weight of the edges represents the dissimilarity measure, i.e., difference in the gray level function, between connected vertices. A segmentation S is a partition of V into components such that each component C corresponds to a connected component in a graph $G' = (V, E')$, where $E' \subseteq E$. In other words, any segmentation is induced by a subset of the edges in E .

The selected method follows the general concept of graph-based methods and measures the evidence of a boundary between two regions by computing (a) intensity differences across the boundary and (b) intensity differences between neighbouring pixels within each region. The internal difference $\text{Int}(C)$ of a component $C \in V$ is defined to be the largest weight in the minimum spanning tree of the component. The difference $\text{Dif}(C_1, C_2)$ between two components $C_1, C_2 \in V$ is defined to be the minimum weight edge connecting the two components.

The input to the algorithm is a graph $G = (V, E)$, with n vertices and m edges. The output is a segmentation of V into r components $S = (C_1, \dots, C_r)$. The steps of the algorithm are as follows:

- (1) Sort E into $\Pi = (o_1, \dots, o_m)$, by non-decreasing edge weight.
- (2) Start with a segmentation S_0 , where each vertex v_i is in its own component.
- (3) Repeat step 4 for $q = 1, \dots, m$.

- (4) Construct S_q given S_{q-1} as follows. Let v_i and v_j denote the vertices connected by the q th edge in the ordering, i.e., $o_q = (v_i, v_j)$. If v_i and v_j are in disjoint components of S_{q-1} and the weight $w(o_q)$ is small compared to the internal difference of both those components (this is defined by a a-priori set parameter k for the algorithm (Felzenszwalb & Huttenlocher, 2004)), then merge the two components otherwise do nothing.
- (5) Return S_m

It has to be noted here that the segmentation method could be applied directly to the whole radiographic image, instead of only the regions given by local thresholding. Experiments showed that such a preprocessing step improves the total results of detecting the suspicious segments.

4. Defect classification

The defect detection step in weld radiographs provides a set of objects that correspond to one of the following classes: non-defect (false alarm), worm holes, porosity, linear slag inclusion, gas pores, lack of fusion and cracks. In order to classify each of the obtained objects, a set of geometrical and texture-based features is extracted which is then used as input to a multi-class classifier.

4.1. Feature definition

The experienced users use both geometrical and intensity-based attributes to evaluate a segment in radiographic images. A system that would try to identify defects should exploit both information sources as well. This is the main observation that leads us to the definition of features that follows.

Initially a set of eight geometric features is defined, which have been used successfully in the literature (e.g., Liao, 2003; Da Silva et al., 2005). These are the relative position to the weld bead, the aspect ratio, the length/area ratio, the area/bounding rectangle ratio, the roundness, the rectangle ratio, the Heywood diameter and the relative angle to the weld bead. These features and their calculation formulas are given in Table 1.

Then, a set of 35 intensity-based descriptors are defined for each segment, many of them have also been used (see, e.g., Mery & Berti, 2003). These correspond to mean value and standard deviation of the gray level values of all pixels in the object and a set of 33 features obtained using the co-occurrence matrix approach for texture description.

Table 1
Geometric features.

Nr.	Name	Symbol/ formula	Explanation
G1	Position	$P = h/H$	H : width of weld bead h : distance of object from middle of weld bead
G2	Aspect ratio	L/e	L : the big axis e : the small axis
G3	Length/Area ratio	e/A	A : area of the object
G4	Area/bound. rect. ratio	A/A_r	A_r : the area of the minimum rectangle that includes the object
G5	Roundness	$\frac{p^2}{4\pi A}$	p : perimeter of the object
G6	Rectangle or 'box' ratio	$\frac{W}{H}$	W/H : the width/height of minimum rectangle that includes object
G7	Heywood diameter	d_H	the diameter of a circle having an equivalent area to that of the object
G8	Angle	θ	Angle of major object axis with line vertical to weld bead

The co-occurrence matrix \mathbf{P}_{kl} is defined as follows. The element $P_{kl}(i,j)$ of this matrix for a window is the number of times, divided by N_T , that gray-levels i and j occur in two pixels separated by that distance and direction given by the vector (k,l) or $(-k,-l)$, where N_T is the number of pixels pairs contributing to \mathbf{P}_{kl} .

Texture features correspond to eleven (11) measurements, i.e., angular second moment, contrast, correlation, sum of squares, inverse different moment, sum average, sum entropy, entropy, difference variance, difference entropy obtained using the co-occurrence matrix that was calculated for three interpixel distances ($d = 1, 2, 3$) (Haralick, Dinstein, & Shanmugam, 1973). These features and their calculation formulas are presented in Table 2. The features can be calculated along different directions, e.g., $(0,180)$, $(-90,90)$ degrees etc.

4.2. Feature selection

It is clear that the features that we mentioned in the previous section require a lot of computational power, especially the texture-based ones. There is a need for a trade-off between efficiency and effectiveness that becomes more clear for images of several tens of MBytes, which is typical for radiography. We need to use only those features that provide useful information and reject the rest, since we expect a great amount of redundancy in the information carried by the features. For this purpose we employed a feature selection method.

Feature selection refers to a method that selects a subset of original features based on an evaluation criterion. For a data set with N features, there exist 2^N candidate subsets. Even for a moderate N , the search space increases exponentially and very soon becomes prohibitive for exhaustive search.

To avoid the computationally intractable exhaustive feature selection we have used a Sequential Backward Selection (SBS) method, which is employed along with a classifier, as described in next subsection. It selects from an initial set of input variables those variables that are mostly related to the output and contain the causality to the output.

Table 2
Texture features.

Nr.	Name	Formula
T1	Angular 2nd moment	$\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [p(i,j)]^2$
T2	Contrast	$\sum_{n=0}^{N_x-1} n^2 \sum_{i=1}^{N_x} \sum_{j=1, i-j =n}^{N_y} p(i,j)$
T3	Correlation	$\frac{1}{\sigma_x \sigma_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (ij \cdot p(i,j) - \mu_x \mu_y)$
T4	Sum of squares	$\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (i-j)^2 \cdot p(i,j)$
T5	Inverse difference moment	$\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \frac{1}{1+(i-j)^2} \cdot p(i,j)$
T6	Sum average	$\sum_{i=2}^{2N_x} i \cdot p_{x+y}(i)$
T7	Sum variance	$-\sum_{i=2}^{2N_x} p_{x+y}(i) \log [p_{x+y}(i)]$
T8	Sum entropy	$-\sum_{i=2}^{2N_x} (i - T_7) p_{x+y}(i)$
T9	Entropy	$-\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} p(i,j) \log p(i,j)$
T10	Difference variance	$\text{var}(P_{x+y})$
T11	Difference entropy	$-\sum_{i=0}^{N_x-1} p_{x-y}(i) \log p_{x-y}(i)$

Note: $p(i,j) = P_{kl}(i,j)$, $p_x = \sum_{j=1}^{N_y} p(i,j)$, $p_y = \sum_{i=1}^{N_x} p(i,j)$, $p_{x+y}(k) = \sum_{j=1}^{N_y} \sum_{i=1}^{N_x} p(i,j)$, $i+j=k$

The SBS starts by constructing a multi-class classifier that uses all input variables. A popular method to train and test is the 3-Cross Validation (3-CV) resampling technique and thus the procedure of training and testing the ANN is repeated three times. Each time, the classifier is trained using 2/3 randomly chosen subjects of the available dataset and tested using the remaining subjects. Each of the trained classifiers is evaluated using the mean value of the classification accuracies (A_1, A_2, A_3) obtained in the corresponding 3-CV training and testing sets. The 3-CV technique outputs a fitness value (%):

$$F = (\bar{A}_1 + \bar{A}_2 + \bar{A}_3)/3 \quad (2)$$

for the initial classifiers that use all input variables. Next, the procedure subtracts one variable from the total number of variables and constructs the 3-CV classifiers that use the remaining variables as input. Each variable is subtracted from the total number of variables and the classifiers are trained with the remaining ones. The input variables set that yields the best average of mean accuracies $\bar{A}_1, \bar{A}_2, \bar{A}_3$ in the 3-CV sets is the one chosen and is assigned the fitness value F_{low} , similarly as when using all input variables. The procedure continues similarly and the classifier that uses M inputs is derived from the one that uses $M + 1$ inputs by subtracting the least informative variable by means of mean accuracy in the 3-CV training and testing sets and the set of M variables is assigned a fitness value. This process is repeated until only one variable remains. The best set of 3-CV classifiers and the corresponding input variables set are selected based on the values of fitness function. The variables that are included in the selected set of variables are the ones considered to be the most robust ones for the classification using the given classifier.

4.3. Classification

Given the feature vector, the next step concerns the classification. There are several multi-class classifiers that can be employed but the ones selected must be able to cope with non-linearity and eventually high dimensionality. Among the state of the art algorithms that satisfy these requirements are the Support Vector Machines (SVM) and the Artificial Neural Networks (ANN).

SVM is a promising classification method initially developed for a binary classification problem (Crammer & Singer, 2002). The SVM draws an optimal hyper-plane (determined by \mathbf{w}, b) in a high dimensional space that defines a boundary that maximizes the margin between data samples that belong to the two classes, so as it generalizes well in unknown data. The decision function for a feature vector \mathbf{X} :

$$f(\mathbf{X}) = \mathbf{w} \cdot \phi(\mathbf{X}) + b \quad (3)$$

where ϕ is a mapping of feature vectors to the high dimensional space. The kernel function of the SVM defines the mapping. We chose to use a newly developed multi-class SVM (Crammer & Singer, 2002) instead of using a binary SVM and adopting techniques for incorporating it in a multi-class problem, e.g., one versus all or all versus all. The optimization problem is solved using an efficient cutting plane algorithm that exploits sparseness and structural decomposition.

The aforementioned SVM is going to be compared with an Artificial Neural Networks (ANN), which is a feedforward ANN of one input layer fed with the set of input variables, one hidden layer of adjustable number of hidden neurons and one output layer of one neuron (Haykin, 1999). The output layer has seven neurons, the maximum activated neuron gives the winner class. The tansig and logsig are used as activation functions in the input and output layer, respectively, while the ANN can be trained using a method such as back-propagation (Haykin, 1999).

5. Experiments

The goals of our experiments are: (a) to verify the validity of our approach, (b) to investigate the effect of reduced feature set with the SBS method in efficiency and quality of results, and (c) to compare the results obtained by state of the art classifiers.

For our experiments a total of 24 radiographs of ship welds provided by Technic Control Co. (Poland) were used. The welds were sized 100 mm \times 400 mm and the radiographs featured a resolution equal to 50 $\mu\text{m}/\text{pixel}$ at depth of 16 bits. An experienced NDT consultant annotated the set of 24 images and an equal number of Re-

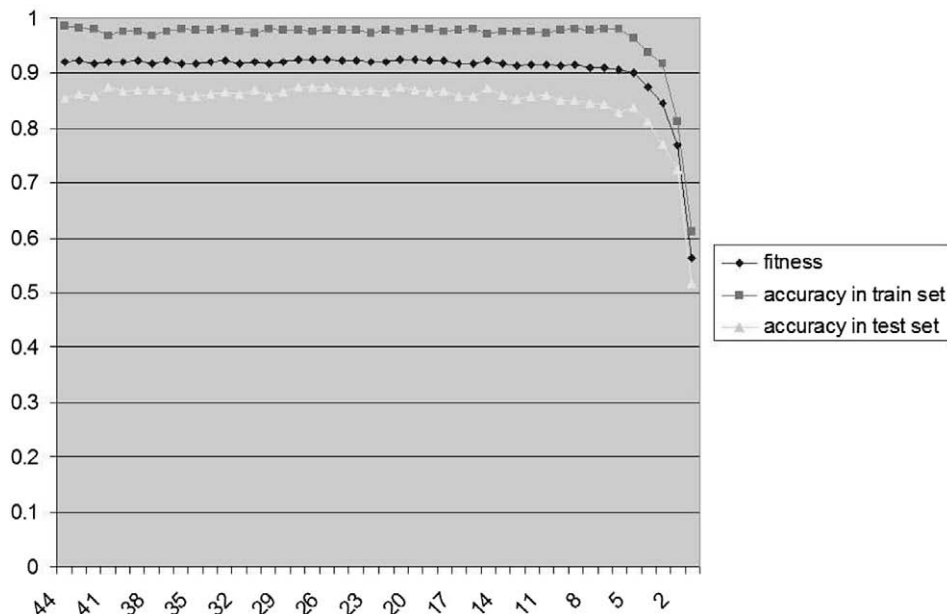


Fig. 3. The Sequential Backward Selection method for feature selection. The diagram depicts the fitness and the accuracy in training and testing set. The horizontal axis is the number of features.

gions of Interest (ROIs) were acquired which contained the frequently met defects of worm holes, linear slag inclusion, gas pores, lack of fusion, crack and porosity.

For preprocessing the Sauvola local thresholding method was applied using the parameters $N = 51$, and $K = 0.05$. After local thresholding, a morphological closing filter was applied to remove a

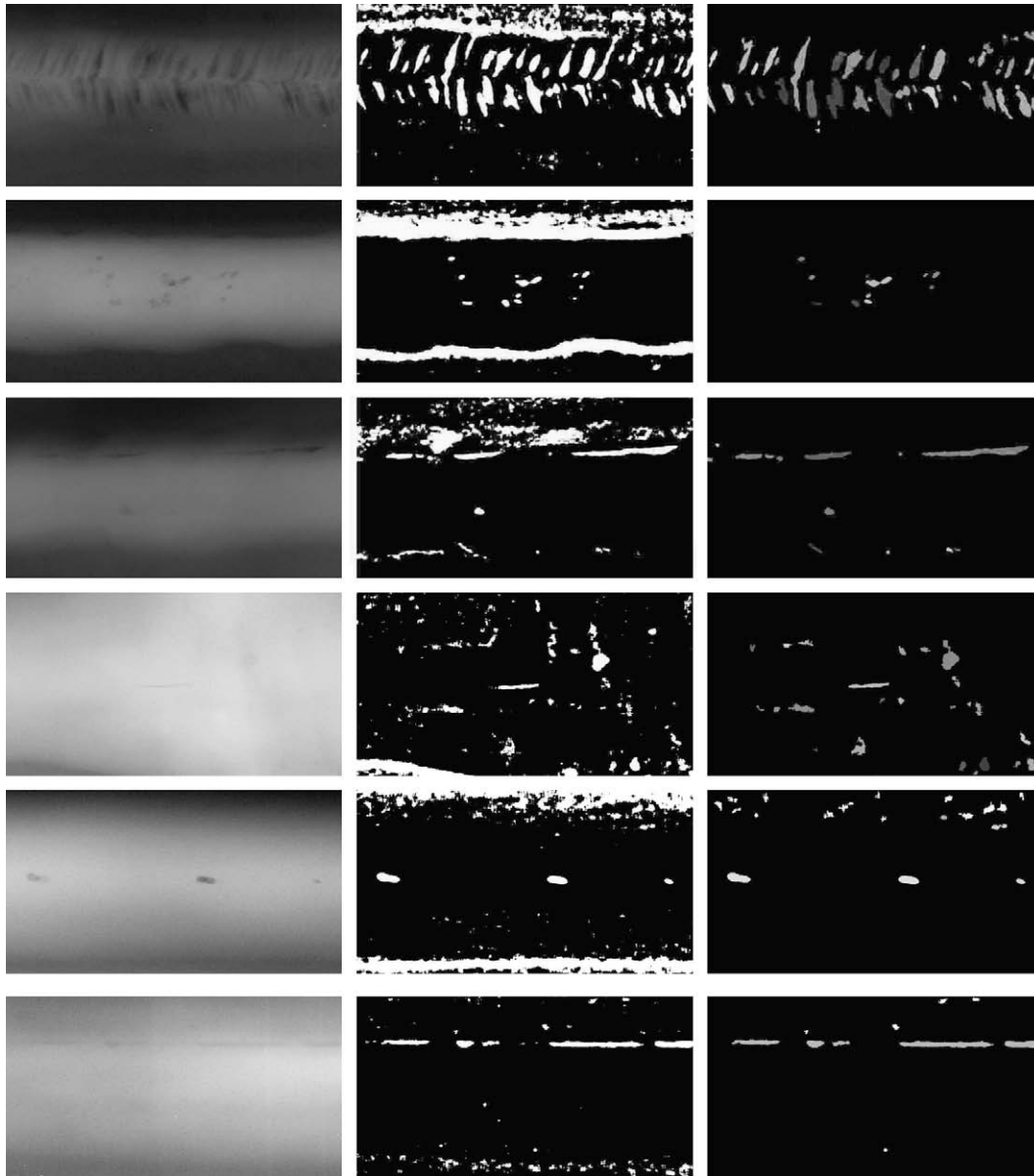


Fig. 4. Original, thresholded and segmented image for welds that contain: 1st row: worm holes (defects almost vertical to the weld), 2nd row: porosity (circle spots over the whole area), 3rd row: linear slag inclusion (linear defects at the top), 4th row: crack (linear in the middle), 5th row: gas pores (ovals in middle), and 6th row: lack of fusion (linear at the top).

Table 3

Mean value of sensitivity, specificity and accuracy of SVM and Neural Network in 3-fold cross validation training sets.

Metric	Classifier	Non-defect	Worm hole	Linear slag	Porosity	Gas pores	Lack of fusion	Crack
Sensitivity	SVM-43	85.17	94.82	95.45	93.69	100.00	74.71	88.73
	SVM-7	78.88	80.98	55.11	80.27	100.00	23.33	74.59
	ANN-43	99.55	99.91	99.91	99.74	99.83	99.65	99.65
	ANN-7	99.14	99.91	99.57	99.65	99.82	99.65	99.91
Specificity	SVM-43	98.00	100.00	99.13	98.72	99.29	97.55	96.29
	SVM-7	85.22	90.91	87.78	87.94	95.23	69.64	90.69
	ANN-43	94.32	100.00	100.00	96.88	100.00	98.40	100.00
	ANN-7	94.28	98.94	100.00	95.41	100.00	99.49	98.49

Overall accuracy: SVM-43:90.42, SVM-7:70.45, ANN-43:98.51, ANN-7:98.00.

number of artifacts in the resulted thresholded image and get a smoothed binary image. Parameters N , and K were manually tuned to the selected values after extensive experiments on the available image data. The object detection module performed well and in average 90% of the defects were detected. This is mainly due to the use of the local thresholding method that performs well even when there is no uniform background in the images. The graph-based segmentation returned each of the detected objects as a different segment colored randomly. We excluded from further processing the detected segments outside the weld bead and the ones with an area greater or smaller than a minimum (50 pixels) and maximum area (3500 pixels), correspondingly. Examples of the original images and the results of thresholding and segmentation are presented in Fig. 4.

Using the set of 11 ROIs, a total of 411 segments were extracted that corresponded to worm holes (85 cases), porosity (94 cases), linear slag inclusion (42 cases), gas pores (13 cases), lack of fusion (57 cases), crack (26 cases) and non-defects, i.e., false positives (94 cases).

The texture features were averaged along horizontal, vertical and diagonal directions. Given all features, the feature selection method performed as depicted in Fig. 3, for the case of the ANN. It becomes clear that the fitness function remains almost the same until we reach the last seven features, when a rapid drop is observed. There-

fore it is estimated to be safe to keep only these features without significant influence in our results. These features are namely the G1, G4, G6 (geometric) and T1 for distance = 1, T5 for distance = 1, T7 for distance = 2 and T10 for distance = 3. It becomes clear from this analysis that the fusion of both texture and geometric features gives better results than only one type of features. This supports our claim that the employment of features from both geometric and texture domain give better results than features from a single source. Similar results are expected with the same feature selection method using an SVM instead of an ANN.

For classification into our seven classes we compared the performance of the Support Vector Machines, the Neural Network and a k -NN classifier. The latter is still used as a reference method in the related literature, (e.g., Warren Liao, 2008, 2009). For the SVM the polynomial kernel gave the best results, although not significantly different by the ones given by arctan and Gaussian kernels. The neural network used one hidden layer with variable number of neurons, ranging from 4 to 20 and was trained using back-propagation with adaptive learning rate and momentum. For the k -NN various values for k were tested and for $k = 1$ we had the best results.

For training and testing the SVM and Neural Network classifiers we used the resampling technique of 3-fold cross validation, i.e., the classifier was trained and evaluated iteratively three times. To this end, three equal sized disjoint subsets were randomly allocated in the set of 383 cases, containing samples from each class. Each time, training was done using the two subsets (training set) and evaluation using the remaining subset (testing set).

We have trained the SVM and ANN classifiers using 43 and 7 features. To avoid favoring highly populated classes equal number

Table 4

Mean accuracy for all classifiers in 3-fold cross validation testing sets using 43 and 7 features.

	SVM-43	SVM-7	ANN-43	ANN-7	k -NN 43	k -NN 7
Accuracy	84.19	63.74	85.40	84.18	60.34	55.23

Table 5

Confusion matrix for SVM-43.

Classified/real	Non-defect	Worm hole	Linear slag	Porosity	Gas pores	Lack of fusion	Crack
Non-defect	0.78	0.00	0.00	0.02	0.00	0.09	0.12
Worm hole	0.00	0.95	0.00	0.00	0.00	0.00	0.00
Linear slag	0.06	0.02	0.93	0.02	0.00	0.00	0.00
Porosity	0.04	0.02	0.07	0.96	0.00	0.00	0.00
Gas pores	0.02	0.00	0.00	0.00	0.92	0.06	0.12
Lack of fusion	0.05	0.00	0.00	0.00	0.08	0.59	0.15
Crack	0.04	0.00	0.00	0.00	0.00	0.26	0.62

Table 6

Confusion matrix for SVM-7.

Classified/real	Non-defect	Worm hole	Linear slag	Porosity	Gas pores	Lack of fusion	Crack
Non-defect	0.77	0.00	0.00	0.01	0.00	0.20	0.27
Worm hole	0.05	0.72	0.26	0.11	0.00	0.00	0.00
Linear slag	0.00	0.00	0.39	0.01	0.00	0.00	0.00
Porosity	0.03	0.28	0.35	0.85	0.00	0.00	0.00
Gas pores	0.08	0.00	0.00	0.01	1.00	0.33	0.12
Lack of fusion	0.05	0.00	0.00	0.00	0.00	0.13	0.19
Crack	0.01	0.00	0.00	0.00	0.00	0.33	0.42

Table 7

Confusion matrix for ANN-43.

Classified/real	Non-defect	Worm hole	Linear slag	Porosity	Gas pores	Lack of fusion	Crack
Non-defect	0.77	0.00	0.00	0.02	0.00	0.11	0.15
Worm hole	0.00	0.95	0.00	0.02	0.00	0.00	0.00
Linear slag	0.06	0.01	0.89	0.02	0.00	0.00	0.00
Porosity	0.02	0.03	0.11	0.93	0.00	0.00	0.00
Gas pores	0.01	0.00	0.00	0.00	0.92	0.00	0.00
Lack of fusion	0.09	0.00	0.00	0.00	0.00	0.85	0.38
Crack	0.04	0.00	0.00	0.00	0.08	0.04	0.46

Table 8
Confusion matrix for ANN-7.

Classified/real	Non-defect	Worm hole	Linear slag	Porosity	Gas pores	Lack of fusion	Crack
Non-defect	0.76	0.00	0.00	0.02	0.08	0.13	0.15
Worm hole	0.01	0.97	0.04	0.03	0.00	0.00	0.00
Linear slag	0.03	0.00	0.80	0.03	0.00	0.00	0.00
Porosity	0.05	0.03	0.15	0.91	0.08	0.00	0.00
Gas pores	0.01	0.00	0.00	0.00	0.77	0.00	0.04
Lack of fusion	0.08	0.00	0.00	0.00	0.08	0.83	0.19
Crack	0.05	0.00	0.00	0.00	0.00	0.04	0.62

Table 9
Confusion matrix for *k*-NN-43.

Classified/real	Non-defect	Worm hole	Linear slag	Porosity	Gas pores	Lack of fusion	Crack
Non-defect	0.64	0.01	0.02	0.00	0.00	0.26	0.19
Worm hole	0.02	0.64	0.37	0.25	0.00	0.00	0.00
Linear slag	0.03	0.19	0.33	0.19	0.00	0.00	0.00
Porosity	0.03	0.16	0.28	0.56	0.00	0.04	0.00
Gas pores	0.02	0.00	0.00	0.00	0.69	0.00	0.00
Lack of fusion	0.20	0.00	0.00	0.00	0.31	0.69	0.04
Crack	0.06	0.00	0.00	0.00	0.00	0.02	0.77

Table 10
Confusion matrix for *k*-NN-7.

Classified/real	Non-defect	Worm hole	Linear slag	Porosity	Gas pores	Lack of fusion	Crack
Non-defect	0.61	0.00	0.04	0.02	0.15	0.48	0.27
Worm hole	0.00	0.74	0.17	0.15	0.00	0.00	0.00
Linear slag	0.04	0.14	0.37	0.17	0.00	0.00	0.00
Porosity	0.04	0.12	0.41	0.66	0.00	0.07	0.00
Gas pores	0.01	0.00	0.00	0.00	0.54	0.00	0.15
Lack of fusion	0.26	0.00	0.00	0.00	0.08	0.30	0.38
Crack	0.04	0.00	0.00	0.00	0.23	0.15	0.19

of samples from each class was used, by random oversampling the minority classes. In Table 3 we provide some metrics that indicate the training quality. These metrics are total accuracy (fraction of cases that are correctly classified), sensitivity per class (fraction of cases in a class that were correctly classified) and specificity per class (fraction of cases not belonging to a class that were correctly not classified to this class).

From Table 3 we infer that the training of the SVM when using reduced feature set seems not as good as the training of the ANN using the same features, which can affect its generalization ability. This can be expected since the features were selected using an ANN and the comparison is rather unfair but we present it for reference purposes, to compare performance on a common feature set. For better comparison we should apply the SBS feature selection using an SVM classifier but the extreme computational requirements render this approach practically infeasible (several weeks are required for the SVM in contrast to several hours for the ANN).

The overall accuracy on the cross validation test sets for the SVM, ANN and *k*-NN classifiers using full and reduced feature sets are given in Table 4. More details for each classifier performance in testing sets are given in Tables 5–10, where the confusion matrices are presented.

From the tabulated results and according to our research goals stated at the beginning of this section we can make several observations. Our method is promising and it becomes clear that the generalization ability of both Neural Network and SVM classifier is very high (both around 85% accuracy). Their results are comparable but we cannot say which is best because the difference is small and although we have performed many experiments to find

the optimal tuning we cannot claim that some further optimization of either method is not possible. Clearly the ANN is trained much faster than the multi-class SVM and for this reason it can be a preferable option. Regarding the *k*-NN classifier it is clear that, as expected, the results are significantly inferior in comparison to the other classifiers which perform much better in the weld classification feature space.

The difficult cases seem to be the cracks, the lack of fusion and the non-defects. The SVM and the ANN using full feature set recognised correctly 62% and 46% of the cracks, 59% and 85% of the lack of fusion, 78% and 77% of the non-defects correspondingly. From the confusion matrices we notice that many cracks are mistaken for lack of fusion and vice versa, which is reasonable considering their visual similarity in radiographs. The non-defects due to their intra class variability may be mistaken as instances of any class.

The results obtained in testing sets using the reduced feature set provide almost as high overall accuracy as the full feature set for the case of the ANN but were not as good for the case of the SVM mainly because the feature selection procedure favors the ANN, as mentioned earlier. Some variations in the sensitivity of specific classes are possible (e.g., cracks or gas pores), which reflect the changed structure of the feature space. The savings in feature computation were approximately 80% when we used the reduced set of seven features. The full set processing of a region of approximately 30 MB approximately four minutes using a Intel Dual Core 2 GHz PC with 4 GB memory. A good trade-off in production has to consider the cost of time and how important these defects are for the quality and safety of the produced components.

6. Conclusions

We have presented promising results for a novel system for multi-class defect detection and classification in weld radiographs using both geometric and texture features to capture the visual properties. For preprocessing we have used local thresholding followed by graph-based segmentation. The defects were classified using the state of the art multi-class SVM, and Neural Network classifiers. We have also investigated the effect of feature selection and showed that it is possible to reduce computation time significantly without seriously affecting overall accuracy. As feature definition and classification techniques evolve the proposed overall strategy is expected to provide better results as well.

In the future, we aim to investigate the method in many more images and eventually create a public dataset which will be able to be used for algorithm evaluation.

References

- Anuncia, S. M., & Saravanan, R. (2006). Non destructive testing using radiographic images – A survey. *Insight: Non-Destructive Testing and Condition Monitoring*, 48(10), 592–597.
- British Standard (1998). Weldings and allied processes-classification of geometric imperfections in metallic materials-part 1: Fusion welding bs en iso 6520-1.
- Carrasco, M. A., & Mery, D. (2004). Segmentation of welding defects using a robust algorithm. *Materials Evaluation*, 62(11), 1142–1147.
- Crammer, K., & Singer, Y. (2002). On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2, 265–292.
- Da Silva, R. R., Siqueira, M. H. S., De Souza, M. P. V., Rebello, J. M. A., & Caloba, L. P. (2005). Estimated accuracy of classification of defects detected in welded joints by radiographic tests. *NDT and E International*, 38(5), 335–343.
- Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2), 167–181.
- Haralick, R. M., Dinstein & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, 610–621.
- Hayes, C. (1997). ABC's of nondestructive weld examination. *Weld*, 76(5), 46–51.
- Haykin, S. (1999). *Neural Networks: A comprehensive foundation*. Prentice Hall.
- Hernandez, S., Saez, D., Mery, D., Da Silva, R., Calba, L. P., & Rebello, J. M. A. (2006). Statistical tools for weld defect evaluation in radiographic testing. In *World conference on NDT – WCNDT06*.
- Jain, A. (1997). Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2), 153–158.
- Lashkia, V. (2001). Defect detection in X-ray images using fuzzy reasoning. *Image and Vision Computing*, 19(5), 261–269.
- Liao, T. W. (2003). Classification of welding flaw types with fuzzy expert systems. *Expert Systems with Applications*, 25(1), 101–111.
- Liao, T. W., & Li, Y. (1998). An automated radiographic ndt system for weld inspection: Part ii – Flaw detection. *NDT and E International*, 31(3), 183–192.
- Mery, D., & Berti, M. A. (2003). Automatic detection of welding defects using texture features. *Insight: Non-Destructive Testing and Condition Monitoring*, 45(10), 676–681.
- Mirapeix, J., Garcia-Allende, P. B., Cobo, A., Conde, O. M., & Lspez-Higuera, J. M. (2007). Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks. *NDT and E International*, 40(4), 315–323.
- Naccredine, N., Tridi, M., Hamami, L., & Ziou, D. (2006). Statistical tools for weld defect evaluation in radiographic testing. In *Proceedings of 12th European conference on non-destructive testing – ECNDT 2006*.
- Naccredine, N., Zemat, M., Belaifa, S. S., & Tridi, M. (2005). Qweld defect detection in industrial radiography based digital image processing. *Transactions on Engineering Computing and Technology*, 2, 145–148.
- Nockemann, C., Heidt, H., & Thomsen, N. (1991). Reliability in ndt: Roc study of radiographic weld inspections. *NDT and E International*, 24(5), 235–245.
- Shafeek, H. I., Gadelmawla, E. S., Abdel-Shafy, A. A., & Elewa, I. M. (2004). Automatic inspection of gas pipeline welding defects using an expert vision system. *NDT and E International*, 37(4), 301–307.
- Sofia, M., & Redouane, D. (2002). Shapes recognition system applied to the non destructive testing. In *Proceedings of 8th European conference on non-destructive testing – ECNDT 2002*.
- Sood, S., Blakeley, B., & Rebuffel, V. (2006). Shapes recognition system applied to the non destructive testing. In *Proceedings of 12th European conference on non-destructive testing – ECNDT 2006*.
- Vilar, R., Zapata, J., & Ruiz, R. (2009). An automatic system of classification of weld defects in radiographic images. *NDT and E International*, 42(5), 467–476.
- Wang, G., & Liao, T. W. (2002). Automatic identification of different types of welding defects in radiographic images. *NDT and E International*, 35(8), 519–528.
- Warren Liao, T. (2008). Classification of weld flaws with imbalanced class data. *Expert Systems with Applications*, 35(3), 1041–1052.
- Warren Liao, T. (2009). Improving the accuracy of computer-aided radiographic weld inspection by feature selection. *NDT and E International*, 42(4), 229–239.