International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:2, 2007

Quantitative analysis of weld defect images in industrial radiography based invariant attributes

N. Nacereddine, M. Tridi, S. S. Belaïfa and M. Zelmat

***Abstract*—**For the characterization of the weld defect region in the radiographic image, looking for features which are invariant regarding the geometrical transformations (rotation, translation and scaling) proves to be necessary because the same defect can be seen from several angles according to the orientation and the distance from the welded framework to the radiation source. Thus, panoply of geometrical attributes satisfying the above conditions is proposed and which result from the calculation of the geometrical parameters (surface, perimeter, etc.) on the one hand and the calculation of the different order moments, on the other hand. Because the large range in values of the raw features and taking into account other considerations imposed by some classifiers, the scaling of these values to lie between 0 and 1 is indispensable. The principal component analysis technique is used in order to reduce the number of the attribute variables in the aim to give better performance to the further defect classification.

***Keywords*—**Geometric parameters, invariant attributes, principal component analysis, weld defect image.

I. Introduction

When an image was segmented into object and background, the stage coming after is to extract the objects from the background to compute the values of the most adapted attributes to differentiate the objects from each other. It is what we call the quantitative analysis of the images. Often, the aim of these measurements is to constitute a base of descriptors which will be exploited during classification.

In an ideal case, each object in a binarized image is represented like only one connected region. However, in practice, we often find on this image noise, spots and small hanging branches. Some methods of post-processing must be employed to eliminate all the useless components. We have used in [1] a post-processing which is based on morphological filtering. When each object is extracted, the attribute measurement procedure is then implemented.

The radiogram images of the welded joints often contain defects which we must identify and quantify in order to decide on their acceptability by referring to non destructive testing standards and codes [2]. After the segmentation of a radiographic image providing a description in term of regions (weld defect and background), the problem is then to interpret their contents. It is thus a question of determining effective attributes which permit to characterize these defect regions and to even recognize them like class elements easily identifiable. In industrial radiography, we can obtain radiograms on which weld defects, if they exist, can have various sizes and orientations. For an example, a crack is identified as crack whatever its size and its orientation may be, and an inclusion is recognized as being an inclusion in spite of its position and its dimension [3].

A major problem in the recognition of such defects is that these defects can be viewed from several angles and this, according to the orientation and the distance of the irradiated welded joint in regard to the radiation source.

Therefore, with an aim to compare a given defect image with another which is stored in a data base, the methods which are not based on invariants must search in a multidimensional parameter space, i.e. to carry out several versions obtained by transformations, applied to the model image to see whether a version coincides with the original image. The purpose of the invariant geometrical attributes is thus to facilitate research, by making a correspondence between the observed defect images and the stored or referenced images from the already computed invariants.

To characterize a given weld defect represented by its boundary or its region, the simplest attributes which be computed are the area and the perimeter [4]. The latter cannot be used because of their sensitivity to geometric transformations. For this reason, we will employ geometric characteristics invariant regardless geometric transformations of translation, rotation and scaling. Panoply of attributes satisfying the above conditions will be proposed in this paper. These geometric invariant attributes will follow from the calculation of geometric parameters (area, perimeter, etc.) on the one hand and the calculation of the various order moments, on the other hand. They will be implemented on binarized images [1] issued from real radiographic films of welded joints. After, we will show the invariance accuracy of the different proposed attributes.

In order to reduce the computational time required for the classification stage it is necessary to select attributes; thus the classifier only works with non-correlated attributes that provide defect detection information. There are a variety of methods for evaluating the performance of the computed attributes.

The main idea behind the principal component analysis (PCA) is to represent multidimensional data with fewer number of variables retaining main features of the data. It is inevitable that by reducing dimensionality some features of the data will be lost. It is hoped that these lost features are comparable with the “noise” and they do not tell much about underlying population [5].

The method PCA tries to project multidimensional data to a lower dimensional space retaining as much as possible variability of the data.

This technique is widely used in many areas of applied statistics. It is natural since interpretation and visualisation in a fewer dimensional space is easier than in many dimensional space. Especially if we can reduce dimensionality to two or three then we can use various plots and try to find structure in the data. PCA is one of the techniques such factorial analysis, used for dimensional reductions.

For this purpose, in this work, the principal component analysis technique will be used to reduce the number of the attribute variables. In other words, by this technique, the initial data can be replaced by new data in which the same observations appear, but described by variables in smaller number.

II. Morphological Attributes

After each connected component which represents an object of interest (weld defect) is isolated, its geometric parameters : Area (*A*), perimeter (*P*) [6]., centre of gravity $G(\bar{x},\bar{y})$, angle of orientation (α), principal axes of inertia, width (*W*) and length (*L*) of the minimal surrounding rectangle, maximal diameter (*Dmax*), radius of maximal inscribed circle (*Rmax*), semi major and semi-minor axes (*a,b*) of the image ellipse [8] (see Fig. 1) are computed.

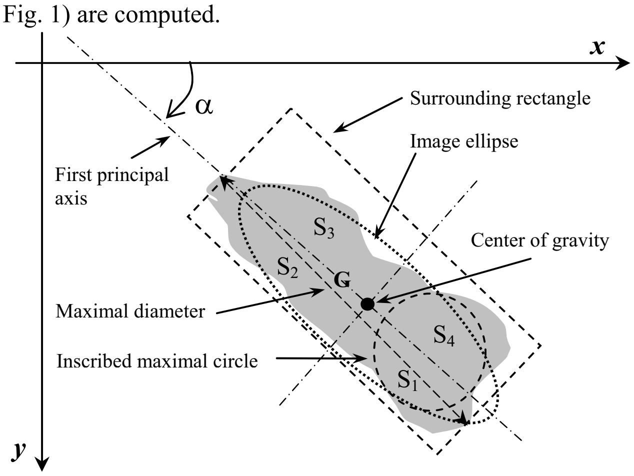

Fig. 1 Illustration of the geometric parameters

III. Geometric Invariant Attributes

The geometric attributes which we will define below are invariants regardless geometric transformations of translation, rotation and scaling.

Compactness	$Comp = 4\pi A / P^2$
Elongation	$Elong = L / W$
Rectangularity	$Rct = A/(L.W)$
Anisometry	$Ani = a / b$
Symmetry	$Sym = SymH \times SymV$ where *SymV* and *SymH* are given by: *if (S4+S3)<(S1+S2);SymV=(S4+S3)/(S1+S2);* *else SymV=(S1+S2)/(S4+S3)* *if (S2+S3)<(S1+S4); SymH=(S2+S3)/(S1+S4);* *else SymH=(S1+S4 (S2+S3)*
Lengthening index	$I_a = \pi \times D^2_{\max} / 4A$
Deviation index to inscribed circle	$I_r = 1 - \pi R^2_{\max} / A$
Invariant moments Φ1,Φ2	$\Phi_1 = \eta_{20} + \eta_{02}$ $\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$ where $\begin{bmatrix} \mu_{2,0} = \sum^n (x_i - \bar{x})^2 & -\mu_{1,1} = \sum^n (x_i - \bar{x})(y_i - \bar{y}) \\ -\mu_{1,1} = \sum^n (x_i - \bar{x})(y_i - \bar{y}) & \mu_{0,2} = \sum^n (y_i - \bar{y})^2 \end{bmatrix}$ is the covariance matrix of the object, and $\eta_{pq} = \mu_{pq} / \mu_{00}^{1+(p+q)/2}$ $p+q = 2,3,...$ the normalized central moments [9].

IV. Results and Discussion

A. Graphic illustration of geometric parameters on weld defect images

We present in Fig. 2 some radiographic film images and their associated weld defects [1] useful in the computation of the invariant geometric attributes and the construction of training and testing sets in the defect classification step.

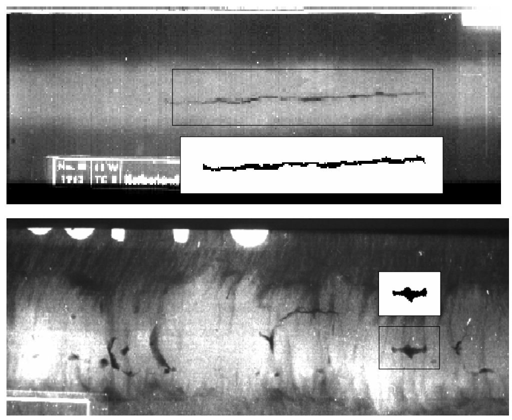

Fig. 2 Some radiographic films and their associated weld defects used in the attribute measurement

B. Relationship between the proposed invariant attributes and weld defect types

Compactness (Comp): Its value is included in [0,1]. It has little values for sharp defects (crack, lack of fusion) and it has values near to 1 for spherical defects (porosity, tungsten inclusion, etc.).

Elongation (Elong): It describes the occupied area in the bounding box of defect. Big values of this attribute characterize longitudinal defects (crack, lack of fusion, lack of penetration, elongated porosity, undercut, etc.).

Rectangularity (Rct): Its value is included in [0,1]. It is equal to 1 for a rectangle. It characterizes rectangular defects (lack of penetration).

Anisometry (Ani): It depends on the direction of principal axis of defect. Its value is proportional to defect lengthening.

Symmetry (Sym): Its value is included in [0,1]. The value 1 describes a perfectly symmetrical shape. The asymmetrical aspect of defect (slag inclusion, warm holes, etc.) can be related by little values of this attribute.

Lengthening index (Ia): Big values of this indicia put in obviousness fine and rectilinear cracks.

Deviation index to inscribed circle (Ir): The indicia value is maximal (near to 1) for lengthened defects and minimal (near to 0) for round defects.

Invariant moments (Φ_1,Φ_2): They gives measures in relation with the pixel spreading in comparison with the centre of mass.

C. Geometric invariance of the proposed attributes

In order to show the invariance performance of the proposed attributes, we have used a set of test images representing a binary image of weld defect (see Fig. 3). Usual geometric transformations are applied on this image: rotation (10° and 20°), reduction by scale change (60% and 80%), their combination (rotation of 15° and reduction of 70%) and mirror effect). Invariant attributes are computed and compared between those of original image and those of its transforms. The logarithm of Φ_1 and Φ_2 are taken to reduce the dynamic range.

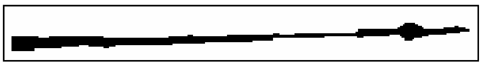

Fig. 3 Original binary image of a weld

As shown in histograms (see Fig. 4), the results for the geometric transformed images are in reasonable agreement with the invariants computed for the original image. The major cause of error can be attributed to the digital nature of the data. This invariance property is very important in the case of our work because we can obtain, in radiographic testing of weld joints, discontinuities having various sizes and orientations.

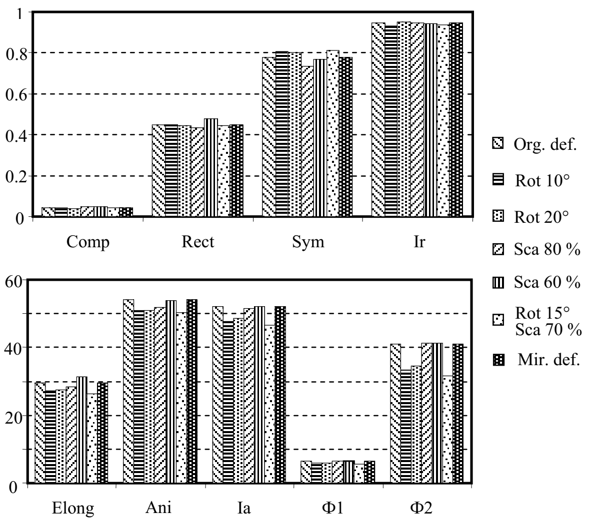

Fig. 4 Attribute values of Fig. 3 and its geometric transforms

D. Normalization of the attribute values

Because the different raw geometric attributes have values ranging from the order of 0 to 200, the features were rescaled to lie between 0 and 1, to avoid the effects of the larger features "swamping" those of the smaller features and possible numerical errors caused by a large range in values. This was done for each data using the maximum and minimum values for each attribute as observed in the data set, i.e.

$$\Gamma_i = \frac{P_i - \min(P_i)}{\max(P_i) - \min(P_i)}; i = 1..9 \tag{1}$$

with P_1 = Compactness, P_2 = Elongation, ... , $P_9 = \Phi_2$ and Γ_1 = rescaled compactness, ..., Γ_9 = rescaled Φ_2

The choice of *max* and *min* values to use for scaling data is difficult. The most obvious choices are the maximum and minimum values observed over the entire data set. However, in this experiment we are attempting to train a classifier on one data set and test it on a completely unseen test set. Therefore, it should be made sure that the values max and min represent really the extreme cases of the attribute in question, and that they are related to a physical significance of the shape of the defects to characterize.

E. Dimensionality reduction of the attribute vector based principal component analysis (PCA)

With an aim of reducing the number of variables representing the attributes, the first result interesting to analyze is the correlation matrix of the attribute variables (see Table I). However, it would be hazardous to use only the table of the correlations to eliminate the variables which presents a great correlation. It can generate a loss of information. This is why, panoply of the variable reduction methods proposed in the literature like as principal component analysis (PCA), factorial analysis, etc. Let us add the fact that the number of measurement for each variable plays a significant role in the correlation calculation between the different attribute variables.

In our study, we investigate the use of PCA relevant features from the morphologic feature. PCA is a statistical tool, which is useful to extract dominant features (principal components) from the set multivariate data. They explain the maximum amount of variance possible by linear transforms by projecting the data into ortho-normal sub-spaces. In our case, PCA will enable us to reduce the dimension of the feature vector and the extracted features should contain the most relevant information. To obtain eigenvectors, the database is formed into a column vector, Γ_n, whose length N is depending on the number of individuals used (weld defect). For M features, we will have an array matrix Γ with the size of $M \times N$. Therefore, we have

$$\Gamma = [\Gamma_1, \Gamma_2, \cdots, \Gamma_M] \tag{2}$$

The mean of the column vector Ψ is defined by:

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n \tag{3}$$

The subtracted training set is represented as matrix:

$$\Phi = [\Phi_1, \Phi_2, \cdots, \Phi_M] \text{ with } \Phi_i = \Gamma_i - \Psi \tag{4}$$

The covariance matrix is calculated using

$$C = \Phi\Phi^T \tag{5}$$

and the correlation matrix is then deduced by

$$R = (diag(C)^{1/2})C(diag(C)^{1/2}) \tag{6}$$

The eigenvector of matrix C as $\vec{v}_i$ with corresponding eigenvalues can be computed by

$$C\vec{v}_i = \lambda_i \vec{v}_i \quad (1 \le I \le L \le M) \tag{7}$$

Any weld defect can be identified as a linear combination of the eigenvectors. The principal components for any weld defect are defined by:

$$P = (\Gamma - \Psi)[\vec{v}_1^T, \vec{v}_2^T, \cdots, \vec{v}_L^T] \tag{8}$$

The matrix P with the size of $N \times L$ represents the database into the axis corresponding to the eigenvector. The values of this matrix are the new features that can be used for classification and recognition purposes.

By examining the initial database eigenvalues (see Fig. 5), we remark that the four first initial components gives more that 97 % of information on entire observations. For this purpose, the first matrix representing the feature variables (invariant attributes)/the individuals (weld defects) was transformed in another matrix where the data is projected in orthonormal sub-spaces with four principal components (see Table II). It is pointed out that these components are variables without physical meaning and are not directly observable.

TABLE I : CORRELATION MATRIX OF ATTRIBUTES VARIABLES

	Cmp	Elg	Rct	Anis	Sym	Ia	Ir	Φ_1	Φ_2
Cmp	1.00								
Elg	-0.66	1.00							
Rct	0.53	-0.28	1.00						
Anis	-0.64	0.98	-0.41	1.00					
Sym	0.00	0.18	0.34	0.14	1.00				
Ia	-0.64	0.89	-0.34	0.89	0.04	1.00			
Ir	-0.95	0.62	-0.45	0.59	0.06	0.59	1.00		
Φ_1	-0.62	0.90	-0.58	0.94	-0.02	0.85	0.56	1.00	
Φ_2	-0.43	0.82	-0.53	0.87	-0.05	0.74	0.38	0.94	1.00

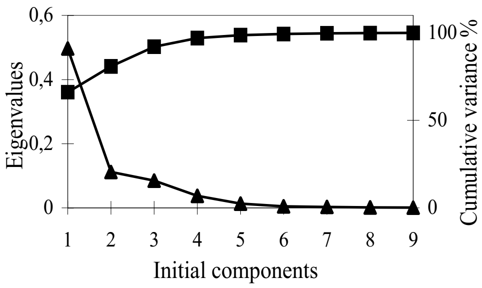

Fig. 5 Initial eigenvalues and cumulative variance percentage

TABLE II
DATABASE INTO THE FOURTH PRINCIPAL AXES

	Pr.Cp1	*Pr.Cp 2*	*Pr.Cp 3*	*Pr.Cp 4*
D 1	0.16	-0.36	-0.61	-0.21
D 2	0.07	-0.23	-0.82	-0.54
D 3	0.24	-0.30	-0.68	-0.32
D 4	0.04	-0.16	-0.78	0.05
D 5	0.84	-0.40	-0.13	-0.13
D 6	0.89	-0.33	-0.11	-0.45
D 7	0.67	-0.62	0.05	-0.06
D 8	0.27	-0.26	-0.74	-0.11
D 9	0.06	-0.25	-0.80	-0.03
D 10	-0.06	-0.15	-0.98	0.29
D 11	0.47	-0.39	-0.41	0.08
D 12	-0.00	-0.26	-0.81	0.19
D 13	-0.31	-0.10	-1.04	0.02
D 14	-0.03	-0.23	-0.82	-0.11
D 15	-0.99	0.40	-0.40	-0.07
D 16	-0.79	0.23	-0.66	-0.22
D 17	-1.05	0.50	-0.16	0.11
D 18	-1.01	0.53	-0.30	-0.13
D 19	-0.94	0.34	-0.47	-0.09
D 20	-0.40	-0.02	-0.66	-0.16
D 21	-0.40	-0.03	-0.71	-0.22
D 22	-0.39	-0.15	-0.39	0.06
D 23	-0.28	-0.23	-0.19	0.24
D 24	-0.18	-0.26	-0.35	0.12
D 25	0.52	-0.09	-0.58	0.04

V. CONCLUSION

Concerning the quantitative analysis of the weld defect images thus extracted, the major problem remains how to build a set of attributes which characterize the most accurately possible these defect regions, while taking in account the specificities of the defects that they represent, the subjectivity and the risks of their interpretation. This is why, it may be that only one attribute plays a decisive role in the discrimination of two defect classes, really distinct, either by the cause of their occurrence or by the severity of the codes and standards in their interpretation.

We have proposed in this work, a set of attributes which are based on the geometric characteristics of the extracted weld defect regions. We have showed the invariance of these attributes in relation to usual geometric transformations. The attribute values thus calculated were normalized, in order to adapt them for the classification problem. The variable reduction of these attributes must be done judiciously, by taking into account the number of individuals (defects) and the detailed analysis of the correlations between the various variables. The principal component analysis (PCA) technique permits to reduce the raw attribute vector to give a vector with four (04) non correlated components which are more suitable to use in the classification stage.

The objective in this work remains then, the research of the attributes which characterize the weld defect images according to two criteria:

- Maximal discrimination : Different defect shapes give different values of a given attribute.
- Minimal redundancy : An attribute don't vary in the same way that another for a given defect shape.

In further work, we will try to classify the weld defects in some classes representing the principal shape class defects usually met in practice. Thus, the four principal components values will be introduced as input data to supervised or unsupervised classifiers.

REFERENCES

[1] N. Nacereddine, "Automated method implementation for detection and classification of weld defects in industrial radiography," M.S. thesis, Dept. Automation and Signal Processing. Boumerdes Univ., Boumerdes, Algeria, 2004.

[2] "Guide for the Non-destructive Inspection of Welds". ANSI, American Welding Society, B1,10-86, USA, 1987.

[3] N. Nacereddine, R. Drai and A. Benchaala, "Weld defect extraction and identification in radiograms based neural networks," in Proc. IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, Crete, Greece, June 2002, pp 38-43.

[4] M. Mjahed, "Reconnaissance de formes planaires d'une scène d'image à partir d'une analyse morphologique par la technique des réseaux de neurones," TELECOM'97, Fès, Maroc, 1997.

[5] W.J. Krzanowski and F.H.C. Marriout FHC. Multivariate analysis. Vol 1. Kendall's library of statistics 1994.

[6] A.K. Jain, R.P.W. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 22, n° 1, January 2000.

[7] L. Dorst and A.W.M. Smeulders, "Length Estimators for Digitized Contours," Computer Vision, Graphics and Image Processing, vol. 40, pp.311–333, 1987.

[8] J. G. Postaire, De l'image à la décision. Edition Dunod. 1987.

[9] M.K. Hu, "Visual Pattern Recognition by Moments Invariants," IRE Trans. Info. Theory, vol. IT-8, 1962.

[10] A.K. Jain. *Fundamentals of digital image processing*. Prentice-Hall 1989.