PCA and Clustering based Weld Flaw Detection from Radiographic Weld Images

Dr.V. Vaithiyanathan[1], Anishin Raj M.M[2], Dr.B. Venkatraman[3]
[1]Assoc Dean Research, School of Computing, SASTRA University, Thanjavur-613401, India
[2]Asst Prof., Department of ICT, SASTRA University, Thanjavur-613401, India
[3]Assoc Director, Radiological Safety and Environmental Group, IGCAR, Kalpakkam-603102, India
[1]vvn@it.sastra.edu, [2]anishinraj@gmail.com, [3]bvenkat@igcar.gov.in

***Abstract*—Pattern recognition of weld defects can be done by partitioning the data points into clusters. The weld defect arises due to various types of adverse conditions and carelessness at the time of welding which leads to major disasters at the later stage. This paper proposes a method for detection of Burn Through, Slag Inclusion and Lack of Penetration weld defects in radiographic images which follow a specific pattern for each type. The image is preprocessed using Histogram equalization and segmented using Region growing methodology. The feature dimension reduction is performed using PCA, which is an unsupervised dimension reduction methodology which chooses the dimensions with the largest variance. K-Means clustering is used for clustering a cohesive group from a set of patterns after performing dimensionality reduction using Principal Component Analysis. The result of classification using this technique is providing a very accurate classification of weld defect.**

Keywords-PCA, K-Means Clustering, Weld Defect, Radiography

I. INTRODUCTION

This paper proposes a method for identifying weld flaws in radiographic images using PCA and K- Means clustering algorithm. This work mainly focuses on weld defects such as Lack of Penetration, Slag Inclusion and Burn Through. This type of weld defect arises due to lack of expertise of welding employ and also due to the presence of non metallic elements present in the atmosphere at the time of welding [1][2]. The radiographic images which are basically dark in colour are pre-processed using histogram equalization technique for the enhancement of the image. The segmentation of radiographic weld defective region is performed using Region growing methodology. Feature selection using PCA reduces the data size by retrieving the most invariant predominant features when the size of a data is too large for an algorithm to compute. These features act as an identity to an image and it is enough to process only these features and not the entire image for pattern identification. This paper uses K-Means clustering for identifying the weld defect in the radiographic image after performing the dimensionality reduction using PCA. The results of implementation clearly show that the weld flaws are being clustered into three different clusters with high accuracy in fewer numbers of iterations.

II. PRE-PROCESSING OF RADIOGRAPHIC IMAGE

This paper utilizes Histogram equalization for image enhancement [3][4]. The Histogram of any given image is the representation of any digital image as a discrete function of the number of pixels and the intensity values. Histogram is very useful in understanding the distribution of intensities of the image. If the histogram of a digital image is concentrated around a particular range of intensity values, then the corresponding image will be unclear. The histogram of a dark image will be clustered towards the low intensity values and vice versa. In this paper Histogram equalization is applied only if the image is low contrast image. To have a very clear, high contrast and a quality image the histogram should be distributed around the entire spectrum and the process of creation of distributed histogram is called histogram equalization. Histogram equalization can be viewed as the transformation which maps every input intensity to an output intensity value [5].

The histogram normalization is carried out using the equation mentioned below, where the intensity levels are discrete, where the total number of pixels is MN and the total number of possible intensity values is represented by L. Thus the discrete form of the transformation can be expressed as below.

$$t_k = f(i_k) = \frac{L-1}{MN}\sum_{p=0}^{k}\eta_p \quad (1)$$

Where i_k is the input intensity, t_k the transformed intensity and η the number of pixels having intensity value i_k.

III. REGION GROWING SEGMENTATION

Region growing is an algorithm that groups pixels or sub regions into a complete large region based on some specific criteria for development. The process begins with defining a set of seed points and allowing the regions to grow from these pixels by appending neighboring points to these seed points that satisfy the membership criteria. The seed points needs to be generated where the membership criteria may not have been clearly defined [3]. In such cases the image is grouped into clusters having similar properties and the pixel in each cluster having properties close to the centroid of the cluster is assigned as a seed. Better results are obtained by taking into consideration both the descriptors and the connectivity. The connectivity is an important issue because when the intensities values have a random arrangement the segmentation would result in a nonsensical image. Another major criterion to take into consideration is the cessation. The process must cease when all pixels obeying the membership rule are grouped to their respective seed points [4] [6]. The algorithm is mentioned below:

Step1: For every pixel in the seed array *s(x, y)*, reduce all the 8-connected pixels to a single pixel and label such points as 1 and name all other pixels as 0.

Step 2: Generate an image from the input image that satisfy the predicate q, such that $f_q(x,y)=1$, at the co-ordinate (x,y).

Step 3: Generate an image by appending all the points in f_q having value one is appended to their corresponding seed points in s.

Step 4: Distinct region label should be assigned for each connected component in the newly created image.

These regions in the resultant image obtained after applying Region growing method will be segmented properly.

IV. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis is one of the key techniques used in image processing [7] [8], which is mainly used for representing data in a lower dimensional plane. The reduced data will be oriented in the path of highest variance without any loss of data. PCA is the most important tool for dimensionality reduction where the covariance of the data is computed followed by finding the Eigen values and Eigen vectors of the covariance matrix [9][10][11].

An ordered orthogonal basis can be created by ordering the Eigen vectors in the downward order of Eigen values thereby projecting the data set in the orientation of highest significant energy.

V. K-MEANS CLUSTERING

K-Means clustering comes under the category of Partitional clustering where the number of clusters is being specified in advance before computation. A hierarchical dendogram will be generated in top down fashion which is more divisive than bottom up hierarchies as in the case of agglomerative clustering. K-Means clustering is having similarities with Forgy's clustering but K-Means clustering uses only two passes rather than iterative nature of Forgy's clustering[7][12][13].The K-Means Clustering algorithm is as follows:

Step 1: Randomly choose K out of given N patterns and assign them as initial cluster centres, This selection could also be done by picking first K out of given N patterns as selection of the K clusters are essential in partition of data and output may vary depending on it.

Step 2: Now allocate the remaining (N-K) patterns to one of the K centre such that the each allocated pattern is closest to its respective K cluster.

Step 3: Compute the centres according to the most recent alignment of the patterns.

Step 4: Repeat the allocation process of the N patterns according to its closest centres.

Step 5: If no further change in alignment of patterns to clusters is observed during two successive iterations then cease the process or else continue the process from Step 3.

VI. RESULTS AND DISCUSSION

The weld radiographic image of Burn Through is shown in Figure1 and its corresponding Region growing based segmented image is show in Figure2.The radiographic images containing weld flaws adopted in this paper for the experiments include images developed at IGCAR, simulated images and images from NDT resource center organization web site. Fifteen simulated images from each class of weld defect are being used in this experiment. The experimental results show that the weld flaws are identified using K- Means clustering after dimensionality reduction using Principal Component Analysis. The feature reduction using PCA is depicted in the Table1 where the image is being reduced to a lower dimension for computational efficiency without sacrificing the data content. The result of K- Means clustering is depicted in the Figure 3 where the green, red and blue colored dots show Burn Through, Lack of Penetration and Slag Inclusion respectively.

TABLE 1
Dimension reduction using PCA for various weld defect radiographic images

Dimension reduction	**Cluster 1**		**Cluster 2**		**Cluster 3**	
PCA	**Burn Through**		**Lack of Penetration**		**Slag Inclusion**	
μ1	1.9210	1.8688	0.1254	0.0651	1.4058	0.3039
μ2	0.3601	0.3663	0.0667	0.0000	1.3547	0.2762
μ3	0.2557	0.2773	0.1238	0.0000	1.2861	0.2739
μ4	0.1151	0.0693	0.1238	0.0000	1.3784	0.3124
μ5	0.0674	0.0581	0.1238	0.0000	1.3508	0.3241
μ6	0.0520	0.0460	0.1238	0.0000	1.2258	0.3084
μ7	0.0000	0.0000	0.1238	0.0000	1.0985	0.3459
μ8	0.0000	0.0000	0.1238	0.0000	1.0151	0.3546
μ9	0.0000	0.0000	0.0000	0.0000	1.2794	0.3550
μ10	0.0000	0.0000	0.2143	0.1286	1.2700	0.3524
μ11	0.0000	0.0000	0.2516	0.1960	1.2195	0.3787
μ12	0.0000	0.0000	0.0000	0.0000	1.3273	0.3592
μ13	0.0000	0.0000	0.2507	0.1302	1.3534	0.3107
μ14	0.0000	0.0000	0.1714	0.001	1.1965	0.2777
μ15	0.0000	0.0000	0.2644	0.0309	1.0021	0.3082

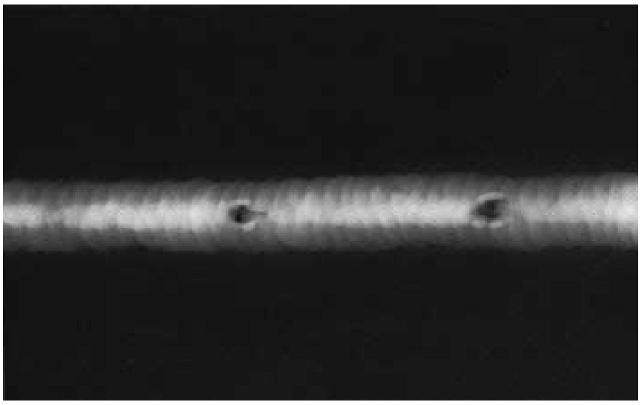

Fig 1.Weld Radiographic image of Burn Through

Fig 2. Resultant Segmented image of Figure1 using Region growing

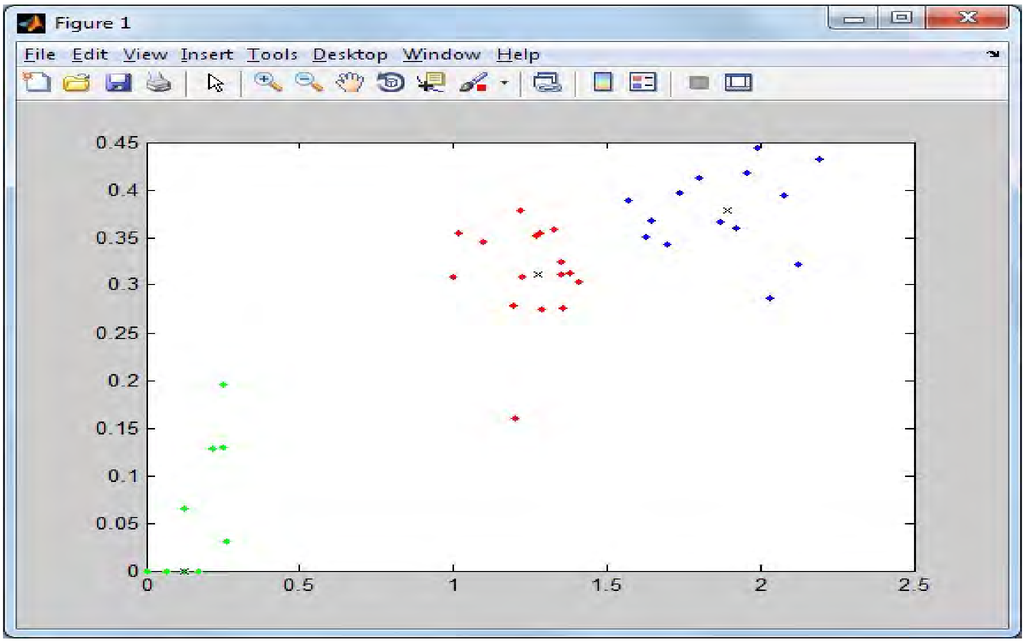

Fig 3.Result of K- Means Clustering (Green-Burn Through, Red- Lack of Penetration, Blue- Slag Inclusion)

VII. CONCLUSION

This paper proposes a method for identification of weld defect through K- Means clustering. The image is made more interpretable by applying histogram equalization which makes the intensities spread out to the entire range of intensity values thereby making the histogram of the resultant image as flat as possible. The radiographic weld images is being segmented using Region growing approach which makes use of a seed point which is allowed to grow by assimilating as many neighboring pixels as possible which satisfies a criteria for membership. The PCA generates a set of orthogonal axis of projection known as Principal components in the descending order of the variance of the input data set. The images having similar Eigen faces have similar properties or features. These values are clustered using K-Means clustering which provides a good accurate result of recognition of weld defect.

REFERENCES

[1] D.-M. Li and Y.-M. Li ,"Detection of welding flaws from radiographic images with fuzzy clustering methods", *Fuzzy Sets and Systems* 108, pp. 145-158, 1999.

[2] D.D Wang and Jinwu Xu, "Fault Detection based on evolving LVQ Neural network".,*Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (1),* Beijing, China, pp.255-260, 1996.

[3] R.C.Gonzalez and R. E.Woods , *Digital Image* Processing.,Prentice-Hall of India Private Limited, New Delhi,2005.

[4] S.Jayaraman, S. Esakkiarajan and T Veerakumar, *Digital Image Processing*,Tata McGraw Hill Education Private Limited. New Delhi.,2009.

[5] J. Alex Stark ,"Adaptive Image Contrast Enhancement Using Generalizations of Histogram Equalization",.*IEEE Transactions on Image Processing*, Vol. 9, No. 5, pp.889-896,2000.

[6] Yian-Leng Chang and Xiaobo li , "Adaptive image region-growing",*IEEE Transactions on Image Processing,* Volume: 3, Issue: 6, pp. 868 – 872,1994.

[7] V. Susheela Devi and M Narasimha Murty ,*Pattern Recognition: An Introduction*,Universities Press,2011.

[8] Xingfu Zhang and Xiangmin Ren , " Two Dimensional Principal Component Analysis Based Independent Component Analysis for Face Recognition", *Proceedings of IEEE Conference on Multimedia Technology*, pp.934 – 936,2011.

[9] Lihong Zhao and Zikui Guo , "Face Recognition method based on Adaptively Weighted Block-Two Dimensional Principal Component Analysis**"** *Proceedings of Third IEEE Conference Computational Intelligence, Communication Systems and Networks*, pp. 22 – 25,2011.

[10] Yisu Zhao, Xiaojun Shen, Nicolas D. Georganas and Emil M. Petriu, "Part-based PCA for Facial Feature Extraction and Classification" *Proceedings of IEEE Conference on Haptic Audio visual Environments and Games*,pp. 99 – 1 04,2009.

[11] Katsuhiro Honda, Ryoichi Nonoguchi, Akira Notsu and Hidetomo Ichihashi ,"PCA-guided K-Means Clustering with Incomplete Data", *Proceedings of IEEE Conference on Fuzzy Systems* , Taipei, Taiwan, pp.1710 – 1714,2011.

[12] Kong Dexi and Kong Rui ," A Fast and Effective Kernel-Based K-Means Clustering Algorithm" ,*Proceedings of Third IEEE Conference on Intelligent System Design and Engineering Applications (ISDEA)*, pp. 58 – 61,2013.

[13] Vinod Kumar Dehariya, Shailendra Kumar Shrivastava and R. C. Jain **, "**Clustering Of Image Data Set Using K-Means And Fuzzy K-Means Algorithms" ,*Proceedings of IEEE Conference on Computational Intelligence and Communication Networks*, pp. 386 – 391,2010.