CHEMICAL ENGINEERING TRANSACTIONS
VOL. 46, 2015

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Peiyu Ren, Yancang Li, Huiping Song

ISBN 978-88-95608-37-2; **ISSN** 2283-9216

DOI: 10.3303/CET1546115

A Data Field Clustering Method for Classification of Concrete Dam Cracks

Long Zhao*[a], LinFeng Jiang[b]

[a] State Key Laboratory of software Engineering, Wuhan University, Wuhan, China
[b] QiLu University of Technology, Shandong, China
zxcvbnm9515@163.com

Crack detection based on digital image processing is more and more widely applied in the maintenance of concrete dam diseases. However, due to the complexity of the crack image, it is difficult to achieve high accuracy of the crack classification. To improve the shortcomings and deficiencies in cracks extraction and classification algorithm under crack detection system this article focuses on the application of general data field to effectively solve the problem of crack classification. We propose a new data field clustering method for classification of concrete dam cracks. Clustering is an important step when building a classifier for dam crack. Clustering is a process of discovering densely populated regions. In the data space, it groups a set of data in a way that maximizes the similarity within clusters and minimizes the similarity between two different clusters. The mutual information (MI) of two grids is a measure of the grid's mutual dependence. This definition is useful in the field of clustering, because it gives a way to quantify the relevance between different grids. A data field clustering method for classification of dam cracks adopts the potential values within grid and potential values between different grids. Well-known crack classification methods are compared with our method. The experimental results show that the proposed method has an obvious increase on the precision and interpretability.

1. Introduction

Using digital image processing technology to identify and classify the cracks in the dam can help us to fully grasp the dam crack information, and provide a reference for dam maintenance management and disaster warning and forecast. Crack category is one of the most important information of the crack. After the recognition and differentiation of the concrete crack, some appropriate methods are needed to determine the type of crack. Different kinds of cracks are different to the dam body. Based on the characteristics of various types of cracks. This paper extracts the feature points that can effectively distinguish the cracks from the projection method based on the generalized data field, and obtains better classification results. In order to overcome the over-segmentation problem appeared after crack extracting, this paper proposes the concept of general data field, which is based on the characteristic that the ratio between grayscale change steep area and total area is statistically less in dam crack images, and serves the increase and decrease of gradient entropy after region merging as the criterion of judging region merging. Afterwards, combined with that criterion, this paper designs an adaptive algorithm which can calculate the threshold of watershed edge segmentation, and that algorithm can solve the over-segmentation problem in crack image extracting effectively. Clustering is one of the effective ways to solve the problem of crack classification. Clustering uses some evaluation criteria to choose a most significant grids and reduce the dimensions. This thesis has mainly studied the classification of crack images. The contribution of grid is shown by the sum of potential values in data field. In order to meet the crack parameters needs of maintenance, the length of transverse and longitudinal cracks and the minimum circumscribed rectangle area of block and reticular cracks have been selected and calculated. The rest of this paper is organized as follows. Section 2 discusses the related classification methods introduced for crack detection. Section 3 explains the data field and the steps of clustering. Our method is given in Section 4. The experiment of rival clustering methods are analyzed in Section 5. A conclusion is drawn in Section 6.

Please cite this article as: Zhao L., Jiang L.F., 2015, A data field clustering method for classification of concrete dam cracks, Chemical Engineering Transactions, 46, 685-690 DOI:10.3303/CET1546115

2. Related work

The dam crack has some characteristics which can be used in the classification and damage degree of crack. The geometrical features such as length width ratio area and density are also proposed. K Ohno and M Ohtsu think the crack mode of cracking in concrete is normally changing from tensile mode to shear mode at impending failure. The classification of the cracks of the radial basis function neural network based on is designed. Li, H. proposed a method for blade crack classification based on the signals monitored by using a squared envelope spectrum. In the case of no intervention, accurate, automatic, real-time extraction of dam crack type and severity information. After extracting, crack images are always prone to cracked and incomplete. At the same time, traditional watershed algorithm has the advantage that it can maintain weak edge information good. Wen-Pei Sung et al proposed artificial neural networks to detect the damage of the dam in real time, and the better classification results are obtained, but the operation space and the speed of the algorithm are much slower. The new compound error function of BP neural network is designed, and the improvement of the training speed of neural network is improved. The distribution density is the percentage of the pixels in the image of dam crack. This definition can be understood as a one-way crack distribution density is small, and the distribution density of mesh cracks. But if there are two values of the noise points and the length of the crack length, the distribution density of the cracks is very close to the distribution density of the crack, which is more or less. So the distribution density of the characteristic value is simple and easy to use, but it can easily lead to the wrong type of some crack. Ioannis Valavanis and Dimitrios Kosmopoulos use the geometric and texture features to classify the damaged image. Noorsuhada Md No et al used the relationship between average frequency and RA value indicated clear trend with respect to crack classifications. Hence, this paper applies general data field to extracting of crack.

3. Definition of generalized data field

Data field sets each data in the dataset as an energy source which radiates its energy into the space and then generates a data field. The mutual effect among different data object is indicated by a field strength function, which might take various kinds of forms, such as nuclear form and gravitational form. The effect power at one place comes from different source could be overlaid, whose superposed result is named as potential value. Nuclear field strength on x of data field from y is calculated as follows.

$$\varphi_{(x)} = e^{-(\frac{||x-y||}{\sigma})^2} \tag{1}$$

The first derivation of field strength on x of data field from y is:

$$F_{(x)} = (y - x) * m * e^{-\left(\frac{||x-y||}{\sigma}\right)^2} \tag{2}$$

Impact factor is a parameter which controls the distance of mutual influence between two data objects. By defining an appropriate impact factor, data field model could describe the data distribution rather well. The clustering centers are the points which have local maximum potential values:

$$\hat{\varphi}(x) = \frac{1}{|H|}\sum_{i=1}^{n} m_i \times K\left(H^{-1}(x - X_i)\right) \tag{3}$$

Where $K(x)$ is a multivariate potential function, H is a positive-definitive 1×d matrix. n is the feature number of D. H is a non-singular constant matrix. Set h_j is the j_{th} data points of H.$h_j = \sum_{i=1}^{n}(E(x) - E(x_i))^2$.
The data field measurement is calculated by a grid-based importance measure algorithm. The potential function accord the data points are estimated within the grid. The potential values of each data point are calculated in different grids, and then integrate all the potential values to calculate the weight of each grid.

3.1 The steps of crack clustering

The input and output of this method are as follows: Input:Multidimensional data items' feature vectors, user-specific grid number, user-specific impact factor parameter, noise thread function (optional).Output: clusters of cracks. This method is divided into the following steps: First, feature space is divided into grids, each data point is put into a grid to form a grid-based data space. Second, an adaptive generalized data field is built to calculate the potential value of each data point, and the distribution of the potential values is calculated based on the characteristics grid. Third, according to the distribution of characteristics space is calculated. The edge of cluster starts from the clustering center, the absolute value of first derivative potential value stops increasing. Partition feature space into grids and assign data items to grids, initialize impact factor, calculate the potential value and first derivative value. The last step is searching the clustering centers. Search points where first derivative potential value equals 0. Accord to the steps above, we can detect the edges of cracks. Search the neighborhood of clustering centers and mark all candidate grids. Filter candidate grids and detect full clusters with Flood-Fill algorithm.

Calculate impact factor σ.

$$\sigma = \max_{1 \le i \le d} s_i \times ifp \quad (4)$$

Where s_i represents the length traverse of each grid and it marks all these grids which contains local maximum potential value as candidate grids. Calculate potential value and first derivative potential value according to equator.

$$MFP_{ij-1} > 0 \&\& MFP_{ij+1} < 0, 1 \le j < k-1 \quad (5)$$

where MFP_{ij-1} is the first derivative potential value of grid M_j on dimension *i*. A candidate grid could be a grid which contains clustering center (center grid in short) only if it existed to be candidate grid on all dimensions. For each center grid$\mathrm{Center}_x = (v_1, v_2, \dots, v_d)$, search its neighborhood and mark grids M_j which satisfies:

$$MFP_{ij} \ge MFP_{ij+1} \quad (6)$$

Where $v_i, 1 \le v_i \le k$, $1 \le i \le d$ is the location of Grid M_j on the dimension i of feature space, and M_{j+1} satisfies the same condition as well. Calculate noise thread t using the noise thread function$f_{[min(den_y)]}$, where den_y is the quantity of data objects inside grid inside edges. Filter all grids whose den_y<t. Using flood-fill algorithm to finds out all connected area in grids. Each connected area corresponds to a cluster in the original feature space

After preprocessing and segmentation of the image, the pixels of the crack region are 1, and the pixels of the background region are 0, and we use this feature to analyze all kinds of crack images. The specific method is as follows: (1) the digital image can be expressed in matrix. Target recognition systems usually extract features that have the following characteristics: the characteristic values of different samples from the same category should be very close, and the characteristics of samples from different classes should be different, so it is not relevant to extract information from the original data, so it is easy to distinguish between the various components of the original data. Feature selection and extraction of the basic tasks: first, to find the most discriminative description of the model, the two is to reduce the dimension of the description data. Practice has proved that the feature extraction is essential when the dimension of the data space is large. When the number of samples is small, too many features can also reduce the performance of classification and the complexity of computer. So it is very important to choose the most representative features.

We were on the potential function of the image features are as follows: the image pixel statistics distribution histogram and Fourier descriptors as the basic feature. Aiming at the characteristics of the crack image, we have to carry out the vector processing of the crack profile in the training set, thus separating the single crack area. To get the crack area, we put all the cracks in the data field. So the data field changing trends of different kinds of cracks are obtained. In the data field, the point of the distance between the smaller class and the larger class is obtained.

3.2 The classification standard of cracks

Dam cracks can be divided into vertical and horizontal cracks according to the distribution. The longitudinal cracks along the dam axis, located in the central crest appears in a few, on the downstream side near the crest. Another common crack of the dam is the transverse crack, that is, the crack is perpendicular to the dam axial.

Fine cracks occur on the surface, in a regular or irregular network. It is caused by shrinkage of concrete (or other cement). Although the fine cracks do not affect the structural integrity of the concrete, does not affect its durability and wear resistance, but it is very conspicuous, the impact is beautiful.

Both vertical and horizontal cracks will damage the integrity of the dam and reduce the bearing capacity of the dam. Transverse joints will be cut or crack seepage dam body, when the depth is larger will cause leakage of the dam, thereby endangering the safety of the dam. Therefore, we should pay special attention to it. Special attention should be paid to the cracks in the downstream side of the dam. If there is a large gap in the seam, and there is a wrong sign on the upper and lower seam, it is also the initial stage of the landslide. The dam is in high temperature and low water level operation state, it is recommended to study the operating conditions, to grasp the temperature and water level changes in the production of panel defects and impact. Strengthen the dam safety monitoring, strengthen the monitoring of the seam change and seepage flow, analyze the data, and control the working state of the dam. According to the number of pixels and the nearest neighbor rule, the matrix is used to make the grid block, determine the direction of the crack and the direction of the connection, use the horizontal and vertical direction as the projection direction, and the two value matrix is projected on the X axis and the Y axis. Contains longitudinal binary image, the projection uniform distribution on the Y axis, the concentration distribution in the X axis; and transverse images contain, projection uniform distribution on the x-axis, concentrated on the Y axis; for the cracking and crack point projection uniformity on the X axis and the Y axis. The aggregation degree of the two valued image pixels is usually used to describe the image pixels, the noise is less, but the number of connected domains of the unidirectional crack is more likely to be the same. So we adopt general data field to calculate the domains of cracks. The method of calculating the number connected domains in data space is able to distinguish the type of crack.

4. Experiment

Matlab is a high performance and powerful computing and simulation software. In the experiment, 180 kinds of crack images are used in the experiment, including 180 kinds of crack images, 100 images are used for training samples and 80 images are used for testing samples.

Figure 1: Tree crack

GA method which is proposed by E. Salari and X. Yu and SOM method which is proposed by Mathavan et al were used to compare with our classification method. Genetic algorithm using is simple and easy to operate. The chromosome length was 20, the population size was 17, and the maximum evolutionary algebra was set to 30, and the fitness function was set as the absolute value of the difference between predicted data and actual data.

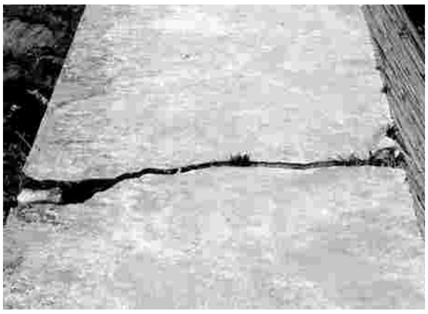

Figure 2: Horizontal crack

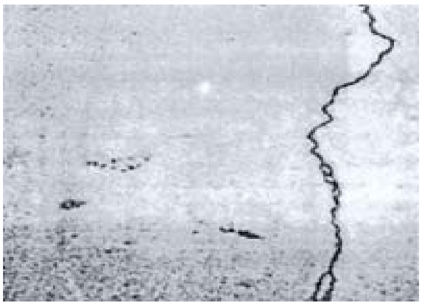

Figure 3: Vertical crack

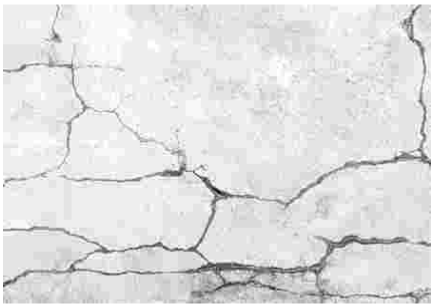

Figure 4: Mesh crack

Table 1 shows the important factor of experiment environments.

Table 1: Experiment Environment

Computers	HP xw6600 Workstation
Operation System	Windows 7 Ultimate
Software Platform	Matlab R2014b
Toolbox	LIBSVM-3.18

Different crack types occupy crack object as primes have relatively large differences, the transverse cracking and longitudinal cracking and other linear cracks often account for the whole dam image for a very small proportion. The unit area of crack object pixel number is less. Since the dam crack image has become the two value image matrix containing cracks, in each grid, a two value matrix is formed in general data field. The basic idea of the algorithm is to calculate crack geometry data center coordinates, and to coordinate geometry data field center as the center, then calculate external rectangular cracks in the number of pixels and the ratio of the rectangular area is rectangular crack pixels distribution density. If calculated density is smaller than given threshold, to expand outside the rectangle side length, then calculate the distribution of crack density, until the calculated value is larger than the threshold.

4.2 The results of cluster

Traverse all data objects and assign them into grids, record the quantity and average feature value of image in each grid. Represent the original feature space with these grids, each grid could be viewed as a data object with mass *m* and feature vector *loc*.

Table 2: The right number of three methods

Type	Test number	SOM	GA	our method
Horizontal	20	17	18	19
Vertical crack	20	19	19	20
Mesh crack	20	18	20	20
Tree crack	20	17	18	19

The effectiveness of our algorithm has been proved through the experiment. Our theoretical analysis and experimental observations reveal that our approach is the method of choice by offering a simple yet effective method and give a better understanding of crack classification problem for dam images. Through extraterritorial rectangle two pixels between the longest distances, short axis is to point to the joint normal to the long axis of the rectangle with the largest connected domain of secant line length. The ratio R reflect crack of linear features: in addition, inter pixel connectivity plays the important role of the moving target in the image boundary and region pixel is determined.

5. Conclusion

Inspired by field theory in physics, Deren Li et al proposed data field model to describe the interaction among data objects. Similar to physics, each data object is viewed as a particle with certain mass and radiates its data energy to the whole data field in order to demonstrate its existence and action in the tasks of spatial data mining. This paper adopts the method of crack classification based on the generalized data field. By the experiment, the characteristic value of the modified potential function can be used as the basis for distinguishing between the cracks and the cracks. Our theoretical analysis and experimental observations reveal that our approach is an effective clustering method and give a better understanding of the clustering on crack images. Our method is efficient and detect clusters of arbitrary shape and insensitive to the outliers. Considering the fact that the distance of mutual effect between data objects in data field ,this method could be optimized by merge each if grids to be a large grid when calculating the potential and first derivative potential value. The data field mutual information is calculated in the grid. The effectiveness of this algorithm has been proved through a series of experiments. It is insensitive to the order of input images. The overall performance of this method is better than the other algorithms.

Acknowledgement

This work was supported by the funds of Shandong provincial water conservancy scientific research and technology promotion project. The project number is SDSLKY201320 (Research on hidden danger intelligent warning system of water conservancy security based on big data).

Reference

Li D.R., Wang S.L., Gan W.Y., and Li D.Y. 2011. Data Field for Hierarchical Clustering.Int. J. Data Warehous. Min. 7, 4 (October 2011), 43-63. DOI: 10.4018/jdwm.2011100103

Li H., Zhang X., & Xu F. 2013. Experimental investigation on centrifugal compressor blade crack classification using the squared envelope spectrum. Sensors, 13(9), 12548-12563. DOI: 10.3390/s130912548.

Mathavan S., Rahman M., & Kamal K. (2015). Use of a self-organizing map for crack detection in highly textured pavement images. Journal of Infrastructure Systems, 21.DOI: 10.1061/(ASCE)IS.1943-555X.0000237

Nor N.M., Ibrahim A., Bunnori N.M., & Saman H.M. (2013). Acoustic emission signal for fatigue crack classification on reinforced concrete beam. Construction & Building Materials, 49(6), 583–590. DOI: 10.1016/j.conbuildmat. 2013.08.057

Ohno K., & Ohtsu M. (2010). Crack classification in concrete based on acoustic emission. Construction & Building Materials, 24(12), 2339–2346. DOI: 10.1016/j.conbuildmat.2010.05.004.

Salari E. and Yu X. 2011. Pavement distress detection and classification using a Genetic Algorithm. In Proceedings of the 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR '11). IEEE Computer Society, Washington, DC, USA, 1-5. DOI: 10.1109/AIPR.2011.6176378

Sung W.P., Shih M.H., and Sui C.H. 2009. Digital-Image-Correlation Technique versus Infinitely Small Element Technique for Crack Analysis of Pipe with Crevice. InProceedings of the 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery - Volume 05 (FSKD '09), Vol. 5. IEEE Computer Society, Washington, DC, USA, 105-109. DOI: 10.1109/FSKD.2009.419

Valavanis I., and Kosmopoulos D. 2010. Multiclass defect detection and classification in weld radiographic images using geometric and texture features. Expert Syst. Appl. 37, 12 (December 2010), 7606-7614. DOI: 10.1016/j.eswa.2010.04.082.

Zhao L., Wang S., & Lin Y. (2014). A new filter approach based on generalized data field. Lecture Notes in Computer Science, 8933, 319-333. DOI: 10.1007/978-3-319-14717-8_25