Research Journal of Applied Sciences 11 (12):1436-1439, 2016
ISSN: 1815-932X

Construct Self-Similarity Matrix Based on Fractal Method

Israa Hadi and Firas Sabar Miften
Faculty of Information Technology, University of Babylon, Iraq

Abstract: Self-similarity is that property of being invariant with different scale. The most of the researches used fractal dimension to calculate the self-similarity. In this study, we present a new algorithm, based on matching rang and domain fractal to find self-similarity properties of the data sets which can be used in data mining such as clustering and classification. This research focuses on two main points. Firstly, ranges and domains matching technique is used to extract self-similarity features from the images. Secondly, using self-similarity features to building the self-similarity matrix. The experimental results show that the images from same class are grouped to gather.

Key words: Fractal, self-similarity, fractal dimension, PIFS, experimental

INTRODUCTION

In image retrieval, similarity metrics are widely active. Several image processing approaches have been proposed lately to find solutions to the dilemma of content based image retrieval. The fractal which stands for a modern theory suggested through last century, enriches a novel method to the problem. The mathematical description of fractal has been clarified by authority of Mandelbrot (1982) which realizes a fractal is a collection thatthe dimension of Hausdorff Besicovich severely outreaches this topological dimension. Though, it represents a very abstract explanation. In general, it is possible to identify the fractal as a fragmentized geometric form which may be distributed into segments where each part is (at the very least nearly) a reduced-size version of that whole. Fractals may be seen as in general self-identicalas well as independent of scale (Leodkin *et al*., 2009). The conception of self-similarity in which an entityseems to be similar at various scales or objects whose small pieces resemble the whole as example this is the Sierpinski triangle in Fig. 1 illustrate self-similarity. In this mathematical object each little piece is an exact smaller copy of the whole object (Mandelbrot, 1982; Kaye, 2008; Mon, 1986).

Image similarity measurement is a main process to distinguish the place through measuring the image similarity. The basic to measure the image similarity is to construct a vector or matrix which can label the individual characteristic of image and recognize it from others. Most researchers developed techniques to mining images based on fractal theory which is used the fractal dimension as the features. The extracted features are used as input to mining algorithms.

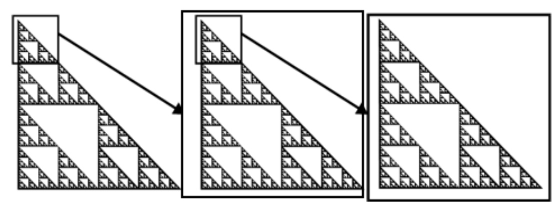

Fig. 1: Sierpinski triangle illustrate self-similarity

Shih (2014) extracted fractal dimension as features to cluster images. He suggested that using dimension of fractal for image sequencing can develop the sufficiency in content-based image retrieval. Wenping and Yingde (Liu and He, 2010) used the dimension of local fractal in the form of features to recognized the dissimilarity between on the one hand the tree leaf and on the other hand natural background. Liu *et al*. (2014) introduced fractal theory as image properties extraction. Fuzzy sequencingis used with neural network for processing characteristics. Tasoulis *et al*. (2014) present a classification tool for image of computer-assisted based on fractal as well as fuzzy sequencing for the quantification of rate of the IPF (Idiopathic Pulmonary Fibrosis) in images. Wein and Blake (1996) used the technique of clustering operation on image domain mass with the clusters constituted by the use of k-d trees.

Partitioned Iterated Function Systems (PIFS): Fractal image compression used PIFS which is applied to a single image. Suppose we are dealing with gray scale image I that has size W×W. The image is partitioned into n non-overlapping block size b×b is named as range blocks be represented by $R = \{R_1, R_2, ..., R_n\}$ note that $I = U^n_{i=1}R_i$.

Corresponding Author: Israa Hadi, Faculty of Information Technology, University of Babylon, Babylon, Iraq

Let D be the set of all possible blocks in the same image 1 which overlapping block of size b×b is named as domain blocks be represented by D = {D_1, D_2, .., D_m}. Now for each range block R_i: I∈{1, 2, 3, ..., n} must find a suitably matched domain block D = {D_1, D_2, ..., D_m} then save transformation parameters thus, obtained is called the fractal codes of image 1.

MATERIALS AND METHODS

Our contribution upgrade PIFS to use in recognition patterns where most of the researcher used a fractal dimension. PIFS applied on single image while upgrade PIFS applied on multi images where range and domain blocks extract from all images and each range block machining with all domain blocks which extract from all images. Output of upgrade PIFS is not transformation parameters but the index of image that has range block and the index of image that has domain block. finally, count the number of occurs indexes to gather.

All images go through two stages. The first known domain pool which is constituted by partitioning images into overlapping block fixed size and the second is ranges is formed from partitioning images into non-overlapping block fixed size. As far as the domain is concerned, the overlapping square results from image partitioning, the group D of domains involves all square degrees enriched by sides size 16. The domain amount double the range size. Consequently, it subsamples of 2×2 pixels to arrive a decreased domain by the identical sum of pixels like the range.

Regarding the range, the image gets partitioning to the form of squares of a non-overlapping that has sides size 8. Therefore, concerning every range, the algorithm attempts to detect a domain which provides the minimum error smaller. With respect to all recently invented ranges the procedure is re-played, i.e., becoming suitable domains have been looked for ranges. The encoding terminations in case there seems no ranges that still uncovered as well as save the image name. This, in turn, includes the block of the domain which making suitable the current range. RMSE is used to find the best fitting between rang and domain (Fig. 2).

Self-similarity matrix is computed as result from matching operation where every entry represents the number of times matching between images. Algorithm 1 read K of images while the output is k×k matrix called Similarity matrix M where entity of M(t_1.t_2) represent the number of matching range from image t_1 with domain from image t_2.

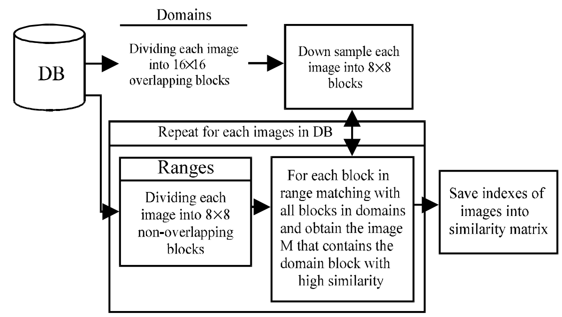

Fig. 2: Search operation to find the best matching between range and domains

Algorithm 1: Calculate Similarity matrixamong images
Input: k face images.
Output: MSimilarity matrix of size k×k
Begin
/*Building domain Pool D*/
For each image I_i: i∈{1, 2, 3,..., k}
Dividing I_i images into overlapping blocks D = {D_{i1}, D_{i2}, D_{i3}, ..., D_{im}}. The size of each domain is 2b×2b then down sample to b×b size. Where m is number of domains in image and b×b is chosen value.
End for
/*Matching range with domain Pool D*/
For each image I_i: i∈{1, 2, 3, ..., k}
Dividing I_i image into non-overlapping range blocks R = {R_{k1}, R_{k2}, R_{k3}, ..., R_{km}} that size b×b. Where n is number of ranges in image.
End for
For each range R_{ij}∈R: i∈{1, 2, 3,..., k} and j∈{1, 2, ..., n}
For each domain D_{tl}∈D: t∈{1, 2, ..., k} and l∈{1, 2, ..., m}
If RMSE (R_{ij}.D_{tl}) is smallest
M(i.t) = M(i.t)+1
End for
End for
ReturnM

The algorithm read k images as input to algorithm. To build domain Pool D in steps 1 and 2 all images are divided into overlapping blocks with size 2b×2b because size of domain block must be double size of range block. Down sample must achieve on size of domain to facility the matching between ranges and domains with equal size. Average four pixels is used as down sample. Each domain indexed by image number which has this domain and domain number. Save all domains in pool domain D to using it later. With new round in step 3 take images again to dividing into non-overlapping blocks with size b×b to get ranges R. Each rang in R compare with all domains in domain pool to find high similarity domain then recorded i the index of image that contain range and t index of image that contain domain. The output of algorithm is M similarity matrix of size k×k where k is number of input images. Rows in M represent i index of image that contain range and column represent t index of image that contain domain. The entity M(i, t) is number how many ranges in image i machining with domains in image (Fig. 3). The output matrix is not symmetric.

The entity M (2,3) means there are 12 ranges in image 2 machining 12 domains in image 3

No	1	2	3	4	5	6	7	8	9	10	11	12
1	0	12	15	3	2	14	9	6	2	3	1	3
2	17	0	12	6	6	19	9	3	1	4	1	0
3	15	19	0	4	2	8	10	5	0	7	1	2
4	4	7	2	0	13	9	10	3	2	4	0	1
5	1	4	3	9	0	6	3	8	9	8	1	1
6	16	14	12	8	2	0	7	2	4	2	1	3
7	10	6	9	8	1	13	0	5	4	5	1	1
8	3	2	5	5	2	4	0	0	14	10	3	2
9	3	8	1	8	6	7	7	9	0	9	1	4
10	10	4	2	4	9	4	2	6	3	0	1	0
11	1	4	2	1	1	1	3	1	1	2	0	9
12	1	4	0	1	0	2	1	2	1	1	2	0

Fig. 3: What is entities of matrix M

The experimental data: The two database is used in this study. The first is the FERET database of images. This database contains a total of 11338 facial images. These have been collected via photographing 994 samples at different angles, during the course of 15 meetings that took place 1993 and 1996. The color FERET is mostly a color copy of the authentic Facial Recognition Technology (FERET) database that was discharged in 20001 and encompassed 14051 grayscale images. The database aimed at developing, examine and appreciate face recognition algorithms. The total images during the color FERET database amount 512 in the form of 768 pixels. Besides, the files located through PPM-format.

Second, faces of the ORL database involves a group of face images gathered from April 1992-1994 by help of the lab. Ten dissimilar images forevery of 40 discrete subjects are identified. Concerning particular subjects, these images have gathered from various times, ranging from lighting, the expressions of face (open/closed eyes, smiling/not smiling) to the information of faces (no glasses/glasses). All of the images have considered in opposite to a dark analogous basis associating the subjects in avertical, ahead position (with endurance for certain side action). Additionally, the image size is 92×112 pixels with 256 grey ranks per pixel.

RESULTS AND DISCUSSION

In this study, we involved two databases: ORL and FERET as mentioned in this study 3. Fractal self-similarity technique was used to extract features from the databases. For ORL database choosing ten different images of each of 40 subjects and total images are 400. While the technique is implemented on FERET database on the two rounds. First choosing different images for 739 subjectsand total images are 7762. In second round choosing different images for 994 subjects and total images are 11338. The results of the proposed method showed the images from same class are grouped. Table 1 the images from 1-10 which belong to one object have high similarity.

This study presents assessing how the proposed method performs. The B-cubed estimates the recall and precision for eachitem in a cluster on a certaindata-set. The precision is the numbers ofitems in the same cluster belong to the same cluster. The recall returns the numbers of items of the same group are given to the samecluster. let be a set of objects and C is a cluster on O. Let $L(o_i)$ ($1 = i = n$) be the class of o_i given by ground truth and $C(o_i)$ be the cluster ID of o_i in C. Then, for objects, o_i and o_j, correctness of the relative between tow objects (o_i and o_j) in same cluster C is assumed by Eq. 1 (Han *et al*., 2011).

$$\text{Correctness}\left(o_i, o_j\right) = \begin{cases} 1 \text{ if } L(o_i) = L(o_j) \Leftrightarrow C(o_i) = C(o_i) \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

B-cubed precision is given by Eq. 2:

$$\text{Precision BCubed} = \frac{\sum_{i=1}^{n} \frac{\sum_{o_j: i \neq j, C(oi) = C(o_j)} \text{Correctness}(o_i, o_j)}{\left\| \left\{ o_j \middle| i \neq j, C(o_i) = C(o_j) \right\} \right\|}}{n} \tag{2}$$

B-cubed recall is given by Eq. 3

$$\text{Recall BCubed} = \frac{\sum_{i=1}^{n} \frac{\sum_{oj: i \neq j, L(o_i) = C(oj)} \text{Correctness}(o_i, o_j)}{\left\| \left\{ o_j \middle| i \neq j, C(o_i) = C(o_j) \right\} \right\|}}{n} \tag{3}$$

Table 1: Similarity matrix represent the similarity between images

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	12	15	3	2	14	9	6	2	3	1	3	0	0	0	10	0	0	2	0	4	0	2	2	0	0	0	0	0	1
2	17	0	12	6	6	19	9	3	1	4	1	0	1	0	3	0	0	1	0	2	0	0	0	1	1	2	1	0	0	2
3	15	19	0	4	2	8	10	5	0	7	1	2	2	0	2	1	1	0	0	2	0	0	1	0	1	0	1	0	0	0
4	4	7	2	0	13	9	10	3	2	4	0	1	1	0	1	0	2	0	1	0	1	3	2	1	1	2	0	0	0	1
5	1	4	3	9	0	6	3	8	9	8	1	1	0	1	1	2	1	1	1	1	2	2	2	0	2	0	0	0	0	0
6	16	14	12	8	2	0	7	2	4	2	1	3	1	0	0	2	0	1	1	2	2	0	1	1	0	0	0	0	1	2
7	10	6	9	8	1	13	0	5	4	5	1	1	0	0	1	2	0	1	1	1	1	0	2	1	2	0	1	1	1	0
8	3	2	5	5	2	4	0	0	14	10	3	2	3	0	1	1	1	0	1	0	0	0	1	3	0	1	2	0	0	0
9	3	8	1	8	6	7	7	9	0	9	1	4	0	3	2	0	2	0	0	2	0	2	2	1	0	0	1	1	0	2
10	10	4	2	4	9	4	2	6	3	0	1	0	2	1	0	1	3	0	3	1	5	2	6	1	0	0	2	2	0	1
11	1	4	2	1	1	1	3	1	1	2	0	9	18	4	5	2	5	7	2	12	1	1	1	1	4	0	1	1	1	1
12	1	4	0	1	0	2	1	2	1	1	2	0	5	22	2	19	7	10	10	3	1	2	2	0	0	0	0	0	0	0

Table 2: Clustering accuracy

Database	No. of images	Subjects	Precision (%)	Recall (%)
ORL	400	40	95	96
FERET1	7762	739	90	92
FERET2	11338	994	87	85

Table 3: Comparison of the proposed method results with other method

Methods	Precision (%)	Recall (%)
K-mean with histogram	75	84
K-mean with the proposed method	95	96

The precision and recall are presented to assess how the proposed method feature extraction with K-mean the results as shown in Table 2. The results of the proposed method feature extraction with K-mean were compared with the results that were obtained from K-mean method with histogram feature extraction. The experimental results showed that our approach yielded 95% precision and 96% recall as shown in Table 3.

CONCLUSION

The study presents a Fractal method to extract self-similarity features for clustering. This method yields the results of 95 and 96% for clustering precision and recall, respectively. In addition, the proposed method proved the possibility of the use fractal self-similarity features in image clustering without the use of fractal dimension. To sum up, the proposed method is very efficient for image clustering. It will be also useful for the clustering of other data.

REFERENCES

Han, J., M. Kamber and J. Pei, 2011. Data Mining: Concepts and Techniques. 3rd Edn., Morgan Kaufmann Publishers, USA., ISBN-13: 9780123814791, Pages: 744.

Kaye, B.H., 2008. A Random Walk Through Fractal Dimensions. 2nd Edn., John Wiley & Sons, Hoboken, New Jersey, USA., ISBN:3-527-29078-8, pp: 136-426.

Leodkin, M.A., T.A. Lebedkina and A. Jacques, 2009. Multifractal Analysis of Unstable Plastic Flow. Nova Science Publishers, New York, Hauppauge, ISBN9781606924686,.

Liu, G., N. Chen, C. Ou, Y. Liao and Y. Yu, 2014. Image feature extraction based on multifractal theory. Proceeding of the International 2014 IEEE Workshop on Electronics Computer and Applications, May 8-9, 2014, IEEE , New Taipei, Taiwan, ISBN:978-1-4799-4565-8, pp: 1023-1026.

Liu, W. and Y. He, 2010. Fractal features and fuzzy clustering based tree leaf segmentation from images with the complex background. Proceeding of International Conference on Computer, Mechatronics, Control and Electronic Engineering, August 24-26, 2010, IEEE, New Taipei City, Taiwan, ISBN:978-1-4244-7957-3, pp: 317-321.

Mandelbrot, B.B., 1982. The Fractal Geometry of Nature. 1st Edn., W.H. Freeman, San Francisco, CA., ISBN: 0716711869.

Mon, K.K., 1986. Self-similarity and fractal dimension of a roughening interface by Monte Carlo simulations. Phys. Rev. Lett., 57: 866-1986.

Shih, A.Z., 2014. An examination of color image clustering by using fractal signatures. Proceedings of the 2014 International Conference on Machine Learning and Cybernetics, July 13-16, 2014, IEEE, New Taipei City, Taiwan, ISBN:978-1-4799-4216-9, pp: 410-413.

Tasoulis, S.K., I. Maglogiannis and V.P. Plagianakos, 2014. Fractal analysis and fuzzy c-means clustering for quantification of fibrotic microscopy images. Artif. Intell. Rev., 42: 313-329.

Wein, C.J. and I.F. Blake, 1996. On the performance of fractal compression with clustering. IEEE Trans. Image Process., 5: 522-526.