Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ARCHAEOLOGICAL FRAGMENTS CLASSIFICATION BASED ON RGB COLOR AND TEXTURE FEATURES

[1]NADA A. RASHEED, [2]MD JAN NORDIN
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
E-mail: [1]nadaar@siswa.ukm.edu.my, [2]jan@ukm.edu.my

ABSTRACT

Artifacts are often found in archaeological excavation sites mixed with each other randomly. Therefore, classifying them manually is a difficult task and time consuming because they commonly exceed thousands of fragments. Thus, the aim of this study is to find a solution for classification of ancient pottery into groups by computer assistance. This is a preparatory stage for the next phase, which is the reconstruction of the archaeological fragments with high accuracy. To solve this problem, several steps must be taken, which are image segmentation via a proposed algorithm, and cluster the fragments into groups based on color and texture features. We proposed a novel algorithm that relies on the intersection of the RGB color between the archaeological fragments, and extraction of texture features from fragments based on Gray Level Co-occurrence Matrix (GLCM) that include Energy, Contrast, Correlation and Homogeneity. Finally, by using both proposed algorithm for classifying the color feature, and Euclidean distance for classifying the texture feature, we can classify the fragments with a high accuracy. The algorithm was tested on a pottery database, and it achieved a success rate almost 95%, so we would like to point out that by using the proposed algorithms we achieved promising results.

Keywords: *RGB Color, Texture Feature, Archaeological Objects, GLCM, Intersection*

1. INTRODUCTION

The archaeology is the scientific study of the last remnants of humanitarian civilization, which means studying the lives of ancient peoples by recovering and analyzing of the material culture and environmental data, which have left behind them. As well as, it provides an important information on dating, trade, technological achievements, population movements, and also for wars and conquests, this on one hand.

On the other hand, the world has witnessed great development in the performance of computers as the use of image processing and pattern recognition techniques. That encouraged researchers to attempt to solve the problem of reconstruction of fractured objects from a large collection of randomly mixed fragments via proposed system instead of assembled manually, which is a tedious and time consuming task. Such as reconstructed the torn documents [1], jigsaw puzzles [2],[3], archaeological pottery fragments [4], wall paintings [5],[6] glass plate photographs [7].

Artifacts are often found in a fractured state, and the process of manual classification may require a great deal of time and tedious work. Classifying these fragments is a challenging task, especially if an artifact object consists of thousands of fragments. Hence, it is worthy to come up with an automatic solution for the cluster of the archaeological pieces accurately and reassembling them to original form. Thus, the aim of this work is to propose a new algorithm to find a solution that categorizes ancient pottery fragments into groups depending on their color and texture. This is a preparatory stage for the next phase, that represented by reconstruction the archaeological fragments. The second aim is improving the accuracy of the results. So the most important contributions of this paper assist the archaeologists to reduce both time and manual effort required, also reduce the number of employees. By depending on the intersection of RGB colors between the fragments, we obtain the color features of an artifact. As well as by depending on GLCM [8] that includes Entropy, Contrast, Correlation and Homogeneity, we obtain the texture feature. To classify the features, we proposed new algorithm

Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for classifying the fragments according to the color feature, while we relied on Euclidean distance to classify the texture feature. After evaluation this algorithm, we achieved promising results.

This paper is structured into several sections. An overview of the reconstruction of fragmented objects is presented in Section 2, the structure search is drawn in Section 3, and an important analysis is presented in Section 4. Finally, Section 5 summarizes and highlights the most important conclusions.

2. LITERATURE REVIEW

Numerous researches challenged the task of achieving the accurate reassembly of the archaeological fragments and returning them to their original form because of the high value in information that represents past civilizations and cultures.

Ying & Gang [9] proposed an approach for the classification of ancient ceramic fragments by relying on surface texture features that were extracted by using Gabor wavelet transformation and classify by applying a non-supervision kernel-based fuzzy clustering algorithm. Their approach produces a 40 dimension eigenvector and achieved over 70%.

Smith et al. [10] suggested a method for the classification of thin ceramic fragments depending on the features of color and texture on the basis of total different geometry (TVG) energies of the image. The features are classified by the fast vector quantization algorithm, and this will classify the fragment by finding the closest neighbors within the database. Their proposed method was applied on 98 fragments in different bowls, vases, and plates. The database included two groups that were very similar in color and texture, hence it is difficult to apply the classification. The result achieved by this method was 76%. The classification was correct when using the SIFT features. The proposed method classified the fragments when using TVG features with a success of 75%.

Karasik & Smilansky [11] proposed a technique based on profile morphological analysis for the classification of ceramic. Their method depended on the mathematical representation of the profile for ceramic fragments on the basis of three functions (radius, tangent, and curvature). The authors employ the Cluster Analysis (CA) method to cluster and classify the fragments by using Discriminate Analysis (DA). The proposed algorithm has been applied on 358 fragments to reconstruct five types of vessels, and the results showed that the fragments were classified correctly 94.8%.

Makridis & Daras [12] focused on the automatic classification of fragments of pottery and ceramic that contain high or little textual information, also to enhance the classification accuracy, both the front and back view characteristics for the potteries were considered. Local features are extracted based on color and texture, which is converted to the global vector. The process of classifying the fragments was accomplished through the application of the K-Nearest Neighborhood classifier (KNN). The results after testing the model on the pottery database with a total of 62 fragments achieved a success rate 70.97%. After testing the model on the ceramic database of a total of 46 fragments, it achieved a success rate 78.26%.

Piccoli et al. [13] provided two complementary approaches for automatic classification of pottery sherds. The first one was characterized by a focus on the profile, and the other approach considers the visual surface feature. In order to extract the features, the authors relied on the points represent important information on the edge of the fragment, which will be used for the purpose of the matching process by applying the rotation invariant feature method. The second approach relied on local features that are extracted from a visual surface, such as color and texture, and also the standard deviation, michelson contrast, kirsch edge map and local binary patterns. They were classified by using K-Nearest Neighborhood and the results achieved in the overall classification accuracy according to the criteria sherd type, production technique, and Chronology were 65.99%, 93.96% and 55.65% respectively.

3. MATERIALS & METHOD

The proposed technique consists of several sets of procedures; each one performs a specific job, which can be summarized in the following structure, As shown in Figure 1 below. Moreover The algorithms are implemented in MATLAB R2014, and performed on a standard laptop computer (Intel Core i5 2.5GHz with 8GB RAM).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

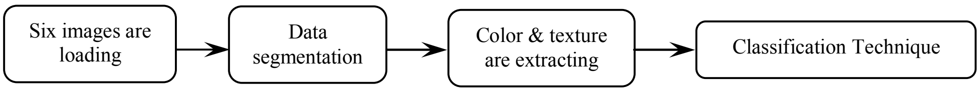


Figure 1: Structure of the Proposed Algorithm

3.1 Image acquisition and pre-processing

The first procedure loads six image files (with a JPG format), each one has dimensions of (300 x 210) pixels. In order to enhance the selection of the important features for recognition, the proposed data segmentation procedure is applied to extract the fragments from the background. The algorithm for this is as follows:

Step1: Separate the value of each pixel into three values (Red, Green, Blue).
Step2: Calculate the average and standard deviation for the rows of each matrix
Step3: Fit a Normal Distribution to the data.
Step4: Calculate the Threshold value by computing the 95% Confidence Interval for the distribution parameters.
Step 4.1: Compare each pixel with the Threshold.
Step 4.2: If it is less than or equal Threshold, then the value of pixel changes to 256.
Step 4.3: Else, the pixel has the same value.
Step5: Performs median filtering of the three matrices (Red, Green, Blue), where each output pixel contains the median value in the 5-by-5 neighborhood around the corresponding pixel in the input image.
Step6: Regroup the three matrices into one matrix.

3.2 Feature extraction

There is always a temptation to include more features in a method in the hope of improving performance. Normally, human classification of fragments is based on the color and texture. Hence, in this study the features were extracted from fragments depending on their color and texture.

3.2.1 Color feature extraction

Usually an object has a uniform color all over its surface, so when be broken into fragments, each fragment will be have the original color of the object feature. Hence, to obtain the color feature, we rely on the mathematical method that includes the intersection of colors between the fragments. The proposed algorithm includes the following steps:

Step1: Determine the values of the three colors in each pixel.

Step2: For each image, all the surface points are stored in an array which is obtained as a set of values arranged in a parallel manner.

$$C = \{\forall P(Red_{ij}, Green_{ij}, Blue_{ij})\} \quad (1)$$

Where i=1… n, j=1… m and (n, m) are the dimensions of the image matrix.

Step3: A denotes the set of the all point values, returning the first image with every element a vector of three colors, and B is another set that represents all point values that are returned to the second image.

$$A = \{\forall P_1(Red_{ij}, Green_{ij}, Blue_{ij})\} \quad (2)$$

$$B = \{\forall P_2(Red_{ij}, Green_{ij}, Blue_{ij})\} \quad (3)$$

Where P_1 and P_2 denotes the number of the color elements in the set A (first image), and set B (second image).

Step4: Obtain the set S, which represents the results from the intersection of colors between the two images:

$$S = A \cap B \quad (4)$$

Step5: This procedure is repeated for another two images, until we obtain all intersections between all images, each element of this set is a vector representing the values of three colors.

3.2.2 Texture feature extraction

As we know that the image is a set of pixels, we can define the texture as an entity consisting of a group of pixels, for this reason, we utilized the GLCM.
Initially, the RGB image must be converted to a grayscale intensity image. Then a new matrix is created, which is composed of the probability value defined by P($i,j|d,\theta$). It expresses the probability of the couple pixels at θ direction (where θ represented by the angles 0º, 45º, 90º, 135º), and d is the distance that specifies the relationship between the intersected pixel and its neighbors (where d represented by [0 1], [-1 1], [-1 0], [-1 -1]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which corresponding 0º, 45º, 90º, 135º respectively). As shown in Figure 2.

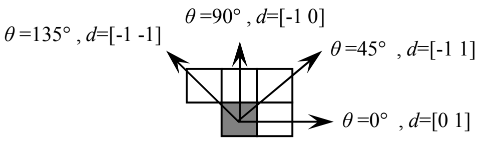


Figure 2: Measurement of GLCM for four distances d and four angles θ.

On the basis of the matrix obtained, the texture features were extracted from the archaeological fragments. Four GLCM texture features (Contrast, Correlation, Energy, and Homogeneity) are given as in equations 5,6,7,8 respectively.

$$Contrast = \sum_{i,j}^{k} P_{i,j}(i-j)^2 \qquad (5)$$

Where i and j are pixels, K is represented (row or column) dimension of a square matrix, and $P_{i,j}$ is the probability of pixel pairs.

$$Correlation = \sum_{i,j}^{k} \frac{(i-\mu_i)(j-\mu_j)P_{i,j}}{\sigma_i \sigma_j} \qquad (6)$$

$$Energy = \sum_{i,j} P_{i,j}^2 \qquad (7)$$

$$Homogeneity = \sum_{i,j} \frac{P_{i,j}}{1+|(i-j)|} \qquad (8)$$

3.3 Classification Technique

To classify the fragments successfully, two fundamental procedures must take place: the proposal of an effective algorithm to classify images based on color feature vectors, and measurement of the similarity between the textures of six images by using the Euclidean distance equation. The first procedure is classifying the fragments based on the color features, to accomplish this, we proposed a novel algorithm with a high success rate. The algorithm is given as the following:

Step1: R= [Color feature for first image,......, Color feature for sixth image]

Step2: Sort color feature for the first image.

Step3: Compare the max value in the first column feature with the values that intersect.
If the value of the corresponding column of the other image is the highest value:

Group and saved the result.

Else
If the value of the corresponding column of the other image is greater than the value of the first image:

Do not group.

Step3.1:Transfer to the second highest value, which represents the intersection with the other image.
Step3.2: Repeat this procedure.
Step3.3: Return to the first step until all the values in the first column are finished.

Step4: Repeat the procedure until all columns that represents rest the images are complete.

The second procedure classifies the texture features between the target image and the image database that represents the inputs for the second method. The feature vectors of the target image are compared with all the database vectors, and the similarity matching is calculated via the minimum Euclidean distance, which is given below:

$$D(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad (9)$$

Where q_i, p_i indicates to the values of two features for images A and B, and n is the vector size. The output of the second classification is an images sorted by the minimal distances.

4. RESULT & DISCUSSION

In order to evaluate the performance of the proposed algorithms, several tests were conducted on the image database. Where the database consist of two sets, the first one are the images obtained from a website [14], which consists of 80 ceramic fragments, most of them with high textures. The second set is ceramic fragments obtained by using a Nikon camera with a resolution of 24MP, which consists of 15 fragments. In this experiment, six images were selected and downloaded from the database. Then the image segmentation algorithm was applied for each image that has been loaded in

Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

order to obtain the fragments without a background. As shown in Figure 3, Figure 4.

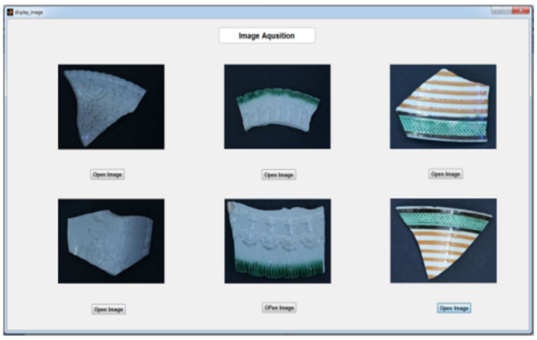


Figure 3: Image acquisition from the datasets

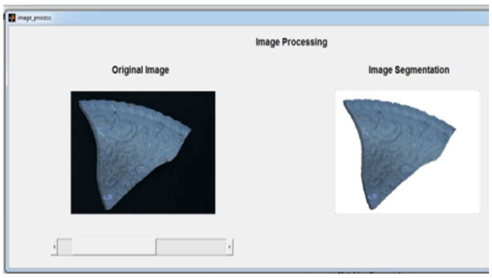
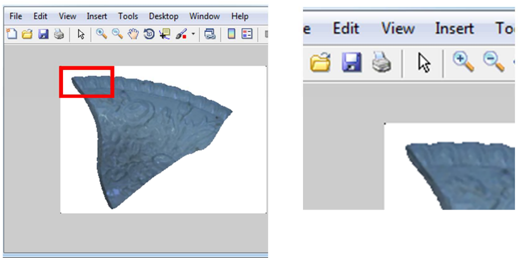


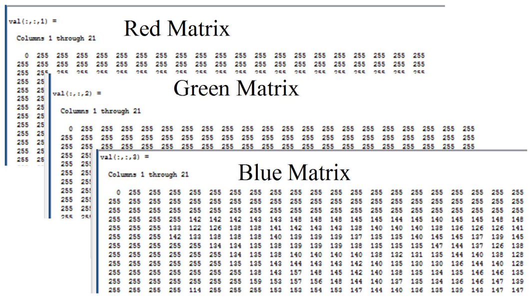
Figure 4: Applying the image segmentation algorithm.

On the basis of the aforementioned results, vectors of colors and texture features were extracted from the fragment image by applying the proposed algorithms that are described in the feature extraction section.

As depicted in Figure 5, Figure 6 below, the color matrix is separated into three matrices, and consequently a new matrix was created with vector elements that consists of three values (Red, Green, Blue) that indicate to the value of each pixel.



(a) (b)



(c)

Figure 5: (a) Demonstrates the image after segmentation.
(b) Part of image.
(c) Image data represented by three matrices.

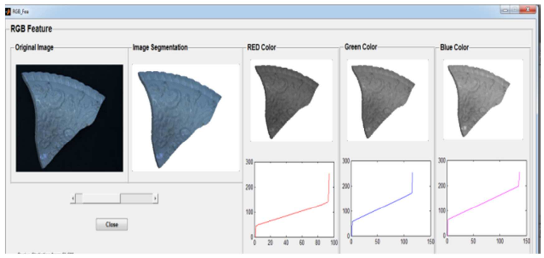


Figure 6: RGB Color feature extraction

After the completion of all the steps, Table 1 summarizes the results of the experiment, which refers to the intersection of colors between the six images. We take into consideration the application of the classification algorithm.

The first column is sorted in ascending order, which represents the intersection of the colors of the first image with the remaining five images, as shown in the Table 2 below. Additionally, we compare the highest value of 7659 with the corresponding value in column 4. We can observe that the value of 7659 is the highest in column 4, which means this image is grouped with image 4. The max value in column one is less than the highest value in the corresponding column (as shown in column 5), where the highest value is equal to 4224. It turns out that the value is less than 10737 in the same column. This means this image is not grouped with the corresponding image. Then the second value is taken, this procedure is repeated by taking all the remaining columns until all images have been processed.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TABLE 1: DEMONSTRATES THE INTERSECTION BETWEEN THE 6 IMAGES COLOR.

	Image1	Image2	Image3	Image4	Image5	Image6
Image1	0	3300	3003	**7659**	**4224**	2025
Image2	3300	0	2967	3588	**10737**	1953
Image3	3003	2967	0	2964	3687	14355
Image4	7659	3588	2964	0	4308	2154
Image5	4224	10737	3687	4308	0	2526
Image6	2025	1953	14355	2154	2526	0

TABLE 2: DEMONSTRATES THE INTERSECTION BETWEEN THE FIRST IMAGE WITH 5 COLOR IMAGES.

Image1	Index
0	1
2025	6
3003	3
3300	2
4224	5
7659	4

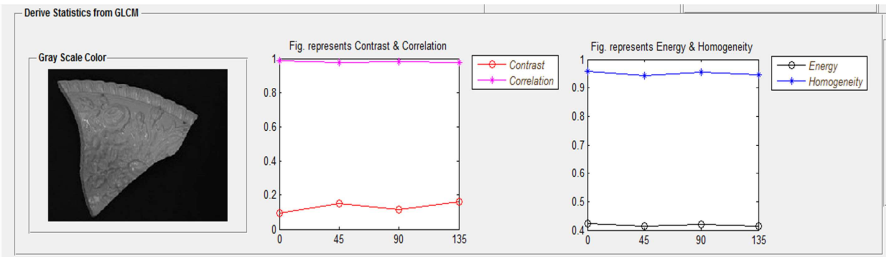


Figure 7: Gray texture feature extraction.

For the purpose of strengthening the features and obtaining the best results, we have utilized texture feature extraction of the fragments. As depicted in Figure 7, the image was converted to the gray type and the four features were calculated. The first and second Figs (inside Figure 7) represent the four texture features according to the angles (0º,45º, 90º,135º). They are near each other in almost all values. Therefore, Tables 3 and 4 demonstrate the texture features, where the first column indicates the image number and the remaining columns indicate the four features according to the four angles.

TABLE 4: DEMONSTRATES THE CORRELATION (R) AND CONTRAST (C) RESULTS ACCORDING TO THE ANGLES.

Image No.	0º		45º		90º		135º	
	R	C	R	C	R	C	R	C
1	0.9853	0.0966	0.9772	0.1504	0.9826	0.1144	0.9753	0.1633
2	0.9761	0.2336	0.9663	0.3305	0.9776	0.2193	0.9633	0.3604
3	0.9827	0.1657	0.9445	0.5311	0.9530	0.4496	0.9464	0.5123
4	0.9918	0.0538	0.9832	0.1100	0.9881	0.0778	0.9848	0.0993
5	0.9769	0.1803	0.9644	0.2772	0.9802	0.1547	0.9676	0.2526
6	0.9794	0.1810	0.9395	0.5339	0.9524	0.4193	0.9449	0.4862

TABLE 5: DEMONSTRATES THE ENERGY (E) AND HOMOGENEITY (H) RESULTS ACCORDING TO THE ANGLES.

Image No.	0º		45º		90º		135º	
	E	H	E	H	E	H	E	H
1	0.4233	0.9588	0.4142	0.9443	0.4210	0.9566	0.4144	0.9455
2	0.2231	0.9140	0.2022	0.8875	0.2184	0.9150	0.2008	0.8837
3	0.2017	0.9328	0.1756	0.8529	0.1801	0.8636	0.1763	0.8551
4	0.3714	0.9783	0.3623	0.9663	0.3680	0.9737	0.3638	0.9681
5	0.2681	0.9312	0.2520	0.9083	0.2699	0.9410	0.2544	0.9122
6	0.3463	0.9358	0.3253	0.8758	0.3297	0.8884	0.3264	0.8816

Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

On the basis of the aforementioned experiment, the Euclidean distance in Eq. 9 is applied for the purpose of classifying the target image and the dataset (the remaining five images). Towards the completion of the experiment, we combine the results of color with the results of texture and displayed in a separate window. As shown in Figure 8, the six images on the left represent the images that have been uploaded and used in the dataset, and the classification process is shown on the right side.

In this experiment, this classification technique has achieved a 100% accuracy to categorize the images into three different groups. Through 15 attempts, and by using 90 fragments that have been chosen randomly, the proposed algorithm was able to classify 85 fragments into separate groups. It failed for only 5 fragments. Hence, the success rate was 94.4%, and the percentage of failure was 5.6%.

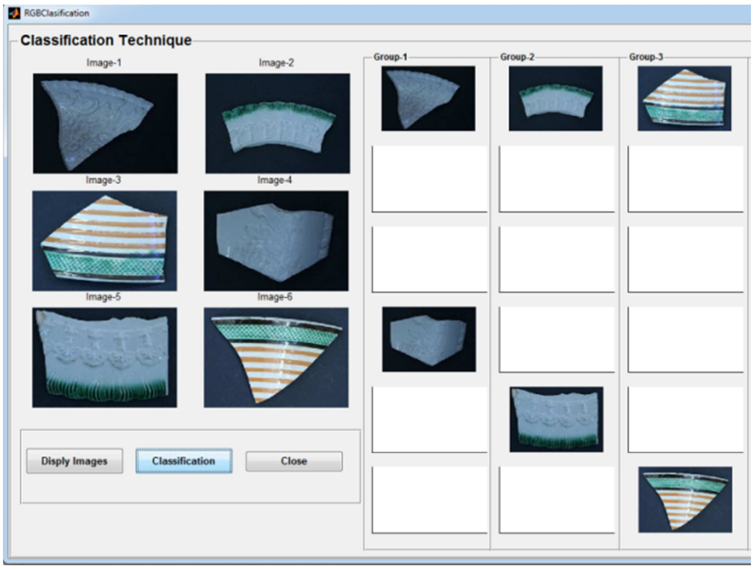


Figure 8: Classification technique depending on Color and Texture.

For the purpose of comparing the performance of the proposed algorithm with previous studies, the algorithm was applied on the fragments that adopted by previous studies [14], Table 5 illustrates the ratios that achieved.

Table 5: Illustrates The Ratios That Achieved

Previous studies	Results
Smith et al. (2010)	76%
Makridis & Daras (2012)	78.26%
Proposed method	94.4%

As shown in Table 5, the proposed method achieved a high percentage because depended on the intersection of the colors between two similar pottery fragments, which is the core to recognize the similar fragments.

5. CONCLUSION

One of the major problems that has not yet been resolved in the field of computer vision is the reassembling of broken or torn objects such as archaeological objects, documents, paleontology and art conservations. This study includes several tasks. Initially, the image segmentation algorithm was applied for the purpose of enhancing the images. Then the image features were extracted using the proposed color algorithm, and the texture feature was extracted by GLCM, which include Energy, Contrast, Correlation and Homogeneity. Finally, the extracted features were classified by combining the results of the proposed algorithm that performs classification on the basis of the image color, and the results that were obtained by the Euclidean distance that classify the texture features. Achieved impressive results through the adoption of the intersection of colors in addition to the effectiveness of the proposed classification algorithm which gave amazing results can be adopted for the classification of other patterns such as retrieve images from a large database. Afterwards, the algorithm was tested on a pottery database, and achieved success rate almost 95%, where the proposed algorithm categorized 85 fragments into groups. In conclusion, we would like

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to point out that by using the proposed algorithms we achieved promising results.

6. ACKNOWLEDGMENTS

The researchers wish to thank Universiti Kebangsaan Malaysia (UKM) and Ministry of Education, Malaysia for supporting this work by research grants ERGS/1/2012/STG07/UKM/02/9 and FRGS/1/2014/ICT07/UKM/02/2.

REFERENCES:

[1] F. Richter, X. Ries, and R. Lienhart, "A Graph algorithmic framework for the assembly of shredded documents," IEEE International Conference on Multimedia and Expo, July. 2011, pp. 1 - 6.

[2] M. Makridis, and N. Papamarkos, "A new technique for solving puzzles," IEEE Trans. Systems, Man, And Cybernetics, Part B: Cybernetics, Vol. 40, No. 3, May. 2010, pp. 789 - 797.

[3] C. Gallagher, "Jigsaw puzzles with pieces of unknown orientation," in Proc. CVPR, 2012, pp. 382-389.

[4] G. Oxholm, and K. Nishino, "A flexible approach to reassembling thin artifacts of unknown geometry," The ScienceDirect interface, Journal of cultural heritage, Vol. 14, No. 1, 2012, pp.51–61.

[5] C. Papaodysseus, P. Rousopoulos, D. Arabadjis, M. Exarhos, and M. Panagopoulos, "Reconstruction of c. 1650 B.C, fragmented wall paintings by exploitation of the thematic content," International Journal of Imaging and Robotics, Vol.5, No.11, January, 2011, pp. 1-7.

[6] H. Shin, C. Doumas, T. Funkhouser, S. Rusinkiewicz, K. Steiglitz, A. Vlachopoulos, T. Weyrich, "Analyzing and simulating fracture patterns of theran wall paintings," ACM Journal On Computing And Cultural Heritage, Vol. 5, No.3, Article 10, October, 2012, pp. 1-10.

[7] F. Stanco, D. Tanasi, and G. Gallo, "Virtual restoration of fragmented glass plate photographs of archaeological repertoires," In Virtual Archaeology Review, Vol.2, No.3, April, 2011, pp.141-144.

[8] R. M. Haralick, K Shanmugan, and I. Dinstein, "Texture features for image classification," IEEE Trans, Systems, Man Cybern, Vol. 3, 1973, pp. 610-621.

[9] L. Ying, and W. Gang, "Kernel fuzzy clustering based classification of ancient-ceramic fragments," IEEE, In Proceedings of the Conference on Information Management and Engineering, April, 2010, pp.348-350.

[10] P. Smith, D. Bespalov, A. Shokoufandeh, and P. Jeppson, "Classification of archaeological ceramic fragments using texture and color descriptors," IEEE, Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June, 2010, pp. 49-54.

[11] A. Karasik, and U. Smilansky, "Computerized morphological classification of ceramics," Journal of Archaeological Science, Vol. 38, No. 10, 2011, pp.2644-2657.

[12] M. Makridis, and P. Daras, "Automatic classification of archaeological pottery sherds" ACM Journal on Computing and Cultural Heritage, December, 2012, Vol. 5, No. 4, pp.1-21.

[13] C. Piccoli, P. Aparajeya, G. Th. Papadopoulos, J. Bintliff, F. F. Leymarie, P. Bes, M. Enden, J. Poblome, and P. Daras, "Towards the automatic classification of pottery sherds: two complementary approaches," Computer Applications and Quantative Methods in Archaeology Conference (CAA), Perth, Western Australia, Vol. 41, March, 2013, pp,1-16.

[14] Ceramic Sherd Database 2010. mst.cs.drexel.edu/datasets/Main/ACVA201, with permission of Drexel Computer Science and NEC Labs.