

Classification System for Wood Recognition Using K-Nearest Neighbor with Optimized Features from Binary Gravitational Algorithm

Ishak Taman, Nur Atika Md Rosid, Mohd Safirin Karis, Saipol Hadi Hasim, Amar Faiz Zainal Abidin, Nur Anis Nordin, Norhaizat Omar, Hazriq Izzuan Jaafar, Zailani Ab Ghani, and Jefery Hassan

Abstract—Woods species recognition is a texture classification difficulty that has been studied by many researchers years ago. The species of the wood are identified by the proposed classification using the textural type that can be observed on the structural features for example the colour of the woods, weight, texture and other features. Any mistakes on texture recognition will affect the economic benefits of wood production where it is an important basis for an identification of woods. Besides, to guide a person to be skilled in wood recognition, it will take a long time and the result the wood sample can be biased. These kinds of problem can be a motivation to develop a system that can recognize the wood effectively. This project will try to integrate both attempts by proposing a classification system consists of feature extractor, classifier and optimizer. The project proposes a classification system using Gray Level Co-Occurrence Matrix (GLCM) as feature extractor, K-Nearest Neighbor (K-NN) as classifier and Binary Gravitational Search Algorithm (BGSA) as the optimizer for GLCM's feature selection and parameters. For this project, images of wood knot from CAIRO UTM database are used for benchmarking the proposed system performance. The result shows that the proposed approach can perform as good as previous literatures with fewer features used as input for the classifier.

Keywords—Wood recognition, k-nearest neighbor, binary gravitational search, computational intelligence.

I. INTRODUCTION

MUCH of previous literatures attempt to use different kind of classifier to benchmark the system performance [1-12]. There also a few literatures that attempt to study the application of optimization technique in feature reduction [3-12]. A literature has been written by Rubiyah Yusof [3] where the used of Gabor Filter is proposed to optimally

Saipol Hadi Hasim, Amar Faiz Zainal Abidin, are with the Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johore 81310, Malaysia.

Ishak Taman, Norhaizat Omar, Zailani Ab Ghani, and Jefery Hassan, are with the Department of Electrical Engineering, Politeknik Ibrahim Sultan, 81700 Pasir Gudang, Johore, Malaysia.

Nur Anis Nordin, is with the *Faculty of Business Management*, Universiti Teknologi MARA, 40000 Shah Alam, Malaysia. (e-mail: nuranishnordin@gmail.com).

Hazriq Izzuan Jaafar, is with the Faculty of Electrical Engineering, Universiti Teknikal Malaysia, Melaka 76100, Malaysia. (e-mail:@utem.edu.my).

localized the characteristic of the images and multiply the single image of wood into two images. Then GLCM will be used as feature extraction of the image produce from Gabor Filter.

Y. T. Jing [2] proposed a method by using one-dimensional GLCM which to improve the performance of classification. It is proven that one-dimensional GLCM can be used on other applications and it is more efficient for calculations. In different literature, the same authors have make a comparative study of feature extraction methods which are GLCM and Gabor Filter with three other classifiers (k-NN, Verification-Based and Covariance Matrix).

In another literature, Y. T. Jing [4] has made a comparative study between GLCM, Gabor filters, Covariance Matrix, Verification-Based and k-NN. In this implementation, the covariance matrix is used to calculate between images which are named as feature images. Two-dimensional images or matrices generated by a feature a feature extraction algorithm is set as the feature images. The results have shown that the covariance matrix which is produced by using feature images generated by the Gabor filters has the best accuracy compared to the others.

Another comprehensive study has being done by Prasetyo [5] on the relationship of different classifiers and feature extraction methods. The features extraction methods that have being used are GLCM, Linear Binary Pattern (LBP), Wavelet, Ranklet, Granulometry, and Law's Masks. The classification methods used are ANN, k-NN, Support Vector Machine (SVM), Linear Discrete Classifier (LCD), and Quadratic Discriminant Classifier (QDC). The performance based on the computation time and classification rate have being studied ny the author in this literature.

Other literature that has been done is by Bremananth R [6]. The author proposed a wood species recognition system where the wood images is collected using a digital camera. After that, the acquired images are resized and conversion to grayscale. GLCM is used to extract the features of the images and correlation technique is used for the classification of the wood species.

M. Nasirzadah [7] is another author that proposed a wood recognition for 37 different tropical woods species based on local binary pattern. This method is used to improve the accuracy of classification for woods recognition and a result

has been reported that classification accuracy of LBP and LBP-HF using minimum distance classifier.

B. H. Wang [8] proposed a method for wood recognition based on GLCM where the images of wood stereogram were used as the research's object and applied with GLCM where four directions were used as the pixel direction of GLCM.

This paper proposed the application of BGSA [13] in selecting the features of GLCM. The objective of this paper is to optimize the features of GLCM and the result obtained should be better or the same as the previous literatures but also reduce the classification time. In this paper, the proposed model is similar to [14, 15]. The result obtained indicates that there is a slight improvement where less features is taken as input for k-NN.

II. METHODOLOGY

The proposed model is similar to [14, 15] where the model is shown in Fig. 1. The model consists of four main components which are database of wood texture images, GLCM as features extraction method, k-NN as classifier, and BGSA as optimization method.

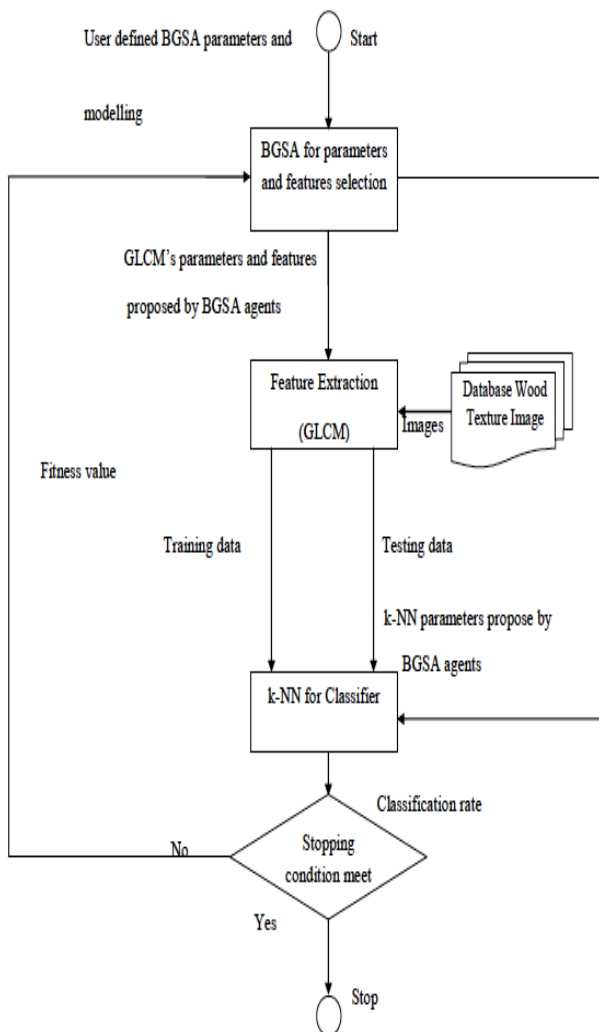


Fig.1 The proposed model for classification of wood.

The wood texture images are taken from CAIRO wood database where the wood samples are obtained from the Forest Research Institute of Malaysia (FRIM). The size of each image is 768 x 576 pixels where there are 25 types of woods consist of the images of the wood texture. The total numbers of images used in this experiment are 1250 images where 1000 images are used for training data and the other 250 images are for testing data. Figure 2 shows the scientific names and images of the wood species in the database used.

GLCM [9, 16] is use as the feature extraction method as the proposed approach. The GLCM is generated by combining the total numbers of gray pixel pairs of the images. There are four orientations that can be focused on when generating the matrix. There are two parameters that being used in the extraction method which are θ, d , and bin whereas θ is refer to the angle of between the pixel of interest and its neighbor.

There are four orientations that can be focused on when generating the matrix. The orientations used for this approach are 0° direction, 45° direction, 90° direction and 135° direction. The transition of the grey-level between the pixel of interest and its neighbors is tabulated for GLCM table. A spatial distance is represented by d which signify the number of pixels between the reference pixels i and the neighboring pixel j and bin is the number of gray-level use. The MATLAB source code is not created by own but the source code provided by previous literature has being used [17].

As mentioned earlier, The BGSA is used for determining the optimized parameters values in GLCM where the GLCM's parameters values is varying based on Equation (1) and Equation (2).

$$d = \{1, 2, 3, 4, 5, 6, 7, 8\} \quad (1)$$

$$bin = \{2, 4, 8, 16, 32, 64, 128, 256\} \quad (2)$$

The k-NN is used as the classifier due to its ease and simplicity. The classifier will classify a test data or training samples based on the highest class of the k-nearest neighbor of the test data or training samples. The average Euclidean distance is used as a tie-break if there is more than one of highest class where the highest class with the least value of the average Euclidean distance will be chosen. In k-NN, only one parameter that needs to be set by the user which is k parameter. Any possible value of k can be attempted to identify the best classification performance but the process could be boring and take a lot of times. BGSA also used to find optimized parameter value in k-NN, k for a suitable range of time period. The proposed approach limits the value of k from integer value 1 to 16.

$$1 \leq k \leq 16 \text{ and } k \in \mathbb{N}^+ \quad (3)$$

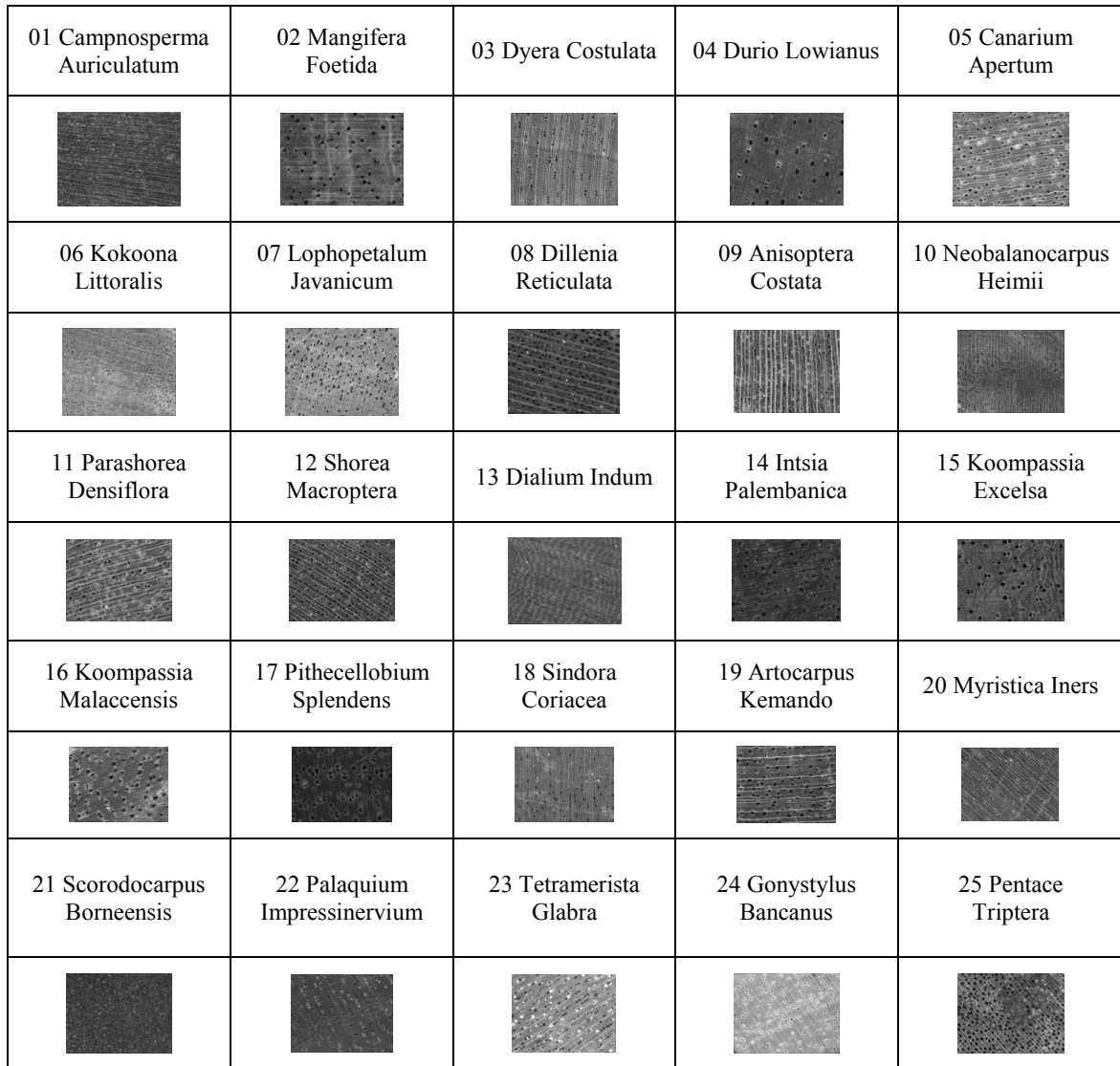


Fig. 2 Example images of each wood species.

TABLE I
FEATURES EXTRACTED FROM TEXTURE IMAGE

Feature Number	Feature	Feature Number	Feature
1	Autocorrelation	12	Sum of squares
2	Contrast	13	Sum of average
3	Correlation (Matlab)	14	Sum of variance
4	Correlation [18, 19]	15	Sum of entropy
5	Cluster prominence	16	Difference variance
6	Sum variance	17	Difference entropy
7	Sum entropy	18	Information measure of correlation (Info A)
8	Entropy	19	Information measure of correlation (Info B)
9	Homogeneity (Matlab)	20	Inverse difference
10	Homogeneity [19]	21	Inverse difference normalized
11	Maximum probability [19]	22	Inverse difference moment normalized

Gravitational Search Algorithm is inspired by the metaphor of gravitational interaction between masses [18]. According to the gravitational law, each particle is attracted to another particle by the gravitational force [19, 20]. GSA is considered as the objects where it will attract to each other by the gravitational force and the objects will move to the objects that heavier in mass. BGSA is the discrete version of the GSA where the dimension is taken only in position of 0 or 1 and the particle position is updated according to the velocity of mass [18].

In this project, each particle in BGSA is represented as the candidate solution of the optimization problem as shown in the Equation (4).

$$s = [d, bin, features, k]^T \quad (4)$$

Where:

- 1st 4 bits for k in k-NN (0000 = 1, 0001 = 2, ... 1111 = 16)
- Next 3 bits for d (000 = 1, 001 = 2, ... 111 = 8)
- Next 3 bits for bin (000 = 2, 001 = 4, 010 = 8, ..., 111 = 256)
- Next 4 bits for θ . Bit 00 for $\theta = 0^\circ$, bit 01 for $\theta = 45^\circ$, bit 10 for $\theta = 90^\circ$ and bit 11 for $\theta = 135^\circ$.
- Last 23 bits for features selections (1 = selection, 0 = off). Bit 15 is for the first feature number and bit 16 is for the second feature number and so on.

The fitness formula for each agent is shown as in the Equation (5).

$$f(s) = \frac{100\% \times T}{N} \quad (5)$$

Where:

- T = number of test data which has been classified correctly
- N = the total number of test data

Based on the model proposed above, the algorithm for BGSA can be found in [13]. The algorithm has been explained clearly with great details in [13], therefore, there is no need for redundant explanation on the implementation of BGSA's algorithm in this paper.

III. IMPLEMENTATION AND EXPERIMENTAL RESULTS

TABLE II
PARAMETERS OF BGSA

Parameters	Value
Number of agents, h	10
Number of iterations, ω	250
B	0.7
E	10

For the case study, BGSA parameters used are as listed in Table II. The results that obtained by this experiment will be compared with the previous literatures which have been done by Prasiteyo et al [27] because of the experiments that have been done by [27] having all the information that are needed

to make a comparison with the results that obtained in this experiments. The comparison will be shown in Table III.

TABLE III
COMPARISON OF RESULTS OBTAINED BY OURS AND PRASETIYO

Parameter	Ours	Prasiteyo [27]					
k-NN	1	2	3	5	7	9	11
d	7	[1, 2, 3, ..., 9, 10]					
bin	8	256					
Angle	[0°, 45°, 90°, 135°]	[0°, 45°, 90°, 135°]					
Total No. of Features	40	520					

According to the table above, it is clearly seen that implementation of BGSA in the methodology could help the features extraction to optimized the process of parameters and features selection of GLCM and k-NN. The complexity of experimental process also is reduced when all the parameters of GLCM and k-NN are chosen by BGSA where it will choose the most efficient values. The number of features use as the input for classifier also decrease by 92.31% by assuming all the possible combinations of k , d , bin , θ and number of features according to Table III.

In order to find the accuracy of classification problems, a global confusion matrix is used. Table for the confusion matrix in this experiment is shown in Table IV. The vertical axis of the confusion matrix represents predicted classes and horizontal axis represents actual defined classes.

In the matrix, the total actual number of classes in each type of woods is the sum of the row which is shown in the table. The values in the diagonal elements are correctly classified classes and have been highlighted with yellow color while the others are the wrong classifications. The global best is calculated by total up all the correct classifications and divided by the total number of classes. The classes of the wood images are classified in numbers according to the species that have been shown in Fig. 2.

From the confusion matrix, we can see that class 1 is the *Camposperma Auriculatum* species have 18 images that are correctly predicted under that wood species and 32 images are wrongly classified. For class 25 which is *Pentance Triptera*, the correctly predicted images are 32 and 18 images have being misclassified. The rest species of wood that are correctly classified can be read in diagonal elements.

The wood species class 9 which is *Anisoptera Costata* has the best classification rate where the rate is 98% correctly being classified. However for wood species class 3 (*Dyera Costulata*), class 11 (*Parashorea Densiflora*) and class 13 (*Dialium Indum*) have the worst classification rate which is 0% which are none of the species are correctly classified. There are also a few species of wood that are having confusion during the classification process due to similarities of information in the features of texture.

TABLE IV
GLOBAL CONFUSION MATRIX

Predicted Class	Actual Class																									Total	
	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		
1	18	17	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	2	0	9	14	8	0	0	0	72	
2	0	19	0	0	0	0	0	0	0	0	2	1	5	0	6	0	7	5	0	1	0	0	0	0	0	46	
3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3	0	5	
4	16	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	17	0	0	0	55	
5	0	0	0	0	39	3	6	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0	52	
6	0	0	5	0	0	23	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	1	0	65	
7	0	0	10	0	11	23	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	88	
8	2	0	0	0	0	0	0	35	0	0	3	7	0	0	0	24	0	1	0	0	0	0	0	0	0	72	
9	0	0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	57	
10	1	0	0	4	0	0	0	0	0	9	0	0	8	0	0	0	0	3	0	3	0	7	0	0	8	43	
11	0	0	0	4	0	0	0	0	0	5	0	0	1	8	5	0	0	4	3	0	0	0	0	0	0	30	
12	0	3	0	0	0	0	0	6	0	0	24	0	33	0	6	0	0	4	0	0	0	0	0	0	0	76	
13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	8	0	0	9	
14	1	2	0	0	0	0	0	0	0	0	1	0	8	0	1	0	0	2	0	0	0	0	0	0	2	17	
15	1	0	0	1	0	0	0	6	0	0	1	10	1	16	24	0	0	0	0	0	0	0	0	0	0	5	65
16	0	0	0	7	0	0	0	1	0	0	5	0	16	16	10	1	0	5	0	1	0	0	0	0	0	2	64
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43	0	0	0	0	0	0	0	0	43	
18	1	4	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	11	0	0	0	0	0	40	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	37	0	0	0	0	0	0	41	
20	0	4	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	23	0	3	0	0	0	0	34	
21	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	38	
22	8	1	0	1	0	0	0	2	0	0	0	3	3	0	0	0	0	9	0	1	0	15	0	0	1	44	
23	0	0	0	0	0	2	0	0	0	0	3	6	12	0	0	0	0	0	0	0	0	23	0	0	0	46	
24	0	0	35	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	60	
25	0	0	0	0	0	0	0	1	10	0	40	0	4	1	0	0	0	0	0	0	0	0	0	0	32	88	
Total	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	1250	

Fig. 3 shows the global best (% error) versus the time iteration. From the figure it is shown that the rate of error is decreasing from the first iteration to the last iteration. The error rate for the first iteration is 43.20% and decrease to 36.64% until the last iteration.

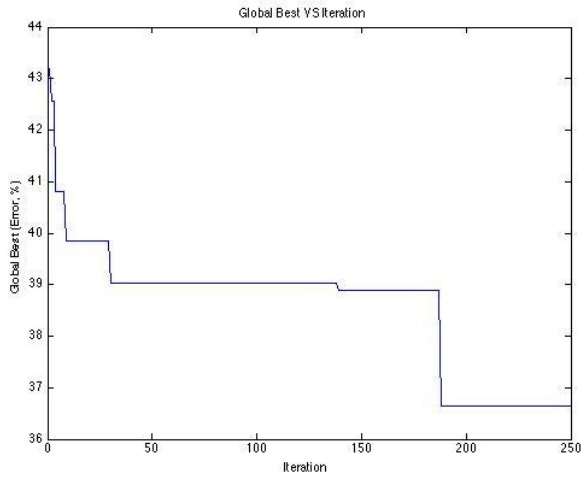


Fig. 3 Convergence curve of the experiment.

IV. CONCLUSION

In this paper, a texture classification system is developed to be used in wood species recognition. The experiment has been carried out by extracting the features of wood using GLCM and k-NN as the classifier. BGSA has been implemented in both methods in order to optimize the parameter values and features selection of GLCM and k-NN. After the comparison has been made, the proposed method shows that BGSA can reduce the difficulty of the classification processes since all the parameters of the features extraction method and classifier

are chosen by BGSA. The optimizer also enables to decrease the number of features use as input for the classifier.

REFERENCES

- [1] J. Y. Tou, P. Y. Lau, Y. H. Tay. "Computer Vision-based Wood Recognition System." Proceeding of International Workshop on Advanced Image Technology, pp. 197-202, 2007.
- [2] J. Y. Tou, Y. H. Tay, P. Y. Lau, "One-dimensional Grey-level Co-occurrence Matrices for Texture Classification." International Symposium on Information Technology, pp. 1-6, 2008.
- [3] R. Yusof, N. R. Rosli, M. Khalid. "Tropical Wood Species Recognition Based on Gabor Filter." International Congress on Image and Signal Processing, pp. 1-5, 2009.
- [4] J. Y. Tou, Y. H. Tay, P. Y. Lau. "A Comparative Study for Texture Classification Techniques on Wood Species Recognition Problem." Fifth International Conference on Natural Computation, Vol. 5, pp. 8-12, 2009.
- [5] Prasetyo, M. Khalid, R Yusof, F. Meriaudeau. "A Comparative Study of Feature Extraction Methods for Wood Texture Classification." Sixth International Conference on Signal-Image Technology and Internet-Based System, pp. 23-29, 2010.
- [6] R. Bremananth. "Wood Species Recognition System." World Academy of Science, Engineering and Technology, pp. 873-879, 2009.
- [7] M. Nasirzadeh, A. Arab Khazael, M. Khalid. "Woods Recognition System Based on Local Binary Pattern." Second International Conference on Computational Intelligence, Communication and Networks, pp. 308-313, 2010.
- [8] B. H. Wang, H. J. Wang, H. N. Qi. "Wood Recognition Based on Grey-Level Co-Occurance Matrix." International Conference on Computer Application and System Modelling, pp. VI 269-VI 272, 2010.
- [9] M. Khalid, E. L. Y. Lee, R. Yusof, M. Nadaraj. "Design of an Intelligent Wood Species Recognition System." International Journal of Simulation Systems, Science & Technology, Vol. 9, no. 3, pp. 9-19, 2008.
- [10] R. M. Harlick, K. Shanmugam, I. Dinstein. "Textual Features of Images Classification." IEEE Transaction on Systems, Man and Cybernetics, Vol. 3, no. 6, 1973.
- [11] M. K. Sayadi, R. Ramezani, N. Gaffari-Nasab. "A Discrete Firefly Meta-heuristic with Local Search for Makespan Minimization in Permutation Flow Shop Scheduling Problem." International Journal of Industrial Engineering Computations 1, pp. 1-10, 2010.

- [12] Y. L. Lew. "Design of an Intelligent Wood Recognition System for the Classification of Tropical Wood Species." M. E. Thesis, Universiti Teknologi Malaysia, 2005.
- [13] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi. "BGSA: Binary Gravitational Search Algorithm." *Natural Computing*, vol. 9, pp. 727-745, 2010.
- [14] M. Khalid, U. Khairuddin, R. Yusof. "Optimized Feature Selection for Improved Tropical Wood Species Recognition System." *ICIC Express Letters, Part B: Applications*, vol. 2, no. 2, pp. 441-446, 2011.
- [15] K. Ahmed, G. Karim, B. M. Mihammed, B. Nacera, A. Mohamed. "A Hybrid Approach for Automatic Classification of Brain MRI Using Genetic Algorithm and Support Vector Machine." *Leonardo Journal of Science*, vol. 17, p. 71-82, 2010.
- [16] D. A. Clausi. "An Analysis of Co-Occurrence Texture Statistics as a Function Grey Level Quantization." *Canada Journal of Remote Sensing*, vol. 28, no. 1, pp. 45-62, 2002.
- [17] Uppuluri. "Matlab Codes for GLCM Texture Features." URL: <http://goo.gl/LOCjK> [Date Accessed: 15th December 2011].
- [18] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi. "GSA: A Gravitational Search Algorithm." *Information Science*, Vol. 179, no. 13, pp. 2232-2248, 2009.
- [19] D. Holliday, R. Resnick, J. Walker, *Fundamental of Physics*, 4th ed. New York: John Wiley & Sons, 1993.
- [20] Schutz, *Gravity from Ground Up*, Cambridge University Press, 2003.
- [21] M. Yusof, A. F. Z. Abidin, M. N. A. Salam, K. Khalil, J. A. A. Mukred, M. K. Hani, Z. Ibrahim. "A Binary Particle Swarm Optimization Approach for Buffer insertion in VLSI Routing." *International Journal of Innovative Management, Information & Production*, vol. 2, no. 3, pp. 34-39, 2011.