

# Clustering the Tropical Wood Species Using Kohonen Self-Organizing Map (KSOM)

Azlin Ahmad

Faculty of Computer and Mathematical Sciences  
Universiti Teknologi MARA (UiTM)  
Shah Alam, Malaysia  
azlin@tmsk.uitm.edu.my

Rubiyah Yusof

Center of Artificial Intelligence and Robotics (CAIRO)  
Universiti Teknologi Malaysia (UTM)  
Kuala Lumpur, Malaysia  
rubiyah@ic.utm.my

**Abstract**—This paper discusses on how the Kohonen Self-Organizing Map (KSOM) is used as a tool to cluster and classify the tropical wood species. Wood features have been extracted through the use of two features extractors; Basic Grey Level Aura Matrix (BGLAM) and Statistical Properties of Pores Distribution (SPPD) techniques from the wood images. The wood dataset is trained and tested separately using KSOM algorithm with different parameters such as the number of epochs and map sizes in order to find the best topological network for clustering and classifying the wood data. The clustering results are analyzed and the best result is selected based on common KSOM performance measurement; topological error and quantization error. The number of cluster performed by KSOM is 61 clusters, while the number of overlapped cluster varies for each map. From the results, the 23x23 map size has produced the lowest number of overlapped clusters with the minimum value of topological error and quantization error.

**Keywords**—clustering, Kohonen Self-Organizing Map (KSOM), tropical wood species

## I. INTRODUCTION

Wood is the most important natural source in tropical countries especially Malaysia and commonly, the wood production is designated for timber industries. The International Tropical Timber Organization (ITTO) has categorized that 60% of its members comprise of tropical countries, which are known as the producer countries and the remaining are the consumer countries that involved the non-tropical countries [1]. Malaysia is one of the wood producer countries, which has the world's largest area of certified tropical forest. The total amount of wood exports of certified timber products is between 72,000 and 84,000 cu m annually or 6,000 and 7,000 cu m monthly [2], and the number has tremendously increased every year. The total exports of the wood and the wood products from January to August 2012 were RM 13.18 billion [3]. The wood is designated for wood industry, such as building and housing, house design, furniture, handicraft, boat, construction and carving. Some of these woods are indigenous species and can only be found in the certain areas of this country.

In the early 1990s at least 1/3 of log exports from Malaysia were illegal due to some reasons, including the high demand for wood and shortage of supply from Permanent Forest Reserves and State Land and the existence of illegal wood-based mills, which is still in operation [4]. Normally this illegal wood trade is carried out through wood

smuggling and the illicit operations in other countries [5]. The Malaysian firms are complicit in illegal harvesting. In Kalimantan, the wood is sometimes smuggled across the border and then shipped as "Malaysian" wood [4]. The Malaysia Timber Council has estimated that currently there are approximately 176 million m<sup>3</sup> total merchantable volume of wood in the production forest. Furthermore, assuming that the logging continues in the year 2004 with the production rate of 22 million m<sup>3</sup>, the total merchantable volume will be depleted within 8 years of by 2014. To mitigate the deforestation, the Compensatory Forestry Plantation Project was launched in 1982 and it was reported that the total of 50,000 hectares has been established. Yet, this is hardly a sufficient mitigating factor because of the 15 years maturity period while the annual rate of deforestation is about 400,000 hectares. Regardless, a commercial forest cannot replace the destroyed virgin rain forest with its accompanying rich biodiversity, which took millions of years to form.

The National Forestry Act (1984) was amended by the Malaysia Parliament on 4<sup>th</sup> August 1993 where heavier penalties would be imposed on forest offenses to deter illegal logging. However, the enforcement has again proved the weakest link due to the lack of manpower to effectively police vast tracts of the rain forest [7]. The detection and classification process is normally done manually by the wood experts who have vast experiences and knowledge in this area. They normally used dichotomous key; the traditional way of classifying the wood species by looking at the tree barks and the pattern of the wood cross-section. The task has become harder caused by some influential factors; including (1) wood reaction, (2) fungus attack, (3) site condition and (4) weather and light condition. Besides that, the other main problem is the overlapped issues among the wood species due to: -

- Some of the wood species have different features even though they are in the same species
- Some different species might have similar features with other species

These have made the wood species classification to become harder and complex. Nevertheless, the authorities have to check and verify every operation and export transaction to prevent the illegal logging. Hence, it is an urge to have an automated wood recognition system to help the authority bodies to control the situation.

## II. PREVIOUS WORKS

Khalid et.al [7] has developed an automatic wood recognition system based on image processing, feature extraction and artificial neural networks. This system is able to classify 30 tropical wood species accordingly based on macroscopic wood anatomy. The wood features are extracted using grey level co-occurrence matrix (GLCM) and Backpropagation Neural Network (BPNN) is applied as a classifier to train and classify the wood data according to its species. While Nasirzadeh et.al [8] used LBP histogram to extract features from the enhanced wood images, which help to determine the classification between the various wood species. The recognition and classification are performed using nearest neighbor classifier in the computed space with Chi-square as a dissimilarity measure.

Hence, the wood recognition work is continued by focusing on the main problem in the wood images; the lack of discriminative features of the texture images [9]. Gabor filter is chosen as a feature extractor to generate multiple processed images from a single image, in order to obtain more wood features. Gabor filter has increased the performance of the BPNN network training by producing higher accuracy compared to the previous work. Another work done by Yusof et.al [10] proposed a new mutation operation for faster feature selection by Genetic Algorithm (GA) based on exclusiveness of the allele. The standard exclusiveness in GA preserves the fit chromosomes that are evaluated using the fitness function. In the same way, the highest fit allele will be preserved and the fitness of the allele is evaluated based on the frequency of occurrences. This method has increased the optimal convergence rate for feature selection while maintaining the classification accuracy. Then, Khalid et.al [11] has developed a pre-classification stage to solve the nonlinearity of the tropical wood species separation boundaries problem using K-Means Clustering and Kernel Discriminant Analysis (KDA). The K-means clustering allows the system to compute the wood database in a respective cluster instead of computing the entire database to classify a wood species while the dimension reduction enables the wood samples to be represented accurately in a lower-dimensional space. As a result, by adding the pre-classification stage has increased the performance of the network tremendously.

The selected intelligent technique, Kohonen Self-Organizing Map (KSOM) has also been widely used as a visualization tool for dimensionality reduction. Its unique topology preserving property can be used to visualize the relative mutual relationships among the data. It has been applied to organize and visualize vast amount of textual information, for example Welfaremap and WEBSOM [12]. The main advantage of KSOM is for the topology preservation of the input space, which makes similar topics appear closely on the map. Most of these applications however are based on 2D grids and map.

## III. METHODOLOGY

### A. Data Acquisition

For this research, the wood samples are obtained from Forest Research Institute of Malaysia (FRIM), which comprises of 52 tropical wood species in cubic form where each species has 5 sample cubes. As a preparation for feature extraction process, the wood images are captured and pre-processed by using the homomorphic filter to enhance, sharpen and flatten lighting variations in the images. For this research, two feature extraction methods have been used to extract the features from wood images; the same methods used by Khairuddin et.al [15] in their work. The two feature extractors are Basic Grey Level Aura Matrix (BGLAM) and Statistical Properties of Pores Distribution (SPPD) to extract the wood features. This combination has produced 157 wood features where 136 features are from BGLAM and the other 21 is from SPPD. The total number of wood samples is 5040, which represents 52 tropical wood species. BGLAM can uniquely represent the wood images by applying the homomorphic image of wood since it can characterize the co-occurrence probability distributions of gray levels at all possible displacements configurations. Qin and Yang [16] had proved in their research that BGLAM could give the necessary and sufficient information to differentiate between images. Then SPPD technique is implemented to both black and white pores images, separately, where only distinct pores are acknowledged as characteristics of a species. Fig. 1 shows the images of ‘Palaquiumridleyi’ or Bitis.

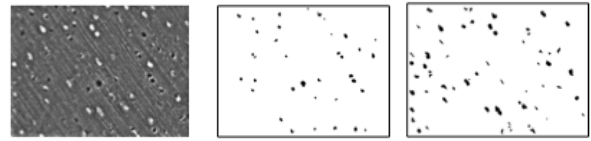


Figure 1. The images of wood (a) homomorphic image of wood,

(b) black pores image of wood and (c) white pores image of wood

There are only 10 features are obtained from black pores and another 10 features from white pores images, which are:-

- Mean size of pores and corresponding standard deviation (2 features)
- Mean distance between pores and corresponding standard deviation (2 features)
- Number of small, medium and large pores (3 features)
- Number of pairs and solitary pores (2 features)□
- Number of pores per square mm (1 feature)

There is another feature that has been obtained from the original image; the grey level feature. Thus, the total number of wood features that have been extracted using SPPD is 21 features.

### B. Data Clustering

After completing the feature extraction process, wood data is trained and tested using KSOM algorithm. In this experiment, the number of input  $x$  is the number of wood features; 157 features while the number of cluster nodes are determined by doing several experiments, arranged into an

array to form a two-dimensional topological map. The KSOM algorithm has several steps to follow during the training and testing process, which are [17]: -

- Step 0: Initialize weights  $w_{ij}$ , set topological, neighborhood and set learning rate parameters
- Step 1: While stopping condition is false, do Steps 2-8
- Step 2: For each input vector  $x$ , do Steps 3-5
- Step 3: For each  $j$ , compute:
 
$$D(j) = \sum_i (w_{ij} - x_i)^2$$
- Step 4: Find index  $J$  such that  $D(J)$  is minimum
- Step 5: For all units  $j$  within a specified neighborhood of  $J$  and for all  $I$ , update weight  $w_{ij}$ :

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha [x_i - w_{ij}(\text{old})]$$

- Step 6: Update learning rate and reduce the radius of topological neighborhood at specified times
- Step 8: Test stopping condition

The full set wood data is split into two sets; 70% is allocated for training and the remaining 30% is for the testing process. The training data set is then being trained by using KSOM network through the use of different number of rough and fine-tune epochs. The selection of epoch number is done randomly, same with the number of cluster nodes in KSOM output layer. The KSOM output nodes are arranged in two-dimensional topological map. The training process is done iteratively until it converges and satisfies the stopping condition.

In this research, the network performance is measured using two common measures; which are topological error and quantization error. The quantization error is used to measure the average distance between each data vector and its best matching unit or winning unit [18]. The value of quantization error shows the average of the distance from the input vector to the prototype in the output layer. If the quantization error value is small, it shows that the input vector is closer to its prototype. While topological error is used to measure the proportion of all data vectors for which first and second-best matching unit or winning unit are not adjacent vectors. Therefore, the lower the topographic error, the better KSOM preserves the topology. Several experiments are performed using different map sizes to investigate the effectiveness of KSOM in clustering the wood data set and to determine the most suitable map size for the data set. The clustering results are examined and analyzed by looking at the number of clusters and number of overlapped clusters produced by KSOM.

The overlapped clusters are formed due to the similarity features among the wood species. The percentage of overlapped clusters is calculated for every map sizes by using the following formula:

$$\% \text{ of overlapped} = \frac{\text{total number of error cluster}}{\text{total cluster}} \times 100 \quad (3)$$

Where the *total cluster* is referring to the number of output unit and the *total number of error cluster* is representing the number of overlapped clusters in the map.

#### IV. RESULTS AND ANALYSIS

The aim of doing the experiments repeatedly is to find the best combination of parameters for the network. This is to ensure that the KSOM network will be able to cluster and classify the tropical wood data accurately. All experiments produced 61 clusters and from the result, 23x23 is the most suitable map size for clustering the tropical wood species.

Table 1. Clustering Results of Wood Data using KSOM

Characteristics	Results
The most suitable map size	23x23
No of Epochs	50000
Quantization Error	0.28
Topological Error	0.02
No of Clusters produced by KSOM	61
No of Overlapped Clusters	11

Table 1 shows the result for KSOM in clustering and classifying the wood species. The KSOM network started to converge and generate the lowest quantization and topological error at 50000<sup>th</sup> iterations (as shown in Fig. 2).

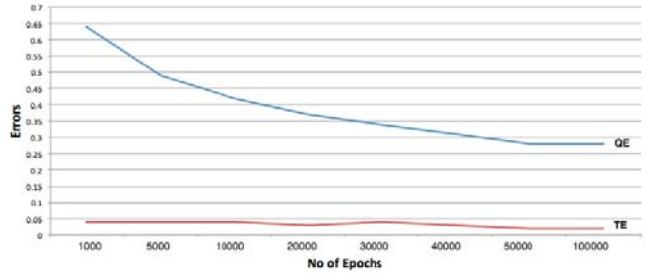


Figure 2. The Quantization Error and Topological Error for 23x23

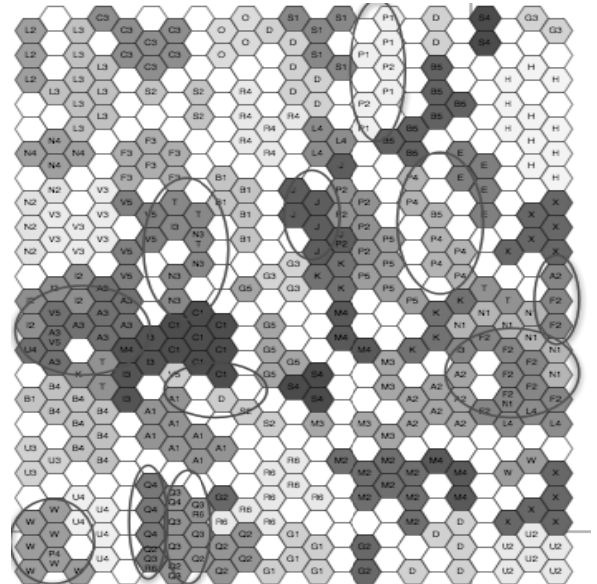


Figure 3. The 11 overlapped clusters by KSOM

This map size has produced the lowest percentage of overlapped clusters; 18.03% (or 11 overlapped clusters; as shown in Fig. 3) and minimum errors, where the quantization error is 0.28 and topological error is 0.02. Based on the result, KSOM clusters the wood data according to the wood

features; which is the pores size. This is shown clearly in Fig 4, where wood species with same pores size are arranged into the same cluster. Generally, the pores size for tropical wood species can be categorized into three sizes; (1) small, (2) medium-sized and (3) large, as shown in Table 2 and Figure 6.

Table 2. Wood Species based on Pores Sizes

Pore Sizes	Wood Species
Small	Mataulat
Medium-Sized	balau, bintangor, bitis, chengal, gerutu, giam jelutong, kapur, kasai, kekatong, keledang, keranji, kulim, machang, medang, melunak, perupok, redbalau
Large	bintangor, durian, gerutu2, kapur, kasai, keledang, keruing, machang, merantibakau, redbalau, rubberwood, sesendok

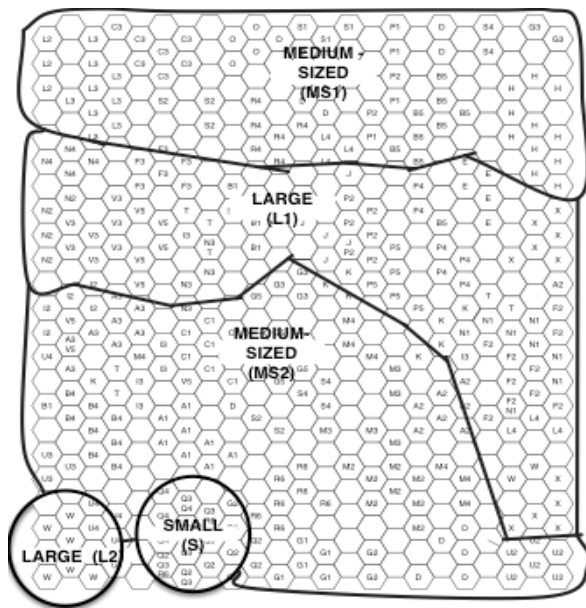


Figure 4: Clustering result according to pores size (Small, Medium-Sized and Large)

## V. CONCLUSION

Wood recognition systems have provided a huge advantage by helping the authorities to determine the originality and the quality of the tropical wood. The reason for adapting and applying different techniques and methods used in the previous researches is to increase the accuracy of the wood recognition system. The use of intelligent techniques in image processing, feature extraction and clustering and classification phases has improved the performance of existing wood recognition system. In the preliminary experiments, KSOM has been used as a clustering tool to cluster thousands of wood data into its original clusters, thus it helps in identifying the error data, which has been misallocated in other clusters. The KSOM able to cluster and classify the wood data according to the wood pore sizes even though there are 11 overlapped clusters have been formed. Further, some extension will be made by hybridizing the KSOM with optimization technique to solve the overlapped clusters problem.

## ACKNOWLEDGMENT

The authors would like to thank Malaysian Ministry of Science, Technology and Innovation for funding this research under Technofund research grant (TF0106C213).

## REFERENCES

- [1] Blaser, J., Sarre, A., Poore, D. & Johnson, S. (2011). Status of Tropical Forest Management 2011. ITTO Technical Series No 38. International Tropical Timber Organization, Yokohama, Japan.
- [2] MC Wong: Environmental Governing -- A Decade of Reaction to the Call for Saving the Forest Community and Forest of Sarawak: *The Life of Forest Peoples and Our Consumption Earth Day, Tokyo and Osaka*, 20/21 April 2002 IDEAL (Institute for Development of Alternative Living), Sibul, Sarawak, Malaysia
- [3] Oon, W. W., & Orini, H. N. (2002). Trends in Malaysian Forest Policy. pp 12-28
- [4] Tacconi L., Obidzinski K., Agung F.: Learning Lessons to Promote Forest Certification and Control Illegal Logging in Indonesia; CIFOR, 2004
- [5] Shahrin, A., & Abdul, B. (n.d.). Malaysian Raw Wood Export Ban Report.
- [6] Khalid, M., Lew, E., Lee, Y. I., & Yusof, R. (2008). DESIGN OF AN INTELLIGENT WOOD SPECIES RECOGNITION SYSTEM. *Most*, 9(3), 9–17.
- [7] Nasirzadeh, M., Khazael, a. A., & Khalid, M. Bin. (2010). Woods Recognition System Based on Local Binary Pattern. *2010 2nd International Conference on Computational Intelligence, Communication Systems and Networks*, (2), 308–313. doi:10.1109/CICSyN.2010.27
- [8] Yusof, R., Rosli, N. R., & Khalid, M. (2009). Tropical Wood Species Recognition Based on Gabor Filter. *2009 2nd International Congress on Image and Signal Processing*, 1–5. doi:10.1109/CISP.2009.5302660
- [9] Rubiyah Yusof, U. (2012). A New Mutation Operation for Faster Convergence in Genetic Algorithm Feature Selection. *International Journal of Innovative Computing, Information and Control(IJICIC)*, 8(10(B)), 7363–7379.
- [10] Khalid, M., Yusof, R., Salwa, A., & Khairuddin, M. (2011). Tropical Wood Species Recognition System based on Multi-Feature Extractors and Classifiers, (November).
- [11] Wan, W., & Fraser, D. (1993). M2dSOM: Clustering and Classification of Remotely Sensed Imagery by Combining Multiple Kohonen Self-organizing Maps and Associative Memory, (41), 2464–2467.
- [12] Forkan, F., & Shamsuddin, S. M. H. (2008). Kohonen-Swarm Algorithm for Unstructured Data in Surface Reconstruction. *2008 Fifth International Conference on Computer Graphics, Imaging and Visualisation*, 5–11. doi:10.1109/CGIV.2008.58
- [13] Uswah Khairuddin, Rubiyah Yusof, Marzuki Khalid and Florian Cordova, Optimized Feature Selection for Improved Tropical Wood Species Recognition System, ICIC Express letters, Part B: Applications, An International Journal of Research and Surveys, Volume 2, Number 2, April 2011, pp 441-446.
- [14] Qin, X., & Yang, Y.-H. (2005). Basic gray level aura matrices: theory and its application to texture synthesis. *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, (4), 128–135 Vol. 1. doi:10.1109/ICCV.2005.43
- [15] Kohonen, T. Self-organizing maps. Berlin: Springer Verlag, 2001.
- [16] Uriarte, E. A., & Martin, F. D. (2005). Topology Preservation in SOM, 19–22.
- [17] Zalhan Mohd Zin, Marzuki Khalid, Ehsan Mesbahi & Rubiyah Yusof, “3D Visualization of SOM for Data Clustering”, Proceedings of the JSST International Conference on Modeling Simulation Technology (JSST2011), Tokyo, Japan, 22<sup>nd</sup> – 23<sup>rd</sup> Oct 2011