K-MEANS SEGMENTATION AND CLASSIFICATION OF SWIETENIA MAHAGONI WOOD DEFECTS

Sri Rahayu[1] ; Dwiza Riana[2*] ; Anton[3]

Program Studi Ilmu Komputer
Universitas Nusa Mandiri
www.nusamandiri.ac.id
[1])srirahayu.rry@nusamandiri.ac.id; [2*])dwiza@nusamandiri.ac.id; [3])anton@nusamandiri.ac.id
(*) Corresponding Author

Abstract— The potential and usefulness of wood to meet the needs of human life are not in doubt. Demands us to continue to maintain the quality. Wood quality is closely related to wood defects. Manual defect checks in the wood industry are unreliable because they are prone to human error, For example, due to acute symptoms of headaches and tired eyes, technology in the form of image processing can help identify wood defects Swietenia Mahagoni. In this case, the method used is Euclidean distance with a ratio of k-means segmentation and thresholding on 42 images of wood defects consisting of 3 types of defects, namely growing skin defects, rotting knots, and healthy knots, every 14 images with data sharing. training for 30 images and testing for 12 images. The results of the k-means segmentation are then extracted on 6 features including metric, eccentricity, contrast, correlation, energy, and homogeneity using the Gray Level Co-occurrence Matrix (GLCM) extractor and classified by calculating the closest distance using the euclidean distance between the results of data feature extraction. testing of the value of feature extraction in the training data which is used as a previous database. It is the smallest value that indicates the type of defect. The success calculation is presented in the confusion matrix calculation and gets a success or accuracy value of 91.67%.

Keywords: Swietenia Mahagoni, Wood Defect, Euclidean Distance, GLCM, K-Means.

Intisari—*Potensi dan kebermanfaatan kayu untuk memenuhi kebutuhan hidup manusia tidak diragukan lagi. Menuntut kita untuk terus menjaga kualitasnya. Kualitas kayu berkaitan erat dengan cacat kayu. Pemeriksaan kecacatan yang dilakukan secara manual di industri kayu tidak dapat diandalkan karena rentan terhadap human error, misalnya karena gejala akut sakit kepala dan mata lelah dapat dibantu oleh teknologi berupa image processing dalam melakukan klasifikasi cacat kayu swietenia mahagoni. Dalam hal ini metode yang digunakan adalah euclidean distance dengan segmentasi k-means pada 42 citra cacat kayu yang terdiri dari 3 jenis cacat yaitu cacat kulit tumbuh, cacat mata kayu busuk dan cacat mata kayu sehat masing – masing 14 citra dengan pembagian data training sebanyak 30 citra dan testing sebanyak 12 citra. Hasil segmentasi k-means kemudian dilakukan ekstraksi pada 6 fitur diantaranya metric, eccentricity, contrast, correlation, energy dan homogenity dengan menggunakan ekstraktor Gray Level Co-occurrence Matrix (GLCM) dan diklasifikasi dengan menghitung jarak paling dekat menggunakan euclidean distance antara hasil ekstraksi fitur data testing terhadap nilai ekstraksi fitur pada data training yang dijadikan database sebelumnya. Nilai paling kecillah yang menunjukkan jenis cacatnya. Perhitungan keberhasilan disajikan dalam perhitungan confution matrix dan mendapat nilai keberhasilan atau akurasi sebesar 91,67%.*

Kata Kunci: *Swietenia Mahagoni, Cacat Kayu, Euclidean Distance, GLCM, K-Means.*

INTRODUCTION

Indonesia's natural wealth lies in its abundant produce, both from the green expanse and the blue expanse of the motherland's sea. Based on data from the Central Statistics Agency (BPS), Indonesia's plywood exports to Sakura Country in 2018 reached 478 thousand tons with a value of US$ 570.17 million equivalent to Rp 7.9 trillion, the largest compared to other countries (BPS, 2019). Indonesia's total plywood export volume in 2018 reached 1.35 million tons, growing 9.88% compared to the previous year(Kurniawan, 2020). Likewise, the export value rose 34.25% to the US $ 1.64 billion or equivalent to Rp. 23 trillion(BPS, 2019). Every citizen of the earth will never be separated from their needs from the results of Indonesia's wealth that comes from this green expanse.

Human primary needs, one of which is the need for a board to make a place to live (Pantow et al., 2018)(Bimo & Maryam, 2020). As happened in North Sulawesi, which makes traditional Manadonese houses to live in, they are also traded in various types, including knock-down wooden houses consisting of Manado wooden houses on stilts (RKPM), modern minimalist wooden houses

(RKMM), mixed wall wooden houses (RKMT), and gazero Yuyu(Pantow et al., 2018).

In addition to housing needs, wood is also used for transportation such as (Lanoeroe et al., 2005) fishing boat made of mahogany, a member of the Meliaceae tribe which includes 50 genera and 550 species of woody plants, including large trees with tree heights reaching 35–40 meters and diameters reaching 125 centimeters(Patria & Pribadi, 2017). To help build houses and boats, of course, the wood adhesive is needed, this can also be made from tannins derived from mahogany bark as a solution from the depleting phenol and resorcinol raw materials(Santoso & Abdurachman, 2016).

Mahogany bark extract containing tannin compounds (catechin derivatives) can react with formaldehyde and additives (catalysts and extenders) to form copolymers for wood adhesives(Lestari et al., 2019). Mahogany is one of the leading commodities. Prioritize development in the HTI program, because of its fast growth, it is also famous for its decorative veneers and plywood(Allo, 2016).

The researchers above are very straightforward in explaining the importance and potential of wood, especially mahogany wood. Seeing this, mahogany wood production needs to be preserved so that its quality is maintained. Speaking of wood quality, a wood examiner who was met in a wood storage area in West Java once said that the quality of the wood was seen from the defects of the wood. Defects which are abnormalities in the wood can affect the quality and content of the wood(Ardhiansyah et al., 2019).

If viewed from the benchmark, then to maintain the quality of the wood, because it is related to the defects that exist in the wood itself, it is necessary to at least identify the types of defects that exist in it. It is known that manual inspection of defects in the wood industry is unreliable because it is prone to human error, for example, due to acute symptoms of headaches and eye fatigue(Ardhiansyah et al., 2019). Technological assistance is needed in this handling. Image-based technology known as image processing can now be used to solve various problems such as identification or recognition by classifying them. Since the 20th century, this image processing technology has been gradually used in the field of wood defects. Extracting wood surface defects is the key to quickly identify the types of wood defects, one of which is the grayscale co-occurrence matrix, which successfully identifies three types of wood defects(Xiang et al., 2016). This study was written to segment and classify mahogany (Swietenia Mahagoni) defects by utilizing an image processing technology approach, namely using the Euclidean distance method which was previously segmented using k-means to identify the types of defects. So it is hoped that wood testers can be helped in recognizing the defects that exist in logs of logs and determining their quality.

MATERIALS AND METHOD

A study will never be carried out without material, namely the processed dataset. In this study, the dataset used is private in the form of an image of a Swietenia Mahagoni wood defect that was taken using a 13MP smartphone camera in sunlight at around 12 noon western Indonesia time in a wood storage area in West Java with a distance of 30 cm. A total of 42 images were used consisting of 3 types of wood defects, namely growth bark defects, rotten knot defects, and healthy wood defects, each with 14 images with 30 training data distribution and 12 testing images. So that each training data is 10 images for each type of defect and 4 images for testing data for each type of defect.

The following is an example of an image of a wood defect for each type of defect:

Table 1. Image Sampling

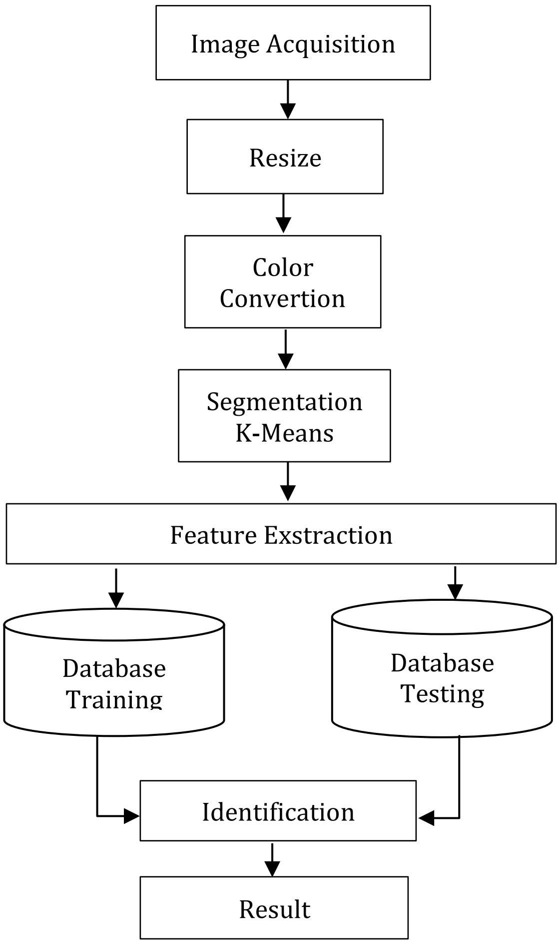
No	Type of Defects	Example Image
1	Growing Skin	
2	Rotten Knots	
3	Healthy Wood Eye	

Source : (Rahayu et al., 2021)

In table 1, you can see an example of each defect in the Swietenia Mahagoni wood that we will process.

The stages of research carried out on the dataset above are described in the chart below :

P-ISSN: 1978-2136 | E-ISSN: 2527-676X
Techno Nusa Mandiri : Journal of Computing and Information Technology
As an Accredited Journal Rank 4 based on Dirjen Risbang SK No. 85/M/KPT/2020



Source : (Rahayu et al., 2021)

Figure 1. Research Method

In Figure 1 above, we can see what stages of research were carried out, in detail the following is an explanation :

1. Image Acquisition
   Image acquisition was carried out by taking own data using a 13MP smartphone camera in sunlight at around noon western Indonesia time in a wood storage area in West Java with a distance of 30 cm between the object and the camera.
2. Resize
   Because the size of the acquired image is too large so that the image processing is old, the size is reduced from 4160 x 1968 pixels to 832 x 394 pixels.
3. Data Training & Testing
   The distribution of training and testing data in this study is 60/40. The total data amounted to 42 images then divided by 60% of the training data, namely 30 images, each type of defect contained 10 images. Then for testing 40% data, there are 12 images with each image per type of defect is 4 images.
4. Color Conversion
   Before segmentation is carried out, first the image is converted to color space into another color. To perform k-means segmentation, the color space of the image is changed from RGB to L*a*b. This color space gives more differences between colors because of their independent devices device (Raval et al., 2017).
5. Segmentation
   Segmentation is carried out to separate the defective area from other areas. Segmentation is done using k-means. The basic idea of this technique is to find the center of each possible data group and then group each data into one of these groups based on the distance(Ariani, 2020)(Premana et al., 2020). The results of this segmentation are continued to the next stage, namely feature extraction.
6. Feature Extraction
   After segmenting, the next step is feature extraction using the Gray Level Co-occurrence Matrix (GLCM) extractor on 6 features including metric, eccentricity, contrast, correlation, energy, and homogeneity. This feature extraction stage aims to determine the characteristics or patterns of the image(Kusanti & Haris, 2018). The results of feature extraction of training data are stored as a database, while the results of feature extraction of testing data will later be displayed during testing.
7. Identification
   The value of the feature extraction of the testing data has calculated the distance from the value of the training data features in the training database using the Euclidean distance. The most minimal Euclidean distance shows the closest distance for the recognition process(Utami, 2016). For example, given two feature vectors p and q, then the distance between the two feature vectors p and q is determined in the following equation:

$$P = (p_1, p_2, \dots, p_n) \quad \text{.......} \quad (1)$$

$$Q = (q_1, q_2, \dots, q_n) \quad \text{.......} \quad (2)$$

$$d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n = q_n)^2} \quad \text{.......} \quad (3)$$

$$= \sqrt{\sum_{i-1}^{n} (p_1 - q_1)^2}$$

Source : (Karimah et al., 2015)

Definition :

P = image P

Q = image Q

d = measure of distance between query image P and Q that is in the database.

p = feature vector of image P
q = feature vector of image Q

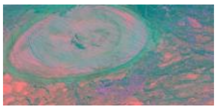
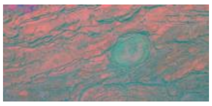
RESULT AND DISCUSSION

From the stages of the research method carried out at each stage, the results of each image processing from the original image, then color conversion, segmentation, feature extraction until identified, are described in the following discussion :

1. Result of Color Conversion

In addition to segmentation using thresholding, segmentation is also done using k-means which is converted to the L*a*b color space. Here are the results of the color conference :

Table 2. The Result of Color Conversion

No	Image RGB	Image L*a*b
1		
2		
3		

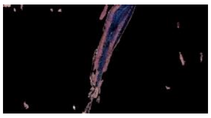
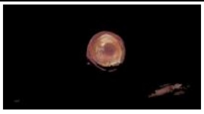
Source : (Rahayu et al., 2021)

In table 2, it can be seen the results of the color conversion to L*a*b, the defect area looks darker than the other areas. The parts that are not affected by defects tend to be colorful but rhythmic as if they are the same part.

2. Result of Segmentation

The image from the color conversion to L*a*b is then segmented with k-means. The result is as follows :

Table 3. The Result of Segmentation

No	Image RGB	Result of Segmentation
1		
2		
3		

Source : (Rahayu et al., 2021)

In table 3, it can be seen that in the segmented image using k-means, the defect area is raised and other areas are colored in black.

3. Result of Feature Extraction

Feature extraction is carried out on the k-means segmentation image resulting from the training data feature extraction and then a database is created, along with the average value of the extraction results :

Tabel 4. Average of Feature Extraction

No	Metric	Eccentricity	Contrast	Correlation	Energy	Homogeneity
1	0,4656	0,9146	0,1525	0,7801	0,8537	0,9761
2	0,2587	0,8518	0,5533	0,9143	0,3978	0,8883
3	0,3112	0,8012	0,1931	0,9319	0,7133	0,9544

Source : (Rahayu et al., 2021)

Table 4 shows the average value of the feature extraction results for each type with 6 features including metric, eccentricity, contrast, correlation, energy, and homogeneity with the GLCM method.

4. Result of Identification

After the training database is formed, the testing data is identified by the Euclidean distance which calculates the closest distance from the testing data features value to the training data feature value, the smallest value will indicate the type of defect. The identification results are presented in the following confusion matrix table :

Table 5. Confusion Matrix

	Prediction				Total
	Type	1	2	3	
Actual	1	3	1	0	4
	2	0	4	0	4
	3	0	0	4	4

Source : (Rahayu et al., 2021)

From table 5 we can calculate the success or accuracy of the identification carried out, namely

P-ISSN: 1978-2136 | E-ISSN: 2527-676X
Techno Nusa Mandiri : Journal of Computing and Information Technology
As an Accredited Journal Rank 4 based on Dirjen Risbang SK No. 85/M/KPT/2020

type 1 (growing bark) 3/4 = 75%, type 2 (rotten knot) 4/4 = 100%, and type 3 (healthy knot) 4/ 4 = 100% if the average is taken then the success or accuracy is 91.67%.

Research on the identification of wood defects has also been carried out by several other researchers, including a study entitled Wood Defects Classification Using GLCM Based Features And PSO Trained Neural Network to identify node defects on wood surfaces with datasets from the Oulu University website, Finland by extracting GLCM features and applying PSO. on Neural Network classification by producing MSE value on training 0.3483 and accuracy on testing 78.26% (Qayyum et al., 2016).

The study entitled Study on Image Recognition and Classification of Wood Skin Defects Based on BOW Model classified bark defects by extracting features using the HOG (Histogram of Oriented Gradient) and SVM methods on 4 types of bark defects Dead knot, Slipknot, Crack and Hole got an average accuracy of 85.4% (Yang & Wang, 2018).

When compared with previous research, the current research has a higher success rate. In future research, this research will be developed by developing more datasets, one of which is the augmented dataset technique.

CONCLUSION

The large role of wood to meet the needs of human life, especially mahogany requires us to continue to maintain its quality. One thing that can be done to help wood testers is to identify the type of wood defect because with manual vision alone there is a risk of mistakes. Image processing technology in this case is the solution. By doing segmentation using k-means then feature extraction is carried out on 6 features including metric, eccentricity, contrast, correlation, energy, and homogeneity using the Gray Level Co-occurrence Matrix (GLCM) extractor and identified by calculating the closest distance using the Euclidean distance between the results. feature extraction data testing on the value of feature extraction on the training data that was used as the previous database. The type of wood defect is indicated by the distance value that is closest or smallest to the difference in the value of the number of each feature with the training data. This method was applied to a dataset in the form of an image of a Swietenia Mahagoni wood defect that was taken using a 13MP cellphone camera, 42 images were used with 3 types of wood defects, namely growth bark defects, rotten knot defects, and healthy wood defects, each with 14 images. with the distribution of training data as many as 30 images and testing as many as 12 images. The results of the research conducted are very good, namely getting a success rate or accuracy of 91.67%.

REFERENCE

Allo, M. K. (2016). Kondisi Sifat Fisik dan Kimia Tanah pada Bekas Tambang Nikel serta Pengaruhnya terhadap Pertumbuhan Trengguli dan Mahoni. *Jurnal Hutan Tropis*, *4*(2), 207–217.

Ardhiansyah, A., Tavita, G. E., & M, I. A. (2019). Identifikasi Jenis Cacat Kayu Bulat Jati (Tectona Grandis Linn. F.) Pada Areal Pemanenan Di Kph Jember. *Jurnal Hutan Lestari*, *7*(1), 525–531. https://doi.org/10.26418/jhl.v7i1.32377

Ariani, D. R. (2020). *Identifikasi Penyebab Kecelakaan Lalu Lintas Dengan Menggunakan Metode K-Means*. *8*, 320–325.

Bimo, W. A., & Maryam, S. (2020). *Mekanisme Pembiayaan Pemilikan Rumah IB Maslahah Dengan Akad Murabahah Di Bank BJB Syariah Kantor Cabang Bogor*. *8*(1), 41–50.

BPS. (2019, October). *Ke Mana Tujuan Ekspor Kayu Lapis Indonesia ? 267*, 2019.

Karimah, F. U., Ernawati, & Andreswari, D. (2015). RANCANG BANGUN APLIKASI PENCARIAN CITRA BATIK BESUREK BERBASIS TEKSTUR DENGAN METODE GRAY LEVEL CO-OCCURRENCE MATRIX DAN EUCLIDEAN DISTANCE. *Jurnal Teknologi Informasi*, *11*(April), 64–77.

Kurniawan, A. (2020). *Ekspor Plywood Asal Kaltim Tetap Berlabu ke Amerika Serikat di Tengah Pandemi*. Https://Klikhijau.Com/.

Kusanti, J., & Haris, N. A. (2018). Klasifikasi Penyakit Daun Padi Berdasarkan Hasil Ekstraksi Fitur GLCM Interval 4 Sudut. *Jurnal Pengembangan IT (JPIT)*, *03*(01), 1–6.

Lanoeroe, S., Kesaulija, E. M., & Rahawarin, Y. Y. (2005). Pemanfaatan Jenis Tumbuhan Berkayu sebagai Bahan Baku Perahu Tradisional oleh Suku Yachai di Kabupaten Mappi. *Biodiversitas*, *6*(3), 212–216.

Lestari, A. S. R. D., Hadi, Y. S., Hermawan, D., Santoso, A., & Pizzi, A. (2019). Physical and mechanical properties of glued-laminated lumber from fast-growing tree species using Mahagony tannin adhesive. *Wood and Fiber Science*, *51*(2), 1–12.

Pantow, F. S., Tumbuan, W. J. F. A., & Ogi, I. W. J. (2018). Analisis Pemasaran Relasional Terhadap Loyalitas Pelanggan Rumah Kayu Manado Pada Pt. Woloan Permai Perkasa. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*, *6*(3), 1618–1627. https://doi.org/10.35794/emba.v6i3.20321

Patria, A. A., & Pribadi, T. W. (2017). Analisis Teknis dan Ekonomis Pembangunan Kapal Ikan Tradisional Ukuran <10 GT Berbahan Kayu Utuh Dengan Teknologi Laminasi Kayu Mahoni. *Jurnal Teknik ITS*, *6*(1). https://doi.org/10.12962/j23373539.v6i1.22839

Premana, A., Bhakti, R. M. H., & Prayogi, D. (2020). Segementasi K-Means Clustering Pada Citra Menggunakan Ekstraksi Fitur Warna Dan Tekstur. *Information Technology Journal of UMUS*, *2*(1), 89–97.

Qayyum, R., Kamal, K., Zafar, T., & Mathavan, S. (2016). Wood Defects Classification Using GLCM Based Features And PSO Trained Neural Network. *International Conference on Automation and Computing (ICAC)*, 3–7.

Rahayu, S., Riana, D., & Anton. (2021). Identifikasi Cacat Kayu Swietenia Mahagoni Menggunakan Euclidean Distance Dengan Perbandingan Segmentasi K-Means Dan Thresholding. *Jurnal Techno Nusa Mandiri*, *18*.

Raval, K., Shukla, R., & Shah, A. K. (2017). Color Image Segmentation using FCM Clustering Technique in RGB , L * a * b , HSV , YIQ Color spaces. *European Journal of Advances in Engineering and Technology*, *4*(3), 194–200.

Santoso, A., & Abdurachman, A. (2016). Karakteristik Ekstrak Kulit Kayu Mahoni Sebagai Bahan Perekat Kayu. *Jurnal Penelitian Hasil Hutan*, *34*(4), 269–284. https://doi.org/10.20886/jphh.2016.34.4.269-284

Utami, E. (2016). Penggunaan Principal Component Analysis dan Euclidean Distance untuk Identifikasi Citra Tanda Tangan Use of Principal Component Analysis and Euclidean Distance to Identify Signature Image. *IPTEK-KOM, 16*(1), 1–16.

Xiang, Z. Y., Qin, Z. Y., Ying, L., Quan, J. L., & Wei, C. Z. (2016). Identification of Wood Defects Based on LBP Features. *Proceedings of the 35th Chinese Control Conference*, 4202–4205.

Yang, F., & Wang, Y. (2018). Study on Image Recognition and Classification of Wood Skin Defects Based on BOW Model. *2018 International Conference on Mechanical, Electrical, Electronic Engineering & Science (MEEES 2018)*, 80–84. https://doi.org/10.2991/meees-18.2018.16