Indonesian Ethnicity Recognition Based on Face Image Using Gray Level Co-occurrence Matrix and Color Histogram

Dinda Mareta Putriany\textsuperscript{a}, Ema Rachmawati\textsuperscript{b}, Febryanti Sthevanie\textsuperscript{c}

School of Computing, Telkom University, Bandung, Indonesia

E-mail: \textsuperscript{a}dindaputriany@student.telkomuniversity.ac.id, \textsuperscript{b}emarachmawati@telkomuniversity.ac.id, \textsuperscript{c}sthevanie@telkomuniversity.ac.id

Abstract. Facial and ethnic recognition became a popular research topic. Different from face recognition, ethnic recognition classifies faces according to the general features of certain ethnic groups. Ethnic recognition in face image is increasingly becoming a necessity and is used in various fields. This paper proposed the Indonesian ethnic recognition system based on periorbital features on facial images. We use the five largest ethnic groups in Indonesia in this study, namely Banjar, Bugis, Javanese, Malay, and Sundanese. Gray Level Co-occurrence Matrix and Color Histogram were used as methods, and Random Forest was used as a classifier. Based on - cross-validation tests with optimal k values, the model achieved 98.65% accuracy.

1. Introduction
The human face is a biometric key that provides demographic information on gender, age, and ethnicity useful in facial recognition systems \cite{1}. Facial recognition is a popular research topic because it has a wide range of applications \cite{2}. For two decades, this research has been studied to understand how humans perceive and determine a given image's ethnicity because ethnicity plays an important role in applying facial images \cite{1,3}. Based on Potensi Desa (PODES) in 2014, Indonesia has more than 250 ethnicities. As many as 71.78\% of Indonesia regions have a heterogeneous composition of citizens and have historically shown that their development is spread heterogeneously with high diversity \cite{4}, ethnic recognition is increasingly difficult for humans to do. Ethnic recognition can be used in population data collection activities to assist an agency in making decisions. It can also be used to identify crime suspects to narrow the scope of searches in the database \cite{5} and improve the security of biometric systems in protecting personal and business information by limiting and monitoring access to data \cite{6}. In this study, we propose an Indonesian ethnic identification system based on the periorbital face, namely the perioocular region (the area around the eyes including the eyebrows) to the nose as a Region of Interest (ROI), because it is the most dominant and important area on the face \cite{6,7} and is an effective biometric approach. We used ethnic Banjar, Bugis, Javanese, Malay, and Sundanese in this study because these are the five largest ethnic groups in Indonesia. We used a combination of textural and color approaches because they are very important for effective facial information \cite{8}. Specifically, we used the GLCM method to capture texture, Color Histogram to capture colors, and Random Forest as a classifier.

2. Proposed System
This section describes our proposed system. An overview of our system is shown in Figure 1. The proposed Indonesian ethnic recognition system consists of three steps, namely preprocessing, feature...
extraction, and classification. First, the RGB image is entered into the system and then preprocessing the image. Second, each preprocessed image is performed feature extraction using GLCM and Color Histogram. Third, the results of feature extraction are classified using Random Forest as the classifier. The classification results are used for system analysis.

![Figure 1. System overview.](image)

2.1. Dataset
The dataset used in this study belongs to the Multimedia Laboratory, Faculty of Informatics, Telkom University [9]. The dataset contains images of the five largest ethnic groups in Indonesia, consisting of Banjar, Bugis, Javanese, Malay, and Sundanese. The image used for the classification is approximately 400 x 200 pixels. The number of each class is 485 for Banjar, 465 for Bugis, 440 for Javanese, 455 for Malay, and 445 for Sundanese, with a total of 2290 periorbital images.

![Figure 2. Image samples from ethnic (a) Banjar (b) Bugis (c) Javanese (d) Malay (e) Sundanese.](image)

2.2. Preprocessing
The purpose of preprocessing is to make the data more ideal before the feature extraction process. At this step, the image with the RGB color space is then converted into a grayscale image for GLCM feature extraction and converted into an HSV image for Color Histogram feature extraction. The preprocessing scheme is shown in Figure 3.

2.3. Gray Level Co-occurrence Matrix (GLCM) Feature Extraction.
The GLCM is calculated as a second-order histogram of a gray image. GLCM is a matrix whose dimensions depend on the intensity of the gray level G in an image. The feature extraction scheme using GLCM is shown in Figure 4.
The first step is a quantized grayscale image. The goal is to convert the grayscale value of an image and classify it into $G$ levels (degrees of gray) expressed in integers. In the Co-occurrence stage, the number of occurrences of pixel intensity value level with another pixel intensity is counted within one distance (distance = 1) and the orientation of $0^\circ$, $45^\circ$, $90^\circ$, and $135^\circ$. In this step, it will produce 4 (four) co-occurrence matrices. In the Symmetric step, the co-occurrence matrix is added to the transposition matrix, aiming to make the matrix symmetrical on the diagonal. The normalization step divides each set of pixel pairs by the total number of pixel pairs used. After that, a GLCM feature extraction process uses 7 (seven) components measured from the Haralick Feature:

- **Contrast**
  Used to measure how far the intensity of a pixel differs from its neighboring pixels throughout the image. When the contrast is 0, then the image does not change the pixel intensity (constant). The contrast value increases exponentially. Contrast can be calculated by equation (1) [8]:

$$\text{Contrast} = \sum_{i,j=0}^{G-1} (i - j)^2 p(i,j) \quad (1)$$

- **Energy**
  Energy is a value that shows how regular the intensity of the pixels in the image. The more organized an image, the energy value is getting higher, and vice versa. Energy can be calculated by equation (2) [11].

$$\text{Energy} = \sqrt{\sum_{i,j=0}^{G-1} p(i,j)^2} \quad (2)$$
• Homogeneity
Homogeneity is a value that shows how close the intensity of a pixel with neighboring pixels. The higher the homogeneity value, then the image's differences pixel intensity is getting closer with value ranges 0 to 1 [11].

\[
Homogeneity = \sum_{i,j=0}^{G-1} \frac{p(i,j)}{1 + (i-j)^2}
\] (3)

• Correlation
Correlation is a value that measures how big the relationship between two pixels. Correlation is the value of the gray linear dependence between pixels in two different directions [12].

\[
Correlation = \sum_{i,j=0}^{G-1} p(i,j) \left[ \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i^2\sigma_j^2} \right]
\]

With \( \mu_i = \sum_{i,j=0}^{G-1} i \times p(i,j) \), \( \mu_j = \sum_{i,j=0}^{G-1} j \times p(i,j) \),
\( \sigma_i^2 = \sum_{i,j=0}^{G-1} p(i,j)(i-\mu_i)^2 \), dan \( \sigma_j^2 = \sum_{i,j=0}^{G-1} p(i,j)(j-\mu_j)^2 \) (4)

• Dissimilarity
Dissimilarity is a measure of the dissimilarity of an image. The value will be high if the image has random pixels, and vice versa will be low if the image pixels are uniform. The value increases linearly. Dissimilarity can be calculated by equation (5) [11].

\[
Dissimilarity = \sum_{i,j=0}^{G-1} |i - j|p(i,j)
\] (5)

• Angular Second Moment (ASM)
ASM is a measure of the homogeneity of an image. ASM has a high value when the image has great homogeneity and vice versa. It can be calculated by equation (6) [11].

\[
ASM = \sum_{i,j=0}^{G-1} p(i,j)^2
\] (6)

• Entropy
Entropy is used to measure the complexity or abnormality of an image. Entropy can be calculated by equation (7) [11].

\[
Entropy = \sum_{i,j=0}^{G} p(i,j) \ log_2 p(i,j)
\]

In the equation, \( i \) and \( j \) is the appearance of a pixel pair, \( p(i,j) \) is the GLCM normalized value, and \( G \) is the gray level.

2.4. Color Histogram Feature Extraction
This extraction uses the HSV color space, commonly used and effective in extracting color features for facial recognition [8][13]. HSV also has a better performance in distinguishing colors when compared to RGB. There are 3 (three) steps: color quantization, normalization, and color histogram construction. Color quantization aims to reduce the possible number of colors to reduce the computation process, and the required process will be easier. Quantization divides each color pixel by a specific range of values.
Pixels are grouped into n groups of color pixel value ranges, called bins. Each bin in the color histogram will contain a statistical value in the form of a percentage of the number of pixels. Then, each group was normalized to simplify the color distribution values on the histogram. Besides, normalization also aims to make the histogram value remain the same in images with the same color distribution with different image sizes.

2.5. Classification
Random Forest (RF) algorithm was used as a classifier in this study. The RF classification scheme is shown in Figure 5. This classification process aims to determine an image's class into 5 (five) ethnic classes, namely Banjar, Bugis, Javanese, Malay, and Sundanese. There are 2 (two) processes, namely training and testing. RF conducts a training process on several random and different data samples (bootstrap). As much as one-third of the sample data from the training set is used as Out-of-Bag (OBB) data, namely, data that is not selected as a bootstrap to calculate errors and determine important variables. The rest is used as In-Bag data to form trees. Each tree can provide predictions with the same or different results. The final prediction is selected based on majority voting.

![Random forest classification scheme](image)

**Figure 5.** Random forest classification scheme [18].

3. Experimental Result
This evaluation step uses stratified k-fold cross-validation, a k-fold variation that divides data into k partitions by ensuring that each partition has the same class proportions [14]. There are k repetitions, and the data divided into k partitions by making one part of the training data while the other becomes test data. Each fold has at least been validated once. The final result is the accuracy of the model performance obtained from the average of each iteration. In this study, there are 4 (four) parameter tuning scenarios: 1) The Gray Level Co-occurrence Matrix (GLCM) parameter tuning; 2) Color Histogram parameter tuning; 3) Combining the GLCM and Color Histogram methods; 4) Comparison of classification and cross-validation methods. In the scenario of tuning parameters 1, 2, and 3 using the Random Forest classification method with 100 trees and $k = 10$ for cross-validation.

3.1. Parameter Tuning Results

3.1.1. Gray Level Co-occurrence Matrix (GLCM) Parameter Tuning Results. There are 7 (seven) global features of the Haralick Feature, namely contrast, energy, homogeneity, correlation, dissimilarity, ASM, and entropy, used in this parameter tuning. These seven features are features commonly used by researchers in facial recognition research. The features that produce good accuracy will be selected in this Indonesian ethnic recognition system. The scenario of using the Haralick Feature and the results is given in Table 1.
### Table 1. Results of GLCM parameter tuning.

| Scenario          | Haralick Feature                                | Accuracy (%) |
|-------------------|------------------------------------------------|--------------|
| Scenario-1        | Contrast                                       | 57.90        |
| Scenario-2        | Energy                                         | 46.90        |
| Scenario-3        | Homogeneity                                    | 47.03        |
| Scenario-4        | Correlation                                    | 56.64        |
| Scenario-5        | Dissimilarity                                  | 54.72        |
| Scenario-6        | ASM                                            | 46.68        |
| Scenario-7        | Entropy                                        | 56.46        |
| Scenario-8        | Contrast, energy, homogeneity, correlation [15] | 78.16        |
| Scenario-9        | Contrast, energy, correlation, entropy [16][17] | 82.05        |
| Scenario-10       | Energy, homogeneity, correlation, entropy [13]  | 79.43        |
| Scenario-11       | Contrast, energy, homogeneity, correlation, entropy [12] | 82.71        |
| Scenario-12       | Contrast, homogeneity, correlation, ASM, entropy | 82.80        |
| Scenario-13       | Contrast, energy, homogeneity, correlation, dissimilarity, entropy | 82.53        |
| Scenario-14       | Contrast, homogeneity, correlation, dissimilarity, ASM, entropy | 82.53        |
| Scenario-15       | Contrast, energy, homogeneity, correlation, dissimilarity, ASM, entropy | 81.57        |

This scenario aims to find the Haralick feature that gets the best accuracy results. Scenario-1 to Scenario-7 is a scenario to get the accuracy of each feature. Scenario 8 to scenario 11 is based on previous research. Scenario-12 is a combination of 4 (four) features that has the highest accuracy. Scenario-13 and Scenario-14 are based on the highest accuracy of a combination of 6 (six) features. Scenario-15 is a combination of 7 (seven) features. The Haralick feature tuning parameter results show that the highest accuracy is 82.80%, obtained by Scenario-12 with a combination of contrast, energy, homogeneity, correlation, ASM, and entropy features. It shows that 5 (five) features can produce good accuracy than 7 (seven) features.

### Table 2. Results of Color Histogram parameter tuning.

| Channel | Number of Bins | Accuracy (%) |
|---------|----------------|--------------|
| H       | 4              | 64.72        |
|         | 8              | 72.93        |
|         | 16             | 80.17        |
|         | 4              | 80.66        |
| S       | 8              | 91.00        |
|         | 16             | 93.41        |
|         | 4              | 85.98        |
| V       | 8              | 93.62        |
|         | 16             | 95.41        |
| H, S    | 16;16          | 95.59        |
| H, V    | 16;16          | 97.16        |
| S, V    | 16;16          | 98.12        |
| H, S, V | 16;16;16       | 98.08        |

The results show that each parameter of the H, S, and V channels with 16 bins gives better accuracy than using 4 and 8 bins. The tuning parameters' results with the best accuracy, with 98.12%, on 16 bins and a combination of S-V channels. Channel S is the percentage of white light added to the base color.
Therefore, light base colors have high saturation intensity, pastel colors have low saturation intensity, and monochrome colors (black and white) have no saturation. Channel V refers to the light intensity in the image. It can be said that the S-V channel has a significant impact because the human face image has a slight color variation so that information can be taken based on the level of color saturation (light-dark) and retrieve information based on the level of light intensity in the image. It was also stated in [18] that S and V channels have greater power to differentiate ethnicity.

3.1.3. Results of Combining GLCM Method and Color Histogram. In this study, the GLCM and Color Histogram methods are combined. This combination determines how much the performance increases if the two methods are used in the Indonesian ethnic recognition system. Besides, it also compares the use of the holistic features and the periorbital features of the two methods. Holistic facial features are features that are contained in the face as a whole. Periorbital features are features of the periocular region (the area around the eyes, including the eyebrows) to the nose. The best parameters from GLCM with Scenario-12 and Color Histogram with 16 bin and S-V channels were used. The result of combining feature extraction methods are given in Table 3.

| Features       | Accuracy (%) | Vector Length |
|----------------|--------------|---------------|
| Holistic       | 96.12        | 52            |
| Periorbital    | 98.73        | 52            |

Combining the GLCM method and the Color Histogram with a vector length of 20 + 32 = 52 obtained 96.12% accuracy for holistic facial features and 98.73% for periorbital features. The best performance is in the periorbital features. Thus, periorbital features are better than the holistic features of the face in ethnic recognition [6][7][19]. Table 5 shows that combining texture and color features can improve the ethnic recognition system's performance rather than only its texture or color features.

Table 4. Performance of the feature extraction method on periorbital features.

| Methods       | Accuracy (%) |
|---------------|--------------|
| GLCM          | 82.80        |
| Color Histogram| 98.12        |
| Combined      | 98.73        |

3.1.4. Results of Comparison of Classification Methods and Parameter Tuning Cross-Validation. In this case, we compare the classification method and cross-validation parameters. The proposed classifier is compared with the Support Vector Machine (SVM) [20]. Using 200 trees (n-estimators = 200) as Random Forest Parameters. Uses the values $10^{-1}$ for regularization, 1 for gamma, and the Polynomial kernel as SVM parameters. These parameters are obtained from hyperparameter tuning using Grid Search. Also, using $k = 10$ for the cross-validation parameter. The results of the comparison of classification methods are given in Table 5.

Table 5. Results of comparison of classification methods.

| Classifiers | Accuracy (%) |
|------------|--------------|
| Random Forest | 98.60        |
| SVM         | 92.27        |

The results show that the Random Forest classification method's accuracy gets the highest accuracy with 98.60%. It shows that Random Forest has better than SVM [20]. Furthermore, looking for the optimal k value in cross validation using the optimization method, namely the elbow method. This method uses an elbow as a cut-off point to select a point where the curve appears to be bent, particularly from high to low slopes, flat or close to flat, or in another direction. The elbow point is the optimal point of several decisions whose value is no longer increasing rapidly and is no longer proportional to further
(stable) increase [21]. Figure 6 shows that the $k$ value in cross-validation is in the range of 2 to 20 with each accuracy result. In the search for values using Random Forest and SVM, both elbows are at a point 6 (six) with an accuracy of 98.65% and 91.75%, respectively.

Figure 6. Graph of optimal $k$ values in cross-validation.

![Figure 6](image1)

Figure 7. Accuracy of each fold.

![Figure 7](image2)

3.2. Analysis of Test Results.

We conduct the testing using cross-validation on 2290 periorbital image by incorporating the optimal color and texture feature parameter. We use GLCM parameters, namely contrast, homogeneity, correlation, ASM, and entropy; S-V channel and 16 bins for Color Histogram; 200 trees for Random Forest. Based on optimal $k$ at point 6, the accuracy is 98.65%. This value is obtained from the calculation of the average accuracy of each iteration of the fold. The accuracy details for each fold are shown in Figure 7. The evaluation results for the overall performance are obtained from the average score for each iteration. These results were measured using accuracy, precision, recall, and f1-score [22], shown in Table 6.

| Table 6. Results of the overall system performance evaluation. |
|-------------------------------------------------------------|
|                               | Accuracy | Precision | Recall | F1-Score |
| Average (%)                 | 98.65%   | 98.66%    | 98.65% | 98.65%   |

Table 7. Euclidean distance feature comparison between misclassified data with actual and predictive classes.

| Table 7. Euclidean distance feature comparison between misclassified data with actual and predictive classes. |
|-------------------------------------------------------------|
|                               | GLCM Feature | Distance Average of | Combined Feature |
| Class                     |              |                      |                  |
| Actual                    | 90.35        | 37.636,24            | 37.636,52        |
| Predicted                 | 65.88        | 34.361,40            | 34.361,57        |

Misclassification in the recognition system is common. A misclassified image has features similar to the image features in the prediction class. To determine the feature similarity, we calculate the distance.
of the actual class and the predicted class using Euclidean Distance. For example, there is an image of the Javanese class, classified into the Sundanese class. Table 7 compares the distance of feature, the actual class feature, and the predicted class feature using the Euclidean Distance.

Table 7 shows that the Java class image classified as a Sundanese class has an average distance of the GLCM and Color Histogram features in the prediction class, which is smaller than the average distance of the GLCM and Color Histogram features in the actual class. Thus, the Javanese class is classified as a Sundanese class because the feature distance is closer to the Sundanese class. Also, the combined feature's mean distance is similar to the average distance of the Color Histogram feature. Thus, the Color Histogram feature is more dominant and affects the classification process.

4. Conclusion

Based on the results of testing and analysis, this study obtained an optimal accuracy of 98.65% with a precision of 98.66%, recall, and f1-score of 98.65%, respectively, in $k = 6$. The best parameters on GLCM are 5 (five) Haralick Features, namely contrast, homogeneity, correlation, ASM, and entropy. The best parameter on the Color Histogram is the S-V channel with 16 bins. The best parameter of Random Forest is 200 trees. The suggestion for further research is to expand the class of Indonesian ethnicity to several other classes.

References

[1] Momin H and Tapamo J R 2016 A comparative study of a face components based model of ethnic classification using gabor filters Appl. Math. Inf. Sci. 10 2255–65
[2] Li Z, Zhang Q, Duan X and Wang Y 2018 A novel semantic approach for multi-ethnic face recognition Int. J. Pattern Recognit. Artif. Intell. 32
[3] Huri K, Omid David E and Netanyahu N S 2018 DeepEthnic: multi-label ethnic classification from face images Lecture Notes in Computer Sci. (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) vol 11141 LNCS (Springer International Publishing) pp 604–12
[4] Dokhi M S, Theodora H S, Wulansari I Y H and Dwi W S N 2016 Analisis kearifan lokal ditinjau dari keragaman budaya PDSPK Kemendikbud RI 1–67
[5] Prasad P and Gupta A 2018 Moving object tracking and detection based on kalman filter and saliency mapping Advances in Intelligent Systems and Computing vol 542 pp 639–46
[6] Mohammad A S and Al-Ani J A 2017 Towards Ethnicity Detection Using Learning Based Classifiers 2017 9th Comput. Sci. Electron. Eng. Conf. CEEC 2017 - Proc. pp 219–24
[7] Rasyid H F, Ramadhani K N and Sthevanie F 2018 Mongoloid and non-mongoloid race classification from face image using local binary pattern feature extractions 2018 6th Int. Conf. on Information and Communication Technol., ICoICT 2018 (IEEE) pp 329–32
[8] Rehman A, Khan G, Siddiqi A, Khan A and Khan U G 2018 Modified texture features from histogram and gray level co-occurrence matrix of facial data for ethnicity detection 5th Int. Multi-Topic ICT Conf.: Technologies For Future Generations, IMTIC 2018 - Proc. (IEEE) pp 1–6
[9] Suyanto S, Arifianto A, Sirwan A and Rizaendra A P 2020 End-to-End speech recognition models for a low-resourced Indonesian Language 2020 8th Int. Conf. on Information and Communication Technol., ICoICT 2020
[10] Al-Kilidar S H S and George L E 2017 Texture recognition using co-occurrence matrix features and neural network J. Theor. Appl. Inf. Technol. 95 5949–61
[11] Abbas Z, Rehman M U, Najam S and Danish Rizvi S M 2019 An efficient gray-level co-occurrence matrix (GLCM) based approach towards classification of skin lesion Proc. - 2019 Amity Int. Conf. on Artificial Intelligence, AICAI 2019 (IEEE) pp 317–20
[12] Bianconi F, Smeraldi F, Abdollahyan M and Xiao P 2017 On the use of skin texture features for gender recognition: An experimental evaluation 2016 6th Int. Conf. on Image Processing Theory, Tools and Applications, IPTA 2016
[13] Nazir A, Ashraf R, Hamdani T and Ali N 2018 Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor 2018 Int. Conf. on Computing, Mathematics and Engineering Technologies pp 1–6

[14] Yadav S and Shukla S 2016 Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification Proc. - 6th Int. Adv. Comput. Conf. IACC 2016 pp 78–83

[15] Sukiman T S A, Suwilo S and Zarlis M 2019 Feature extraction method GLCM and LVQ in digital image-based face recognition Sink. J. dan Penelit. Tek. Inform. 4 pp 1–4

[16] Yahia S and Salem Y Ben 2016 3D face recognition using local binary pattern and grey level co-occurrence matrix 2016 17th Int. Conf. Sci. Tech. Autom. Control Comput. Eng. 1 pp 328–38

[17] Dixit U D and Shirdhonkar M S 2018 Face-based document image retrieval system Procedia Comput. Sci. 132 pp 659–68

[18] Du H, Salah S H and Ahmed H O 2014 A color and texture based multi-level fusion scheme for ethnicity identification Mobile Multimedia/Image Processing, Security, and Applications 2014 vol 9120 p 91200B

[19] Wang C, Zhang Q, Liu W, Liu Y and Miao L 2019 Facial feature discovery for ethnicity recognition Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 9

[20] Kremic E and Subasi A 2016 Performance of random forest and SVM in face recognition Int. Arab J. Inf. Technol. 13 pp 287–93

[21] Syakur M A, Khotimah B K, Rochman E M S and Satoto B D 2018 Integration k-means clustering method and elbow method for identification of the best customer profile cluster IOP Conf. Ser. Mater. Sci. Eng. 336

[22] Sokolova M, Japkowicz N and Szpakowicz S 2006 Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation AI 2006 Adv. Artif. Intell. 4304