ISSN 2355-0082

Ethnicity Classification Based on Facial Features Using Viola-Jones Algorithm

Irham Surya Pratama[1], Felix Indra Kurniadi[2]
[1, 2] School of Engineering and Technology: Informatics Engineering, Universistas Tanri Abeng, Jakarta, Indonesia
irham.surya@student.tau.ac.id
felixindra@tau.ac.id

Accepted on April 29, 2020
Approved on June 15, 2020

***Abstract*—Ethnicity information is an aspect of human identity. Ethnicity holds the same importance in human identity as gender or age. In this paper, we will propose a new approach to classify Indonesian Races, especially Western Indonesian (Java and Sumatra) and Eastern Indonesian (Papua and Nusa Tenggara). We implement two major steps for this research, the first step is face detection using the Viola-Jones Algorithm and the second step is the classification process using Support Vector Machine and K-Nearest Neighbour. Our proposed approaches give good results, with the best result is 81% using YCbCr color space.**

***Index Terms*—Ethnicity Classification, K-Nearest Neighbour, SVM, Viola-Jones Algorithm**

I. Introduction

The human has many properties such as ocular, periocular, gait, voice, and fingerprint—a Biometric system used for personal recognition, gender classification, and ethnicity classification/ ethnicity detection [1]. Three crucial aspects of soft biometric, especially in face regions, are ethnicity, gender, and age.

Ethnicity plays a significant role in biometric recognition. Classification of ethnicity has a great impact on surveillance, advertisement, and social media profiling. Even though many challenges in ethnicity classification, we cannot forget the importance of ethnicity classification for health care, educational, and socioeconomic status study [2].

In the current year, the use of demographic data such as ethnicity for a facial image has increased, especially in automatic detection using machine learning and computer vision approaches. The practical applications range from law enforcement and disaster victim identification [3].

Several works have tried to propose a new framework for ethnicity classification. Heng et al., implement a hybrid supervised learning using Convolutional Neural Network and Image ranking engine algorithm. The proposed method gives a significant improvement in overall accuracy [2].

Mohammad et al. proposed the Fusion of Local Binary Pattern and Histogram of Gradient (HOG). In this research, there are three crucial steps. The first step is preprocessing using the CLAHE filter before applying the dlib landmark to solved the difficulty of the light condition. The second step is extracting features using fusion LBP and HOG. Furthermore, the last step is the classification process using Support Vector Machine (SVM) and Multilayer Perceptron (MLP) [1].

Jilani et al. proposed geometric feature extraction with SVM as the classification process. They are also implementing dimensional reduction using Principal Component Analysis (PCA) and Partial Least Square Regression (PLR). This result of this research shows a promising result [3].

Achkar et al. proposed a different approach; they segment the face area using the Viola-Jones algorithm. This step was done to focus on the face. Furthermore, the next step is extracting features using the facial-geometric area. The last step is the classification process using Multi-Layer Perceptron [4].

According to previous research, especially Achkar et al., we proposed the same approaches as Achkar et al. The differences between our approach with Achkar et al. are in the feature extraction process, we used maximum likelihood estimation (MLE) of the images rather than using geometric facial features. Another difference is that our classification process was done using SVM because, according to the previous research, most of the research using SVM as the classifier process compared to MLP.

We are also using color images compared to grayscale images to our approach, and Based on Dhivakar [5] research on skin color recognition, the best color space for human color skin is YCbCr.

This paper separated into four sections. Section I is the introduction of the research statement and the background of the research. Section II will explain the proposed method more thoroughly. Section III will explain the result of the paper, and the last section will explain our conclusion toward our proposed method.

II. METHODOLOGY

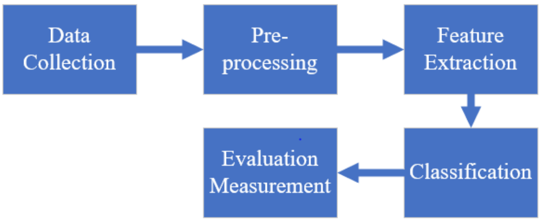

Fig. 1. Our Research Methodology

A. Data Collection

First, confirm that you have the correct template for your paper size. This template is for International Journal of New Media Technology (IJNMT). It has been tailored for output on the A4 paper size. Data was taken using a mobile phone camera with 8 Megapixels resolution. We did not use any camera filter, and we took the data in the same place to minimize any different images. We took 48 total images, which consists of 24 people from Java, and Sumatra region, and the other 24 people from the Nusa Tenggara and Papua region. The sample consists of three different angles: front face, side-left angle, and side-right angle.

TABLE I. DATASET

Class	Total
Western Indonesian (Java, and Sumatra region)	24
Eastern Indonesian (Nusa Tenggara and Papua region)	24
Total	48

B. Preprocessing

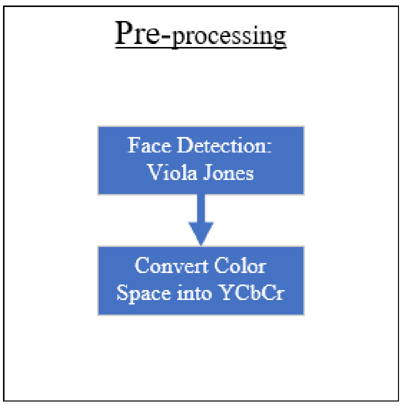

Fig. 2. Our Research Methodology

The first stage of the preprocessing is finding the face area using the Viola-Jones algorithm. The Viola-Jones algorithm consists of four-stage: The first stage is the Haar Feature, creating an integral image, AdaBoost and cascading classifiers. The Viola-Jones algorithm utilizes a Haar feature to represent the images. In the next part, after getting the Haar feature images, the Viola-Jones algorithm finds the integral image at location x, y, which contains the sum of the pixels.

$$I' = \sum_{x' \le x, y \le y'} I(x', y') \quad (1)$$

This phase was resulting in enormous rectangle features that can be computed in sub-windows. The next step is finding the set of features using Adaboost. After finding several weak features and then the Viola-Jones cascades the complex model into one classifier. The purpose of cascading is to reject non-face and face images [6]. Figure 2 represents the images after cropping using the Viola-Jones algorithm.

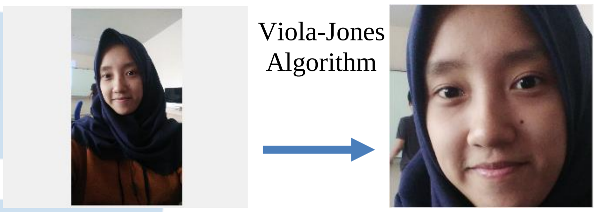

Fig. 3. Face Detection Using Viola-Jones Algorithm

The next phase is converting the RGB images into YCbCr color spaces. The converting images color images RGB is vital because the RGB is not suitable to perform any face detection or color detection type. We did not use RGB color space because of the character RGB colorspace. The character of RGB color space is merging between luminance and chrominance. Handling this issue, we implemented the YCbCr color space, which consists of Xerox YES color space [7].

$$Y = 0.253R + 0.684G + 0.063B \quad (2)$$

$$E = 0.5R - 0.5G \quad (3)$$

$$S = 0.25R + 0.25G - 0.5B \quad (4)$$

Where Y represents the luminance component and E, S represents chrominance components. R, G, B represent Red Green Blue color.

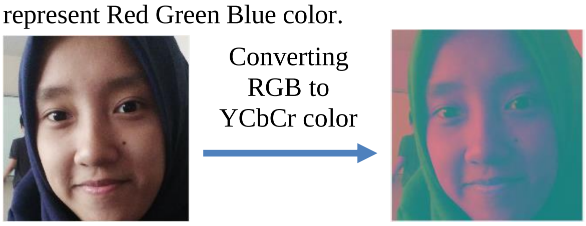

Fig. 4. Converting RGB into YCbCr Color Space

ISSN 2355-0082

C. Feature Extraction

In the feature extraction, we used the Maximum Likelihood Estimation (MLE) to extract the features of the images according to the likelihood function. The MLE selects the set of values of the images and maximizes the likelihood.

D. Classification Process

In the classification process, we implemented the Support Vector Machine (SVM) as a binary linear classifier. The algorithm classifies data based on the optimum hyperplane, which linearly separated the class. Given the training set:

$$\{(X_i, y_i)\}_{1 \leq i \leq n}, X_i \epsilon \mathbb{R}^s, y_i \epsilon \{+1, -1\} \tag{5}$$

SVM finds the hyperplane by solving this equation:

$$\begin{cases} \min_{w_b} \frac{1}{2} ||W||^2 \\ y_i(w.x + b) \end{cases} \tag{6}$$

For $i = 1,..n$ are the observation, w is the weight, and b is the bias. The {+1} represent Western ethnic Indonesia and {-1} represent Eastern ethnic Indonesia [3].

E. Evaluation Measurement

Finding our proposed method is good compared to other classifiers; we evaluated our result in accuracy, precision, and recall [8].

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} * 100\% \tag{7}$$

$$precision = \frac{tp}{tp + fp} * 100\% \tag{8}$$

$$recall = \frac{tp}{tp + fn} * 100\% \tag{9}$$

Where tp is true positive, fn is false negative, tn is true negative and fp is false positive. To understand the tp, fp, fn and tn, Figure 5 will explain more thoroughly.

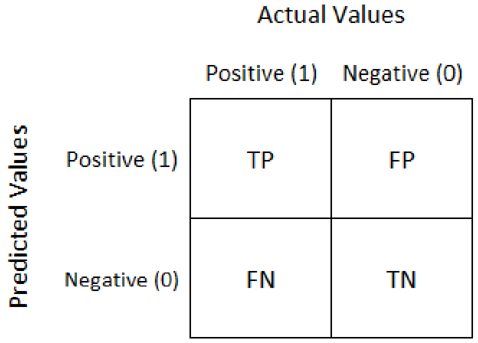

Fig. 5. Confusion Matrix [9]

III. Experiment Result

In this experiment, we compared two things our proposed method on RGB color spaces and YCbCr color spaces. We also compared three different Kernel in Support Vector Machine to find our dataset is linear, Gaussian, or radial using Linear Kernel, Gaussian Kernel, and Radial Basis Function (RBF Kernel). Another comparison is using K-Nearest Neighbour. We are also separate our dataset into a 70% training set and 30% for the test set. We separated the images manually.

Table 2 and Table 3 show the evaluation between RGB color spaces and YCbCr color space.

TABLE II. THE EVALUATION OF RGB COLOR SPACE

Classifier	Accuracy	Precision	Recall
Svm-Gaussian Kernel	50%	50%	100%
Svm -Linear Kernel	**75%**	**75%**	**75%**
Svm – RBF Kernel	50%	50%	100%

TABLE III. THE EVALUATION OF YCBCR COLOR SPACE

Classifier	Accuracy	Precision	Recall
Svm-Gaussian Kernel	50%	0%	0%
Svm -Linear Kernel	**81%**	**86%**	**75%**
Svm – RBF Kernel	50%	0%	0%

According to Table 2 and Table 3 results using YCbCr color space significantly improve our result, which proves that the RGB color space is not good enough to handle chrominance and luminance of the images. The result also shows that our dataset is linearly separable.

Table 4 and Table 5 show the comparison of SVM with KNN in each color space.

TABLE IV. THE COMPARISON CLASSIFIER IN RGB COLOR SPACE

Classifier	Accuracy	Precision	Recall
Svm – Linear Kernel	**75%**	**75%**	**75%**

KNN	63%	63%	63%

TABLE V. THE COMPARISON CLASSIFIER IN YCBCR COLOR SPACE

Classifier	Accuracy	Precision	Recall
Svm-Linear Kernel	**81%**	**86%**	**75%**
KNN	75%	83%	63%

Table 4 and Table 5 shows that the proposed classifier, which is SVM with Linear Kernel, has better accuracy, precision, and recall measure compared KNN.

IV. CONCLUSION

In this paper, we want to classify Indonesia's ethnicity into two different ethnicities (Western and Eastern Indonesia). We implement the Viola-Jones algorithm to separate between face and non-face from images, and we convert the colorspace from RGB into YCbCr color space. The next step is to extract the images into Maximum Likelihood Estimation (MLE) and creating a model using the SVM algorithm. The result shows that the YCbCr gives better results compared to RGB color spaces, and SVM gives better results compared to the KNN classifier. We also know our dataset is linear data, which is a good measure for the next experiment to improve the accuracy based on the linear approach.

REFERENCES

[1] A. S. Mohammad and J. A. Al-Ani, "Towards ethnicity detection using learning based classifiers," in *2017 9th Computer Science and Electronic Engineering (CEEC)*, Colchester, Sep. 2017, pp. 219–224, doi: 10.1109/CEEC.2017.8101628.

[2] Z. Heng, M. Dipu, and K.-H. Yap, "Hybrid Supervised Deep Learning for Ethnicity Classification using Face Images," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, Florence, May 2018, pp. 1–5, doi: 10.1109/ISCAS.2018.8351370.

[3] S. K. Jilani, H. Ugail, A. M. Bukar, A. Logan, and T. Munshi, "A Machine Learning Approach for Ethnic Classification: The British Pakistani Face," in *2017 International Conference on Cyberworlds (CW)*, Chester, Sep. 2017, pp. 170–173, doi: 10.1109/CW.2017.27.

[4] R. Achkar, G. Haidar, M. el Assal, D. Habchy, D. al Ashi, and T. Maylaa, "Ethnicity Recognition System using Back Propagation Algorithm of an MLP," in *2019 Fourth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, Beirut, Lebanon, Jul. 2019, pp. 1–5, doi: 10.1109/ACTEA.2019.8851071.

[5] B. Dhivakar, C. Sridevi, S. Selvakumar, and P. Guhan, "Face Detection and Recognition Using Skin Color," *Communication and Networking*, p. 7, 2015.

[6] M. Zervos, "Multi-camera face detection and recognition applied to people tracking," Ecole Polytechnique Fédérale de Lausanne, Switzerland, 2013.

[7] E. Saber and A. M. Tekalp, "Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions," *Pattern Recognition Letters*, vol. 19, no. 8, pp. 669–680, Jun. 1998, doi: 10.1016/S0167-8655(98)00044-0.

[8] J. Brownlee, "Classification Accuracy is not Enough Measures You Can See," 2014. https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/ (accessed Apr. 06, 2020).

[9] S. Narkhade, "Understanding Confusion Matrix," *Toward Datascience*, May 09, 2018. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62 (accessed May 12, 2020).