Journal of Engineering Science and Technology
Vol. 15, No. 4 (2020) 2149 - 2162
© School of Engineering, Taylor's University

# FACE DETECTION USING THE VIOLA JONES METHOD WITH SEGMENTATION OF SKIN COLOR ON FACE IMAGES

MUHAMMAD FACHRURROZI*, SAPARUDIN, KEREL KHALIF AFIF

Informatics Engineering Department, Faculty of Computer Science, Universitas Sriwijaya
Jalan Raya Palembang Prabumulih KM. 32 Indralaya, Ogan Ilir, Sumatera Selatan, Indonesia
*Corresponding Author: mfachrz@unsri.ac.id

**Abstract**

Face detection is a step used to look for faces in an image. The process of finding faces in this study uses the Viola-Jones method that detects faces in the color segmentation images. Skin color segmentation is used to remove information in images that have no potential as a face. Reducing unnecessary information on the image can improve the performance of face detection. One area that has potential as a face on the human body is skin color. Based on the test results with single face image data, the values of recall and precision were 94.11% and 99.83% and the accuracy of detection accuracy using Intersection over Union was 0.6963 with 6.14% faster processing time than the Viola-Jones only.

Keywords: Face detection, Intersection over Union, Skin color segmentation, Viola-Jones.

# 1. Introduction

Face detection is the first step in the process of recognizing and extracting facial features. Apart from faces, there are also more specialized detection processes such as iris detection, ear detection, and tongue detection [1]. But most research is face detection. Face detection aims to determine whether there are faces or not in an image. The face detection process is very important because good detection quality is very influential in subsequent processes such as facial feature extraction and use it in facial recognition processes [2]. Face detection results are needed for further processing because the face is one of the biological information that exists in humans in addition to fingerprints, eyes, and sounds [3].

Face detection consists of several processes. To detect faces, features are used as a detector. The features used are characteristics of an object such as color, shape, and texture [4]. Therefore, the first process starts from the training process on Haar-like features using AdaBoost learning on the Viola-Jones method. Then the feature is saved into a JSON file to be used as a detector in a face detection system. The second process, the system developed is carried out the testing process of the detectors that have been trained. This second process obtained the value of recall and precision. After that, the third process, the system will detect faces in an image. This process produces an accuracy of detection accuracy using Intersection over Union (IoU).

In the third process, the image used for detection is the image of skin color segmentation. The segmentation process is intended to simplify and change the presentation of the image to something to be more easily analysed, by removing image information that is not skin color. The image used for the process of skin color segmentation uses images with the YCbCr color space. YCbCr's color space is used because it has the best performance in detecting skin tones with lighting components and color components [5]. But, the YCbCr color space can also still detect colours that are not skin color (objects of the same color as skin color).

The use of skin color segmentation can make it easier for the Viola-Jones method to find facial objects faster because of calculations on one of the steps in the Viola-Jones method, namely the cascade classifiers. In this step, cascade classifiers can be eliminated faster and will not do any further calculations on the sub-window of skin color segmentation results.

This research contributes to the results of detection accuracy and processing time from face detection by the Viola-Jones method on the image of skin segmentation results to be used in further research.

# 2. Face Detection

## 2.1. Skin color segmentation

Skin color play an important role in detecting faces in color images [6]. The skin color segmentation is a segmentation process with a regional approach that works by analysing the color values of each pixel in the image by uses the thresholding method with the YCbCr color space in this study. Because segmentation is done at each pixel, the result of segmentation can produce noise at every pixel that is not skin color with a value still within the threshold.

The stages in skin color segmentation in the YCbCr color space can be seen as follows.

1. Determine the RGB image that is the object of detection and the threshold value used as the range of color values to be segmented (skin color).
2. Convert an RGB image into an YCbCr image.
3. Filter the colours of each pixel in the image based on the threshold value. If the pixel value falls within the range of the threshold value then the pixel is skin color, on the other hand, the pixel is categorized as a background by changing the pixel color to black.

## 2.2. Thresholding

Thresholding is one of the important techniques in image segmentation for classification of foreground and background objects [7]. The threshold value in this study is adjusted to the color space used, the YCbCr color space. The range of threshold values in the YCbCr color space is determined by the components Y (Luminance), Cb (Chroma blue), and Cr (Chroma red). The range of threshold values for skin color used in this study uses a range of values in one study [5]. The range of values for skin color in Y component is 0 - 255, Cb component is 134 - 172, and the Cr component is 78 - 126.

## 2.3. YCbCr

YCbCr color space is included in the color model used in television transmissions. In this color model, the luminance/lighting component is separated from the color component. Component (Y) represents luminance/lighting, and chrominance information is divided into two different color components. The Cb color component represents the difference between the blue component and the reference value, while Cr represents the difference between the red component and the reference value [5]. Because the Y component is separate from the color component, the YCbCr color space has a high level of skin color detection [8-9]. The skin color in the YCbCr color space can be seen in the face example in Fig. 1.

**Fig. 1. YCbCr image.**

## 2.4. Viola-Jones

Viola-Jones method is a method used in object detection which has a fairly high level of accuracy [10]. Some of the advantages of using the Viola-Jones method are as follows:

1. Robust - has a high detection rate and a low false-positive rate.
2. Real-time - For applications of at least two frames per second must be processed.
3. Face detection only (not recognition) - the goal is to separate the facial area and not the face.

Viola-Jones method combines four main keys namely Haar-like feature, integral image, AdaBoost learning, and Cascade classifier [10].

1. Haar-like features - difference between the sum of pixels in the dark area and sum of pixels in the light area as shown in Fig. 2.
2. Integral image - the technique of calculating feature values quickly by changing the value of each pixel into a new image representation.
3. AdaBoost learning - used to improve classification performance with simple learning to combine many weak classifiers into strong classifiers. A weak classifier is a correct answer with a level of truth that is less accurate [10].

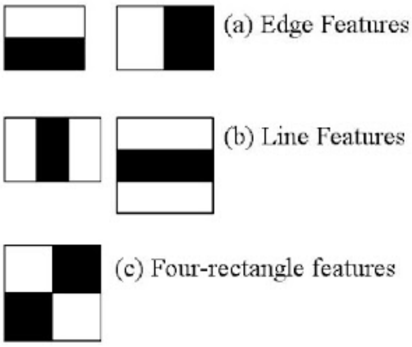


**Fig. 2. Haar-like features.**

A weak classifier is stated in the Eq. (1):

$$h_j(x) = \begin{cases} 1, P_jF_j(x) < P_jF_j(x) \\ 0, else \end{cases} \quad (1)$$

where *hj* is the value of a weak classification between 0 and 1 that meets the conditions.

Weak classifier with their weights combined to produce strong classifiers, Viola-Jones suggests the brute force method, which is to determine the weak classifier by evaluating each feature in all training data to find the feature with the best performance. But this is thought to be the cause of the length of the training procedure.

4. Cascade classifier - a method for combining complex classifiers in a multilevel structure that can increase the speed of object detection by focusing on the image area that has the opportunity. The combination of cascade classifier can increase detection rates with low false-positive rates [10]. Cascade classifier consists of several stages, each stage there are some weak classifiers and is shown in Fig. 3.

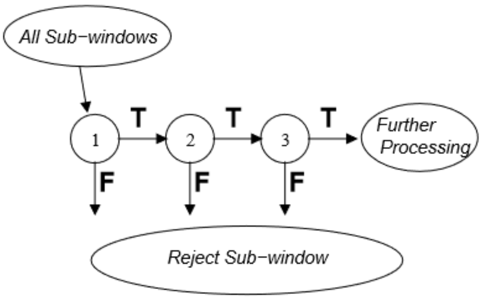


**Fig. 3. Cascade classifier.**

## 2.5. Intersection over Union

Intersection over Union is an evaluation method to measure the accuracy of the detection of an object. The application of Intersection over Union in the evaluation requires two bounding boxes [11].

1. The ground-truth bounding boxes (manual bounding boxes).
2. The predicted bounding boxes (bounding boxes from this research).

This evaluation method is used to find out how accurate the detection accuracy is obtained with the actual detection (right in the face area) as in Fig. 4.

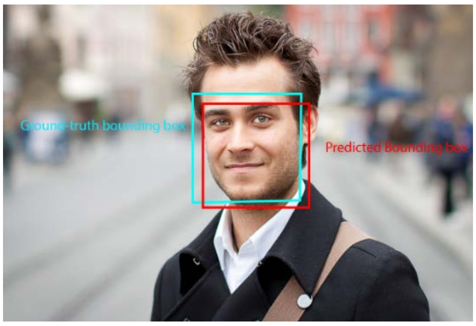


**Fig. 4. Intersection over Union face detection.**

The Intersection over Union value can be calculated by calculating the area values of the two bounding boxes. Calculating the values of the two bounding boxes can be seen in Eq. 2.

$$IoU = \frac{AoO}{AoU} \tag{2}$$

where *AoO* is the results of the total area of the incision in the bounding box and *AoU* is the result of the combined total area of the bounding box.

## 2.6. OpenCV

OpenCV is an open source computer vision and machine learning software ability. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception. OpenCV leans mostly towards real-time vision applications [12]. The implementation of Viola-Jones method is provided by OpenCV with a highly optimized code to run in real-time face detection. OpenCV method in face detection uses Haar-cascade models to detect faces in an image.

## 3. Methodology

Figure 5 is a diagram of the training process. Where this process consists of two sub processes, namely training classifiers and testing classifiers.

The classifiers training process is carried out to calculate the value of the best features and stored together with the weights on the features that have been trained, while the classifiers testing process is carried out testing of the testing data to get the value of recall and precision.

The steps in the training classifiers process are as follows:

1. Training data with positive and negative classes are input into the system.
2. All training data is processed into a new data representation using integral images.
3. Train all training data on each feature created with AdaBoost learning.
4. Every training one stage cascade classifier, the best feature data is stored.

The steps in the testing classifiers process are as follows:

1. Testing data with positive and negative classes are input into the system.
2. All test data is processed into new data representations using integral images.
3. Classify all testing data with cascade classifiers.
4. Showing test results in the form of recall and precision.

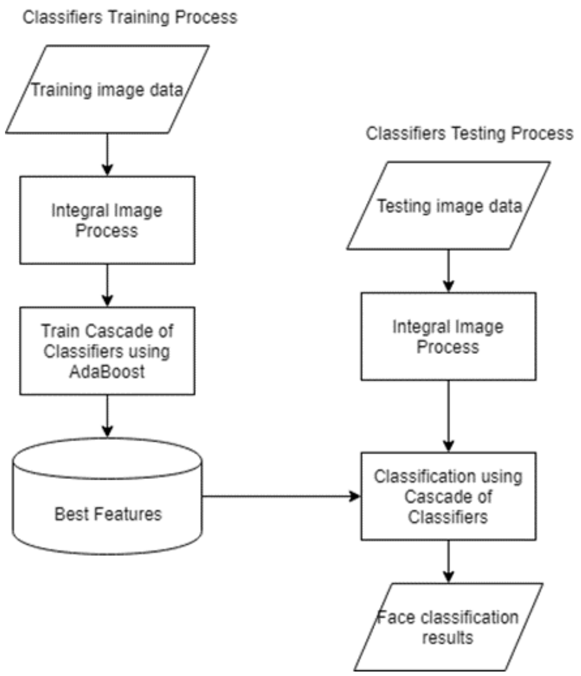


**Fig. 5. Stages of training process.**

Figure 6 is a diagram in the process of detecting faces and Intersection over Union testing in an image. The process is done by counting each sub-window of various sizes.

The steps in face detection in an image are as follows:

1. Single-face image input into a program.
2. Convert RGB image color space to YCbCr.
3. Segment skin tones on YCbCr images.
4. The results of segmentation are converted to grey color space.
5. The gray image is calculated integral image.
6. Perform the calculation process in each sub-window using cascade classifiers.
7. Combine overlapping face positions.
8. Draw a rectangle at each face position.
9. Input ground-truth bounding box
10. Calculate Intersection over Union value
11. Show face detection and Intersection over Union results.

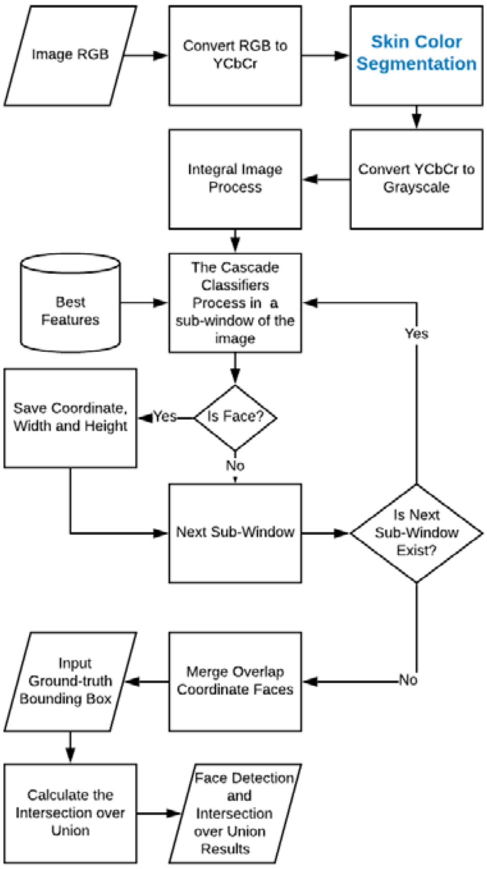


**Fig. 6. Face detection process.**

## 4. Implementation and Results

### 4.1. Training process

The training data used amounted to 9375 image data with 1176 positive classes and 8199 negative classes. Data in the form of images with grayscale color space and measuring 24 x 24 pixels as shown in Fig. 7.

While the features created have a minimum and maximum size of 4 and 10 (units of pixels). With these requirements, training data produces 44052 features that are formed. The program interface for data training can be seen in Fig. 8.


**Fig. 7. Data training.**

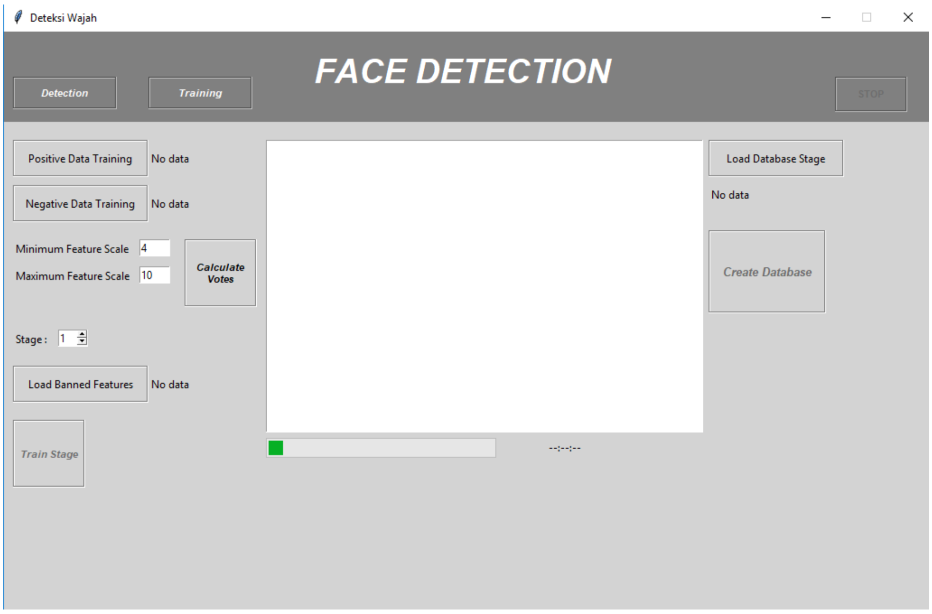


**Fig. 8. Interface of training process program.**

The training process is carried out using hardware with the following specifications:

1. AMD Quad-Core processor A12-9700P.
2. AMD Radeon R8 M445DX.
3. RAM 8 GB DDR4.
4. Hard Disk 1000 GB.

The training process produced 3612 of the best features with a total of 44 stage cascade classifiers. The training process time for 3612 features is 36 days.

## 4.2. Testing process

The face detection interface in the program can be seen in Fig. 9. In the interface, the testing process is carried out in this study.

The testing process for testing classifiers uses data testing with criteria such as training data with a total data of 21620 data with 628 positive classes and 20992 negative classes. All testing data is calculated using a detector in the form of Haar-like features that have been trained and then calculated to get the value of recall and precision using 4 attributes, namely True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The test results for the testing classifier are shown in the chart in Fig. 9 which shows the results of each stage of the testing data.

Based on the chart in Fig. 10, the recall and precision values at the last stage are 94.11% and 99.83%.

Table 1 shows the comparison of detection rates. The detection rates obtained in this study are high enough for application in the Viola-Jones method. The highest detection rate is obtained by the segmentation method and the watershed algorithm.

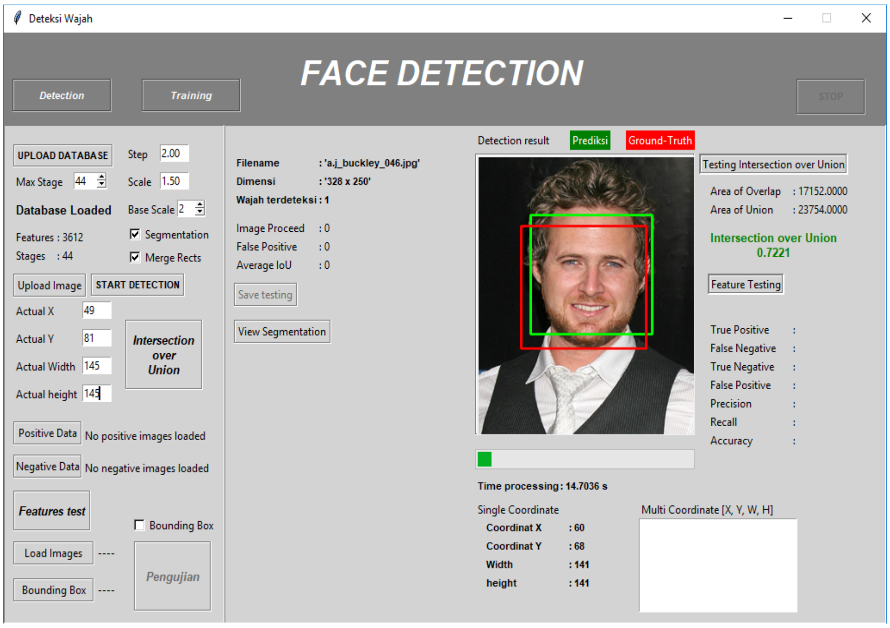


**Fig. 9. Face detection program.**

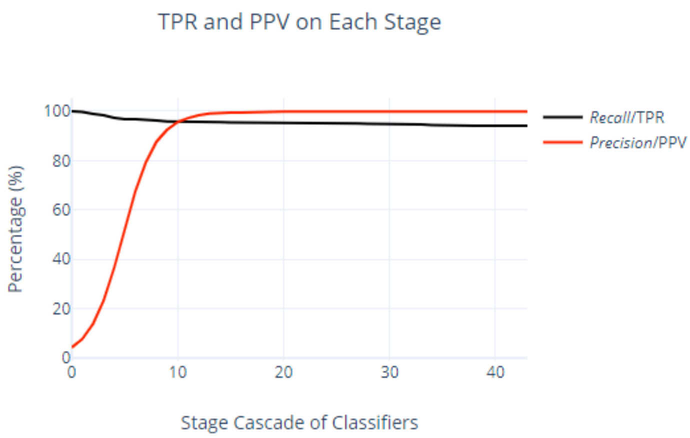


**Fig. 10. Recall and precision charts.**

**Table 1. Comparison of detection rates.**

| Previous research | Method | Detection rate [%] |
|---|---|---|
| **Shobana et al. [13]** | Skin segmentation Pre-Filter | 96.00% |
| **Basbrain et al. [14]** | Modification of Viola-Jones | 95.00% |
| **Rewar and Lenka [9]** | Skin Based Models (YCbCr) | 90.00% |
| **Kumar [15]** | Skin Based Models (YCbCr) | 80.00% |
| **Yadav and Nain [16]** | Hybrid Skin Color Model | 97.85 % |
| **Proposed Method** | Viola-Jones | 94.11 % |

The accuracy of the detection testing process uses 100 single-face images with RGB color space and various sizes. The testing process was carried out with three detection methods namely the detection method carried out in this study (Segmentation & Viola-Jones), Viola-Jones method only and *OpenCV*. These methods get a pretty good accuracy value of Intersection over Union with 0.6963 for segmentation and Viola-Jones, 0.6375 for Viola-Jones, and 0.7172 for *OpenCV*. Comparison of the value of Intersection over Union can be seen in Fig. 11.

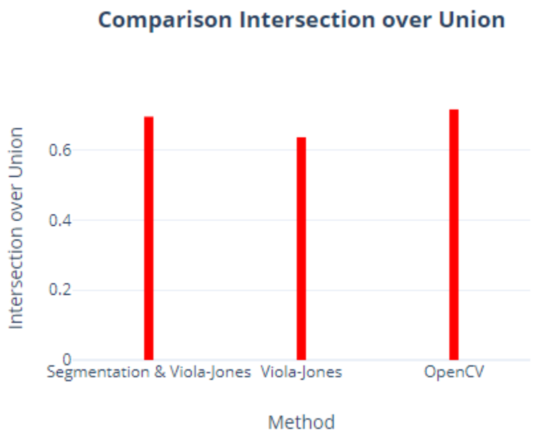


**Fig. 11. Comparison of Intersection over Union (IoU).**

In Fig. 11, it can be seen that the Viola-Jones method with skin color segmentation has a higher accuracy value than the Viola-Jones method alone. In addition to the accuracy of detection accuracy, detection of 100 images also obtained several results, namely Viola-Jones with segmentation can detect as many as 100 faces in image data correctly, Viola-Jones can detect 93 faces in image data correctly with 6 images not detected and 7 false positive, and *OpenCV* can detect as many as 99 faces in image data with 8 false positives.

The following are some images that detected false positive/undetected in Viola-Jones and OpenCV methods which can be seen in Figs. 12 and 13.

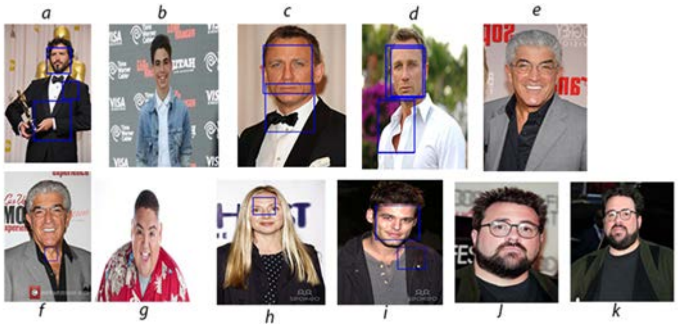

**Fig. 12. Viola-Jones results.**

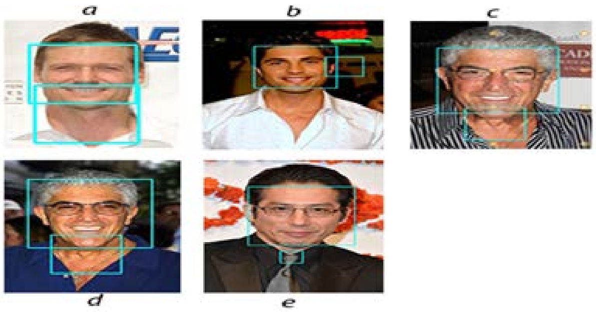

**Fig. 13. Image OpenCV results.**

In addition to the level of detection accuracy, the processing time is also obtained. Figure 14 shows the processing time of the Viola-Jones method reduced by the Viola-Jones processing time by skin color segmentation.

Based on Fig. 14, it was found that the detection process using Viola-Jones & segmentation get a faster time in processing. The percentage of time efficiency can be seen in Fig. 15.

In Fig. 15, the time efficiency obtained is quite large, which is 6.14% faster than the Viola-Jones method.

Based on the testing process that has been carried out, the recall and precision values are 94.11% and 99.83%, the accuracy of the detection accuracy of Intersection over Union (IoU) is 0.6963, and the time efficiency is 6.11% faster than the Viola-Jones method.


**Fig. 14. Viola Jones time against Viola Jones & segmentation.**

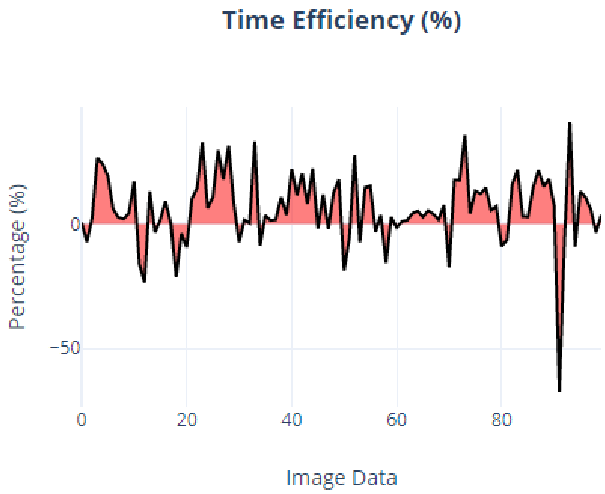


**Fig. 15. Time efficiency.**

## 5. Conclusions

From the results obtained from the testing process, it can be concluded that:

- The system can detect faces with high recall and precision values.
- Viola-Jones method and skin color segmentation get greater Intersection over Union accuracy value than the Viola-Jones method and faster processing time. Skin color segmentation reduces unnecessary information in the image which makes the process that occurs in the Viola-Jones method faster in the face classification process.
- Skin color segmentation images can improve accuracy and better performance with an Intersection over Union value of more than 0.5.

## Nomenclatures

| | |
|---|---|
| $F_j$ | Feature to j |
| FN | False negative |
| FP | False positive |
| $h_j$ | A weak classifier |
| $P_j$ | Parity to j |
| TN | True negative |
| TP | True positive |
| $x$ | Sub-image dimensions |

**Abbreviations**

| | |
|---|---|
| AOA | Area of Overlap |
| AOU | Area of Union |
| IOU | Intersection over Union |
| JSON | JavaScript Object Notation |
| RGB | Red, Green, Blue |
| YCbCr | Luminance, Chroma Blue, Chroma Red |

## References

1. Saparudin; Erwin; and Fachrurrozi, M. (2017). Tongue segmentation using active contour model. *Proceedings of 2017 IOP Conference Series: Materials Science and Engineering*. Palembang, Indonesia, 1-5.
2. Fachrurrozi, M.; Badillah, C.F.; Saparudin.; Erlina, J.; Erwin.; Mardiana.; and Lazuardi, A. (2017). The grouping of facial images using agglomerative hierarchical clustering to improve the CBIR based face recognition system. *Proceedings of 2017 International Conference on Data and Software Engineering (ICoDSE)*. Palembang, Indonesia, 1-6.
3. Fachrurrozi, M.; Erwin.; Saparudin.; and Mardiana. (2017). Multi-object face recognition using content based image retrieval (CBIR). *Proceedings of 2017 International Conference on Electrical Engineering and Computer Science (ICECOS)*. Palembang, Indonesia, 1-5.
4. Erwin.; Fachrurrozi, M.; Fiqih, A.; Bahardiansyah, R.S.; Algani, R.; and Primanita, A. (2017). Content based image retrieval for multi-objects fruits recognition using K-means and K-nearest neighbor. *Proceedings of 2017 International Conference on Data and Software Engineering (ICoDSE)*. Palembang, Indonesia, 1-6.
5. Lyon, D.; and Vincent, N. (2009). Interactive embedded face recognition. *Journal of Object Technology*, 8(1), 13-53.
6. Mohsin, H.; and Abdullah, S.H. (2018). Human face detection using skin color segmentation and morphological operations. *Journal of Al-Nisour University College*, 7, 63-80.
7. Senthilkumaran, N.; and Vaithegi, S. (2016). Image segmentation by using thresholding techniques for medical images. *Computer Science & Engineering: An International Journal*, 6(1), 1-13.
8. Kumar, A.; Kaur, A.; and Kumar, M. (2018). Face detection techniques: a review. *Artificial Intelligence Review*, 52, 927-948.

9. Rewar, E.; and Lenka, S.K. (2013). Comparative analysis of skin color based models for face detection. *Signal & Image Processing: An International Journal*, 4(2), 69-75.
10. Viola, P.; and Jones, M. (2001). Robust real-time object detection. *International Journal of Computer Vision*, 1-25.
11. Rosebrock, A. (2016). Intersection over union (IoU) for object detection. Retrieved July 10, 2019, from http://pyimagesearch.com.
12. Opencv (2019). About opencv. Retrieved November 13, 2019, from https://opencv.org/about/.
13. Shobana, L.; Yekkala, A.K.; and Eajaz, S. (2006). Face detection using skin segmentation as pre-filter. *Proceedings of International Conference on Advances in Pattern Recognition*. Kolkata, India, 51-55.
14. Basbrain, A.M.; Gan, J.Q.; and Clark, A. (2017). Accuracy enhancement of the Viola-Jones algorithm for thermal face detection. *Proceedings of International Conference on Intelligent Computing*. Liverpool, United Kingdom, 71-82.
15. Kumar, A. (2014). An empirical study of selection of the appropriate color space for skin detection: a case of face detection in color images. *Proceedings of 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*. Ghaziabad, India, 725-730.
16. Yadav, S.; and Nain, N. (2016). A novel approach for face detection using hybrid skin color model. *Journal of Reliable Intelligent Environments*, 2(2), 145-158.