

Mobile Application to Identify Indonesian Flowers on Android Platform

Tita Karlita, Achmad Basuki, Lakmi Makarti

Program Studi Teknik Informatika
Departemen Informatika dan Komputer
Politeknik Elektronika Negeri Surabaya
Kampus PENS, Jalan Raya ITS Sukolilo, Surabaya 60111
Tel: (031) 594 7280; Fax: (031) 594 6114
Email:tita@eepis-its.edu, basuki@eepis-its.edu, laxmimakarti@gmail.com

Abstract

Although many people love flowers, they do not know their name. Especially, many people do not recognize local flowers. To find the flower image, we can use search engine such as Google, but it does not give much help to find the name of local flower. Sometimes, Google cannot show the correct name of local flowers. This study proposes an application to identify Indonesian flowers that runs on the Android platform for easy use anywhere. Flower recognition is based on the color features using the Hue-Index, shape feature using Centroid Contour Distance (CCD), and the similarity measurement using Entropy calculations. The outputs of this application are information about inputted flower image including Latin name, local name, description, distribution and ecology. Based on tests performed on 44 types of flowers with 181 images in the database, the best similarity percentage is 97.72%. With this application, people will be expected to know more about Indonesia flowers.

Keywords: Indonesian flowers, android, hue-index, CCD, entropy.

1. INTRODUCTION

It can be found various flowers that are endemic or non-endemic in Indonesia. However, Indonesian people are lack the ability to recognize the flowers of their country. In fact, there are many books or websites that provide the information about Indonesian flowers. That kind of information source will be easy to use if the people want to know the information about the flowers that their names have been already known. Nonetheless, it will not be the same if they want to know the information about the flowers that they do not even know the name but they have the picture of the flowers. For example, when someone sees a flower on the side of the street, and he/she

wants to know the detail informations of that flower, it's difficult when he/she does not know its name.

Previously there have been some studies to create the software associated with the identification of flower, as it has been done by Sani and Yoga (2008). They used the moment of color as color feature and centroid-contour distance as shape feature to perform recognition on flowers[1]. Additionally, Andrianto (2008) also made a web applications for Image Retrieval to identify flowers with histogram intersection method and integral projection[3].

Nevertheless, both of those applications only shows output as group of images that similar to image input without providing information about the flower itself. To identify the Indonesian flowers, it needs some processes for increasing the recognition performance and classifying the major candidate to show the flower's name. Therefore, this study proposes an application to identify Indonesian flowers on the Android platform to make it easier when the user wants to take a picture of flower and searching for its information immediately.

2. RELATED WORKS

Content-Based Image Retrieval (CBIR) is a technique in image processing that usually used for image retrieval. On CBIR there are three features that widely used, they are colour features, shape features and texture features. In Image Retrieval, image features are extracted as vector and being saved in database. These vectors will be compared for similarity with features vector of input image. The similarity is measured by calculating the distance between input image features and image features on database using certain method [2].

Each flower has its own colour and shape. Therefore, this study performed flower identification based on colour and shape features of the flower. Rahmad et al (2008) has conducted a study of Hue-index, which is a component of the HSV colour model stating natural colour regardless of exposure. The results showed that the image recognition using histogram Hue-Index gives better accuracy than the RGB colour histogram[4].

Centroid-Contour Distance (CCD) method which is conducted by Sani and Yoga (2008) produces shape feature that calculates the distance R between centroid of a shape and the edges corresponding to the multiple limitations of the angle of α . In Figure 1, the distance R is the distance between point C (centroid) to the point P (the edges of the object) and the angle α is the angle formed between the adjacent two lines of R and have the same multiples for each angle formed. For example, if the value of α multiples eg 10° , there will be distance of 36 R (R_1, R_2, \dots, R_{36}) [1].

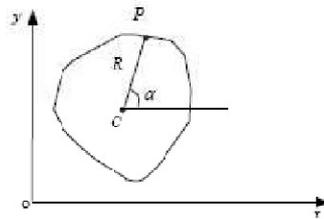


Figure 1. Centroid-Contour Distance approach[1]

3. ORIGINALITY

This study performs colour features extraction using Hue-Index as it was practiced by Rahmad and Basuki (2008). While the shape features was done with a slightly modified of CCD method performed by Sani and Yoga (2008). CCD method that is performed in this study is done by saving three-point range against the edge of the central point in every angle of θ . This is done because the flower image allows having more than one point on every angle of θ .

This application is able to produce a flower image search process to generate information about the name of the flower. Calculating the mode of flower appearance frequency in search results does it. The combination of colour histogram features and shape features are able to improve the searching performance.

4. SYSTEM DESIGN

This study aims to build application to identify Indonesian flowers on the Android platform with a combination of Hue-Index and Centroid-Contour Distance respectively represent colour features and shape features of flowers. Figure 2 is the flowchart diagram for this application.

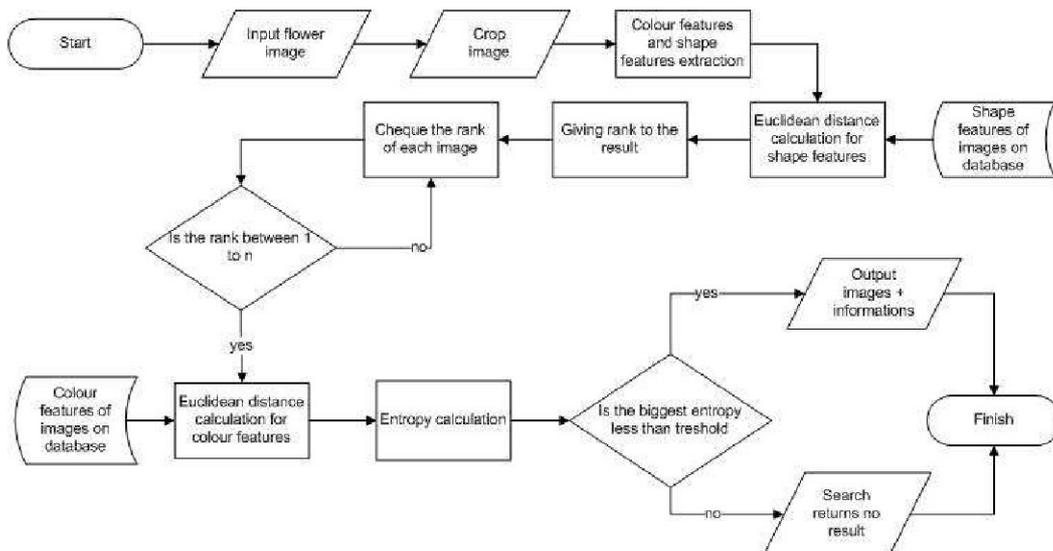


Figure 2. Flowchart diagram

Before the process of computation toward input image being performed, the image has to be cropped. This process is done to provide better results for taking only the foreground. Actually, programs can do this process automatically, but that way would slow down the process speed. Because of this matter, this study using the cropping feature that already provided by Android operating system to do the cropping process. Once the image is truncated, then the image is ready for identifying process including colour conversion, features extraction and image similarity measurement.

4.1. RGB to HSV Conversion

This process aims to convert the image color model of RGB color model to HSV color model. Once the image is inputted, color conversion is performed. Firstly, take the R, G, B value of each pixel of the image. After that, it can be converted to HSV color model using the formula below[6]:

$$\begin{aligned}
 H &= \arccos\left(\frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B))}}\right) \\
 S &= 1 - 3\frac{\min(R+G+B)}{R+G+B} \\
 V &= \frac{1}{3}(R+G+B)
 \end{aligned} \tag{1}$$

Where R is the red value, G the green value and B is the blue value of RGB colour model. While H is Hue, S is Saturation and V is Value or Intensity of HSV colour model.

4.2. Colour Features Extraction Using Hue-Index Method

Histogram is the number of occurrences of an event. In this paper, the event in question is the event of color appearance in the image. After the conversion from RGB to HSV colour, features are extracted from each pixel. Then histogram is made where each pixel contributes its Hue value in the calculation of the color histogram. For this application, the Saturation value and the Value / Intensity are not used because they do not represent the color of the image. Hue values from 0 to 360 so that the colour histogram obtained is one-dimensional array of 361 long data. For example, there is an image with 10x10 pixels of size each has a value of Hue as in Figure 3(a). Checking on each pixel and calculating the number of pixels with a certain value of Hue. Hue-Index values are shown in Figure 3(b).

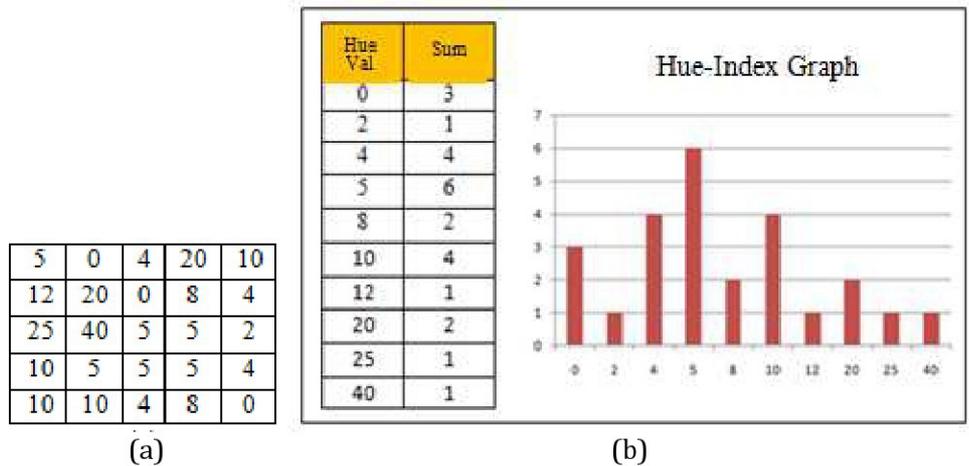


Figure 3. (a) Illustration of Hue value in each pixel of image (b) Hue-Index graph

4.3. Shape Features Extraction

Having obtained the vectors of colour feature then performed the extraction of shape features. Firstly, the image is edge detected to obtain object boundary. After that, Centroid-Contour Distance method is applied to get vectors of shape feature.

4.3.1. Edge Detection

Edge detection in an image is a process that results in the boundary of image objects. A point (x, y) is said to be the edge of an image when the point is having a high difference with its neighbors point. In this application, edge detection using Sobel method. This method takes the principle of laplacian and gaussian known as the function to generate the High Pass Filter. The advantages of this Sobel method are the ability to reduce noise before doing the calculation of edge detection. Sobel edge detection method multiplies each pixel of the image with vertical and horizontal Sobel kernel. Figure 4. is the result of edge detection using Sobel method.

$$\text{Vertical kernel} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{Horizontal kernel} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

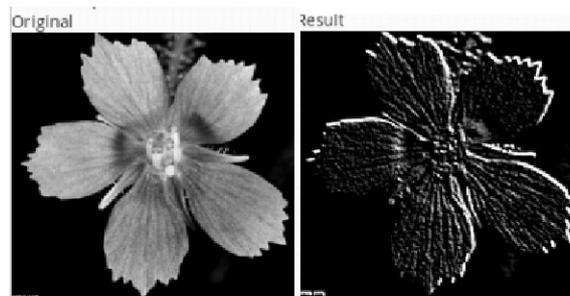


Figure 4. Edge detection using Sobel method

4.3.2 Centroid-Contour Distance

After the edge detection on the image, then performed calculations of CCD algorithm as follows:

1. Determine the polar coordinates for the point P (r, θ). Where r starts from one and the distance is increasing by 1 pixel, $r = r + 1$. Initial value of θ is set equal to 0° . Then used a multiple angle θ of 10° to 360° .
2. Calculate the coordinate position of the point P on each multiple angles θ . So that in every distance r there are 36 points of each angle θ . The formula below is used to determine the position of a point P.

$$\begin{aligned} X &= X_0 + r * \text{Cos}\theta \\ Y &= Y_0 + r * \text{Sin}\theta \end{aligned} \quad (2)$$

Where:

X = Position of point P on the X-axis

Y = Position of point P on the Y-axis

X_0 = Position of midpoint on the X-axis

Y_0 = Position of midpoint on the Y-axis

r = Multiple of distance ($r = 1, 2, 3, \dots, \text{et seq}$)

θ = Multiple of angle ($\theta = 10, 20, \dots, 360$)

Only then every position coordinates of the point P to be checked against its nearest neighboring pixels. If the point P and its neighbors are white, then the point P is an edge. When the point P is the edge of the object, then the value of r from that point will be saved as shape feature. Figure 5 shows an illustration of the CCD method calculation, which is applied to flower image that has been edges detected.

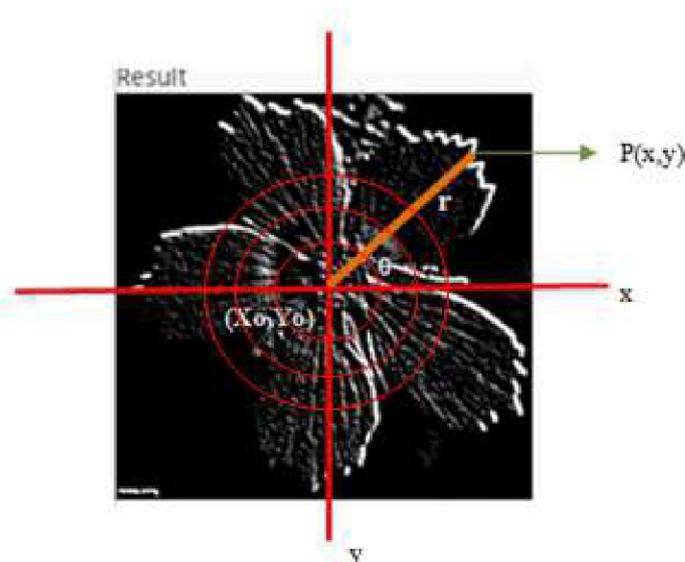


Figure 5. Illustration of CCD calculation.

4.3. Normalization

Normalization process is performed to convert the values of histogram value between 0 and 1. This is necessary so that the images that have different sizes but the same features still have the same histogram. Normalization obtained by the following formula.

$$P(Hue) = \frac{H(Hue)}{\text{Max}(H(Hue))} \quad (3)$$

Figure 6 is a graph of Hue-Index, which has been normalized. From the graph in Figure 6, it can be seen that the histogram is still the same, only the values are minimized.

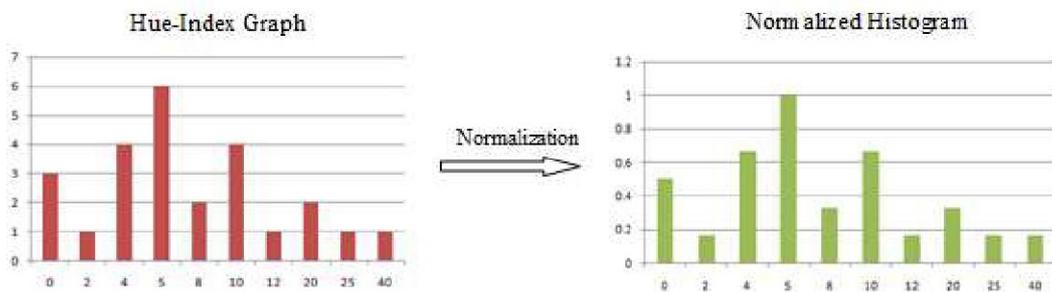


Figure 6. Normalized histogram graph.

4.4. The Distance of Image Similarity Calculation

This process calculates the distance of similarity between input image and the images in the database. The smaller the distance the image is more similar to the input image. Distance calculations in this application use Euclidean Distance. This method is implemented by the following formula.

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Where,

d = distance between two images using Euclidean Distance calculation

x_i = data number i in the input image histogram

y_i = data number i in the histogram of images in database

n = size of histogram data

4.5. Entropy Calculation

Taking the data, which is rank 1 to certain rank from the sorted Euclidian Distance calculation result, does entropy calculation. For example, it was taken the data rank 1 to 6. Of all the images that are considered to have potential as an output, the entropy value is calculated with the following formula.

$$\text{Entropy}(S) = -p_+ \log_2 p_+ - p_- \log_2 p_- \quad (5)$$

Where (p_+) is the sum of the average value of image group Euclidian distance that has positive solution or support on the input data. While (p_-) is the sum of the average value of image group Euclidian distance that have negative solution or does not support the input data.

5. EXPERIMENT AND ANALYSIS

There are two kinds of tests performed on this application. The first is an image similarity measurement testing. While the second is the flower identification test showed identification based on what features give better results. Trials conducted on 44 types of flowers with 181 images in the database. Each type of flower on average have 5 images with different exposure and positions.

5.1. Image Similarity Measurement Testing

The test is performed to determine how to measure the similarity between the input image and the output image. Image similarity calculation is done based on the features of each image. Due to the color feature and shape feature model has the same value that is between 0 and 1, so that the test is performed on one feature only that is color feature. Figure 7 is an image search result with the smallest Euclidean distance value based on the input image. Table 1 shows the result of image similarity measurement test that has been done.

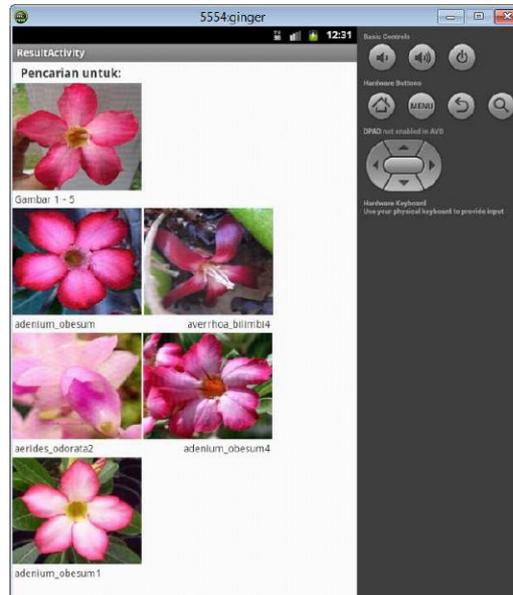


Figure 7. Search result of 5 most similar images

Table 1.Result of image similarity measurement test

Testing Method	Similarity (%)
Euclidean Distance	50%
Euclidean Distance + Entropi	93.18%

5.1.1. Euclidean Distance Method Testing

Identify the flowers based on the color features by measure the similarity using Euclidean distance method produces similarity percentage of 50%. It can be said that the image with the smallest Euclidean value is not always an appropriate output. This is made possible because of the lighting in images that affect the color of the flowers. Another possibility is that the flowers position also affects the formation of histogram features.

5.1.2. Entropy Method Testing

Looking at the test results of image similarity measurement using Euclidian distance, it turns out there is often more than one image corresponding to the input images are ranked from 1 to 5. As seen in Figure 7, pictures of flowers input, i.e. Adeniumobesum appears 3 times out of 5 output images. Based on this matter, the approach is to calculate the entropy value based on dataset of the top 5 ranked images. This strategy increases the similarity accuracy up to 93.18%. As already mentioned, one type of image had an average of 5 images with different positions and exposure. The pictures may not have the lowest value of Euclidean distance, but it will often appear at top rank for the corresponding input image.

5.2. Flower Identification Testing

The result that considered as the output is 1 type of flower that has the biggest entropy value and less than the specified threshold. This refers to the right information that appears based on flower images that is inputted. Tests carried out in 3 stages. The first test is based on color feature, the second test is based on shape feature and the last is a testing based on a combination of color features and shape features. Table 2 shows the similarity percentage of each test results.

Table 2.Flower identification testing result

Testing	Similarity Percentage Result
Colour features	93.18%
Shape features	9.09%
Colour features and Shape features	97.72%

5.2.1 ColorFeatures Testing

From the test results in Table 2, the test based on color features has the similarity level of 93.18%, which is pretty good. This is because the flower coloris very diverse. But there are still some errors due to similar flower color. Figure 8 shows one of the error outputs, where the two types of flowers have almost the same color.

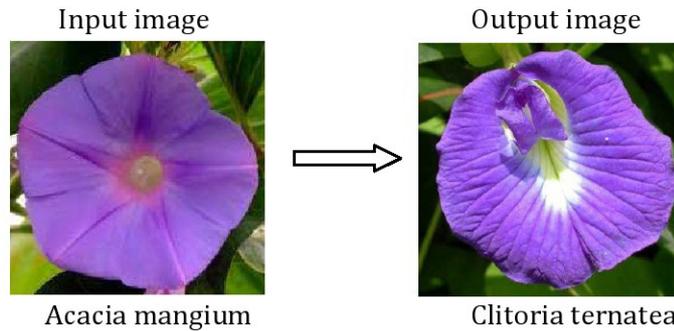


Figure 8. Error occurrence of identification based on colour features

5.2.2 Shape Features Testing

While the results of trials using shape features only produced similarity percentage of 9.09%. As we know most of the flower does have a similar shape. Thus it is difficult to detect flowers only based on its shape. However, input image of flower for testing by shape features was still being ranked 1 to 9. Out of the 44 types of images tested, only 1 image that is ranked more than 9.

5.2.3 Combination of Shape Features and ColorFeatures Testing

From the above two results, it is performed identified based on shape features first, and then taking the flowers that have rank of 1 to 9. It is intended to reduce the range for next search. Identification based on color features only performed on this 9 images. Shape features are done first so they do not ruin the good similarity result by colour features. In this way, the level of flowers similarity increased up to 97.72%. Flowers in testing based on color feature that gives false output can be resolved by doing shape features identification prior to color features. Some flowers that do not produce proper output due to the flower was ranked more than 9 when identified by shape features, so it cannot be detected for the identification by colour features. Figure 9 shows the display of output result.

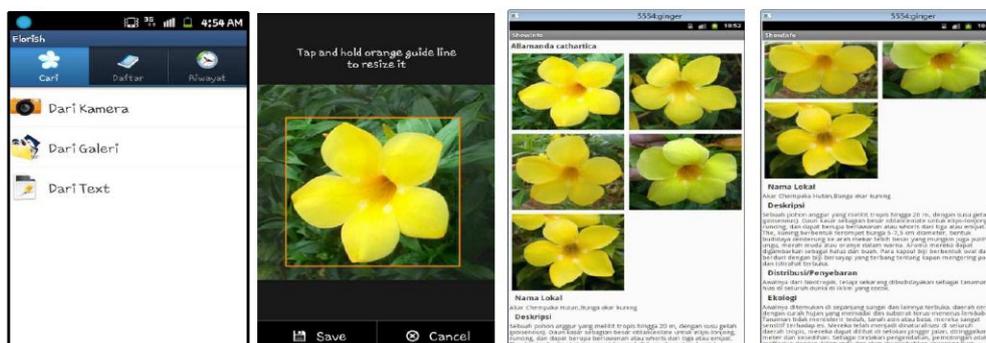


Figure 9. Display of output result

6. CONCLUSION

From the study and the results of trials that have been performed, it can be concluded that:

1. Calculation of the entropy of the dataset derived from Euclidean distance calculation with a range of rank 1 to a certain rank can improve the accuracy of similarity between the input image and the output image.
2. The use of shape features with CCD method helps to improve the percentage of similarity that is produced by the colour feature using Hue-index method.
3. The accuracy of the output may vary depends on the position of the image, the size of database, the number of rank taken for the calculation of entropy, and the crop process.
4. Based on tests performed on 44 types of flowers with 181 images in the database, the similarity percentage between the input image and output image of the best at 97.72%.

The weakness of this application is that the accuracy may vary due to output images that are not in the top rank when identifying by shape features will not be detected as an output even though in terms of identification by colour features it appears as the correct output. In addition, because of the process of cropping images was manually done by the user through the features of the Android, the possibility of error when cutting an image that would cause the output does not give suitable results.

For future work, it would be good if this application is developed with the addition of the cutting process images programmatically, but it does not slow down the search computing process. Also, the shape features can be improved or given the additional feature that is texture features.

REFERENCES

- [1] S. Muhamad Isa and Yoga Pradana, **Flower Image Retrieval Berdasarkan Color Moments, Centroid-Contour Distance dan Angle Code Histogram**, Universitas Tarumanagara, Jakarta, 2008.
- [2] V. N. Gudivada and V. V. Raghavan, **Special issue on content-based image retrieval systems** - guest eds, IEEE Computer. 28(9), 1995.
- [3] Andrianto Susilo, **Web Image Retrieval untuk identifikasi Bunga dengan Pengelompokan Content Menggunakan Ciri Warna dan Bentuk**, EEPIS, 2008
- [4] Cahya Rahmad, Achmad Basuki, **Image Retrieval Menggunakan Hue Index Histogram**, Jurnal Teknologi & Manajemen Informatika, Vol. 6 No. 2, Mei 2008.
- [5] S. Sural, G. Qian and S. Pramanik, **Segmentation and Histogram Generation Using the HSV Color Space for Image Retrieval**, Michigan State University, USA, 2002.
- [6] Achmad Basuki, **Aplikasi Pengolahan Citra: Deteksi Warna**, Politeknik Elektronika Negeri Surabaya, 2006.