

## Region Based Image Retrieval Using Ratio of Proportional Overlapping Object

Agus Zainal Arifin, Rizka Wakhidatus Sholikhah\*, Dimas Fanny H. P., Dini Adni Navastara

Informatics Department, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember,  
Jl. Raya ITS Kampus Sukolilo, Surabaya, 60111, Indonesia

\*Corresponding author, e-mail: rizka.wakhidatus15@mhs.if.its.ac.id

### Abstract

*In Region Based Image Retrieval (RBIR), determination of the relevant block in query region is based on the percentage of image objects that overlap with each sub-block. But in some images, the sizes of relevant objects are small. It may cause the object to be ignored in determining the relevant sub-blocks. Therefore, in this study we proposed a system of RBIR based on the percentage of proportional objects that overlap with sub-blocks. Each sub-block is selected as a query region. The color and texture features of the query region will be extracted by using HSV histogram and Local Binary Pattern (LBP), respectively. We also used shape as global feature by applying invariant moment as descriptor. Experimental results show that the proposed method has average precision with 74%.*

**Keywords:** Percentage overlapping object, RBIR, Invariant moment, HSV, LBP

Copyright © 2016 Universitas Ahmad Dahlan. All rights reserved.

### 1. Introduction

In the last few decades, Content Based Image Retrieval (CBIR) has become popular research. CBIR is one of the techniques to retrieve image from large image databases based on their visual similarity [1]. Searching images using their content has many advantages than using their annotation text, because not every images have annotation and not every annotations can resemble the images well. Therefore, CBIR is able to overcome the weakness of text-based image retrieval method.

Query by Example (QBE) is one of the query techniques in CBIR system that gives an example image as query. The features of query will be extracted and compared with features from image databases. The feature can be divided as global feature and local feature. Global feature uses overall image to extract their feature without consider the users interest [2, 3]. While the local feature only uses part of image that users need. Some of the researches find that using local feature based on their region is more effective to satisfy what users require [1], [4-6]. Local feature extraction from query image based on region is known as Region Based Image Retrieval (RBIR).

In RBIR, some of the regions are created in query image. Not every region is relevant to determine the user interest. Therefore users have to define the Region of Interest (ROI) in image query so that the irrelevant region can be eliminated. ROI chosen by the user is more capable to express user interest, but it becomes less effective if users have to deal with a lot of query. Another approach is creating ROI by system [1, [4]. The method divides image into sub-blocks in certain size (e.g. 3x3) than determine ROI from every sub-blocks which is overlap with the object [1]. The object is obtained from segmenting between foreground and background. Another research to automatically defined ROI is using Region Important Index (RII) and Saliency Region Overlapping Block (SROB) [4]. However if the image content small object, sometimes the sub-block cannot be detected and become irrelevant sub-block. That can lead to define wrong ROI or selected region. Therefore it is necessary to have a method that can be adapted well in the different size of object.

In this paper, we proposed system RBIR based on the percentage of proportional objects that overlap with sub-block to determine selected region. To find the similarity, we use color and texture as local feature and shape as global feature. This method is expected to improve image relevancy compared to existing methods.

## 2. Research Method

The Wang's image datasets are used in this paper (can be downloaded in <http://wang.ist.psu.edu/docs/related/>). This dataset usually used to measure performance of image retrieval method. There are 1000 images, which consists of 10 categories. Each category has 100 images. Step by step of proposed method as we can see in Figure 1 will be described in the next subsection.

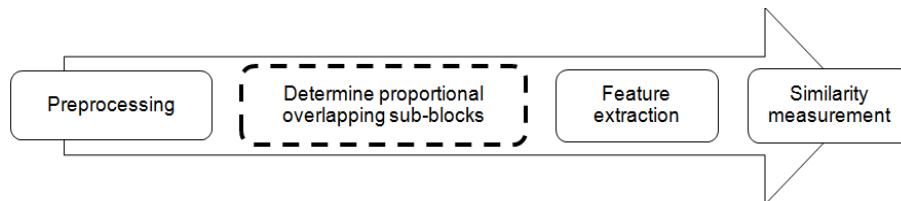


Figure 1. Design System

### 2.1. Preprocessing

The first step, image query is blurred by Gaussian filter to reduce noise. Then convert the result image to gray scale. In this research, segmentation is done by using the edge descriptor. The use of edge descriptor to simplify the process of segmentation has also been done in previous studies [1]. Sobel filter is used to extract the proper edge of objects from the gray scale image. This stage produces black and white image with edges of objects, the result are shown in Figure 2.

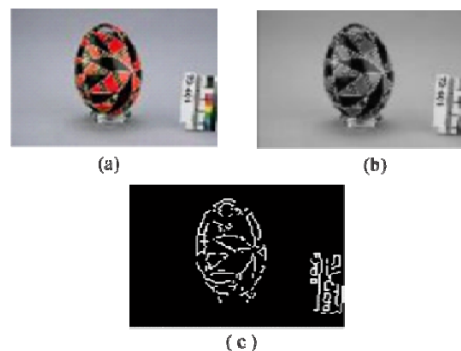


Figure 2. Preprocessing: Sobel filter, (a) Original image, (b) gray scale image and (c) image after applying Sobel filter

However, there are still any gaps between edges in the same object. To overcome this issue, dilation is used, so that the gap between edges can be reduced. To perform dilation, line-shape of structuring element which is 5 pixels wide is used. Due to space between edges that have vary in position, we use four kind of line structuring element with different angels, i.e.  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ . After dilation process, filling are perform to get fully segmented image.

Dilation process by applying 4 structuring element has side effect in object size. Compare to original object, the result has bigger object than the original image. Erosion using circle with size 3 as structuring element are performed to minimize those effects as shown in Figure 3.

### 2.2. Determine Proportional Overlapping Sub-blocks

To determine the region as the query, image query is divided into fixed size  $n \times n$ . In this paper we use  $3 \times 3$  as shown in Figure 4. From the previous research [4, 6], the best size to divide images is  $5 \times 5$  but has higher computation than  $3 \times 3$ .

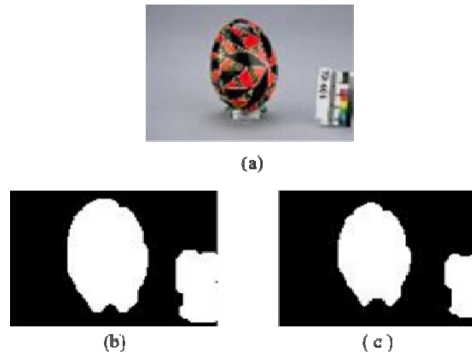


Figure 3. Erosion using circle, (a) Original image, (b) Image after dilation and filling and (c) Image after erosion

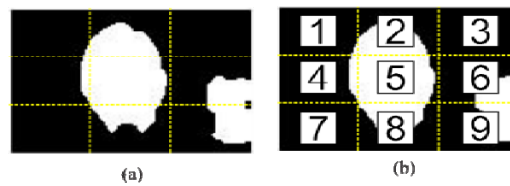


Figure 4. Determination region as query, (a) Divided image into fixed size 3 x 3, (b) Give the identity number in each sub-blocks

Commonly, there is relevant and irrelevant sub-block in an image. The irrelevant sub-block will be eliminated, because they do not have information to describe what user needs and can cause error in image retrieval. To determine whether the sub-block is relevant or not, our proposed method are used.

Black and white image that already segmented between object and background, than label are given in every object. However, not all images can be segmented well. There are some images that still contain noise. For example if the background image is not homogenous, sometimes in segmentation process a part of background can be detected as object. So filtering relevant object of the image is necessary in this step. The idea is based on assumption that if the detected object has smaller area than average area of object in an image and its location is far from the center of image, then it can be classified as noise or irrelevant object. By calculating area of object and minimum distance between every pixel in object and center of image, we can determine the relevancy of object. The distance between pixels  $(P_x, P_y)$  in object  $i$  and center of image  $(C_x, C_y)$  are done by using Euclidean distance as shown in Equation 1.

$$d(i) = \arg \min (d(P, C)) \quad (1)$$

Where  $d(P, C)$  can be calculated from Equation 2.

$$d(P, C) = \sqrt{(C_x - P_x)^2 + (C_y - P_y)^2} \quad (2)$$

Area in every objects is calculated based on their label, for example if the image contains two objects the result will show two area size. Then we combine the result between area and pixel distance of every objects to the center of image. From image  $I$ , the object  $i=(1,2,3,\dots, k)$  having area  $L_i$  more than  $\alpha$  and distance  $d(i)$  less than  $\beta$ , can be determined as relevant object. Where  $\alpha$  can be calculated by Equation 3 and  $\beta$  has a value 0.4.

$$\alpha = (\sum_{i=1}^k L_i) / 5 \quad (3)$$

After obtaining the relevant object, the next step is compute the percentage of object area in every sub-blocks. From area  $Lo_{i,j}$  that describe as area object  $L_i$  overlap with sub-block  $b_j$ , the ratio of proportional area overlapping object  $\tau_{(i,j)}$  can be calculated by Equation 4.

$$\tau_{(i,j)} = \frac{Lo_{i,j}}{L_i} \quad (4)$$

This method used threshold  $\tau_{(i,j)} = 0.1$ . So that if sub-block contains area overlapping object more than 0.1 it can be indicated as relevant sub-block. By using the ratio of area overlapping object, the relevant object that has small size can still be detected. Every relevant sub-blocks will be included into list of query sub-block that we called Saliency Region Overlapping Block (SROB). For example in Figure 4, the SROB of image query is sub-block 2, 5, 6, 8 and 9. Then the selected query sub-block is used in the next step to extract local feature of query image.

### 2.3. Feature Extraction

Local feature and global feature are combined to get better retrieval results. Local feature of image query and image database are extracted based on the selected sub-blocks. This research, uses color and texture as local feature, while global feature uses shape of object as descriptor.

#### 2.3.1. Color

Colors are commonly used as feature descriptor, because naturally human visualization can easily distinguish image by its color. In this paper, we use HSV histogram to extract color feature. HSV is chosen because this model can be superior color space, compare to RGB [7]. HSV color model divides color into 3 component, H (Hue), S (saturation), and V (Value).

The Hue component represents the type of color e.g. red, yellow, green, etc. Hue are described by specific position in color wheel, red starting at 0 degrees, green at 120 degrees and blue at 240 degrees. Complementary colours are in-between: yellow is at 60 degrees, cyan is at 180 degrees, and magenta is at 300 degrees. Saturation component represents how white the color is. Saturation has value between 0 and 1. The color shows pure color when the saturation value is 1, and will be diluted by white when saturation is decreased. Value component also describe as brightness. This component measure how black the color is. When value component are decrease, the blackness of the color are increased.

Every component of HSV model can be computed from RGB (Red, Green, Blue) model by calculating maximum RGB  $M$ , minimum RGB  $m$ , and delta  $d$  between  $M$  and  $m$  as shown in Equation 5-10 [7].

From this purpose, color is extracted from every selected sub-blocks in image query and image database. The feature is represented by histogram from each component (H, S, V). Than the value is normalized between 0 and 1.

$$M = \max (R, G, B) \quad (5)$$

$$m = \min (R, G, B) \quad (6)$$

$$d = M - m \quad (7)$$

$$S = \frac{d}{M} \quad (8)$$

$$V = M \quad (9)$$

$$H = \frac{1}{6} \cdot \begin{cases} \frac{G-B}{d}, & \text{if } M = R \text{ and } m = B \\ 1 + \frac{R-B}{d}, & \text{if } M = G \text{ and } m = B \\ 2 + \frac{B-R}{d}, & \text{if } M = G \text{ and } m = R \\ 3 + \frac{G-R}{d}, & \text{if } M = B \text{ and } m = R \\ 4 + \frac{R-G}{d}, & \text{if } M = B \text{ and } m = G \\ 5 + \frac{B-G}{d}, & \text{if } M = R \text{ and } m = G \end{cases} \quad (10)$$

### 2.3.2. Texture

Texture has an important role to describe the surface of an object and its relationship with the surrounding area [4]. Local Binary Pattern (LBP) [8] is known as good texture descriptor with higher performance result. Some of the applications are used LBP as descriptor [4], [6], [9].

LBP describes a pixel value based on its neighbor gray level pixel. Given a central pixel  $c$  that has gray value  $g_c$  and gray value from its neighbours  $g_p$ , the LBP can be calculated using Equation 11.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (11)$$

$P$  is the total number of neighbours and  $R$  is radius of the neighbourhood. After the LBP of each pixel in selected sub-block are calculated, then the histogram is created to represent the texture in every selected sub-block.

### 2.3.3. Shape

Moment invariants are important shape descriptors in computer vision. There are two types of shapes descriptors: contour-based shape descriptors and region-based shape descriptors. Regularly, in the most popular type invariant moment used is contour-based shape descriptor [10].

Hu invariant moments is basic to measure similarity between the template or database image. Hu's Seven Moments Invariants are invariant under translation, changes in scale, and also rotation. So it describes the image despite of its location, size, and rotation. In this research we use invariant moment to extract the shape feature. The seven mome invariant are used based on normalized central moments [11]. They are useful for image scaling, translation and rotation. To calculate them, formula in Equation 12 can be used.

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} + \eta_{12})^2 + (3\eta_{21} - \mu_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \mu_{03})^2 \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} - 3\eta_{03})[(3(\eta_{30} - \eta_{12})^2 - (\eta_{21} + \mu_{03})^2)] \\ \phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \mu_{03})^2)] \end{aligned} \quad (12)$$

Translation invariance can be achieved by shifting the polynomial basis to the object centroid as sown in Equation 13.

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q I(x, y) dx dy \quad (13)$$

In Equation 13, image described as  $I(x, y)$  as a piecewise continuous bounded function, variable  $p$  and  $q$  is positive integer [10]. Where  $x_c = \frac{m_{10}}{m_{00}}$ ,  $y_c = \frac{m_{01}}{m_{00}}$  are the coordinates of the object centroid. Scaling invariants is obtained by normalization of each moment. For the scale

invariants the formula is represented as Equation 14. The moment  $\eta_{pq}$  can be computed using the centroid of the image  $I(x,y)$  that equivalent to the  $m_{pq}$  whose center has been shifted to centroid of the image.

$$\eta_{pq} = \frac{m_{pq}}{m_{pq}^\gamma}, \quad \gamma = \left\lceil \frac{(p+q)}{2} \right\rceil + 1 \quad (14)$$

$$p, q = 0, 1, 2, 3, \dots$$

#### 2.4. Similarity Measurement

Similarity between image query and image database is computed by using Euclidean distance for every feature. For local feature, the distance is calculate for each selected sub-blocks of query and image database. Assume that image query has  $n$  selected sub-blocks. Selected sub-blocks query are represented by  $b_q = \{b_1, b_2, \dots, b_n\}$  and  $b_{img} = \{b_1, b_2, \dots, b_n\}$  represented as selected sub-blocks in image database. The distance only compute in the same sub-blocks index. For example, to find distance in sub-block 1, the euclidean distance is performed between  $b_{q,1}$  and  $b_{img,1}$ . So that the distance between sub-block query and sub-block image database  $d(b_{q,i}, b_{img,i})$  can be calculate using Equation 15.

$$d(b_{q,i}, b_{img,i}) = \sqrt{(fq_i - fimg_i)^2} \quad (15)$$

The final distance of local feature is described as average distance for every sub-blocks as shown in Equation 16.

$$d(b_q, b_{img}) = \frac{\sum_{i=1}^n d(b_{q,i}, b_{img,i})}{n} \quad (16)$$

For global feature, the similarity is calculated by applying Euclidean distance between image query and image database directly. The total distance from image query and image database is a combination of three feature distances. For every feature, we assign weight that are multiplied by distance value. Given weight of feature color  $w_c$ , weight of feature texture  $w_t$ , and weight of feature shape  $w_s$ , the total distance can be calculate by Equation 17.

$$d(q, img) = w_c \cdot d_c + w_t \cdot d_t + w_s \cdot d_s \quad (17)$$

In this research the optimal distance that we used is 0.1, 0.4, 0.5 for  $w_c$ ,  $w_t$ ,  $w_s$  respectively. After obtaining all the distance to image database, then performed the squencing by ascending order. The smallest the distance, indicate that two images has higher similarity.

### 3. Results and Discussion

#### 3.1. Results

##### 1. Performance Measure Using Precision-Recall

The proposed method is evaluated using precision. Precision is commonly used to evaluate the performance of IR system. Based on Table 1, precision can be calculated by Equation 18.

$$precision = \frac{TP}{TP+FP} \quad (18)$$

Where  $TP$  (True Positive) shows the number of relevant images that can be retrieved by system,  $TN$  (True Negative) is the number of relevant images that cannot be retrieved by system,  $FP$  (False Positive) is the number of irrelevant images that system retrieve and  $FN$  (False Negative) is the number of irrelevant images that doesn't retrieve by the system.

For each query, the results are shown based on their ranking. The ranking are compute from the distance between query and image in database. Table 2 shows precision with a different  $k$ , where  $k$  describe as number of retrieval images.

Table 1. Confusion Matrix

	Relevant	Not relevant
Retrieve	TP	FP
Unretrieve	TN	FN

Table 2. Precision in Variety of k

Label query	Precision			
	k=5	k=10	k=15	k=20
Africa	0,60	0,80	0,67	0,70
Beaches	0,40	0,40	0,40	0,45
Buildings	0,80	0,50	0,47	0,45
Bus	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
Dinosaur	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
Elephant	<b>1,00</b>	0,90	0,80	0,70
Flowers	<b>1,00</b>	0,90	0,87	0,75
Horse	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>
Mountains	0,60	0,70	0,60	0,50
Food	<b>1,00</b>	0,90	0,80	0,80
Average	0,84	0,81	0,76	0,74

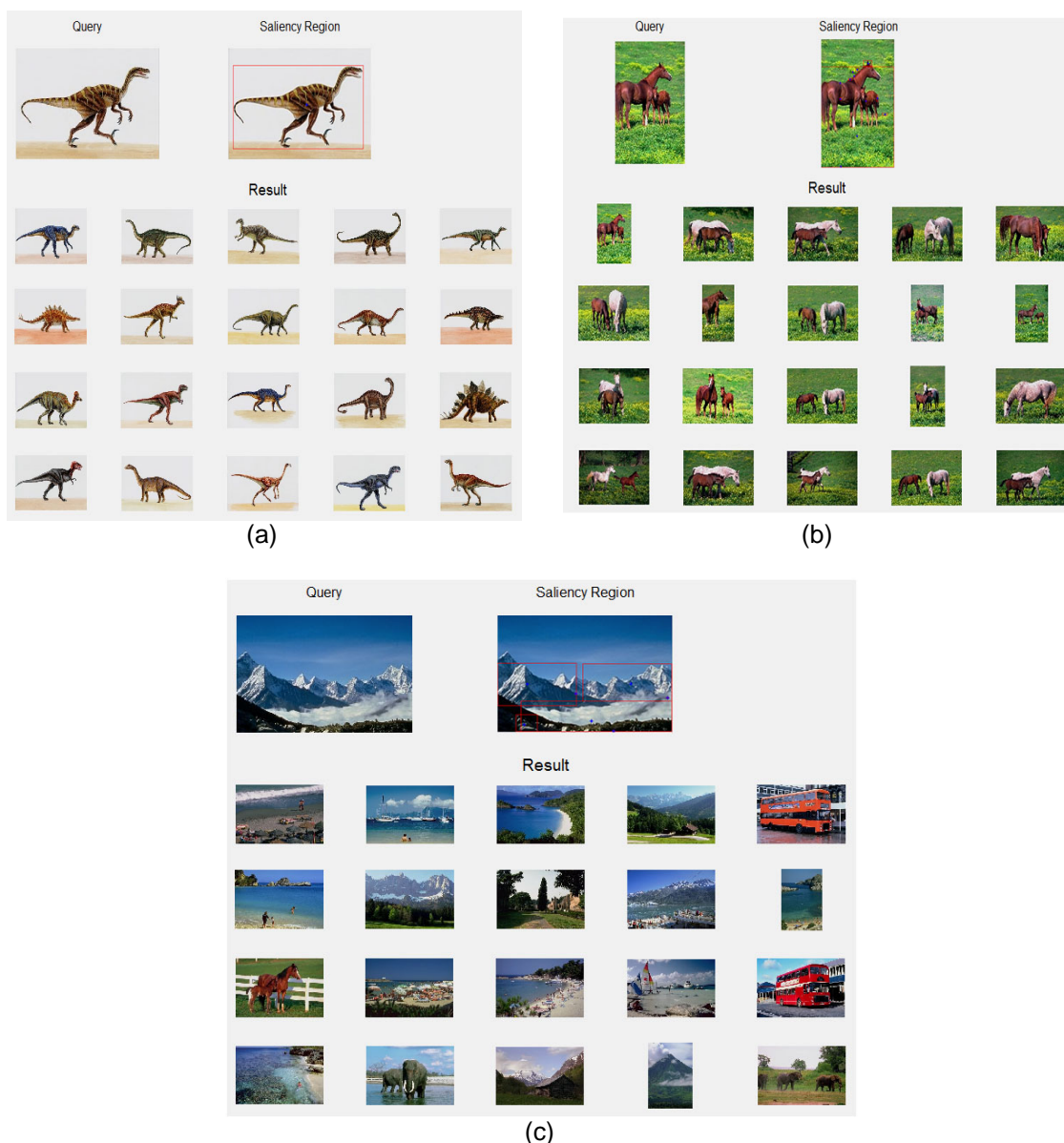


Figure 5. Example query and results for retrieval k=20, (a) Query and results of category “Dinosaur”, (b) Query and results of category “Horses”, (c) Query and results of category “Mountain”

From Table 2, it can be seen the average precision for  $k$  with a value of 5, 10, 15, and 20 are 0.84, 0.81, 0.76, and 0.74. When  $k$  increases the values of precision will decrease. For all variations of document return, Bus, Dinosaur and Horse has the highest precision with 1.00 (100%). Example of the query and result can be seen in Figure 5. Experiments were also conducted by using standard 11 recall. The results can be seen in Figure 6 that shows the precision-recall of the proposed method.

## 2. Weight Parameter

The measurement of the similarity between the query image with the images in the database is used three kind of features, where each feature has a weight. The weight of the features are used to indicate the percentage of the effect of these features to determine similarity. The weights for the color, texture, and shape features are symbolized by  $w_c$ ,  $w_t$ , and  $w_s$ , respectively. We performed experiments using different weights, to obtain the optimal parameter values. Table 3 shows the average value of precision for  $k=20$  with different weights. Based on the Table 3, the value of the highest precision is obtained when the weights are  $w_c$ ,  $w_t$  and  $w_s$  of 0.10, 0.40 and 0.50, respectively, and also when the values are 0.35, 0.35 and 0.30. Between those two set-weight values, in this research we choose the first one because in  $k=5$ , 10 and 15 the first value-set is out perform than the second one.

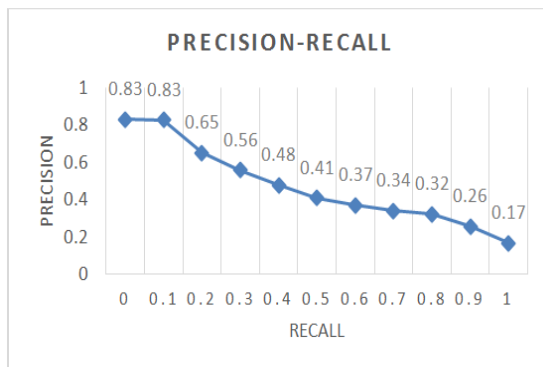


Figure 6. Performance of the proposed method

Table 3. Weight Parameter

$w_c$	$w_t$	$w_s$	Precision
0.40	0.50	0.10	0.73
0.10	0.40	0.50	<b>0.74</b>
0.50	0.10	0.40	0.61
0.10	0.50	0.40	0.73
0.35	0.35	0.30	<b>0.74</b>
0.30	0.35	0.35	0.72
0.35	0.30	0.35	0.72

## 3.2. Discussion

From the experimental results, there are several points that could be discussed. In Table 2, it can be seen that for category "Dinosaur" in Figure 5(a) has a precision value of 1.0 (100%). This is because the "Dinosaur" has a background image that is relatively homogeneous, consisting only of one color. It affects the segmentation results in the determination of the selected region for local features extraction. The segmentation results can also affect the determination of the shape which is used as a global feature. Category which also produces precision value 1.0 is "Horse" in Figure 5(b). Although the Horse has a background consisting of a variety of colors, but have a different color combination of other colors in the dataset, as well as texture. Another category is "Bus", in this category the Bus has similar shape in every image on this category that can be identify well using the feature shape.

This method can also be used in a heterogeneous background image for example in Figure 5 (b). However, because the background is heterogeneous, the selected saliency region is the overall picture of a horse image. It is because when segmenting using edge descriptor, grass is also detected as object.

The other category such as "Beach" and "Mountain", both have similarities in color feature. In a dataset that we use, both categories are dominated by white and blue colors. This can cause errors when retrieve the images, as can be seen in Figure 5(c). Beside that the scanary image has no fixed shape, so that the global feature that has higer weight cannot give much effect.



This method has a weakness in the images that has no clear boundary between foreground and background, as in images of scenery. Whereas in the images with clear object, the method can produce good precision value. This weakness is based on the selection of the selected region are taken from the results of image segmentation. For objects that segmented properly, the selected region obtained will be better in representing the object.

Image consisting of multiple objects at the same time can also affect the determination of the selected region. Some of the images that have not homogeneous background are often detected as objects. Therefore, the determination of detected object relevancy is important to do.

In image retrieval, selected image query can affect the value of precision. In research conducted by Vemina and Jacob [1] mentioned the results of testing against several methods. Tests were performed using the same dataset with the dataset used in this study [1]. For some category like "Africa", "Bus", "Elephant", "Dinosaur", and "Food" our proposed method produces better precision value. But for the other category the precision value of our proposed method is lower than the Vemina and Jacob Method [1]. Nonetheless query that is used in previous method [1] is unknown, thus allowing for the difference between the image query conducted in this study with the experiments that have been conducted by Vemina and Jacob.

In this study, the determination of the weight parameter for each feature has an impact on the value of precision of the search results. In Table 3 it can be seen that the weight variations result the different precision. For this proposed method, the shape and texture features can be better descriptor than the color feature in some category. This is because in some category the object has common in color, so that in this case color cannot be used as good descriptor. We also find that texture can give a larger effect than the shape, because when the  $w_t$  is set with low value the precision can be drop about 0.13.

#### 4. Conclusion

The proposed method based on proportional overlapping region to choose relevant selected region has average precision 74 % in  $k=20$  and has maximum precision 100%. However, this method has a weakness in the images that has no clear boundary between foreground and background, as in images of scenery. For further research, it will be optimized in determining the weigh parameter. And will also optimize edge descriptor segmentation by pruning unnecessary edge from irrelevant object.

#### References

- [1] R Vemina, KP Jacob. A Sub-block Based Image Retrieval Using Modified Integrated Region Matching. *IJCSI: International Journal of Computer Science*. 2013; 10(1): 686-692.
- [2] XY Wang, YJ Yu, HY Yang. An Effective Image Retrieval Scheme Using Color, Texture and Shape Features. *Computer Standards & Interfaces*. 2011; 33(1): 59-68.
- [3] XY Wang, ZF Chen, JJ Yun. An effective method for color image retrieval based on texture. *Computer Standards & Interfaces*. 2012; 34(1): 31-35.
- [4] S Nishant, V Tyagi. Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching. *Information Sciences*. 2014: 212-224.
- [5] M Subrahmanyam, RP Maheshwari, R Balasubramanian. Local maximum edge binary patterns: A new descriptor for image retrieval and object tracking. *Signal Processing*. 2012; 92(6): 1467-1479.
- [6] Guizhi Li. Improving Relevance Feedback in Image Retrieval by Incorporating Unlabelled Images. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(7): 3634-3640
- [7] V Chernov, J Alander, V Bochko. Integer-based accurate conversion between RGB and HSV color spaces. *Computer and Electrical Engineering*. 2015; 46: 328-337.
- [8] T Ojala, M Pietikainen, T Maenpaa. Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 2002; 24(7): 971-987.
- [9] Z Guo, L Zhang, D Zhang. A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*. 2010; 19(6): 1657-1663.
- [10] M Rizon, H Yazid, P Saad, AYM Shakaff, AR Saad, MR Mamat, S Yaacob, H Desa, M Karthigayan. Object detection using geometric invariant moment. *American Journal of Applied Sciences*. 2008; 2(6): 1876-1878.
- [11] MK Hu. Visual Pattern Recognition by Moment Invariants. *IRE Transations on Information theory*. 1962: 179-187.