

Full Length Research Paper

An alternative approach for unsupervised cluster based image retrieval

P. Pattanasethanon^{1*} and B. Attachoo²

¹Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand.

²Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand.

Accepted 2 August, 2012

In this research, we proposed an alternative method to retrieve images from its clusters which is determined by its color histogram. In other approach, it is called cluster based image retrieval model. One common problem with conventional content based image retrieval (CBIR) system is that the output data contains too many non-relevant feedbacks due to the indistinct symmetry of semantic gap. In this paper, two clustering techniques are performed to solve the problem mentioned above; including the misconception of semantic gap, called k-means and hierarchical clustering technique. The proposed method shows a promising retrieval procedure in terms of runtime errors reduction and accuracy. The experimental results determine that this alternative image retrieval model does not only retrieve images to the homogeneity of its features but it is also practically sufficient.

Key words: Content based image retrieval, color histogram, K-means, clustering image retrieval.

INTRODUCTION

Content based image retrieval (CBIR) systems have been one of the highlighted topics which correlate with the images' semantics. According to Ritendra et al. (2008), some images are indescribable, some does not have enough information, and some has homogeneous meaning. But what most research seems to lack, perhaps disregard, is the solution for semantic gap misconception. Within the image, the so called phrase "Semantic Gap" has vaguely perplexed the comprehension between computer and human in terms of conception (Huang et al., 2010). In order to retrieve an image, we need to consider its featured characteristics since these components will collate its mutual characteristics with the query image from the database. However, with more features, the system might face in-process ambiguity which slows down the computation process.

In this paper, we proposed an alternative model for unsupervised clustering based image retrieval using histogram based in both red green blue (RGB) and hue-saturation-value (HSV) color space. The aim of this research is to improve an accurate retrieval system and reduce runtime error in the retrieval features from the

query and database. With several features defectively matched, the system displays more non-relevant feedbacks on the interface in return (Ozden and Polat, 2007). By implementing clustering techniques to the system, the solution for non-relevant feedback is asserted (Rui et al., 1998). The clustering method sorts images of similar semantics together, thus, we obtain more relevant feedbacks at the output as an asset.

The algorithms of our alternative model include clustering, constructing histograms indices, and retrieving. The implementation of clustering algorithms creates a bound of comprehension in terms of semantic gap correlating human and computer itself to raise the conception of retrieval system. Two types of clustering methods involved in this research are k-means and hierarchical clustering. We construct color histograms in order to cluster them into 10 clusters (Amato and Lecce, 2003). Each cluster can be represented by group vectors; each consists of different values. In the retrieving procedure, the query image is compared to every group vector.

Our model is designed to reduce runtime error for the retrieval speed with accurate retrieval efficiency and More relevant feedbacks on the output. Feedbacks of highest similarity measures are retrieved as the group vector corresponds to the clusters. Since only one conformation

*Corresponding author. E-mail: april_it2005@hotmail.com.

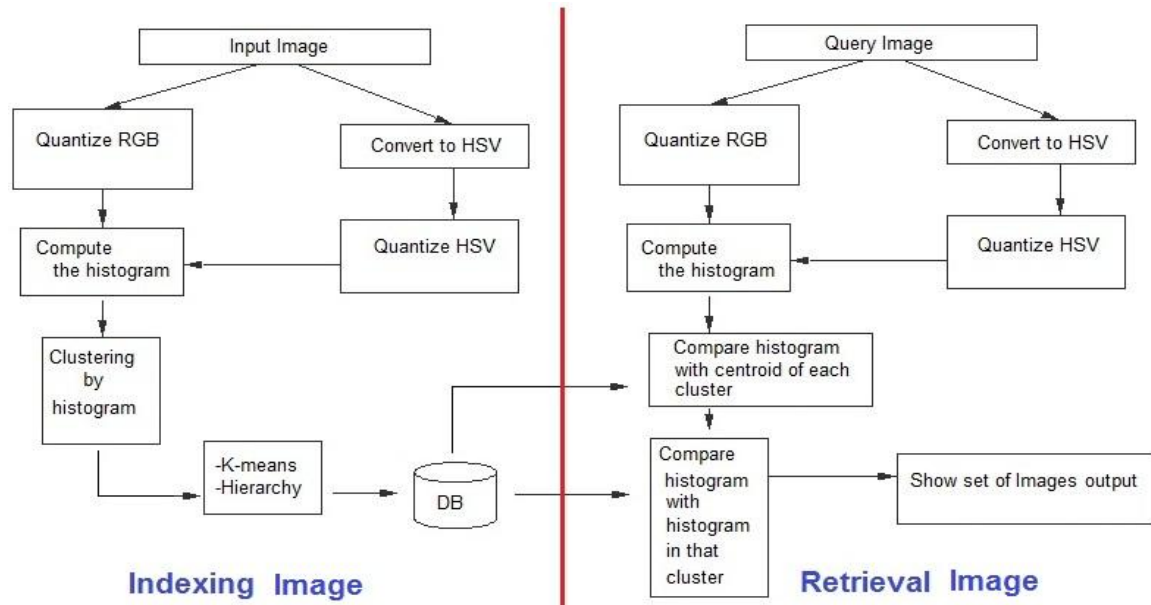


Figure 1. Flowchart of image retrieval system.

is required, the process of looping through the entire database is unnecessary.

This research is organized into 5 parts. The first part we discuss about the basis of color; the conception and how it functions. The second part describes clustering algorithms that are included in our research. After knowing how these algorithms function, evaluation is required. The third part demonstrates how clustering quality and matching paradigm determines which algorithm is more sufficient than the other. Then the experimental results are discussed, and finally, the conclusion.

COLOR EXTRACTION

In 1991, Swain and Ballard (1991) proposed a retrieval system by referring to the images' color histograms. This characteristic feature of the image is simple enough to compute and compare. Nevertheless, the impediment of this technique refers to be the uncertainty of the feature. For instance, two distinctively different images may contain an indistinguishable histogram. Thus, the image retrieval result performs ineffective results. In this paper, we use the semantics of color histogram to quantize the different color spatial characteristics. Then follow by clustering images into categories that each discernible image is classified (Yeelar and Narisara, 2008).

The steps for image clustering by color histogram are shown in Figure 1. The process of color indexing includes quantizing the RGB and HSV space. After the quantizing process, we computed the histograms in order to categorize those using k-means and hierarchy clustering. In the retrieval process the query image will disintegrate

its color characteristics for both models; RGB and HSV, and then quantizing process is performed to compute the histograms. Centroid comparison is denoted from the histograms. Once the homogeneous equivalent feature is discovered, solving for similarity measures within the cluster is performed at the next step.

RGB TO HSV CONVERSION

According to Gonzalez and Richard (2002), color values of RGB model from a standard RGB value consist of color, light, and brightness in the image. Since this research uses both RGB and HSV model for color classification, the conversion of RGB to HSV can be described as presented by the following equation.

$$H = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\}$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)]$$

$$V = \frac{1}{3} (R + G + B)$$

Where H is Hue, S is saturation value, V is brightness

QUANTIZATION

The normality of color values in images is stored in an RGB model. The total amount of 24 bits is being used to

store the color; 8 bits each for Red, Green, and Blue. It will require a profusely high amount of memory and computation speed. Therefore, quantization detaches the color layers and decreases the amount of data being used. Colors in RGB model are quantized into $(8 \times 8 \times 8) = 512$ bins. In HSV color space, quantization of hue requires the most attention. The hue circle consists of the primary color red, green, and blue separated by 120 degrees. A circular quantization at 20 degree steps sufficiently separated the hues such that the three primaries and yellow, magenta and cyan are represented each with three sub-divisions. Saturation and value are each quantized to three levels yielding greater perceptual tolerance along these dimensions. As a consequence, H is quantized to 18 levels and S and V are quantized to 3 levels. Therefore, the quantized HSV space has $18 \times 3 \times 3 = 162$ histogram bins (Linh and Reiner, 2005).

COLOR HISTOGRAM

$$D_k = |x_i^{(n)} - c_k|^2$$

Histograms in images denote the brightness, darkness, depth, and height. Theoretically, every image has a particular histogram for image indexing. Consider I as color image with the size. The position is the amount of pixels of image I . The histogram can then be computed as the following equation.

$$h_{c_i}(I) = \frac{H_{c_i}(I)}{(n_1 \times n_2)}$$

Where h_{c_i} is the histogram of color c_i in image I . n_2 is the length of the image. n_1 is the width of the image the denominator $n_1 \times n_2$ is the amount of pixels within the certain image. By doing so, the problem of unequal-sized image is resolved which when the color histogram of different sizes are compared, the histogram vector from the image is obtained.

IMAGE CLUSTERING

There are two techniques that are being used in this research which include k-means and hierarchical clustering. The clustering procedures in this work are as described below.

K-means clustering

As earlier mentioned, k-means clustering method are method that deals with minimizing average squares Euclidean distance (Tapas et al., 2002). The K-means algorithm of our work is first performs by a partition step which detaches 'n' objects in the data set of 'k' categories. It is partition into k groups and each group is substituted with its average squared Euclidean distance as follows: is the Minimum distance of the centroid and the object. c_k is the centroid of group k^{th} , generated by random. k is the cluster number. n is the

number of images in the group; $1, 2, 3, \dots, n$. $x_i^{(n)}$ is the image vector i in group n . Convergence of the centroid in each group to settle, the new centroid (C_{new}) is calculated as follow:

$$C_{new} = \frac{\bar{x} + C_{old(x)}}{2}$$

Where $C_{old(x)}$ is the old centroid value and $\bar{x}$ is the image vector. These are also known as the centroids or means of the categories. The idea is that clusters should not overlap by virtue of the clustering process for which we know what sort of data should be placed in the same cluster. A measure of how well the centroids represent the member of their clusters is the residual sum of square (RSS) or uncertain radius, the squared distance of each vector from its centroid sum of overall vectors:

$$RSS_k = \sum_{\bar{x} \in \omega_k} |\bar{x} - \bar{\mu}(\omega_k)|^2$$

Where $\bar{\mu}(\omega_k)$ is the center or centroids of the image data of cluster ω_k . RSS is an objective function in K-means and the key is to minimize it. A measure of how well the centroids represent their data is subjected to the fact that minimizing RSS is equivalent to minimizing the average squared distance for which 'n' is fixed. The member in the data set is determined by the shortest average squared Euclidean distance from the centroid. The shortest path will discriminates the data as the step consecutively repeat its iteration and stop as the centroid is constant and stable. The process of k-means algorithm in this paper is as shown in the Figure 2. The first step is to arbitrarily set the 'k' into perspective number of groups. For instance, in this our works we set to 10 cluster, $k=10$. Then follows by initializing the centroid for each cluster and find the shortest path for average square Euclidean distance in order to eliminate the RSS from the objective function.

Hierarchical clustering

The hypothesis of distance hierarchy tree is a concept hierarchy structure. Exploiting this mechanism also better facilitates the representation and computation of the distance between categorical values. A concept hierarchy consists of two parts: a node set and link set (Murthy et al., 2010). This paper extends the distance hierarchy structure with link weight. Each link has a weight that represents a distance.

Suppose, a point X in a distance hierarchy consists of two parts, an anchor and a positive real value offset, denoted as $X(N,d)$, that is, $\text{anchor}(X) = N$ and $\text{offset}(X) = d$. The anchor is leaf node and the offset represents distance from the root of hierarchy to the point. A point X is an ancestor of Y if X is in the part from Y to the root of the hierarchy.

Let $X(N_x, d_x)$ and $Y(N_y, d_y)$ be two points, the distance between X and Y can be defined as:

$$|X - Y| = d_x + d_y - 2d_{LCP(X,Y)}$$

Where $LCP(X,Y)$ is the least common point of X and Y in the distance hierarchy. $d_{LCP(X,Y)}$ is the distance between the least common point and the root.

Hierarchical clustering technique can be easily expressed in binary tree graph or dendrogram. It represents the relationship of the clusters, sub clusters, and level of clusters.

CLUSTERING QUALITY

In clustering algorithms, it is always wise to evaluate the quality. Two common values in quality measures are Purity and Entropy.

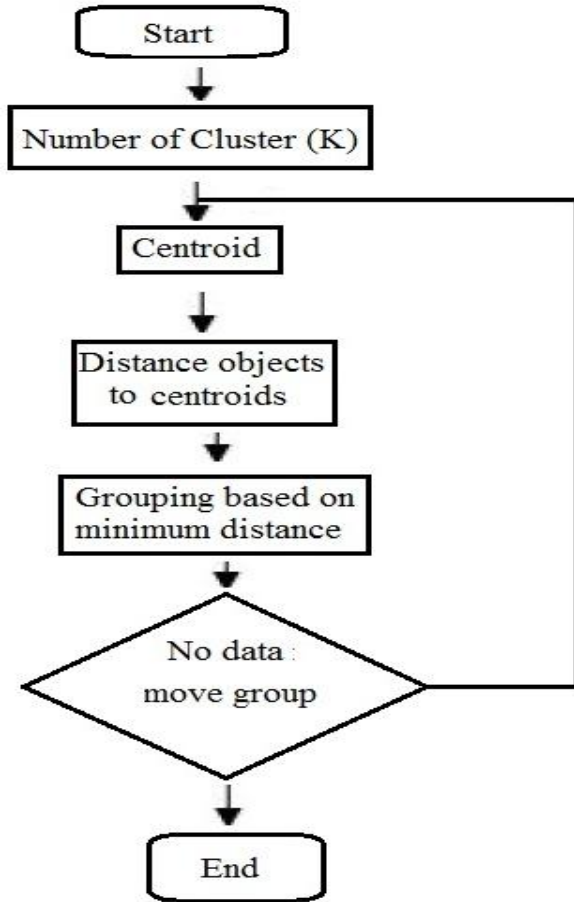


Figure 2. Flow chart of K-means algorithm.

For instance, assume that we have n sets of images which are clustered into c clusters denote by $1, \dots, c$. The images are also clustered in accordance with the clustering algorithm which are m clusters, denote by, $j=1, \dots, m$.

The purity of the image can be computed as follows (Chen et al., 2003).

$$p(c_j) = \frac{1}{|c_j|_{k=1, \dots, c}} \max |c_j, k|$$

Where c_j , k is the c_j clusters that are the member of k . $|c_j|$ is the member of cluster j . Each cluster may consist of several members from other categories. The purity measures are the ration of the members within the main category divided by all members of the cluster. The range of purity measures will be between $[\frac{1}{c}, 1]$. The higher the purity, the better quality is represented.

Entropy is another measure which is determined from the dispersion of the categories within the cluster. As shown in the equation below.

$$h(c_j) = \frac{1}{\log c} \sum_{k=1}^c \frac{c_{jk}}{c_j} \log \frac{c_{jk}}{c_j}$$

Owing to the fact that it measures the dispersion of the categories in the cluster, it displays a more accurate result of quality measures than purity. Entropy is normalized in the range of $[0, 1]$. In comparison to purity, entropy nearest to one denotes that it is the member of every cluster.

SIMILARITY MEASURE

Similarity measure is a method to compare the characteristics of the image. This is one of the procedures in image retrieval. The similarity of two images can be obtained from comparing characteristic vectors from the image (Chen et al., 2003). It can practically be computed by distance measurement. However, in this research, we compare each histogram, instead. For instance, I and J are the images of 512 bins that are being compared. The function (I) and (J) are the two components of the image that will be compared in this case. As shown in the equation.

$$D^2(I, J) = \sum_{i=1}^{512} |f_i(I) - f_i(J)|^2$$

The overall comparison will then be solved by R-squared or correlation r . The correlation r is one of the most common and most useful statistics. A correlation is a single number that describes the degree of relationship between two variables. In this paper, the correlation r is a correlation between query image and retrieved images. When I is a query image and J is a retrieved image. They are reduced to the matrices of the same size. The correlation r is defined as follows:

$$r = \frac{\sum_m \sum_n (I_{mn} - \bar{I})(J_{mn} - \bar{J})}{\sqrt{(\sum_m \sum_n (I_{mn} - \bar{I})^2)(\sum_m \sum_n (J_{mn} - \bar{J})^2)}}$$

Where $\bar{I}$ is the average of matrix element I . $\bar{J}$ is the average of matrix element J .

RETRIEVAL EFFICIENCY

Muller et al. (2001) proposed a research called "Performance Evaluation in Content-Based Image Retrieval: Overview and proposals". The research refers to two value 'precision' and 'recall' as a reference factor for retrieval efficiency. This method evaluates the effectiveness from the amount of relevant feedback and the query image. Precision denotes how precise the system is by considering the relevant feedback and the retrieved images. This will show the preciseness within the cluster. The equations below express the computation of precision and recall.

$$\text{Recall} = \frac{\text{Number of relevant image retrieved}}{\text{The total number of relevant image in databased}}$$

$$\text{Precision} = \frac{\text{Number of relevant image retrieved}}{\text{The total number of image retrieved}}$$

COMPARING WITH OTHER RESEARCH

Recently, Murthy et al. (2010) launch a research name "Content Based Image Retrieval Using Hierarchical and K-means Clustering Techniques". The first block diagram in Figure 3 and Table 1 are the steps of procedures in their algorithms and the algorithm of this work. Some experimental results of the previous work and the present works are compared in Table 7 and Figure 10, respectively.

EXPERIMENTAL RESULTS

The data in this research were downloaded from <http://wang1.ist.psu.edu/> with 1000 of corel images which

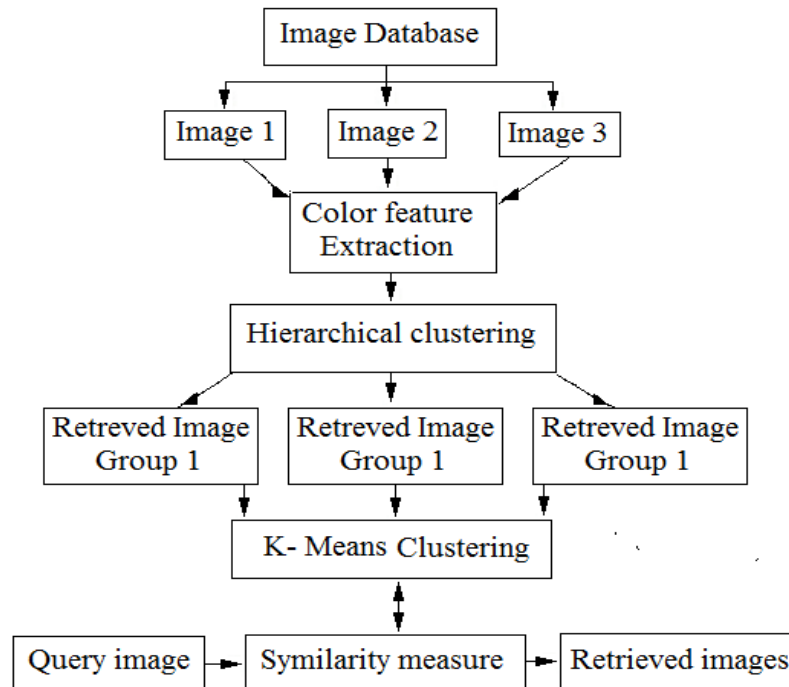


Figure 3. Flowchart of other research.

Table 1. Comparing with other research.

Item	This research	Other research
1. Problem statement	Relevance feedback	Relevance feedback
2. Data	http://wang1.ist.psu.edu/	http://wang1.ist.psu.edu/
3. Methodology	- Color histogram - K-means or Hierarchy	- Color histogram - K-means - Hierarchy
4. Similarity measure	- Euclidian distance - Correlation coefficient	Total time (No evaluated)
5. Retrieval efficiency	Precision, Recall	Total time (No evaluated)

can be sub-divided into 10 categories considering the semantics and visualization. There are 100 images as members of each cluster (Table 2), and some sample images of each group are shown in Table 3.

Our operating system for this experiment is processor core 2, clock speed 2.79 GHz, RAM 2 GB, Microsoft window vista running MATLAB 7.6 programming for the procedures. As a cluster based image retrieval model, the prominent step is to cluster all images before we launch RGB to HSV color space conversion. RGB and HSV color spaces are quantized into 512 bins and 162 bins, respectively. The two clustering techniques

mentioned; k-means and hierarchical clustering, detach these data into 10 distinctive clusters. By comprehending the semantics of each cluster, the experimental results manifest that the first group obtained 40 images clustered by RGB k-means model. The second cluster obtained 54 images and the third cluster obtained 79 images, respectively, while, RGB hierarchical model clustered the first cluster into 121 images out of 1000. The second cluster obtained 695 images and third cluster obtained 15 images, respectively. On the other hand, we obtained 55 images from the first cluster by HSV model, 62 images for the second cluster, and 154 images for the third

Table 2. Original image category in 10 groups.

Group no.	Name	Image amount	Item
1	Africans and Villages	100	0 - 99
2	Beach	100	100 -199
3	Buildings	100	200 -299
4	Buses	100	300 -399
5	Dinosaurs	100	400 - 499
6	Elephants	100	500 - 599
7	Flowers	100	600 - 699
8	Horses	100	700 - 799
9	Mountain and Glaciers	100	800 - 899
10	Food	100	900 - 999

Table 3. A samples of images in each group.

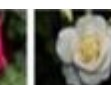
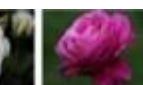
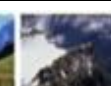
Group no.	A sample of images in each group					
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

Table 4. Comparison of new images account of 10 categories in with difference clustering method.

Cluster	RGB		HSV	
	K-means	Hierarchical	K-means	Hierarchical
1	40	121	55	32
2	54	695	62	182
3	79	15	154	69
4	198	22	163	67
5	89	18	109	480
6	220	10	32	67
7	76	23	45	26
8	130	44	120	42
9	20	31	201	23
10	94	21	59	12

**Figure 4.** Some image set of cluster 1 using RGB k-means clustering.

cluster. Finally, by performing HSV hierarchical model on the image database, we obtained 32 images for the first cluster, 182 images for the second cluster, and 69 images for the third cluster etc. As the results exhibit, clustering with k-means technique reveals a convergent homogeneity of cluster members which contrasts with hierarchical clustering technique. The divergence of cluster members in hierarchical technique has substantiated that k-means clustering technique is more sufficient as a clustering algorithm.

Figures 4 to 7 express the practical paradigm of how clustering methods work in both color spaces for cluster 1; RGB: K-means, HSV: K-means, RGB: Hierarchy, HSV: Hierarchy. In which case, the average value of each group is the factor that represents the characteristics of the cluster. From Table 5, the sequence of purity after clustering the data; RGB K-means, HSV K-means, RGB Hierarchy and HSV Hierarchy displays the mean values of purity as 0.774, 0.707, 0.676, and 0.613, respectively. Distinctive value approaches 1 indicate a higher



Figure 5. Some image set of cluster 1 using RGB hierarchical clustering.

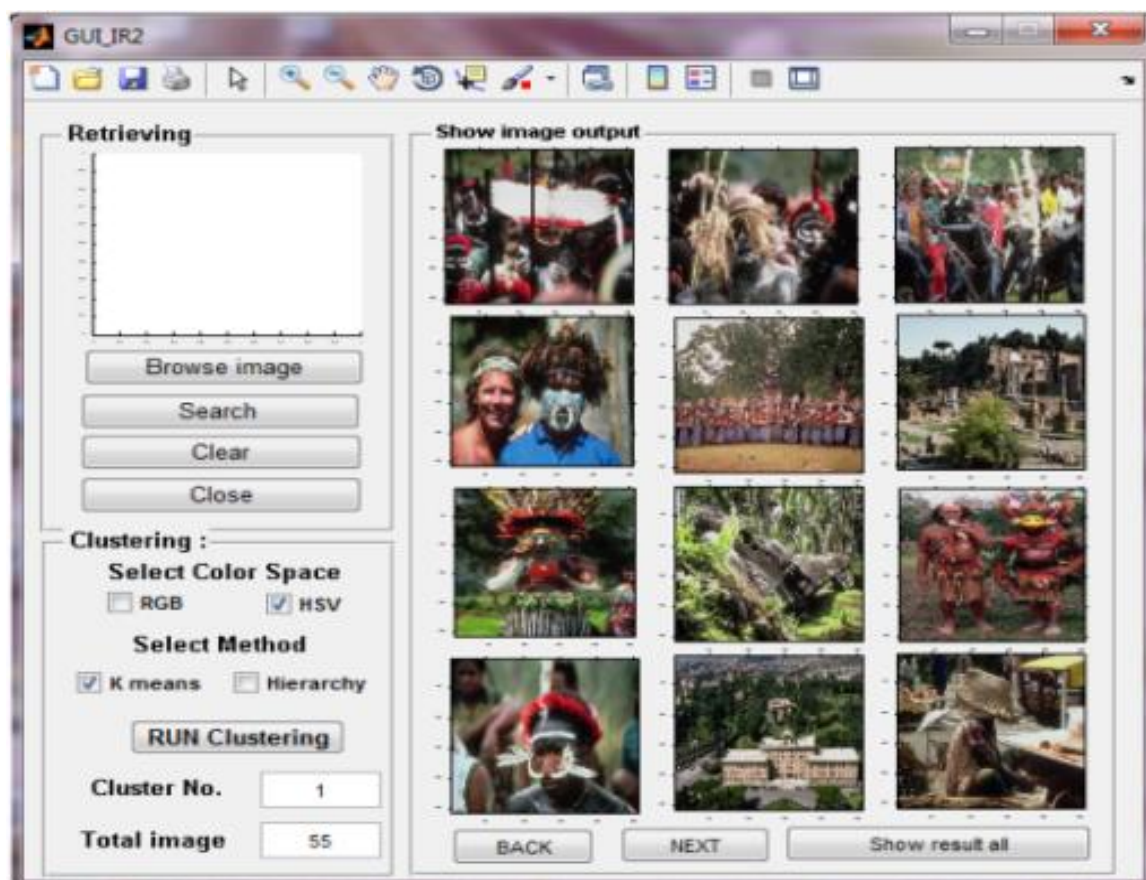


Figure 6. Some image set in cluster 1 using HSV k-means clustering.

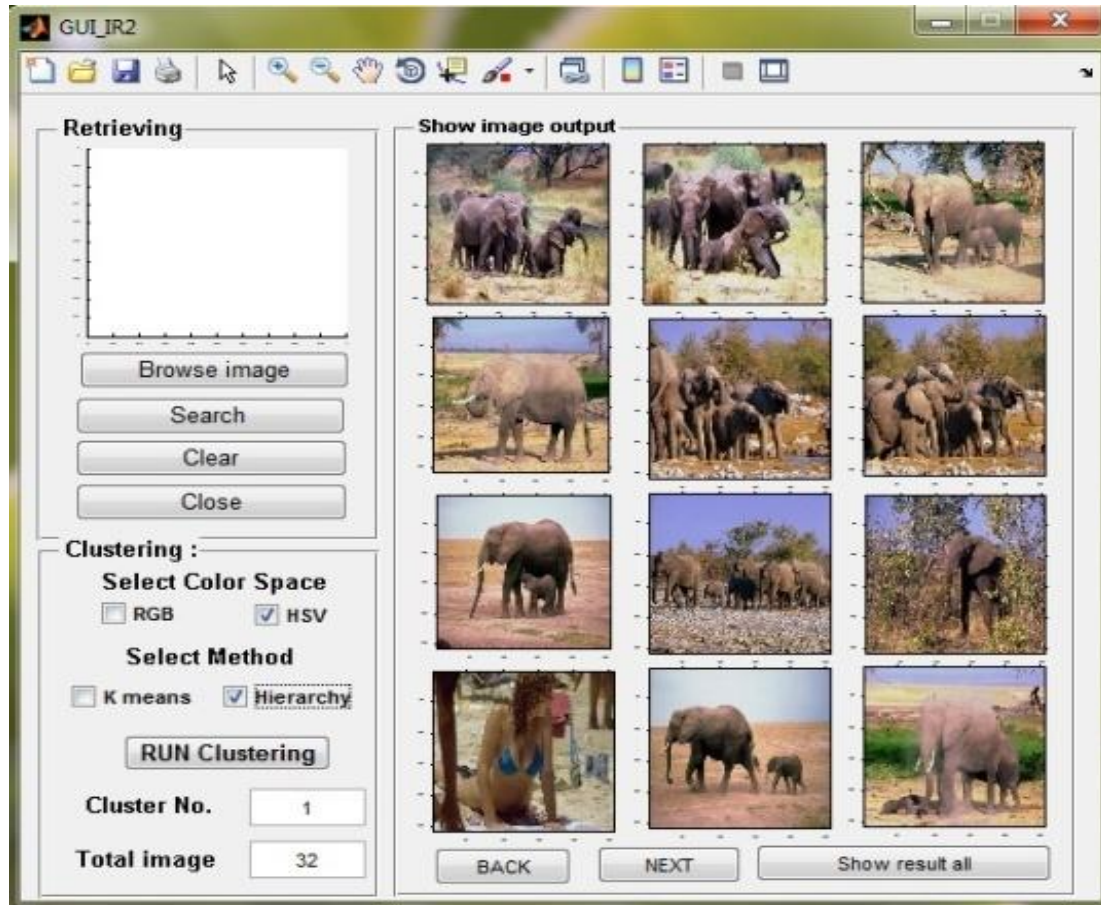


Figure 7. Some image set in cluster 1 using HSV hierarchical clustering.

Table 5. Comparison of the purity in each category with difference clustering method.

Cluster	RGB		HSV	
	K-Means	Hierarchical	K-Means	Hierarchical
1	0.756	0.544	0.334	0.449
2	0.208	0.234	0.342	0.862
3	0.978	0.634	0.746	0.164
4	0.458	0.498	0.677	0.986
5	0.990	0.180	0.865	0.510
6	0.874	0.750	0.609	0.450
7	0.866	0.990	0.832	0.984
8	0.980	0.986	0.785	0.990
9	0.950	0.970	0.897	0.198
10	0.680	0.980	0.985	0.420
Mean	0.774	0.676	0.707	0.613

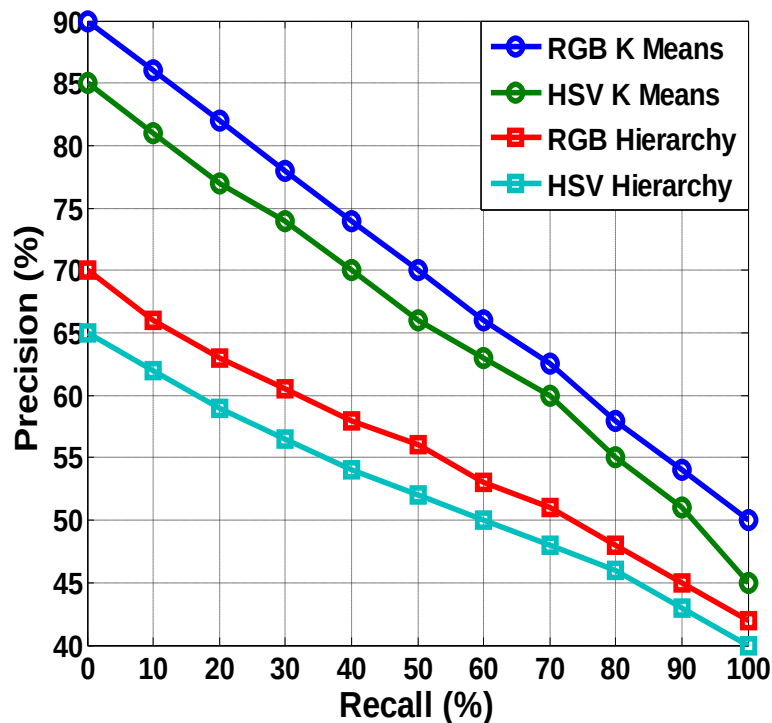
homogeneity in terms of cluster. From Table 6, the entropy of RGB K-means, HSV K-means, RGB Hierarchy and HSV Hierarchy, indicates distinguish quality of the clustering method. The mean values of entropy are 0.265, 0.386, 0.366 and 0.433, respectively. The results

of clustering quality are as shown in Tables 5 to 6. By considering the purity and entropy, the observation demonstrates a sufficient clustering quality in k-means than that of hierarchical clustering technique.

Retrieval efficiency is evaluated based on two values,

Table 6. Comparison of the entropy of each group with difference clustering method.

Cluster	RGB		HSV	
	K-Means	Hierarchical	K-Means	Hierarchical
1	0.240	0.422	0.820	0.620
2	0.920	0.846	0.560	0.248
3	0.090	0.386	0.250	0.844
4	0.475	0.440	0.190	0.058
5	0.120	0.842	0.098	0.450
6	0.130	0.520	0.876	0.550
7	0.220	0.098	0.230	0.048
8	0.110	0.084	0.500	0.082
9	0.084	0.120	0.108	0.754
10	0.264	0.110	0.120	0.684
Mean	0.265	0.386	0.366	0.433

**Figure 8.** Precision – recall curve for different clustering methods.

precision and recall. The results vividly respond with anticipated relevant feedbacks. The effectiveness of the retrieval system can be computed as follows. Arbitrarily select 100 images from the 10 clusters, 10 images per cluster. The interface can display 10 images. Five query images are arbitrarily selected from each cluster which these queries will be averaged into one value in order to display in the PR graph in Figure 8. The graph demonstrates the results from the total of 10 values.

Figure 8 demonstrates the distinguish correlation between precision and recall regarding the alternative

model retrieval system by clustering. In cases where the images are quantized and clustered into RGB K-means, HSV K-means, RGB Hierarchy, and HSV Hierarchy, respectively. The exactness result from quantization indicates that RGB K-means is more accurate in terms of color quantization than HSV K-means. Likewise, RGB Hierarchy algorithm performed a more accurate color quantizing result than HSV Hierarchy algorithm. Based on the effectiveness evaluation of the retrieval system, we can consider the results illustrating an equal recall value but different precision value. The greater precision

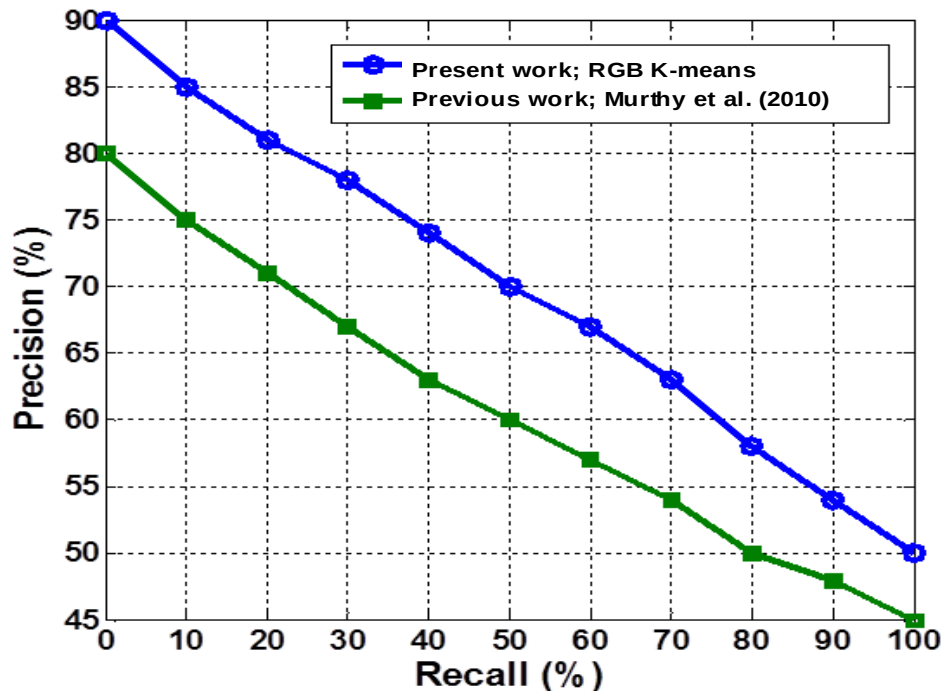


Figure 9. Comparison of precision – recall curve of present work and Murthy et al. (2010).

value obtained implies that it has better efficiency. On the basis of typical recall calculation, the effectiveness of any retrieval system is analyzed at recall = 50%. Figure 8 shows that the precision of RGB K-mean is 70%; HSV K-means is 66%; RGB Hierarchy is 56%; and HSV Hierarchy is 52%, respectively. After we compare the 4 retrieval models, the retrieval efficiency of RGB K-Means algorithm indicates the highest performance of all four while HSV Hierarchy indicates the lowest performance.

Figure 9 shows the correlation of precision and recall in present and previous works to estimate the effectiveness of the retrieval system in RGB color space based on Murthy et al. (2010) model which can be compared to our work referencing the recall point at 50%. From our alternative model, the precision obtained in RGB color space and K-means is 70%, which is 10% greater than that of the previous results from Murthy et al. (2010) Precision = 60%. The result has, therefore, manifest a higher performance in terms of retrieval effectiveness.

Table 7 summarizes the precision and recall value between present and previous work. The precision value obtained from the image retrieval model using k-mean clustering in this work is higher than those of hierarchy method. While in the previous work, precision value is 87%, present work shows as high as 95% precision. In case of image retrieval using Hierarchical clustering, the precision of our work is 93 and 85% for the previous works, respectively. In addition, the recall value of image retrieval using K-mean in our work is 65%. In another work, the recall value is slightly lesser, 60%. For image

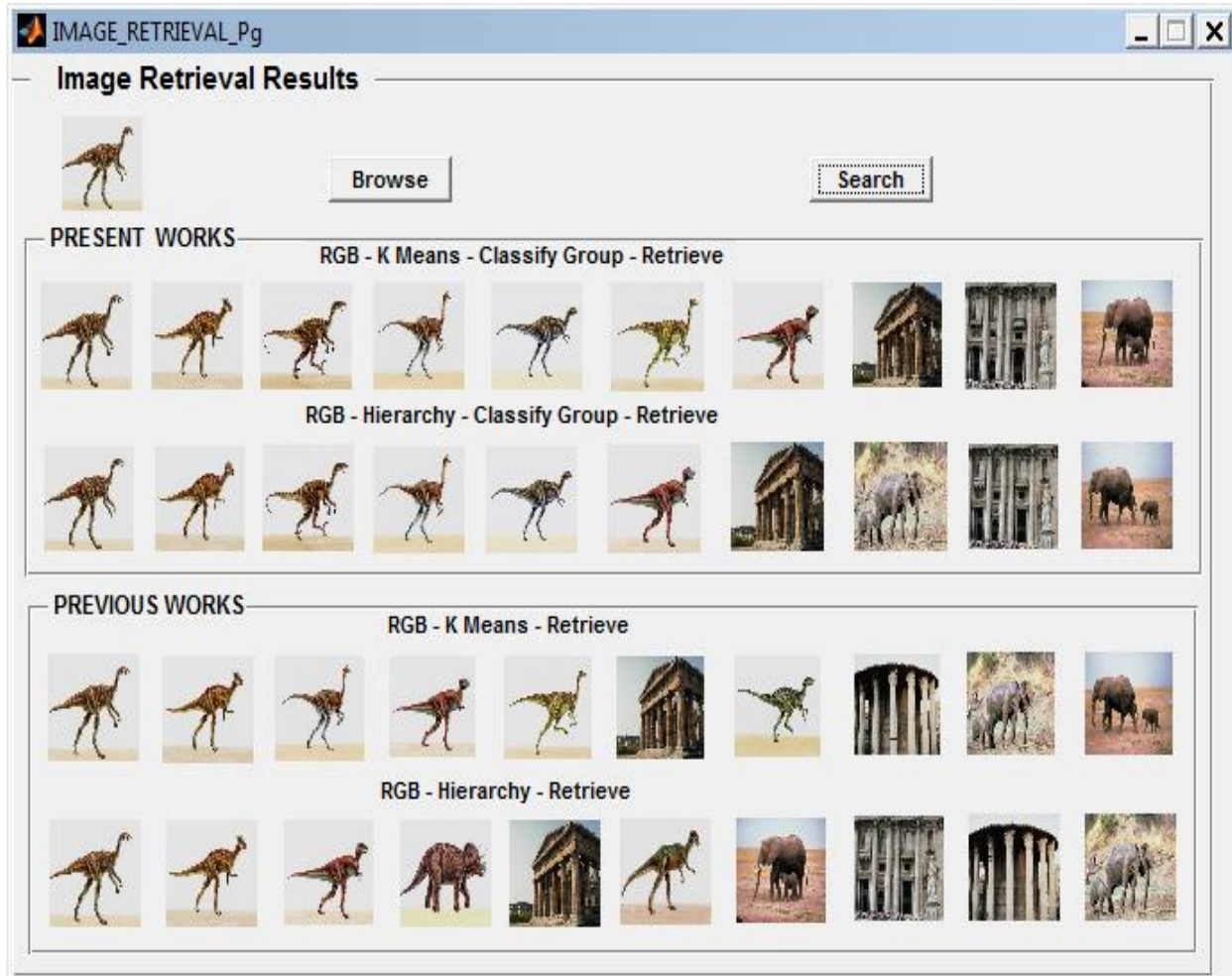
retrieval using hierarchy clustering, the recall value of the present work is 60 and 45% for previous work. In Figure 9, some results of retrieval image from our present work are compared with another work, which is implied by the GUI window of MATLAB programming. The retrieved image and number of relevance output image is consistent, which the precision and recall values are distinctively shown in Table 7.

Conclusion

This paper proposed an alternative model for cluster based image retrieval using color histogram. We cluster the color histograms in order to increase the efficiency of the indices and the retrieval system. The procedure of constructing the indices begins with quantizing the histogram in both RGB and HSV model. This is to organize the signal in order to construct the histograms which are the color characteristics of the images. With clustering as the dominant methodology, two techniques namely K-means and Hierarchy clustering techniques are performed collaterally. The retrieval process uses the same method redundantly until the indistinctive data approaches. In this procedure, we compare two images which are the query and the image from our database with homogeneous characteristics. The representative of each group or the centroid with homogeneous histogram is compared. The experimental results of this research indicate that the highest retrieval efficiency is

Table 7. Precision and Recall comparison for RGB color space.

Method (Reference Table 1)	Precision (%)		Recall (%)	
	K- Means	Hierarchy	K- Mean	Hierarchy
Present work	70	56	50	50
Other work	60	50	50	50

**Figure 10.** Some results from image retrieval comparison between present and previous works.

RGB K-means algorithm while the lowest performance is HSV Hierarchy algorithm. This brings about a rather more accurate result and faster speed with reduced runtime error than other models because the equilibrium data derives from the representation of each cluster in the database. Therefore, the contrast result or result with heterogeneous histogram is incessantly eliminated. For this reason, this paper not only proposed a model with retrieval speed but also sufficiently reveals almost all the possible relevant on the output interface. We also hope that the future work would include developing other useful features such as shape or region for image feature for

image retrieval, since, color is the only characteristic feature in this research. Therefore, errors may occur at some point.

REFERENCES

- Amato A, Lecce VD (2003). Edge Detection Techniques in Image Retrieval: The Semantic Meaning of Edge. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communication. pp. 143-148.
- Chen Y, Wang JZ, Krovetz R (2003). Content-Based Image Retrieval by Clustering. Proceeding of the 5th ACM SIGMM Int. Workshop Multimedia Inf. Retrieval 1:193-200.

- Gonzalez R, Richard W (2002). Digital Image Processing. 2 ed, Prentice Hall. United States.
- Huang MY, Gao K, Chan L (2010). A Review of Region-Based Image Retrieval. J. Signal. Proc. Syst. 59:143-161.
- Linh VT, Reiner L (2005). Compact color descriptors for color based image retrieval. Signal Proc. 85:233-246.
- Muller H, Muller W, David McG S, Thierry P (2001). Performance Evaluation Content Based Image Retrieval: Overview and Proposals. Pattern Recogn. Lett. 22:593-601.
- Murthy V, Vamsidhar E, Rao P, Raju S (2010). Content Based Image Retrieval using Hierarchical and K- means Clustering Techniques. Int. J. Eng. Sci. Tech. 2(3):209-212.
- Ozden M, Polat E (2007). A color image segmentation approach for content-based image retrieval. Pattern Recogn. 40:1318-1325.
- Ritendra D, Dhiraj J, Jia Li, James WZ (2008). Image retrieval: Ideas Influences and trend of the new age. ACM Computing Surveys. 40(2):5:1-5:60.
- Rui Y, Huang TS, Ortega M, Mehrotra S (1998). Relevance Feedback: A Power Tool for Interactive content-Based Image Retrieval. IEEE Trans. Circ. Video. Technol. 8(5):644-655.
- Swain M, Ballard D (1991). Color Indexing. Int. J. Comput. Vision 7:11-32
- Tapas K, David MM, Nathan SN, Chistine DP, Ruth S, Angela Y, Wu (2002). An Efficient K-Means Clustering Algorithm: Analysis and Implementation. IEEE Trans. Pattern. Anal. Mach. Intell. 24(7):881-891.
- Yeelar N (2008). Image retrieval by color based clustering technique. M.Sc. thesis, Chiang Mai University, Chiang Mai, Thailand.