Content Based Image Retrieval Using Two Color Feature Extraction

F A Mufarroha¹, D R Anamisa¹ and A G Hapsani²

¹Departement of Informatic Engineering, University of Trunojoyo Madura, Bangkalan – Madura, Indonesia
²Faculty of Computer Science, University of Brawijaya, Malang - East Java, Indonesia

Abstract. Content Based Image Retrieval (CBIR) is a process to search for an image based on the content or features that are inside. Nowadays, many image retrieval applications have been made to meet the needs, so this application can provide convenience in terms of the introduction and search for an image. In this research, we used 10 different objects as image retrieval consists of Bicycle, Cow, Flower, Frangipani, Grape, Horse, Lovebird, Orange, Strawberry, Tree. These objects can be expressed in 10 classes. Our aim using these objects is viewed from the color of every object and the object of a different kind. From that point of view we built a CBIR system by utilizing the main features of the object (image). The main feature is the color feature. In this research, the main process is the extraction of color features with the color histogram and color moments. So in this research it will produce feature extraction by measuring the similarity in the library image. This measurement is done by calculating the closest distance using the euclidean distance method. The data library used in this research is 10 pieces of data, the test data is 50 pieces with 5 pieces for each of these classes. After testing using data and methods described above, the results of accuracy are obtained that the application of the color moments method gets better results than the color histogram method.

Keywords— content based image retrieval; feature extraction; color histogram; color moment;

1. Introduction

In the current development, context based techniques or languages are generally called information retrieval (IR). IR becomes impractical because of the large database size and subjective judgments, this means in the image with text. Based on the description above, it takes a more accurate technique in image search, namely search based on content (Content Based Image Retrieval (CBIR)). To do a search, CBIR extracts features to bring up the features in an image, such as color features, shapes and textures.

In everyday life, the objects that we can meet are very numerous and varied so that in this research we will classify several objects so that objects can be identified. The main advantage of the CBIR approach is the transition to the use of traditional keyword-based approaches to the automatic retrieval process. Some of the applications of CBIR technology include applications such as fingerprint identification, biodiversity information systems, digital libraries, crime prevention, medicine, historical research, and others [1]. To overcome the process in CBIR, the researcher first selects the
method in extracting features. It has been explained previously that to get the features of an object, we
can take the characteristics of an image both color features [2] [3], form [4], texture [5], or a
combination too [6]. In several types of feature extraction techniques, there are available how to
extract an information in the image by focusing on reflected information [7]. In research [8],
researchers used one of the color spaces by doing a combination of the RGB color space.
In the selection, researchers finally applied two color feature extraction methods, such as the color
histogram and Color Moments. Both of these methods are compared to find out the comparison of
accuracy levels. Color feature extraction used is a color histogram, because this feature is obtained
from the pixel intensity value diagram [9]. While the second feature extraction method, uses Color
Moments. The selection of the Color Moments method for the extraction of color features is due to
Color Moments can give rise to color characteristics with three moments, such as the mean, standard
deviation and skew [10].
Based on the two feature extraction methods above, researchers will build a CBIR system on several
objects that the researchers have chosen before. The system starts by extracting objects using two
color feature extractions (color moment and color histogram). Features that have been obtained will be
processed with similarity measurements by applying the most distance calculation method such as
euclidean distance to get the classification results.

2. Methods
In this research, researchers built a system in image search. The workings of a system that constructs
researchers can be seen in Figure 1.

![Figure 1. Overview System](image_url)

The system starts with the image request process. The feature will be taken by using color feature
extraction. When the color feature has been generated, the closest distance search is made to the class
features in the database. The researcher determined 10 different objects as 10 classes used in the
process of measuring similarities. These objects, such as: bicycle, cow, flower, frangipani, grape,
horse, lovebird, orange, strawberry, and tree. In determining the 10 classes, the researcher conducted
an analysis based on the colors of the 10 objects above. Color based analysis is done by looking at the
color differences of the 10 objects, where researchers assume that the color is 10, the object has the
opposite color. Therefore, this research selects objects based on color difference analysis, with the aim
that researchers apply the color feature extraction method to the object.

2.1. Image retrieval
There are two ways that can be done in retrieving an image or image retrieval.

- Context Based is data retrieval by referring to the semantic content and this is related to the
  image, usually related to image descriptions such as keywords [11] [12] [13].
- Content Based is data retrieval by referring to image features such as color, texture, shape, or
  combination. And commonly referred to as Content Based Image Retrieval (CBIR).

The beginning of the use of the term Content-Based Image Retrieval in the literature was Kato to
describe experiments into automatic retrieval [14]. Kato is taken from an image from a database based
on color and shape features. In Content Based Image Retrieval, visual characteristics of each image
input are extracted and described as multi-dimensional vector features. These features are then stored
in a database. To perform image retrieval, the user gives an image query on the search system and then performs the feature extraction process to query the image so that the process is obtained by the vector feature value. Vector feature values obtained from image queries will be compared in common with vector feature values contained in the database. Images that are considered similar are images that have the same vector feature value.

Figure 2. Illustration Content Based Image Retrieval

In general, the workings of content based image retrieval can be illustrated in accordance with figure 2. The query image is inputted into the CBIR system then the feature extraction process is carried out by the system. The extracted image is then matched with the image data in the database or library and the image that is considered similar is then displayed as an output image.

2.2. Color histogram
Color histogram is a graph that describes the spread of pixel intensity values of an image or a certain part of the image. With a histogram, we can know the relative frequency of occurrence of the intensity in the image. In addition, information about brightness and contrast can be found through histograms. Color histogram is the most widely used feature to present the color characteristics of an image. The color that is often used is RGB (Red, Green, Blue) [15] [16]. If each color is discounted to m intervals, the total number of discretions is m^3. A color histogram H (M) is a vector (h1, h2, ..., hn) where hn shows the number of pixels in image M in bin n [17]. The vector feature is stored as an index of the image. In the picture below it is the result of a color histogram calculation. Figure 3 shows the different degrees of red, green, and blue colors.

Figure 3. The color histogram calculation results on the image;
(a) orange (b) grape (c) strawberry

2.3. Color moment
Color moments are a process of color features by measuring image fermentation. When calculating, a moment is generated from a measurement for color similarity between images. Then it is compared to indexed image values in the database for image retrieval. In the color moment there is an assumption that the color distribution in the image can be interpreted as a possible distribution. There is a unique value in Probable Distribution which will
later be used as a color based image feature (example: normal distribution is distinguished by the mean and variance). There are 3 main moments in the image color distribution, namely Mean (equation 1), Standard Deviation (equation 2) and Skewness (equation 3) [6] [18] [19].

\[
\mu_{cm} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} x_{ij}^{cm} \tag{1}
\]

\[
\sigma_{cm} = \left[ \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij}^{cm} - \mu_{cm})^2 \right]^{1/2} \tag{2}
\]

\[
\theta_{cm} = \left[ \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij}^{cm} - \mu_{cm})^3 \right]^{1/3} \tag{3}
\]

Where \( \mu_{cm} \) is the average color value in the image, \( \sigma_{cm} \) is the square root of the variant and \( \theta_{cm} \) is a measure of the degree of asymmetry. While \( N \) is the total number of pixels in the image and \( x_{ij}^{cm} \) is the value of the \( i \) color component at pixel \( j \).

2.4. Similarity measurement

The last stage in the retrieval system is the search for similarities between query images and features of images that have been stored in the database. In search of similarities or in other words is the measurement of similarity, the researcher implements the euclidean distance method. The euclidean distance method or is the distance between two points that can be measured and produced based on two vectors.

Based on the two vectors, the distance between the two vectors can be calculated using equation 5 [20]. Whereas for \( n \) dimensions of vector space, the euclidean distance distance is determined using Equation 6. Where \( d \) is a measure of the distance between querying images in the library.

\[
d(A,B) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{5}
\]

\[
d(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{6}
\]

3. Experiment and results

In this research, it will describe the trial stages and also the results of the trials that have been conducted. In this research, 50 trial data have been tested. 50 test data contains 5 pieces of data for each object or class. Whereas it has been explained in session 2 that this study has 10 classes that have been previously defined in each class representing what objects.

Based on the 50 test data above, two color feature extraction methods were applied, namely the color histogram and color moment. Some results of the trial by applying the color moment are shown in Table 1. Whereas when implementing the color histogram, some results of the trial can be seen in Table 2. The results of the test on 50 test data show that the application of the color moments method is better than using the color method histogram. However, some results indicate that there is test data that is correct in the right class in the color moment method but not in the color histogram method and vice versa as well.

**Table 1. Data result with color moments**

| Object data testing | Training and class | Exp. Class | Valid resultclass |
|--------------------|--------------------|-----------|------------------|
| ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image4.png) | Min | Class | Valid |
Table 2. Data result with color histogram

| Object | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | Class | Exp. | Valid |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|------|-------|
| Bycycle_1 | 1230 | 3638 | 12443 | 6496 | 7023 | 1724 | 8755 | 13126 | 3020 | 7214 | 1230 | 1 | Y |
| Cow_2    | 3418 | 2912 | 9966 | 4083 | 4679 | 3251 | 11432 | 10940 | 2032 | 4724 | 2032 | 9 | N |
| Flower_3 | 5900 | 4522 | 8327 | 3931 | 5020 | 5720 | 13497 | 9591 | 4378 | 4949 | 3931, | 4 | N |
| Frangipani_2 | 8373 | 6177 | 5310 | 2516 | 2886 | 8170 | 16537 | 6581 | 6025 | 3043 | 2516, | 4 | Y |
| Grape_4  | 5819 | 4199 | 7364 | 2359 | 2027 | 5520 | 14031 | 8557 | 3478 | 2179 | 2027, | 5 | Y |
| Horse_4  | 5243 | 4325 | 8639 | 3866 | 3888 | 4979 | 13075 | 9763 | 3627 | 3975 | 3627, | 9 | N |
| Lovebird_2 | 32133 | 30438 | 25119 | 27148 | 28150 | 31710 | 38376 | 25991 | 30467 | 27477 | 2511, | 3 | Y |
| Orange_2 | 12080 | 9478 | 5815 | 7023 | 6801 | 11614 | 19761 | 5217 | 9516 | 6599 | 5217, | 8 | Y |
| Strawberry_2 | 4208 | 3173 | 9159 | 3799 | 3662 | 3843 | 12279 | 10050 | 1636 | 3881 | 1636, | 9 | Y |

4. Conclusion

The purpose of this study is to make an approach in data search based on the characteristics of the image itself. Characteristics of images can be taken from the characteristics of color, shape, texture or even the combination. In this study, an approach was carried out by applying two color feature extraction methods to obtain the characteristics of the image based on the characteristics of the colors possessed by the object. The trial was conducted using 50 test data where there were 5 pieces of data in each class. In this study, 10 classes have been established in which the class is divided based on the analysis of the color differences of the objects that the researcher selected. The object is like bicycle, cow, flower, frangipani, grape, horse, lovebird, orange, strawberry, and tree. From the results of the trials that have been conducted, it shows that the application of the color moments method gets better results than the color histogram method. There are some weaknesses in this study is the amount of learning data and test data used are still categorized a little. For further research, it is necessary to make improvements by applying different methods and increasing the amount of learning data to improve the trial results. One of the studies that will be carried out in the next research is trying to apply feature extraction and selecting other objects. The feature extraction that might be used in the future is applying form and texture feature extraction because color feature extraction has been applied in this study.
5. References

[1] R. d. S. Torres and A. X. Falcão, "Content-Based Image Retrieval: Theory and Applications," RITA. vol. 13, no. 2, pp. 161-185, 2006.

[2] A. Singla and M. Garg, "CBIR Approach Based On Combined HSV, Auto Correlogram, Color Moments and Gabor Wavelet," International Journal Of Engineering And Computer Science, vol. 3, no. 10, pp. 9007-9012, 2014.

[3] F. A. Mufarroha, F. Utaminingrum, and W. F. Mahmudy, "Segmentation Algorithm to Determine Group for Hand Gesture Recognition," Journal of Telecommunication, Electronic and Computer Engineering, vol. 9, pp. 13-17, 2017.

[4] F. A. Mufarroha and F. Utaminingrum, "Hand Gesture Recognition using Adaptive Network Based Fuzzy Inference System and K-Nearest Neighbor," International Journal of Technology, vol. 8, no. 3, pp. 559-567, 2017.

[5] S. M. B and B. M. Naik, "Content Based Image Retrieval Using Texture Structure Histogram and Texture Features," International Journal of Computational Intelligence Research, vol. 13, pp. 2237-2245, 2017.

[6] V. Anusha, V. U. Reddy, and T. Ramashri, "Content Based Image Retrieval Using Color Moments and Texture," International Journal of Engineering Research & Technology (IJERT), vol. 3, no. 10, pp. 9007-9012, 2014.

[7] S. Kumar and A. Chauhan, "Feature Extraction Techniques Based on Color Images," in Special Conference Issue: National Conference on Cloud Computing & Big Data, 2013, pp. 208-214.

[8] J. Poco, A. Mayhua, and J. Heer, "Extracting and Retargeting Color Mappings from Bitmap Images of Visualizations," IEEE Trans. Vis. Comput. Graph., vol. 24, no. 1, pp. 637-646, 2018.

[9] S. Saad, H. Kumar, and T. K. Tewari, "Efficient Content Based Image Retrieval using SVM and Color Histogram," in Proceedings of 2017 Tenth International Conference on Contemporary Computing (IC3), Noida, India, 2017, vol. 1, pp. 1-3.

[10] M. P. S. Jadhav and D. S. Jadhav, "Video Summarization Using Higher Order Color Moments (VSUHCM)," vol. 45, Procedia Computer Science, 2015, pp. 275-281. [Online]. Available.

[11] Y. Mistry, D. T. Ingle, and M. D. Ingle, "Content based image retrieval using hybrid features and various distance metric," Journal of Electrical Systems and Information Technology, vol. 5, no. 3, pp. 874-888, 2018.

[12] K. N. Sukhia, M. M. Riaz, A. Ghafoor, S. S. Ali, and N. Iltat, "Content-based histopathological image retrieval using multi-scale and multichannel decoder based LTP," Biomedical Signal Processing and Control, vol. 54, p. 101582, 2019/09/01/ 2019.

[13] Content based image retrieval using image features information fusion, 51, 2019.

[14] T. Kato, "Database architecture for content-based image retrieval," in image storage and retrieval systems, 1992, vol. 1662: International Society for Optics and Photonics, pp. 112-123.

[15] I. Leichter, M. Lindenbaum, and E. Rivlin, "Mean Shift tracking with multiple reference color histograms," Computer Vision and Image Understanding, vol. 114, no. 3, pp. 400-408, 2010.

[16] Y. Yalman, "Histogram based perceptual quality assessment method for color images," Computer Standards & Interfaces, vol. 36, no. 6, pp. 899-908, 2014.

[17] C. A. Hussain, D. V. Rao, and T. Praveen, "COLOR HISTOGRAM BASED IMAGE RETRIEVAL," International Journal of Advanced Engineering Technology, pp. 63-66, 2013.

[18] Pushpalatha.K.R and Chaitra.M, "Color Histogram based Image Retrieval – A Survey," International Journal of Advanced Research in Computer Science, vol. 4, pp. 119-126, 2013.

[19] J. K. Patil and R. Kumar, "Plant Leaf Disease Image Retrieval Using Color Moments," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 2, pp. 36-42, 2013.

[20] D. Kurchaniya and P. K. Johari, "Analysis of Different Similarity Measures in Image Retrieval Based on Texture and Shape," International Research Journal of Engineering and Technology (IRJET), vol. 4, no. 4, pp. 2137-2143, 2017.