COMPUTOLOGY: Journal of Applied Computer Science and Intelligent Technologies
Volume 2, Issue 2, Jul-Dec 2022, pp. 23-32
DOI: 10.17492/computology.v2i2.2203
https://www.journalpressindia.com/cjacsit

A Content Based Image Retrieval Mechanism Based on Primitive Features

*Sumit Kumar**

ABSTRACT

Content-based image retrieval (CBIR) is one of the most practiced areas in recent times. In a CBIR system, each image is represented using its primitive features like color, texture, and shape. These primitive features are extracted in various ways to form the final feature vector and used for image retrieval based on a selected similarity measurement. The size of a feature vector is also one of the deciding factors in computation time. Hence, smaller-sized feature vectors with comparable results are always advised. In this paper, to make the system closer to human perception, we have used the HSV counterpart. As all the pixel values do not contain vital information and unnecessarily increase computational overhead. Hence, in this work, we have used mid-rise quantization. In this work, we have used local statistical parameters to evaluate the color information. For texture extraction, GLCM is employed, and the shape feature is extracted using adaptive tetrolet transformation. We have validated our system using various widely accepted benchmark databases. The retrieved results are capable enough to demonstrate its improvement over other works.

Keywords: Content-Based Image Retrieval (CBIR); Mid-Rise Quantization; GLCM; Tetrolet;

1.0 Introduction

In our digital age, smartphones, and cameras are the most used equipment as they are available to almost everyone. This eventually produces a high volume of multimedia data like video, audio, and images. But, in today's environment, everyone is very keen to have selfies all the time. This makes images a major part of multimedia data. On a local or distant server that a third party owns, this created data is stored. However, in either scenario, it might be used shortly. Thus, it becomes challenging to locate retrievable images that are comparable to a query image. The first stage of image retrieval is descriptive, meaning that either the owner of the image or the CBIR service provider has given each image a description. This method is no longer adequate for a number of reasons. Since so many photographs are currently available, explaining them is almost impossible. This procedure could go longer than expected. Second, every individual thinks differently. As a result, several techniques might be found to illustrate or annotate a single photograph. Different people have different ways of conceptualizing and thinking. Text-based image retrieval (TBIR) is the name of this procedure [1].

Academics and researchers begin concentrating on retrieval methods based on the characteristics of the image to get around this. The official name of this method is content-based image retrieval (CBIR). [3, 2].

**Assistant Professor, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India (E-mail: sumitvarshney68@gmail.com)*

In CBIR, all of the images in the chosen image database are mined for their basic visual image attributes. There are two steps to the CBIR process. In the initial stage, one of the primitive visual features. These primitive visual features are texture, shape, and color. Now, the selection of feature is selected based on the realization of the application. When CBIR came into picture, only feature based works are executed. But, as the volume of images start growing then various combinations are indulged. All of the photos utilized in the retrieval system have the same features retrieved from them. Global and local image feature extraction methods are used in CBIR.

The chosen visual feature is taken from the entire image for global feature extraction. Although this extraction method is quick, the characteristics produced are not closely related to the true core of the image. Local picture features become essential as a result. In local image feature extraction, the image is first separated into a number of non-overlapping blocks. From each block, the desired feature is then extracted individually. Finally, all the values are merged to create the feature vector. There is always a trade-off between the time needed and the system's efficiency because local picture feature extraction requires some processing time. As a result, choosing the characteristics becomes exceedingly difficult. The second stage of the CBIR process is finding the similar images based on a selected distance based similarity measurement. The selection of similarity measurement depends many times on the area of application.

1.1 Motivation

For an efficient CBIR system, the actual content of the image must be closely correlated with the extracted features. As one know, there are two ways to extract the features and both have their pros and cons. In this work, we have tried to take the advantage for both. Hence, for the color features, we have extracted various statistical features from block wise. For texture and shape feature, we have extracted global features. We have deployed GLCM so that, we can extracted texture features in various directions. For the shape feature, we have tried to get the object of the image in such a way that it can approximate detect the prominent object. Therefore, in this paper, we have used tetrolet transformation to approximate the object. Later, various statistical parameters are used to form the shape based features.

1.2 Contribution of the Paper

The major contributions of this paper is as follows:

- Color, texture, and shape based features are incorporated.
- Both directionality and orientation-based features are employed for texture information.
- To make the system closer to human perception, we have converted the input image into the HSV counterpart.
- For texture feature, GLCM, and shape feature is extracted using adaptive tetrolet transformation.
- The proposed system has been tested on various benchmark databases.

2.0 Preliminaries

We have briefly discussed the techniques indulged in various feature extraction processes.

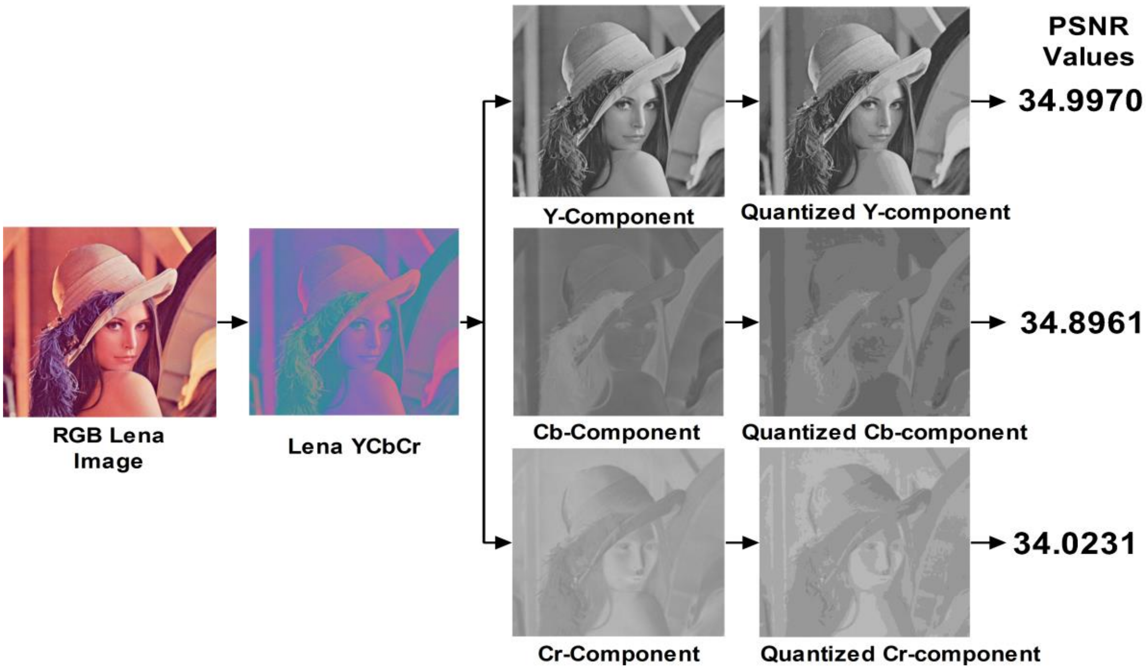
2.1 Mid-rise quantization

Since we adopted 16-bin quantization [4] in this suggested work, the entire range of pixel values has been represented using only 16-pixel values. Here the pixel intensities are divided into 0-15, 16-31, and carry on. There are three photos in figure 1. Here, we have demonstrated PSNR values of all three components of a YCbCr Image. A PSNR value reflects the similarity between two comparable images. As one can see from the figure, various PSNR values are on the higher end. This shows that we are able to preserve the vital content of the image even after using minimal pixel intensities.

Figure 1: Various Quantized Value

$$Q = \begin{cases} 8 & \text{if } q \in \{0-15\} \\ 24 & \text{if } q \in \{16-31\} \\ 40 & \text{if } q \in \{32-47\} \\ 56 & \text{if } q \in \{48-63\} \\ 72 & \text{if } q \in \{64-79\} \\ 88 & \text{if } q \in \{80-95\} \\ 104 & \text{if } q \in \{96-111\} \\ 120 & \text{if } q \in \{112-127\} \\ 136 & \text{if } q \in \{128-143\} \\ 152 & \text{if } q \in \{144-159\} \\ 168 & \text{if } q \in \{160-175\} \\ 184 & \text{if } q \in \{176-191\} \\ 200 & \text{if } q \in \{192-207\} \\ 216 & \text{if } q \in \{208-223\} \\ 232 & \text{if } q \in \{224-239\} \\ 248 & \text{otherwise} \end{cases}$$

Figure 2: PSNR Values for the Quantized Image



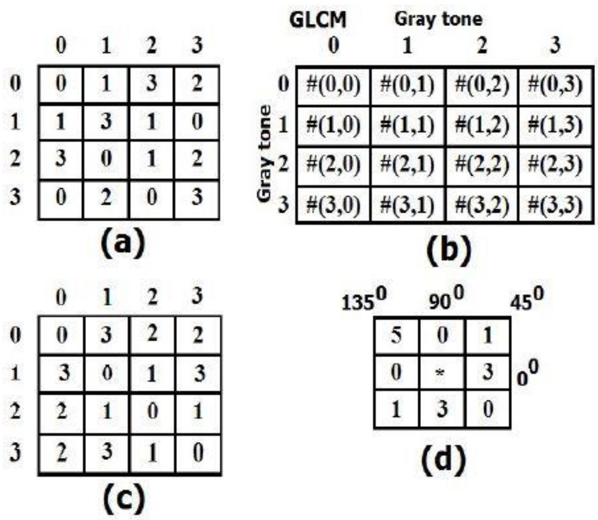
The reconstructed and reference images are sufficiently similar, as evidenced by the PSNR value.

2.2 Grey level co-occurrence matrix (GLCM)

In this section, we have summarized the use of GLCM [5] in this paper. GLCM is one of the second-order statistical parameters that has a long history of use in digital image processing. In addition to calculating the frequency of a particular pixel pair in a picture, it establishes the spatial relationship between pixels [?].

Assume that IMG is the input image and that CELLh and CELLv represent the horizontal and vertical cells, respectively. Now that the image's gray level has been quantized, there is a predetermined number of gray levels, G. Let The total number of cells in the input picture IMG, which is structured in a row-column way, is given by multiplying Lhor, which represents the horizontal spatial domain, by Lver, which represents the vertical spatial domain. Image's final GLCM matrix for the four different directions [?] Our method used a unit distance from any pixel to assess the spatial relationship. Figure 3 for horizontal direction is a graphic representation of the GLCM technique. The two parameters (i) relative distance (d) and (ii) relative direction (ϕ), where ϕ = 0o, 45o, 90o, 135o, are used to evaluate GLCM. Now, the probability of conditional co-occurrence can be calculated as

Figure 3: (a) Input Picture that is a 4 × 4 Matrix, (b) the Basic GLCM Matrix, (c) Spatial Occurrence for the Horizontal Direction, and (d) the Input



$$GLCM = \frac{1}{4}\{C_{0^\circ,d} + C_{45^\circ,d} + C_{90^\circ,d} + C_{135^\circ,d}\} \quad (1)$$

where d=1 in this scheme. where $C_{u,v}$ can be defined as:

$$C_{u,v} = \frac{O_{u,v}}{\sum \sum_{u,v=1}^{G} O_{u,v}} \quad (2)$$

where G is the quantization level or gray level and Ou, v is the frequency of the pixel pair (u,v) or (v,u).

2.3 Adaptive tetrolet transformation

Only rows and columns are used in the two directions of the discrete wavelet transformation (DWT). An image's geometry cannot be fully covered by just two orientations. directionlets, Curvelets, and Counterlets are just a few of the transform domain techniques that can be used to effectively incorporate a picture's local geometry. To effectively depict an image in this suggested manner, we used Adaptive tetrolet transformation (ATT). A unique variation of the standard Haar wavelet is the tetrolet. Golomb introduces tetrominoes in his paper. Jens Krommweh [6] proposed the ATT, which uses tetrominoes to represent an image effectively.

The low-pass sub-bands of an image IMG is divided into number of blocks BLKu,v of size 4 × 4. For each block BLKu,v, all 117 possible covering cov = 1, 2, 3....117 are implemented, and for each tiling t∗ 4 low-pass sub-bands (LP) and 12 high-pass sub-bands (HP) are obtained. These sub bands can be calculated as:

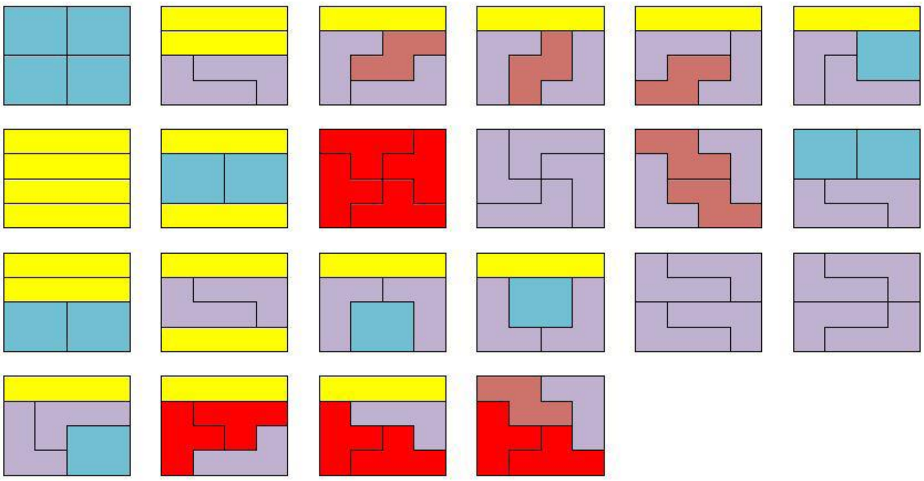
$$A^{r,(t)} = (IMG^{r,(t)}[G])_{s=0}^{3}$$

$$IMG^{r,(t)}[G] = \sum_{(u,v)\in IMG_s^{(t)}} \in [0, L(u,v)] IMG^{r-1}[u,v]$$

Only low-pass sub-bands are taken into consideration for the additional decomposition. All HP are simultaneously kept separately for texture analysis at each level.

Figure 4: 5 Basic Tetrominoes, O-I-T-S-L Tetrominoes

Figure 5: 22 Fundamental Tiling for 4 × 4 Board

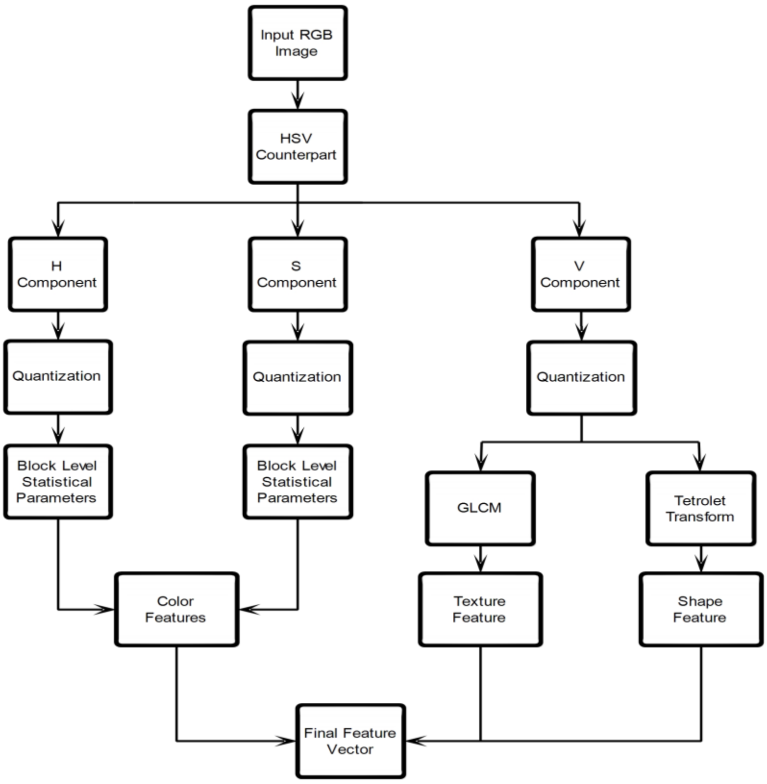


3.0 Proposed System

CBIR is one of the fatest growing area of research where various features like color, texture, and shape are extracted from the image in different ways. As, we have discussed earlier, in this paper we are considering only the gray scale images so here we have only texture and shape features. It is quite obvious that local image features are more associated with the actual content but its computational time is more hence to keep all these in mind, authors have proposed a novel work where global and local features are combinedly used to provide comparable results. The overall feature extraction process has been divided into two parts. In the first part, texture feature are extracted based on LBP and GLCM/. Second part comprises of shape features based on Tetrolet transformation and EDH. Finally, all the extracted primitive visual features are combined to established the final feature vector of the respective input image. The underline algorithm has been step-wise eleborated in Algo 1. The same algorithmic steps has been depicted in Figure 6.

Algorithm 1 Algorithimic Steps to Extract Features
Input: Input Gray Image QI .
Output: Final feature vector f vI .
Parameter: Size of QI is M × N × 3,
Step-1: Select a query input image QI of size M × N × 3.
Step-2: Convert the RGB image to the HSV counterpart.
Step-3: Break down into its fundamental elements.
Step-4: Analyze the various statistical parameters from the H and S components at block level for color feature extraction.
Step-5: To extract the Texture feature, we have evaluated GLCM parameters from the intensity component.
Step-6: For the shape features, we have incorporated multiple levels of adaptive tetrolet transform followed by some statistical parameters.
Step-7: Steps 4 through 6 should be combined in order to generate the final feature vector.

Figure 6: Feature Extraction Process



4.0 Results and Discussion

The experiment was carried out in the Matlab environment, which was set up on a computer running Windows 8 with an Intel(R) Core(TM) i7-4770 CPU clocked at 3.4 GHz and 4GB of RAM. We have taken data from four commonly used benchmark databases to validate our solution. Here are some quick facts regarding those databases: We used four common image databases, specifically:

1. GHIM-10000 (DB-1) [7]: It includes 10,000 unique images divided into 20 categories, including firecrackers, buildings, vehicles, horses, and insects, among others.
2. Sorted Produce-1400 (DB-2) [8]: It has 1400 unique photographs of 14 numerous kinds of food, including potatoes, apples, and more.
3. Oliva and Torralba Scene (OT-Scene) (DB-3) [9]: There are 2688 distinct photographs divided into 8 different categories of scenic beauty.
4. Coral 1000 (DB-4) [10]: has 1000 unique images in 10 categories, including dinosaurs, beaches, buses, and roses.

Figure 7: Some Pictorially Results from Various Image Databases

(a)

(b)

Three well-known and recognized parameters—precision, recall, and F-score—have been used to assess the suggested system's integrity. Out of all images that are recovered, precision is the percentage of images that are successfully retrieved. Recall can be summed up as telling you how many accurate images were found among all the pictures in a given category in the relevant database. They can be stated as follows:

$$Precision = \frac{X_{DB} \cap R_{DB}}{X_{DB}}$$

$$Recall = \frac{X_{DB} \cap R_{DB}}{R_{DB}}$$

The harmonic mean of precision and recall is known as the F-score. It provides you with a value that indicates the system's overall performance. This implies that the retrieval performance increases with increasing value.

$$F\_Score = \frac{2 \times Precision \times Recall}{(Precision + Recall)}$$

We have displayed some findings pictorially from the DB-1 in figure 7. Similarly, we have illustrated some results from the DB-4 image benchmark database in figure 7. In the Table 1, we have shown various results based on top-20 retrieved images of various parameters.

Table 1: Benchmark Databases' F-score, Precision, and Recall Averages

Database	Precision	Recall	F-score
DB-4	80.25	16.05	26.75
DB-3	84.70	4.82	9.14
DB-2	87.25	17.45	29.08
DB-1	68.50	2.76	5.27

Table 2: Result Comparison for Corel-1000 Image Database

	Precision	Recall	F-Score
[11]	71.04	14.208	23.68
[12]	53.04	10.608	17.68
[13]	48.81	9.762	16.27
[14]	79.5	15.9	26.5
[15]	76.9	15.38	25.63333
[16]	69.2	13.84	23.06667
[17]	75.59	15.118	25.19667
[18]	66.81	13.362	22.27
Proposed	80.25	16.05	26.75

Now, for the comparative study, the designed CBIR framework have compared our results with various similar papers in the Table 2. There are total eight works which we have taken into consideration for the comparison.

5.0 Conclusion

The unique CBIR approach introduced in this work depends on local feature extraction, requiring that each feature be retrieved from its corresponding component. To account for human vision, the RGB input image is first transformed into an HSV color image. An image feature is now extracted from its assigned component in order to decrease the overlap of picture information and improve the retrieval effectiveness. As a result, both color components are used to extract color information as well as the shape and texture elements of intensity. For the local color feature extraction, both chrominance components are split into blocks of n × n size, and a few statistical parameters are assessed. GLCM and adaptive tetrolet transformation are combined with statistical parameters for the shape and texture aspects. The final feature vector is created by merging all of the retrieved features. Utilizing the Euclidean distance as a measure of similarity, this feature has been utilized to find related photos. Four well-known benchmark picture databases have been used to validate the described system.

References

1. Duan, L., Li, W., Tsang, I. W. H., & Xu, D. (2011). Improving web image search by bag-based reranking. IEEE Transactions on Image Processing, 20(11), 3280-3290.
2. Varish, N., Kumar, S., & Pal, A. K. (2017). A novel similarity measure for content based image retrieval in discrete cosine transform domain. Fundamenta Informaticae, 156(2), 209-235.
3. Kumar, S., & Pal, A. K. (2016). A CBIR scheme using active contour and edge histogram descriptor in YCbCr color space. IJCTA, 9(41), 889-898.
4. Kumar, S., Pradhan, J., & Pal, A. K. (2021). Adaptive tetrolet based color, texture and shape feature extraction for content based image retrieval application. Multimedia Tools and Applications, 80(19), 29017-29049.
5. Kumar, S., Pradhan, J., & Pal, A. K. (2017, December). A CBIR scheme using GLCM features in DCT domain. In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-7). IEEE.
6. Krommweh, J. (2010). Tetrolet transform: A new adaptive Haar wavelet algorithm for sparse image representation. Journal of Visual Communication and Image Representation, 21(4), 364-374.
7. Liu, G. H., Yang, J. Y., & Li, Z. (2015). Content-based image retrieval using computational visual attention model. pattern recognition, 48(8), 2554-2566.
8. Kumar, S., Pal, A. K., Islam, S. H., & Hammoudeh, M. (2021). Secure and efficient image retrieval through invariant features selection in insecure cloud environments. Neural Computing and Applications, 1-26.
9. Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision, 42, 145-175.
10. Li, J., &Wang, J. Z. (2006, October). Real-time computerized annotation of pictures. In Proceedings of the 14th ACM international conference on Multimedia (pp. 911-920).
11. Vimina, E. 1., & Divya, M. O. (2020). Maximal multi-channel local binary pattern with colour information for CBIR. Multimedia Tools and Applications, 79(35-36), 25357-25377.
12. Dubey, S. R., & Mukherjee, S. (2020). Ldop: local directional order pattern for robust face retrieval. Multimedia Tools and Applications, 79, 6363-6382.
13. Dubey, S. R. (2019). Local directional relation pattern for unconstrained and robust face retrieval. Multimedia Tools and Applications, 78(19), 28063-28088.

14. Varish, N., Kumar, S., & Pal, A. K. (2017). A novel similarity measure for content based image retrieval in discrete cosine transform domain. Fundamenta Informaticae, 156(2), 209-235.
15. Raghuwanshi, G., & Tyagi, V. (2016). Texture image retrieval using adaptive tetrolet transforms. Digital Signal Processing, 48, 50-57.
16. Vimina, E. R., & Jacob, K. P. (2019). Feature fusion method using BoVW framework for enhancing image retrieval. IET Image Processing, 13(11), 1979-1985.
17. Khokher, A., & Talwar, R. (2017). A fast and effective image retrieval scheme using color-, texture-, and shape-based histograms. Multimedia Tools and Applications, 76, 21787-21809.
18. Galshetwar, G. M., Waghmare, L. M., Gonde, A. B., & Murala, S. (2017). Edgy salient local binary patterns in inter-plane relationship for image retrieval in diabetic retinopathy. Procedia computer science, 115, 440-447.