International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2390291

# EDBC Algorithm used for Content-Based Image Retrieval

**Vishma Kumar Karna, Shatendra Dubey**

Department of Information Technology, NRI Institute of Information Science and Technology, Bhopal, India

ARTICLEINFO

**Article History:**

Accepted: 15 April 2023

Published: 07 May 2023

**Publication Issue**

Volume 9, Issue 3

May-June-2023

**Page Number**

47-56

ABSTRACT

The tremendous increase and om- nipresent accessibility of graphic documents on the network led to the high interest in research on content-based image retrieval (CBIR). This has ce- mented the approach for a massive sum of innovative procedures and schemes, and growing curiosity in allied fields to upkeep such projects. Existing associ- ated theories include efficient Content-based Image Retrieval (CBIR) frame by enacting the content- based image, K-means and hybrid clustering is func- tional over combined lineament vector of information images, texture features. In similar cases it is tight in expressing the user's semantic Intention knowledge to permit information distribution and reuse, models ought to be managed within repositories, where they might be retrieved upon users' queries. There is still a lack of adequate tools for incisive/handling visual content. In this paper, a novel algorithm Efficient Density-based Clustering Algorithm (EDBC) is sug- gested for content-based image retrieval technique that will enhance scalability and lower maintenance costs significantly, enhance the efficacy of software development.

**Keywords:** Content-based image retrieval, EDBC, Clustering.

## I. INTRODUCTION

Billions of people are projected to the network sharing and browsing photos, with the stellar rep- utation of digital devices embedded with cam- eras and fast growth of Internet technology. The omnipresent contact to both digital pictures and the Internet shacks bright light on many emerging applications based on image search. To retrieve vital visual documents to a textual or visual query efficiently from a large-scale visual corpus is the target of image search [1].

Digital images and video or visual objects are becoming as significant as traditional textual based information, as the information technology pro- liferates throughout our society. This occurrence has several explanations: demilitarization in the expertise of satellite and imaging, the beginning of World Wide Web (WWW) as an infrastructure for digital communications, the approaching conjunc- tion of computers and television, and the upsurge in usage and availability of digital cameras and video recorders [2].

Currently, researches in the field of investiga- tion, organization, and retrieval of images from the sizeable visual database were most common amid image processing researches. Some of the practical reasons for this phenomenon are, the problems related with retrieval of an image is of more significant interest as digital cameras take a cheap image and vast aggregates of new images become accessible on the web each day. However, the difference is that still there exists a lack of ad- equate tools for searching/handling graphical con- tent. Secondly, as it reaches a substantial amount of unsettled challenging queries in understanding the image, the difficulty in retrieval is also motivating at a deeper level. Finally, graphic records offer a different testing ground in assessing the ideas for handling the image, where to assign stern assumptions about scene or imaging conditions or testing an algorithm on fewer pictures alone are not satisfactory [3]. Utilizing considerable progress in the available quantity of data related to visual state, there occurs a real need for systems to record and deliver reclamation from libraries containing digital pictures and video. Concerning the inter- ests of users', recovery of images through search from an extensive database of images using visual contents was utilized. This technique is known as Content-based image retrieval. A most precise content-based image retrieval outcome and pre- processing of the dimension database of an image using Efficient Density-based Clustering (EDBC) for operational recovery system for the picture. An efficient and more precise Content-based image retrieval technique through EDBC mechanism was projected for enhancing Scalability.

## II. LITERATURE REVIEW

For the past three decades, CBIR has been a dy- namic area of research owing to an extensive range of presentations in techniques related to retrieval of images. The verse "content-based" denotes the ex- ploration procedure that assesses the exact contents of pictures rather than utilizing outdated techniques of image explanation for recovery. The word "con- tent" in this context states the consistency, color, figure, or any additional data that might be obtained from the source itself [4].

According to the data mentioned above [[5], [6]], there are numerous categories of image recovery methods. Dissimilar attention based indicators and descriptors have been projected for extraction of features in a retrieval procedure [[7], [8], [9] ].

CBIR employs low-level attributes like shape, surface, dimension, etc. which portrays the con- tent of an image. However, finding a single best illustration of material is nearly impossible that may isolate diverse classes with very much char- acterized boundaries. Moreover CBIR which mea- sures the distance is not so effective while de- scribing low-level attributes. To overcome this, a new approach was established where CBIR might be achieved by implying a genetic algorithm that optimizes the loads gained from feature extrac- tions of a picture and integrating it with machine learning techniques. One such machine learning practice is clustering, which handles the ambiguity, uncertainty, and imprecision when accessing the multimedia data that addresses the constraints of both traditional distance-based metrics and classi- fier grounded recovery methodologies [10].

A CBIR system which combines the custom of Support Vector Machine (SVM) with color histogram was projected to achieve an effective procedure for retrieval of an image. A two-level ap- proach was utilized by initially filtering the appro- priate images applying SVM trailed by color based matching. According to the period of recovery and correctness of regained image, the evaluation of the suggested procedure was established [11].

Microstructure Descriptor (MSD) was developed by integrating colors and orientation of edges which seamlessly depicted the features of an im- age. To repossess the descriptions successfully, the technique assimilated color, texture, form and spatial information. Conversely, this methodology was insufficient for overall attributes of image and was incapable of exploiting relations amid sites of different objects [12].

An innovative system of CBIR based on a Scale-

Invariant Features Transform (SIFT) descriptor, a hue descriptor, and soft obligation were developed. The SIFT was utilized for recognizing crucial points, while local coverings around them were designated by using SIFT and hue descriptors. The different language was generated for every descriptor, and it was quantized through algorithm related to k-means clustering. In place of a stable assignment, a soft assignment was employed in this model to overcome the loss in quantization which might decrease the performance of recovery [13]. A new context [14] by examining the Particle Swarm Optimization Algorithm (PSOA) was estab- lished. It was noted that, to find the resemblances among the inquiry image and pictures from the catalog, this procedure extracted three sorts of attributes from every image, likely color, texture, and form. It employed an appropriate distance measure for every kind of structure utilized. The PSOA was assimilated to uplift the suggested prac- tice through discovering out adjacent significant groupings midst structures and their consistent re-semblance capacities [15], [16].

An inventive process [17] for CBIR by inte- grating color, spatial, and texture attributes of the source image was recognized. A structure trajec- tory was established by exploiting every structure. The CENsus transform hISTogram (CENTRIST) characteristic was employed for spatial assembly, and a principal component analysis (PCA) was practiced on CENTRIST for curbing the measure- ments. This algorithm incorporates diverse den- sity (DD) and multiple instance learning (MIL) to achieve objective occurrences. It produced im- proved fallouts when associated with contemporary CBIR techniques [18].

An outline for CBIR by creating a biased dom- inant color (DC) descriptor was established in which; the descriptor allocates biasing to every DC in the source with the purpose of extracting seman- tic features. This overcame the shortcomings of dominant color descriptor (DCD) and diminished the consequence of imaging background while a pictorial matching decision was thereby increasing the performance [19], [20].

For large-scale image retrieval, various binary structures at the level of indexing were entrenched where multi – IDF arrangement correlates be- tween structures. The procedure of embedding via Hamming was employed as matching verification method. With the purpose of lessening the conse- quence of improper recognition and enhance the exactness of graphical corresponding, SIFT de- scriptive words were unified with binary structures [21]. A CBIR system that incorporates an affine instant to define invariants occurring in indigenous zones of the source to obtain repossession of image was proposed. The generated instants were unified through a BoVW structure to make comprehensive attribute vectors. A system of three distinct strategy elements was utilized, where affine instants are totaled firstly, followed by invariants calculation over the consequences of actual source and in the final stage, the progression of standardization was implemented to upsurge the series of invariants. The second stage aims in refining the previous phase, whereas the final stage advances the out- comes of preceding segment [22].

A novel method of image representation, which is based on the construction of histograms over two rectangular regions of an image was proposed. Division of the image into two rectangular areas at the time of creation of histograms adds the spatial information to the BoVW model. The pro- posed image representation uses separate visual words for upper and lower rectangular regions of an image [23]. An extensive evaluation of visual descriptors for the content-based retrieval of remote-sensing (RS) images were presented. The assessment includes global hand-crafted, local hand-crafted, and convolutional neural networks (CNN) features coupled with four different content-based image retrieval schemes. The content of RS images might be quite heterogeneous, ranging from images containing fine-grained textures to coarse-grained ones or images containing objects [24].

## III. PROPOSED METHODOLOGY

A "Content-based image retrieval" system to search the similar images to a query image from an extensive image database using Efficient Density- based Clustering for efficient picture retrieval sys- tem was proposed. This proposed method was a dynamic and more accurate Content-based image retrieval method through the EDBC Mechanism for enhancing Scalability. The experiments were implemented in Java and conducted on an Intel Core 2.93GHz, 4GB RAM with Windows 10. The flow of data through the proposed technique is displayed in Figure 1.

Let the data set to be processed be denoted as D, the algorithm's clustering radius Eps , and the minimum number of objects in the neighborhood MinPts . Then, the following are some basic concepts of the algorithm [25].

1) Eps neighbouring area:
let p be the centre of a sphere in the dataset D . For data within the radius Eps of the object's area, a collection of points contained in the sphere is,

$$NEps(p) = \{ q \in D \mid dist(p, q) \leqslant Eps\} \quad (1)$$

The neighboring area is shown in Figure 2.

2) Density:
At the position of the data point in the data set D, the number of points, Num, contained in the neighborhood Eps is its density.

3) Core point:
At the position of data point p in the data set D, if the density (Num) in the neighborhood Eps satisfies Num $\geqslant$ MinPts, it is called a core point.

4) Border point:
At the position of the data point p in the data set D, if the density in the neighborhood Eps satisfies Num $\leqslant$ MinPts , but it is inside the sphere, it is called a border point.

5) Noise point:
All the objects other than the core and border points in D.

6) Direct density-reachable:
Given objects ,p, q $\in$ D, where there is a core point, and this is inside the Eps neighborhood of q, it is said that from
p to q is direct density-reachable, i.e.,
q $\in$ NEps(p), |NEps(p)| $\geqslant$ MinPts, as shown in Figure 3.

7) Density-reachable:
Given objects p1, p2, ....pn $\in$ D , where
p1 = q , pn = q , if pi+1 is direct density-
reachable from, then p is density-reachable from q. The definition is shown in Figure 4.

8) Density-connected:
Given objects p, q $\in$ D, if there is a point o $\in$ D that is density-reachable from p and q, then p and q are density-connected. The
definition is shown in Figure 5.

The various concepts were represented in the form of figures as based on the def- initions mentioned above, the idea of the EDBC algorithm is as follows. The search can start from the neighboring Eps area of any compelling data point. Given enough data points in the neighborhood ($\geqslant$ MinPts ) the cluster will expand or otherwise, a data point is temporarily marked as noise. This point can later be found in other neigh- borhoods Eps and characterized as part of a cluster. The neighborhood is also part of the group if a data object in a cluster is marked as a core. Thus, all the points found in the neighborhood, as well as the core neighborhood, are added to the group. This process is repeated until density-connected groups are entirely found. Finally, new and untreated points are retrieved and processed to find deeper clusters or noise. After all the objects in data set D are checked, the algorithm ends.

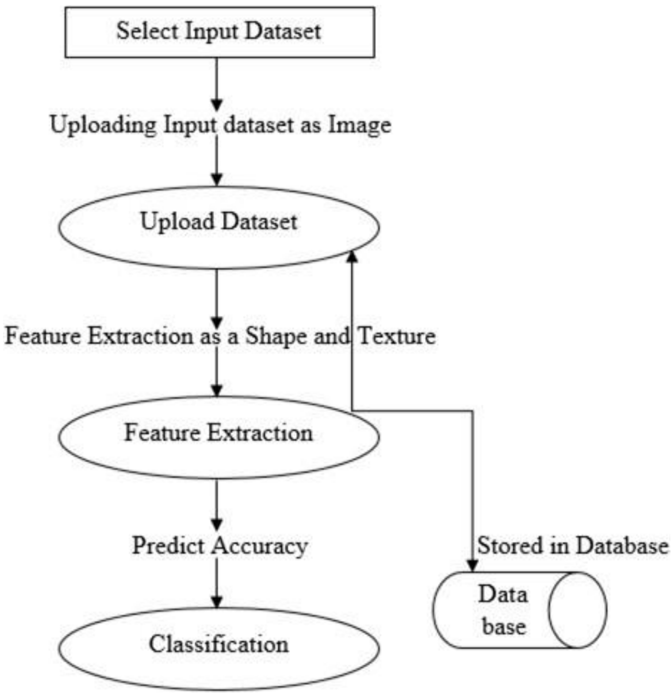


Fig. 1. Data Flow diagram

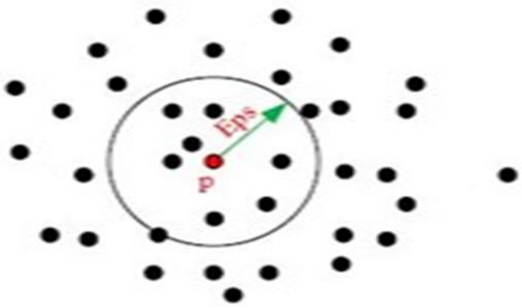


Fig. 2. Neighborhood definition

A. EDBC Algorithm:

According to previous clarification of simple ideas and notions of the EDBC algorithm, its processing flow can be summarized through subse- quent steps. Here, supposing the spatial dataset D, given clustering radius Eps , the minimal amount of neighboring objects MinPts, and the current assortment of entities as N1.

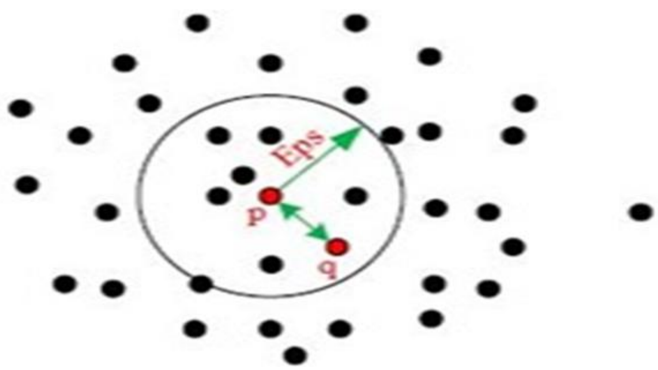


Fig. 3. Definition of direct density- reachable

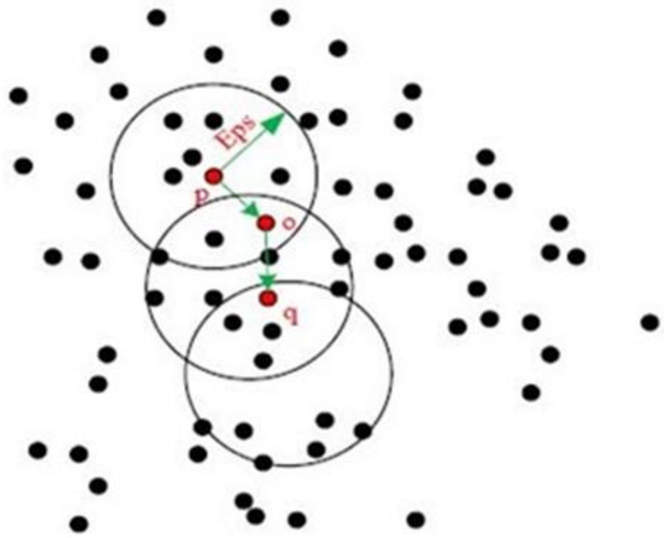


Fig. 4. Definition of density-reachable

The detailed implementation steps can be sum- marized in Algorithm 1.

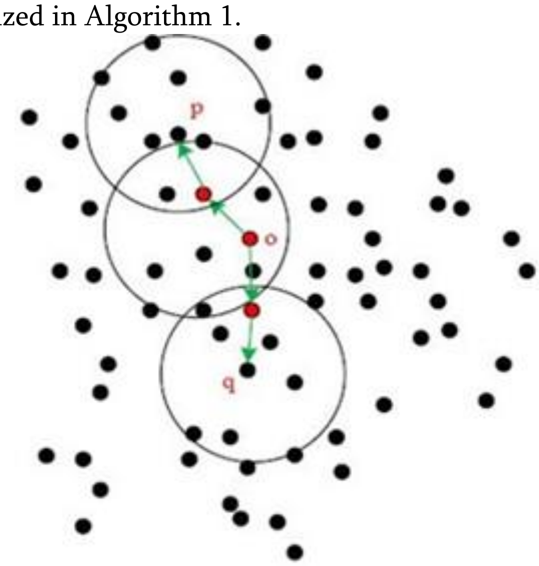


Fig. 5. Definition of density-connected

**Algorithm 1** : EDBC algorithm implementation steps.

**Input**: Data Set *D* to be processed.
**Output**: Cluster that satisfies the Clustering requirements.

**Parameters**:Clustering Radius *EPS*, Minimal number of neighbouring points *MinPts*.

```
1: Main function ()
2: {
3:  DensityCluster(D, Eps, MinPts )
4: {
```

```
ClusterNum = 0
for each unchecked point M in D do
   set M as checked
   NeighbourResult = NeighbourSearching(M
   ,Eps )
   if sizeof(NeighbourResult)= MinPts then
      ClusterNum = UpdateClusterNum
      expandNeighbourPart(M, NeighbourResult,
                              ClusterNum, Eps,
      MinPts )
   else
      set M as NOISE .
   end if
end for
}
}
Subfunction()
expandNeighbourPart(M, NeighbourResult,
   ClusterNum, Eps, MinPts )
NeighbourSearching(M, Eps )
```

*A.* Architecture

The architecture of this study comprises various modules of image retrieval that are utilized in a specific combination. The various modules that correspond to the system are

- Data set collection module
- Pre-Processing
- Testing and Training Data set
- Feature Extraction
- Classification

1) Module Description: The modules mentioned above refer to various subsets of architecture system which aids in retrieving the images through this recovery process. These modules are explained as,

**a. Dataset Collection:**

This module is utilized in collecting the unstructured datasets like images that can be employed for content with quality based retrieval process through data mining techniques.

**b. Pre-Processing:**

In this module, Noise Filtering is used to filter the unnecessary data from an image. It is also utilized in removing various types of noises from the images. In most of the cases, this feature is interactive. In general, Pre-Processing is the modification of an image by changing the brightness values of pixels to enhance the visual impact of images.

**c. Testing and Training Data set:**

This module deals with the testing of the contents that are being processed in the pre-processor unit. These data sets are then trained to possess the attributes that are to be utilized for retrieving the image.

**d. Feature Extraction:**

To describe the content of an image, the various features such as shape, texture, color, etc. are generally utilized where the features of the image can be classified into primitives. Here, the information from the raw data which is more relevant for discrimination amid the classes is extracted.

Extraction of features has to be made over low within-class variability and high between-class variability. Also, the redundant information has to be discarded during extraction.

**e. Classification:**

This deals with the process of utilizing samples of known informational classes, i.e., training sets to classify pixels of unknown identity. This type of classification deals with examining a large number of unknown pixels and divides it into some classes based on natural groupings present in image values. By employing all these modules, the accuracy of image recovery is in the course of development and ultimately provides an output with enhanced. Figure 6 depicts the architecture of this technique.

**Description:**

In this above the architecture diagram, to collect the unstructured data set like images for content with quality-based retrieval images using data mining approach Noise Filtering is used to filter the unnecessary information from an image. It is also used to remove various types of noises from the images. Mostly this feature is interactive. Pre-processing is the modification of image by changing the pixel brightness values to improve its visual impact Features such as shape, texture, colour, etc. are used to describe the content of the image. Image features can be classified into primitives. Extract the infor- mation from the raw data that is most relevant for discrimination between the classes. Extract features with low within-class variability and high between class variability discard redundant information. The process of using samples of known informational classes (training sets) to classify pixels of unknown identity. In this type of classification is a method which examines a large number of unknown pixels and divides it into number of classes based on natural groupings present in the image values.

Ultimately, our approach provides more accuracy.

## IV.IMPLEMENTATION

Firstly, to assess the developed process and equate it with advanced methods, 3 clusters of data sets are selected and tested in experiments. In these experiments, the suggested scheme is associated with every approach utilizing the similar dataset where every source was recognized as an inquiry image. If a recovered source is the same class as of the inquiry image, then, the method is identified as successful; otherwise, it fails. Here, the precision is utilized as a performance measure.

$$Precision = \frac{N}{N+M}$$

Where N = Number of Relevant Images and M = Number of Irrelevant Images.

The investigational outcomes and comparisons are presented in Table 1. Here, the table displays the comparison of fallouts for three image extrac- tion techniques. Figure 7 Represents the evaluation of the EDBC method with former methods.

TABLE I
COMPARISONS WITH THE OTHER METHODS

| | Image 1 | Image 2 | Image 3 |
|---|---|---|---|
| EDBC Method | 87.37% | 92.6% | 95.93 % |
| Others Method | 55% | 67.16% | 77% |

## V. RESULTS AND DISCUSSIONS

The outcomes obtained revealed that content- based image and data related to repossession of text enhances the scalability, lowers the maintenance costs significantly and progresses the efficacy of software development. The various interpretations achieved through the tests conducted were depicts the group of pictures obtained in the dataset.

This might be achieved over the suggested prac- tice of image retrieval employing EDBC procedure. The course of an assortment of an image from the available cluster of imaginings. This collection has to be made from the designated group only. It can be noted that an image should be selected by the user which acts as the input for the extraction procedure.

This section seems like the configuration win- dow in which, the operator has to inject his query or might search an image through this algorithm. The extraction of images according to the content through the data set. Here, three set of examples

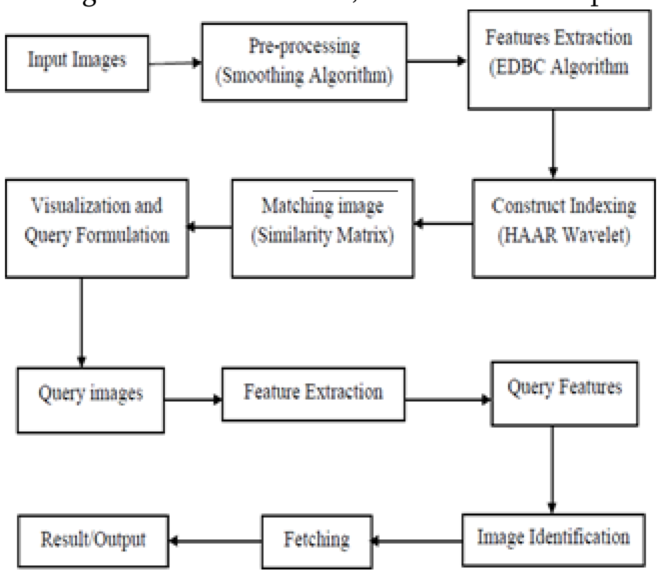


Fig. 6. Architecture diagram

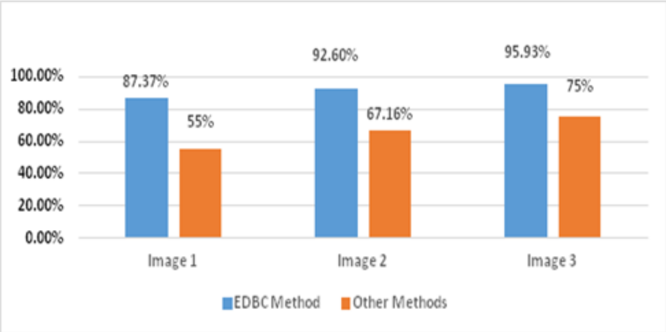


Fig. 7. The measure of EDBC method with former methods

were selected for the content based extraction of images from the given data set.

The Figure 8 characterizes the graphical demon- stration of content-based image extraction, which displays the assessment amid the entire sum of descriptions and the amount of extracted images.

From the obtained results, it was evident that, from this system of image recovery, the opera- tor can acquire a selective array of search data retrieved through the extraction process from the entire data set.

## VI. CONCLUSIONS

In early years, CBIR is one of the most dynamic study areas. Much such initial research effort fo- cused on determining the “best” attribute represen- tations of an image. Retrieval was executed as a totaling of resemblances of the discrete represen- tation of feature with secure weights. Whereas this method of computer-centric establishes the source of CBIR, the worth of such organizations was restricted as a measure of exertion in expressing high-level perceptions using low-level attributes and human discernment subjectivity. In this paper, we have introduced an efficient and more accurate content with texture-based image retrieval method through Efficient Density-based Clustering Algo- rithm (EDBC) mechanism for enhancing scalabil- ity. Here, Local features are extracted by Dense Matrix and global features are extracted by the color histogram. A massive sum of experiments has validated the proficiency and efficacy of the suggested methodology.

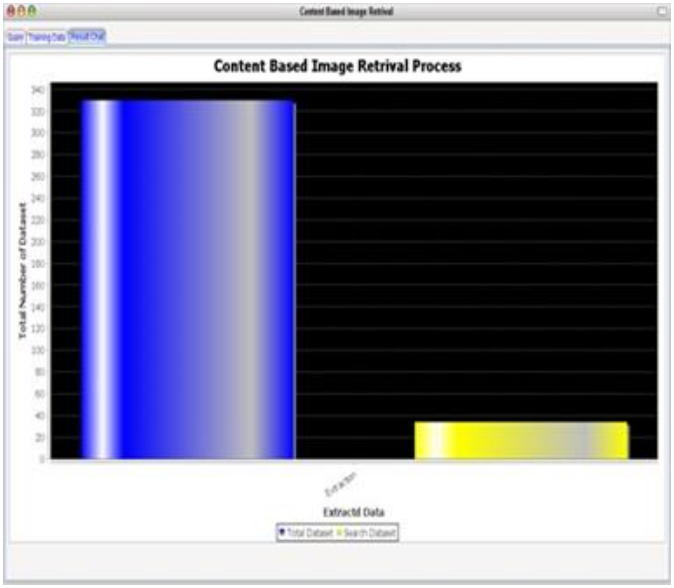


Fig. 8. Comparison graphs of total and search data sets

## VII. FUTURE WORK

In the future, researchers can use different kinds of features along with existing features and ma- chine learning techniques like neural network clas- sification in the successive recovery of media. One such direction for further investigation of this method is to customize this in handling other media types like video and audio through texture-based retrieval technique.

## VIII. REFERENCES

[1]. W. Zhou, H. Li, and Q. Tian, “Recent advance in content-based image retrieval: A literature survey,” arXiv preprint arXiv:1706.06064, 2017.

[2]. D. S. Shete, M. Chavan, and K. Kolhapur, “Content based image retrieval,” Internation- alJournal of Emerging Technology and Ad- vanced Engineering, vol. 2, no. 9, pp. 85–90, 2012.

[3]. S. Nirmal, “Content-based image retrieval techniques,” 3rd National Conference, Com- puting For Nation Development, year=2009.

[4]. J.-M. Guo, H. Prasetyo, and J.-H. Chen, “Content-based image retrieval using error diffusion block truncation coding features,” IEEE Transactions on Circuits and Systems for

Video Technology, vol. 25, no. 3, pp. 466–481, 2015.

[5]. T. Mahmood, T. Nawaz, R. Ashraf, M. Shah, Z. Khan, A. Irtaza, and Z. Mehmood, “A sur- vey on block based copy move image forgery detection techniques,” in Emerging Technolo- gies (ICET), 2015 International Conference on. IEEE, 2015, pp. 1–6.

[6]. T. Mahmood, T. Nawaz, Z. Mehmood, Z. Khan, M. Shah, and R. Ashraf, “Forensic analysis of copy-move forgery in digital im- ages using the stationary wavelets,” in Innova- tive Computing Technology (INTECH), 2016 Sixth International Conference on. IEEE, 2016, pp. 578–583.

[7]. D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International jour- nal of computer vision, vol. 60, no. 2, pp. 91–110, 2004.

[8]. J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide-baseline stereo from maximally stable extremal regions,” Image and vision computing, vol. 22, no. 10, pp. 761–767, 2004.

[9]. N. Dalal and B. Triggs, “Histograms of ori- ented gradients for human detection,” in Com- puter Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Confer- ence on, vol. 1. IEEE, 2005, pp. 886–893.

[10]. E. Naaz and T. A. Kumar, “Enhanced content based image retrieval using machine learning techniques,” in Innovations in Informa- tion, Embedded and Communication Systems (ICIIECS), 2017 International Conference on. IEEE, 2017, pp. 1–12.

[11]. S. Saad, H. Kumar, and T. K. Tewari, “Ef- ficient content based image retrieval using svm and color histogram,” in Contemporary Computing (IC3), 2017 Tenth International Conference on. IEEE, 2017, pp. 1–3.

[12]. G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, “Image retrieval based on micro-structure de- scriptor,” Pattern Recognition, vol. 44, no. 9, pp. 2123–2133, 2011.

[13]. N. S. Mansoori, M. Nejati, P. Razzaghi, and S. Samavi, “Bag of visual words approach for image retrieval using color information,” in Electrical Engineering (ICEE), 2013 21st Iranian Conference on. IEEE, 2013, pp. 1–6.

[14]. R. Saxena and S. Dey, “Light weight access control mechanism for mobile-based cloud data storage,” Int. J. Next Gener. Comput., vol. 9, no. 2, 2018. Online]. Available: http://perpetualinnovation.net/ojs/ index.php/ijngc/article/view/424

[15]. B.-M. Chang, H.-H. Tsai, and W.-L. Chou, “Using visual features to design a content- based image retrieval method optimized by particle swarm optimization algorithm,” Engi- neering Applications of Artificial Intelligence, vol. 26, no. 10, pp. 2372–2382, 2013.

[16]. A. Nazir, R. Ashraf, T. Hamdani, and N. Ali, “Content based image retrieval system by using hsv color histogram, discrete wavelet transform and edge histogram descriptor,” in Computing, Mathematics and Engineering Technologies (iCoMET), 2018 International Conference on. IEEE, 2018, pp. 1–6.

[17]. R. Saxena and S. Dey, “On-demand integrity verification technique for cloud data storage,” Int. J. Next Gener. Comput., vol. 9, no. 1, 2018. Online]. Available: http://perpetualinnovation.net/ojs/ index.php/ijngc/article/view/425

[18]. G.-L. Shen and X.-J. Wu, “Content based image retrieval by combining color, texture and centrist,” 2013.

[19]. A. Talib, M. Mahmuddin, H. Husni, and L. E. George, “A weighted dominant color descriptor for content-based image retrieval,” Journal of Visual Communication and Image Representation, vol. 24, no. 3, pp. 345–360, 2013.

[20]. D. C. G. Pedronette, J. Almeida, and R. D. S. Torres, “A scalable re-ranking method for

content-based image retrieval," Information Sciences, vol. 265, pp. 91–104, 2014.

[21]. L. Zheng, S. Wang, and Q. Tian, "Coupled binary embedding for large-scale image re- trieval," IEEE transactions on image process- ing, vol. 23, no. 8, pp. 3368–3380, 2014.

[22]. E. G. Karakasis, A. Amanatiadis, A. Gaster- atos, and S. A. Chatzichristofis, "Image mo- ment invariants as local features for content based image retrieval using the bag-of-visual- words model," Pattern Recognition Letters, vol. 55, pp. 22–27, 2015.

[23]. Z. Mehmood, F. Abbas, T. Mahmood, M. A. Javid, A. Rehman, and T. Nawaz, "Content-based image retrieval based on visual words fusion versus features fusion of local and global features," Arabian Journal for Science and Engineering, pp. 1–20, 2018.

[24]. Z. Mehmood, S. M. Anwar, and M. Altaf, "A novel image retrieval based on rectangular spatial histograms of visual words," Kuwait Journal of Science, vol. 45, no. 1, 2018.

[25]. M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al., "A density-based algorithm for discov- ering clusters in large spatial databases with noise."

**Cite this article as :**

Vishma Kumar Karna, Shatendra Dubey , "EDBC Algorithm used for Content-Based Image Retrieval", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.47-56, May-June-2023. Available at doi : https://doi.org/10.32628/CSEIT2390291
Journal URL : https://ijsrcseit.com/CSEIT2390291