Content-Based Image Retrieval Techniques: A Survey

M Sivakumar1*, N M Saravana Kumar2, N Karthikeyan1
1Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India
2Department of Computer Science and Engineering, M. Kumarasamy College of Engineering, Karur, Tamil Nadu, India
Email: *siva.recursion@gmail.com

Abstract. In these eras, digital day's cell smartphones and cameras are gaining a titanic reputation, resulting in the speedy growth of virtual images that are available throughout the internet. With an exponential boom within the image sizes, databases in different media lie an extremely good project for massive scale image search. Image retrieval from databases or the internet wishes an unsuccessful and effective technique because of the explosive increase of digital snapshots. Image recovery is known to be a comprehensive research area, particularly in content-based image retrieval (CBIR). CBIR retrieves comparable photographs from a huge photograph database primarily based on photo features, which has been totally lively study vicinity recently. The content materials that may be derived from images which include color, shade, texture, shape, are used for retrieving an image from the database. This paper will provide a survey and discuss the contemporary literature of various sorts of image retrieval (IR) structures and comparisons among them.

Keywords: Image retrieval, Image Filtering, Content-based image retrieval, image recovery, local binary pattern

1. Introduction
In the world of the internet, a need arises for storing a large amount of information. The advancements in the technological world like communication and the development in technical resources made a huge demand in the digital market. The e-data throughout the world brought platform for technical frauds and the hackers a chance to theft. Encryption is commonly used for solving threats in technological surroundings without decrypting. The robustness of content-based image retrieval offers several ways for intruders to target and attack images' features. CBIR's main advantages are that no special algorithm is needed to make the image more secured and for retrieving the image because the encryption and feature extraction is done individually. Some studies show that the bases on FCA implement the diverse social images [1].

Content-Based Image Retrieval (CBIR), as shown in Figure 1, is often known as image quality query (IQQ). CBIR is the image recovery process from a dataset or library of digital images according to digital visual material such as color, structure, texture, and spatial layout to represent and index [3] the image. The quest results in the content of the image rather than metadata being evaluated. Metadata comprises tags, keywords, and image property details. Using deep learning and neural network, the image can be retrieved [5] [6].

In many images, the texture character is an essential one. Photographs with microimages can achieve them. In a wide variety of graphic and image recognition applications, texture may play an important role. Since the 1960s, textural analysis has therefore become a focus of intense study. But
several of the techniques proposed were not adequately successful for actual textures. During the last few years, local binary patterns (LBPs) [2] were become highly descriptive and computationally efficient; this is the texture descriptors at local images, which have contributed to major developments in the application of texture techniques to various problems and applications [8]. The research on retrieving images starts from 2-D textures and space-time (dynamic) textures and moves based on 3-D textures. Figure 1 shows the content-based image retrieval [9].

Some system uses entropy-based image retrieval techniques [10]. It is not an efficient method to develop the system because entropy may change. Re-collect images based on attributes derived directly from the image metadata rather than keywords or annotations. Furthermore, the classification of photographs is specifically and implicitly related to different fields and personal, social, and occupational applications. Images raise the curiosity, illumination, elaboration, awareness, and clarity in the contact process.

![Figure 1: Content-Based Image Retrieval](image)

1.1 Texture Extraction
Homogeneity property is in the visual pattern, which is nothing but texture. The appearance of only one color or intensity does not result in this property. It has almost every surface, including nuclei, plants, rocks, hair, textiles, and is a natural property. It provides vital knowledge about the structural arrangement of surfaces and their interactions with their surroundings. In particular, it is possible to distinguish unique textures in the image with a two-dimensional grey level difference, which models texture. The relative luminosity of pixel pairs is measured to approximate the degree of contrast, regularity, curiosity, and directionality. The texture classification process involves two phases:
- Learning phase
- Recognition phase.

1.2 Color Extraction
Analyzing each image applied to the dataset and a color histogram [7] that indicates the proportion of each color in the image is measured. A colored space is required to be complete, compact, uniform, and user-oriented. To improve the color histogram effects on quantization, the color moments were proposed. The basis of this system is mathematical; by using moments, we define the color distribution. To improve the fast search on large-scale images, the system proposed the color set with a color histogram.
1.3 Shape Extraction
After images have been segmented into regions or objects, shape features are typically represented. Form’s classification state-of-the-art approaches can be separated into either boundaries or regions. The ultimate types are points, lines, planes, and conical parts, including Ellipses, Circles, and Parables. The outline is aligned with the other image images in a dataset.

Some shape descriptors include:
1.3.1 Curvature Scale Space (CSS): The MPEG-7 specification uses the CSS descriptor to include multi-size points that are in zero-crossing in a planar curve. The function’s dimension vectors differ in this regard for various contours images, such that two CSS descriptors must be matched by a specific matching algorithm.

1.3.2 Tensors Scales Descriptors (TSD): This is a tensor scale pattern that works on a morphometric parameter that unites local structured thickness, direction, and anisotropy representativeness. TSD is a descriptive form. In other words, the tensor is represented at every point of the photograph by the largest ellipse (2D) in the same homogenous area. The effect of TSD is to delete the parameters of the tensor scale for the first image and measure the histogram for the orientation in an ellipse. A correlation-based gap is used to compare TSDs.

2. Application of CBIR

2.1 Medical diagnosis
CBIR is an active area where it provides powerful services for biomedical information services nowadays. Three major domains will use CBIR strategies immediately: teaching, research, and diagnostics. For teaching the student using searching tools about the studied cases to locate the techniques. Several studies have image information for different type’s data. For instance, researchers are permitted to implement mining techniques to uncover peculiar image pattern between text and image material (e.g., care notes and patient records). The diagnostic method can also be aided by similarity queries dependent on image content descriptors. In their professional decision-making procedures, physicians typically use related cases for case-based purposes.

2.2 Biodiversity Information Systems
For biodiversity research, biologists collect several forms of data, including spatial data and pictures of life; some systems Biodiversity information system (BIS) help researchers develop textual, picture-based geographic inquiries. It will help the scientist to develop ecosystem species. An image is an input for the system, and the machine will produce the output with input images. This inquiry will be paired with textual and spatial forecasts by means of ‘drainage showing where fish with ‘wide eyes’ are joined to fins of fish that take the photographic form of the fins.’ In this sector, there are examples of initiatives.

2.3 Digital Libraries
A variety of digital libraries support image-based facilities. One example is the Digital Butterfly Museum that seeks to establish a Taiwanese digital butterfly collection. This Digital Library provides an image retrieval [4] module focused on color, texture, and patterns based on content. Zhu et al. offer a digital library for image retrieval focused on content promoting geography. In a separate image sense, Zhu et al. the framework handles air pictures that can be found by texture descriptors. Cross-referencing with a Geographical Name Information System (GNIS) gazetteer will include place names linked to the recovered photos. Some papers define the architecture for satellite imagery and video data from the heterogeneous archive array in this same domain. Other initiatives protect the numerous concepts of the CBIR. For example, while research is centered on new search techniques to improve CBIR systems' performance, another common topic is the offer of image descriptors.
3. Basic Local Binary Pattern (Lbp) Operators Overview

The LBP is expected to have two complementary elements, a trend, and its intensity locally. The pixel of an image is marked by the threshold of a binary number in each pixel's neighborhood. For texture definition, the distribution of the computed LBP mark over one area is then used. Figure 3 of the original LBP operator uses the core value of a threshold in a neighborhood of 3 to 3[1]. Pixel weights compound threshold values and an LBP code is given by summing the result. The contrast indicator C was taken from extracting from that of those above (or equal to) the middle pixel the total average grey level pixels are lesser than the threshold values. If the same value (1 or 0) is set to every neighbor in the center pixel, the value of C shall be 0.

A 3x3 neighborhood with the centrality of each pixel is labeled by the LBP operator as shown in Figure 2 and Figure 3, and the results are known as a binary integer. For marking with a binary integer, the decimal number is used. The extracted numbers in binary were known as LBP code. Whereas the operator uses pixel intensity content, the LDP operator uses neighborhood pixel edge response values and encrypts the image structure.

\[ \begin{array}{ccc}
  m3 & m2 & m1 \\
  m4 & X & m0 \\
  m5 & m6 & m7 \\
\end{array} \]

\[ \begin{array}{ccc}
  b3 & b2 & b1 \\
  b4 & X & b0 \\
  b5 & b6 & b7 \\
\end{array} \]

**Figure 2:** Edge Response and LDP Binary Bit Positions

- \[ \begin{array}{ccc}
  85 & 32 & 26 \\
  53 & 50 & 10 \\
  60 & 38 & 45 \\
\end{array} \] \[ \begin{array}{ccc}
  313 & 97 & 503 \\
  537 & X & 399 \\
  161 & 97 & 303 \\
\end{array} \] \[ \begin{array}{ccc}
  0 & 0 & 1 \\
  1 & X & 1 \\
  0 & 0 & 0 \\
\end{array} \]

**Figure 3:** LDP Code with k=3

LBP or C is used to extract the texture detection. In [4], the LBP operator is expanded to include districts of varying sizes. The use of a circular neighborhood and bilinear pixel-coordinate interplays in to develop the neighborhood of all pixel’s number and radius. The multi-scale LBP (MLBP) is developed by various radii operators. A supplementary comparison test can be taken from the grey variation of the local neighborhood. Several studies show that the pattern of LBP appears more on compared with others. With the help of this, the uniform patterns are established with a view to reducing pattern numbers. There was a broad use of uniform shapes, such as in the face definition, to minimize the vector’s length. In paper [10], the LBP operator has suggested completed modeling and a related completed texture classification scheme (CLBP). The variations between locals are separated into two supplementary components: the signs (like LBP) and the proportions. The complementary contrast LBP measures, the magnitude variable, offer an efficient substitute. In addition, they provide details on the strength of the image.

One of the most often used methods is the local binary pattern (LBP). It is also used to evaluate texture. The picture can be seen as a micro pattern composition that the LBP operator can explain well.
Break the panel into cells. Compare the pixel to each one of its eight neighbors for each pixel in a cell as represented in Figure 4. This gives a binary number of 8 digits.

4. Spatial Domain LBP Variants
The LBP approach can be changed to create it appropriate for the desires of various styles troubles due to its flexibility. To improve the LBP robustness and strength, many systems were developed. Some variants ought to more than a class, but here some of the best among them are taken into consideration. The selection of the suitable technique for the given system has a reply on factors of the device used.

5. Preprocessing
In many implementations, the input image must be pre-processed before extraction of the LBP function. For this reason, Gabor multi-scale filtering and rim detection were explicitly used. Earlier than the LBP estimation of face credibility, Gabor filtering was commonly used. One explanation for this is that Gabor and LBP-based methods provide additional facts: LBP tracks details at the time encoding with a larger spectrum of scales by Gabor filter. It tells about of appearance, as an example, [11] suggested extracting LBP characteristics from images generated by filtering an image with gabor different in scale and orientation with 40 filters.

Several studies show side detection had used previously for computation of LBP to decorate the records. Unique snapshots have low in-sensitive than gradient images. Possibly the first one changed into [13] presenting neighborhood part styles (LEPs) for use with capabilities of color for coloration of texture retrieval. To derive LEP and discover strong edges, the Sobel edge detection and thresholding techniques are used with LBP-like computation. [14] Proposed a methodology focused on the shooting with multi-scale warmth matrices of the intrinsic structural details on facial appearances. Warm kernels do well to characterize topological facial structural details. LBP histograms are then used for face classification calculated for non-overlapping blocks.

6. Neighborhood Sampling and Topology
A crucial issue that made the LBP flexible to distinctive various problems is that the in topology community LBP may be distinctive, relying on the wishes of the software given. The LBP extractions functions are typically finished in a round or rectangular community. A round community is vital in
particular for rotation invariant operators. In a few applications, face popularity, invariance isn't always essential; however, anisotropic records may be essential. It explores the use of different group topologies and encoding of their LBP research editions for the scientific image. An operator in an elliptic group with Quinary Encoding (EQP) had a reasonable efficiency. [12] LBP with elliptical sampling used rotation-invariant collectively for circular sampling in four directions to get records of anisotropic and iso tropics. A multi form LBP (Ms-LBP) was rendered using these types of an operator in the distinct step of a picture pyramid. The drawbacks of the technique are that large samples of the extracted macrostructures are wanted. [15] Taken into consideration special methods of using bit strings for the encoding of pixel patches, it records the pixel-based descriptors. So suggested an LBP (TPLBP) three-patch and an LBP four-patch (FPLBP).

An additional patch in a hoop of r-radius is considered for each pixel in TPLBP, uniformly arranged. Then, values are equivalent to those of the important patch for patches pair in the circle with an exact distance. The value of an unmarried bit is about in line with the two patches is just like the primary patch. The code produced will have S bits according to the pixel. In FPLBP, two jewelry focused at pixel had been applied in preference to one ring in tPLBP. Table 1 shows the various image retrieval methods.

Table 1: Comparison of Different Image Retrieval Techniques

| Ref. No. | Year | Methods Used                  | Merits                                      | Demerits                                    | Performance Metrics |
|---------|------|--------------------------------|---------------------------------------------|---------------------------------------------|---------------------|
| [11]    | 2017 | Clustering-based geometrical structure retrieval | Accurately retrieve the geometric structure | Performance of C-GSR method based on the selection of cutoff distance | Retrieved size (for Cylinder_0 canonical scatterer) = 19.5 Real size of Cylinder_0 canonical scatterer = 20 Retrieved size (for the Real size of Dihedral_0 canonical scatterer = 30 |
| [12]    | 2018 | Mind Camera                    | High average precision                      | Feedback is affected by the accuracy of tags | Average precision (@ top 25 returned images) = 0.892 |
| [4]     | 2018 | Convolutional Neural Network   | Achieved reasonable classification accuracy | CNN cannot be trained well without having enough number of data set and it is difficult to have enough data size for few cases | Retrieval error rate (test dataset) = 0.36% Retrieval error rate (real wafers) = 3.7% Retrieval time = 0.13 sec/wafer map |
| [8]     | 2018 | Unsupervised multi code hashing method | Does not required any labeled annotated image | By increasing hash bits (b), both the retrieval time and the memory size required for storing the hash codes increase | Recall (@b=8) =61.12% Time (@b=8) = 10 ×10−4s Storage (@b=8) =Complexity (@b=8) =0.061 KB Recall (@b=32) =65.29% Time (@b=32) = 41 ×10−4s Storage (@b=32) =Complexity (@b=32) =0.068 KB |
| [14]    | 2018 | Dynamic match                  | More effective                              | Additional time required for                | Mean Average Precision (mAP) (for Holiday's |


| Reference | Year | Methodology                                                                 | Evaluation                                                                 | Results                                                                 |
|-----------|------|-------------------------------------------------------------------------------|---------------------------------------------------------------------------|------------------------------------------------------------------------|
| [5]       | 2019 | Personalized Sketch-Based Image Retrieval                                    | Has strong generalization ability                                          | Mean Average Precision (MAP) = 0.6449                                   |
|           |      |                                                                               | Once the data relationship or data volume cannot meet the requirements, personalized SBIR will cause poor training results |                                                                          |
| [2]       | 2019 | Heterogeneity aware multi resolution Local Binary Pattern                    | Preserve more texture information                                          | Balanced Accuracy (BAC) = 0.8613                                        |
|           |      |                                                                               | Low F1 measure                                                             | F1 measure = 0.7665                                                    |
| [13]      | 2019 | Deep learning framework                                                     | High accuracy                                                              | Average error rate = 10.08% Time/second = 5.12 s Time efficiency = 0.176 |
|           |      |                                                                               | Mostly misidentification is occurred in those images with multiple categories of features | Classification accuracy = 98.56%                                       |
| [1]       | 2019 | Multi-modal procedure                                                        | Existing methods combined really smart and reliable route                  | P@20 (for First flickr images) = 0.4                                  |
|           |      |                                                                               | The weak point of multimodal procedure is the automatic choice of relevant images | P@20 (for text-based cluster images) = 0.4                              |
|           |      |                                                                               |                                                                          | C@20 (for First flickr images) = 0.2727                               |
|           |      |                                                                               |                                                                          | C@20 (for text-based cluster images) = 0.3636                           |
|           |      |                                                                               |                                                                          | C@20 (for visual cluster images) = 0.5455                             |
Comparisons of various methods in image retrieval with respect to retrieval time and retrieval accuracy are shown in Figure 5 and Figure 6.

7. Conclusion
In this article, an in-depth evaluation of image recovery based on changed strategies has studied and evaluated. It indicates that all the researchers expressed numerous methods for image retrieval, which will beautify the overall performance of the image relative to traditional approaches. Based on the analysis, it's far recognized that the multi-modal system has higher performance than the opposite techniques. The primary project of multi-modal manner is to select the applicable and non-relevant images as carefully as possible to humans' behavior, which is still a long way from being accomplished. After images have been segmented into regions or objects, shape features are typically represented. Form’s classification state-of-the-art approaches can be separated into either boundaries or regions. The ultimate types are points, lines, planes, and conical parts, including Ellipses, Circles, and Parables. The outline is aligned with the other image images in a dataset.

Re-collect images based on attributes derived directly from the image metadata rather than keywords or annotations. Furthermore, the classification of photographs is specifically and implicitly
related to different fields and personal, social, and occupational applications. Images raise the curiosity, illumination, elaboration, awareness, and clarity in the contact process.

References

[1] Benavent. X, Castellanos. A, De Ves. E, García-Serrano. A, & Cigarrán. J, "FCA-based knowledge representation and local generalized linear models to address relevance and diversity in diverse social images," Future Generation Computer Systems, 100, 250-265, (2019).

[2] Erfankhah. H, Yazdi. M, Babaie. M, & Tizhoosh. H. R, "Heterogeneity-Aware Local Binary Patterns for Retrieval of Histopathology Images," IEEE Access, 7, 18354-18367, (2019).

[3] Lai. H. P, Visani. M, Boucher. A, & Ogier. J. M, "Unsupervised and semi-supervised clustering for large image database indexing and retrieval," In 2012 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (pp. 1-6). IEEE, (2012).

[4] Nakazawa. T, & Kulkarni. D. V, "Wafer map defect pattern classification and image retrieval using convolutional neural network," IEEE Transactions on Semiconductor Manufacturing, 31(2), 309-314, (2018).

[5] Qi. Q, Huo. Q, Wang. J, Sun. H, Cao. Y, & Liao. J, "Personalized Sketch-Based Image Retrieval by Convolutional Neural Network and Deep Transfer Learning," IEEE Access, 7, 16537-16549, (2019).

[6] Janani M and Dr. Manicka Chezian. R, “A Survey on Content-Based Image Retrieval System,” International Journal of Advanced Research in Computer Engineering & Technology, Volume 1, Issue 5, pp 266, July 2012.

[7] Raza. A, Dawood. H, Dawood. H, Shabbir. S, Mehboob. R, & Banjar. A "Correlated primary visual texton histogram features for content base image retrieval”, IEEE Access, 6, 46595-46616, (2018).

[8] Reato, T, Demir. B, & Bruzzone, L, “An Unsupervised Multicode Hashing Method for Accurate and Scalable Remote Sensing Image Retrieval”, IEEE Geoscience and Remote Sensing Letters, 16(2), 276-280, (2018).

[9] Singh. N, Singh. K, & Sinha. A. K, “A novel approach for content based image retrieval,” Procedia Technology, 4, 245-250, (2012).

[10] Wu. H, Li. Y, Bi. X, Zhang. L, Bie. R, & Wang. Y, (2018). Joint entropy-based learning model for imageretrieval. Journal of Visual Communication and Image Representation, 55, 415-423.

[11] Wu. J, Chen. Y, Dai. D, Chen. S, & Wang. X, “Clustering-based geometrical structure retrieval of manmade target in SAR images,” IEEE Geoscience and Remote Sensing Letters, 14(3), 279-283, (2017).

[12] Vikas Tondar and Pramod S. Nair, “A Comparative Studyon Clustering Algorithms using Image Data,” International Journal of Computer Applications ISSN: 0975 – 8887, Vol.133, No.17, 2016.

[13] Xiang. J, Zhang. N, Pan. R, & Gao. W, “Fabric imageretrieval system using hierarchical search based on a deep convolutional neural network,” IEEE Access, 7, 35405-35417, (2019).

[14] Yang. J, Liang. J, Shen. H, Wang. K, Rosin. P. L, & Yang. M. H, “Dynamic match kernel with deep convolutional features for image retrieval,” IEEE transactions on Image Processing, 27(11), 5288-5302, (2018).

[15] Wazarkar S and B.N. Keshavamurthy, “A Survey on Image Data Analysis through Clustering Techniques for Real World Applications,” J. Vis. Commun. Image R, 2018.