

# An Optimized Image Retrieval using Clustering and Classification Techniques

Megha Jogiya<sup>1</sup>, Ms. Rauki Yadav<sup>2</sup>

<sup>1</sup>PG Scholar, Department of Computer Engineering, Bhagwan Mahavir College of Engineering & Technology, India

<sup>2</sup>Assistant professor, Department of Computer Engineering, Bhagwan Mahavir College of Engineering & Technology, India

**Abstract** — Content Based Image Retrieval (CBIR) is a process in which for a given query image, similar images will be retrieved from a large image database based on their content similarity. The content of image refer to its features or attributes or parameters which are mathematically determined from a digital image. In this approach the images retrieved may not exactly match with the visually similar or semantically similar images. Semantic similarity refers how far the user expectation meets the retrieval. Content Based Image Retrieval gained its importance from early 1980's and still got lot of scope for the research community to find more sophisticated methods to improve the retrieval.[6] Content Based Image Retrieval got its significance in domain specific applications such as biomedical and satellite imaging etc.[7] In this paper we presented an exhaustive literature review of CBIR from its inception to till date, with all the new approaches that has included in this process. We present our review on benchmark image databases, color spaces which are used for implementation of CBIR process, image content as color, texture and shape attributes, and feature extraction techniques, similarity measures, feature set formation and reduction techniques, image indexing applied in the process of retrieval along with various classifiers with their effect in retrieval process, effect of relevance feedback and its importance in retrieval.

**Keywords:** Image Retrieval, Clustering, Classification, Feature Extraction, Optimization.

## I. INTRODUCTION

Image retrieval is a technique, which uses visual contents to search images from large scale image databases according to user's interests. The main goal of the feature based image retrieval is to find images which are similar to query image visually without using any textual descriptions for the image.[2] It presented a picture indexing and abstraction approach for pictorial database. The Pictorial database consists of picture objects and

picture relations. It construct picture indexes, operations are formulated to perform picture object clustering and classification, searching. Feature extraction is the basis of content based image retrieval machine. [4]

## II. PROBLEM DEFINITION

Feature based image retrieval; means query image is given and based on feature of query image the similar image is retrieved. The two critical bottlenecks for performance and quality in system are semantic gap and retrieval time. The semantic gap is difference between machine and human perception of visual data. The semantic gap can be tackled by using better feature extraction method that works how human perception work, and building modules that model human visual perception better. One of the major challenges that I am going to face during my research will be to develop such technique or algorithms which would be able to search the similar image quickly and efficiently, so image should be retrieved from large image database

## III. FEATURE EXTRACTION

Images are usually represented by the visual features such as color, shape and texture. There are mainly two approaches for feature extraction in content based image retrieval.[9] (1) Feature extraction in spatial domain It is based on statistical calculation on the image. Many of the spatial domain methods suffer from insufficient number of features and also sensitive to noise. (2) Feature extraction in transform domain include the use of Discrete Cosine transform Multi-resolution methods such as Gabor filters.[11] Wavelet transform, curve let and Contour-let transform for feature representation[4]. Shape features aim to capture the

geometric form of objects or regions within an image.[17] They are particularly useful when the visual content is characterized more by the outline or overall form rather than color or texture [15].

#### IV. GENERAL CONTENT BASED IMAGE RETRIEVAL SYSTEM

**General Content Based Image Retrieval System**  
Early research into image retrieval had begun, as far back as four decades ago. Both the database management system and computer vision communities started working on image retrieval in the early 70's. Here feature extraction process of database image and query image is done. As the pre-processing step all the feature of database image are extracted and kept in separate file finally feature vector of query image is generated and its compared with feature vector of all database images and distance is calculated, finally the distances are sorted in ascending order and images with least distance is calculated. The most important part of any CBIR technique is its feature extraction method and matching and retrieval process.[2]

#### V. LITERATURE SURVEY

Clustering and classification techniques significantly enhance CBIR systems by improving retrieval efficiency and accuracy. Integrating deep learning, optimized clustering, and intelligent classification forms the foundation for next-generation image retrieval systems. Ongoing research focuses on improving scalability, accuracy, and semantic understanding.[1] Karthikeyan, P. Aruna the proposed system concentrates on content-based image retrieval (CBIR) for image clustering to give additional effect to the document clustering approach. It suggests two kinds of indexing keys, major color sets (MCS) and distribution block signature (DBS) to prune away the irrelevant images to given query image. Major colour sets are related with color information while distribution block signatures are related with spatial information. After successively applying these filters to a large database, only small amount of high potential candidates that are somewhat similar to that of query image are identified. Then, the system uses quad modelling method (QM) to set the initial weight of two-dimensional cells in query image according to each major color and retrieve more similar images through similarity association

function associated with the weights. The proposed system evaluates the system efficiency by implementing and testing the clustering results with DBscan and K-means clustering algorithms. Experiment shows that the proposed image clustering algorithm performs with an average efficiency of 94.4% for various image categories.[5]

Ying Liua, Dengsheng Zhanga, Guojun Lua, Wei-Ying Ma attempts to provide a comprehensive survey of the recent technical achievements in high-level semantic-based image retrieval. Major recent publications are included in this survey covering different aspects of the research in this area, including low-level image feature extraction, similarity measurement, and deriving high-level semantic features. We identify five major categories of the state-of-the-art techniques in narrowing down the semantic gap": (1) using object ontology to define high-level concepts; (2) using machine learning methods to associate low-level features with query concepts; (3) using relevance feedback to learn users" intention; (4) generating semantic template to support high-level image retrieval;(5) Fusing the evidences from HTML text and the visual content of images for WWW image retrieval. In addition, some other related issues such as image test bed and retrieval performance evaluation are also discussed. Finally, based on existing technology and the demand from real-world applications, a few promising future research directions are suggested. [7]

Ravi Shankar K.C, Prasad B.G, Gupta S.K, Biswas K.K, have proposed method of CBIR using dominant color region based indexing. The approach is based on the fact that our world is dominant by visual information and a tremendous amount of such information is being added day by day. It would be impossible to cope with this explosion of visual data, unless they are organized such that we can retrieve them efficiently and effectively. The main problem in organizing and managing such visual data is indexing, the assignment of a synthetic descriptor which facilitate its retrieval. It involves extracting relevant entities or characteristics from images as index keys. Then a representation is chosen for the keys and specific meaning is assigned to it. Color is an important cue for content based image retrieval system. Authors have proposed a technique to index and store

images based on dominant color regions. Features like region size and location of the region are extracted and used as similarity measures. Images with similarity indices are stored as an image cluster in a hash table. A prototype of the retrieval system is developed using JAVA [10]

### VII. PROPOSED SYSTEM

First image is selected from standard image database like coral. Then some preprocessing is done on image to reduce noise and unwanted features. Then color and texture features of the images are calculated from processed image and stored in the database. The texture data is selected and based on that data clustering is done using efficient clustering algorithm.

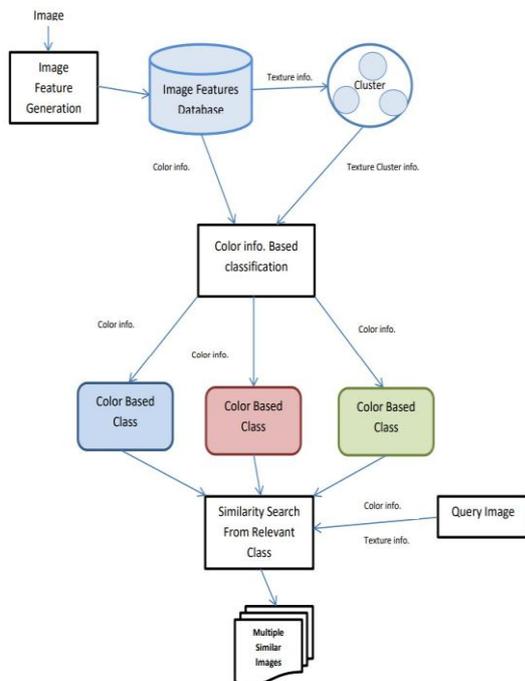


Fig 1 Block Diagram of Application

Then second level clustering is done on colour features using soft clustering algorithm using fuzzy logic.[4] When query image is presented the feature of the images are extracted. The similarity search is done using the cluster centre. So it will take less time compared to searching whole database.

### VIII. CLUSTERING ALGORITHM

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in

other groups (clusters). K-mean clustering is a centroid based clustering algorithm's-means clustering is a method of vector quantization originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition N observations into k clusters in which each observation belongs to the cluster with the nearest mean.[16] Given a set of observations (x1, x2, ..., xn), where each observation is a dimensional real vector, k-means clustering aims to partition the n observations into k sets (k ≤ n) S = {S1, S2, ..., Sk} so as to minimize the within-cluster sum of squares. [3]

Where  $\mu_i$  is the mean of points in  $S_i$ .

$$\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

Fuzzy C is soft clustering techniques. In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. [15] Thus, points on the edge of a cluster, may be in the cluster to a lesser degree than points in the center of cluster. The degree of belonging,  $w_k(x)$ , is related inversely to the distance from x to the cluster center as calculated on the previous pass. It also depends on a parameter m that controls how much weight is given to the closest center. The fuzzy c-means algorithm is very similar to the k-means algorithm.[18]

$$w_k(x) = \frac{1}{\sum_j \left( \frac{d(\text{center}_k, x)}{d(\text{center}_j, x)} \right)^{2/(m-1)}} \cdot$$

### X. CONCLUSION

Feature based image retrieval is very broad subject. Each application area requires different extraction techniques and matching procedure. The Algorithm implemented here uses only color feature for image extraction and it searches images in whole database so it will take more time. Texture information is also used to retrieve more accurate result. To reduce the retrieval time for similar images clustering algorithms will be applied on image feature database

### REFERENCE

[1] M.E. Elalami, Supporting image retrieval

- framework with rule base system, Knowledge-Based Systems 24 (2) (2019)
- [2] P. Howarth, S. Rüger, Evaluation of texture features for content-based image retrieval, in: Proceedings of the International Conference on Image and Video Retrieval, July 2014, pp. 326–334.
- [3] A.Kannan<sup>1</sup> Dr.V.Mohan<sup>2</sup> Dr.N.Anbazhagan<sup>3</sup>, “Image Retrieval Based on Clustering and Non-Clustering Techniques using Image Mining”- Int J Engg Techsci Vol 1(1) 2020,54-61
- [4] Ela Yildizer, Ali Metin Balci, Tamer N. Jarada, Reda Alhajj ,“Integrating wavelets with clustering and indexing for effective content-based image retrieval”, (2021)
- [5] M. Karthikeyan\*, P. Aruna “Probability based document clustering and image clustering using content-based image retrieval” (2022)
- [6] Zhang D, and Lu G , “Evaluation of Similarity measurement for image retrieval” , in proceedings of IEEE Neural network and signal processing, 2013, pp-928-938.
- [7] Ying Liua,\*, Dengsheng Zhanga, Guojun Lua ,Wei-Ying Ma “Asurvey of content- based image retrieval with high-level semantics” (2017)
- [8] Ali Mosleh<sup>1</sup>, Farzad Zargari<sup>2</sup>, Reza Azizi - “Texture Image Retrieval Using wavelet Transform” (2018)
- [9] Markus Koskela, Jorma Laaksonen, and Erkki Oja -“Comparison Of Techniques For Content-based Image Retrieval” Conference on Image Analysis (SCIA 2019). Bergen, Norway, pages 579-586
- [10] Ravishankar , K.C.Prasad, B.G. Gupta, S.K. Biswas, “Dominant colour region based indexing for CBIR”, International conference on image analysis and processing,2016
- [11] Urszula Marmol - “Use Of Gabor Filters For Texture Classification Of Airborne Images And Lidar Data” - Archives of Photogrammetry, Cartography and Remote Sensing, Vol. 22, 2018, pp. 325-336 ISSN 2083-2214
- [12] Ray-I Chang 1, Shu-Yu Lin 1,2, Jan-Ming Ho 2, Chi-Wen Fann 2, and Yu-Chun Wang 2-“A Novel Content Based Image Retrieval System using K-means/KNN with Feature Extraction” - DOI: 10.2298/CSIS120122047C
- [13] C. Lakshmi Devasena, M. Hemalatha – “A Hybrid Image Mining Technique using LIM-based Data Mining Algorithm” International Journal of Computer Applications (0975 – 8887) Volume 25– No.2, July 2019
- [14] Swapnalini Pattanaik, D.G. Bhalke – “Efficient Content based Image Retrieval System using Mpeg-7 Features”- International Journal of Computer Applications (0975 – 8887) Volume 53– No.5, September 2022
- [15] Harikrishna Narasimhan, Purushothaman Ramraj “Contribution-Based Clustering Algorithm for Content-Based Image Retrieval
- [16] Jiawei Han, Micheline Kamber, Jian Pei, “Data Mining: Concepts and Techniques” The Morgan Kaufmann series in Data Management Systems.
- [17] Vlad Atanasiu , Peter Fornaro “On the utility of Colour in shape analysis: An introduction to Colour science via palaeographical case studies” Heliyon Volume 9, Issue 10, October 2023
- [18] S. S. Balamurugan, S. S. Babu, et al., Optimized Content-Based Image Retrieval Using Hybrid Clustering and Deep Learning Classification, (2021)