International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Review Article)

Conceptional review of the Content-based Image retrieval

Varsha Kiran Patil *, Shristi Manoj Dhamange, Shraddha Anil Bhandurge and Samruddhi Udaysinh Gaikwad

Department of Electronics and Telecommunications AISSMS Institute of Information Technology, Pune, India.

International Journal of Science and Research Archive, 2023, 10(01), 316–329

Publication history: Received on 30 July 2023; revised on 09 September 2023; accepted on 12 September 2023

Article DOI: https://doi.org/10.30574/ijsra.2023.10.1.0733

Abstract

This article examines the process of the Content-based Image Retrieval (CBIR) system , the features and steps involved, the methods employed, and the applications. CBIR aims to identify, sort, and manage images on the basis of the requirements posed to the system. The system cross checks the requirements using Machine learning processes against a given database and reduces the manual effort of sifting through all pictures individually.

Incorporating references to the different methods of feature extraction, this article emphasizes understanding which features are considered important and the characteristics of features that are searched for. It argues the advantages of multiple methods while suggesting how each method is suitable when employed for a specific purpose. The process of indexing is also highlighted in this article. These processes are particularly useful in this advancing world where huge databases of images need to be handled on a daily basis in various applications and where manual methods are simply not feasible.

Keywords: Feature selection; Content-based image retrieval; Machine learning; Spatial features; Database; Indexing

1. Introduction

Efficient image search requires a powerful image index and libraries. Keyword or text indexing is commonly used in several fields including geographic information systems, multimedia libraries, CAD/CAM systems, criminal detection, and visual arts archives. The text description is linked to the database as keywords or free text.

Text-based systems struggle to generate graphic keywords and classify images, requiring human administrators to do so. This can be time-consuming and subjective, limiting image searches unless specific templates or keywords are used.

Describing images based on visual properties like colors, textures, shapes, and layout is more useful than text-based indexing for certain applications. It can lead to easier visual searches.

Visual feature-based indexing techniques are used in CBIR to create test images and identify characteristics that match images stored in a database.

1.1. Image retrieval system

Compared to conventional text-based image retrieval, CBIR is significantly different. A search is conducted initially using visual feature similarity. Choosing or making a representative example or examples, then looking for pictures similar to those examples would constitute creating a query for such a search.

* Corresponding author: Varsha K. Patil

Copyright © 2023 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

Interaction between the database and the user via a visual interface improves the user's ability to evaluate retrieval results and define and improve queries based on the evaluations.

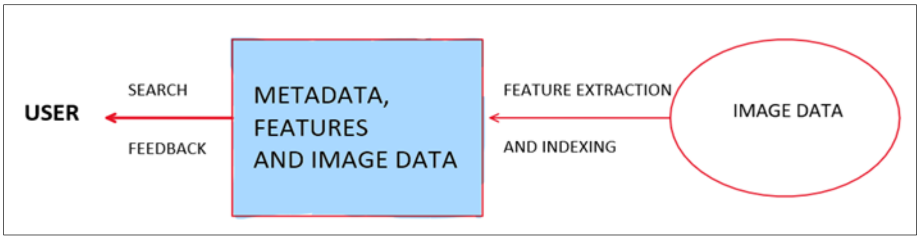

Figure 1 The Block diagram of the image retrieval

A metadata database that stores image features serves as the brain of this system. In order to calculate similarity, these content metadata are extracted during retrieval. The "closing of the loop," which links the creation of queries to data browsing, is then made possible by the user interface. Figure 1 shown below highlights the process of a method for obtaining pictures based on certain specifications.

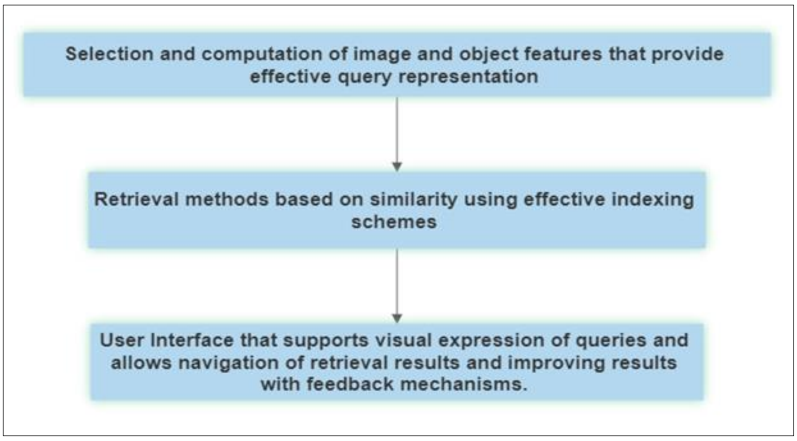

Figure 2 The methods to achieve image retrieval

2. Feature extraction in CBIR

2.1. Feature Extraction

The method of extracting features for content-based image retrieval is creating a basic representation (either alphanumeric or simply numerical) of some characteristics of digital images and thereby gaining information about the given image which can be further used for categorization and classification and paves the way for future computation. It is a way to reduce its dimensions. A feature can be related to visual quality, as well as to how one perceives a picture or an object, or a spatial aspect or characteristics.

A feature might be a composite representation of many characteristics or it can be related to a single property. Features are categorized as general-purpose or industry-specific categories. The domain-related features are created expressly for a specified application, whereas the general features can be utilized in any situation. Each feature has a close relationship with the type of data that it collects. Depending on the application at hand and the type (degree) of information needed, one feature may be preferred to another.

2.1.1. The basis of feature extraction

As a picture carries information at several levels, using multiple characteristics simultaneously is a need for creating efficient algorithms to find pictures depending on what's in them different feature representations can be used to describe any visual quality. Visual characteristics might be low-level ones like color, form, or texture to intermediate-level ones like spatial linkages and areas to high-level ones like object localization and identification or picture classification. Selecting the best feature is called feature selection.

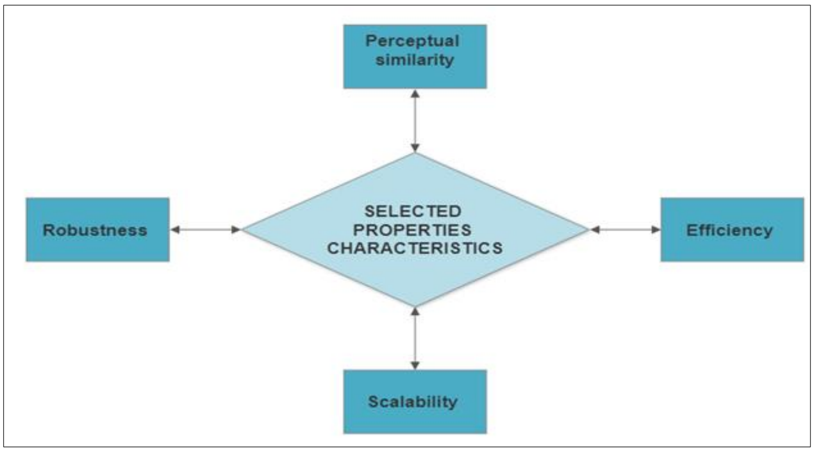

Figure 3 Flowchart for Image feature selection

A flowchart of the characteristics the image features extracted must possess in order to have a certain feasibility for further processes.

- Perceptual similarity: Only when two pictures are not "similar" is the feature gap between them high;
- Efficiency: They are quick to compute;
- Economy: They are compact in size to maintain retrieval effectiveness;
- Scalability: The system's performance is unaffected by the database's size;
- Robustness: Retrieval of database images should not be impacted by modifications to the imaging settings (such as illuminations, geometric transformations, etc.).

3. Classification of features and their Categorisation

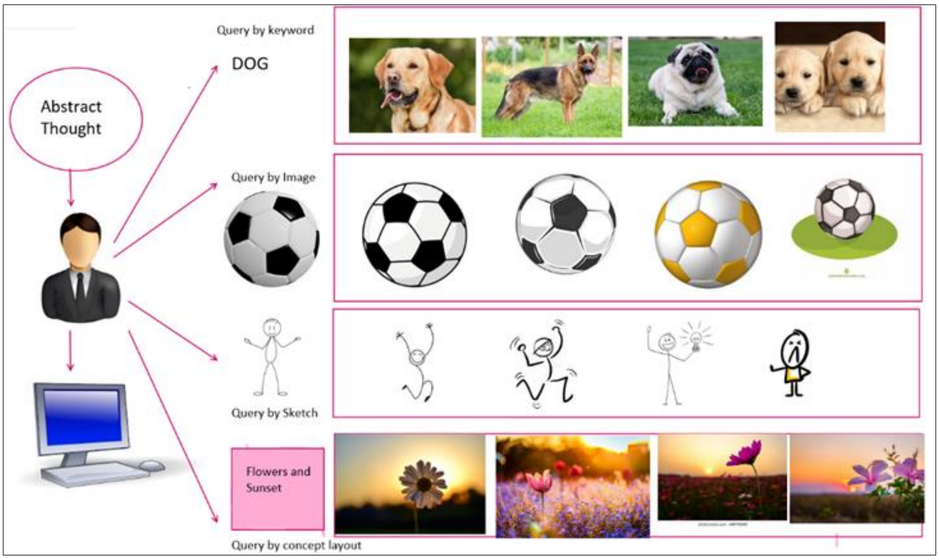

Figure 4 Image retrieval on basis of queries

The above figure shows the categorization of photographs based on their methods and how images are retrieved on the basis of the queries selected by the user.

The utilization of smartphones, digital cameras, and the internet has surged due to technological advancements, which have led to a rise in the saving and exchanging of multimedia data. No information has been left out of the paraphrased text. Finding a suitable image from the past can be a difficult research task. Search engines often rely on text-based methods which depend on captions as input. To search for an image, the user inputs a query using words or phrases that are compared with those in the archive's keywords. Despite the development of automatic image annotation systems using various factors such as texture, edges, color, shape-related information, and spatial layout, the process of retrieving images can still result in irrelevant images. This is due to the discrepancies in manual annotation and human visual perception, which can also cause errors in the procedure for retrieval. Content-based image retrieval (CBIR) overcomes these problems by analyzing the visual elements in the search image and matching them with images in the archive based on the similarity in appearance between the picture feature vectors. When it comes to image retrieval, low-level visual characteristics like shape, color, texture, and spatial arrangement are analyzed. Matching these traits to information from the query plays a key role in organizing the output. QBIC and Simplicity are two models of retrieval of images that rely on low-level visual semantics. Various fields have utilized CBIR and feature extraction methods for image analysis, including medical imaging, remote monitoring, crime investigation, video evaluation, military monitoring, and textile manufacturing. The fundamental principles and workings of picture retrieval are outlined in Figure 4.

When it comes to image retrieval systems, it is necessary to have a mechanism that can locate and organize comparable images in a database with little to no involvement from humans. The choice of graphic elements for such systems depends on the end user's requirements, according to literature sources. For such systems, having a feature representation that distinguishes between different features is crucial. To achieve more durability and distinctiveness in the features, it is vital to combine low-level visual features, even if it requires more computational resources. However, improper feature selection can decrease the image retrieval model's performance. Machine learning algorithms can use image feature vectors as input by training and testing models in order to enhance content-based image retrieval performance. Recent years have seen deep neural networks (DNN) being the subject of image retrieval trends, despite their high computational cost. This part aims to give a summary of current research trends challenging CBIR and feature representation.

3.1. Color features

Color is a vital low-level visual feature that can be easily distinguished by the human eye. Images of objects in the actual world that fall within the range of what humans can see and can be distinguished from each other based on color differences. These color differences remain consistent and are not noticeably impacted by changes in scale, position, or orientation within an image. By using a dominant color descriptor (DCD), it is possible to summarize an image's overall color information using only a few key representative colors .The DCD is a descriptive tool for MPEG-7 color that utilizes a concise, user-friendly format to portray the characteristic color distribution and attributes.

A new method for CBIR (content-based image retrieval) has been suggested by Shao et al. [26] which employs the MPEG-7 descriptor. The technique involves picking out eight significant colors from an image, measuring features with a histogram intersection algorithm, and making similarity computations less complicated.

Classical techniques for image retrieval using labels and annotations may not meet customer requirements, leading researchers to focus on obtaining pictures that are relevant to the content. [27]. Danum proposed A technique that uses a compact visual descriptor that can adapt based on the context of the image through a clustering process in two stages. Experiments were conducted using the COIL-100 image library, and it was demonstrated that the proposed approach is effective.

[18] Wang and colleagues introduced a technique that utilizes color and texture characteristics to retrieve images, which accurately mimics the way humans perceive visual data. Combining both color and texture characteristics is a strong feature set for retrieving color images, but it comes with a significant computational expense. The suggested approach achieved better image retrieval results compared to conventional methods but did not increase the number of feature dimensions beyond other techniques. However, comparing low-level features in pairs can still be a hindrance.

The property of invariant descriptors being complete has been the subject of research by various groups. [23] Moment functions are Zernike and pseudo-Zernike polynomials. form a horizontal basis and can be used to represent images. These functions have descriptors that are mutually independent and retain orthogonality and rotation invariance. Zhang and colleagues have found that there are fewer pseudo-Zernike moments affected by image noise compared to Zernike moments. They developed a novel method to obtain a comprehensive set of pseudo-Zernike moment invariants that have demonstrated superior pattern recognition capabilities compared to other methods.

In their study, Guo and colleagues (19) introduced a fresh method for indexing images that rely on features derived from the (BTC) block truncation coding method known as error diffusion. Generation of a descriptor for image features utilizes a bitmap image and two color quantizers using vector quantization (VQ) that were created by EDBTC. To determine how similar a query image is to an image in a database, two features, namely Bit Pattern Histogram Feature

(BHF), and Color Histogram Feature (CHF) have been developed. These features are based on the VQ-indexed bitmap image and VQ-indexed color quantizer and are used to determine the resemblance between the two images. By measuring the distance between BHF and CHF, one can determine the resemblance between the two images. According to the results, the suggested method is more effective than the antecedent BTC image indexing and other available image retrieval methods. The EDBTC has a strong capability for compressing images and indexing them for CBIR. No information has been left out in this paraphrased text.

In their study, Liu and colleagues (23) introduced a fresh approach to studying images based on regions, which involves the use of a decision tree called DT-ST. This method relies on image segmentation key methods for machine learning and aims to address the challenge of feature discretization in decision trees by creating semantic templates from the low-level capabilities to classify different regions in an image. The proposed method involves a hybrid tree that addresses issues related to noise and fragmentation in tree structures for improved accuracy and reduced misclassification. This technique enables users to search for images using labels and regions, resulting in higher retrieval precision in comparison to conventional CBIR methods. Experiments indicate that the proposed technique bridges the difference between low-level and high-level features outperforming the C4.5 & ID3 algorithms in image semantic learning.

In their study, Islam and colleagues (28) developed an advanced vector quantization algorithm based on color that can categorize image components automatically. Unlike traditional vector quantization algorithms, this algorithm utilizes variable feature vectors such as dominant color descriptors. Additionally, it includes a unique stopping criterion and splitting method that enables the algorithm to understand the ideal number of clusters and prevent excessive fragmentation of region clusters.

In their research, JieXian and colleagues (20) introduced a new method called multi-scale distance coherence vector (MDCV). The reason for developing this technique was that certain shapes can have similar descriptors, and the DCV algorithm might not entirely separate unwanted data. The MDCV method uses a Gaussian function and creates the contour curve of the image and can maintain its effectiveness even when the image transforms such as translation, scaling, and rotation.

The summary reveals that there are multiple low-level features available. However, color moments are not considered effective as they fail to represent all image regions. Histogram-based features demand significant computational power, and DCD requires fewer computations because of its low dimensions and demonstrates better performance for region-based image retrieval.

3.2. Texture features

The selection of regions or objects in images is dependent on texture, which is a crucial characteristic. It applies to various types of images, including satellite images, photomicrographs, and aerial photographs. This research introduces a collection of textural features that depend on grey-tone spatial dependencies and are easy to compute. The study also shows how these features can be used for categorizing data from three tasks, including photomicrographs of various sandstones, 1:20 000 panchromatic aerial images with several types of land-use, and Earth Resources Technology Satellite (ERTS) multispecialty imagery with a variety of land-use categories. Two decision-making processes were used in the study: a min-max decision process and a piecewise linear decision rule. The data was split into a training set and a test set for each experiment. The identification accuracy of the test set was determined to be 89% for the photomicrographs, 83% for the satellite imagery, and 82% for the aerial photographic imagery. Based on these findings, it can be concluded that the textural features utilized in this study have the widespread potential for various applications related to image classification.

Papakostas and colleagues conducted research on four datasets - TRIESCH I, JAFFE, ORL, and COIL - to showcase the effectiveness of wavelet moments in distinguishing between different classes. They used wavelet moments (WMs) in two configurations, WMs-1 and the WMs-2, to enhance the classification abilities of wavelet moments. The performance of WMs was compared to other moment family like Fourier-Mellin, Zernike, Legendre, and pseudo-Zernike, using varying percentages of the datasets. The results indicated that the proposed model, WMs, performed better than the other moment descriptors.

Using Corel datasets, Liu et al. [23] assessed the proposed model for image retrieval, MSD. Based on the HSV, RGB, and Lab color spaces, the retrieval efficiency of the model is assessed. The results show that the model works. performs superiorly in the Lab color and HSV spaces. In MSD, 72 color and orientation quantization scales have been employed for picture retrieval. The suggested model performs better than other models, such as Gabor MTH.

In a different study, 10,000 color photos [22] were gathered to assess the suggested model for texture-based image retrieval. In terms of mean accuracy value and recall rate, the suggested model outperforms earlier models, and it is found demonstrating that the color co-occurrence matrix is superior to the grey co-occurrence matrix.

Irtaza and Jaffar [24] used the Corel image gallery, which has 10,900 photos split into two sets: Corel A (1,000 images, 10 categories), and Corel B (9,900 images), to evaluate their suggested model (SVM-based architecture) for category image retrieval. To evaluate the retrieval capacity of SVM, the tests consisted of comparing the average accuracy and recollection rates of the proposed approach with those of other widely used retrieval systems. The outcomes demonstrated that the suggested model performed better in terms of consistency and efficiency of image retrieval than alternative approaches.

In another study, Fadaei et al. [25] tested their model (LDRP) for content-based image retrieval, using Brodatz comprised of 112 grayscale & Vistex datasets containing 54 color images, The study included comparing the performance and average precision rates of the proposed model with other prior methods. Semantic representation of texture features depends on the forms of the objects in the photographs, despite the fact that they have a greater semantic significance than color features.

3.3. Shape features

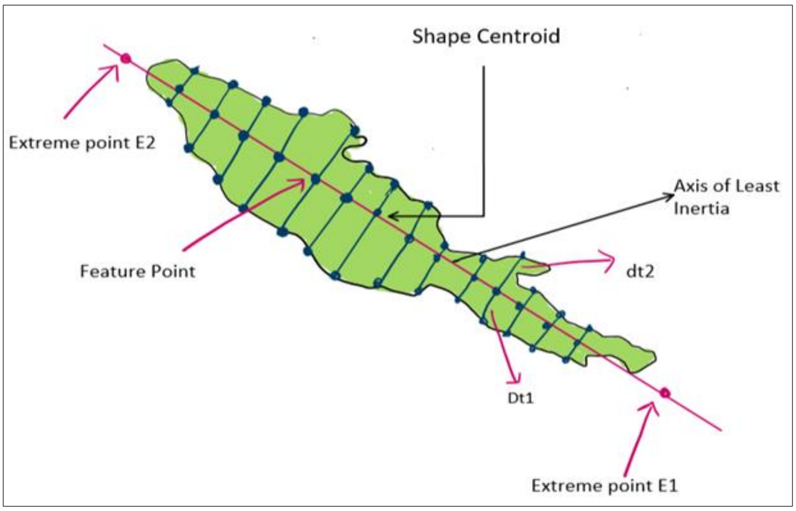

Figure 5 Illustration of shape feature extraction on a given image shape

The characteristics of the shape of an object are referred to as visual features, which include its shape, perimeter, and diameter. These features can be categorized into two groups, based on the object's border and regional characteristics. Methods for extracting image shape features are divided into two types: region-based and boundary-based. While shape features are easy to understand and are intuitive, they lack a solid mathematical model and may not provide reliable results for targets that are not well-defined in shape. Additionally, comprehensive target descriptions require high levels of computational and storage resources. The accuracy of shape feature extraction is also dependent on the quality of pre-segmentation. The reflected shape information of the target may not always be identical to human intuition.

A shape descriptor, in general, is a group of integers that are generated to characterize a certain form characteristic. A description makes an effort to characterize form in a manner that is compatible with human intuition (task-specific needs). decent retrieval

A shape descriptor needs to be capable to locate perceptually comparable shapes in a database quickly if it is to be accurate. These descriptors frequently resemble a vector. The shape descriptors should adhere to these standards:

- The descriptions should accurately give the information content by being as thorough as they may be.
- The descriptors have to be compactly recorded and stored. The descriptor size of the vector shouldn't be excessive.
- In order to avoid an large execution time, the estimation of the distance between descriptors should be straightforward.

The following areas of use that gain from shape feature extraction and representation:

- Shape retrieval is the process of searching a huge database for all shapes that are close enough to a query shape. Usually, only the first few shapes that are closest to the query are chosen, or all shapes similar to the query.
- Shape recognition and classification: identifying which representative class is the most comparable or whether a particular form matches a model successfully.
- Shape registration and alignment involve changing one shape to make it fit another shape, as wholly or partially and closely as possible.
- Creating a shape with less number of components (triangles, segments, points, etc.) while maintaining a comparable appearance is known as shape approximation and implication.

There are 3 main classification methods used here:

- Approaches based on regions and contours: This is the most typical and ubiquitous classification, and MPEG-7 recommends it. Instead of using shape interior points, it makes use of shape border points. Different approaches are further classified into global and structural approaches under each class. Whether the form is represented as a whole or in segments or parts determines the subclass (primitives).
- Domains of space and transformation: Although approaches in the domain feature match shape on a feature basis, methods match shapes based on points (point features) in the space domain.
- Preserving information (IP) and not preserving information (NIP): IP methods enable a precise reconstruction a form from its description, in contrast to NIP approaches that can only achieve partial ambiguous reconstruction. IP is not necessary for the purpose of object recognition.

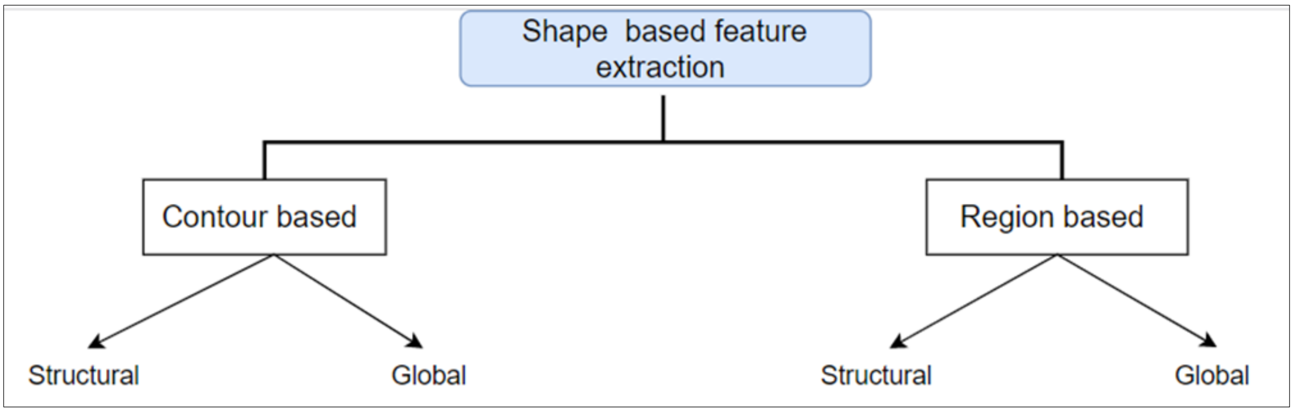

Figure 6 Types of shape-based feature extraction illustrated via a flowchart

3.4. Spatial features

Various methods have been developed to represent the arrangement of objects within a 2D image space. The BoVW model is a well-known approach that uses histograms to represent images, but it does not consider spatial arrangements. SPM is another technique that captures spatial features, but it is not effective for handling scaling and rotations. Zafar and colleagues introduced a method that includes spatial information in the Bag of Visual Words model by calculating geometric correlations among similar visual features. Ali and colleagues proposed a Hybrid Geometric Spatial Image Representation (HGSIR) that creates histograms for rectangular, circular, and triangular regions of images. These methods were evaluated using multiple datasets, and the results demonstrate that they perform better than other state-of-the-art methods for accurately classifying images.

In their study, Khan and colleagues introduced a method called Pairs of Identical Visual Words (PIWs) to represent the spatial distribution of global visual words. This method takes into account the relationships among similar visual words, resulting in histograms that provide detailed information about intertype visual word relationships. This approach has numerous benefits, such as incorporating global information, efficient extraction of spatial information, and simplifying complexity, ultimately leading to improved classification rates. Anwar and team presented a model that utilized symbol recognition to enhance the BoVW model with spatial information. To achieve rotation invariance, circular tiling were used to modify angle histograms. This model was tested on multiple datasets and was found to be effective in achieving rotation invariance. Khan and colleagues offered a solution for capturing both local and global spatial distributions of words through a soft spatial histogram. Their approach was tested on various datasets, and the researchers found that it enhanced the overall performance. Ali and his team suggested a different method for representing the global spatial distribution by constructing histograms using orthogonal vectors between PIWs. The effectiveness of their technique was assessed through three satellite scene datasets.

3.5. Local Feature Extraction

In an image, local features refer to unique structures such as edges, points, or small patches of the image that are distinguishable due to their texture, color, or brightness from the surrounding area. The crucial aspect is that these features are noticeable and stand out from their surroundings, regardless of their actual meaning. Examples of local features are corner pixels, blobs, and edge pixels.

4. Vector Space Model (VSM)

The Vector Space model is a mathematical approach that represents written works as numerical vectors, based on the frequency of phrases. The VSM, a well-liked model for information retrieval, shows documents as vectors in a high-dimension space where every dimension is connected to a word from the lexicon.

The VSM is predicated on the idea that a document's meaning may be derived from the distribution of its terms and that papers with related content will have related term distributions.

To begin using the VSM, a group of documents must first be pre-processed by being tokenized, stemmed, and stop word-free. The next is to construct a matrix of documents and terms, where row corresponds to a term and column is corresponded to a document. The matrix contains the frequency of every term in every document (for e.g., TF-IDF).

The query is also pre-processed and presented as a vector in the same space as the documents. A similarity score is then computed between each query vector and the document vector using a cosine similarity metric. After documents are rated on the basis of how similar they are to the query, the highest-rated papers are returned as the search results.

The VSM has various benefits, including its ease of use, potency, and capacity for handling enormous document collections. The assumption of "bag of words", which disregards word order and context, and the issue of term sparsity, where numerous terms occur in a small number of documents, are some of its drawbacks. More complex models that consider the semantic links in between words and documents, including probabilistic models or neural models, can be used to overcome these restrictions.

5. Parameters Vector Space Model (VSM)

This model is used in various NLP applications, including text categorization, information retrieval, and clustering. To determine the similarity between two texts, their vector representations are compared. In the VSM, each document or query is represented as an M-dimensional vector, where M is the total number of distinct terms across all documents and queries. The i-th index of the vector indicates the score of a vector's i-th phrase.

Two important variables for scoring in CBIR are Term-Frequency (TF) and Inverse-Document-Frequency (IDF). In CBIR, attribute categories are assigned weights to show the significance of each attribute, such as color being more important than texture or brightness being more important than size. These weights come from a relevance feedback process in which a human evaluator is involved.

Various methods such as term frequency (TF), inverse document frequency (IDF), and term frequency-inverse document frequency (TF-IDF) can be employed to compute weights.

5.1. Term Frequency (tf)

Term frequency (TF) is a popular measure used in CBIR systems. It is the ratio of the number of times a particular term occurs in an image to the total number of terms occurring in the image. This measure is used to score terms based on their relevance to the image. For example, if an image contains the word “cat” 10 times, and all other words in the image only occur once, then the term frequency of “cat” would be 10/11. This indicates that “cat” is a highly relevant term for the image.

TF= (Number of times the term appears in the document)/(Total number of terms in the document)

5.2. Inverse Document Frequency (IDF)

IDF is a measure used in information retrieval, CBIR, and text mining. It is a measure of how significant a specific term is to a document collection. The IDF is calculated by dividing the total documents in the collection by the documents carrying the term and then getting the log of the result.

IDF = log((Number of the documents in the corpus)/(Number of the documents in the corpus contain the term))

5.3. Term Frequency-Inverse Document Frequency (TF-IDF)

It is a statistic that is used to measure a word's significance in a document or many documents. It is calculated by dividing the frequency of a word by the inverse of the frequency of the same word in the entire collection of documents. The result is a score that indicates how important a word is relative to the entire set of documents.

TF-IDF is the multiplication the TF and IDF

$$\text{TF-IDF} = \text{TF} \times \text{IDF}$$

5.4. Similarity Measures for Document Vectors

The level of similarity between two document vectors is assessed using similarity measures. Cosine similarity and Euclidean distance are the two most often used similarity measures in CBIR.

5.5. Cosine similarity

The cosine of the angle between any two non-zero vectors in an inner product space is used to measure how similar the vectors are.

It is a commonly used technique in CBIR systems to compare the similarity of two images. This can be done by extracting the feature vectors from the photo and then calculating the cosine similarity between those vectors.

$$\cos(a,b) = ab/(||a||\ ||b||)$$

5.6. Euclidean Distance

Euclidean distance is a popular measure of similarity used in CBIR. It measures the straight line or Euclidean distance between points in a feature space, where the feature space is derived from the properties of the image. For example, if an image has a certain color, shape, or texture, its feature space would include these properties. The Euclidean distance between two images is then calculated by taking the square root of sum of the squared differences between the feature values of the two images.

6. Feature Selection

The problem of selecting features is brought on by the propensity for data collectors to obtain as much information as they can. Not all characteristics or qualities, meanwhile, are essential for certain learning tasks, like grouping. When class labels are provided, it is simpler to choosing features those are pertinent to the classes when using supervised learning, particularly for classification problems. It is unclear how each feature should be chosen for clustering tasks or for unlabeled data without class labels. Certain aspects might be unnecessary, unimportant, or merely tangentially significant. Finding the optimum set of pertinent features to show the data's natural clusters based on the selected criterion, clustering is the aim of selecting features.

It is essential to assess each feature's ability to discriminate in feature selection processes. Many techniques can be utilized for this, with similarity measure being the most popular one. For selecting features in image retrieval, the evaluation of gene expression data, and the evaluation of medical image feature data, researchers have also used additional methodologies such max divergence and intercluster interactions. In our work, we evaluate the various characteristics of feature-data relations using three criteria. While inter-cluster and inter-cluster affinity describe the relationship between features and classes, mutual data is utilized to calculate the association between features and data. These criteria can be used to assess the feature's discrimination and description power, respectively.

The procedure of choosing significant attributes for the underlying categories is known as feature selection for clustering. Additional categories that can be used to separate these approaches include global versus local and wrapper (i.e., with feedback) versus filter (i.e., without feedback - blind). While local strategies choose features for each specific cluster, global approaches choose features for the entire data set.

7. Indexing in CBIR

The process of transforming an original image into a tree structure with inserted image features involves several steps. Firstly, the Binary Threshold technique is used to remove the image background. Next, the image features are divided into two parts: the Global section and the Regional section. The colors in the Global section are simplified using a quantization technique. All the features are then calculated and stored in the data structure, as described in the Global feature. K-Mean Clustering is applied to the Binary Threshold (T) resultant image using a value of k set to 15. Finally, a labeling algorithm is employed to spatially detect the regions.

To find related images to a particular image from a database is the goal of image indexing (i.e., a pattern image). Each image has a distinctive quality. So, by comparing the features that are derived from the photos, image indexing may be put into practice. The criteria for comparing photographs may be based on characteristics like color, intensity, form, position, texture, and other aspects described above. There are two sorts of current image indexing techniques:

- Written (manual)
- Based on content (automated)

7.1. Text-based

Simple strategies are used, and specific keywords are provided for each image while keeping in mind the user's perspective. Among them are: Caption indexing; keyword additions; Classification, common subject titles, etc.

This indexing has the drawback of being labor-consuming. Inter indexer consistency issues are more likely to arise when indexing things, things, and things than when indexing text.

7.2. Content-based

This approach, sometimes referred to as automated indexing, involves indexing photographs based on their content, such as color, form, orientation, texture, spatial relationship, etc. Software handles this type of indexing; algorithms that can distinguish between colors, shapes, textures, etc. are created

7.3. Image indexing model

Figure 7 depicts the image processing procedures from the original picture through the insertion of image characteristics into the tree structure.

- Firstly, Using the Binary Threshold approach, we split the item from the picture backdrop.
- We divide image features into two parts: Global section, region section
- The colours in the Global section have been made simpler using the quantization technique.
- All features are calculated and stored into the data structure in the manner described in the Global feature.
- Used the binary threshold (T) resultant image from K-Mean Clustering. Here we are utilising the value of k \sis 15.
- Using the labelling algorithm to detect the regions spatially.

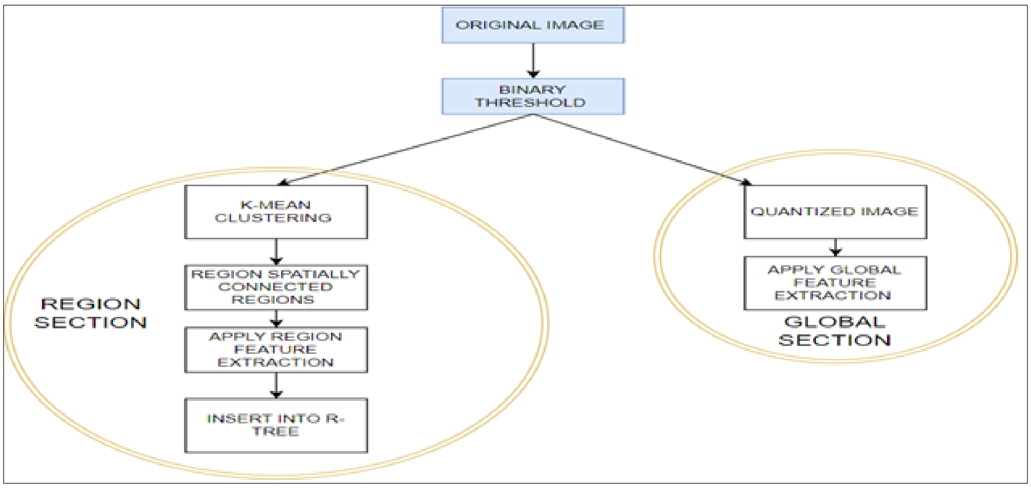

Figure 7 The process of indexing and its sections: Global Region

8. K- mean clustering algorithm

There are many machine learning methods useful for the CBIR such as Support vector machine, random forest, KNN etc. Here we will study KNN.

The technique of grouping things based on a tight relationship or shared qualities is known as clustering. The characteristics might be attribute values, relationships between the objects, or a combination of the two, and the objects can be tangible or abstract things. Similarity patterns are discovered within a set of objects using clustering.

- K-means clustering is an unmonitored training approach to cluster analysis. The basic concept involves identifying clusters with the same target category and making predictions for new data items by assuming they belong to the nearest cluster centre.
- Suppose we have m feature vectors (a1, a2…am) that all belong to the same class "C," and we know that they belong to "z" clusters where z < m.
- We may employ a minimum distance classification to separate the clusters if they are clearly distinct from one another.
- To begin the K-means clustering process, we first initialise the means of "z" clusters, which are μ1, μ2…μz.
- One way to do this is to assign random numbers to them.
- We then determine the membership of "a" by calculating the | |a-μi| |, where (i=0, 1…..z).
- The minimum distance determines "a's" membership in its respective cluster. This process is repeated for all "m" feature vectors with following sub steps.
    - Put "z" points into the area that represents the things that will be grouped together. The first group centroids are these positions.
    - Each object should be put in the group with the closest centroid.
    - Recalculate where the "z" centroids are after all the items have been assigned.
    - When the centroids stop moving, repeat steps 2 and 3 once more. This shows that the groups have come together and that the algorithm has found the best answer

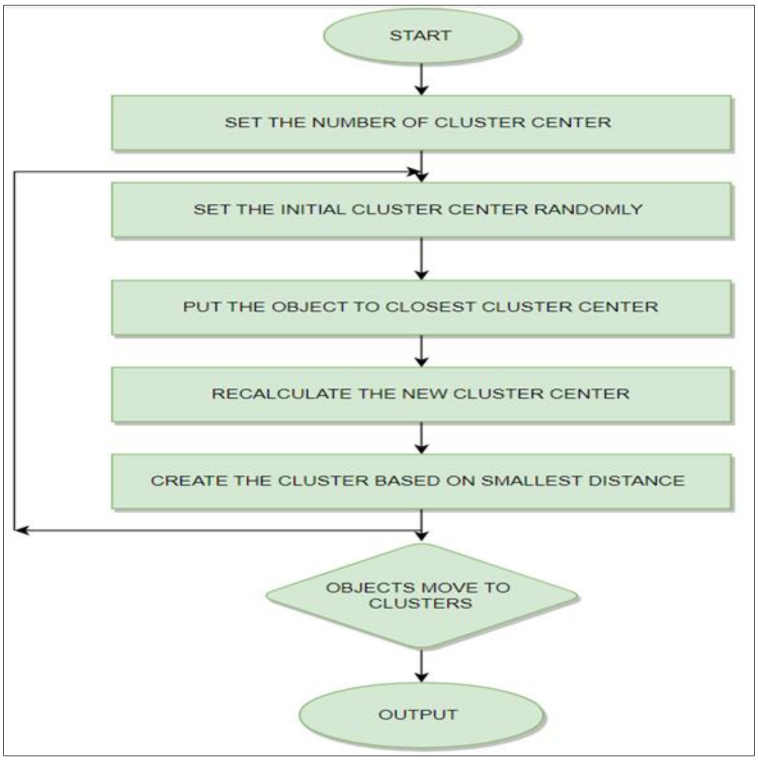

Figure 8 Steps involved in the K-means clustering algorithm

9. Applications of CBIR

Many real-world applications of the CBIR idea exist, and the main ones fall into one of the following categories.

Applications in medicine

- Retrieving images from remote sensors
- Naturally retrieved images
- Applications in forensics
- Security-related software
- Business applications
- Miscellaneous applications

Many possible uses for CBIR technology exist, including automatic face recognition for criminal investigations and the detection of online copyright breaches. Shape recognition software can also be used for medical diagnosis, including the detection of tumors and the measurement of internal organs, as well as for spotting errors and flaws in industrial automation. Journalism, marketing, remote sensing for information about the environment, forecasting the weather, and satellite photo monitoring are some further potential uses. Fashion, graphic design, copyright databases, engineering and building design, art museums, museums, archaeology, and online image searches CBIR technology also has applications in mapmaking, forensic analysis for biometric identification matching and crime detection analysis of security systems, radar engineering for target detection and identification and guiding of missiles and airplanes, and robotic systems for motion control via visual cues and object recognition are other fields in which it can be used.

10. Conclusion

The paper therefore helps create an idea and provide a basic understanding of Content based image retrieval and the processes involved in executing it. It provides a streamlined idea of the internal processes and the features of such systems which aims to help future developers gain better understanding of the basic concepts. Content based image retrieval has wide ranging implications and applications in future as it helps manage large chunks of visual data and can pave a definite way to guide data scientists and aid multiple big data systems. It will minimize human efforts to a great degree and help make error free perceptions and analysis. CBIR techniques will be invaluable in areas such as medicine, social media, security systems, identification purposes and managing public and private databases.

Compliance with ethical standards

Acknowlegement

I would like to express my utmost gratitude to our guide and esteemed faculty Varsha K. Patil and other teaching faculty of AISSMS Institute of Information technology. Their support, encouragement and mentorship has been an invaluable input to pave a path for our research and boost our morale and ideas. I would also like to thank Aditya Patil of BITS, Goa who helped us in editing, formatting and research purposes. His input has been a precious addition to our efforts.

I would also like to thank our college, AISSMS Institute of Information technology for giving us the opportunity to learn under their guidance and mentored us to this stage. We hope our paper makes a meaningful and significant contribution to the field.

Disclosure of conflict of interest

There is no conflict with any existing studies or research papers as well as anything on the academic published sphere.

References

[1] M. Alrahhal and K. P. Supreethi, Content-Based Image Retrieval using Local Patterns and Supervised Machine Learning Techniques, 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 118-124, doi: 10.1109/AICAI.2019.8701255.

[2] Duan, Guoyong & Yang, Jing & Yang, Yilong. (2011). Content-Based Image Retrieval Research. Physics Procedia. 22. 471–477. 10.1016/j.phpro.2011.11.073.

[3] Dhana Sekaran S, N. Herald Anantha Rufus, D. Anand, R. Senthil Rama, Archana Kumar, A.S. Vigneshwar, Evolutionary Optimization with Deep Transfer Learning for Content based Image Retrieval in Cloud Environment, 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), pp.826-831, 2022.

[4] D. Raja, M. Karthikeyan, Content-based Image Retrieval using Reptile Search Algorithm with Deep Learning for Agricultural Crops, 2022 7th International Conference on Communication and Electronics Systems (ICCES), pp.1038-1043, 2022.

[5] Ramamurthy B, M. Prasanna, CBMIR: Content-Based Medical Image Retrieval Using Hybrid Texture Feature Extraction Method, 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), pp.1-7, 2022.

[6] Devendra Walkoli, Vinit Sali, Saurabh Patil, Rohan Sonawane, Bodireddy Mahalakshmi, Content-Based Image Retrieval using SIFT and CNN, 2021 Asian Conference on Innovation in Technology (ASIANCON), pp.1-5, 2021.

[7] Niveditha Arunkumar, A Ranjith Ram, CBIR Systems: Techniques and Challenges, 2020 International Conference on Communication and Signal Processing (ICCSP), pp.0141-0146, 2020.

[8] Mayank R. Kapadia, Chirag N. Paunwala, Content-Based Image Retrieval Techniques and Their Applications in Medical Science, Biomedical Signal and Image Processing with Artificial Intelligence, pp.123, 2023.

[9] Subramanian Manoharan, Lingamuthu Velmurugan, Venkatesan Chandran, Perumal Sasikumar, Content-Based Image Retrieval Using Colour, Gray, Advanced Texture, Shape Features, and Random Forest Classifier with Optimized Particle Swarm Optimization, 2022 International Journal of Biomedical Imaging, 10.1155/2022/3211793.

[10] Jindal, Sonika & Seth, Nitika. (2017). A Review on Content-Based Image Retrival. International Journal of Computers & Technology. Vol 15. 7498-7503. 10.24297/ijct.v15i14.5640.

[11] Jadhav, Shital & Patil, Sonal & Solunke, Hiralal. (2021). Comprehensive Review of Content based Image Retrieval. International Journal of Computer Applications. 183. 37-40. 10.5120/ijca2021921633.

[12] Khassaf, Nuha & Hameed, Shaimaa. (2020). Image Retrieval based Convolutional Neural Network. Al-Mustansiriyah Journal of Science. 31. 43. 10.23851/mjs.v31i4.897.

[13] Yakin, Syamsul & Hasanuddin, Tasrif & Kurniati, Nia. (2021). Application of content-based image retrieval in digital image search system. Bulletin of Electrical Engineering and Informatics. 10. 1122-1128. 10.11591/eei.v10i2.2713.

[14] Macarthur, Sean & Brodley, Carla & Shyu, Chi-Ren. (2000). Relevance Feedback Decision Trees in Content-Based Image Retrieval.

[15] S.Rubini, R.Divya, G.Divyalakshmi, Mr T.M. Senthil Ganesan (2018). CONTENT BASED IMAGE RETRIEVAL (CBIR). International Research Journal of Engineering and Technology (IRJET). e-ISSN: 2395-0056 p-ISSN: 2395-0072.

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

[17] N. Ali, D. A. Mazhar, Z. Iqbal, R. Ashraf, J. Ahmed, and F. Zeeshan, Content-based image retrieval based on late fusion of binary and local descriptors, International Journal of Computer Science and Information Security (IJCSIS), vol. 14, no. 11, 2016.

[18] X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, Content-based image retrieval by integrating color and texture features, Multimedia Tools and Applications, vol. 68, no. 3, pp. 545–569, 2014.

[19] 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.

[20] Z. Jiexian, L. Xiupeng, and F. Yu, Multiscale distance coherence vector algorithm for content-based image retrieval, The Scientific World Journal, vol. 2014, Article ID 615973, 13 pages, 2014.

[21] G. Papakostas, D. Koulouriotis, and V. Tourassis, Feature extraction based on wavelet moments and moment invariants in machine vision systems, in Human-Centric Machine Vision, InTech, London, UK, 2012.

[22] Latif, Afshan & Rasheed, Aqsa & Sajid, Umer & Jameel, Ahmed & Ali, Nouman & Ratyal, Naeem Iqbal & Zafar, Bushra & Dar, Saadat & Sajid, Muhammad & Shehryar, Tehmina. (2019). Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review. Mathematical Problems in Engineering. 2019. 10.1155/2019/9658350.

[23] Liu, Guang-Hai & Li, Zuo-Yong & Zhang, Lei & Xu, Yong. (2011). Image retrieval based on micro-structure descriptor. Pattern Recognition. 44.2123-2133. 10.1016/j.patcog.2011.02.003.

[24] Irtaza, Aun & Jaffar, Arfan. (2014). Categorical image retrieval through genetically optimized support vector machines (GOSVM) and hybrid texture features. Signal, Image and Video Processing. 9. 10.1007/s11760-013-0601-8.

[25] Fadaei, Sadegh & Amirfattahi, Rassoul & Ahmadzadeh, M.R.. (2017). Local Derivative Radial Patterns: A New Texture Descriptor for Content-Based Image Retrieval. Signal Processing. 137. 10.1016/j.sigpro.2017.02.013.

[26] Shao, Hong & Wu, Yueshu & Cui, Wen-cheng & Zhang, Jinxia. (2008). Image Retrieval Based on MPEG-7 Dominant Color Descriptor. Proceedings of the 9th International Conference for Young Computer Scientists, ICYCS 2008. 753-757. 10.1109/ICYCS.2008.89.

[27] Duanmu, Xiaoyin. (2010). Image Retrieval Using Color Moment Invariant. ITNG2010 - 7th International Conference on Information Technology: New Generations. 200-203. 10.1109/ITNG.2010.231.

[28] Islam, M.M., Zhang, D. and Lu, G., 2008, December. Automatic categorization of image regions using dominant color-based vector quantization. In 2008 Digital Image Computing: Techniques and Applications (pp. 191-198). IEEE.

[29] Khan R, Barat C, Muselet D, Ducottet C. Spatial orientations of visual word pairs to improve bag-of-visual-words model. In: Proceedings of the British Machine Vision Conference. BMVA Press; 2012. p. 89–1.