Journal of
Image Processing & Pattern
Recognition Progress

ISSN: 2394-1995
Volume 12, Issue 1, 2025
January–April
DOI (Journal): 10.37591/JoIPPRP

https://journals.stmjournals.com/joipprp

Research JoIPPRP

# Content-based Image Retrieval: Recent Trends and Techniques

Monika Varshney[1], Raj Kumar[2,*]

***Abstract***

*Due to the affordability of digital devices and the accessibility of internet technologies, a vast number of multimedia databases have been established for various applications. These image databases increase the need for efficient picture retrieval search strategies that meet user requirements. When compared to other systems, the content-based image retrieval (CBIR) system is one of the most widely used systems for retrieving images from enormous databases. Much work has been done in the last few years to enhance CBIR techniques, with an emphasis on closing the semantic gap between high-level and low-level characteristics. Owing to the growing body of research in this area, the present study examines, evaluates, and contrasts, the most advanced techniques currently used in the field of CBIR. This research analyzes the data and gives a summary. It presents new low-level feature extraction techniques, machine learning algorithms, similar metrics, and a comparison of the state-of-the-art techniques' performance with potential future developments.*

**Keywords:** Content-based image retrieval, feature extraction, color feature, texture feature, shape feature, SIFT, SURF, similarity metrics, machine learning, convolution neural network, SVM

## INTRODUCTION

Digital image growth is accelerating due to the quick development of new digital technologies and multimedia. To locate images in the database, a lot of interest is paid. Creating a productive method for extracting images from these massive datasets is much needed. It is difficult to retrieve particular images from a vast image database. Text-based retrieval has been utilized in recent years to retrieve images from databases [1]. However, two main issues with text-based picture retrieval are human perception and image annotation. An excellent substitute for conventional text-based image searching is content-based image retrieval (CBIR) [2, 3].

An effective visual-content-based strategy for searching and retrieving comparable visual images from sizable digital image collections is developed using CBIR techniques. In numerous computer vision and artificial intelligence-related application domains, CBIR techniques are extensively employed [4]. The majorities of CBIR approaches that have been suggested automatically extract low-level elements (such as color, texture, and forms) and compare them to determine how similar two images are. However, there is often a semantic gap between high level and low-level features, meaning that low level traits are unable to adequately convey high level notions.

The most crucial component of image retrieval systems is featuring extraction. In general, features fall into two categories: local features and global

*Author for Correspondence
Raj Kumar
E-mail: rajkumar.citd@gmail.com

[1]Assistant Professor, Department of Computer Science, Dharm Samaj College, Raja Mahendra Pratap Singh University, Aligarh, Uttar Pradesh, India
[2]Assistant Professor, Department of Computer Science, Dr. Bhimrao Ambedkar University, Agra, Uttar Pradesh, India

Received Date: November 30, 2024
Accepted Date: January 16, 2025
Published Date: February 07, 2025

**Citation:** Monika Varshney, Raj Kumar. Content-based Image Retrieval: Recent Trends and Techniques. Journal of Image Processing & Pattern Recognition Progress. 2025; 12(1): 1–32p.

features, which include color, texture, shape, and spatial information as well as local features, which vary depending on the feature extraction technique employed. The whole image is represented by the global feature. Due to their fast feature extraction and similarity computations, these characteristics are advantageous [5]. They were unable to distinguish the object in the picture from the background, though. Local features are better suited for object detection and image retrieval when compared to global characteristics [6]. The process of identifying and naming an object [7] in a picture is known as object recognition [6–8].

Local features are the focal points or specific areas within an image, such as borders, blobs, and corners. They withstand changes in background color, scale, rotation, translation, clutter, and partial occlusions with resilience [6].

**Feature Extraction in CBIR**

The first and most crucial step in CBIR is feature extraction, which aims to translate human perception into a numerical description. It involves identifying and removing the image's most pertinent and expressive elements. Local and global feature extraction techniques are divided into two categories. To enhance CBIR performance, the extracted features are utilized in either supervised or unsupervised machine learning techniques [7].

Deep learning is being used more and more in image retrieval to reduce running time costs [9]. The rise in dimensionality is an additional issue that hinders CBIR performance. High dimensional features are typically produced during the numerical representation of visual image content. This problem's solution is dimensionality decrease [10–12]. Using similarity measurements, CBIR performance is determined. Achieving high system accuracy is feasible by choosing the right similarity measure. The image dataset selection affects the effectiveness of CBIR, which can be measured using a variety of metrics such as precision, recall, and running time [9, 11, 13–20].

Various surveys that assess the current state-of-the-art and contain various new avenues have been carried out to find recent trends in the field of CBIR. Much research identified significant obstacles and reviewed the state-of-the-art [21–25].

The Organization of Paper is as Follows:

A brief overview of the CBIR system architecture is provided; a review of recent CBIR studies based on previous years' surveys and image retrieval techniques is covered; comparison of several CBIR methodologies in performance evaluation; examining the most recent machine learning algorithms employed in CBIR; and then the similarity measures.

## CBIR ARCHITECTURAL DESIGN

A user submits a query image, which is referred to as an online process. In parallel, dataset images are reset before the query image is sent, which is referred to as an offline process. The process of feature extraction is now applied to the input image, converting the visual image into a numerical format. In the feature matching phase, the features that were extracted from the input image are now compared to the dataset. A few similarity metrics are also observed in order to determine how similar the input image and the dataset image are to each other. The image is retrieved if the input image matches the provided dataset as shown in Figure 1.

## LOW-LEVEL FEATURES

As we previously discussed, the most crucial aspect of CBIR is feature extraction and picture selection, which reflect the semantic information of images [26]. These features are further separated into local features [6], which are determined by calculating certain important places, such as corners, edges, and blobs, and global features, which include color, texture, shape, and spatial information. The features of the entire image are provided via global features. Global aspects, such as the information between nearby pixels, disregard local information. Invariants like illumination, viewpoint, background clutter, and occlusion do not affect global features.

 

Journal of Image Processing & Pattern Recognition Progress
Volume 12, Issue 1
ISSN: 2394-1995

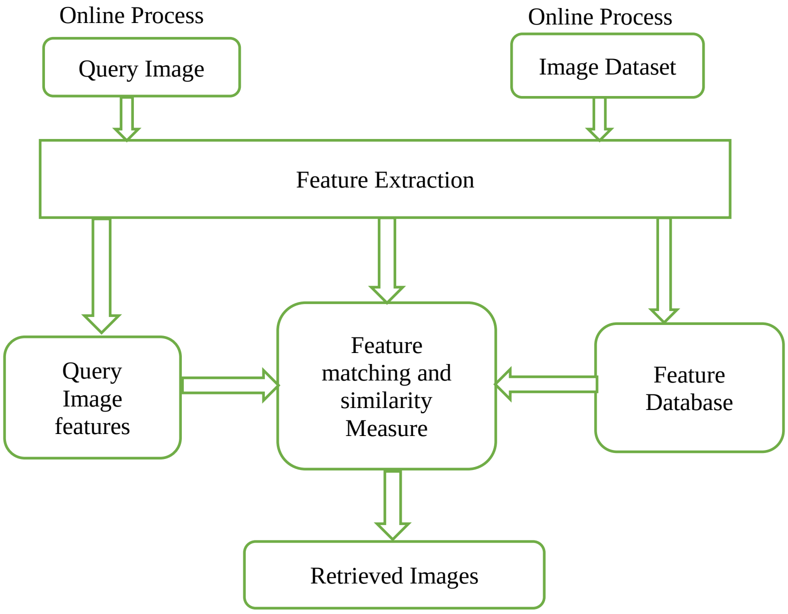


**Figure 1.** CBIR system architecture.

## Color Features

One of the CBIR System's most significant low-level and potent features is color [27–30]. The color properties of a picture are determined by the intensity values of a pixel in any of its color planes. Different color spaces used in image retrieval are utilized to determine color features. RGB, HSV, HIS, YCbCr, and LUV are the color spaces most frequently utilized in CBIR. The color content of a picture can be represented in a variety of ways, including dominant color descriptors [30], color histograms [31], color correlograms [32], color co-occurrence matrices [30], color moments [33] and many more. Since color features are unaffected by translation, rotation, and scale changes, they are regarded as powerful features [13, 34].

### *Color Space*

Numerous color spaces, including RGB, CMYK, YCbCr, YUV, HSV, HIS, YCgCb, and Lab planes, are widely used and well-liked [35]. The RGB color space, commonly used in digital imaging, is based on the primary color's red, green, and blue. Each of these colors is represented by values that range from 0 to 255. Cyan Magenta Yellow, or CMY, is a secondary hue.

Color is represented through the HSV or CIE Lab and LUV color spaces, which are grounded in the way humans perceive vision [36]. A more reliable method of representing images is the Hue Saturation Value (HSV) plane, which is mostly utilized for color characteristics. The YUV color space is utilized by the commonly used video formats PAL, NTSC, and SECAM. Here, UV stands for color information, while Y is the brightness component. V is between 0 and 157, U is between 0 and 112, and Y is between 0 and 255. The altered form of YUV space is called YIQ (Yellow in-phase Quadrature) color space.

The scaled representation of YUV space is called YCbCr, where Y represents the luminance component, which ranges from (16-235), and Cb and Cr represent the chrominance components, which range from (16-240). Min-Max-Difference-Hue D is the difference between the min and max values in the HMMD color space, where min and max represent the RGB minimum and maximum values, respectively. Compared to HSV and HIS color spaces, it is more efficient.

### *Color Histogram*

The color histogram is a popular, straightforward, and effective color feature [37, 32, 38]. It describes the image's three color channels' probabilistic intensity distribution. Methods based on color histograms are simple to use and independent of scale and orientation. Among the many benefits of color histograms are their speed, resilience, low memory usage, and suitability for global color descriptors. It is unable to contain high dimensional feature vectors and characterize the spatial properties of images, though. Images' histograms are computed by measuring how many pixels of each color appear in the image and in distinct bins. Additionally, histograms based on global color and block color histograms are suggested by Singh and Kaur [37].

***Color Coherence Vector***

The inability of color histograms to express spatial information is one of its main disadvantages. Singh and Kaur suggested the Color Coherence Vector (CCV) extension histogram [37], which divides each histogram bin according to local characteristics [27–29]. The number of coherent pixels compared to incoherent pixels is stored in a CCV. CCVs distinguish between coherent and incoherent pixels to produce unique color histograms. Spatial color histograms, annular color histograms, and angular color histograms are enhanced versions of Color Contrast Vectors (CCVs).

***Color Co-occurrence Matrix (CCM) [30, 38]***

Co-occurrence matrix of colors (CCM), the traversal of neighboring pixel color differences in a picture is represented by Qiu and Kong [30, 38]. The probability that each pixel and its neighboring pixel will have the same color is determined using CCM.

A color co-occurrence matrix (CCM) can be created by representing each image with four scan pattern motifs, which can then be further subdivided into four two-dimensional matrices. These matrices can be used to record color variations in images and produce scan pattern motifs.

***Color Correlogram [32]***

Color Correlogram captures the distribution of both color and spatial information and provides information on how the spatial correlation of color pairs varies with distance. Additionally, it offers the spatial details of an image's pixels. Information about the spatial distribution of comparable hues is provided by the Color Auto Correlogram [32], which can simplify computation. The k-th entry in a color correlogram, which is a table organized by color pairs, represents the likelihood of encountering a pixel of color j at a distance k from a pixel of color i in the image. Let I stand for the complete set of pixels in the image, and let Ic(i) stand for the set of pixels of color c(i). The color correlogram is then defined as follows:

$$\gamma(i,j)^{(k)} = Pr_{p1\in c(I),p2\in I}\,[p2 \in I_{c(j)}\ |p1-p2| = \ k] \tag{1}$$

Where $\mathrm{i,j} \in \{1,2,\ldots,\mathrm{N}\}, \mathrm{k} \in \{1,2,\ldots,\mathrm{d}\}$, and $|\mathrm{p1}-\mathrm{p2}|$ is the distance between pixels p1 and p2.

***Color Difference Histogram (CDH)* [38, 39]**

Color Difference Histogram (CDH) determines the consistent color differences between two pixels in La* b* color space if the backgrounds differ in terms of color and edge orientations [38]. Compared to histograms and correlograms, CDH is more effective.

***Dominant Color Descriptor DCD***

Color descriptors was proposed by MPEG-7 [40, 41, 42]. It can be used in object-based image retrieval. This method is based on precisely describing the prominent colors of an image or region of interest. Its main emphasis is on dominant colors. it extracts dominant colors from image. DCD consists of two primary elements: the representative colors and their respective percentages. It utilizes color characteristics to identify the key colors and their proportions in an image. The image is divided into eight coarse segments, and the quantized color is the mean value of each segment. The closest colors are then combined using clustering. Ultimately, five main colors and their respective percentages have been determined.

 

Journal of Image Processing & Pattern Recognition Progress
Volume 12, Issue 1
ISSN: 2394-1995

***Color Moments***

Color moments is a statistical metric selected to depict the image's color details [33]. It provides two condensed representations of the image's pixel distribution information. The average information about the image's pixel distribution is provided by the first order moment, or mean. Second order moment (variance) gives the closeness of the pixel distribution about mean color estimation, while the third order moment (skewness) provides an effective and efficient representation of images.

$$\mu_i = \frac{1}{N}\sum_{j=1}^{n} P_{ij} \tag{2}$$

$$\sigma_i = (\frac{1}{N}\sum_{j=1}^{n}((P_{ij}-\mu_i)^2)^{\frac{1}{2}} \tag{3}$$

$$S_i = (\frac{1}{N}\sum_{j=1}^{n}((P_{ij}-\mu_i)^3)^{\frac{1}{3}} \tag{4}$$

Similarity between two image distributions is defined as the sum of weighted differences between the moments of two distributions.

$$d_{mom}\,(H,I) = \sum_{i=1}^{r} w_{i1}\,\left|E_i^1 - E_i^2\right| + w_{i2}\left|\sigma_i^1 - \sigma_i^2\right| + w_{i3}\,|S_i^1 - S_i^2| \tag{5}$$

where (H, I) are the two image distribution components, i is the current channel index (1=H, 2=S, 3=V), r is the number of channels, here 3, $E_i^1$, $E_i^2$, are the first order moments of two image distributions, $\sigma_i^1$, $\sigma_i^2$ are the second order moments of two image distributions, Si1 and Si2 represent the third-order moments of the two image distributions, while wi denotes the corresponding weights for each moment. Pairs of images are ranked based on $d_{mom}$ values. Images with lower DMoM values are ranked higher and are more like each other than those with higher DMoM values.

The following procedure is used to calculate moments. For efficiency, all pictures are first resized to the same size. Probability distributions serve as the foundation for color moments. Therefore, the size of the image should not affect the comparison's outcome. The three-color moments were calculated for the query image using the formula provided earlier. After then, the computations are carried out once again for our database photographs. After assigning suitable weights, the $d_{mom}$ value is determined, and the pictures are arranged in ascending order of this value. The result images are those that have the lowest $d_{mom}$ values. Thus, color moments can be applied as a method for color-based image comparison.

**Texture Features**

Texture is another important low-level feature of an image used for image retrieval [29]. Texture refers to a property of an image that is commonly utilized in image processing and computer vision. Texture features are used in identifying objects and regions of interest in an image. These features can identify repetitive patterns or structures showing granular details of an image. Approaches for extracting texture features from an image can be categorized as, statistical methods and structural approaches. In structural approach, the surface pattern is repeated whereas in statistical approach surface patterns are not regularly repeated in the same pattern. Some widely used algorithms used for texture analysis are gray-level co-occurrence matrix (GLCM), Gabor wavelet transform [45], local binary patterns (LBP) [44, 45], Tamura Texture feature [46], steerable pyramid, and edge histogram descriptor.

***Co-Occurrence Matrix***

Co-occurrence matrix can be used to extract texture information from an image. Texture features such as entropy, contrast, and homogeneity can be derived from an image using a co-occurrence matrix [47].

***Tamura Texture Feature***

Tamura *et al*. proposed another texture feature which is a computational approximation that is based on the study if human visual perception of texture [46]. Basically, there are six textual features:

directionality, line-likeness, coarseness, contrast, roughness and regularity for feature calculations. Directionality focuses to measure the total degree of direction as mono directional or bidirectional. Line likeliness refers to the form of an object, while coarseness represents the fundamental texture characteristic. It varies as the scale and frequency of repetition change. The contrast is used to represent the dynamic grey levels. Roughness is typically selected for tactile textures, while regularity refers to the variations in the placement pattern.

***Steerable Pyramid***

The steerable pyramid model is applied in texture recognition by breaking down an image into multiple directional sub-bands and a low-pass residual [48]. This process decomposes the image into several groups of oriented sub-bands along with one decimated low-pass sub-band. Texture features are then derived from the results of these oriented filters [49].

***Discrete Wavelet Transform (DWT)***

For texture analysis and image classification, wavelet transform methods are used. In wavelet transform [50, 51], an image is decomposed into four different frequency bands implemented by a low-pass filter (L) and high-pass filter (H). Gabor filter was found to be very effective filter among the various wavelet transform filters. Due to its effectiveness Gabor filter is widely used in different types of applications [52]. A 2-D discrete wavelet transform is efficiently computed by first applying the filter bank to each column of the image, followed by applying the filter bank to each row of the resulting coefficients [53].

Figure 2 shows the standard decomposition of an image into four frequency bands (LL, HL, LH, HH). In the first phase on decomposition, one low-pass and three high-pass sub images are formed. In the second phase, the low-pass sub-image is further broken down into one low-pass and three high-pass sub-images. Sub images are considered for the purpose of quick filtering. The horizontal, vertical, and diagonal sub-images provide details about variations in brightness.

***Gabor Wavelet Transform (GWT)***

Gabor filters are a set of wavelets, each of which captures energy at a particular frequency and direction, providing a localized frequency representation that can be used for texture analysis. The Gabor Wavelet Transform (GWT) is a widely utilized method for extracting features in texture analysis. This function is derived by modulating a Gaussian with a complex sinusoid, characterized by the sinusoid's frequency (ω) and the standard deviations (σx and σy) of the Gaussian envelope [18]:

$$\Psi(x,y) = \frac{1}{2\Pi\sigma_x\sigma_y} e^{\left[-\left(\frac{1}{2}\right)\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi j\omega x\right]} \tag{6}$$

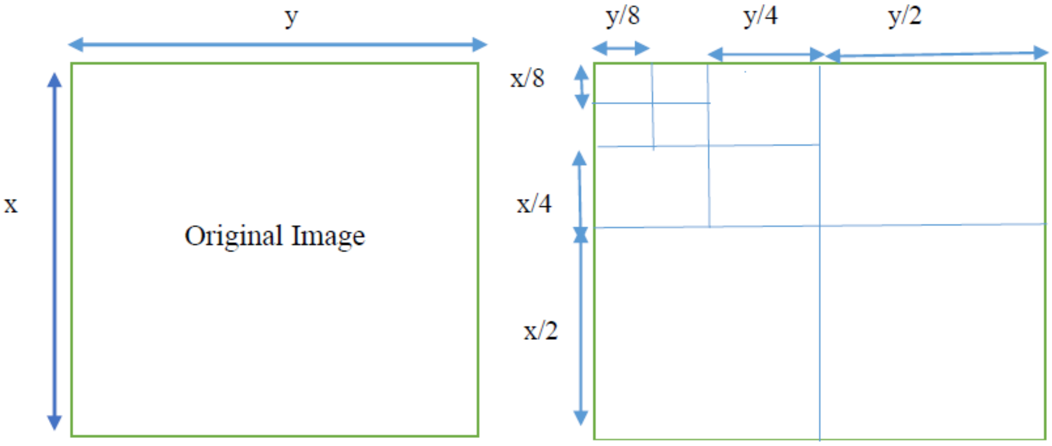

**Figure 2.** Standard wavelet decomposition [53].

The Gabor functions do not lead to an orthogonal decomposition, meaning the Gabor wavelet derived from the wavelet transform is redundant. By performing dilation and rotation on the function ψ(x, y), the resulting Gabor wavelets are produced as follows:

$$\Psi(x,y) = a^{-m}\Psi(x_1,y_1) \tag{7}$$

The Gabor filter response is obtained by convolving the Gabor window with the image I. Here, x1 and y1 are defined as:

x1=a–m (x cos(Ɵk)+y sin(Ɵk))

y1=a–m (–x sin(Ɵk)+y cos(Ɵk))

where Ɵ=nπ/K, with m ranging from 0 to S–1 and n ranging from 0 to K–1. K represents the number of orientations, and S denotes the number of scales.

$$G_{mn}(x,y) = \sum_s \sum_t I(x-s, y-t)\psi^*_{mn}\ (s,t) \tag{8}$$

The image's features are derived by applying the filter at various orientations and scales, resulting in an array of magnitudes. To capture the texture similarity, the mean and standard deviation of these transformed coefficients are computed. These values represent the texture characteristics of the region [54, 55].

$$\mu_{mn} = \sum_x \sum_y G_{mn}\ (x,y) \tag{9}$$

$$\sigma_{mn} = \sqrt{\sum_x \sum_y (|G_{mn}(x,y) - \mu_{mn}|)^2} \tag{10}$$

**Local Binary Pattern**

Local Binary Pattern [53] [Ojala] is one of the methods used as a texture feature in image retrieval for feature extraction of gray scale images. LBP is resistant to changes in rotation and scale. It is also quicker and computationally simpler to implement. These techniques can capture local details with greater accuracy and precision when compared to other descriptors. It describes the spatial configuration of a local image texture and when combined with variance measures it is said to be more powerful tool for rotation invariant texture analysis. The first use of Local binary Patterns proposed in Sarwar [56]. DFT of LBP histogram is a variant of LBP.

It describes the texture pattern of an image, and the Fourier transform of LBP examines the relationships between the orientations of each pattern. Penatti [57] employed histogram intersection as a similarity metric and talked about effective feature descriptors and operations for image retrieval, such as Local Binary Patterns (LBP) [58, 59, 60, 61], Local Ternary Patterns (LTP) [62, 63, 60], Local Derivative Patterns (LDP) [64, 59, 60], and Local Tetra Patterns (LTTP) [65, 60].

In recent years, various versions of LBP descriptors have been introduced for extracting texture features from color images. One such color descriptor, Local Color Vector Binary Pattern (LCVBP), was proposed by Lee *et al.* [66]. OC-LBP operator is an effective version of LBP. It reduces the dimensionality of LBP features. Different color models have been proposed by Zhu *et al*. as the extensions of the OC-LBP which is applied on three channels of color image [32]. Another local color descriptor is Quaternion local ranking binary pattern (QLRBP). In this operator local information could not be completely described.

Therefore, performance of this method is not good for image retrieval. A color version of LBP operator is LBPC to extract texture color of color images.

 

Two additional local texture descriptors for image retrieval have been introduced: Color Radial Mean Completed Local Binary Pattern (CRMCLBP) and Prototype Data Model (PDM).

These techniques are built on the Radial Mean Local Binary Pattern (RMLBP), which is resilient to noise. These two new features were more efficient compared to earlier operators. Both newly proposed methods were enhanced using an adaptive feature weighting algorithm based on Particle Swarm Optimization. Feature weighting has been considered as a very effective approach in enhancement of the retrieval accuracy.

## Shape Feature

Another crucial visual component of the CBIR system is shape. One of the fundamental characteristics used to characterize the content of images is shape, which can also be utilized to convey potent information [67, 68]. One of the low-level characteristics used to identify items is shape. Semantic information is carried by shape [69, 70].

There are two types of shape representation techniques: region-based and boundary-based, as classified in Figure 3 [71, 72]. Region-based techniques extract features based on the entire region, whereas boundary-base techniques just use the region's contour or outside boundary and disregard its inside. The Fourier Descriptor [73] and Moment Invariants [74] are the most effective representations for extracting features.

### *Global Contour-Based Shape Descriptors*

Mathematical procedures called global contour-based shape descriptors are used to depict an object's general shape. These descriptors can be used to compare and categorize various forms because they are based in the object’s outlines or contours.

The Fourier descriptor is an illustration of a global contour-based shape descriptor. This technique breaks down an object's shape into a set of frequencies using the Fourier transform.

This can be used to create a shape signature that accurately depicts the object's general shape. The shape distribution descriptor, shape context descriptor, and centroidal profile descriptor are further techniques [68, 75].

#### *Fourier Transforms (FD)*

Many applications involving form representation, especially character recognition, have successfully employed the FD. Due to their attractive characteristics, including easy derivation, straightforward normalization, and noise resistance, they are very popular in a variety of applications [76]. A complex vector created from the boundary coordinates (xn, yn), where n ranges from 0 to N–1, is processed using the Fourier transform to yield the FD.

#### *Wavelet Descriptor (WD)*

Wavelet transformations were initially employed by the WD to describe how planar closed curves look like. This method divides the object into many, varying-sized pieces using a variety of resolutions. Higher resolution components contain global information, while lower resolution components contain more detailed local information [72, 77]. Wavelet descriptors are unique, noise-invariant, noise-insensitive [78], and stable against boundary fluctuations, among other benefits.

### *Structural Contour-based Shape Description Methods*

Structural shape representation is another member of the form analysis family. The structural technique divides shapes into boundary pieces called primitives.

 

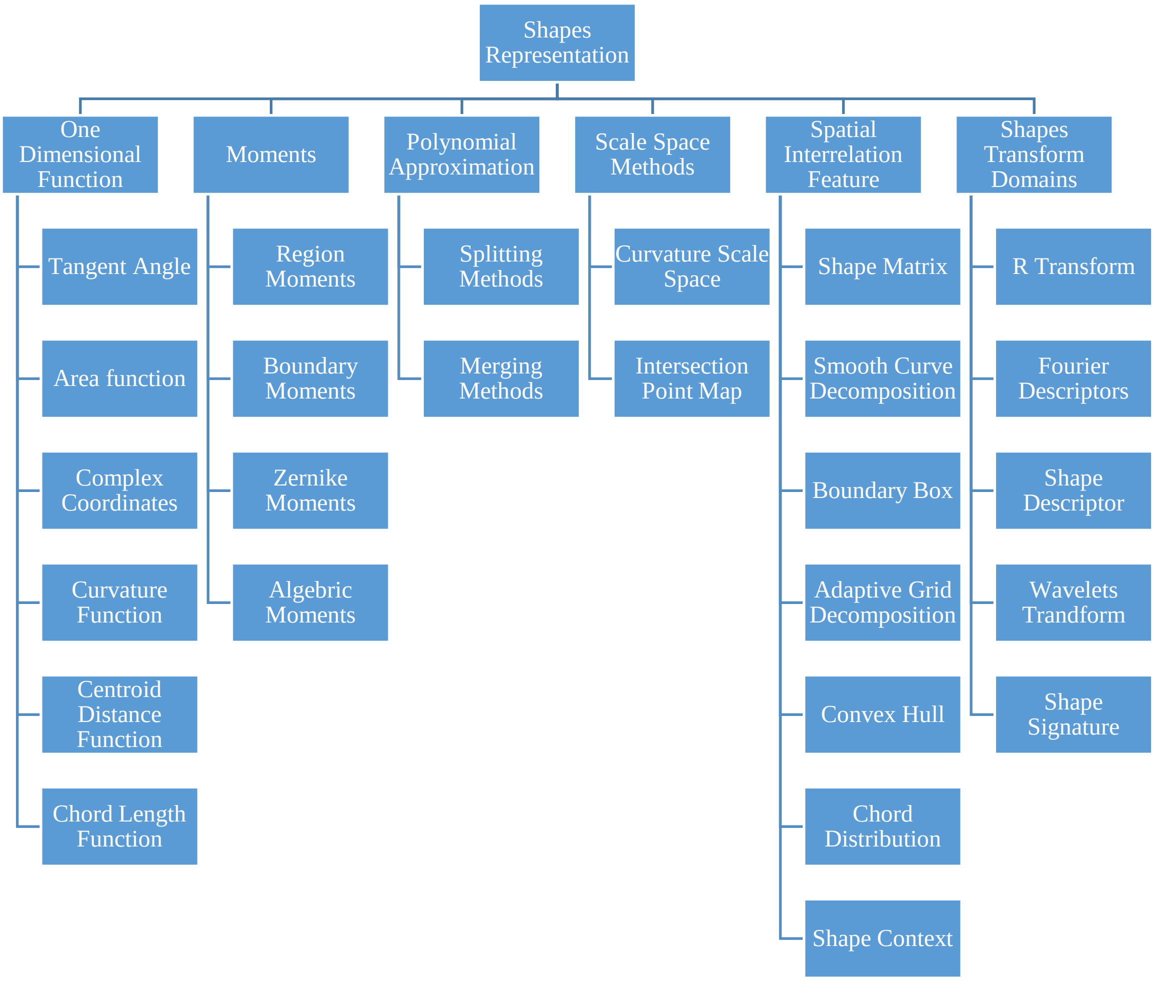


**Figure 3.** Shape representation and description method classification.

To represent shapes, different structural approaches employ different primitives and arrange them in different ways. Common boundary decomposition methods include polygonal approximation, curvature decomposition, curve fitting, and chain coding [68].

*Global Region-based Shape Descriptors*

Rather than relying solely on boundary information, region-based methods take into account every pixel that is accessible within the shape region prior to generating the shape representation. Shape description makes use of moment descriptors. Another way to classify region-based approaches is global or structural.

As previously mentioned, segmentation is a component of the structural technique. The shape matrix, Zernike and image moments and Angular Radial Transform (ART) techniques are examples of the global approach.

*Structural region-based shape descriptor*

The structural region-based shape representation is an important addition to the collection of shape analysis methods. The medial axis and convex hull are examples of the structural approach.

**Spatial Information**

Spatial information is absent from the low level elements previously addressed. The arrangement of elements in a two-dimensional image is known as a spatial feature. For instance, the histogram of an image would be identical if we took two distinct sections with two different spatial contents. One of the finest techniques for capturing the spatial characteristics of images is spatial pyramid matching [76–81]. The drawback of spatial features is that they involve intricate calculations.

**Combined features based CBIR**

From the early studies in CBIR system it was found that using only a single low level feature such as either only color, or shape or texture may be inefficient. It may fail to retrieve image accurately and efficiently which is similar to query image. To produce efficient results, combination of features like color-texture [37], color-shape etc. were used. Combining multiple features will improve the performance of CBIR and would give more accurate results as compared to using a single feature. Fast color-spatial feature based information retrieval methods were proposed by Lin *et al.* [82]. Kavitha Rao *et al*. proposed a color-texture features of image sub-blocks [42]. Charles *et al.* introduced a novel mesh color texture pattern for Information retrieval system [35].

In order to recover images, Shrivastava and Tyagi presented a CBIR system with a feed forward architecture that consists of three steps [34]. The first stage was retrieving the relevant N images from the dataset, which contained M images, based on color attributes. To determine the color features, the authors employed a color histogram. In the following step, the texture feature is calculated using the Gabor filter, and P relevant photos are extracted from the N image subset. Ashraf [83] developed a new CBIR method that relies on extracting color and texture data. The final step involves calculating the Fourier descriptor as shape features to get K relevant photos from P. Four extraction techniques were applied: co-occurrence matrices, wavelet moment, color moment, and color histogram.

Ashraf [83] integrated the k-means clustering algorithm with particle swarm optimization (PSO), a stochastic method. They also proposed a color-texture information retrieval approach utilizing Gaussian copula models of Gabor wavelets by Li, *et al*. [32].

A new CBIR system was proposed, combining color, texture, and shape features. To extract these characteristics, the system utilized the color auto correlogram for color, Gabor transform for texture, and wavelet transform for shape. Image analysis at a single resolution level may lose some valuable details [84].

Thus, a new multi-resolution analysis technique was created by Srivastava and Khare that examines photos at several levels, with each level obtaining information that the previous level missed [13]. This method is based on using the Local Binary Pattern (LBP) descriptor to extract texture features and Legendre moments to capture form features from the texture data at various resolution levels. While LBP focuses on local feature extraction, it contributes to a powerful feature vector when local features are integrated with global ones. Latif *et al.* proposed an invariant CBIR system to texture rotation and color change [3]. The suggested method creates a 360-degree feature vector by concatenating color and texture characteristics. Images are quantized through a color histogram and then converted to the HSV color space to capture color details. To ensure invariance to lighting changes, only the Hue and Saturation channels are utilized. For texture analysis, rotation-invariant features are obtained using rotated local binary patterns (RLBP).

As a multistage CBIR technique, an effective framework for information retrieval. In order to reduce the search space and, consequently, the computational cost, color features were retrieved in the first stage using color moment, which was calculated by calculating the mean and standard deviation for each channel in RGB color space [85]. The new sub-dataset created from the first step is used in the second stage to extract texture and shape (edge) properties from the images. Texture information was extracted using LBP, while edge information was extracted using a Canny edge detector.

A novel method for the selection of seed points for dominant color-based image retrieval technique. The proposed dominant color descriptor was evaluated using four image datasets in the experiments, leading to improved results [86].

 

However, the semantic gap remained because the same color information could be assigned to images in different semantic classes, so the suggested method needs to be fused with other feature extraction methods (shape, texture, and spatial information). To reduce the semantic gap, a CBIR system that merges color and edge features to form a feature descriptor. For color extraction, the authors used color histogram [87].

The YCbCr color space's canny edge histogram was employed for edge extraction. The authors used the Haar wavelet, which is faster in terms of computations, to speed up the calculation of the discrete wavelet in order to achieve better enhancement [88]. In order to determine the semantic class of the photos, the suggested method employed an artificial neural network (ANN), which requires more time for both training and testing. The efficacy of the system was demonstrated by the stated mean precision and recall findings, which used Manhattan distance as its similarity measurement. However, it also suffers from the lack of spatial information, and no information about its computational cost efficiency is available.

A novel CBIR system is presented, which is based on extracting global and local texture features in both frequency and spatial domains and color features in spatial domain [89]. Images are first filtered using a Gaussian filter to lessen noise effects, and then GLCM is used to extract global texture features in the spatial domain. Color features are obtained by using a quantized histogram in the RGB color space. The Gabor filter is used to extract local texture information in order to improve retrieval performance.

A subjective methodology for the CBIR system was created based on the combination of low-level features (color and texture) [90]. Color features were obtained through color moments in the HSV color space, and texture characteristics were derived using Discrete Wavelet Transform (DWT) and Gabor wavelet methods. For further enhancement, the color and edge directivity descriptor were calculated and included in the feature vector, with dimensions of 1×250. The larger feature vector dimension gives more accurate retrieval results.

A unique CBIR method that benefits from mixing color, shape, and texture [91]. Color characteristics were extracted using the Canny edge histogram and the DWT transform in the YCbCr color space, and texture features were extracted using GLCM. Shape features were obtained through the Canny edge detection method in the RGB color space. The suggested method used simulated annealing (SA) in conjunction with a genetic algorithm (GA), which raises the fitness number and improves the quality of the solution.

A framework for retrieving color texture images, which employs Shearlet domain modeling and the Copula multivariate model [92]. The Non-Subsample Shearlet Transform (NSST) dependencies between various sub-bands are modeled using the suggested framework, Gaussian Copula, while the coefficients are marginally modeled using non-Gaussian models. Six distinct schemes have been introduced to model NSST coefficients, based on the four neighboring types defined earlier. To explore the similarities within the proposed retrieval framework, the Kullback-Leibler Divergence (KLD) is calculated in different scenarios, comparing both Gaussian Copula and non-Gaussian functions.

A new technique for retrieving color-textured images on a Multivariate Generalized Gamma Distribution manifold (MG_D) [93]. Defined the MG_D, which accurately represents the statistical dependence between the dual-tree complex wavelet transform (DTCWT) of the color components, using Gaussian copula theory. The key contribution of this study is to approach the MG_D as a Riemannian manifold and introduce the geodesic distance (GD) as a method to measure Riemannian similarity on it. We do a geometrical investigation of the MG_D manifold using information geometry methods, allowing us to obtain two appropriate approximations of the GD. Experiments are carried out on five well-known color texture databases using the RGB color model within the CBIR framework.

 

The combined primitive characteristics are used to retrieve images in the conventional CBIR framework. However, since one feature may obscure another, this type of fusion is never effective. A hierarchical architecture to address this problem, taking into account characteristics like color, texture, and shape in a single hierarchy [94].

The hierarchical system's primary focus is on choosing the best order for visual feature retrieval. Semantic-based picture retrieval has thus been implemented. After identifying the region of interest using a salience map at the first level, shape characteristics based on edge histogram descriptors are added. Based on the directionality of the Tamura features, we have suggested a novel directional texture feature extraction method for the second hierarchy.

Furthermore, although human visual perception is not sensitive to every color, color is thought to be another primordial aspect of an image. The salient colors in the image can be used to visualize it, and this study develops a color image quantization-based method.

Kumar and Singh introduced LBPCd, a new color and texture descriptor, along with ULBPCd, its low-dimensional counterpart [27]. By using cross-correlation to express interaction between three color components and encode the first-order differences between different interacting elements, the descriptor LPBCd captures local information.

Very high speed is offered by the low-dimensional ULBPCd variant of the LBPCd without sacrificing classification performance. Combining two complementary descriptors, CH and CDH, substantially improves the discriminative qualities of the LPBCd and ULBPCd descriptors. The resulting effective descriptors are called LBPCd+CH+CDH and ULBPCd+CH+CDH.

**Local Features**

Global features are robust to noise and have good computation time. Their failure to take into account the local information of nearby pixels is a drawback, nevertheless. Variations in occlusion, illumination, viewpoint, and background clutter can also affect global characteristics.

Conversely, local features take into account the image's local areas. These feature descriptors use certain patches or key points, small portions of the entire image to identify sections of the image in order to extract features. Noise generally affects local attributes. SIFT [95], Color SIFT, HOG [96], SURF [97], BRIEF, LBP [56] and others are some well-known local feature descriptors [98].

***Scale-Invariant Feature Transform***

One of the well-known local descriptors that the Scale-Invariant Feature Transform (SIFT) [95]. The SIFT method transforms the image into a collection of local feature vectors. Being unique is the goal of these feature vectors. Because of its variation in lighting, SIFT can accurately detect objects even in the presence of clutter. SIFT is not affected by rotation or scaling.

However, low lighting and well located key spots in an image are two areas where the SIFT descriptor's performance is poor. For this reason, the SIFT descriptor loses its discriminative efficacy. Another drawback of SIFT is that it requires a lot of memory and has a high computational cost. The local derivative pattern (LDP) as a novel technique for CBIR to build the feature descriptor because of the high dimensionality of SIFT [99]. The BRISK descriptor was suggested in order to enhance CBIR's performance.

A novel method in which the author combined visual words with binary robust invariant scalable key points (BRISK) and SIFT [100]. The drawbacks of poor lighting and incorrect key point localization are lessened by using BRISK.

 

A CBIR system based on SIFT and local intensity order pattern (LIOP) [101]. The SIFT disadvantage of scenes with low contrast and illumination was addressed in this method by using LIOP. In contrast to SIFT, LIOP performs poorly when it comes to scale change. An alternative method that combines the two descriptors to create visual words based on the bag-of-visual-words (BoVW), a vector of defined length that is used to describe local features.

The semantic gap between an image's high-level and low-level features will be lessened by this fusion. Following the fusing of visual words, K-means clustering is used. Each image's visual histogram is calculated using k-means clustering. Support vector machines (SVMs), which are trained using a histogram, are used to classify the images.

***Speeded-Up Robust Features***

A new local feature descriptor, known as Speeded-up Robust Features (SURF), was introduced as an innovative interest point detector that is both scale- and rotation-invariant to lessen the high dimensionality and complexity issue in SIFT [97]. A dimensionality reduction strategy was employed to deactivate high dimensionality, which will impair feature computing performance. Because SURF uses an indexing technique based on the Laplacian sign, it takes less time to compute and match features than SIFT, making it faster and more reliable.

But SURF does not do well in rotation. A new CBIR system based on SURF and rapid retina key point (FREAK) [102]. While SURF is more resilient to changes in scale and illumination, FREAK performs better in classification.

By combining the two descriptors, visual words based on the BoVW technique will be created, minimizing the meaning difference between an image's high-level and low-level properties. Following the fusion process, the fused visual words are subjected to K-mean clustering, and each image's visual word is then subjected to a histogram computation. An SVM is trained using these histograms in order to categorize the photos according to their semantics. There is no color information provided by FREAK or SURF.

***Local Binary Pattern (LBP)***

LBP, which is founded on the qualitative level for local patterns, was first presented by Sarwar [56]. The center pixel serves as a threshold for LBP's comparison of the center pixel and its eight surrounding neighbors. The values of all eight neighboring pixels are chosen, and each is converted to a binary value based on the value of the central pixel.

The final LBP value for the current pixel is then produced by multiplying these binary values by appropriate weights, usually in power of two, and adding them all. The process for calculating LBP characteristics is shown in Figure 4. In contrast to the impact of lighting, LBP offers benefits in terms of faster performance and greater reliability.

Due of its invariance to monotonic grayscale transformations, LBP is resilient. The computation of LBP is straightforward. Though, its loss of global spatial information is a drawback.

The LBP variance (LBPV) was introduced as a way to get beyond the limitations of LBP. Following a local variant LBP, LBPV performs a global rotation invariant [103]. A similarity measurement-based approach was suggested to speed up the suggested matching and decrease feature size. A CBIR system that reduces the semantic gap by utilizing LBPV and LIOP features [104]. Two small visual dictionaries are created using these two feature descriptors, and they are then concatenated to create a single, large visual dictionary. Principal component analysis (PCA) was employed to minimize the size, the histogram was computed, and SVM was trained using LBPV and LIOP [105].

 

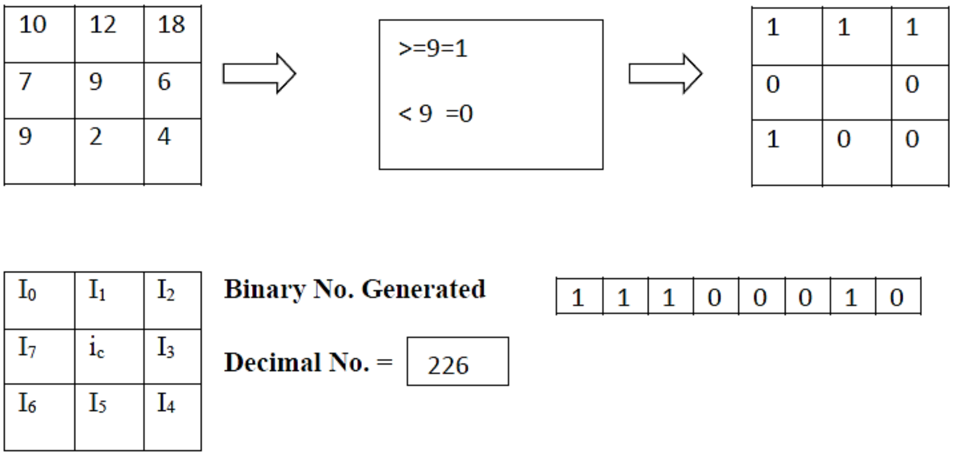


**Figure 4.** The procedure to compute LBP features.

***Histogram of Oriented Gradient (HOG)***

The Histogram of Oriented Gradients (HOG), a unique spatially normalized descriptor, surpasses other popular feature descriptors like wavelets [96]. HOG establishes the shape and appearance of the local object based on the distribution of local intensity gradients or the directions of the object's edge, even in the absence of specific knowledge on matching gradients or edge positions. HOG divides a whole image into discrete spatial units called "cells" or "regions", and then accumulates edge orientations across the pixels of the cells or local 1D histogram of gradient directions. The histogram entries are combined to create the image representation.

Additionally, the HOG normalizes all of the block cells using the energy it gathers, which is a measure of the local histogram over bigger region "blocks". HOG is therefore more resistant to lighting and shadows. HOG has been effectively applied in numerous applications, particularly object identification, for the past 10 years. A CBIR system was proposed that uses HOG to extract global features and SURF to extract local features [106]. HOG is employed because it provides more spatial information, which enhances retrieval efficiency, whereas SURF works better in images with a clear background, low illumination, and noise. One huge vocabulary is created by concatenating the two visual vocabularies created by the two descriptions. Better retrieval outcomes are obtained with a wider vocabulary, but the computational cost goes up. Consequently, a percentage of the retrieved features was taken. After applying K-means++ clustering to the fused feature vector, each image's histogram was determined. For classification, SVM was utilized.

There is no spatial information about nearby pixels provided by HOG. It merely provides the pixel's orientation for analysis. This restriction is addressed by the Co-occurrence Histogram of Oriented Gradient (CoHOG) [107]. By combining the advantages of CoHOG and SURF to address each other's shortcomings, developed a CBIR system to close the semantic gap [108]. To lower the computational cost, CoHOG relies on local discriminant analysis to reduce the dimension of each CoHOG and SURF feature vector. By highlighting the pertinent and irrelevant photos that were found during the database search, relevance feedback helps the user improve specificity (accuracy) and sensitivity (recall).

***Harris Corner Detector***

In 1988, the Harris detector was first presented as a reliable corner detector [109]. It is utilized for camera calibration, picture matching, video stabilization, and tracking and is regarded as a reference technique [110]. Points with significant intensity variation can be found by examining the eigenvalues of the autocorrelation matrix (sometimes referred to as the structure tensor) data. Although the Harris detector is resistant to noise, scale, and rotation, its computational cost is a drawback. Based on the enhanced Harris detector [111] and SURF detector and descriptor, developed a CBIR approach to extract encrypted images from the cloud [112].

 

To increase efficiency and decrease the amount of time needed for searching, developed searchable indexes for feature vectors using the Local Sensitive Hash function.

***Features from Accelerated Segment Test (FAST) [113]***

FAST overcomes the computational cost limitation of Harris. Here, repeatability refers to the capacity to identify an interest point under many transformations. FAST favors corner detection over edge detection since a corner is a point of interest due to its two-dimensional intensity change.

In order to describe characteristics in cloud computing, suggested a CBIR approach based on FAST [114]. Before being stored on the cloud, these attributes are encrypted by stream chipper to protect the privacy of the photographs.

Faster search times are achieved by using the Locality Sensitive Hash algorithm. Additionally, in an effort to stop unauthorized users from retrieving photographs, the cloud server uses a watermark-based protocol to add a unique watermark to retrieved images that are in encrypted form to the query user. Because FAST requires less computing time, the suggested CBIR was able to retrieve equivalent images more efficiently. However, FAST does not work well in environments with high levels of noise. Its reliance on threshold values is another drawback.

## PERFORMANCE EVALUATION

In a CBIR system, key parameters are defined and assessed based on precision, recall, and response time. The precision and recall rates for image retrieval can be expressed using the following formulas:

$$\text{Precision} = \frac{R}{N} \tag{11}$$

$$\text{Recall} = \frac{R}{M} \tag{12}$$

The letter N refers to the total number of images retrieved in the query, while R represents the images linked to the example in the result. Additionally, M denotes the images connected to the example in the test set S (saturation).

According to comparison research, the combined feature has the highest accuracy. According to comparison study, precision and memory were significantly increased by the combination or hybrid features technique, which combines texture, color, and shape as shown in Table 1.

## MACHINE LEARNING

Feature representation and similarity assessment play a vital role in determining the image retrieval effectiveness of a CBIR system. Various works have been done by researchers in the past few years. A variety of techniques have been proposed but even then, it remains one of the most challenging problems in the ongoing CBIR research, and the main reason for it is the semantic gap issue that exists between the low-level image pixels captured by machines and high level semantic concepts perceived by humans. A key challenge in Artificial Intelligence, from a broader viewpoint, is how to develop and train machines that are as intelligent as humans to effectively handle real-world tasks.

Machine Learning is a promising approach that seeks to tackle this challenge over the long term [115]. Many significant developments in machine learning techniques have been made in the last few years. A subfield of machine learning that has lately become well-known is called deep learning. A set of machine learning algorithms known as "deep learning" uses deep architectures made up of several non-linear transformations in an effort to represent high-level abstractions in data.

**Table 1.** Comparison study for color, texture, shape and combined features.

| **Research Paper** | **Feature(s)** | **Feature extraction methods** | **Dataset** | **Number of Sample Images** | **Accuracy** |
|---|---|---|---|---|---|
| Region-based image retrieval with high-level semantics using decision tree learning (Liu *et al.*, 2008) [1] | Color | High-level semantics using decision tree learning | Corel | 500–1000 | 76.8% |
| Automatic categorization of image regions using dominant color based vector quantization (Islam *et al.*, 2008) [8] | Color | Dominant color-based vector quantization | Corel | 500 | 97.67% |
| Image Retrieval Using Both Color and Texture Features (Kong, 2009) [38] | Texture | Gray-level co-occurrence matrix (GLCM), and color co-occurrence matrix (CCM) | Corel | 1000 | 44% |
| Color Feature Extraction for CBIR (Mishra HK, 2011) [36] | Color | Average Mean of RGB components | Corel | 300 and 10000 | 50% with 300 images 30% with 1000 images |
| Fusion of color, shape, and texture features for content-based image retrieval (Wang *et al.*, 2014) [45] | Color | Color Histogram | Corel | 1000 | 80% |
| Information fusion in content based image retrieval: A comprehensive overview. Inf Fusion (Piras and Giacinto, 2017) [116] | Color | Fusion of color and texture features | Corel | 500 | 61.3% |
| An efficient technique for retrieval of color images in large databases. (Shrivastava and Tyagi, 2015) [34] | Color Texture Shape | Color Histogram Gabor Filter Fourier Descriptor | Corel | | 76.9% |
| Content-based image retrieval using PSO and k-means clustering algorithm (Younus *et al.*, 2015) [79] | Color Texture | Color Moment Color Histogram Wavelet Moment Co-occurrence Matrix | WANG | 1000 | 73.5% |
| Content based image retrieval scheme using color, texture and shape features. (Zhao *et al*., 2016) [11] | Color Texture Shape | CDE CLCM Hue Moments | Wang | | 90.5% |
| Content-based image retrieval using color, texture and shape features (Ponomarev *et al*., 2016) [68] | Color Texture Shape | Color auto correlogram Gabor Transform Wavelet Transform | Corel | | 83% |
| Integration of wavelet transform, local binary patterns and moments for content-based image retrieval. (Srivastava and Khare, 2017) [13] | Texture Shape | Wavelet Transform/LBP Legendre Moments | Corel 1k Corel 5k Corel 10k | | 99.9% 56.7% 35.3% |
| An effective method for color image retrieval based on texture (Wang *et al.*, 2012) [117] | Texture | SED | Corel 1K Corel 10K | 1000 10000 | 88.62% |
| Design of Feature Extraction in Content-Based Image Retrieval (CBIR) using Color and Texture [118] | Texture | Wavelet Transform, Tammura Texture | Corel | 1000 | 76.35% |

Journal of Image Processing & Pattern Recognition Progress
Volume 12, Issue 1
ISSN: 2394-1995

| | | | | | |
|---|---|---|---|---|---|
| Content-based image retrieval in DCT compressed domain with MPEG-7 edge descriptor and genetic algorithm (Phadikar *et al.*, 2018) [119] | Color Texture | Color Histogram Color Moment MPEG-7 Edge descriptor | Corel | 1000 | 98.5% |
| An efficient seed points selection approach in dominant color descriptors (DCD). (Pavithra and Sree Sharmila, 2019) [81] | Color | Dominant Color Descriptor | Wang Corel-10 K | 1000 | 75.3% 41.3% |
| Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor, (Nazir *et al.*, 2018) [49] | Color Texture | Color Histogram DWT EDH | Corel | | 73.5% |
| Boosting content based image retrieval performance through integration of parametric & nonparametric approaches. (Rana *et al.*, 2019) [120] | Color Texture Shape | Color Moment Ranklet Transformation Invariant Moment | Corel Corel | 5k 10k | 67.4% 67.9% |
| MDCBIR-MF: Multimedia data for content-based image retrieval by using multiple features. (Ashraf *et al.*, 2020) [86] | Color Texture | Color Moment DWT/Gabor Filter/CEDD | Corel Corel | 5k 1k | 87.5% 79.8% |
| Content-based image retrieval using color, shape and texture descriptors and features. (Alsmadi, 2020) [87] | Color Texture Shape | GLCM DWT Canny Edge Histogram | Corel | | 90.1% |
| An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model. (Jabeen *et al.*, 2018) [97] | SURF FREAK | | Corel | 1k | 86% |
| Scene analysis and search using local features and support vector machine for effective content-based image retrieval. (Sharif *et al.*, 2019) [95] | SIFT BRISK | | Corel Corel | 1k 5k | 84.3% 57.3% |
| Content-based image retrieval based on visual words fusion versus features fusion of local and global features. (Mehmood *et al.*, 2018 [78] | HOG SURF | | Corel Corel | 5k 1k | 80.8% 60.6% |
| Boosting the performance of the BoVW model using SURF–CoHOG-based sparse features with relevance feedback for CBIR. (Baig *et al.*, 2020) [100] | CoHOG SURF | | Corel | 1k | 86.4% |
| Content based image retrieval by using color descriptor and discrete Wavelet transform (Ashraf *et al.*, 2020) [83] | Combined features | Image features information fusion | Corel | 1000 | 90% |

**Deep Learning**

Deep learning is a subset of machine learning that involves the use of multiple layers of processing stages within hierarchical structures to analyze patterns and learn representations or features. It encompasses various research domains, such as neural networks, graphical models, optimization, pattern recognition, and signal processing. The deep supervised backpropagation convolutional network for digit recognition [121]. It has become a valuable research topic in the recent years in both computer vision and machine learning where deep learning achieves state-of-the art results for a variety of tasks. Deep learning algorithm includes convolutional neural network (CNN), deep neural network (DNN), deep belief network, and Boltzmann machine, with CNN exhibiting outstanding performance in computer vision applications such as face recognition, object detection and semantic segmentation [122] and specifically in the CBIR domain [3]. The deep convolutional neural networks (CNNs) introduced by Hinton were initially applied to the image classification challenge in the ImageNet competition, achieving remarkable results.

**Convolutional Neural Network (CNN)**

Convolutional Neural Networks (CNNs) have transformed machine learning, particularly in areas like computer vision and image processing. They excel in tasks such as semantic segmentation, object detection, and image classification. CNNs also minimize the reliance on manual feature extraction, a process that was once both time-intensive and prone to mistakes. Instead, they automatically learn hierarchical feature representations from data.

State-of-the-art performance in visual identification tasks has been made possible by this self-learning ability, which has also made it easier to apply CNNs to a variety of applications, from video analysis and natural language processing to medical diagnostics and autonomous vehicles. CNNs are the perfect choice for the Content-Based Medical Image Retrieval (CBMIR) framework because of their adaptability and effectiveness in handling challenging pattern recognition tasks. This framework provides a potent tool for extracting significant patterns from complex medical pictures [123, 124, 66].

The individual neurons in a convolutional neural network (CNN), a form of feed-forward artificial neural network (CNN), are tiled so that they react to areas of the visual field that overlap [121, 125]. They are invariants of Multi-layer Perceptrons (MLPs), which are made with minimal preprocessing in mind, and are inspired by biology.

The recognition of images and videos makes extensive use of these models. In contrast to other techniques for feature extraction and classification, convolutional neural networks require minimal pre-processing. The network at CNN creates its own filters. Additionally, CNN is unaffected by rotation, scaling, or translation [126]. One of CNN's disadvantages is that it relies on labeled data [126]. CNNs also eliminate the need for manual feature extraction [127].

CNNs consist of three types of layers, convolutional layer, pooling layer and fully connected layer, Figure 5 shows an example of CNN [128].

The model's lowest layers are made up of the max-pooling and alternating convolution layers. The upper layers, on the other hand, are fully connected and match a conventional multi-layer perceptron, which combines logistic regression and hidden layers.

The collection of all features mappings at the layer below is the input to the first fully-connected layer. While the intermediary layers (pooling layer) serve to down sample the volume of incoming inputs, the convolutional layer applies filters on input images in order to learn features. The final layer, also called the fully connected layer, is responsible for making predictions about the class or label of the input image. The final CNN layer is the only one that is entirely connected.

***Convolutional Layer***

By repeatedly applying a function to sub-regions of the entire image, we are able to create a feature map. This is primarily accomplished via convolution of the input image using a linear filter, followed by the addition of a bias term and the application of a non-linear function [125].

The k-th feature map can be denoted as $h_k$ at a given layer, whose filters we can determine by the bias $b_k$ and weights $W^k$, then we can obtain the feature map by the given equation:

$$h^{ij}_{k} = \tanh(w^k \times x)_{ij} + b_k) \tag{13}$$

 

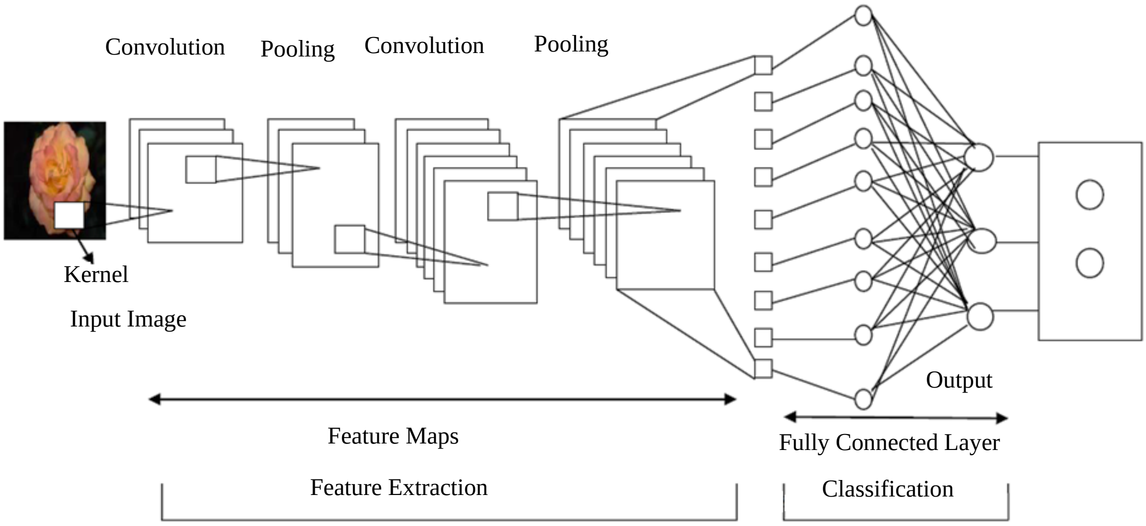


**Figure 5.** Convolutional neural network (CNN) architecture.

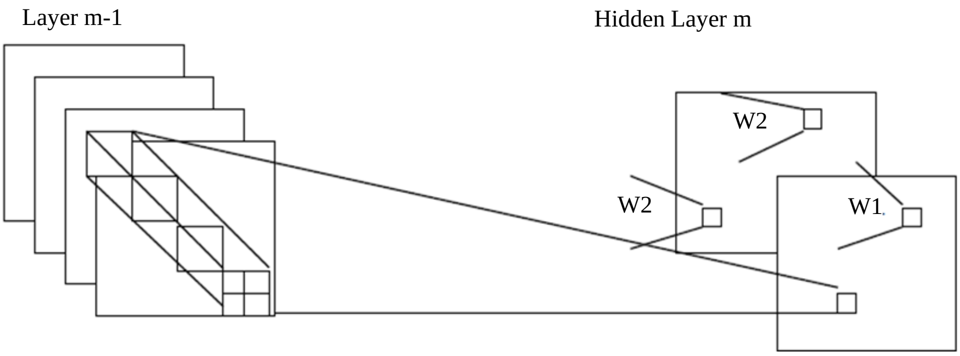


**Figure 6.** Convolutional layer.

The previous layer (m–1) consists of four feature maps, while the hidden layer (m) contains two feature maps, labeled h0 and h1 (Figure 6). The pixels of layer (m–1) that lie within their 2×2 receptive field in the layer below (colored squares) are used for the computation of the pixels in the feature maps h0 and h1 (blue and red squares). It can be observed that how all four input feature maps are spanned by the receptive field. The input feature maps are identified by the primary dimensions, while the pixel coordinates are represented by the remaining two dimensions.

When we combine it all as shown in Figure 6, at layer m the weight that connects each pixel of the $k^{th}$ feature map with the pixel of the $l^{th}$ layer at layer (m–1) and at coordinates (i, j) is denoted [125].

***Max-Pooling***

A key idea in CNNs is max-pooling, which is a type of non-linear down sampling. Each sub-region of the input image is assigned a maximum value once it has been divided into a collection of non-overlapping rectangles. The following justifies the usage of max-pooling in vision: Removing non-maximal values reduces the computation of upper levels. Let us say a convolutional layer is cascaded with a max-pooling layer.

A single pixel can transform the incoming image in eight different ways. If max-pooling is performed across a 2×2 region, three of the eight alternative configurations result in exactly the same output at the convolutional layer. For max-pooling over a 3×3 zone, this rises to 5/8 [125, 115]. This offers a type of translation in variance.

 

In terms of retrieval accuracy, the developed algorithms have demonstrated amazing response, outperforming several modern CBIR techniques. Furthermore, these algorithms stand out for their efficiency and speed in picture retrieval tasks [115]. With the advent and advancement of Deep Learning Neural Networks, CBIR has experienced an improved performance. CNN models are used to extract the deep features of an image more accurately and efficiently.

The utilization of deep models enables the extraction of both higher-level and low-level features from images, effectively reducing the semantic gap, as mentioned [129].

**Support Vector Machine**

The performance of CBIR has been constrained by various factors [5], including the subjective nature of human perception [57], the similarity between visual features, and challenges related to the semantic gap in queries [130–132]. To address these challenges, interactive relevance feedback has been introduced, which facilitates user-system interaction. Relevance feedback is a supervised active learning method where users provide positive and negative feedback to enhance the system's accuracy [133, 134]. Initially, the system retrieves a list of images ranked based on predefined similarity metrics. The user then marks the images as either relevant or irrelevant. The system refines the results based on this feedback and presents an updated list to the user. This process repeats until the user is satisfied with the retrieved images [135].

The performance of the CBIR system will be lowered by the constraints of the typical relevant feedback of CBIR, which include the imbalance of training-set problems, categorization problems, limited information from user problems, and insufficient training set problems.

As a result, an improved relevance-feedback approach that assists user queries by selecting representative images and ranking the retrieved images based on their weights [133]. The support vector machine (SVM) has been adapted for the purpose of supervised learning and is used to support the learning process to reduce the semantic gap between the user and the CBIR system [132, 136, 118].

Support vector machine is widely recognized as one of the most effective techniques for optimizing predicted outcomes [137, 101]. Initially as a kernel-based approach for regression and classification tasks, SVM has gained significant attention in the fields of data mining, pattern recognition, and machine learning [130]. This is largely due to its outstanding generalization capability, ability to find optimal solutions, and strong discriminative power. SVM has demonstrated its effectiveness in addressing real-world binary classification problems. The superiority of SVMs over other supervised learning techniques has been demonstrated [66]. SVMs have become a widely used classification method in recent years due to their solid theoretical foundation and excellent ability to generalize.

Support vector machines establish decision functions by directly learning from the training data, optimizing the existing margin between decision boundaries in a high-dimensional space known as the feature space.

This classification approach reduces the training data's classification mistakes and improves generalization; in other words, SVMs' classification abilities differ greatly from those of other methods, especially when the number of input data is small. SVMs are an effective method for regression analysis and data classification. One key benefit of SVMs is that they identify a small subset of support vectors during the training process, which typically represents only a fraction of the entire dataset.

This collection of support vectors, which is made up of a tiny data set, represents a specific classification problem. SVM is not so popular when the data sets are very large because some SVM implementations demand huge training time or in other cases when the data sets are imbalanced, the accuracy of SVM is poor, we have presented some techniques when the data sets are imbalanced.

 

## SIMILARITY MEASURES

Retrieval performance not only depends on best visual feature extraction, but also on effective selection of similarity measures [26, 129]. The similarity measurement between two images is measured numerically, and determines which images are considered most relevant to the query image and should be returned from the dataset. The similarity metric influences the system's computational complexity and indirectly assesses the CBIR's accuracy. To obtain relevant results, various researchers use different methods to measure similarity.

The similarity measure can be divided into distance measure and similarity metric [138]. Some of the well-known similarity measures are Manhattan, Euclidean [29], Chi-Square, and Mahalanobis distance. The choice of evaluation parameter depends on the chosen coordinate system.

Among these methods, Manhattan and Euclidean distances are the simplest, calculating the distance between points using Cartesian coordinates.

Chi-Square is a statistical approach that measures the distance based on feature metrics. For distances in multivariate space, the Mahalanobis distance is used. Some distance techniques focus on finding the shortest path between two points within a manifold using the shortest path algorithms, for example, the geodesic distance. The Minkowski distance calculates distances in N-Dimensional space.

The most widely used distance function for numerical attributes calculates the shortest distance between two points in an N-dimensional space.

### Euclidean Distance [29]

Euclidean distance which is defined in the following formula:

$$ED\left(M^k, M^t\right) = \sqrt{\sum_{i=1}^{n}(M_i^k - M_i^t)^2} \tag{14}$$

Where, $M^k$ and $M^t$ are image query and image database respectively, $i$ is a feature range. Closer distance represents the higher similarity between images.

### Minkowski Family

Measures similarity between two matrices in N-dimensional space. It is widely used in CBIR as a dissimilarity measure because it is simple in implementation and computation. The mathematical expression is:

$$L_p(X, Y) = \left(\sum_{l=1}^{N} |X_l - Y_l|^p\right)^{\frac{1}{p}} \tag{15}$$

Where X and Y are two vectors in *RN* and *N* is the total dimensionality of the Euclidean Space. Minkowski distance is also named $Lp$-Norm. While $L1$ is known as Manhattan or City block or Taxi-cab distance. It is calculated by making the value of $p$ equal to 1 in Eq. (1). It is variant to coordinate system rotation but robust against reflection and translation. $L1$ which is also named Chessboard or Chebyshev distance is another member in the Minkowski family.

$$L_\infty (X, Y) = \max_{l=1}(|x_l - y_l|) = \lim_{p \to \infty}\left(\sum_{l=1}^{N} |X_l - Y_l|^p\right)^{\frac{1}{p}} \tag{16}$$

On a chessboard which represents a 2-D space, $L1$ is the minimum number of movements that a king needed to move between two squares. The final member in Minkowski family is the fractional distance, which is the result if $p$ is less than one and $p \in (0,1)$. It is the most robust against noise as compared to other members in the Minkowski family.

 

Chi-square distance is another statistical method for measuring similarity between feature metrics. It is widely used for computing the difference between histogram functions [139]. Its mathematical expression is shown below [120]:

$$Chi - square = \sum_{i=1}^{N} \left( \frac{(x_i - y_i)^2}{(x_i + y_i)} \right) \tag{17}$$

Chi-square is used for shape classification, local descriptors matching, boundary detection near duplicate image identification and finally texture and object categories classification.

Histogram intersection distance is another distance metric which is used in image retrieval and other computer vision algorithms such as segmentation, clustering, classification and codebook creation. It is reported to be robust to distraction in object's background, image resolution variation, occlusion and viewpoint variation as proposed and defined as [140]:

$$Histogram\ Intersection\ Distance = \sum_{i=1}^{N} \min(M_i, S_i) \tag{18}$$

Where $S$ and $M$ are two histograms with $N$ bins.

Mahalanobis Distance is used to calculate the separation between points in a multivariate space by accounting for correlations. Mahalanobis distance, which is defined as:

$$Mahalanobis\ Distance = \sqrt{(v-m)^T (C^{-1})(v-m)} \tag{19}$$

Where $T$ represents matrix transpose, $v$ is the feature vector, $m$ is the mean row vector and $C$ is the covariance matrix. The calculation of this distance becomes expansive for high dimensional data because of the computation of the covariance matrix.

Canberra distance defined by Lance and Williams for unsigned numbers [138]. The modified version for signed numbers is proposed, and is defined as [141]:

$$Canberra\ distance = \sum_{i=1}^{N} \left( \frac{|x_i - y_i|}{|x_i + y_i|} \right) \tag{20}$$

where x and y are vectors with real values. This metric is reported to be sensitive to values near zero while it is appropriate when difference in sign indicates difference in classes [141].

Squared Chord is another distance metric which is used in image retrieval and it is unsuitable for negative values feature vectors [142], it is defined as:

$$Squared\ Chord\ distance = \sum_{i=1}^{N} (\sqrt{x_i} - \sqrt{y_i})^2 \tag{21}$$

Unlike the previously described distances, which measure dissimilarity, the cosine distance is used to measure similarity. When assessing similarity, a higher value indicates that the images are more similar to the query image. Cosine distance evaluates the angle between two vectors, as described below [142]:

$$Cosine\ distance = \frac{(X.Y)}{|X|.|Y|} \tag{22}$$

Choosing the right similarity measure is a still a big challenge for researchers. Many researchers made this decision through experiments [143], the authors made a comparison between Canberra, Chi-square, Manhattan and Euclidean distances. They found that using Euclidean distance as a similarity

 

Journal of Image Processing & Pattern Recognition Progress
Volume 12, Issue 1
ISSN: 2394-1995

measure achieved higher precision. While included L1, L2, weighted L1, Chi Square, Square Chord and Extended Canberra in the experiments [144].

They found that using weighted L1 was the best choice to the proposed descriptor because weighted L1 gave better results in terms of computational complexity and accuracy. Although L1 and L2 have lower complexity than Square Chord, Canberra and Chi Square, the retrieval accuracy was also lower because both are noise sensitive and did not consider the neighbor bins.

Other researchers believed that features should not be equally treated when measuring the similarity between query image and dataset images. Assigned weights to color, texture and shape features based on their experience. Choosing the appropriate similarity measure is still an open research area which needs more exploring [11].

## IMAGE DATABASES

This section introduces several popular image databases that are commonly used in CBIR algorithms [117, 145]. Various nature scenes, faces, animals, buildings, people, retinal, and many other types of images are included in the databases, along with color, grayscale, texture, form, segmentation, and medical images. These images will assist in enhancing various aspects of CBIR to optimize the retrieval process as shown in Figure 7.

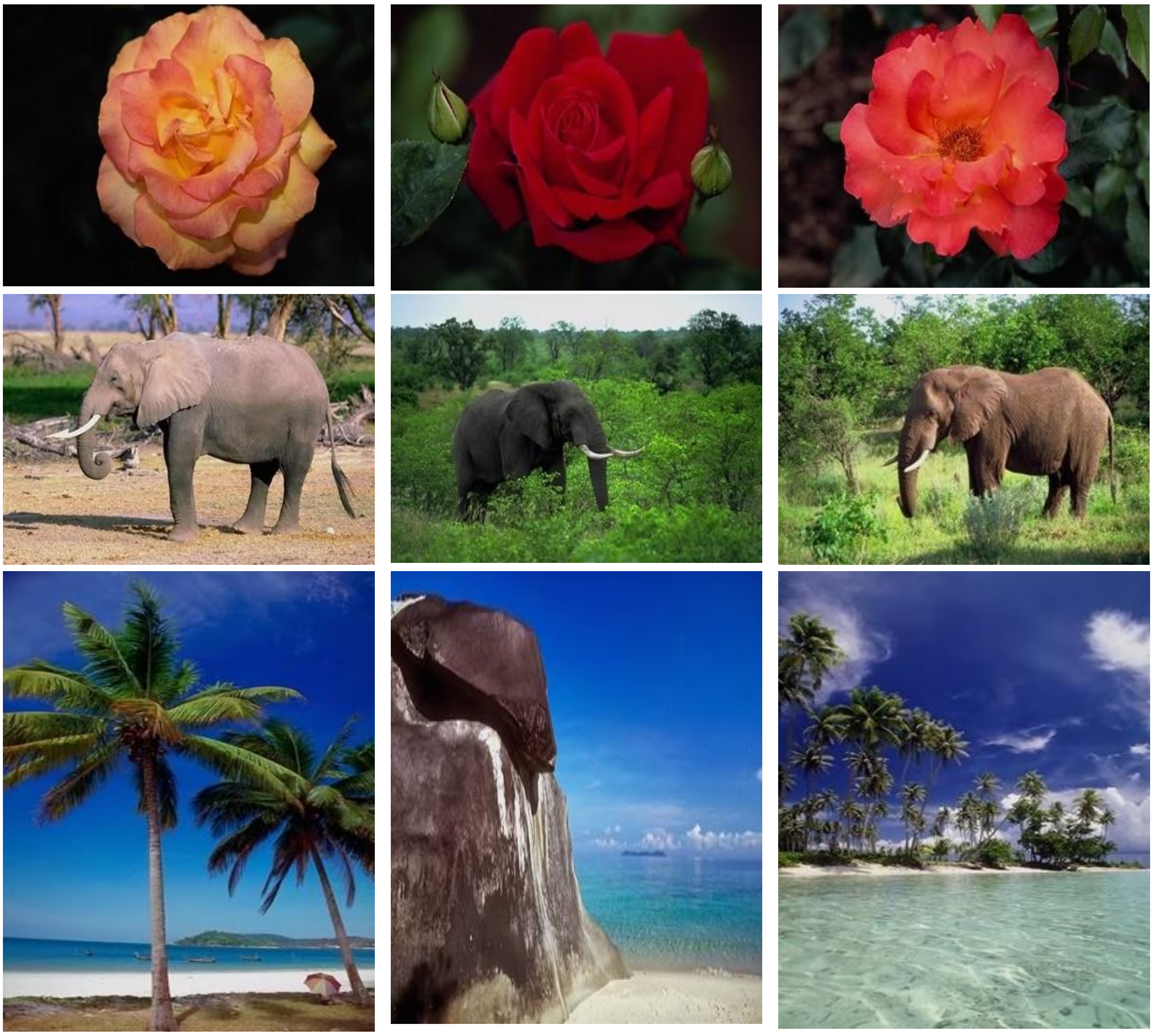

**Figure 7.** Example image datasets of corel and wang.

**COREL Database**

This database, developed by James J. Wang and Jai Li from the University of Pennsylvania State University and Stanford, contains 10,000 test images, along with a subset of 1,000 images in various sizes, all in .jpg format. There are three categories in the Corel database: Corel-1K, Corel-5K, and Corel-10K. For example, Corel-1K includes 10 categories with 100 images each, Corel-5K has 50 categories with 100 images each, and so on. Each image is labeled by 1–5 keywords. The Corel 5K dataset is typically split into three sections: 4000 images for training, 500 images for validation, and 499 images for testing. In total, there are 4500 images for training and 499 for validation.

**Coil-100Database**

This database comprises of comprises of 7200 images. The images in Coil-100 dataset are taken by the Computer Science Department in the Centre of Research on Intelligent Systems. The entire database consists of 100 objects, each having 72 different poses. The images of the 100 objects are taken at the 5-degrees pose internals.

**WANG Database**

This database was created by James G. Wang, Jaili, and Giowiderhold at the University of Pennsylvania State University, Stanford, and includes 1000 multiclass photos in.jpg format organized into 10 groups of 100 images each.

**Holidays Database**

The dataset contains a wide range of scene types, including natural, man-made, water, and fire effects, and the images are of high resolution. It comprises 1491 images, organized into 500 groups, with each group representing a unique scene. The query image is the first image in each group, and the other images in the group are the ones that were successfully retrieved.

**Oxford5k**

This dataset includes 5062 images, 15.886 million descriptors, and 55 queries. The images depict buildings from Oxford and are all in an "upright" orientation, as they are intended for display on the web.

**Flickr60k and Flickr1M**

Two distinct collections of random photos are created using the Flickr60k dataset, which is used to learn the quantization centroids and the HE parameters (median values). The Flickr60k set consists of 67,714 images and 140 million descriptors, while the Flickr1M set contains 1 million images and 2.072 billion descriptors.

For these tasks, we utilized 5 million and 140 million descriptors, respectively. The Flickr1M dataset serves as a collection of distractor images for large-scale image search. In comparison to the Holidays dataset, Flickr contains some biases, such as a higher number of low-resolution images and a greater variety of human photos.

**CONCLUSION**

In this review work we have provided a thorough overview of CBIR. We provided a quick overview of CBIR and its many low-level characteristics, including color, texture, and shape, in the first part. Additional local characteristics, such as SIFT, SURF, and LBP, are provided. Owing to certain drawbacks with these features when used alone, combination features were added to produce more precise and effective results. Despite a number of approaches, the semantic gap problem, which arises between the low-level image pixels that machines capture and the high-level semantic concepts that humans perceive, remains one of the most difficult issues in the ongoing CBIR research.

 

The adoption of different machine learning algorithms is the solution to this issue. The usage of CNN, SVM, and other machine learning algorithms for classification in the CBIR system has been surveyed in this work.

**The Challenges and Future Research Directions**

There are still a lot of obstacles to overcome with the rise of big data and the application of deep learning methods. Over the last 10 years, CBIR has advanced significantly in theory, technology, and application. These difficulties provide some possible avenues for future CBIR study, like increasing the size of datasets, developing efficient learning techniques for small-scale image retrieval, improving database search efficiency, and implementing automated machine learning and neural architecture search.

The most advanced deep neural network architectures available today are artificially created, which has been a drawback given the speed at which machines are developing their computational capabilities and the gap in human understanding. Recent breakthroughs in deep learning neural networks have been achieved to address this challenge. Automated Machine Learning (AutoML) and Neural Architecture Search (NAS) are two such models.

**REFERENCES**

1. Liu Y, Zhang D, Lu G. Region-based image retrieval with high-level semantics using decision tree learning. Pattern Recognit. 2008 Aug 1; 41(8): 2554–70.
2. Jenni K, Mandala S, Sunar MS. Content based image retrieval using colour strings comparison. Procedia Comput Sci. 2015 Jan 1; 50: 374–9.
3. Latif A, Rasheed A, Sajid U, Ahmed J, Ali N, Ratyal NI, Zafar B, Dar SH, Sajid M, Khalil T. Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review. Math Probl Eng. 2019; 2019(1): 9658350.
4. Shrivastava N, Tyagi V. Corrigendum to "Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching". [Information Sciences. 2014; 259: 212–224]. Inf Sci. 2017 Dec 1; 421: 273.
5. Datta R, Joshi D, Li J, Wang JZ. Image retrieval: Ideas, influences, and trends of the new age. ACM Comput Surv. 2008 May 8; 40(2): 1–60.
6. Halawani A, Teynor A, Setia L, Brunner G, Burkhardt H. Fundamentals and Applications of Image Retrieval: An Overview. Datenbank-Spektrum. 2006 Aug; 18(14–23): 6.
7. Bansal M, Kumar M, Kumar M. 2D object recognition techniques: state-of-the-art work. Arch Comput Methods Eng. 2021 May; 28(3): 1147–61.
8. Islam MM, Zhang D, Lu G. Automatic categorization of image regions using dominant color based vector quantization. In 2008 IEEE Digital Image Computing: Techniques and Applications. 2008 Dec 1; 191–198.
9. Markowska-Kaczmar U, Kwaśnicka H. Deep learning—A new era in bridging the semantic gap. In: Bridging the Semantic Gap in Image and Video Analysis. Cham: Springer; 2018; 123–59.
10. Jégou H, Perronnin F, Douze M, Sánchez J, Pérez P, Schmid C. Aggregating local image descriptors into compact codes. IEEE Trans Pattern Anal Mach Intell. 2011 Dec 13; 34(9): 1704–16.
11. Zhao M, Zhang H, Sun J. A novel image retrieval method based on multi-trend structure descriptor. J Vis Commun Image Represent. 2016 Jul 1; 38: 73–81.
12. Jain M, Jégou H, Gros P. Asymmetric hamming embedding: taking the best of our bits for large scale image search. In proceedings of the 19th ACM international conference on multimedia. 2011 Nov 28; 1441–1444.
13. Srivastava P, Khare A. Integration of wavelet transform, local binary patterns and moments for content-based image retrieval. J Vis Commun Image Represent. 2017 Jan 1; 42: 78–103.
14. Das R, Thepade S, Bhattacharya S, Ghosh S. Retrieval architecture with classified query for content based image recognition. Appl Comput Intell Soft Comput. 2016; 2016(1): 1861247.
15. Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. Int J Comput Vis. 2001 May; 42: 145–75.

16. Yang HF, Lin K, Chen CS. Supervised learning of semantics-preserving hash via deep convolutional neural networks. IEEE Trans Pattern Anal Mach Intell. 2017 Feb 9; 40(2): 437–51.
17. Fei-Fei L, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. In 2004 IEEE conference on computer vision and pattern recognition workshop. 2004 Jun 27; 178–178.
18. Griffin G, Holub A, Perona P. Caltech-256 object category dataset. Pasadena: Technical Report 7694, California Institute of Technology; 2007 Apr.
19. Ashraf R, Ahmed M, Ahmad U, Habib MA, Jabbar S, Naseer K. MDCBIR-MF: multimedia data for content-based image retrieval by using multiple features. Multimed Tools Appl. 2020 Apr; 79(13): 8553–79.
20. Wang JZ, Li J, Wiederholdy G. Simplicity: Semantics-sensitive integrated matching for picture libraries. In International Conference on Advances in Visual Information Systems. Berlin, Heidelberg: Springer; 2000 Nov 2; 360–371.
21. Alzu'bi A, Amira A, Ramzan N. Content-based image retrieval with compact deep convolutional features. Neurocomputing. 2017 Aug 2; 249: 95–105.
22. Tarawneh AS, Celik C, Hassanat AB, Chetverikov D. Detailed investigation of deep features with sparse representation and dimensionality reduction in CBIR: A comparative study. Intell Data Anal. 2020 Jan 1; 24(1): 47–68.
23. Latif A, Rasheed A, Sajid U, Ahmed J, Ali N, Ratyal NI, Zafar B, Dar SH, Sajid M, Khalil T. Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review. Math Probl Eng. 2019; 2019(1): 9658350.
24. Kumar N, Mishra B, Bali V. A novel approach for blast-induced fly rock prediction based on particle swarm optimization and artificial neural network. In Proceedings of International Conference on Recent Advancement on Computer and Communication: ICRAC 2017. Singapore: Springer; 2018; 19–27.
25. Tian D. Support Vector Machine for Content-based Image Retrieval: A Comprehensive Overview. J Inf Hiding Multim Signal Process. 2018 Nov; 9(6): 1464–78.
26. Bunte K, Biehl M, Jonkman MF, Petkov N. Learning effective color features for content based image retrieval in dermatology. Pattern Recognit. 2011 Sep 1; 44(9): 1892–902.
27. Kumari B, Kumar R, Singh VK, Pawar L, Pandey P, Sharma M. An efficient system for color image retrieval representing semantic information to enhance performance by optimizing feature extraction. Procedia Comput Sci. 2019 Jan 1; 152: 102–10.
28. Kumar S, Pal AK, Varish N, Nurhidayat I, Eldin SM, Sahoo SK. A hierarchical approach based CBIR scheme using shape, texture, and color for accelerating retrieval process. J King Saud Univ-Comput Inf Sci. 2023 Jul 1; 35(7): 101609.
29. Iqbal K, Odetayo MO, James A. Content-based image retrieval approach for biometric security using colour, texture and shape features controlled by fuzzy heuristics. J Comput Syst Sci. 2012 Jul 1; 78(4): 1258–77.
30. Qiu G. Color image indexing using BTC. IEEE Trans Image Process. 2003 Jan; 12(1): 93–101.
31. Flickner M, Sawhney H, Niblack W, Ashley J, Huang Q, Dom B, Gorkani M, Hafner J, Lee D, Petkovic D, Steele D. Query by image and video content: The QBIC system. Computer. 1995 Sep; 28(9): 23–32.
32. Li C, Huang Y, Zhu L. Color texture image retrieval based on Gaussian copula models of Gabor wavelets. Pattern Recognit. 2017 Apr 1; 64: 118–29.
33. Duanmu X. Image retrieval using color moment invariant. In 2010 IEEE 7th international conference on information technology: new generations. 2010 Apr 12; 200–203.
34. Shrivastava N, Tyagi V. An efficient technique for retrieval of color images in large databases. Comput Electr Eng. 2015 Aug 1; 46: 314–27.
35. Charles YR, Ramraj R. A novel local mesh color texture pattern for image retrieval system. AEU-Int J Electron Commun. 2016 Mar 1; 70(3): 225–33.
36. Mishra HK, Narula MS, Shah MV. Color Feature Extraction for CBIR. Int J Eng Sci Technol. 2011 Dec; 3(12): 8357–8365.

 

37. Singh C, Kaur KP. A fast and efficient image retrieval system based on color and texture features. J Vis Commun Image Represent. 2016 Nov 1; 41: 225–38.
38. Kong FH. Image retrieval using both color and texture features. In 2009 IEEE International Conference on Machine Learning and Cybernetics. 2009 Jul 12; 4: 2228–2232.
39. Liu GH, Yang JY. Content-based image retrieval using color difference histogram. Pattern Recognit. 2013 Jan 1; 46(1): 188–98.
40. Talib A, Mahmuddin M, Husni H, George LE. A weighted dominant color descriptor for content-based image retrieval. J Vis Commun Image Represent. 2013 Apr 1; 24(3): 345–60.
41. Bella MI, Vasuki A. An efficient image retrieval framework using fused information feature. Comput Electr Eng. 2019 May 1; 75: 46–60.
42. Kavitha C, Rao BP, Govardhan A. Image retrieval based on color and texture features of the image sub-blocks. Int J Comput Appl. 2011 Feb 28; 15(7): 33–7.
43. Do MN, Vetterli M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. IEEE Trans Image Process. 2002 Feb; 11(2): 146–58.
44. Sotoodeh M, Moosavi MR, Boostani R. A novel adaptive LBP-based descriptor for color image retrieval. Expert Syst Appl. 2019 Aug 1; 127: 342–52.
45. Wang XY, Zhang BB, Yang HY. Content-based image retrieval by integrating color and texture features. Multimed Tools Appl. 2014 Feb; 68(3): 545–69.
46. Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. IEEE Trans Syst Man Cybern. 1978 Jun; 8(6): 460–73.
47. Wu J, Wei Z, Chang Y. Color and texture feature for content based image retrieval. Int *J Digit* Content Technol Appl. 2010; 4(3): 43–49.
48. Ashraf R, Ahmed M, Jabbar S, Khalid S, Ahmad A, Din S, Jeon G. Content based image retrieval by using color descriptor and discrete wavelet transform. J Med Syst. 2018 Mar; 42(3): 44.
49. Nazir A, Ashraf R, Hamdani T, Ali N. Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor. In 2018 international conference on computing, mathematics and engineering technologies (iCoMET). 2018 Mar 3; 1–6.
50. Singh N, Singh K, Sinha AK. A novel approach for content based image retrieval. Procedia Technol. 2012 Jan 1; 4: 245–50.
51. Murala S, Gonde AB, Maheshwari RP. Color and texture features for image indexing and retrieval. In 2009 IEEE International Advance Computing Conference. 2009 Mar 6; 1411–1416.
52. Zhang D, Wong A, Indrawan M, Lu G. Content-based image retrieval using Gabor texture features. The First IEEE Pacific-Rim Conference on Multimedia (2000 International symposium on Multimedia Information Processing) . 2000; 392-395.
53. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. 1996 Jan 1; 29(1): 51–9.
54. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell. 2002 Jul; 24(7): 971–87.
55. Guo Z, Zhang L, Zhang D. Rotation invariant texture classification using LBP variance (LBPV) with global matching. Pattern Recognit. 2010 Mar 1; 43(3): 706–19.
56. Sarwar A, Mehmood Z, Saba T, Qazi KA, Adnan A, Jamal H. A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine. J Inf Sci. 2019 Feb; 45(1): 117–35.
57. Penatti OAB, Valle E, Torres RS. Comparative study of global color and texture descriptors for web image retrieval. J Vis Commun Image Represent. 2012;23(2):359–80. doi: 10.1016/j.jvcir. 2011.11.002.
58. Gai S, Yang G, Zhang S. Multiscale texture classification using reduced quaternion wavelet transform. AEU-Int J Electron Commun. 2013 Mar 1; 67(3): 233–41.
59. Zhang D, Lu G. Review of shape representation and description techniques. Pattern Recognit. 2004 Jan 1; 37(1): 1–9.
60. Ping Tian D. A review on image feature extraction and representation techniques. Int J Multimed Ubiquitous Eng. 2013 Jul; 8(4): 385–96.

 

61. Srivastava P, Binh NT, Khare A. Content-based image retrieval using moments of local ternary pattern. Mob Netw Appl. 2014 Oct; 19: 618–25.
62. Murala S, Maheshwari RP, Balasubramanian R. Local tetra patterns: a new feature descriptor for content-based image retrieval. IEEE Trans Image Process. 2012 Apr 3; 21(5): 2874–86.
63. Zhang B, Gao Y, Zhao S, Liu J. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. IEEE Trans Image Process. 2009 Nov 3; 19(2): 533–44.
64. Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process. 2010 Feb 17; 19(6): 1635–50.
65. Liu M, Yang L, Liang Y. A chroma texture-based method in color image retrieval. Optik. 2015 Oct 1; 126(20): 2629–33.
66. Lee SH, Choi JY, Plataniotis KN, Ro YM. Local color vector binary pattern for face recognition. In 2011 18th IEEE International Conference on Image Processing. 2011 Sep 11; 2997–3000.
67. Marxsen S, Hilbert D, Laubenbacher R, Sturmfels B, David H, Laubenbacher RC. Theory of algebraic invariants. Cambridge: Cambridge University Press; 1993.
68. Ponomarev A, Nalamwar HS, Babakov I, Parkhi CS, Buddhawar G. Content-based image retrieval using color, texture and shape features. Key Eng Mater. 2016 Mar 18; 685: 872–6.
69. Estellers V, Zosso D, Bresson X, Thiran JP. Harmonic active contours. IEEE Trans Image Process. 2014;23(1):69–82. doi: 10.1109/TIP.2013.2286326. PubMed: 24144662.
70. Shakeel N, Shakeel S. Context-free word importance scores for attacking neural networks. J Comput Cogn Eng. 2022 Sep 27; 1(4): 187–92.
71. Chen R, Pu D, Tong Y, Wu M. Image-denoising algorithm based on improved K-singular value decomposition and atom optimization. CAAI Trans Intell Technol. 2022 Mar; 7(1): 117–27.
72. Meng J, Li Y, Liang H, Ma Y. Single-image dehazing based on two-stream convolutional neural network. Journal of Artificial Intelligence and Technology. 2022 Jun 30; 2(3): 100–10.
73. Srivastava D, Singh SS, Rajitha B, Verma M, Kaur M, Lee HN. Content-based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation and Evaluation Parameters. IEEE Access. 2023 Aug 25; 11: 95410–95431.
74. Shahid AR, Khan S, Yan H. Contour and region harmonic features for sub-local facial expression recognition. J Vis Commun Image Represent. 2020 Nov 1; 73: 102949.
75. Zhang Q, Xiao J, Tian C, Chun-Wei Lin J, Zhang S. A robust deformed convolutional neural network (CNN) for image denoising. CAAI Trans Intell Technol. 2023; 8(2): 331–342.
76. Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR’06), New York, USA. 2006; 2: 2169–2178.
77. Naeem M, Ashraf R, Ali N, Ahmad M, Habib MA. Bottom up approach for better requirements elicitation. In Proceedings of the International Conference on Future Networks and Distributed Systems - ICFNDS’17, New York, USA. 2017; 1–4.
78. Mehmood Z, Abbas F, Mahmood T, Javid MA, Rehman A, Nawaz T. Content-based image retrieval based on visual words fusion versus features fusion of local and global features. Arab J Sci Eng. 2018 Dec; 43(12): 7265–84.
79. Younus ZS, Mohamad D, Saba T, Alkawaz HM, Rehman A, Al-Rodhaan M, Al-Dhelaan A. Content-based image retrieval using PSO and k-means clustering algorithm. Arab J Geosci. 2015 Aug; 8(8): 6211–6224. doi: 10.1007/s12517- 014-1584-7
80. Pavithra LK, Sharmila TS. An efficient framework for image retrieval using color, texture and edge features. Comput Electr Eng. 2018 Aug; 70: 580–593. https://doi.org/10.1016/j. compeleceng.2017. 08.030
81. Pavithra LK, Sree Sharmila T. An efficient seed points selection approach in dominant color descriptors (DCD). Cluster Comput. 2019 Dec; 22(4): 1225–1240. https://doi.org/10.1007/s10586-019- 02907-3
82. Lin CH, Huang DC, Chan YK, Chen KH, Chang YJ. Fast color-spatial feature based image retrieval methods. Expert Syst Appl. 2011 Sep 1; 38(9): 11412–20.

 

83. Ashraf R, Ahmedm M, Jabbar S, Khalid S, Ahmad A, Din S, Jeon G. Content based image retrieval by using color descriptor and discrete Wavelet transform. J Med Syst. 2018 Mar; 42(3): 44. https://doi.org/10.1007/s10916-017-0880-7
84. Jacobs CE, Finkelstein A, Salesin DH. Fast multiresolution image querying. In Proceedings of the 22nd annual conference on Computer graphics and interactive techniques - SIGGRAPH '95, New York, USA. 1995; 277–286.
85. Tadi Bani N, Fekri-Ershad S. Content-based image retrieval based on combination of texture and colour information extracted in spatial and frequency domains. Electronic Library. 2019 Aug; 37(4): 650–666. https://doi.org/10.1108/EL-03-2019-0067
86. Ashraf R, Ahmed M, Ahmad U, Habib MA, Jabbar S, Naseer K. MDCBIR-MF: Multimedia data for content-based image retrieval by using multiple features. Multimed Tools Appl. 2020 Apr; 79(13–14): 8553–8579. https://doi.org/10.1007/ s11042-018-5961-1
87. Alsmadi MK. Content-based image retrieval using color, shape and texture descriptors and features. Arab J Sci Eng. 2020 Apr; 45(4): 3317–3330. https://doi.org/10.1007/s13369- 020-04384-y
88. Sadegh Etemad, Maryam Amirmazlaghani. Color texture image retrieval based on Copula multivariate modeling in the Shearlet domain. Eng Appl Artif Intell. 2021; 102: 104256. https://doi.org/10.1016/j.engappai.2021.104256.
89. Zakariae Abbad, Ahmed Drissi El Maliani, Said Ouatik El Alaoui, Mohammed El Hassouni, Mohamed Tahar Kadaoui Abbassi. Exploring multivariate generalized gamma manifold for color texture retrieval. Pattern Recognit. 2023; 143: 109748. https://doi.org/10.1016/j.patcog.2023. 109748.
90. Sumit Kumar, Arup Kumar Pal, Naushad Varish, Irfan Nurhidayat, Eldin Sayed M, Soubhagya Kumar Sahoo. A hierarchical approach based CBIR scheme using shape, texture, and color for accelerating retrieval process. J King Saud Univ Comput Inf Sci. 2023; 35(7): 101609. https://doi.org/10.1016/j.jksuci.2023.101609.
91. Low DG. Distinctive image features from scale-invariant keypoints. Int J Comput Vis. 2004; 60(2): 91–110. https://doi.org/10.1023/B:VISI.0000029664. 99615.94
92. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005; 1: 886–893.
93. Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-Up Robust Features (SURF). Comput Vis Image Underst. 2008 Jun; 110(3): 346–359. https:// doi.org/10.1016/j.cviu.
94. Montazer GA, Giveki D. Content based image retrieval system using clustered scale invariant feature transforms. Optik (Stuttg). 2015 Sep; 126(18): 1695–1699. https://doi.org/10.1016/j.ijleo. 2015.05. 002
95. Sharif U, Mehmood Z, Mahmood T, Javid MA, Rehman A, Saba T. Scene analysis and search using local features and support vector machine for effective content-based image retrieval. Artif Intell Rev. 2019 Aug; 52(2): 901–925. https:// doi.org/10.1007/s10462-018-9636-0.
96. Yousuf M, Mehmood Z, Habib HA, Mahmood T, Saba T, Rehman A, Rashid M. A novel technique based on visual words fusion analysis of sparse features for effective content-based image retrieval. Math Probl Eng. 2018; 2018: 2134395(13p). https://doi.org/10.1155/2018/2134395
97. Jabeen S, Mehmood Z, Mahmood T, Saba T, Rehman A, Mahmood MT, Rubin DL. An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model. PLoS One. 2018 Apr; 13(4): e0194526. https://doi.org/10.1371/journal. pone. 0194526
98. Capitanio JP, Del Rosso LA, Calonder LA, Blozis SA, Penedo MC. Behavioral effects of prenatal ketamine exposure in rhesus macaques are dependent on MAOA genotype. Exp Clin Psychopharmacol. 2012 Jun; 20(3): 173–180.
99. Watanabe T, Ito S, Yokoi K. Co-occurrence histograms of oriented gradients for human detection. IPSJ Trans Comput Vis Appl. 2010; 2: 39–47. https://doi.org/10.2197/ ipsjtcva.2.39

 

100. Baig F, Mehmood Z, Rashid M, Javid MA, Rehman A, Saba T, Adnan A. Boosting the performance of the BoVW model using SURF–CoHOG-based sparse features with relevance feedback for CBIR. Iranian Journal of Science and Technology, Transactions of Electrical Engineering. 2020 Mar; 44(1): 99–118. https://doi.org/10.1007/s40998-019-00237-z
101. Harris C, Stephens M. A combined corner and edge detector. In Proceedings of the Alvey Vision Conference 1988, Manchester, UK. 1988; 23.1–23.6.
102. Sánchez J, Monzón N, Salgad A. An analysis and implementation of the harris corner detector. Image Process Line. 2018 Oct; 8(9): 305–328. DOI: 10.5201/ipol.2018.229
103. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). 2005 Jun 20; 1: 886–893.
104. Sotoodeh M, Moosavi MR, Boostani R. A novel adaptive LBP-based descriptor for color image retrieval. Expert Syst Appl. 2019 Aug 1; 127: 342–52.
105. Wu J, Wei Z, Chang Y. Color and texture feature for content based image retrieval. Int J Digit Content Technol Appl. 2010; 4(3): 43–49.
106. Besiris D, Zigouris E. Dictionary-based color image retrieval using multiset theory. J Vis Commun Image Represent. 2013 Oct 1; 24(7): 1155–67.
107. Li H, Qin J, Xiang X, Pan L, Ma W, Xiong NN. An efficient image matching algorithm based on adaptive threshold and RANSAC. IEEE Access. 2018; 6: 66963–66971. https://doi.org/10.1109/ACCESS.2018. 2878147
108. Qin J, Li H, Xiang X, Tan Y, Pan W, Ma W, Xiong NN. An encrypted image retrieval method based on harris corner optimization and LSH in cloud computing. IEEE Access. 2019; 7: 24626–24633. https://doi.org/10.1109/ACCESS.2019.2894673
109. Rosten E, Porter R, Drummond T. Faster and better: A machine learning approach to corner detection. IEEE Trans Pattern Anal Mach Intell. 2010 Jan; 32(1): 105–119. Hannover, Germany. https://doi.org/10.1109/ TPAMI.2008.275
110. Sharmi N, Mohamed Shameem P, Parvathy R. Content-based image retrieval using fast machine learning approach in cloud computing. In Communications in computer and information science. Singapore: Springer; 2018; 837: 434–444.
111. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. Comput Intell Neurosci. 2018; 2018: 7068349(13p). DOI: 10.1155/2018/ 7068349
112. Zechao Hu, Bors Adrian G. Co-attention enabled content-based image retrieval. Neural Networks. 2023; 164: 245–263. https://doi.org/10.1016/j.neunet.2023.04.009.
113. Patil A, Rane M. Convolutional neural networks: an overview and its applications in pattern recognition. In Smart Innovation, Systems and Technologies, 195(Insights into Imaging). 2021; 21–30. DOI: 10.1007/978-981-15-7078-0_3
114. Chaudhry S, Chandra R. Unconstrained face detection from a mobile source using convolutional neural networks. Lecture Notes in Computer Science. Including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics. Cham: Springer; 2016 Oct; 9948: 567–576. https://link.springer. com/chapter/10.1007%2F978-3-319-46672-9_63
115. Giriraj Gautam, Anita Khanna. Content Based Image Retrieval System Using CNN based Deep Learning Models. Procedia comput sci. 2024; 235: 3131–3141.
116. Piras L, Giacinto G. Information fusion in content based image retrieval: A comprehensive overview. Inf Fusion. 2017 Sep 1; 37: 50–60.
117. Wang XY, Chen ZF, Yun JJ. An effective method for color image retrieval based on texture. Comput Stand Interfaces. 2012 Jan 1; 34(1): 31–5.
118. Sakhare SV, Nasre VG. Design of feature extraction in content-based image retrieval (CBIR) using color and texture. Int J Comput Sci Inform. 2012;1(2):157–61. doi: 10.47893/IJCSI.2012. 1029.
119. Phadikar BS, Phadikar A, Maity GK. Content-based image retrieval in DCT compressed domain with MPEG-7 edge descriptor and genetic algorithm. Pattern Anal Appl. 2018 May; 21: 469–89.

 

120. Rana SP, Dey M, Siarry P. Boosting content based image retrieval performance through integration of parametric & nonparametric approaches. J Vis Commun Image Represent. 2019 Jan; 58(3): 205–219. https://doi. org/10.1016/j.jvcir.2018.11.015
121. Ouhda Mohamed, El Asnaoui Khalid, Ouanan Mohammed, Aksasse Brahim. Content-Based Image Retrieval Using Convolutional Neural Networks. Advances in Intelligent Systems and Computing, vol 756 (AISC 756). Cham: Springer; 2017; 463–476. https://doi.org/10.1007/978-3-319-91337-7_41.
122. Pele O, Werman M. The quadratic-chi histogram distance family. Lecture Notes in Computer Science. Including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics. Berlin, Heidelberg: Springer; 2010; 6312(2): 749–762. DOI: 10.1007/978-3-642-15552-9_54
123. Elwakil AS. Fractional-order circuits and systems: An emerging interdisciplinary research area. IEEE Circuits Syst Mag. 2010;10(4):40–50. doi: 10.1109/MCAS.2010.938637.
124. Ivrlač MT, Nossek JA. Toward a circuit theory of communication. IEEE Trans Circuits Syst I. 2010;57(7):1663–83. doi: 10.1109/TCSI.2010.2043994.
125. Şaban Öztürk. Convolutional neural network based dictionary learning to create hash codes for content-based image retrieval. Procedia comput sci. 2021; 183: 624–629.
126. Sergyan S. Color histogram features based image classification in content-based image retrieval systems. In 2008 6th International Symposium on Applied Machine Intelligence and Informatics, Herlany, Slovakia. 2008; 221–224.
127. Cha S-H. Comprehensive Survey on Distance Similarity measures between probability density functions. International Journal of Mathematics Model Methods Applied Sciences. 2007; 1(4): 300–307. https://www. naun.org/main/NAUN/ijmmas/mmmas-49.pdf
128. Swain MJ, Ballard DH. Color indexing. Int J Comput Vis. 1991; 7(1): 11–32. https://doi.org/10.1007/BF00130487
129. 129Manjunath BS, Ohm JR, Vasudevan VV, Yamada A. Color and texture descriptors. IEEE Trans Circuits Syst Video Technol. 2001 Jun; 11(6): 703–15.
130. Tao DC, Tang XO, Li XL, Wu XD. Asymmetric bagging and random space for support vector machines-based relevance feedback in image retrieval. IEEE Trans Pattern Anal Mach Intell. 2006 Jul; 28(7): 1088–1099.
131. Qin T, Zhang XD, Liu TY, Wang DS, Ma WY, Zhang HJ. An active Feedback framework for image retrieval. Pattern Recognit Lett. 2008; 29(5): 637–6461.
132. Crucianu M, Ferecatu M, Boujemaa N. Relevance feedback for image retrieval: A short survey. Report of the DELOS2 European Network of Excellence, FP6, 1–21.
133. Hasler P, Lande TS. Overview of floating-gate devices, circuits, and systems. IEEE Trans Circuits Syst II Analog Digital Signal Process. 2001;48(1):1–3.
134. Chang JS, Facchetti AF, Reuss R. A circuits and systems perspective of organic/printed electronics: Review, challenges, and contemporary and emerging design approaches. IEEE J Emerg Sel Top Circuits Syst. 2017;7(1):7–26. doi: 10.1109/JETCAS.2017.2673863.
135. Reddy AH, Chandra NS. Local oppugnant color space extrema patterns for content based natural and texture image retrieval. AEU-Int J Electron Commun. 2015 Jan 1; 69(1): 290–8.
136. Rui Y, Huang TS, Ortega M, Methrotra S. Relevance feedback: A power tool in Interactive content-based image retrieval. IEEE Trans Circuits Syst Video Technol. 1998; 8(5): 644–655.
137. Palepu Pavani, Sashi Prabha T. (Anil Neerukonda Institute of Technology and Sciences). Content Based Image Retrieval Using Machine Learning Approach. Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013. Advances in Intelligent Systems and Computing, vol 247. Cham: Springer; 2014; 173–179.
138. Lance GN, Williams WT. Mixed-data classificatory programs i-agglomerative systems. Aust Comput J. 1967; 1(1): 15–20. https:// www.semanticscholar.org/paper/Mixed-Data-Classificatory-Programs-I-Agglomerative-Lance-Williams/ 4072b6353dd761ba0c219a4b86 0073f08bb3db27

139. Rui HS, Ruger DS, Liu H, Huang Z. Dissimilarity measures for content-based image retrieval. In 2008 IEEE International Conference on Multimedia and Expo, Hannover, Germany. 2008; 1365–1368.
140. 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. 2018; 6: 46595–46616. https://doi.org/10. 1109/ACCESS.2018.2866091
141. Jégou H, Douze M, Schmid C. Hamming embedding and weak geometry consistency for large scale image search, extended version. Available from: https://inria.hal.science/inria-00548651/document/.
142. Philbin J, Chum O, Isard M, JS, Zisserman A. Object retrieval with large vocabularies and fast spatial matching. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Minneapolis, MN, USA. 2007; 1–8.
143. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, et al. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 2020 Sep; 408: 189–215. https://doi.org/10.1016/j.neucom.2019.10.118
144. Sun B-Y, Huang D-S, Fang H-T. Lidar signal denoising using least-squares support vector machine. IEEE Signal Process Lett. 2005 Feb; 12: 101–104.
145. Raheja JL, Mishra A, Chaudhary A. Indian sign language recognition using SVM. Pattern Recogn Image Anal. 2016 Apr; 26: 434–441.

 