IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 6 July 2023, accepted 19 July 2023, date of publication 25 August 2023, date of current version 7 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3308911

SURVEY

Content-Based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation, and Evaluation Parameters

DIVYA SRIVASTAVA[1], (Associate Member, IEEE), SHASHANK SHESHAR SINGH[2], B. RAJITHA[3], (Senior Member, IEEE), MADHUSHI VERMA[1], MANJIT KAUR[4], (Senior Member, IEEE), AND HEUNG-NO LEE[5], (Senior Member, IEEE)

[1]School of Computer Science Engineering and Technology (SCSET), Bennett University, Greater Noida 201310, India
[2]Department of Computer Science and Engineering (CSED), Thapar Institute of Engineering and Technology, Patiala 147004, India
[3]Department of Computer Science and Engineering (CSED), MNNIT Allahabad, Prayagraj 211004, India
[4]School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana 506371, India
[5]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

Corresponding author: Heung-No Lee (heungno@gist.ac.kr)

This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2023-2021-0-01835) supervised by the Institute of Information & Communications Technology Planning & Evaluation (IITP).

ABSTRACT In the era of massive data production through the internet and social media, the volume of images generated is immense. Storing and retrieving relevant images efficiently pose significant challenges. Content-based image retrieval (CBIR) has emerged as a prevalent method for retrieving relevant images based on query images from large image collections. CBIR relies on three fundamental elements: the selection, extraction, and representation of features. This paper delves into a comprehensive survey of these crucial aspects. This paper begins by investigating the significance and wide-ranging applications of CBIR. It subsequently delves into an intricate analysis of feature selection, encompassing attributes such as color, texture, shape, and descriptors. Following this, the paper navigates through sections dedicated to feature extraction techniques and their subsequent representation. Furthermore, this paper includes an assessment of recent research articles and the methodologies they employ within the realm of CBIR. Significantly, CBIR has witnessed a notable expansion to incorporate deep learning techniques in recent times. The survey presents an overview of these recent methods and their integration into CBIR frameworks. This paper concludes by offering an extensive outline of 215 articles, encompassing a wide range of analyses conducted within the field of CBIR. Finally, this paper also outlines potential research directions for the future. It sheds light on areas where CBIR can continue to evolve and enhance its capabilities.

INDEX TERMS Content-based image retrieval, text-based image retrieval, deep learning, image processing, image retrieval, color feature, texture feature, shape feature, key point descriptors, speed up robust feature.

I. INTRODUCTION

Every day, terabytes of images are shared and stored on the internet, contributing to the creation of a vast image database. Retrieving relevant images from such a massive dataset is a challenging and demanding task, leading to new avenues of research in multimedia. Text-based image retrieval (TBIR) and content-based image retrieval (CBIR) are the two primary methods used to retrieve images based on text and content, respectively. TBIR relies on the textual data (metadata) associated with the image. This textual data can be manually added or generated automatically using various

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

tools. However, TBIR faces two significant issues regarding image annotation: 1) manually annotating the information is time-consuming and requires significant effort, 2) there can be variations in human interpretation of the annotated information. For instance, the same image of a beach could be interpreted differently, leading to various labels like palm trees, ocean, boats, sunbathing, vacation spot, and so on. Automatic annotation of images is limited by a significant semantic gap. For instance, when searching for an image of a "Jaguar," the system might retrieve images of both the animal and the automobile, as depicted in Figure 1. Furthermore, there are instances where formulating a precise query for a specific image content becomes challenging. As shown in Figure 2, a textual query for the given image could take various forms such as a) "shaded rose," b) "rose with different colors," c) "petals with different colored rose images," and so on. This reliance on text-based retrieval confines the scope to image annotations. In cases where the annotations and textual queries do not align, the results can be inaccurate.

To address the limitations of TBIR, a solution emerged that leverages visual content for image retrieval, known as CBIR. In CBIR, images are categorized based on their actual visual content, incorporating features such as shape, color, and texture. A significant step in the history of CBIR can be traced back to the conference on database applications of pictorial applications hosted in Florence in the year 1979 [1]. This conference centered around the advancement of integrated databases containing both textual and visual elements, along with their diverse applications. Subsequently, in 1992, the US National Science Foundation (USNSF) organized a workshop focused on the utilization of visual information in management systems across domains like education, industry, weather forecasting, entertainment, and healthcare [2]. The foundational principles of CBIR are explored in [3]. The present paper aims to consolidate the state-of-the-art methodologies, evaluation parameters, datasets, and deep learning approaches employed in the domain of CBIR.

The main highlights of the paper are:

- **TBIR and CBIR Comparison:** The paper delves into the distinction between TBIR and CBIR, discussing their respective applications, advantages, and drawbacks.
- **Unique Approach:** Unlike existing surveys, this paper provides an in-depth understanding of the manual extraction of features, as opposed to automatic feature selection. Few surveys explore the nuances of manually selecting and extracting image features.
- **Domain Knowledge Consideration:** This section stands out by focusing on the often overlooked domain knowledge of images. Existing surveys tend to neglect the significance of domain knowledge in comprehending an image.
- **Image Content Handling:** The paper covers the selection, extraction, and representation of image content.
- **Evaluation Measures and Parameters:** The most frequently employed evaluation methods and parameters are discussed, accompanied by their equations and usage.
- **CBIR with Deep Learning:** An overview of various deep learning techniques used in CBIR is presented. To enhance clarity, these techniques are summarized using a diagrammatic flow which illustrates the different approaches.

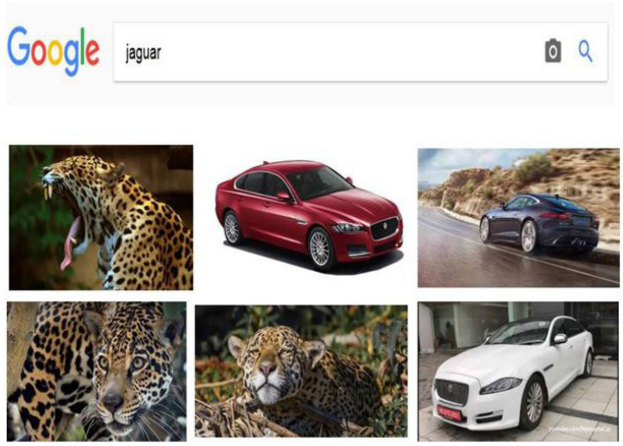

FIGURE 1. Images retrieved for Jaguar as a textual query (source: https://images.google.com/).

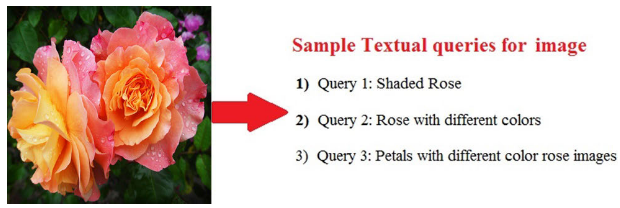

FIGURE 2. Sample of query statements for the given rose (source: https://pixabay.com/).

The remaining sections of the paper are structured as follows: Section II elaborates on CBIR methods. Section III conducts an in-depth analysis of image content. Section IV introduces evaluation techniques for CBIR methods. Section V explores the integration of Deep Learning into CBIR. Finally, Section VI provides a conclusion summarizing the paper's key findings and contributions.

II. CONTENT-BASED IMAGE RETRIEVAL

While CBIR may not be a recent research area, the exponential growth of image databases, diverse usage scenarios, and broad applications continue to keep it vibrant and active. In today's context, the internet witnesses the storage and retrieval of terabytes of images daily using various search engines. However, prior to the storage and retrieval of images, possessing domain knowledge is crucial for effective implementation.

A. DOMAIN KNOWLEDGE

In visual search, knowledge of the image's domain plays a vital role, as it helps to reduce the gap between the feature description and the semantic interpretation of the image. While well-defined laws are not available for domain knowledge, several papers [4], [5], [6], [7], [8], [9], [10], [11] have presented various methods to acquire knowledge about the domain of the images, as summarized below.

1) **Laws of Syntactic Equality and Similarity**: These laws establish relationships among images based on their pixel content or sometimes their actual contents, without considering how they visually or physically appear. For instance, if two images share specific shades of brown in their lower parts, they might be classified as outdoor scenes and differentiated from other images. Hatano [4] employed RGB color space to determine syntactic similarity between query and stored images.
2) **Equality and Similarity based on Human Perception**: This approach relies on how humans perceive color similarity, and it's well-defined by color spaces like CIE-Lab and Munsell. Rossin [5], Siddiqi and Kimia [6], and Treisman et al. [7] explored various techniques rooted in human perception to describe image equality and similarity.
3) **Geometric and Topological Pattern-based Equivalence**: This principle defines both similarities and differences in patterns within a space. Objects might share similar geometries but differ in physical properties. For instance, two images might contain objects that are geometrically similar, such as circles or rectangles, positioned in similar ways. However, these objects could differ in terms of color, texture, or other perceptual attributes. Chang and Hsu [8] and Tagare et al. [9] delved into various geometric and topological rules.
4) **Category-Based Rules**: This rule categorizes images based on common properties. For example, images of faces would share properties like eyes, nose, lips, and ears.
5) **Man-made Customs**: This involves using artificial conventions to define equality among images. For example, rules could state that images with straight lines and perpendicular corners depict indoor scenes. In [11], it was established using rules for the fashion, textile, and clothing industries.

B. NARROW V/S BROAD DOMAIN

Images can be divided into two types: narrow domain and broad domain, which helps in their semantic description [12]. Domain knowledge assumes a pivotal role in image retrieval by bridging the gap between the technical feature description and the semantic interpretation of an image. A common approach to image retrieval involves initially narrowing down images from a broad domain to a more defined semantic scope. In this context, a narrow domain pertains to limited variability in recording circumstances, viewing angles (frontal/side), and object appearances within the image content. For example, the "Faces_easy" category in the Caltech-101 dataset has a narrow domain because all the faces are shown from the front with a clear background. Consequently, within a narrow domain of images, there exists a distinct and well-defined semantic description. But if those same faces were in a crowd or had different backgrounds or angles, the domain would be broad, meaning there's a lot of variation in how the same thing appears. The internet has lots of images, and that's a great example of a broad domain (for more information refer [13] and [14]).

C. CBIR APPLICATIONS

In literature, applications of CBIR are categorized into three broad categories: search by association, search by a specific image, and search by category. These categories are determined by the user's purpose when using the CBIR system.

1) *Search by Association.* This type of search refers to situations where the user doesn't have a specific search goal but is interested in finding interesting items. This kind of search involves refining the search iteratively based on the user's feedback about relevance [15], [16]. An early instance of this type of search is discussed in [17], where a basic idea was introduced about a visual interface for accessing any image in a database. In this example, the query image was a rough sketch, indicating that the user isn't aiming for a specific search but is looking for interesting and similar matches to the query image.
2) *Search by a Specific Image.* In this category, users are focused on searching for a particular image. This specific image could have been anything - a particular image that the user had in mind, an image of the same object for which the user already had an image, a target image, and so on. For example, Flickner et al. [18] discussed searching art catalogs, while Cox et al. [19] discussed searching stamps, catalogs, industrial components, and art.
3) *Search by Category.* This type of search focuses on retrieving images belonging to a specific category. In this scenario, the user provides an image as a starting point and aims to find other images that also belonged to the same category. For instance, Huet and Hancock [20] delved into image retrieval based on specific line patterns. Srivastava et al. [21] explored pattern-based image retrieval, with a focus on various types of stripe patterns in clothing (horizontal, vertical, diagonal, etc.). Jain and Vailaya [22] presented an approach for classifying trademarks, while Srivastava et al. [23] discussed the classification of objects such as faces, sunflowers, bikes, airplanes, cameras, and more.

Currently, CBIR is extensively used in the fashion industry to filter clothing patterns, fabric types, etc. CBIR also finds applications in various domains including medical imaging systems, satellite imagery, search engines (such as Google, Bing, Yahoo, etc.), and more. The effectiveness of CBIR relies heavily on the content of images being used. "Content" refers to the aspects of images like color, texture, shape, etc. The following section discusses different types of content, how they are extracted, and how they are represented.

III. CONTENT OF THE IMAGES

The contents of an image refer to its distinct features, and this forms a crucial element of CBIR. It involves three key stages: feature selection, extraction, and representation. The primary objective of a content-based retrieval system is to identify the visual attributes that characterize an image, video, audio, and more. Such attributes encompass color, shape, texture, and keypoint descriptors, all of which are chosen for integration into the CBIR process. The choice of features heavily relies on the nature of the image dataset. For instance, datasets like the Brodatz dataset, describable textures datasets (DTD), and VisTex predominantly contain texture-based images. In these cases, texture features provide more precise information than color and shape features.

Similarly, datasets like the Wang dataset highlight the significance of color features in the CBIR process. Shape features, on the other hand, are essential when identifying specific shapes within images, irrespective of their color and texture. For example, datasets containing logos and traffic signs require shape features due to their distinct shapes. Hence, selecting the right features is a pivotal aspect of CBIR. The extraction of chosen features and their subsequent representation also play a vital role. The comparison between query image and the images in database depends on the similarity between their features. Various methods, such as color models, texture feature extraction, local binary patterns, etc., are employed for feature extraction. The selected features are then represented in various forms, including histograms and n-layered arrays (termed component vectors). The subsequent subsections delve comprehensively into color, texture, and shape features, exploring their extraction and representation.

A. COLOR FEATURE

Color is a fundamental and extensively used visual aspect in CBIR. This is because color is multidimensional, unlike the one-dimensional nature of images. It's our perception of electromagnetic radiation through our eyes. It's represented as a three-part vector, commonly referred to as a color space. Understanding color involves its lightness, hue, and saturation. Hue signifies the primary color that shapes a distinct visual impression for us. The intensity of color depends on how much white light is mixed in. For example, colors without white light are fully saturated, while colors like pink (a blend of red and white) are less intense. The powerful use of color in computer graphics has been studied by MacDonald [24]. In the domains of business, science, and industries, Weale et al. [25] explored color applications. Color features are also employed to pinpoint areas of interest or divisions within images. Researchers [26], [27], [28], [29], [30] investigated the utilization of color features in CBIR by focusing on specific regions of interest or divisions. To integrate color as a visual element in image processing, color features are extracted using diverse color models/color spaces [31], which is discussed in the subsequent subsection.

1) EXTRACTION OF COLOR FEATURES

Color features are extracted using various color models. These models offer diverse methods for representing and comprehending colors, each tailored to specific contexts and applications. The selection of a color model holds significant importance in computer vision algorithms. Different color models possess distinct characteristics. Table 1 provides a comparison of color models for extracting color features in CBIR. For instance, when applying a standard segmentation algorithm to an RGB model, it might group a sky image with various shades of blue into one segment. In contrast, a color model sensitive to shading might divide the same sky image into separate regions based on the different shades of blue. As a result, no single color model is universally applicable, as different applications may favor different color models. Various color models are discussed below:

- **Coloriemetric Color Model.** This model employs human perception to determine the visual similarity of two colors. It's a theoretical construct, and the Commission Internationale de l'Éclairage (CIE) [32] has introduced various standard observer color-matching functions derived from different experiments. CIE model's initial experiment dates back to the nineteenth century and was conducted by Thomas [33]. Subsequent researchers have validated its principles. CIE conceptualizes colors as vectors within a three-dimensional space, consisting of the luminance component Y and two additional components, X and Z [34]. CIE XYZ is a device-independent color model, making it particularly useful where consistent color representation matters, regardless of device characteristics. However, this model's effectiveness in image processing techniques is limited, as it isn't well-suited for quantitatively assessing color perception [35].
- **RGB Color Model.** RGB color model represents the primary colors of red, green, and blue. This model is commonly used in imaging and video devices where the color spectrum is formed by combining these three primary colors. In RGB color model, along with the primary colors of red, green, and blue, there's also a reference point for white in RGB color space. The model is primarily defined by the Maxwell triangle [36], illustrated in Figure 3. This model finds its main application in situations where color addition is essential. This is because different combinations of red, green, and blue lights are added to produce a wide array of colors. RGB is predominantly employed to display images in electronic systems such as computer monitors, TV screens, digital photography, and more.

TABLE 1. Comparison of color models for color feature extraction in CBIR.

Color Model	Representation	Color Space	Advantages	Disadvantages
Colorimetric	Tristimulus Values	CIE XYZ	Physiologically relevant	Device-dependent
RGB	Red, Green, Blue	Additive	Widely used in digital imaging	Sensitive to lighting conditions
HSV	Hue, Saturation, Value	Cylindrical	Intuitive interpretation	High memory requirement
CMYK	Cyan, Magenta, Yellow, Key (Black)	Subtractive	Effective for printing applications	Limited support in digital systems

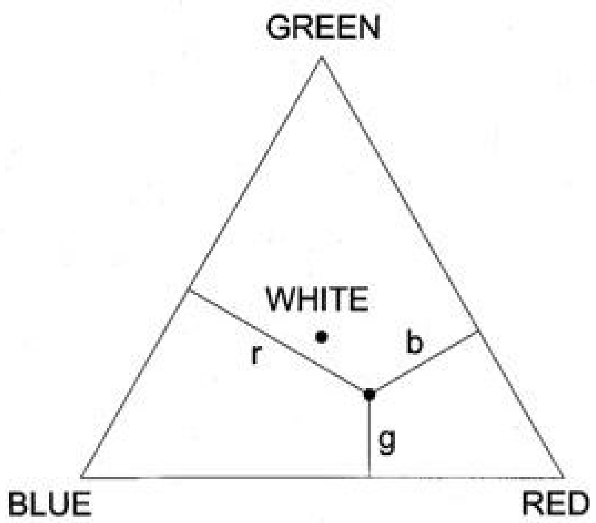

FIGURE 3. Maxwell triangle [37].

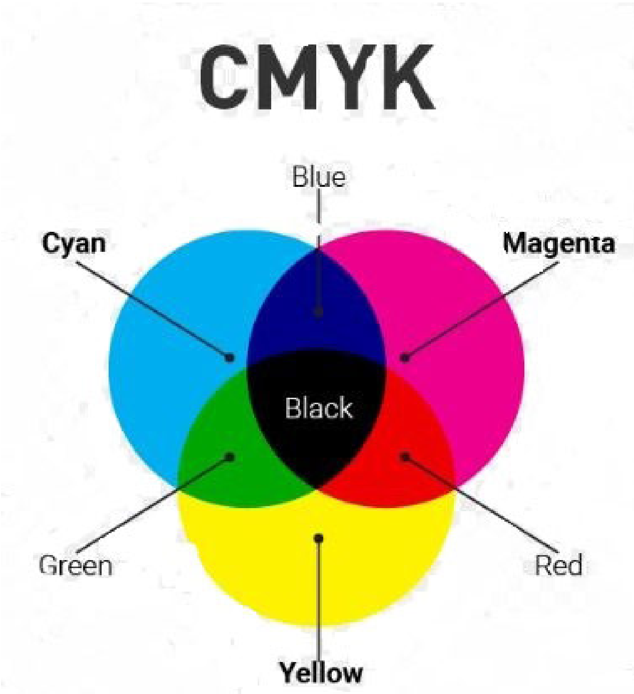

FIGURE 4. CMYK model [38].

- **CMYK Model.** CMYK stands for Cyan (C), Magenta (M), Yellow (Y), and Key (K). This color model primarily functions on a light background, often white. It's extensively employed in printing because the ink in printing subtracts colors from white light. Due to this subtractive nature, it's also referred to as a subtractive model. The process involves subtracting red, green, and blue from white light. When red is subtracted, it yields cyan. Subtracting green results in magenta, while subtracting blue produces yellow. When cyan, magenta, and yellow are combined, the colors darken, eventually culminating in a black color. The ''Key'' (K) component in CMYK denotes black, and this combination is displayed as 'K' in the CMYK model, as illustrated in Figure 4.
- **HSV Model.** This color model represents hue, saturation, and value. This model is part of non-uniform color spaces and is comparable to models like hue saturation intensity (HSI), hue chroma value (HCV), hue saturation brightness (HSB), and hue saturation lightness (HSL). It's visualized as a double cone, as shown in Figure 5. The central axis of cone denotes the value (intensity), ranging from 0% at black to 100% at white. Saturation also increases as you move away from black, as depicted in the Figure 5. In this model, hue signifies the primary color that creates a distinctive visual impression. The amount of white light mixed with the hue determines the color's saturation. Colors without white light are fully saturated, while colors like pink (a mix of red and white) are less saturated. HSV model is widely used in image processing as it provides a broad spectrum of colors and facilitates quantitative assessment of color values. Researchers like Sural et al. [39] employed HSV model histograms for image segmentation and retrieval. Additionally, an integration of gray-level co-occurrence matrix and HSV color [41] was utilized in CBIR.

2) REPRESENTATION OF COLOR FEATURE

Color holds significant importance in CBIR, and its representation is equally crucial. Extracted features are stored in an n-dimensional array known as a feature vector. Depending on its intended use, this feature vector can be depicted in various manners. The diverse color models, their representations, and their applications were explored in detail in [42]. An effective method for representing color for image retrieval was introduced in [43]. There are multiple strategies for handling color features, including histograms, moments, and names. These strategies are elaborated upon below:

- **Color Histogram.** Swain and Ballard introduced the color histogram for image matching [44]. This method

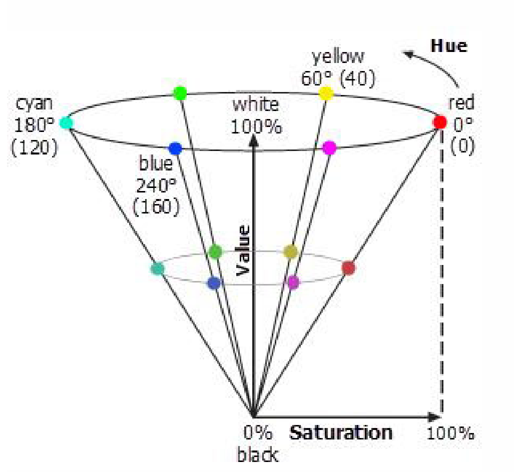

FIGURE 5. HSV model [40].

comprises three separate histograms, each corresponding to a primary color (red, green, and blue). It employs bins to quantize color distribution. A histogram gains more discriminative power with an increase in the number of bins used. One advantage of representing color features with histograms is their independence from translations and rotations around the viewing axis. Histograms remain relatively unaffected by slight changes in an image's perspective, scale, and orientation. They prove valuable when spatial arrangement details hold less significance. However, since histograms don't account for spatial arrangement, two distinct images can share the same histogram. Stricker et al. [46] analyzed various images, as depicted in Figure 6, with similar histograms.

- **Color Moments.** In 1995, Stricker and Orengo [46] introduced color moments as a method to assess the similarity of color images for image and video retrieval. Color moments provide three key snapshots of the probability distribution for each primary color. The first-order moment (mean) offers information about the average color of the image. The second and third-order moments reflect the variance and skewness of the color channel. This approach is highly efficient and effective for characterizing color distribution in images [46]. Color moments comprise only 9 values (3 moments for each primary color), making it a compact representation of color features. However, due to this compactness, it may occasionally have lower discriminative power. Yu et al. [47] incorporated color texture moments for CBIR.
- **Color Names.** Color names are frequently referred to as "basic hues" [48]. Each hue is associated with a fundamental color in accordance with the naming convention. It essentially involves dividing the color space and assigning color names to each segment. Liu et al. [49] discussed area-based image retrieval using high-level semantic color names. Mojsilovic et al. [50] explored image similarity and retrieval structure of colors and vocabulary.

FIGURE 6. Two images with similar histograms [45].

- **Color Sets.** Smith and Chang [65] introduced the color set approach for color image retrieval. This technique is employed to identify specific regions within a color image. The premise of this approach is that distinctive image regions contain a small number of dominant colors that are interconnected. An edge is established to determine whether a pixel belongs to a distinct region. This method finds application in VisualSEEK applications [51], [52], [53].
- **Color-Spatial Model.** All the previously mentioned techniques provide global color properties of the image but fail to offer spatial information about the image. Several methods were proposed to combine color features with spatial relations. Some of these techniques are outlined below.

  1) *Color Feature Extraction of Sub-blocks.* In this approach, the image is divided into sub-blocks, and subsequently, the color feature is extracted from each sub-block. Chua et al. [54] proposed a rapid signature-based color spatial image retrieval technique. Dividing the image into sub-blocks and then calculating the color feature for each sub-block made this approach highly efficient in terms of computation and storage. Hsu et al. [55], Rao et al. [56], and Chinque et al. [57] also employed a combination of spatial color methods for CBIR.
  2) *Color Coherence Vector (CCV).* In this method, each pixel is classified as either coherent or non-coherent. A pixel that is part of a large connected component is termed coherent, while a pixel that is not part of such a component is considered non-coherent. The challenge in CCV approach lies in determining whether a component is substantial. Pass et al. [58], Roy and Mukherjee [59], Al-Hamami and Al-Rashdan [60], and Ravani et al. [61] employed CCV for CBIR.
  3) *Color Correlogram.* The color correlogram, introduced by Huang et al. [62], encodes color spatial information into co-occurrence matrices, and its applications were explored in [63]. Handling numerous co-occurrence matrices poses a

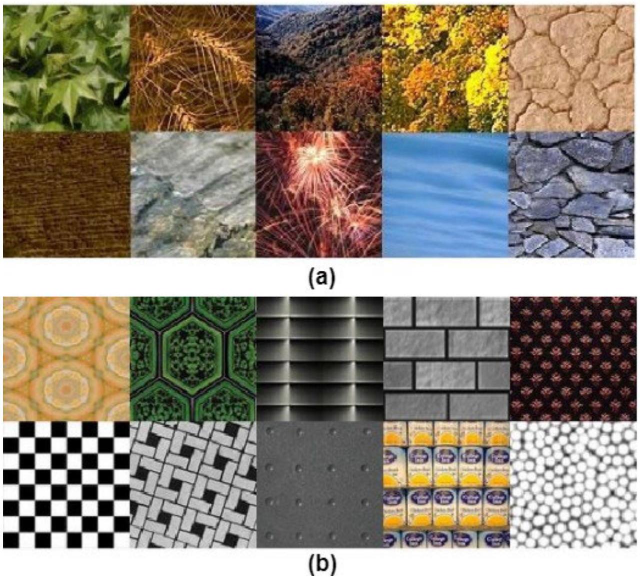

FIGURE 7. **(a) Examples of Natural textured surfaces (b) Examples of Artificially created surfaces [68].**

significant challenge with this method. Therefore, only the main diagonal of the co-occurrence matrices is calculated, stored, and referred to as the auto correlogram [64].

B. TEXTURE FEATURE

Texture features characterize the visual appearance and material attributes of an image. It involves the spatial arrangement of elements in various relative positions [66]. Spatial texture arrangement defines how texture features are organized within an image, reflecting properties like uniformity, coarseness, regularity, smoothness, directionality, and more [67]. Two images are said to possess similar textures when their spatial pixel arrangements match. Texture is categorized into two types: *(a) natural textured surfaces* and *(b) artificial textured surfaces*. Natural textures, also known as stochastic textures, lack a specific pattern and are common in real-world images like wood, water, grass, and paper, as shown in Figure 7(a). Artificially generated textures are structured textures with texels consistently positioned in specific patterns such as checks, stripes, and dots, as depicted in Figure 7(b). Various texture models in literature describe attributes like roughness, consistency, linearity, directionality, and granularity.

Texture features can be obtained through various methods [69], such as perceptual, statistical, local binary pattern, structural, transform, and MPEG-7 texture descriptor. Several studies [70], [71], [72] offered an overview of different techniques for extracting texture characteristics. Figure 8 presents an overview of these methods.

1) PERCEPTUAL METHOD

There's quite a long list of perceptual texture features, but among them, the six textural features introduced by Tamura et al. [75] in 1978 are widely used. These commonly used six features are coarseness, contrast, directionality, line-likeness, regularity, and roughness. Some methods [76], [82] employed these features for CBIR.

1) *Coarseness.* This characteristic of texture pertains to the dimensions of tiny constituents composing the texture. It helps to estimate how much the texture elements repeat. It's also closely tied to scale since it measures the rate of various spatial changes. If an image has a few large texels, it's considered coarse, while an image with many small texels is seen as having a fine texture [73]. In addition to the coarseness defined by Tamura, researchers have also defined coarseness using various methods. A comparison of these methods was provided in [74].
2) *Contrast.* This texture characteristic describes the quality of the image rather than its pattern. A highly contrasted image is one where the various texels are easily noticeable and distinguishable. Contrast is used to highlight pixels where two textures (pixels) have the same structure but differ in brightness levels. Tamura et al. defined contrast in terms of standard deviation and kurtosis [75]. Contrast is employed to accentuate pixels exhibiting similar structural patterns but varying in terms of their brightness levels. Tamura et al. introduced a definition of contrast based on kurtosis and standard deviation [75].
3) *Directionality.* Directionality is a global texture property that estimates the prominent direction or directions within an image. It's calculated by considering the shape and spatial arrangement of texels. Tamura et al.'s approach focused on the total number of pixel directions rather than the specific type of direction [75].
4) *Line-Likeness.* Line-likeness is a feature that complements the directionality characteristic. When the alignment of a specific edge matches that of its neighboring edges, it's recognized as a line. The determination of line-likeness involves utilizing a directional co-occurrence matrix [73].
5) *Regularity.* This feature measures the randomness [76] in the arrangement of texture elements within an image. The more random the placement of texture elements in an image, the less uniformity in the image, and vice versa.

2) STATISTICAL METHOD

The statistical method uses the intensity values to determine the texture properties of an image. It can be categorized into first-order, second-order, and third-order (or more pixels) statistics. Julesz et al. [83] were the first to study statistical methods in relation to human perception. Haralick and Shanmugam defined the statistical method using the Gray-Level Co-occurrence Matrix (GLCM) [84], which is widely used for texture feature extraction.

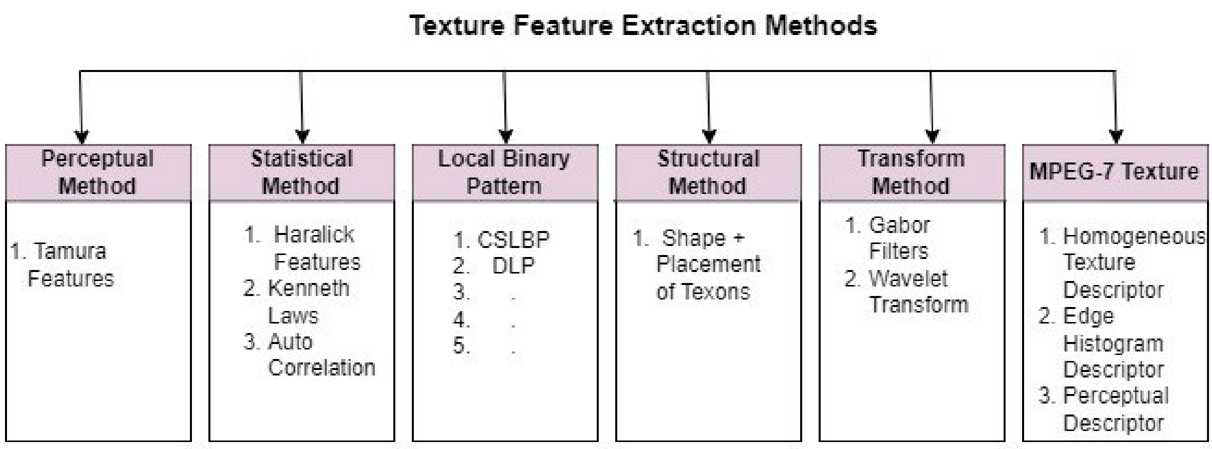

FIGURE 8. **Texture feature extraction methods.**

GLCM captures the spatial relationship between the intensity values of a reference pixel and its neighboring pixels. Haralick defined 14 statistical texture features using GLCM. Among these 14 features, contrast, correlation, energy, and homogeneity are the most commonly used [73]. GLCM is calculated using two parameters: angle (θ) and distance (d). Various combinations of angle (θ) and distance (d) for computing texture features are studied in [85] to detect patterns in fabrics. Srivastava et al. [21] discussed pattern-based image retrieval using GLCM.

In 1980, Kenneth Laws introduced a set of convolutional filters and three primary masks to detect edges, spots, textures, and waves in textures (known as Laws textures). Features defined include regularity, fineness, coarseness, density, frequency, uniformity, linearity, roughness, phase, granularity, randomness, contrast, smoothness, and directionality. The applications of Kenneth textures were discussed in [86] and [88]. Another technique for texture characterization was auto-correlation [89]. It defines the fineness/coarseness of texture based on repeating patterns of texture elements. Typical texture auto-correlation exhibits peaks and lows [90], [91]. For coarse textures, auto-correlation decreases slowly, while for fine textures, auto-correlation decreases rapidly. A comparative study among various statistical methods was presented in [92].

3) LOCAL BINARY PATTERN

Local Binary Pattern (LBP) [93] is a widely recognized method for extracting texture features. It establishes a relationship between a reference pixel and its neighboring pixels. In this technique, each pixel is treated as a reference pixel, and information is extracted based on its neighbors. The process involves subtracting the value of the central pixel from each adjacent pixel. If the difference between the two is greater or equal, the assigned value is 1; otherwise, it's 0. The 0s and 1s are merged using the power of two to generate a singular value for the central pixel.

LBP was adapted to be rotation invariant [94]. Multi-Block LBP (MBLBP) calculates average powers using a 2×3 mask, over which LBP is computed. MBLBP showed an 8% improvement over the original LBP [95]. Heikkila et al. designed a center symmetric LBP (CSLBP) [97], which surpasses LBP in terms of computational efficiency by reducing the descriptor length. CSLBP examines four pairs of diagonal opposite pixels, generating a binary pattern of four parts. CSLBP demonstrated high accuracy in local-based image similarity [98].

Tan and Triggs introduced the local ternary pattern (LTP) [99], utilizing an edge radius to analyze neighboring and central pixels, particularly for face recognition. LTP outperformed LBP and CSLBP under uniform lighting variations [99]. For non-uniform lighting variations, the center symmetric local ternary pattern (CSLTP) is more effective [101]. The paired rotation invariant and noise-tolerant (BRINT) approach is a variation of LBP, resilient against lighting changes, rotation, and noise [102].

Chakraborty et al. [100] and Zang et al. [103] proposed descriptors: local directional gradient pattern (LDGP) and local derivative pattern (LDP), respectively, for facial recognition. The local ternary co-occurrence pattern (LTCoP), a combination of LTP and LDP, has been applied in medical image retrieval [104]. Sobel-LBP, an enhancement of LBP [105], improved image edges using the Sobel operator. Another edge-based descriptor, the directional local extrema pattern (DLEP) [106], involved edges at specific directions and has been used for CBIR. Murala and Wu designed local tetra pattern (LTrP) [107], which used the vertical and horizontal directional neighborhoods of each pixel to form a tetra pattern, later transformed into a paired pattern for CBIR.

Various other texture descriptors, such as local extrema co-occurrence pattern [108], focus symmetric local binary co-occurrence pattern [109], and multi-goal local extrema peak valley pattern [110], were utilized in different image retrieval contexts. The local neighborhood difference

TABLE 2. Qualitative analysis of LBP and its variants.

Descriptor	Robust Against			
	Expression	Illumination	Background	Noise
LBP [93]	Good	Poor	Poor	Poor
CSLBP [97]	Good	Poor	Poor	Poor
BRINT [102]	Average	Poor	Poor	Good
MBLBP [95]	Average	Poor	Poor	Good
LDP [103]	Average	Average	Poor	Poor
LTrP [107]	Poor	Poor	Poor	Poor

pattern [111] was proposed for large-scale natural and texture image retrieval. Neighbourhood gradient hexa pattern [112] and quadruple local pattern (LQP) [113] descriptors were suggested for face recognition and retrieval.

In the literature, LBP and its variations are predominantly used for facial recognition. These approaches are commonly applied to handle variations in facial expression, illumination, background, and noise. Texture descriptors are also widely employed in medical imaging. Murala et al. proposed the local mesh pattern [114] for indexing and retrieving biomedical images. It was based on a local pattern formed by an array of neighboring pixels. Another descriptor used in clinical image retrieval was the peak valley edge pattern [115], which extracted directional edge information using first-order derivatives. These two approaches were combined in [116] to create the local mesh peak valley edge pattern for indexing and retrieving Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) images. Table 2 shows the qualitative analysis of LBP and its variants.

4) STRUCTURAL METHOD

The structural method is particularly suited for images with large texture elements, known as macro-textured images. In these cases, the focus is initially on the size of the texture elements and then on other properties. This method is effective for textures that are either artificially created with a regular pattern or intentionally placed in a specific arrangement. It consists of two main aspects: first, extracting the texture elements, and second, deducing the rules governing their placement within the image.

- *Texture element extraction.* A region within an image that exhibits a consistent dark intensity is referred to as a texture element. Numerous methods exist in the literature for extracting such texture elements. Voorhees and Poggio [117] defined these elements as "blobs" and extracted them by convolving the image with Laplacian of Gaussian (LoG) filters at various scales. Several articles, including [118], [119], [120], [121], [122], discussed various approaches employed for the extraction of image texture components.
- *Inference of the position rule.* In [123], authors employed Voronoi diagram to describe the arrangement of texture elements within an image. The complete steps to construct Voronoi diagram were outlined in [123]. Tuceryan et al. and Torodovic et al. discussed texel-based segmentation using the Voronoi diagram in [124] and [125], respectively.

5) TRANSFORM METHOD

The Transform Method [126] involved examining the image in the frequency domain. As texture often exhibits repetitive patterns, frequency domain analysis is valuable for texture analysis in an image. Fourier transform, Gabor transform, and wavelet transform are commonly used transformation techniques in the literature [127] for image recognition. Texture features can be extracted using neighborhood Fourier transform as proposed by Zhou et al. [128] for image classification.

Gabor transform [129] is employed for local analysis of texture in the spatial domain. Essentially, a Gabor filter captures specific frequency content in the image within a local region in a particular direction. Gabor transform has been extensively used in studies from [130], [131], and [132] for texture-based image retrieval. Gabor is suitable for stationary signals as it relies on spatial correlation using Fourier analysis. Gabor Filters have also been utilized in medical image retrieval using GNU image tool [133]. Wavelet Transform [134] is another technique that works efficiently for both non-stationary and stationary signals. It provides four pieces of information about the image: horizontal edges, vertical edges, diagonal edges, and the approximation part. There are several wavelet transforms in the literature. Notable among them are the Haar wavelet [135], dual tree wavelet [136], [137], Daubechies wavelet [138], [139], and complex wavelet [140], [141]. Traina et al. [142] designed an application for medical image retrieval using Daubuchies, Haar, and Gabor wavelet transforms.

6) MPEG-7 TEXTURE DESCRIPTOR

The MPEG-7 texture descriptor [143] includes three types of descriptors: Homogeneous Texture Descriptor (HTD), Edge Histogram Descriptor (EHD), and Perceptual Browsing Descriptor (PBD). HTD assumes homogeneous texture within an image or region and provides a quantitative representation based on spatial-frequency measurements. EHD offers information about five types of edges: four directional edges and one non-directional edge. It utilizes an 80-bin histogram to depict the local edge distribution within an image. EHD is suitable for images with non-homogeneous textures, like natural images, clip art, and sketches. PBD is utilized in applications where perceptually significant features are employed for TBIR [145], [146].

C. SHAPE FEATURES

In image processing, the shape of an object refers to its geometric structure. In 2D images, it's the outline that encloses the object, while in 3D images, it encompasses the mathematical representation of the object's surface. A comprehensive analysis of various techniques to represent and describe shapes was presented in [147], summarized in Figure 9.

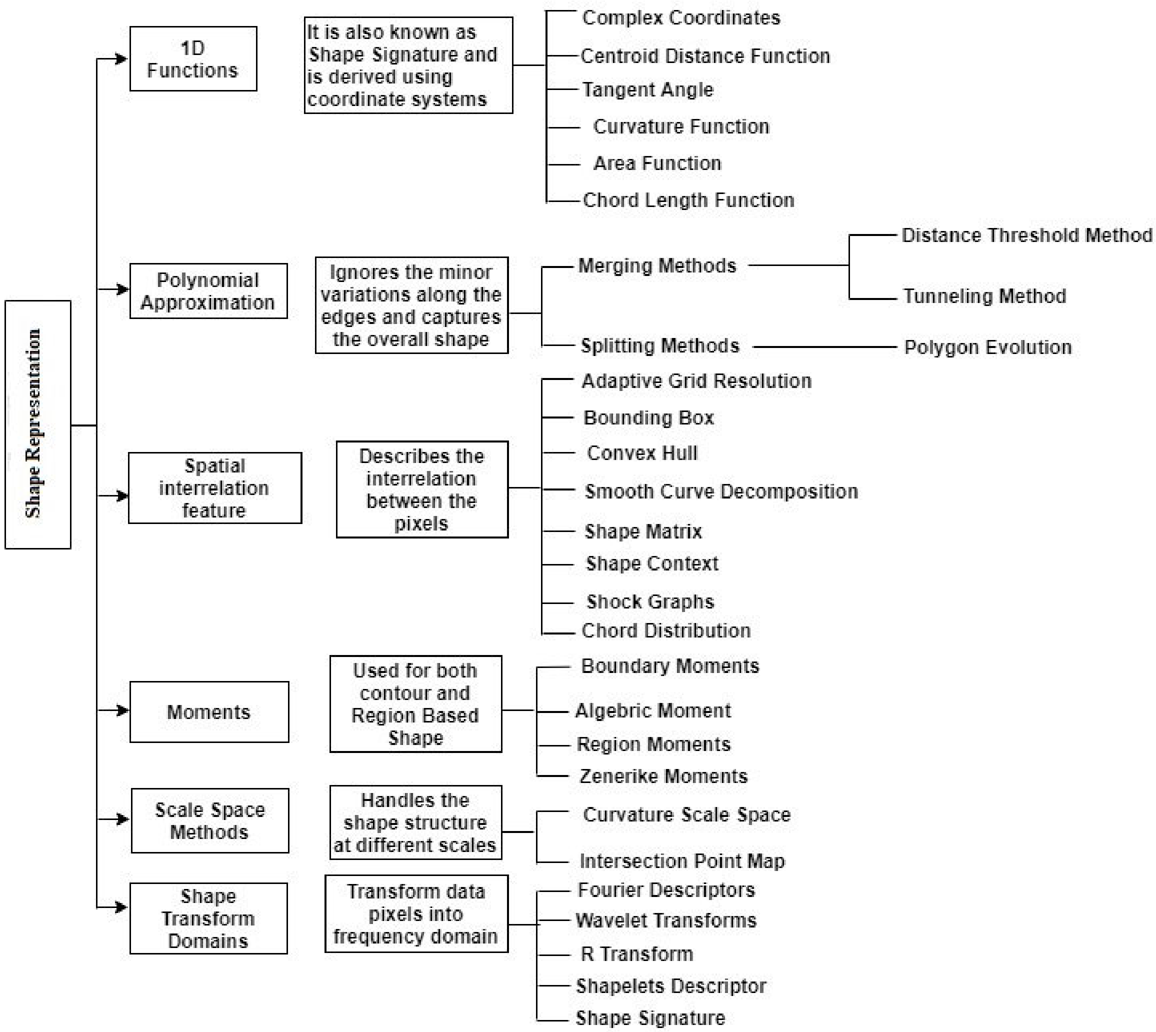

FIGURE 9. Shape description techniques [147].

Typically, shape features are useful for objects with distinct shapes, such as logos, brand names, and traffic signs. Extracting and representing shape features are crucial in many CBIR applications that deal with well-defined shapes. Research on shape measures for CBIR, using shape features on a dataset containing 500 brand names, was discussed in [148]. Shape characteristics are often derived from second invariants [149], Fourier features [150], [151], boundary segments [152], contour shapes [153], and more.

D. INTEREST POINT DESCRIPTORS

In recent literature, alongside color, texture, and shape features, several interest point (keypoint) descriptors such as Scale Invariant Feature Transform (SIFT) [163], Speed Up Robust Feature Transform (SURF) [164], and Principal Component Analysis-SIFT (PCA-SIFT) [165] have been widely used for image analysis and retrieval. Keypoint descriptors operate in three stages: detecting interest points in an image, describing those interest points, and comparing images based on their interest points. David Lowe introduced SIFT [167] for the detection and description of local features in images. SIFT identifies interest points in an image, typically corner points or distinctive points. These points are then analyzed at various scales to obtain the final keypoints. Features are computed for each keypoint by calculating the orientation histogram of the gradients of the image around the keypoint location, resulting in a 128-dimensional feature vector. This vector is calculated for all keypoints detected. SIFT is scale, rotation, translation, and partially illumination invariant [163]. Similar analyses of SIFT and its variations, including Affine SIFT (ASIFT), PCA-SIFT, Color Invariant SIFT (CSIFT), SURF, and Global SIFT (GSIFT), can be found in [169]. However, SIFT can be computationally expensive; SURF [164] is an advancement over SIFT. It employs Gaussian approximations for faster computation of interest points. SURF is preferred over SIFT for its robustness to scale, rotation, translation, and illumination changes, as well as its computational efficiency [170]. These descriptors find

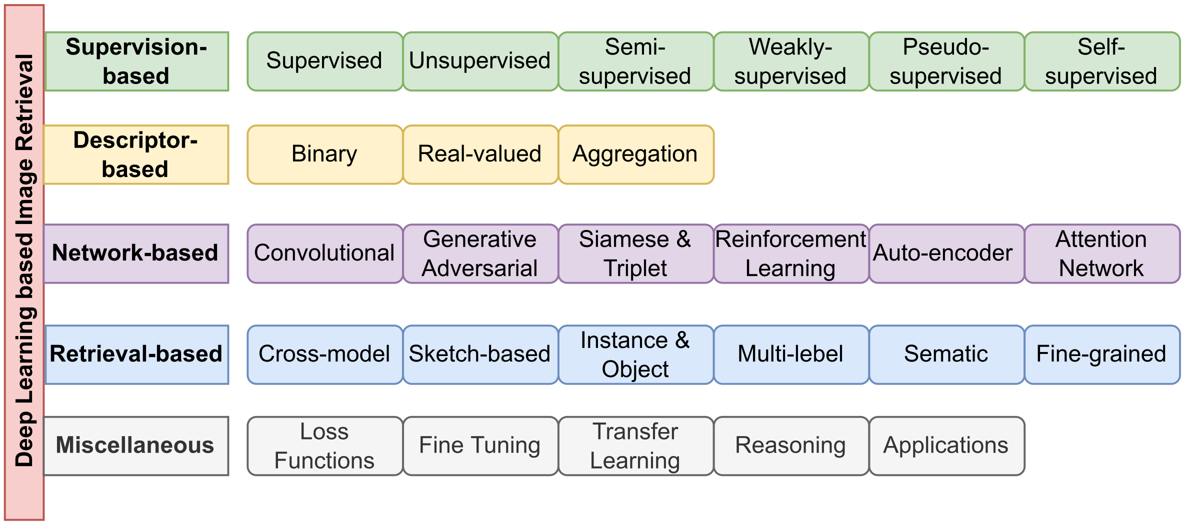

FIGURE 10. Visualizing the taxonomy of deep learning-based image retrieval methods.

applications in object recognition [171], image classification [172], and retrieval.

A fusion of SURF and LBP features was proposed by Srivastava et al. [23] for image characterization containing different types of single objects. They used SURF keypoints to create a circular structure and extract the Region of Interest (ROI) in an image. Subsequently, LBP features were computed for the ROI. Nouman et al. [173] proposed a novel approach for CBIR using a combination of SIFT and SURF. They addressed scaling and rotation issues and employed a bag of visual words to describe features. Darn et al. [174] introduced a camera-based image retrieval system using Scale and Rotation Invariant Features (SIRF) and compared them with other descriptors like SIFT, SURF, and Oriented FAST and Rotated BRIEF (ORB). A fusion of SURF and dominant color features was used for image retrieval on a mobile platform by [175].

E. CBIR UTILIZING COMBINATION OF COLOR, TEXTURE, SHAPE, INTEREST POINT DESCRIPTORS

As research in the field of CBIR advanced, experts began focusing on combining two or more low-level features to enhance system performance. The combination of multiple features proved to be more effective than using individual features in CBIR systems [77], [78], [79], [80], [81], [82]. For instance, Huang et al. [177] proposed combining color moments and Gabor texture features for image retrieval. They assigned appropriate weights to color and texture features and calculated similarity using the Euclidean distance. The study compared RGB moments, HSV color moments, Gabor filters, and a combination of color moments and Gabor features for retrieval.

Later, approaches emerged where fixed weights were assigned to color and texture features, and in some cases, weights were dynamically assigned based on the image itself [155]. Wang et al. [156] integrated color, texture, and shape for CBIR. Savita et al. [157] introduced a method using dominant colors, texture, and shape features. Instead of utilizing all color features, they employed K-dominant color features along with texture and shape for image retrieval. This approach outperformed existing methods like LBP and color models.

Prasad et al. [30] combined color, shape, and location information for region-based image retrieval systems. Similarly, another region-based approach using color, shape, and texture features was proposed in [158]. In the context of region-of-interest based natural image retrieval, a combination of color and texture features was employed [159]. However, it is required to design more efficient retrieval systems with the help of recently developed methods such as [160] and [161].

Table 3 shows the comparative study of CBIR using color, texture, shape, and descriptor using various features including feature selection, feature extraction, dataset used, and key findings.

IV. EVALUATION TECHNIQUES AND PARAMETERS

Performance evaluation of CBIR approaches play a pivotal role in assessing the effectiveness and efficiency of these methods. Early assessments of CBIR systems often involved presenting the results of one or more sample queries. This practice was convenient as it could showcase positive outcomes by using query images that yielded excellent results. However, this approach lacks objectivity as it doesn't provide a standardized performance measure or a means of comparing different CBIR systems.

Recently, researchers have utilized various performance metrics like precision, recall, specificity, F-measure, and class accuracy to establish more rigorous and objective evaluation criteria for CBIR approaches. Precision measures the ratio of relevant images retrieved to the total number of retrieved

TABLE 3. Comparative study of CBIR techniques using color, texture, shape, and descriptor.

Year	Author(s)	Feature Selection	Feature Extraction	Dataset Used	Key Findings
2002	J Han et al. [176]	Color	Fuzzy Color Histogram (FCH)	Different color images	1) FCH was proposed for color images and performed better than the conventional color histogram approach. 2) Results were computed on various images: flowers, cars, lions, etc.
2010	Huang et al. [177]	Color and Texture	Color moment and Gabor Texture	Different color images	1) Fixed weights were assigned to each element, and similarity was calculated using combined color and texture features. 2) Comparison of RGB moment, HSV color moment, and Gabor filter were presented for image retrieval system.
2011	Yue et al. [155]	Color and Texture	HSV Model and GLCM	Different color images	1) Color and texture elements were assigned random weights based on the image. 2) Both components were initially isolated and then compared with the features of images in the dataset. They were subsequently categorized into color and texture elements separately.
2014	Bhosale et al. [171]	Color, Texture and Shape	Dominant colors, Spatial Texture and Pseudo-Zernike	Different color images	1) Combined trio of color, texture, and shape. 2) The combined feature extraction showed better visual inclination than the single component recovery, leading to improved retrieval results.
2012	N. Shrivastava et al. [178]	Color, Shape and Texture	Color moments, statistical approach for texture and region-based for shape	Corel	1) First, images were compared based on color feature similarity. 2) Images retrieved based on color were further improved by matching their texture and shape features for retrieval.

TABLE 3. *(Continued.)* **Comparative study of CBIR techniques using color, texture, shape, and descriptor.**

Year	Author(s)	Feature Selection	Feature Extraction	Dataset Used	Key Findings
2013	N. Shrivastava et al. [179]	Texture	LBP	Outex, Curet, UIUC	1) Compared LBP and its different variations, including CSLBP, CLSP (Complete local structure pattern), and RLSP (Robust local structure pattern). 2) RLSP demonstrated superior performance compared to CSLBP and CLSP across all images utilized in the experiment.
2014	N. Shrivastava et al. [180]	Color and Texture	ROI and LBP	MPEG-7 CCD and Corel	1) Conducted a comparative study of different ROI techniques for image retrieval. 2) Integrated ROI and LBP for improved and expedited image retrieval systems.
2012	Chauhan et al. et al. [181]	Color, Shape and Texture	Color moments, GLCM, ROI	Corel	1) Implemented a three-layered feed-forward architecture in the order of color-texture-shape. 2) Utilized relevance feedback for better results.
2014	N. Shrivastava et al [182]	Color, Texture and Shape	Color histogram for color, Gabor filter, and Fourier descriptor	Corel	1) Employed a three-step process for image retrieval: initially based on color, followed by texture, and finally by shape features. 2) Used outcomes of each phase as input for the subsequent phase, resulting in a gradual reduction in the dataset's size at each step.
2015	Karakasis et al [183]	Moment invariants descriptors	Affine moments invariants	Different color images	1) Introduced an image retrieval framework that utilized affine image moment invariants as descriptors for local image regions. 2) Effectively addressed the challenge of handling transformations in CBIR.

TABLE 3. ***(Continued.)*** **Comparative study of CBIR techniques using color, texture, shape, and descriptor.**

Year	Author(s)	Feature Selection	Feature Extraction	Dataset Used	Key Findings
2015	Guo et al. [184]	Color and Texture	ODBTC, CCF, BPF	COREL, BRODATZ, VISTEX	1) Introduced organized variable block truncation coding (ODBTC) to generate the image content descriptor. 2) Utilized ODBTC to create two components: color co-occurrence feature (CCF) and bit pattern feature (BPF), both used for image retrieval.
2015	Kavitha H. et al. [185]	Texture	Edge and orientation	COIL-100	1) Presented a combined approach that incorporated both edge features and the distribution of angle directions of objects for CBIR.
2015	Lee et al. [175]	Descriptor and Color	SURF and Dominant color	Different types of images	1) Developed a novel image search strategy that involved analyzing images using a combination of SURF (SpeedUp Robust Feature) and DCD (Dominant Color Descriptor) techniques. 2) The mobile image search framework operated smoothly on iPhones and could easily locate natural color images.
2016	Nouman Ali et al [173]	Descriptors	SIFT and SURF	Corel	1) Combined SIFT and SURF for CBIR. 2) Addressed the issue of scale and rotation of images to some extent.
2018	Chauhan et al [157]	Color, Shape and Texture	K dominant colors, LBP	Different color images	1) Introduced a novel approach that employed K-dominant color for CBIR, combining color, shape, and texture features.

images, while recall quantifies the ratio of relevant images retrieved to the total number of relevant images in the database. Specificity focuses on the correct rejection of non-relevant images. F-measure balances precision and recall, providing a single score that incorporates both aspects of retrieval quality. Class accuracy is often employed for multi-class CBIR systems, assessing the accuracy of retrieval for each class independently. In addition to these metrics, evaluation methodologies like ROC curves and precision-recall curves provide a visual representation of system performance across different thresholds. These advancements have enabled a more standardized and unbiased assessment of CBIR methods, fostering meaningful comparisons and promoting advancements in the field. However, the choice of metrics should be aligned with the specific goals and requirements of the application to ensure a comprehensive evaluation of CBIR systems. Some of these metrics can be computed using the following formulas:

$$\text{Precision} = \frac{Tp}{Tp + Fp} \tag{1}$$

$$\text{Recall/Sensitivity} = \frac{Tp}{Tp + Fn} \tag{2}$$

$$\text{Specificity} = \frac{Tn}{Tn + Fp} \tag{3}$$

$$\text{F-measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Accuracy} = \frac{Tp}{Tp + Tn} \tag{5}$$

In these equations, Tp represents true positives, Fp represents false positives, Tn represents true negatives, and Fn represents false negatives. The abbreviations used provide a concise way to represent these evaluation metrics in the context of CBIR performance assessment.

However, in the field of image retrieval, precision and recall metrics may not always provide a comprehensive view of performance. To address this, researchers have introduced additional measures that are built upon precision and recall, such as Precision@K, Discounted Cumulative Gain (DCG), Normalized DGC (NDGC), etc.

Precision gives the proportion of retrieved similar images (compared to the query image) to the total number of retrieved images, without considering their arrangement. However, precision alone doesn't reveal whether these similar images are located at the beginning or scattered across the retrieval list.

One commonly used measure, Precision@K [210], is often employed to assess the performance of image retrieval methods. Precision@K quantifies the number of relevant images retrieved within the top K positions for a given query image. It can be calculated as:

$$\text{Precision@K} = \frac{S@K}{K} \tag{6}$$

Here, K represents the number of retrieved images up to the K^{th} position, and $S@K$ is the count of relevant (similar) images retrieved within those top K positions in response to the query image. In essence, Precision@K helps evaluate the quality of retrieval within a specific subset of the retrieved images, focusing on the top K positions.

Indeed, both precision and Precision@K lack the ability to differentiate whether the retrieved relevant images are concentrated at the top positions or scattered throughout the list. To address this limitation, the concept of DCG was introduced as a measure to assess the performance of the retrieval process in terms of the positions of the retrieved relevant images. DCG evaluates whether the relevant images retrieved are clustered at the beginning or distributed further down in positions.

In DCG, images that appear at the top positions are given higher importance, while those appearing at lower positions are penalized. The formula for calculating DCG at a specific position x is given by Equation 7:

$$DCG_x = \sum_{l=1}^{q=x} \frac{rel_l}{\log_2(l+1)} \tag{7}$$

Here, rel_l signifies the relevance of the retrieved image in relation to the query image, while q represents the position at which the image is found in the list of results.

However, DCG alone is unable to justify performance since the results can vary greatly based on the number of retrieved images. Therefore, the raw DCG is normalized to account for the total number of retrieved images, resulting in NDCG. NDCG is calculated as the ratio of DCG to the Ideal DCG (IDCG) at position p, as shown in Equation 8. IDCG represents the highest possible DCG when the retrieved images are ranked in descending order of their relevance.

$$NDCG = \frac{DCG_p}{IDCG_p} \tag{8}$$

NDCG provides a more comprehensive measure of retrieval performance by considering both the relevance of retrieved images and their positions, while also accounting for the effect of the number of retrieved images.

Certainly, the effectiveness of CBIR systems relies on the accuracy and speed of retrieving relevant images. Various techniques are employed for comparing the similarity between images. Some commonly used methods include Manhattan, Euclidean, Chi-Square, and Mahalanobis distance. The choice of evaluation parameter depends on the chosen coordinate system and manifold.

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. Additionally, some distance methods are centered around shortest algorithms, finding the shortest path between two points within a manifold. An example of such a method is the geodesic distance. The Minkowski distance calculates distances in N-Dimensional space. Table 4 presents the names

TABLE 4. Description of evaluation techniques used in CBIR.

Technique	Description
Manhattan Distance	Calculates distance between two points by summing absolute differences of Cartesian coordinates.
Euclidean Distance	Measures shortest distance between two points in N-dimensional space.
Chi-Square Distance	Statistical method for measuring similarity between feature metrics.
Mahalanobis Distance	Calculates distance between points in multivariate space considering correlations.
Geodesic Distance	Distance along manifold using shortest path algorithm.
Minkowski Distance	Measures similarity between two matrices in N-dimensional real space.
Bhattacharya Distance	Measures distinctiveness of classes based on standard deviation differences.
Hausdorff Distance	Measures similarity between overlapped objects.
Chessboard Distance	Also known as Chebyshev or L Infinity Distance, calculates longest distance on one axis.
Chamfer Distance	Iterative closest-point matching for similarity in segmented images.

TABLE 5. Summary of state-of-art approaches for image retrieval using supervised deep learning techniques.

Author(s)	Objective	Dataset Used	Deep Learning Model Applied
A. Krizhevsky etal. [189]	ImageNet Classification	ILSVRC-2010 and ILSVRC-2012	CNN
R. Xia et al. [205]	Image Retrieval	-	CNN Hashing (Feature Based)
Sun etal. [206]	Face Verification	Labeled Face in the Wild (LFW)	ConvNets
V. Erin Liong et al. [207]	Compact Binary Code Learning	-	Supervised Deep Hashing (SDH)
Karpathy & Fei-Fei [208]	I2DGA	Flickr8K, Flickr30K & MSCOCO M	CNN & RNN (Multi modal)
Li et al. [209]	Social Image Analysis	MIRFlickr and NUS-WIDE	DCE

and descriptions of some popular evaluation techniques frequently used in CBIR.

V. CBIR USING DEEP LEARNING AND FUTURE OUTLOOK

A. CBIR USING DEEP LEARNING

The exponential growth in the volume of images available on the web has introduced a significant challenge in locating identical or similar images in response to given query images. This challenge becomes even more intricate when relying on manual feature extraction techniques. To address this issue, deep learning techniques have emerged as a viable solution. In recent years, there has been a notable transition from manual feature extraction and representation methods to learning-based approaches, commonly referred to as deep learning methods [186]. These techniques facilitate the automatic learning of abstract features from the data itself.

Diverse architectures were proposed to cater to nature of the data being processed. For instance, artificial neural networks (ANNs) have proven effective for 1-dimensional data [187], [188], while convolutional neural networks (CNNs) are well-suited for image data. Recurrent neural networks (RNNs) find utility in time-series data analysis [191], [192]. The landscape of deep learning techniques applied to image retrieval has been enriched with a multitude of state-of-the-art approaches. These encompass a range of learning paradigms, including supervised learning, unsupervised learning, semi-supervised learning, self-supervised learning, and network-based learning utilizing architectures such as neural networks, convolutional networks, artificial networks, attention networks, Siamese networks, and triplet networks, among others.

Figure 10 visually illustrates the landscape of deep learning-based image retrieval methods. Some of these approaches tailor the use of deep learning techniques based on the specific nature of retrieval tasks, such as sketch-based, multi-label, object-based, and symmetric-based retrieval. This surge of interest and exploration has led to remarkable advancements in the field of image retrieval, with deep learning at its core. Table 5 provides an overview of recent cutting-edge approaches that employ supervised deep learning techniques for image retrieval. Furthermore, Figure 10 offers a classification of prominent deep learning-based image retrieval techniques into five categories: supervision-based, descriptor-based, network-based, retrieval-based, and miscellaneous.

Finally, it's important to note that the effectiveness of deep learning in image retrieval is closely tied to the availability of large-scale datasets. The selection of datasets aligns with the specific characteristics of the data type under consideration. Table 6 provides insights into various large-scale datasets utilized for deep learning-based image retrieval, detailing the year of introduction, image type, and the number of images (both testing and training) associated with each dataset.

B. FUTURE DIRECTIONS

This section outlines several significant trajectories that hold promise for the future of CBIR.

The performance of CBIR systems is undeniably influenced by the quality of images within the database. When images contain factors such as noise, poor visibility, or inadequate texture, CBIR systems can exhibit reduced accuracy and relevance in retrieving matching images. These challenges arise due to the distortion or loss of critical visual information, hindering CBIR's ability to accurately assess and compare images based on their content. As a result,

TABLE 6. Large scale dataset used in deep learning techniques.

Dataset	Year	Type of Images	Testing	Training
MNIST [193]	1998	Handwritten Digit Images	60,000	10,000
CIFAR-10 [194]	2009	Object Category Images	50,000	10,000
NUS-WIDE [195]	2009	Scene Images	97,214	65,075
SUN397 [196]	2010	Scene Images	397 100	754 8,000
MIRFlicker-1M [197]	2010	Scene Images	1 M	
SVHN [198]	2011	House Number Images	73257	26032
ILSVRC2012 [199]	2012	Object Category Images	1.2 M	50,000
UT-ZAP50K [200]	2014	Shoes Images	42,025	8,000
Yahoo-1M [201]	2015	Clothing Images	1,011,723	112,363
Fashion MNIST [202]	2017	Fashion Product images	60,000	10,000
Google Landmarks [203]	2017	Landmark Images	1 M	-
Google Landmarks v2 [204]	2020	Landmark Images	5 M	-

CBIR systems may struggle to effectively match queries with relevant images, leading to suboptimal performance.

In addition to image quality, the storage of CBIR data presents another set of challenges. The efficient management of image data is crucial for maintaining fast retrieval times and optimizing system resources. Storing large volumes of images requires careful consideration of storage structures, indexing methods, and retrieval algorithms. The data storage approach directly impacts retrieval speed, scalability, and resource utilization. Ensuring that images are organized, indexed, and stored in a manner that facilitates quick and accurate retrieval is a complex task. This involves finding a balance between storage efficiency and retrieval performance while accommodating various image types, sizes, and feature representations.

Addressing these challenges requires a multi-faceted approach that combines advancements in image enhancement techniques, feature extraction methods, and storage solutions. Researchers and practitioners continue to work on developing innovative strategies to enhance image quality, mitigate visibility issues, and effectively manage data storage for CBIR systems, ultimately aiming to improve the overall performance and reliability of these systems.

In the near future, several strategies could be employed to address the aforementioned challenges and enhance the performance of CBIR systems: Geng et al. [160] presented a hybrid CNN for image denoising, which may contribute to preprocessing noisy images to improve CBIR performance. Ahmad [212] introduced a deep image retrieval approach using ANN interpolation and similarity measurement-based indexing, which may advance CBIR's efficiency in retrieving relevant images. The work by Meng et al. [213] on visibility restoration using CNN can be adapted to enhance the visibility of query images in CBIR. Wang et al. [215] explored spiking neural networks, and their approach can potentially be incorporated into feature extraction or similarity measurement stages of CBIR. Hsiao and Chung [161] presented an AI-infused semantic model for question generation, which can be adapted to assist in formulating more effective queries in CBIR systems. Furthermore, Geng et al. [211] discussed scheduling strategies for automated storage and retrieval systems, which can provide insights into optimizing the retrieval process in CBIR systems by efficiently managing the storage and retrieval of image data. Incorporating these findings into CBIR research may led to enhanced performance and capabilities.

There are many other such pathways include delving into cross-modal retrieval, fostering collaboration across disciplines, further refining deep learning techniques for enhanced feature extraction, advancing personalized retrieval approaches, and tackling important ethical and privacy considerations. These potential future directions underscore CBIR's evolving nature and the opportunities it presents across various domains.

1) **Cross-Modal Retrieval:** Expanding CBIR to include different modalities such as text, audio, and 3D data will enable more comprehensive content retrieval and support complex queries involving multiple data types.
2) **Interdisciplinary Collaboration:** Collaborating across fields like computer vision, machine learning, and domain-specific knowledge will lead to more holistic CBIR solutions that address real-world challenges effectively.
3) **Deep Learning Advancements:** Further enhancing deep learning techniques, including refining network architectures and transfer learning strategies, will boost CBIR's feature extraction, representation, and matching capabilities.
4) **Personalized Retrieval:** Developing user-centric retrieval methods that adapt to individual preferences and contexts will enhance user satisfaction and the usability of CBIR systems.
5) **Ethical and Privacy Considerations:** Addressing privacy concerns, bias, and fairness issues associated with CBIR technologies is crucial to ensure their responsible and ethical deployment in various applications.

VI. CONCLUSION

As the digital landscape continually generates an overwhelming volume of images via online platforms and social media, the task of efficiently storing and retrieving relevant images remains a formidable endeavor. In response, CBIR has emerged as a prominent approach, enabling the retrieval of pertinent images from vast collections based on query images. This paper has provided a comprehensive survey of CBIR's fundamental elements, specifically focusing on feature selection, extraction, and representation. By delving into the significance and applications of CBIR, analyzing feature attributes, exploring extraction techniques, and considering recent advancements, this survey highlights the dynamic evolution of CBIR. Notably, the integration of deep learning techniques has expanded CBIR's horizons, leading to innovative methods and frameworks.

However, the efficacy of CBIR systems is inherently influenced by the quality of images stored within their databases. When images suffer from issues like noise, poor visibility, or inadequate texture, CBIR systems can experience reduced accuracy and relevance in retrieving appropriate matches. These challenges arise due to the alteration or absence of potential visual information, hindering CBIR's ability to precisely evaluate and compare images based on their content. Consequently, CBIR systems might encounter difficulties in connecting queries with fitting images, resulting in suboptimal performance.

Furthermore, the storage of CBIR data introduces another layer of complexity. Efficiently managing image data is pivotal for maintaining significant retrieval speeds and optimizing system resources. Dealing with large image volumes necessitates thoughtful deliberation on storage structures, indexing methodologies, and retrieval algorithms. The chosen data storage approach significantly affects retrieval speed, scalability, and resource utilization. Maintaining the balance between storage efficiency and retrieval performance while accommodating diverse image attributes is a multifaceted undertaking.

The future scope involves addressing these challenges through multimodal approaches that seamlessly integrate image enhancement techniques, feature extraction methods, and innovative storage solutions. Researchers will continue to strive to improve image quality, address visibility issues, and simplify data storage within CBIR systems. Together, these efforts will strengthen CBIR systems' performance, boosting their reliability and effectiveness.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare no conflict of interest. No human or animal subjects have been studied in this article. This article surveyed the content-based image retrieval methods and challenges.

REFERENCES

[1] A. Blaser, Ed., *Data Base Techniques for Pictorial Applications* (Lecture Notes in Computer Science), vol. 81. Florence, Italy: Springer, Jun. 1979.

[2] R. Jain, "NSF workshop on visual information management systems," *ACM SIGMOD Rec.*, vol. 22, no. 3, pp. 57–75, Sep. 1993.

[3] F. Long, H. Zhang, and D. D. Feng, "Fundamentals of content-based image retrieval," in *Multimedia Information Retrieval and Management*. Berlin, Germany: Springer, 2003, pp. 1–26.

[4] H. Hatano, "Image processing and database system in the national museum of western art; an integrated system for art research," *Inspel*, vol. 30, no. 3, pp. 259–267, 1996.

[5] P. L. Rosin, "Edges: Saliency measures and automatic thresholding," *Mach. Vis. Appl.*, vol. 9, no. 4, pp. 139–159, Feb. 1997.

[6] K. Siddiqi and B. B. Kimia, "Parts of visual form: Computational aspects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 3, pp. 239–251, Mar. 1995.

[7] A. Treisman, P. F. Cavanagh, B. Ramachandran, and R. V. S. von der Heydt, "Form perception and attention: Striate cortex and beyond," in *Visual Perception: The Neurophysiological Foundation*. Washington, DC, USA: American Psychological Association, 1990, pp. 273–316.

[8] S.-S. Chang and A. Hsu, "Image information systems: Where do we go from here?" *IEEE Trans. Knowl. Data Eng.*, vol. 4, no. 5, pp. 431–442, Oct. 1992.

[9] H. Tagare, F. M. Vos, C. C. Jaffe, and J. S. Duncan, "Arrangement: A spatial relation between parts for evaluating similarity of tomographic section," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 9, pp. 880–893, Sep. 1995.

[10] J. M. Corridoni, A. Del Bimbo, and P. Pala, "Image retrieval by color semantics," *Multimedia Syst.*, vol. 7, no. 3, pp. 175–183, May 1999.

[11] T. K. Lau and I. King, "Montage: An image database for the fashion, textile, and clothing industry in Hong Kong," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 1998, pp. 410–417.

[12] A. W. M. Smeulders, M. L. Kersten, and T. Gevers, "Crossing the divide between computer vision and data bases in search of image databases," in *Visual Database Systems 4 (VDB4)*. Boston, MA, USA: Springer, 1998, pp. 223–239.

[13] P. G. B. Enser, "Progress in documentation pictorial information retrieval," *J. Documentation*, vol. 51, no. 2, pp. 126–170, Feb. 1995.

[14] M. J. Swain, "Searching for multimedia on the world wide web," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst.*, vol. 1, Jun. 1999, pp. 32–37.

[15] A. Hiroike, Y. Musha, A. Sugimoto, and Y. Mori, "Visualization of information spaces to retrieve and browse image data," in *Proc. Int. Conf. Adv. Vis. Inf. Syst.* Berlin, Germany: Springer, 1999, pp. 155–163.

[16] G. Frederix, G. Caenen, and E. J. Pauwels, "Panoramic, adaptive and reconfigurable interface for similarity search," in *Proc. Int. Conf. Image Process.*, vol. 3, 2000, pp. 222–225.

[17] T. Kato, T. Kurita, N. Otsu, and K. Hirata, "A sketch retrieval method for full color image database-query by visual example," in *Proc. 11th IAPR Int. Conf. Pattern Recognit.*, vol. 1, 1992, pp. 530–533.

[18] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: The QBIC system," *Computer*, vol. 28, no. 9, pp. 23–32, 1995.

[19] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian image retrieval system, PicHunter: Theory, implementation, and psychophysical experiments," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 20–37, Jan. 2000.

[20] B. Huet and E. R. Hancock, "Line pattern retrieval using relational histograms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 12, pp. 1363–1370, Dec. 1999.

[21] D. Srivastava, B. Rajitha, S. Agarwal, and S. Singh, "Pattern-based image retrieval using GLCM," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 10819–10832, Aug. 2020.

[22] A. K. Jain and A. Vailaya, "Shape-based retrieval: A case study with trademark image databases," *Pattern Recognit.*, vol. 31, no. 9, pp. 1369–1390, Sep. 1998.

[23] D. Srivastava, R. Bakthula, and S. Agarwal, "Image classification using SURF and bag of LBP features constructed by clustering with fixed centers," *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 14129–14153, Jun. 2019.

[24] L. W. MacDonald, "Using color effectively in computer graphics," *IEEE Comput. Graph. Appl.*, vol. 19, no. 4, pp. 20–35, Jul./Aug. 1999.

[25] R. A. Weale, "Color in business, science and industry," *Brit. J. Ophthalmol.*, vol. 60, no. 3, p. 232, 1976.

[26] C.-S. Fuh, S.-W. Cho, and K. Essig, "Hierarchical color image region segmentation for content-based image retrieval system," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 156–162, Jan. 2000.

[27] T. Chen, L.-H. Chen, and K.-K. Ma, "Colour image indexing using SOM for region-of-interest retrieval," *Pattern Anal. Appl.*, vol. 2, no. 2, pp. 164–171, Jun. 1999.

[28] J. Alnihoud, "Content-based image retrieval system based on self organizing map, fuzzy color histogram and subtractive fuzzy clustering," *Int. Arab J. Inf. Technol.*, vol. 9, no. 5, pp. 452–458, 2012.

[29] N. Shrivastava and V. Tyagi, "Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching," *Inf. Sci.*, vol. 259, pp. 212–224, Feb. 2014.

[30] B. G. Prasad, K. K. Biswas, and S. K. Gupta, "Region-based image retrieval using integrated color, shape, and location index," *Comput. Vis. Image Understand.*, vol. 94, nos. 1–3, pp. 193–233, Apr. 2004.

[31] K. N. Plataniotis and A. N. Venetsanopoulos, *Color Image Processing and Applications*. Springer, 2013.

[32] R. G. Kuehni, "Industrial color difference evaluation," CIE, Vienna, Austria, CIE Publication 116, 1995.

[33] T. Young, "II. The Bakerian lecture. On the theory of light and colors," *Philos. Trans. Roy. Soc. London*, vol. 92, pp. 12–48, 1832.

[34] G. Wyszecki and W. S. Stiles, *Color Science*, vol. 8. New York, NY, USA: Wiley, 1982.

[35] C. A. Poynton, *A Technical Introduction to Digital Video*. Toronto, ON, Canada: Prentice-Hall, 1879. [Online]. Available: http://www.inforamp.net/-poynton/Poynton-Digital-Video.html

[36] J. Maxwell, "On the theory of three primary colours," in *The Scientific Papers of James Clerk Maxwell* (Cambridge Library Collection—Physical Sciences), W. Niven, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2016, pp. 445–450, doi: 10.1017/CBO9780511698095.025.

[37] [Online]. Available: https://it.kgsu.ru/C_GRAPH/c_gr005.html

[38] [Online]. Available: https://www.zekagraphic.com/16-most-used-graphic-design-terms-about-color/

[39] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Proc. Int. Conf. Image Process.*, vol. 2, 2002, pp. 11–15.

[40] W.-L. Chen, C.-D. Kan, C.-H. Lin, Y.-S. Chen, and Y.-C. Mai, "Hypervolemia screening in predialysis healthcare for hemodialysis patients using fuzzy color reason analysis," *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 1, 2017, Art.no. 1550147716685090, doi: 10.1177/1550147716685090.

[41] D. John and S. T. Tharani, "Content-based image retrieval using HSV-color histogram and GLCM," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 2, no. 1, pp. 4613–4617, 2014.

[42] S. Sergyn, "Content-based image classification," in *Proc. 5th Slovakian-Hungarian Joint Symp. Appl. Mach. Intell. Inform.*, Jan. 2007, pp. 427–434.

[43] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and H. Shin, "An efficient color representation for image retrieval," *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 140–147, Jan. 2001.

[44] M. J. Ballard, D. H. Swainfand, and M. Swain, "Indexing via color histograms," *IEEE Trans. Image Process.*, pp. 261—273, 1990.

[45] G. Pass and R. Zabih, "Comparing images using joint histograms," *Multimedia Syst.*, vol. 7, no. 3, pp. 234–240, May 1999.

[46] M. A. Stricker and M. Orengo, "Similarity of color images," *Proc. SPIE*, vol. 2420, pp. 381–392, Mar. 1995.

[47] H. Yu, M. Li, H.-J. Zhang, and J. Feng, "Color texture moments for content-based image retrieval," in *Proc. Int. Conf. Image Process.*, vol. 3, 2002, pp. 929–932.

[48] E. B. Goldstein and J. Brockmole, *Sensation and Perception*. Boston, MA, USA: Cengage Learning, 2016.

[49] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "Region-based image retrieval with high-level semantic color names," in *Proc. 11th Int. Multimedia Modelling Conf. (MMM)*, 2005, pp. 180–187.

[50] A. Mojsilovic, J. Kovacevic, J. Hu, R. J. Safranek, and S. K. Ganapathy, "Matching and retrieval based on the vocabulary and grammar of color patterns," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 38–54, Jan. 2000.

[51] J. R. Smith and S.-F. Chang, "Tools and techniques for color image retrieval," *Proc. SPIE*, vol. 2670, pp. 426–437, Mar. 1996.

[52] J. R. Smith and S.-F. Chang, "Querying by color regions using the VisualSEEk content-based visual query system," *Intell. Multimedia Inf. Retr.*, vol. 7, no. 3, pp. 23–41, 1997.

[53] J. R. Smith and S.-F. Chang, "VisualSEEk: A fully automated content-based image query system," in *Proc. 4th ACM Int. Conf. Multimedia*, 1996, pp. 87–98.

[54] T. S. Chua, K.-L. Tan, and B. C. Ooi, "Fast signature-based color-spatial image retrieval," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst.*, Jun. 1997, pp. 1–12.

[55] W. Hsu, S. T. Chua, and H. H. Pung, "An integrated color-spatial approach to content-based image retrieval," in *Proc. 3rd ACM Int. Conf. Multimedia*, 1995.

[56] A. Rao, R. K. Srihari, and Z. Zhang, "Spatial color histograms for content-based image retrieval," in *Proc. 11th Int. Conf. Tools With Artif. Intell.*, 1999, pp. 183–186.

[57] L. Cinque, G. Ciocca, S. Levialdi, A. Pellicanò, and R. Schettini, "Color-based image retrieval using spatial-chromatic histograms," *Image Vis. Comput.*, vol. 19, no. 13, pp. 979–986, Nov. 2001.

[58] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proc. 4th ACM Int. Conf. Multimedia*, 1996, pp. 65–73.

[59] K. Roy and J. Mukherjee, "Image similarity measure using color histogram, color coherence vector, and Sobel method," *Int. J. Sci. Res.*, vol. 2, no. 1, pp. 538–543, 2013.

[60] A. Al-Hamami and H. Al-Rashdan, "Improving the effectiveness of the color coherence vector," *Int. Arab J. Inf. Technol.*, vol. 7, no. 3, pp. 324–332, 2010.

[61] R. Ravani, M. R. Mirali, and M. Baniasadi, "Parallel CBIR system based on color coherence vector," in *Proc. 17th Int. Conf. Syst., Signals Image Process.*, 2010, pp. 519–521.

[62] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 762–768.

[63] J. Huang and R. Zabih, *Color-Spatial Image Indexing and Applications*. Ithaca, NY, USA: Cornell Univ., 1998.

[64] F. Mahmoudi, J. Shanbehzadeh, A.-M. Eftekhari-Moghadam, and H. Soltanian-Zadeh, "Image retrieval based on shape similarity by edge orientation autocorrelogram," *Pattern Recognit.*, vol. 36, no. 8, pp. 1725–1736, Aug. 2003.

[65] J. R. Smith and S.-F. Chang, "Single color extraction and image query," in *Proc. Int. Conf. Image Process.*, vol. 3, 1995, pp. 1–4.

[66] D. H. Ballard, *CM Brown Computer Vision*. Hoboken, NY, USA: Prentice-Hill, 1982.

[67] M. D. Levine, *Vision in Man and Machine*. New York, NY, USA: McGraw-Hill, 1985.

[68] Accessed: 2013. [Online]. Available: https://www.cs.auckland.ac.nz/~georgy/research/texture/thesis-html/node5.html

[69] T. R. Reed and J. M. H. Dubuf, "A review of recent texture segmentation and feature extraction techniques," *Comput. Vis. Image Understand.*, vol. 57, no. 3, pp. 359–372, May 1993.

[70] A. Materka and M. Strzelecki, "Texture analysis methodsa review," Inst. Electron., Tech. Univ. Lodz, Brussels, Belgium, Rep. COST B11, 1998, pp. 9–11.

[71] J. Zhang and T. Tan, "Brief review of invariant texture analysis methods," *Pattern Recognit.*, vol. 35, no. 3, pp. 735–747, Mar. 2002.

[72] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 7, no. 3, pp. 1–22, Apr. 2008.

[73] V. Tyagi, *Content-Based Image Retrieval—2017, Ideas, Influences, and Current Trends: Texture Feature*. Singapore: Springer, 2017, ch. 8.

[74] J. Chamorro-Martinez and P. Martinez-Jimenez, "A comparative study of texture coarseness measures," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1337–1340.

[75] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 6, pp. 460–473, Jun. 1978.

[76] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 7, pp. 722–733, Jul. 1996.

[77] N. Abbadeni, "Computational perceptual features for texture representation and retrieval," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 236–246, Jan. 2011.

[78] A. R. Rao and G. L. Lohse, "Identifying high level features of texture perception," *CVGIP: Graph. Models Image Process.*, vol. 55, no. 3, pp. 218–233, May 1993.

[79] Q. Iqbal and J. K. Aggarwall, "Applying perceptual grouping to content-based image retrieval: Building images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 1999, pp. 42–48.

[80] J. Chen, T. N. Pappas, A. Mojsilovic, and B. E. Rogowitz, "Adaptive perceptual color-texture image segmentation," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1524–1536, Oct. 2005.

[81] S. K. Saha, A. K. Das, and B. Chanda, "CBIR using perception based texture and colour measures," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 2, 2004, pp. 985–988.

[82] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 5, pp. 1264–1274, Sep./Oct. 1989.
[83] B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, "Inability of humans to discriminate between visual textures that agree in second-order statistics—Revisited," *Perception*, vol. 2, no. 4, pp. 391–405, Dec. 1973.
[84] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[85] S. Singh, D. Srivastava, and S. Agarwal, "GLCM and its application in pattern recognition," in *Proc. 5th Int. Symp. Comput. Bus. Intell. (ISCBI)*, Dubai, United Arab Emirates, Aug. 2017, pp. 20–25, doi: 10.1109/ISCBI.2017.8053537.
[86] K. I. Laws, "Textured image segmentation," Ph.D. dissertation, Univ. Southern California Los Angeles Image Processing INST, Tech. Rep. USCIPI-940, 1980.
[87] K. I. Laws, "Rapid texture identification," *Proc. SPIE*, vol. 238, pp. 376–381, Dec. 1980.
[88] K. I. Laws, "Goal-directed textured-image segmentation," *Proc. SPIE*, vol. 548, pp. 19–26, Apr. 1985.
[89] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.
[90] H.-C. Lin, L.-L. Wang, and S.-N. Yang, "Extracting periodicity of a regular texture based on autocorrelation functions," *Pattern Recognit. Lett.*, vol. 18, no. 5, pp. 433–443, May 1997.
[91] L. G. Brown and H. Shvaytser, "Surface orientation from projective foreshortening of isotropic texture autocorrelation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 6, pp. 584–588, Jun. 1990.
[92] J. J. van der Sanden and D. H. Hoekman, "Review of relationships between grey-tone co-occurrence, semivariance, and autocorrelation based image texture analysis approaches," *Can. J. Remote Sens.*, vol. 31, no. 3, pp. 207–213, Jan. 2005.
[93] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.
[94] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[95] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block LBP representation," in *Proc. Int. Conf. Biometrics*. Berlin, Germany: Springer, 2007, pp. 11–18.
[96] S. Murala, Q. M. J. Wu, R. Balasubramanian, and R. P. Maheshwari, "Joint histogram between color and local extrema patterns for object tracking," *Proc. SPIE*, vol. 8663, Mar. 2013, Art. no. 86630T.
[97] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with center-symmetric local binary patterns," in *Computer Vision, Graphics and Image Processing*. Berlin, Germany: Springer, 2006, pp. 58–69.
[98] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, Mar. 2009.
[99] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
[100] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local directional gradient pattern: A local descriptor for face recognition," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 1201–1216, Jan. 2017.
[101] R. Gupta, H. Patil, and A. Mittal, "Robust order-based methods for feature description," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 334–341, doi: 10.1109/CVPR.2010.5540195.
[102] L. Liu, Y. Long, P. W. Fieguth, S. Lao, and G. Zhao, "BRINT: Binary rotation invariant and noise tolerant texture classification," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3071–3084, Jul. 2014.
[103] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, Feb. 2010.
[104] S. Murala and Q. M. Jonathan Wu, "Local ternary co-occurrence patterns: A new feature descriptor for MRI and CT image retrieval," *Neurocomputing*, vol. 119, pp. 399–412, Nov. 2013.
[105] S. Zhao, Y. Gao, and B. Zhang, "Sobel-LBP," in *Proc. 15th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2008, pp. 2144–2147.
[106] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Directional local extrema patterns: A new descriptor for content based image retrieval," *Int. J. Multimedia Inf. Retr.*, vol. 1, no. 3, pp. 191–203, Oct. 2012.
[107] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, May 2012.
[108] M. Verma, B. Raman, and S. Murala, "Local extrema co-occurrence pattern for color and texture image retrieval," *Neurocomputing*, vol. 165, pp. 255–269, Oct. 2015.
[109] M. Verma and B. Raman, "Center symmetric local binary co-occurrence pattern for texture, face and bio-medical image retrieval," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 224–236, Oct. 2015.
[110] M. Dey, B. Raman, and M. Verma, "A novel colour- and texture-based image retrieval technique using multi-resolution local extrema peak valley pattern and RGB colour histogram," *Pattern Anal. Appl.*, vol. 19, no. 4, pp. 1159–1179, Nov. 2016.
[111] M. Verma and B. Raman, "Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 11843–11866, May 2018.
[112] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local gradient hexa pattern: A descriptor for face recognition and retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 171–180, Jan. 2018.
[113] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local quadruple pattern: A novel descriptor for facial image recognition and retrieval," *Comput. Electr. Eng.*, vol. 62, pp. 92–104, Aug. 2017.
[114] S. Murala and Q. M. J. Wu, "Local mesh patterns versus local binary patterns: Biomedical image indexing and retrieval," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 3, pp. 929–938, May 2014.
[115] S. Murala and Q. M. J. Wu, "Peak valley edge patterns: A new descriptor for biomedical image indexing and retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 444–449.
[116] S. Murala and Q. M. J. Wu, "MRI and CT image indexing and retrieval using local mesh peak valley edge patterns," *Signal Process., Image Commun.*, vol. 29, no. 3, pp. 400–409, Mar. 2014.
[117] H. Voorhees and T. Poggio, "Detecting textons and texture boundaries in natural images," in *Proc. 1st Int. Conf. Comput. Vis.*, vol. 59, 1987, pp. 364–367.
[118] D. Blostein and N. Ahuja, "Shape from texture: Integrating texture-element extraction and surface estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 12, pp. 1233–1251, Dec. 1989.
[119] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–10, Nov. 2012.
[120] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.
[121] S. W. Zucker, "Toward a model of texture," *Comput. Graph. Image Process.*, vol. 5, no. 2, pp. 190–202, Jun. 1976.
[122] K. S. Fu, "Syntactic pattern recognition," in *Applications*. Berlin, Germany: Springer, 1977.
[123] N. Ahuja, "Dot pattern processing using Voronoi neighborhoods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 3, pp. 336–343, May 1982.
[124] M. Tuceryan and A. K. Jain, "Texture segmentation using Voronoi polygons," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 2, pp. 211–216, Feb. 1990.
[125] S. Todorovic and N. Ahuja, "Texel-based texture segmentation," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 841–848.
[126] M. A. Georgeson, "Spatial Fourier analysis and human vision," in *Tutorial Essays in Psychology: Guide to Recent Advances* vol. 2, 1979, pp. 39–88.
[127] A. Rosenfeld and J. S. Weszka, "Picture recognition," in *Digital Pattern Recognition*. Berlin, Germany: Springer, 1980, pp. 135–166.
[128] F. Zhou, J. F. Feng, and Q. Y. Shi, "Texture feature based on local Fourier transform," in *Proc. Int. Conf. Image Process.*, vol. 2, 2001, pp. 610–613.
[129] S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on Gabor filters," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1160–1167, Oct. 2002.
[130] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[131] D. Zhang, A. Wong, M. Indrawan, and G. Lu, "Content-based image retrieval using Gabor texture features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 3656, pp. 13–15, 2000.
[132] D. Srivastava, S. Goel, and S. Agarwal, "Pipelined technique for image retrieval using texture and color," in *Proc. 4th Int. Conf. Power, Control Embedded Syst. (ICPCES)*, Mar. 2017, pp. 1–6.
[133] H. Müller, A. Rosset, J.-P. Vallee, and A. Geissbuhler, "Comparing features sets for content-based image retrieval in a medical-case database," *Proc. SPIE*, vol. 5371, pp. 99–109, Apr. 2004.
[134] J. C. Goswami and A. K. Chan, *Fundamentals of Wavelets: Theory, Algorithms, and Applications*, vol. 233. Hoboken, NJ, USA: Wiley, 2011.
[135] P. Porwik and A. Lisowska, "The Haar-wavelet transform in digital image processing: Its status and achievements," *Mach. Graph. Vis.*, vol. 13, nos. 1–2, pp. 79–98, 2004.
[136] N. Kingsbury, "The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement," in *Proc. 9th Eur. Signal Process. Conf. (EUSIPCO)*, Rhodes, Greece, vol. 811, Sep. 1998, pp. 1–4.
[137] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
[138] I. Daubechies, *Ten Lectures on Wavelets* (CBMS-NSF Regional Conference Series in Applied Mathematics), vol. 61. Philadelphia, PA, USA: SIAM, 1992.
[139] C. Vonesch, T. Blu, and M. Unser, "Generalized Daubechies wavelet families," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4415–4429, Sep. 2007.
[140] J.-M. Lina, "Image processing with complex Daubechies wavelets," *J. Math. Imag. Vis.*, vol. 7, Jun. 1997, Art. no. 211223.
[141] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Texture image retrieval using new rotated complex wavelet filters," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 35, no. 6, pp. 1168–1178, Dec. 2005.
[142] A. J. M. Traina, C. A. B. Castanon, and C. Traina, "MultiWaveMed: A system for medical image retrieval through wavelets transformations," in *Proc. 16th IEEE Symp. Comput.-Based Med. Syst.*, Jun. 2003, pp. 150–155.
[143] Y. M. R. Ro, M. K. Kim, H. K. K. Kang, B. S. M. Manjunath, and J. K. Kim, "MPEG-7 homogeneous texture descriptor," *ETRI J.*, vol. 23, no. 2, pp. 41–51, Jun. 2001.
[144] C. S. W. Won, D. K. P. Park, and S.-J.-P. Park, "Efficient use of MPEG-7 edge histogram descriptor," *ETRI J.*, vol. 24, no. 1, pp. 23–30, Feb. 2002.
[145] K.-L. Lee and L.-H. Chen, "An efficient computation method for the texture browsing descriptor of MPEG-7," *Image Vis. Comput.*, vol. 23, no. 5, pp. 479–489, May 2005.
[146] P. Wu, B. S. Manjunath, S. Newsam, and H. D. Shin, "A texture descriptor for browsing and similarity retrieval," *Signal Process., Image Commun.*, vol. 16, nos. 1–2, pp. 33–43, Sep. 2000.
[147] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognit.*, vol. 37, no. 1, pp. 1–19, Jan. 2004.
[148] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee, "Shape measures for content-based image retrieval: A comparison," *Inf. Process. Manag.*, vol. 33, no. 3, pp. 319–337, 1997.
[149] M. Yang, K. Kpalma, and J. Ronsin, "Shape-based invariant feature extraction for object recognition," in *Advances in Reasoning-Based Image Processing Intelligent Systems*. Berlin, Germany: Springer, 2012, pp. 255–314.
[150] D. Zhang and G. Lu, "A comparative study of Fourier descriptors for shape representation and retrieval," in *Proc. 5th Asian Conf. Comput. Vis.*, 2002, p. 35.
[151] D. Zhang and G. Lu, "A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval," *J. Vis. Commun. Image Represent.*, vol. 14, no. 1, pp. 39–57, Mar. 2003.
[152] C. Esperanca and H. Samet, "A differential code for shape representation in image database applications," in *Proc. Int. Conf. Image Process.*, vol. 1, 1997, pp. 1545–1559.
[153] K. Kpalma and J. Ronsin, "Multiscale contour description for pattern recognition," *Pattern Recognit. Lett.*, vol. 27, no. 13, pp. 1545–1559, Oct. 2006.
[154] S. M. Singh and K. Hemachandran, "Content-based image retrieval using color moment and Gabor texture feature," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, p. 299, 2012.
[155] J. Yue, Z. Li, L. Liu, and Z. Fu, "Content-based image retrieval using color and texture fused features," *Math. Comput. Model.*, vol. 54, nos. 3–4, pp. 1121–1127, Aug. 2011.
[156] X.-Y. Wang, Y.-J. Yu, and H.-Y. Yang, "An effective image retrieval scheme using color, texture and shape features," *Comput. Standards Interfaces*, vol. 33, no. 1, pp. 59–68, Jan. 2011.
[157] S. Chauhan, R. Prasad, P. Saurabh, and P. Mewada, "Dominant and LBP-based content image retrieval using combination of color, shape and texture features," in *Progress in Computing, Analytics and Networking*. Singapore: Springer, 2018, pp. 235–243.
[158] N. Shrivastava and V. Tyagi, "Region based image retrieval using integrated color, texture and shape features," in *Information Systems Design and Intelligent Applications*. New Delhi, India: Springer, 2015, pp. 309–316.
[159] J. Zhang, C.-W. Yoo, and S.-W. Ha, "ROI based natural image retrieval using color and texture feature," in *Proc. FSKD*, 2007, pp. 740–744.
[160] S. Geng, L. Wang, D. Li, B. Jiang, and X. Su, "Research on scheduling strategy for automated storage and retrieval system," *CAAI Trans. Intell. Technol.*, vol. 7, no. 3, pp. 522–536, 2022, doi: 10.1049/cit2.12066.
[161] I. H. Hsiao and C. Y. Chung, "AI-infused semantic model to enrich and expand programming question generation," *J. Artif. Intell. Technol.*, vol. 2, no. 2, pp. 47–54, 2022.
[162] M. Brown and D. G. Lowe, "Invariant features from interest point groups," in *Proc. BMVC*, vol. 4, 2002, pp. 398–410.
[163] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[164] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 404–417.
[165] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun./Jul. 2004, p. II.
[166] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.
[167] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, Art. no. 11501157.
[168] C. Schmid and K. Mikolajczyk, "A performance evaluation of local descriptors," in *Proc. ICPR*, 2004, vol. 60, no. 2, pp. 91–110.
[169] J. Wu, Z. Cui, V. S. Sheng, P. Zhao, D. Su, and S. Gong, "A comparative study of SIFT and its variants," *Meas. Sci. Rev.*, vol. 13, no. 3, 2013, Art. no. 122131.
[170] J. Luo and O. Gwun, "A comparison of SIFT, PCA-SIFT and SURF," *Int. J. Image Process.*, vol. 3, no. 4, pp. 143–152, 2009.
[171] B. B. Swapnali, S. K. Vijay, and K. HarpaleVarsha, "Feature extraction using surf algorithm for object recognition," *Int. J. Tech. Res. Appl.*, vol. 2, no. 4, pp. 197–199, 2014.
[172] D. Srivastava, B. Rajitha, and S. Agarwal, "An efficient image classification using bag-of-words based on SURF and texture features," in *Proc. 14th IEEE India Council Int. Conf. (INDICON)*, Dec. 2017, pp. 1–6.
[173] N. Ali, K. B. Bajwa, R. Sablatnig, S. A. Chatzichristofis, Z. Iqbal, M. Rashid, and H. A. Habib, "A novel image retrieval based on visual words integration of SIFT and SURF," *PLoS ONE*, vol. 11, no. 6, Jun. 2016, Art. no. e0157428.
[174] Q. B. Dang, V. P. Le, M. M. Luqman, M. Coustaty, C. D. Tran, and J-M. Ogier, "Camera-based document image retrieval system using local features–comparing SRIF with LLAH, SIFT, SURF and ORB," in *Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2015, pp. 1211–1215.
[175] Y.-H. Lee and Y. Kim, "Efficient image retrieval using advanced SURF and DCD on mobile platform," *Multimedia Tools Appl.*, vol. 74, no. 7, pp. 2289–2299, Apr. 2015.
[176] J. Han and K.-K. Ma, "Fuzzy color histogram and its use in color image retrieval," *IEEE Trans. Image Process.*, vol. 11, no. 8, pp. 944–952, Aug. 2002.
[177] Z.-C. Huang, P. P. K. Chan, W. W. Y. Ng, and D. S. Yeung, "Content-based image retrieval using color moment and Gabor texture feature," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, vol. 2, Jul. 2010.
[178] N. Shrivastava and V. Tyagi, "Multistage content-based image retrieval," in *Proc. CSI 6th Int. Conf. Softw. Eng. (CONSEG)*, Sep. 2012, pp. 1–4.
[179] N. Shrivastava and V. Tyagi, "An effective scheme for image texture classification based on binary local structure pattern," *Vis. Comput.*, vol. 30, no. 11, pp. 1223–1232, Nov. 2014.
[180] N. Shrivastava and V. Tyagi, "A review of ROI image retrieval techniques," in *Proc. 3rd Int. Conf. Frontiers Intell. Comput., Theory Appl. (FICTA)*. Cham, Switzerland: Springer, 2015, pp. 509–520.
[181] N. Chauhan and M. Goyani, "Enhanced multistage content-based image retrieval," *Int. J. Comput. Sci. Mobile Comput.*, vol. 2, no. 5, pp. 175–179, 2013.

[182] N. Shrivastava and V. Tyagi, "An efficient technique for retrieval of color images in large databases," *Comput. Electr. Eng.*, vol. 46, pp. 314–327, Aug. 2015.
[183] E. G. Karakasis, A. Amanatiadis, A. Gasteratos, and S. A. Chatzichristofis, "Image moment invariants as local features for content based image retrieval using the bag-of-visual-words model," *Pattern Recognit. Lett.*, vol. 55, pp. 22–27, Apr. 2015.
[184] J.-M. Guo and H. Prasetyo, "Content-based image retrieval using features extracted from halftoning-based block truncation coding," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1010–1024, Mar. 2015.
[185] H. Kavitha and M. V. Sudhamani, "Content-based image retrieval using edge and gradient orientation features of an object in an image from database," *J. Intell. Syst.*, vol. 25, no. 3, pp. 441–454, Jul. 2016.
[186] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. CVPR Workshops*, 2014, pp. 806–813.
[187] M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, nos. 14–15, pp. 2627–2636, 1998.
[188] F. Amato, A. Lopez, E. M. Pena-Mendez, P. Vanhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," *J. Appl. Biomed.*, vol. 11, no. 2, pp. 47–58, 2013.
[189] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.
[190] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[191] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Gated feedback recurrent neural networks," in *Proc. ICML*, 2015, pp. 2067–2075.
[192] M. Sundermeyer, R. Schluter, and H. Ney, "LSTM neural networks for language modeling," in *Proc. Conf. Int. Speech Commun. Assoc.*, 2012, pp. 194–197.
[193] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[194] A. Krizhevsky, "Learning multiple layers of features from tiny images," M.S. thesis, MIT, NYU, Univ. Tront, Toronto, ON, Canada, Apr. 2009.
[195] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "NUS-WIDE: A real-world web image database from National University of Singapore," in *Proc. ACM Int. Conf. Image Video Retr.*, Jul. 2009, p. 19.
[196] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3485–3492.
[197] M. J. Huiskes, B. Thomee, and M. S. Lew, "New trends and ideas in visual concept detection: The MIR Flickr retrieval evaluation initiative," in *Proc. ICMIR*, 2010, pp. 527–536.
[198] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *Proc. NIPS*, 2011, pp. 1–9.
[199] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[200] A. Yu and K. Grauman, "Fine-grained visual comparisons with local learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 192–199.
[201] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 27–35.
[202] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.
[203] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, "Large-scale image retrieval with attentive deep local features," in *Proc. ICCV*, 2017, p. 3456 3465.
[204] T. Weyand, A. Araujo, B. Cao, and J. Sim, "Google landmarks dataset v2—A large-scale benchmark for instance-level recognition and retrieval," in *Proc. CVPR*, 2020, p. 2575 2584.
[205] R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan, "Supervised hashing for image retrieval via image representation learning," in *Proc. AAAI*, 2014, pp. 1–7.
[206] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 1891–1898.
[207] V. E. Liong, J. Lu, G. Wang, P. Moulin, and J. Zhou, "Deep hashing for compact binary codes learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2475–2483.
[208] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3128–3137.
[209] Z. Li, J. Tang, and T. Mei, "Deep collaborative embedding for social image understanding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2070–2083, Sep. 2019.
[210] D. Srivastava, B. Rajitha, and S. Agarwal, "Content-based image retrieval for categorized dataset by aggregating gradient and texture features," *Neural Comput. Appl.*, vol. 33, no. 19, pp. 12247–12261, Oct. 2021, doi: 10.1007/s00521-020-05614-y.
[211] M. Zheng, K. Zhi, J. Zeng, C. Tian, and L. You, "A hybrid CNN for image denoising," *J. Artif. Intell. Technol.*, vol. 2, no. 3, pp. 93–99, Apr. 2022.
[212] F. Ahmad, "Deep image retrieval using artificial neural network interpolation and indexing based on similarity measurement," *CAAI Trans. Intell. Technol.*, vol. 7, no. 2, pp. 200–218, Jun. 2022, doi: 10.1049/cit2.12083.
[213] J. Meng, Y. Li, H. Liang, and Y. Ma, "Single image dehazing based on two-stream convolutional neural network," *J. Artif. Intell. Technol.*, vol. 2, no. 3, pp. 100–110, Jun. 2022.
[214] G. S. Vieira, A. U. Fonseca, N. M. Sousa, J. P. Felix, and F. Soares, "A novel content-based image retrieval system with feature descriptor integration and accuracy noise reduction," *Expert Syst. Appl.*, vol. 232, Dec. 2023, Art. no. 120774.
[215] X. Wang, M. Zhong, H. Cheng, J. Xie, Y. Zhou, J. Ren, and M. Liu, "SpikeGoogle: Spiking neural networks with GoogLeNet-like inception module," *CAAI Trans. Intell. Technol.*, vol. 7, no. 3, pp. 492–502, Sep. 2022, doi: 10.1049/cit2.12082.