Hindawi
Mathematical Problems in Engineering
Volume 2019, Article ID 9658350, 21 pages
https://doi.org/10.1155/2019/9658350

Review Article

Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review

Afshan Latif,[1] Aqsa Rasheed,[1] Umer Sajid,[1] Jameel Ahmed,[2] Nouman Ali,[1] Naeem Iqbal Ratyal,[3,4] Bushra Zafar,[5,6] Saadat Hanif Dar,[1] Muhammad Sajid,[3] and Tehmina Khalil[1]

[1]*Department of Software Engineering, Mirpur University of Science and Technology (MUST), Mirpur-10250 (AJK), Pakistan*
[2]*Department of Electrical Engineering, RIPHAH International University, Islamabad 75300, Pakistan*
[3]*Department of Electrical Engineering, Mirpur University of Science and Technology (MUST), Mirpur-10250 (AJK), Pakistan*
[4]*Department of Computer Systems Engineering, Mirpur University of Science and Technology (MUST), Mirpur-10250 (AJK), Pakistan*
[5]*Department of Computer Science, Government College University, Faisalabad 38000, Pakistan*
[6]*Department of Computer Science, National Textile University, Faisalabad 38000, Pakistan*

Correspondence should be addressed to Nouman Ali; nouman.ali@live.com

Received 10 April 2019; Revised 20 July 2019; Accepted 24 July 2019; Published 26 August 2019

Academic Editor: Marek Lefik

Multimedia content analysis is applied in different real-world computer vision applications, and digital images constitute a major part of multimedia data. In last few years, the complexity of multimedia contents, especially the images, has grown exponentially, and on daily basis, more than millions of images are uploaded at different archives such as Twitter, Facebook, and Instagram. To search for a relevant image from an archive is a challenging research problem for computer vision research community. Most of the search engines retrieve images on the basis of traditional text-based approaches that rely on captions and metadata. In the last two decades, extensive research is reported for content-based image retrieval (CBIR), image classification, and analysis. In CBIR and image classification-based models, high-level image visuals are represented in the form of feature vectors that consists of numerical values. The research shows that there is a significant gap between image feature representation and human visual understanding. Due to this reason, the research presented in this area is focused to reduce the semantic gap between the image feature representation and human visual understanding. In this paper, we aim to present a comprehensive review of the recent development in the area of CBIR and image representation. We analyzed the main aspects of various image retrieval and image representation models from low-level feature extraction to recent semantic deep-learning approaches. The important concepts and major research studies based on CBIR and image representation are discussed in detail, and future research directions are concluded to inspire further research in this area.

1. Introduction

Due to recent development in technology, there is an increase in the usage of digital cameras, smartphone, and Internet. The shared and stored multimedia data are growing, and to search or to retrieve a relevant image from an archive is a challenging research problem [1–3]. The fundamental need of any image retrieval model is to search and arrange the images that are in a visual semantic relationship with the query given by the user. Most of the search engines on the Internet retrieve the images on the basis of text-based approaches that require captions as input [4–6]. The user submits a query by entering some text or keywords that are matched with the keywords that are placed in the archive. The output is generated on the basis of matching in keywords, and this process can retrieve the images that are not relevant. The difference in human visual perception and manual labeling/annotation is the main reason for generating the output that is irrelevant [7–10]. It is near to impossible to apply the concept of manual labeling to

existing large size image archives that contain millions of images. The second approach for image retrieval and analysis is to apply an automatic image annotation system that can label image on the basis of image contents. The approaches based on automatic image annotation are dependent on how accurate a system is in detecting color, edges, texture, spatial layout, and shape-related information [11–13]. Significant research is being performed in this area to enhance the performance of automatic image annotation, but the difference in visual perception can mislead the retrieval process. Content-based image retrieval (CBIR) is a framework that can overcome the abovementioned problems as it is based on the visual analysis of contents that are part of the query image. To provide a query image as an input is the main requirement of CBIR and it matches the visual contents of query image with the images that are placed in the archive, and closeness in the visual similarity in terms of image feature vector provides a base to find images with similar contents. In CBIR, low-level visual features (e.g., color, shape, texture, and spatial layout) are computed from the query and matching of these features is performed to sort the output [1]. According to the literature, Query-By-Image Content (QBIC) and SIMPLicity are the examples of image retrieval models that are based on the extraction of low-level visual semantic [1]. After the successful implementation of the abovementioned models, CBIR and feature extraction approaches are applied in various applications such as medical image analysis, remote sensing, crime detection, video analysis, military surveillance, and textile industry. Figure 1 provides an overview of the basic concepts and mechanism of image retrieval [14–16].

The basic need for any image retrieval system is to search and sort similar images from the archive with minimum human interaction with the machine. According to the literature, the selection of visual features for any system is dependent on the requirements of the end user. The discriminative feature representation is another main requirement for any image retrieval system [17, 18]. To make the feature more robust and unique in terms of representation fusion of low-level visual features, high computational cost is required to obtain more reliable results [19, 20]. However, the improper selection of features can decrease the performance of image retrieval model [12]. The image feature vector can be used as an input for machine learning algorithms through training and test models and it can improve the performance of CBIR [1, 2]. A machine learning algorithm can be applied by using training-testing (either through supervised or through unsupervised) framework in both cases. The recent trends for image retrievals are focused on deep neural networks (DNN) that are able to generate better results at a high computational cost [21–23]. In this paper, we aim to provide a compressive overview of the recent research trends that are challenging in the field of CBIR and feature representation. The basic objectives of this research study are as follows: (1) How the performance of CBIR can be enhanced by using low-level visual features? (2) How semantic gap between the low-level image representation and high-level image semantics can be reduced? (3) How important is image spatial layout for image retrieval and representation? (4) How machine learning-based approaches can improve the performance of CBIR? (5) How learning can be enhanced by the use of deep neural networks (DNN)?

In this review, we have conducted a detailed analysis to address the abovementioned objectives. The recent trends are discussed in detail by highlighting the main contributions, and upcoming future challenges are discussed by keeping the focus of CBIR and feature extraction. The structure of the paper is as follow: Section 2 is about color feature, Section 3 is about texture features, Section 4 is about shape features, Section 5 is about spatial features, Section 6 is about low-level feature fusion, Section 7 is about local feature, commonly used dataset for CBIR and overview to basic machine learning techniques, Section 8 is about deep-learning-based CBIR, Section 9 is about feature extraction for face recognition, Section 10 is about distance measures, Section 11 is about performance evaluation criteria for CBIR and feature extraction techniques, while the last Section 12 points towards the possible future research directions.

2. Color Features

Color is considered as one of the important low-level visual features as the human eye can differentiate between visuals on the basis of color. The images of the real-world object that are taken within the range of human visual spectrum can be distinguished on the basis of differences in color [24–27]. The color feature is steady and hardly gets affected by the image translation, scale, and rotation [28–31]. Through the use of dominant color descriptor (DCD) [24], the overall color information of the image can be replaced by a small amount of representing colors. DCD is taken as one of the MPEG-7 color descriptors and uses an effective, compact, and intuitive format to narrate the indicative color distribution and feature. Shao et al. [24] presented a novel approach for CBIR that is based on MPEG-7 descriptor. Eight dominant colors from each image are selected, features are measured by the histogram intersection algorithm, and similarity computation complexity is simplified by this.

According to Duanmu [25], classical techniques can retrieve images by using their labels and annotation which cannot meet the requirements of the customers; therefore, the researchers focused on another way of retrieving the images that is retrieving images based on their content. The proposed method uses a small image descriptor that is changeable according to the context of the image by a two-stage clustering technique. COIL-100 image library is used for the experiments. Results obtained from the experiments proved that the proposed method to be efficient [25].

Wang et al. [26] proposed a method based on color for retrieving image on the basis of image content, which is established from the consolidation of color and texture features. This provides an effective and flexible estimation of how early human can process visual content [26]. The fusion of color and texture features offers a vigorous feature set for color image retrieval approaches. Results obtained from the experiments reveal that the proposed method retrieved images more accurately than the other traditional methods.

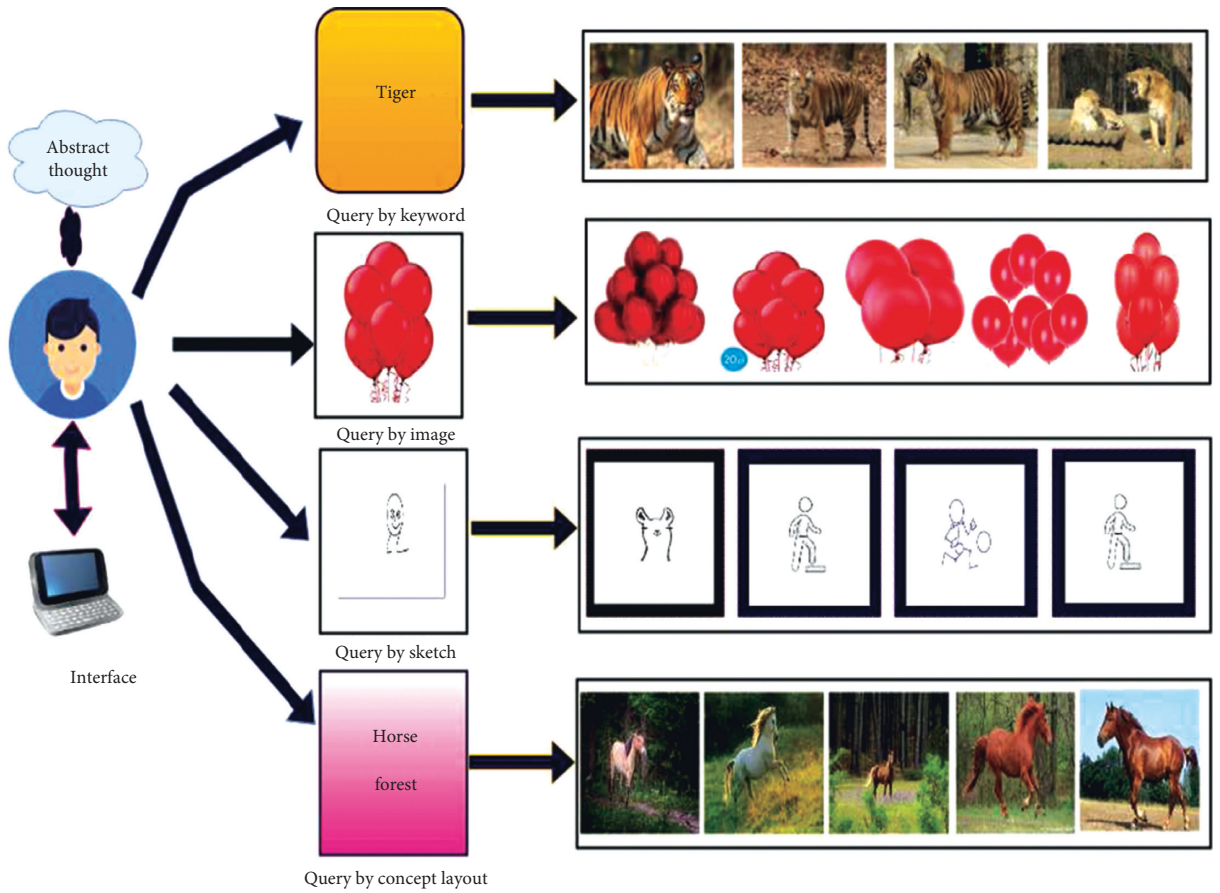

Figure 1: Pictorial representation of different concepts of image retrieval [6].

However, the feature dimensions are not higher than other approaches and require a high computational cost. A pairwise comparison for both low-level features is used to calculate similarity measure which could be a bottleneck [26].

Various research groups carried out a study on the completeness property of invariant descriptors [27]. Zernike and pseudo-Zernike polynomials which are orthogonal basis moment functions can represent the image by a set of mutually independent descriptors, and these moment functions hold orthogonality and rotation invariance [27]. PZMs proved to be more vigorous to image noise over the Zernike moments. Zhang et al. [27] presented a new approach to derive a complete set of pseudo-Zernike moment invariants. The link between pseudo-Zernike moments of the original image and the same shape but distinct orientation and scale images is formed first. An absolute set of scale and rotation invariants is obtained from this relationship. And this proposed technique proved to be better in performance in recognizing pattern over other techniques [27].

Guo et al. [28] proposed a new approach for indexing images based on the features extracted from the error diffusion block truncation coding (EDBTC). To originate image feature descriptor, two color quantizers and a bitmap image using vector quantization (VQ) are processed which are produced by EDBTC. For assessing the resemblance between the query image and the image in the database, two features Color Histogram Feature (CHF) and Bit Pattern Histogram Feature (BHF) are introduced. The CHF and BHF are calculated from the VQ-indexed color quantizer and VQ-indexed bitmap image, respectively. The distance evaluated from CHF and BHF can be used to assess the likeliness between the two images. Results obtained from the experiments show that the proposed scheme performs better than former BTC-based image indexing and other existing image retrieval schemes. The EDBTC has good ability for image compression as well as indexing images for CBIR [28].

Liu et al. [29] proposed a novel method for region-based image learning which utilizes a decision tree named DT-ST. Image segmentation and machine learning techniques are the base of this proposed technique. DT-ST controls the feature discretization problem which frequently occurs in contemporary decision tree learning algorithms by constructing semantic templates from low-level features for annotating the regions of an image. It presents a hybrid tree which is good for handling the noise and tree fragmentation problems and reduced the chances of misclassification. In semantic-based image retrieval, the user can query image through both labels and regions of images. Results obtained from the experiments conducted to check the effectiveness of the proposed technique reveal that this technique provides higher retrieval accuracy than the traditional CBIR techniques and the semantic gap between low- and high-level

features is reduced to a significant level. The proposed technique performs well than the two effectively set decision tree induction algorithms ID3 and C4.5 in image semantic learning [29]. Islam et al. [30] presented a supreme color-based vector quantization algorithm that can automatically categorize the image components. The new algorithm efficiently holds the variable feature vector like the dominant color descriptors than the traditional vector quantization algorithm. This algorithm is accompanied by the novel splitting and stopping criterion. The number of clusters can be learned, and unnecessary overfragmentation of region clusters can be avoided by the algorithm through these criteria.

Jiexian et al. [31] presented a multiscale distance coherence vector (MDCV) for CBIR. The purpose behind this is that different shapes may have the same descriptor and distance coherence vector algorithm may not completely eliminate the noise. The proposed technique first uses the Gaussian function to develop the image contour curve. The proposed technique is invariant to different operations like translation, rotation, and scaling transformation.

2.1. Summary of Color Features. There are various low-level color features, and the performance of color moments is not good as it can represent all the regions of the image. Histogram-based color features require high computational cost while DCD performs better for region-based image retrieval and is computationally less expensive due to low dimensions. A detailed summary of the abovementioned color features [24–31] is represented in Table 1.

3. Texture Features

Papakostas et al. [32] performed their experiments on four datasets, namely, COIL, ORL, JAFFE, and TRIESCH I in order to show the discrimination power of the wavelet moments. These datasets are divided into 10, 40, 7, and 10 classes. For the evaluation of the proposed model (WMs), two different configurations of wavelets WMS-1 and WMs-2 are used where the former uses cubic B-spline and the other uses the Mexican hat mother wavelets. By keeping only effective characteristics in feature selection approach greatly improves the classification capabilities of the wavelet moments. The performance of the proposed model is compared with Zernike, pseudo-Zernike, Fourier–Mellin, and Legendre and with two others by using 25, 50, 75, and 100 percent of the entire datasets, and each moment family behaves differently in each dataset. Classification performance of the moment descriptors shows the better results of the proposed model (wavelet Moments and moment invariants). For the evaluation of the proposed model (MSD) for image retrieval, Liu et al. [33] perform experiments on Corel datasets as there are no specific datasets for content-based image retrieval (CBIR). Corel-5000 and Corel-10000 are used with 15000 images, and HSV, RGB, and Lab color space are used to evaluate the retrieval performance. On both datasets Corel-5000 and Corel-10000, the average retrieval and recall rates of the proposed model using different color quantization level and texture orientation quantization levels are evaluated and our proposed model performs better on HSV and Lab color space and poor on RGB color space. For getting good results between storage space, retrieval accuracy, and speed, 72 color and orientation quantization levels are used in MSD and 6 for image retrieval. The average retrieval and recall ratios of MSD are compared with other methods like Gabor MTH on Corel datasets because these algorithms are developed for image retrieval for the evaluation of MSD and the results show that our proposed model (MSD) outperforms other models.

10,000 color images [34] were collected from public resources of natural scenes such as landscapes, peoples, and textures in order to perform their experiments for image retrieval based on texture. Generally, for retrieval results, properties such as smoothness, regularity, distribution, and coarseness are considered while used additionally the color information with these properties. The precision comparison between the proposed model (color co-occurrence matrix) and the gray-level co-occurrence matrix method provides results to evaluate the proposed model. The comparison shows that the color co-occurrence matrix is better than the gray-level co-occurrence matrix because of the additionally added property (color information). For CBIR [35], Corel, COIL, and Caltech-101 datasets (those datasets are chosen that have images grouped in the form of semantic concepts) containing 10908, 7200 images, and 101 image categories for respective datasets are used. The mean precision and recall rates obtained by the proposed method (embedded neural network with bandlet transform) on top 20 retrievals are compared with the other standard and with the state-of-the-art retrieval systems. The mean precision and recall rates obtained by the proposed method are 0.820 and 0.164 on top 20 retrievals. These results show that the research presented in [35] clearly outperformed other models in terms of mean precision value and recall rate.

With Corel image gallery containing 10900 images for categorical image retrieval, Irtaza and Jaffar [36] conducted experiments to show the effectiveness of the proposed model (SVM-based architecture; Figure 2 represents an example of binary classification while using SVM). The Corel image gallery is divided into two sets Corel A having 1000 images that are divided into ten categories and Corel B that has 9900 images. The mean precision and recall rates obtained by the proposed method on the top 20 retrievals are compared with other standard retrieval systems. Different numbers of returned images are used to show the retrieval capacity of SVM and it shows consistent results. Thus, the results and comparison show that the proposed model has better results and is more consistent in image retrieval. Fadaei et al. [38] performed experiments on Brodatz and Vistex datasets for content-based image retrieval containing 112 grayscale and 54 color images for respective datasets. The distance between the query image and dataset image is calculated, images that have minimum distance are retrieved, and then the precision and recall rates are calculated. The results of the proposed models are compared with other prior methods. The retrieval time of Brodatz is longer than that of Vistex database because Brodatz has more images than Vistex; thus, it needs

Table 1: A summary of the performance of color features.

Author	Application	Method	Dataset	Accuracy
Duanmu [25]	Image retrieval	Color moment invariant	COIL-100	0.985
Wang et al. [26]	Content-based image retrieval	Integrated color and texture features	Corel	0.613
Zhang et al. [27]	Object recognition	Complete set of pseudo-Zernike invariants	COIL-100	—
Guo et al. [28]	Content-based image retrieval	Error diffusion block truncation coding features	Corel	0.797
Shao et al. [24]	Image retrieval	MPEG-7 dominant color descriptor	Corel	0.8964
Liu et al. [29]	Region-based image retrieval	High-level semantics using decision tree learning	Corel	0.768
Islam et al. [30]	Automatic categorization of image regions	Dominant color-based vector quantization	Corel	0.9767
Jiexian et al. [31]	Content-based image retrieval	Multiscale distance coherence vector algorithm	MPEG-7 image database	0.97

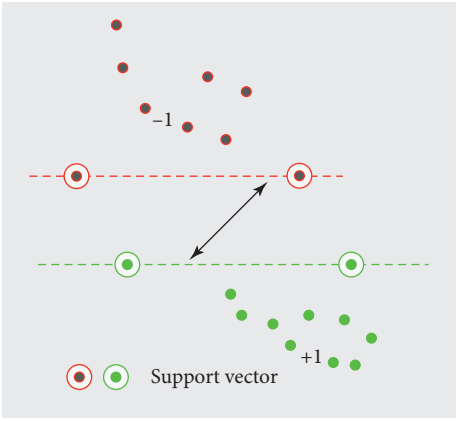

Figure 2: Example of SVM-based classification [37].

time for more feature matching and processing. The dimension of the feature vector for the proposed model is 3124 which is higher than that of other methods. Retrieving time of the proposed model is slower in feature matching and faster in feature extraction although the dimension of the feature vector is high. Comparison and results show that the proposed model (LDRP) has better performance and average precision rates and is faster in feature extraction and slower in feature matching.

3.1. Summary of Texture Features. There are various low-level texture features and they can be applied in different domains of image retrieval. As they represent a group of pixel, therefore they are semantically more meaningful than color features. The main drawback of texture features is the sensitivity to image noise and their semantic representation also depends on the shapes of objects in the images. A detailed summary of the abovementioned texture features [32–36, 38, 39] is represented in Table 2.

4. Shape Features

Shape is also considered as an important low-level feature as it is helpful in identification of real-world shapes and objects. Zhang and Lu [15] presented a comprehensive review of the application of shape features in the domain of image retrieval and image representation. Region-based and contour-based are the main classifications of shape features [14]. Figure 3 presents a basic overview of the classification of shape features. Trademark-based image retrieval [41] is one of the specific domains where shape features are used for image representation.

5. Spatial Features

Image spatial features are mainly concerned with the locations of objects within the 2D image space. The Bag of Visual Words (BoVW) [42] is one of the popular frameworks that ignore image spatial layout while representing the image as a histogram. Spatial Pyramid Matching (SPM) [43–45] is reported as one of the popular techniques that can capture image spatial attributes but is insensitive to scaling and rotations. Zafar et al. [46] presented a method to encode the respective spatial information for representing the histogram of the BoVW model. This is initiated by the calculation of the universal geometric correlation between the sets of similar visual words corresponding to the center of the image. Five databases are used for assessing the performance of the proposed scheme based on respective spatial information. Ali et al. [47] proposed Hybrid Geometric Spatial Image Representation (HGSIR) by using image classification-based framework. The base of this is the compound of different histograms calculated for the rectangular, triangular, and circular areas of the images. To assess how well the presented approach performs, five datasets are used for this. And the results show that this research performs better than the state-of-the-art methods concerning how accurately images are classified. In another research, Zafar et al. [48] presented a novel technique for representing images that includes the spatial information to the reversed index of the BoVW model. The spatial information is attached by computing the universal corresponding spatial inclination of visual words in a gyration-invariant fashion. The geometric correlation of similar visual words is calculated. This is done by computing an orthogonal vector corresponding to every single point in the triplets of similar visual words. The histogram of visual words is

Table 2: A summary of the performance of texture features.

Authors	Datasets	Purpose	Model	Performance/accuracy
Papakostas et al. [32]	COIL, ORL, JAFFE, TRIESCH I	Wavelet moments and their corresponding invariants in machine vision system	Wavelet moments and moment invariants	Classification performances on (100%) percent of entire data are 0.3083, 0.2425, 0.1784, and 0.1500, respectively, for datasets
Wang et al. [34]	Corel-1000 and Corel-10000	Image retrieval	SED	Similarity between query image and image database is 3.9198, 9.92209, and 8.86239 for dragons, busses, and landscapes, and there will be high precision rate when the query image has noteworthy regions or texture
Liu et al. [33]	Corel datasets (Corel-5000 and Corel-10000)	Image retrieval	MSD	Average retrieval precision and recall ratios on Corel-5000 and Corel-10000 are 55.92%, 6.71% and 41.44%, 5.48%
Lasmar and Berthoumieu [40]	Vistex, Brodatz, ALOT	Texture image retrieval	GC-MGG and GC-MWbl	Improvement in average retrieval rate on Brodatz (EB2) by our model is 6.86% and 5.23%, respectively, with Daubechies filter db4 and dual-tree complex wavelet transform
Fadaei et al. [38]	Brodatz and Vistex	Content-based image retrieval	LDRP	80.81% and 91.91% are average precision rates of the first-order LDRP($P = 6$, $K = 4$) for the respective datasets

computed based on the size of orthogonal vectors that provides information about the respective position of the linear visual words. For the evaluation of the presented method, four datasets are used. Ali [49] proposed two techniques for representing the images. The base of these techniques is the histogram of triangles that incorporates the spatial information to the reversed index of BoF representation. An image is divided into two or four triangles which are assessed individually for calculating the histograms of triangles for two levels: level 1 and level 2. Two datasets are used for evaluating the results of the presented technique. Experimental results show that the proposed technique performs well while retrieving images.

Khan et al. [50] proposed PIW (Pairs of Identical visual word, the set of all pairs of VWs of the same type) to represent global spatial distribution (histogram orientation of segments formed by PIW). Khan et al. [50] just considered relationships among similar visual words so histograms that are produced by each word type compose powerful details of intratype visual words relationships. The advantages of this approach over others are as follows: it enables infusion of global information, powerful geometric transformation, efficient extraction of spatial information, reduces complexity, and improves classification rate by adding distinguishing information. Anwar et al. [51] presented a model by using symbol recognition (symbol recognition is performed by using scale-invariant feature transform-based BoVW). To add spatial information to BoVW, circular tilings are used and modify angles histograms of an existing method (proposed by Rahat) to make them rotation invariant as they are not rotation invariant before. Then these modified angles are merged with circular tilings which get an increased rate of classification and it reduces the computation complexity. Anwar et al. [52] performed experiments on various datasets belonging to different categories (as they have different backgrounds) to verify the proposed model and to verify rotation invariant of images in coins; authors rotated coin images to an extreme extent. Khan et al. [53] proposed a global and local relative spatial distribution of visual words over an image named soft pairwise spatial angle-distance histogram to include distance and angle information of visual words. The aim is to provide efficient representation capable of adding relative spatial information and by performing experiments on classification tasks on MSRC-2, 15Scene, Caltech-101, Caltech-256, and Pascal VOC 2007 datasets, so authors concluded that the proposed method performs well and improves overall performance. In order to acquire rotation invariance efficiently, Ali et al. [54] proposed to represent global spatial distribution by constructing histograms based on the computation of the orthogonal vector between PIWs. For the evaluation of the presented method, three satellite scene datasets are used.

6. Low-Level Feature Fusion

Ashraf et al. [55] presented a CBIR model that is based on color and discrete wavelet transform (DWT). For the retrieval of similar images, the low-level feature color, texture,

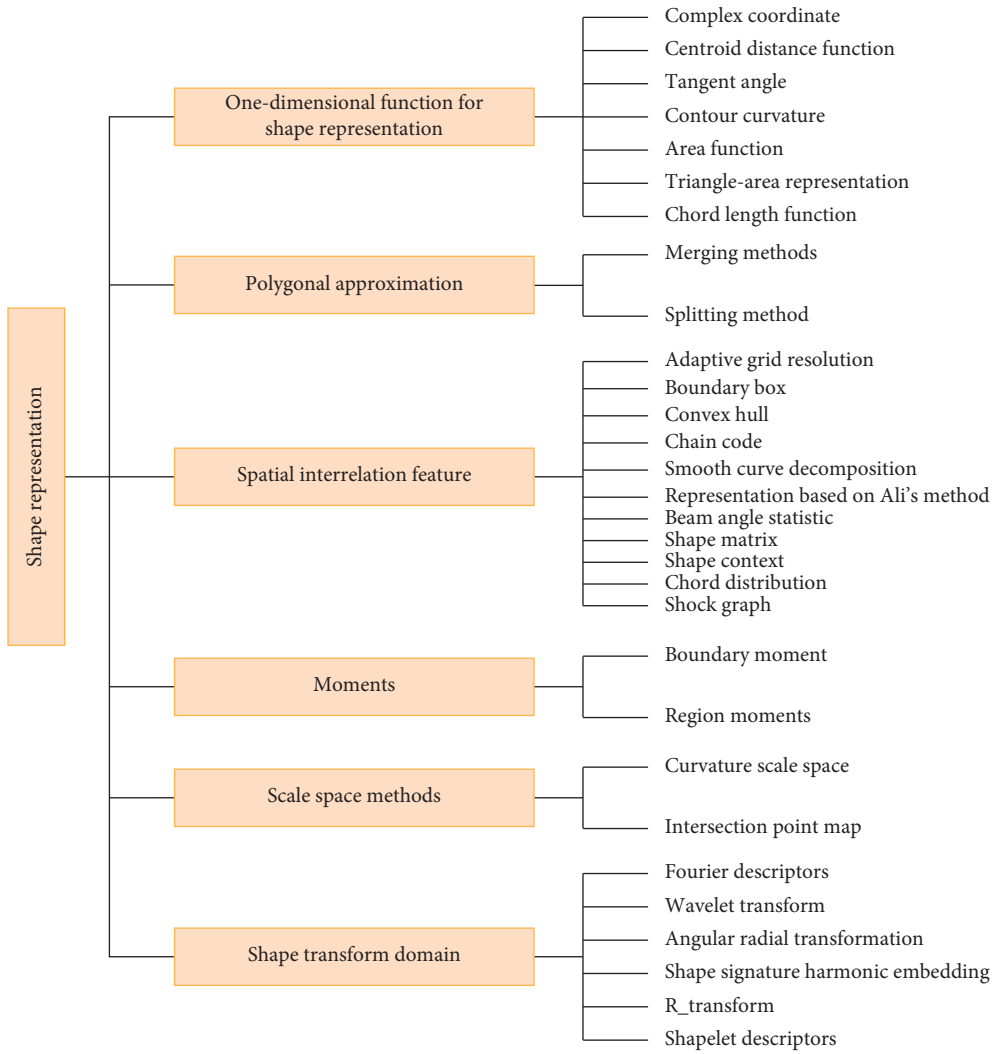

Figure 3: An overview of shape-based feature extraction approaches [14, 15].

and shape are used. These features play a significant role in the retrieval process. Different types of features and feature extraction technique are discussed and scenarios in which feature extraction technique is good are explained [55]. To prepare the eigenvector information from the image [55], color edge detection and discrete wavelet approaches are used. The color space RGB and YCbCr are used to extract the color features. The researchers in [55] transformed RGB images to YCbCr color space to extract the meaningful information. The YCbCr transformation is selected in this case because the human visual system can view different colors and brightness sensitivity. In YCbCr, the Y represents the luminance while the color is represented by Cb and Cr. The output of YCbCr is dependent on two factors, while in case of RGB, the output image is dependent on the intensity of R, G, and B, respectively. The YCbCr color space is also used to solve the color vibration problem. To extract the edge features, the Canny edge detector is used. The viewfinder ensures that this special feature responds to the opponent and then provides the best shape in any size. In order to retrieve the query image, the color and edge-based features are extracted to compute the feature vector. If there is a small distance between the query image and repository image, the correlated image from the database is selected to match with the image that is passed in query. To reduce the computational steps and enhance the search, the color features are also incorporated with histogram and the Haar Wavelet transform was applied. And then for image retrieval, the artificial neural network (ANN) is applied; then, its performance is measured against the existing CBIR system. The result shows that this method has a better performance than the others [55].

Ashraf et al. [56] presented a new CBIR technique that uses the combination of color and texture features to extract the local vector which is used as a featured vector. Color moments are used to extract the color feature, and for the texture feature, the discrete wavelet transform and Gabor wavelet methods are used. To enhance the feature vector color and edge, directory descriptor is also used in the feature vector. Then, this method is compared with all other existing CBIR methods and good performance is achieved [56] in terms of precision and recall values.

Mistry et al. [57] conducted a study on CBIR by using hybrid features and various distance metrics. In this paper,

the hybrid features combine three different features descriptors which consist of spatial features, frequency, binarized statistical image features (BSIF), and color and edge directivity descriptors (CEDD). Features are extracted by using BSIF, CEDD, HSV color histogram, and color moment. Features that are extracted by using HSV histogram contain color quantization and color space conversion and histogram computation. Feature extraction by using the BSIF includes conversion of RGB to grayscale image and patch selection from grayscale image. It also includes subtraction of mean value from the components. Feature extraction by using the CEDD process includes HSV color two-stage fuzzy linking system. Feature extraction using the color moment process first converts the RGB into its component and then finds out the mean and standard deviation for each component. The stored features are then compared with the query image feature vector. Minimum distance by using the distance classifiers results in the comparison and then the image is retrieved. Different experiments are performed on that approach, and the results show that this approach significantly performs better than the existing methods [57].

Ahmed et al. [58] conducted a study on CBIR by using image feature information fusion. In this technique, the fusion between the extracted spatial color features with shape features extracted and object recognition takes place. Colors with shape together can differentiate the object more accurately. Spatial color feature in the feature vector increases the retrieval of the image. In the proposed method, RGB color is used to extract the color feature while the gray-level images are used to extract the object edges and corner in the formation of shape. The detection of corner and edges from the shape creates more powerful descriptor. Shape detection conforms the better understanding of object or image. Shape image detection on the basis of edges and corner formation combining with the color produces more accurate result for retrieval or detection of image. For selecting the high variance component, the dimension reduction takes place on the feature vector. Then, the compact data features are the input of Bag of Word (BoW) for quick indexing or retrieval of image. The results of the experiment performed based on this technique show that it outperforms the existing CBIR technique [58].

Liu et al. [59] proposed a method for classifying and searching an image by fusing the local base pattern (LBP) and color information feature (CIF). For deriving the image descriptor, the LBP extracts the textural feature. But the LBP has not good performance for the color feature descriptor. Both the color feature and textural feature are used for the efficient retrieval of the color image from a large set of database. In this proposed method, a new color feature CIF with the LBP-based feature is used for image retrieval as well as for classification. CIF and LBP both together represent the color and textural information of an image. Several experiments are performed using a large set of database, and the results show that this method has good performance for retrieval and classification of the images [59].

Zhou et al. [60] conducted a study on collaborative index embedding. This work explores the potential of unifying indexing of SIFT feature and the deep convolutional neuron network (d-CNN) for the retrieval of image. To check the shared image-level neighborhood structure and to implicitly integrate the CNN and SIFT features, index the collaborative index embedding algorithms proposed which continuously update the index file of CNN and SIFT features. After continuous iteration of the embedding index, the CNN embedded index is used for the online query, which shows the efficient retrieval accuracy with 10 percent more than the original CNN and SIFT index. The results of the extensive experiment performed based on this method show that it achieves higher performance in the retrieval [60].

Li et al. [61] studied on the color texture feature image which is based on the Gaussian copula model of Gabor wavelets. He proposed an efficient method for the retrieval of the image in the color and texture context by using the Gaussian copula model which is based on Gabor wavelets. Gabor filter is considered as a linear filter which is used for signal analysis. Orientation and the frequency representation of Gabor filter are resembled with the human visual system and it is particularly used for texture image retrieval and the copula model is used to capture the dependence structure in the variable where dependencies exist. Gabor wavelets are used to decompose the color image; after decomposition, three types of dependencies exist in decomposed subbands of Gabor wavelet. These three dependencies are directional dependence, color dependence, and scale dependence. After the decomposition, existence dependencies are analyzed and captured by using the Gaussian copula method. There are three types of schemes developed for Gaussian copula, and accordingly, four Kullback–Leibler distances (KLD) are introduced for color retrieval image. Several experiments are performed using the datasets ALOT and STex, and the results show that it performs better than the several state-of-the-art retrieval methods [61].

Bu et al. [62] studied on CBIR by using color and texture features by combining the color and texture features extracted from the image using Multi-Resolution Multi-Directional (MRMD) filters. MRMD filters are used as simple and it can be independent to low- and high-frequency features, and it produces efficient multiresolution multidirectional analyses. HSV color space is used as its characteristics are very close to the human visual system. Local and global features are extracted from the domain of low- and high-frequency in each color space. Several experiments are performed by comparing the precision VS recall of the retrieval and the feature dimension vector. The results show that this method has significant improvement over the existing techniques [62]. A detailed summary of the abovementioned low-level feature fusion for CBIR is represented in Table 3.

Nazir et al. [63] conducted a study on CBIR by fusing the color and texture features. Since retrieving the image from a large set of databases is a challenging task, researchers proposed many techniques to overcome this challenge. Nazir et al. [63] used both the color and texture features to retrieve the image. The previous research shows that by retrieving the image using a single feature does not provide good results and using multiple features for image retrieval

Table 3: A summary of the performance of fusion feature-based approaches for CBIR.

Author	Dataset	Images/classes	Techniques	Applications	Precision
Nazir et al. [63]	Corel 1-K	1000 images which are divided into 10 classes	HSV color histogram, discrete wavelet transform, and edge histogram descriptor	Content-based image retrieval	0.735
Ashraf et al. [56]	Corel 1000	It contains 10 categories. Each category contains 100 images with different size	Multimedia data for content-based image retrieval by using multiple features	Content-based image retrieval	0.875
Mistry et al. [57]	Wang	Dataset contains 1000 images from 10 different classes	Hybrid features and various distance metric	Content-based image retrieval	0.875
Ahmed et al. [58]	Corel-1000	Dataset contains 1000 image splitted into 10 categories. Each category consists of 100 images	Image features information fusion	Content-based image retrieval	For Africa and building categories, the precision is 0.90
Liu et al. [59]	Brodatz, Vistex	Brodatz consisting of 1856 and 600 texture image Vistex consisting of 640 and 864 texture images. Each class in Brodatz and Vistex consists of 16 similar images	Fusion of color histogram and LBP-based features	Texture-based images retrieval	0.841 and 0.952

seems to be a better option. The color feature is extracted using the color histogram while the texture feature is extracted using discrete wavelet transform (DWT) and by edge histogram descriptor. In the extraction of color features, the color space of the image describes the color array. HSV color space is used for color feature, as reported the hue and saturation is very close to the human visual system. The DWT is used for texture feature extraction because it is very efficient for nonstationary signal. It varies for both the frequency and spatial range. Here, the author applied "Daubechies dbl" wave as it gives very efficient result than the others. Edge histogram descriptor is used to depict only the distribution of local edges in the image. EDH is used to find the most relevant image from the database and it performed some computational steps, and at last, EDH is calculated for the image. Different experiments are used to determine this technique; as a result, it performs better than the existing CBIR system [63].

7. Local Feature-Based Approaches

Kang et al. [64] conducted a study on image similarity assessment technique based on sparse feature representation. To automatically interpret the similar things in different images is the main reason behind similarity assessment. Information fidelity problem is taken as the image similarity assessment problem. For gathering information available in the reference image and estimating the amount of information that can be collected from the test image, a feature-based approach is proposed [64]. This feature-based approach will basically assess the similar things between two images. A descriptor dictionary is learned to extract different features points and the corresponding descriptor from an image to understand the information available in the image. Then sparse representation is used to formulate the image similarity assessment problem. The proposed scheme is applied to three popular applications which are image copy-move detection, retrieval, and recognition that are properly formulated to sparse representation problem. Several public datasets such as Corel-1000, COIL-20, COIL-100, and Caltech-101 are used for simulation and obtaining the desired results [64].

Zhao et al. [65] proposed cooperative sparse representation in two opposite directions for semisupervised image annotation. According to the recent research studies [8], sparse representation is effective for many computer vision problems and its kernel version has powerful classification capability. They focused on cooperative SR application in the semisupervised image annotation which may increase the number of labeled images in the training image classifiers for future use. A set of labeled and unlabeled images is provided, and the usual SR methodology which is also known as forward SR is used to represent each unlabeled image with many other labeled images, and after that, the unlabeled image is annotated according to the label image annotations. In backward SR approach, the annotation process is completed and labels are assigned to the images that are without semantic description. The main focus is on the contribution of backward SR to image annotation. To evaluate the complementary nature between two SRs in the opposite direction, a semisupervised method called cotraining is adopted which builds a unique learning model for improved image annotation in kernel space. Results of the experiment show that two SRs are different and independent. Co-KSR results better with an image annotation with high performance improvement over other state-of-the-art semisupervised classifiers such as TSVM, GFHF, and LGC. Therefore, the proposed Co-KSR method can be an effective method for semisupervised image annotation. Figure 4 represents an overview of automatic image annotation. Different high-level semantics are assigned to image through image annotation framework.

Thiagarajan et al. [66] conducted a study on supervised local sparse coding of subimage features for image retrieval. After being widely used in image modeling, sparse

representation is now being used in applications of computer vision. The features that differentiate one image from the other must be extracted for retrieving and classifying images. To perform supervised local sparse coding of larger overlapping regions, a feature extraction approach is proposed which uses multiple global/local features. A method is proposed for designing dictionary and supervised local sparse coding of subimage heterogeneous features. Experimental results show that proposed features outperform the spatial pyramid features obtained using local descriptors. Hong and Zhu [67] proposed a novel ranking method with QBME for retrieving images faster which is based on a novel learning framework. The current QBME approach uses all examples individually and then combines their results in which on each increment of query example their computational time also increases. First, the semantic correlation, which is learned using sparse representation, of image data in the training process is explored. A semantic correlation hypergraph is constructed to model the relationship between images in the dataset. A prelearned semantic correlation is used after constructing SCHG to estimate the linking value among images. Second, a multiple probing strategy is proposed to rank the images with multiple query examples. The current QBME method accepts one input example at a time, but in the proposed method, all input examples are processed at the same time. Therefore, the proposed scheme shows effectiveness in terms of speed and retrieval performance. Wang et al. [68] carried out a study on retrieval-based face annotation by weak label regularized local coordinate coding. To detect a human face from an image and annotate it according to the image automatically is important to many real-world applications. A framework is provided to address the problems in mining massive web facial images available online. For a given query image, first using content-based image retrieval, top "n" images from web facial image databases are retrieved and then their labels are used for auto annotations. This method has two main problems that are (1) how to match the query image and images placed in the archive and (2) how similar labels can be assigned to the images that are not correlated with each other. A WLRLCC technique is proposed which exploits the principle of both local coordinate coding and graph-based weak label organization. To evaluate this proposed study, experiments were conducted on many different web facial image databases. The result proves this technique to be effective. For further improving the efficiency and scalability, an offline approximation scheme (AWLRLCC) is proposed. This is better in maintaining the comparable results and takes less time to annotate images.

Srinivas et al. [69] carried out a study on content-based medical image retrieval using dictionary learning. For grouping large medical datasets, a clustering method using dictionary learning is proposed. A K-SVD groups similar images into the clusters using dictionaries. An orthogonal matching pursuit (OMP) algorithm is used to match a query image with the existing dictionary to identify the dictionary with the sparsest representation. For retrieving the images that are similar to the query images, the images included in the cluster associated with this dictionary are compared

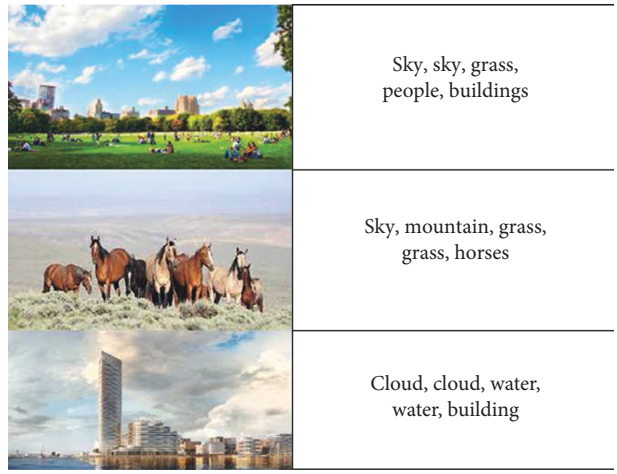

Figure 4: Example of image annotation [19].

using similarity measure. The best thing about this approach is that it does not require training and works well on different medical databases. An images database named IRNA is used for evaluating the performance of the proposed method. Results demonstrate that the proposed method efficiently retrieves image from medical databases.

Mohamadzadeh and Farsi [70] conducted a study on content-based image retrieval system via sparse representation. Several multimedia information processing systems and applications require image retrieval which finds query image in image datasets and then represents as required. Studies show that the images are retrieved in two ways, i.e., text-based and content-based image retrieval. The purpose of the retrieval systems is to retrieve the image automatically according to the query. But many researchers are attracted towards the speed and accuracy with which the images are retrieved automatically. The proposed scheme uses sparse representation to retrieve images. The goal is to present a CBIR technique involving IDWT feature and sparse representation. The color spaces that are considered include HSI and CIE-L*a*b*. The P (0.5), P (1), and ANMRR metrics of the proposed scheme and existing methods have been computed and compared. The datasets that are used to obtain metrics are Flower, Corel, ALOI, Vistex, and MPEG-7. The results of the experiments show that the proposed method has higher retrieval accuracy than the other conventional methods with the DALM algorithm for S plane. This proposed method has high performance than other methods for five datasets, and the size of the feature vector and storage space are reduced and image retrieval is improved.

Mainly two different approaches are used for the query to retrieve the images: one is text-based and the other one is through the image-based search. Image-based retrieval systems rely on models such as BoVW, and CBIR is one important application of BoVW with the aim of providing the similar image related to the query. Consider the image retrieval system when a user cannot provide an exemplar image instead only a sketch, and the raw counter is available that is called sketch-based image retrieval (SBIR). SBIR uses the edges or counter image for retrieval, and hence, it is

difficult compared to CBIR. Li et al. [71] proposed a novel sketch-based imaged retrieval using product quantization with sparse coding to construct the codebook. In this method, the desired image sketch is drawn and features are extracted using the state-of-the-art local descriptors. Then by using product quantization and sparse coding, authors [71] encoded the features into the optimized codebook and then encode the sketch features using quantization residual to improve the representation ability. Hence, this method can be efficiently computed and good performance is achieved compared to several popular SBIR. Due to the product quantization, its benefit is that it can be quickly implemented.

Image retrieval is a technique to browse, search, and retrieve the image for a large set of database. It provides convenience to human lives [72]. Machine learning is effectively increasing the quality of retrieval. Machine learning is also efficiently used for image annotation, image classification, and image recognition. Many different techniques are used to retrieve the image using color and texture features. It is difficult for simple feature extraction technique to obtain the high-level semantics information of target information; hence, for this solution, many different models are proposed which contribute to extract the semantic information of the target image. Due to advancement in machine learning, deep learning has appeared in many fields of modern life. In the deep learning also, different techniques are presented. It is to be mention that the sparse representation model is based on the foundation of sparse representation. However, the high quality of the image retrieval result is obtained from a large number of learning instances. But with the wastage of many human resources, it also occupies much computing resources. To solve this problem, the authors proposed the sparse coding-based few learning instances model for image retrieval. This model combines cross-validation sparse coding representation, sparse coding-based instance distance, and improved KNN model that reduce the number of learning instances by deleting some nonuseful even mistaken learning instances and selecting the optimized learning instances while preserving the retrieval accuracy.

According to Duan et al. [73], face recognition gained high attention in computer vision. In the last two decades, many face recognition methods are introduced. There are two main procedures for face recognition: one is to extract the discriminative feature from the face so that it can separate face image of different person and the second is that the face matching is to design effective classifiers to recognize different person. A large number of face recognition methods are proposed in the last few years, which are mainly classified into holistic and local feature representation. Generally, the local feature has better performance than the holistic feature because of robustness and stableness to local change in image feature description. Most of the local feature representations need strong prior knowledge. Because of this feature of the contextual information, the authors propose a context-aware local binary feature learning (CA-LBFL) method for face reorganization. It takes the context-aware binary code directly from the raw pixels and then compared it to with existing model that learns the feature code individually. The proposed method [73], CA-LBFL, takes the contextual information of adjacent bits by limiting the number of bitwise changes in each descriptor and obtains more robust local binary features. A detailed summary of the abovementioned local feature for CBIR is represented in Table 4. Figures 5–7 represent images that are randomly selected from the benchmarks that are commonly used to evaluate the performance of CBIR, while Figure 8 provides an overview of commonly used techniques of machine learning for CBIR framework and Figure 9 is about the key disciplines of machine-human interactions.

As discussed in section 5, histogram-based image description extracts local features and then encodes them. This process requires a precomputed codebook, also known as visual vocabulary. If there are n numbers of image datasets, separate codebook is required to be computed for every case and this process requires high computational cost [77]. In case of a limited number of training samples, the computed codebook can be biased and it can degrade the performance of the BoVW model. When the precomputed codebook from any dataset is applied for online/new set of images, the discriminating ability of codebook decreases [77]. To overcome this limitation, the authors proposed a novel implicit codebook transfer method for visual representation [77]. The proposed approach is different from the previous research as it is based on a prelearned codebooks based on nonlinear transfer. In this case, the local features are reconstructed on the basis of nonlinear transformation and implicit transformation is possible. This approach provides the use of prelearned codebooks for new visual applications through implicit learning. The proposed research is validated through several standard image benchmarks, and experimental results demonstrate the effectiveness and efficiency of this implicit learning [77].

The authors [78] proposed a novel fine-grained image classification model by using a combination of codebook generation with low-rank sparse coding (LRSC). Class-specific and generic codebooks are computed by applying optimization on accumulative reconstruction error, the sparsity constraints, and incoherence of codebook. The proposed research [78] is different from the baseline approach of BoVW image classification model that is based on the computation of a generic codebook by using all images from the training set. The local features that lie within a spatial region are encoded jointly through LRSC. The similarity among the local features is obtained through LRSC approach as this provides more discriminating fine-grained image classification [78].

According to [79], image visual features play a vital role in autonomous image classification. However, in computer vision applications, the appearance of the same view in the images of different classes often results in visual features inconsistently. The construction of explicit semantic space is an open computer vision research problem. To deal with visual features inconsistently and construction of explicit semantic space, the authors proposed structured weak semantic space for image classification problem [79]. To handle the limitation of weak semantic space, exemplar

Table 4: A summary of the performance of local feature-based approaches for CBIR.

Author	Application	Method	Dataset	Accuracy
Kang et al. [64]	Image similarity assessment	Feature-based sparse representation	COIL-20	0.985
Zhao et al. [65]	Semisupervised image annotation	Cooperative sparse representation	ImageCLEF-VCDT	—
Thiagarajan et al. [66]	Image retrieval	Supervised local sparse coding of sub image feature	Cambridge image dataset	0.97
Hong and Zhu [67]	Transductive learning image retrieval	Hypergraph-based multiexample ranking	Yale face dataset	0.65
Wang et al. [68]	Retrieval-based face annotation	Weak label regularized local coordinate coding	Databases "WDB," "ADB"	—
Srinivas et al. [69]	Content-based medical image retrieval	Dictionary learning	ImageCLEF dataset	0.5
Mohamadzadeh and Farsi [70]	Content-based image retrieval system	Sparse representation	Flower dataset, Corel dataset	—
Li et al. [71]	Sketch-based image retrieval	SBIR framework based on product quantization (PQ) with sparse coding	Eitz benchmark dataset	0.98
Duan et al. [73]	Face recognition	Context-aware local binary feature learning	LFW, YTF, FERET	0.846

Figure 5: Randomly selected images from Corel-1500 image benchmark [74].

Figure 6: Randomly selected images from some of the classes of Caltech-101 image benchmark [75].

Figure 7: Randomly selected images from 15Scene image benchmark [43].

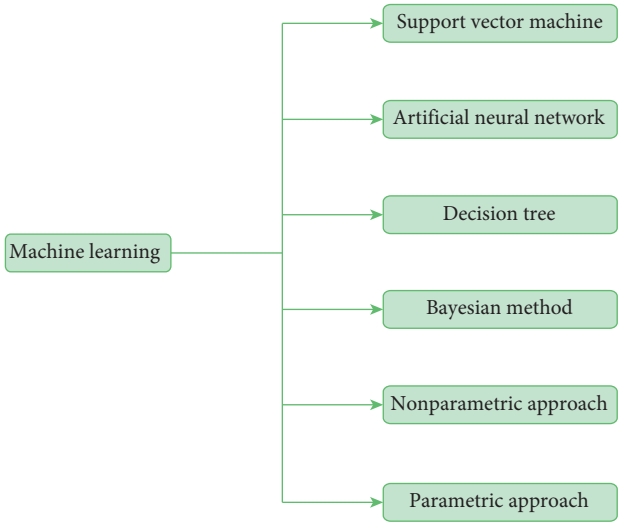

Figure 8: An overview of basic machine learning techniques for CBIR [1, 76].

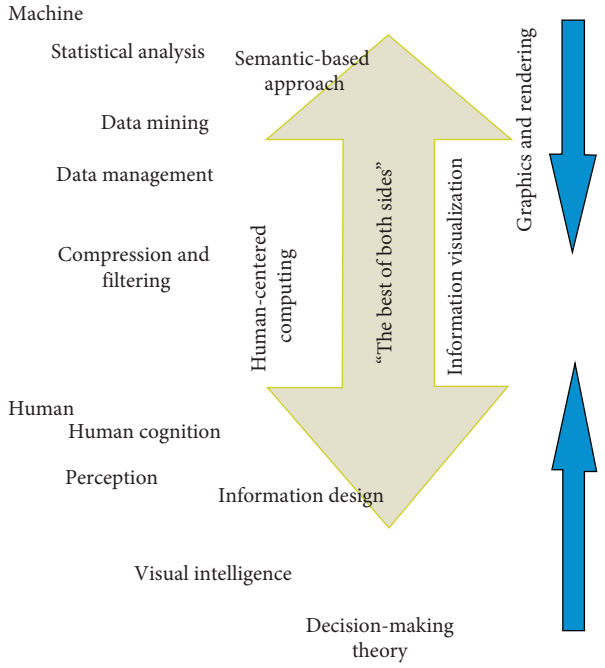

Figure 9: An overview of the key disciplines of machine-human interactions [76].

classifier is trained to discriminate between training images and test images. The structured constraints are considered to construct the weak semantic space and this is obtained by applying a low-rank constraint on the outputs of exemplar classifiers with a sparsity constraint. An alternative optimization technique is applied to obtain the learning of exemplar classifiers. Various visual features are combined to obtain efficient learning of exemplar classifier [79].

According to [80], object-centric-based categorization for image classification is more reliable as compared to the approaches that are based on division of the image into subregions like SPM. To find the location of an object within the image is an open problem for computer vision research community. According to [80], the performance of image classification model degrades if the available semantic information within the image is ignored. The authors proposed a novel approach for object categorization through Semantically modeling the Object and Context information (SOC). A prelearned classifier is applied by computing correlations of each candidate region with high confidence scores, and these regions are grouped as a cluster for object selection. The other areas of the images in which there is no object are treated as the background. This approach provides a unique and discriminative feature for object categorization and representation [80].

According to [81], supervised learning is mostly used for categorization and classification of digital images. Supervised learning is dependent on labeled datasets, and in some cases, when there are too many images, it is difficult to manage the labeling process. To handle this problem, the authors proposed a novel weak semantic consistency constrained (WSCC) approach for image classification. In this case, the extreme circumstance is obtained by considering each image as one class. Through this approach, learning of exemplar classifier is used to predict weak semantic correlations [81]. In case when there is no available labeled information, the images are clustered through the weak semantic correlations and images within the one cluster are assigned the same midlevel class. The partially labeled images are used to constrain the process of clustering and they are assigned to various midlevel classes on the basis of visual semantics. In this way, the newly assigned images are used for classifier learning and the process is repeated till convergence. The experiments are performed by using semisupervised and unsupervised image classification [81].

8. CBIR Research Using Deep-Learning Techniques

Searching for digital images from lager storage or databases is often required, so content-based image retrieval (CBIR) also known as query-based image retrieval (QBIR) is used for image retrieval. Many approaches are used to resolve this issue such as scale-invariant transform and vector of locally aggregated descriptor. Due to most prominent results and with a great performance of the deep convolutional neural network (CNN), a novel term frequency-inverse document frequency (TF-IDF) using as description vector the weighted convolutional word frequencies based on CNN is proposed for CBIR. For this purpose, the learned filers of convolutional layers of convolution neuron model were used as a detector of the visual words, in which the degree of the visual pattern is provided by the activation of each filter as tf part. Then three approaches of computing the idf part are proposed [82]. By providing powerful image retrieval techniques with a better outcome, these approaches concatenate the TF-IDF with CNN analysis for visual content. To prove the proposed model, the authors conduct experiment on four image retrieval datasets and the outcomes of the experiments show the existence of the truth of the model. Figure 10 represents an example of image classification-based framework using the DNN framework.

In order to handle the large scale, Shi et al. [83] proposed a hashing algorithm that extracts features from images and learns their binary representations. The authors model the pairwise matrix and an objective function with deep-learning framework that learns the binary representations of images. Experiments are conducted on thousands of histopathology images (on 5356 skeletal muscle and 2176 lung cancer images with 4 types of diseases) to indicate the trustworthiness of the proposed algorithm. The efficiency of the proposed algorithms is achieved with 97.94% classification accuracy.

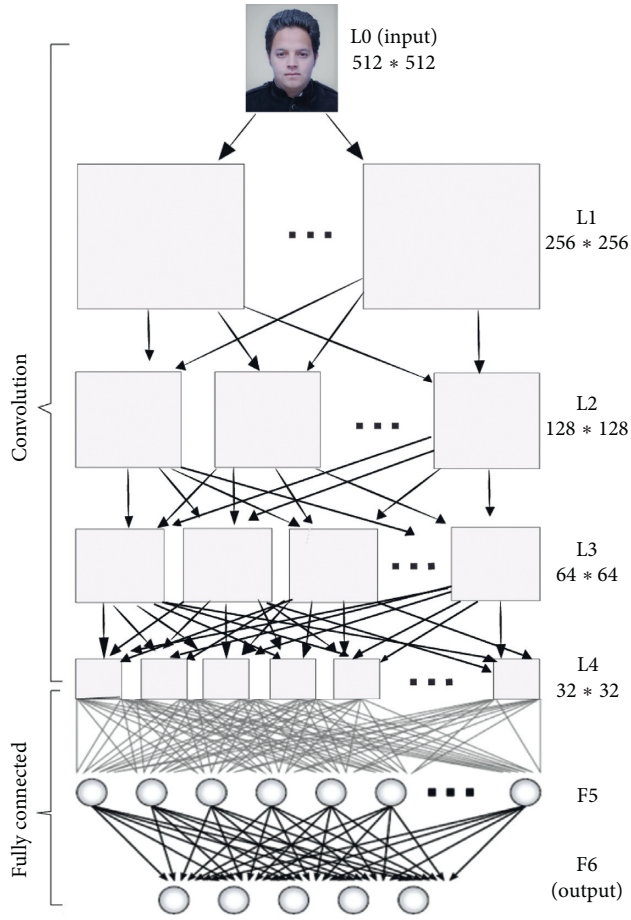

Figure 10: Example of image classification-based framework using DNN framework.

Zhu et al. [84] proposed unsupervised visual hashing approach known as the semantics assisted visual hashing (SAVH). This system uses two components that are offline learning and online learning. In offline learning firstly, the image pixel is transformed into mathematical vector representation by extracting the visual and texture feature. Then, text enhancing the visual graph is extracted with the assistance of topic hypergraph, and the semantics information is extracted from the text information and then the hash code of image is learned which preserves the correlation of image between the semantics and images, and then at the last, the hash function code is generated within the linear aggressive model. These desirable properties match the requirement of real application scenarios of CBIR [84].

In computer vision applications, the use of CNN has shown a remarkable performance, especially in CBIR models. Most of the CNN models get the features in the last layer using a single CNN with order less quantization approach and its drawback is they limit the utilization of intermediate convolutional layer for identifying local image pattern. So, in this paper, a new technique is identified as bilinear CNN-based architecture. This method used two parallel CNN models to extract the feature without the prior knowledge of the semantics of image content. The feature is

directly extracted from the activation of the convolutional layer rather than reducing very low-dimensional feature. The experiment on this approach gives a very important conclusion: This model reduces the image representation to the compact length as it used different quantized levels to extract the feature, so it is remarkable to boost the retrieval performance and the search time and storage cost. Secondly, the bilinear CRB-CNN is very effective in learning a very complex image having different semantics. Ten milliseconds is needed to extract the feature from the image and search from the database and very small disk size is needed to represent and store the image. And at the end, end-to-end tanning is applied without any other metadata, annotations, tags which conformed the capability of CRB-CNN to extract the feature from only the visual information in CBIR task. This technique also applies on the large-scale database image to retrieve the image and showed a high retrieval performance [85].

For efficient image search, hashing function gains efficient attention in CBIR [86]. Hashing function gives a similar binary code to the similar content of the image which maps the high-dimensional visual data into low-dimensional binary space. This approach is basically depending upon the CNN. It is to be assumed that the semantic labels are represented by the several latent layer attributes (binary code) and classification also depends upon these attributes. Based on this approach, the supervised deep hashing technique constructs a hash function from a latent layer in the deep neurons network and the binary code is learned from the objective functions that explained about the classification error and other desirable properties in the binary code. The main feature of the SSDH is that it unifies retrieval and classification in a single model. SSDH is scalable to large-scale search, and by slight modification in the existing deep network for classification, SSDH is simple and easily realizable [86]. A detailed summary of the abovementioned deep-learning-based features for CBIR is represented in Table 5.

Effective image analysis and classification of the visual information using discriminative information is considered as an open research problem [91]. Many research models are proposed using different approaches either by combining views by graph-based approach or by using transfer learning. It is difficult from the existing methods to compute the discriminative information at the image borders and to find similarity consistency constraint. The authors [91] proposed a multiview label sharing method (MVLS) for this open research problem and tried to maintain and retain the similarity. For visual classification and representation, optimization over the transformation and classification parameters is combined for transformation matrix learning and classifier training. Results on MVLS with different six views (no intra-view and no inter-view plus no intra-view) and nine views (combination of intra-view and inter-view) are conducted. Experimental results are compared with several state-of-the-art research and results shows the effectiveness of the proposed MVLS approach [91].

For the understanding of images and object categorization, methods like CNN and local feature have shown good performance in many application domains. The use of CNN models is still challenging for precise categorization of object and in the case with limited training information and labels. To handle the semantic gap, the smooth constraints can be used, but the performance of the CNN model degrades due to the smaller size of the training set. The authors [92] proposed a multiview algorithm with few labels and view consistency (MVFL-VC). Both labeled and unlabeled images are used together for the image view consistency with multiview information. The discriminative power of the learned parameter is also enhanced by unlabeled training images. To evaluate the proposed algorithm, experiments are conducted on different datasets. The proposed MVFL-VC algorithm can be used with other image classification and representation techniques. The algorithm is tested on unlabeled and unseen datasets. The results of experiments and analysis reveal the effectiveness of the proposed method [92].

The extraction of domain space knowledge can be beneficial to reduce the semantic gap [93]. The authors proposed multiview semantics representation (MVSR), which is a semantics representation for visual recognition. The proposed algorithm divides the images on the basis of semantic and visual similarities [93]. Two visual similarities for training samples provide a stable and homogenous perception that can handle different partition techniques and different views. The proposed research based on MVSR is more discriminative than other semantics approaches as the semantic information is computed for future use from each view and from separate collection of images and different views. Different publicly available image benchmarks are used to evaluate this research, and the experimental results show the effectiveness of MVSR. The result demonstrated that MVSR improved classification performance in terms of precision for image sets with more visual variations.

9. Feature Extraction Techniques for Face Recognition

Face recognition is one of the important applications of computer vision and is used for the identity of a person on the basis of facial features and is considered as a challenging computer vision problem due to complex nature of facial manifold. In the study [94], the authors proposed a pose- and expression-invariant algorithm for 3D face recognition. The pose of the probe face image is corrected by employing an intrinsic coordinate system (ICS)-based approach. For feature extraction, this study employed region-based principal component analysis (PCA). The classification module was implemented by using Mahalanobis Cosine (MahCos) distance metric and weighted Borda count method through re-ranking stage. The methodology is validated by using two face recognition datasets that are GavabDB and FRGC v2.0.

In another 3D face recognition algorithm [95], the authors employed a two-pass face alignment method capable of handling frontal and profile face images using ICS and a minimum nose-tip-scanner distance-based approach. Face recognition in multiview mode was performed using PCA-based features employing multistage unified classifier and

Table 5: A summary of the performance of deep-learning-based approaches for CBIR.

Authors	Datasets	Purpose	Model	Accuracy
Krizhevsky et al. [87]	ILSVRC-2010 and ILSVRC-2012	Image classification	CNN	37.50% top-1 and 17.00% top-5 error rate on ILSVRC-2010 and 15.3% top-5 error rate on ILSVRC-2012
Sun et al. [88]	LFW (Labeled Face in the Wild)	Face verification	ConvNets DeepID	97.45% accuracy
Karpathy and Fei-Fei [89]	Flickr8K, Flickr30 K and MSCOCO	Generation of descriptions of image regions	CNN and multimodal RNN	Encouraging results
Li et al. [90]	MIRFlickr and NUS-WIDE	Social image understanding	DCE	The performance of CBIR 0.512 on MIRFlickr and 0.632 NUS-WID with $k = 1000$
Kondylidis et al. [82]	INRIA Holidays, Oxford 5k, Paris 6k, UK Bench	Content-based image retrieval	CNN based tf-idf	Improved results
Shi et al. [83]	5356 skeletal muscle and 2176 lung cancer images with four types of diseases	Histopathology image classification and retrieval	PDRH algorithm	97.49% classification accuracy and MAP (97.49% and 97.33%)

SVM. The performance of the methodology is corroborated using four image benchmarks that are GavabDB, Bosphorus, UMB-DB, and FRGC v2.0.

In a recently published research [96], the authors introduced a novel approach for alignment of facial faces and transformed pose of face acquisition into aligned frontal view based on the three-dimensional variance of the facial data. The facial features are extracted using Kernel Fisher analysis (KFA) in a subject-specific perspective based on iso-depth curves. The classification of the faces is performed by using four classification algorithms. The methodology is tested on GavabDB and FRGC v2.0 3D face databases.

In another recently proposed research [97], the authors proposed a deeply learned pose-invariant image analysis algorithm with applications in 3D face recognition. The face alignment in the proposed methodology was accomplished using a nose-tip heuristic-based pose learning approach followed by a coarse-to-fine alignment algorithm. The feature extraction module is employed through a deep-learning algorithm using AlexNet. The classification is performed using AlexNet and SVM in separate experiments employing GavabDB, Bosphorus, UMB-DB, and FRGC v2.0 3D face databases.

In [98], a hybrid model to age-invariant face recognition has been presented. Specifically, face images are represented by generative and discriminative models. Deep networks are then used to extract discriminative features. The deeply learned generative and discriminative matching scores are then fused to get final recognition accuracies. The approach is suitable to recognize face images across a variety of challenging datasets such as MORPH and FG-Net.

In [99], demographic traits including age group, gender, and race have been used to enhance the recognition accuracies of face images across challenging aging variations. First, the convolutional neural networks are used to extract age-, gender-, and race-specific face features. These features in conjunction with deeply learned features are used to recognize and retrieve face images. The experimental results suggest that recognition and retrieval rates can be enhanced significantly by demographic-assisted face features.

In [100], facial asymmetry-based anthropometric dimensions have been used to estimate the gender and ethnicity of a given face image. A regressive model is first used to determine the discriminative dimensions. The gender- and ethnic-specific dimensions are subsequently applied to train a neural network for the face classification task. The study is significant to analyze the role of facial asymmetry-based dimensions to estimate the gender and race of a test face image.

Asymmetric face features have been used to grade face palsy disease in [101]. More specifically, the generative adversarial network (GAN) has been used to estimate the severity of facial palsy disease for a given face image. Deeply learned features from a face image are then used to grade the facial palsy into one of the five grades according to benchmark definitions. A matching-scores space-based face recognition scheme has been presented in [102]. Local, global, and densely sampled asymmetric face features have been used to build a matching-scores space. A probe face image can be recognized based on the matching scores in the proposed space. The study is very significant to analyze the impact of age on facial asymmetry.

The role of facial asymmetry-based age group estimation in recognizing face images across temporal variations has been studied in [103]. First, the age group of a probe face image is estimated using facial asymmetry. The information learned from the age group estimation is then used to recognize face image across aging variations more effectively.

In [104], data augmentation has been effectively used to recognize face images across makeup variations. The authors used six celebrity-famous makeup styles to augment the face datasets. The augmented datasets are then used to train a deep network. Face recognition experiments show the effectiveness of the proposed approach to recognize face images across artificial makeup variations across a variety of challenging datasets. More recently, the impact of asymmetric left and asymmetric right face images on accurate age estimation has been studied in [105]. The study analyses how accurate the age estimation is influenced by the left

and right half-face images. The extensive experimental results suggest that asymmetric right face images can be used to estimate the exact age of a probe face image more accurately.

3D face recognition is an active area of research and underpins numerous applications [94–97]. However, it is a challenging problem due to the complex nature of the facial manifold. The existing methods based on holistic, local, and hybrid features show competitive performance but are still short of what is needed [94–97]. Alignment of facial surfaces is another key step to obtain state-of-the-art performance. Novel and accurate alignment algorithms may further enhance face recognition accuracies. On the other hand, deep-learning algorithms successfully employed in various image processing applications are needed to be explored to improve 3D face recognition performance.

In the above-presented studies [98–105], handcrafted and deeply learned face features have been introduced for robust face recognition. The experimental results suggest that deeply learned face features can surpass the performance of handcrafted features. The results have been reported on aging datasets such as MORPH, FG-Net, CACD, and FERET. In future, the presented studies can be extended to analyze the impact of deeply learned densely sampled features on face recognition performance. Moreover, new datasets such as LAP-1 and LAP-2 can also be used for face recognition and age estimation.

10. Distance Measures

Different distance measures are applied on the feature vectors to compute the similarity among the query images and the images placed in the archive. The distance measure is selected according to the structure of the feature vector and it indicates the similarity. The effective image retrieval is dependent on the type of applied similarity as it matches the object regions, background, and objects in the image. According to the literature [76], it is a challenging task to find the adequate and robust distance measure. A detailed summary of the popular distance measures that are commonly used in CBIR is referred to the article [76]. Figure 11 represents the concept of top-5 to top-25 image retrieval results on the basis of search by query image.

11. Performance Evaluation Criteria

There are various performance evaluation criteria for CBIR and they are handled in a predefined standard. It is important to mention here that there is no single standard rule/criterion to evaluate the CBIR performance. There are set of some common measures that are reported in the literature. The selection of any measure among the criteria mentioned below depends on the application domain, user requirement, and the nature of the algorithm itself. The following performance evaluation criteria are commonly used.

11.1. Precision and Recall. Precision (P) and recall (R) are commonly used for performance evaluation of CBIR research. Precision is the ratio of the number of relevant images within the first k results to the total number of images that are retrieved and is expressed as follows: precision (P) is equivalent to the ratio of relevant images retrieved to the total number of images retrieved (N_{TR}):

$$P = \frac{\mathrm{tp}}{N_{\mathrm{TR}}} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp}}, \tag{1}$$

where tp refers to the relevant images retrieved and fp refers to the false positive, i.e., the images misclassified as relevant images.

11.2. Recall. Recall (R) is stated as the ratio of relevant images retrieved to the number of relevant images in the database:

$$R = \frac{\mathrm{tp}}{N_{\mathrm{RI}}} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fn}}, \tag{2}$$

where tp refers to the relevant images retrieved, N_{RI} refers to the number of relevant images in the database. N_{RI} is obtained as tp + fn, where fn refers to the false negative, i.e., the images that actually belonged to the relevant class, but misclassified as belonging to some other class.

11.3. F-Measure. It is the harmonic mean of P and R; the higher F-measure values indicate better predictive power:

$$F = 2\frac{P \cdot R}{P + R}, \tag{3}$$

where P and R refer to precision and recall, respectively.

11.4. Average Precision. The average precision (AP) for a single query k is obtained by taking the mean over the precision values at each relevant image:

$$\mathrm{AP} = \frac{\sum_{k=1}^{\mathrm{NRI}} (P(k) \times R(k))}{\mathrm{NRI}}. \tag{4}$$

11.5. Mean Average Precision. For a set of queries S, the mean average precision (MAP) is the mean of AP values for each query and is given by

$$\mathrm{MAP} = \frac{\sum_{q=1}^{S} \mathrm{AP}(q)}{S}, \tag{5}$$

where S is the number of queries.

11.6. Precision-Recall Curve. Rank-based retrieval systems display appropriate sets of top-k retrieved images. The P and R values for each set are demonstrated graphically by the PRcurve. The PRcurve shows the trade-off between P and R under different thresholds.

Many other evaluation measures have also been proposed in the literature as averaged normalized modified retrieval rank (ANMRR) [106]. It has been applied for

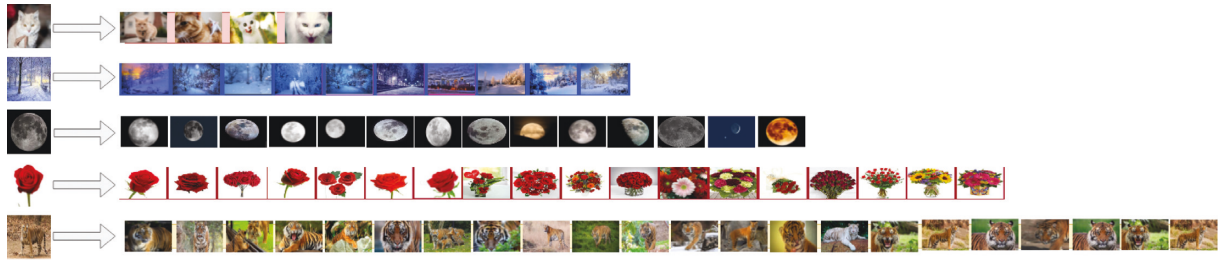

Figure 11: Example of top-5 to top-25 image retrieval results on the basis of search by query image.

MPEG-7 color experiments. ANMRR produces results in the range [0-1], where smaller values indicate better performance. Mean normalized retrieval order (MNRO) proposed by Chatzichristofis et al. [107] used a metric to represent the scaled-up behavior of the system without bias for top-k retrievals. For more details on performance evaluation metrics, the readers are referred to the article [76].

12. Conclusion and Future Directions

We have presented a comprehensive literature review on different techniques for CBIR and image representation. The main focus of this study is to present an overview of different techniques that are applied in different research models since the last 12–15 years. After this review, it is summarized that image features representation is done by the use of low-level visual features such as color, texture, spatial layout, and shape. Due to diversity in image datasets, or nonhomogeneous image properties, they cannot be represented by using single feature representation. One of the solutions to increase the performance of CBIR and image representation is to use low-level features in fusion. The semantic gap can be reduced by using the fusion of different local features as they represent the image in the form of patches and the performance is enhanced while using the fusion of local features. The combination of local and global features is also one of the directions for future research in this area. Previous research for CBIR and image representation is with traditional machine learning approaches that have shown good result in various domains. The optimization of feature representation in terms of feature dimensions can provide a strong framework for the learning of classification-based model and it will not face the problems like overfitting. The recent research for CBIR is shifted to the use of deep neural networks and they have shown good results on many datasets and outperformed handcrafted features subject to the condition of fine-tuning of the network. The large-scale image datasets and high computational machines are the main requirements for any deep network. It is a difficult and time-consuming task to manage a large-scale image dataset for supervised training of a deep network. Therefore, the performance evaluation of a deep network on a large-scale unlabeled dataset in unsupervised learning mode is also one of the possible future research directions in this area.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, 2012.

[2] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262–282, 2007.

[3] T. Khalil, M. U. Akram, H. Raja, A. Jameel, and I. Basit, "Detection of glaucoma using cup to disc ratio from spectral domain optical coherence tomography images," *IEEE Access*, vol. 6, pp. 4560–4576, 2018.

[4] S. Yang, L. Li, S. Wang, W. Zhang, Q. Huang, and Q. Tian, "SkeletonNet: a hybrid network with a skeleton-embedding process for multi-view image representation learning," *IEEE Transactions on Multimedia*, vol. 1, no. 1, 2019.

[5] W. Zhao, L. Yan, and Y. Zhang, "Geometric-constrained multi-view image matching method based on semi-global optimization," *Geo-Spatial Information Science*, vol. 21, no. 2, pp. 115–126, 2018.

[6] W. Zhou, H. Li, and Q. Tian, "Recent advance in content-based image retrieval: a literature survey," 2017, https://arxiv.org/abs/1706.06064.

[7] A. Amelio, "A new axiomatic methodology for the image similarity," *Applied Soft Computing*, vol. 81, p. 105474, 2019.

[8] C. Celik and H. S. Bilge, "Content based image retrieval with sparse representations and local feature descriptors: a comparative study," *Pattern Recognition*, vol. 68, pp. 1–13, 2017.

[9] T. Khalil, M. Usman Akram, S. Khalid, and A. Jameel, "Improved automated detection of glaucoma from fundus image using hybrid structural and textural features," *IET Image Processing*, vol. 11, no. 9, pp. 693–700, 2017.

[10] L. Amelio, R. Janković, and A. Amelio, "A new dissimilarity measure for clustering with application to dermoscopic images," in *Proceedings of the 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–8, IEEE, Zakynthos, Greece, July 2018.

[11] S. Susan, P. Agrawal, M. Mittal, and S. Bansal, "New shape descriptor in the context of edge continuity," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 2, pp. 101–109, 2019.

[12] L. Piras and G. Giacinto, "Information fusion in content based image retrieval: a comprehensive overview," *Information Fusion*, vol. 37, pp. 50–60, 2017.

[13] L. Amelio and A. Amelio, "Classification methods in image analysis with a special focus on medical analytics," in

Machine Learning Paradigms, pp. 31–69, Springer, Basel, Switzerland, 2019.

[14] D. Ping Tian, "A review on image feature extraction and representation techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 385–396, 2013.

[15] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004.

[16] R. Datta, J. Li, and J. Z. Wang, "Content-based image retrieval: approaches and trends of the new age," in *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 253–262, ACM, Singapore, November 2005.

[17] Z. Yu and W. Wang, "Learning DALTS for cross-modal retrieval," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 1, pp. 9–16, 2019.

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

[19] N. Ali, *Image Retrieval Using Visual Image Features and Automatic Image Annotation*, University of Engineering and Technology, Taxila, Pakistan, 2016.

[20] B. Zafar, R. Ashraf, N. Ali et al., "Intelligent image classification-based on spatial weighted histograms of concentric circles," *Computer Science and Information Systems*, vol. 15, no. 3, pp. 615–633, 2018.

[21] G. Qi, H. Wang, M. Haner, C. Weng, S. Chen, and Z. Zhu, "Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 2, pp. 80–91, 2019.

[22] U. Markowska-Kaczmar and H. Kwaśnicka, "Deep learning––a new era in bridging the semantic gap," in *Bridging the Semantic Gap in Image and Video Analysis*, pp. 123–159, Springer, Basel, Switzerland, 2018.

[23] F. Riaz, S. Jabbar, M. Sajid, M. Ahmad, K. Naseer, and N. Ali, "A collision avoidance scheme for autonomous vehicles inspired by human social norms," *Computers & Electrical Engineering*, vol. 69, pp. 690–704, 2018.

[24] H. Shao, Y. Wu, W. Cui, and J. Zhang, "Image retrieval based on MPEG-7 dominant color descriptor," in *Proceedings of the 9th International Conference for Young Computer Scientists ICYCS 2008*, pp. 753–757, IEEE, Hunan, China, November 2008.

[25] X. Duanmu, "Image retrieval using color moment invariant," in *Proceedings of the 2010 Seventh International Conference on Information Technology: New Generations (ITNG)*, pp. 200–203, IEEE, Las Vegas, NV, USA, April 2010.

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

[27] H. Zhang, Z. Dong, and H. Shu, "Object recognition by a complete set of pseudo-Zernike moment invariants," in *Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 930–933, IEEE, Dallas, TX, USA, March 2010.

[28] J. M. Guo, H. Prasetyo, and J. H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 466–481, 2015.

[29] Y. Liu, D. Zhang, and G. Lu, "Region-based image retrieval with high-level semantics using decision tree learning," *Pattern Recognition*, vol. 41, no. 8, pp. 2554–2570, 2008.

[30] M. M. Islam, D. Zhang, and G. Lu, "Automatic categorization of image regions using dominant color based vector quantization," in *Proceedings of the Digital Image Computing: Techniques and Applications*, pp. 191–198, IEEE, Canberra, Australia, December 2008.

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

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

[33] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor," *Pattern Recognition*, vol. 44, no. 9, pp. 2123–2133, 2011.

[34] X.-Y. Wang, Z.-F. Chen, and J.-J. Yun, "An effective method for color image retrieval based on texture," *Computer Standards & Interfaces*, vol. 34, no. 1, pp. 31–35, 2012.

[35] R. Ashraf, K. Bashir, A. Irtaza, and M. Mahmood, "Content based image retrieval using embedded neural networks with bandletized regions," *Entropy*, vol. 17, no. 6, pp. 3552–3580, 2015.

[36] A. Irtaza and M. A. Jaffar, "Categorical image retrieval through genetically optimized support vector machines (GOSVM) and hybrid texture features," *Signal, Image and Video Processing*, vol. 9, no. 7, pp. 1503–1519, 2015.

[37] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

[38] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, "Local derivative radial patterns: a new texture descriptor for content-based image retrieval," *Signal Processing*, vol. 137, pp. 274–286, 2017.

[39] X. Wang and Z. Wang, "A novel method for image retrieval based on structure elements' descriptor," *Journal of Visual Communication and Image Representation*, vol. 24, no. 1, pp. 63–74, 2013.

[40] N.-E. Lasmar and Y. Berthoumieu, "Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2246–2261, 2014.

[41] Z. Hong and Q. Jiang, "Hybrid content-based trademark retrieval using region and contour features," in *Proceedings of the 22nd International Conference on Advanced Information Networking and Applications-Workshops AINAW 2008*, pp. 1163–1168, IEEE, Okinawa, Japan, March 2008.

[42] N. Ali, K. B. Bajwa, R. Sablatnig et al., "A novel image retrieval based on visual words integration of SIFT and SURF," *PLoS One*, vol. 11, no. 6, Article ID e0157428, 2016.

[43] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Volume 2 (CVPR'06)*, pp. 2169–2178, IEEE, New York, NY, USA, June 2006.

[44] Z. Mehmood, S. M. Anwar, N. Ali, H. A. Habib, and M. Rashid, "A novel image retrieval based on a combination of local and global histograms of visual words," *Mathematical Problems in Engineering*, vol. 2016, Article ID 8217250, 12 pages, 2016.

[45] M. Naeem, R. Ashraf, N. Ali, M. Ahmad, and M. A. Habib, "Bottom up approach for better requirements elicitation," in *Proceedings of the International Conference on Future Networks and Distributed Systems*, p. 60, ACM, Cambridge, UK, July 2017.

[46] B. Zafar, R. Ashraf, N. Ali et al., "A novel discriminating and relative global spatial image representation with applications in CBIR," *Applied Sciences*, vol. 8, no. 11, p. 2242, 2018.

[47] N. Ali, B. Zafar, F. Riaz et al., "A hybrid geometric spatial image representation for scene classification," *PLoS One*, vol. 13, no. 9, Article ID e0203339, 2018.

[48] B. Zafar, R. Ashraf, N. Ali, M. Ahmed, S. Jabbar, and S. A. Chatzichristofis, "Image classification by addition of spatial information based on histograms of orthogonal vectors," *PLoS One*, vol. 13, no. 6, Article ID e0198175, 2018.

[49] N. Ali, K. B. Bajwa, R. Sablatnig, and Z. Mehmood, "Image retrieval by addition of spatial information based on histograms of triangular regions," *Computers & Electrical Engineering*, vol. 54, pp. 539–550, 2016.

[50] R. Khan, C. Barat, D. Muselet, and C. Ducottet, "Spatial orientations of visual word pairs to improve bag-of-visual-words model," in *Proceedings of the British Machine Vision Conference*, pp. 89–91, BMVA Press, Surrey, UK, September 2012.

[51] H. Anwar, S. Zambanini, and M. Kampel, "A rotation-invariant bag of visual words model for symbols based ancient coin classification," in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*, pp. 5257–5261, IEEE, Paris, France, October 2014.

[52] H. Anwar, S. Zambanini, and M. Kampel, "Efficient scale- and rotation-invariant encoding of visual words for image classification," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1762–1765, 2015.

[53] R. Khan, C. Barat, D. Muselet, and C. Ducottet, "Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model," *Computer Vision and Image Understanding*, vol. 132, pp. 102–112, 2015.

[54] N. Ali, B. Zafar, M. K. Iqbal et al., "Modeling global geometric spatial information for rotation invariant classification of satellite images," *PLos One*, vol. 14, no. 7, Article ID e0219833, 2019.

[55] R. Ashraf, M. Ahmed, S. Jabbar et al., "Content based image retrieval by using color descriptor and discrete wavelet transform," *Journal of Medical Systems*, vol. 42, no. 3, p. 44, 2018.

[56] R. Ashraf, M. Ahmed, U. Ahmad, M. A. Habib, S. Jabbar, and K. Naseer, "MDCBIR-MF: multimedia data for content-based image retrieval by using multiple features," *Multimedia Tools and Applications*, pp. 1–27, 2018.

[57] Y. Mistry, D. Ingole, and M. Ingole, "Content based image retrieval using hybrid features and various distance metric," *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, pp. 878–888, 2017.

[58] K. T. Ahmed, M. A. Iqbal, and A. Iqbal, "Content based image retrieval using image features information fusion," *Information Fusion*, vol. 51, pp. 76–99, 2018.

[59] P. Liu, J.-M. Guo, K. Chamnongthai, and H. Prasetyo, "Fusion of color histogram and LBP-based features for texture image retrieval and classification," *Information Sciences*, vol. 390, pp. 95–111, 2017.

[60] W. Zhou, H. Li, J. Sun, and Q. Tian, "Collaborative index embedding for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1154–1166, 2018.

[61] C. Li, Y. Huang, and L. Zhu, "Color texture image retrieval based on Gaussian copula models of Gabor wavelets," *Pattern Recognition*, vol. 64, pp. 118–129, 2017.

[62] H. H. Bu, N. Kim, C. J. Moon, and J. H. Kim, "Content-based image retrieval using combined color and texture features extracted by multi-resolution multi-direction filtering," *Journal of Information Processing Systems*, vol. 13, no. 3, pp. 464–475, 2017.

[63] A. Nazir, R. Ashraf, T. Hamdani, and N. Ali, "Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor," in *Proceedings of the 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pp. 1–6, IEEE, Sukkur, Pakistan, March 2018.

[64] L.-W. Kang, C.-Y. Hsu, H.-W. Chen, C.-S. Lu, C.-Y. Lin, and S.-C. Pei, "Feature-based sparse representation for image similarity assessment," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 1019–1030, 2011.

[65] Z.-Q. Zhao, H. Glotin, Z. Xie, J. Gao, and X. Wu, "Cooperative sparse representation in two opposite directions for semi-supervised image annotation," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4218–4231, 2012.

[66] J. J. Thiagarajan, K. N. Ramamurthy, P. Sattigeri, and A. Spanias, "Supervised local sparse coding of sub-image features for image retrieval," in *Proceedings of the 2012 19th IEEE International Conference on Image Processing (ICIP)*, pp. 3117–3120, IEEE, Melbourne, Australia, September-October 2012.

[67] C. Hong and J. Zhu, "Hypergraph-based multi-example ranking with sparse representation for transductive learning image retrieval," *Neurocomputing*, vol. 101, pp. 94–103, 2013.

[68] D. Wang, S. C. Hoi, Y. He, J. Zhu, T. Mei, and J. Luo, "Retrieval-based face annotation by weak label regularized local coordinate coding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 550–563, 2014.

[69] M. Srinivas, R. R. Naidu, C. S. Sastry, and C. K. Mohan, "Content based medical image retrieval using dictionary learning," *Neurocomputing*, vol. 168, pp. 880–895, 2015.

[70] S. Mohamadzadeh and H. Farsi, "Content-based image retrieval system via sparse representation," *IET Computer Vision*, vol. 10, no. 1, pp. 95–102, 2016.

[71] Q. Li, Y. Han, and J. Dang, "Sketch4Image: a novel framework for sketch-based image retrieval based on product quantization with coding residuals," *Multimedia Tools and Applications*, vol. 75, no. 5, pp. 2419–2434, 2016.

[72] H. Wu, R. Bie, J. Guo, X. Meng, and S. Wang, "Sparse coding based few learning instances for image retrieval," *Multimedia Tools and Applications*, vol. 78, no. 5, pp. 6033–6047, 2018.

[73] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Context-aware local binary feature learning for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1139–1153, 2018.

[74] J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," in *Proceedings of the 14th ACM International Conference on Multimedia*, pp. 911–920, ACM, Santa Barbara, CA, USA, October 2006.

[75] G. Griffin, A. Holub, and P. Perona, *Caltech-256 Object Category Dataset*, California Institute of Technology, Pasadena, CA, USA, 2007, https://authors.library.caltech.edu/7694/.

[76] A. Alzu'bi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: a comprehensive study," *Journal of*

Visual Communication and Image Representation, vol. 32, pp. 20–54, 2015.

[77] C. Zhang, J. Cheng, J. Liu, J. Pang, Q. Huang, and Q. Tian, "Beyond explicit codebook generation: visual representation using implicitly transferred codebooks," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5777–5788, 2015.

[78] C. Zhang, C. Liang, L. Li, J. Liu, Q. Huang, and Q. Tian, "Fine-grained image classification via low-rank sparse coding with general and class-specific codebooks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 7, pp. 1550–1559, 2016.

[79] C. Zhang, J. Cheng, and Q. Tian, "Structured weak semantic space construction for visual categorization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3442–3451, 2017.

[80] C. Zhang, J. Cheng, and Q. Tian, "Semantically modeling of object and context for categorization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 4, pp. 1013–1024, 2018.

[81] C. Zhang, J. Cheng, and Q. Tian, "Unsupervised and semi-supervised image classification with weak semantic consistency," *IEEE Transactions on Multimedia*, 2019.

[82] N. Kondylidis, M. Tzelepi, and A. Tefas, "Exploiting tf-idf in deep convolutional neural networks for content based image retrieval," *Multimedia Tools and Applications*, vol. 77, no. 23, pp. 30729–30748, 2018.

[83] X. Shi, M. Sapkota, F. Xing, F. Liu, L. Cui, and L. Yang, "Pairwise based deep ranking hashing for histopathology image classification and retrieval," *Pattern Recognition*, vol. 81, pp. 14–22, 2018.

[84] L. Zhu, J. Shen, L. Xie, and Z. Cheng, "Unsupervised visual hashing with semantic assistant for content-based image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, pp. 472–486, 2017.

[85] A. Alzu'bi, A. Amira, and N. Ramzan, "Content-based image retrieval with compact deep convolutional features," *Neurocomputing*, vol. 249, pp. 95–105, 2017.

[86] H.-F. Yang, K. Lin, and C.-S. Chen, "Supervised learning of semantics-preserving hash via deep convolutional neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 437–451, 2018.

[87] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2012.

[88] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891–1898, Columbus, OH, USA, June 2014.

[89] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3128–3137, Boston, MA, USA, June 2015.

[90] Z. Li, J. Tang, and T. Mei, "Deep collaborative embedding for social image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, pp. 2070–2083, 2018.

[91] C. Zhang, J. Cheng, and Q. Tian, "Multiview label sharing for visual representations and classifications," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 903–913, 2018.

[92] C. Zhang, J. Cheng, and Q. Tian, "Multiview, few-labeled object categorization by predicting labels with view consistency," *IEEE Transactions on Cybernetics*, vol. 49, no. 11, pp. 3834–3843, 2019.

[93] C. Zhang, J. Cheng, and Q. Tian, "Multiview semantic representation for visual recognition," *IEEE Transactions on Cybernetics*, pp. 1–12, 2018.

[94] N. I. Ratyal, I. A. Taj, U. I. Bajwa, and M. Sajid, "3D face recognition based on pose and expression invariant alignment," *Computers & Electrical Engineering*, vol. 46, pp. 241–255, 2015.

[95] N. Ratyal, I. Taj, U. Bajwa, and M. Sajid, "Pose and expression invariant alignment based multi-view 3D face recognition," *KSII Transactions on Internet & Information Systems*, vol. 12, no. 10, 2018.

[96] N. I. Ratyal, I. A. Taj, M. Sajid, N. Ali, A. Mahmood, and S. Razzaq, "Three-dimensional face recognition using variance-based registration and subject-specific descriptors," *International Journal of Advanced Robotic Systems*, vol. 16, no. 3, article 1729881419851716, 2019.

[97] N. Ratyal, I. A. Taj, M. Sajid et al., "Deeply learned pose invariant image analysis with applications in 3D face recognition," *Mathematical Problems in Engineering*, vol. 2019, Article ID 3547416, 21 pages, 2019.

[98] M. Sajid and T. Shafique, "Hybrid generative–discriminative approach to age-invariant face recognition," *Journal of Electronic Imaging*, vol. 27, no. 2, article 023029, 2018.

[99] M. Sajid, T. Shafique, S. Manzoor et al., "Demographic-assisted age-invariant face recognition and retrieval," *Symmetry*, vol. 10, no. 5, p. 148, 2018.

[100] M. Sajid, T. Shafique, I. Riaz et al., "Facial asymmetry-based anthropometric differences between gender and ethnicity," *Symmetry*, vol. 10, no. 7, p. 232, 2018.

[101] M. Sajid, T. Shafique, M. Baig, I. Riaz, S. Amin, and S. Manzoor, "Automatic grading of palsy using asymmetrical facial features: a study complemented by new solutions," *Symmetry*, vol. 10, no. 7, p. 242, 2018.

[102] M. Sajid, I. A. Taj, U. I. Bajwa, and N. I. Ratyal, "The role of facial asymmetry in recognizing age-separated face images," *Computers & Electrical Engineering*, vol. 54, pp. 255–270, 2016.

[103] M. Sajid, I. A. Taj, U. I. Bajwa, and N. I. Ratyal, "Facial asymmetry-based age group estimation: role in recognizing age-separated face images," *Journal of Forensic Sciences*, vol. 63, no. 6, pp. 1727–1749, 2018.

[104] M. Sajid, N. Ali, S. H. Dar et al., "Data augmentation-assisted makeup-invariant face recognition," *Mathematical Problems in Engineering*, vol. 2018, Article ID 2850632, 10 pages, 2018.

[105] M. Sajid, N. Iqbal Ratyal, N. Ali et al., "The impact of asymmetric left and asymmetric right face images on accurate age estimation," *Mathematical Problems in Engineering*, vol. 2019, Article ID 8041413, 10 pages, 2019.

[106] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley & Sons, Hoboken, NJ, USA, 2002.

[107] S. A. Chatzichristofis, C. Iakovidou, Y. S. Boutalis, and E. Angelopoulou, "Mean normalized retrieval order (MNRO): a new content-based image retrieval performance measure," *Multimedia Tools and Applications*, vol. 70, no. 3, pp. 1767–1798, 2014.