IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 13, 2019, accepted September 24, 2019, date of publication October 18, 2019,
date of current version December 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2948266

Greedy Learning of Deep Boltzmann Machine (GDBM)'s Variance and Search Algorithm for Efficient Image Retrieval

MUDHAFAR JALIL JASSIM GHRABAT[1], GUANGZHI MA[1], ZAID AMEEN ABDULJABBAR[2], MUSTAFA A. AL SIBAHEE[3], AND SAFA JALIL JASSIM[4]

[1]School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
[2]Computer Science Department, College of Education for Pure Sciences, University of Basrah, Basrah 61004, Iraq
[3]School of Computer Science, Shenzhen University, Shenzhen 518060, China
[4]College of Engineering, University of Mustansiriyah, Baghdad 10001, Iraq

Corresponding author: Guangzhi Ma (maguangzhi@hust.edu.cn)

ABSTRACT Despite extensive research on content-based image retrieval, challenges such as low accuracy, incapability to handle complex queries and high time consumption persist. Initially, a preprocessing technique is introduced in this study, a technique that uses a median filter to remove noise to achieve improved accuracy and reliability. Then, Fourier and circularity descriptors are extract in an effective manner correspondent to the texture and affine shape adaptation features. In addition, various descriptors, such as color histogram, color moment, color autocorrelogram and color coherency vector, are extracted as the invariant color features. The multiple ant colony optimization (MACOBTC) approach is implemented with whole features to find relevant features. Finally, the relevant features are utilized for the greedy learning of deep Boltzmann machine classifier (GDBM). The proposed approach obtains effective performance and accurate results on four datasets and is analyzed with various parameters such as accuracy, precision, recall, Jaccard, Dice, and Kappa coefficients. The GDBM provides a 25% increase in accuracy compared with existing techniques, such as the a priori classification algorithm.

INDEX TERMS Image retrieval, image preprocessing, image feature, shape feature, SIFT descriptor, Boltzmann machine classification.

I. INTRODUCTION

Information retrieval refers to extracting information from images. The retrieval process is often conducted on information, text, and images that are used in various applications, such as wide baseline matching, computer vision tasks, pattern recognition, recognition and tracking of objects, recognition of texture, image reconstruction and retrieval, video data mining, robot localization, camera motion recognition, and recovery of object groups. Image retrieval completely depends on recommendations or suggestions on the query of users. For example, query expansion, which is entirely automatic, provides recommendations for the concerns of users by using search engines, such as Google, Bing, and Yahoo. The main issue in information retrieval is its requirement for a considerable number of computations during a query. Document expansion aims to address the mismatch problem from the opposing viewpoint [1] by including related indexing terms of document illustration in the search index. Image retrieval is generally categorized into the following types: Text-based image retrieval (TBIR), Content-based image retrieval(CBIR), Semantic-based image retrieval. In TBIR, the images are annotated and interpreted by a textual description depending on their textual query similarity during information retrieval. In CBIR, the most visually similar images for the query are retrieved. In Semantic-based image retrieval, the keywords or meta-data are matched with the query for retrieval.

Typical TBIRs, which use the same image but which is subjective, are limited because of manual annotation. CBIRs aim to overcome this drawback. TBIRs search for images that are most significant and related to the given query images from a large image dataset. By contrast, CBIRs identify visual content by extracting features of low-level from the images

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

and measuring their similarity degrees. They also describe image contents by their colors, textures, and shapes. Notably, the textual features of TBIRs have high computational efficiency and are invariant to rotation and scales. In contrast, CBIRs use wavelet transform, such as classic discrete wavelet transform that separates images into low or high resolution and vertical, diagonal or horizontal. They also utilize high- and low-pass filters [2].

Feature extraction is a challenging feature eliciting task in image retrieval because of the variances in the textures, shapes, and colors of images. Numerous methods are used for feature extraction, but they are problematic. For example, image segmentation involves human perception. The use of shape-based features is relatively reliable and intuitive but lacks mathematical foundations with target deformation. The use of low-level visual features fails to capture image semantics. A histogram of oriented gradients (HOG) [3] is related to the scale-invariant feature transform (SIFT) and the context of shape differs from the computed dense grid of regular images. The HOG requires local contrast normalization to improve accuracy.

Local binary pattern (LBP) [4] is a widespread technique for extracting texture features because of its computational simplicity and discriminative power. The LBP operator is robust to monotonic grayscale changes. In image retrieval, relevant features are optimally selected by classifier. The k-means clustering algorithm is commonly used among many classification techniques. Clustering techniques are utilized to label images in a database according to color, size, and shape. This labeling eases the retrieval process but results in problems such as inaccuracy and high complexity.

Several traditional techniques are adapted to retrieve information from images. However, these approaches have problems such as ineffective feature extraction, misclassification, and inaccurate segmentation. An effective image retrieval technique is proposed in this work to overcome these problems. The main objective of this study is to distinguish query images from database images. We consider extraction features (color, texture and shape) from image color features. The color feature gives us information about the spatial arrangement of color or intensities in an image or selected region of an image; the texture feature gives us information about intensity ranges in an image or the selected region of an image; the shape feature provides us information about structure and statistical features in an image or the selected region of an image.

In this study, we emphasize classifying an image with the optimized features by using a classification algorithm. The processes in our work are as follows. Initially, we introduce a novel approach that relies on a trained classifier to classify with the selected features and retrieve relevant data on the basis of greedy learning of deep Boltzmann machine (GDBM) technique. Subsequently, features, such as shape, color, and texture, provide a complementary improvement and can robustly detect objects due to their invariance to scale, rotation, noise, and illumination. Finally, k-means clustering, an average evolutionary search algorithm, is used for dimensionality reduction.

To address significant image retrieval, we present three main contributions of this work. First, we propose a GDBM classifier, and an MACOBTC optimization is utilized for image retrieval. Furthermore, we ensure efficient reuse of the image features; the texture, shape and color features are used to get optimal feature set to achieve high accuracy. Second, we overcome the problems of previous works such as minimal accuracy and high time consumption. The time consumption can be decreased by using the deep Boltzmann machine approach. The retrieval speed increases and time consumption decreases. These drawbacks can be solved in this proposed work by developing an efficient image retrieval approach that incorporates the features of the color histogram, color autocorrelogram, color moments, color coherency vector, texture extraction, affine shape adaptation, SIFT descriptors, and Fourier descriptor to increase the classification rate due we introduced a major innovation in the novel approach that relies on a trained classifier classifying an image with the selected features to retrieve the relevant data based on deep Boltzmann machine (GDBM) to reduce time consumption.

Ultimately, in our work, the results show that our proposed method is more efficient. This study indicates the major innovation of the proposed method by comparing with other recent methods.

There is a key point to make in this work. A local quantized extreme pattern (LQEP) is an existing technique that fails to encode highly spatially structured information. To overcome this, a novel approach is introduced. In our work, features such as texture, shape, and color make complementary improvement and are also proficient in strongly identifying objects because of their invariance to scale, rotation, noise, and illumination. This is a major innovation of our proposed work. In our study, after undergoing k-means clustering, an average evolutionary search algorithm is used for dimensionality reduction.

Figure 1, presents a general diagram of image retrieval. Here, a large collection of images are referred to as an image database. This database is considered as a system where the image features are stored and integrated, where an image got queried and feature extraction took place by developing combinations of variables such as color, spate, texture, etc. From there, necessary features are selected, thus enhancing dimensionality reduction. On the other hand, features are selected from the feature database. Usually, the performance of image retrieval system depends upon the similarity-based measurement of feature vector of the image database and query-based feature vector. Thus, the two selected feature vectors undergo similarity measures, where it verifies the retrieved images from the database are similar to the queried image [4].

The remainder of this paper is systematized as follows. Section II offers a detailed evaluation of the existing research work related to image retrieval. Section III shows a detailed description of the proposed approach. Section IV illustrates

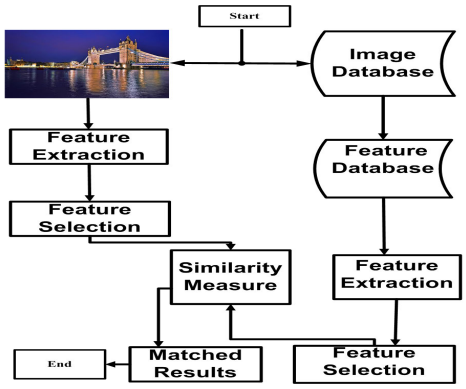

FIGURE 1. General diagram of image retrieval.

the performance and provides comparative analyses of the proposed approach. Section V provides the conclusions and future work directions.

II. RELATED WORK

Image retrieval was familiarized as early as 1987 by S.E. Madnick. In the research field, image retrieval plays a significant role. Discrepancy feature extraction techniques were implemented depending on aspects such as texture, spatial, and color layout. Feature extraction can greatly reserve important image information and provide better image categorization and retrieval.

The literature [5] presented a different model of neural network that depends on the Gaussian–Bernoulli deep Boltzmann machine (GDBM) to predict optimized conditions. The research work was performed to predict the condition of a compressor in a certain field. The outcomes obtained at the end of the research indicate that the proposed model provides enhanced and effective performance when compared with the older traditional techniques.

The literature [6] proposed a new method for image retrieval. By gathering numerous descriptors based on regions, a very small-sized and global fixed-length depiction was produced for each image. Contrary to the prevailing studies that already employed deep networks that are pretrained as a black box for producing the features, the work proposed controlled a deep designed and well-trained network to perform a particular action in the image retrieval process. The main role of the work was to control the ranked layout for studying the convolution process and the projection weights, which were utilized for building the features of a region.

The literature [7] performed a theoretical analysis on the Gaussian-binary restricted Boltzmann machines (GRBMs) considering the density models. The major part of the research work was to demonstrate that GRBMs could be expressed as a controlled combination of the Gaussians. The limitations and capabilities of the model provided enhanced insight. The GRBM also learned meaningful features without using a regularization term. The authors in [8] developed a SIFT descriptor for representing the image and the matching in image retrieval. They normalized elongated adjacent areas, conducted transformation to the affine scale space and improved the SIFT descriptor by a polar histogram alignment bin. The main downside of the developed method is its high computational complexity. Thus, the accuracy of its performance needs to be improved.

Reference [9] surveyed CBIRs with feature extraction techniques and their applications. The performance was dependent on the properties inherent in the images. CBIRs automatically retrieve color, texture and shape features. These features sometimes include text-based and visual features. Numerous methodologies have also been introduced to extract visual features and characterize features from different perspectives. Texture feature extraction is divided into the following categories: statistical techniques, structural techniques, and spectral techniques. CBIRs allow agencies to focus on technology transfer or dissemination of best practices in various fields. These techniques are utilized in commercial applications, such as search engine optimization, web mining and image retrieval. Among existing feature extraction techniques in image retrieval, color histogram collapses the order of bins when histograms have many dimensions.

Reference [10] proposed an image retrieval approach using a mixture of texture and shape features. Fuzzy C-means clustering and k-means clustering were employed to segment the images. Tamura textures, shape moment invariant features and Haralick features were utilized to extract shape and texture information of images. The extracted features were combined with other features. The results showed that the combined features perform better in retrieval than the separate features. However, the proposed method has a high time consumption.

Reference [11] reviewed CBIRs using color and texture features. CBIRs were adopted to extract only low-level features from images, such as color and texture. The combination of color and texture features was referred to as image retrieval with the use of a wavelet-based color histogram, and the similarity function was assigned as the distance function. These wavelet transforms were used to reduce the processing time. Notably, traditional techniques have problems, such as inaccurate feature extraction, high time consumption, and ineffectiveness. Reference [12] proposed an efficient image retrieval system based on a bag-of-features. Two quantitative and qualitative benchmark datasets were utilized, and patch- and image-based integrations of SIFT, LBP and HOG were applied in the feature extraction process. The results showed that the proposed integration of LBP and SIFT demonstrated excellent performance but has computational complexities.

Reference [13] proposed an automatic classification of medicinal images for CBIR. Grouping of visual structures from image content, image retrieval in medical applications (IRMA) classification codes and multiclass support vector machine (SVM) classifiers was adopted as classification methods. These methods were implemented with classes, such as modality facets, to provide automatic classification results with error rates. However, the resulting dataset was unevenly distributed, and many image classes had similar

or overlapping content. Consequently, the classification results were inaccurate.

Reference [14] Investigated the use of neural codes (i.e., descriptors) in the retrieval of images. Numerous standard retrieval benchmarks were utilized to test the competitive performance of neural codes. Once the convolutional neural network was trained for an unrelated grouping task, the network was reinstructed with a dataset of images similar to those encountered in the test. Discriminative dimensionality was reduced during matching dataset pairs to increase the performance of principal component analysis (PCA) compressed neural networks. The training and testing datasets differed from each other, while the constitutional neural network (CNN) was trained for the task of classification. The retrieval dataset was utilized to obtain relevant photographs classified by CNN. Dimensionality reduction was applied by adjusting the size of the network layer and was employed to produce codes rather than post hoc processes. The results showed that CNN-based neural codes retain a small number of informative results and are thus ineffective.

Reference [15] utilized a feed-forward back propagation neural network for CBIR. The neural network was adopted to train the features in the database, and the back propagation algorithm was applied to train image features. Then, the features were compared with relevant features and similar query images. The feed-forward neural network was used to improve the retrieval time and recall rate with vastly effectual and exact classification. However, the method is inefficient and slow when used with deep learning networks and low-level image descriptors.

Reference [16] surveyed 2D image retrieval with the multiple dimensions and the images having multi modality from sources comprising a varied group of medical information. Image search techniques were used in the CBIR, which was considered to discover images similar to the specified query. The similarity in the properties of visual components was adopted to identify the similarity of two images. Image features were used to reduce sensory and gaps semantic, and Euclidean range was used to assessment the features in terms of vector, metrics and range. However, the image retrieval technique is inapplicable to multidimensional and multi modality-based images.

Reference [17] presented an application of image retrieval called color directional local quinary pattern. The method was utilized to extract texture features using the colors red, green and blue (RGB). Corel-5000 and MIT-color datasets were adopted to evaluate the color texture features. The results showed that this method performs better than existing approaches, but it has high time consumption. Reference [18] surveyed the difficulties in learning similarity-conserving binary codes for an effective similarity search in large-scale image collections. The variances in different PCA directions and nonorthogonal relaxations were obtained using random orthogonal transformation. Minimization was utilized to refine the initial orthogonal transformation and thereby reduce the quantization error. However, this method has a quick increase in computational time when the training size becomes bigger. Reference [19] combined conditional random fields (CRFs) and Boltzmann machine shape priors for image labeling. The performance of the model in complex facial images from the dataset was evaluated. In the face segmentation process, local and global structures were combined into the CRF and GDBM of the GLOC model. The results showed that the method reduces the errors in face labeling, but it has high complexity in the computational process.

Reference [20] identified the color difference histogram by CBIR. The histogram was utilized to define the image features for retrieval of images, and the color difference histogram used the perceptually uniformed histogram values. The algorithm was applied to determine low-level features (i.e., features that do not require any segmentation of image or training model). Multitext histogram was utilized to obtain the uniform color differences perceptually in feature illustration. However, the method was inaccurate for the retrieval process. Reference [21] investigated a deep learning framework for CBIR tasks. The main goal of this proposed framework was to overcome the problem of long-lasting essential representation of features in CBIRs. The pretrained model helps to capture information that is high-semantic in the raw pixels. In this pretrained deep model, similarity learning can improve retrieval performance while applying feature representation in a new domain. The classification also considerably improves the retrieval performance. However, this model has difficulty in evaluating diverse datasets to obtain the semantic gap of information retrieval.

Reference [22] proposed a CBIR technique to retrieve satellite images from a database. First, the images were segmented into various parts by using the unsupervised J-seg algorithm. Second, a region-based representation technique was constructed for each image, and the textural features of each image were extracted using a statistical approach. Last, the images were classified by using a Bayesian classifier probabilistic approach. The proposed system efficiently segmented multispectral thematic map images, and the region-level description is understandable for users. However, the system had computational complexity and high computational time consumption. Reference [23] proposed a new approach to visual data mining for CBIR. The diversity and visually interactive (DiVI) method were used to improve similar queries with diversity. The proposed DiVI minimized the semantic gap in CBIR by using the activity of users and increased the applicability and acceptance of similarity. However, it lacked accuracy in image retrieval.

Reference [24] presented a technique for CBIR to generate image content descriptors. The method exploited the benefits of low-complexity that are ordered dither block truncation coding to compress image blocks into their corresponding bitmaps and quantized images in the encoding stage. Two image features, color cooccurrence and the bit pattern, were proposed to index images directly generated from the data streams that were encoded deprived of executing the

decoding process. The proposed method provided superior image retrieval results on existing block truncation coding and offered an improved and effective descriptor for indexing images. However, the method was limited by explicit semantic gaps.

Reference [25] suggested a rule- and classifier-based scene graph parser for image retrieval. The main goal was to overcome the issues of semantically complex queries, including the attributes and relationships of images in image retrieval systems. A query graph with relations and attributes was obtained as the best model by using the parsers' output. The output was utilized to generate 3D scenes. However, this method had difficulty in understanding the interdependency of objects in scenes. Reference [26] investigated probable techniques for the total local features that are deep for generating compressed descriptors globally for the retrieval of images. First, the different distributions of pairwise similarities in deep learning and traditional hand-engineered features were determined. Second, existing aggregation methods were re-evaluated carefully. The results showed that the sum pooling-based aggregation method provided better deep convolutional features than shallow features. This method could bear overfitting risks and improve system efficiency. However, the model was costly.

Reference [27] incorporated deep CNN with hash functions to learn feature representations jointly and mappings to hash codes. Optimization problems such as multivariate ranking measures and nonsmooth learning were solved using an effective surrogate loss-based scheme. The proposed method outperformed other hashing methods in terms of ranking quality. However, the method was limited by the tight coupling of the CNN. Reference [28] proposed a framework of supervised learning towards generating compact and hash codes that are bit-scalable directly from the raw images. Hashing learning was considered an issue of normalized learning of similarity. Hence, the images of training were organized in a group of triplet samples. Adjacency consistency was enforced by introducing a regularization term. Then, deep CNN was adopted to train the model. The proposed framework achieves superior performance in reidentifying persons in surveillance. The disadvantages of this work were the explicit semantic gap and the low effectiveness of the system.

Reference [29] presented a simple but effectual framework of deep learning for creating binary codes that are hash-like for fast image retrieval. Hash-like functions and domain-specific image representations were learned by adding a latent-attribute layer in deep CNN. The proposed method was highly scalable and did not rely on the pairwise similarity of data. The proposed framework also exhibited an improved efficiency in large-scale datasets. However, it had low reliability in image retrieval. Reference [30] introduced a new feature descriptor called local quantized extreme pattern (LQEP) for CBIR. The directional relations among the pixel center and its neighbors were collected using standard local quantized patterns, and guiding data depending on the local extreme in various directions for a given center pixel was collected using the directional local extreme pattern (DLEP). The LQEP and DLEP concepts were then integrated to offer an LQEP for the process of image retrieval. The color histogram of the RGB was incorporated with LQEP to generate the feature vector. The results showed that the proposed technique offered better results than existing techniques in terms of evaluation metrics. However, this method failed to encode highly spatial structure information.

An innovative study presented [31] an unsupervised visual hashing method known as semantic-assisted visual hashing. This approach was utilized to associate modalities involved. CBIR via sketching was presented [32]. In this technique, images are annotated manually to infer keywords for retrieving an image via a text-based search. This approach matched legal sketches and mug shots from an image gallery but failed to discriminate the unaltered images from color images or edge representations.

The CBIR retrieval method was used [33] to extract images that were similar in feature as query images from image databases. This method was practical in crime prevention, face detection, image fusion, and digital libraries. In a digital library, it retrieved the relevant book name, novel name and author name but lacked effectiveness in certain situations. A technique adopted the aptitude visual hashing to enhancing visually without explicit semantic labels, but this method failed to exhibit effectiveness when additional content CBIR [34], such as features for querying images from large-scale datasets, was implemented. For industrial and commercial uses, texture information can be developed to improve method accuracy, possibly enhancing the color extent of images. The threshold parameters must be fine-tuned, and feedback is offered to overcome disadvantages and realize improvement.

A CBIR technique [35] that uses shape, color and feedback relevance was utilized for the semantically correct representation of color and shape.

Hashing based methods are widely utilized in retrieving large scale cross image models. Several conventional methods have been designed upon binary supervision which transforms the complex relationship of multi-label into similar or dissimilar. But few methods have been developed where the rich semantic data were implied in multi-label data. This, in turn, provides improved accuracy on searching results.

Reference [56] constructed an approach based on multi-level semantic supervision generation by discovering label relevance. This paper had been designed with a deep hashing structure in order to cross retrieve the multi-label images with text. In such a case, it can capture the binary similarity and the complex semantic structure simultaneously. This research had claimed on resolving the problem of conventional deep-crossed hashing method models which was not sufficiently utilizing the rich semantic information in multi-label data. Here, the multi-level supervision on the basis of label co-occurrence has been adopted in order to ensure the preservation of exact semantic similarities by learnt hash codes i.e.

data points having common class labels are same as with data points with less common labels. The phases involved in the work are deep feature-based learning, generating supervision, hash codes. This methodology learnt image and text-based features by two perfectly designed deep NNs respectively. Finally, the compressed hash codes along with each modalities' distinctive features were learned spontaneously in one framework. This entire structure had provided a guarantee in learning the optimal features for particular cross-modal retrieval tasks. Experimentation had been conducted on a cross model dataset named as MIRFlickr-25k by comparing it with CMFH, CCA, SCM, STMH, DCMH and SePH.

Reference [57] stated that hash coding is widely utilized in approximating the Nearest Neighbour based search for image retrieval in large-scale. Also, several deep hashing methods had been developed and provided improved performance over conventional feature-based learning techniques. This is due to the evaluation of pairwise similarity on semantic labels, where it has been assigned in a hard way, i.e., if no class label has been shared then the pairwise similarity will be '0', else it will be '1'. But these similarity definitions could not impact the ranking for paired images with multiple labels. Thus, this research had been developed based on the hashing method to improve the ability to retrieve the multi-label images. Here, a pairwise similarity measure had been calculated on normalizing the semantic labels. Based on that, pairwise similarity had been categorized as hard and soft. In addition to that, Mean Square Error loss and Cross-entropy loss are added respectively for improved feature-based learning and hash coding.

The literature [29] presented deep semantic classification for multi-image retrieval. In this reference, a deep convolution network is merged with hash functions for performing hash codes map. Handcrafted features are tricky to present in the suggested method. The suggested method was done using three path deep hash function, semantic classification supervision, surrogate wastage optimization. A participation image is transformed also provided to convolution layers, fully linked layers in deep hash functions. Problems in SVM was solved by semantic classification supervision.

Reference [58] presented an effective deep learning framework to create binary codes by utilizing (CNN) to retrieve considerable scale images. in the existence of data labels, hidden layers are utilized to present the latent idea for the dominant class labels. The hash codes also images are learned in a point-wise technique instead pairwise in else supervised of image retrieval. With unpretentious alterations in deep CNN, this approach preference1% to 30% rise retrieval precision on two data sets.

The proposed methodology deals with the shape descriptor, which must be robust against geometric attacks to images. However, the technique does not provide outstanding accuracy in classifying image labels. From this review of existing approaches, we identify major disadvantages in image retrieval; high computational complexity, inaccurate feature extraction, classification, explicit semantic gap, low reliability, ineffective retrieval of images.

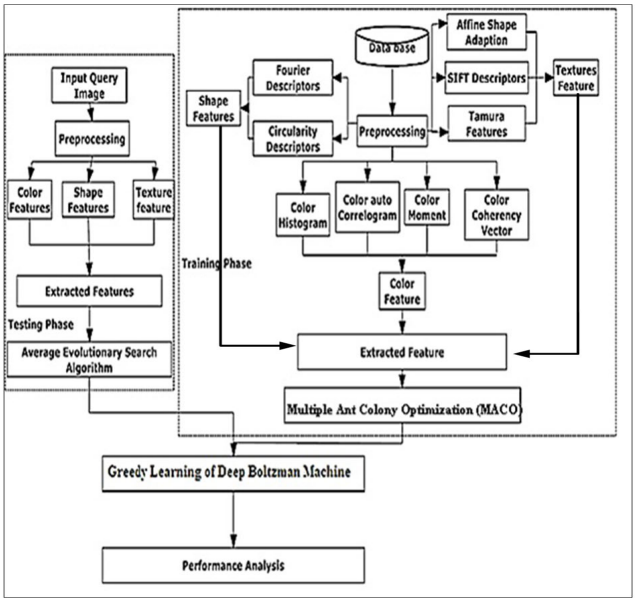

FIGURE 2. Overall work flow preprocessing.

The method proposed in this current study solves the aforementioned issues through the following ways. A preprocessing technique with the median filter is used to remove noise for improved accuracy and reliability. Shape, color and texture features are extracted using different methods. From the extracted features, dimensionality reduction is performed using the average evolutionary search algorithm. Retrieval of images using four datasets is effective because of the relevant data always being classified speedily by the trained classifier.

III. THE PROPOSED METHODOLOGY

The proposed image illustration approach depends on the implementation of the image retrieval and classification technique to classify images in a database, as shown in Figure 2. First, images are obtained from the dataset, and noise removal is performed at the preprocessing step using a median filter. Second, color, shape and texture features are extracted from the preprocessed images for useful information analysis. Third, extracted features that contain relevant information on the image object are optimized. Finally, the selected features are classified with the trained classifier to retrieve relevant data from the dataset.

A. PREPROCESSING

In this study, inputs are collected from the dataset, image noises are eliminated, and the image size is decreased to improve image quality. The median filter is utilized to reduce the input image size. The standard median filter, which is also called a median smoother, is a rank selection filter that eliminates noise simply by varying the assessment luminance of the center pixel within the frame. This method eliminates blurred images and thin line information at low noise

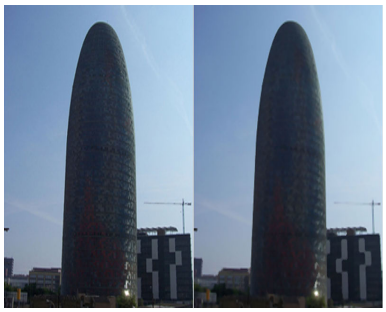

FIGURE 3. Dataset image (a) Input image (b) Filtered image.

concentrations. The filtered image S= $S(i, j)$ from SMF is defined by the equation:

$$S = (i, j) = Median(k, l)\epsilon W_{m,n} D(i + k, j + 1)), \quad (1)$$

where W is a window matrix with size $m \times n$, D is the noisy input image and I, J, K are the indexes of the image. Figure 3 shows a sample image of the preprocessing test. The processing technique ensures that the filtered image acquires better accuracy than the input image.

B. FEATURE EXTRACTION

In this step, feature extraction, which is also called dimensionality elicitation, is employed to retrieve appropriate and informative features. The image features are extracted in terms of color, shape, and texture features. Different methods of feature extraction are used for the effective elicitation of an image. The shape feature helps to provide the shape of the objects in the retrieved image. Similarly, the color and texture features help to provide the object's color and texture present in the retrieved image.

C. COLOR HISTOGRAM

A color histogram depicts the distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in an individual static list of color ranges, and it spans the image color space and the set of all possible colors. The histogram, which represents all colors and levels of occurrence in an image, is used instead of the image. A histogram reveals important facts regarding an image, the distinguish this step. First for each gray level, the number of pixels is added correspondingly to obtain the total number of pixels constituting the image. Second, an appropriate threshold can be identified easily through the histogram, which is convenient to determine. Third, image brightness can be determined by observing the histogram and spatial distribution of the values. Ultimately an image histogram can help rapidly recognize processing operations that are suitable for a specific image. Utilizing color, the information can be computed faster than using other variants. Also using this, we can accurately retrieve the images despite the manipulation of orientation, size and position of a certain image. In the above equation

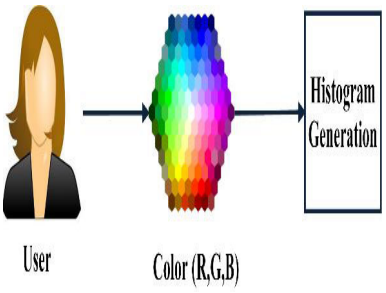

FIGURE 4. Histogram generation.

TABLE 1. Simulation output of colour features.

S. No	Feature values
1	7
2	0
3	5
4	66
5	119
6	32
7	53
8	86
9	1
10	6

histogram image:

$$\|HG_1 - HG_1'\| = \sum_{i}^{m} \mid HG_1(i) - HG_1'(i) \mid, \quad (2)$$

where $HG_1[i]$- is the histogram of image i,$HG_1'[i]$ represents similar histogram.

The image color space is discretised such that n distinct (discretised) colors are observed. Figure.4 shown in a color histogram H is a vector $(h_1, h_2, \ldots, h_n)$ in which each bucket H_i contains the number of pixels of color i in the image. For a given image G_1, color histogram HG_1 is a compact summary of the image. The most analogous images of G_1 are identified, and image G_1' is returned along with the similar color histogram HG_1' by querying the database of images. Characteristically, color histograms are compared by using either the sum of squared differences (L1 distance) or the sum of the absolute value of differences (L2 distance). Here, L2 distance is used. Thus, differences are assumed to be weighted consistently through dissimilar color buckets for effortlessness. As shown in Table 1. These color feature values give the extracted intensity information about the query image.

D. COLOR AUTOCORRELOGRAM

This technique assumes that the spatial correlation of dual colors changes with the change in their distance. From the image pixel, the color histogram is taken and acquires probability to capture the color distribution in an image and is excluded from other spatial correlations. Let I be an$M \times M$ image [31]. The colors in I are quantized into n colors $D_1, D_2, \ldots D_n$. For the pixel $(A) = (x, y)\varepsilon I$, let I_p denote the

color. Let $I_c =^{\Delta} \{A \mid I(p) = C\}$. For pixels $A_1 = (x_1, y_1)$ and $A_2 = (x_2, y_2)$, $\mid A \mid = max\ (\mid x_1 - x_2 \mid, \mid y_1 - y_2 \mid)$. We present Color autocorrelogram provides the contribution of individual color intensity for the entire image [36]. These Correlogram highly tolerate and adapts the huge changes in shape or appearance caused in viewing positions, camera zooms, etc.

E. COLOR MOMENTS

This technique measures different images on the basis of their color features. The calculation uses the color moment to determine the color similarity between images [36]. In our study, we avoid analyzing the complete color distribution. We improve the speed of image retrieval. The characterization of single-dimensional color distributions with first three moments seems to be more robust and computed really faster than other conventional histogram methods. The color moment can be measured:

$$\mu = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} Q_{ij}}{MN}, \tag{3}$$

where Q_{ij} indicates the pixel values of the *ith* row and *jth* the column, M refers to the total number of rows in the image and N denotes the total number of columns in the image Standard Deviation(S_d).

$$StandardDeviation = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} - ((Q_{i,j} - \mu)^2)}, \tag{4}$$

where Q_i, j indicates the pixel value of the ith row and jth column, M is the number of rows of an image and N is the number of columns of the image, μ denotes the mean of the image pixel. Skewness is used to measure the asymmetry distribution among the Skewness is used to measure the asymmetry distribution amongst the pixels of an image:

$$Skewness = 3 \sqrt{\frac{1}{M} \sum_{i=1}^{M} - ((Q_{i,j} - M_i)^2)}, \tag{5}$$

where, $Q_{i,j}$ indicates the pixel value of the ith row, M is the number of rows of an image, and M_i denotes the mean of the image.

F. COLOR COHERENCY VECTOR

Color coherence is defined as the pixel degree of color, which is the very large, similarly colored region members. The coherent pixels are a portion of a substantial neighboring region, whereas incoherent pixels are not. A coherence vector of color signifies the process of classification for individual color in the image [37]. We present color coherency vector to separate coherent from incoherent pixels, CCV provide a better distinction than other traditional color histograms. A DB with 15000 number of images can be processed using CCV's in under 2s. Thus a CCV can provide remarkable responses for gaining color histograms.

G. TEXTURE EXTRACTION

The texture feature is extracted from the images. Here, three methods are used for texture extraction. Related texture extraction works are the learning of collection, organization, analysis, and interpretation of the image [37]. We utilized as it intends to capture the repetitive and granular patterns of surfaces presented within an image.

H. AFFINE SHAPE ADAPTATION

This approach is applied to adopt the shape of the smoothing kernels present in an affine group of smoothing kernels iteratively to the local image structure of a definite image point [38]. In our method, the achievement is the reduction in computational cost.

I. SIFT DESCRIPTORS

SIFT includes the feature descriptor and feature detector. The detector retrieves the frames or key points from an image. The image shape is extracted by a shape descriptor, and it uses two techniques [39]. In our proposed SIFT descriptors, As it extracts the significant key points from an image for feature description, it is well suited for matching different images and the objects. Their extracted features are remarkable irrespective to change in scaling, orientation, or any other transformations. Thus it is well suited for image retrieval especially for face description.

J. FOURIER DESCRIPTOR

Each image of the dataset is converted into its image component as an abstract representation. The Fourier descriptor is rendered invariant against translation, scale rotation and starting point [40]. In our method one of the advantages is the automatic retrieval process instead of the traditional context-based approach which requires greater time consumption. It is one of the best attributes according to its characteristics with rotation, translation, scaling, and the starting point.

Properties of Fourier descriptor consider the M contour points of an image component as a direct function $x(n) = (x_1(n), x_2(n))$. The discrete complex function $U(n)$ is expressed as:

$$U(n) = x_1(n) + iX_2(n). \tag{6}$$

$U(n)$ is transformed into the frequency domain by using a discrete Fourier transform (DFT). Inverse discrete Fourier transform (IDFT) does not have any loss. DFT and IDFT are defined as $U(n)$ and $a(k)$.

$$U(n) = \sum_{k=0}^{N-1} a(k) e^{j2\pi kn/N}, \quad 0 \le n \le N-1, \tag{7}$$

where $U(n)$ is the discrete Fourier transform, which varies from 0 to N-1, e represents an exponential function of $2\pi kn$, and N is the total number of pixels and k varies from 0 to $N-1$.

$$a(k) = \frac{1}{N} \sum_{n=0}^{N-1} u(n) e^{-j2\pi kn/N} 0 \le n \le N-1, \tag{8}$$

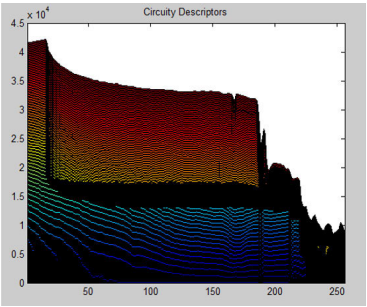

FIGURE 5. Circulatory descriptors.

TABLE 2. Best features.

Sal. No	Best Feature values
1	3.833×10-4
2	0.1547
3	0.0028
4	1.1429
5	1.1429
6	0.0028
7	1.1429
8	1.1429
9	0.0028
10	1.1429

$a(k)$ is the inverse discrete Fourier transform where N is the total number of pixels in the image. n is the index of pixels that varies from 0 to $N-1$. The coefficient a (k) is termed as the Fourier descriptor. It represents the discrete shape contour in the Fourier domain.

K. CIRCULATORY DESCRIPTOR

A circulatory descriptor is used for the shape description of an image. Shape prediction is improved and becomes fast with this technique [36]. We present these features which are considered to be scalable and rotatable with simple intensity comparisons that make the image resolution a distinct one. It can be successfully implemented in processing various illumination changes. As these attributes show superior advantages as defined, they are chosen from others. The values of best feature values are mentioned in Table 2. As shown Figure 5. The Best features are extracted the unique information about the query which are the best with reduced dimension.

In this work, different approaches have been used to effectively process image and enhance the extraction of features on color, texture, and shape.

IV. OPTIMIZATION PROCESS

Usually, DBACO is used for feature optimization. In this study, Multiple ant colony optimization (MACOBTC) is used for selecting best features for color Images. Here, the output are the resultant optimized features with reduced dimensions. The proposed method MACOBTC presents a technique of BTC depending on the multiple ant colony optimizations (MACO) to acquire images that have better visual quality. In this study, we use a modified algorithm as per the features that are optimized. Thus, the input is the features, and the output is the reduced dimension size.

The (ACO) ant colony optimization algorithm is a method that is probabilistic for the purpose of resolving issues related to computation, which in turn can be minimized for the identification of good paths over the graph. Following the ACO idea, this MACO process is capable of performing in a competitive manner, thereby solving the issues regarding the discrete domain due to its structure space of the random binary search that is unique.

The MACO solution is signified by a binary bit string by means of selecting each node from both values that are positive, i.e., 0 or 1. Rather than allocating an ant for searching all positions, an ant is assigned for each position to minimize the time for determining the extracted common bitmap because each common bitmap position that is independent has a block distortion effect.

Each position has only two choices for the common bitmap that is 0 or 1. After this, the individual ant is modified randomly to 0 or 1, and the initial bitmap that is common is constructed once all of these ants have been initialized. The ant is selected each time by changing the values to the opposite. The pheromone for this ant on behalf of selecting 0 or 1 is altered by calculating the estimation value for the resultant block.

By relating the restructured pheromone value of this ant, the best among 0 or 1 is identified. Once all the ants' pheromones are updated, the solution of near-optimal is created in relation to the pheromone matrices that are updated. Although the binary matrix initialization in each loop is random, the value of one single position is reversed; therefore, the final optimal outcomes are not affected. This method is detailed as below.

V. GREEDY LEARNING OF DEEP BOLTZMAN MACHINE

Generally, there will be less interest in complex learning, which is a totally connected machine of Boltzmann. In spite of this, deep multi-layer Boltzmann machine learning is considered. As shown in Figure 6, the succeeding layer in the left panel is responsible for capturing the complicated outputs of the preceding layer, correlating the higher-order hidden features to the activities at the lower layer in the right panel. Deep Boltzmann machines are usually fascinating due to various factors.

Firstly, as like the network of deep belief, this, in turn, has the potential of internal depiction learning which becomes complex increasingly that is regarded as the most prominent manner of resolving the speech and object recognition issues.

Then, the representation of high-level could be made from the huge supply of the sensory labels that are unlabeled and the labeled data can be formerly utilized for fine-tuning the desired task at the hand in a slight manner.

At last, unless like the network of deep belief, the procedure of inference that is approximate along with the first bottom-up will be integrated into that of the top-down

Algorithm 1 MACOBTC for Color Image

Input: Extracted features with 1xm dimension f_ij
Output: Optimized features C^*_{ij}.
step 1: Two mean values, f_Handf_L, which are defined in Equation (1):

$$f(x) = \begin{cases} k = \dfrac{1}{M * N} \sum_{i=1}^{M} \sum_{j=1}^{N}, & fij \\ f_H = \dfrac{1}{n_H} \sum F_{ij}, & \text{if, } f_{ij} \geq k, \\ f_L = \dfrac{1}{n_L} \sum F_{ij} & othewrwise, \end{cases} \tag{9}$$

where u is the block is mean value and f_{ij} is the pixel value located at the position (i, j) in each features.

step 2: Generate three pairs of quantization values, i.e., (x_{RH}, x_{RL}), (x_{GH}, x_{GL}), (x_{BH}, x_{BL}), and a random binary vector. $C = (C_{ij}|C_{ij} \in (0, 1), 1 \leq i \leq m, 1 \leq j \leq n)$.
As an initial common bitmap, calculate the initial MSE using Equation (2):

$$MSE = \sum_{Cij=0} (f_{ij} - f_L)^2 + \sum_{Cij=0} (f_{ij} - f_H)^2, \tag{10}$$

where $y_{ij} = (R_{ij}, G_{ij}, B_{ij})$ are the original values from the three channels at the same position.

step 3: Attain the image Z by interchanging the tagged pixel as 1 with y_H and 0 with y_L. Then, a value of vector C is selected and alter it into the contradictory rate to make a transitory vector. Recreate the Z image by interchanging the pixel labeled as 1 with y'_H and 0 with y'_L and acquire the image Z'.

step 4 : Estimate the consistent new MSE_{ij} for this vector by using Equation (2). Generate two pheromone conditions, δ_{ij} (0) and δ_{ij} (1), and each component in these vectors is selected δ_{ij} $(k) = 0.5(k \in 0, 1)$ because each ant located in vector C has the same probability of selecting path "0" or path "1" in the initial state.

step 5: Calculate the pheromone that is incremental $\Delta\delta_{ij}(k)$ as per the Equation (3), and update the primary pheromone vectors δ_{ij} (0) and $\delta_{ij}(1)$ using Equation (4):

$$\Delta\delta_{ij}(k) = \begin{cases} \dfrac{1}{MSE}, & if\ C_{ij} = k, \\ \dfrac{1}{newMSE_{ij}}, & othewrwise, \end{cases} \tag{11}$$

where $k \in 0, 1, i \in [1, m], j \in [i, n]$.

$$\delta_{ij}(k) = \delta_{ij}(k) + \Delta\delta_{ij}(k), \tag{12}$$

where $k \in 0, 1, i \in [1, m], j \in [i, n]$.

step 6: Update the common image features according to the following rule:

$$c*_{ij} = \begin{cases} 1, & if\ \delta_{ij}(1) > \delta_{ij}(0), \\ 0, & othewrwise, \end{cases} \tag{13}$$

where $i \in [1, m], j \in [i, n]$.

step 7: Replicate Steps 3 through 6 until all of the standards in vectors C are distributed. Likewise, all of the image features are handled in the same way. So far, the near-optimal common optimized feature $C*$ for the entire features has been produced.

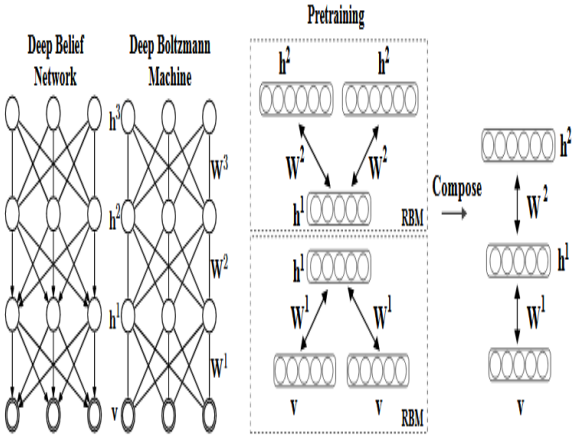

FIGURE 6. A three-layer DBM.

feedback. This will allows the deep Boltzmann machine for propagating in a good manner uncertainly and therefore this will deal much robustly with the inputs that are ambiguous.

The features of training and testing mean values are depicted as shown in the Figure 6. The energy of the status $(u, g1, g2)$ is defined as:

$$E(v, h^1, h^2, \Psi) = v^T\ w^1\ h^1 - h^1\ w^2\ h^2, \tag{14}$$

where $\Psi = (w^1, w^2)$. The probability that a pattern assigned to a vision vector w is:

$$p(w, \Psi) = \frac{1}{z(\Psi)} \sum h^1, h^2 exp(-E(v, h^1, w^2, \Psi)). \tag{15}$$

The conditional ranking distributions over a visional and two sets of invisible units of measure are given by logistic functions:

$$p(h^1_j = 1|v, h^2) = \sigma(\sum_i w^1_{ij} v_i + \sigma \sum_m s^1_{im} h^2_j), \tag{16}$$

$$p(h^1_m = 1|h^1) = \sigma(\sum_j w^1_{im} h^1_i), \tag{17}$$

$$p(u_i = 1|h^1) = \sigma(\sum_j w^1_{ij} h_i). \tag{18}$$

As to the estimated learning of the maximum likelihood, still there is a need for applying learning process for overall Boltzmann machines labeled above, however, it might be somewhat unhurried, predominantly once the unknown elements form layers which become progressively isolated as of the noticeable elements. Here is, though, a rapid technique to modify the pattern constraints to the sensible standards that are depicted in the further part.

A. GREEDY LAYERWISE PRETRAINING OF DBM

A greedy method was introduced, that is having the layer-by-layer learning un-supervised process that involves a one layer GDBM's stack learning at a time. Afterward, the GDBM's stack has been well-read, the entire stack could be observed as a distinct model of probabilistic, named a " DBN".

Remarkably, mentioned the method is not the deep Boltzmann machine. The upper two categories of will form into a Greedy Learning of Deep Boltzman Machine (GDBM) which is considered as the pattern of graphical view that is undirected, however, the bottom layers that are from the generative directed pattern.

Once, the initial GDBM in the stack is being learned, this generative model could be inscribed as:

$$p(v, \Psi) = (\sum_{h^1} p(h^1; w^1))p(v/h^1; w^1), \tag{19}$$

where $p(h^1; w^1) = \sum_v p(h^1, v; w^1)$ is an implicit preceding through h^1 well-defined in the constraints.

If the GDBM second in the stack trace a substitutes $p(h^1; w^1)$ by $p(h^1;^2)\sum_{h^2} p(h^1, h^2; w^2)$, and if it performs correctly, $p(h^1; w^2)$ will turn out to be an improved model of the combined subsequent dispersal over h^1, wherever the combined posterior is just is not factorial combination of the factorial posteriors for all situations of training, $\frac{1}{N}\sum_n p(h^1, |v_n; w^1)$, as the second GDBM is substituting $p(h^1; w^1)$ via means of a superior pattern, it would be possible to infer $p(h^1; w^1, w^2)$ via averaging the two patterns of h^1 which can be finished roughly by means of $1/2$ h^1 bottom-up and $1/2$ w^2 top-down. By means of bottom-up w^1 and w^2 top-down long for aggregate to double-counting the indication since g^2 is reliant on u.

To modify DBM pattern parameters, greedy layer-by-layer pre-training technique is presented on learning the GDBM's stack, however by a minor alteration that is presented to eradicate the problem of double-counting once bottom-up and top-down impacts are consequently united. For the GDBM of lower-level, the tie the visible-to-hidden weights and input are being doubled, at the right panel. In this altered GDBM using parameters that are tied, the restrictive allocations over the visible and hidden conditions are well-defined as:

$$p(h_j^1 = 1|v) = \sigma(\sum_i w_{ij}^1 + \sum_i w_{ij}^1), \tag{20}$$

$$p(v_i^1 = 1|h^1) = \sigma(\sum_j (w_{ij}^1 h_j). \tag{21}$$

Learning of contrastive divergence performs better and the enhanced GDBM is worthy at renovating its data that are trained. On the other hand, for the GDBM top-level the number of invisible units is doubled. The uncertain disseminations for this pattern proceeds the method:

$$p(h_j^1 = 1|h^2) = \sigma(\sum_m w_{jm}^1 h_{m1}^2 + \sum_m w_{jm}^1 w_m^2), \tag{22}$$

$$p(h_m^2 = 1|h^1) = \sigma(\sum_j (w_{jm}^2 h_j^1). \tag{23}$$

Once these two elements are collected to custom a particular scheme, the over-all feedback impending into the initial hidden layer is split that in turn leads to the respective conditional distribution over h^1:

$$p(h_j^1 = 1|v, h^2) = \sigma(\sum_i v_{ij}^1 v_i + \sum_m w_{im}^2 h_m^2). \tag{24}$$

The restrictive allocations over u and g^2 persist the similar as distinct in the above equations. Perceive that the conditional dispersals that are well-defined by the self-possessed technique are accurately the similar uncertain dispersals well-defined by means of the DBM.

Consequently, we will avidly pre-train the two improved GDBMs, the deep Boltzmann machines that are directionless models using symmetric weights. Once training a stack greedily of these two GDBMs, only alteration is required to the initial and the latter GDBM at the stack. For the entire in-between GDBMs, their connecting weights are simply halved in both ways once constituting them to custom a deep Boltzmann machine. Greedily pre-training loads of a DBM in this method helps two determinations.

Specifying a data vector on the noticeable neurons, the respective neuron in the hidden layer can be triggered in a particular bottom-up pass by magnifying the bottom-up feedback to recompense the deficiency of top-down response. The fast approximate inference is employed by utilizing the mean-field method, which congregates much earlier than the initialization in a random way. Hybridizing ACO and BTC in optimization is a novelty and implementing greedy learning for Boltzmann is also a novelty. We provide betterment of result in terms of accuracy, precision, recall, and sensitivity.

VI. PERFORMANCE ANALYSIS

This section discusses the performance of the proposed technique in the dataset. A relative analysis of the suggested dataset is performed to evaluate the suggested technique and the existing a priori algorithm. The metrics are true positive (TP), true negative (TN), accuracy, false positive (FP), precision, false negative (FN), specificity, sensitivity, and recall.

The coefficients, such as Jaccard, Kappa, and Dice, are analyzed. Notably, CNN classification provides effective results. The time consumption of the system has been reduced to approximately 14.57 ms per image, where the time consumption of the existing system is 18.31 ms per image [41].

A. DATASET DESCRIPTION

In this work, European 1M,[1] Flickr,[2] Corel 1K[3] and LFW.[4] datasets are used in the simulation. The Corel 1K dataset is also utilized to analyze the performance of the proposed method.

B. PARAMETER DESCRIPTIONS

Accuracy is defined as the closeness of a measured value to the standard values. This variable is termed as the weight arithmetic mean of precision and inverse precision [42].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{25}$$

[1] http://image.ntua.gr/iva/datasets/ec1m/.
[2] http://image.ntua.gr/iva/datasets/flickr_logos
[3] https://blog.csdn.net/garfielder007/article/details/51483759.
[4] http://vis-www.cs.umass.edu/lfw/

Sensitivity is also called TP and used to measure the ratio of positives that are correctly identified features in the image [42].

$$Truepositiverate = \frac{TP}{TP + FN}. \quad (26)$$

Precision is a basic measure for evaluating the performance of classification techniques. In image retrieval, precision is the fraction of retrieved images that are relevant to the queries images. Precision is calculated as follows:

$$Precision(X(I_i)) = \frac{A}{B}. \quad (27)$$

Average retrieval precision is calculated as:

$$ARP = \frac{1}{|\ Ds\ |}\sum_{i=1}^{|Ds|} X(I_i). \quad (28)$$

where Ds is the total number of correctly retrieved images.

Recall measures the prediction capability of models and is mainly used to select the instance of a certain class from a dataset [42]. This parameter is also called sensitivity and calculated as follows:

$$Racall(T(I_i)) = A/(Total\ number\ of\ relevant\ images\ ) \quad (29)$$

Average retrieval rate is calculated as;

$$ARR = \frac{1}{|\ Ds\ |}\sum_{i=1}^{|Ds|} T(I_i), \quad (30)$$

where A is the number of relevant images retrieved, B is the total number of images retrieved, ARP is the average retrieval precision, TP is the number of true positive images, TN is the number of true negative images, FP is the number of false positive images, FN the number of false negative images, and Ds is the number of all images in the database.

C. JACCARD, DICE AND KAPPA COEFFICIENTS

Jaccard coefficient distance measures the dissimilarity between images 'A' and 'B'. This coefficient is calculated as follows:

$$J_C(A, B) = \frac{|\ A \cap B\ |}{|\ A \cup B\ |}, \quad (31)$$

where the Jaccard coefficient J_c ranges between 0 and 1. The value is 1 when the two images are identical; the value is 0 when the two images are completely different.

Dice coefficient measures the spatial overlap level between two regions that are labeled as similar over the average volume of these regions:

$$D(A, B) = \frac{2|\ A \cap B\ |}{|\ A\ | + |\ B\ |}. \quad (32)$$

Kappa coefficient measures the difference between the observed agreements of two images. A high kappa coefficient indicates a high classification rate. The coefficient is calculated as follows:

$$KappaCoef = \frac{((n * sumA_{i+})) - su(A_{i+}A_{+i}))}{n^2 - sum(A_{i+}A_{+i})}. \quad (33)$$

TABLE 3. The comparison a priori algorithm and GDBM.

Parameters	A priori algorithm	GDBM
TP	97	882
TN	45	14
FP	638	3
FN	120	1

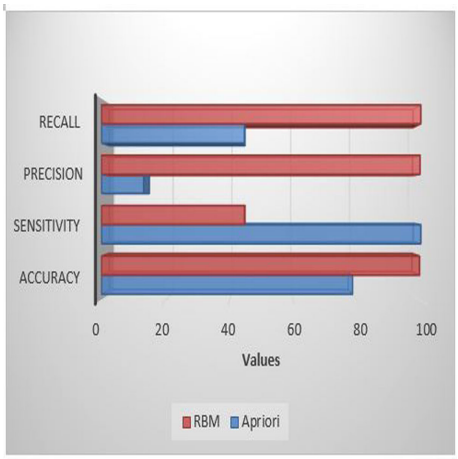

FIGURE 7. Performance analysis.

Each dataset is used to analyse the result and compared with existing a priori algorithm.

D. TRUE POSITIVE, TRUE NEGATIVE, FLASE POSITIVE AND FLASE NEGATIVE

The TP value can be estimated when the positive class in the actual label is correctly predicted as the positive class. Similarly, the TN value can be approximated when the negative class is correctly predicted as the negative class. The FP value can be estimated when the positive class in the actual class is incorrectly predicted as the negative class. Similarly, the FN value can be approximated when the negative class in the actual class is incorrectly predicted as the positive class shown in Table 3.

E. ACCURACY, PRECISION AND RECALL

Figure 6 shows the performance of the suggested approach in the European 1 M dataset in terms of accuracy, precision and recall values. These parameter values are compared with those of the a priori algorithm. Accuracy, sensitivity, precision, and recall are increased by 25%, 55%, 86.7%, and 55.2%, respectively. These results are better than those of the existing a priori algorithm. The best results of the proposed approach can be obtained by the posterior distribution for an unbiased sample.

F. JACCARD, DICE AND KAPPA

Figure 7 illustrates the performance measures of the existing and proposed classification algorithms. Jaccard, Dice and kappa values are increased by 84.15%, 72.67%, and 19.41%, respectively. This result shows that the proposed

TABLE 4. Result of the a priori algorithm and GDBM.

Parameters	A GDBM	Apriori Algorithm
Sensitivity	99.8867	44.7005
Specificity	82.3529	6.5886
Precision	99.6610	13.1973
Recall	99.8867	44.7005
Jaccard Coefficient	0.9956	0.1578
Dice Coefficient	0.9978	0.2726
Kappa Coefficient	0.9945	0.7948
Accuracy	99.5000	77.8908

TABLE 5. Comparison for flicker dataset.

Flickr Logos dataset						
Class	Precision			Recall		
	SURF [42]	SIFT [42]	Proposed	SURF [42]	SIFT [42]	Proposed
Google	0.88	0.84	0.96	0.29	0.28	0.33
FedEx	0.84	0.78	0.91	0.28	0.26	0.34
Porsche	0.82	0.78	1	0.27	0.26	0.325
Red Bull	0.84	0.78	0.892	0.28	0.26	0.334
Starbucks	0.82	0.96	1	0.27	0.32	0.326
Intel	0.82	0.78	0.874	0.27	0.26	0.31
Sprite	0.84	0.74	0.92	0.26	0.25	0.298
DHL	0.74	0.74	0.81	0.25	0.25	0.285
Vodafone	0.82	0.72	0.92	0.25	0.24	0.296
NBC	0.76	0.68	0.805	0.25	0.22	0.318
Average	0.82	0.78	0.9071	0.27	0.26	0.319

technique is better than the existing algorithm because the existing a priori algorithm requires more datasets for training to obtain improved results. However, the proposed algorithm improves accuracy mechanically and does not require additional datasets.

The results in Table 4 show that the performance of TP, TN FP, FN, accuracy, precision, recall, Jaccard, Dice and Kappa in the proposed GDBM better than the existing a priori algorithm.

VII. RESULTS AND DISCUSSION

The comparative analysis of the proposed and existing techniques using various datasets is discussed in this section. Table 5 presents a comparison between the proposed and existing techniques using the Flickr Logos dataset. Table 6 shows the comparison between the proposed and existing techniques utilizing the Corel 1K dataset.

Table 7 illustrates the classifier comparison on the Corel 1K dataset. Table 8 illustrates the comparison between the proposed and existing techniques using the LFW dataset. The results indicate that our proposed technique produces better results than the existing technique.

A. COMPARISON ON DATASET EUROPEAN 1M

The European 1M dataset comprises 909,940 geo-tagged images from 22 European cities scuttled from Flickr with geographic queries covering a window of each city center.

Figure 8. Depicts the simulation results on the European 1M dataset. These results indicate that the proposed method is better than the existing method in terms of precision, recall, TP, TN, FP, and FN. The proposed method predicts the image correctly from the dataset where the image is retrieved.

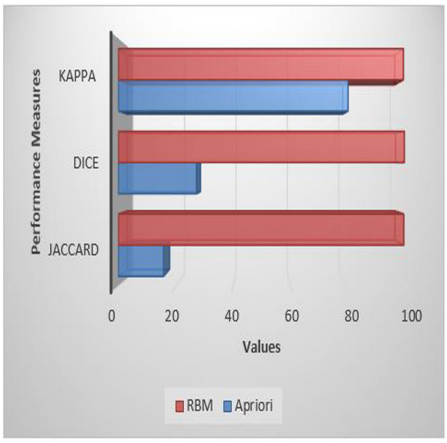

FIGURE 8. Performance measures Jaccard, Dice and kappa.

FIGURE 9. Simulation results for the European 1M dataset of Torre Agbar.

The TP and TN values of the proposed technique prove the images recognized correctly and acquire better retrieval of images.

B. COMPARISON ON DATASET FLICKR

In the Flickr Logos 27 dataset, 10 classes are selected randomly (FedEx, Google, Red Bull, Starbucks, Porsche, Intel, Sprite, Vodafone, NBC, and DHL) for the training stage. Figure 9 shows the precision and recall values of the existing SURF and SIFT [42] techniques and the proposed method. Ten classes of the test dataset images are considered. Table 5 shows the average values of precision and recall and illustrates that our proposed work extracting important features by multiple ant colony optimizations (MACO) has the best results among the compared methods. Table 5 and Figure 9 illustrate a comparative analysis of the Flickr Logos dataset for the proposed method and other existing methods, namely, SIFT and SURF, in terms of precision and recall.

In Table 5, the suggested technique illustrates a substantial enhancement related to other existing methods in terms of precision and recall. For the Google class, the precision values are 0.88, 0.84 and 0.96 for SURF, SIFT and the suggested method (GDBM), respectively.

TABLE 6. Comparison for the Corel 1K dataset.

Corel 1K dataset								
Semanticcategory	Performancemeasures	fusion of visual words	GMM-EM	SIFT LBP	Color texture	EODH-color SIFT	HOG-LBP	Proposed
Africa	Precision	74.19	72.5	57	72.6	74.6	55	100
	Recall	14.83	14.5	11.4	16.1	14.92	11	100
Beach	Precision	75.38	65.2	58	59.3	37.8	47	100
	Recall	15.07	13.04	11.6	20.3	7.56	9.4	68.0272
Buildings	Precision	75.82	70.6	43	58.7	53.9	56	100
	Recall	15.16	14.12	8.6	19.1	10.78	11.2	100
Buses	Precision	81.59	89.2	93	89.1	96.7	91	100
	Recall	16.31	17.84	18.6	12.6	19.34	18.2	100
Dinosaurs	Precision	100	100	98	77.2	99	94	83
	Recall	20	20	19.6	10.9	19.8	18.8	98.8095
Elephants	Precision	96.7	70.5	58	99.3	66	49	99
	Recall	19.34	14.1	11.6	16.3	13.2	9.8	100
Flowers	Precision	93.21	94.8	83	70.2	92	85	100
	Recall	18.64	18.96	16.6	12.9	18.4	17	100
Horses	Precision	85.25	91.8	68	92.8	87	52	91
	Recall	17.05	18.36	13.6	14.4	17.4	10.4	100
Mountains	Precision	80.47	72.25	46	85.6	58.5	37	82
	Recall	16.09	14.45	9.2	23.6	11.7	7.4	100
Food	Precision	81.32	78.8	53	56.2	62.2	55	97
	Recall	16.26	15.76	9.2	14.8	12.44	11	100
	Recall	16.26	15.76	9.2	14.8	12.44	11	100

TABLE 7. classifier comparison.

Class	Base Method	DNN	Neuro-Fuzzy	Proposed GDBM
People	57	76	43	100
Beaches	51	61	88	100
Buildings	44	45	69	100
Buses	79	80	81	100
Dinosaurs	99	100	100	83
Elephant	58	76	62	99
Flower	60	89	92	100
Horses	53	79	33	91
Mountains	38	38	80	82
Food	63	72	86	97

TABLE 8. Comparison for the Corel 1K dataset.

Coral 1,000 dataset	
Method	Average Precision Rate
Guoping Qiu [60]	0.595
Gahroudi, Mahdi Rezaei [61]	0.396
Yu, F-X [62]	0.717
Silakari, Sanjay [63]	0.56
Lin, Chuen-Horng [64]	0.727
Jhanwar, N et al. [65]	0.526
Huang, Po-Whei [66]	0.532
Lu, Tzu-Chuen [67]	0.665
Chiang, Te-Wei [68]	0.533
Guo, Jing-Ming2015 [69]	0.797
Guo, Jing-Ming2016 [70]	0.792
Proposed model	**0.854**

Similarly, the recall values for SIFT, SURF, and the offered method are 0.29, 0.28 and 0.33, respectively. The value of the proposed method is improved compared with those of the other approaches. Comparison for the Flickr dataset recall value demonstrates a better image retrieval in terms of color, texture and shape feature extraction.

C. COMPARISON ON DATASET COREL 1K

In this section, we consider 10,800 images, more than the previous work (Zhang et al. 2012). The Corel 1K repository

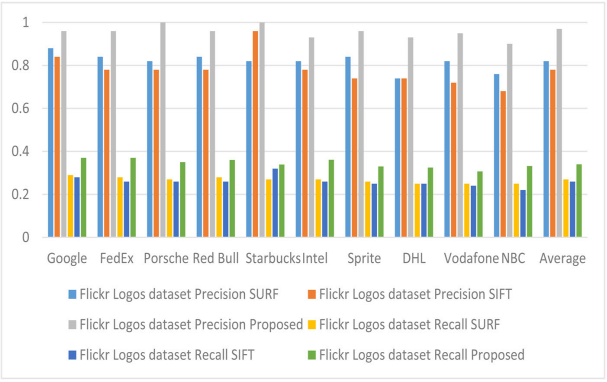

FIGURE 10. Comparative analysis for flicker dataset.

covers approximately 1,000 images in 10 semantic groups, and the respective semantic group covers 100 images with a size resolution of 256 × 384.

Figure 10 shows the image model from every semantic group of the image source Corel 1K. A set of 700 random images from the Corel 1K repository image are chosen as a training set, and the residual 300 arbitrary imageries are used as an assessment set to check the recommended method. Table 6 and Figure 10 illustrate a comparative analysis of the Corel 1K dataset for the proposed method and the other traditional approaches.

As indicated in Table 6 and Figure 10, the proposed method shows a noteworthy enhancement related to other existing methods in terms of precision and recall. Table 6 depicts the performance measures for the Corel 1K dataset. For the Africa semantic category, the precision and recall values are 74.19 and 14.83 for the recommended approach based on the fusion of visual words, 72.5 and 14.5 for GMM-EM, 57 and 11.4 for SIFT-LBP, 72.6 and 16.1 for color texture, 74.6 and 14.92 for EODH-color SIFT, 55 and 11 for HOG-LPB and 100 and 100 for the proposed method, respectively.

FIGURE 11. Results of the Corel 1K dataset.

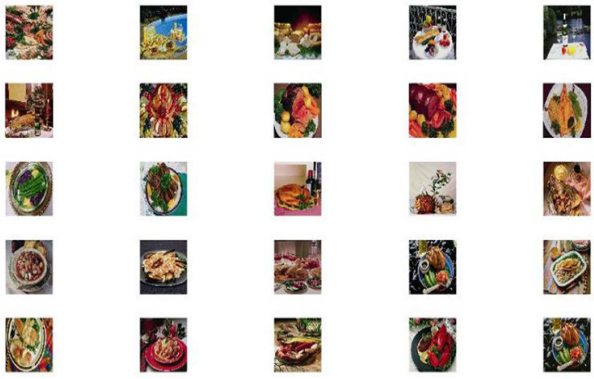

FIGURE 12. Results of the LFW dataset.

The proposed method achieves better outcomes than existing methods in the Corel 1K dataset on the extraction of NBPC features, which preserves the global spatial features. The proposed method achieves better outcomes than existing methods in the Corel 1K dataset on the extraction of NBPC features, which preserves the global spatial features, as shown below in Table 7, 8.

Comparison for the Corel 1K dataset of proposed with existing approaches using Average precision rate (APR). Table 8 is comparison for the Corel 1K dataset of proposed with existing approaches using average precision rate (APR). The proposed method achieves excellent outcomes than existing methods in the Corel 1K dataset. We achieved excellent results due the color histogram, color auto correlogram, color moments, color Coherency Vector, texture extraction, affine shape adaptation, SIFT Descriptors, Fourier Descriptor, Circulatory with a novel relevant feature selection are utilized to train a Greedy Learning of Deep Boltzmann Machine classifier (GDBM) that assist improvement performance Precision.

D. COMPARISON ON DATASET LFW

The dataset covers more than 13,000 images of faces collected from the web. Each face has been considered by the name of the person visualized. The comparative analysis is reviewed here to determine its characteristics Figure 11. Table 9 illustrates a comparative analysis of the LFW dataset. Table 9 and Figure 12 compare the accuracy of our proposed GDBM method with that of existing methods. The proposed method's accuracy is 98.82, which is improved compared with other traditional methods.

The accuracy of the proposed method is high due to its ideal face detection approach. Thus, the LWF dataset is preferred for its effective accuracy. The proposed method achieves better performance due to its better classification accuracy. Effective feature extraction in terms of color, texture, and shape is the main reason to acquire better results compared to existing LWT datasets .

Figure 13 depicts the execution time for the four proposed datasets. The execution time on the European dataset is 23.6992. For the Corel dataset, the execution time is 24.8954. The execution time for the Flickr dataset is 71.55. The LFW dataset attains an execution time of 8.2824. The execution

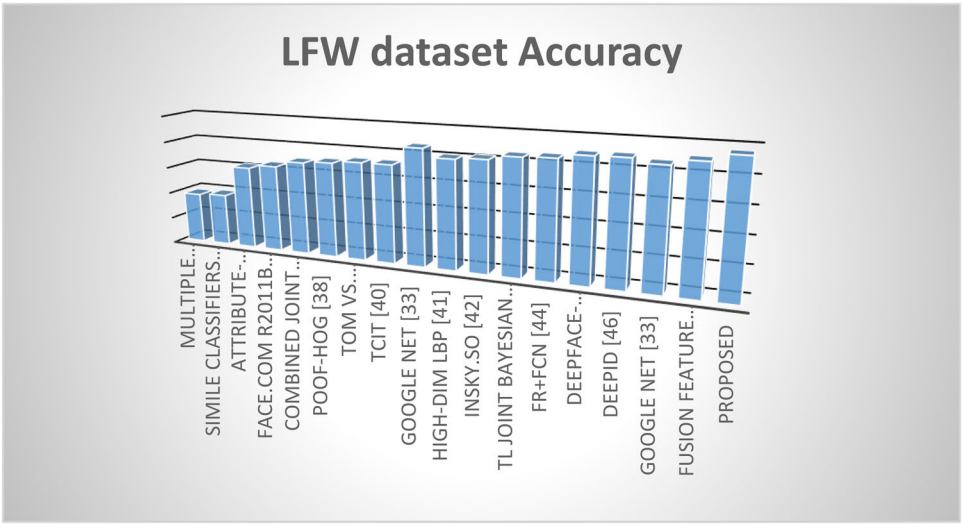

FIGURE 13. Comparison accuracy with existing and proposed GDBM methods.

TABLE 9. Comparison for the LFW dataset.

LFW dataset	
Method	Accuracy
Multiple Le+Comp [43]	84.45
Simile Classifiers [44]	84.72
Attribute-Predict [45]	90.57
Face.com R2011b [46]	91.3
Combined Joint Bayesian [46]	92.42
Poof-Hog [47]	92.8
Tom vs. Pete+Attribute [48]	93.3
TCIT [48]	93.33
Google Net [47]	96.7
High-Dim LBP [49]	95.17
Insky.so [50]	95.51
TL Joint Bayesian [51]	96.33
FR+FCN [50]	96.54
DeepFace-Ensemble [53]	97.35
DeepID [53]	97.45
Google Net [54]	96.7
Fusion feature [55]	97.68
proposed GDBM methods	98.92

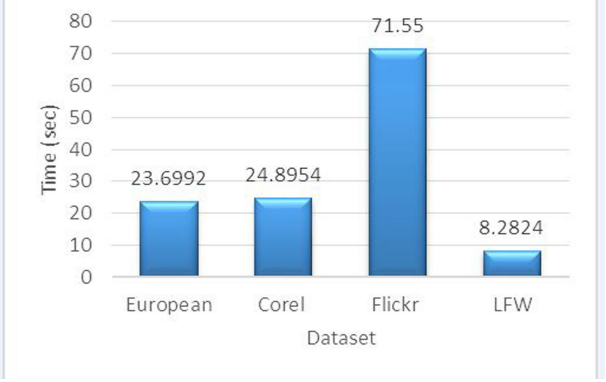

FIGURE 14. Simulation time of the proposed method.

time the European dataset is 23.6992 because it needs to retrieve 1,000 images from the dataset.

Similarly, the Corel dataset has an execution time of 24.8954. The Flickr dataset has an execution time of 71.55 due to its large quantity of data (approximately 4,500 images). LFW achieves a shorter execution time due to the retrieval of 900 images only. Within this studying, the time consumption of the system has been reduced to about 3.75 ms per image, better than that of the existing system [41].

VIII. CONCLUSION

The image retrieval is one of the challenging research areas for retrieving an image effectively. This study overcomes several drawbacks of previous works, such as low accuracy, inability to handle complex queries, high time consumption. In this work, several novelties are presented in the classification and optimization processes to improve image retrieval by predicting an image correctly.

Initially, the median filter is used to preprocess the input image, and the resulting image is utilized to extract features, such as color, shape, and texture, including color histogram, color moments, color correlogram, coherency vectors, Fourier descriptor, circulatory descriptor, SIFT descriptors, affine shape adaptation. A correlation-based neighborhood binary pattern is introduced as a type of color histogram which can represent all colors and levels of occurrence of an image. Then, all the extracted features are optimally selected by the multiple ant colony optimization (MACOBTC) approach to obtain the best feature subset, which is used to build classifier by using the proposed novel GDBM.

Subsequently, the GDBM is used to classify the querying image, and the improved outputs are used to retrieve the most similar images from the dataset. The performance of the GDBM is examined and compared with the existing algorithms. To achieve this objective, the accuracy, precision, recall and sensitivity values are fully examined with four datasets.

The performance of the GDBM is distinguished over the existing techniques because of its remarkable accuracy, precision, and recall. Chiefly, the GDBM provides a 25% increase in accuracy compared with existing techniques, such as the a priori classification algorithm.

Presently, in this study, the proposed method has been tested on 5000 datasets. The experimental results suggest the proposed method very useful in assisting the medical field, i.e., in diagnose sections which completely depend on image retrieval. In future, this proposed method is highly probably utilized for the big data in medical field.

ACKNOWLEDGMENT

This study is done in the Chinese Education Ministry Key Lab of Image Processing and Intelligent Control, with the help of professor Chih-Cheng Hung in the Center for Machine Vision and Security Research of Kennesaw State University in the United States. The work is supported by the National Key $R-D$ Program of China(Grant No. 2017YFC0112804) and by the National Natural Science Foundation of China (Grant No. 81671768).

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

AVAILABILITY OF DATA AND MATERIAL

The data used to support the findings of this study are available from the corresponding author upon request.

FUNDING

Not applicable.

REFERENCES

[1] Z. Xia, N. N. Xiong, A. V. Vasilakos, and X. Sun, "EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing," *Inf. Sci.*, vol. 387, pp. 195–204, May 2017.
[2] M. Paulin, J. Mairal, M. Douze, Z. Harchaoui, F. Perronnin, and C. Schmid, "Convolutional patch representations for image retrieval: An unsupervised approach," *Int. J. Comput. Vis.*, vol. 121, no. 1, pp. 149–168, 2017.
[3] R. Kapoor, R. Gupta, L. H. Son, S. Jha, and R. Kumar, "Detection of power quality event using histogram of oriented gradients and support vector machine," *Measurement*, vol. 120, pp. 52–75, May 2018.

[4] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.
[5] J. Wang, K. Wang, Y. Wang, Z. Huang, and R. Xue, "Deep Boltzmann machine based condition prediction for smart manufacturing," *J. Ambient Intell. Hum. Comput.*, vol. 10, no. 3, pp. 851–861, 2019.
[6] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, "Deep image retrieval: Learning global representations for image search," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 241–257.
[7] J. Melchior, N. Wang, and L. Wiskott, "Gaussian-binary restricted Boltzmann machines for modeling natural image statistics," *PLoS ONE*, vol. 12, no. 2, 2017, Art. no. e0171015.
[8] K. Liao, G. Liu, and Y. Hui, "An improvement to the SIFT descriptor for image representation and matching," *Pattern Recognit. Lett.*, vol. 34, no. 11, pp. 1211–1220, 2013.
[9] A. Khokher and R. Talwar, "Content-based image retrieval: Feature extraction techniques and applications," in *Proc. Int. Conf. Recent Adv. Future Trends Inf. Technol. (iRAFIT2012)*, 2012, pp. 9–14.
[10] K. Iqbal, M. Odetayo, A. James, R. Iqbal, N. Kumar, and S. Barma, "An efficient image retrieval scheme for colour enhancement of embedded and distributed surveillance images," *Neurocomputing*, vol. 174, pp. 413–430, Jan. 2016.
[11] M. P. H. Kumar and D. N. I. Modi, "Survey on content based image retrieval system using color and texture," Tech. Rep., 2017.
[12] J. Yu, Z. Qin, T. Wan, and X. Zhang, "Feature integration analysis of bag-of-features model for image retrieval," *Neurocomputing*, vol. 120, pp. 355–364, Nov. 2013.
[13] E. Uwimana and M. E. Ruiz, "Automatic classification of medical images for content based image retrieval systems (CBIR)," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, 2008, pp. 788–792.
[14] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 584–599.
[15] A. Nagathan and M. I. Manimozhi, "Content-based image retrieval system using feed-forward backpropagation neural network," *Int. J. Comput. Sci. Netw. Secur.*, vol. 14, no. 6, pp. 70–77, 2014.
[16] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1025–1039, 2013.
[17] S. K. Vipparthi and S. K. Nagar, "Color directional local quinary patterns for content based indexing and retrieval," *Hum.-Centric Comput. Inf. Sci.*, vol. 4, May 2014, Art. no. 6.
[18] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916–2929, Dec. 2013.
[19] A. Kae, K. Sohn, H. Lee, and E. Learned-Miller, "Augmenting CRFs with Boltzmann machine shape priors for image labeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2019–2026.
[20] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognit.*, vol. 46, no. 1, pp. 188–198, 2013.
[21] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 157–166.
[22] S. G. Sanu and P. S. Tamase, "Satellite image mining using content based image retrieval," *Int. J. Eng. Sci.*, vol. 13928, 2017.
[23] L. F. D. Santos, R. L. Dias, M. X. Ribeiro, A. J. M. Traina, and C. Traina, "Combining diversity queries and visual mining to improve content-based image retrieval systems: The divi method," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2015, pp. 357–362.
[24] J.-M. Guo and H. Prasetyo, "Content-based image retrieval using features extracted from halftoning-based block truncation coding," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1010–1024, Mar. 2015.
[25] S. Schuster, R. Krishna, A. Chang, L. Fei-Fei, and C. D. Manning, "Generating semantically precise scene graphs from textual descriptions for improved image retrieval," in *Proc. 4th Workshop Vis. Lang.*, 2015, pp. 70–80.
[26] A. Babenko and V. Lempitsky, "Aggregating local deep features for image retrieval," in *Proc. IEEE Int. Conf Comput. Vis.*, Jun. 2015, pp. 1269–1277.
[27] F. Zhao, Y. Huang, L. Wang, and T. Tan, "Deep semantic ranking based hashing for multi-label image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1556–1564.
[28] R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang, "Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4766–4779, Dec. 2015.
[29] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jun. 2015, pp. 27–35.
[30] L. K. Rao and D. V. Rao, "Local quantized extrema patterns for content-based natural and texture image retrieval," *Hum.-Centric Comput. Inf. Sci.*, vol. 5, Sep. 2015, Art. no. 26.
[31] L. Zhu, J. Shen, L. Xie, and Z. Cheng, "Unsupervised visual hashing with semantic assistant for content-based image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 2, pp. 472–486, Feb. 2017.
[32] P. Shádmó, P. Pozsegovics, Z. Vámossy, and S. Sergyán, "Sketch4match—Content-based image retrieval system using sketches," in *Proc. IEEE 9th Int. Symp. Appl. Mach. Intell. Inform. (SAMI)*, Jan. 2013, pp. 183–188.
[33] T. Avinash and V. Bansal, "PATSEEK: Content based image retrieval system for patent database," in *Proc. ICEB*, 2004, pp. 1167–1171.
[34] V. N. Kumar, N. Gupta, and S. Singh, "Content based image retrieval," in *Proc. Interface Conf. Adv. Comput.*, 2016, pp. 6–10.
[35] Y. Mussarat, S. Mohsin, I. Irum, and M. Sharif, "Content based image retrieval by shape, color and relevance feedback," *Life Sci. J.*, vol. 10, pp. 593–598, Apr. 2013.
[36] V. Vandana and S. Jindal, "CBIR system using color moment and color auto-correlogram with block truncation coding," *Int. J. Comput. Appl.*, vol. 161, no. 9, pp. 1–7, 2017.
[37] P. Greg, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proc. ACM Multimedia*, vol. 96, Nov. 1996, pp. 65–73.
[38] M. Dmytro and F. Radenovic, "Learning discriminative affine regions via discriminability," *CoRR*, 2017.
[39] N. Neeru and L. Kaur, "Modified SIFT descriptors for face recognition under different emotions," *J. Eng.*, vol. 2016, Jan. 2016, Art. no. 9387545.
[40] H. Fan and H. Zhu, "Separation of vehicle detection area using Fourier descriptor under Internet of Things monitoring," *IEEE Access*, vol. 6, pp. 47600–47609, 2018.
[41] C. Iakovidou, N. Anagnostopoulos, M. Lux, K. Christodoulou, Y. Boutalis, and S. A. Chatzichristofis, "Composite description based on salient contours and color information for CBIR tasks," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3115–3129, Jun. 2019.
[42] M. Alkhawlani, M. Elmogy, and H. Elbakry, "Content-based image retrieval using local features descriptors and bag-of-visual words," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 9, pp. 212–219, 2015.
[43] U. Sharif, Z. Mehmood, T. Mahmood, M. A. Javid, A. Rehman, and T. Saba, "Scene analysis and search using local features and support vector machine for effective content-based image retrieval," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 901–925, 2018.
[44] Z. Lu, J. Yang, and Q. Liu, "Face image retrieval based on shape and texture feature fusion," *Comput. Vis. Media*, vol. 3, no. 4, pp. 359–368, 2017.
[45] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2707–2714.
[46] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 365–372.
[47] Q. Yin, X. Tang, and J. Sun, "An associate-predict model for face recognition," in *Proc. CVPR*, Jun. 2011, pp. 497–504. [Online]. Available: http://doi.ieeecomputersociety.org
[48] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 566–579.
[49] T. Berg and P. Belhumeur, "Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 955–962.
[50] T. Ber and P. N. Belhumeur, "Tom-vs-pete classifiers and identity-preserving alignment for face verification," in *Proc. BMVC*, 2012, p. 7.
[51] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3025–3032.
[52] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, "A practical transfer learning algorithm for face verification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2013, pp. 3208–3215.

[53] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Recover canonical-view faces in the wild with deep neural networks," 2014, *arXiv:1404.3543*. [Online]. Available: https://arxiv.org/abs/1404.3543
[54] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.
[55] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1891–1898.
[56] Z. Ji, W. Yao, W. Wei, H. Song, and H. Pi, "Deep multi-level semantic hashing for cross-modal retrieval," *IEEE Access*, vol. 7, pp. 23667–23674, 2019.
[57] Z. Zhang, Q. Zou, Y. Lin, L. Chen, and S. Wang, "Improved deep hashing with soft pairwise similarity for multi-label image retrieval," *IEEE Trans. Multimedia*, to be published.
[58] H. Kim, C. Lee, J. Lee, J. Kim, T. Yu, G. Chung, and J. Kim, "An explicit numerical algorithm for surface reconstruction from unorganized points using Gaussian filter," *J. Korean Soc., Ind. Appl. Math.*, vol. 23, no. 1, pp. 31–38, 2019.
[59] G. Qiu, "Color image indexing using BTC," *IEEE Trans. Image Process.*, vol. 12, no. 1, pp. 93–101, Jan. 2003.
[60] M. R. Gahroudi and M. R. Sarshar, "Image retrieval based on texture and color method in BTC-VQ compressed domain," in *Proc. 9th Int. Symp. Signal Process. Appl.*, Feb. 2007, pp. 1–4.
[61] F.-X. Yu, H. Luo, and Z.-M. Lu, "Colour image retrieval using pattern co-occurrence matrices based on BTC and VQ," *Electron. Lett.*, vol. 47, no. 2, pp. 100–101, Jan. 2011.
[62] S. Silakari, M. Motwani, and M. Maheshwari, "Color image clustering using block truncation algorithm," 2009, *arXiv:0910.1849*. [Online]. Available: https://arxiv.org/abs/0910.1849
[63] C.-H. Lin, R.-T. Chen, and Y.-K. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image Vis. Comput.*, vol. 27, no. 6, pp. 658–665, 2009.
[64] N. Jhanwar, S. Chaudhuri, G. Seetharaman, and B. Zavidovique, "Content based image retrieval using motif cooccurrence matrix," *Image Vis. Comput.*, vol. 22, no. 14, pp. 1211–1220, 2004.
[65] P. W. Huang and S. K. Dai, "Image retrieval by texture similarity," *Pattern Recognit.*, vol. 36, no. 3, pp. 665–679, 2003.
[66] T.-C. Lu and C.-C. Chang, "Color image retrieval technique based on color features and image bitmap," *Inf. Process. Manage.*, vol. 43, no. 2, pp. 461–472, 2007.
[67] T.-W. Chiang and T.-W. Tsai, "Content-based image retrieval via the multiresolution wavelet features of interest," *J. Inf. Technol. Appl.*, vol. 1, no. 3, pp. 205–214, 2006.
[68] J.-M. Guo, H. Prasetyo, and J.-H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 466–481, Mar. 2015.
[69] J.-M. Guo, H. Prasetyo, and N.-J. Wang, "Effective image retrieval system using dot-diffused block truncation coding features," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1576–1590, Sep. 2015.

MUDHAFAR JALIL JASSIM GHRABAT received the B.S. degree in computer science from Al-Mustansiriya University, Baghdad, Iraq, in 2005, and the master's degree in information technology from SRM University, Chennai, India, in 2015. He is currently pursuing the Ph.D. degree with the Center for Biomedical Imaging and Bioinformatics, Huazhong University of Science and Technology, Wuhan, China. His current research interests include machine learning, data mining, and image processing.

GUANGZHI MA received the B.S. degree in computer science from the Dalian University of Science and Technology, Dalian, China, in 1985, and the master's and Ph.D. degrees in computer science from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1988 and 2009, respectively. He is currently an Associate Professor of computer science with the Computer School, HUST, where he is also a member of the Center for Biomedical Imaging and Bioinformatics. His research interests include data mining, artificial intelligence, image processing, and net computing.

ZAID AMEEN ABDULJABBAR received the bachelor's and master's degrees in computer science from Basrah University, Iraq, in 2001 and 2006, respectively, and the Ph.D. degree in computer engineering from HUST, China, in 2016. His research interests include cloud security, searchable encryption systems, similarity measures, the Internet of Things, secure computation, biometric, and soft computing. He received the best paper award in the 11th International Conference on Green, Pervasive, and Cloud Computing. He served as a Reviewer for several prestigious journals and a PC Member of more than 25 international conferences.

MUSTAFA A. AL SIBAHEE was a Lecturer with the Department of Communication Engineering, Iraq University College, Basrah, Iraq, in 2018. He is currently a Researcher with the Shenzhen Institute, Huazhong University of Science and Technology, Shenzhen, China. His research interests include similarity measures, the Internet of Things, secure computation, biometric, and soft computing.

SAFA JALIL JASSIM received the B.C. degree in electrical engineering from Mustansiriya University, Iraq, Baghdad, in 2011, and the master's degree in electrical engineering from Anna University, Chennai, India, in 2016. Her current research interests include machine learning, big data, and control of information systems.